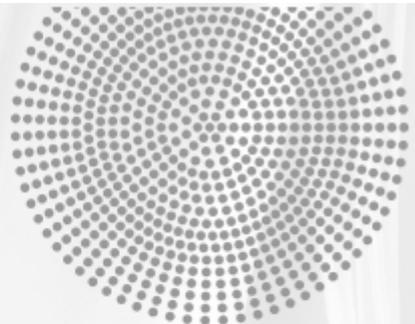




SENADA



# PROSIDING

SEMINAR NASIONAL SAINS DATA (SENADA)



▶ ▶ ▶ **IMPLEMENTASI SAINS DATA DALAM PENGUATAN  
EKONOMI DIGITAL KREATIF UNTUK MENGHADAPI  
RESESI GLOBAL** ◆



5 JUNI 2023

PROGRAM STUDI SAINS DATA,  
FAKULTAS ILMU KOMPUTER,  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR,  
SURABAYA, INDONESIA



## **DEWAN REDAKSI**

### **PELINDUNG**

**Rektor** : Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU

### **PEMBINA:**

**Dekan** : Prof. Dr. Ir. Novirina Hendrasarie,

### **PENANGGUNG JAWAB:**

**Koorprodi** : Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU

### **KETUA TIM REDAKSI:**

Aviolla Terza Damaliana, S.Si., M.Stat

### **WAKIL KETUA TIM REDAKSI:**

Trimono, S.Si., M.Si

### **SEKRETARIS**

Kartika Maulida Hindrayani, S.Kom., M.Kom  
Wahyu Kyestiati Sumarno, S.Pd., M.Ed., M.Pd

### **EDITOR**

Tresna Maulana Fahrudin. S.ST., MT  
Amri Muhaimin, S.Stat., M.Stat., MS

### **SECTION EDITOR**

Prismahardi Aji Riyantoko, S.Si., M.Si

### **REVIEWER INTERNAL**

Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU  
Dr. Ir. Mohammad Idhom., SP., S.Kom., MT  
Sugiarto, S.Kom., M.Kom  
Dr. Ir. Kartini, S.Kom., MT

### **ALAMAT:**

Seminar Nasional Sains Data (SENADA)  
Program Studi Sains Data  
Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional “Veteran” Jawa Timur  
Jalan Raya Rungkut Madya Gunung Anyar, Surabaya 60294  
Email : [senada@upnjatim.ac.id](mailto:senada@upnjatim.ac.id)

No.	Judul Artikel Dan Penulis	Hal
1.	55 - Pengaruh Adanya Podcast Terhadap Kemampuan Berbahasa Indonesia Pada Mahasiswa	Hal 01-07
	Ilmatius Sa'diyah, Bagus Rizky Prasetyo	
2.	58 - Menganalisis Pergerakan Harga Saham PT Indofood CBP Sukses Makmur Tbk - Sebuah Tinjauan Sebelum dan Sesudah Masa Pandemi	Hal 08-16
	Iqbal Ramadhan Anniswa	
3.	59 - Monitoring Suhu Pada Tanaman Seledri Menggunakan App Inventor	Hal 17-25
	Andreas Lois Figo Banggut, Aries Boedi Setiawan, Andrijani Sumarahinsih	
4.	61 - Monitoring Kelembaban Tanah dengan Menggunakan Metode Fuzzy Sugeno pada Bunga Krisan Berbasis IoT	Hal 26-33
	Yosefa Stefania Fono, Aries Boedi Setiawan, Delila cahya Permatasari	
5.	64 - Analisis Pengaruh Data Saham Provinsi Terhadap Ekonomi Indonesia	Hal 34-46
	Allan Ruhui Fatmah Sari, Adhisa Shilfadianis Iffadah, Edelin Fortuna, Nayya Ramadhani Putri Widjanarko, Yunita Nur	
6.	74 - Implementasi Metode Clustering SOM Dalam menganalisis Alasan Tidak KB di Kabupaten Temanggung	Hal 47-58
	Mufida Fauziah Faiz, Achmad Fauzan	
7.	75 - Implementasi Algoritma Naïve Bayes Classifier Dalam Klasifikasi Review Program Career Preparation Eduwork	Hal 59-67
	Adinda Alifatul Jannah, Achmad Fauzan	
8.	81 - Implementasi Multiple Linear Regression dalam Menganalisis Jarak Luncur Guguran Awan Panas	Hal 68-74
	Rahma Elis Setiawati, Rohmatul Fajriyah	
9.	82 - Penerapan Metode Fuzzy Sugeno Pada Otomatisasi Oven Pengereng Ikan Berbasis IoT	Hal 75-83
	Sebastianus Pehan Makin, Narchrowie, Subairi	
10.	83 - Monitoring Kualitas Air Aquarium Menggunakan Metode Fuzzy Mamdani Berbasis IoT	Hal 84-93
	Kornelius Setiawan Bu'u	
11.	84 - Penerapan Database Relasional pada Perpustakaan Berbasis Website dan Kontrol Akses	Hal 94-100
	Sandria Amelia Putri, Ahmad Wafi Faturrahman, Meisya Vira Amelia, Alya Setya Paramita, Farhan Dequika P.P.A., Kartika Maulida Hindrayani	
12.	85 - Analisis Prediksi Peluang Pendapatan Biaya Aset Sewa PT. KAI DAOP 6 Yogyakarta Dengan Menggunakan Metode Markov	Hal 101-105
	Nabila Azzahra Haris Putri, Abdullah Ahmad Dzikrullah	
13.	89 - Monitoring Suhu dan Kelembaban Pada Penetas Telur Otomatis Menggunakan Metode Fuzzy Sugeno Berbasis IoT	Hal 106-115
	Hari Yoal, Wahyu Dirgantara, Subairi	

No.	Judul Artikel Dan Penulis	Hal
14.	90 - Pendeteksian Spam pada E-mail menggunakan Pendekatan Natural Language Processing	Hal 116-121
	Ikbar Athallah Taufik, Dimas Dzaky Daniswara, Amri Muhaimin	
15.	91 - Analisis Topic Modelling pada Ulasan Aplikasi Shopee di PlayStore Menggunakan Latent Dirichlet Allocation (LDA)	Hal 122-133
	Sahat Renaldi. S, Primus Akbar Atnanda, Amri Muhaimin	
16.	102 - Klasifikasi Abjad SIBI (Sistem Bahasa Isyarat Indonesia) menggunakan Mediapipe dengan metode Deep Learning	Hal 134-141
	Muhammad Alfian Pratama, Muhammad Reza Erfit, Nadiya Mujahidatul Farhani, Ignatius Arvantya Hartono, Maryamah Maryamah	
17.	104 - Perspektif Wisatawan Mancanegara (Wisman) Terhadap Pariwisata Indonesia menggunakan Latent Dirichlet Allocation (LDA)	Hal 142-152
	Rafik Septiana, Muhammad Hanif, Christeigen Theodore Suhalmi, Faqih Syamil, Arya Duta Gordra Sumitro Putra, Maryamah Maryamah	
18.	105 - Speech Emotion Recognition (SER) dengan Metode Bidirectional LSTM	Hal 153-161
	Nicholas Juan Calvin Pradiptamurty, Hafiyah Khayyiroh Shafro, Mohammad Sihabudin Al Qurtubi, Giovanni Alberta Tambahjong, Qothrotunnidha' Almaulidiyah, Maryamah Maryamah	
19.	106 - Perspektif Opini Pelanggan pada Aplikasi MyPertamina menggunakan Metode Latent Dirichlet Allocation (LDA)	Hal 162-171
	Ilham Ahmad Kamil, Anisyaul Fitria, Riri Dwi Setyawati Jatiningtyas, Elfira Rahma Putri, Mutiara Afifah, Maryamah Maryamah	
20.	107 - Sistem Rekomendasi Pencarian Indekos di Surabaya Menggunakan Random Forest	Hal 172-177
	Theresa Agnes Virnauli Sinaga, Lucia Bellanie Debra, Jasmine Taj Ariva, Devi Rizky Aditya, Vivia Faustine Gunawan, Maryamah Maryamah	
21.	110 - Optimasi Pengolahan Data Facemask di Teachable Machine melalui Analisis Nilai Epoch dan Learning Rate	Hal 178-183
	Dimas Dzaky Daniswara, Ikbar Athallah Taufik, Prismahardi Aji Riyantoko	
22.	112 - Peramalan Lonjakan Kasus Harian Covid-19 Di Indonesia Dengan Model Arima	Hal 184-189
	Aviolla Terza Damaliana, Amri Muhaimin , Prismahardi Aji Riyantoko	
23.	113 - Pengaruh Nilai Tukar Rupiah terhadap Dolar Amerika Serikat dengan Jumlah Wisatawan Asia Tenggara	Hal 190-199
	Vannesa Nathania, Amirah Rizky Ramadhanti, Trimono	

No.	Judul Artikel Dan Penulis	Hal
24.	114 - Evaluasi Kualitas Pelayanan Kesehatan di Kota Surabaya - Studi Kasus pada Puskesmas dan Rumah Sakit	Hal 200-212
	Irma Octavia Chaniago, Halim Arif Cahyono, Fikrie Hartanta Sembiring, Regita Putri Permata	
25.	115 - Analisis Perubahan Tingkat Pengangguran di Kabupaten/Kota Kalimantan Barat Tahun 2010-2018	Hal 213-221
	Arindra Harris Abdillah, Muchamad Risqi, Trimono,	
26.	116 - Analisis Pengaruh COVID-19 Terhadap Jumlah Pengangguran di Jawa Timur Menggunakan Metode Uji Mean	Hal 222-229
	Larasati Romadhani Yunita Putri, Mirechelin Kristanaya, Trimono	
27.	117 - Analisis Dampak Covid-19 terhadap Nilai Ekspor Provinsi Jawa Timur dengan Uji Mean	Hal 230-234
	Affa Lelira Ibrahim, Deannisa Syafira Putri, Trimono	
28.	118 - Analisis Perbandingan Tingkat Kemiskinan Di Jawa Timur Selama Dan Sesudah Krisis Pangan	Hal 235-240
	Shafira Amanda Putri, Nabilah Selayanti, Trimono	
29.	119 - Analisis Keterkaitan Jumlah Rumah Ibadah berdasarkan Penganutnya	Hal 241-252
	Ajeng Puspa Wardani, Maulidya Prastita Syah, Trimono	
30.	120 - Statistika Deskriptif Pada Analisis Ketimpangan Kemiskinan (Studi Kasus Data Kemiskinan Wilayah Perkotaan dan Pedesaan Seluruh Provinsi di Indonesia periode tahun 2021-2022)	Hal 253-259
	Muhammad Nashif Farid, Mohammad Sufa Ammar Habibi, Trimono	
31.	121 - Analisis Kesamaan Rata-rata Indeks Harga Konsumen (IHK) Menurut Kelompok dan Sub Kelompok pada Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau dengan Kelompok dan Sub Kelompok 9 Pendidikan	Hal 260-265
	Melya Vebryanti, Wahyu Melinda Permanasari, Trimono	
32.	122 - Peramalan Penumpang Pelabuhan Terbesar Indonesia Efek Pandemi COVID-19 - Pendekatan Pemodelan Seasonal Autoregressive Integrated Moving Average dan Intervensi	Hal 266-279
	Regita Putri Permata, Rifdatun Ni'mah, Irma Octavia Chaniago	
33.	124 - Analisis Sentimen Persepsi Publik Terhadap UPN “Veteran” Jawa Timur Menggunakan Metode SVM, Naïve Bayes, dan Multinomial Logistic Regression	Hal 280-289
	Rahmatul Amanillah, Sahat Renaldi. S, Prismahardi Aji Riyantoko	
34.	125 - One-Shot Learning Menggunakan Siamese Neural Network Untuk Pendeteksian Wajah	Hal 290-295
	Muhammad Dzeaulfath, Ardy Octavian Winarta, Prizka Rismawati Arum	

No.	Judul Artikel Dan Penulis	Hal
35.	126 - Model Prediksi Kepadatan Lalu Lintas - Perbandingan Antara Algoritma Random Forest dan XGBoost	Hal 296-303
	Fitri Indah Sari, Ellexia Leonie Gunawan, Chelsea Ayu Adhigiadany , Angela Lisanthoni	
36.	127 - Adaptive Synthetic Support Vector Machine Multiclass untuk mengklasifikasikan Imbalance data pada Sentimen kenaikan Bahan Bakar Minyak	Hal 304-312
	Ismatullah, Fatkhurokhman Fauzi, Indah Manfaati Nur	
37.	128 - Analisis dan Simulasi Teori Antrian pada Indomaret Darmo Surabaya	Hal 313-319
	Thesion Marta Sianipar, Halim Arif Cahyono, Amalia Nur Alifah	
38.	129 - Regresi Data Panel untuk Memodelkan Tingkat Kemiskinan di Provinsi Papua	Hal 320-326
	Dwi Rahmadini, Ina Alfidayanti, M. Al Haris	
39.	130 - Analisis Kesamaan Rata-rata Indeks Harga Konsumen (IHK) Menurut Sub Kelompok Makanan dengan Sub Kelompok Pendidikan	Hal 327-335
	Melya Vebryanti, Wahyu Melinda Permanasari, Trimono	
40.	131 - Analisis Sentimen Brand Ambassador Artis Korea Selatan pada Produk Indonesia dengan Lexicon	Hal 336-344
	Galuh Etha Pratiwi, Tiani Wahyu Utami, Rochdi Wasono	



# Pengaruh Adanya Podcast Terhadap Kemampuan Berbahasa Indonesia Pada Mahasiswa

Ilmatus Sa'diyah<sup>1</sup>, Michael Alexander Justin Audison Sibarani<sup>2</sup>,

<sup>1,2</sup>Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[ilmatus.sisfo@upnjatim.ac.id](mailto:ilmatus.sisfo@upnjatim.ac.id)

<sup>2</sup>[20083010019@student.upnjatim.ac.id](mailto:20083010019@student.upnjatim.ac.id)

Corresponding author email: [ilmatus.sisfo@upnjatim.ac.id](mailto:ilmatus.sisfo@upnjatim.ac.id)

**Abstract:** *With the development of the era, human need for technology is also inseparable. One of them is podcasts. Podcasts are audio files available and broadcast online. The presence of podcasts is a new medium in helping the implementation of education today. However, its influence on students' Indonesian language skills is still a question. Therefore, researchers are trying to find answers through this research. The method used is a questionnaire with a Likert scale as the answer model. Validity testing is done through SPSS and response analysis is done through the results of filling in the Likert scale answers. After the research was carried out, it was found that most of the respondents agreed with the hypothesis that had been proposed by the researcher. Thus, the researcher concluded that the existence of a podcast had an effect on Indonesian language skills.*

**Keywords:** *podcast, Indonesian, language skills*

**Abstrak:** Semakin berkembangnya zaman kebutuhan manusia akan teknologi juga tidak terpisahkan. Salah satunya adalah podcast. Podcast merupakan file audio yang tersedia dan disiarkan secara online. Kehadiran podcast menjadi media baru dalam membantu penyelenggaraan pendidikan saat ini. Namun, pengaruhnya terhadap kemampuan berbahasa Indonesia mahasiswa masih menjadi pertanyaan. Oleh karena itu, peneliti berusaha mencari jawabannya melalui penelitian ini. Metode yang digunakan adalah kuisioner dengan skala likert sebagai model jawaban. Pengujian validitas dilakukan melalui SPSS dan analisa respon dilakukan melalui hasil pengisian jawaban skala likert. Setelah penelitian dilakukan, didapati bahwa sebagian besar responden menyetujui hipotesis yang telah diajukan oleh peneliti. Sehingga, peniliti membuat kesimpulan bahwa adanya podcast berpengaruh terhadap kemampuan berbahasa Indonesia.

**Kata kunci:** Podcast, Bahasa Indonesia, keterampilan berbahasa

## I. PENDAHULUAN

Teknologi telah berkembang pesat dari masa ke masa. Terhitung sampai abad ke-21, penciptaan teknologi seakan telah menjelma menjadi ajang perlombaan. Seluruh negara berlomba untuk menciptakan teknologi yang terbaik demi kemajuan umat manusia dan juga ego masing-masing untuk menjadi kiblat dunia. Teknologi juga telah menjadi kebutuhan dasar yang diperlukan oleh seluruh umat manusia. Hal ini didukung dengan adanya teknologi diseluruh aspek kehidupan manusia. Sosial, ekonomi, pendidikan, kesehatan, dan hobi pun telah memanfaatkan adanya teknologi. Beberapa contoh yang dapat kita temukan antara lain, hadirnya sosial media, berkembangnya aplikasi konsultasi kesehatan, integrasi sistem pemerintahan melalui aplikasi, hingga sistem keamanan dengan pengenalan wajah. Membahas mengenai teknologi rasanya kurang lengkap apabila tidak mengulik mengenai podcast.

Seiring berjalannya kehidupan manusia di era teknologi digital, podcast menjadi salah satu dari sekian banyak alat yang dimanfaatkan [1]. Podcast sangat mudah ditemui melalui media internet. Bahkan, terdapat sebuah aplikasi khusus yang membuatnya dapat digunakan melalui smartphone kita. Hal ini yang membuat podcast menjadi media pendidikan perguruan tinggi semakin populer. Selain itu, kondisi saat ini yang membuat pembelajaran harus dilaksanakan secara daring sejak mewabahnya virus Covid-19 menjadikan penggunaan podcast semakin relevan. Tujuan penggunaan podcast



sebagai media pendidikan dapat dikatakan beragam. Mulai dari pembawa konten utama perkuliahan, sarana pendukung media pendidikan lainnya, hingga menjadi sarana diskusi bagi dosen dan mahasiswa [2]. Namun, pengaruhnya dalam meningkatkan kemampuan berbahasa Bahasa Indonesia bagi mahasiswa masih dipertanyakan. Padahal kemampuan berbahasa yang baik merupakan suatu hal yang seharusnya dimiliki oleh mahasiswa.

Oleh karena itu, peneliti memunculkan dua rumusan masalah pada penyusunan artikel ini. Pertama, bagaimanakah cara podcast meningkatkan kemampuan berbahasa Bahasa Indonesia mahasiswa? Kedua, bagaimanakah cara podcast dapat memberikan dampak yang positif bagi perkembangan pendidikan? Rumusan kedua dimunculkan sebab peneliti beranggapan bahwa semakin berkembangnya teknologi, maka cara konvensional dalam penyelenggaraan pendidikan sudah sepantasnya mendapatkan penyegaran. Hal ini dilakukan untuk mengubah suasana kelas yang bagi sebagian mahasiswa identik dengan kata “membosankan” menjadi lebih menarik.

Peneliti bertujuan untuk menjabarkan mengenai kemungkinan podcast untuk meningkatkan kemampuan berbahasa Bahasa Indonesia mahasiswa dan mencari cara lain bagi podcast untuk memberikan dampak terhadap perkembangan pendidikan, khususnya di Indonesia. Peneliti berharap agar penelitian yang dilakukan mendapat luaran positif mengenai pengaruh adanya podcast terhadap pendidikan, khususnya kemampuan berbahasa mahasiswa.

## II. KAJIAN PUSTAKA

Dokumen Podcast adalah salah satu dari sekian banyak platform digital populer di dunia. Podcast dapat didefinisikan sebagai file audio yang tersedia dan disiarkan secara online melalui beberapa perangkat, antara lain komputer pribadi, telepon pintar, tablet, dan lain lain [3]. Podcast digunakan oleh masyarakat secara luas untuk berbagai keperluan, di antaranya berita, pendidikan, hiburan dan hobi. Dalam dunia pendidikan podcast sering digunakan para dosen untuk memperkuat materi yang akan mereka berikan ke mahasiswa [2]. Akses dapat dilakukan di segala situasi. Hal ini dapat terjadi karena podcast dapat diputar berulang kali selama kita telah mengunduh filenya atau memiliki koneksi internet untuk memutarinya.

Adanya podcast juga memberikan mahasiswa kemampuan untuk belajar dengan gaya mereka sendiri [4]. Bahkan, podcast terbukti cukup efektif digunakan pada beberap studi yang telah dilakukan seperti halnya sport manajemen [2], keperawatan [5], dan matematika [6]. Salah satu studi yang dilakukan oleh Angeulina Popova dengan judul “Bagaimana podcast bisa mendukung keterlibatan mahasiswa dalam proses pembelajaran?” menjelaskan bahwa kreatifitas mahasiswa dapat difasilitasi melalui podcast. Hal ini dilakukan melalui konten podcast yang dibuat oleh mahasiswa (student-generated content). Bahkan, kolaborasi pembuatan podcast yang dilakukan oleh mahasiswa dapat memberikan motivasi, refleksi diri dan pembelajaran langsung melalui praktik [7].

Bahasa merupakan media komunikasi manusia dalam penyampain gagasan dan informasi baik secara lisan, maupun tulisan. Dalam penyampaiannya terkandung makna yang dapat diterima berbeda oleh lawan bicara tergantung pemilihan kata, serta tinggi dan rendahnya nada yang kita gunakan. Terkadang perilaku seseorang dapat dinilai hanya dengan cara mereka menggunakan bahasa [8]. Sementara itu, kemampuan menurut KBBI merupakan kesanggupan, kecakapan, maupun kekuatan. Kemampuan memiliki kata inti mampu yang berarti sanggup untuk melakukan sesuatu [8]. Sehingga, kemampuan berbahasa dapat diartikan sebagai kesanggupan, kecakapan, maupun kekuatan dalam mengolah dan menyampaikan gagasan atau informasi baik secara lisan, maupun tulisan. Kemampuan berbahasa Indonesia dapat dikategorikan dalam empat golongan, antara lain kemampuan mendengarkan atau menyimak, berbicara, membaca dan menulis. Kemampuan menulis dapat dipadukan dengan membaca, sehingga terbentuklah kemampuan tulisan. Sementara itu kemampuan lisan merupakan perpaduan dari kemampuan mendengarkan dan berbicara [9].

**III. METODE PENELITIAN**

Penulis menggunakan metode kuantitatif dalam penelitian yang dilakukan. Metode kuantitatif dilakukan guna meneliti suatu populasi, maupun mencari sampel. Teknik pengambilan biasanya dilakukan secara acak. Hasil yang didapatkan bertujuan untuk membuktikan hipotesis yang sebelumnya telah ditentukan. Penyebaran kuisioner akan dilakukan guna memenuhi data yang dibutuhkan dengan ketentuan pengisi kuisioner merupakan mahasiswa aktif yang pernah menggunakan podcast. Kuisioner akan disebarakan melalui media sosial guna mendapatkan data yang acak dan menjaring data sebanyak mungkin dalam waktu yang singkat. Kuisioner berupa 8 pertanyaan pilihan ganda dan 2 pertanyaan bersifat deskriptif. Pertanyaan pilihan ganda berupa skala likert dengan nilai yang digunakan berkisar antara 1 sampai 5. Nantinya “1” menyatakan sangat tidak setuju, sedangkan “5” menyatakan sangat setuju. Kuisioner tersebut akan disebarakan selama 3 hari yang dimulai dari 17 Juni 2021 hingga 19 Juni 2021.

Sampling kuota digunakan sebagai metode sampling pada penelitian ini. Singkatnya, melalui sampling ini peneliti harus mendapatkan jumlah sampel hingga kuota yang ditentukan telah terpenuhi. Peneliti menentukan jumlah minimal agar kuota terpenuhi merupakan 50 orang. Terdapat beberapa alasan yang menguatkan peneliti sehingga menggunakan teknik tersebut. Pertama, peneliti tidak menemukan data yang secara rinci menyatakan mengenai jumlah pengguna aktif podcast yang berstatus sebagai mahasiswa, khususnya di kota Surabaya. Sehingga, penghitungan sampel secara pasti belum dapat dilakukan. Kedua, adanya urgensi mengenai penyelesaian penelitian ini yang membuat peneliti harus segera mendapatkan data dengan waktu terbatas. Ketiga, terdapat sumber yang mengatakan bahwa sampel penelitian dapat dikatakan layak apabila berkisar antara 30 hingga 500 [10]. Setelah sampel telah didapatkan, peneliti akan melakukan uji validitas untuk membuktikan kebenaran dari instrumen penelitian yang dimanfaatkan. Pengujian yang dilakukan menggunakan SPSS sebagai media ujinya.

Terakhir, peneliti akan melakukan analisis data skala likert untuk melihat respon yang diberikan responden terhadap kuisioner [11]. Rumus yang akan digunakan adalah:

1. Perhitungan skor masing-masing skala di tiap butir pertanyaan

$T \times P_n$	T = Keseluruhan responden yang memilih skor
	$P_n$ = Opsi skor yang dipilih pada skala likert

2. Perhitungan index tiap butir pertanyaan

$\% = \frac{\text{Total skor}}{Y} \times 100$	Y = Jumlah responden x skor tertinggi
	Total skor = $\Sigma$ Total responden yang memilih skor x pilihan skor

3. Interval untuk masing-masing skala

$$I = \frac{100}{\text{Jumlah skala likert}}$$

**IV. HASIL DAN PEMBAHASAN**

Pengisian kuisioner ditutup pada Sabtu, 19 Juni 2021 dengan jumlah responden sebanyak 60 orang. Sebanyak 66,7% responden berjenis kelamin perempuan, sedangkan 33,3% sisanya adalah laki-laki. Mayoritas dari responden berasal dari angkatan 2019 sebanyak 51,7% diikuti oleh angkatan 2020 sebanyak 43,3% dan sisanya sebanyak 5% berasal dari angkatan 2018 dan 2017. Semua data yang diperoleh kemudian diunduh untuk diolah menggunakan aplikasi SPSS. Perlu diperhatikan bahwa data yang diolah merupakan 8 pertanyaan kuisioner yang menggunakan skala likert sebagai jawabannya. Hasil dari pengolahan tersebut diperlihatkan dalam Gambar 1.

		Correlations								Total
		X1	X2	X3	X4	X5	X6	X7	X8	
X1	Pearson Correlation	1	.375**	.271*	.153	.080	.120	.172	.245	.431**
	Sig. (2-tailed)		.003	.036	.243	.545	.361	.189	.059	.001
	N	60	60	60	60	60	60	60	60	60
X2	Pearson Correlation	.375**	1	.522**	.346**	.327*	.329*	.408**	.328*	.664**
	Sig. (2-tailed)	.003		.000	.007	.011	.010	.001	.011	.000
	N	60	60	60	60	60	60	60	60	60
X3	Pearson Correlation	.271*	.522**	1	.537**	.482**	.425**	.400**	.284*	.756**
	Sig. (2-tailed)	.036	.000		.000	.000	.001	.002	.028	.000
	N	60	60	60	60	60	60	60	60	60
X4	Pearson Correlation	.153	.346**	.537**	1	.673**	.324*	.332**	.301*	.723**
	Sig. (2-tailed)	.243	.007	.000		.000	.012	.010	.019	.000
	N	60	60	60	60	60	60	60	60	60
X5	Pearson Correlation	.080	.327*	.482**	.673**	1	.545**	.459**	.331**	.754**
	Sig. (2-tailed)	.545	.011	.000	.000		.000	.000	.010	.000
	N	60	60	60	60	60	60	60	60	60
X6	Pearson Correlation	.120	.329*	.425**	.324*	.545**	1	.406**	.220	.625**
	Sig. (2-tailed)	.361	.010	.001	.012	.000		.001	.091	.000
	N	60	60	60	60	60	60	60	60	60
X7	Pearson Correlation	.172	.408**	.400**	.332**	.459**	.406**	1	.660**	.716**
	Sig. (2-tailed)	.189	.001	.002	.010	.000	.001		.000	.000
	N	60	60	60	60	60	60	60	60	60
X8	Pearson Correlation	.245	.328*	.284*	.301*	.331**	.220	.660**	1	.623**
	Sig. (2-tailed)	.059	.011	.028	.019	.010	.091	.000		.000
	N	60	60	60	60	60	60	60	60	60
Total	Pearson Correlation	.431**	.664**	.756**	.723**	.754**	.625**	.716**	.623**	1
	Sig. (2-tailed)	.001	.000	.000	.000	.000	.000	.000	.000	
	N	60	60	60	60	60	60	60	60	60

\*\* Correlation is significant at the 0.01 level (2-tailed).

\* Correlation is significant at the 0.05 level (2-tailed).

**Gambar 1.** Hasil Uji Skala Likert

Untuk dapat melakukan uji validitas, data yang kita gunakan selanjutnya adalah yang terletak pada kolom total. Data tersebut disebut sebagai Rhitung dan nantinya dibandingkan dengan Rtabel. Karena responden berjumlah 60 orang, maka nilai yang digunakan adalah 0,25 untuk taraf signifikansi 5%. Data dianggap valid apabila Rhitung > Rtabel. Hasil perbandingan ditampilkan dalam Tabel 1.

**Tabel 1.** Hasil Perbandingan

Variabel	Rhitung	Rtabel	Keterangan
<b>X1</b>	0,431	0,25	<b>Valid</b>
<b>X2</b>	0,664	0,25	<b>Valid</b>
<b>X3</b>	0,756	0,25	<b>Valid</b>
<b>X4</b>	0,723	0,25	<b>Valid</b>
<b>X5</b>	0,754	0,25	<b>Valid</b>
<b>X6</b>	0,625	0,25	<b>Valid</b>
<b>X7</b>	0,716	0,25	<b>Valid</b>
<b>X8</b>	<b>0,623</b>	<b>0,25</b>	<b>Valid</b>

Dari pengolahan tabel tersebut, peneliti dapat mengetahui bahwa instrumen yang dimanfaatkan dalam penelitian ini dinyatakan valid. Selanjutnya, peneliti akan melakukan penghitungan interval masing-masing skala berdasarkan rumus yang sebelumnya telah ditetapkan. Hasilnya dapat dilihat pada Tabel 2.

**Tabel 2.** Perhitungan Interval dan Interpretasi Jawaban

Interval	Interpretasi Jawaban
0% - 19,99%	Sangat Tidak Setuju
20% - 39,99%	Tidak Setuju
40% - 59,99%	Ragu-Ragu
60% - 79,99%	Setuju
80% - 100%	Sangat Setuju

Tahapan berikutnya adalah mengolah jawaban yang telah didapat dari kuisisioner menggunakan rumus yang telah ditentukan. Hasil pengolahan tersebut ditampilkan pada Tabel 3.

**Tabel 3.** Pengolahan Kuesioner

Pernyataan	Skala Likert	Hasil	Keterangan
------------	--------------	-------	------------



	1	2	3	4	5		
<b>Pengguna merasa podcast sangat mudah diakses</b>	0	1	6	25	28	86,67 %	<b>Sangat Setuju</b>
<b>Pengguna merasa podcast memberikan pengalaman yang menarik ketika pertama kali digunakan</b>	0	0	11	32	17	82%	<b>Sangat Setuju</b>
<b>Pengguna sering kali menggunakan podcast sebagai media edukasi</b>	0	6	20	20	14	74%	<b>Setuju</b>
<b>Pengguna merasa senang menggunakan podcast sebagai media edukasi dibandingkan media lainnya. (buku, video, dll)</b>	1	8	24	18	9	68,67%	<b>Setuju</b>
<b>Pengguna merasa penyampaian konten melalui podcast lebih mudah untuk dipahami</b>	0	4	20	23	13	75%	<b>Setuju</b>
<b>Pengguna merasa adanya podcast dapat membantu pengguna dalam meningkatkan kemampuan menyimak/mendengarkan</b>	0	1	6	29	24	85,33%	<b>Sangat Setuju</b>
<b>Pengguna merasa adanya podcast dapat membantu pengguna dalam meningkatkan kemampuan berbicara</b>	0	2	15	27	16	79%	<b>Setuju</b>
<b>Pengguna merasa adanya podcast dapat membantu dalam meningkatkan kosakata pengguna</b>	0	2	11	32	15	80%	<b>Sangat Setuju</b>

Setelah pengolahan data selesai dilakukan, didapati bahwa pertanyaan yang menjadi fokus dalam penelitian ini mendapatkan respon yang positif dari keseluruhan responden. Hal ini ditunjukkan dengan sebanyak 85,33% pengguna merasa adanya podcast membantu dalam meningkatkan kemampuan mendengar, 79% pengguna merasa bahwa podcast dapat membantu mereka meningkatkan kemampuan berbicara, dan 80% dari seluruh responden sangat setuju bahwa podcast dapat membantu mereka dalam meningkatkan kosakata. Seluruh temuan ini tentunya menjawab keraguan peneliti terhadap kemampuan podcast dalam meningkatkan kemampuan berbahasa Indonesia. Selain data yang diperoleh di atas peneliti juga merangkum beberapa jawaban mengenai 2 pertanyaan lainnya. Pertanyaan tersebut dapat dilihat dalam Lampiran 1.

Pertama, menurut mereka cara kerja podcast yang dapat dimainkan berulang kali layaknya sebuah lagu dapat melatih kemampuan mendengar. Melalui kemampuan mendengar yang cukup baik, secara tidak langsung mempengaruhi responden dalam mencerna setiap kosa kata baru yang mereka



terima. Sebagian responden juga mengatakan bahwa jenis konten yang disampaikan turut berpengaruh dalam peningkatan kemampuan berbahasa mereka. Bahkan, ada responden yang mempraktikkan pembicaraan dalam podcast tersebut menggunakan gaya bicaranya sendiri untuk melatih kemampuan berbicara dan memperkaya kosakatanya. Namun, beberapa responden menekankan bahwa efektivitasnya bertambah apabila didengarkan beberapa kali.

Kedua, sebagian besar responden beranggapan bahwa banyaknya konten yang dapat dieksplor melalui podcast dapat memberikan wawasan yang luas bagi pendengarnya dan menjadi keuntungan bagi perkembangan pendidikan. Selain itu, podcast yang berbentuk audio membuatnya sangat fleksibel dan lebih hemat dalam hal penggunaan kuota. Apalagi aplikasi penyedia podcast sudah bisa diakses dengan mudah di Indonesia. Beberapa responden mengatakan bahwa dampak positif podcast dapat dirasakan apabila tujuan pembuatannya terarah dengan baik dan pembicaranya merupakan orang yang berkualitas. Namun, ada responden yang berpendapat bahwa podcast mungkin lebih sulit untuk diterima karena pola pendidikan kita yang membutuhkan asupan visual daripada audio.

## V. KESIMPULAN

Melalui penelitian ini, peneliti mengambil kesimpulan bahwa podcast turut berpengaruh dalam peningkatan kemampuan berbahasa Indonesia. Konten yang terdapat didalamnya dan cara kita dalam mendengarkan suatu podcast yang akhirnya dapat memperluas wawasan kita. Podcast juga dapat menjadi suatu media pendidikan yang menarik apabila direncanakan dan dieksekusi dengan matang. Dalam hal ini sangat penting apabila sang pembicara dalam podcast benar-benar mengerti mengenai apa yang dia sampaikan. Maka dari itu, apabila institusi pendidikan sudah mulai menggunakan podcast sebagai media mereka ada baiknya untuk mempertimbangkan hal-hal diatas untuk memberikan pengalaman pertama yang baik bagi pelajarnya.

Peneliti juga menyampaikan bahwa data yang disampaikan dalam penilitan ini mungkin memiliki banyak kekurangan karena sebagian besar respondennya merupakan mahasiswa Surabaya. Untuk itu penelitian lebih lanjut mengenai kawasan lain, seperti kota ataupun provinsi akan sangat dianjurkan. Tentunya dengan cakupan responden yang lebih luas supaya hasil yang diperoleh sangat representatif. Selain itu, peneliti juga menyarankan kepada para pembaca untuk mencoba menggunakan podcast sebagai sarana pendidikan maupun pengembangan diri. Hal ini didasari oleh hasil yang peniliti dapatkan melalui jawaban para responden mengenai pengaruh adanya podcast terhadap kemampuan berbahasa Indonesia mereka.

## UCAPAN TERIMA KASIH

## REFERENSI

1. Merhi, M. I. (2015). Factors influencing higher education students to adopt podcast: Anempirical study. *Computers and Education*, 83, 32–43. <https://doi.org/10.1016/j.compedu.2014.12.014>
2. Rockhill, C., Pastore, D., & Johnston, D. (2019). The effectiveness of podcasts in sport management education. *Journal of Hospitality, Leisure, Sport and Tourism Education*, 25, 100211. <https://doi.org/10.1016/j.jhlste.2019.100211>
3. Stiffler, D., Stoten, S., & Cullen, D. (2011). Podcasting as an instructional supplement to online learning: A pilot study. *CIN - Computers Informatics Nursing*, 29(3), 144–148. <https://doi.org/10.1097/NCN.0b013e3181fc3fdf>
4. Lonn, S., & Teasley, S. D. (2009). Podcasting in higher education: What are the implications for teaching and learning? *Internet and Higher Education*, 12(2), 88–92. <https://doi.org/10.1016/j.iheduc.2009.06.002>
5. Hargett, J. L. (2018). Podcasting in Nursing Education: Using Commercially Prepared Podcasts to Spark Learning. *Teaching and Learning in Nursing*, 13(1), 55–57. <https://doi.org/10.1016/j.teln.2017.08.003>



6. Susilowati, R.D., Utama, Nuqhty F. (2020). Application of Podcasts on Spotify as a Mathematics Learning Media in the Middle of Covid-19 Pandemic. *Jurnal RIset Pendidikan dan Inovasi Pembelajaran Matematika*. 4(1): 68—78
7. Popova, A., & Edirisingha, P. (2010). How can podcasts support engaging students in learning activities? *Procedia - Social and Behavioral Sciences*, 2(2), 5034–5038. <https://doi.org/10.1016/j.sbspro.2010.03.816>
8. Setyawan, F. H. (2016). Meningkatkan Kemampuan Berbahasa Anak Usia Dini Melalui Model Pembelajaran Audio Visual Berbasis Android. *Jurnal PG-PAUD Trunojoyo*, 3(2),92–98.
9. Robbi, B., & Salim, A. (n.d.). *Kemampuan Berbahasa Indoensia Mahasiswa Asing*. 1(3), 49–56.
10. Sugiyono. (2015). *METODE PENELITIAN PENDIDIKAN Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: CV. ALVABETA.
11. Editor, T. (n.d.). *Pengertian Skala Likert dan Contoh Cara Hitung Kuesionernya*. Diedit.Com. Retrieved June 19, 2021, from <https://www.diedit.com/skala-likert/>



# Menganalisis Pergerakan Harga Saham PT Indofood CBP Sukses Makmur Tbk: Sebuah Tinjauan Sebelum dan Sesudah Masa Pandemi

Iqbal Ramadhan Anniswa<sup>1</sup>,

<sup>1,2</sup>Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[21083010111@upnjatim.ac.id](mailto:21083010111@upnjatim.ac.id)

Corresponding author email: [21083010111@upnjatim.ac.id](mailto:21083010111@upnjatim.ac.id)

**Abstract:** This study aims to determine the share price of Indofood CBP Sukses Makmur TBK (ICBP) using three different methods, namely Long Short-Term Memory (LSTM), Logistic Regression and Support Vector Machine (SVM). Daily stock price data from October 2010 to January 2023 are used for input in this study. The LSTM method is used to predict ICBP's stock price by taking information from several previous periods, while Logistic Regression and SVM are used to predict stock price trends.

**Keywords:** stock, investors, LSTM (Long Short Term Memory), SVM (Support Vector Machine), Logistic Regression

**Abstrak:** Penelitian ini bertujuan untuk mengetahui harga saham Indofood CBP Sukses Makmur TBK (ICBP) dengan menggunakan tiga metode berbeda yaitu Long Short-Term Memory (LSTM), Regresi Logistik dan Support Vector Machine (SVM). Data Harian harga saham dari periode oktober 2010 sampai dengan januari 2023 digunakan untuk input dalam penelitian ini. metode LSTM ini digunakan untuk memprediksi harga saham ICBP dengan mengambil informasi dari beberapa periode sebelumnya, sedangkan Regresi logistik dan SVM Digunakan untuk prediksi tren harga saham.

**Kata kunci:** saham, investor, LSTM (Long Short Term Memory), SVM (Support Vector Machine), Regresi Logistik

## I. PENDAHULUAN

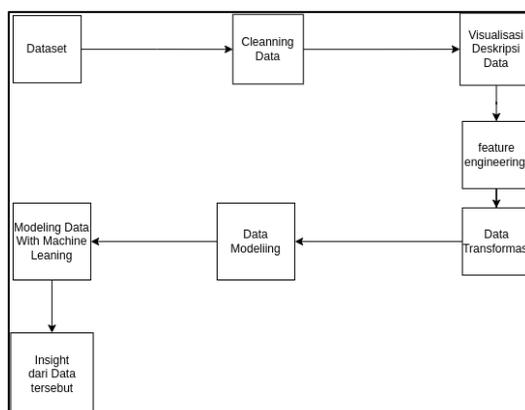
Pandemi COVID-19 telah menimbulkan dampak yang signifikan pada sektor ekonomi di seluruh dunia, termasuk pasar saham. Harga saham perusahaan-perusahaan terkemuka terdampak oleh fluktuasi pasar yang tidak menentu dan kondisi ketidakpastian yang terjadi akibat pandemi ini. Dalam situasi seperti ini, investor dan trader perlu membuat keputusan investasi yang tepat dengan mempertimbangkan risiko dan potensi keuntungan yang ada [3].

Oleh karena itu, penelitian yang bertujuan untuk menganalisis harga saham Indofood CBP Sukses Makmur Tbk (ICBP) pada periode 2010 hingga 2023 menjadi penting. Penelitian ini memfokuskan pada masa pandemi COVID-19 dan juga pra-COVID, yang membantu memperoleh gambaran yang lebih lengkap mengenai tren harga saham ICBP selama periode tersebut. Penelitian ini menggunakan tiga metode berbeda, yaitu Long Short-Term Memory (LSTM) [1], Regresi Logistik, dan Support Vector Machine (SVM) untuk memprediksi tren harga saham ICBP selama periode tersebut.

Dalam pendahuluan ini, akan diuraikan tentang latar belakang pentingnya penelitian ini di tengah kondisi yang penuh ketidakpastian baik dalam kondisi pandemi COVID-19 maupun di masa New Normal. Penelitian ini dapat memberikan pandangan yang lebih jelas bagi investor dan trader dalam mengambil keputusan investasi pada saham ICBP. Hal ini penting karena keputusan investasi yang tepat sangat diperlukan dalam menghadapi situasi penuh ketidakpastian seperti pandemi COVID-19 dan memperoleh keuntungan yang maksimal.

## II. METODE PENELITIAN

Penelitian ini mengadopsi tiga metode analisis yang berbeda, yaitu Long Short-Term Memory (LSTM), Regresi Logistik, dan Support Vector Machine (SVM), untuk memprediksi pergerakan harga saham ICBP selama periode 2010 hingga 2023. Metode LSTM digunakan untuk memodelkan hubungan antara harga saham ICBP dengan faktor-faktor eksternal seperti kondisi pasar, nilai tukar mata uang, dan faktor-faktor lainnya yang dapat mempengaruhi harga saham. Metode Regresi Logistik berfokus pada analisis korelasi antara faktor-faktor internal dan eksternal perusahaan dengan harga saham, sehingga dapat memberikan pandangan yang lebih jelas mengenai faktor-faktor yang mempengaruhi tren harga saham ICBP. Metode SVM, pada sisi lain, digunakan untuk mengoptimalkan prediksi harga saham ICBP dengan mempertimbangkan faktor-faktor risiko dan volatilitas pasar.



**Gambar 1.** Metode penelitian

### A. Dataset

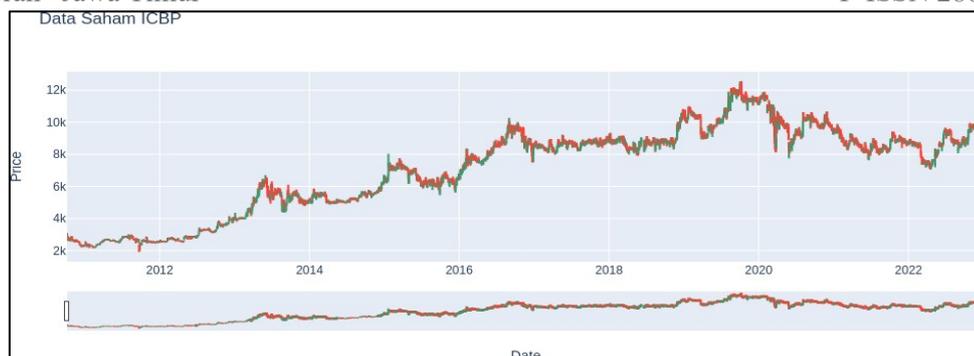
Dataset yang digunakan dalam penelitian ini adalah data historis harga saham Indofood CBP Sukses Makmur Tbk (ICBP) dari periode 7 Oktober 2010 hingga 7 Januari 2023. Data tersebut diambil dari platform Kaggle dan terdiri dari 3197 observasi dengan 6 variabel, yaitu harga pembukaan, harga penutupan, harga tertinggi, harga terendah, volume perdagangan, dan harga penyesuaian (adjusted close price). Data ini akan digunakan sebagai input dalam proses analisis menggunakan metode LSTM, Regresi Logistik, dan SVM untuk memprediksi tren harga saham ICBP pada masa pandemi COVID-19 dan masa new normal.

### B. Cleaning Data

Sebelum dilakukan analisis, data historis harga saham ICBP perlu melalui tahap data cleaning terlebih dahulu. Tahap ini dilakukan untuk memastikan data yang digunakan dalam penelitian berkualitas dan akurat. Proses data cleaning meliputi penghapusan data kosong (missing values) dan data outlier yang mungkin mempengaruhi hasil analisis, serta penyesuaian skala data agar dapat digunakan dalam proses analisis menggunakan metode LSTM, Regresi Logistik, dan SVM. Data cleaning dilakukan menggunakan software Python dan library Pandas untuk mempermudah proses pengolahan data. Setelah melalui tahap ini, data yang dihasilkan siap untuk digunakan dalam analisis dan prediksi tren harga saham ICBP di masa yang akan datang.

### C. Visualisasi Deskripsi data

Visualisasi data deskriptif dilakukan untuk memberikan gambaran yang lebih jelas mengenai data historis harga saham ICBP dari periode 2010 hingga 2023. Data yang sudah melalui tahap data cleaning dapat diolah menggunakan software Python dan library Matplotlib untuk menghasilkan visualisasi yang informatif dan mudah dipahami. Visualisasi data deskriptif yang dihasilkan meliputi grafik tren harga saham ICBP, volume perdagangan saham ICBP, serta perbandingan dengan indeks harga saham lainnya.

**Gambar 2.** Visualisasi harga dari semua variabel

pada gambar 2 tersebut menunjukkan kenaikan harga saham dari tahun 2010 hingga 2023 ini mengalami naik dan turun. kenaikan lebih tajam terjadi pada tahun 2019. kenaikan harga saham pada tahun 2019 bisa dilihat dari kenaikan rata-rata setiap tahun pada semua variabel. Kenaikan harga saham ICBP dapat disebabkan oleh beberapa faktor seperti permintaan yang tinggi untuk bahan pokok masyarakat saat pandemi. Untuk melihat apakah harga saham berasal dari kuartil akhir, dapat dilihat dengan memperhatikan perbandingan harga saham sebelum masa kuartil terakhir dan pada kuartil akhir pada keempat variabel yang diamati. Terlihat bahwa semua variabel mengalami kenaikan harga pada kuartir akhir.

#### D. Featuring Engineer

Pada penelitian ini, dilakukan feature engineering untuk meningkatkan akurasi prediksi model. Feature engineering dilakukan dengan menambahkan beberapa fitur baru berdasarkan karakteristik data historis harga saham ICBP dan faktor-faktor eksternal yang mempengaruhinya, seperti fluktuasi pasar dan kebijakan pemerintah bisa dilihat pada gambar (3) dan (4).

	timestamp	open	low	high	close	volume
0	2010-10-07	2975	2850	3100	2975	570465000
1	2010-10-08	2850	2825	3000	2850	72267000
2	2010-10-11	2725	2600	2850	2725	102116000
3	2010-10-12	2775	2650	2875	2775	50820000
4	2010-10-13	2800	2750	2825	2800	20092000
...	...	...	...	...	...	...
3192	2023-01-02	9975	9975	10200	10150	1201400
3193	2023-01-03	10100	10100	10300	10175	3169800
3194	2023-01-04	10225	10025	10250	10100	3973100
3195	2023-01-05	10100	9875	10225	10225	3774400
3196	2023-01-06	10225	10100	10300	10175	6288100

**Gambar 3.** Raw performant csv

	timestamp	open	low	high	close	volume	day	month	year	is_quarter_end	open-close	low-high	target
0	2010-10-07	2975	2850	3100	2975	570465000	7	10	2010	0	0	-250	0
1	2010-10-08	2850	2825	3000	2850	72267000	8	10	2010	0	0	-175	0
2	2010-10-11	2725	2600	2850	2725	102116000	11	10	2010	0	0	-250	1
3	2010-10-12	2775	2650	2875	2775	50820000	12	10	2010	0	0	-225	1
4	2010-10-13	2800	2750	2825	2800	20092000	13	10	2010	0	0	-75	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
3192	2023-01-02	9975	9975	10200	10150	1201400	2	1	2023	0	-175	-225	1
3193	2023-01-03	10100	10100	10300	10175	3169800	3	1	2023	0	-75	-200	0
3194	2023-01-04	10225	10025	10250	10100	3973100	4	1	2023	0	125	-225	1
3195	2023-01-05	10100	9875	10225	10225	3774400	5	1	2023	0	-125	-350	0
3196	2023-01-06	10225	10100	10300	10175	6288100	6	1	2023	0	50	-200	0

**Gambar 4.** Data telah dimanipulasi

Melalui feature engineering, beberapa fitur baru telah ditambahkan pada data historis harga saham ICBP dan faktor-faktor eksternal untuk meningkatkan akurasi prediksi model.

#### E. Data Transformasi

Berikutnya, kami melakukan evaluasi performa model yang telah dimodelkan dengan menggunakan metode `roc_auc_score` dan menghitung tingkat keakuratan model dengan menggunakan library Scikit-Learn pada Bahasa pemrograman Python. Untuk memperoleh hasil modeling yang lebih optimal, feature engineering dilakukan dengan menggunakan metode Z-score(1) untuk mengubah fitur-fitur yang digunakan dalam model regresi logistik dan SVM. Selain itu, fitur tersebut juga diubah menjadi bentuk yang dapat dimodelkan menggunakan LSTM dengan menggunakan metode Min Max(2) pada library Scikit-Learn. Dengan demikian, diharapkan dapat meningkatkan akurasi prediksi model dan mendapatkan hasil yang lebih baik.

$$Z \text{ score} = (x - \text{mean}) / \text{std} \quad (1)$$

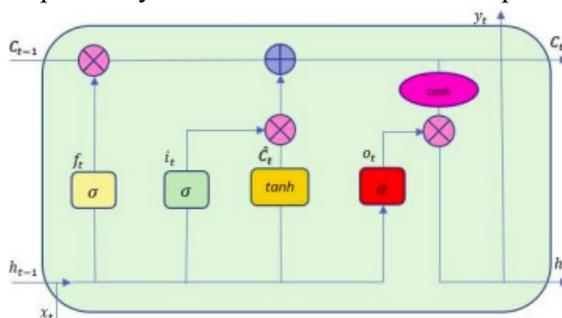
$$\text{Minmax} = (x - \text{min}) / (\text{max} - \text{min}) \quad (2)$$

#### F. Data Modeling

Untuk memodelkan Data setelah melakukan Data Transformasi, harus diperhatikan apakah data transformasi ini sesuai dengan data yang kita modelkan nanti. Berikut metode analisis yang digunakan dalam data modeling pada Penelitian ini

1. Metode LSTM:

metode ini digunakan untuk memprediksi harga saham ICBP dengan memodelkan hubungan antara harga saham dengan faktor-faktor eksternal seperti fluktuasi pasar, sentimen investor dan kebijakan pemerintah [1]. LSTM dapat menangkap pola jangka panjang dalam waktu yang lebih lama. LSTM menggunakan blok memory-cell yang terdiri dari input gate, forget gate dan output gate untuk mengganti lapisan RNN agar dapat mengatasi masalah vanishing gradient pada RNN. Pada LSTM informasi dapat disimpan dalam waktu panjang karena dapat mempelajari long-term dependency. arsitektur LSTM bisa dilihat pada gambar 5.



Gambar 5. Arsitektur LSTM

pada gerbang pertama pada LSTM disebut dengan forget gate, dimana terdapat proses memilah informasi yang ada pada *cell state* menggunakan persamaan 1 (1). informasi akan dibuang dari *cell state* jika *forget gate* bernilai 0, sebaliknya informasi akan disimpan apabila *cell state* nya pada *forget gate* bernilai 1.

$$f_t = \sigma((W_f \cdot x_t)(U_f \cdot h_{t-1}) + b_f) \tag{1}$$

penambahan informasi yang berguna ke *cell state* dilakukan pada gerbang kedua yaitu *input gate*. informasi tersebut diatur menggunakan fungsi sigmoid dan menyaring nilai yang disimpan, prosesnya mirip dengan forget gate yang menggunakan persamaan 2 dan tanh perhitungannya dengan menggunakan persamaan 3. *Cell state* dihasilkan dari nilai output kedua lapisan yang telah digabungkan.

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{2}$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{3}$$

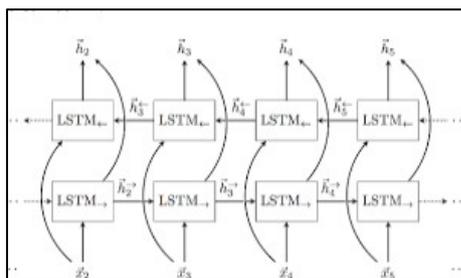
kemudian dilakukan proses perbaruan nilai *cell state* dengan menggunakan persamaan 4. Selanjutnya akan melewati output gate, yang dimana didalamnya nilai Output *cell state* akan ditentukan pada lapisan sigmoid digunakan untuk memiliki output berdasarkan *cell state* yang ada dengan menggunakan persamaan 5, dan selanjutnya akan diteruskan ke lapisan tanh menggunakan persamaan 6.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \tag{4}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \tag{5}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{6}$$

LSTM memiliki kelemahan hanya memproses kata satu arah saja, yaitu dari awal sampai akhir. dalam penelitian ini mempertimbangkan untuk menggunakan bidirectional LSTM agar dapat membaca kata dari 2 arah, yaitu awal sampai akhir dan akhir sampai awal. sesuai arsitektur bidirectional LSTM pada gambar 6, bidirectional LSTM memiliki 2 lapisan LSTM yang terpisah satu untuk maju dan satu untuk mundur sesuai dengan persamaan 7.



**Gambar 6.** Arsitektur Bidirectional LSTM

$$\mathbf{h}_t^{BiLSTM} = \mathbf{h}_t^{forward} \oplus \mathbf{h}_t^{backward} \quad (7)$$

## 2. Metode Regresi Logistik

Regresi Logistik merupakan metode statistik yang digunakan untuk memprediksi probabilitas terjadinya sebuah kejadian pada suatu variabel dependen berdasarkan variabel independen[5]. Dalam penelitian ini, Regresi Logistik digunakan untuk memprediksi pergerakan tren harga saham ICBP dengan mempertimbangkan faktor-faktor eksternal seperti fluktuasi pasar dan kebijakan pemerintah sebagai variabel independen. Hasil prediksi Regresi Logistik dapat memberikan pandangan yang lebih jelas bagi investor dan trader dalam mengambil keputusan investasi pada saham ICBP.

Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai 0 dan 1, sehingga mengikuti distribusi Bernoulli sebagai berikut:

$$f(y^i) = \pi_i^{y^i} (1 - \pi_i)^{1-y^i}$$

Dimana:

$\pi_i$  = peluang kejadian ke-i

$y^i$  = peubah acak ke-i yang terdiri dari 0 dan 1

Bentuk model regresi logistic dengan satu variable predictor adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

Untuk mempermudah menaksir parameter regresi, maka  $\pi(x)$  pada persamaan diatas ditransformasikan sehingga menghasilkan bentuk logit regresi logistik, sebagai berikut:

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

## 3. Support Vector Machine (SVM)

digunakan untuk mengoptimalkan prediksi harga saham ICBP dengan memperhatikan faktor-faktor risiko dan volatilitas pasar. SVM adalah metode machine learning yang memanfaatkan teknik kernel untuk memisahkan data ke dalam kelas-kelas yang berbeda. SVM dapat menangani data yang kompleks dan berguna untuk memprediksi pergerakan harga saham dalam jangka waktu yang lebih pendek.

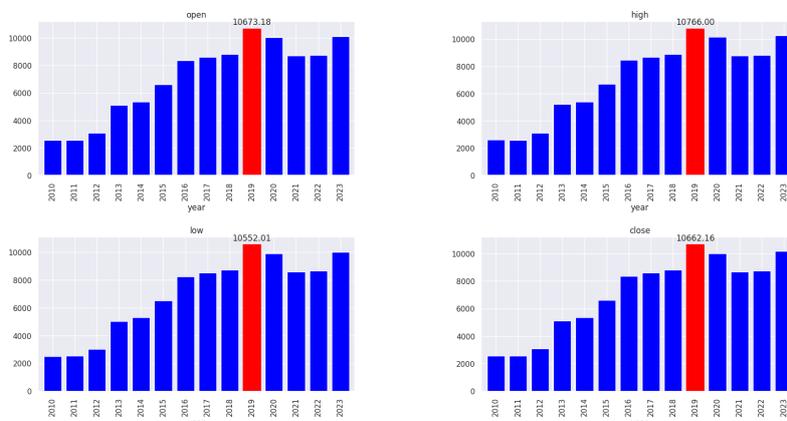
## 4. XGBoost

salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi pada data dengan tingkat akurasi yang tinggi. XGBoost Classifier memiliki kemampuan untuk melakukan klasifikasi pada data yang memiliki fitur berjumlah banyak dan heterogen, serta mampu menangani data yang tidak seimbang (imbalanced data) dengan baik. Algoritma ini juga mampu mengatasi overfitting pada model dengan cara melakukan regularisasi pada setiap iterasi boosting.

## III. HASIL DAN PEMBAHASAN

hasil dan evaluasi dilakukan dengan menggunakan metode dengan alur penelitian seperti pada

gambar 1 menghasilkan yang bagus untuk perusahaan tersebut dengan menghasilkan harga saham yang naik, hal tersebut bisa dilihat dengan pada gambar (3) dengan visualisasi gambar tersebut bahwa kita bisa melihat rata-rata dan pergerakan harga saham setiap fitur dalam satu tahun dengan periode 2010 hingga 2023.



**Gambar 7.** Visualisasi pendapatan dalam per tahun dari 2010-2023

pada gambar tersebut bahwa kita bisa mengetahui bahwa harga saham tertinggi jatuh pada tahun 2019 dan naik lagi dengan posisi kedua pada tahun 2023. dengan begitu bahwa perusahaan ICBP ini sedang mendapatkan kesuksesan tinggi dalam menghadapi tahun 2023 ini. selain dalam saham kuartil tersebut ialah juga membuktikan bahwa juga mengalami kenaikan yang itu mendukung pada visual dengan cara memanipulasi data tersebut tanggal saham dicatat untuk mengetahui apakah tanggal tersebut termasuk dalam kuartir akhir atau tidak. setelah itu baru kita hitung rata-rata tersebut dengan mengelompokkan berdasarkan nilai kuartir tersebut apakah bernilai “1” untuk ya atau bernilai “0” untuk tidak

	open	low	high	close	volume
is_quater_end					
0	7133.17426	7046.315171	7212.067168	7133.496947	6.845577e+06
1	7152.63764	7056.518727	7235.354869	7146.724719	6.252429e+06

**Gambar 8.** Tabel rata-rata pada tiap fitur dengan mengelompokkan kuartir akhir

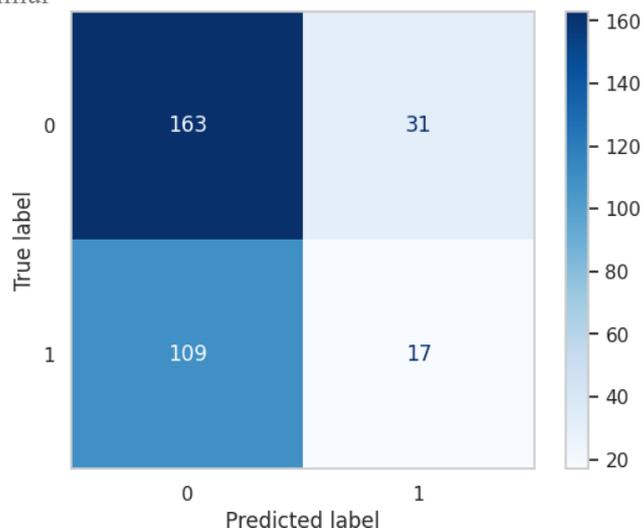
setelah memanipulasi feature harga saham dengan membagikan dengan posisi apakah dalam kuartir terakhir ataupun tidak. setelah itu kita melakukan modeling data dengan algoritma Support Vector Machine dengan akurasi 51% untuk datatest kita dan 72 % untuk model data latih kita yang bisa dilihat pada gambar.

```

Training Accuracy : 0.7291183172550132
Valid Accuracy : 0.5102683685157912
    
```

**Gambar 9.** Hasil train dan test untuk pemodelan Linear Logistik dan SVM

dengan menggunakan model machine learning Support Vector Machine bisa kita mendapatkan output seperti pada gambar.



**Gambar 10.** Hasil output untuk model Machine Learning yang lebih cocok

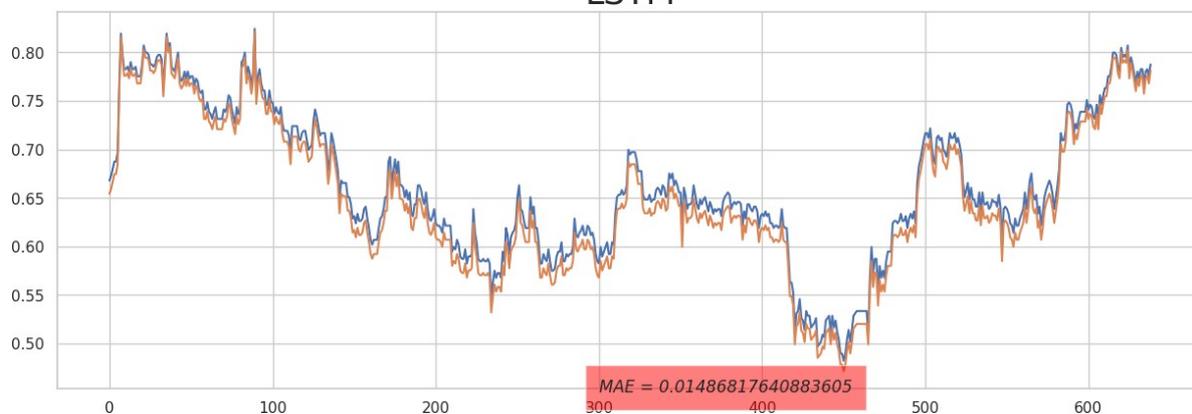
Hasil pada Gambar 10 menunjukkan hasil dari penggunaan tiga atribut yaitu "open-close", "low-high", dan "is\_quarter\_end" dalam proses visualisasi menggunakan Confusion Matrix Display. Atribut "open-close" digunakan untuk mengukur perbedaan antara harga saham pada saat pembukaan dan penutupan, sementara atribut "low-high" digunakan untuk mengukur perbedaan antara harga saham terendah dan tertinggi. Selain itu, atribut "is\_quarter\_end" juga dimasukkan sebagai salah satu fitur karena memiliki potensi untuk memberikan sinyal penting dalam memprediksi pergerakan harga saham.

Dalam proses visualisasi, digunakan Confusion Matrix Display untuk memberikan representasi visual tentang seberapa baik model dapat memprediksi pergerakan harga saham. Hal ini dilakukan dengan membagi hasil prediksi menjadi empat kategori yaitu true positive, false positive, true negative, dan false negative. Dengan menggunakan Confusion Matrix Display, para investor dan perusahaan dapat memperoleh pemahaman yang lebih baik tentang bagaimana model berkinerja dalam memprediksi pergerakan harga saham.

Dengan memasukkan tiga atribut tersebut dan menggunakan Confusion Matrix Display sebagai alat visualisasi, diharapkan dapat memberikan informasi yang lebih akurat dan komprehensif dalam memprediksi pergerakan harga saham ICBP.

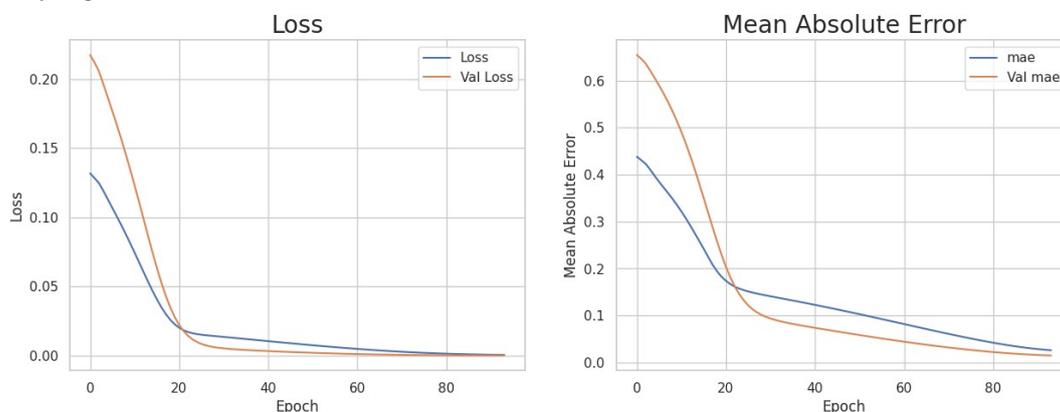
Setelah mengidentifikasi faktor-faktor krusial yang mempengaruhi pergerakan harga, mari kita lanjutkan dengan menjelajahi hasil dari jalur line plot yang dapat memberikan wawasan yang menakjubkan tentang tren harga dan prediksi masa depannya. Dengan menggunakan algoritma LSTM tersebut kita bisa memodelkan data yang ada untuk memprediksi harga saham yang terjadi. Hasil tersebut bisa kita lihat pada gambar 11. Dari hasil prediksi menggunakan LSTM, terlihat bahwa harga saham PT Indofood mengalami kenaikan. Hal ini dapat menjadi pertimbangan bagi para petinggi perusahaan untuk mengambil keputusan yang tepat, serta bagi para investor untuk melakukan strategi investasi yang sesuai.

## Prediksi Harga Saham PT Indofood CBP Sukses Makmur Tbk LSTM



**Gambar 11.** Line Chart laju harga saham pada PT Indofood CBP Sukses Makmur TBK

Hasil visualisasi menggunakan metode LSTM menunjukkan bahwa model tersebut mampu memberikan hasil prediksi yang akurat, dengan nilai MAE yang hanya sebesar 0.012%. Hal ini menunjukkan bahwa model tersebut mampu memprediksi pergerakan harga saham dengan sangat baik. Selain itu, untuk melihat performa model secara detail, dilakukan visualisasi hasil loss function dan MAE pada setiap epoch. Dari hasil visualisasi tersebut, terlihat bahwa loss function model cenderung menurun secara signifikan setiap epoch, sementara nilai MAE semakin mendekati nol. Hal ini menunjukkan bahwa model telah berhasil dilatih dengan baik dan mampu memberikan hasil prediksi yang konsisten dan akurat.



**Gambar 12.** Visualisasi nilai Loss Function dan juga Mean Absolute Error

Berikut hasil visualisasi pada gambar 12 menggunakan metode LSTM yang menunjukkan model tersebut mampu memberikan hasil prediksi yang akurat, dengan nilai MAE yang hanya sebesar 0.012%. Hal ini menunjukkan bahwa model tersebut mampu memprediksi pergerakan harga saham dengan sangat baik.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penggunaan model LSTM pada data historis harga saham ICBP dari periode 2010 hingga 2023 dengan menggunakan library TensorFlow dapat memberikan hasil prediksi yang akurat. Transformasi data dengan menggunakan metode logaritmik dan normalisasi data dengan metode MinMaxScaler pada library Sklearn juga dapat meningkatkan akurasi model. Evaluasi kinerja model menggunakan metode ROC-AUC dan MAE menunjukkan hasil yang baik dengan nilai MAE sebesar 0.012%. Hal ini menjadi bahan pertimbangan bagi para petinggi perusahaan dalam mengambil keputusan terkait pergerakan harga saham di masa depan.

**UCAPAN TERIMA KASIH**

Kami mengucapkan terima kasih yang sebesar-besarnya atas presentasi yang telah disampaikan mengenai penggunaan model LSTM pada data historis harga saham ICBP. Presentasi tersebut sangat informatif dan memberikan wawasan yang lebih mendalam tentang penggunaan teknologi dalam menganalisis data keuangan.

Kami sangat mengapresiasi upaya yang telah dilakukan untuk melakukan penelitian ini dan memberikan hasil yang sangat bermanfaat bagi para investor dan perusahaan. Kami berharap penelitian ini dapat menjadi landasan yang kuat bagi para investor dalam mengambil keputusan investasi di masa depan.

Sekali lagi, terima kasih banyak untuk presentasi yang telah disampaikan dan semoga hasil penelitian ini dapat memberikan manfaat yang lebih besar bagi dunia keuangan.

**REFERENSI**

1. Dloifur Rohman Alghifari; Mohammad Edi and Lutfi Firmansyah, Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia Bidirectional LSTM Implementation for Sentiment Analysis Against Grab Indonesia Services. Jurnal Manajemen Informatika (JAMIKA), Volume 12 Nomor 2 Edisi Oktober 2022, pp.11..
2. Putri Anis Qori, Dhama Sekar Oktafani, Iqbal Kharisudin, Analisis Peramalan dengan Long Short Term Memory pada Data Kasus Covid-19 di Provinsi Jawa Tengah, Jurusan Matematika Universitas Negeri Semarang: Semarang, Indonesia, 2022, pp. 1–7.
3. Eky Kriswahyuni, Syahril Effendi, *FAKTOR-FAKTOR MEMPENGARUHI HARGA SAHAM TERHADAP PERUSAHAAN TRANSPORTASI DAN LOGISTIK BURSA EFEK INDONESIA*; Universitas Putera Batam: Batam, Indonesia, Volume 5 No.2, 2021; pp. 1–11.
4. Marco Peixeiro, Time Series Forecasting Python; Manning, Shelter New York, United States of America, 2022, Volume 1, pp.31-86).
5. Yumira Adriani Tampil, Hanny Komalig, Yohanis Langi, Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado, Universitas Sam Ratulangi, Manado, Indonesia, [varlangi@gmail.com](mailto:varlangi@gmail.com), 2017.



# Monitoring Suhu Pada Tanaman Seledri Menggunakan *App Inventor*

Andreas Lois Figo Banggut<sup>1</sup>, Aries Boedi Setiawan<sup>2</sup>, Andrijani Sumarahinsih<sup>3</sup>

<sup>1,3</sup> *Andreas L.F. Banggut (Jurusan Teknik Elektro, Universitas Merdeka Malang)*

<sup>1</sup> [andrelfb24@gmail.com](mailto:andrelfb24@gmail.com)

<sup>3</sup> [andrijani.sumarahinsih@unmer.ac.id](mailto:andrijani.sumarahinsih@unmer.ac.id)

<sup>2</sup> *Aries Boedi Setiawan (Jurusan Teknik Elektro, Universitas Merdeka Malang)*

Corresponding author email: [aries@unmer.ac.id](mailto:aries@unmer.ac.id)

**Abstract:** *In its growth period, plants need many supporting factors such as appropriate air temperature. The frequent incidence of crop failure is caused by weather that causes the temperature to not match the ideal temperature needed by plants. Celery plant is one of them because it is vulnerable to weather. So a temperature monitoring system for celery plants was created which was monitored using App Inventor on a smartphone. Starting from reading the DHT11 sensor which is then processed by ESP32 which then sends the data to the App inventor on the smartphone, then ESP32 gives commands to the output which then works based on the conditions that occur from the results of sensor readings with PWM fuzzy sugeno. The parameter being monitored is the temperature of the celery plant. The reading results will be displayed on App Inventor as an Internet of Things (IoT) support system.*

**Keywords:** *celery plant, DHT11 sensor, fuzzy sugeno method, App Inventor.*

**Abstrak:** Dalam masa pertumbuhannya tanaman membutuhkan banyak faktor pendukung seperti suhu udara sesuai. Seringnya kejadian gagal panen yang diakibatkan oleh cuaca yang menyebabkan suhu tidak sesuai dengan suhu ideal yang dibutuhkan tanaman. Tanaman seledri adalah salah satunya karena rentan terhadap cuaca. Sehingga dibuatlah sistem monitoring suhu pada tanaman seledri yang dimonitoring menggunakan *App Inventor* pada *smartphone*. Dimulai dari pembacaan sensor DHT11 yang kemudian diproses oleh ESP32 yang kemudian mengirimkan data tersebut ke *App inventor* pada *smartphone*, kemudian ESP32 memberi perintah kepada output yang kemudian bekerja berdasarkan kondisi yang terjadi dari hasil pembacaan sensor dengan PWM fuzzy sugeno. Parameter yang dimonitoring adalah suhu pada tanaman seledri. Hasil pembacaan tersebut akan di tampilkan pada *App Inventor* sebagai sistem pendukung *Internet of Things (IoT)*.

**Kata kunci:** Tanaman seledri, sensor DHT11, *App Inventor*, metode fuzzy sugeno

## I. PENDAHULUAN

Kondisi iklim yang kurang stabil contohnya musim penghujan yang muncul secara tiba-tiba, media tanam yang sedikit akibat pembangunan yang semakin banyak, menjadi sebuah alasan banyak tanaman yang gagal panen. Pembuatan *greenhouse* merupakan salah satu solusi yang awalnya bertujuan sebagai pemenuhan akan kebutuhan bahan pangan yang terus-menerus dengan menghiraukan musim [1]. Pada proses tumbuhnya tanaman dipengaruhi oleh beberapa faktor dilingkungan yaitu suhu udara, kelembaban pada tanah, cahaya matahari, pH tanah, dll. Jika pada prosesnya tumbuhan tidak mendapatkan parameter yang menjadi syarat pertumbuhannya maka menjadi faktor gagalnya proses pertumbuhan pada tanaman tersebut.

Sebagai salah satu negara agraris Indonesia sangat kaya akan sumber daya alam yang berlimpah terutama pada sektor pertanian. Seledri termasuk jenis sayuran yang dibudidayakan di Indonesia. Seledri (*Apium Graveolens L.*) ialah tumbuhan yang memiliki fungsi yaitu menjadi obat serta sayuran dalam kebutuhan sehari-hari. Selain itu juga dipakai sebagai penyedap makanan. Seledri berfungsi untuk menambah nafsu makan dan penurun tekanan darah oleh masyarakat tradisonal. Seledri merupakan jenis tanaman subtropis, dengan suhu 15-24 °C agar pertumbuhannya bagus [2].



*Internet of things (IoT)* ialah sebuah sistem dengan tujuan menjadikan internet dalam konektivitas secara *real time* sebagai manfaatnya, sehingga proses kendali pada perangkat keras yang banyak berbasis komunikasi bisa dilakukan. *Internet of Things (IoT)* dapat diartikan menjadi sarana penyambungan sebuah alat (*things*) yang tidak manusia jalankan secara langsung pada internet oleh perangkat sebagai alat *controlling* pada peralatan elektronik atau memonitoring sebuah tanaman yang dikontrol secara *wireless* dengan internet pada jarak yang jauh. Tetapi *IoT* tidak saja berkaitan pada kendali jarak jauh, namun juga mampu difungsikan untuk berbagi informasi, menampilkan informasi dengan *real time* di internet. *User* sangat dibutuhkan sebagai pengatur proses berjalannya sistem kerja alat dengan langsung. Keuntungannya juga untuk memudahkan manusia sehingga bekerja lebih efisien dan cepat. Jadi teknologi ini sangat baik jika mampu diaplikasikan pada kehidupan sehari-hari [3].

Sistem IoT menampilkan informasi tentang perangkat elektronik yang dikendalikan oleh sebuah mikrokontroler. Mikrokontroler merupakan pengendali dan pengontrol alat atau perangkat elektronik. Beberapa contoh mikrokontroler diantaranya Arduino UNO dan NodeMCU ESP32. Arduino UNO merupakan papan mikrokontroler yang digunakan pada proyek-proyek elektronika. NodeMCU ESP32 adalah mikrokontroler yang merupakan pengembangan dari mikrokontroler NodeMCU yang telah mensupport jaringan internet sebagai pengembangan *IoT*.

Logika fuzzy mempunyai variabel keanggotaan dengan kisaran 0 ke 1. Logika fuzzy menampilkan sampai mana sebuah nilai adalah benar dan sampai mana sebuah nilai adalah salah. Logika fuzzy merupakan sebuah logika dengan memetakan suatu nilai input kepada suatu nilai output yang kontiniu. Fuzzy dijelaskan berdasarkan derajat keanggotaan dan derajat kebenaran. Oleh sebab itu nilai akan dikonversi setengah benar dan setengah salah secara bersamaan [4].

Dalam upaya penelitian ini akan akan merancang prototipe “Monitoring suhu pada tanaman seledri menggunakan *App Inventor*” sebagai sistem monitoring suhu.

## II. KAJIAN PUSTAKA

### Monitoring suhu

Pada penelitian ini dilakukan monitoring suhu ruang pada tanaman seledri berdasarkan parameter suhu yang dibutuhkan oleh tanaman seledri itu sendiri. Pada dasarnya proses monitoring ini bertujuan untuk mengetahui suhu udara disekitar tanaman seledri menggunakan *smartphone* android. Yang nantinya apabila suhu melebihi atau kurang dari parameter suhu yang dibutuhkan tanaman seledri maka secara otomatis kipas dan heater akan bekerja berdasarkan suhu yang dibaca oleh sensor.

### Tanaman Seledri

Seledri adalah jenis tanaman semusim dengan bentuk rumput atau semak. Seledri terdiri atas akar, batang dan tangkai daun. Seledri mempunyai daun tipis, rapuh, dengan bentuk belah ketupat miring, yang Panjangnya 2-8cm dan lebarnya 2-5cm, ujung dan pangkalnya meruncing, panjang tangkai anak daunnya 1-3 cm. Seledri mempunyai usia tanam sekitar 8-16 minggu sesuai varietasnya. Pertumbuhan seledri dikatakan maksimal jika jumlah tangkai, batang, dan daunnya kelihatan banyak [5]. Seledri dapat tumbuh dan berkembang dengan baik pada ketinggian diatas 0-1200 Mdpl, oleh karena seledri cocok dibudidayakan di dataran rendah ataupun tinggi, dengan kelembaban berkisar 80% sampai 90% serta cukup akan sinar matahari. Dalam pertumbuhannya tanaman seledri memerlukan suhu minimal 24-30 °C, tetapi pertumbuhan seledri dimaksimalkan lagi jika berada pada daerah pegunungan dengan suhu berkisar antara 18-24 °C.

### NodeMCU ESP32

ESP32 merupakan mikrokontroler yang dibuat oleh *Espireessif System* ialah hasil perkembangan dari ESP8266. Mikrokontroler ini telah dilengkapi *Bluetooth* dan *WiFi* berbentuk *chip* yang



mensupport dalam pembuatan sistem aplikasi *Internet of Things* [6]. ESP32 mempunyai kelebihan yang cukup sedikit jika dibandingkan dengan ESP8266 yang diedarkan pada pasaran, akan tetapi ESP32 lebih rumit. Mulai daripada segi *Wi-Fi* dan *Bluetooth*, ESP32 sudah terintegrasi dengan *system on ship*, ESP32 juga mempunyai pin GPIO berjumlah 32 pin yang lebih banyak dari ESP8266.

### **Arduino IDE**

Arduino IDE (*Integrated Development Environment*) merupakan perangkat lunak yang dipakai dalam pemrograman di arduino, yang berfungsi sebagai aplikasi pemrograman *board* arduino. Arduino berfungsi sebagai editor teks dalam hal mengedit, memasukan dan juga mengecek kebenaran kode program. Selain itu dapat dipakai untuk upload program ke papan arduino. Kelebihan lainnya adalah arduino IDE dapat digunakan dalam sistem operasi linux, Mac OS bahkan Windows. Pada *Software* Arduino IDE terdapat environment tertulis pada Java. Bahasa C++ atau C yang digunakan juga terdapat *compiler* [7]. Fungsi lainnya pada dasarnya adalah membuat *code* atau *script* pada banyak jenis *microprocessor* atau *board* yang mendukung.

### **Driver mosfet**

Mosfet (*Metal Oxide Semiconductor Field Effect Transistor*) adalah jenis transistor yang mempunyai impedansi masukan (*gate*) begitu tinggi oleh karena itu mosfet dapat digunakan sebagai saklar elektronik (metode *switching*) [8]. Jenis mosfet yang dipakai pada penelitian ini adalah driver mosfet D4184 Dual.

### **App Inventor**

*App Inventor* merupakan aplikasi web terbuka yang pada mulanya diciptakan oleh *Google*, yang sekarang berada dibawah naungan *Massachusetts Institute of Technology* (MIT). *App Inventor* memberikan manfaat bagi pemula dalam pemrograman untuk membuat aplikasi *software* pada android. *App Inventor* memberikan tampilan grafik *interface*, menyerupai *interface* pemakai pada *Star Logo* TNG dan *Scratch*, yang memudahkan pemakai dalam men-*drag-and-drop visual object* dalam membuat aplikasi yang dapat bekerja pada sistem android. *App Inventor* memakai teknik *visual programming*, memiliki bentuk berupa *puzzle-puzzle* yang tersusun dan mempunyai logika tersendiri [9].

### **Sensor DHT11**

Sensor DHT11 ialah sensor yang dipakai mengukur suhu dan kelembaban. Sensor DHT11 lebih unggul karena mempunyai tingkat presisi yang bagus dan fitur kalibrasi yang tersimpan di OTP program memori dengan kepresisian yang tinggi [10].

### **Mini AC Portable**

AC Mini Portable adalah perangkat elektronik yang berguna untuk menyalurkan hawa dingin menggunakan bantuan berupa air sebagai perangkat pendingin sehingga suhu yang berada dalam mini greenhouse bisa turun meskipun secara perlahan [11].

### **Mini Heater**

Electrical Heating Element atau elemen pemanas listrik sering diaplikasikan pada kehidupan manusia, entah pada rumah tangga maupun mesin dan peralatan pada pabrik/perusahaan. Electrical Heating Element memiliki jenis yang beragam berdasarkan tempat, media dan fungsinya. Panas yang dihasilkan Electrical Heating Element ini berasal dari pita atau kawat dengan ketahanan listrik yang tinggi [12].

### Mudul AC light Dimmer

Modul ini mayoritas juga dipakai dalam proyek yang berhubungan dengan heater, motor, ataupun lampu dimana modul AC light dimmer memiliki fitur zero cross detector yang mengatur jumlah tegangan bolak balik dengan mengubah sinyal AC menjadi sinyal potongan-potongan [13].

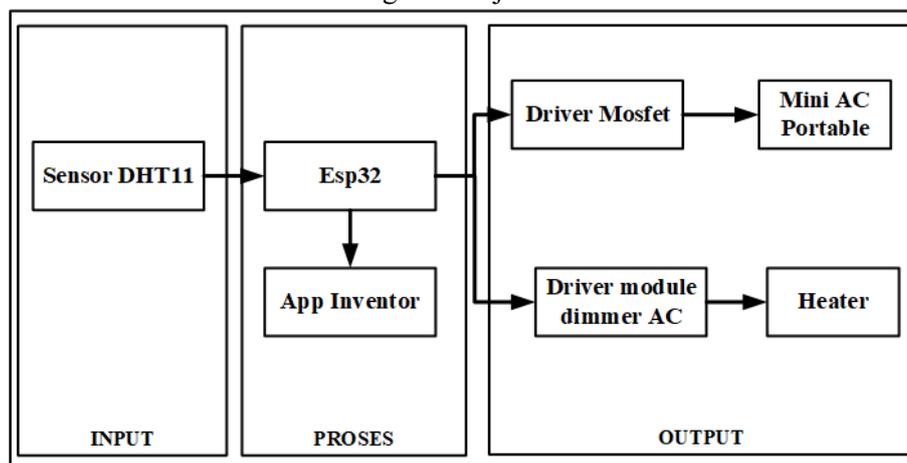
### Fuzzy Sugeno

Fuzzy Sugeno merupakan Penalaran dengan metode Sugeno hampir sama dengan penalaran Mamdani, hanya saja output (konsekuen) sistem tidak berupa himpunan Fuzzy, melainkan berupa konstanta atau persamaan linear. Michio Sugeno mengusulkan penggunaan singleton sebagai fungsi keanggotaan dari konsekuen. Singleton adalah sebuah himpunan Fuzzy dengan fungsi keanggotaan yang pada titik tertentu mempunyai sebuah nilai 1 dan 0 di luar titik tersebut. Berdasarkan model Fuzzy tersebut, ada tahapan-tahapan yang harus dilakukan dalam implementasi metode Sugeno yaitu sebagai berikut:

- ✓ Pembentukan Himpunan Fuzzy  
Variabel input akan diproses ke dalam himpunan fuzzy dengan mengambil nilai-nilai tegas dan menentukan derajat keanggotaannya.
- ✓ Aplikasi fungsi Implikasi  
Menentukan aturan yang akan digunakan pada fungsi implikasi dengan menggunakan operator fuzzy IF dan THEN.
- ✓ Defuzzyfikasi  
Mengubah nilai keanggotaan himpunan fuzzy menjadi nilai keluaran dengan cara menghitung rata-ratanya [14].

### III. METODE PENELITIAN

Pada bagian ini merupakan tampilan dari Blok Diagram kerja alat, *flowchart*, dan skematik rangkaian. Berikut adalah Gambar Blok Diagram Kerja alat.



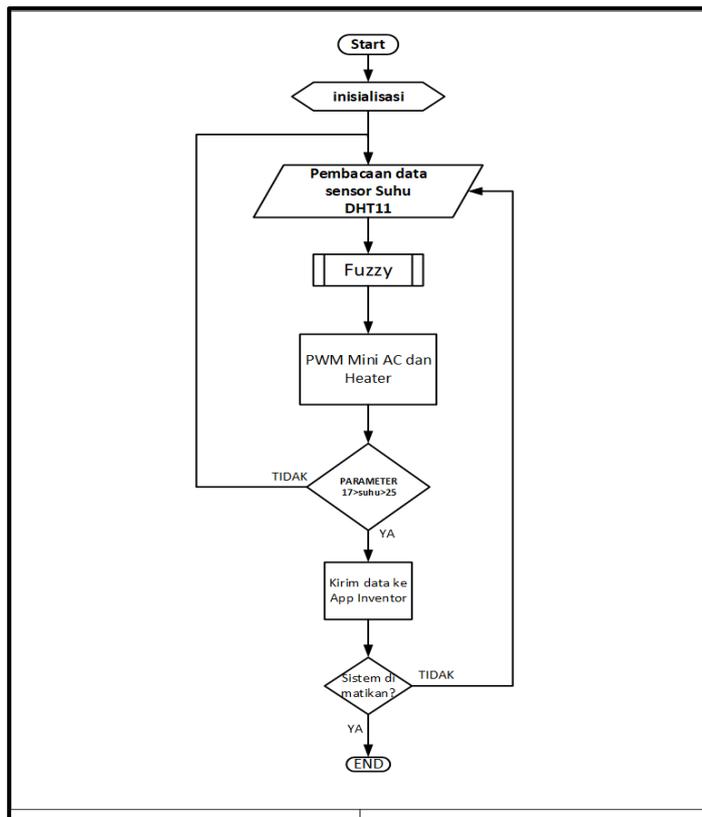
**Gambar 1.** Diagram Blok kerja Alat

Pada Gambar 1 merupakan tampilan Diagram Blok kerja alat yang terbagi menjadi Input, Proses dan Output. Pada bagian input terdapat sensor DHT11. Pada bagian proses diantaranya Esp32 dan *App inventor* sebagai perangkat atau IoT untuk monitoring dan mengontrol proses kerja mini greenhouse melalui jaringan internet. Pada bagian output diantaranya *mini AC portable* dan *heater*.

Fungsi dan tugas dari masing-masing komponen pada diagram blok diatas adalah Sensor DHT11 digunakan untuk mengukur tingkat suhu udara disekitar tanaman seledri. NodeMCU ESP32 yang bertugas untuk mengontrol dan mengolah data sensor yang akan dikirimkan data input dari sensor yang digunakan menuju output untuk kemudian di kirim ke Android melalui *App inventor*. *Heater*

berfungsi menaikkan suhu udara apabila suhunya terlalu rendah atau berada dibawah 18 °C. *Mini AC portable* berfungsi menurunkan suhu udara apabila suhunya terlalu tinggi atau berada di atas 24 °C.

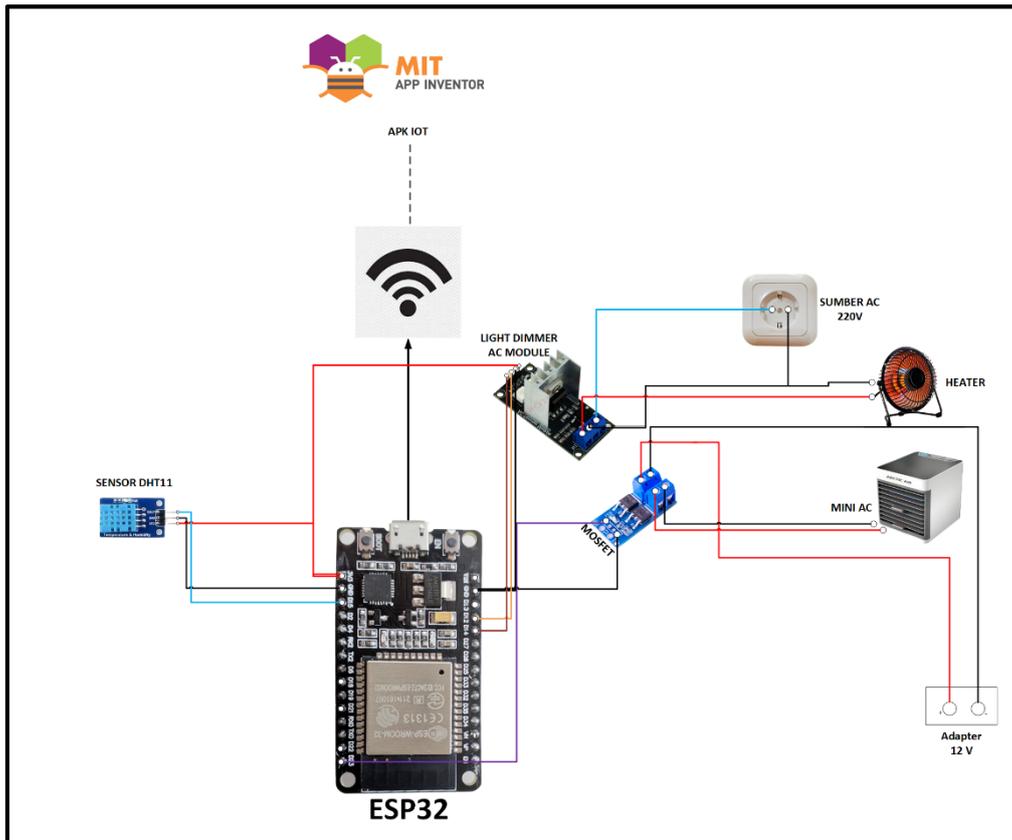
*Flowchart* merupakan gambaran dari system kerja pada software. Berikut adalah Gambar dari *flowchart* pada penelitian ini.



**Gambar 2.** Diagram Blok kerja Alat

Gambar 2 merupakan perangkat lunak yang bermula dari proses inisialisasi, pembacaan data sensor untuk membaca nilai suhu udara pada mini greenhouse. Dilanjutkan metode fuzzy untuk mengatur PWM mini AC dan Heater. Jika parameter input data suhu sudah sesuai maka akan kirim data monitoring ke App inventor jika tidak maka PWM dari mini AC dan Heater akan diatur kemudian dilakukan pembacaan data sensor ulang. Apabila sistem ingin dimatikan maka sistem akan berhenti, jika tidak maka dilakukan pembacaan sensor ulang.

Skematik rangkaian merupakan gambaran dari sistem pengkabelan pada penelitian ini. Berikut adalah Gambar dari skematik rangkaian pada penelitian ini.



**Gambar 3.** Diagram Blok kerja Alat

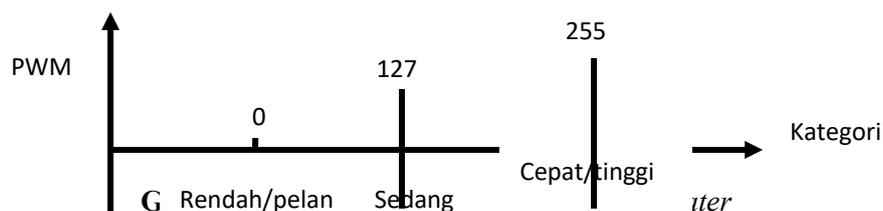
Gambar 3 merupakan gambar skematik rangkaian pada sistem pada penelitian ini yang mana input pin DHT11 yaitu pin VCC akan dihubungkan ke pin 3,3 V ESP32, pin GND dihubungkan ke pin GND ESP32, dan pin Data Output DHT11 dihubungkan ke pin D15 ESP32. Pin mosfet yaitu GND dihubungkan ke pin GND ESP32, pin PWM/Trigger dihubungkan ke pin D23 ESP32, pin Vin+ dan Vin- dihubungkan ke power supply, dan pin Vout+ dan Vout- dihubungkan ke *mini AC portable*. Pin modul AC *light Dimmer* yaitu pin GND dihubungkan ke pin GND ESP32, pin VCC dihubungkan ke pin 3,3 V ESP32, pin *ZeroCross* dihubungkan ke pin D12 ESP32, pin PWM dihubungkan ke pin D14 ESP32, pin Vin+ dan Vin- dihubungkan ke sumber AC, dan pin Vout+ dan Vout- dihubungkan ke *mini heater*.

Pada penelitian ini menggunakan metode *fuzzy sugeno* untuk mengontrol PWM pada output. Berikut merupakan Tabel input derajat keanggotaan suhu pada fuzzy sugeno.

**Tabel 1.** Input derajat keanggotaan fuzzy suhu

No	Fungsi keanggotaan input	Range
1	Sangat Dingin	2-10 °C
2	Dingin	8-19 °C
3	Normal	18-24 °C
4	Panas	23-34.5 °C
5	Sangat panas	32-45 °C

Tabel 1 diatas merupakan tabel yang berisi derajat keanggotaan suhu yang digunakan dalam menentukan kategori suhu dan putaran pwm output nantinya. Berikut merupakan nilai PWM yang digunakan pada output *mini AC* dan *heater*.



Setelah penentuan input dan output maka langkah pembuatan *rules*. Berikut adalah tampilan *rules fuzzy* yang digunakan pada penelitian ini.

```

1. If (Suhu is sangat_dingin) then (mini_AC is PELAN)(mini_heater is TINGGI) (1)
2. If (Suhu is dingin) then (mini_AC is PELAN)(mini_heater is SEDANG) (1)
3. If (Suhu is normal) then (mini_AC is PELAN)(mini_heater is RENDAH) (1)
4. If (Suhu is panas) then (mini_AC is SEDANG)(mini_heater is RENDAH) (1)
5. If (Suhu is sangat_panas) then (mini_AC is CEPAT)(mini_heater is RENDAH) (1)
  
```

Gambar 5. *Rules fuzzy*

#### IV. HASIL DAN PEMBAHASAN

Dalam penelitian meneliti tentang hasil monitoring suhu pada tanaman seledri sekaligus menampilkan hasil perbandingan pwm pada Arduino IDE dan simulasi pada matlab. Berikut merupakan tabel hasil perbandingan PWM pada Arduino IDE dan simulasi pada Matlab.

Tabel 2. Input derajat keanggotaan fuzzy suhu

NO	Input Suhu	Output PWM serial monitor		Output PWM fuzzy matlab		Selisih
		Mini heater	Mini AC	Mini heater	Mini AC	
1	5	255	0	255	0	0
2	10	127	0	27	0	0
3	15	127	0	127	0	0
4	19	42	0	42.3	0	0.3
5	20	0	0	0	0	0
6	21	0	0	0	0	0
7	22	0	0	0	0	0
8	23	0	27.8	0	27.2	0.4
9	24	0	127	0	127	0
10	25	0	127	0	127	0
11	30	0	127	0	127	0
12	35	0	255	0	255	0
<b>Total</b>		551	663.8	551.3	663.2	0.7
<b>rata-rata</b>		45.91	55.31	45.94	55.26	0.058

Dari Tabel diatas dilakukan 12 percobaan sehingga dihasilkan selisih antara PWM mini heater pada Arduino IDE dan Pada Matlab adalah 0.3 dan selisih antara PWM mini AC pada Arduino IDE dan pada Matlab adalah 0.4 dengan rata-ratanya 0,058. Data hasil monitoring tersebut akan ditampilkan pada *App Inventor* di *smartphone*. Berikut merupakan tampilan *App Inventor* di *smartphone*.



**Gambar 5.** *Rules fuzzy*

Pada gambar diatas merupakan tampilan *App Inventor* yang menunjukkan hasil monitoring suhu di sekitar tanaman seledri adalah 27 °C. Hasil pembacaan tersebut akan dikelola lagi oleh ESP32 yang menentukan apakah suhu tersebut sesuai dengan range atau parameter yang dibutuhkan oleh tanaman seledri, karena suhu tersebut adalah 27 °C maka sudah melebihi suhu normal yang nantinya ESP32 akan mengontol mini AC untuk menurunkan suhu menjadi normal dengan kisaran 18-24 °C dengan PWM mini AC 127.

## V. KESIMPULAN

Kesimpulan dari penelitian ini adalah *metode fuzzy sugeno* dapat diimplementasikan pada sistem kontrol suhu pada tanaman seledri menggunakan mini AC dan heater dengan selisih rata-rata 0.03, dan *App inventor* dapat digunakan sebagai aplikasi monitoring suhu pada tanaman seledri.

## UCAPAN TERIMA KASIH

Ucapan Terima kasih kepada orang-orang yang telah membantu saya sehingga artikel ini dapat ditulis dengan baik dan benar.

## REFERENSI

1. M. Yusniar, "Smart greenhouse tanaman seledri berbasis raspberry pi menggunakan internet of things (IoT).," 2021
2. T. & B. I. I. Pranata, "penerapan logika fuzzy pada sistem penyiraman tanaman otomatis berbasis mikrokontroler," *Jurnal Coding sistem Komputer Universitas Tanjung Pura*.
3. W. R. A. & S. I. Kurniasih, "Sistem Keamanan Pintu dan Jendela Rumah Berbasis IoT," *Jurnal Riset Sistem Informasi dan Teknik Informatika (Jurasic)*, pp. 266-274, 2020.
4. I. B. E. Putra, "Perencanaan Penyiraman Otomatis Bertenaga Surya Berbasis Arduino Uno untuk Tanaman Bibit Jenitri," *Jurnal SinarFe7*, pp. 427-432, 2018.
5. M. Yusniar, "smart greenhouse tanaman seledri berbasis raspberry pi menggunakan Internet of Things (IoT)," 2021.
6. H. D. Nur Isnianto, "Sistem Telemonitoring KWH Meter menggunakan modul Wi-Fi ESP8266 berbasis Arduino Uno," *Jurnal Rekayasa Elektrika*, vol. 15, pp. 25-33, 2019.



7. M. Fauzan Zarkashie, "Rancang bangun sistem pengukuran kualitas air untuk keperluan Higiene Sanitasi berbasis Arduino Uno," 2021.
8. I. Aqidatul, "Rancang bangun dan Analisis Inverter Full Bridge 1 Fasa dengan Berbagai Variasi Input Menggunakan SPWM (Sinusoidal Pulse Width Modulation)," 2017.
9. A. Kadir, "Panduan praktis mempelajari Aplikasi mikrokontroler dan Pemrogramannya menggunakan Android.," 2011.
10. R. Syam, Dasar-dasar Teknik Sensor, Makassar, 2013.
11. B. B. S. A. & F. H. B. Prasetya, "Shield Mamdani pada tanaman Tomat Hidroponik," 2019.
12. Centralindo, T., "Heater," 2011.
13. A. Kurniawan, "Dimmer PWM Arduino," 2018.
14. D. Rahakbauw, "penerapan Logika fuzzy Metode Sugeno berdasarkan data persediaan dan jumlah permintaan (studi kasus : Pabrik Roti Ambon)," *Jurnal Teknik Informasi dan Sistem Informasi*, vol. 9, pp. 121-134, 2015.



# Monitoring Kelembaban Tanah dengan Menggunakan Metode Fuzzy Sugeno pada Bunga Krisan Berbasis IoT

Yosefa Stefania Fono<sup>1</sup>, Aries Boedi Setiawan<sup>2</sup>, Delila cahya Permatasari<sup>3</sup>

<sup>1,3</sup>Yosefa Stefania Fono (Jurusan Teknik Elektro, Universitas Merdeka Malang)

<sup>1</sup>[yosefastefaniafono12@gmail.com](mailto:yosefastefaniafono12@gmail.com)

<sup>3</sup>[delila.permatasari@unmer.ac.id](mailto:delila.permatasari@unmer.ac.id)

<sup>2</sup>Aries Boedi Setiawan (Jurusan Teknik Elektro, Universitas Merdeka Malang)

Corresponding author email: [aries@unmer.ac.id](mailto:aries@unmer.ac.id)

**Abstract:** Land as the main factor in agriculture must be considered properly to achieve the desired results. One of the people's hobbies is farming which requires a certain time and watering schedule, such as the cultivation of chrysanthemums. Soil moisture for the growth of this flower requires 70-80 percent. Plants can wilt if moisture is insufficient, while excessive moisture can reduce root volume for growth. Manual watering is sometimes less effective due to time constraints, climate and soil conditions. For that we need a device that can monitor soil moisture automatically and help water the plants. One of the information and communication technology innovations in agriculture is the use of the Internet of Things. This can be done to monitor soil moisture which is a growing medium for horticultural plants by using the Internet of Things. To determine the steps or soil handling it will be very helpful to know the soil moisture value. In this study, the Sugeno fuzzy method is used which is simpler and more precise as a decision maker.

**Keywords:** soil moisture, chrysanthemum flower, IoT, Fuzzy Sugeno

**Abstrak:** Tanah sebagai faktor utama dalam pertanian harus diperhatikan dengan baik Untuk mencapai hasil yang diinginkan. Salah satu hobi masyarakat adalah bercocok tanam yang membutuhkan waktu dan jadwal penyiraman tertentu seperti pada budidaya bunga krisan. kelembaban tanah untuk pertumbuhan bunga ini membutuhkan 70-80 persen. Tanaman dapat layu jika kelembaban tidak mencukupi, sedangkan kelembaban yang berlebihan dapat mengurangi volume akar untuk tumbuh kembang. Penyiraman manual terkadang kurang efektif karena keterbatasan waktu, iklim dan kondisi tanah. Untuk itu dibutuhkan perangkat yang dapat memantau kelembaban tanah secara otomatis dan membantu menyiram tanaman. Salah satu inovasi teknologi informasi dan komunikasi di bidang pertanian adalah penggunaan Internet of Things. Hal ini dapat dilakukan untuk memantau kelembaban tanah yang merupakan media tumbuh tanaman hortikultura dengan menggunakan Internet of Things. Untuk menentukan langkah-langkah atau penanganan tanah akan sangat membantu dengan mengetahui nilai kelembaban tanah. Pada penelitian ini digunakan metode fuzzy Sugeno yang lebih sederhana dan presisi sebagai pengambil keputusan.

Kata kunci: Kelembaban Tanah, Bunga Krisan, IoT, Fuzzy Sugeno

## I. PENDAHULUAN

Bertanam adalah suatu kegiatan untuk merawat atau membudidayakan suatu tanaman yang sangat banyak digemari para penduduk Indonesia. Kegiatan bercocok tanam ini tidak hanya dilakukan oleh masyarakat yang tinggal di pedesaan tetapi juga oleh masyarakat yang tinggal di perkotaan. Penghuni metropolitan menjadikan penggarapan bunga sampingan [1]. Kegiatan menanam dapat dilakukan di pekarangan rumah bagi masyarakat yang tinggal di perkotaan. Mereka tidak hanya dapat menyalurkan hobinya, tetapi juga dapat mempercantik rumah dan menciptakan lingkungan yang asri.

Setiap tanaman akan menyerap kelembaban yang cukup dari tanah untuk perkembangannya. Tanaman akan tampak tidak segar jika media tanam kering dan kelembaban tanah berada di bawah titik batas tertentu selain itu, jika kadar air tanah tidak diperlukan, akan menyebabkan kadar oksigen



dalam tanah turun sehingga nafas akar terhambat, terganggu, dan zat beracun terbentuk. berkurangnya volume akar yang mengangkut air dan suplemen melalui akar. Krisan akan tumbuh subur saat kelembapan tanahnya berkisar 70-80%[2].

Tingkat kesuburan tanah dipengaruhi oleh jumlah air yang ada di dalam tanah. Fotosintesis sangat tergantung pada ketersediaan air. Tujuan menyediakan air bagi tanaman adalah untuk memastikan bahwa mereka memiliki cukup air yang tersedia untuk fotosintesis. Penyiraman manual dinilai tidak efisien karena memakan banyak waktu dan tenaga, serta pemilik tidak bisa membiarkan tanaman terlalu lama karena bisa kehabisan air dan mati[3] [4].

Dengan adanya permasalahan tersebut, maka akan dirancang alat penyiram tanaman yang dapat bekerja secara otomatis berdasarkan kondisi lingkungan. Kelembaban tanah merupakan salah satu parameter desain yang dapat dikontrol oleh alat penyiram tanaman ini. Dalam penelitian ini hanya memanfaatkan aplikasi Blynk yang sudah tersedia untuk Android, dan sistem otomatis yang memanfaatkan teknologi IoT dan dapat memantau dari jarak jauh. Teknologi IOT menyimpan program dalam mikrokontroler yang berfungsi sebagai pengontrol rangkaian elektronik[5].

Logika fuzzy Sugeno digunakan dalam perancangan alat penyiraman otomatis berbasis IoT ini untuk mengontrol sistem kontrol yang membuat keputusan menjadi lebih sederhana dan presisi [6] Permasalahan yang akan diuraikan adalah merancang alat penyiram tanaman otomatis menggunakan *Internet of Things (IOT)* dengan latar belakang tersebut. Tujuannya agar dapat merancang alat penyiram tanaman otomatis yang menggunakan air untuk menghasilkan lebih banyak bunga krisan dengan kualitas dan kuantitas yang lebih tinggi.

## II. KAJIAN PUSTAKA

### Bunga Krisan

Bunga krisan atau nama latinnya *Chrysanthemum* adalah salah satu bunga yang banyak dibudidayakan dan memiliki ragam jenisnya. Tanaman ini memang dapat dijadikan sebagai sumber atau sarana untuk memperoleh penghasilan yang menjanjikan bagi petani bunga krisan. Kelembaban udara bunga krisan yang baik adalah antara 60% - 70%. Jika kelembaban terlalu tinggi, penguapan air tanaman akan menjadi lebih kecil dalam waktu singkat, dan tanaman akan selalu terlihat segar. jangkauan dan membutuhkan kelembaban tinggi untuk rooting awal 90-95%. Tanaman muda sampai dewasa membutuhkan 70-80%[7].

### NodeMCUEsp32

Nodemcu esp32 adalah mikrokontroler yang sangat baik digunakan karena memiliki wifi dan bluetooth dibandingkan dengan esp8266 dan sangat mudah digunakan dalam sistem iot untuk memonitoring suhu dan kelembaban paa bunga krisan. Dengan *WiFi* modul terintegrasi ke dalam chip mikrokontroler, *Bluetooth dual-mode*, dan fitur hemat daya untuk meningkatkan fleksibilitas, mikrokontroler ESP32 menyediakan sistem berbiaya rendah dan berdaya rendah[8].

### NodeMCU ESP32

ESP32 merupakan mikrokontroler yang dibuat oleh *Espiressif System* ialah hasil perkembangan dari ESP8266. Mikrokontroler ini telah dilengkapi *Bluetooth* dan *WiFi* berbentuk *chip* yang mensupport dalam pembuatan sistem aplikasi *Internet of Things* [6]. ESP32 mempunyai kelebihan yang cukup sedikit jika dibandingkan dengan ESP8266 yang diedarkan pada pasaran, akan tetapi ESP32 lebih rumit. Mulai daripada segi *Wi-Fi* dan *Bluetooth*, ESP32 sudah teritegrasi dengan *system on ship*, ESP32 juga mempunyai pin GPIO berjumlah 32 pin yang lebih banyak dari ESP8266 .

### Arduino IDE



Arduino IDE (*Integrated Development Environment*) merupakan perangkat lunak yang dipakai dalam pemrograman di arduino, yang berfungsi sebagai aplikasi pemrograman *board* arduino. Arduino berfungsi sebagai editor teks dalam hal mengedit, memasukan dan juga mengecek kebenaran kode program. Selain itu dapat dipakai untuk upload program ke papan arduino. Kelebihan lainnya adalah arduino IDE dapat digunakan dalam sistem operasi linux, Mac OS bahkan Windows. Pada *Software* Arduino IDE terdapat environment tertulis pada Java. Bahasa C++ atau C yang digunakan juga terdapat *compiler*. Fungsi lainnya pada dasarnya adalah membuat *code* atau *script* pada banyak jenis *microprocessor* atau *board* yang mendukung[9].

### Sensor Kelembaban Tanah

*Soil Moisture Sensor HW-390* adalah perangkat dua..probe yang mengukur tingkat kelembaban dengan melewatkan arus listrik melalui tanah dan membaca resistansinya. Oleh karena itu, nilai sensor akan lebih tinggi (resistensi lebih tinggi) di tanah kering dibandingkan..dengan tanah dengan kandungan udara yang lebih tinggi (resistensi lebih rendah) Untuk menampilkan tingkat kelembaban tanaman dan mengintegrasikan kelembaban tanah, sensor ini sangat berguna[10].

### IoT (*Internet of Things*)

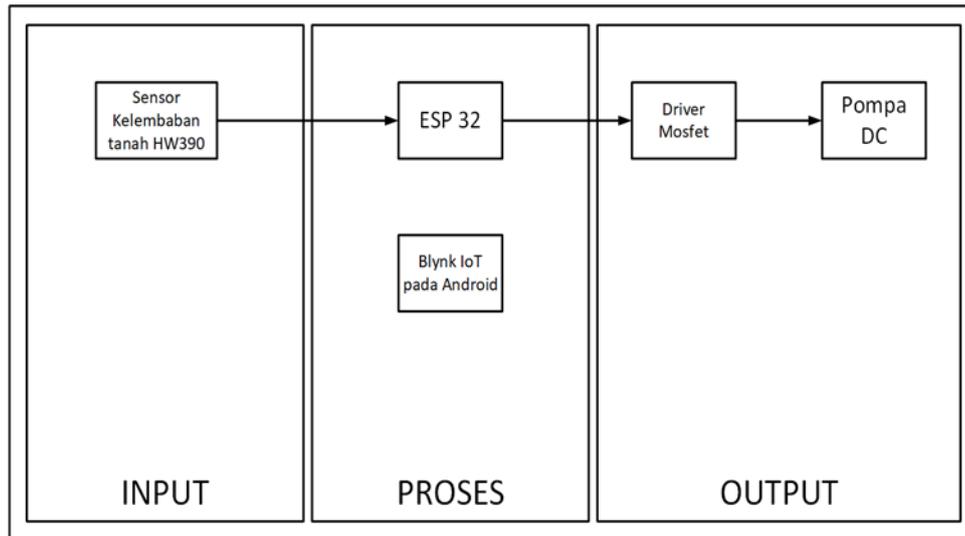
Menurut (*Burange & Misalkar, 2015*) IoT adalah struktur di mana objek, orang diberikan identitas eksklusif dan kemampuan untuk memindahkan data melalui jaringan tanpa memerlukan dua arah manusia kemanusia yaitu sumber untuk tujuan atau interaksi manusia kemanusia. IoT adalah pengembangan ilmiah yang sangat menjanjikan yang mengoptimalkan perangkat berdasarkan sensor cerdas dan cerdas yang bekerja bersama melalui Internet[11].

### Fuzzy Sugeno

Metode *fuzzy Sugeno* adalah metode inferensi *fuzzy* untuk aturan yang dinyatakan dalam bentuk *IF-THEN*. Metode *fuzzy* dikembangkan oleh *Sugeno* yang telah terbukti efektif ketika berurusan dengan sistem nonlinier kompleks yang sangat menantang untuk dianalisis dan disintesis Model ini menggunakan fungsi keanggotaan tunggal. fungsi keanggotaan tunggal adalah dimana satu nilai crisp memiliki derajat keanggotaan 1 dan nilai crisp lainnya memiliki derajat keanggotaan 0. Dalam metode ini, setiap keluaran yang dibentuk dari aturan *IF THEN* bukanlah himpunan *fuzzy*, tetapi persamaan konstan atau linier[12].

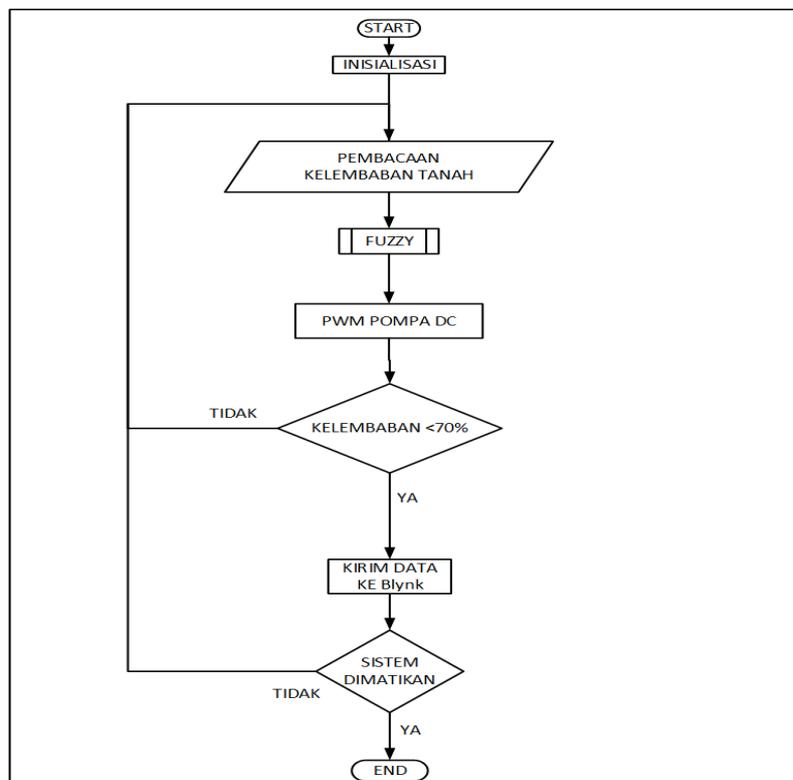
## III. METODE PENELITIAN

Pada bagian ini merupakan tampilan dari Blok Diagram kerja alat, *flowchart*, dan Diagram Fuzzy . Saat membuat desain rancangan alat sangat membutuhkan diagram blok yang menggambarkan tahapan atau proses yang dilakukan Berikut adalah Gambar Blok Diagram Kerja alat.

**Gambar 1.** Diagram Blok kerja Alat

Gambar 1 merupakan tampilan diagram blok alat yang terbagi menjadi Input, Proses dan Output. Pada bagian input terdapat sensor kelembaban tanah Hw-390 untuk mengukur kelembaban tanah. Blok proses terdapat Esp32 Sebagai Mikrokontroler utama dan sebagai modul *Wifi* untuk mengirimkan data, dan juga *Aplikasi Blynk* untuk mengontrol dan monitoring secara real time sehingga *user* tidak perlu mengecek secara manual jika user tidak berada dirumah. Pada bagian output terdapat *driver mosfet* sebagai pengendali pompa Dc.

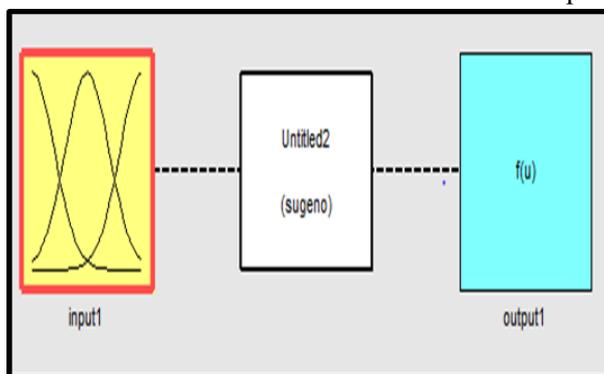
*Flowchart* merupakan rangkaian alur pada sistem kerja rangkaian. Berikut adalah Gambar dari *flowchart* pada penelitian ini.

**Gambar 2.** Flow Chart

Gambar 2 merupakan perangkat lunak yang bermula dari proses inisialisasi, pembacaan data sensor kelembaban tanah. Kemudian masuk pada proses fuzzy untuk mengatur PWM pompa dc . Jika data kelembaban tanah sudah sesuai maka akan kirim data monitoring ke aplikasi pada android jika

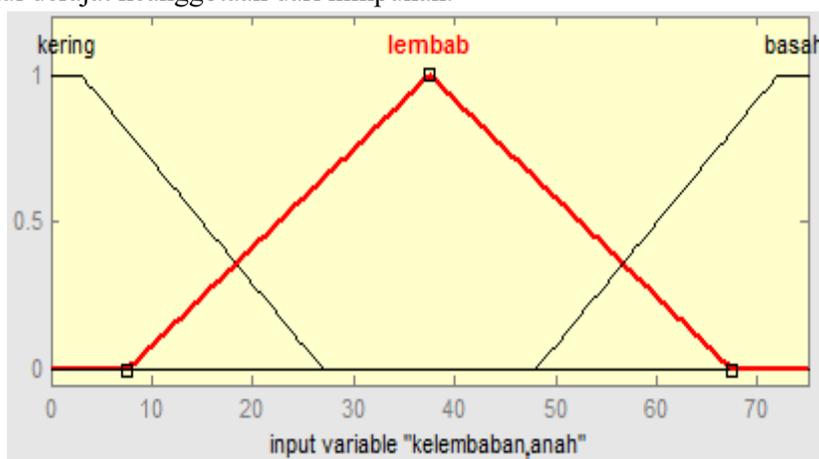
tidak maka PWM pompa dc akan diatur kemudian dilakukan pembacaan ulang jika sudah selesai maka sistem akan dimatikan dan selesai.

Diagram Fuzzy digunakan untuk mengubah nilai data sensor menjadi himpunan *fuzzy* selama penelitian dan pembangunan rumah kaca Ini diproses segera, menghasilkan *fuzzy IF-THEN* yang menggambarkan hubungan antara masukan dan keluaran. Berikut Merupakan gambar diagram fuzzy



**Gambar 3.** Diagram Fuzzy sugeno

Untuk mendapatkan fungsi keanggotaan pada penelitian ini digunakan variabel kelembaban tanah dengan bentuk gabungan kurva segitiga, kurva linear turun dan kurva linear naik seperti Gambar dan untuk dapat mengetahui derajat keanggotaan dari himpunan seperti Gambar 4 dan untuk dapat mengetahui derajat keanggotaan dari himpunan.



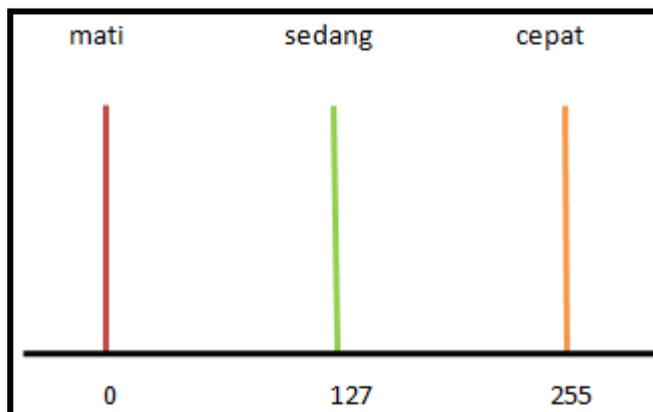
**Gambar 4.** Input variabel kelembaban Tanah

Gambar 4 merupakan *domain range* kelembaban tanah dengan fungsi sebagai batasan nilai yang akan digunakan pada mikrokontroler untuk mengambil keputusan dari hasil pengukuran sensor kelembaban tanah *HW-390* dengan batasan *range* dari 0-75%

**Tabel 1.** Domain Range kelembaban tanah

No	Input kelembaban tanah	Range(%)
1	Kering	[3-27 ]
2	Lembab	[37.5-67.5]
3	Basah	[48-72]

Tabel 1 diatas merupakan *Domain range* kelembaban tanah. Kemudian untuk mengatur PWM pompa Dc untuk penyiraman otomatis dengan sistem fuzzy dapat mengatur pompa dc sesuai dengan kondisi kelembaban tanah. Berikut ini merupakan tampilan output pada Pwm pompa dc untuk mengatur mati hidup air dalam penyiraman otomatis. Tampilan fungsi keanggotaan output pompa dc pada Gambar 5



**Gambar 5.** Fungsi Keanggotaan Output pompa dc

Gambar 5 merupakan tampilan output pwm pompa dc yaitu mati ,sedang dan cepat dengan nilai 0, pelan dengan nilai 127, dan sangat cepat dengan nilai 255.

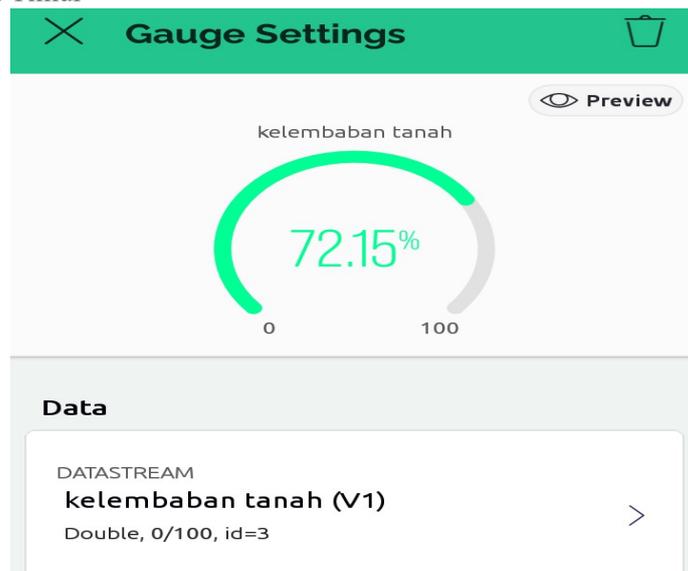
#### IV. HASIL DAN PEMBAHASAN

Pada Hasil pembahasan ini menjelas mengenai data yang diambil dan dicantumkan berupa nilai-nilai. Berikut merupakan tabel perbandingan output Pwm pompa dc pada arduino ide dan fuzzy matlab.

**Tabel 2.** Perbandingan output Pwm pompa dc pada arduino ide dan fuzzy matlab

NO	Input kelembaban tanah	Output PWM serial monitor	Output PWM fuzzy matlab	Selisih
		Pompa dc	Pompa dc	
1	65	0	0	0
2	60	0	0	0
3	45,20	140	140	0
4	42	140,2	140	0,2
5	35,60	127	127	0
6	10	202,50	203	0,5
7	7,80	255	255	0
8	1.50	255	255	0
<b>Total</b>		<b>1119.7</b>	<b>1120</b>	<b>0,7</b>
<b>Rata-rata</b>		<b>139,96</b>	<b>140</b>	<b>0,08</b>

Dari Tabel 4.5 Pengujian di atas yang dilakukan sebanyak 8 kali dengan mendapatkan hasil rata-rata selisih sebesar 0.08. Data hasil monitoring tersebut akan ditampilkan pada *Aplikasi blynk* yang ada di *smartphone*. Berikut merupakan tampilan *Aplikasi Blynk* di *smartphone*.



**Gambar 6.** Tampilan Aplikasi *Blynk* pada *Smartphone*

Gambar 6 menunjukkan nilai kelembaban tanah 72,15% yang merupakan hasil data yang di terima pada aplikasi *blynk* pada *android*.

## V. KESIMPULAN

Kesimpulan dari penelitian ini adalah metode fuzzy sugeno dapat diimplementasikan pada sistem kontrol kelembaban tanah pada bunga krisan menggunakan pompa dc dengan selisih rata-rata 0.08, dan Aplikasi blynk pada smartphone dapat digunakan sebagai aplikasi monitoring kelembaban tanah pada bunga krisan.

## UCAPAN TERIMA KASIH

Ucapan Terima kasih kepada orang-orang yang telah membantu saya sehingga artikel ini dapat ditulis dengan baik dan benar.

## REFERENSI

1. Al-Bahadly, I., & Thompson, J. (2015, December). Garden watering system based on moisture sensing. In 2015 9th international conference on sensing technology (ICST) (pp. 263-268). IEEE.
2. Affandy, I., & Raharja, W. K. (2021). Pemanfaatan internet of things untuk telemonitoring rumah kaca tanaman krisan. *Jurnal Ilmiah Teknologi dan Rekayasa*, 26(2), 79-93.
- A. W. Suhardi Sitti Nur Faridah, “Kinerja Sistem Kontrol Kadar Air Tanah Pada Operasi Sistem Irigasi,” *Sprink. J Agri Techno*, vol. 6, no. 2, pp. 1–9, 2014.
3. Nurdiana, N., & Perawati, P. (2021). Monitoring Kelembaban Tanah Pada Penyiram Tanaman Otomatis. *Jurnal Tekno*, 18(1), 9-15.
4. Wilanda, A., Pasaribu, F. N., & Amelia, A. (2021). RANCANG BANGUN ALAT PENYIRAM TANAMAN DAN MONITORING OTOMATIS BERBASIS INTERNET OF THINGS (IoT). *Prosiding Konferensi Nasional Social & Engineering Polmed (KONSEP)*, 2(1), 122-129.
5. Kurniawan, W. A. (2020). Implementasi Sistem Penyiraman Tanaman Otomatis Dan Monitoring Berbasis Internet Of Things Menggunakan Fuzzy Logic Control (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
6. Nuryanto, H. (2007). Budi Daya Tanaman Krisan. *Ganeca Exact*.
7. Muliadi, M., Imran, A., & Rasul, M. (2020). Pengembangan tempat sampah pintar menggunakan ESP32. *Jurnal Media Elektrik*, 17(2), 73-79.



8. Hansen, D., & Hoendarto, G. (2017). PERANCANGAN PERANGKAT PENYIRAM TANAMAN OTOMATIS MENGGUNAKAN MIKROKONTROLER ARDUINO UN. INTEKSIS, 4(2).
9. Saputro, I. A., Suseno, J. E., & Widodo, C. E. (2017). Rancang bangun sistem pengaturan kelembaban tanah secara real time menggunakan mikrokontroler dan diakses di web. Youngster Physics Journal, 6(1), 40-47.
10. Asmara, R. K. P. (2020). Rancang Bangun Alat Monitoring Dan Penanganan Kualitas Ait Pada Akuarium Ikan Hias Berbasis Internet Of Things (IOT). Jurnal Teknik Elektro Dan Komputer TRIAC, 7(2), 69-74.
11. Sitio, S. L. M. (2018). Penerapan Fuzzy Inference System Sugeno untuk Menentukan Jumlah Pembelian Obat (Studi Kasus: Garuda Sentra Medika). Jurnal Informatika Universitas Pamulang, 3(2), 104.



# Analisis Pengaruh Data Saham Provinsi Terhadap Ekonomi Indonesia

Allan Ruhui Fatmah Sari<sup>1</sup>, Adhisa Shilfadianis Iffadah<sup>2</sup>, Edelin Fortuna<sup>3</sup>, Nayya Ramadhani Putri Widjanarko<sup>4</sup>, Yunita Nur<sup>5</sup>

<sup>1,2,3,4,5</sup>Fakultas Ilmu Komputer, Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[21083010007@student.upnjatim.ac.id](mailto:21083010007@student.upnjatim.ac.id), <sup>2</sup>[21083010016@student.upnjatim.ac.id](mailto:21083010016@student.upnjatim.ac.id), <sup>3</sup>[21083010087@student.upnjatim.ac.id](mailto:21083010087@student.upnjatim.ac.id),

<sup>4</sup>[21083010075@student.upnjatim.ac.id](mailto:21083010075@student.upnjatim.ac.id), <sup>5</sup>[21083010107@student.upnjatim.ac.id](mailto:21083010107@student.upnjatim.ac.id)

Corresponding author email: [21083010007@student.upnjatim.ac.id](mailto:21083010007@student.upnjatim.ac.id).

**Abstract:** *This study uses a quantitative method by conducting parametric tests. The purpose of this research is to analyze the influence of each province's stock data on the economy in Indonesia. The data in this study is in the form of monthly data from November 2022 - January 2023 and the research data source is <https://www.ojk.go.id>. The results showed that the variable frequency, value (billions) and individual investor share ownership had an effect of 99.8% in November 2022, 99.7% in December 2022 and 99.7% in January 2023, while the rest was influenced by other variables but not included in the estimation model.*

**Keywords:** *Economy, Indonesian provincial stocks, multiple linear regression*

**Abstrak:** Penelitian ini menggunakan metode kuantitatif dengan melakukan uji parametrik. Tujuan dilakukan penelitian ini yaitu untuk analisis pengaruh data saham tiap provinsi terhadap perekonomian di Indonesia. Data dalam penelitian ini berupa data bulanan dari bulan November 2022 - Januari 2023 dan sumber data penelitian dari <https://www.ojk.go.id>. Hasil penelitian menunjukkan bahwa variabel frekuensi, nilai (miliar) dan kepemilikan saham investor individu memberikan pengaruh sebesar 99,8% bulan November 2022, 99,7% bulan Desember 2022 dan 99,7% bulan Januari 2023, sedangkan sisanya dipengaruhi oleh variabel lain namun tidak dimasukkan dalam model estimasi.

Kata kunci: Ekonomi, Saham Provinsi Indonesia, Regresi Linear Berganda

## I. PENDAHULUAN

Perkembangan ekonomi negeri dapat dibuktikan berdasarkan pada pertumbuhan pasar modalnya. Pasar modal mempunyai peranan penting untuk perekonomian suatu negeri sebagai fasilitas pemerataan pemasukan. Pasar modal merupakan aktivitas yang berkaitan dengan penawaran umum dan perdagangan efek yang dibentuk serta lembaga dan profesi yang berkaitan dengan efek. Tujuan dari pasar modal negara Indonesia berfokus untuk memusatkan dana berasal dari warga supaya dapat menyalurkan ke dalam sektor yang lebih produktif. Manfaat dari pasar modal adalah selaku fasilitas pengalokasian dana untuk mengalihkan dana tersebut dari pemberi pinjaman terhadap peminjam (Bila Niawaradila, Gendro Wiyono, Alfiatul Maulida, 2021).

Pasar Modal menyediakan berbagai alternatif investasi lainnya, seperti menabung di bank, asuransi, saham, surat utang (obligasi), reksa dana, membeli emas, dan berbagai instrumen derivatif dari efek atau surat berharga. Pasar Modal juga bertindak sebagai penghubung antara investor dan perusahaan lainnya. Saham adalah sebuah obligasi yang mempunyai nilai tinggi dalam perdagangan. Saham merupakan surat berharga sebagai fakta bahwa sekuritas tersebut dimiliki dalam industri. Tujuannya investor melaksanakan investasi saham adalah menciptakan keuntungan (return) capital gain ataupun deviden yang maksimal. Dividen yang dibagi oleh industri penerbit saham ataupun selisih positif dari harga saham saat dibeli maupun dijual itu merupakan asal dari return.

Penanaman saham yang dilakukan oleh investor bukan untuk sekuritas tunggal saja, namun juga melaksanakan penganekekaragaman portofolio, sebab risiko portofolio bisa jadi lebih kecil dari resiko



rata-rata tertimbang tiap-tiap sekuritas tunggal. Secara universal resiko bisa jadi dikurangi dengan mencampurkan sebagian sekuritas tunggal dalam wujud portofolio agar bisa meminimalkan resiko tanpa wajib mengurangkan return saham yang diterima. Sehingga investor wajib mengenali diversifikasi risiko tersebut. Harga saham merupakan cerminan dari nilai industri sehingga bisa jadi bahan pertimbangan dasar untuk investor. Harga saham yang terus membesar menjadi penilaian besar pula terhadap industri, kebalikannya jika harga saham tersebut rendah maka penilaian industrinya juga rendah.

Pengertian saham secara sederhana merupakan suatu alat bukti atas pihak kepemilikan dari sebuah perusahaan yang di investasikan. Bentuk dari saham biasanya adalah lembaran kertas yang menyatakan kepemilikan surat berharga tersebut adalah pemilik dari perusahaan yang membuat surat. Secara umum, ada banyak metode dan teknik yang telah dikembangkan dalam penilaian perusahaan, di antaranya adalah dengan pendekatan penilaian laba, penilaian aktiva, pendekatan harga saham. Pendekatan lain yang banyak dilakukan oleh investor dalam melakukan analisis penilaian saham untuk menilai prospek perusahaan adalah analisis fundamental secara top-down.

Salah satu indikator dalam menentukan apakah seseorang akan melakukan investasi atau menabung adalah tingkat suku bunga. Suku bunga merupakan harga yang harus dibayarkan apabila terjadi pertukaran antara satu rupiah sekarang dengan satu rupiah. Tingkat bunga yang cenderung tinggi akan mengakibatkan para pemilik modal lebih memilih untuk menanamkan modalnya di bank. Jika suku bunga deposito terus meningkat, maka terjadi kecenderungan para pemilik modal untuk mengalihkan dananya ke deposito daripada menanamkan modalnya di pasar modal dengan memperkirakan tingkat keuntungan dan faktor risiko yang rendah. Hal ini berdampak negatif terhadap harga saham dimana harga saham di pasar modal akan mengalami kemerosotan secara signifikan. Dengan alasan tingkat keuntungan yang diharapkan saham lebih kecil dibandingkan dengan keuntungan dari tingkat suku bunga. Sehingga terjadinya penurunan permintaan terhadap harga saham dan harga saham akan terus mengalami penurunan seiring dengan kenaikan suku bunga yang terjadi. Bank Indonesia memiliki kebijakan menentukan suku bunga yang dikenal dengan BI Rate. Tingkat suku bunga sektor keuangan atau dikenal dengan tingkat suku bunga bebas resiko (risk free) sering digunakan sebagai panduan investor yang meliputi tingkat suku bunga bank sentral dan tingkat suku bunga deposito. Tidak stabilnya suku bunga menimbulkan para investor lebih tertarik untuk berinvestasi di bank yang memiliki risiko kecil dibandingkan berinvestasi di saham. Penurunan ketertarikan investor berinvestasi dalam bentuk saham akan mengakibatkan volume transaksi saham di Bursa Efek Indonesia berkurang dan menimbulkan nilai saham LQ45 melemah. Dengan demikian suku bunga dan keuntungan ialah variabel yang sangat berpengaruh terhadap keputusan para investor, di mana mempengaruhi terhadap keinginan investor dengan suku bunga yang rendah untuk melakukan investasi portofolio di pasar modal.

Pada tahun 2021 PT Bursa Efek Indonesia (BEI) menyatakan kontribusi pasar modal Indonesia terhadap pertumbuhan ekonomi nasional masih punya ruang besar untuk bertumbuh, mengingat usaha pencapaian pasar modal terhadap perekonomian nasional masih sekitar 48-50 persen. Direktur Utama BEI, Inarno Djajadi menuturkan, kapitalisasi pasar saham Indonesia terhadap Produk Domestik Bruto (PDB) masih rendah dibandingkan dengan negara di ASEAN. Seperti kontribusi pasar modal di Singapura terhadap perekonomian nasional telah mencapai di atas 100 persen. Melansir data pada Badan Pusat Statistik (BPS), perekonomian Indonesia pada tahun 2020 diukur berdasarkan PDB atas dasar harga berlaku mencapai Rp 13.432,2 triliun dan PDB per kapita memperoleh jumlah Rp 56,9 juta atau USD 3.911,7. Sementara itu, kapitalisasi pasar saham BEI tercatat sebesar Rp 7.329 triliun pada 11 Februari 2021 (Agustina Melani, 2021)

Pada tahun 2019 terdapat peneliti tentang Analisis Faktor-Faktor Yang Mempengaruhi Harga Saham Industri Makanan Dan Minuman Di Indonesia. Penelitian ini dilakukan dua orang mahasiswa Universitas Muhammadiyah Sumatera Utara yakni Murviana Koto dan Rashid Ridho. Peneliti



bertujuan untuk melihat bagaimana pengaruh dari beberapa faktor yang diduga memberikan pengaruh kuat pada harga saham perusahaan makanan dan minuman di Indonesia. Faktor-faktor yang akan diteliti pada penelitian ini adalah total asset untuk melihat ukuran perusahaan, laba perusahaan (net income), Earning per Share (EPS) yang mewakili rasio pasar dan suku bunga yang mewakili kondisi makro ekonomi. Peneliti menggunakan model regresi linier untuk metode kuadrat terkecil biasa OLS (ordinary least square method). Analisis data dilakukan menggunakan software Eviews 8. Berdasarkan hasil penelitian yang sudah diteliti, variabel total aset, laba, suku bunga, dan earning per share secara bersamaan memberikan pengaruh sebesar 71,5% kepada harga saham, sisanya sebesar 28,6 %, dijelaskan oleh variabel lain yang tidak dimasukkan ke dalam model estimasi. Hasil pengujian yang dilakukan secara parsial memperlihatkan bahwa hanya EPS dan suku bunga yang berpengaruh signifikan, walaupun keduanya berada pada arah yang berlawanan. Faktor EPS memberikan pengaruh positif sedangkan suku Bunga memberikan pengaruh negatif terhadap harga saham. Sedangkan, total asset dan laba memberikan pengaruh yang tidak signifikan terhadap harga saham (Murviana Koto, Rashid Ridho, 2019).

Penelitian tentang pengaruh suku bunga dan inflasi terhadap harga saham pada perusahaan perbankan yang terdaftar di LQ45 bursa efek Indonesia pernah dilakukan oleh Yuni Rachmawati, Universitas Tridinanti Palembang tahun 2019. Penelitian dilakukan untuk mengetahui pengaruh inflasi terhadap harga saham, pengaruh suku bunga terhadap harga saham, pengaruh suku bunga dan inflasi terhadap harga saham di perusahaan perbankan yang terdaftar di LQ45 BEI pada tahun 2015-2017. Hasilnya adalah Inflasi berpengaruh negatif dan signifikan terhadap harga saham perusahaan dengan taraf signifikan 5%. Hasil hipotesis nilai signifikansi (0,011) lebih kecil dari taraf signifikansi (0,05). Kemudian, suku bunga berpengaruh negatif dan signifikan terhadap harga saham yang mana nilai signifikansi (0,000) lebih kecil dari taraf signifikansi (0,05). Inflasi dan suku bunga secara simultan berpengaruh dan signifikan terhadap harga saham ditunjukkan dengan nilai signifikansi (0,000) lebih kecil dari taraf signifikansi (Yuni Rachmawati, 2019). Dengan adanya penelitian ini, diharapkan dapat mengetahui pengaruh frekuensi transaksi saham dan kepemilikan saham investor individu terhadap nilai total transaksi saham Indonesia dari masing-masing provinsi. Sehingga dapat menginterpretasikan faktor mana yang harus dioptimalkan untuk meningkatkan pertumbuhan ekonomi Indonesia melalui sektor saham.

## II. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan uji parametrik yang bertujuan untuk menganalisis pengaruh data saham provinsi terhadap perekonomian di Indonesia. Data untuk penelitian ini diperoleh dari website Otoritas Jasa Keuangan (<https://www.ojk.go.id>), mengambil data statistik bulanan pasar modal di bulan November, Desember 2022 dan Januari 2023. Sampel data pada penelitian ini adalah 34 provinsi di Indonesia yang telah terdaftar dan *termapping* oleh Kustodian Sentral Efek Indonesia (KSEI).

### 1.1. Variabel Penelitian dan Definisi Operasional

Variabel dependen pada penelitian ini adalah data nilai (miliar), yaitu total transaksi saham dari 34 provinsi di Indonesia. Variabel Independennya adalah data frekuensi transaksi saham (X1) dan kepemilikan saham investor individu (X2).

### 1.2. Metode Analisis

Metode analisis yang digunakan pada penelitian ini adalah metode analisis regresi linier berganda. Metode ini merupakan metode statistik yang melakukan prediksi seberapa besar pengaruh antara variabel dependen dengan dua variabel independen. Alasan mengapa peneliti menggunakan metode ini adalah karena secara konsep penelitian ini bertujuan untuk



memprediksi apakah transaksi saham dapat mempengaruhi perekonomian di Indonesia atau tidak dan data yang digunakan memiliki lebih dari satu variabel independen.

### 1.2.1. Scatter plot

Scatter plot adalah tipe grafik yang digunakan untuk memvisualisasikan suatu data ditandai dengan titik-titik untuk mewakili setiap variabelnya. Metode ini adalah salah satu cara yang baik untuk mengetahui nonlinieritas, outlier, dan pergerakan dalam hubungan biaya. Penelitian ini menggunakan scatter plot untuk melihat persebaran data frekuensi dengan data nilai (miliar) dan data kepemilikan saham investor individu dengan data nilai (miliar).

### 1.2.2. Korelasi pearson

Metode korelasi pearson adalah jenis korelasi untuk mengukur hubungan antara dua variabel yang berjenis data kontinu. Berikut adalah rumus dari metode korelasi Pearson:  
Metode ini digunakan peneliti untuk mengukur kuat atau lemah hubungan secara linier dari dua variabel yaitu variabel frekuensi (X1) dengan variabel nilai (Y) dan variabel kepemilikan (X2) dengan variabel nilai (Y). Jika hasil perhitungan dari dua variabel mendekati angka 1 atau -1, maka korelasi bisa dikatakan kuat, sebaliknya semakin mendekati 0 maka korelasi dari kedua variabel dikatakan lemah.

### 2.2.3. Model Regresi Linear Berganda

Regresi Linier Berganda merupakan model perhitungan regresi yang memprediksi pengaruh antara variabel dependen dengan dua variabel independen. Bentuk umum dari model ini adalah  $Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$ , dimana  $b_1, b_2, \dots, b_n$  adalah koefisien parameter dari variabel dependen,  $x_1, x_2, \dots, x_n$  adalah variabel independen, Y adalah variabel dependen, e (epsilon) adalah error random. penelitian ini jika dimasukkan kedalam model dapat digambarkan menjadi Y adalah variabel nilai,  $x_1$  adalah variabel frekuensi,  $x_2$  variabel kepemilikan saham investor individu.

### 2.2.4. Uji F

Uji F pada metode analisis regresi linier berganda digunakan untuk mengetahui apakah secara bersama-sama variabel independen berpengaruh terhadap variabel dependen. Metode ini bekerja dengan membuat hipotesis nol dan hipotesis alternatifnya, setelah membuat hipotesis tentukan tingkat signifikansi. Secara umum tingkat signifikansi alfa sebesar 5% atau 0,05. Kesimpulan umum dari hasil hipotesis adalah jika  $F_{hitung} > F_{tabel}$  maka  $H_0$  ditolak dan  $H_1$  diterima, begitu juga sebaliknya.

### 2.2.5. R-Squared

R-Squared merupakan metode yang memperlihatkan seberapa besar kesesuaian atau pengaruh dari variabel independen terhadap variabel dependen. berikut adalah rumus dari R-Squared:

$$r^2 = \frac{SSR}{SST}$$

nilai R-Squared akan terus bertambah mengikuti banyaknya dari variabel independen. R-squared memiliki nilai yang berkisar antara 0-1 dengan penjelasan semakin mendekati angka 1 maka hasil R-Squared dapat dikatakan baik, begitu juga sebaliknya.

### 2.2.6. Uji T

Uji T adalah metode yang digunakan untuk mengetahui apakah masing-masing variabel independen berpengaruh terhadap variabel dependen. Pengujian ini membuat hipotesis dan hipotesis alternatifnya. Umumnya  $H_0$  akan diterima jika tidak terdapat pengaruh signifikansi antara variabel independen dengan variabel dependen secara simultan, begitu juga sebaliknya jika terdapat pengaruh yang



signifikan dari variabel independen terhadap variabel dependen berarti  $H_0$  ditolak dan  $H_1$  diterima. Peneliti menggunakan pengujian ini untuk melihat apakah ada pengaruh yang signifikan dari variabel independen frekuensi dan kepemilikan saham investor individu terhadap variabel dependen.

### 2.2.7. Uji Asumsi Regresi

Beberapa asumsi yang harus dipenuhi dalam analisis regresi linier berganda, antara lain :

#### a. Asumsi residual berdistribusi normal menggunakan Jarque-Bera

Uji Asumsi Jarque-Bera adalah salah satu uji normalitas goodness of fit. Uji normalitas dilakukan untuk mengetahui apakah variabel dependen dan independen dalam model regresi linear berdistribusi normal atau tidak, tetapi ada pengecualian terhadap data saat menguji distribusi normal, menurut Damodar N Gujarati (2006:148) teorema limit sentral jika jumlah sampel besar ( $n > 30$ ), maka distribusi sampel akan mendekati dianggap normal. Penelitian ini memiliki sampe 34  $> 30$  maka asumsi normalitas dapat diabaikan karena  $> 30$ .

#### b. Asumsi residual tidak terjadi autokorelasi menggunakan Durbin Watson

Uji Durbin Watson adalah uji autokorelasi yang menilai apakah ada autokorelasi pada residual. Penyebab adanya autokorelasi adalah adanya kesalahan mengabaikan variabel yang penting atau adanya bentuk fungsi yang tidak tepat. Uji autokorelasi ini memiliki syarat dan ketentuan, uji dapat dikatakan tidak terjadi autokorelasi jika  $du < d < 4-du$ .

#### c. Asumsi residual tidak terjadi heteroskedastisitas menggunakan Breusch Pagan

Heteroskedastisitas adalah keadaan terjadinya ketidaksamaan varian dari error untuk semua uji setiap variabel independen pada model linier regresi. Uji Breusch Pagan adalah salah satu metode uji untuk mendeteksi adanya heteroskedastisitas dalam suatu model. Suatu data dapat dikatakan tidak heteroskedastisitas apabila tidak terdapat pola tertentu dan tidak menyebar di atas maupun di bawah angka nol pada sumbu y. Apabila uji ini tidak terpenuhi, maka model linier regresi dapat dinyatakan tidak valid dalam prediksinya.

#### d. Memeriksa apakah variabel independen memenuhi asumsi tidak terjadi multikolinieritas menggunakan nilai VIF.

Multikolinieritas merupakan uji yang melihat apakah semua variabel memiliki korelasi yang sempurna atau tidak. Peneliti memeriksa multikolinieritas dengan melihat nilai VIF. Nilai VIF adalah suatu faktor yang mengukur seberapa banyak varian dari koefisien regresi untuk setiap variabel yang dipengaruhi oleh interaksi dengan variabel lain dalam model. Nilai VIF yang baik adalah dibawah 10, semakin tinggi nilai VIF maka terdapat multikolinieritas yang signifikan.

## III. HASIL DAN PEMBAHASAN

### 3.1 Statistika Deskriptif

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa karakteristik dalam statistika deskriptif ini yaitu jumlah sampel ( $N$ ), rata-rata sampel (*mean*), nilai maksimum, nilai minimum dan standar deviasi untuk masing-masing variabel pada masing-masing bulan.

**Table 1. Statistika Deskriptif Data Saham**

Novemb er 2022	Jumla h sampe l (N)	Rata-rata sampel (Mean)	Nilai Maksimum	Nilai Minimum	Standar Deviasi
Frekuensi (X1)	34	1.352897e+ 06	1.853987e+ 07	9.956000e+ 03	3.392305e+0 6



	Kepemilikan saham investor individu (X2)	34	23133.558824	591702	28	101198.437941
	Nilai (Miliar) (Y)	34	10936.617647	231886	29	39800.339051
<b>Desember 2022</b>	Frekuensi (X1)	34	1.092984e+06	1.443402e+07	9.407000e+03	2.677039e+06
	Kepemilikan saham investor individu (X2)	34	21527.176471	540430	26	92445.372121
	Nilai (Miliar) (Y)	34	10797.558824	246091	20	42186.541254
<b>Januari 2023</b>	Frekuensi (X1)	34	1.187139e+06	1.573276e+07	1.014600e+04	2.909264e+06
	Kepemilikan saham investor individu (X2)	34	21318	534115	26	91362.311929
	Nilai (Miliar) (Y)	34	8190.382353	171450	24	29474.304341

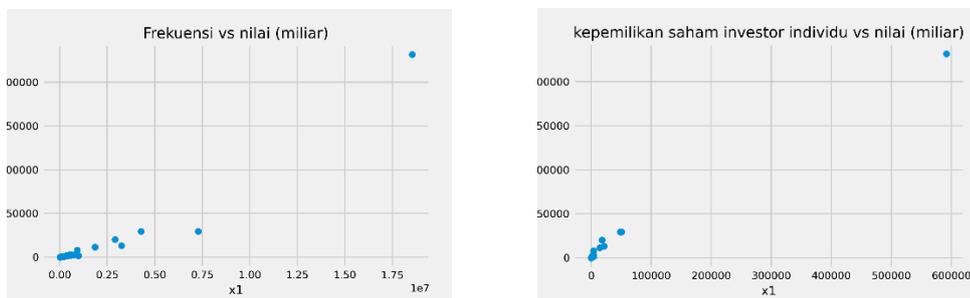
Pada tabel 1 diatas menunjukkan bahwa jumlah data yang digunakan pada penelitian ini sebanyak 34 sampel data yang mana mewakili provinsi di Indonesia. Berdasarkan hasil perhitungan didapatkan pada setiap bulannya frekuensi memiliki nilai rata-rata  $1.352897e+06$  pada bulan November 2022,  $1.092984e+06$  pada bulan Desember 2022, dan  $1.187139e+06$  pada bulan Januari 2023. Nilai minimum frekuensi tertinggi terdapat pada bulan Januari 2023 dengan nilai  $1.014600e+04$  dan nilai maksimum tertinggi frekuensi terdapat pada bulan November 2022 dengan nilai  $1.853987e+07$ . Sedangkan standar deviasi frekuensi tertinggi terdapat pada bulan November 2022 dengan nilai  $3.392305e+06$ . Pada variabel kepemilikan saham investor individu nilai rata-ratanya yaitu 23133.558824 pada bulan November 2022, 21527.176471 pada bulan Desember 2022, dan 21318 pada bulan Januari 2023. Untuk nilai minimum dan maksimumnya sendiri, variabel ini memiliki nilai tertinggi pada bulan November 2022 dengan nilai 28 dan 591702. Sedangkan standar deviasi kepemilikan saham investor individu tertinggi terdapat pada bulan November 2022 dengan nilai 101198.437941. Variabel nilai (miliar) sendiri memiliki rata-rata yaitu 10936.617647 pada bulan November 2022, 10797.558824 pada bulan Desember 2022, dan 8190.382353 pada bulan Januari 2023. Nilai minimum variabel nilai (miliar) tertinggi terdapat pada bulan November 2022 dan nilai maksimumnya terdapat pada bulan Desember 2022 dengan nilai 246091. Sedangkan untuk standard deviasinya sendiri tertinggi terdapat pada bulan Desember 2022 dengan nilai 42186.541254.

Sehingga didapatkan informasi bahwa rata-rata, nilai maksimum, nilai minimum dan standar deviasi tertinggi terdapat pada variabel frekuensi. Dimana pada semua variabelnya mengalami penurunan pada bulan Desember 2022 dan meningkat kembali pada Januari 2023. Serta mayoritas memiliki nilai tertinggi pada bulan November 2022. Meskipun ada beberapa nilai minimum dan maksimum tertinggi terdapat pada bulan Desember 2022 dan Januari 2023.

### 3.2 Penilaian Goodness of Fit Suatu Model

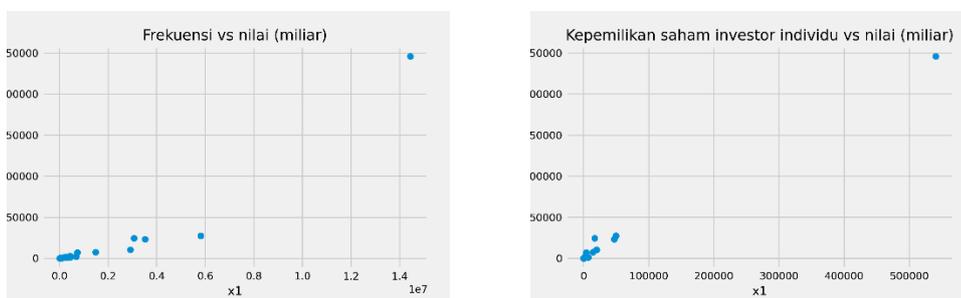
#### 3.2.1 Scatter Plot

##### 1. Bulan November 2022



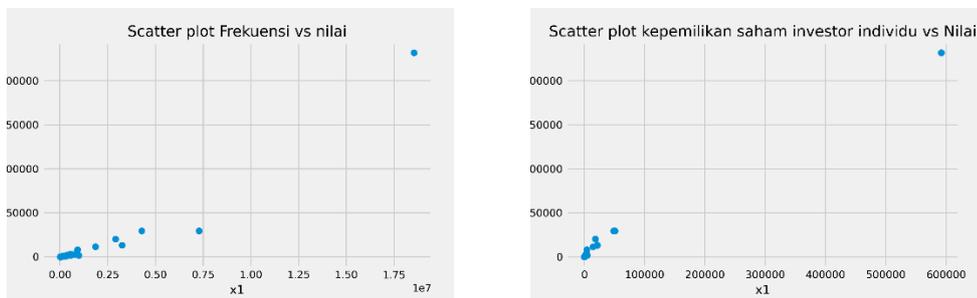
**Gambar 1.** Scatter Plot Variabel X1 X2 Terhadap Y Bulan November 2022

##### 2. Bulan Desember 2022



**Gambar 2.** Scatter Plot Variabel X1 X2 Terhadap Y Bulan Desember 2022

##### 3. Bulan Januari 2023



**Gambar 3.** Scatter Plot Variabel X1 X2 Terhadap Y Bulan Januari 2023

Pada ketiga gambar di atas, menunjukkan bahwa selama tiga bulan pada gambar sebelah kiri titik-titik data membentuk garis linier sehingga dapat disimpulkan bahwa frekuensi memiliki hubungan yang linier positif dengan nilai transaksi saham. Pada gambar sebelah kanan menunjukkan bahwa titik-titik data membentuk garis linier sehingga dapat disimpulkan bahwa kepemilikan saham investor secara individu memiliki hubungan yang linier positif dengan nilai transaksi saham. Dari seluruh gambar tersebut diduga bahwa tanda koefisien parameter akan bertanda positif.

#### 3.2.2 Uji Korelasi Pearson

**Table 2. Hasil Uji Korelasi Pearson Data Saham**

Bulan	Frekuensi	Kepemilikan Saham Investor Individu
November 2022	0.96058	0.99646
Desember 2022	0.94482	0.99740
Januari 2023	0.94643	0.99744



Pada tabel diatas dapat dilihat bahwa korelasi pearson variabel frekuensi memiliki nilai 0.96058 pada bulan November 2022, 0.94482 pada bulan Desember 2022, dan 0.94643 pada bulan Januari 2023 sehingga nilai tertinggi korelasi yang didapatkan variabel frekuensi ini yaitu terdapat pada bulan November 2022. Sedangkan pada variabel kepemilikan saham investor individu memiliki nilai 0.99646 pada bulan November 2022, 0.99740 pada bulan Desember 2022, dan 0.99744 pada bulan Januari 2023 sehingga nilai tertinggi korelasi pearson yang didapatkan oleh variabel ini yaitu terdapat pada bulan Januari 2023.

Secara umum, kedua variabel tersebut memiliki nilai korelasi pearson yang tinggi yang mana berarti hubungan antara variabel dependen dan independen memiliki hubungan yang sangat erat.

### 3.2.3 Hasil Analisis Berganda

Dari data yang diperoleh, selanjutnya dianalisis dengan metode regresi linear berganda menggunakan python. Berdasarkan output dari dua variabel independen frekuensi transaksi saham dan kepemilikan saham investor individu terhadap nilai transaksi saham ditunjukkan pada tabel berikut:

**Table 3. Model Regresi Linear Berganda Bulan November 2022**

<b>November 2022</b>	<b>Coefficient</b>	<b>Std Error</b>	<b>t</b>	<b>P&gt; t </b>
<b>Const</b>	378.9214	388.324	0.976	0.337
<b>X1</b>	0.0024	0.000	8.153	0.000
<b>X2</b>	0.3165	0.010	32.191	0.000

Dari hasil tabel analisis regresi bulan November 2022 di atas, memperoleh persamaan regresi linier berganda sebagai berikut:

$$\text{Nilai Transaksi Saham} = 378,9214 + 0,0024 \text{ Frekuensi Transaksi Saham} + 0,3165 \text{ Kepemilikan Saham Investor Individu}$$

Berdasarkan model di atas dapat disimpulkan bahwa pada model regresi variabel frekuensi transaksi saham dan kepemilikan saham investor individu berimplikasi positif terhadap nilai transaksi saham. Sehingga kenaikan pada kedua variabel tersebut akan meningkatkan nilai transaksi saham yang mana dapat memberikan dampak baik terhadap pertumbuhan ekonomi Indonesia.

**Table 4. Model Regresi Linear Berganda Bulan Desember 2022**

<b>Desember 2022</b>	<b>Coefficient</b>	<b>Std Error</b>	<b>t</b>	<b>P&gt; t </b>
<b>Const</b>	15.7244	490.521	0.032	0.975
<b>X1</b>	0.0019	0.000	4.358	0.000
<b>X2</b>	0.4034	0.013	31.638	0.000

Dari hasil tabel analisis regresi bulan Desember 2022 di atas, memperoleh persamaan regresi linier berganda sebagai berikut:

$$\text{Nilai Transaksi Saham} = 15,7244 + 0,0019 \text{ Frekuensi Transaksi Saham} + 0,4034 \text{ Kepemilikan Saham Investor Individu}$$

Berdasarkan model di atas dapat disimpulkan bahwa pada model regresi variabel frekuensi transaksi saham dan kepemilikan saham investor individu berimplikasi positif terhadap nilai transaksi saham. Sehingga kenaikan pada kedua variabel tersebut akan meningkatkan nilai transaksi saham yang mana dapat memberikan dampak baik terhadap pertumbuhan ekonomi Indonesia.

**Table 5. Model Regresi Linear Berganda Bulan Januari 2023**

Januari 2023	Coefficient	Std Error	t	P> t
<b>Const</b>	214.8711	346.685	0.620	0.540
<b>X1</b>	0.0020	0.000	7.047	0.000
<b>X2</b>	0.2608	0.009	28.370	0.000

Dari hasil tabel analisis regresi bulan Januari 2023 di atas, memperoleh persamaan regresi linier berganda sebagai berikut:

$$\text{Nilai Transaksi Saham} = 214,8711 + 0,0020 \text{ Frekuensi Transaksi Saham} + 0,2608 \text{ Kepemilikan Saham Investor Individu}$$

Berdasarkan model di atas dapat disimpulkan bahwa pada model regresi variabel frekuensi transaksi saham dan kepemilikan saham investor individu berimplikasi positif terhadap nilai transaksi saham. Sehingga kenaikan pada kedua variabel tersebut akan meningkatkan nilai transaksi saham yang mana dapat memberikan dampak baik terhadap pertumbuhan ekonomi Indonesia.

### 3.2.4 Uji F

Uji statistik F ini pada dasarnya digunakan untuk menunjukkan apakah semua variabel independen yang dimasukkan dalam model mempunyai pengaruh secara bersama-sama terhadap variabel dependennya. Hasil perhitungan uji F dapat dilihat pada tabel berikut :

**Table 6. Hasil Analisis Uji F**

Bulan	F	Probabilitas
<b>November 2022</b>	6890.	8.77e-42
<b>Desember 2022</b>	4793.	2.40e-39
<b>Januari 2023</b>	4692.	3.33e-39

Berdasarkan tabel diatas didapatkan nilai uji F dan probabilitasnya pada bulan November 2022 adalah 6890 dan 8.77e-42. Pada bulan Desember 2022 nilai uji F dan probabilitasnya adalah 4793 dan 2.40e-39 dan pada bulan Januari 2023 nilai uji F dan probabilitasnya adalah 4692 dan 3.33e-39. Sehingga, bulan November 2022 memiliki nilai uji F tertinggi dan probabilitas terendah.

Secara keseluruhan didapatkan kedua variabel independen Frekuensi dan Kepemilikan saham investor individu yang dimasukkan dalam model adalah dengan tingkat signifikansi dibawah 0,05 yaitu sebesar 8.77e-42, 2.40e-39, dan 3.33e-39. Hal ini dapat disimpulkan bahwa variabel Frekuensi dan Kepemilikan saham investor individu mempengaruhi variabel Nilai (miliar) secara signifikan pada setiap bulannya.

### 3.2.5 Uji Parsial (t test)

Uji parsial digunakan untuk mengetahui pengaruh masing-masing variabel independen Frekuensi dan Kepemilikan saham investor individu terhadap variabel dependen Nilai (miliar).

**Table 7. Hasil Analisis Uji Parsial t-test**

Bulan	Variabel	t	p >  t
<b>November 2022</b>	Const	0.976	0.337
	X1	8.153	0.000
	X2	32.191	0.000
<b>Desember 2022</b>	Const	0.032	0.975
	X1	4.358	0.000
	X2	31.638	0.000
<b>Januari 2023</b>	Const	0.620	0.540
	X1	7.047	0.000

X2	28.370	0.000
----	--------	-------

Berdasarkan tabel diatas didapatkan pada bulan November 2022 nilai T-stats untuk variabel frekuensi adalah 8.153 atau p-value 0.000 dan untuk variabel kepemilikan saham investor individu didapatkan nilai T-stats sebesar 32.191 atau p-value 0.000. Dengan menggunakan alfa 0,05 didapatkan t tabel sebesar 2.0395 yang berarti bahwa kedua variabel independen dalam bulan November 2022 berpengaruh secara signifikan terhadap nilai total transaksi saham. Pada bulan Desember 2022 nilai T-stats untuk variabel frekuensi adalah 4.358 atau p-value 0.000 dan untuk variabel kepemilikan saham investor individu didapatkan nilai T-stats sebesar 31.638 atau p-value 0.000. Dengan menggunakan alfa 0,05 didapatkan t tabel sebesar 2.0395 yang berarti bahwa kedua variabel independen dalam bulan Desember 2022 berpengaruh secara signifikan terhadap nilai total transaksi saham. Sedangkan pada bulan Januari 2023 nilai T-stats untuk variabel frekuensi adalah 7.047 atau p-value 0.000 dan untuk variabel kepemilikan saham investor individu didapatkan nilai T-stats sebesar 28.370 atau p-value 0.000. Dengan menggunakan alfa 0,05 didapatkan t tabel sebesar 2.0395 yang berarti bahwa kedua variabel independen dalam bulan Januari 2023 berpengaruh secara signifikan terhadap nilai total transaksi saham

Sehingga dapat disimpulkan bahwa pada ketiga bulan yang diteliti dari kedua variabel independen Frekuensi dan Kepemilikan saham investor individu yang dimasukkan dalam model adalah dengan tingkat signifikansi dibawah 0,05 sebesar 0.000. Hal ini dapat disimpulkan bahwa variabel Frekuensi dan Kepemilikan saham investor individu mempengaruhi variabel Nilai (miliar) secara signifikan.

### 3.2.6 Uji Koefisien Determinasi ( $R^2$ )

Koefisien determinasi digunakan untuk menguji goodness-fit dari model regresi. Hasil perhitungan koefisien determinasi dapat dilihat pada tabel berikut:

**Table 8. Hasil Analisis Uji Koefisien Determinasi ( $R^2$ )**

Bulan	R-squared	Adj. R-squared
November 2022	0.998	0.998
Desember 2022	0.997	0.997
Januari 2023	0.997	0.996

Berdasarkan tabel diatas, didapatkan pada bulan November 2022 nilai R-Square Adjusted sebesar 0.998 yang artinya 99,8% nilai total transaksi saham dipengaruhi oleh 8 variabel yang diteliti serta 0,1% lainnya dipengaruhi oleh variabel lainnya yang tidak diketahui atau tidak diteliti. Pada bulan Desember 2022 nilai R-Square Adjusted sebesar 0.997 yang artinya 99,7% nilai total transaksi saham dipengaruhi oleh dua variabel yang diteliti serta 0,1% lainnya dipengaruhi oleh variabel lainnya yang tidak diketahui atau tidak diteliti. Dan pada bulan Januari 2023 nilai R-Square Adjusted sebesar 0.996 yang artinya 99,6% nilai total transaksi saham dipengaruhi oleh dua variabel yang diteliti serta 0,1% lainnya dipengaruhi oleh variabel lainnya yang tidak diketahui atau tidak diteliti. Sehingga dapat disimpulkan bahwa pada ketiga bulan yang diteliti semuanya hanya dipengaruhi oleh variabel lainnya hanya sebesar 0.1%.

## 3.3 Pengujian Asumsi Klasik

### 3.3.1 Uji Normalitas

Uji normalitas digunakan untuk menguji apakah dalam model regresi, data residual memiliki distribusi normal. Pengujian normalitas data secara analisis statistic yang digunakan dalam penelitian ini yaitu dengan menggunakan pengujian Jarque-Bera. Secara multivarians pengujian normalitas data dilakukan terhadap nilai residual. Data yang berdistribusi normal ditunjukkan

dengan nilai signifikansi diatas 0,05 (Ghozali,2013). Hasil pengujian normalitas pada pengujian terhadap 34 data dalam ketiga bulan yang diteliti terlihat pada tabel berikut:

**Table 9. Hasil Uji Normalitas**

Pengujian	November 2022	Desember 2022	Januari 2023
Jarque-Bera	65.61218218103708	354.85833470107326	30.233811007914234
Chi <sup>2</sup> twotail prob	5.655821136340989e- 15	8.77994288937541e- 78	2.7215189093445674e- 07
Skew	1.8418513676478863	3.305807666913395	1.0656489129556157
Kurtosis	8.722308598592882	17.379665466800247	7.098668470358563

Berdasarkan tabel diatas didapatkan bahwa pada bulan November 2022 JBnya sebesar 65.612, bulan Desember 2022 sebesar 354.858, dan pada bulan Januari 2023 sebesar 30.234. Sehingga dengan menggunakan tabel chisquare sebagai pembandingnya yang didapatkan dengan nilai sebesar 47.40 maka pada ketiga bulan tersebut hanya pada bulan Januari 2023 saja H0 diterima yang berarti pada data bulan November dan Desember 2022 data tidak berdistribusi normal.

Akan tetapi pada penelitian ini uji normalitasnya akan menggunakan asumsi *Central Limit Theorem*. Menurut Damodar N Gujarati (2006:148) teorema ini dapat digunakan jika jumlah sampel besar ( $N > 30$ ), maka distribusi sampel akan dianggap mendekati normal. Jadi, dapat disimpulkan bahwa meskipun hasil dari pengujian normalitas menunjukkan sebagian data berdistribusi tidak normal, namun dikarenakan sampel dalam penelitian ini sebanyak 34 lebih dari 30 sesuai dengan *Central Limit Theorem* maka data dianggap berdistribusi normal.

### 3.3.2 Uji Multikolinieritas

Uji multikolinieritas digunakan untuk menguji apakah model regresi yang ditemukan ada korelasi antar variable bebas (independen). Jika variabel independen saling berkorelasi, maka variabel ini tidak orthogonal. Untuk mengetahui apakah terjadi mmultikolinieritas dapat dilihat dari nilai *Variance Inflation Factor* (VIF) yang terdapat pada masing-masing variabel dalam ketiga bulan yang diteliti seperti terlihat pada tabel berikut:

**Table 10. Hasil Uji Multikolinearitas**

Bulan	Feature	VIF
November 2022	X1	7.816266
	X2	7.816266
Desember 2022	X1	6.934689
	X2	6.934689
Januari 2023	X1	7.816266
	X2	7.816266

Berdasarkan tabel diatas, dapat diketahui bahwa nilai VIF X1 dan X2 pada bulan November 2022 sebesar 7.816, pada bulan Desember 2022 sebesar 6.934, dan pada bulan Januari 2023 sebesar 7.816. Untuk suatu model regresi dinyatakan bebas dari multikolinieritas adalah jika mempunyai nilai VIF lebih kecil dari 10. Jika dilihat dari perhitungan pada tabel maka diperoleh bahwa semua variabel bebas memiliki nilai VIF dibawah 10. Sehingga dapat disimpulkan bahwa antar dua variabel tidak saling memengaruhi sehingga tidak terjadi multikolinieritas .

### 3.3.3 Uji Autokorelasi

Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode-1 (sebelumnya). Jika terjadi korelasi, maka dinamakan ada masalah autokorelasi. Untuk melihat ada tidaknya autokorelasi

adalah dengan menggunakan uji Durbin-Watson (Ghozali, 2013). Berikut merupakan hasil perhitungan uji autokorelasi pada tabel berikut ini:

**Table 11. Hasil Uji Autokorelasi**

Bulan	Durbin-Watson
November 2022	1.6250268787132272
Desember 2022	1.8487326586197759
Januari 2023	2.005986694570869

Dari hasil pengolahan data di atas terlihat bahwa nilai Durbin-Watson pada bulan November 2022 adalah 1.625, bulan Desember 2022 sebesar 1.849, dan pada bulan Januari 2023 adalah sebesar 2.006, dengan  $n = 34$  dan  $k = 2$ . Nilai Durbin-watson berdasarkan tabel dengan derajat kebebasan sebesar 5% adalah  $dL = 0,94973$ ;  $dU = 1,2617$ ;  $4-dL = 3,05$ ;  $4-dU = 2,7383$ . Dengan nilai Durbin Watson pada bulan November 2022 adalah 1.625, sehingga berada diantara  $dU < d < 4-dU$  yang berarti tidak ada autokorelasi. Durbin-watson pada bulan Desember 2022 adalah 1.849 sehingga berada diantara  $du < d < 4-dU$  yang berarti tidak terjadi autokorelasi. Dan Durbin Watson pada bulan Januari 2023 adalah 2.006 sehingga berada diantara  $dU < d < 4-dU$  sehingga tidak terjadi autokorelasi. Sehingga dapat disimpulkan bahwa dari ketiga bulan yang diteliti tidak ada yang terjadi autokorelasi.

### 3.3.4 Uji Heteroskedastisitas

Uji heteroskedastisitas bertujuan untuk terdapat perbedaan residual antara satu pengamatan ke pengamatan lainnya. Untuk melihat ada tidaknya heteroskedastisitas adalah dengan menggunakan uji Breush-Pagan. Berikut merupakan hasil perhitungan uji heteroskedastisitas pada tabel berikut ini:

**Table 12. Hasil Uji Heteroskedastisitas**

Bulan	Breush-Pagan	P-Value	Hasil
November 2022	12.128	0.002	Gagal Tolak – Tidak terjadi
Desember 2022	8.307	0.015	Gagal Tolak – Tidak terjadi
Januari 2023	14.969	0.000	Gagal Tolak – Tidak terjadi

Dikarenakan Statistik Uji Breush-Pagan memiliki nilai seperti tabel di atas dimana lebih kecil dari  $\chi_{0.05,33} = 47.404$  dapat disimpulkan bahwa  $H_0$  gagal tolak yang artinya bahwa asumsi residual tidak terjadi heteroskedastisitas.

## IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut ini :

1. Hasil penelitian dapat diketahui bahwa semua variabel mengalami penurunan pada bulan Desember 2022 namun terjadi peningkatan pada bulan Januari 2023. Walaupun terjadi penurunan dan kenaikan di tiap bulannya, kepemilikan saham investor secara individu memiliki pengaruh linear positif dengan nilai transaksi saham.
2. Hasil penelitian menunjukkan bahwa variabel frekuensi transaksi saham dan kepemilikan saham investor individu memberikan pengaruh sebesar 99,8% bulan November 2022, 99,7% bulan Desember 2022 dan 99,7% bulan Januari 2023, sedangkan sisanya dipengaruhi oleh variabel lain namun tidak dimasukan dalam model estimasi.

Pengaruh frekuensi transaksi saham dan kepemilikan saham investor individu terhadap jumlah nilai transaksi saham yang tinggi dapat membantu meningkatkan pertumbuhan ekonomi Indonesia secara signifikan dan optimal.



## REFERENSI

1. D. S. Sampurna, "ANALISIS PENGARUH FAKTOR-FAKTOR EKONOMI MAKRO TERHADAP IHSG DI BURSA EFEK INDONESIA (BEI)," *Stei Ekonomi*, Vols. 25, No. 1, pp. 54-73, Juni, 2016.
2. E. D. ., P. K. Kurnia Rahmadani, "PENGUNGKAPAN LINGKUNGAN PERUSAHAAN: DITINJAU DARI MANAJEMEN LABA DAN MEKANISME TATA KELOLA PERUSAHAAN YANG BAIK," *Jurnal Kajian Akuntansi dan Bisnis Terkini*, Vols. vol. 2, no. 1, pp. 94-107, Maret 2021.
3. Y. Rachmawati, "Pengaruh Inflasi dan Suku Bunga Terhadap Harga Saham Pada Perusahaan Perbankan Yang Terdaftar Di LQ45 Bursa Efek Indonesia," *Jurnal Media Akuntansi*, Vols. Vol. 1, No. 1, pp. 66-79, September 2018.
4. S. A. M. M. C. Dr. Meiryani, "MEMAHAMI UJI HETEROSKEDASTISITAS DALAM MODEL REGRESI," 06 Agustus 2021.
5. A. Hidayat, "Regresi Linear Berganda: Penjelasan, Contoh, Tutorial," *statistikian.com*, Januari 2018. [Online]. Available: <https://www.statistikian.com/2018/01/penjelasan-tutorial-regresi-linear-berganda.html>. [Accessed 28 April 2023].
6. A. Melani, "IHSG Bakal Sentuh 7.948 pada 2023, Ini Faktor Pendorongnya," *liputan6.com*, 13 Maret 2023. [Online]. Available: <https://www.liputan6.com/saham/read/5231216/ihsg-bakal-sentuh-7948-pada-2023-ini-faktor-pendorongnya>. [Accessed 28 April 2023].
7. A. Melani, "Peluang Besar Kontribusi Pasar Modal terhadap Ekonomi RI," *liputan6.com*, 12 Februari 2021. [Online]. Available: <https://www.liputan6.com/saham/read/4481350/peluang-besar-kontribusi-pasar-modal-terhadap-ekonomi-ri>. [Accessed 28 April 2023].
8. R. R. Murviana Koto, "Analisis Faktor-Faktor Yang Mempengaruhi Harga Saham Industri Makanan Dan Minuman Di Indonesia," *Prosiding FRIMA*, no. No. 2 , 30-03-2019.



# Implementasi Metode *Clustering* SOM Dalam menganalisis Alasan Tidak KB di Kabupaten Temanggung

Mufida Fauziah Faiz<sup>1</sup>, Achmad Fauzan<sup>2</sup>,

Program Studi Statistika, FMIPA UII

Corresponding author email: [achmadfauzan@uii.ac.id](mailto:achmadfauzan@uii.ac.id)

**Abstract:** Birth is the ability to create an actual population or produce. The total fertility rate is the average number of children a woman has during her childbearing years, between the ages of 15 and 49. This total birth rate is an important indicator and a strategic indicator to determine how much a country or all countries manage their population through family planning (KB) programs. . Many people do not use family planning programs for certain reasons. Identification of reasons for not using family planning in regions or sub-districts is needed as an effort to control population growth. So, it is important that clustering is carried out in sub-districts whose residents do not use family planning for various reasons. The Kohonen self-organizing map (SOM) is used as a grouping method. SOM is a high-dimensional data analysis method, does not require assumptions and can produce information presentation. Using data on 14 variables, reasons for couples of childbearing age (PUS) not using family planning for certain reasons in Temanggung Regency are derived from data from the Office of Population Control, Family Planning, Women's Empowerment and Child Protection (DPPKBPPPA). From the data obtained 2 clusters which have their own characteristics. The clusters formed are cluster 1 of 2 districts, and cluster 2 of 18 districts.

**Keywords:** Family Planning, Self Organizing Maps (SOM), Clusters.

**Abstrak:** Kelahiran adalah kemampuan untuk menciptakan populasi atau berproduksi yang sebenarnya. Tingkat kesuburan total adalah rata-rata jumlah anak yang dimiliki wanita di masa suburnya yaitu antara usia 15 dan 49. Angka kelahiran total ini merupakan indikator penting dan indikator strategis untuk menentukan seberapa besar suatu negara atau seluruh negara mengelola penduduknya melalui program Keluarga Berencana (KB). Banyak masyarakat yang tidak menggunakan program KB dikarenakan alasan tertentu. Identifikasi alasan tidak menggunakan KB di daerah atau kecamatan sangat diperlukan sebagai upaya untuk mengendalikan pertumbuhan penduduk. Jadi, penting apabila dilakukan *clustering* pada kecamatan yang penduduknya tidak menggunakan KB karena berbagai alasan. *Self-organizing map* (SOM) Kohonen digunakan sebagai metode pengelompokan. SOM adalah suatu metode analisis data berdimensi tinggi, tidak memerlukan asumsi dan dapat menghasilkan penyajian informasi. Menggunakan data 14 variabel alasan Pasangan Usia Subur (PUS) tidak menggunakan KB dikarenakan alasan tertentu di Kabupaten Temanggung yang berasal dari data Dinas Pengendalian Penduduk, Keluarga Berencana, Pemberdayaan Perempuan dan Perlindungan Anak (DPPKBPPPA). Dari data didapatkan 2 *cluster* yang memiliki ciri – ciri tersendiri. *Cluster* terbentuk yaitu *cluster* 1 dari 2 kecamatan, dan *cluster* 2 dari 18 kecamatan.

**Kata kunci:** Keluarga Berencana, *Self Organizing Maps* (SOM), *Cluster*.

## I. PENDAHULUAN

Kelahiran adalah kemampuan untuk menciptakan populasi atau berproduksi yang sebenarnya. Tingkat kesuburan total adalah rata-rata jumlah anak yang dimiliki wanita selama masa suburnya yaitu antara usia 15 dan 49. Angka kelahiran total ini merupakan indikator penting dan indikator strategis untuk menentukan seberapa besar suatu negara atau seluruh negara mengelola penduduknya melalui program Keluarga Berencana (KB).

Hasil survei Badan Pusat Statistik (BPS) tahun 2020, angka kelahiran total yaitu sejumlah 2.10. Jadi, untuk rata-rata dari perempuan di Indonesia akan melahirkan dua anak pada masa reproduksinya. Pandemi Covid-19 yang sudah berlangsung dari Maret tahun 2020 berdampak pada angka kelahiran total. Prediksi dari sejumlah lembaga akan terjadi lonjakan angka kelahiran total selama pandemi covid-19. Ini dikarenakan akibat penurunan penggunaan alat kontrasepsi dan keterbatasan layanan kesehatan yang berlangsung selama masa *pandemic* [1].



Keluarga berencana adalah program nasional yang ditujukan untuk mengurangi angka kelahiran dan membatasi pertumbuhan penduduk di negara tersebut. Misalnya di Amerika Serikat ada program keluarga berencana yaitu *Planned Parenthood*. Program KB untuk menciptakan kesejahteraan ekonomi, stabilitas, kemajuan, sosial dan spiritual bagi penduduk. Program yang tercantum dalam UU No. 10/1992 salah satunya yaitu KB, diatur dan dikendalikan oleh Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN). Salah satu bentuk KB yaitu penggunaan alat kontrasepsi untuk menunda dan mencegah kehamilan. Alat kontrasepsi umum yang digunakan antara lain kondom, alat kontrasepsi suntik, *implant*, Pil, KB, IUD, dan alat kontrasepsi permanen [2].

Dari 14 variabel yang digunakan penulis membuat deskriptif data yang digunakan untuk melihat gambaran umum dari dampak atau akibat yang ada apabila pasangan usia subur apabila tidak menggunakan program KB. Dengan diketahui gambaran umum dari dampaknya tersebut selanjutnya melakukan pengelompokan wilayah di kecamatan Temanggung dengan menggunakan profilisasi di setiap *cluster*. Banyak data dan banyak variabel akan membutuhkan analisis dalam membentuk profilisasi akibat atau dampak Pasangan Usia Subur tidak menggunakan KB ini.

Penggunaan metode SOM dikarenakan keunggulan sebagai alat visualisasi dan analisis data yang berdimensi tinggi. Namun, jaringan ini juga dapat digunakan untuk pengelompokan, pengurangan dimensi, klasifikasi, kuantisasi vektor, dan penambahan data.

Berdasarkan kajian Bunga Rahayu [3], dari data 13 variabel dampak longsor di 34 provinsi di Indonesia pada 2008 hingga 2018 dari data publikasi DIBI BNPB. Karakteristik ditemukan di setiap *cluster*, yaitu *Cluster 1* menunjukkan pengaruh kerusakan jalan yang paling kuat, *Cluster 2* tidak berpengaruh signifikan semua karakteristik, *Cluster 3* yaitu pengaruh sangat signifikan pada semua karakteristik, meliputi, korban luka-luka, tertimpa dan terluka, Rumah Rusak Berat, Puskesmas, Tempat Ibadah Rumah Rusak Ringan, Lembaga Pendidikan, jumlah kejadian, korban meninggal hilang dan jembatan Sawah Rusak. Namun, *cluster 4* memiliki dampak terbesar pada rumah rusak sedang. Sedangkan pada kasus ini digunakan 14 variabel dari data pasangan usia subur yang tidak menggunakan KB karena alasan tertentu dan didapatkan 2 *cluster* dimana *cluster* pertama beranggotakan 2 kecamatan, sedangkan *cluster* kedua beranggotakan 18 kecamatan.

Menurut Kohonen dan Fausett, strategi pembelajaran yang digunakan oleh SOM adalah unsupervised learning, yang menyiratkan topologi terorganisir ke dalam unit kelas atau *cluster*. Akibatnya, pendekatan SOM akan digunakan dalam investigasi *clustering* ini. Oleh karena itu pada disini akan menggunakan analisis *clustering* dengan metode SOM.

## II. METODE PENELITIAN

### 2.1. Data dan Sumber Data

Data yang digunakan dalam laporan kerja praktik adalah data dari Pasangan Usia Subur (PUS) yang tidak menggunakan KB, data ini merupakan data sekunder yang didapat dari Dinas Pengendalian Penduduk, Keluarga Berencana, Pemberdayaan Perempuan dan Perlindungan Anak (DPPKBPPPA).

### 2.2. Definisi Variabel

No	Variabel	Keterangan	Satuan
1	Kecamatan	Variabel kecamatan yaitu Kecamatan yang ada di Temanggung. Di Temanggung terdapat 20 Kecamatan yaitu (Kecamatan Bejen, Bejen, Bansari, Candioto, Gemawang, Jumo, Kaloran, Kandangan, Kedu, Kledung, Kranggan, Ngadirej	Jumlah Kecamatan



		o,Parakan,Pringsurat,Selopampang,Temanggung, Tembarak, Tlogomulyo, Tretep Dan Wonoboyo).	
2	Ingin Hamil	Variabel Ingin hamil merupakan keterangan bahwa pasangan usia subur itu Ingin hamil jadi tidak menggunakan KB.	Orang
3	Kesehatan	Variabel Kesehatan merupakan keterangan bahwa pasangan usia subur itu mengalami gangguan Kesehatan apabila menggunakan KB.	Orang
4	Efek Samping	Variabel Efek Samping merupakan keterangan bahwa pasangan usia subur itu mengalami efek samping apabila menggunakan KB, dan efek samping ini yang berbeda beda dari tiap pasangan usia subur.	Orang
5	Tempat Pelayanan Jauh	Variabel Tempat Pelayanan Jauh merupakan keterangan bahwa pasangan usia subur itu ada keinginan menggunakan KB tetapi pelayanan yang jauh, ini biasanya untuk daerah yang pelosok.	Orang
6	Mahal	Variabel Mahal merupakan keterangan bahwa pasangan usia subur itu ada keinginan menggunakan KB tapi terhalang oleh harga KB yang mahal.	Orang
7	Tidak Cocok	Variabel Tidak cocok merupakan keterangan bahwa pasangan usia subur tidak cocok menggunakan KB biasanya terjadi perubahan susasana hati dan kecemasan.	Orang
8	Menolak	Variabel menolak merupakan keterangan bahwa pasangan usia subur itu menolak menggunakan KB.	Orang
9	Agama	Variabel Agama merupakan	Orang

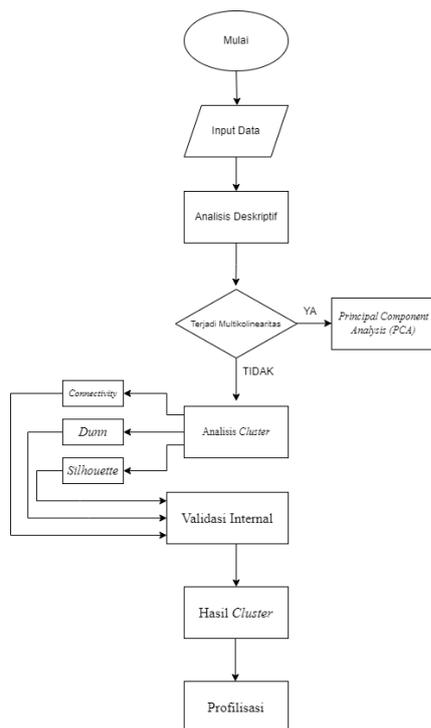


		keterangan bahwa pasangan usia subur itu menolak karena dalam agama tidak diperbolehkan menggunakan KB.	
10	Pasangan Jauh	Variabel Pasangan jauh merupakan keterangan bahwa pasangan usia subur itu lagi berhubungan jarak jauh karena bekerja jadi tidak menggunakan KB.	Orang
11	Menopause	Variabel Menopause merupakan keterangan bahwa Wanita atau Ibu rumah tangga tersebut sudah berusia 45 sampai 55 jadi sudah tidak bisa mempunyai anak lagi dan tidak perlu menggunakan KB.	Orang
12	Tidak tersedia	Variabel tidak tersedia merupakan keterangan bahwa jenis KB yang diinginkan tidak tersedia di tempat tersebut.	Orang
13	Tidak ada petugas	Variabel tidak ada petugas merupakan keterangan bahwa tidak ada petugas yang membantu pelayanan KB.	Orang

### 2.3. Metode Penelitian

Pada penelitian menggunakan metode *Clustering Self Organizing Map* (SOM) untuk melakukan pengelompokan Pasangan Usia Subur yang tidak menggunakan KB di Kabupaten Temanggung. Metode metode *Clustering Self Organizing Map* (SOM) ini dapat dilakukan dengan menggunakan *software R Studio*.

Untuk mendapatkan hasil yang diinginkan dari penelitian ini, maka terdapat langkah-langkah yang harus dilakukan. Langkah analisis menggunakan metode *Clustering Self Organizing Map* (SOM) dilihat melalui diagram berikut ini.



**Gambar 1.** Diagram Alir

**Gambar 1.** Merupakan diagram alir dan dijabarkan Langkah – Langkah dengan lebih rinci seperti berikut ini.

1. Mulai
2. Tahapan penginputan data “Pasangan Usia Subur yang tidak menggunakan KB”
3. Setelah data diinputkan, kemudian melakukan analisis deskriptif dari data tersebut yang secara umum yaitu menunjukkan kateristik dari variabel yang ada didalam data tersebut.
4. Kemudian melakukan pengecekan Multikolinearitas, karena dalam data tersebut terjadi multikolinearitas dan dilakukan *Principal Component Anaysis* (PCA).
5. Setelah itu melakukan Analisis *Cluster* dengan melakukan validasi *cluster* yaitu menentukan *cluster* terbaik dengan menggunakan *Connectivity*, *Dunn*, dan *Silhouette*. [4]

Validasi internal meliputi: konektivitas, nilai *Silhouette* dan indeks *Dunn*.

- *Connectivity*

Konektivitas menyatakan tingkat hubungan *cluster*, yang ditentukan oleh jumlah tetangga terdekat.

$$\text{Conn}(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)} \quad (1)$$

dengan,

$nn_{i(j)}$  = pengamatan tetangga terdekat I ke j dan L

$nn_{i(j)}$  = sebagai parameter yang menentukan jumlah tetangga yang berpartisipasi dalam pengukuran

- Nilai *Silhouette*

Nilai *Silhouette* adalah rata-rata dari nilai *Silhouette* untuk setiap titik data. Skor *Silhouette* mengukur tingkat kepercayaan dalam proses pengelompokan setiap rangkaian pengamatan, membentuk *cluster* baik ketika nilai indeks mendekati 1 maupun sebaliknya ketika indeks mendekati -1.

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

dengan,

$a_{(i)}$  = jarak dari rata-rata i dan seluruh pengamatan yang lain

$b_{(i)}$  = jarak dari rata-rata  $i$  dengan pengamatan pada *cluster* yang paling dekat

- Indeks *Dunn*

Indeks *Dunn* yaitu perbandingan jarak terpendek dari data observasi *cluster* yang berbeda dengan jarak terjauh dalam *cluster*.

Indeks *Dunn* didapatkan dari hasil bagi  $d_{min}$  dengan  $d_{max}$

$$Dunn = \frac{d_{min}}{d_{max}} \quad (3)$$

dengan,

$d_{min}$  = jarak yang paling kecil antara pengamatan pada *cluster* yang berbeda - beda

$d_{max}$  = jarak yang paling besar pada setiap *cluster* data

6. Selanjutnya membuat *plot* dari hasil *cluster* terbaik.
7. Langkah terakhir melakukan profilisasi dari *cluster* yang terbentuk dan menampilkannya dalam bentuk data *frame*.
8. Selesai.

### III. HASIL DAN PEMBAHASAN

#### 3.1. Analisis Deskriptif

Sebelum melakukan *clustering* dengan menggunakan metode SOM, melakukan analisis deskriptif untuk mengetahui gambaran umum dari data yang ingin digunakan.

**Tabel 1.** Statistika Deskriptif

Variabel	Minimum	Median	Mean	Maksimum
Ingin Hamil	476	869.5	1059.7	2352
Kesehatan	83	273.5	330.2	1061
Efek Samping	46	221.5	279.65	1028
Tempat Pelayanan Jauh	0	1	2.15	7
Mahal	0	5	4.15	8
Tidak Cocok	7	39.5	52.55	234
Menolak	9	47.5	62.5	189
Agama	11	32	41.25	119
Pasangan Jauh	50	98.5	125.7	356
Menopause	31	128.5	130.5	305
Tidak Tersedia	0	2	2.15	13
Tidak Ada Petugas	0	0	0.45	4



**Gambar 2.** Grafik Banyaknya PUS yang tidak menggunakan KB

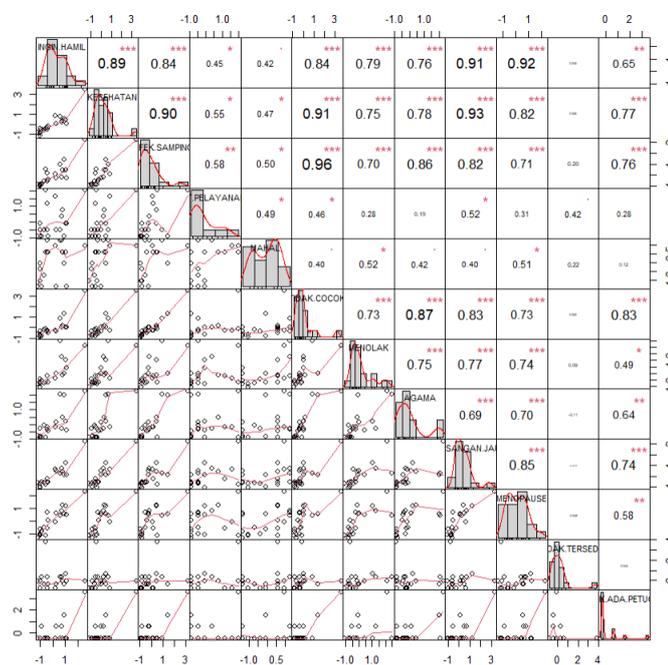
Berdasarkan **Gambar 2.** grafik tersebut menunjukkan data banyaknya pasangan usia subur (PUS) yang tidak menggunakan KB dengan alasan tertentu di Temanggung. Dimana terlihat bahwa

untuk kecamatan temanggung sendiri merupakan kecamatan dengan pasangan usia subur yang tidak menggunakan Program KB karena alasan tertentu paling banyak, contohnya karena alasan ingin hamil, kecamatan temanggung memiliki frekuensi paling banyak dikarenakan kecamatan temanggung merupakan kecamatan yang merupakan pusat kota sehingga mayoritas pasangan memilih untuk tinggal diperkotaan karena fasilitas umum yang memadai. Untuk kecamatan kledung merupakan alasan pasangan usia subur tidak menggunakan KB paling besar untuk kategori tempat pelayanan jauh, karena kecamatan kledung terletak di dataran tinggi jadi masyarakat apabila ingin menggunakan pelayanan KB tempatnya jauh dari permukiman penduduk.

### 3.2. Uji Asumsi

#### 4.2.1 Uji Multikolinearitas

Multikolinearitas adalah korelasi yang ada pada variabel bebas. Jika multikolinearitas terjadi, maka akan mempengaruhi hasil *cluster*. Terdapat koefisien korelasi yang melebihi 0.8 maka hal itu menandakan terjadinya multikolinearitas.



Gambar 3. Matriks Korelasi

Beberapa variabel memiliki multikolinearitas, namun terdapat juga beberapa variabel yang tidak mengandung multikolinearitas. Koefisien korelasi yang menyebabkan adanya multikolinearitas sehingga perlu dilakukan sebuah metode untuk menghilangkan multikolinearitas. Salah satunya dengan mereduksi variabel menggunakan PCA.

#### 4.2.2 Principal Component Analysis (PCA)

PCA atau analisis komponen utama merupakan salah satu metode untuk mengatasi gejala multikolinearitas pada analisis *cluster*. Untuk menentukan seberapa banyak komponen utama yang terbentuk, dilihat dari nilai *eigen value* yang lebih dari satu. Berikut adalah nilai *eigen value* yang dihasilkan

Tabel 3. Nilai Eigen

Komponen	Nilai Eigen
1	7.783
2	1.133

Besarnya nilai *eigen* pada komponen satu adalah 7.783 dan komponen dua adalah 1.133, dikarenakan dua komponen dengan nilai *eigen* lebih dari satu, maka akan terdapat dua faktor utama yang digunakan dalam analisis *cluster*.

**Tabel 4.** Hasil PCA

Kecamatan	Comp 1	Comp 2
Bansari	-1.55	0.14
Bejen	-1.44	-0.25
Bulu	0.21	0.15
Candiroto	-0.42	-0.09
Gemawang	-0.54	-0.08
Jumo	-0.91	-0.06
Kaloran	0.81	-0.46
Kandangan	0.69	0.39
Kedu	1.09	0.44
Kledung	-0.71	-0.06
Kranggan	0.85	0.34
Ngadirejo	1.04	0.28
Parakan	0.98	-0.11
Pringsurat	0.85	0.80
Selopampang	-1.39	-0.03
Temanggung	4.40	-0.63
Tembarak	-0.65	0.07
Tlogomulyo	-1.32	-1.34
Tretep	-1.03	-0.43
Wonobojo	-0.98	-0.23

Dimana nilai kombinasi linear untuk komponen utama pertama adalah sebagai berikut :

$$y_1 = -1.55 \text{ Bansari} - 1.44 \text{ Bejen} + 0.21 \text{ Bulu} - 0.42 \text{ Candiroto} - 0.54 \text{ Gemawang} - 0.91 \text{ Jumo} + 0.81 \text{ Kaloran} + 0.69 \text{ Kandangan} + 1.09 \text{ Kedu} - 0.71 \text{ Kledung} + 0.85 \text{ Kranggan} + 1.04 \text{ Ngadirejo} + 0.98 \text{ Parakan} + 0.85 \text{ Pringsurat} - 1.39 \text{ Selopampang} + 4.40 \text{ Temanggung} - 0.65 \text{ Tembarak} - 1.32 \text{ Tlogomulyo} - 1.03 \text{ Tretep} - 0.98 \text{ Wonobojo}.$$

Dan kemudian nilai kombinasi linear untuk komponen utama kedua adalah sebagai berikut :

$$y_2 = 0.14 \text{ Bansari} - 0.25 \text{ Bejen} + 0.15 \text{ Bulu} - 0.09 \text{ Candiroto} - 0.08 \text{ Gemawang} - 0.06 \text{ Jumo} - 0.46 \text{ Kaloran} + 0.39 \text{ Kandangan} + 0.44 \text{ Kedu} - 0.06 \text{ Kledung} + 0.34 \text{ Kranggan} + 0.28 \text{ Ngadirejo} - 0.11 \text{ Parakan} + 0.80 \text{ Pringsurat} - 0.03 \text{ Selopampang} - 0.63 \text{ Temanggung} - 0.07 \text{ Tembarak} - 1.34 \text{ Tlogomulyo} - 0.43 \text{ Tretep} - 0.23 \text{ Wonobojo}.$$

### 3.3. Penentuan Jumlah Cluster

Dalam menentukan jumlah *cluster* menggunakan (3) metode validasi internal, yaitu dengan Indeks *Dunn* tertinggi, *Silhouette* yang terlihat mendekati 1, dan *Connectivity* yang terendah untuk membentuk jumlah *cluster*.

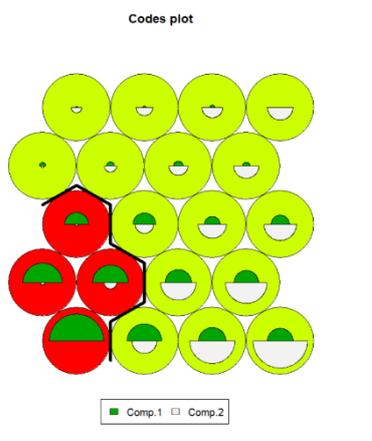
**Tabel 5.** Nilai *Score* validasi internal

Cluster	Connectivity	Dunn	Silhouette
2	3.6401	0.1610	0.6151
3	14.8409	0.0517	0.3891
4	17.4337	0.1563	0.4678
5	20.2556	0.1421	0.3580

**Tabel 5.** Merupakan validasi dari nilai indeks *Dunn* tertinggi yaitu 0,1610 pada *cluster* 2, nilai indeks konektivitas terendah yaitu 3,6401 pada *cluster* 2, dan nilai indeks *Silhouette* yang terlihat berdekatan 1 yaitu sebesar 0,6151 pada *cluster* 2. Dari hasil validasi *cluster* memberikan nilai *cluster* terbaik yaitu 2 (dua). Jadi, disini menggunakan jumlah *cluster* 2 dalam teknik *clustering* untuk mengelompokkan PUS tanpa KB dengan alasan yang pasti di kecamatan yang ada di Kabupaten

### 3.4. Pengelompokan (*Clustering*) dengan SOM

Proses algoritma SOM menggunakan program R untuk mendapatkan model SOM yang menghasilkan *plot* berikut dari data yang dilakukan oleh PCA.



**Gambar 4.** Diagram kipas

**Gambar 4.** Merupakan *plot hexagonal* dengan *grid* 4x5. Bagan kipas menunjukkan distribusi variabel pada peta. Pengertian pada algoritma SOM terjadi apabila diagram sudah memiliki warna dan terbatas pada vektor yang digambarkan pada *plot* pemetaan. 2 warna berbeda menunjukkan status dari masing-masing *cluster* yang dihasilkan. Tabel berikut memberikan informasi lebih lanjut mengenai *cluster* yang terbentuk dan karakteristik dari *cluster*.

Berikut hasil dari *clustering* kecamatan yang didapatkan dengan menggunakan program R.

**Tabel 6.** Anggota *cluster*

<i>Cluster</i>	Jumlah Anggota	Anggota <i>Cluster</i>
1	2	Kecamatan Kaloran dan Kecamatan Temanggung
2	18	Kecamatan (Bejen, Bansari, Bulu, Candiroto, Gemawang, Jumo, Kandangan, Kledung, Kedu, Kranggan, Ngadirejo, Parakan, Pringsurat, Selopampang, Tretep, Tlogomulyo, Tembarak Dan Wonoboyo)

**Tabel 6.** Merupakan informasi dari anggota *cluster*, anggota *cluster* pertama yaitu ada 2 kecamatan yaitu kecamatan temanggung dan kecamatan kaloran, sedangkan untuk *cluster* 2 yaitu 18 kecamatan yang lain Kecamatan (Bejen, Bansari, Bulu, Candiroto, Gemawang, Jumo, Kandangan, Kledung, Kedu, Kranggan, Ngadirejo, Parakan, Pringsurat, Selopampang, Tretep, Tlogomulyo, Tembarak Dan Wonoboyo).

Di dalam 16 kecamatan pada Kabupaten Temanggung, teridentifikasi dua *cluster* dengan metode *self-organizing map (SOM)*, yang mencerminkan alasan rata-rata pasangan usia subur yang berdekatan. Dari tabel di atas terlihat bahwa kabupaten relatif dekat berdasarkan 14 variabel yang digunakan.

### 3.5. Profilisasi *Cluster*

Karena jumlah anggota dan anggota dari setiap *cluster* sudah diketahui, maka dihitung rata-rata dari *cluster* untuk mengetahui ciri - ciri setiap *cluster* (profilisasi).

**Tabel 7.** Profilisasi Hasil Perhitungan Rata-Rata *Cluster*

Variabel	<i>Cluster 1</i>	<i>Cluster 2</i>
Ingin Hamil	1765,5	981,2222
Kesehatan	789	279,2222

Efek Samping	667,5	236,5556
Tempat Pelayanan Jauh	6,5	1,666667
Mahal	7	3,833333
Tidak Cocok	136,5	43,22222
Menolak	107	57,55556
Agama	77,5	37,22222
Pasangan Jauh	263,5	110,3333
Menopause	214	121,2222
Tidak Tersedia	1	2,277778
Tidak Ada Petugas	2	0,277778

**Tabel 7.** Merupakan hasil profilisasi dari analisis *Cluster SOM* yang dapat digunakan untuk mengetahui alasan Pasangan Usia Subur (PUS) tidak mengikuti program Keluarga Berencana (KB) yang dijalankan oleh pemerintah. Jika dibandingkan dari hasil perhitungan rata-rata pada kedua *cluster*, secara sekilas maka terlihat bahwa *cluster* pertama lebih tinggi atau dominan dibandingkan *cluster* kedua dikarenakan hasil perhitungan rata – ratanya lebih besar dibandingkan dengan *cluster* kedua.

Pada *cluster* 1 merupakan kelompok dari kecamatan temanggung dan kaloran yang merupakan kecamatan yang paling menonjol atau tertinggi yang tidak mengikuti program KB yang dijalankan oleh pemerintah. Sedangkan *cluster* kedua yaitu 18 kecamatan yang ada di Temanggung yang memiliki tingkat terendah dalam mengikuti program KB yang dijalankan oleh pemerintah jika dibandingkan dengan kelompok lainnya.

#### IV. Kesimpulan

Dari pembahasan yang sudah dijabarkan, disimpulkan sebagai berikut:

1. Dari data alasan Pasangan Usia Subur tidak menggunakan KB, dari hasil analisis deskriptif didapatkan hasil untuk variabel ingin hamil, Kesehatan, efek samping, tempat pelayanan jauh, tidak cocok, agama, pasangan jauh, menopause dan tidak ada petugas paling banyak berada pada kecamatan Temanggung karena kecamatan Temanggung terletak pada pusat Kabupaten Temanggung, jadi padat penduduknya. Untuk variabel tempat pelayanan jauh yaitu di kecamatan Kledung karena kecamatan kledung terletak di dataran tinggi jadi masyarakat apabila ingin menggunakan pelayanan KB tempatnya jauh dari permukiman penduduk. Untuk variabel mahal paling banyak yaitu di kecamatan gemawang, kandang dan parakan. Variabel menolak paling banyak yaitu di kecamatan parakan. Dan untuk variabel tidak tersedia paling banyak yaitu di kecamatan kledung.
2. Dari hasil analisis *cluster SOM* yang telah dilakukan, diperoleh pengelompokan sebanyak dua *cluster*. Dimana *cluster* pertama beranggotakan 2 kecamatan, sedangkan *cluster* kedua beranggotakan 18 kecamatan.
3. Dari hasil profilisasi dari analisis *Cluster SOM* yang dapat digunakan oleh penulis untuk mengetahui alasan Pasangan Usia Subur (PUS) tidak mengikuti program Keluarga Berencana (KB) yang dijalankan oleh pemerintah. Jika dibandingkan dari hasil perhitungan rata-rata pada kedua *cluster*, secara sekilas maka terlihat bahwa *cluster* pertama lebih tinggi atau dominan dibandingkan *cluster* kedua dikarenakan hasil perhitungan rata – ratanya lebih besar dibandingkan *cluster* kedua

#### UCAPAN TERIMA KASIH

#### REFERENSI

1. L. K. "Angka Kelahiran Indonesia Terus Turun," 13 February 2023. [Online]. Available: [https://data.kompas.id/data-detail/kompas\\_statistic/63fdc5cde0dac278b0fa595f](https://data.kompas.id/data-detail/kompas_statistic/63fdc5cde0dac278b0fa595f).



2. R. C. Swari, "Sebenarnya, Program Keluarga Berencana (KB) Itu Apa, Sih?," 05 January 2022. [Online]. Available: <https://hellosehat.com/seks/kontrasepsi/program-keluarga-berencana-kb/>.
3. B. Rahayu, "dspace.uin.ac.id," 29 August 2019. [Online]. Available: <https://dspace.uin.ac.id/bitstream/handle/123456789/16023/13611109.pdf?sequence=17&isAllowed=y>. [Accessed 9 May 2023].
4. A. Pratama, "rpubs.com," [Online]. Available: <https://rpubs.com/apriliyanuspratama/928038>. [Accessed 8 May 2023].
5. A. D. Andriani, "etheses.uin-malang.ac.id," 18 May 2018. [Online]. Available: <http://etheses.uin-malang.ac.id/12554/1/13650130.pdf>. [Accessed 2 May 2023].
6. D. F. Atika, "Analisis Klaster untuk," *Jurnal Ilmiah Populer*, vol. 01, p. Juny, 2019.
7. S. N. Kapita and . S. D. Abdullah, "PENGELOMPOKKAN DATA MUTU SEKOLAH MENGGUNAKAN JARINGAN," (*Jurnal Informatika dan Komputer*), vol. 3, pp. 56-61, 2020.
8. A. R. Rayhani, W. Astuti, Z. Shufila and E. Widodo, "Implementasi Self Organizing Map dalam Pengelompokan Kabupaten di Jawa Barat berdasarkan Kasus Covid-19," *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 2, pp. 174-183, 2021.
9. Q. LINA, "Self Organizing Maps (SOM) dengan Software R & Kohonen [Part 2]," medium.com, 17 Juny 2019. [Online]. Available: <https://medium.com/@16611110/self-organizing-maps-som-dengan-software-r-kohonen-part-2-ac338f0b9452>. [Accessed 2 May 2023].
10. M. H. "Apa Itu Principal Component Analysis (PCA)?," www.megabagus.id, 28 September 2019. [Online]. Available: <https://www.megabagus.id/apa-itu-principal-component-analysis-pca/>. [Accessed 11 May 2023].
11. M. Rais, R. Goejantoro and S. Prangga, "Optimalisasi K-Means Cluster dengan Principal Component Analysis," *Jurnal EKSPONENSIAL*, vol. 12, pp. 129 -136, 2021.
12. T. Shafira, "Menghilangkan Multikolinieritas dengan PCA (Principal Component Analysis)," tamarashafiraa.medium.com, 06 July 2020. [Online]. Available: <https://tamarashafiraa.medium.com/menghilangkan-multikolinieritas-dengan-pca-principal-component-analysis-2ea0939f8178>. [Accessed 09 May 2023].
13. A. REISLA RAYHANI, "Self Organizing Map (SOM) Menggunakan R Studio," medium.com, July 2023. [Online]. Available: <https://medium.com/@18611042/self-organizing-map-som-menggunakan-r-studio-3fb497462290>. [Accessed 09 May 2023].
14. Y. Anis and R. Isnanto, "Penerapan Metode Self-Organizing Map(SOM) Untuk Visualisasi Data Geospasial Pada Informasi Sebaran Data Pemilih Tetap (DPT)," *Jurnal Sistem Informasi Bisn*, vol. 01, p. 48, 21 March 2014.
15. A. F. Khairati, A. A. Adlina, G. F. Hertono and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161-170, 2019.
16. M. D. I. ZZATI, "Clustering Menggunakan Self Organizing Maps (SOM)," medium.com, 12 July 2020. [Online]. Available: <https://medium.com/@17611037/clustering->



- menggunakan-self-organizing-maps-som-695bedadc139. [Accessed 01 May 2023].
17. N. R. Simamora, "Analisis Komponen Utama untuk Menganalisis Indikator IPM di Jawa Timur," rpubs.com, 3 January 2023. [Online]. [Accessed 1 May 2023].
  18. F. Hestuningtias and A. H. Primandari, "Pengelompokan Kecamatan Di Kabupaten Merangin Berdasarkan Produksi Tanaman Perkebunan Tahun 2021 Menggunakan Agglomerative Hierarchical Clustering," *Emerging Statistics and Data Science Journal*, vol. 1, no. 1, pp. 137 - 147, 2023.
  19. D. Hedyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk," (*Journal Information Engineering and Educational Technology*), vol. 02, pp. 49 - 54, 2021.
  20. B. Informatika, N. Hasan and G. B. Sulisty, "Penerapan Metode Principle Component Analysis (PCA) untuk Clustering," *Bianglala Informatika*, vol. 8, pp. 58 - 62, 2020.
  21. O. Tazkiyah, T. and A. D, "Self Organizing Map (SOM)," 2016. [Online]. Available: <https://statslab-rshiny.fmipa.unej.ac.id/JORS/SOM/>.



# Implementasi Algoritma *Naïve Bayes Classifier* Dalam Klasifikasi *Review Program Career Preparation Eduwork*

Adinda Alifatul Jannah<sup>1</sup>, Achmad Fauzan<sup>2</sup>,

<sup>1, 2</sup> Prodi Statistika, FMIPA Universitas Islam Indonesia

Corresponding author email: [achmadfauzan@uii.ac.id](mailto:achmadfauzan@uii.ac.id)

**Abstract:** The development of digital technology in the industrial era 5.0 is directly proportional to the increasingly strict job selection qualifications. Degrees and education do not guarantee someone to get a job if they are not equipped with critical skills and global skills so that careful career preparation is needed. Eduwork as a provider of skills improvisation, mentoring, and job distribution courses launched the Career Preparation Program as a form of guidance service for improvising skills of job seekers. This study aims to identify feedback and ratings from users of the newly launched Career Preparation Program by Eduwork using *Naïve Bayes Classifier* algorithm implemented by sentiment analysis. Based on the research, the performance produced from the test results carried out on a ratio of 70:30 between train data compared to test data resulted in an accuracy rate of 92.5%, with a precision rate is 1 While in the Recall Class a value of 96.72% was produced (positive class: positive). The majority of Career Preparation Program users provide positive feedback and are satisfied with Eduwork services.

**Keywords:** *Naïve Bayes Classifier*, Sentiment analysis, Review, Career preparation program

**Abstrak:** Perkembangan teknologi digital di era industri 5.0 berbanding lurus dengan kualifikasi seleksi kerja yang semakin ketat. Gelar dan pendidikan tidak menjamin seseorang memperoleh pekerjaan jika tidak dibekali dengan *critical skill* dan *global skill* sehingga diperlukan persiapan karir yang matang. Eduwork sebagai penyedia kursus improvisasi *skill*, mentoring, dan penyaluran kerja meluncurkan *Career Preparation Program* sebagai bentuk layanan bimbingan untuk improvisasi *skill* para pencari pekerjaan. Penelitian ini bertujuan untuk mengidentifikasi *feedback* serta penilaian dari pengguna *Career Preparation Program* yang baru diluncurkan oleh Eduwork menggunakan analisis sentimen yang diimplementasikan dalam algoritma *Naïve Bayes Classifier*. Berdasarkan penelitian, diperoleh performa yang di hasilkan dari hasil pengujian yang dilakukan terhadap perbandingan 70:30 antar data *train* berbanding data *test* dihasilkan nilai akurasi sebesar 92.5%, dengan nilai presisi 1 Sementara pada *Class Recall* dihasilkan nilai 96,72% (*positive class: positive*). Para pengguna *Career Preparation Program* mayoritas memberikan *feedback* positif dan puas dengan pelayanan Eduwork.

**Kata kunci:** *Career preparation program*, *Naïve Bayes Classifier*, Analisis sentimen, Review

## I. PENDAHULUAN

Persaingan dunia kerja di era digitalisasi industri 5.0 berkembang semakin ketat. Hal tersebut sejalan dengan perkembangan teknologi digital yang menuntut *skill* dan profesionalisme pekerja. Banyak perusahaan besar yang telah menggunakan perkembangan teknologi kecerdasan buatan (*artificial intelligence*) dengan menggunakan robot atau mesin dalam proses industri. Hal tersebut membawa pengaruh besar dalam penurunan angkatan kerja. Menurut data BPS, di Bulan Agustus 2022 terdapat 8.425.931 orang atau 5.86% jumlah pengangguran terbuka di Indonesia. Proporsi pengangguran terbuka mayoritas alumni sekolah kejuruan sebanyak 9.42%, lulusan SMA 8.57%, lulusan SMP 5.95%, lulusan universitas dan diploma 9.39%, dan sisanya lulusan sekolah dasar [1]. Angka tersebut mengalami penurunan pada Februari 2023 menjadi 7.99 juta orang dengan proporsi tertinggi berasal dari Provinsi Banten [2].

Gelar dan pendidikan tidak menjamin seseorang memperoleh pekerjaan jika tidak dibekali dengan *critical skill* dan *global skill* (terutama kemampuan komunikasi dan bahasa asing). Angka pengangguran terbuka yang lebih besar dari lapangan kerja yang ada ternyata tidak optimal untuk menemukan sumber daya manusia yang sesuai kualifikasi dari perusahaan. Perusahaan akan memberi rangkaian seleksi yang banyak untuk menentukan kualifikasi sumber daya manusia yang dapat bersaing dan memiliki profesionalitas [3]. Berdasarkan survei *Talent Development* pada *Engineering Career Center* (ECC) mengenai permasalahan para pencari kerja saat *apply* pekerjaan di perusahaan,



diperoleh beberapa permasalahan, seperti minimnya kemampuan *public speaking*, kepercayaan diri saat wawancara, kebingungan mencari informasi di dunia kerja, dan kurangnya pengetahuan dalam pembuatan *Curriculum Vitae* [4]. Dari survei diatas diperoleh informasi bahwa kesulitan kerja tidak hanya dialami oleh pekerja berpengalaman yang telah diterima disuatu perusahaan, namun juga oleh para pencari kerja.

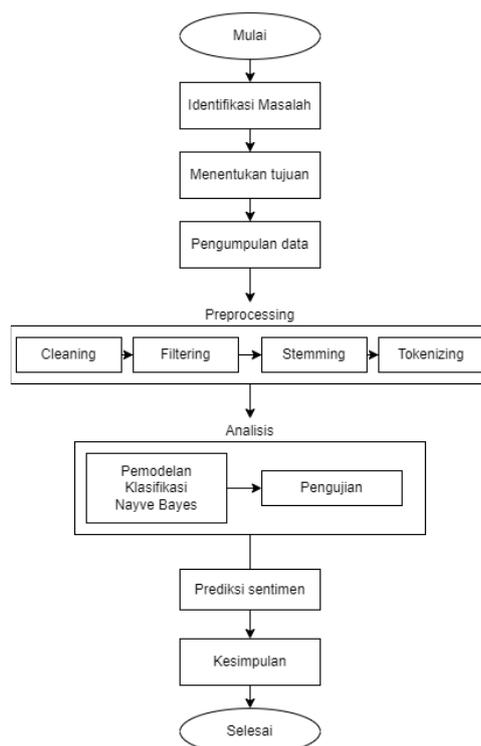
Diperlukan persiapan karir atau *career development* menuju rangkaian seleksi perusahaan dan pengembangan skill agar lebih unggul dari pencari kerja yang lain sangat dibutuhkan. Karena marak digunakan dalam meningkatkan kualitas SDM, banyak program-program *career development* yang diadakan oleh lembaga-lembaga kursus dan penyedia kegiatan mentoring, salah satunya adalah Eduwork. Eduwork merupakan *platform e-learning* yang membantu pengembangan kemampuan dan menyediakan layanan *mentoring* serta penyaluran kerja. Eduwork memiliki banyak program unggulan yang dapat dipilih dalam menelusuran minat, improvisasi kemampuan, bahkan *career development* yang baru-baru ini diluncurkan.

*Career Preparation Program* merupakan program berbentuk *career development* sebagai upaya mempersiapkan dan memberikan layanan bimbingan kepada pada pencari kerja untuk menjadi sumber daya yang unggul dari para pencari kerja lainnya. Program ini menjawab permasalahan para pencari kerja dalam mempelajari persiapan seleksi pekerjaan. Para pencari kerja akan dibimbing langsung dengan praktisi profesional secara privat. Program ini baru diperkenalkan ke masyarakat pada tahun 2023 sehingga dibutuhkan banyak masukan dan kecenderungan masyarakat dalam menerima adanya program tersebut. Dalam proses analisis data, digunakan *text mining* jenis analisis sentimen untuk mengidentifikasi *review* pengguna *Career Preparation Program*. Analisis sentimen mengelompokkan *review* dalam ungkapan positif, negatif, bahkan netral. Beberapa pendekatan sentimen yang marak diimplementasikan antara lain *Naïve Bayes Classifier*, SVM, K *Nearest Neighbor*, *Lexicon*, dan masih banyak pendekatan lainnya [5]. Pendekatan sentimen yang ada memiliki beberapa perbedaan pada proses pengklasifikasian, seperti *Naïve Bayes Classifier* yang melibatkan sedikit kuantitas data *train* dengan proses pengklasifikasian jauh lebih cepat dari yang lainnya dalam tiga kelompok, yaitu *positive*, *negative*, atau netral [6] Hal tersebut menjadi faktor penggunaan analisis sentimen pada penelitian ini dengan memperhatikan kuantitas data yang sedikit.

Berdasarkan penelitian Permadi [7], diperoleh informasi performa akurasi klasifikasi dan proporsi kategori hampir 90% pada *review* restoran di Singapura. Penelitian yang dilakukan Hayuningtyas dan Retnosari [8] menyatakan nilai akurasi *review* pendapat publik akan penggunaan bahasa yang baik dalam Taman Mini Indonesia Indah dengan analisis *Naïve Bayes* sebesar 70%. Dari beberapa hasil analisis terdahulu, dalam penelitian ini peneliti ingin meneliti *review* program untuk mengetahui umpan balik serta penilaian dari pengguna *Career Preparation Program* yang baru diluncurkan oleh Eduwork.

## II. METODE PENELITIAN

Data pada penelitian ini berbentuk primer berdasarkan wawancara pada 10 pengguna pertama *Career Preparation Program* Eduwork. Data dihimpun pada 10-16 Februari 2023. Data berbentuk teks dengan konsep *text mining* dan dianalisis menggunakan analisis sentimen dengan pendekatan *Naïve Bayes Classifier*. Hanya terdapat satu variabel *review* dalam proses analisis ini.

**Gambar 1.** Tahapan Penelitian

### 3.1 *Text Mining*

Penggalian teks merupakan proses dengan tujuan mengklasifikasi segala bentuk informasi yang berbentuk teks untuk menentukan pola kata yang bervariasi [9]. *Text Mining* biasanya berupa *preprocessing* data dalam penelitian untuk mentransformasikan data mentah menjadi data yang rapi dan tertata. *Preprocessing* data dalam *text mining* terdiri atas *cleaning*, *filtering*, *stemming*, dan *tokenizing*. Pada penelitian ini tidak dilakukan *case folding* karena seluruh data tidak menggunakan huruf kapital [10].

1. *Cleaning*: Membersihkan data dari karakter atau simbol-simbol yang berpengaruh pada proses analisis.
2. *Filtering*: Menghapus teks atau kata tidak penting berdasarkan *stopword*.
3. *Stemming*: Pengecekan kata untuk diselaraskan sehingga mengurangi daftar kata.
4. *Tokenizing*: Pemisahan teks menggunakan spasi sebagai entitas yang siap di analisis.

### 3.2 Analisis Sentimen

Analisis sentimen merupakan bentuk *text mining* untuk mengidentifikasi, ekstraksi, dan mengolah opini atau *review* untuk memperoleh klasifikasi dan umpan balik dari pandangan masyarakat terhadap sesuatu yang dapat dinilai. Sentimen mengacu pada permasalahan yang telah ditentukan yang biasanya didasarkan pada subjek penelitian [11].

### 3.3 Algoritma *Naïve Bayes Classifier*

Algoritma ini menghasilkan pengklasifikasian bersifat *supervised learning* yang menggunakan pendekatan peluang atau probabilitas beberapa tahun yang akan datang dari informasi masa lalu. Model yang digunakan bersumber pada teorema Bayes seperti persamaan (1).

$$P(A|X) = \frac{P(A) \cdot P(X|A)}{P(X)} \quad (1)$$

Persamaan (1) memuat nilai  $X$  sebagai atribut,  $A$  menyatakan kelas,  $P(A|X)$  sebagai probabilitas  $A$  jika  $X$  telah terjadi,  $P(X|A)$  merupakan probabilitas  $X$  jika  $A$  telah terjadi, probabilitas kejadian  $A$  dinyatakan dalam  $P(A)$ , dan  $P(X)$  adalah probabilitas kejadian  $X$  [12].

Algoritma ini juga dapat menentukan kategori setiap dokumen menggunakan rumus:

$$C^* = \arg \max_{C \in \mathcal{C}} p_{\mu} p(W_k | C_i) \quad (2)$$

Dengan nilai  $A^*$  dalam persamaan (2) merupakan kategori yang mempunyai peluang kemunculan dan seluruh nilai di sisi kiri adalah nilai data *training* yang dimaksimalkan [13].

### III. HASIL DAN PEMBAHASAN

Peneliti melihat terlebih dahulu data yang digunakan dalam penelitian dalam sebuah *wordcloud*. *Wordcloud* merupakan salah satu bentuk visualisasi yang digunakan untuk menggambarkan kata dalam sebuah matriks agar lebih mudah dipahami dan menarik [14]. *Wordcloud* data review produk *Career Preparation Program* Eduwork seperti pada gambar berikut.



Gambar 2 *Wordcloud* Data Penelitian

*Wordcloud* diatas menggambarkan beberapa kata yang paling banyak disebutkan dalam *review* adalah *mentor*, *interview*, *baik*, *seru*, *materi*, *simulasi*, dan *portofolio*. Kata ‘*mentor*’ menggambarkan bahwa kegiatan ini dilayani oleh mentor profesional dibidangnya. Adanya mentor menjadi poin utama dalam *Career Preparation Program* ini. Keberadaan mentor sangat penting dalam program ini. Kata dengan kuantitas terbanyak kedua ialah ‘*interview*’. Hal ini menjelaskan bahwa salah satu kendala yang dialami pengguna dan dapat diselesaikan oleh pembinaan dari *Career Preparatin Program* adalah *interview* atau wawancara kerja. Materi *interview* menjadi materi terpenting dalam program ini. Kata berikutnya adalah ‘*baik*’ dan ‘*seru*’ yang menandakan bahwa mentor yang membersamai pengguna bersifat baik dan menyampaikan materi pembelajaran dengan suasana kelas yang seru. Selanjutnya terdapat kata ‘*simulasi*’ yang menandakan keunggulan materi *interview* disertai dengan simulasi agar para pengguna lebih mendalami materi dan merasakan atmosfer wawancara yang sebenarnya. Kata terakhir yang memiliki kuantitas besar adalah kata ‘*portofolio*’ yang menandakan kesulitan tertinggi kedua yang dialami pengguna sehingga memutuskan mendaftarkan dirinya ke *Career Preparation Program* ini. Pembuatan portofolio sangat penting agar unggul dalam persaingan seleksi kerja. Pembuatan portofolio menjadi program unggulan lain setelah adanya simulasi *interview*. Terdapat banyak kata lain dengan kuantitas yang rendah. Kata-kata yang banyak disebutkan memiliki konotasi positif sehingga mayoritas pengguna memberikan *feedback* positif jika diperhatikan dari *wordcloud* yang ada.

#### 1. *Text Preprocessing*

Tahap ini mempersiapkan data dengan mentransformasikan teks yang tidak terstruktur agar terstruktur dan baku tanpa tambahan apapun. Tahap ini mempersiapkan proses berikutnya yaitu *text mining* dengan analisis sentimen untuk memperoleh informasi berdasarkan *feedback* yang ingin dicapai. *Text preprocessing* yang dilakukan adalah *cleaning*, *stemming*, *filtering*, dan *tokenizing*. Jika data yang dimiliki tidak dipersiapkan menggunakan *text preprocessing*, maka akan memperlambat analisis dan memberi informasi yang tidak optimal.

## 2. *Cleansing*

Proses *cleansing* dilakukan untuk menghilangkan karakter, tanda baca, ataupun simbol mengganggu (tanpa berdampak) pada proses analisis sentimen, salah satunya menghilangkan titik (.), tanda koma (,), atau karakter yang bukan berupa huruf dan angka. Proses ini mempersingkat waktu analisis dan mempermudah algoritma dalam klasifikasi [15].

**Tabel 1** Proses *Cleansing*

Sebelum	Sesudah
memberi byk informasi belajar, mulai dari pembuatan cv, wawancara, smp dgn seluruh rangkaian pencarian pekerjaan.	memberi byk informasi belajar mulai dari pembuatan cv wawancara smp dgn seluruh rangkaian pencarian pekerjaan

## 3. *Stemming*

Pada tahapan *stemming*, data diperbaiki agar tidak terdapat kesalahan pengedaan atau kata singkatan sebagai upaya penyeragaman kata dan pengecekan kata agar lebih ringkas berdasarkan kata baku yang mempunyai kesamaan arti namun berbeda bentuk penulisan, termasuk imbuhan yang dihilangkan [16].

**Tabel 2** Proses *Stemming*

Sebelum	Sesudah
memberi byk informasi belajar mulai dari pembuatan cv wawancara smp dgn seluruh rangkaian pencarian pekerjaan	beri banyak informasi belajar mulai dari buat cv wawancara sampai dengan seluruh rangkaian cari kerja

## 4. *Filtering*

Tahapan berikutnya adalah filterisasi sebagai upaya untuk menghilangkan teks atau kata yang tidak penting dalam proses sentimen. Dalam penelitian ini, *stopword* yang dijadikan acuan berasal dari kamus *stopwords*. Nantinya kata-kata yang telah diseragamkan akan diklasifikasikan untuk mengikuti proses analisis atau dihilangkan [17].

**Tabel 3** Proses *Filtering*

Sebelum	Sesudah
beri banyak informasi belajar mulai dari buat cv wawancara sampai dengan seluruh rangkaian cari kerja	beri informasi belajar buat cv wawancara rangkaian cari kerja

## 5. *Tokenizing*

Pada tahap *preprocessing* yang terakhir, dilakukan pemisahan kata menjadi beberapa potongan kata atau token. Karakter yang kurang dari dua dan bukan huruf atau angka akan dieliminasi. Pola ini digunakan untuk melihat pola *review* yang bernilai dalam *text matrix* yang akan diolah. Tahap ini telah disesuaikan dengan tahap-tahap sebelumnya dan hanya menyisakan kata-kata yang menunjang analisis [18].

**Tabel 4** Proses *Tokenizing*

Sebelum	Sesudah
beri informasi belajar buat cv wawancara rangkaian cari kerja	“beri” “informasi” “belajar” “buat” “cv” “wawancara” “rangkaian” “cari” “kerja”

## 6. *Skenario Pengujian*

Dengan 10 *review* masa percobaan peluncuran *career preparation program*, dilakukan skenario perbandinga data *training* dengan data *testing* yang merupakan sebuah tahap yang tidak dapat terlewat dalam mengembangkan model *machine learning*. Pembagian data ini bertujuan untuk menguji sejauh mana kinerja model *machine learning* dalam memprediksi hasil yang akurat pada *new dataset* yang akan ada di masa mendatang. Data *training* adalah bagian data yang berguna sebagai

pelatih model *machine learning*. Data *training* merupakan *dataset* yang sudah diketahui labelnya atau target *outputnya*, sehingga model *machine learning* dapat mempelajari pola pada data tersebut dan menghasilkan model yang dapat melakukan prediksi dengan akurat [18-19]. Contoh penggunaan perbandingan ini dapat diperhatikan dalam tabel 5 dan 6.

**Tabel 5** Contoh data *training*

No	Class	Keywords
1	Positif	Baik (4), Seru (4), Asik (2)
2	Negatif	Kaku (2), Dadakan (2)
3	Netral	Interview (8), Mentor (7)

**Tabel 6** Contoh data *testing*

Kaku	Interaksi	Seru
------	-----------	------

### 7. Analisis Algoritma *Naïve Bayes Classifier*

Pembagian ini menggunakan perbandingan 70:30 dan siap dianalisis menggunakan algoritma *Naïve Bayes Classifier*. Data *training* membentuk peluang dan diujikan pada tabel peluang oleh data *testing*. Tabel 7 menyatakan proporsi pembagian data yang menguji *machine learning*.

**Tabel 7** Proporsi Perbandingan Data

No	Class	Data Training	Data Testing
1	Positif	703	107
2	Negatif	12	107
3	Netral	81	107

Dari penerapan *Naïve Bayes Classifier*, diperoleh asumsi probabilitas untuk setiap  $P(X)$ . Asumsi yang dimiliki adalah *review* positif dan negatif dari pengguna *Career Preparation Program*.

**Tabel 8** Probabilitas Klasifikasi

Positif		Negatif	
Positif	703	Negatif	12
P (positif)	0.983	P (negatif)	0.017

Contoh penggunaan pada kata mentor yang terdiri atas 7 kata data *training*, diperoleh nilai  $P(\text{mentor}|\text{Positif}) = 0.875$  dan  $P(\text{mentor}|\text{negatif}) = 0.0765$  menggunakan rumus analisis pada persamaan (1).

### 8. Analisis Sentimen dengan *Confusion Matrix*

Setelah dibuat proporsi data, dilakukan klasifikasi berdasarkan *output actual* dan *prediction* data. Penggunaan *confusion matrix* mengidentifikasi banyaknya prediksi benar dan salah pada setiap kelas negative dan kelas positif.

**Tabel 9** *Confusion Matrix*

	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP	FP
<i>Negative</i>	FN	TN

Keterangan [7]:

- TP berarti *True Positive* yang menyatakan kuantitas data bernilai *actual* positif diklasifikasi dalam positif.
- TN berarti *True Negative*, yang menyatakan kuantitas data bernilai *actual* negatif diklasifikasi dalam negatif.
- FN berarti *False Negative*, yang menyatakan kuantitas data bernilai *actual* negatif tetapi diklasifikasi salah.



- FP berarti *False Positive*, yang menyatakan kuantitas data bernilai *actual* positif tetapi diklasifikasi salah.

Jika telah diperoleh angka yang menyusun *confusion matrix*, maka nilai tersebut digunakan untuk mengukur nilai *accuracy*, *precision*, dan *recall*. Berdasarkan analisis yang telah dilakukan, *review* program dengan *class 0* terklasifikasi *True* berjumlah 328 kata, disamping itu *class 1* dengan klasifikasi benar sebanyak 106 kata. Akurasi pada klasifikasi benar adalah 93.75%. Sedangkan klasifikasi salah pada *class 0* dan *class 1* tidak lebih dari 30 kata dengan akurasi 92%. Berdasarkan informasi berikut, diperoleh performa yang di hasilkan dari hasil pengujian yang dilakukan terhadap 70% data *train* dengan perbandingan 30% data *test* menghasilkan *accuracy* senilai 92.5%, dengan nilai presisi 1 Sementara pada *Class Recall* dihasilkan nilai 96,72% (*positive class: positive*).

#### IV. KESIMPULAN

1. Berdasarkan *sentiment analysis* menggunakan pendekatan *Naïve Bayes Classifier* yang telah dilakukan, diperoleh probabilitas klasifikasi positif sebesar 0.983 dan klasifikasi negatif sebesar 0.017 dengan nilai akurasi pada pengujian proporsi antara data *train* berbanding data *test* 70:30 sebesar 92.5%. Peluang *review* positif sebesar 98.3% menyatakan banyaknya *review* positif yang diberikan oleh pengguna *Career Preparation Program*. Hal tersebut bernilai cukup besar dan menyatakan bahwa model yang digunakan dapat menganalisis performa *review* pengguna *Career Preparation Program*.
2. Emosi dasar yang dihasilkan paling banyak adalah positif dan kepercayaan terhadap produk *Career Preparation Program Eduwork*

#### UCAPAN TERIMA KASIH

Penulis berterima kasih pada Eduwork (PT. Talenta Sinergi Group) yang telah memberikan izin penelitian serta seluruh responden penelitian yang merupakan pengguna *Career Preparation Program Eduwork*. Rasa terima kasih juga penulis ucapkan pada panitia Seminar Nasional Sains Data (SENADA) 2023 yang telah mengadakan seminar nasional dan melakukan publikasi pada makalah yang penulis buat. Tidak lupa rasa terima kasih pada segenap rekan yang sudah membantu proses penelitian sehingga makalah ini dapat diselesaikan dengan baik.

#### REFERENSI

1. BPS, "Pengangguran Terbuka Menurut Pendidikan Tertinggi yang Ditamatkan 1986 - 2022," Badan Pusat Statistik, 9 Januari 2023. [Online]. Available: <https://www.bps.go.id/statictable/2009/04/16/972/pengangguran-terbuka-menurut-pendidikan-tertinggi-yang-ditamatkan-1986---2022.html>. [Accessed 20 Februari 2023].
2. A. Ahdiat, "Jumlah dan Tingkat Pengangguran Terbuka di Indonesia (Februari 2019-Februari 2023)," Katadata Media Network, 5 Mei 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/05/05/awal-2023-ada-79-juta-pengangguran-di-indonesia>. [Accessed 7 Mei 2023].
3. Akhlanudin, "Gema Uhamka," Universitas Muhammadiyah Prof. Dr. Hamka, 16 Juni 2016. [Online]. Available: <https://gema.uhamka.ac.id/2016/06/16/kenapa-lulusan-perguruan-tinggi-makin-susah-mendapat-pekerjaan/#:~:text=Susah%20terserapnya%20lulusan%20perguruan%20tinggi,manajemen%20perusahaan%2C%E2%80%9D%20kata%20Lilis..>
4. R. A. Ariestanie, "Training "Upgrading Self Value" untuk Mempersiapkan Mahasiswa dan Alumni dalam Menghadapi Proses Seleksi," in *Proceeding Indonesia Career Center Network Summit IV*, Samarinda, 2019.
5. P. A. Permatasari, Linawati and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media



- Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. XX, no. 2, pp. 177-186, 2021.
6. N. T. Romadhoni, I. Santoso and S. Budilaksono, "Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line," *Jurnal IKRA-ITH Informatika*, vol. 3, no. 2, pp. 1-9, 2019.
  7. V. A. Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *Jurnal Buana Informatika*, vol. 11, no. 2, pp. 141-151, 2020.
  8. R. Y. Hayuningtyas and R. Sari, "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan Naive Bayes dan PSO," *Jurnal Techno Nusa Mandiri*, vol. 16, no. 1, pp. 37-42, 2019.
  9. A. V. Sudiantoro and E. Zuliarso, "Analisis Sentimen Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," in *Prosiding SINTAK*, Semarang, 2018.
  10. B. Gunawan, H. S. Pratiwi and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika*, vol. 4, no. 2, pp. 113-118, 2018.
  11. E. Sipayung, H. Maharani and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem Informasi*, vol. 8, no. 1, pp. 958-965, 2016.
  12. S. M. Salsabila, A. A. Murtopo and N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier," *Jurnal Minfo Polgan*, vol. 11, no. 2, pp. 30-35, 2022.
  13. D. Darwis, N. Siskawati and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131-145, 2021.
  14. M. Pradana, "Penggunaan Fitur Wordcloud dan document term matrix dalam Text Mining," *Jurnal Ilmiah Informatika*, vol. 8, no. 01, pp. 38-43, 2020.
  15. A. F. Rozi and A. S. Purnomo, "Analisis Sentimen Untuk Respon Masyarakat Terhadap Universitas (Studi Kasus : Universitas Mercu Buana Yogyakarta)," *Jurnal Information System & Artificial Intelligence*, vol. 1, no. 2, pp. 53-60, 2021.
  16. N. Saputra, "Analisis Sentimen dengan Preprocessing Kata," *Jurnal Dinamika Informasi*, vol. 7, no. 1, pp. 45-57, 2018.
  17. L. Ardiani, H. Sujaini and Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, no. 2, pp. 183-190, 2020.
  18. R. Julianto, E. D. Bintari and Indrianti, "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naive Bayesian Classification," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 3, no. 1, pp. 23-30, 2017.
  19. Y. Astuti, I. R. Wulandari, A. R. Putra and N. Kharomadhona, "Naive Bayes untuk Prediksi Tingkat Pemahaman Kuliah Online Terhadap Mata Kuliah Algoritma Struktur Data," *Jurnal Edukasi dan Penelitian Informatika*, vol. 8, no. 1, pp. 28-32, 2022.



20. D. Untari, "Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree dan Algoritma C4.5," *Jurnal Informasi & Teknologi*, vol. 2, no. 1, pp. 23-29, 2020.



# Implementasi Multiple Linear Regression dalam Menganalisis Jarak Luncur Guguran Awan Panas

Rahma Elis Setiawati<sup>1</sup>, Rohmatul Fajriyah<sup>2</sup>,

<sup>1,2</sup> Program Studi Statistika, FMIPA UII

Corresponding author email: [20611163@students.uui.ac.id](mailto:20611163@students.uui.ac.id), [b966110101@uui.ac.id](mailto:b966110101@uui.ac.id)

**Abstract:** Mount Merapi is one of the most active volcanoes in Indonesia. The Multiple Linear Regression method used in this study is due to the presence of several factors or independent variables that can influence the dependent variable, namely the sliding distance of hot cloud fall. In addition to using descriptive statistics to understand data in general, in this study also used the Multiple Linear Regression method with the help of R Studio software. The results of this study show that the magnitude of the amplitude and duration of the slide affect the glide distance of hot cloud avalanches. The greater the amplitude and duration, the more the sliding distance of the hot cloud fall. The amplitude and duration variables also had an influence of 67.2% on the glide distance variable. However, another 32.8% were influenced by other variables not observed in the study. The importance of research is due to the high level of activity of Mount Merapi is very dangerous for the surrounding community. The hope of the research conducted aims to provide information with a high level of accuracy also helps volcanologists, governments, and the public in taking appropriate action in anticipating future hot cloud avalanche disasters.

**Keywords:** Regression, Amplitude, Slide Distance, R

**Abstrak:** Gunung Merapi merupakan salah satu gunung berapi di Indonesia yang paling aktif. Erupsi Gunung Merapi sangat berbahaya bagi pemukiman padat yang mengelilinginya. Penelitian ini memberikan rekomendasi bagi para ahli vulkanologi, pemerintah, dan masyarakat dalam mengantisipasi bencana guguran awan panas di Gunung Merapi. Metode Regresi Linear Berganda digunakan pada penelitian ini disebabkan oleh adanya beberapa faktor atau variabel bebas yang dapat memberi pengaruh pada variabel terikat, yaitu jarak luncur guguran awan panas. Selain menggunakan statistika deskriptif untuk memahami data secara umum, dalam penelitian ini juga digunakan metode *Multiple Linear Regression* dengan bantuan perangkat lunak R Studio. Hasil dari penelitian ini menunjukkan bahwa besaran amplitudo dan durasi luncuran mempengaruhi jarak luncur guguran awan panas. Semakin besar amplitudo dan durasi, maka jarak luncur guguran awan panas semakin meningkat. Variabel amplitudo dan durasi juga memberikan pengaruh sebesar 67,2% terhadap variabel jarak luncur. Namun, 32,8% lainnya dipengaruhi variabel lain yang tidak diamati dalam penelitian ini. Pentingnya penelitian disebabkan oleh tingginya tingkat aktivitas Gunung Merapi sangat berbahaya bagi masyarakat sekitarnya. Harapan dari penelitian yang dilakukan bertujuan memberi informasi dengan tingkat akurat tinggi juga membantu para ahli vulkanologi, pemerintah, dan masyarakat dalam mengambil tindakan yang tepat dalam mengantisipasi bencana guguran awan panas di masa depan.

**Kata kunci:** Regresi, Amplitudo, Jarak Luncur, R

## I. PENDAHULUAN

Gunung Merapi dianggap sebagai salah satu gunung yang berbahaya karena telah terjadi erupsi setiap 2 hingga 5 tahun sekali menurut catatan modern. Erupsi Gunung Merapi memiliki potensi bahaya yang tinggi bagi pemukiman padat di sekitarnya. Gunung Merapi telah mengalami 68 kali letusan sejak tahun 1548. Terdapat dua kota besar, yaitu Yogyakarta dan Magelang yang berada kurang dari 30 km dari puncak gunung tersebut. Di lereng Gunung Merapi terdapat pemukiman pada ketinggian 1700 m dan jaraknya hanya 4 km dari puncak. Karena aktivitasnya yang sangat tinggi, Gunung Merapi termasuk salah satu bentuk *Decade Volcanoes* selain 16 gunung api se-dunia. [1]

Awan panas pada Gunung Merapi memiliki jenis guguran. Awan panas tipe ini disebabkan oleh keruntuhan kubah di puncak gunung yang tidak stabil, yang kemudian menyebabkan longsoran. Ketidakstabilan kubah merupakan akibat dari tingginya tekanan gas yang berasal dari dalam gunung yang memicu gas naik ke atas dan saling berguguran keluar melalui terusan kepundan. Kecepatan awan panas ditentukan oleh gaya gravitasi pada kubah lava yang mengalami keruntuhan. [2]

Hingga saat pembuatan jurnal ini, Gunung Merapi berada pada status siaga level III. Secara umum, statistik terbagi menjadi dua jenis, yaitu statistik inferensia dan statistik deskriptif [3]. Statistik deskriptif berfungsi dalam menggambarkan karakteristik sebuah data dengan menghitung ukuran-ukuran seperti mean dan median untuk mengukur pusat data, serta standar deviasi untuk mengukur variasi data. [4]

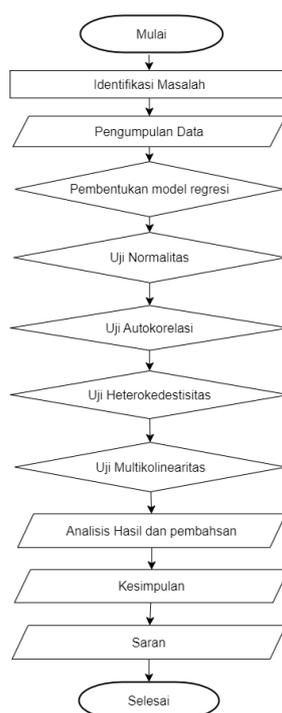
Amplitudo dan durasi luncuran adalah dua faktor yang mempengaruhi jarak luncur guguran awan panas. Amplitudo mengacu pada besar kecilnya gelombang atau getaran pada luncuran awan panas, sedangkan durasi luncuran mengacu pada waktu yang dibutuhkan untuk luncuran awan panas dari puncak gunung hingga mencapai lereng gunung.

Penelitian tentang hubungan antara besaran amplitudo dan durasi luncuran terhadap jarak luncur guguran awan panas Gunung Merapi menggunakan metode *Multiple Linear Regression* sangat penting untuk dilakukan. Model dari metode tersebut dapat mengidentifikasi hubungan linear antar beberapa variabel independen dengan variabel dependen.

## II. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang berasal dari Balai Penyelidikan dan Pengembangan Teknologi Kebencanaan Geologi (BPPTKG). Data yang digunakan merupakan Data Harian Guguran Awan Panas Gunung Merapi Periode Januari 2021 – April 2022

Dalam penelitian ini, selain menggunakan statistik deskriptif untuk memperoleh gambaran umum data, digunakan pula metode *multiple linear regression* atau regresi linear berganda dengan bantuan perangkat lunak R Studio.



**Gambar 1.** Diagram Alir Penelitian

### 2.1. Analisis Deskriptif

Statistik deskriptif merupakan suatu teknik untuk mengumpulkan serta menyajikan data hingga tersaji informasi yang bermanfaat. [5] Tujuan dari statistik deskriptif adalah untuk menjelaskan atau menggambarkan objek penelitian melalui data yang diperoleh dari populasi atau sampel. [6] Biasanya, statistik deskriptif menyajikan data dalam bentuk ukuran pemusatan, seperti mean. [7] Selain menggunakan ukuran pemusatan, data juga dapat diwakili dengan bentuk lain seperti diagram Pareto atau tabel untuk membantu memahami informasi yang disajikan. [8]

## 2.2. Multiple Linear Regression

Regresi berganda atau yang juga dikenal sebagai *multiple regression*, adalah sebuah model regresi sebagai media perhitungan minimal dua variabel bebas dalam melakukan prediksi atau analisis. Model regresi melibatkan dua bentuk variabel, yakni variabel bebas (variabel independen) dan variabel terikat (variabel dependen). Variabel yang berfungsi sebagai penyebab atau prediktor dalam model analisis bebas merupakan variabel independen yang marak disebut variabel bebas, sedangkan variabel terikat atau respon didefinisikan sebagai nilai dibawah pengaruh variabel bebas, biasanya dikenal dengan variabel dependen (variabel yang ingin diprediksi). [9]

*Multiple linear regression* dengan  $q$  variabel bebas mengacu pada persamaan model (1) berikut.

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_p x_{jq} - \varepsilon_j \quad (1)$$

dengan :

$y_j$  : variabel dependen untuk pemngamatan ke- $j$ ,  $j=1,2,\dots,z$ .

$\beta_0, \beta_1, \beta_2, \dots, \beta_q$  : parameter.

$X_{j1}, X_{j2}, \dots, X_{jq}$  : variabel independen.

$\varepsilon_i$  : *error* (sisa) pengamatan ke- $i$ , berasumsi *normal distribution* saling identik serta bebas dengan rata-rata nol dan varians  $\sigma^2$ .

Pada notasi *matrix* dalam persamaan (1) dapat ditulis dalam bentuk persamaan (2).

$$Y = X \beta + \varepsilon \quad (2)$$

dengan :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_z \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,q-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2,q-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{z1} & x_{z2} & \dots & x_{z,q-1} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} \text{ dan } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_z \end{pmatrix}$$

$Y$  adalah *vector* variabel terikat ( $n \times 1$ ).

$X$  adalah *matrix* variabel bebas ( $n \times (p - 1)$ ).

$\beta$  adalah *vector* parameter ( $p \times 1$ ).

$\varepsilon$  adalah *vector error* ( $n \times 1$ ).

### 2.2.1. Uji Normalitas Residual

Dalam penelitian ini, metode Jarque-Bera digunakan untuk mengevaluasi perbedaan kemiringan dan kurtosis data. Distribusi chi-kuadrat dengan derajat kebebasan dua digunakan dalam pengujian Jarque-Bera. Menurut Winarno (2015), Uji Jarque-Bera digunakan untuk memeriksa ada tidaknya kenormalan distribusi pada data dengan membandingkan skewness dan kurtosis dari data dengan nilai yang diharapkan pada distribusi normal. Distribusi hasil uji Jarque-Bera menggunakan  $Z^2$  dengan derajat kebebasan sebesar 2. Data berdistribusi normal, jika hasil uji Jarque-Bera menunjukkan nilai probabilitas lebih dari 1%. Begitupun apabila probabilitas kurang dari 1%, data dianggap tidak memiliki distribusi normal. Dewi et al. (2014) menambahkan untuk mengevaluasi apakah residual data (variabel pengganggu) yang ada dalam model regresi dikatakan berdistribusi normal, digunakan uji normalitas.

### 2.2.2. Uji Asumsi non Autokorelasi

Uji autokorelasi digunakan untuk mengevaluasi adanya keterkaitan antara nilai data pada satu periode dengan periode sebelum nilai tersebut. Artinya, uji autokorelasi melibatkan analisis regresi guna menentukan pengaruh variabel bebas pada variabel terikat, dan diperlukan untuk memastikan tidak adanya signifikansi korelasi antar pengamatan pada periode tertentu dengan data pada periode sebelumnya. [10] Uji autokorelasi hanya diperlukan untuk data yang disusun secara berurutan dalam waktu atau *time series*.



Durbin-Watson adalah metode yang digunakan pada penelitian ini untuk memprediksi keberadaan korelasi pada model *multiple linear regression*. Durbin-Watson mengevaluasi keberadaan autokorelasi dengan menghitung rasio dari selisih nilai residual antara dua observasi yang berdekatan dengan keseluruhan nilai residual. Nilai rasio tersebut kemudian dibandingkan dengan nilai kritis yang tergantung pada jumlah variabel independen dan jumlah observasi. Apabila nilai rasio yang dihasilkan lebih kecil dari nilai kritis bawah, maka terdapat indikasi adanya autokorelasi positif. Begitupun apabila rasio lebih tinggi dari titik kritis atas, maka terdapat indikasi adanya autokorelasi negatif. Namun, jika nilai rasio berada di antara nilai kritis bawah dan atas, maka tidak ada cukup bukti untuk menyimpulkan keberadaan autokorelasi.

### 2.2.3. Uji Asumsi Heterokedastisitas

Heteroskedastisitas terjadi ketika variasi dari kesalahan (*error*) dalam model regresi tidak konstan atau ada perbedaan dalam variasi antar kesalahan satu dengan kesalahan yang lain. Dalam melakukan uji ini, peneliti dapat memeriksa apakah ada perbedaan yang tidak konsisten dalam variansi residual antara pengamatan-pengamatan yang berbeda. Homoskedastisitas adalah salah satu asumsi dalam model regresi yang menyatakan bahwa variansi residual atau kesalahan prediksi adalah sama di seluruh level variabel independen atau faktor-faktor prediktor yang digunakan dalam model. [11]

### 2.2.4. Uji Asumsi Multikolinearitas

Uji multikolinearitas adalah suatu metode yang digunakan untuk mengevaluasi keberadaan signifikansi korelasi antar variabel independen dalam model *multiple linear regression*. Uji multikolinearitas berfungsi dalam menentukan apakah ada atau tidaknya hubungan yang kuat antara variabel independen yang berakibat hubungan antar variabel bebas dan variabel terikat dalam model regresi kurang maksimal. Korelasi yang tinggi antara variabel bebas dapat memengaruhi validitas model regresi dan menyebabkan kesulitan dalam menginterpretasi koefisien regresi.

Dampak adanya multikolinieritas pada model *multiple linear regression* adalah :

1. Meskipun penaksir OLS masih dianggap sebagai BLUE (Best Linear Unbiased Estimator), namun memiliki varian dan kovarians yang besar yang mengakibatkan sulitnya untuk memperoleh estimasi yang akurat.
2. Ketika variansi dan kovariansi penaksir OLS meningkat, hal ini akan menyebabkan interval estimasi menjadi lebih luas dan hasil uji t menjadi lebih sempit. Dampak dari hal tersebut adalah kemungkinan variabel independen menjadi tidak signifikan secara statistik dalam memberi pengaruh pada variabel terikat.
3. Meskipun melalui uji-t setiap variabel terikat secara terpisah tidak signifikan dalam mempengaruhi variabel bebas, nilai koefisien determinasi ( $R^2$ ) masih bisa tinggi.

Metode nilai *variance inflation factor* (VIF) dan nilai toleransi (TOL) digunakan sebagai pendeteksi adanya multikolinieritas dalam model *multiple linear regression*. Jika nilai *Variance Inflation Factor* (VIF) dengan hasil diatas 10, untuk itu, hal tersebut menandakan adanya indikasi multikolinieritas dalam model. Sebaliknya, apabila *Tolerance* (TOL) bernilai 1, hal tersebut menunjukkan multikolinieritas dalam model tidak ada.

## III. HASIL DAN PEMBAHASAN

### 3.1. Statistika Deskriptif

Analisis deskriptif pada data guguran awan panas pada periode januari 2021 – April 2022 dengan menggunakan variabel Amplitudo, Durasi, dan Jarak Luncur pada tabel 1.

**Tabel 1.** Statistika Deskriptif Guguran Awan Panas

Deskripsi	Amplitudo (mm)	Durasi (Detik)	Jarak Luncur (m)
-----------	----------------	----------------	------------------

Minimum	10	47,0	300
Kuartil 1	27	99,0	1100
Median	37	124,0	1500
Mean	38	132,9	1520
Kuartil 3	50	152,0	1800
Maksimum	90	702,3	9000

Berdasarkan **Tabel 5.1.** terdapat 3 variabel yang digunakan, yakni amplitudo (mm), Durasi (Detik), dan Jarak Luncur (m). Pada kolom deskripsi terdapat variabel minimum yang menunjukkan nilai minimum dari ketiga variabel yang digunakan. Nilai kuartil 1 atau disebut juga kuartil atas merupakan kuartil yang membagi 25% urutan data terkecil. Median merupakan nilai tengah data setelah data disusun berurutan. Sementara itu, mean (rata-rata) menyatakan jumlah nilai data dibandingkan dengan jumlah total data. Kuartil 3 atau kuartil atas menyatakan nilai tengah antar median dengan nilai maksimal data. Sedangkan nilai maksimum adalah nilai tertinggi dalam kelompok data.

### 3.2. Estimasi Model Regresi Linear Berganda

Diperlukan sebanyak tiga kali proses penghapusan outlier untuk menghasilkan data yang terbebas dari outlier. Dari data yang baru, outlier sudah tidak terdeteksi sehingga penelitian dapat dilanjutkan ke alur berikutnya yakni pembentukan estimasi model regresi linear berganda. Model pertama yang terbentuk adalah sebagai berikut.

$$\text{Jarak Luncur} = 375,5007 + 3,0502 \text{ Amplitudo} + 7,7223 \text{ Durasi}$$

*Coefficients* menunjukkan koefisien estimasi untuk setiap variabel dalam model *multiple linear regression*. Koefisien estimasi menunjukkan besarnya pengaruh setiap variabel bebas terhadap variabel terikat. Dalam model diatas, terdapat tiga koefisien estimasi, yaitu koefisien untuk intercept, Amplitudo, dan Durasi. Dalam model diatas, nilai koefisien untuk Amplitudo adalah 3,0502 dan untuk Durasi adalah 7,7223. Hal ini menunjukkan bahwa setiap peningkatan satu mm pada Amplitudo akan meningkatkan Jarak Luncur sebanyak 3,0502 m, sementara setiap peningkatan satu detik pada Durasi akan meningkatkan Jarak Luncur sejauh 7,7223 m.

#### 3.2.1. Uji Normalitas Residual

Pada peneitian ini, nilai probabilitas Jarque-Bera yang dihasilkan sebesar  $2.2e-16$  yang artinya peluang Jarque-Bera  $< 1\%$  menyatakan data yang ada tidak berdistribusi normal. Karena pada estimasi model regresi ini uji asumsi normalitasnya tidak terpenuhi, maka dilakukan transformasi pada residual dilakukan dalam bentuk akar kuadrat sebanyak 2 kali untuk memenuhi asumsi normalitas. Dengan menggunakan tingkat kepercayaan 99% dan dilakukan transformasi sebanyak dua kali, maka asumsi normalitas telah terpenuhi.

#### 3.2.2. Uji Asumsi non Autokorelasi

Dalam model diatas, autokorelasi perlu diujikan jika data berupa data runtun waktu atau time series. Autokorelasi merujuk pada pengaruh nilai observasi sebelumnya terhadap nilai pada sampel atau observasi tertentu. Dalam penelitian ini, nilai DW berada diantara *range* DU dan DL, sehingga dapat disimpulkan bahwa asumsi non-autokorelasi tidak terpenuhi.

#### 3.2.3. Uji Asumsi Heterokedastisitas

Jika p-value tes kurang dari beberapa tingkat signifikansi (yaitu  $\alpha = 0,01$  seperti yang telah ditentukan oleh peneliti) maka kami menolak hipotesis nol dan menyimpulkan bahwa heteroskedastisitas hadir dalam model regresi.

Dengan menggunakan tingkat kepercayaan 99%, pada penelitian ini didapatkan nilai P-Value test pada uji asumsi heterokedastisitas yakni sebesar 0,09166, dimana nilai yang ada lebih besar dari  $\alpha = 0,01$  seperti yang telah ditentukan oleh peneliti. Maka dapat disimpulkan bahwa variansi galat bersifat homoskedastisitas.

### 3.2.4. Uji Multikolinieritas

Dalam deteksi multikolinieritas pada *multiple linear regression*, digunakan metode nilai VIF TOL. Jika hasil VIF yang diperoleh diatas 10, data yang ada menunjukkan bahwa terdeteksi multikolinieritas dalam model.

**Tabel 2.** Tabel Nilai VIF

Amplitudo	Durasi
1,056575	1,056575

Dari **Tabel 2.**, dapat disimpulkan bahwa asumsi non multikolinieritas terpenuhi karena dari hasil perhitungan VIF pada variabel yang digunakan menghasilkan nilai  $< 10$ .

### 3.3. Pemilihan Model Terbaik

Dengan menggunakan metode *forward selection*, dapat diketahui bahwa model terbaik membutuhkan variabel Amplitudo dan variabel Durasi.

**Tabel 3.** Forward Selection

Selection Summary						
Step	Variable Entered	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	data.t2\$Durasi..detik.	0.6692	0.6684	4.3638	37.7165	0.2522
2	data.t2\$Amplitudo..mm.	0.6720	0.6703	3.0000	36.3418	0.2515

Dan jika digunakan metode *backward elimination* akan menghasilkan bahwa untuk mendapatkan model terbaik adalah dengan mengikutsertakan kedua variabel pada **Tabel 3.** dimana itu diartikan bahwa tidak ada variabel yang harus dibuang untuk mendapatkan model regresi terbaik. Sehingga didapatkan model regresi sebagai berikut.

$$\text{arak Luncur} = 0,47633 + 0,09427\text{Amplitudo} + 1,62514\text{Durasi} + e$$

Dalam persamaan regresi linear berganda di atas, variabel Amplitudo dan Durasi memiliki koefisien positif. Artinya, peningkatan Amplitudo dan Durasi akan berdampak positif pada Jarak Luncur. Koefisien regresi pada variabel Amplitudo sebesar 0,09427 menunjukkan bahwa pada setiap peningkatan 1mm Amplitudo akan meningkatkan Jarak Luncur sebesar 0,09427m. Sementara itu, koefisien regresi variabel Durasi sebesar 1,62514 menunjukkan pada setiap peningkatan Durasi 1 detik akan meningkatkan Jarak Luncur sebesar 1,62514m.

### 3.4. Koefisien Determinasi ( $R^2$ )

Didapatkan *Determination coefficient value* ( $R^2$ ) bernilai 0,672, nilai tersebut menandakan variabel Amplitudo dan variabel Durasi berpengaruh pada variabel Jarak Luncur senilai 67,2%. Nilai 32,8% sisanya tidak terpengaruh oleh variabel tersebut atau dipengaruhi oleh variabel lain yang tidak diobservasi.

## IV. KESIMPULAN

1. Variabel Amplitudo dan Durasi mempunyai koefisien positif. Artinya, peningkatan Amplitudo dan Durasi berdampak positif pada Jarak Luncur. Dengan demikian dapat disimpulkan bahwa setelah dilakukan beberapa uji dan hasilnya dapat disimpulkan semakin besar Amplitudo dan Durasi akan semakin jauh pula Jarak Luncur Guguran Awan Panas yang dihasilkan oleh Gunung Merapi.



2. Variabel Amplitudo dan variabel Durasi berpengaruh terhadap variabel Jarak Luncur senilai 67,2%. Sedangkan 32,8% lainnya merupakan pengaruh dari variabel di luar persamaan atau variabel lain yang tidak diobservasi.
3. Prediksi yang dihasilkan untuk setiap peningkatan 1mm Amplitudo akan meningkatkan Jarak Luncur sebesar 0,09427m. Sementara itu, setiap peningkatan Durasi selama 1 detik akan meningkatkan Jarak Luncur sejauh 1,62514m.

**UCAPAN TERIMA KASIH****REFERENSI**

1. bencanapedia.id. (2021, 10 16). *Gunung Merapi*. Diambil kembali dari bencanapedia.id: [http://www.bencanapedia.id/Gunung\\_Merapi#:~:text=Gunung%20Merapi%20adalah%20gunung%20termuda%20dalam%20rangkaian%20gunung,aktivitas%20vulkanik%20di%20sepanjang%20bagian%20tengah%20Pulau%20Jawa](http://www.bencanapedia.id/Gunung_Merapi#:~:text=Gunung%20Merapi%20adalah%20gunung%20termuda%20dalam%20rangkaian%20gunung,aktivitas%20vulkanik%20di%20sepanjang%20bagian%20tengah%20Pulau%20Jawa).
2. farah, f. (2016, December 02). *Pengertian Awan Panas Aliran, Hembusan, dan Jatuhan*. Diambil kembali dari ilmugeografi.com: <https://ilmugeografi.com/geologi/pengertian-awan-panas>
3. Walpole, R. (1995). *Pengantar Statistika Edisi Ke-3*. Jakarta: PT. Gramedia Pustaka Utama.
4. Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics Principles & Methods*. USA: John Wiley & Sons.
5. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists 9th Ed*. USA: Pearson.
6. Sugiyono. (2007). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
7. Fauzy, M. (2009). *Dalam Metode Penelitian Kuantitatif*. Semarang: Walisongo Press.
8. Kuswanto. (2012). *Dalam Statistik Untuk Pemula dan Orang Awam*. Jakarta: Laskar Aksara.
9. Setia Ningsih, H. D. (2019, Januari). *Penerapan Metode Suksesif Interval pada Analisis Regresi*. Diambil kembali dari ejurnal.ung: <https://ejurnal.ung.ac.id/index.php/jjom/article/view/1742/1270>
10. gramedia.com. (2021). *Pengertian Uji Asumsi dan Jenis-Jenisnya*. Diambil kembali dari gramedia.com: <https://www.gramedia.com/literasi/uji-asumsi/>
11. gramedia.com. (2021). *Pengertian Uji Asumsi dan Jenis-Jenisnya*. Diambil kembali dari gramedia.com: <https://www.gramedia.com/literasi/uji-asumsi/>



# Penerapan Metode Fuzzy Sugeno Pada Otomatisasi Oven Pengering Ikan Berbasis IoT

Sebastianus Pehan Makin<sup>1</sup>, Narchrowie<sup>2</sup>, Subairi<sup>3</sup>,

<sup>1,2,3</sup> Jurusan Teknik Elektro, Universitas Merdeka Malang

<sup>1</sup>[gasparmakin1998@gmail.com](mailto:gasparmakin1998@gmail.com)

<sup>2</sup>[nachrowie@unmer.ac.id](mailto:nachrowie@unmer.ac.id)

<sup>3</sup>[subairi@unmer.ac.id](mailto:subairi@unmer.ac.id)

Corresponding author email: [nachrowie@unmer.ac.id](mailto:nachrowie@unmer.ac.id)

**Abstract:** Salted fish is a typical food of Indonesian society. Almost all people consume food made from salted fish, fishery products are still processed and preserved by salting. The process of drying salted fish can use a variety of sources, one of which is to leave the fish in the sun. Drying with the help of sunlight does not require large costs and although this drying can change the nature of fresh fish meat, the nutritional value is not relatively changed. So the role of sunlight is very important, the uncertain and uncertain weather conditions become one of the problems in the process of drying the salted fish. Many salted fish processing industries in the region but still use traditional methods in drying. Based on these problems, the researcher aims to make an Automatic Salted Fish Dryer Oven with Internet Of Things (IoT) based Fuzzy Logic Method, which is used to dry salted fish without the influence of weather and to minimize the contact of insects that land on the fish and speed up the production process from days to 4-8 hours. with a moisture content in the fish as much as 40% or with a weight of 200 grams of 1000 grams of initial weight of fish.

**Keywords:** Fuzzy Logic, Internet Of Things

**Abstrak:** Ikan Asin merupakan makanan khas masyarakat Indonesia. Hampir semua masyarakat mengkonsumsi makanan yang terbuat dari ikan asin, produk perikanan masih di olah dan di awetkan dengan cara pengasinan. Proses pengeringan ikan asin dapat menggunakan beragam sumber, salah satunya adalah dengan membiarkan ikan di bawah sinar matahari. Pengeringan yang dilakukan dengan bantuan sinar matahari tidak memerlukan biaya besar dan meskipun pengeringan ini dapat mengubah sifat daging ikan segar, namun nilai gizi tidak relatif berubah. Maka peran sinar matahari sangat penting, keadaan cuaca yang tidak menentu dan pasti menjadi salah satu masalah dalam proses pengeringan ikan asin tersebut. Banyak industri pengolahan ikan asin di wilayah ini namun masih menggunakan metode tradisional dalam pengeringannya. Berdasarkan permasalahan tersebut, maka peneliti bertujuan untuk membuat Oven Pengering Ikan Asin Otomatis dengan Metode *Fuzzy Logic* berbasis *Internet Of Things (IoT)*, yang digunakan untuk mengeringkan ikan asin tanpa pengaruh cuaca dan untuk meminimalkan kontak serangga yang hinggap pada ikan tersebut serta mempercepat proses produksi dari waktu sehari-hari menjadi 4-8 jam. dengan kadar air pada ikan sebanyak 40% atau dengan berat 200 gram dari 1000 gram berat awal ikan.

**Kata kunci:** media sosial, ujaran kebencian, komentar youtube

## I. PENDAHULUAN

Ikan asin merupakan makanan khas masyarakat Indonesia. Hampir semua lapisan masyarakat mengkonsumsi makanan yang terbuat dari ikan asin, dimana ikan yang di asinkan dan kemudian melalui proses pengeringan[1]. asin merupakan hasil yang diolah secara tradisional dengan membersihkan ikan lalu menambahkan garam dengan kadar air yang di tetapkan sesuai Standar Nasional Indonesia yaitu sekitar 20%-30%. Dari berat total ikan dengan kadar air ikan setelah dikeringkan maksimal 40%. [2].

Proses pengeringan ikan asin dapat menggunakan beragam sumber, salah satunya adalah dengan membiarkan dibawah sinar matahari, keadaan cuaca yang tidak menentu dan pasti menjadi salah satu masalah dalam proses peneringan ikan asin tersebut. Proses produksi dan pembuatan ikan asin akan menjadi lebih lambat jika ikan asin tidak mendapatkan sinar matahari yang cukup [3]. Banyak industri pengolahan ikan asin di wilayah ini namun masih menggunakan metode tradisional dalam proses pengeringannya. Proses pengeringan yang membutuhkan waktu yang lama dan dengan cuaca yang terkadang tidak menentu, sehingga dapat menghambat produktifitas masyarakat.



Berdasarkan permasalahan tersebut, maka peneliti bertujuan untuk membuat Oven Pengering Ikan Asin Otomatis dengan Metode *Fuzzy Logic* berbasis *Internet Of Things* (IoT), yang digunakan untuk mengeringkan ikan asin tanpa pengaruh cuaca dan untuk meminimalkan kontak serangga yang hinggap pada ikan tersebut. Alat ini terhubung dengan aplikasi *android* dan dapat menampilkan informasi proses pengeringan ikan asin, dimana informasi terkait pengeringan ikan asin ditampilkan pada aplikasi *android* dengan memanfaatkan internet of things, sehingga pengguna dapat memantau dan mengontrol proses pengeringan ikan menggunakan aplikasi yang terpasang pada *smartphone* [5]. Disini tampilan monitoring sistem kerja alat melalui aplikasi *blink* yang berperan penting karena menjadi sarana pemantauan pengeringan ikan asin dan pengendalian komponen didalam alat pengering ikan asin tersebut. Suhu yang terdapat pada oven pengering ikan mencapai 50<sup>0</sup>C.

## II. KAJIAN PUSTAKA

### **Sensor Load Cell**

*Load Cell* merupakan komponen inti yang terdapat pada timbangan digital. Secara umum *Load Cell* digunakan untuk menghitung massa dari suatu benda. Sebuah sensor *Load Cell* tersusun dari beberapa konduktor, *strain gauge*, dan jembatan *wheatstone*. *Load Cell* merupakan sensor yang cukup populer untuk pengukuran *Force* terutama untuk pengukuran berat (timbangan elektronik) [6]. Penggunaan *Load Cell* sangat simple dan sangat memudahkan dalam implementasinya. Pada alat ini sensor *Load Cell* digunakan untuk mengetahui berat ikan pada saat proses pengeringan. *Load Cell* yang digunakan untuk mengukur berat dari ikan pada alat oven pengering [7].

### **NodeMCUEsp32**

Esp32 merupakan suksesor atau penerus dari Esp8266 yang memiliki banyak fitur tambahan dan keunggulan dibandingkan generasi sebelumnya. Pada Esp32 terdapat inti CPU serta *Wi-Fi* yang lebih cepat, GPIO yang lebih banyak, dan dukungan terhadap *Bluetooth* 4.2, serta konsumsi daya yang rendah. Esp32 sendiri tidak jauh berbeda dengan Esp8266 yang familiar dipasaran, hanya saja Esp32 lebih kompleks dibandingkan Esp8266. Pada 2016 Espressif meluncurkan produk terbarunya yang bernama Esp32. Esp32 hadir tidak untuk menggantikan Esp8266, namun memberikan perbaikan pada semua lini. Tak hanya mempunyai dukungan konektivitas *WiFi*, akan tetapi juga *Bluetooth* membuatnya lebih serbaguna dan cocok dalam menangani proyek yang besar.[8].

### **Sensor Thermocouple**

*Thermocouple* merupakan salah satu sensor yang paling umum digunakan untuk mengukur suhu karena relatif murah tetapi akurat yang dapat beroperasi pada suhu panas maupun dingin. Beberapa kelebihan *Thermocouple* yang membuatnya menjadi populer ialah responnya yang cepat terhadap perubahan suhu dan pula rentang suhu operasionalnya yang luas yaitu berkisar diantara-200<sup>0</sup>Chingga2000<sup>0</sup>C. Selain respon yang cepat dan rentang suhu yang luas, *Thermocouple* juga tahan terhadap guncangan atau getaran serta praktis digunakan. Sensor suhu *Thermocouple* mempunyai nilai *output* yang kecil dengan *noise* yang tinggi, sehingga memerlukan rangkaian pengkondisi sinyal supaya nilai *output* dapat dibaca dengan baik.[9]

### **Driver MOSFET**

MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) adalah sebuah komponen semikonduktor yang dapat digunakan sebagai saklar elektronik dan penguat sinyal pada perangkat elektronik. MOSFET umumnya memiliki 3 kaki, yaitu *Gate* (G), *Drain* (D), dan *Source* (S). Dari segi bahan semikonduktor pembuatnya, MOSFET terbagi atas 2 bagian yaitu MOSFET *N-type* dan MOSFET *P-type* [8].

### **IoT (Internet of Things)**

Sistem internet of things pada penelitian ini diterapkan pada sistem pengiriman informasi suhu dan data berat dari oven pengeringan ikan yang dapat dimonitoring dengan aplikasi Blynk pada perangkat android berguna.

### Heater

Heater adalah pemanas yang memanfaatkan arus listrik AC frekuensi tinggi yang dialirkan kepada benda kerja berupa batang penghantar yang akan menghasilkan medan elektro magnetik di sekitar benda kerja tersebut, sehingga menghasilkan arus eddy yang akan membuat molekul – molekul dari benda logam yang terdapat disekitar medan elektro magnetik mengeluarkan panas dan meleburkan benda itu sendiri.

### SSR (Solid State Relay)

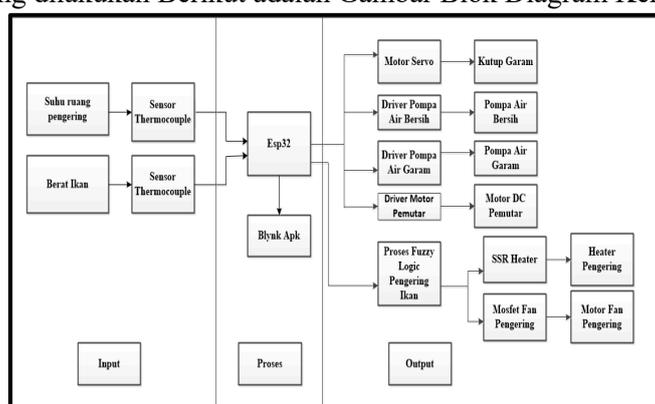
SSR (*Solid State Relay*) merupakan *relay* yang tidak memiliki kumparan dan kontak sesungguhnya dan sebagai penggantinya maka digunakan alat penghubung semikonduktor seperti transistor bipolar, mosfet, scr, atau triac. SSR ditetapkan sebagai kontrol *ON-OFF* dimana arus beban dilakukan oleh satu atau lebih semikonduktor. SSR tidak memiliki bagian yang berputar, *relay* tersebut tahan terhadap guncangan dan getaran serta tertutup rapat terhadap kotoran dan kelembaban.[11]

### Fuzzy Sugeno

Metode *fuzzy Sugeno* adalah metode inferensi *fuzzy* untuk aturan yang dinyatakan dalam bentuk *IF-THEN*. Metode *fuzzy* dikembangkan oleh *Sugeno* yang telah terbukti efektif ketika berurusan dengan sistem nonlinier kompleks yang sangat menantang untuk dianalisis dan disintesis Model ini menggunakan fungsi keanggotaan tunggal. fungsi keanggotaan tunggal adalah dimana satu nilai crisp memiliki derajat keanggotaan 1 dan nilai crisp lainnya memiliki derajat keanggotaan 0. Dalam metode ini, setiap keluaran yang dibentuk dari aturan *IF THEN* bukanlah himpunan *fuzzy*, tetapi persamaan konstan atau linier[12].

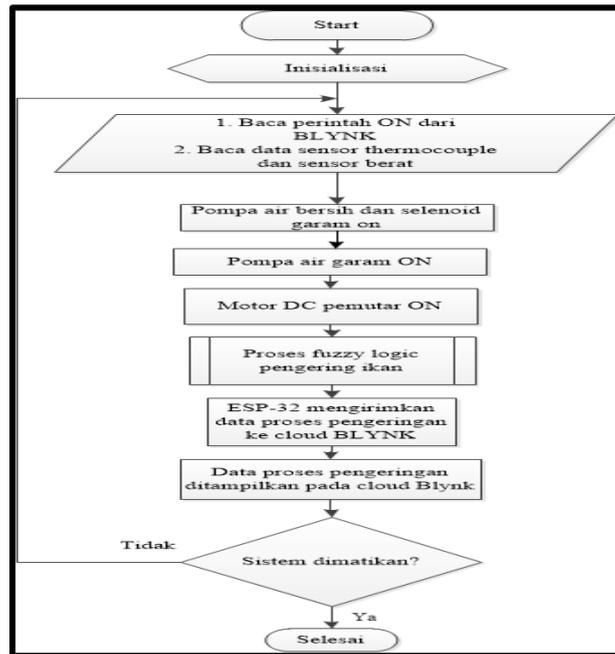
## III. METODE PENELITIAN

Pada bagian ini merupakan tampilan dari Blok Diagram kerja alat, *flowchart*, dan Diagram Fuzzy . Saat membuat desain rancangan alat sangat membutuhkan diagram blok yang menggambarkan tahapan atau proses yang dilakukan Berikut adalah Gambar Blok Diagram Kerja alat.

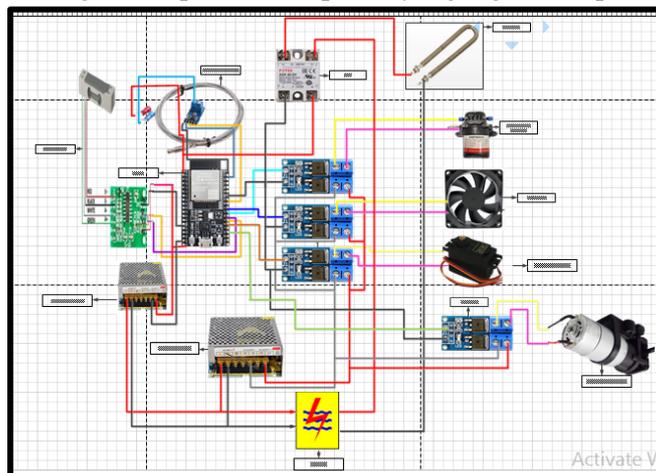


**Gambar 1.** Diagram Blok kerja Alat

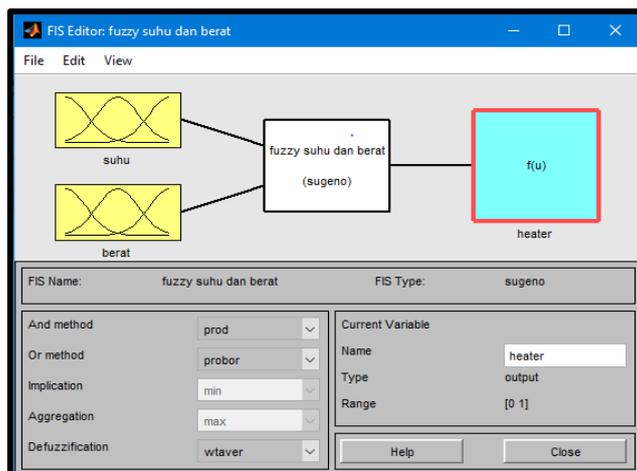
Gambar 1 merupakan bagian dari diagram blok yang berisi komponen-komponen yang digunakan pada penelitian ini seperti sensor thermocouple sebagai sensor suhu, sensor loadcell untuk menghitung berat ikan dan esp 32 sebagai mikrokontroler untuk menghubungkan dengan internet dan mengirimkan data melalui aplikasi Blynk kemudian terdapat motor servo untuk membuka katup garam menuju penampung air bersih serta heater pada bagian output untuk menghasilkan suhu panas pada ruang pemanas Berikut adalah Gambar dari *flowchart* pada penelitian ini.

**Gambar 2.** Flow Chart

Gambar 2 merupakan perangkat lunak yang bermula dari proses inialisasi, pembacaan data sensor sensor thermocouple dan load cell. Kemudian proses pencampuran garam dengan air bersih serta pompa ar garam akan bekerja untuk mengalirkan air garam pada ruang pengeringan dialnjututkan dengan proses pengeringan ikan asin dengan metode pengambilan keputusan menggunakan metode fuzzy sugeno serta mengirimkan data suhu dan berat menggunakan aplikasi blynk, sistem akan dimatikan dan selesai apabila berat ikan turun sesuai dengan setpoint.berikut adalah gambar diagram blok atau skema perancangan komponen- komponen yang digunakan pada penelitian ini.

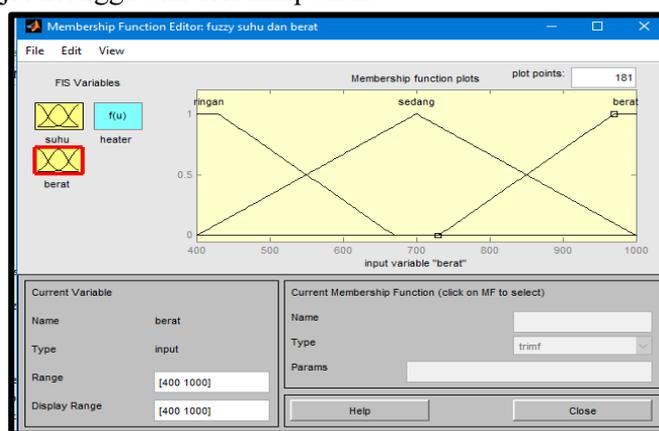
**Gambar skema perancangan alat**

Pada gambar diatas berfungsi untuk memberi petunjuk perancangan komponen elektronika dan Diagram Fuzzy digunakan untuk mengubah nilai data sensor menjadi himpunan *fuzzy* selama penelitian dan pembangunan rumah kaca Ini diproses segera, menghasilkan *fuzzy IF-THEN* yang menggambarkan hubungan antara masukkan dan keluaran. Berikut Merupakan tampilan fuzzy sugeno pada matlab



**Gambar 3.** Fuzzy sugeno

Untuk mendapatkan fungsi keanggotaan pada penelitian ini digunakan variabel kelembaban tanah dengan bentuk gabungan kurva segitiga, kurva linear turun dan kurva linear naik seperti Gambar dan untuk dapat mengetahui derajat keanggotaan dari himpunan seperti Gambar 4 dan untuk dapat mengetahui derajat keanggotaan dari himpunan.



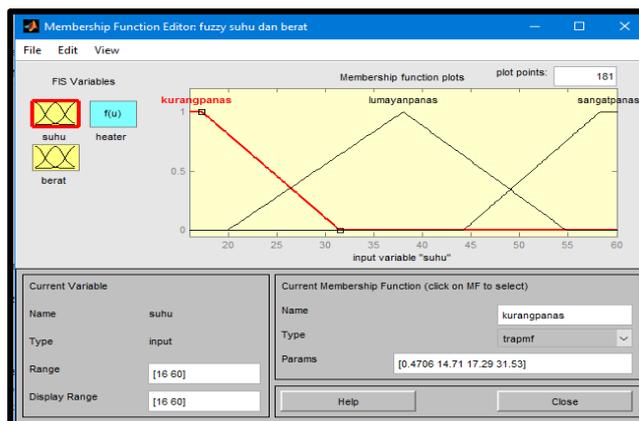
**Gambar 4.** Input variabel berat

Gambar 4 merupakan *domain range input berat* dengan fungsi sebagai batasan nilai yang akan digunakan pada mikrokontroler untuk mengambil keputusan dari hasil pengukuran sensor loadcell dengan batasan *range* dari 430-1270 gram.

**Tabel 1.** *Domain Range input berat*

No	Membership Functions Input	Range(g)
1	Ringan	[430 670]
2	Sedang	[400 700]
3	Berat	[730 1270]

Tabel 1 diatas merupakan *Domain range* input berat Kemudian untuk mengatur PWM pada heater serta kipas dc untuk pengeringan ikan asin otomatis dengan sistem fuzzy dapat mengatur suhu panas pada oven pemanas agar sesuai dengan setpoint. Berikut adalah domain range input suhu.


**Gambar 5.** domain range input suhu

Dari Gambar diatas dapat dijelaskan seperti pada Tabel berikut.

**Tabel 2.** domain range input suhu

No	Membership Functions Input	Range(°C)
1	Kurang Panas	[0.47 31.53]
2	Lumayan Panas	[19.88 54.82]
3	Sangat Panas	[44.16 75.84]

*Output* yang digunakan dalam penelitian ini merupakan nilai PWM (*pulse Width Modulation*) untuk mengatur suhu dalam ruang oven pengering ikan asin. *Membership Functions Output* PWM sebagai berikut :

1. Pemanas
  - a. *Low* = 65
  - b. *Medium* = 127
  - c. *High* = 255

#### IV. HASIL DAN PEMBAHASAN

Pada Hasil pembahasan ini menjelas mengenai data yang diambil dan dicantumkan berupa nilai-nilai. Berikut merupakan Tabel perbandingan output Pwm heater atau pemanas pada arduino ide dan fuzzy matlab.

**Tabel 3.** Perbandingan output Pwm pompa dc pada arduino ide dan fuzzy matlab

Suhu	Berat	Esp32	Matlab	Selisih	Error
		PWM Pemanas	PWM Pemanas		
30	1000	255	255	0	0%
45	700	121.25	121	0.25	0.20%
60	400	65	65	0	0%
<b>Rata-rata</b>					<b>0.06%</b>

Pada Tabel 3. merupakan hasil dari perbandingan kerja alat dan matlab dengan rata-rata nilai error 0.06%. Untuk mencari rata-rata error digunakan rumus sebagai berikut :

$$\text{rata-rata error} = \frac{\sum \text{Hasil Penjumlahan Error}}{\text{Banyaknya Data}}$$

$$\text{rata-rata error} = \frac{0+0.20+0}{3} = 0.06$$

contoh perhitungan *Fuzzy Sugeno* dengan *input* Suhu 60° dan Berat 400 gram.

dengan suhu =60° maka termasuk dalam kategori sangat panas, sedangkan berat 400 gram termasuk dalam kategori ringan. Untuk perhitungan kategorising sangat panas pada *input* Suhu berbentuk trapesium terpotong dimana nilai Suhu =60° berada pada *range*  $x > 58,24$  sehingga derajat keanggotaan kategori sangat panas =1. Untuk perhitungan kategori ringan pada *input* Berat berbentuk trapesium terpotong dengan nilai berat =400 gram berada pada *range*  $x < 430$  sehingga derajat keanggotaan ketegori ringan=1. Nilai derajat keanggotaan pada *input* Suhu :  $\mu$  kurang panas = 0

$\mu$  lumayan panas = 0

$\mu$  sangat panas = 1

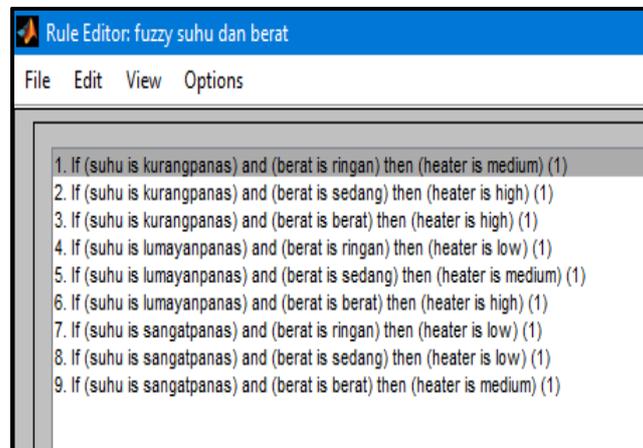
Nilai derajat keanggotaan pada *input* berat :

$\mu$  ringan = 1

$\mu$  sedang = 0

$\mu$  berat = 0

Adapun *rule base* pada *fuzzy* yang di rancang dengan *input* Suhu dan Berat serta *output* PWM *Heater* ditunjukkan pada Gambar dibawah ini.



**Gambar 6.** *Rule Base Fuzzy*

Pada setiap rule menggunakan fungsi min atau minimal dimana akan memilih nilai terkecil dari setiap rule yang telah dibuat.

$$Rule1 = \min(\mu_{kurangpanas}, \mu_{ringan}) = \min(0, 1) = 0$$

$$Rule2 = \min(\mu_{kurangpanas}, \mu_{sedang}) = \min(0, 0) = 0$$

$$Rule3 = \min(\mu_{kurangpanas}, \mu_{berat}) = \min(0, 0) = 0$$

$$Rule4 = \min(\mu_{lumayanpanas}, \mu_{ringan}) = \min(0, 1) = 0$$

$$Rule5 = \min(\mu_{lumayanpanas}, \mu_{sedang}) = \min(0, 0) = 0$$

$$Rule6 = \min(\mu_{lumayanpanas}, \mu_{berat}) = \min(0, 0) = 0$$

$$Rule7 = \min(\mu_{sangatpanas}, \mu_{ringan}) = \min(1, 1) = 1$$

$$Rule8 = \min(\mu_{sangatpanas}, \mu_{sedang}) = \min(1, 0) = 0$$

$$Rule9 = \min(\mu_{sangatpanas}, \mu_{berat}) = \min(1, 0) = 0$$

selanjutnya proses defuzzifikasi dengan rumus *fuzzy* sugeno.

PWM

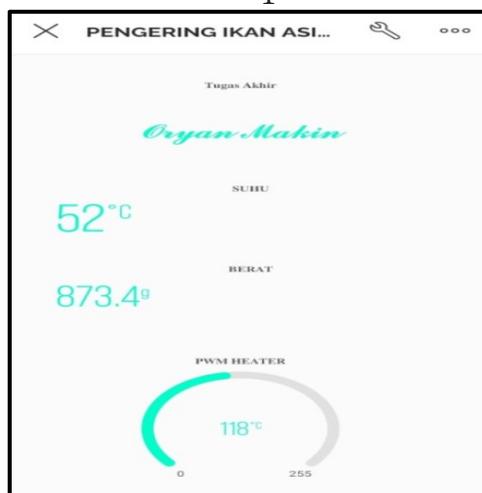
$$= \frac{(r1 * md) + (r2 * hg) + (r3 * hg) + (r4 * lw) + (r5 * md) + (r6 * hg) + (r7 * lw) + (r8 * lw) + (r9 * md)}{r1 + r2 + r3 + r4 + r5 + r6 + r7 + r8 + r9}$$

Dimana *output* PWM *Heater* berupa kategori *Low*=65, *Medium*=127 dan *High*=255.

PWM

$$= \frac{(0 * 127) + (0 * 255) + (0 * 255) + (0 * 65) + (0 * 127) + (0 * 255) + (1 * 65) + (0 * 65) + (0 * 127)}{0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0}$$

$$PWM = \frac{(0) + (0) + (0) + (0) + (0) + (0) + (65) + (0) + (0)}{0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0}$$
$$PWM = \frac{65}{1} = 65$$



**Gambar 7.** Tampilan Aplikasi *Blynk* pada *Smartphone*

pada Gambar diatas tampilan data pada aplikasi *Blynk* berupa Suhu, Berat dan PWM Heater.

## V. KESIMPULAN

Kesimpulan dari penelitian ini adalah metode fuzzy sugeno dapat diimplementasikan pada pengering ikan otomatis dengan sistem internet of things bekerja sangat baik dengan mempercepat proses pengeringan ikan dari waktu yang lama hingga berhari-hari menjadi 3-8 jam dengan kadar air 40% proses pengeringan ikan asin dapat dimonitoring dengan aplikasi Blynk melalui *android* atau *smartphone*.

## UCAPAN TERIMA KASIH

Ucapan Terima kasih kepada orang-orang yang telah membantu saya sehingga artikel ini dapat ditulis dengan baik dan benar.

## REFERENSI

1. Fitriyanti, L. (2021). Analisis Pendapatan Nelayan Dengan Sistem Pengeringan Ikan Asin Lendra Sebelum Covid-19 Di Pangandaran, Kelurahan Pananjung Kecamatan Pangandaran Kabupaten Pangandaran. *AGRISIA-Jurnal Ilmu-Ilmu Pertanian*, 14(1).
2. Yuliati, S., Kalsum, L., Junaidi, R., Fadarina, H. C., Azizah, R. R. R., Utami, W. A., & Ningrum, G. M. (2020). Rancang Bangun Tray Dryer Sistem Hybrid (Surya-Heater) Untuk Pengeringan Ikan Asin. *Kinetika*, 11(2), 10-18.
3. Lukman, M. F., Arifin, S., & Islamiyah, M. (2022). Rancang Bangun Alat Pengering Ikan Asin Otomatis Berbasis Arduino Uno. *Jurnal Ilmiah Teknologi Informasi Asia*, 16(1), 37-44.
4. Darmawan, R., Wiryawan, B., & Wahyuningrum, P. I. (2020). Status Sumberdaya Ikan Ekor Kuning (*Caesio Cuning*) pada Musim Peralihan di Perairan Karimunjawa: Suatu Pendekatan Menggunakan Spawning Potential Ratio. *ALBACORE Jurnal Penelitian Perikanan Laut*, 4(1), 21-32.
5. Gusdi, A. T., & Sipahutar, Y. H. (2021). Penerapan Sanitation Standard Operation Procedures (SSOP) dan Good Manufacturing Practice (GMP) dalam Pengolahan Fillet Ikan Ekor Kuning (*Caesio Cuning*) Beku. *PELAGICUS*, 2(3), 117-126.
6. Setiawan, A., & Purnamasari, A. I. (2019). Pengembangan Smart Home Dengan Microcontrollers ESP32 Dan MC-38 Door Magnetic Switch Sensor Berbasis Internet of



- Things (IoT) Untuk Meningkatkan Deteksi Dini Keamanan Perumahan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(3), 451-457.
7. Yudhanto, A., Sari, A. P., Rachman, N., & Subairi, S. (2020). Implementasi Sensor Thermocouple Berbasis Telemetry Untuk Mengukur Thermal Pembakaran Propelan Roket. *JASIEK (Jurnal Aplikasi Sains, Informasi, Elektronika dan Komputer)*, 2(1), 38-46.
  8. Mandayatma, E. (2018). Peningkatan Resolusi Sensor Load Cell Pada Timbangan Elektronik. *Jurnal Eltek*, 16(1), 37-50.
  9. Hidayat, F., & Krismadinata, K. (2019). Rancang bangun VVVF inverter 3 FASA untuk Operasi Motor Induksi Tiga fasa DENGAN Antarmuka Komputer. *INVOTEK: Jurnal Inovasi Vokasional Dan Teknologi*, 19(2), 47-56.
  10. Herlan, A., Fitri, I., & Nuraini, R. (2021). Rancang Bangun Sistem Monitoring Data Sebaran Covid-19 Secara Real-Time menggunakan Arduino Berbasis Internet of Things (IoT). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 5(2), 206-212.
  11. SARIFATULLAH, M. (2020). PERANCANGAN SISTEM PROTEKSI THERMAL PADA MOTOR INDUKSI 3 FASA BERBASIS KONTROL ARDUINO MENGGUNAKAN JARINGAN IoT. *Jurnal Online Mahasiswa (JOM) Bidang Teknik Elektro*, 1(1).
  12. Puspitasari, M. D. M., & Maulidina, M. (2022). Rancang Bangun Intensor (Induktor Heater) Menggunakan Thermal Sensor Berbasis Mikrokontroler Arduinodalam Mengolah Logam. *Nusantara of Engineering*, 5(1), 1-7

# Monitoring Kualitas Air Aquarium Menggunakan Metode Fuzzy Mamdani Berbasis IoT

Kornelius Setiawan Bu’u<sup>1</sup>, Nachrowie<sup>2</sup>, Elta Sonalitha<sup>3</sup>,

<sup>1, 2, 3</sup> Jurusan Teknik Elektro, Universitas Merdeka Malang

<sup>1</sup>[kobuu96@gmail.com](mailto:kobuu96@gmail.com)

<sup>2</sup>[nachrowie@unmer.ac.id](mailto:nachrowie@unmer.ac.id)

<sup>3</sup>[elta.sonalitha@unmer.ac.id](mailto:elta.sonalitha@unmer.ac.id)

Corresponding author email: [nachrowie@unmer.ac.id](mailto:nachrowie@unmer.ac.id)

**Abstract:** Aquarium is an aesthetic container as well as a place for guppy farming, there are several obstacles that are often experienced by aquarium fish farmers, namely in controlling water quality in this case is controlling the level of water turbidity and water pH in the aquarium. The level of turbidity of the water negatively affects the aesthetic value of the aquarium and the bioata in it. High turbidity levels can also affect the ability of fish gills to absorb dissolved oxygen and high turbidity levels cause the development of fungi or fungi *Branchiomyces Sanguinis* in fish insang and can produce moss in aquarium containers, causing fish health to be disrupted so that fish are susceptible to disease and even death. with the internet of things-based water quality monitoring system using the Blynk application can control pH values and turbidity via android or user smartphones such as pH value status and turbidity value as well as automatic water draining and filling status and the fuzzy mamdani method as a decision making support when draining seen from the pH value and turbidity value to produce a total or amount of water that must be drained 25%, 50% or 75% of the total amount of water in the aquarium.

**Keywords:** pH, Turbidity, Blynk, IoT, fuzzy Mamdani.

**Abstrak:** Aquarium adalah sebuah wadah estetik juga sebagai tempat pada pembudidayaan ikan guppy, terdapat beberapa kendala yang sering dialami oleh para pembudidaya ikan aquarium yaitu dalam mengontrol kualitas air dalam hal ini adalah pengontrolan tingkat kekeruhan air dan pH air pada aquarium pada rentang pH 6-8 atau netral. Tingkat Kekeruhan air dengan rentang 0-10 ntu selebihnya berdampak negatif terhadap nilai estetik aquarium serta bioata yang ada didalamnya. Tingkat *turbidity* yang tinggi juga dapat mempengaruhi kemampuan insang ikan menyerap oksigen terlarut dan tingkat kekeruhan yang tinggi menyebabkan berkembangnya jamur atau *fungi Branchiomyces Sanguinis* di insang ikan serta dapat menghasilkan lumut pada wadah akuarium sehingga menyebabkan kesehatan ikan terganggu sehingga ikan mudah terkena penyakit bahkan kematian. dengan adanya sistem monitoring kualitas air berbasis *internet of things* menggunakan aplikasi *Blynk* dapat mengontrol nilai pH dan *turbidity* melalui android atau smartphone pengguna seperti status nilai pH dan nilai turbidity serta status pengurusan dan pengisian air secara otomatis dan metode *fuzzy mamdani* sebagai pendukung pengambilan keputusan pada saat melakukan pengurusan dilihat dari nilai pH dan nilai *turbidity* untuk menghasilkan total atau banyaknya air yang harus dikuras 25%, 50% atau 75% dari total banyaknya air pada aquarium.

**Kata kunci:** pH, Turbidity, Blynk, IoT, fuzzy Mamdani

## I. PENDAHULUAN

Budidaya ikan hias pada aquarium telah menjadi hobi yang banyak digemari pada semua kalangan masyarakat terutama para pembudidaya ikan hias jenis guppy, kualitas air aquarium sangatlah penting untuk menunjang kehidupan ikan guppy diantaranya nilai pH dan tingkat kekeruhan air atau Turbidity.

Ikan guppy (*Poecilia reticulata*) merupakan ikan aquarium yang sangat populer dan digemari, ikan ini mempunyai tubuh yang cantik dan menawan warnanya yang sangat bervariasi dan menarik perhatian[1].

Mengatur dan mengontrol kualitas air dalam hal ini tingkat kekeruhan air dan pH air dalam aquarium bukanlah hal yang mudah untuk dilakukan karena Kekeruhan yang tinggi berpengaruh cukup besar terhadap ikan dan dinding aquarium. Penelitian ini dilakukan untuk memonitoring serta mengontrol kualitas air secara otomatis, meliputi pH dan kekeruhan air aquarium, insang sebagai



organ pernapasan ikan sangat peka terhadap kualitas air sehingga pengontrolan kualitas air bermanfaat untuk membatasi tumbuh kembangnya parasit serta jamur pada organ pernapasan ikan serta tumbuhnya lumut pada tank akuarium. Dengan adanya bakteri dapat menyebabkan kesehatan pada ikan terganggu hingga ikan akan mudah terserang penyakit[2][3].

Maka dari itu penelitian dilakukan untuk membuat suatu *prototype* yang berguna untuk mengatur konsentrasi *pH* dan tingkat kekeruhan pada air secara otomatis dan dapat dimonitoring. Mikrokontrolernya menggunakan *Esp32* dan sensornya menggunakan sensor *pH* air dan sensor *Turbidity* serta outputnya menggunakan pompa *pH up* dan pompa *pH down* dan pompa air pengisian dan pengurasan dengan sistem monitoring berbasis *IoT* menggunakan aplikasi *Blynk* untuk menampilkan nilai *pH* dan tingkat kekeruhan air atau *Turbidity* serta status pengurasan air dan pengisian air pada aquarium ditampilkan pada android atau smartphone pengguna dan metode *fuzzy mamdani* sebagai pendukung pengambilan keputusan pada saat melakukan pengurasan dan pengisian air berdasarkan input nilai *pH* dan *turbidity* yang telah disimulasikan melalui aplikasi *Matlab*.

## II. KAJIAN PUSTAKA

### Esp32

ESP32 adalah sebuah modul yang memiliki fungsi Wi-Fi untuk menghubungkan mikrokontroler dengan internet. Modul ini juga dilengkapi dengan port Micro Universal Serial Bus (USB) yang dapat digunakan untuk sumber daya dan pemrograman[2]. Modul ESP32 menggunakan tegangan catu daya sebesar 3.3V dan memiliki beberapa mode Wi-Fi, seperti Access Point, Station, dan Both. Modul ini terdiri dari beberapa bagian, termasuk prosesor, memori, dan pin GPIO yang jumlahnya bervariasi tergantung pada tipe ESP32. Kemampuan modul ini untuk bekerja secara mandiri mirip dengan mikrokontroler, sehingga tidak memerlukan mikrokontroler lain yang sejenis[3]

### Sensor pH

pH meter adalah sensor yang dapat menentukan konsentrasi ion hidrogen. Nilai pH 7 menunjukkan bahwa larutan atau cairan dalam keadaan netral. Jika nilai pH lebih rendah dari 7 menunjukkan larutan atau cairan bersifat asam, sebaliknya nilai pH yang lebih tinggi dari 7 menunjukkan bahwa larutan atau cairan tersebut bersifat basa. Untuk sensor pH ini skala yang digunakan adalah 1-14.[5]

### Sensor Turbidity

Sensor kekeruhan berfungsi untuk mendeteksi kekeruhan pada air atau zat cair dengan memanfaatkan sifat optik air akibat pantulan cahaya dan membandingkan cahaya yang dipantulkan dengan cahaya yang akan datang, kekeruhan adalah kondisi air yang tidak jernih dan disebabkan oleh partikel (padatan tersuspensi) yang biasanya tidak terlihat oleh mata manusia, mirip dengan gas di udara. Semakin banyak partikel dalam air, semakin tinggi kekeruhan air. Pada sensor kekeruhan, semakin tinggi tingkat kekeruhan air maka perubahan tegangan keluaran sensor [6].

### Pompa air 12 VDC

Pompa air elektrik ini biasa digunakan dalam berbagai aplikasi, termasuk penggunaan di akuarium. Dalam konteks akuarium, pompa air elektrik berfungsi untuk mengalirkan dan mengedarkan air di dalam akuarium. Hal ini penting untuk menjaga sirkulasi air yang baik, memastikan adanya oksigenasi yang cukup bagi ikan dan organisme akuatik lainnya, serta membantu menjaga kondisi air yang optimal.

### Fuzzy mamdani

Pada tahun 1975, Ebrahim Mamdani mengembangkan metode fuzzy yang disebut Fuzzy Mamdani secara matematis logika ini sehingga mudah untuk dimengerti. Proses pemecahan masalah dengan fuzzy logic Mamdani untuk menghasilkan solusi yang terbaik dan akurat melalui beberapa proses yaitu pembentukan himpunan fuzzy; implementasi fungsi implisit; komposisi aturan; Defuzzifikasi.

### Internet of Things

Blynk digunakan untuk mengontrol Arduino, ESP32, ESP8266 dan modul serupa melalui Internet. dan blynk juga termasuk dalam Aplikasi sederhana yang mudah dipelajari oleh pemula maupun tingkat

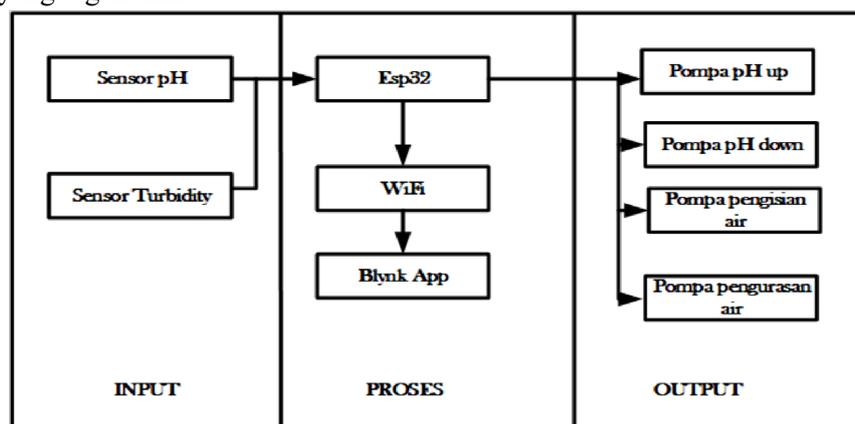
profesional, Blynk adalah sebuah aplikasi yang tidak bergantung dengan modul atau perangkat tertentu. Dari aplikasi ini, kita bisa mengontrol semuanya dari kejauhan, dimanapun posisi kita, selama kita memiliki koneksi internet [6]. Ini berbasis IoT (Internet of Things).

### III. METODE PENELITIAN

Metode Pada metode penelitian menjelaskan tentang perancangan alat dan variabel yang diteliti serta penerapan metode penentuan keputusan guna mendukung penelitian ini.

#### DIAGRAM BLOK

Pada perancangan alat diperlukan suatu blok diagram guna mempermudah untuk menjelaskan proses kerja dari alat yang digunakan.

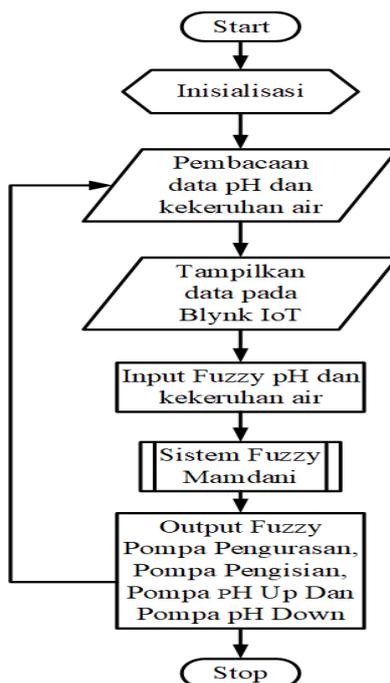


**Gambar 1.**Diagram blok

Sistem ini menggunakan dua jenis sensor, yaitu sensor pH dan sensor Turbidity, sebagai input. Data dari kedua sensor ini akan diproses oleh mikrokontroler. Mikrokontroler berfungsi untuk mengatur kerja output, seperti mengontrol pompa pH up yang berisi larutan asam dan pompa pH down yang berisi larutan basa, guna mengatur tingkat konsentrasi pH yang terlarut dalam air. Selain itu, jika tingkat kekeruhan air melebihi batas normal, mikrokontroler juga akan mengatur pompa air untuk mengatur sirkulasi air. Sistem ini juga dilengkapi dengan fitur monitoring yang memungkinkan pengguna untuk memantau sistem melalui aplikasi Blynk pada perangkat Android. Aplikasi ini memungkinkan pengguna untuk terhubung dengan mikrokontroler dan mendapatkan informasi tentang kondisi air, seperti tingkat pH dan kekeruhan. Dengan demikian, pengguna dapat melakukan pemantauan dan pengaturan sistem secara praktis melalui aplikasi Blynk.

#### FLOWCHART

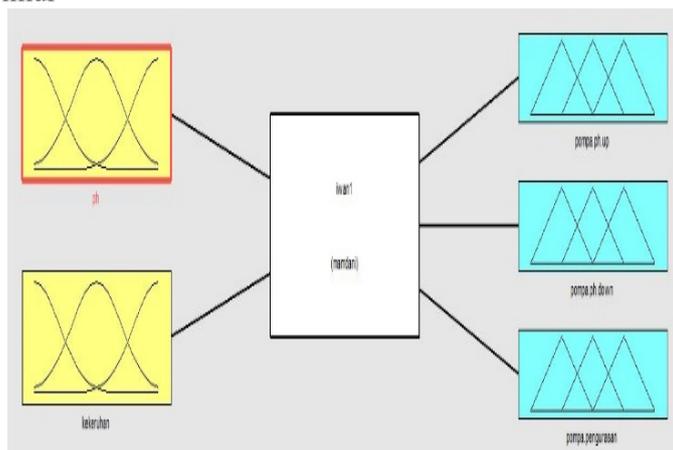
*flowchart* adalah penjelasan singkat tentang bagaimana suatu sistem bekerja dari awal hingga akhir

**Gambar 2.***flowchart*

Flowchart di atas menggambarkan proses awal sistem yang melibatkan inisialisasi port dan sistem yang terhubung ke mikrokontroler ESP32. Setelah inisialisasi, mikrokontroler akan membaca data analog dari sensor pH dan sensor Turbidity. Data yang diterima dari sensor pH akan menjadi acuan untuk proses selanjutnya, yang akan diolah menggunakan metode logika kabur (fuzzy logic) atau fuzzy Mamdani. Setelah itu, sistem akan melakukan evaluasi berdasarkan data dari sensor Turbidity. Jika tingkat kekeruhan air mencapai target atau melebihi batas normal, maka proses pengurasan air akan dilakukan. Pada proses ini, pompa pengurasan akan diaktifkan untuk menguras air dari akuarium. Setelah proses pengurasan selesai, pompa pengisian air akan diaktifkan untuk mengisi air kembali ke dalam akuarium. Setelah proses pengisian selesai, sistem akan kembali berjalan seperti semula dan memulai siklus monitoring dan pengendalian air berdasarkan kondisi yang terdeteksi. Flowchart tersebut memberikan gambaran visual tentang urutan langkah-langkah yang terjadi dalam sistem ini, mulai dari inisialisasi hingga pengendalian air berdasarkan data sensor dan metode logika fuzzy mamdani yang digunakan.

### **FUZZY MAMDANI**

Metode *fuzzy mamdani* sebagai pendukung pengambilan keputusan dalam sistem pengontrolan kualitas air dengan mengelola inputan dari tiap sensor yang kemudian diproses lebih lanjut sesuai aturan *fuzzy* untuk menggerakkan output berupa pompa air pengurasan dan pompa *pH up* serta *pH down*. Pengujian ini dibuat menggunakan *software matlab*. Membership function input berfungsi sebagai untuk menghitung derajat keanggotaan dari masing-masing variabel pada input.



**Gambar 3.** membership functions input dan output pada membership functions input terbagi atas 2 yaitu pH dan kekeruhan terbagi menjadi 3 bagian yaitu bersih, keruh dan keruh sekali

**Tabel 1.** Domain Range input kekeruhan

no	Membership functions input	Range (ntu)
1	Bersih	0-10
2	Keruh	8-23
3	Keruh sekali	20-26

Dari Tabel diatas dapat dinyatakan dalam persamaan sebagai berikut:

$$\mu_{bersih}[x] = \begin{cases} 0; & x \geq 10 \\ \frac{10-x}{10-5}; & 5 < x < 10 \\ 1; & 0 \leq x \leq 5 \end{cases}$$

$$\mu_{keruh}[x] = \begin{cases} 0; & x \leq 8 \text{ atau } x \geq 23 \\ \frac{x-8}{15-8}; & 8 < x < 15 \\ 1; & x = 15 \\ \frac{23-x}{23-15}; & 15 < x < 23 \\ 0; & x \geq 23 \end{cases}$$

$$\mu_{sangat\ keruh}[x] = \begin{cases} 0; & x \leq 20 \\ \frac{20-x}{26-20}; & 20 < x < 26 \\ 1; & x \geq 26 \end{cases}$$

Domain range dari membership functions input kekeruhan sebagai batasan nilai yang akan digunakan pada mikrokontroler untuk mengambil keputusan dari hasil pembacaan sensor turbidity dan Domain Range Input pH terbagi atas 5 variabel seperti pada tabel 2

**Tabel 2.** Domain Range sensor pH

No	Membership functions input	Range pH
1	Sangat asam	[0 4.5]
2	Asam	[4 6]
3	ideal	[5.5 7.5]
4	Basa	[7 9]
5	Sangat basa	[9.5 14]

Dari Tabel diatas dapat dinyatakan pada persamaan berikut :

$$\mu_{\text{sangat asam}}[x] = \begin{cases} 0; & x \geq 4.5 \\ \frac{4.5 - x}{2.16 - 0.2}; & 0.2 < x < 2.16 \\ 1; & 0 \leq x \leq 0.2 \end{cases}$$

$$\mu_{\text{asam}}[x] = \begin{cases} 0; & x \leq 4 \text{ atau } x \geq 6 \\ \frac{x - 4}{4.5 - x}; & 4 < x < 5 \\ 1; & x = 5 \\ \frac{6 - x}{6 - 5}; & 5 < x < 6 \\ 0; & x \geq 6 \end{cases}$$

$$\mu_{\text{ideal}}[x] = \begin{cases} 0; & x \leq 5.5 \\ \frac{x - 5.5}{6 - 5.5}; & 5.5 < x < 6 \\ 1; & 5.8 \leq x \leq 6 \\ \frac{7.5 - x}{7.5 - 7}; & 7 < x < 7.5 \\ 0; & x \geq 7.5 \end{cases}$$

$$\mu_{\text{basa}}[x] = \begin{cases} 0; & x \leq 7 \\ \frac{8 - x}{8 - 7}; & 7 < x < 8 \\ 1; & x = 8 \\ \frac{x - 7}{8 - 7}; & 7 < x < 8 \end{cases}$$

$$\mu_{\text{sangat basa}}[x] = \begin{cases} 0; & x \leq 9.5 \\ \frac{9.5 - x}{9.5 - 9}; & 9.5 < x < 14 \\ 1; & 9.5 \leq x \leq 14 \end{cases}$$

### Membership function output

Output yang dilakukan dalam penelitian ini terbagi atas 2 yaitu output pwm dan output waktu, Output pwm (pulse width modulation) digunakan untuk mengontrol kecepatan putar pompa Ph up dan pompa Ph down, sedangkan output waktu digunakan untuk pompa pengurasan air. Membership functions output pompa ph down terdiri dari sedikit, sedang dan lama seperti

**Tabel 3.** *Domain Range output pwm*

No	Membership functions input	Pwm
1	Mati	0-60
2	Sedang	50-160
3	Cepat	150-255

Tabel 3 merupakan domain range output pompa pH up dan pompa pH down Dinyatakan pada persamaan berikut :

$$\mu_{\text{mati}}[x] = \begin{cases} 0; & x \geq 60 \\ \frac{60-x}{60-30}; & 30 < x < 60 \\ 1; & 0 \leq x \leq 30 \end{cases}$$

$$\mu_{\text{sedang}}[x] = \begin{cases} 0; & x \leq 50 \text{ atau } x \geq 160 \\ \frac{50-x}{70-50}; & 50 < x < 70 \\ 0; & 50 \leq x \leq 70 \\ \frac{70-x}{140-70}; & 70 < x < 140 \\ 1; & 70 \leq x \leq 140 \end{cases}$$

$$\mu_{\text{cepat}}[x] = \begin{cases} 0; & x \leq 150 \text{ atau } x \geq 255 \\ \frac{150-x}{165-150}; & 150 < x < 255 \\ 1; & 165 \leq x \leq 255 \end{cases}$$

Membership functions output waktu pompa pengurasan adalah Membership functions output pwm pompa pengurasan berfungsi untuk mengalirkan air dari aquarium menuju ke tempat pembuangan dengan domain range yang berbeda dari setiap membership functions domain range dari output waktu pompa pengurasan seperti pada Tabel 4

**Tabel 4.** *Domain Range output waktu pompa pengurasan*

No	Membership functions input	waktu
1	Tidak ada	0
2	Cepat	5-200
3	Lama	190-432

Dari domain range output waktu pompa pengurasan dapat Dinyatakan pada persamaan berikut :

$$\mu_{\text{tidak ada}}[x] = 0;$$

$$\mu_{\text{cepat}}[x] = \begin{cases} 0; & x \leq 5 \text{ atau } x \geq 200 \\ \frac{200-x}{200-5}; & 5 < x < 200 \\ 1; & 5 \leq x \leq 150 \end{cases}$$

$$\mu_{\text{lama}}[x] = \begin{cases} 0; & x \leq 190 \text{ atau } x \geq 432 \\ \frac{190-x}{220-190}; & 190 < x < 220 \\ 1; & 220 \leq x \leq 432 \end{cases}$$



Pompa pengisian air akan bekerja bergantian dengan pompa pengurasan air untuk mengalirkan sirkulasi air

#### IV. HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan didapat dari beberapa percobaan alat sesuai dengan beberapa kondisi air dilihat dari nilai pH dan tingkat kejernihan air yang mendapatkan respon kerja alat seperti pada Tabel

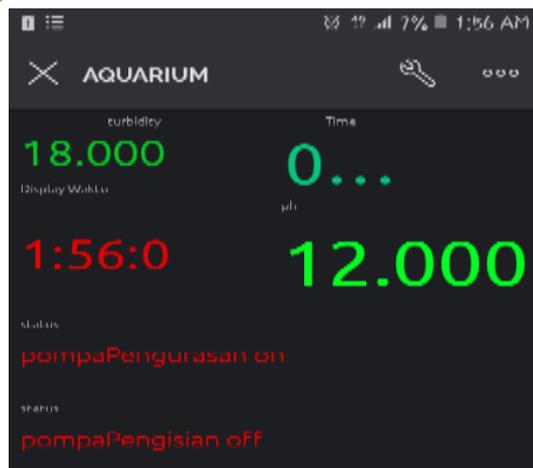
**Tabel 5.** Respon kerja alat

kondisi	Nilai ph	Nilai turbidity	Respon alat
bersih	7	1	Pompa ph up off Pomp ph down off Pompapengurasan off Pompa pengisian off
Air keruh	7	14	Pompa ph up off Pomp ph down off Pompapengurasan on Pompa pengisian on
Air bersih	3	1	Pompa ph up on Pomp ph down off Pompapengurasan off Pompa pengisian off
Air bersih	12	1	Pompa ph up off Pomp ph down on Pompapengurasan off Pompa pengisian off

Keterangan

1. buffer pH 6,8 akan dialirkan oleh pompa pH up
2. buffer pH 4,0 akan dialirkan oleh pompa pH down

Sistem akan bekerja dan mengirimkan data seperti tampilan pada aplikasi Blynk pada android pengguna



**Gambar 4.** Tampilan pada Aplikasi *Blynk*

Gambar 6. Menunjukkan tampilan nilai turbidity adalah 18 dan nilai pH adalah 12 dan pada kondisi ini pompa pengurusan akan bekerja dengan status pompa pengurusan on dan pompa pengisian off



**Gambar 5.** Tampilan akuarium

Pada akurium terdapat lampu sebagai tambahan aksesoris dan terdapat 2 buah wadah berwarna merah untuk menampung air bersih dan air kotor hasil pengurusan serta 2 botol air yang berisi buffer pH up dan buffer pH down

## V. KESIMPULAN

sensor-sensor yang digunakan memiliki tingkat keakuratan yang memadai dalam membaca tingkat kekeruhan air dan pH. Sistem monitoring menggunakan aplikasi Blynk juga efektif dalam memantau dan mengendalikan sistem secara real-time. Penggunaan metode Fuzzy Mamdani membantu dalam pengambilan keputusan yang lebih cerdas berdasarkan data yang diperoleh dari sensor-sensor tersebut.

## UCAPAN TERIMA KASIH

Terimakasih untuk semua pihak yang telah membantu dalam penelitian ini sehingga dapat terselesaikan dengan baik kiranya dapat bermanfaat bagi pembudidaya ikan hias pada aquarium dan masyarakat umum.

## REFERENSI

1. Santoso, B., & Arfianto, A. D. (2019). Sistem pengganti air berdasarkan kekeruhan dan pemberi pakan ikan pada akuarium air tawar secara otomatis berbasis mikrokontroler atmega 16. *Jurnal Ilmiah Teknologi dan Informasi ASIA* Vol, 8(2).
2. Budi Prijo Sembodo and Novendra Geofanda Pratama, “Smart Aquarium Based Microcontroller”, *Journal of Applied Electrical & Science Technology - University of PGRI Adi Buana Surabaya*, Vol. 03, No. 2, 2021.



3. Daniel Patricko Hutabarat, Rudy Susanto, Bryan Prasetya, Barry Linando, Senanayake Mudiyansele Namal Arosha, “Smart system for maintaining aquascape environment using internet of things based light and temperature controller”, *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 12, No. 1, February 2022, pp. 896~902 ISSN: 2088-8708, DOI: 10.11591/ijece.v12i1.pp896-902.
4. Mohammad Fahmi Suhaimi, Nurul Huda Mat Tahir, Safuan Naim Mohamad, Suzanna Ridzuan Aw, “IoT Based Automatic Aquarium Monitoring System for Freshwater Fish”, *International Journal of Synergy in Engineering and Technology*, Vol. 2 No. 1 (2021) 125-133.
5. Wahyuni Eka Sari , Eko Junirianto , Geofani Fatur Rahman, 2021. “Sistem Pengukuran PH, Kelembapan, dan Suhu Berbasis Internet of Things (IoT)”, *Teknologi Rekayasa Komputer, Politeknik Negeri Samarinda, Samarinda, Indonesia*.
6. Mohammad, S. A., NikhilaChowdary, D. S., & Jebakumar, R. (2022). An Smart Aquaponic System Using IoT. *Journal of Positive School Psychology*, 226-235.
7. Muhamad Fauzan Zarkashie, 2021. “Rancang Bangun Sistem Pengukuran Kualitas Air Untuk Keperluan Higiene Sanitasi Berbasis Arduino Uno”, *Program Studi Fisika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta*.
8. Hutabarat, D. P., Susanto, R., Prasetya, B., Linando, B., & Senanayake, S. M. N. (2022). Smart system for maintaining aquascape environment using internet of things based light and temperature controller. *International Journal of Electrical and Computer Engineering*, 12(1), 896.
9. Pratama, D. R., Wijayanti, H., & Yulianto, H. (2018). Pengaruh warna wadah pemeliharaan terhadap peningkatan intensitas warna ikan guppy (*Poecilia reticulata*). *e-Jurnal Rekayasa dan Teknologi Budidaya Perairan*, 7(1), 775-782.
10. Pratama, D. R., Wijayanti, H., & Yulianto, H. (2018). Pengaruh warna wadah pemeliharaan terhadap peningkatan intensitas warna ikan guppy (*Poecilia reticulata*). *e-Jurnal Rekayasa dan Teknologi Budidaya Perairan*, 7(1), 775-782.
11. Mulyono, S., Qomaruddin, M., & Anwar, M. S. (2018). Penggunaan Node-RED pada Sistem Monitoring dan Kontrol Green House berbasis Protokol MQTT. *TRANSISTOR Elektro Dan Informatika*, 3(1), 31-44.



# Penerapan Database Relasional pada Perpustakaan Berbasis Website dan Kontrol Akses

Sandria Amelia Putri<sup>1</sup>, Ahmad Wafi Faturrahman<sup>2</sup>, Meisya Vira Amelia<sup>3</sup>, Alya Setya Paramita<sup>4</sup>, Farhan Dequika P.P.A.<sup>5</sup>, Kartika Maulida Hindrayani<sup>6</sup>

Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[21083010005@student.upnjatim.ac.id](mailto:21083010005@student.upnjatim.ac.id), <sup>2</sup>[21083010011@student.upnjatim.ac.id](mailto:21083010011@student.upnjatim.ac.id), <sup>3</sup>[21083010018@student.upnjatim.ac.id](mailto:21083010018@student.upnjatim.ac.id),

<sup>4</sup>[21083010046@student.upnjatim.ac.id](mailto:21083010046@student.upnjatim.ac.id), <sup>5</sup>[21083010062@student.upnjatim.ac.id](mailto:21083010062@student.upnjatim.ac.id), <sup>6</sup>[kartika.maulida.ds@upnjatim.ac.id](mailto:kartika.maulida.ds@upnjatim.ac.id)

Corresponding author email: [21083010005@student.upnjatim.ac.id](mailto:21083010005@student.upnjatim.ac.id)

**Abstract:** Leakage and misuse of someone's personal data is basically due to a lack of security from a database. Each database must have a controller to be able to access the database. This control system can be carried out by admins and admins who will provide access to anyone who can access the database so that not just anyone accesses it<sup>1</sup>. Just like control and access to the online library application that our group made, the admin can see the entire database, from the list of library members to the list of available books, and these databases can only be accessed and edited by the admin. This implementation uses a method in the form of SLDC (Software Development Life Cycle). So that the results obtained are that there are two users named admin and members. Admin has the power to access and provide database access, while members can only access a list of available books and borrowing history. This is done so that other than the admin, not just anyone accesses personal data from someone and misuses it.

**Keywords:** Access Control, Online Libraries, Relational Databases, WEB Applications

**Abstrak:** Kebocoran dan penyalahgunaan data pribadi seseorang pada dasarnya karena kurangnya keamanan dari suatu database. Setiap database pasti terdapat pengontrol untuk bisa mengakses database tersebut. Sistem kontrol ini dapat dilakukan oleh admin dan admin yang akan memberikan akses kepada siapa saja yang bisa mengakses database tersebut agar tidak sembarangan orang mengaksesnya. Sama seperti kontrol dan akses pada aplikasi perpustakaan online yang kelompok kami buat yaitu admin bisa melihat keseluruhan database, mulai dari daftar anggota perpustakaan hingga daftar buku yang tersedia dan database-database ini hanya bisa diakses dan diedit oleh admin. Pada implementasi ini digunakan metode berupa SLDC (Software Development Life Cycle). Sehingga didapatkan hasil yaitu terdapat dua user yang bernama admin dan anggota. Admin berkuasa untuk mengakses dan memberikan akses database sedangkan anggota hanya dapat mengakses list dari buku yang tersedia dan riwayat meminjamannya. Hal ini dilakukan agar selain admin tidak sembarang orang mengakses data pribadi dari seseorang dan menyalahgunakannya.

**Kata kunci:** Kontrol Akses, Perpustakaan Online, Database Relasional, Aplikasi WEB

## I. PENDAHULUAN

Perpustakaan merupakan wadah penting dalam mempelajari ilmu pada masyarakat. Fungsi utamanya adalah menyediakan akses ke berbagai sumber informasi, termasuk buku, jurnal, majalah, dan materi audiovisual. Perpustakaan juga berperan dalam mendukung pendidikan [1] [2], penelitian, pengembangan budaya, dan hiburan [3]. Proses bisnis pada perpustakaan biasanya adalah pengelolaan buku, total buku, denda peminjaman, pengkategorian buku seperti buku mata pelajaran, buku cerita, dan lain-lain [4]. Sebelum terdapat dukungan teknologi, pengelolaan perpustakaan tersebut dengan menggunakan lisan, maupun catatan di kertas sehingga dapat disimpulkan belum terautomasi [5]. Hal ini mengakibatkan kesalahan-kesalahan seperti catatan yang hilang, maupun aset perpustakaan yang tidak terdata atau hilang.

Dengan adanya dukungan teknologi perpustakaan dapat diakses secara online sehingga memudahkan dalam penyimpanan data dan informasi [6]. Keamanan dari *database* harus benar-benar



diperhatikan [7]. Jika tidak, dapat terjadi berbagai hal yang tidak diinginkan seperti kebocoran dan penyalahgunaan data. *Third party application* juga dapat mengetahui data seorang pengguna jika keamanan *database* tersebut tidak aman. Setiap *database* pasti terdapat pengontrol untuk bisa mengakses *database* tersebut. Sistem kontrol ini dapat dilakukan oleh admin dan admin yang akan memberikan akses kepada siapa saja yang bisa mengakses *database* tersebut agar tidak sembarangan orang mengaksesnya [8].

Penelitian terdahulu oleh Triyono dkk [9] membahas mengenai sistem keamanan sebuah gedung yang dibuat berbasis WEB. Sistem dapat membantu manusia dengan menjaga keamanan bangunan pada kondisi kritis. Kebakaran, pencurian, dan bencana alam merupakan beberapa hal yang dapat menimbulkan ketidaknyamanan dan kerusakan pada gedung. Pada jurnal lain yang ditulis oleh Lestari dkk [10] mengenai Pendidikan online yang terdesentralisasi dengan manajemen control akses berbasis blockchain. Penyimpanan hak akses akan digantikan dengan teknologi *blockchain*. *Blockchain* merupakan pengimplementasian dari *database* terdistribusi. Keunggulan dari *blockchain* yaitu pada pencegahan serangan *SQL Injection*. Teknik *hash* untuk keamanan penyimpanan data digunakan oleh *Blockchain* alih-alih *SQL*.

Pada artikel ini, penulis membahas mengenai kontrol dan akses untuk aplikasi perpustakaan online. Perpustakaan online menyediakan berbagai macam buku digital. Admin memiliki kewenangan mengakses, memberikan akses, serta mengamankan data pribadi pengguna seperti biodata, buku yang sedang disewa, *wishlist* buku pengguna, dan data lainnya dengan menggunakan metode SLDC (*Software Development Life Cycle*). Tujuannya agar *third party application* tersebut tidak bisa menyalahgunakan data pribadi pengguna. Sehingga, diharapkan pengguna merasa bahwa perpustakaan online ini benar-benar aman dari kebocoran data.

## II. METODE PENELITIAN

Studi literatur dari beberapa artikel terkait kontrol akses *database* dengan pengambilan kata kunci (kontrol akses) dilaksanakan pada tahap metode penelitian. Dalam studi literatur ditemukan beberapa hal yang menyebabkan kebocoran data yang dipengaruhi kontrol akses yang belum tepat penggunaannya. maka dari itu, beberapa teori yang relevan dikaji sehingga mendapatkan bahan-bahan yang tepat untuk diterapkan pada implementasi *database* relasional pada perpustakaan online ini. PHP dan MySQL dipilih dalam pengimplementasian karena mendukung situs web dinamis [11].

Selain itu, metode SLDC (*Software Development Life Cycle*) dalam pembuatan *website*. Tahap-tahap dalam SLDC adalah sebagai berikut:

### 1. Perencanaan (*Planning*)

Pada tahap ini, kami membuat perencanaan dari suatu permasalahan pada kontrol akses pada perpustakaan online. Seperti dalam pembuatan *website* yang akan memudahkan dalam mengelola perpustakaan.

### 2. Menganalisis Kebutuhan (*Require Analysis*)

Pada tahap ini, kami menganalisis apa saja yang dibutuhkan untuk seorang *user* (anggota perpustakaan) apabila ingin meminjam buku pada perpustakaan, seperti mendaftar sebagai anggota perpustakaan.

### 3. Mendesain Sistem (*Design System*)

Apabila semua telah mendapatkan gambaran mengenai sistem yang akan dibuat, maka langkah selanjutnya adalah mendesain sistem sesuai kebutuhan. Desain sistem yang dibutuhkan di sini adalah *database* yang dapat menampung data dari seorang anggota perpustakaan beserta admin yang akan mengelola *database* tersebut.

### 4. Mengimplementasikan Sistem (*Implementation System*)

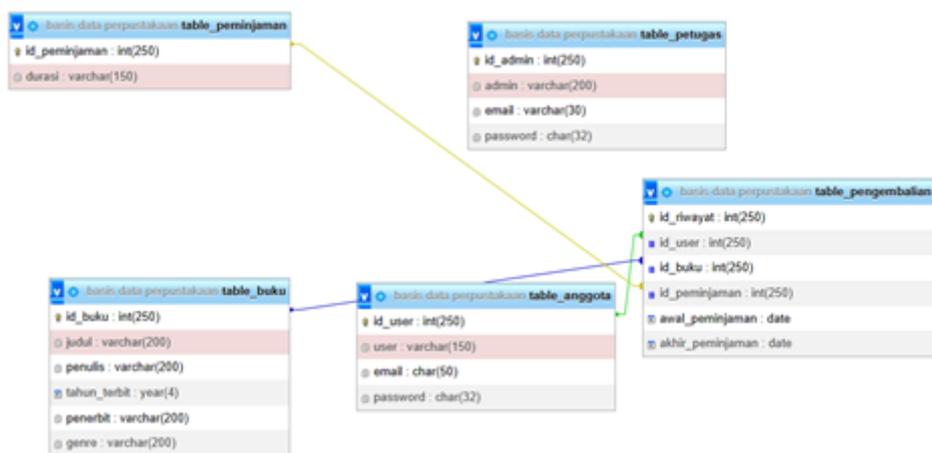
Tahapan ini merupakan tahapan di mana sistem dari aplikasi perpustakaan *online* telah selesai dibangun, diuji coba, dan dapat bekerja secara optimal. *User* yang terdiri dari admin dan anggota perpustakaan sudah menggunakannya secara real.

### 5. Perawatan Sistem (*Maintenance System*)

Apabila sistem sudah terimplementasikan dengan baik, maka sistem tersebut perlu adanya perawatan atau pemeliharaan guna untuk memeriksa data yang sudah tidak terpakai serta memeriksa apakah terjadi masalah *bug* pada sistem tersebut.

## III. HASIL DAN PEMBAHASAN

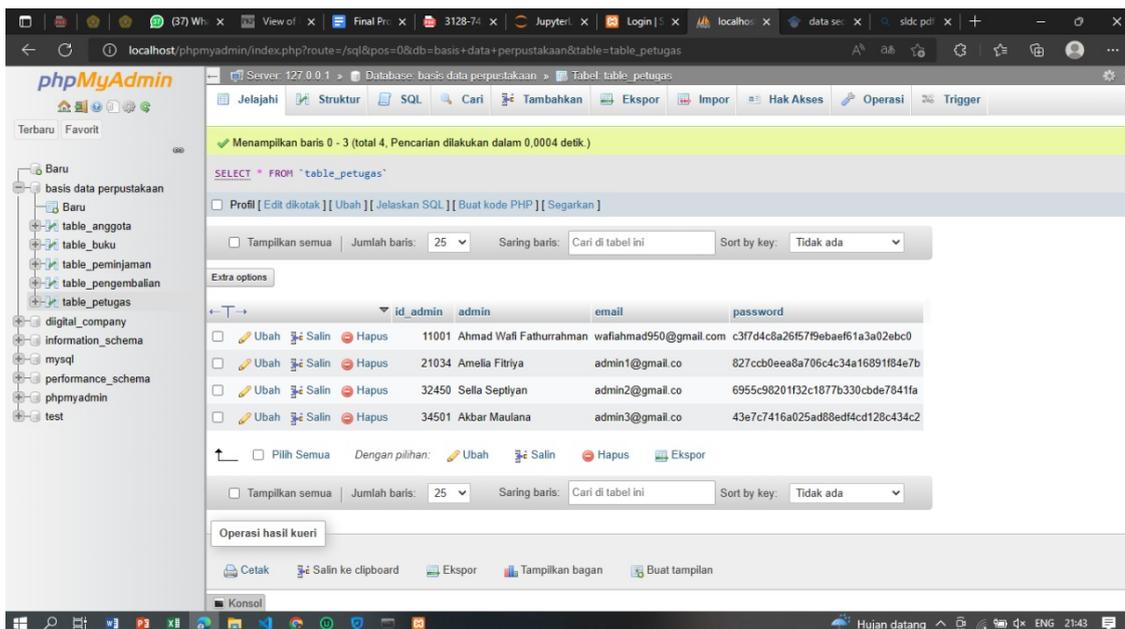
### 3.1. *Relational Table*



Gambar 1 Tampilan *Relational Table*

Pada perancangan relasional tabel terdiri dari lima tabel. Lima tabel yang akan digunakan yakni tabel anggota, tabel buku, tabel peminjam, tabel pengembalian dan tabel petugas.

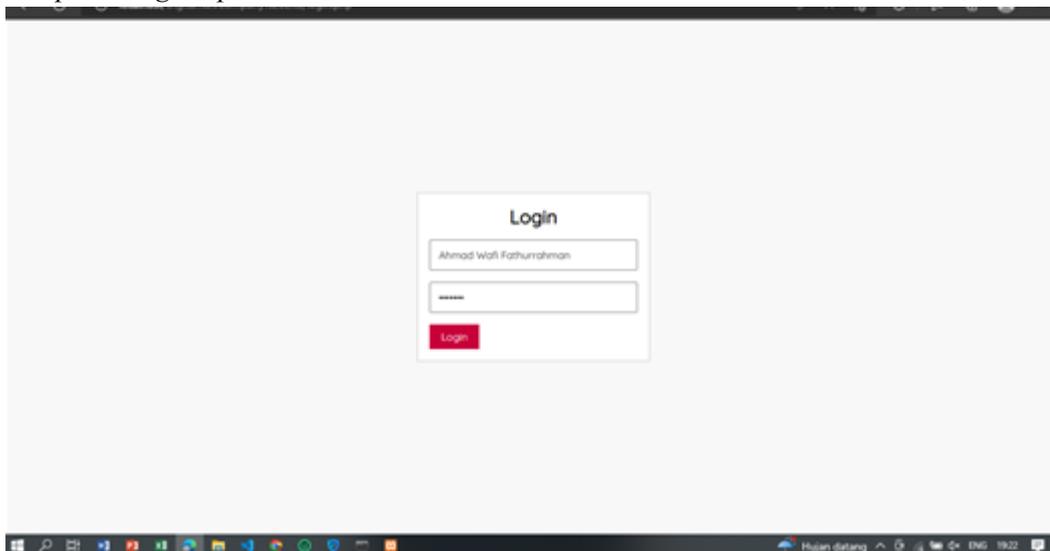
### 3.2. *Kontrol Akses*



Gambar 2 Tampilan Data pada *MySQL*

Ketika login dengan menggunakan akun-akun tersebut, maka kita bisa mengakses *database* (web) sebagai admin. Ketika memasuki web sebagai admin, maka kita bisa melihat keseluruhan *database*, mulai dari daftar anggota perpustakaan hingga daftar buku yang tersedia. *Database-database* ini hanya bisa diakses oleh admin. Bahkan admin juga memiliki kekuasaan untuk mengedit *database* tersebut.

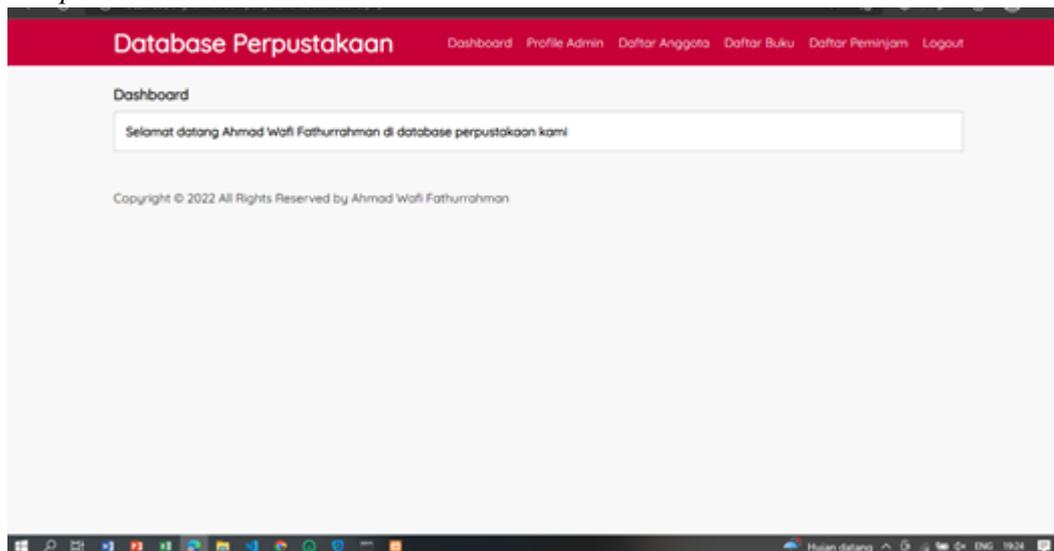
### 3.3. Tampilan Login Aplikasi



**Gambar 1** Tampilan *Login* Aplikasi

Pada saat aplikasi diakses, halaman yang muncul adalah menu *login*. Fitur *login* berfungsi sebagai filter awal untuk masuk ke dalam sistem *database*. Dengan *password* yang bersifat rahasia, maka menu ini dapat menjaga sistem aplikasinya agar user lain tidak dapat masuk ke akun yang tidak seharusnya.

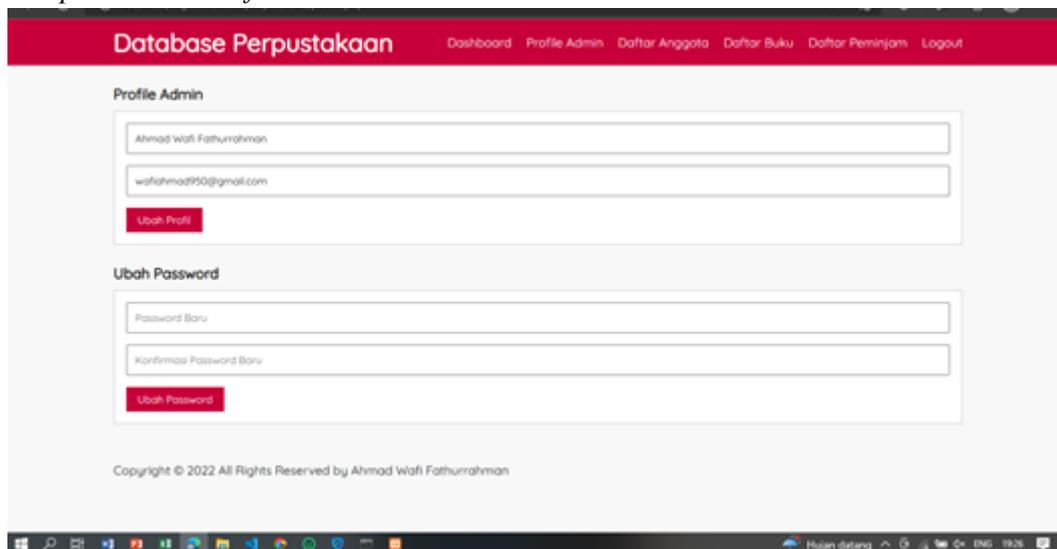
### 3.4. Tampilan Menu Utama



**Gambar 2** Tampilan Menu Utama

Setelah *login* berhasil, maka aplikasi akan memunculkan ucapan selamat datang kepada nama admin dan beberapa menu untuk mengelola *database* perpustakaan. Dalam menu tersebut disajikan

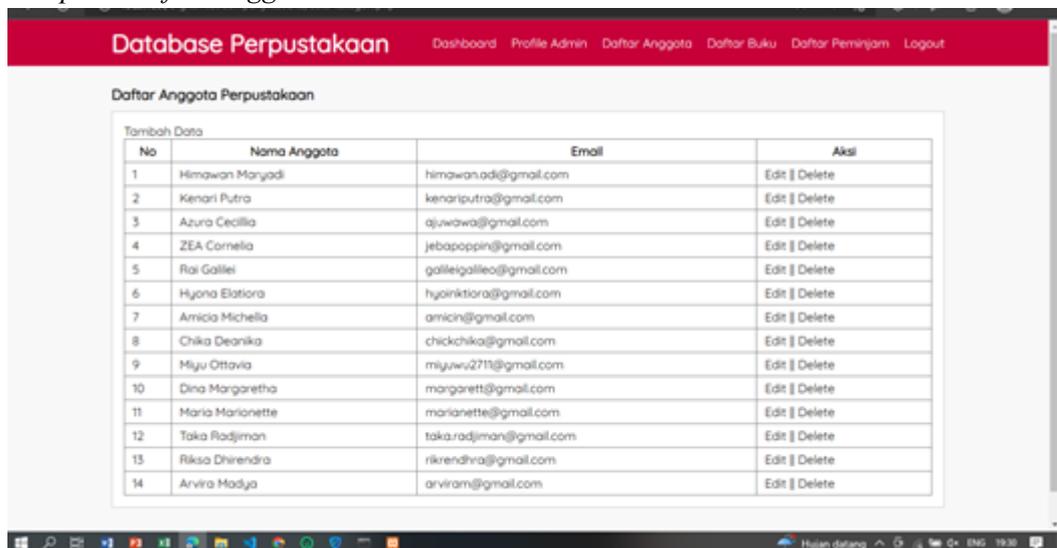
### 3.5. Tampilan Menu Profile Admin



**Gambar 3** Tampilan Menu *Profile* Admin

Pada menu *profile* admin, disajikan nama dan email dari admin. Selain itu, terdapat fitur untuk mengubah profil dan *password* apabila admin tersebut ingin menggantinya.

### 3.6. Tampilan Daftar Anggota



**Gambar 4** Tampilan Daftar Anggota

Pada menu daftar anggota kita disajikan data dari anggota perpustakaan tersebut, seperti nama dan email. Admin di sini diberi fitur untuk menambah, mengubah dan menghapus data anggota perpustakaan apabila diperlukan.

### 3.7. Tampilan Daftar Buku

No	Judul Buku	Nama Penulis	Tahun Terbit	Penerbit	Genre	Aksi
1	Chairil	Hasan Aspahani	2016	Gagas Media	Biografi	Edit   Delete
2	Filosofi Teras	Henry Manampiring	2018	Kompas	Filsafat	Edit   Delete
3	Aku Ini Bintang Jolang	Chairil Anwar	1986	Gramedia	Puisi	Edit   Delete
4	Hujan	Tere Lije	2016	Gramedia	Novel	Edit   Delete
5	Koala Kumal: (Edisi Revisi)	Raditya Dika	2015	Gagas Media	Autobiografi	Edit   Delete
6	Analisis Regresi untuk Penelitian	Prof. Dr. Sugono, M.Si	2018	Deepublish	Literatur	Edit   Delete
7	Bumi Manusia	Pramoedya Ananta Toer	1980	Hasta Mitra	Novel	Edit   Delete
8	Aljabar Linear Elementer	Howard Anton, Chris Forres	2004	Erlangga	Literatur	Edit   Delete
9	Sirkus Pohon	Andrea Hirata	2017	Bentang Pustaka	Novel	Edit   Delete
10	Laskar Pelangi	Andrea Hirata	2005	Bentang Pustaka	Novel	Edit   Delete

Gambar 5 Tampilan Daftar Buku

Pada menu daftar buku, kita juga disajikan data dari buku yang terdapat pada perpustakaan dengan rincian judul buku, nama penulis, tahun terbit, penerbit dan genre. Sama seperti sebelumnya, admin dapat menambah, mengedit dan menghapus data apabila diperlukan<sup>3</sup>.

### 3.8. Tampilan Daftar Peminjaman

No	Peminjam	Id_Buku	Id_Peminjaman	Tanggal Peminjaman	Tanggal Pengembalian	Aksi
1	ZEA Cornelia	83430	140	2022-11-20	2022-12-02	Edit   Delete
2	Chika Deanika	17856	707	2022-11-07	2022-11-13	Edit   Delete
3	Himawan Marjadi	15403	140	2022-11-01	2022-11-14	Edit   Delete
4	Rai Galilei	20134	140	2022-10-15	2022-10-29	Edit   Delete
5	Amicia Michella	18150	707	2022-10-13	2022-10-19	Edit   Delete
6	Himawan Marjadi	12826	301	2022-10-10	2022-11-10	Edit   Delete
7	Anvira Madya	31301	140	2022-10-05	2022-10-18	Edit   Delete
8	Kenari Putra	61875	140	2022-09-30	2022-10-13	Edit   Delete
9	Azura Cecilia	11876	140	2022-09-27	2022-10-11	Edit   Delete
10	ZEA Cornelia	11543	140	2022-09-20	2022-10-13	Edit   Delete
11	Rai Galilei	83430	140	2022-09-15	2022-09-29	Edit   Delete
12	Hjona Elatiara	17856	707	2022-09-10	2022-09-16	Edit   Delete
13	Riksa Dhirendra	15403	140	2022-09-02	2022-09-15	Edit   Delete
14	Chika Deanika	20134	707	2022-09-01	2022-09-07	Edit   Delete
15	Miju Ottavia	18150	301	2022-09-01	2022-10-01	Edit   Delete

Gambar 6 Tampilan Daftar Peminjaman

Terakhir, pada menu daftar peminjaman disajikan data riwayat beberapa anggota perpustakaan yang telah meminjam buku dengan rincian nama peminjam, id buku, id peminjaman, tanggal peminjaman dan tanggal pengembalian. Setelah itu, admin dapat menambah, mengubah dan menghapus data apabila diperlukan.

## IV. KESIMPULAN

Implementasi *database* relasional pada perpustakaan *online* ke dalam web ini dapat diakses oleh semua orang. Perbedaannya adalah *user* dibagi menjadi dua, yaitu terdapat admin dan anggota



perpustakaan. Admin memiliki kuasa untuk melihat dan mengedit *database* sedangkan anggota perpustakaan hanya dapat melihat list buku yang tersedia dan riwayat peminjamannya sendiri.

## UCAPAN TERIMA KASIH

## REFERENSI

- [1] E. Fatmawati, “DUKUNGAN PERPUSTAKAAN DALAM IMPLEMENTASI ‘KAMPUS MERDEKA DAN MERDEKA BELAJAR,’” *Jurnal Pustaka Ilmiah*, vol. 6, no. 2, p. 1076, Jan. 2021, doi: 10.20961/jpi.v6i2.46682.
- [2] A. Akbar, W. O. D. Aplisalita, and L. O. Rusadi, “Fungsi Perpustakaan Sekolah Terhadap Prestasi Belajar Siswa Sekolah Dasar,” *EDUKATIF: JURNAL ILMU PENDIDIKAN*, vol. 3, no. 1, pp. 203–212, Mar. 2021, doi: 10.31004/edukatif.v3i1.286.
- [3] M. Desy Ria and A. Budiman, “PERANCANGAN SISTEM INFORMASI TATA KELOLA TEKNOLOGI INFORMASI PERPUSTAKAAN,” *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 2, no. 1, pp. 122–133, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [4] Q. Jafar Adrian and R. Nur Devija, “Jurnal Manajemen Informatika (JAMIKA) Penerapan Sistem Informasi Administrasi Perpustakaan Menggunakan Model Desain User Experience”, doi: 10.34010/jamika.v1i11.
- [5] A. Rahma and R. Wulandari, “Perbedaan Yang Ada Pada Perpustakaan Konvensional Dengan Perpustakaan Pada Saat Ini,” *Jurnal Perpustakaan Dan Informasi*, vol. 16, 2022.
- [6] D. Wulan Dari and A. Oktarini Sari, “RANCANG BANGUN APLIKASI PERPUSTAKAAN SEKOLAH BERBASIS WEBSITE”, [Online]. Available: [www.bsi.ac.id](http://www.bsi.ac.id)
- [7] G. Susilo<sup>1</sup>, S. Sistem Informasi, S. Bina, P. Magelang, J. Raden, and S. No, “TRANSFORMASI Jurnal Informasi & Pengembangan Iptek’(STMIK BINA PATRIA ) KEAMANAN BASIS DATA PADA SISTEM INFORMASI DI ERA GLOBAL,” 2016.
- [8] Q. Aini *et al.*, “Pengamanan Pengelolaan Hak Akses Web Berbasis Yii Framework,” 2018.
- [9] S. Kasus, D. Jenderal EBTKE, and T. Andik Priyanto, “Jurnal Manajemen Informatika Jakarta Rancang bangun aplikasi informasi pengelolaan keamanan berbasis WEB”, doi: 10.52362/jmijayakarta.v2i1.736.
- [10] N. P. Lestari, Y. Durachman, S. Watini, and S. Millah, “Manajemen Kontrol Akses Berbasis Blockchain untuk Pendidikan Online Terdesentralisasi,” *Technomedia Journal*, vol. 6, no. 1, pp. 111–123, Jul. 2021, doi: 10.33050/tmj.v6i1.1682.
- [11] Yandri, “Perancangan Dan Implementasi Sistem Informasi Laporan Keuangan Apotek Jati Jaya Dengan Menggunakan Bahasa Pemrograman PHP Dan Database MySQL,” *Journal of Computer Science and Information Systems (JCoIns)*, vol. 3, no. 3, pp. 114–135, 2022.



# Analisis Prediksi Peluang Pendapatan Biaya Aset Sewa PT. KAI DAOP 6 Yogyakarta Dengan Menggunakan Metode *Markov*

Nabila Azzahra Haris Putri<sup>1a</sup>, Abdullah Ahmad Dzirkullah<sup>2b</sup><sup>1</sup> Program Studi Statistika, FMIPA UIICorresponding author email: <sup>a</sup>20611195@students.uui.ac.id, <sup>b</sup>adzirkullah@uui.ac.id

**Abstract:** PT. Kereta Api Indonesia (PT. KAI) is a company engaged in land transportation. Lease assets of PT. KAI is a service offered by PT. KAI to other parties to rent assets, which consist of land, buildings, or other infrastructure facilities to be used for a certain period of time. The Markov chain is a random process of information in the future occurring in the present, where the shape of the Markov chain in the future depends on the present. This method is used for making decisions on the amount of income for the next 6 months. If this week goes up, the chance for a change in status will increase by 53.89%, and the predicted chance for a status change will be 46.10%. If it falls this month, the prediction opportunity for predicting status will decrease by 46.13%, and the chance for prediction for status will increase by 53.86%. the probability of the next 6 periods going up is higher than going down. Efforts can be made to increase the income of PT KAI Daop 6 leased assets by being responsible and firm in their collection of asset tenants who have late payments of tenants, in accordance with applicable regulations..

**Keywords:** Leased Assets, Cash In, Markov Method

**Abstrak:** PT. Kereta Api Indonesia (PT. KAI) merupakan perusahaan yang bergerak di bidang transportasi darat. Sewa aset PT. KAI merupakan layanan yang ditawarkan oleh pihak PT. KAI kepada pihak lain untuk menyewa aset, yang terdiri berupa tanah, bangunan, atau fasilitas infrastruktur lainnya untuk digunakan selama jangka waktu tertentu. Rantai Markov merupakan proses acak suatu informasi pada masa depan terjadi pada masa sekarang, dimana bentuk dari rantai Markov dalam masa depan bergantung pada masa kini. Metode ini digunakan untuk pengambilan keputusan jumlah pendapatan selama 6 bulan kedepan. Apabila pekan ini naik maka peluang perubahan status naik yaitu sebesar 53.89%, dan prediksi peluang status turun sebesar 46.10%. Jika pada bulan ini turun maka prediksi peluang prediksi status turun sebesar 46.13%, dan peluang prediksi status naik sebesar 53.86%. peluang 6 periode kedepan naik lebih tinggi dibandingkan turun. Upaya yang dapat dilakukan untuk meningkatkan pendapatan aset sewa PT KAI Daop 6 dengan bertanggung jawab serta tegas dalam tindakan penagihan terhadap penyewa aset yang memiliki keterlambatan pembayaran penyewa, sesuai dengan ketentuan yang berlaku.

**Kata kunci:** Aset Sewa, pendapatan, Metode Markov

## I. PENDAHULUAN

Indonesia merupakan negara besar yang memiliki infrastruktur transportasi yang cukup mengalami perkembangan yang luar biasa beberapa tahun belakangan ini. PT. KAI merupakan salah satu perusahaan yang dipegang oleh BUMN yang bergerak di bidang transportasi darat, baik antar kota maupun antar daerah. PT. KAI memiliki fungsi sebagai penyedia layanan transportasi yang aman, terjangkau, dan efektif bagi masyarakat Indonesia, dengan jenis layanan kereta api antara lain: eksekutif, ekonomi, ekonomi premium, bisnis, luxury, dan panaromic.

PT. KAI tidak hanya bertanggung jawab terhadap transportasi saja, melainkan ada penyewaan aset yang harus dikelola dengan baik. Sewa aset dapat diartikan sebagai suatu perjanjian antara PT. KAI dengan penyewa, dimana PT. KAI menyediakan pemanfaatan aset milik PT. KAI berupa tanah dan bangunan, antara lain: rumah atau sarana lainnya di lingkungan PT. KAI. Penyewa menyewa aset diluar area stasiun yang di sewakan, di gunakan sebagai rumah tinggal keluarga, warung, kios jualan dan lain sebagainya.

Pengelolaan Aset PT. KAI di Unit Penagihan terdiri dilakukan oleh petugas yang memiliki kompeten dan professional dalam menjalankan tugas. Petugas Unit Penagihan harus memiliki pengetahuan dan keterampilan dalam penagihan tunggakan serta pemahaman terkait peraturan dan hukum penagihan, sehingga dapat memberikan pelayanan yang memuaskan dalam proses penagihan.

Teknologi informasi sangat membantu memudahkan petugas dalam tindakan penagihan, didukung oleh sistem manajemen data pelanggan dan pembayaran, sistem pengiriman surat tagihan, sistem telepon interaktif, dan sistem manajemen riwayat tunggakan pelanggan. Teknologi informasi juga menyediakan data pelanggan meliputi: informasi tentang nama pelanggan, nomor identitas, alamat pelanggan, nomor telepon, dan riwayat pembayaran.

Dalam memanfaatkan aset sewa PT. KAI memiliki strategi untuk meningkatkan pemanfaatan dan pendapatan dari aset sewanya. Salah satu strategi yang diterapkan adalah dengan memperluas jangkauan dan meningkatkan kualitas layanan yang diberikan kepada penyewa. Selain itu, PT. KAI juga berupaya untuk melakukan efisiensi dalam manajemen aset sewa agar dapat memaksimalkan potensi pendapatan dari aset tersebut. Maka dari itu peneliti melakukan analisis rantai Markov pada data pendapatan sewa aset KAI Daop 6 untuk memprediksi peluang pendapatan biaya di periode yang akan datang.

## II. METODE PENELITIAN

Data yang digunakan peneliti adalah data sekunder atau data yang diambil dari KAI yaitu data pendapatan uang aset sewa yang masuk ke PT. KAI Daop 6 Yogyakarta dari bulan Januari 2017 sampai dengan Februari 2023.

Metode Rantai Markov berfungsi untuk melihat prediksi peluang transisi pendapatan uang aset sewa per bulannya berdasarkan data yang diketahui sebelumnya. Tahapan yang dilakukan peneliti untuk menyelesaikan penelitian ini yaitu pengumpulan data; melakukan pengkategorian data perbulannya berdasarkan urutan data sebelum dan sesudahnya yaitu naik dan turun; membuat matrix transisi berdasarkan data yang didapat yaitu  $2 \times 2$  dimana *state* 0 = turun dan 1 = naik; dan alur perhitungan untuk menentukan *n-step* atau peluang transisi beberapa bulan kedepan.



**Gambar 1** Diagram alir

Pada *variable* yang digunakan penulis dalam laporan kerja praktik terdapat 3 data variabel di dalamnya yaitu: *variable* bulan ke- dimana pada *variable* ini menghitung dari bulan Januari 2017

sampai dengan Februari 2023; *variable cash in* atau pendapatan yang masuk di KAI Daop 6; dan *variable* status yang dihasilkan dari jumlah pemasukan perbulannya atau dari bulan sebelum dibandingkan dengan bulan setelahnya.

Proses Markov merupakan dasar dari metode simulasi stokastik dimana peristiwa pada masa depan hanya bergantung pada apa yang terjadi pada saat ini dan tidak dapat di pengaruhi oleh peristiwa pada masa lalu. Nama Markov umum dikenal sebagai Markov Monte Carlo. (Kadafi, Nugroho & Novianti, 2017).

Rantai Markov dapat dikenal jika

- Nilai yang mungkin terjadi dari  $X_t$  adalah terhitung
- Mempunyai peluang transisi keadaan “i” (pada waktu t) ke keadaan “j” (pada waktu t+1) adalah  $p_{ij}$  ialah

$$P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_1 = i_1, X_0 = i_0) = P_{ij} \quad (1)$$

### 2.1 Peluang Transisi

Peluang transisi adalah proses Markov dalam perpindahan *state*. Sebagai contoh P menyatakan matriks peluang transisi satu langkah  $P_{ij}$ , seperti matriks dibawah ini.

$$P = \begin{bmatrix} P_{00} & P_{01} & \dots & P_{0j} \\ P_{10} & P_{11} & \dots & P_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ P_{i0} & P_{i1} & \dots & P_{ij} \end{bmatrix} \quad (2)$$

### 2.2 Peluang Transisi n-step

Peluang transisi n-step di definisikan sebagai peluang transisi n-step  $P_{ij}^{(n)}$ , merupakan sebuah kemungkinan sistem bersyarat yang berada dalam keadaan *state* I akan berada pada *state* j setelah melalui proses transisi n-step. Jadi,

$$P_{ij}^{(n)} = P \{ X_{t+n} = j | X_t = i \} \quad (3)$$

Maka,  $P_{ij}^{(n)}$  merupakan probabilitas bersyarat, probabilitas tersebut harus bilangan positif, maka dari itu *state* tersebut harus memiliki sifat sebagai berikut.

1.  $P_{ij}^{(n)} \geq 0$ , untuk seluruh i dan j dimana  $n = 1, 2, 3, \dots$  dst
2.  $\sum_{j=0}^{\infty} P_{ij}^{(n)} = 1$ , untuk seluruh i dimana  $n = 1, 2, 3, \dots$  dst

*Matrix* peluang transisi n-step bisa di dapatkan dengan cara mengalikan peluang transisi sebanyak n.

$$\begin{aligned} P^{(n)} &= P \times P \times P \times \dots \times P = P^n \\ &= P \times P \times P^{n-1} = P^{n-1} \times P \end{aligned}$$

## III. HASIL DAN PEMBAHASAN

### 3.1. Statistika Deskriptif



Gambar 2 Contoh Gambar Tunggal

Pada gambar diatas, merupakan grafik *line chart* yang berisikan rata-rata pendapatan uang aset sewa dari tahun 2017 – 2023. Dilihat dari grafik, bahwa rata-rata pendapatan sempat mengalami



penurunan yang cukup besar pada tahun 2020 hanya sebesar 2.816.168.563, dimana pada tahun tersebut merupakan puncak virus covid-19 yang sedang melanda Indonesia. Rata-rata tertinggi pendapatan uang aset sewa pada tahun 2021 sebesar 5.143.713.033.

### 3.2. Matriks Peluang Transisi

**Tabel 1** Matriks Transisi Sesudah dan sebelum kejadian

Status	Turun	Naik	Jumlah Status
Turun	16	22	38
Naik	23	11	34

Data diatas merupakan jumlah status pendapatan uang masuk aset sewa transisi turun ke turun sebanyak 16, transisi naik ke turun sebanyak 23, transisi turun ke naik sebanyak 22, serta transisi naik ke naik sebanyak 11. Maka ditulis peluang sebagai berikut:

**Tabel 2** Matriks Peluang Transisi

Status	Turun	Naik
Turun	0.42105	0.57894
Naik	0.67647	0.32352

### 3.3. Peluang Transisi Satu Langkah (n-step)

Perhitungan peluang transisi *n-step* dengan matriks 2 x 2 dalam memprediksi peluang transisi jumlah kejadian naik turun nya pemasukan biaya uang aset sewa selama 6 periode/bulan kedepan, maka didapatkan hasil peluang sebagai berikut.

$$P^6 = P \times P \times P \times P \times P \times P$$

$$P = \begin{matrix} 0 & 1 \\ 0.46130 & 0.53869 \\ 1 & 0.46103 & 0.53896 \end{matrix}$$

### 3.4. Hasil Peluang

Jika hasil sebelumnya, 6 periode kedepan diinput kedalam presentase maka di dapatkan hasil sebagai berikut.

**Tabel 4.** Hasil Prediksi Peluang 6 Periode Kedepan

Status	Turun	Naik
turun	46.13%	53.86%
naik	46.10%	53.89%

Berdasarkan dari informasi diatas bahwa *table* tersebut merupakan peluang Peramalan 6 periode kedepan atau 6 bulan kedepan yaitu pada bulan Agustus 2023 berdasarkan dari 74 bulan/periode sebelumnya. Prediksi peluang transisi jumlah pendapatan dalam 6 bulan kedepan pada Agustus 2023 bila pekan ini naik maka peluang perubahan status naik yaitu sebesar 53.89%, kemudian apabila pekan ini naik maka prediksi peluang status turun sebesar 46.10%. Jika pada bulan ini turun maka prediksi peluang prediksi status turun sebesar 46.13%. Peluang prediksi status naik jika bulan ini turun sebesar 53.86%.

Berdasarkan hasil prediksi peluang bahwa presentase peluang transisi jumlah pendapatan periode selanjutnya untuk status naik lebih besar daripada status perpindahan yang lainnya. Hal tersebut dapat dijadikan sebagai gambaran pendapatan mendatang yang status naik, pegawai PT. KAI



dapat berupaya untuk bertanggung jawab serta tegas dalam tindakan penagihan terhadap penyewa aset yang memiliki keterlambatan pembayaran agar memastikan bahwa penyewa membayar aset sewa tepat pada waktunya, sesuai dengan ketentuan yang berlaku.

#### **IV. KESIMPULAN**

Berdasarkan pembahasan yang telah di dapatkan, disimpulkan sebagai berikut:

1. Peluang transisi jumlah pendapatan biaya aset sewa pada periode ke-6 mendatang yaitu bulan Agustus 2023 apabila bulan ini naik maka perubahan ke turun dan naik secara berturut-turut yaitu 46.10% dan 53.89%, dimana terdapat jumlah data sebanyak 74 data dengan peluang 6 periode kedepan naik lebih tinggi dibandingkan turun. Diperoleh rata-rata tertinggi pendapatan pada 6 tahun terakhir tertinggi di tahun 2021 dan terendah pada tahun 2020.
2. Peningkatan pemanfaatan aset juga dapat dilakukan dengan terus mempertahankan kualitas aset dengan melakukan perawatan dan pengecekan rutin aset yang disewakan agar kualitas tetap terjaga dan meningkatkan daya tarik bagi penyewa yang akan menggunakan aset PT. KAI.

#### **UCAPAN TERIMA KASIH**

Puji syukur kehadiran Allah SWT atas segala limpaham rahmat, inayah, taufik, dan hidayahnya sehingga kami dapat menyelesaikan penyusunan makalah dengan sangat sederhana. Semoga dengan makalah ini dapat dipergunakan sebagai salah satu acuan, petunjuk, dan pedoman bagi pembaca makalah ini. Harapan saya semoga makalah ini membantu menambah pengetahuan dan pengalaman bagi para pembaca, sehingga saya dapat memperbaiki bentuk maupun isi makalah ini sehingga kedepannya dapat lebih baik. Kami ucapkan Terima Kasih

#### **REFERENSI**

- [1] Denis, Djoni, Marline (2022). Analisis Rantai Markov Untuk Mengetahui Peluang Perpindahan Merek Kartu Seluler Pra Bayar.
- [2] Derwin Suhartono (2013). Markov Chain
- [3] Fatimatuzzahro, N. S. (2022). Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Barat [Skripsi]. Universitas Islam Indonesia.
- [4] Oky Nasrul (2019). Pemanfaatan Tanah Aset PT Kereta Api Indonesia (Persero) Divisi Regional II Sumatera Barat Oleh Pihak Ketiga.
- [5] Trifena Setiawaty (2020). Pengaruh motivasi dan kopensasi terhadap kinerja karyawan pada PT. Kereta Api Indonesia.
- [6] Yeni Lesmana (2018). Pengaruh Kereta Api Indonesia dalam kebutuhan masyarakat.



# Monitoring Suhu dan Kelembaban Pada Penetas Telur Otomatis Menggunakan Metode Fuzzy Sugeno Berbasis IoT

Hari Yoal<sup>1</sup>, Wahyu Dirgantara, S.T.,M .T.<sup>2</sup>, Ir. Subairi,S.T.,M.T.,IPM<sup>3</sup>

<sup>1,2</sup>Program Studi Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur

Hari Yoa<sup>1,3</sup> (Jurusan Teknik Elektro, Universitas Merdeka Malang)

<sup>1</sup>[hariyoal198@gmail.com](mailto:hariyoal198@gmail.com)

<sup>3</sup>[wahyu.digantara@unmer.ac.id](mailto:wahyu.digantara@unmer.ac.id)

<sup>2</sup>Wahyu Dirgantara (Jurusan Teknik Elektro, Universitas Merdeka Malang)

Corresponding author email: [subairi@unmer.ac.id](mailto:subairi@unmer.ac.id)

**Abstract:** *One of the mainstay businesses of farmers engaged in animal husbandry is raising chickens and ducks. Every year statistical data records that the public's need for chicken meat continues to increase. Indirectly this will have a positive impact on farmers, especially chicken farmers. Chicken farmers will try to increase their chicken breeding business, which aims to prevent the decline in broiler and layer prices. In this research, an automatic egg hatcher has been made by applying the artificial heating equalization method. The egg incubator is made by taking into account the ideal temperature conditions for incubating chicken eggs, which are 35.3 0C - 40.5 0C, with humidity in the machine ranging from 70%-80%. This machine has a capacity for 60 eggs. This egg incubator is a modification of a previously made tool equipped with a fan as air circulation.*

**Keywords:** *Artificial heating, Egg hatching machine, Free-range chicken eggs, Chicken eggs*

**Abstrak:** Salah satu usaha andalan peternak yang bergerak di bidang peternakan adalah beternak ayam dan itik. Setiap tahun data statistik mencatat bahwa kebutuhan masyarakat akan daging ayam terus meningkat. Secara tidak langsung hal ini akan berdampak positif bagi peternak khususnya peternak ayam. Peternak ayam akan berusaha meningkatkan usaha pembibitan ayamnya, yang bertujuan untuk mencegah penurunan harga ayam pedaging dan ayam petelur. Pada penelitian ini telah dibuat alat penetas telur otomatis dengan menerapkan metode pemerataan pemanasan buatan. Mesin penetas telur dibuat dengan memperhatikan kondisi suhu yang ideal untuk mengerami telur ayam yaitu 35,3 °C – 40,5 °C, dengan kelembaban di dalam mesin berkisar antara 70%-80%. Mesin ini memiliki kapasitas untuk 60 butir telur. Mesin penetas telur ini merupakan modifikasi dari alat yang dibuat sebelumnya yang dilengkapi dengan kipas sebagai sirkulasi udara. Dan tidak perlu roller untuk memutar atau membolak-balik telur karena telur sudah dipanaskan.

**Kata kunci:** Pemanasan buatan, Mesin penetasan telur, Telur ayam kampung, Telur ayam

## I. PENDAHULUAN

Bisnis ternak ayam memiliki potensi yang menjanjikan dan meningkatkan taraf hidup peternak, jika pengelolaan dilakukan secara baik. Kondisi diperkuat dengan meningkatnya permintaan unggas setiap bulan secara signifikan dengan banyaknya dibuka restoran dan café yang menyediakan menu berbahan dasar unggas. Hal tersebut didukung dari data Badan Pusat Statistik menyebutkan bahwa produksi daging ayam di provinsi Sulawesi Tengah pada tahun 2020 mencapai 5704 ton. Informasi produksi daging yang dihasilkan mendukung dalam meningkatkan pemasaran baik skala lokal, nasional dan internasional. Untuk memenuhi permintaan konsumen tentunya dibutuhkan teknologi yang dapat mempercepat dan mempermudah dalam penetasan telur [1]. Seiring dengan pertambahan populasi manusia, hal itu berkait erat dengan penyediaan makanan yang juga semakin meningkat, maka kebutuhan ayam pedaging sebagai salah satu sumber makanan hewani, juga semakin bertambah. Reproduksi merupakan proses berkembang biakan untuk setiap makhluk hidup, termasuk unggas yang banyak dibudidayakan sebagai hewan ternak. Semua unggas melakukan reproduksinya dengan cara bertelur.



Salah satu jalan untuk mengatasinya yaitu dengan menggantikan peran mesin penetas telur konvensional yang ditingkatkan kemampuannya menjadi mesin otomatis penetas telur dan bekerja sesuai perintah yang ditanamkan pada mikrokontroler. Perancangan Setelah melakukan proses perkawinan induk jantan dengan induk betina, selang waktu induk betina akan mengeluarkan telur hasil pembuahannya. Jumlah telur yang dihasilkan setiap unggas bervariasi. Ayam dapat menghasilkan 13-20 butir telur. Hal ini dimungkinkan untuk peningkatan jumlah populasi ayam. Dalam sekalimasa bertelur dimana seekor ayam dapat menghasilkan satu butir telur sehari. Induk betina akan mengerami telurnya selama waktu tertentu hingga menetas menjadi anakan. Telur ayam mempunyai lama penetasan normal 21 hari, jika hanya mengandalkan pengeraman alami maka keberhasilan telur menetas hanya 50%-60%. Apabila proses penetasan alami yaitu menggunakan indukan mengalami kendala yang disebabkan oleh induknya sendiri dan lingkungan yang tidak mendukung disebabkan oleh faktor suhu dan kelembaban, dapat merugikan bagi peternak telur [2].

Alat penetas telur adalah sebuah alat yang dapat membantu untuk menetas telur. Alat penetas ini dilengkapi dengan peralatan pendukung untuk mengatur kondisi lingkungannya mirip atau serupa dengan indukan. Box (kotak) alat penetas diusahakan dibuat dari bahan yang anti rayap dan anti air agar lebih awet dan higienis sehingga tidak mempengaruhi kualitas telur yang akan ditetaskan. Pada mulanya alat penetas hanya sebuah alat sederhana yang hanya menggunakan lampu untuk menghasilkan panas dan tanpa instrument-instrument pendukung lainnya dan hanya digunakan oleh peternak-peternak tradisional dengan skala kecil, tapi seiring dengan perkembangan zaman alat penetas telur dikembangkan untuk meningkatkan kemampuan penetasan dan kemudahan melakukan penetasan telur [3].

Dari permasalahan di penelitian bertujuan membuat alat penetas telur otomatis dengan kontrol suhu dan kelembaban yang dikendalikan dengan menggunakan metode fuzzy sugeno sebagai pendukung pengambilan keputusan untuk outputnya adalah tingkat kecerahan lampu dan kipas sebagai pengatur sirkulasi udara. Tujuan penelitian ini yaitu menghasilkan sebuah simulasi sistem pengontrolan suhu pada mesin penetas telur berbasis mikrokontroler Esp32[4] dan mengontrol suhu yang ada pada oven penetas dengan di kontrol dan di monitoring menggunakan Blynk. Blynk merupakan platform baru yang memungkinkan anda untuk dengan cepat membangun interface untuk mengendalikan dan memantau proyek hardware dari perangkat Android. *Blynk* adalah IOT (*Internet of Things*)[4].

Penelitian ini diharapkan manfaat bagi berbagai pihak diantaranya bagi peternak yaitu membantu memperlancar proses pengontrolan suhu yang selama ini dilakukan secara manual ke sistem otomatisasi suhu dan kelembaban sehingga meningkatkan ketelitian, pada mesin penetasan telur otomatis dan berbasis internet of things untuk memonitoring data suhu dan kelembaban secara real time menggunakan aplikasi *Blynk* yang terinstal pada android pengguna, pada pengembangnya diharapkan bisa menjadi bahan acuan atau pembanding untuk penelitian kedepannya.

## II. KAJIAN PUSTAKA

### NodeMCUEsp32

*Esp-32* merupakan suksesor atau penerus dari *Esp-8266* yang memiliki banyak fitur tambahan dan keunggulan dibandingkan generasi sebelumnya. Pada ESP32 terdapat inti CPU serta *Wi-Fi* yang lebih cepat, GPIO yang lebih banyak, dan dukungan terhadap *Bluetooth* 4.2, serta konsumsi daya yang rendah [6]. ESP32 sendiri tidak jauh berbeda dengan ESP8266 yang familiar dipasaran, hanya saja ESP32 lebih kompleks dibandingkan ESP8266. Pada 2016 Espressif meluncurkan produk terbarunya yang bernama ESP32. ESP32 hadir tidak untuk menggantikan ESP8266, namun memberikan perbaikan pada semua lini. Tak hanya mempunyai dukungan konektivitas *WiFi*, akan tetapi juga *Bluetooth* membuatnya lebih serbaguna dan cocok dalam menangani proyek yang besar..

**sensor thermocouple**

*Thermocouple* adalah sensor ukur yang menghasilkan perubahan tegangan karena perbedaan suhu relatif antara *junction* dua logam yang berbeda dan *output junction*, dimana satu logam sebagai referensi dan logam lainnya sebagai pengukur suhu. Dalam hal ini sensor *thermocouple* mengubah panas pembakaran menjadi tegangan listrik [7]. *Thermocouple* merupakan salah satu sensor yang paling umum digunakan untuk mengukur suhu karena relatif murah tetapi akurat yang dapat beroperasi pada suhu panas maupun dingin.

Beberapa kelebihan *Thermocouple* yang membuatnya menjadi populer ialah responnya yang cepat terhadap perubahan suhu dan pula rentang suhu operasionalnya yang luas yaitu berkisar diantara -200°C hingga 2000°C. Selain respon yang cepat dan rentang suhu yang luas, *Thermocouple* juga tahan terhadap guncangan atau getaran serta praktis digunakan. Sensor suhu *Thermocouple* mempunyai nilai output yang kecil dengan *noise* yang tinggi, sehingga memerlukan memerlukan rangkaian pengkondisi sinyal supaya nilai *output* dapat dibaca dengan baik.

**Driver Mosfet**

MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) adalah sebuah komponen semikonduktor yang dapat digunakan sebagai saklar elektronik dan penguat sinyal pada perangkat elektronik. MOSFET umumnya memiliki 3 kaki, yaitu *Gate* (G), *Drain* (D), dan *Source* (S). Dari segi bahan semikonduktor pembuatnya, MOSFET terbagi atas 2 bagian yaitu MOSFET N-type dan MOSFET P-type [8]. MOSFET bekerja secara elektronik mevariasikan sepanjang jalur pembawa muatan (*electron* atau *hole*). Muatan listrik masuk melalui saluran pada *source* dan keluar melalui drain. Lebar saluran dikendalikan oleh tegangan pada *elektrode* yang disebut dengan gate atau gerbang yang terletak antara *source* dan *drain*. Ini terisolasi dari 18 saluran didekat lapisan oksida logam yang tipis. Kapasitas MOS pada komponen ini adalah bagian Utamanya[9].

**Sensor DHT-22**

DHT-22 salah satu sensor yang dapat digunakan untuk melakukan pengukuran suhu dan kelembaban adalah DHT-22. satuan suhu adalah derajat celsius (<sup>0</sup>c) dan satuan kelembaban adalah persen(%) kelembaban adalah jumlah masa uap air yang ada dalam volume di udara, Suhu adalah ukuran kuantitatif terhadap Temperatur panas dan dingin. DHT-22 ditunjukkan pada Gambar 2.4. DHT-22 adalah sensor digital untuk pengukuran kelembaban dan suhu relatif [10].

**IoT Arduino software IDE**

Arduino IDE merupakan kependekan *Integrated Development Enviroenment*, atau secara Bahasa mudahnya merupakan lingkungan terintegrasi yang digunakan untuk melakukan pengembangan. Disebut sebagai lingkungan karena melalui *software* inilah Arduino dilakukan pemrograman untuk melakukan fungsi-fungsi yang dibenamkan melalui sintaks pemrograman. Arduino menggunakan Bahasa pemrograman sendiri yang menyerupai bahasa C bahasa pemrograman Arduino (*Sketch*) sudah dilakukan perubahan untuk memudahkan pemula dalam melakukan pemrograman dari Bahasa aslinya. Arduino IDE (*Integrated Development Environment*) adalah *software* yang digunakan untuk memrogram di Arduino, dengan kata lain Arduino IDE sebagai media untuk memrogram *board* Arduino IDE. Arduino ini berguna sebagai teks editor untuk membuat, mengedit, dan juga mevalidasi kode program. [11].

**Blynk**

*Blynk* merupakan platform baru yang memungkinkan anda untuk dengan cepat membangun *interface* untuk mengendalikan dan memantau proyek *hardware* dari iOS dan perangkat Android [12].

### Fuzzy Logic

*Fuzzy Logic* merupakan suatu cara untuk bisa memetakan suatu input ke dalam suatu output. *Fuzzy Logic* memiliki himpunan *Fuzzy* dimana teori himpunan *Fuzzy* merupakan perkembangan dari teori himpunan klasik. Dimana dengan logika *Fuzzy*, hasil yang keluar tidak akan selalu konstan dengan *input* yang ada. *Fuzzy Logic* merupakan himpunan logika yang dikembangkan dalam hal mengatasi nilai yang berada diantara kebenaran (*true*) dan kesalahan (*false*). Metode *Fuzzy* sugeno dipakai dalam menunjang penelitian yang akan dilakukan [13].

### Dimmer

Dimmer adalah sebuah rangkaian komponen elektronika dari input sinyal AC kemudian sinyal tersebut diproses menjadi sinyal AC Phase maju dari pada sinyal AC inputan, yang menyebabkan Penurunan Daya (Watt) bisa disimpulkan Dimmer berguna menurunkan daya (watt) yang mengakibatkan lampu bisa redup. Dimmer berupa resistor yang digunakan untuk memvariasikan kecepatan suplai udara dari blower dengan cara membatasi arus listrik yang mengalir sehingga putaran blower bervariasi tergantung besar kecilnya angka dimmer. Dimmer digunakan untuk mendapatkan nilai equivalence ratio yang diinginkan. Adapun dimmer yang digunakan dalam pengambilan data disetting mempunyai 5 tingkat kecepatan [14].

### lampu pijar

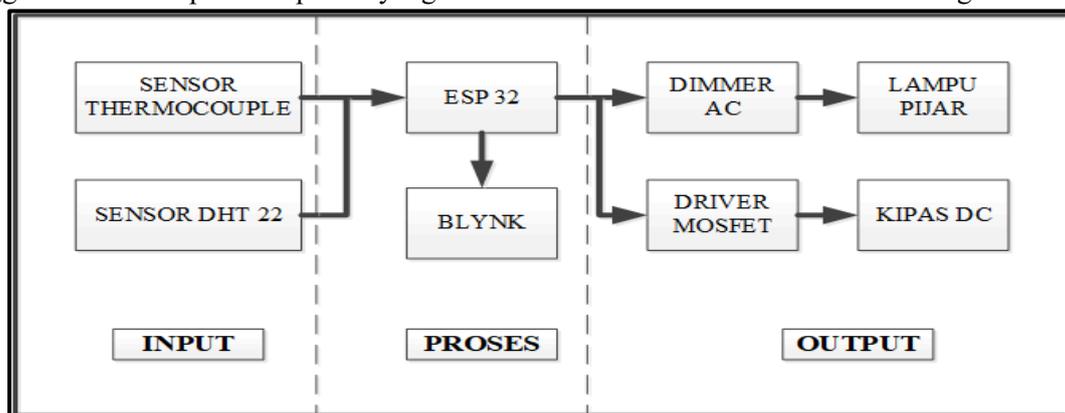
Lampu Pijar adalah salah satu jenis lampu yang sudah tidak asing digunakan oleh masyarakat. Meskipun pemerintah telah menghimbau masyarakat untuk menggunakan lampu hemat energi, namun dalam kenyataannya masih ada masyarakat yang memilih menggunakan lampu pijar sebagai penerangan karena harganya yang murah dan mudah didapat. Lampu Pijar digunakan untuk memanaskan ruang pemanas [15].

### Kipas

Kipas yang bergerak beserta sepasang pipih yang berbentuk magnet pada bagian yang diam Kipas angin menggunakan motor listrik untuk mengubah energi listrik menjadi gerakan baling baling. Dalam Kipas Angin DC terdapat suatu kumparan besi pada bagian. Ketika listrik mengalir pada lilitan kawat dalam kumparan besi, hal ini membuat kumparan besi menjadi sebuah magnet [16].

## III. METODE PENELITIAN

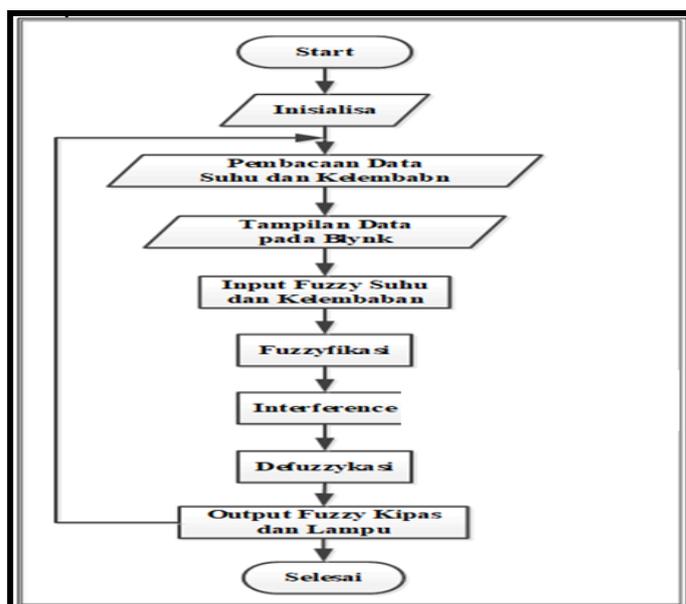
Pada bagian ini merupakan tampilan dari Blok Diagram kerja alat, *flowchart*, dan Diagram Fuzzy Saat membuat desain rancangan alat sangat membutuhkan diagram blok yang menggambarkan tahapan atau proses yang dilakukan Berikut adalah Gambar Blok Diagram Kerja alat.



Gambar 1. Diagram Blok kerja Alat

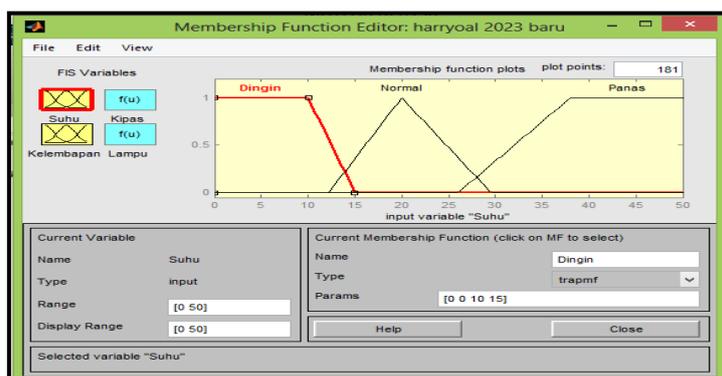
Pada Gambar 3.1. merupakan blok diagram cara kerja alat, dimana terdapat *Input*, *Output* dan Proses. Dimana cara kerja dari masing-masing blok diatas adalah Blok *input* Terdapat sensor *Thermocouple* yang digunakan untuk membaca suhu ruangan penetas telur dan sensor *DHT-22* digunakan untuk membaca data kelembapan pada alat penetas telur dan pada Blok Proses *Esp32* digunakan sebagai mikrokontroler utama dan sebagai modul *wifi* untuk mengirim data, serta *Blynk* untuk memantau proses kerja mesin penetas telur. Blok *Output* Terdapat *Dimer* untuk mengontrol lampu pijar dimana lampu pijar digunakan untuk memanaskan telur dan *Driver Mosfet* untuk mengontrol kipas DC yang di gunakan untuk menormalkan suhu dalam oven pemanas sesuai Setpoint.

*Flowchart* merupakan rangkaian alur pada sistem kerja rangkaian. Berikut adalah Gambar dari *flowchart* pada penelitian ini.



**Gambar 2.** Diagram Alur Perangkat Lunak Alat

Dimana diagram alir diatas merupakan sebuah software yang menunjukkan alur atau cara kerja sistem dengan alat bantu dalam penelitian ini. Dimana diawali dengan start, kemudian dilanjutkan dengan proses inialisasi untuk pin yang digunakan, kemudian dilanjutkan dengan data sensor *thermocouple* dilanjutkan dengan menampilkan data sensor suhu pada Aplikasi *Blynk*, kemudian lanjutkan membaca data sensor suhu dengan suhu 35°C - 45°C lalu masuk ke proses *Fuzzy* dimana outputnya adalah PWM (*Pulse Width Modulation*) dan pencarian Dimmer untuk melanjutkan menentukan suhu. Setelah proses *Fuzzy* maka data pengeringan dikirim ke data *Blynk cloud* dan jika Sistem dimatikan maka proses selesai, sebaliknya jika dan mulai proses. jika tidak maka mulailah dari proses awal.



**Gambar 3.**

**Tabel 1** Fungsi Derajat Keanggotaan Suhu

No	Fungsi keanggotaan input	Range
1	Dingin	0-15 °C
2	Normal	15-30 °C
3	Panas	30-50 °C

**Tabel 2** Fungsi derajat keanggotaan Kelembaban

No	Fungsi keanggotaan input	Range
1	rendah	30-65 %
2	sedang	65-70 %
3	tinggi	70-100 %

Selain suhu dan kelembaban terdapat juga Derajat keanggotaan *Output* kipas dan dimmer lampu

**Tabel 3** Fungsi derajat keanggotaan *output* kipas

No	Fungsi keanggotaan output	Range
1	mati	85
2	pelan	170
3	cepat	255

Selain output kipas terdapat juga Derajat keanggotaan Output dimmer lampu yang berfungsi untuk mengatur tingkat pencerahan lampu

**Tabel 4** Fungsi derajat keanggotaan *output* kipas

No	Fungsi keanggotaan output	Range
1	mati	85
2	pelan	170
3	cepat	255

Range output dimmer lampu terbagi atas beberapa variabel

**Tabel 5** Fungsi derajat keanggotaan output dimmer lampu

No	Fungsi keanggotaan output	Range
1	Redup	35
2	Terang	75
3	Normal	95

#### IV. HASIL DAN PEMBAHASAN

Pada tahap ini dilakukan uji coba dengan menyalakan lampu serta menghidupkan kipas dan tingkat kecerahan lampu berubah sesuai dengan suhu yang berubah berdasarkan output fuzzy.

Uji coba pertama pada ruangan penetas telur dengan kenaikan suhu dari suhu awal 25<sup>0</sup>C dan terus meningkat seperti pada Gambar dibawah ini


**Gambar 4a**

**Gambar 4b**

Gambar 1a merupakan suhu awal pada ruang penetas telur Gambar 1b merupakan suhu akhir pada ruang penetas telur .

Pengujian berhenti pada suhu akhir 35°C pada ruang penetas telur dan berikut adalah hasil perbandingan antara output PWM kipas pada matlab dan arduino seperti pada Tabel 4.19

**Tabel 8** Perbandingan PWM kipas DC

Suhu °C	Kelembaban (%)	PWM Kipas		% error
		Arduino	Matlab	
40	51,30	127	127	0.5
39	51,20	126	110	16
38	50,80	122	98.7	23.3
37	50,90	112	86.2	25.8
36	50,60	90	67.1	22.9
35	50,40	27	32.4	5.4
36	50,50	90	32.4	57.6
35	50,40	27	32.4	5.4
36	50,50	90	32.4	57.6
35	50,60	27	32.4	5.4
<b>Jumlah</b>		<b>83.8</b>	<b>65.1</b>	
<b>Rata-rata</b>		<b>83.8</b>	<b>65.1</b>	
<b>Rata-rata error</b>				<b>21.9</b>

Pada Tabel Merupakan hasil perbandingan nilai antara matlab dan arduino dengan rata-rata eror **21.9** %. Berikut adalah hasil perbandingan antara output dimmer lampu pada matlab dan arduino seperti pada Tabel 4.10

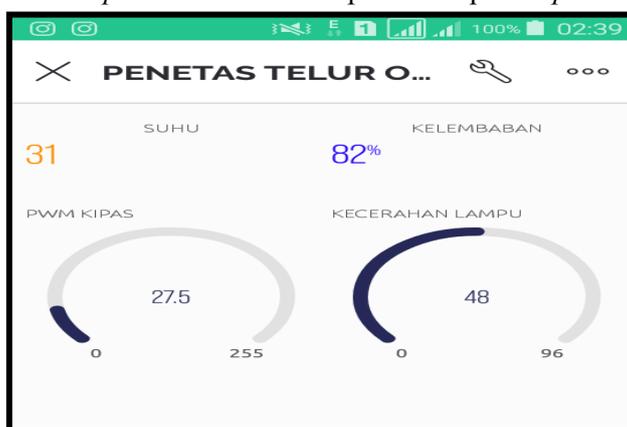
**Tabel 9** Perbandingan Dimmer lampu

Suhu C°	Kelembaban (%)	PWM Dimmer Lampu		error
		Arduino	Matlab	
40.00	51.30	48	14.80	48
39.00	51.20	48	22.3	25,7
38.00	50.80	48	27.3	20,7
37.00	50.90	48	33.2	14.8
36.00	50.60	48	41.5	6.5

35.00	50.40	48	55	7
36.00	50.50	48	55	7
35.00	50.40	48	55	7
36.00	50.50	48	55	7
35.00	50.60	48	41.5	6.5
<b>Jumlah</b>		<b>480.00</b>	<b>400.6</b>	<b>135.4</b>
<b>Rata-rata</b>		<b>48.</b>	<b>64.63</b>	
<b>Rata-rata eror</b>				<b>7.97</b>

Hasil perbandingan pwm dimmer lampu pada arduino dan simulasi matlab yang di lakukan 10 kali percobaan dengan hasil error **7.97%**

*Aplikasi blynk yang ada di smartphone. Berikut merupakan tampilan Aplikasi Blynk di smartphone.*



**Gambar 5.** Tampilan Aplikasi *Blynk* pada *Smartphone*

Gambar 5 menunjukkan Hasil monitoring suhu dan kelembaban dan output pwm kipas dan tingkat kecerahan lampu pada oven pemanas ditampilkan pada aplikasi *Blynk*.

## V. KESIMPULAN

Dengan adanya teknologi mikrokontroler Esp32 dan aplikasi *Blynk* IoT pada sistem penetas telur otomatis berbasis *internet of things* sangat membantu untuk Tidak terjadi gangguan pengiriman data menggunakan *Microcontroller Esp32*.

. Sistem mampu memberikan informasi suhu dan kelembaban Alat penetas telur Dengan menggunakan Aplikasi *Blynk* IoT, dari jarak yang jauh dengan perangkat android

Pada nilai suhu dan kelembaban yang berubah- ubah, Metode *fuzzy sugeno* sebagai pendukung pengambil keputusan bekerja dengan baik dan, untuk membuat bertujuan mengantur nilai suhu 35-40 dan kelembaban 70-100 pada mesin penetas telur otomatis.

## SARAN

Agar tidak terjadi gangguan pengiriman data sebaiknya *microcontroller Esp32* dihubungkan dengan jaringan internet yang stabil dan kuat.

## UCAPAN TERIMA KASIH

Ucapan Terima kasih kepada orang-orang yang telah membantu saya sehingga artikel ini dapat ditulis dengan baik dan benar.

**REFERENSI**

1. Zakaria, A., Sollu, T. S., & Asali, S. (2021). PERANCANGAN SISTEM PENETAS TELUR AYAM BERBASIS SMS GATEWAY. *Infotronik: Jurnal Teknologi Informasi dan Elektronika*, 6(2), 48-55.
2. Jufril, D., Darwison, D., Rahmadya, B., & Derisma, D. (2015). Implementasi Mesin Penetas Telur Ayam Otomatis Menggunakan Metoda Fuzzy Logic Control. *Prosiding Semnastek..*
3. Ahya, R., & Akuba, S. (2018). Rancang bangun alat penetas telur semi otomatis. *Jurnal Teknologi Pertanian Gorontalo (JTPG)*, 3(1), 44-44.
4. Ratag, C. C. (2017). Simulasi Sistem Pengontrolan Suhu Pada Mesin Penetas Telur Berbasis Mikrokontroler At89C52. *Respati*, 8(22).
5. Tamba, S. P., Nasution, A. H. M., Indriani, S., Fadhilah, N., & Arifin, C. (2019). Pengontrolan lampu jarak jauh dengan nodemcu menggunakan Blynk. *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, 2(1), 93-98.
6. Setiawan, A., & Purnamasari, A. I. 2019. Pengembangan smart home dengan microcontrlles ESP32 Dan mc-38 door magnetic switch sesor berbasis internet of things (IOT) untuk meningkatakn deteksi dini keamanan perumahan jurunal (*RESTI rekayasa sistem dan teknolgi informasi*), 3(3), 451-457.
7. Yudhanto, A., Sari A. P., Rachman, N., & Subairi, S. (2020) Implemetasi Sesor Thermocouple Berbasis Telemetry untuk mengukur thermal pembakaran propelan roket. *JASIEK jurnal aplikasi sains, informasi, elektronika dan komputer* 2(1), 38-46.
8. Hidayat, f., & Krismadinata, K. (2019). Rancang bangun VVVF Inverter 3 Fasa untuk oprasi motor induk tiga fasa dengan antar muka komputer. *INVOTEK jurnal inovasi vakasional dan teknologi*, 19(2), 47-56.
9. Herlan, A., Fitri, & Nurani r. (2021). Rancang bangun sistem monitoring data sabaran covi-19 secara real-time menggunakan arduinio berbasis internet of things (IOT). *Jurnal JTIK (jurnal teknologi informasi dan komunikasi)*, 5(2), 206-212.
10. Tambah. S.P., Nasiution, A. H. M., Indriani, S., Fadhilih, N., & Arifin, C. (2019). Pengontrolan lampu jarak jauh dengan nodemcu menggunakan Blynk. *jurnal tekinkom (Teknik Informasi Dan Komputer)*, 2(1), 93-98
11. Nasron, Suroso, Dan Astriana, Rahman Putgri, 2019 “perancangan logika FUZZY untuk sistem pengendali kelembaban tanah dan suhu tanaman”, politeknik negeri Sriwijaya, Indonesia.
12. Suliono, F. D., & Sudarmanta, B. (2020). Pengaruh kecepatan aliran udara dengan pengaturan dimmer pada tekanan udara masuk pada proses gasifikasi sekam padi terhadap pembentukan flamable gas. *JTT (Jurnal Teknologi Terapan)*, 6(1), 62-68.
13. Dirgantara, W., Rabi, A., & Muchlis, C. (2021, January). Implementasi IoT untuk Kontrol dan Monitoring Tingkat Kekeuruhan pada Kolam Ikan Hias Menggunakan Metode Fuzzy Sugeno. In *Seminar Nasional Teknologi Fakultas Teknik 2021*.
14. Adi, P. D. P., Sihombing, V., Siregar, V. M. M., Yanris, G. J., Sianturi, F. A., Purba, W., ... & Prasetya, D. A. (2021, April). A performance evaluation of ZigBee mesh communication on the Internet of Things (IoT). In *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)* (pp. 7-13). IEEE.



15. Subairi, S., Aries Boedi Setiawan, A. B., & Tiwikrama, K. (2020). Pemeriksaan suhu tubuh tanpa kontak langsung sebagai pencegahan Covid-19 untuk pengunjung gedung berbasis IOT.



# Pendeteksian Spam pada E-mail menggunakan Pendekatan *Natural Language Processing*

Ikbar Athallah Taufik<sup>1</sup>, Dimas Dzaky Daniswara<sup>2</sup>, Amri Muhaimin<sup>3</sup>

<sup>1, 2, 3</sup> Program Studi Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

<sup>1</sup>[20083010027@student.upnjatim.ac.id](mailto:20083010027@student.upnjatim.ac.id) <sup>2</sup>[20083010006@student.upnjatim.ac.id](mailto:20083010006@student.upnjatim.ac.id) <sup>3</sup>[amri.muhamin.stat@upnjatim.ac.id](mailto:amri.muhamin.stat@upnjatim.ac.id)

Corresponding author email: [20083010027@student.upnjatim.ac.id](mailto:20083010027@student.upnjatim.ac.id)

**Abstract:** *Natural Language Processing (NLP) is a branch of computer science that deals with the processing of human natural language by machines or computers. In this study, detection was performed on a dataset containing spam and non-spam emails. Email or electronic mail is a common communication medium used on the internet as a means for exchanging information. The NLP approach used in this study involves data preprocessing, such as removing punctuation, irrelevant common words, tokenizing, stemming, and others, as well as classification techniques such as Support Vector Classifier (SVC), Naive Bayes, etc. Among various models tested, one model showed a higher precision rate of 0.98 compared to other models. The study indicates that the NLP approach yields better performance in detecting spam compared to other methods. However, further technological advancements and the development of more complex detection methods are still required to improve the performance and accuracy of email spam detection models.*

**Keywords:** *NLP, Spam, Email, Machine Learning, Detection*

**Abstrak:** Natural Language Processing (NLP) adalah sebuah cabang ilmu komputer yang berkaitan dengan pemrosesan bahasa alami manusia oleh mesin atau komputer, pada penelitian ini telah dilakukan pendeteksian pada sebuah dataset yang berisikan spam dan bukan spam pada email. Email atau surel adalah media komunikasi yang umum digunakan dalam internet sebagai sarana seseorang untuk bertukar informasi. Metode yang digunakan dalam pendekatan NLP ini meliputi preprocessing data, seperti penghapusan tanda baca, kata-kata umum yang tidak relevan, tokenize, stemming, dan lainnya, serta teknik-teknik klasifikasi, seperti Support Vector Classifier (SVC), Naive Bayes, dll. Dari berbagai model yang telah dilakukan uji, terdapat satu model yang menunjukkan angka lebih tinggi dari model lainnya dengan presisi 0,98. Penelitian menunjukkan bahwa pendekatan NLP menghasilkan kinerja yang lebih baik dalam mendeteksi spam dibandingkan dengan metode-metode lain. Namun, peningkatan teknologi dan pengembangan metode deteksi yang lebih kompleks masih diperlukan untuk meningkatkan kinerja dan akurasi dari model deteksi spam email.

**Kata kunci:** NLP, Spam, Email, Pembelajaran Mesin, Pendeteksian

## I. PENDAHULUAN

Perkembangan teknologi saat ini terhadap surat menyurat sudah sangat berkembang, manusia sudah jarang sekali ditemukan mengirim surat melalui kantor pos atau manusia tidak lagi menggunakan surat berbentuk fisik hanya untuk berbicara dengan satu sama lainnya. Surat saat ini bukan lagi berbentuk fisik / hardfile melainkan dalam bentuk digital atau yang bisa disebut dengan E-mail. Adanya e-mail memudahkan manusia untuk mengirimkan surat atas dasar kepentingan tertentu kepada seseorang [1].

Email atau surel adalah media komunikasi yang umum digunakan dalam internet sebagai sarana seseorang untuk bertukar informasi dan kebutuhan pribadi. Kemudahan penggunaannya menjadikan email tetap populer hingga saat ini, bahkan digunakan sebagai sarana untuk memverifikasi identitas pengguna aplikasi dan dapat memastikan informasi data pengguna media sosial seperti Tiktok, Instagram, Twitter, dan Facebook benar-benar sama. Meskipun email memiliki dampak positif yang besar, penggunaannya juga bisa terkena dampak dari sisi negatif

email jika tidak menggunakannya dengan baik dan tepat. Banyak penyalahgunaan email yang dapat merugikan pengguna lain, seperti spam atau email sampah yang berisi iklan, scam [2].

Spam yang berarti *stupid pointless annoying messages* juga memiliki beberapa jenisnya diantaranya seperti, iklan, phishing, virus malware, scam dan yang lainnya [3]. Spam juga dapat dibedakan dengan non-spam, terutama dilihat pada subjek dan body dari surat, subjek berarti judul dari isi surat biasanya pada spam judul berisikan kata-kata promosi seperti “Diskon untuk Anda”, dan body berarti isi atau inti dari surat yang dikirim oleh spammer tersebut. [4],

Perkembangan e-mail juga memiliki masalahnya, dimana e-mail pada akun seseorang bisa terkena spam, hal ini mengakibatkan pengguna tidak mengetahui apakah email tersebut berbahaya atau hanya sekedar sampah. Di dalam spam pada email dapat berupa file sampah, dokumen yang tidak penting dan bahkan spam pada email bisa saja menjadi berbahaya jika di dalamnya terdapat virus atau *malware* yang dapat membahayakan pengguna [5].

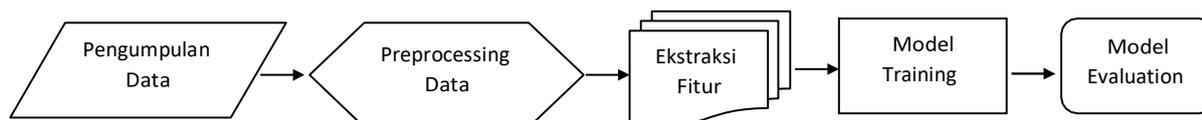
Untuk mengatasi masalah ini, metode deteksi spam email dengan menggunakan pendekatan NLP atau *Natural Language Processing* telah dikembangkan. NLP adalah satu diantara teknologi kecerdasan buatan yang memungkinkan komputer untuk memahami, menganalisis, dan memanipulasi bahasa manusia dengan cara yang mirip seperti manusia. Dalam deteksi spam email, NLP digunakan untuk memeriksa konten pesan email dan mengidentifikasi apakah email tersebut masuk dalam kategori spam atau bukan.

Penjelasan mengenai deteksi spam banyak ditemukan dari berbagai sumber lainnya, dimana mereka mencoba mendeteksi spam dengan bermacam-macam metode untuk menemukan model deteksi yang akurat atau mendapatkan model pengklasifikasian yang presisi, seperti pada [6]. dijelaskan disitu bahwa mereka menemukan tingkat keakuratan model machine learning yang digunakan.

Penelitian yang dilakukan pada artikel ini, membahas mengenai deteksi spam email dengan menggunakan NLP. Selain daripada itu, dalam artikel ini akan membahas juga tentang teknik-teknik NLP yang digunakan dalam deteksi spam email, serta keunggulan dan kelemahan dari metode ini. Penelitian ini diharapkan dapat memberikan manfaat dan kontribusi bagi pengguna email dalam mengatasi masalah spam email dan memaksimalkan produktivitas dalam penggunaan email.

## II. METODE PENELITIAN

Secara umum, setiap pemrosesan pada NLP terdiri dari lima fase utama yaitu, pengumpulan data, pra-pemrosesan data, ekstraksi fitur, pelatihan model, dan evaluasi model. Gambar di bawah menunjukkan alur untuk fase-fase tersebut yang digunakan sebagai metode dalam pendeteksian spam pada e-mail. [7]



**Gambar 1** Alur metode penelitian

Berikut adalah penjelasan metode penelitian yang dapat dilakukan dalam pendeteksian spam email dengan Natural Language Processing (NLP):

### 2.1 Pengumpulan Data

Pertama-tama, data email yang akan digunakan untuk pelatihan dan pengujian algoritma



deteksi spam email harus dikumpulkan. Data ini harus mencakup berbagai jenis email, termasuk email spam dan non-spam. Pengumpulan data dapat dilakukan dengan melakukan crawling pada berbagai sumber atau menggunakan dataset yang tersedia secara publik. Pada Artikel ini data diperoleh dari situs publik yang memberikan akses untuk mengambil dataset [8], Dari dataset terhitung keseluruhan data memiliki 6046, dengan total spam sebanyak 1896, dan ham sebanyak 4150, data yang terhitung dimuali dari 0 sampai 6046 [9].

## 2.2 Data Preprocessing

Setelah data email dikumpulkan, langkah selanjutnya adalah melakukan preprocessing data. Hal ini meliputi cleaning, penghapusan karakter-karakter khusus, konversi semua huruf kecil, penghapusan kata-kata umum, dan lain sebagainya. Data email yang sudah dipreprocessing ini akan membantu dalam membangun model deteksi spam email yang akurat [10].

### 2.2.1 Cleaning data:

Pembersihan pada data diperlukan karena data yang tidak bearturan atau tidak terstruktur dapat mempengaruhi hasil akurasi maka, dilakukanlah pembersihan pada data dari simbol, tanda baca, hashtag, karakter khusus, dan karakter lainnya yang mengganggu struktur kata pada data, juga melakukan konversi teks ke huruf kecil [4]. Kata-kata semisal “Buku” dan “buku” mempunyai makna yang sama, namun apabila tidak dikonversi menjadi huruf kecil, kedua kata tersebut direpresentasikan sebagai dua kata yang berbeda dalam model ruang vektor (yang menghasilkan lebih banyak dimensi) [11].

### 2.2.2 Tokenize:

Tokenize memecah kalimat menjadi kata-kata sehingga membuat dokumen terpecah menjadikannya bagian-bagian yang lebih kecil sehingga pada saat proses analisa nantinya akan lebih mudah [3].

### 2.2.3 Menghilangkan stopwords:

Menghapus kata-kata yang tidak memiliki pengertian jika tidak ada kata lain seperti: dan, saya, atau. [4]

### 2.2.4 Stemming data:

Mengubah kata-kata yang memiliki imbuhan menjadi tidak ada imbuhan atau kembali menjadi kata dasar aslinya.[4]

## 2.3 Feature Extraction

Setelah data email dipreprocessing, langkah selanjutnya adalah melakukan ekstraksi fitur. Fitur-fitur yang dapat digunakan dalam deteksi spam email meliputi kata kunci tertentu, frekuensi kemunculan kata, dan panjang email. Fitur-fitur ini akan menjadi input dalam model deteksi spam email. [4]

## 2.4 Pembuatan Model

Setelah fitur-fitur diekstraksi, selanjutnya adalah membangun model deteksi spam email. Model dapat dibuat dengan menggunakan algoritma Machine Learning seperti Gaussian NB (NaïveBayes), Multinomial NB, Bernouliie NB, Decision Tree, KNeighbors, Support Vector Classifier (SVC), Logistic Regression, dan (SGDC) Stochastic Gradient Descent Classifier. Model ini akan dilatih menggunakan data email yang sudah dipreprocessing dan diekstraksi



fiturnya.

### 2.5 Evaluasi Model

Setelah model selesai dibuat, langkah selanjutnya adalah melakukan evaluasi model. Evaluasi ini bertujuan untuk mengukur kinerja model dalam mendeteksi email spam dan non-spam. Metrik yang dapat digunakan untuk evaluasi model meliputi akurasi, dan presisi [6].

### 2.6 Uji Coba

Setelah model berhasil diuji dan terbukti akurat dalam mendeteksi spam email, langkah selanjutnya adalah mengujinya pada dataset email yang belum pernah dilihat sebelumnya. Uji coba ini akan membantu untuk mengevaluasi kemampuan model dalam mendeteksi spam email secara umum.

Dengan menggunakan metode penelitian di atas, diharapkan dapat menghasilkan model deteksi spam email yang akurat dan dapat membantu pengguna email dalam mengatasi masalah spam email.

## III. HASIL DAN PEMBAHASAN

Deteksi spam email menggunakan Natural Language Processing (NLP) merupakan salah satu metode yang dapat digunakan untuk memfilter email yang masuk ke dalam kotak masuk pengguna. Dalam penelitian ini, dilakukan pengimplementasian beberapa teknik NLP seperti klasifikasi teks dan pemrosesan bahasa alami untuk melakukan deteksi spam email.

Pada tahap awal, dilakukan pengumpulan data email yang bersumber dari web dimana data sudah dalam bentuk dataset, dari dataset tersebut dilakukan import ke dalam bahasa pemrograman yang dilanjut dengan mengimport beberapa library yang dibutuhkan untuk melakukan pengerjaan NLP. Data yang sudah masuk dapat dilakukan analisis sederhana dengan melihat variabel, besarnya data, dan tipe dari data. Data kemudian dilakukan pembersihan untuk menghindari kesalahan dalam melakukan proses NLP, dilakukan juga exploratory data analysis untuk mengetahui diagram dan heat map pada data.

Pemrosesan berikutnya yaitu dengan melakukan preprocessing pada data tersebut seperti konversi ke lowercase, tokenize, menghapus karakter khusus, stopword removal, dan stemming untuk memperoleh data yang lebih bersih dan terstruktur. Selanjutnya, dilakukan proses ekstraksi fitur menggunakan metode TF-IDF untuk mengonversi setiap email menjadi vektor fitur yang dapat digunakan oleh model [3]. Proses ekstrasi yang telah selesai dilanjutkan dengan membagi data menjadi data untuk dilatih dan data untuk percobaan.

Pengerjaan dari tahap-tahap sebelumnya yang telah berhasil dilakukan, membawa pada tahap selanjutnya dimana, setelah data melewati beberapa metode sehingga, didapat data train dan data test maka, dilakukan implementasi model yang mana, mengimplementasikan beberapa algoritma klasifikasi seperti Gaussian NB, Multinomial NB, Bernoulli NB, Decision Tree, KNeighbors, Support Vector Classifier (SVC), Logistic Regression, dan SGDClassifier untuk melakukan pengklasifikasian spam dan bukan spam pada email dari dataset sehingga, dapat dilihat hasil akurasi dan presisi dari masing-masing model.

**Tabel 1.** Tabel Hasil Modeling

Model	Accuracy	Precision
GaussianNB	0,90	0,83



MultinomialNB	0,93	0,89
BernoulliNB	0,94	0,98
SVC	0,95	0,90
KNeighbors	0,39	0,33
Decision Tree	0,88	0,79
Logistic Regression	0,93	0,88
SGDClassifier	0,96	0,90

Evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi, dan presisi. Hasil evaluasi menunjukkan bahwa algoritma Bernoulli Naive Bayes dengan akurasi 0,94 dan presisi sebesar 0,98 memberikan kinerja yang lebih baik dibandingkan dengan algoritma lainnya. Akurasi dan presisi terkecil ada pada model KNeighbors dengan akurasi 0,39 dan presisi sebesar 0,33 sehingga, model ini kurang baik untuk dapat dijadikan model pendeteksian spam pada e-mail.

Dalam penelitian ini, dipertimbangkan juga kelemahan dari teknik deteksi spam email dengan NLP. Salah satu kelemahan utama adalah kurangnya kemampuan untuk memahami konteks email dan konten yang dinamis. Hal ini dapat mengakibatkan kesalahan klasifikasi email yang seharusnya dianggap sebagai spam namun terdeteksi sebagai email biasa dan sebaliknya.

#### IV. KESIMPULAN

Dalam kesimpulannya, pendekatan NLP menjadi salah satu metode yang efektif dalam mendeteksi spam pada email. Dalam pengembangan teknologi di masa depan, pendekatan NLP ini dapat dikembangkan lebih lanjut untuk mengatasi masalah-masalah baru dan meningkatkan keamanan informasi pada email.

Pendekatan NLP dapat memberikan hasil yang lebih akurat dalam pendeteksian spam pada email dibandingkan dengan pendekatan konvensional. Namun, peningkatan teknologi dan pengembangan metode deteksi yang lebih kompleks masih diperlukan untuk meningkatkan kinerja dan akurasi dari model deteksi spam email.

Secara keseluruhan, teknik deteksi spam email dengan NLP dapat memberikan solusi yang efektif untuk mengatasi masalah spam email. Dalam pengembangan teknologi di masa depan, pendekatan NLP ini dapat dikembangkan lebih lanjut untuk mengatasi masalah-masalah baru dan meningkatkan keamanan informasi pada email.

Selain itu, dalam perkembangannya teknik NLP dapat dikembangkan lebih lanjut dengan menggunakan metode deep learning seperti convolutional neural network (CNN) dan recurrent neural network (RNN) untuk meningkatkan akurasi dan efektivitas pendeteksian spam pada email.

#### UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada semua yang telah membantu dalam menyelesaikan artikel ini. Terima kasih kepada orang tua yang memberikan dukungan doa dan memfasilitasi untuk bisa menyelesaikan penulisan ini. Ucapan terimakasih kepada dosen-dosen yang memberikan masukan dan ilmu yang berharga. Terima kasih kepada tim yang selalu mendukung dalam menyelesaikan penulisan ini.

Ucapan terima kasih juga kepada pembaca artikel ini yang telah meluangkan waktunya untuk membaca tulisan ini. Semoga artikel ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu pengetahuan di bidang Sains Data atau bidang yang berkaitan dengan pembahasan tulisan ini

**REFERENSI**

- [1] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, “Machine learning for email spam filtering: review, approaches and open research problems,” *Heliyon*, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.
- [2] F. Z. Ruskanda, “Study on the Effect of Preprocessing Methods for Spam Email Detection,” *Indones. J. Comput.*, vol. 4, no. 1, p. 109, 2019, doi: 10.21108/indojc.2019.4.1.284.
- [3] E. P. Laksono and A. Wicaksono, “Penyaringan Spam email menggunakan,” vol. 5, no. 2, pp. 26–32.
- [4] F. Rahma, A. Z. Farmadiansyah, and A. F. Hidayatullah, “Deteksi Surel Spam dan Non Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes,” *Automata*, vol. 2, no. 2, 2021.
- [5] R. S. Lutfiyani and N. Retnowati, “Implementasi Pendeteksian Spam Email Menggunakan Metode TextMining Dengan Algoritma Naïve Bayes Dan Decision Tree J48,” *J. Komput. dan Inform.*, vol. 9, no. 2, pp. 244–252, 2021, doi: 10.35508/jicon.v9i2.5304.
- [6] S. K. Arts and S. K. Arts, “Performance Evaluation of Machine Learning Algorithms for Email SpamDetection,” pp. 1–4, 2020.
- [7] I. AbdulNabi and Q. Yaseen, “Spam email detection using deep learning techniques,” *Procedia Comput. Sci.*, vol. 184, no. 2019, pp. 853–858, 2021, doi: 10.1016/j.procs.2021.03.107.
- [8] nitisha bharathi, “Email Spam Dataset,” *kaggle*, 2020. <https://www.kaggle.com/datasets/nitishabharathi/email-spam-dataset>.
- [9] N. Q. Fitriyah, H. Oktavianto, and H. Hasbullah, “Deteksi Spam Pada Email Berbasis Fitur Konten Menggunakan Naïve Bayes,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 5, no. 1, pp. 1–7, 2020, doi: 10.32528/justindo.v5i1.3414.
- [10] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [11] M. A. Ghani and H. Sulaiman, “Deteksi Spam Email dengan Metode Naive Bayes dan Particle Swarm Optimization (PSO),” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 11–20, 2023, doi: 10.29408/jit.v6i1.7049.



# Analisis *Topic Modelling* pada Ulasan Aplikasi Shopee di PlayStore Menggunakan *Latent Dirichlet Allocation (LDA)*

Sahat Renaldi. S<sup>1</sup>, Primus Akbar Atnanda<sup>2</sup>, Amri Muhaimin<sup>3</sup>

<sup>1,2,3</sup> Program Studi Sains Data, UPN "Veteran" Jawa Timur

<sup>2</sup>[20083010026@student.upnjatim.ac.id](mailto:20083010026@student.upnjatim.ac.id)

<sup>3</sup>[amri.muhaimin.stat@upnjatim.ac.id](mailto:amri.muhaimin.stat@upnjatim.ac.id)

Corresponding author email: [20083010026@student.upnjatim.ac.id](mailto:20083010026@student.upnjatim.ac.id)

**Abstract:** *Shopee is a popular e-commerce platform in Indonesia. To improve service quality, data analysis is needed to understand user responses and preferences. This study aims to conduct data analysis and topic modelling using the Latent Dirichlet Allocation (LDA) method on 33,896 user review data of the Shopee application on PlayStore. LDA modelling was performed by considering the parameters passes and iterations. Passes with values of 5, 10, 15, and 20 were tested with a combination of iterations values of 50 and 100 and a random state of 142. During the testing, it was found that the highest average coherence score was achieved by using passes 10 and iterations 50. Therefore, passes 10 and iterations 50 were selected as the final LDA model parameters, resulting in seven topics. The topics that emerged from the analysis include user satisfaction, shipping and product suitability, user trust, user experience, service and support, application, and promotional offers. The "application" topic was found to be the most critical and received the highest number of rating 1. High-weighted words on the application topic such as "slow," "heavy," and "crash" indicate that issues related to the application's system performance and functionality need to be addressed and improved.*

**Keywords:** *Shopee, Modelling, Latent Dirichlet Allocation*

**Abstrak:** *Shopee merupakan salah satu platform e-commerce yang populer di Indonesia. Dalam upaya untuk meningkatkan kualitas layanan, analisis data perlu dilakukan untuk memahami respon dan preferensi pengguna terhadap Shopee. Penelitian ini bertujuan untuk melakukan analisis data dan pemodelan topik dengan metode Latent Dirichlet Allocation (LDA) pada 33.896 data ulasan pengguna aplikasi Shopee di PlayStore. Pemodelan LDA dilakukan dengan mempertimbangkan parameter passes dan iterations. Parameter passes dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai iterations 50 dan 100 serta random state 142. Dalam pengujian, ditemukan bahwa nilai rata-rata coherence score tertinggi terdapat pada kombinasi nilai passes 10 dan iterations 50. Nilai passes 10 dan iterations 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik. Adapun topik-topik yang terbantu antara lain topik kepuasan pengguna, topik pengiriman dan kesesuaian barang, topik kepercayaan pengguna, topik pengalaman pengguna, topik layanan dan bantuan, topik aplikasi, dan topik terkait promosi dan penawaran. Topik "aplikasi" menjadi topik yang paling kritis dan mendapatkan rating 1 terbanyak. Kata-kata dengan bobot tinggi pada topik aplikasi seperti "lambat", "berat", dan "buka" menunjukkan bahwa masalah terkait kinerja sistem dan fungsionalitas aplikasi menjadi salah satu isu yang perlu dibenahi.*

**Kata kunci:** *Shopee, Pemodelan, Latent Dirichlet Allocation*

## I. PENDAHULUAN

Belanja online telah menjadi gaya hidup banyak orang di zaman digital sekarang ini. Berbicara tentang belanja online, Shopee merupakan salah satu platform belanja online terkemuka di Indonesia dengan jumlah pengguna aktif yang terus meningkat setiap tahunnya. Berdasarkan hasil survei dalam laporan Status Literasi Digital Indonesia tahun 2021 yang dikeluarkan oleh Kementerian Komunikasi dan Informatika dan Katadata Insight Center (KIC), Shopee merupakan *e-commerce* yang paling banyak digunakan masyarakat di Indonesia [1]. Shopee menyediakan berbagai layanan yang memudahkan pengguna dalam melakukan transaksi pembayaran, pengiriman, maupun pengembalian barang. Shopee juga menawarkan promo menarik dan program loyalitas, seperti melalui *flash sale*, voucher gratis ongkir, *Shopee coins*, dan program lainnya. Hal ini membuat Shopee semakin diminati oleh pengguna dan memperkuat posisi Shopee sebagai salah satu platform belanja online di Indonesia.



Dalam lingkup persaingan bisnis yang semakin kompetitif, Shopee harus selalu memperhatikan kebutuhan dan preferensi pengguna. Ulasan pengguna Shopee yang ada di PlayStore dapat menjadi salah satu sumber informasi bagi perusahaan untuk mengevaluasi kualitas layanan yang diberikan. Diharapkan dari data ulasan yang ada dapat menjadi bahan pembelajaran dan peningkatan pelayanan bagi pengguna. Salah satu cara yang dapat digunakan untuk menganalisis data teks ulasan pengguna adalah dengan *topic modelling*. *Topic modelling* adalah metode yang dapat digunakan untuk mengekstrak topik atau tema dari kumpulan dokumen teks yang besar dan kompleks. Dalam konteks ulasan pengguna Shopee di PlayStore, *topic modelling* dapat digunakan untuk mengidentifikasi topik atau tema yang sering dibicarakan oleh pengguna. Salah satu metode *topic modelling* yang umum digunakan adalah *Latent Dirichlet Allocation (LDA)*. LDA mengasumsikan bahwa setiap dokumen terdiri dari beberapa topik yang berbeda dan setiap kata dalam dokumen tersebut berasal dari salah satu topik tersebut

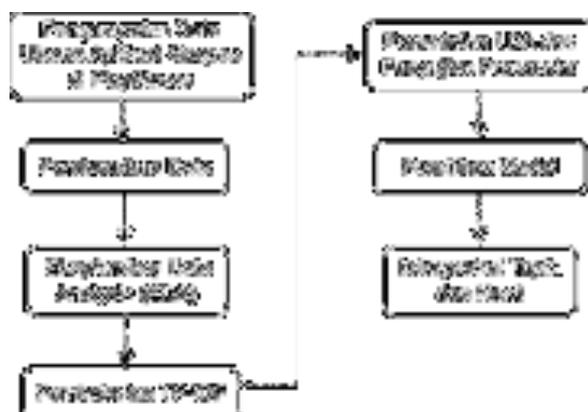
Beberapa penelitian terdahulu terkait *topic modelling* yang menggunakan LDA antara lain penelitian yang dilakukan oleh Habibi et al. (2021) menganalisis *caption* konten terkait Gerakan Masyarakat Hidup Sehat (GERMAS) di Instagram. Berdasarkan hasil analisis konten pada masing-masing segmen topik, ditemukan bahwa topik yang paling dominan terkait dengan GERMAS adalah pola makan hidup sehat [2]. Penelitian yang dilakukan oleh Destarani et al. (2019) memodelkan data keluhan yang dilaporkan melalui situs web keluhan online masyarakat Denpasar menggunakan LDA dengan estimasi parameter Gibbs *sampling*. Jumlah topik yang diperoleh melalui nilai log-likelihood tertinggi adalah sebesar -42.528,84, nilai tersebut berada pada jumlah topik 19. Topik yang sedang tren didasarkan pada probabilitas topik tertinggi, yakni nilai probabilitas topik 0,055 yang dapat diinterpretasikan bahwa banyak penduduk Denpasar mengeluhkan tentang jalan rusak dan meminta untuk memperbaiki jalan [3]. Sutherland et al. (2020) meneliti 104.161 data ulasan online dari pelanggan akomodasi di Korea. Dengan menggunakan model LDA, penelitian ini berhasil mengekstrak dan memvalidasi topik-topik penting yang menjadi perhatian tamu. Implikasi dari temuan ini dapat membantu manajer dalam meningkatkan kepuasan pelanggan [4].

Berdasarkan hasil tinjauan pada penelitian-penelitian sebelumnya, metode *Latent Dirichlet Allocation (LDA)* mampu memodelkan data teks ke dalam topik-topik tertentu. Oleh karena itu, peneliti tertarik untuk menggunakan LDA dalam memodelkan data ulasan pengguna aplikasi Shopee di PlayStore. Penelitian ini menggunakan data ulasan yang terhitung sejak tanggal 1 April 2023 sampai 20 April 2023. Setelah dilakukan pembersihan pada data mentah, didapatkan 33.896 record data yang akan digunakan dalam pemodelan. Penelitian ini mempertimbangkan parameter *passes* dan *iterations* dalam pemodelan dengan random state 142. Untuk memilih *passes* dan *iteration* final yang akan digunakan, terlebih dahulu dilakukan pengujian dengan mengkombinasikan nilai *passes* dan *iterations*. Jumlah topik ( $k$ ) yang akan digunakan dalam model akhir dilihat K-Optimum yang paling sering muncul dari hasil kombinasi parameter. Parameter *passes* 5, 10, 15, 20 akan dikombinasikan dan diuji dengan *iterations* 50, dan 100. Parameter *passes* dan *iterations* yang dipilih sebagai model final adalah parameter dengan rata-rata *coherence score* tertinggi. Setelah dilakukan pemodelan dengan metode LDA, topik-topik yang dihasilkan akan dianalisis dan diinterpretasikan. Hasil analisis tersebut diharapkan dapat memberikan wawasan yang berguna bagi Shopee dalam meningkatkan kualitas layanan dan produk.”

## II. METODE PENELITIAN

Dalam penelitian ini, alur penelitian dimulai dari proses pengumpulan data, pembersihan data, Ekploratory Data Analysis (EDA), pembobotan dengan TF-IDF, pemodelan LDA dan pengujian

parameter, pemilihan model, interpretasi topik dan hasil.



**Gambar 1** Alur Penelitian

### 3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan dari hasil proses crawling data PlayStore pada ulasan aplikasi Shopee dengan menggunakan library *google\_play\_scraper* yang ada di Python. Data ulasan yang diambil terhitung sejak tanggal 1 April 2023 sampai 20 April 2023.

### 3.2 Pembersihan Data

Data teks ulasan aplikasi Shopee yang diperoleh dari proses crawling merupakan data mentah yang masih mengandung unsur-unsur yang tidak diperlukan. Pembersihan data dilakukan sebagai proses persiapan sebelum data dilakukan pemodelan. Pembersihan data teks perlu dilakukan untuk menghilangkan elemen yang tidak relevan atau tidak memberikan makna yang berarti. Tujuannya adalah agar proses pemodelan menjadi lebih efektif dan efisien. Pembersihan data mencakup proses tokenisasi, *replace slank* atau mengganti kata atau frasa yang bersifat slang atau bahasa gaul dengan kata atau frasa yang lebih baku, penghapusan emoji, *stemming*, *case folding*, *stopword*, penghapusan simbol-simbol, tanda baca dan unsur lainnya yang tidak terlalu memiliki kepentingan. Setelah dilakukan pembersihan data, maka data siap digunakan dalam proses pemodelan untuk memetakan data ke dalam topik-topik tertentu. Proses pembersihan data yang dilakukan dalam penelitian ini diilustrasikan seperti pada Tabel 1 dibawah ini.

**Tabel 1.** Tahap Pembersihan

Normalisasi	Sampel
Contoh	Semua seller amanah / sportif dlm mengirimkan barang
Tokenisasi	'Semua', 'seller', 'amanah', '/', 'sportif', 'dlm', 'mengirimkan', 'barang', '
Replace slank	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'mengirimkan', 'barang', ' '
Penghapusan emoji	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'mengirimkan', 'barang'
Stemming	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'kirim', 'barang'
Case folding	semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'kirim', 'barang'
Stopword	'penjual', 'amanah', '/', 'sportif', 'kirim', 'barang'
Simbol dan tanda baca	'penjual', 'amanah', 'sportif', 'kirim', 'barang'



Hasil akhir	penjual amanah sportif kirim barang
-------------	-------------------------------------

### 3.3 *Exploratory Data Analysis (EDA)*

Setelah melalui tahap pembersihan data, pada data teks ulasan aplikasi Shopee kemudian dilakukan analisis data eksploratif untuk mengeksplorasi struktur, pola, dan karakteristik dari data yang ada. Pada tahap ini, dilakukan beberapa analisis seperti analisis perolehan rating, frekuensi kata, dan visualisasi wordcloud untuk memperoleh *insight* atau pemahaman awal mengenai data yang ada.

### 3.4 *Pembobotan TF-IDF*

*Term Frequency - Invers Document Frequency* (TF-IDF) adalah suatu proses untuk memberikan bobot relasi suatu kata (*term*) dengan dokumen. Proses TF-IDF mengkombinasikan dua skema perhitungan bobot, yakni frekuensi kemunculan suatu kata di dalam suatu dokumen dan *inverse* frekuensi dokumen yang memuat kata tersebut. Metode TF-IDF menentukan frekuensi relatif kata-kata dalam dokumen tertentu melalui *invers* proporsi kata di seluruh korpus dokumen. *Invers Document Frequency* adalah metrik untuk menentukan seberapa jarang suatu kata didasarkan pada suatu dokumen [5]. Berikut adalah persamaan untuk menghitung nilai IDF dan bobot ( $w$ ) masing-masing dokumen :

$$IDF_t = \log_{10} (D/df_i) \quad (1)$$

$$W_{d,t} = tf_{d,t} (IDF_t) \quad (2)$$

Keterangan :

$D$  = jumlah dokumen yang berisi term ( $t$ )

$df_i$  = jumlah kemunculan kata terhadap  $D$

$W_{d,t}$  = bobot dokumen ke- $d$  pada kata ke- $t$

$tf_{d,t}$  = frekuensi kata

### 3.5 *Pemodelan LDA dan Pengujian Parameter*

*Latent Dirichlet Allocation* (LDA) merupakan suatu model probabilistik yang digunakan untuk mengenali topik-topik yang terkandung dalam kumpulan dokumen teks. Model ini berdasarkan asumsi bahwa setiap dokumen terdiri dari kombinasi beberapa topik, dan setiap kata dalam dokumen berasal dari salah satu topik tersebut. LDA dapat digunakan untuk menyimpulkan, mengelompokkan, menghubungkan, dan memproses koleksi data besar karena menghasilkan daftar topik yang memiliki bobot untuk setiap dokumen dalam koleksi [6]. Dalam penelitian ini, parameter *passes* dan *iterations* dipertimbangkan untuk pemodelan dengan random state 142. Parameter *passes* mengontrol jumlah iterasi yang dilakukan oleh algoritma untuk melatih model LDA pada dataset yang diberikan. Sementara itu, *iterations* merujuk pada jumlah kali yang dilakukan oleh algoritma Gibbs *sampling* untuk menghasilkan perkiraan topik dari dokumen yang diberikan. Untuk menentukan nilai *passes* dan *iterations* final yang akan digunakan dalam penelitian ini, dilakukan pengujian dengan mengkombinasikan beberapa nilai *passes* dan *iterations*. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100.

### 3.6 *Pemilihan Model*

Jumlah topik ( $k$ ) yang akan digunakan dalam model akhir dilihat dari nilai  $k$ -optimum yang paling sering muncul dari hasil kombinasi parameter. Setelah itu, parameter *passes* dan *iterations* yang memberikan rata-rata *coherence score* tertinggi dipilih sebagai model final (metrik *coherence score* yang digunakan adalah  $C_V$ ). Dengan cara ini, model LDA yang dihasilkan diharapkan mampu memberikan representasi topik yang akurat dan optimal untuk dokumen yang diberikan.

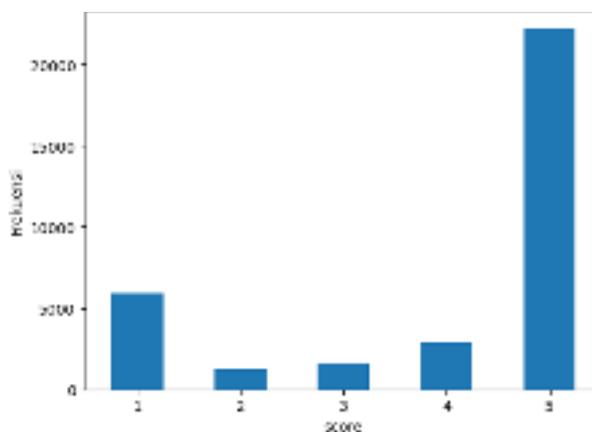
### 3.7 Interpretasi Topik dan Hasil

Setiap topik dalam LDA direpresentasikan sebagai distribusi kata, dan kata-kata yang memiliki probabilitas tertinggi dalam distribusi tersebut dianggap sebagai kata-kata yang paling merepresentasikan topik tersebut. Interpretasi topik dalam LDA dilakukan dengan melihat kata-kata yang paling berkontribusi dalam setiap topik. Setelah didapatkan topik-topik, kemudian model diaplikasikan pada keseluruhan data untuk menggali *insight* yang lebih mendalam.

## III. HASIL DAN PEMBAHASAN

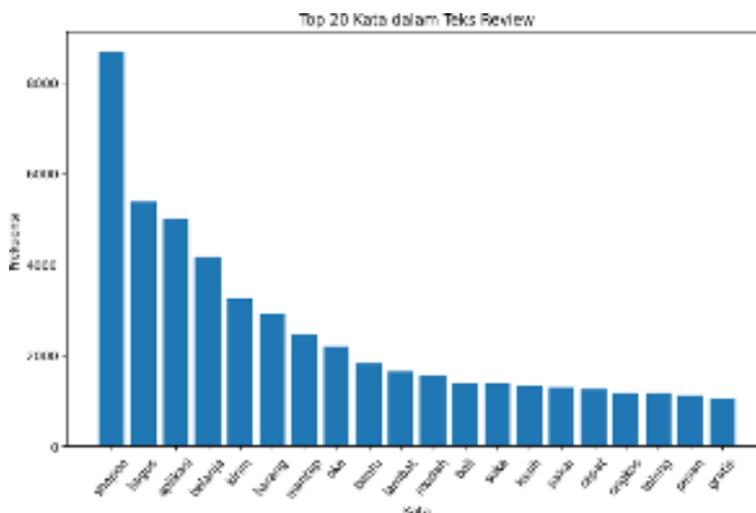
### 3.1. Data Ulasan Pengguna

Penelitian ini menggunakan data ulasan pengguna aplikasi Shopee di PlayStore yang terhitung sejak tanggal 1 April 2023 sampai 20 April 2023. Setelah dilakukan pembersihan pada data mentah, didapatkan 33.896 record data bersih yang akan masuk dalam proses pemodelan.



**Gambar 2.** Rating di PlayStore

Berdasarkan Gambar 2, terlihat bahwa distribusi rating yang pada aplikasi Shopee memiliki puncak tertinggi pada rating 5. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang sangat baik terhadap Shopee. Namun, terdapat juga sejumlah pengguna yang memberikan rating 1 dan rating 2. Meskipun jumlah yang memberikan rating 1 dan rating 2 lebih sedikit dibandingkan dengan pengguna yang memberikan rating 5, perlu perhatian khusus untuk mengevaluasi alasan di balik rating tersebut dan memperbaiki masalah yang mungkin dihadapi oleh pengguna.



**Gambar 3.** Top 20 Kata dengan Frekuensi Tertinggi

Barplot pada Gambar 3 memberikan gambaran mengenai kata-kata yang sering muncul dalam keseluruhan data ulasan aplikasi Shopee di PlayStore yang dapat membantu dalam memahami hal yang diungkapkan pengguna secara umum. Kata ‘shopee’, ‘bagus’, ‘aplikasi’, ‘belanja’, ‘kirim’, merupakan lima kata teratas dengan kemunculan paling sering dalam data. Walaupun frekuensi kemunculan suatu kata dapat memberikan gambaran tentang seberapa sering kata tersebut digunakan dalam dokumen, tetapi ini tidak cukup digunakan untuk memahami topik yang sebenarnya dibahas oleh pengguna. Perlu dilakukan analisis lebih lanjut untuk memahami konteks kata-kata yang digunakan dalam dokumen ulasan tersebut.

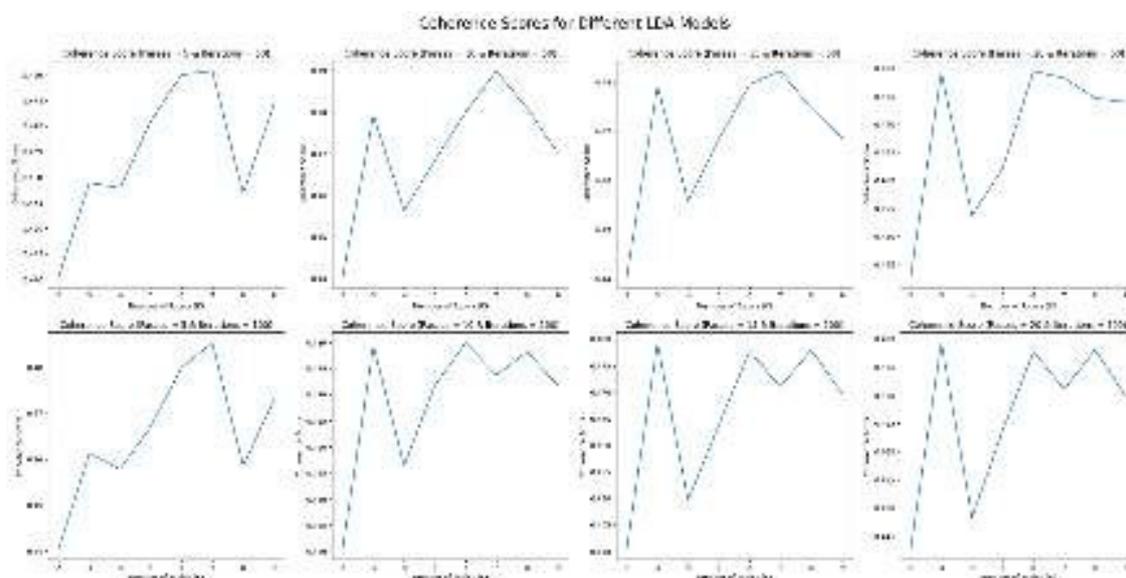


**Gambar 4.** Wordcloud

Visualisasi wordcloud pada Gambar 4 juga menampilkan beberapa kata kunci yang paling sering muncul dalam ulasan pengguna aplikasi Shopee. Dalam wordcloud tersebut juga terlihat hal-hal detail yang sebelumnya tidak diketahui, seperti kata ‘tolong’, ‘kecewa’, ‘lambat’.

#### 4.2 Pemodelan

Penelitian ini mengimplementasikan algoritma *Latent Dirichlet Allocation* (LDA) pada data ulasan pengguna Shopee di PlayStore untuk melakukan topik modelling, yaitu mengidentifikasi topik-topik yang dibicarakan dalam ulasan pengguna dan menganalisis hubungan antar topik. Parameter *passes* dan *iterations* dipertimbangkan dalam pemodelan. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100 serta *random state* 142.



**Gambar 5.** Coherence Score untuk Setiap Parameter Pengujian

Jumlah topik ( $k$ ) yang akan digunakan dalam model akhir dilihat dari nilai  $k$ -optimum yang paling sering muncul dari hasil kombinasi parameter. Setelah itu, parameter *passes* dan *iterations* yang memberikan rata-rata *coherence score* tertinggi dipilih sebagai parameter model final.

**Tabel 2.** Hasil Pengujian

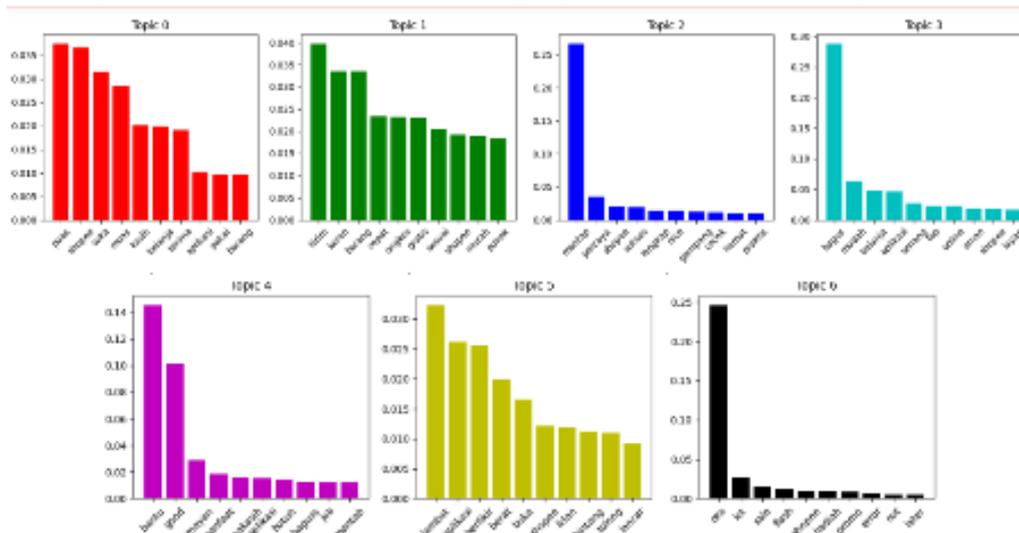
Passes	Iterations	Rata-rata	K-Optimum
5		5	
10	50	10	50
15		0.4686	7
20		0.468	6
5		0.4654	7
10	100	10	100
15		0.4661	3
20		0.4663	3

**Tabel 3.** Coherence Value (Passes = 10, Iterations = 50)

K	Coherence Value
2	0.4405
3	0.4793
4	0.4564
5	0.4684
6	0.4799
7	0.4898
8	0.4809
9	0.4703

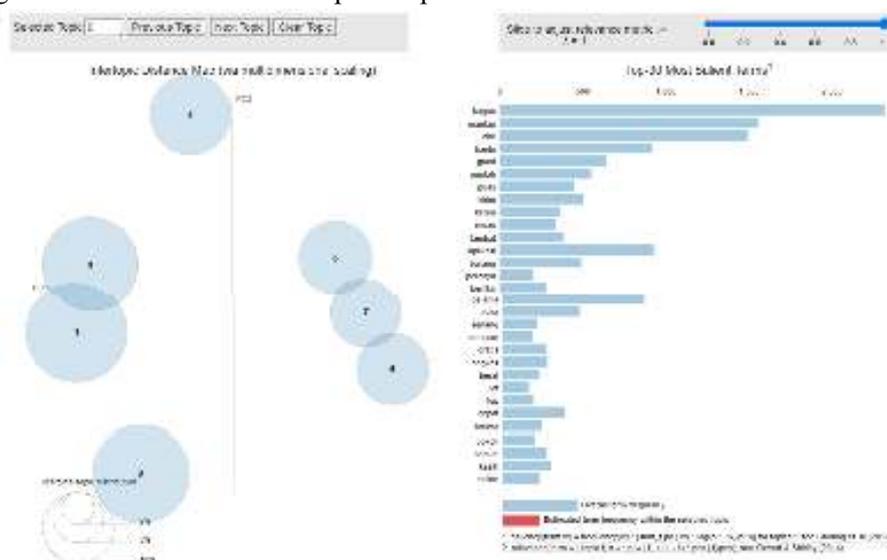
Hasil pengujian seperti yang ditunjukkan pada Tabel 7 menunjukkan bahwa nilai rata-rata *coherence score* tertinggi terdapat pada kombinasi nilai *passes* 10 dan nilai *iterations* 50 dengan nilai rata-rata *coherence score* sebesar 0.47068. *Coherence score* merupakan ukuran kualitas interpretasi model LDA. Oleh karena itu, memilih parameter model yang memberikan nilai rata-rata coherence

score tertinggi adalah langkah yang tepat menurut peneliti. Tabel 7 juga memberikan nilai K-Optimum yang paling sering muncul adalah 7. Oleh karena itu, nilai *passes* sebesar 10 dan nilai *iterations* sebesar 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik.



**Gambar 6.** Topik Utama berdasarkan Probabilitas Kata

Gambar 6 menunjukkan sebaran kata-kata di setiap topik yang terbentuk dan bobot kata-kata yang paling berkontribusi. Bobot ini menunjukkan seberapa penting kata tersebut dalam topik dan dihitung sebagai probabilitas kemunculan kata tersebut dalam topik. Sebagai contoh, untuk topik 0, kata-kata yang paling berkontribusi adalah “puas”, “shopee”, “suka”, “muas”, dan seterusnya, dengan bobot probabilitas tertentu untuk setiap kata. Semakin tinggi nilai probabilitasnya, semakin besar kemungkinan kata tersebut muncul pada topik tersebut.



**Gambar 7.** PyLDAvis secara Umum

Gambar 7 pada panel kiri menampilkan diagram gelembung dengan dimensi yang menggambarkan jarak antara topik dan area gelembung yang sesuai dengan prevalensi relatif topik di dalam korpus. Setiap gelembung mewakili satu topik, semakin besar gelembung maka semakin



tinggi persentase jumlah ulasan dalam korpus tentang topik tersebut. Jika melihat gelembung yang muncul pada tujuh topik yang terbentuk, dapat diketahui bahwa terdapat beberapa topik yang area gelembungnya memiliki ukuran lebih kecil dibandingkan topik lainnya. Oleh karena itu, frekuensi antara topik tersebut dengan topik lainnya sedikit berbeda dalam bobotnya. Selain itu juga terlihat beberapa gelembung yang saling beririsan, fenomena ini menunjukkan adanya kemungkinan bahwa antara topik yang satu dengan topik lainnya memiliki sedikit kesamaan atau kemiripan.

#### 4.3 Interpretasi Topik

Berdasarkan hasil pemodelan yang telah dilakukan, ditemukan bahwa terdapat tujuh topik yang muncul pada data ulasan pengguna Shopee. Tiap topik memiliki sejumlah kata kunci yang memiliki bobot yang berbeda-beda. Bobot kata-kata tersebut menunjukkan tingkat kepentingan kata tersebut dalam konteks topik tertentu. Bobot ini dapat dianggap sebagai representasi numerik dari tingkat signifikansi kata dalam topik tersebut dibandingkan dengan kata-kata di topik lainnya. Bobot yang semakin tinggi menunjukkan semakin kuat hubungan antara kata-kata tersebut dengan topik yang sedang dibahas. Oleh karena itu, analisis bobot kata dalam topik dapat memberikan wawasan tentang topik yang dibicarakan dalam ulasan pengguna Shopee, dan dapat digunakan sebagai dasar untuk mengidentifikasi preferensi pelanggan dalam konteks pengalaman mereka dengan Shopee. Dengan mempertimbangkan bobot kemunculan kata di setiap topik dan diagram PyLDAvis, didapatkan inti dari setiap topik sebagai berikut :

- Topik pertama (0) terkait dengan **kepuasan pengguna**. Kata-kata seperti "puas", "suka", "kasih", dan "terima" menunjukkan kepuasan dari pengguna terhadap Shopee. Kepuasan pengguna merupakan faktor penting dalam bisnis *e-commerce*. Kepuasan pengguna dapat berdampak pada retensi pengguna, loyalitas, dan meningkatkan reputasi bisnis. Sehingga, topik ini perlu diperhatikan sebagai tolok ukur kepuasan pengguna.
- Topik kedua (1) dapat dikaitkan dengan aspek **pengiriman dan kesesuaian barang** yang diterima oleh pengguna. Terdapat kata-kata kunci seperti "kirim", "barang", dan "sesuai" yang menunjukkan bahwa para pengguna memberikan penilaian terhadap kualitas pengiriman dan kesesuaian barang yang mereka terima dari Shopee. Selain itu, kata-kata seperti "cepat", "ongkos", dan "gratis" juga muncul dalam topik ini, mengindikasikan bahwa para pengguna sangat memperhatikan waktu pengiriman dan biaya pengiriman yang dikenakan.
- Topik ketiga (2) menggambarkan **kepercayaan pengguna** terhadap Shopee. Bobot tertinggi pada topik ini adalah kata "mantap", "shopee", dan "percaya". Kata-kata ini merepresentasikan tingkat kepercayaan kepada Shopee. Kata-kata seperti "sukses", "nice", "gampang", dan "cocok" juga muncul pada topik ini, menunjukkan bahwa pengguna memandang Shopee sebagai platform *e-commerce* yang dapat diandalkan.
- Topik keempat (3) dapat diinterpretasikan sebagai topik yang berkaitan dengan **pengalaman pengguna** dalam berbelanja di Shopee. Bobot tertinggi pada topik ini didominasi oleh kata-kata seperti "bagus", "mudah", dan "belanja", yang menunjukkan kesan pengguna terhadap berbelanja di Shopee. Selain itu, kata-kata seperti "aman", "layan", dan "online" memberikan tambahan informasi bahwa topik ini berkaitan dengan pengalaman pengguna dalam berbelanja secara online di Shopee.
- Topik kelima (4) berkaitan dengan **layanan dan bantuan** yang diberikan oleh Shopee. Kata-kata yang paling sering muncul dalam topik ini adalah "bantu", "good", dan "lumayan". Bobot yang tinggi dari kata-kata tersebut menunjukkan bahwa pengguna memberikan penilaian terhadap layanan dan bantuan yang diberikan oleh Shopee. Disamping itu, terdapat kata-kata

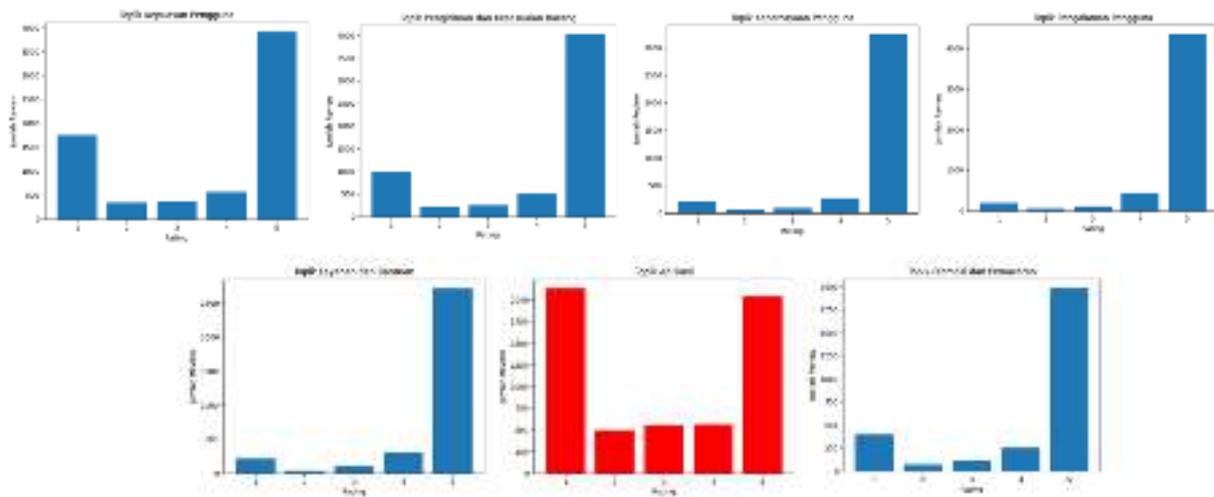
- seperti "manfaat", "terimakasih", dan "butuh" yang juga muncul dalam topik ini, menunjukkan bahwa pengguna merasa terbantu oleh layanan dan bantuan yang diberikan oleh Shopee.
- Topik keenam (5) berkaitan dengan **aplikasi Shopee**. Bobot tertinggi pada topik ini adalah kata "lambat" dan "aplikasi". Hal ini menunjukkan bahwa pengguna mengalami beberapa masalah dalam menggunakan aplikasi Shopee. Topik ini juga diperkuat dengan kata-kata “berat”, “iklan”, “tolong” yang mengindikasikan permasalahan saat menjalankan aplikasi.
  - Topik ketujuh (6) berkaitan dengan **promosi dan penawaran**. Berdasarkan bobot kata-kata yang ditemukan pada topik 6, dapat disimpulkan bahwa pengguna memberikan ulasan tentang promosi dan penawaran yang diberikan oleh Shopee. Kata-kata seperti "oke", "sale", "flash", "hadiah", dan "promo" menunjukkan bahwa pengguna memberikan feedback terkait promosi dan penawaran yang disediakan oleh Shopee.

**Tabel 4.** Kesimpulan Topik

Model	Kesimpulan Topik
Topic 0 : 0.037*"puas" + 0.037*"shopee" + 0.031*"suka" + 0.028*"muas" + 0.020*"kasih" + 0.020*"belanja" + 0.019*"terima" + 0.010*"aplikasi" + 0.010*"pakai" + 0.010*"barang"	Kepuasan Pengguna
Topic 1 : 0.040*"kirim" + 0.034*"keren" + 0.034*"barang" + 0.023*"cepat" + 0.023*"ongkos" + 0.023*"gratis" + 0.020*"sesuai" + 0.019*"shopee" + 0.019*"murah" + 0.018*"pokok"	Pengiriman dan Kesesuaian Barang
Topic 2 : 0.267*"mantap" + 0.035*"percaya" + 0.020*"shopee" + 0.020*"sukses" + 0.014*"lengkap" + 0.014*"nice" + 0.013*"gampang" + 0.011*"cocok" + 0.009*"hemat" + 0.009*"praktis"	Kepercayaan Pengguna
Topic 3 : 0.288*"bagus" + 0.063*"mudah" + 0.049*"belanja" + 0.047*"aplikasi" + 0.027*"senang" + 0.024*"top" + 0.023*"online" + 0.018*"aman" + 0.017*"shopee" + 0.017*"layan"	Pengalaman Pengguna
Topic 4 : 0.145*"bantu" + 0.102*"good" + 0.029*"lumayan" + 0.019*"manfaat" + 0.016*"terimakasih" + 0.015*"aplikasi" + 0.014*"butuh" + 0.013*"baguss" + 0.012*"job" + 0.012*"mantab"	Layanan dan Bantuan
Topic 5 : 0.032*"lambat" + 0.026*"aplikasi" + 0.025*"berfikir" + 0.020*"berat" + 0.017*"buka" + 0.012*"shopee" + 0.012*"iklan" + 0.011*"bintang" + 0.011*"tolong" + 0.009*"lancar"	Aplikasi
Topic 6 : 0.246*"oke" + 0.026*"lot" + 0.015*"sale" + 0.013*"flash" + 0.009*"shopee" + 0.009*"hadiah" + 0.008*"promo" + 0.007*"error" + 0.005*"out" + 0.005*"later"	Promosi dan Penawaran

#### 4.4 Implementasi

Dalam tahap implementasi, model yang telah dibuat dan topik-topik yang terbentuk kemudian diterapkan pada keseluruhan data teks ulasan yang ada. Penerapan model pada data memberikan gambaran yang lebih komprehensif tentang penilaian pengguna terhadap setiap topik. Dari penerapan model pada data, dapat diketahui topik-topik mana saja yang perlu dijadikan perhatian khusus dilihat dari sebaran banyaknya penilaian yang kurang baik atau rating yang rendah yang diberikan pengguna.



**Gambar 8.** Sebaran Rating di Setiap Topik

Hasil dari implementasi model pada keseluruhan data memberikan pandangan yang lebih jelas dan mendalam mengenai topik-topik yang perlu dijadikan perhatian khusus. Dari Gambar 8 dapat disimpulkan bahwa topik aplikasi memiliki intensitas rating 1 paling tinggi, atau dalam pengertian lain pengguna kebanyakan memberikan rating 1 karena terkait dengan isu aplikasi. Kata "lambat", "aplikasi", "berat", "buka", merupakan kata-kata dengan bobot tinggi pada topik aplikasi. Hal ini menunjukkan bahwa aplikasi Shopee perlu diperbaiki dan ditingkatkan dalam hal kinerja sistem dan fungsionalitasnya. Selain itu, perlu diperhatikan bahwa topik kepuasan pengguna dan topik pengiriman dan kesesuaian barang memiliki persentase rating 1 yang juga cukup tinggi, Shopee perlu melakukan evaluasi dan perbaikan dalam meningkatkan tingkat kepuasan pengguna serta dalam hal pengiriman dan kesesuaian produk yang diterima pembeli.

#### IV. KESIMPULAN

Berdasarkan hasil dari proses analisis data dan pemodelan terhadap 33.896 ulasan pengguna aplikasi Shopee di PlayStore yang tercatat mulai dari tanggal 1 April 2023 sampai 21 April 2023 diperoleh kesimpulan sebagai berikut :

1. Distribusi rating yang pada aplikasi Shopee memiliki puncak tertinggi pada rating 5. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang sangat baik terhadap Shopee. Kata "shopee", "bagus", "aplikasi", "belanja", "kirin", merupakan 5 kata teratas dengan kemunculan paling sering dalam data. Disamping itu, dalam wordcloud yang terbentuk juga terlihat hal-hal detail yang sebelumnya tidak tampak, seperti kata "tolong", "kecewa", "lambat" yang mengarah sesuatu yang negatif.
2. Pemodelan LDA dilakukan dengan mempertimbangkan parameter *passes* dan *iterations*. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100 serta *random state* 142. Hasil pengujian menunjukkan bahwa nilai rata-rata coherence score tertinggi terdapat pada kombinasi nilai *passes* 10 dan nilai *iterations* 50 dengan nilai rata-rata *coherence score* sebesar 0.47068. Nilai K-Optimum yang paling sering muncul dari pengujian tersebut adalah 7. Oleh karena itu, nilai *passes* sebesar 10 dan nilai *iterations* sebesar 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik. Adapun topik-topik yang terbantu antara lain topik kepuasan pengguna, topik pengiriman dan kesesuaian



- barang, topik kepercayaan pengguna, topik pengalaman pengguna, topik layanan dan bantuan, topik aplikasi, dan topik terkait promosi dan penawaran.
3. Hasil dari implementasi model pada keseluruhan data, terlihat bahwa topik "aplikasi" menjadi topik yang paling kritis dan mendapatkan rating 1 terbanyak. Lebih lanjut, kata-kata dengan bobot tinggi pada topik aplikasi seperti "lambat", "berat", dan "buka" menunjukkan bahwa masalah terkait kinerja sistem dan fungsionalitas aplikasi menjadi salah satu isu yang perlu diperbaiki dan ditingkatkan.

## UCAPAN TERIMA KASIH

## REFERENSI

1. R. Ameliah, R. Adi Hegara, I. Rahmawati, dan Dkk, “Status Literasi Digital di Indonesia Ringkasan Eksekutif,” *Indeks Literasi Digit. Indones.*, hal. 9, 2021, [Daring]. Tersedia pada: <https://katadata.co.id/StatusLiterasiDigital>.
2. M. Habibi, A. Priadana, A. B. Saputra, dan P. W. Cahyo, “Topic Modelling of Germas Related Content on Instagram Using Latent Dirichlet Allocation (LDA),” vol. 34, no. Ahms 2020, hal. 260–264, 2021, doi: 10.2991/ahsr.k.210127.060.
3. A. R. Destarani, I. Slamet, dan S. Subanti, “Trend Topic Analysis using Latent Dirichlet Allocation (LDA) (Study Case: Denpasar People’s Complaints Online Website),” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 1, hal. 50–58, 2019, doi: 10.26555/jiteki.v5i1.13088.
4. I. Sutherland, Y. Sim, S. K. Lee, J. Byun, dan K. Kiatkawsin, “Topic modeling of online accommodation reviews via latent dirichlet allocation,” *Sustain.*, vol. 12, no. 5, hal. 1–15, 2020, doi: 10.3390/su12051821.
5. A. S. Neogi, K. A. Garg, R. K. Mishra, dan Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, hal. 100019, 2021, doi: 10.1016/j.jjime.2021.100019.
6. J. C. Campbell, A. Hindle, dan E. Stroulia, “Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data,” *Art Sci. Anal. Softw. Data*, vol. 3, hal. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.
7. D. Endriani, “Analisis topic modelling mengenai pemberlakuan pembatasan kegiatan masyarakat menggunakan latent direchlet allocation (LDA),” *Univ. Islam Indonesia.*, hal. 1–79, 2022.



# Klasifikasi Abjad SIBI (Sistem Bahasa Isyarat Indonesia) menggunakan *Mediapipe* dengan metode *Deep Learning*

Muhammad Alfian Pratama<sup>1</sup>, Muhammad Reza Erfit<sup>2</sup>, Nadiya Mujahidatul Farhani<sup>3</sup>,  
Ignatius Arvantya Hartono<sup>4</sup>, Maryamah Maryamah<sup>5</sup>

<sup>1,2,3,4,5</sup> *Teknologi Sains Data, Universitas Airlangga*

<sup>1</sup> [muhammad.alfian.pratama-2020@ftmm.unair.ac.id](mailto:muhammad.alfian.pratama-2020@ftmm.unair.ac.id)

<sup>2</sup> [muhammad.reza.erfit-2020@ftmm.unair.ac.id](mailto:muhammad.reza.erfit-2020@ftmm.unair.ac.id)

<sup>3</sup> [nadiya.mujahidatul.farhani-2020@ftmm.unair.ac.id](mailto:nadiya.mujahidatul.farhani-2020@ftmm.unair.ac.id)

<sup>4</sup> [ignatius.arvantya.hartono-2020@ftmm.unair.ac.id](mailto:ignatius.arvantya.hartono-2020@ftmm.unair.ac.id)

<sup>5</sup> [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

Corresponding author email: [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

**Abstract:** General public knowledge in Indonesia regarding the Indonesian Sign Language System (SIBI) is quite low. This can prevent deaf and mute people from doing activities in public facilities. In this paper, we propose an alphabetical classification of SIBI sign language using the *mediapipe* and the *Deep Learning* method to help deaf and mute people communicate with the public. The methodology of this paper collects a dataset in the form of image data from the hand patterns of each SIBI sign language alphabet by combining a webcam with the help of the *open-cv* library to retrieve image data. With the *mediapipe*, the data is extracted from the coordinates of the landmarks in his hand and then normalized for each coordinate. Then the model is trained with a fully connected layer deep learning algorithm. The experimental results obtained an accuracy of 94.32% and a loss of 15.17% on the training data, an accuracy of 93.52% and a loss of 18.91% on the validation data, and an accuracy of 93.94% on the test data. These results show that the *mediapipe* and the deep learning fully connected layer algorithm successfully detect the right and left hands, and coordinates of landmarks according to the SIBI sign language alphabet correctly.

**Keywords:** Image Classification, SIBI, *Mediapipe*, *Deep Learning*, Fully Connected Layer

**Abstrak:** Pengetahuan masyarakat umum di Indonesia terhadap Sistem Bahasa Isyarat Indonesia (SIBI) cukup rendah. Hal ini dapat menghambat para penyandang tunarungu dan tunawicara untuk beraktivitas di fasilitas publik. Oleh karena itu, penelitian ini mengusulkan Klasifikasi abjad bahasa isyarat SIBI dengan menggunakan *mediapipe* dan metode *Deep Learning* untuk membantu para penyandang tunarungu dan tunawicara berkomunikasi dengan masyarakat umum. Metodologi yang dilakukan adalah pengumpulan dataset berupa data citra dari bentuk pola tangan setiap abjad bahasa isyarat SIBI dengan menggabungkan *webcam* dengan bantuan *library open-cv* untuk mengambil data citra. Dengan *mediapipe*, data tersebut diekstrak koordinat landmark di tangannya dan kemudian dinormalkan setiap koordinatnya. Selanjutnya model dilatih dengan algoritma *deep learning fully connected layer*. Hasil eksperimen mendapatkan akurasi sebesar 94,32% dan *loss* 15,17% pada data training, akurasi 93,52 % dan *loss* 18,91 % pada data validasi, serta akurasi sebesar 93,94% pada data tes. Hasil tersebut menunjukkan bahwa *mediapipe* dan algoritma *deep learning fully connected layer* berhasil mendeteksi tangan kanan dan kiri, koordinat *landmark* sesuai abjad bahasa isyarat SIBI dengan tepat.

**Kata kunci:** Klasifikasi Gambar, SIBI, *Mediapipe*, *Deep Learning*, Fully Connected Layer

## I. PENDAHULUAN

Bahasa isyarat merupakan salah satu cara berkomunikasi utama bagi tunarungu kepada sesama tunarungu atau masyarakat sekitarnya [1]. Setiap negara memiliki bahasa isyarat masing-masing yang berbeda, bahkan banyak daerah yang juga memiliki bahasa ataupun dialek isyarat yang berbeda. Menurut World Federation of the Deaf, terdapat lebih dari 300 bahasa isyarat di seluruh dunia dan 70 juta tunarungu menggunakannya. SIBI merupakan kata yang memiliki makna yang sama dan sinonim diisyaratkan dengan tempat arah dan frekuensi yang sama tetapi dengan penampil yang berbeda.



Beberapa kata yang memiliki makna yang berlawanan (yang tergolong antonim) yang diisyaratkan dengan penampil dan tempat yang sama tetapi arah gerakannya berbeda [2]. SIBI yang telah disepakati menjadi salah satu sarana yang membantu komunikasi antara individu tunarungu dan tunawicara dengan masyarakat secara luas [3]. Jumlah penduduk Indonesia yang mengalami disabilitas mencapai 211.889 orang, di antaranya 6,5% (13.802 orang) adalah tunarungu, dan 2,6% (5.580 orang) adalah tunawicara [3]. Meskipun minat masyarakat Indonesia dalam mempelajari bahasa isyarat tinggi, kurangnya sumber pembelajaran Bahasa Isyarat Indonesia menjadi salah satu hambatan dalam pemahaman masyarakat terhadap bahasa isyarat. Selain itu, kurikulum pendidikan di Indonesia belum sepenuhnya mampu memenuhi kebutuhan individu tunarungu dan tunawicara [3]. Oleh karena itu, terdapat keterbatasan dalam jumlah masyarakat umum yang memiliki kemampuan dalam memahami dan menggunakan bahasa isyarat [4].

*Mediapipe* adalah suatu kerangka kerja yang dikembangkan oleh Google untuk membangun rangkaian proses dalam memproses data persepsi, baik itu dalam format audio maupun video. Diluncurkan pada tahun 2019, kerangka kerja ini menyediakan berbagai solusi machine learning, seperti deteksi wajah, pengenalan gerakan tangan, segmentasi rambut, pemahaman holistik, dan lain sebagainya. Informasi lebih lanjut mengenai solusi-solusi tersebut dapat ditemukan di situs web resmi *Mediapipe* [5]. *Mediapipe* dirancang bagi mereka yang ingin mengimplementasikan kecerdasan buatan kedalam aplikasi yang akan dibangun. *Mediapipe* juga memungkinkan pembangunan aplikasi *crossplatform* yang bisa berjalan di berbagai perangkat keras yang berbeda [6]. Model yang dibangun oleh pengembang Google menggunakan Tensorflow lite memfasilitasi aliran informasi yang mudah diadaptasi dan dimodifikasi melalui grafik [7]. Kerangka kerja perangkat lunak *Mediapipe* dipilih sebagai alat untuk mengembangkan sistem klasifikasi abjad bahasa isyarat ini karena *Mediapipe* memiliki fitur-fitur yang dapat digunakan untuk melakukan pengolahan data multimedia seperti video. Selain itu, *Mediapipe* juga memiliki modularitas yang tinggi, sehingga dapat dengan mudah dikembangkan untuk melakukan klasifikasi abjad bahasa isyarat Indonesia.

Beberapa penelitian pengenalan bahasa isyarat Indonesia telah dilakukan sebelumnya dengan metode dan hasil yang beragam menggunakan model Convolutional Neural Network (CNN) yang merupakan model Inflated 3D dikombinasikan dengan metode transfer learning pada Sistem Bahasa Isyarat Indonesia (SIBI) mendapatkan akurasi 97.5% [8]. Pada tahun 2021, penelitian dilakukan mengenai pengenalan bahasa isyarat. Sebagai contoh, Ilham Rizaldy Widy Putra mengembangkan sistem deteksi simbol bahasa isyarat seperti "Saya," "Kamu," "Dia," "Cinta," "Maaf," dan "Sedih" dengan menggunakan metode Convolutional Neural Network (CNN) [9]. Pada penelitian lainnya, Mohammad Farid Naufal dan timnya menggunakan metode CNN untuk mengklasifikasikan objek peragaan dalam permainan "gunting," "batu," dan "kertas," dan berhasil mencapai akurasi klasifikasi sebesar 97,66% [10]. Darmatasia juga melakukan penelitian pada tahun yang sama dengan fokus pada klasifikasi bahasa isyarat menggunakan Sistem Bahasa Isyarat Indonesia (SIBI) dan metode Gradient-CNN, dan mencapai akurasi sebesar 98% [11]. Namun, penelitian yang telah disebutkan sebelumnya memiliki batasan karena membutuhkan perangkat khusus seperti *Leap Motion Controller* atau kamera kedalaman (*depth camera*) untuk mendapatkan informasi tiga dimensi dari gestur bahasa isyarat. Dengan adanya *Mediapipe*, sebuah kerangka kerja *open source* yang dikembangkan oleh Google, rekonstruksi kerangka tangan tiga dimensi dapat dilakukan menggunakan data dua dimensi yang diperoleh dari kamera RGB biasa yang dapat secara otomatis mengklasifikasikan abjad bahasa isyarat Indonesia berdasarkan video yang diinputkan ke dalam sistem secara langsung.

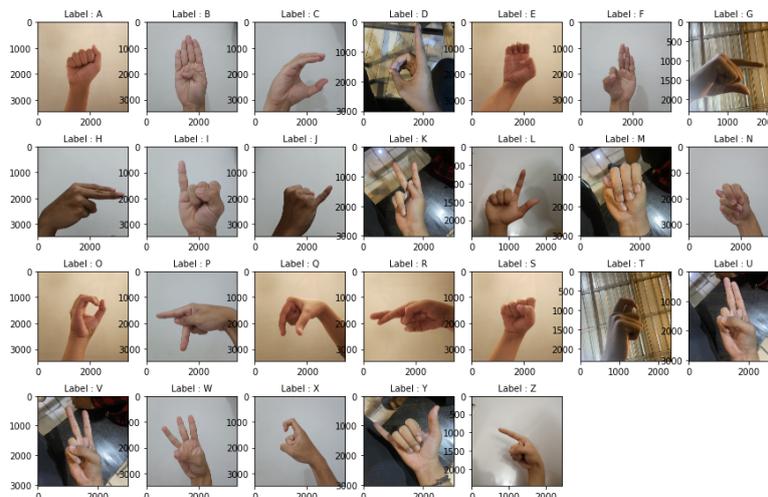
Oleh karena itu, penelitian ini mengusulkan Klasifikasi abjad bahasa isyarat SIBI dengan menggunakan *mediapipe* dan metode *Deep Learning*. Metodologi yang dilakukan adalah pengumpulan dataset berupa data citra dari bentuk pola tangan setiap abjad bahasa isyarat SIBI dengan menggabungkan webcam dengan bantuan library *open-cv* untuk mengambil data citra. Dengan *mediapipe*, data tersebut diekstrak koordinat landmark di tangannya dan kemudian dinormalkan setiap

koordinatnya. Selanjutnya model dilatih dengan algoritma *deep learning fully connected layer*. Peneliti ini dapat mempermudah para penyandang tunarungu dan tunawicara berkomunikasi dengan masyarakat umum.

## II. METODE PENELITIAN

### 3.1 Sumber Data

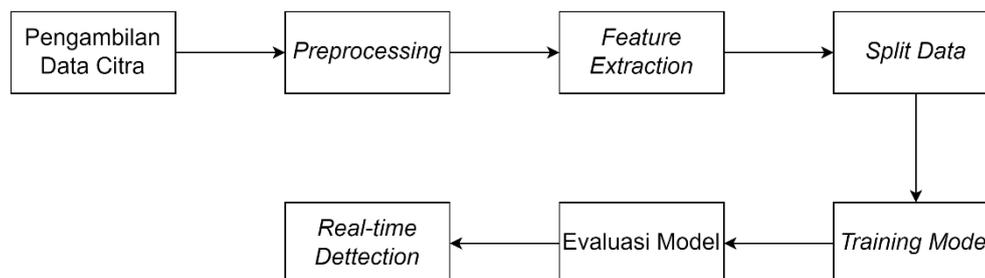
Data yang digunakan merupakan data citra abjad bahasa isyarat SIBI yang diperagakan oleh peneliti. Proses pengambilan data citra dilakukan menggunakan *smartphone* masing-masing dengan menggunakan rasio 1:1 dan melakukan pengambilan data citra sebanyak 5-6 data untuk setiap abjad bahasa isyarat SIBI, sehingga data awal sebanyak 650 data citra untuk semua abjad bahasa isyarat SIBI yang akan digunakan untuk dilakukan analisis klasifikasi. Gambar 1 merupakan tampilan dataset yang berhasil dikumpulkan.



Gambar 1. Tampilan dataset

### 3.2 Metodologi

Alur penelitian yang dilakukan yaitu terdiri dari beberapa alur yang diawali dengan pencarian data hingga melakukan analisis. Berikut merupakan alur dari penelitian yang dilakukan antara lain pengambilan data citra, *data preprocessing*, *feature extraction*, *split data*, *training model*, evaluasi, *real time detection*. Alur penelitian atau metodologi diawali dengan pengambilan data citra sesuai dengan yang dijelaskan pada bagian sumber data, lalu dilakukan data *preprocessing*. Pada proses data *preprocessing* dilakukan penyamaan ukuran *pixel* dan augmentasi data. Setelah dilakukan data *preprocessing*, selanjutnya adalah melakukan ekstraksi fitur menggunakan *mediapipe* khususnya menggunakan *hand solution* untuk ekstraksi koordinat *landmark* pada tangan lalu dilakukan normalisasi untuk setiap koordinatnya. Selanjutnya, adalah melakukan *splitting data* untuk *training* yang digunakan untuk melakukan pelatihan model dan *testing* untuk mengevaluasi model. Lalu, melakukan pelatihan model menggunakan algoritma *deep learning fully connected layer*. Setelah model berhasil dilatih, model digunakan untuk melakukan evaluasi model. Langkah yang terakhir adalah membuat aplikasi *real time detection* untuk melakukan klasifikasi abjad bahasa isyarat Indonesia SIBI secara *real time*.

**Gambar 2.** Metodologi penelitian

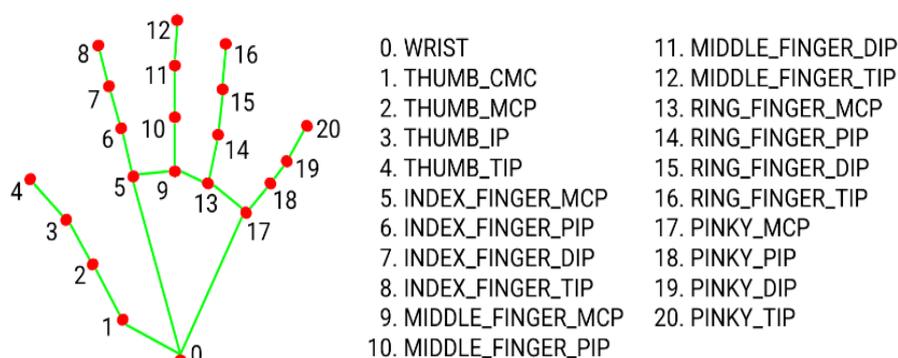
### 3.2.1 Data Preprocessing

Langkah data *preprocessing* yang pertama kali dilakukan adalah menyamakan ukuran *pixel* dari setiap data citra, hal ini dilakukan karena pada saat pengambilan data kamera yang digunakan berbeda-beda sehingga ukuran *pixel* untuk setiap data citra berbeda. Citra dirubah menjadi ukuran 250x250 *pixel*. Tampilan data citra setelah dilakukan penyamaan ukuran *pixel* akan terlihat data citra tidak blur dan masih jelas.

Tahap selanjutnya dari data *preprocessing* adalah melakukan augmentasi data, dikarenakan data yang dikumpulkan kurang untuk dilakukan pemodelan. Pada augmentasi data menggunakan *package ImageDataGenerator* dari *library* Tensorflow. Augmentasi dilakukan dengan memutar citra maksimal 30° khusus untuk huruf G, I, J, dan Q maksimal 15° dikarenakan memiliki pose yang relatif mirip, memperbesar citra maksimal 0.1, membalikan gambar secara horizontal, dan tingkat kecerahan antara 0.2 hingga 0.8. Masing-masing data citra akan menghasilkan 20 data citra baru hasil augmentasi. Sehingga, jumlah data akhir yang diperoleh adalah sebanyak 13.000 data.

### 3.2.2 Feature Extraction

Fitur ekstraksi yang digunakan adalah *mediapipe* menggunakan *hands solution*. Dengan menggunakan fitur ekstraksi ini dapat mengetahui koordinat dari masing-masing *landmark* pada tangan seperti pada Gambar 3.

**Gambar 3.** *Mediapipe hand landmark*

Hasil dari fitur ekstraksi ini akan menghasilkan 21 titik koordinat untuk setiap *landmark* tangan atau 42 titik koordinat untuk dua tangan. Setelah mengetahui masing-masing koordinat dari *landmark* tangan, perlu melakukan normalisasi nilai untuk setiap koordinat. Karena apabila tidak dilakukan normalisasi posisi dari tangan terhadap kamera akan sangat mempengaruhi koordinatnya dan nilai dari masing-masing koordinat akan sangat beragam sehingga akan sulit untuk dilakukan klasifikasi. Proses normalisasi dilakukan dengan cara mengurangi seluruh koordinat *landmark*



dengan nilai pada koordinat *landmark wrist*, hal ini dilakukan untuk menjadikan *landmark wrist* sebagai titik pusat. Selanjutnya, adalah membagi seluruh koordinat *landmark* dengan koordinat *landmark* terjauh atau nilai koordinat terbesar agar rentang nilai seluruh koordinat berada di rentang -1 hingga 1.

Dibalik kemudahan dari fitur ekstraksi menggunakan *mediapipe*, terdapat kelemahan yang ditemukan yaitu *mediapipe* tidak dapat melakukan ekstraksi fitur untuk semua data citra. Ada beberapa faktor yang menyebabkan hal ini terjadi diantaranya adalah pencahayaan dari citra yang kurang dan tidak semua jari terlihat pada gambar. Sehingga, jumlah data citra yang dapat dilakukan fitur ekstraksi menggunakan *mediapipe* ini ada sejumlah 8439 data.

### 3.2.3 Balancing Data

Dikarenakan *mediapipe* tidak dapat melakukan ekstraksi fitur untuk semua data citra, maka data yang didapatkan mengalami *unbalance*. Oleh karena itu, dilakukan *balancing data* menggunakan metode *oversampling* dengan algoritma *SMOTE*. Setelah dilakukan *oversampling* didapatkan jumlah data akhir sebanyak 13078 data dengan jumlah data untuk masing-masing kelas sebanyak 503 data.

### 3.2.4 Split Data

Sebelum melakukan pembuatan model, data harus dipisahkan untuk data *training* dan data *test* dengan jumlah data *training* sebanyak 80% jumlah data keseluruhan dan data *test* sebanyak 20% jumlah data keseluruhan. Proses pemisahan data dilakukan lagi untuk data *training* untuk mendapatkan data validasi yang digunakan pada proses pelatihan model *neural network*, dengan jumlah data *training* sebanyak 80% dari jumlah data *training* sebelumnya dan data validasi sebanyak 20% jumlah data *training*.

### 3.2.5 Model Deep Learning (*fully-connected layer*)

Model *Deep Learning* yang digunakan untuk klasifikasi adalah model *deep learning fully connected layer*. *Activation map* yang dihasilkan dari *feature extraction layer* masih berbentuk *multidimensional array*, sehingga mau tidak mau kita harus melakukan *reshape activation map* menjadi sebuah vektor agar bisa digunakan sebagai *input* dari *fully-connected layer*. Model yang digunakan sangat sederhana dengan 21\*2 *input neuron* yang akan menjadi nilai *vector* untuk mengenali *landmark* yang sudah dijelaskan sebelumnya. Kemudian, metode *dropout* agar menonaktifkan sejumlah *edge* yang terhubung ke semua *neuron* agar tidak terjadi *overfitting* dan juga mempercepat proses *learning*. *Neuron* ini yang akan dihilangkan sementara dipilih secara acak. Pada penelitian ini, peneliti memberikan probabilitas *dropout* sebesar 0,1 dan 0,2.

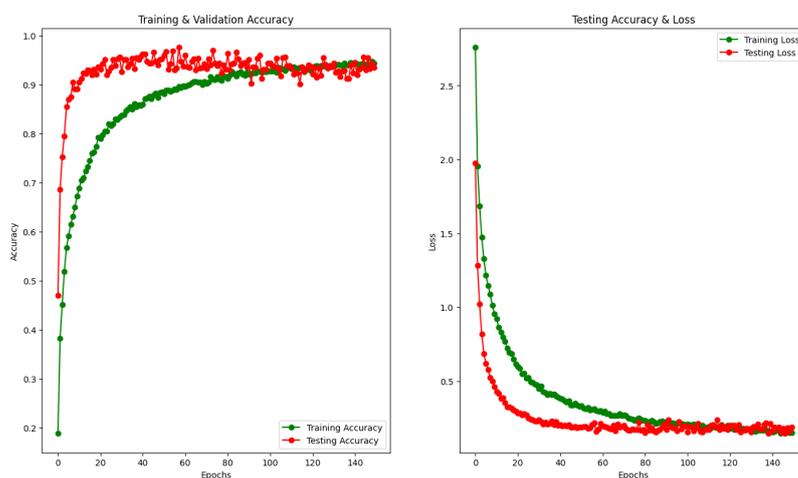
Kemudian adalah proses *dense* untuk menambahkan *layer* yang *fully connected* dengan jumlah unit 512 dan 256 yang berarti jumlah node yang harus ada pada hidden layer menggunakan aktivasi relu. Aktivasi relu menjadi pilihan bagi kita karena sifatnya yang lebih berfungsi dengan baik. Selanjutnya masuk ke tahap klasifikasi dimana, menggunakan aktivasi *softmax* agar yang diklasifikasi pada input bisa sesuai dengan kategori yang telah ditentukan, yakni berjumlah 26 sesuai dengan jumlah huruf yang diklasifikasikan. Setelah arsitektur model *deep learning* telah selesai dirancang, maka didapatkan model yang terbaik yang memiliki jumlah parameter sebanyak 160.026. Untuk menyusun model agar siap untuk dilakukan proses *training*, Penelitian ini, menggunakan *adam optimizer*, *sparse categorical cross entropy* sebagai nilai *loss*, dan nilai akurasi sebagai nilai pengukurannya, untuk *fitting* modelnya ditentukan *epoch* 150 dan *batch size* 128. Metrik evaluasi yang digunakan pada penelitian ini adalah nilai akurasi.

### III. HASIL DAN PEMBAHASAN

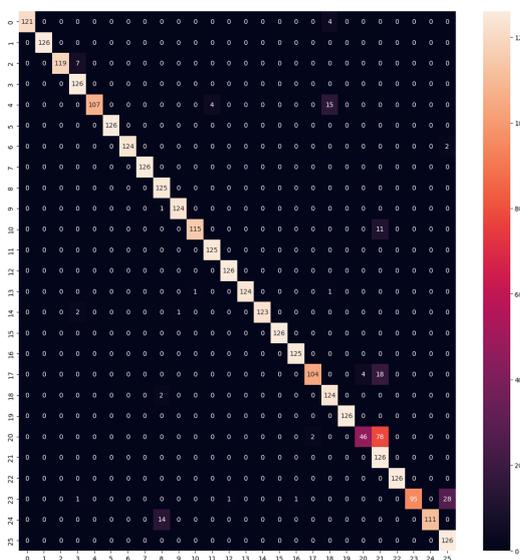
Pada penelitian ini, dilakukan evaluasi model dengan pengukuran performa akurasi dan percobaan *real-time detection*. Nilai akurasi dihitung berdasarkan jumlah data yang benar diprediksi dibandingkan dengan seluruh data dan nilai loss dihitung menggunakan fungsi loss *sparse categorical cross entropy*. Untuk *real-time detection* penelitian ini menggunakan bantuan *library open-cv* sebagai input gambar dari webcam dan *library mediapipe* untuk ekstraksi fitur dari input webcam.

#### 4.1 Evaluasi Model

Setelah pelatihan model selesai dilakukan didapatkan akurasi sebesar 94,32% dan *loss* sebesar 15,17% pada data *training*, lalu pada data validasi akurasi sebesar 93,52% dan *loss* 18,91%. Sedangkan, pada data *test* didapatkan akurasi sebesar 93,94%. Gambar 4. merupakan visualisasikan secara interaktif nilai akurasi dan *loss* pada data *training* dan data *testing*. Dapat dilihat bahwa model cukup stabil pada masing-masing data dan tidak terjadi *overfitting*.



Gambar 4. Plot evaluasi



Gambar 5: Confusion matrix

Dari *confusion matrix* pada Gambar 5. dapat dilihat bahwa semua gambar sudah diklasifikasikan dengan benar kecuali pada beberapa gambar dengan index 17, 20, dan 23, yakni huruf R dan X serta U dan V karena memiliki bentuk pola tangan yang hampir mirip. Dengan akurasi 94% menunjukkan bahwa semua sudah diklasifikasikan dengan baik dan sangat akurat, Namun, ada beberapa huruf yang tidak diklasifikasikan dengan sempurna, yakni huruf dengan *index* 17, 20, dan 21 karena data memiliki bentuk pola tangan yang hampir sama. Jumlah data *test* pada masing-masing huruf dapat dilihat pada variabel support dengan jumlah total pada data test adalah 3270.

#### 4.2 Real-Time Detection

*Real time detection* dilakukan dengan mengambil data citra menggunakan *webcam* dengan bantuan *library open-cv*. Lalu, untuk mendeteksi tangan beserta koordinat *landmark* menggunakan *mediapipe*. Selanjutnya, untuk model untuk klasifikasi bahasa isyarat menggunakan model yang sudah dilakukan *training* sebelumnya. Dengan menggabungkan ketiga aspek tersebut *real time detection* dapat dibangun untuk melakukan klasifikasi abjad bahasa isyarat Indonesia SIBI dengan tampilan seperti pada Gambar 10.



**Gambar 10.** Implementasi *real time detection*

Setelah dilakukan pengujian untuk semua abjad, saat ini sistem masih cukup kesulitan untuk memprediksi abjad dengan bentuk tangan yang hampir mirip seperti huruf X dan T, dan huruf R, U, dan V. Namun, untuk deteksi tangan kanan dan kiri beserta koordinat *landmark mediapipe* bisa melakukannya dengan sangat akurat, cepat, dan ringan.

## IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa klasifikasi abjad Bahasa Isyarat Indonesia menggunakan *Mediapipe* merupakan solusi yang efektif untuk mengenali dan mengklasifikasikan abjad Bahasa Isyarat Indonesia secara otomatis. Dalam pengembangannya, dilakukan proses pra-pemrosesan data dengan pemerataan ukuran piksel dan augmentasi data. Kemudian, dilakukan ekstraksi fitur dengan menggunakan *Mediapipe*, khususnya *hand solution*, untuk mengekstrak koordinat *landmark* pada tangan dan menormalkan setiap koordinat. Selanjutnya, dilakukan pemisahan data untuk pelatihan dan pengujian, dan dilatih menggunakan algoritma *deep learning fully connected layer*.

Dalam tahap evaluasi model didapatkan akurasi sebesar 94,32% dan loss sebesar 15,17% pada data training, lalu pada data validasi akurasi sebesar 93,52% dan loss 18,91 %. Sedangkan, pada data test didapatkan akurasi sebesar 93,94%. Saat divisualisasikan model juga cukup stabil pada masing-masing data dan tidak terjadi *overfitting*. Melalui *confusion matrix*-nya diketahui bahwa semua gambar dapat diklasifikasikan dengan baik kecuali pada huruf R dan X serta U dan V karena memiliki bentuk pola tangan yang hampir mirip. Dari *confusion matrix* tersebut juga diketahui keakuratan model



melalui nilai accuracy, precision, recall, dan f1-score. Dari 4 nilai tersebut diketahui bahwa semua sudah diklasifikasikan dengan baik dan sangat akurat kecuali huruf R, X, dan U karena bentuk pola tangan yang hampir sama. Jumlah data test pada masing-masing huruf dapat dilihat pada variabel support dengan jumlah total pada data test adalah 3270.

Secara keseluruhan, penggunaan Mediapipe dalam klasifikasi abjad Bahasa Isyarat Indonesia menunjukkan potensi yang besar untuk mengembangkan sistem yang lebih canggih dalam mengenali dan mengklasifikasikan gerakan tangan pada Bahasa Isyarat Indonesia. Penelitian ini memiliki implikasi penting untuk pengembangan teknologi dalam bidang Bahasa Isyarat Indonesia, terutama dalam membantu penyandang tuna rungu dan mempermudah komunikasi antara mereka dengan orang lain. Saran penelitian selanjutnya adalah saat pengambilan data lebih baik menggunakan fungsi *capture* pada aplikasi *real time detection* agar data yang digunakan dapat terdeteksi semua fiturnya oleh *mediapipe* dibandingkan melakukan pengambilan data secara manual menggunakan kamera pada *smartphone*.

## REFERENSI

- [1] Imam Suyudi, Sudadio, Suherman. Pengenalan Bahasa Isyarat Indonesia menggunakan Mediapipe dengan Model Random Forest dan Multinomial Logistic Regression. Jurnal Ilmu Siber dan Teknologi Digital (JISTED). Vol 1, No 1, 2022, halaman 65-80. DOI: 10.35912/jisted.v1i1.1899
- [2] Hakim, Lukman, Samino, et al. Kamus Sistem Isyarat Bahasa Indonesia. Edisi Kelima. Jakarta: DIREKTORAT PEMBINAAN SEKOLAH LUAR BIASA, 2008.
- [3] Sindarto SS, Ratnawati DE, Arwani I. Klasifikasi Citra Sistem Isyarat Bahasa Indonesia (SIBI) dengan Metode Convolutional Neural Network pada Perangkat Lunak berbasis Android. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 2022;6(5):2129-2138. Available from: <http://j-ptiik.ub.ac.id>
- [4] Suyudi I, Sudadio S, Suherman S. Pengenalan Bahasa Isyarat Indonesia menggunakan Mediapipe dengan Model Random Forest dan Multinomial Logistic Regression. Jurnal Ilmu Siber dan Teknologi Digital (JISTED). 2022;1(1):65-80. <https://doi.org/10.35912/jisted.v1i1.1899>
- [5] Anam N. Sistem Deteksi Simbol pada SIBI (Sistem Isyarat Bahasa Indonesia) menggunakan Mediapipe dan RESTNet-50 [Tugas Akhir]. Universitas Dinamika; 2022.
- [6] Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang C-L, Grundmann M. MediaPipe hands: On-device real-time hand tracking. arXiv.org. 2020 Jun 18. Tersedia pada: <https://arxiv.org/abs/2006.10214>
- [7] Haldera A, Tayade A. Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. Int J Res Publ Rev. 2021;2(5):9-17. ISSN 2582-7421.
- [8] Suhartjito, Thiracitta, N., Gunawan, H. (2021). Pengenalan bahasa isyarat SIBI menggunakan convolutional neural network yang dikombinasikan dengan transfer learning dan parameter non-trainable. Procedia Computer Science, 179: 72-80.
- [9] Putra IRW. Sistem Deteksi Simbol pada SIBI (Sistem Isyarat Bahasa Indonesia) Menggunakan Convolutional Neural Network. Universitas Dinamika, 2021.
- [10] Naufal MF, et al. Klasifikasi Citra Game Batu Kertas Gunting Menggunakan Convolutional Neural Network. Techno.COM, Vol. 20, No. 1, Februari 2021, hlm. 166-174.
- [11] Darmatasia. Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Gradient-Convolutional Neural Network. Jurnal Instek, Volume 6, Nomor 1, April 2021, Teknik Informatika, Universitas Islam Negeri Alauddin Makassar.



# Perspektif Wisatawan Mancanegara (Wisman) Terhadap Pariwisata Indonesia menggunakan *Latent Dirichlet Allocation (LDA)*

Rafik Septiana<sup>1</sup>, Muhammad Hanif<sup>2</sup>, Christeigen Theodore Suhelim<sup>3</sup>, Faqih Syamil<sup>4</sup>,  
Arya Duta Gordra Sumitro Putra<sup>5</sup>, Maryamah Maryamah<sup>6</sup>

<sup>1, 2, 3, 4, 5, 6</sup> *Teknologi Sains Data, Universitas Airlangga*

<sup>1</sup> [rafik.septiana-2021@ftmm.unair.ac.id](mailto:rafik.septiana-2021@ftmm.unair.ac.id)

<sup>2</sup> [muhammad.hanif-2021@ftmm.unair.ac.id](mailto:muhammad.hanif-2021@ftmm.unair.ac.id)

<sup>3</sup> [christeigen.theodore.suhelim-2021@ftmm.unair.ac.id](mailto:christeigen.theodore.suhelim-2021@ftmm.unair.ac.id)

<sup>4</sup> [faqih.syamil-2021@ftmm.unair.ac.id](mailto:faqih.syamil-2021@ftmm.unair.ac.id)

<sup>5</sup> [arya.duta.gor-2021@ftmm.unair.ac.id](mailto:arya.duta.gor-2021@ftmm.unair.ac.id)

<sup>6</sup> [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

Corresponding author email: [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

**A Abstract:** *The 2022 G20 Bali summit in Indonesia provides an opportunity to revitalize the tourism industry after the Covid-19 pandemic. The perspectives of foreign tourists visiting Indonesia, including the advantages and disadvantages of their tourism experience are an important aspect to improve tourism in Indonesia. In this paper, we propose the perspective of foreign tourists on Indonesian tourism using Latent Dirichlet Allocation (LDA). The first step was scrapping data comments by foreign tourists regarding tourism in Indonesia from Twitter. The next step of this paper is data preprocessing consisting of cleaning data, tokenization, stopwords, and stemming. The LDA method is used to analyze and identify the main topics from the comments until obtaining relevant insights. The results of the research reveal three main topics: tourist satisfaction with Indonesian tourism, foreign tourist activities during their tour, and plans for foreign tourists to visit Indonesia. These findings provide valuable insights for the government in optimizing the tourism sector in the future and supporting post-pandemic economic recovery. In the context of recovery, understanding the strengths and weaknesses identified by foreign tourists can help the government take appropriate steps to increase Indonesia's attractiveness as a tourist destination.*

**Keywords:** *Perspective of International Tourists, Indonesia Tourism, Topic Modelling, Latent Dirichlet Allocation (LDA)*

**Abstrak:** Konferensi Tingkat Tinggi (KTT) G20 2022 di Bali-Indonesia memberikan peluang untuk merevitalisasi industri pariwisata pasca pandemi Covid-19. Penelitian ini menganalisis perspektif wisatawan mancanegara (wisman) yang telah mengunjungi Indonesia, termasuk kelebihan dan kekurangan pengalaman pariwisata mereka. Pada paper ini, kami mengusulkan perspektif wisatawan mancanegara (wisman) terhadap pariwisata Indonesia menggunakan Latent Dirichlet Allocation (LDA). Langkah-langkah yang dilakukan meliputi *scrapping* data dari *platform* Twitter yang berisi komentar wisman terkait pariwisata di Indonesia. Selanjutnya menggunakan Metode LDA untuk menganalisis mengidentifikasi topik utama yang muncul dari komentar. Langkah-langkah pengolahan data dan analisis dilakukan untuk memperoleh wawasan yang relevan. Hasil penelitian mengungkap tiga topik utama: kepuasan wisman terhadap pariwisata Indonesia, aktivitas wisman selama berwisata, dan rencana kunjungan wisman ke Indonesia. Temuan ini memberikan wawasan berharga bagi pemerintah dalam mengoptimalkan sektor pariwisata di masa depan dan mendukung pemulihan ekonomi pasca pandemi. Dalam konteks pemulihan, pemahaman kelebihan dan kekurangan yang diidentifikasi oleh wisman dapat membantu pemerintah mengambil langkah-langkah yang tepat untuk meningkatkan daya tarik Indonesia sebagai tujuan wisata yang diinginkan oleh wisman.

**Kata kunci:** *Perspektif Wisatawan Mancanegara, Pariwisata Indonesia, Pemodelan Topik, Latent Dirichlet Allocation (LDA)*



## I. PENDAHULUAN

Pariwisata merupakan sektor yang sangat potensial untuk memberikan kontribusi bagi kemakmuran ekonomi suatu negara [1]. Pariwisata merupakan salah satu bidang kritis di Indonesia yang mendapat perhatian besar dari pemerintah, karena berpotensi memberikan kontribusi yang signifikan terhadap perekonomian nasional. Pemerintah Indonesia telah melakukan beberapa upaya untuk meningkatkan sektor pariwisata, antara lain memasarkan tempat wisata, menyediakan infrastruktur yang memadai, dan meningkatkan pelayanan pariwisata. Namun, memahami perspektif dan persepsi wisatawan mancanegara (wisman) terhadap lokasi pariwisata Indonesia sangat penting dalam mendongkrak pariwisata. Pengunjung asing memainkan peran penting dalam perluasan industri pariwisata karena mereka sangat berkontribusi terhadap pengeluaran pariwisata, penciptaan lapangan kerja, dan peningkatan pendapatan masyarakat lokal [2]. Oleh karena itu, pemahaman menyeluruh tentang persepsi wisatawan internasional terhadap pariwisata Indonesia dapat menjadi landasan penting untuk mengembangkan tujuan wisata. Untuk mengetahui persepsi pengunjung asing terhadap pariwisata Indonesia, penelitian ini memanfaatkan metode *Latent Dirichlet Allocation* (LDA). LDA adalah metode untuk mengidentifikasi topik dan tema dalam teks yang telah banyak digunakan dalam analisis dan penelitian teks. LDA akan digunakan dalam penelitian ini untuk mengevaluasi evaluasi pengunjung internasional terhadap pariwisata Indonesia untuk menentukan topik utama yang muncul dalam ulasan mereka.

Isu-isu yang dibahas dalam penelitian ini meliputi bagaimana pengunjung internasional memandang pariwisata Indonesia, elemen apa yang membuat mereka khawatir, dan apakah ada perbedaan sikap di antara kelompok turis asing yang berbeda dari negara yang berbeda. Studi ini akan menganalisis penelitian sebelumnya tentang masalah yang sama untuk menjawab pertanyaan-pertanyaan ini. Studi sebelumnya telah dilakukan dalam upaya untuk memahami perspektif wisatawan mancanegara terhadap pariwisata Indonesia, namun penelitian yang menggunakan pendekatan LDA dalam menilai evaluasi wisatawan masih terbatas. Beberapa penelitian sebelumnya telah menemukan bahwa unsur-unsur seperti kualitas layanan, keindahan alam, keragaman budaya, dan aksesibilitas berdampak pada pendapat wisatawan internasional tentang pariwisata. Namun, belum ada penelitian yang mengidentifikasi subjek utama yang muncul dalam ulasan pengunjung internasional tentang pariwisata Indonesia dengan menggunakan pendekatan LDA.

Beberapa studi tambahan menemukan fitur budaya dan keragaman menjadi faktor penarik wisatawan internasional memilih Indonesia sebagai tujuan wisata. Daya tarik terbesar wisatawan internasional adalah kekayaan budaya Indonesia, seperti kehidupan adat, kesenian tradisional, dan upacara keagamaan [3]. Hal ini menunjukkan pentingnya faktor budaya dalam menarik perhatian dan menghasilkan pengalaman unik bagi wisatawan internasional. Meskipun studi ini memberikan wawasan awal yang bermanfaat, mengidentifikasi masalah utama yang muncul dalam penilaian pengunjung internasional terhadap pariwisata Indonesia membutuhkan pendekatan yang lebih sistematis dan menyeluruh. Maka, metode *Latent Dirichlet Allocation* (LDA) digunakan dalam pekerjaan ini sebagai alat analisis yang dapat memberikan wawasan yang lebih dalam<sup>[4]</sup>. Pendekatan LDA akan membantu akademisi untuk secara otomatis menemukan tren topik dalam evaluasi wisatawan mancanegara. Penelitian ini akan menggunakan pendekatan ini untuk menentukan isu-isu utama yang sering muncul dalam review pengunjung asing, seperti “keindahan alam”, “budaya dan tradisi”, “kuliner”, “aktivitas rekreasi”, dan sebagainya. Teknik ini dimaksudkan untuk memberikan pengetahuan yang lebih menyeluruh tentang apa yang dianggap vital dan menarik oleh pengunjung internasional dalam pariwisata Indonesia.

Penelitian ini memiliki implikasi penting bagi pengembangan destinasi pariwisata di Indonesia karena memberikan pengetahuan yang lebih baik tentang perspektif pengunjung internasional terhadap



pariwisata Indonesia. Temuan penelitian ini dapat membantu pemerintah, organisasi pariwisata, dan pelaku sektor pariwisata meningkatkan pengalaman pengunjung internasional selama kunjungan mereka ke Indonesia. Mengetahui hal apa yang paling diapresiasi dan disukai wisatawan mancanegara memungkinkan upaya pengembangan dan promosi pariwisata lebih tepat sasaran dan efektif. Temuan penelitian ini diproyeksikan dapat memberikan kontribusi yang signifikan terhadap pertumbuhan pariwisata Indonesia. Dengan pemahaman yang lebih baik tentang selera dan pendapat wisatawan internasional, pemerintah dan sektor pariwisata dapat mengidentifikasi area yang memungkinkan untuk perbaikan, seperti pembangunan infrastruktur, layanan pariwisata, promosi destinasi, dan pelestarian budaya lokal. Selain itu, penelitian ini dapat membantu mengembangkan rencana pemasaran yang lebih efisien untuk menarik wisatawan mancanegara dan meningkatkan daya saing Indonesia sebagai tujuan wisata internasional. Penelitian ini juga dapat berfungsi sebagai sumber bagi akademisi lain yang tertarik dengan pariwisata dan analisis evaluasi pengunjung. Temuan penelitian ini dapat digunakan untuk memandu penelitian di masa depan tentang subjek yang muncul dalam evaluasi wisatawan mancanegara dari berbagai tempat wisata di seluruh dunia. Hal ini dapat membantu memperluas pemahaman kami tentang selera dan perspektif turis internasional, serta memberikan wawasan yang berguna tentang inisiatif pengembangan pariwisata lain.

Untuk mencapai tujuan penelitian, peneliti mengusulkan perspektif wisatawan mancanegara (wisman) terhadap pariwisata Indonesia menggunakan LDA. Hal pertama yang dilakukan adalah mengumpulkan data ulasan pengunjung internasional dari media sosial Twitter yang mengandung kata kunci. Penelitian ini berfokus pada beberapa pusat lokasi pariwisata Indonesia, yaitu Bali, Pulau Komodo, Labuan Bajo, dan Lombok. Rentang waktu unggahan data tweet yaitu pada bulan Mei tahun 2022. Data akan dianalisis menggunakan pendekatan LDA untuk menentukan subjek utama yang disebutkan dalam evaluasi wisatawan internasional. Selanjutnya, temuan tersebut akan diperiksa secara kualitatif dan statistik untuk mengkarakterisasi dan memahami sudut pandang pengunjung internasional terhadap pariwisata Indonesia. Dengan adanya penelitian ini, para pemangku kepentingan pariwisata, masyarakat lokal, dan wisatawan mancanegara dapat menumbuhkan sinergi dalam pembangunan pariwisata Indonesia yang berkelanjutan, budaya, dan menguntungkan bagi semua pihak yang berkepentingan.

## II. METODE PENELITIAN

Tahapan metodologi yang dilakukan pada penelitian ini terdiri dari beberapa proses yaitu pengumpulan data, preprocessing data, pemodelan topik, data processing dan analisis dari topik yang dihasilkan. Metodologi yang digunakan pada penelitian ini mengikuti proses *text mining* atau proses menambang data berupa teks yang diperoleh melalui dokumen untuk menemukan kata-kata yang berhubungan dengan isi dokumen sehingga dilakukan analisis terhadap hubungan dari sumber teks dapat dilakukan [4]. Penambangan teks ini menganalisis sejumlah data teks tidak terstruktur sehingga dapat mengidentifikasi pola-pola tertentu dan kata kunci dalam data [5].

Penelitian ini menggunakan data yang diperoleh melalui pengambilan data melalui situs media sosial Twitter. Pemrosesan data teks yang diambil dari data Twitter digunakan untuk mengambil kata-kata yang terkait dengan topik pariwisata Indonesia, yang kemudian dilakukan analisis pola-pola tertentu untuk memperoleh informasi yang bermakna dari data tersebut. Data yang diambil berupa *tweet* dari para pengguna aplikasi Twitter. Bahasa yang diambil untuk penelitian ini adalah bahasa Inggris berdasarkan topik-topik terkait yang ditentukan oleh keyword yang telah ditetapkan. keyword yang digunakan adalah beberapa tempat pariwisata yaitu Bali, Pulau Komodo, Labuan Bajo, dan Lombok. Data diambil berdasarkan dalam rentang bulan Mei 2022 dikarenakan pada bulan tersebut adalah angka kunjungan tertinggi pada tahun 2022 kuartal pertama yaitu sebesar 212332 kunjungan.

Data yang digunakan pada penelitian ini diambil menggunakan teknik *web scraping* menggunakan bahasa pemrograman Python dan *library* Selenium. *Library* Selenium akan digunakan untuk



melakukan otomatisasi pada halaman web, seperti melakukan *scrolling* dan melakukan pencarian topik. Setelah itu, halaman web akan diekstrak ke dalam file berekstensi CSV dengan memanfaatkan struktur halaman web, seperti tag HTML dan elemen CSS di dalam elemen unggahan. *Web scraping* atau ekstraksi web merupakan suatu teknik untuk mengekstrak data dari internet dan menyimpan hasilnya ke dalam *file* seperti database, CSV, atau *file* berekstensi lainnya [6]. *Web scraping* adalah proses mendapatkan dokumen tidak terstruktur dari website dalam bentuk *markup language* (HTML). Hasil ekstraksi data akan dianalisis untuk mengambil data tertentu dari halaman tersebut [7]. *Web scraping* adalah teknik mendapatkan informasi secara otomatis dari sebuah website tanpa menyalinnya secara manual [8]. Program akan menganalisis dokumen HTML dari internet dan mendapatkan data berdasarkan tag HTML untuk menggapit informasi yang ingin diambil (membuat *scraping template*). Setelah itu, informasi tersebut akan disimpan ke dalam tabel database, CSV, atau format *file* lainnya. Pengambilan data ini akan menggunakan bahasa pemrograman Python, yakni dengan mengambil data melalui tag HTML dan menyimpannya ke dalam *file* CSV untuk diproses ke tahap selanjutnya.

Setelah mendapatkan data dengan menggunakan *web scraping*, proses selanjutnya adalah *data cleaning*, yaitu proses mengidentifikasi dan menghilangkan kesalahan dalam data<sup>[12]</sup>. Hal ini untuk memastikan bahwa data yang diambil memiliki kualitas yang baik. Pembersihan data dapat digunakan untuk membersihkan data *string* untuk menghilangkan karakter yang tidak relevan, salah eja, dan kesalahan serupa lainnya dalam teks. *Data cleaning* dalam pengertian umum adalah proses menyelidiki data dari ketidakakuratan dan membuat data yang dapat dikelola untuk analisis. Dalam penelitian ini, peneliti akan melakukan pembersihan data pada tipe data *string*, yang mana akan dilakukan pembersihan terhadap kata-kata yang tidak relevan dengan topik, menghapus kesalahan ejaan, menormalkan data, atau teknik pembersihan lainnya. Hal ini penting untuk menjamin data yang telah diperoleh dari Twitter dapat dianalisis dan diekstraksi dengan lebih mudah dan terstruktur. Beberapa teknik yang dilakukan dalam *data cleaning* adalah *case folding* (mengubah huruf kapital menjadi kecil), *stemming* (mengubah suatu kata menjadi kata dasar), dan *tokenization* (memecah kalimat menjadi kumpulan kata).

Data yang telah dikumpulkan akan dilakukan proses pembersihan (*cleaning*) melalui *library* Pandas dan Pyspark. Pyspark merupakan *library* Python yang menggunakan bantuan *tools* berupa Apache Spark. *Library* tersebut digunakan untuk meningkatkan performa dalam melakukan pembersihan data. Hal ini disebabkan oleh besarnya data yang akan digunakan dan proses eksekusi akan memakan waktu yang sangat lama apabila hanya menggunakan bahasa Python saja. Proses pembersihan data ini meliputi penghilangan data yang kosong (*missing data*), penghapusan tanda baca di kalimat, *case folding* (pengubahan huruf kapital menjadi huruf kecil), penghapusan *stopwords* pada kalimat (misalnya of, or, dll), *stemming* (pengubahan suatu kata menjadi kata dasar), dan proses *tokenization* (pemecahan kalimat menjadi kata per kata).

Setelah dilakukan pembersihan data, tahap selanjutnya adalah pencarian topik-topik tersembunyi pada kumpulan tweet dengan menggunakan Latent Dirichlet Allocation (LDA). Penelitian ini menggunakan metode Latent Dirichlet Allocation (LDA) yang merupakan model dalam *machine learning* yang menggunakan probabilitas korpus yang diwakili oleh distribusi data untuk setiap kata dalam dokumen [9]. Latent Dirichlet Allocation (LDA) adalah suatu metode untuk mengolah data dalam jumlah besar dengan asumsi bahwa satu dokumen terdiri dari berbagai topik yang merupakan distribusi kosakata [10]. Latent Dirichlet Allocation (LDA) adalah suatu metode dalam *unsupervised learning* yang digunakan untuk mengelompokkan data menjadi beberapa topik, merangkum, dan mengolah data yang besar [11]. Pada penelitian ini, akan dilakukan pengelompokan data Twitter yang telah diperoleh menjadi beberapa topik berdasarkan sebaran kata-kata tertentu dalam *tweet* pengguna media sosial Twitter sehingga ditemukan informasi terkait perspektif wisatawan mancanegara (wisatawan mancanegara) dari topik tersebut. Latent Dirichlet Allocation dapat dikalkulasi menggunakan persamaan (1):

$$P(z = j|z_t, w_t, d_t, \cdot) \propto \frac{(C_{w_t j}^{WT} + \beta)}{\sum_{w=1}^W (C_{w_t j}^{WT} + W\beta)} \times \frac{(C_{d_t j}^{DT} + \alpha)}{\sum_{d=1}^T (C_{d_t j}^{DT} + T\alpha)} \quad (1)$$

$C_{w_t j}^{WT}$  = Frekuensi sebuah kata muncul dalam topik 1 dan topik 2

$\beta$  = distribusi kata berdasarkan topik (parameter konsentrasi)

$W$  = Panjang kosakata (jumlah token/kata unik dalam dokumen lengkap)

$C_{d_t j}^{DT}$  = Frekuensi dokumen muncul sebagai topik 1 dan topik 2, ketika terasi dimulai

$\alpha$  = Distribusi topik setiap dokumen

$T$  = Jumlah topik

Regex (*Regular Expression*) merupakan gabungan dari dua jenis karakter, yaitu karakter literal dan meta. Literal adalah karakter yang mewakili dirinya sendiri seperti keseluruhan alfabet, huruf kapital, huruf kecil, atau karakter lainnya [13]. Regex juga merupakan salah satu implementasi operasi pencocokan pola untuk tipe data teks atau *string* [14]. Regex adalah kunci untuk melakukan pemrosesan teks secara kuat, fleksibel, dan efisien dengan menggunakan pola notasi umum seperti bahasa pemrograman [15]. Regex memudahkan tahap pra pemrosesan data, khususnya dalam pencarian teks. Hal ini menjadi dasar penggunaan regex untuk tokenisasi sebuah teks karena informasi terkait letak angka, tanda baca, dan karakter di luar *string* dapat diketahui. Setelah letak karakter diketahui, karakter tersebut akan dihapus. Regex juga dapat memisahkan kalimat menjadi kelompok kata dengan memisahkannya dengan spasi. Hal tersebut dapat memudahkan tokenisasi dokumen saat tahap pra pemrosesan data.

Pada analisis ini, dibutuhkan dua *hyperparameter*, yaitu  $\alpha$  dan  $\beta$ . Untuk mengendalikan distribusi topik pada data, dibutuhkan hyperparameter  $\alpha$ , yang mana semakin kecil nilai  $\alpha$  maka data *tweet* akan cenderung memiliki satu topik, dan semakin besar nilai  $\alpha$  maka semakin seragam topik-topik pada data tersebut. Di sisi lain, nilai  $\beta$  berperan dalam mengendalikan distribusi kata-kata pada suatu topik. Nilai  $\beta$  yang semakin kecil menunjukkan bahwa topik cenderung memiliki sedikit kata, sedangkan nilai yang semakin besar menunjukkan bahwa topik-topik memiliki variasi kata yang banyak. Luaran dari LDA adalah probabilitas kata-kata pada tiap topik. Suatu topik pada LDA ( $k$ ) memiliki semua kosakata dari kumpulan dokumen yang diolah. Masing-masing kata pada topik  $k$  tersebut akan memiliki nilai dengan rentang 0 hingga 1. Nilai ini menunjukkan tingkat probabilitas suatu kata untuk muncul pada topik tertentu. Selain probabilitas dari masing-masing kata, LDA juga menampilkan probabilitas dari masing-masing topik pada suatu dokumen. Dengan nilai probabilitas ini, hasil dari penelitian dapat direpresentasikan dalam suatu dokumen atau topik berdasarkan kata-kata yang memiliki probabilitas paling tinggi. Dengan demikian, seluruh informasi dari data unggahan para pengguna Twitter dapat diekstrak untuk mendapatkan topik-topik tersembunyi, yang mana hasil akan direpresentasikan berdasarkan kata-kata yang memiliki probabilitas paling tinggi. Pada penelitian ini, algoritma LDA juga akan dibandingkan dengan algoritma PLSA yang memiliki kemampuan yang sama dalam mengekstrak topik sehingga kita dapat membandingkan keduanya untuk mendapatkan hasil yang terbaik.

### III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan *Latent Dirichlet Allocation* (LDA) untuk menemukan topik-topik tersembunyi (laten) beserta distribusi kata di dalam topik tersebut. Sebagai pembanding, algoritma *Latent Semantic Analysis* (LSA) dan *Probabilistic Latent Semantic Analysis* (pLSA) juga digunakan dalam analisis ini. LDA, LSA, dan pLSA merupakan beberapa model *machine learning* yang umum digunakan pada kasus analisis topik [16,17]. Pada algoritma LSA, matriks dokumen yang berisi istilah-istilah dipecah ke dalam tiga matriks menggunakan proses *singular value decomposition* (SVD) [17]. Algoritma LSA membangun ruang semantik dari korpus (kumpulan dokumen) yang berupa ruang vektor, yang mana ruang semantik ini kemudian dapat digunakan untuk menghitung kesamaan antar kata, kalimat, paragraf, ataupun seluruh dokumen [17,18]. Akan tetapi, algoritma LSA tidak memiliki

model probabilistik untuk kemunculan kata atau istilah sehingga pengelompokan topik lebih sulit untuk diinterpretasikan [16]. Selain itu, algoritma LSA memiliki kekurangan dari sisi interpretabilitas *embeddings* yang dapat menimbulkan subjektivitas pada hasil [19]. Untuk mengatasi limitasi pada algoritma LSA, digunakan algoritma pemodelan topik yang memanfaatkan model probabilistik, yaitu pLSA dan LDA. Kedua model tersebut memodelkan topik laten dan memanfaatkan konsep probabilitas untuk memberikan bobot atau tingkat kepentingan dari setiap kata yang terkelompok pada suatu topik.

Analisis pada penelitian ini difokuskan pada model LDA. Sebagai pembanding, model LSA dan pLSA juga digunakan. Hal ini disebabkan oleh keunggulan model LDA yang dapat menggeneralisasikan dokumen (kumpulan kata) baru dengan lebih baik [16]. Model pLSA merupakan pengembangan dari model LSA yang menyelesaikan masalah interpretabilitas dan subjektivitas [17]. Akan tetapi, model pLSA termasuk *incomplete* karena tidak menyediakan model probabilistik pada tingkat dokumen. Setiap dokumen direpresentasikan sebagai *list* angka (proporsi campuran topik), dan tidak ada model probabilistik generatif untuk angka-angka tersebut. Hal ini menimbulkan jumlah parameter yang dibutuhkan dalam model semakin besar seiring dengan semakin besarnya korpus, yang mana dapat menyebabkan *overfitting* pada model. Selain itu, proses menetapkan probabilitas ke dokumen di luar *training set* juga belum jelas [16]. Singkatnya, model pLSA bukanlah model generatif yang tepat apabila terdapat dokumen baru. Oleh karenanya, model LDA digunakan sebagai solusi dari kelemahan model pLSA. Model LDA memiliki kemampuan pertukaran (*exchangeability*) dari kata dan dokumen menggunakan distribusi Dirichlet, yang memungkinkan proses generatif yang koheren untuk data uji [16].

Tabel 1 merujuk kepada kata-kata penting pada setiap topik yang diperoleh menggunakan metode LSA, sedangkan Tabel 2 merujuk kepada kata-kata penting pada setiap topik yang diurutkan berdasarkan bobot probabilitas atau tingkat kepentingannya, yang diperoleh menggunakan metode pLSA dan LDA.

**Tabel 1.** Hasil Keseluruhan Skenario Topik (Model LSA)

topic_id	word
0	bali; photo; posted; indonesia; com; go; love; speedway; irene; island
1	posted; photo; speedway; com; indonesia; gods; courtesy; peace; island; cd
2	go; want; lets; wanna; let; posted; bad; photo; lombok; need

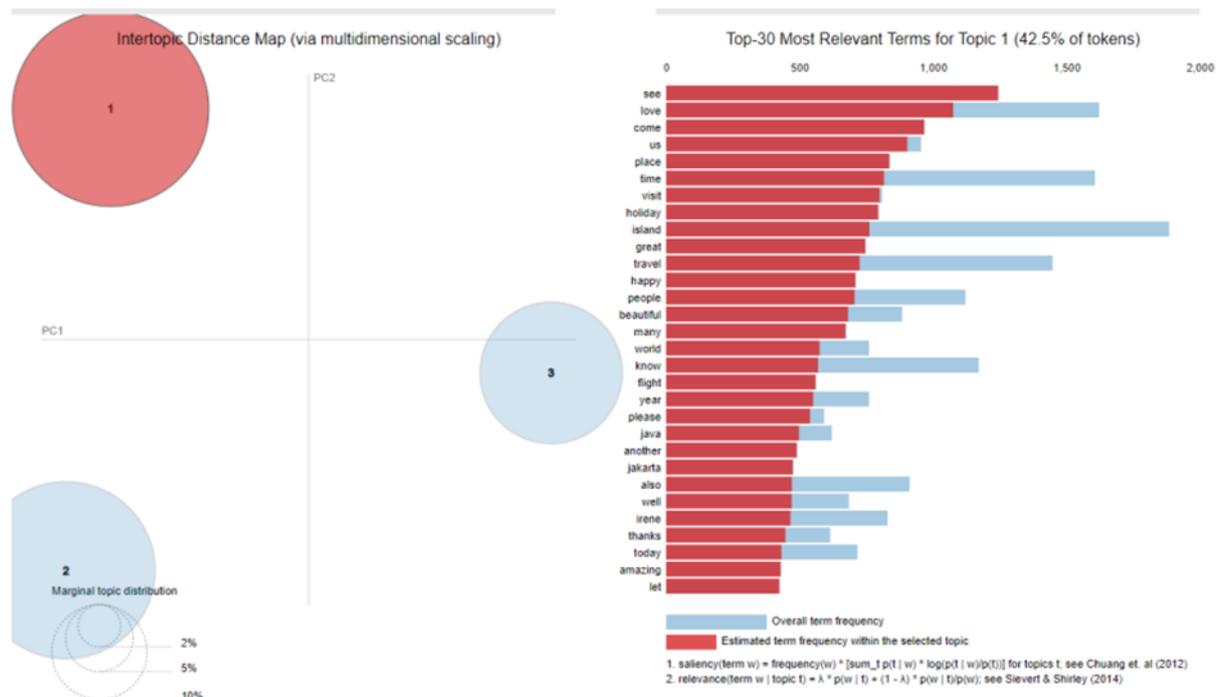
**Tabel 2.** Perbandingan Hasil Keseluruhan Skenario Topik (Model pLSA dan LDA)

pLSA				LDA		
no	word	importance	topic_id	word	importance	word_count
1	flight	0.013002	0	lombok	0.015760	809
2	bali	0.011489	0	like	0.010584	887
3	man	0.009647	0	love	0.007022	668
4	photo	0.009001	0	island	0.006647	553
5	com	0.008833	0	day	0.006442	529
6	cliff	0.008426	0	back	0.006229	530

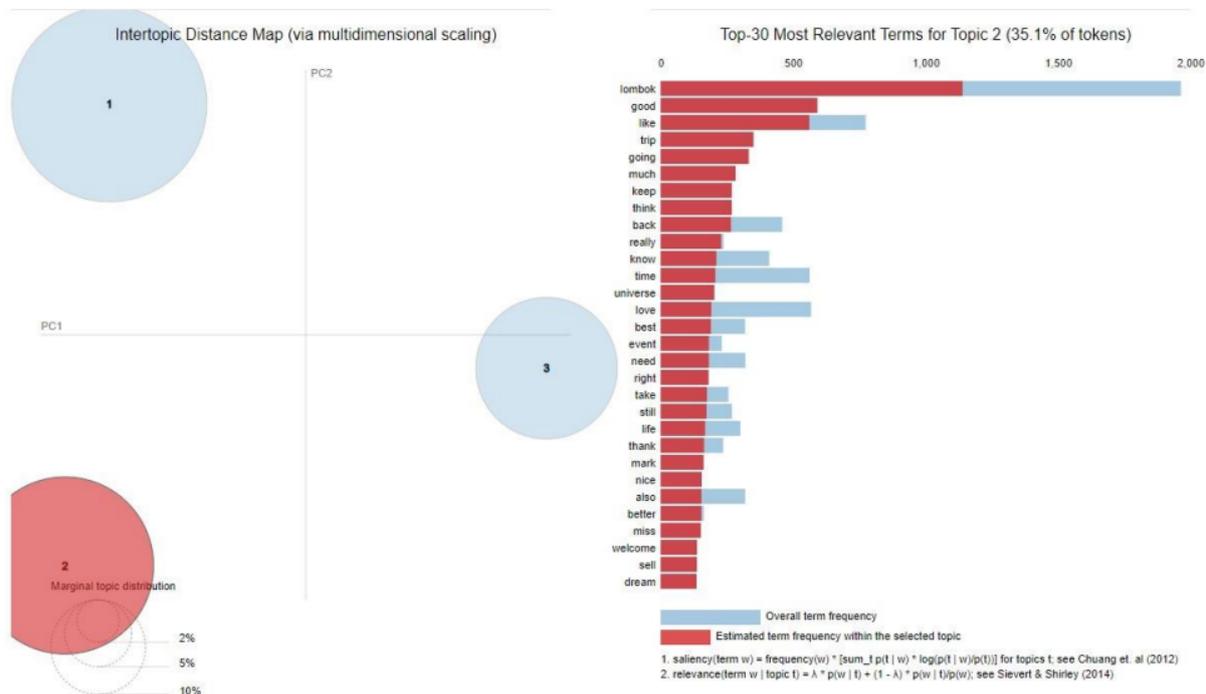


7	island	00.007442	0	time	0.006160	652
8	internet	0.007117	0	see	0.005527	469
9	plane	0.006988	0	know	0.005186	435
10	hotel	0.006988	0	trip	0.004864	404
11	bali	0.014843	1	lombok	0.021229	809
12	year	0.009379	1	photo	0.015976	533
13	com	0.008966	1	posted	0.011335	367
14	indonesia	0.008278	1	check	0.008028	313
15	trip	0.007553	1	good	0.007831	695
16	dom	0.007312	1	keep	0.007531	261
17	day	0.007296	1	universe	0.006517	123
18	friend	0.006948	1	event	0.005966	221
19	place	0.006929	1	listing	0.005816	222
20	week	0.006368	1	morning	0.005391	195
21	home	0.010192	2	lombok	0.004582	809
22	bali	0.009529	2	news	0.003762	225
23	position	0.007653	2	tourism	0.003532	157
24	family	0.7601	2	resort	0.003532	138
25	taste	0.006288	2	project	0.003227	132
26	visit	0.006254	2	flights	0.003020	147
27	kid	0.006074	2	lets	0.003009	149
28	life	0.005999	2	world	0.002936	298
29	beach	0.005961	2	video	0.002892	138

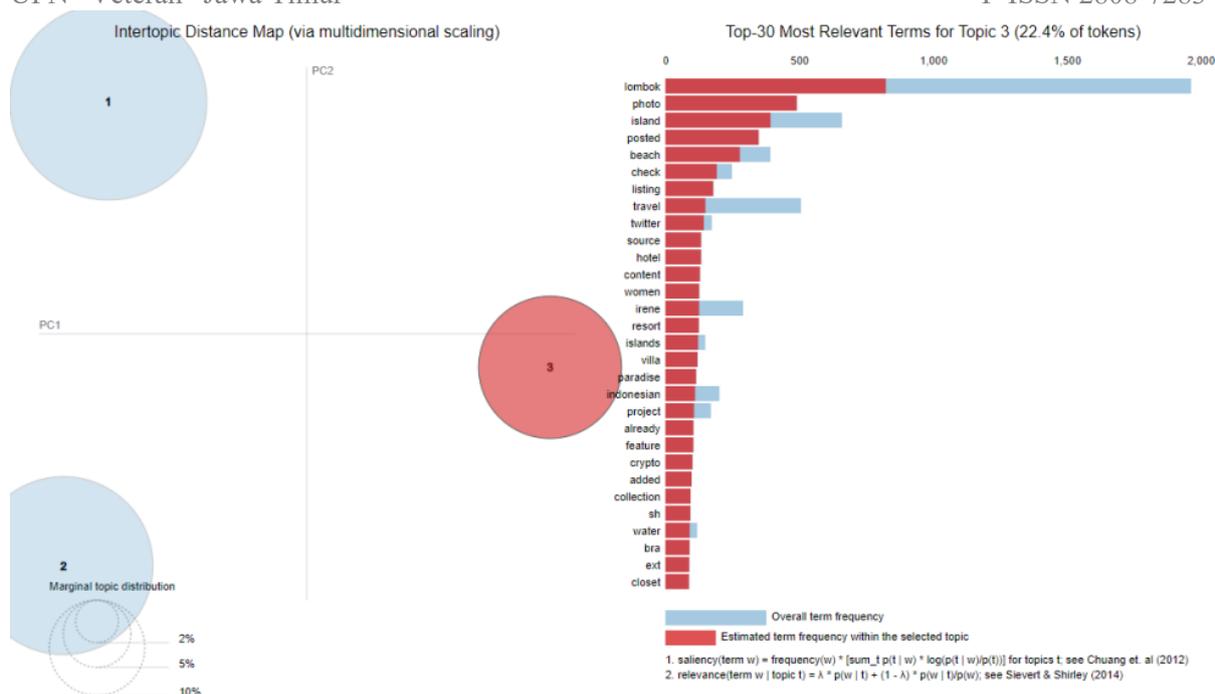
Tahap berikutnya adalah interpretasi topik yang sudah diekstraksi oleh LDA. Interpretabilitas merupakan kunci berharga dalam penentuan kualitas LDA yang diaplikasikan untuk masalah prediksi. Hal ini dikarenakan topik diskusi akan dikaitkan dengan risiko hasil. Matriks  $\beta$  dapat digunakan untuk mendapatkan pemahaman awal terkait jenis topik yang muncul dengan melihat kata-kata yang paling mewakili topik tersebut. Untuk menginterpretasikan topik tersebut dapat digunakan visualisasi berupa bar chart. Berikut merupakan visualisasi untuk penemuan kata dengan metode *Latent Dirichlet Allocation* LDA per topik.



**Gambar 1.** Topic 1 Intertopic Distance Location Map & its Top 30 Most Relevant Terms



**Gambar 2.** Topic 2 Intertopic Distance Location Map & its Top 30 Most Relevant Terms



**Gambar 3.** Topic 3 Intertopic Distance Location Map & its Top 30 Most Relevant Terms

Topik dapat dilihat sebagai konstruksi dasar yang diukur dengan kombinasi istilah yang membentuk topik melalui logika probabilitas. Misalnya, topik 0 mengandung istilah seperti “like”, “good”, “love”, “time”, “back”, “know”, “see”, dan “trip”. Kata-kata tersebut mengacu pada kepuasan dan kecintaan para wisatawan mancanegara (wisman) saat berwisata di Indonesia. Demikian pula pada topik 2 yang berisi “photo”, “posted”, “check”, “good”, “keep”, “event”, dan “listing” yang merujuk kepada kegiatan-kegiatan yang dilakukan wisman di Indonesia. Terakhir, topik 3 mengandung istilah “news”, “tourism”, “resort”, “project”, “flights”, “world”, dan “video” yang merepresentasikan rencana wisman saat mendatangi Indonesia. Pola dari hasil tersebut dapat diidentifikasi dengan jelas melalui pemeriksaan setiap topik LDA.

Probabilitas dari topik yang dimodelkan memberikan informasi kuantitatif terkait tweet yang diunggah wisman mengenai pengalaman mereka ketika menikmati pariwisata Indonesia. Framework LDA memodelkan pariwisata Indonesia sebagai gabungan berkelanjutan dari beberapa topik. Dengan demikian, LDA menyediakan konteks tambahan dalam tingkat variabel terstruktur. Oleh sebab itu, topik LDA probabilitas meningkatkan prediksi serta pemahaman tentang apa pendapat wisman terhadap pariwisata di Indonesia.

#### IV. KESIMPULAN

Metode *Latent Dirichlet Allocation (LDA)* dapat membantu dalam memahami pandangan dan perspektif wisman. Metode ini memiliki keunggulan dapat menggeneralisasikan dokumen (kumpulan kata) baru dengan lebih baik daripada model pLSA yang termasuk *incomplete* karena tidak menyediakan model probabilistik pada tingkat dokumen. Dari hasil pengujian dengan metode LDA, didapati bahwa hasil dari ekstraksi topik LDA mengenai pendapat wisatawan mancanegara (wisman) terhadap pariwisata di Indonesia menunjukkan adanya pola berupa tiga topik. Tiga topik tersebut meliputi topik terkait kepuasan dan kecintaan para wisman saat berwisata di Indonesia, kegiatan yang dilakukan saat berkunjung di Indonesia, dan rencana wisman saat mengunjungi Indonesia. Dari ketiga topik tersebut, didapati bahwa wisman yang berkunjung di Indonesia memiliki kesan yang positif



terhadap pariwisata Indonesia. Seluruh kata-kata di dalam topik menyatakan kepuasan dan kecintaan wisman terhadap pariwisata di Indonesia. Wisman mengabadikan momen kunjungannya ke Indonesia melalui foto sehingga dapat diketahui kegiatan yang dilakukan selama di Indonesia. Kata-kata yang memiliki tingkat probabilitas tinggi menunjukkan tujuan wisman ke Indonesia, yaitu untuk berwisata ataupun melakukan proyek pribadi. Dengan demikian, hasil dari pencarian topik-topik tersembunyi tersebut dapat digunakan untuk membantu pemerintah dan pihak terkait dalam potensi pariwisata di Indonesia terkhusus untuk membuat kebijakan atau membangun fasilitas yang memadai untuk meningkatkan daya tarik pariwisata Indonesia dalam rangka pemulihan ekonomi pasca pandemi Covid-19.

Penelitian ini memberikan wawasan yang bermanfaat tentang pandangan dan perspektif wisman terhadap pariwisata Indonesia dengan menggunakan metode Latent Dirichlet Allocation (LDA). Meskipun demikian, masih ada beberapa area yang dapat dieksplorasi lebih lanjut (misalnya berfokus pada suatu daerah di Indonesia). Pertama, penelitian selanjutnya dapat memperluas cakupan data yang digunakan, dengan melibatkan jumlah sampel yang lebih besar, dari wisman yang berbeda asal dan usia. Dengan mengumpulkan data dari kelompok wisman yang berbeda, hasil dari ekstraksi topik LDA dapat menjadi lebih representatif. Kedua, penelitian selanjutnya dapat memperdalam analisis topik LDA untuk mengidentifikasi lebih banyak topik tersembunyi terkait pariwisata Indonesia yang belum terungkap dalam studi ini. Terakhir, penelitian masa depan dapat mengeksplorasi hubungan antara pandangan dan perspektif wisman dengan faktor-faktor seperti preferensi wisatawan terhadap destinasi wisata, tingkat kepuasan dan pengalaman wisatawan, dan dampak pariwisata pada lingkungan dan masyarakat setempat. Diharapkan penelitian ini dapat membantu lebih memahami pandangan dan perspektif wisman terhadap pariwisata Indonesia dan memberikan masukan untuk pengembangan pariwisata Indonesia.

## REFERENSI

1. A. R. Rahma, "Potensi Sumber Daya Alam dalam Mengembangkan Sektor Pariwisata di Indonesia," *Jurnal Nasional Pariwisata*, vol. 12, April 2020.
2. O. Kaikara, "Tourism Development Strategy," *International Journal Papier*, vol. 1, no. 1, pp. 20-25, 27 Agustus 2020.
3. L. Asy'ari, R. D. Dienaputra, A. Nugraha, R. Tahir, C. U. Rakhman and R. R. Putra, "Kajian Konsep Ekowisata Berbasis Masyarakat Dalam Menunjang Pengembangan Pariwisata: Sebuah Studi Literatur," *Jurnal Ilmiah Pariwisata Agama dan Budaya*, vol. 6, 30 Desember 2021.
4. I. Sutherland, Y. Sim, S. K. Lee, J. Byun and K. Kiatkawsin, "Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation," *Sustainability*, vol. 12, 27 Februari 2020.
5. A. Udgave dan P. Kulkarni, "Text Mining and Text Analytics of Research Articles," *PalArch's Journal of Archeology of Egypt/Egyptology*, no. 17, p. 6, 2020.
6. M. L. C. Chilmi, "Latent Dirichlet Allocation (LDA) Untuk Mengetahui Topik Pembicaraan Warganet Twitter Tentang Omnibus Law," Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2021.
7. H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li and L. Zhao, "Latent Dirichlet Allocation (LDA) and Topic Modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78(11), pp. 15169-15211, 2019.
8. B. Zhao, *Encyclopedia of Big Data*, Virginia: Springer, 2017.
9. E. K. Putri dan T. Setiadi, "Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes," *Jurnal Sarjana Teknik Informatika*, vol. 2, no. 3, 2014.
10. F. Fathonah dan A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin



- Covid-19 Menggunakan Metode Naïve Bayes,” *Jurnal Sains dan Informatika*, vol. 7, no. 2, 2021.
11. D. M. Blei, A. Y. Ng dan M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3, 2003.
  12. M. Turland, *php| architect’s Guide to Web*, 1 ed., Toronto: Marco Tabini, 2010.
  13. D. D. Ayani, H. S. Pratiwi and H. Muhardi, "Implementasi Web Scraping Untuk Pengambilan Data Pada Situs Marketplace," *Jurnal Sistem dan Teknologi Informasi*, vol. 7(4), pp. 2460-3562, 2019.
  14. R. R. Deshmukh and V. Wangikar, "Data Cleaning: Current Approches and Issues," Aurangabad, 2011.
  15. S. Madya, *Metodologi Pengajaran Bahasa dari Era Prametode Sampai Era Pascametode*, Yogyakarta: UNY Press, 2013.
  16. L. Gerald Andrew, “pLSA and LDA.”.
  17. L. . H. Anaya, “Comparing Latent Dirichlet Allocation and Latent Semantic Analysis as Classifiers,” UNIVERSITY OF NORTH TEXAS, 2011.
  18. T. Cvitanic, B. Lee, H. Ik Song, Katherine Fu, and D. Rosen, “LDA v. LSA: A Comparison of Two Computational Text Analysis Tools for the Functional Categorization of Patents,” Georgia Institute of Technology.
  19. J. Xu, “Topic Modeling with LSA, PSLA, LDA & lda2Vec ,” *NanoNets*, Jul. 01, 2021.



# Speech Emotion Recognition (SER) dengan Metode Bidirectional LSTM

Nicholas Juan Calvin Pradiptamurty<sup>1</sup>, Hafiyah Khayyiroh Shafro<sup>2</sup>, Mohammad Sihabudin Al Qurtubi<sup>3</sup>, Giovanny Alberta Tambahjong<sup>4</sup>, Qothrotunnidha<sup>5</sup>, Almaulidiyah<sup>5</sup>, Maryamah Maryamah<sup>6</sup>

<sup>1, 2, 3, 4, 5</sup> *Teknologi Sains Data, Universitas Airlangga*

<sup>1</sup> [nicholas.juan.kalvin-2020@ftmm.unair.ac.id](mailto:nicholas.juan.kalvin-2020@ftmm.unair.ac.id)

<sup>2</sup> [hafiyah.khayyiroh.afro-2020@ftmm.unair.ac.id](mailto:hafiyah.khayyiroh.afro-2020@ftmm.unair.ac.id)

<sup>3</sup> [mohammad.sihabudin.al-2020@ftmm.unair.ac.id](mailto:mohammad.sihabudin.al-2020@ftmm.unair.ac.id)

<sup>4</sup> [giovanny.alberta.tambahjong-2020@ftmm.unair.ac.id](mailto:giovanny.alberta.tambahjong-2020@ftmm.unair.ac.id)

<sup>5</sup> [qothrotunnidha.almaulidiyah-2020@ftmm.unair.ac.id](mailto:qothrotunnidha.almaulidiyah-2020@ftmm.unair.ac.id)

<sup>6</sup> [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

Corresponding author email: [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

**Abstract:** Emotions are a part of humans as a form of response to experienced events. Emotion analysis or known as speech emotion recognition (SER) is a field many researchers are interested in because voice recognition systems can assist in criminal investigations, monitoring, and detection of potentially dangerous events, and assisting the health care system. Therefore, this study proposes the detection of SER using the Bidirectional Long short-term memory (Bi-LSTM) model approach. The dataset used was scraped on the YouTube platform. The dataset is manually labeled then feature extraction is performed using the Mel Frequency Cepstral Coefficients (MFCC). The experiment using the Bi-LSTM method has an AUC ROC value of 0.97 and an f1-score value of 0.878. Based on these results, it can be concluded that the performance of the proposed method succeeded in predicting SER better than other comparison methods. This model also proved to be more precise in classifying human voices based on four types of emotions, namely happy, sad, angry, and neutral.

**Keywords:** Speech Emotion Recognition, Bidirectional Long short-term memory (Bi-LSTM), Audio Classification

**Abstrak:** Emosi merupakan bagian tak terpisahkan dari setiap manusia sebagai bentuk respon terhadap suatu peristiwa yang dialami. Analisis emosi atau dikenal dengan *speech emotion recognition* (SER) menjadi bidang penelitian yang diminati oleh banyak peneliti karena sistem pengenalan suara dapat membantu dalam investigasi kriminal, pengawasan dan deteksi peristiwa yang berpotensi berbahaya, serta membantu sistem perawatan kesehatan. Oleh karena itu, penelitian ini mengusulkan deteksi SER dengan pendekatan model Bidirectional Long short-term memory (Bi-LSTM). Dataset yang digunakan diambil dengan metode *scraping* pada platform youtube. Selanjutnya dataset dilakukan manual label dan dilakukan ekstraksi fitur menggunakan *Mel Frequency Cepstral Coefficients* (MFCC). Metode Bi-LSTM menghasilkan nilai ROC AUC sebesar 0,97 dan nilai *F1-score* sebesar 0,878. Berdasarkan hasil tersebut maka dapat disimpulkan bahwa kinerja metode yang diusulkan, yakni model Bi-LSTM berhasil memprediksi SER lebih baik dibandingkan dengan metode lainnya. Model ini pun terbukti lebih tepat dalam mengelompokkan suara manusia berdasarkan empat jenis emosi yakni senang, sedih, marah, dan netral.

**Kata kunci:** Speech Emotion Recognition, Bidirectional Long Short-Term Memory (Bi-LSTM), Klasifikasi Audio

## I. PENDAHULUAN

Emosi adalah bagian yang tidak terhindarkan dari setiap pribadi seseorang pada saat berkomunikasi [1]. Bentuknya dapat diekspresikan dengan berbagai macam cara yang terkadang tidak dapat diketahui secara langsung [2]. Pada dasarnya setiap manusia mempunyai emosi sebagai bentuk implementasi atau respon yang melibatkan perilaku dan fisiologis yang digunakan untuk menangani sebuah peristiwa yang dialami. Emosi dapat dikatakan sebagai faktor yang mudah diketahui oleh sesama manusia dalam berkomunikasi dan sebagai sarana menambah cita rasa hidup dengan memperkenalkan cara mengekspresikan perasaan mereka dalam bentuk komunikasi [3].

Analisis emosi manusia telah menjadi topik penelitian dalam berbagai disiplin ilmu, seperti Ilmu Kognitif, Psikologi. Berkat difusi media sosial ini dapat menarik minat para ilmuwan



komputer atau mesin cerdas [4]. Akibat perkembangan zaman, interaksi kerjasama manusia-mesin semakin sering ditemui sehingga kebutuhan akan sistem dalam mengerti manusia semakin meningkat, salah satu contohnya adalah mesin dipaksa untuk dapat memahami keadaan emosi dari suara manusia. Ketika mesin dapat mendeteksi emosi manusia dari interaksi yang berlangsung, hal itu dapat berguna untuk mengetahui kondisi psikologis atau emosi dari manusia tersebut [5]. Selain itu, pengenalan emosi yang dilakukan mesin sangat berguna untuk aplikasi yang membutuhkan interaksi manusia-mesin seperti *customer service* [6]. *Customer service* dalam pelayanannya memerlukan respon yang disesuaikan berdasarkan emosi yang terdeteksi. Oleh karena itu, dikenalkan bidang penelitian yang baru, yaitu *speech emotion recognition* (SER).

*Speech emotion recognition* adalah sebuah sistem yang dapat mengklasifikasikan perasaan atau emosi seseorang dari gaya berbicaranya. Berdasarkan teori *speech emotion analysis*, emosi seseorang dapat dirasakan secara nonverbal dari perubahan ritme respirasi, tensi otot yang menggetarkan suara dan mengganti karakteristik akustik, dan lain-lain [7]. *Speech emotion recognition* diyakini dapat meningkatkan kinerja sistem pengenalan suara sehingga membantu dalam investigasi kriminal, pengawasan dan deteksi peristiwa yang berpotensi berbahaya, serta membantu sistem perawatan kesehatan [8]. Untuk dapat mengenali emosi secara efektif dan akurat dari audio ucapan seseorang, audio tersebut perlu diekstraksi dari audio analog ke audio digital sehingga dapat diproses lebih lanjut. Akan tetapi, diferensiasi dari emosi terbukti susah untuk ditentukan karena sifat natural suara yang kompleks. Beberapa penelitian terkait SER telah dilakukan dengan menggunakan metode *Hidden Markov Model* (HMM) dan MFCC berhasil menentukan 4 dari 7 jenis emosi yang digunakan dengan akurasi lebih dari 50% [5], penelitian lain menerapkan algoritma *boosting* untuk SER dalam bahasa Indonesia dan memperoleh akurasi sebesar 65% [9], dan pemodelan dengan berbagai metode RNN menghasilkan Bi-LSTM sebagai model terbaik dalam menghasilkan error yang paling kecil di antara metode RNN lainnya [10].

Salah satu metode ekstraksi suara yang banyak digunakan dalam bidang *speech technology* yakni rekognisi suara maupun ucapan adalah MFCC. MFCC sendiri digunakan untuk melakukan ekstraksi fitur dengan mendapatkan nilai-nilai tertentu sebagai konsep konversi sinyal suara menjadi beberapa parameter tertentu. Penerapan ekstraksi fitur MFCC berdasarkan penelitian terdahulu yang menerapkan ekstraksi fitur MFCC terbukti sangat baik dengan taraf keakuratan tinggi dalam mengekstraksi fitur suara yang merepresentasikan suara manusia, terutama pada *speech recognition* [11].

Berdasarkan permasalahan di atas, paper ini mengusulkan *speech emotion recognition* (SER) menggunakan metode Bi-LSTM dengan ekstraksi fitur menggunakan MFCC. Selain menggunakan metode Bi-LSTM, akan digunakan juga metode LSTM dan CNN untuk digunakan perbandingan dalam menghasilkan metode mana yang paling baik. Dalam mengatasi masalah terkait *imbalanced*, akan digunakan *class weights* sebagai metode dalam mengatasi kelas yang *imbalanced* agar tidak terlalu mengubah informasi data audio yang diperoleh. Oleh karena itu, penelitian ini bertujuan mendapatkan informasi dari *audio mining* untuk mendeteksi SER dengan tepat dan akurat menggunakan pendekatan model Bi-LSTM, LSTM, dan CNN kemudian dievaluasi menggunakan satuan ukur nilai akurasi.

II.

## II. METODE PENELITIAN

### 2.1. Sumber Data

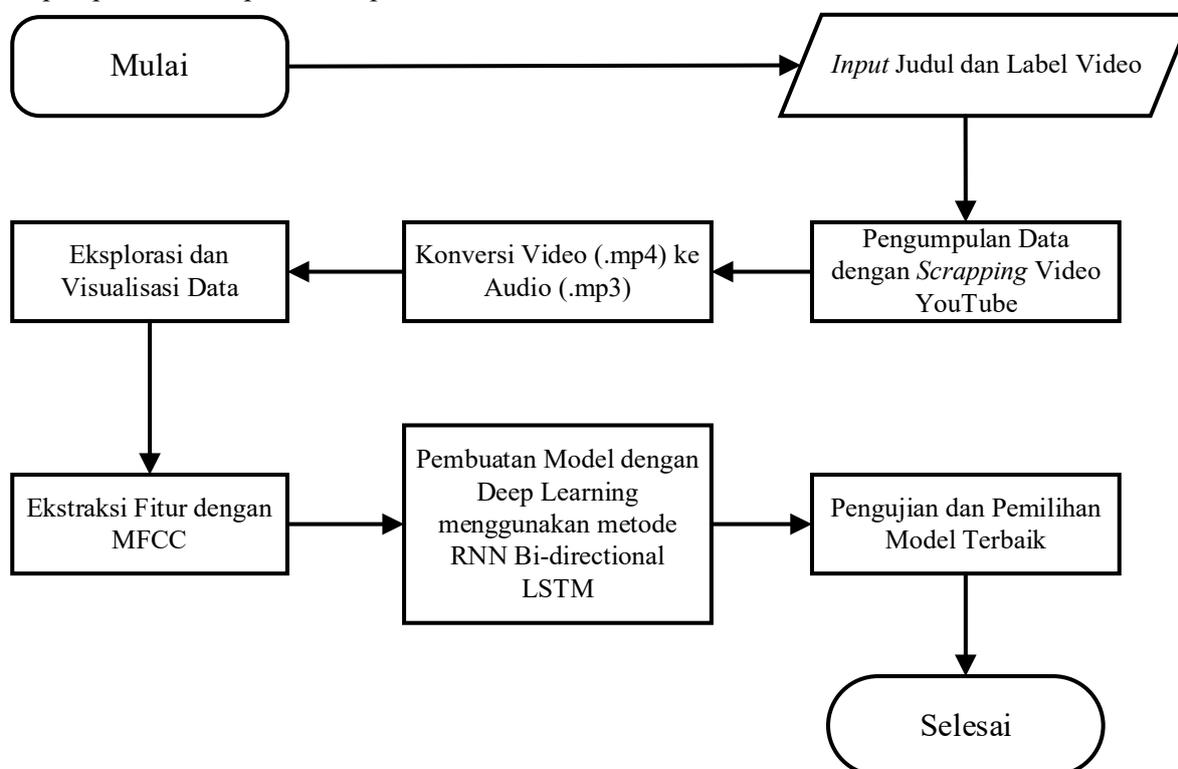
Data *audio* didapatkan dengan melakukan *scraping* atau mengambil data yang tersedia secara *online*. Sumber data yang digunakan adalah dari platform media sosial *YouTube* dimana tersedia konten video dengan jumlah yang sangat banyak. Platform ini memberikan peneliti banyak pilihan data yang dapat digunakan. Peneliti harus jeli dalam memilih video dimana suara terdengar bersih dengan emosi yang jelas. Pemilihan video yang ambigu (sarkas, banyak orang, atau banyak suara lain) tidaklah

dipreferensikan. Ditemukan bahwa video terbaik yang digunakan untuk analisis adalah video bergenre esai video, komentar, atau video log (*vlog*).

Teknik pengambilan video adalah dengan membuat skrip otomatis menggunakan bahasa pemrograman *Python*, yang dilengkapi dengan library utama *yt-dlp*, *ffmpeg* dan *youtube-search*. Library lain yang digunakan untuk pemrosesan data pun juga digunakan. Cara kerja program *scraping* ini adalah dengan mencari video dari platform *YouTube* sesuai dengan judul yang ditentukan oleh peneliti yang dimasukkan ke dalam program. Setelah video ditemukan, program akan meminta peneliti untuk memilih label yang tepat yang merepresentasikan video tersebut. Setelah label ditentukan, barulah program akan mengunduh dan mensegmentasi video menjadi beberapa bagian audio masing-masing berdurasi 3 detik dan diletakan di folder sesuai dengan label yang telah ditentukan.

## 2.2. Metodologi Penelitian

Tahapan analisis dalam penelitian ini adalah studi literatur, penentuan metode dan pengumpulan data, eksplorasi dan visualisasi data, ekstraksi fitur, pembuatan model, dan evaluasi model. *Flowchart* tahapan penelitian dapat dilihat pada Gambar 1.



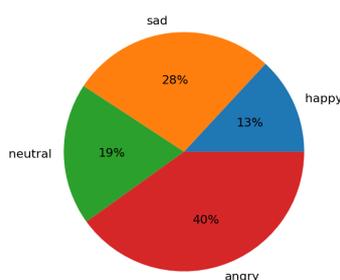
**Gambar 1.** *Flowchart* Penelitian

Adapun penjelasan tahapan analisis secara rinci adalah sebagai berikut:

1. Studi literatur dilakukan untuk menentukan pendekatan terbaik untuk mendapatkan sistem yang efektif dan efisien berdasarkan penelitian dan pendekatan yang telah dilakukan sebelumnya oleh peneliti lain. Peneliti mencari penelitian-penelitian sebelumnya terkait audio *mining* di GoogleScholar dengan kata kunci “Audio Mining”. Setelah itu, menentukan topik untuk audio *mining* berdasarkan kemampuan dan ke-*familiar*-an topik yang dipilih;
2. Menentukan metode dan objek pengumpulan data yang dalam hal ini adalah *scraping* video dari laman YouTube. Peneliti mencari dan mengunduh video yang relevan dengan pengekspresian emosi manusia ketika, marah, senang, netral, dan sedih. Video lalu

disegmentasi per 3 detik lalu diekspor menjadi file suara .mp3. Dalam tahap *scrapping*, peneliti membuat modul dalam format .py. Modul tersebut berfungsi untuk melakukan *scrapping* serta mengubahnya menjadi potongan-potongan audio berdurasi 3 detik yang terkumpul dalam folder khusus. Hal ini dimaksudkan agar dalam proses *scrapping* video, peneliti tidak perlu mengeksekusi kode rumit berulang kali dan hanya perlu meng-*import* modul yang telah dibuat lalu menginisiasinya. Pemotongan audio menjadi 3 detik dilakukan dengan tujuan agar dalam pemrosesan data audio tidak terlalu berat dan ketika audio dibagi menjadi beberapa segmen, tiap segmen memuat informasi yang berbeda-beda sehingga sistem mendeteksi data yang heterogen dari satu audio yang berdurasi lebih dari 3 detik. Dengan demikian, dapat dilakukan proses *train test split* dengan banyak data;

3. Eksplorasi dan visualisasi data (EVD) dilakukan untuk menentukan pendekatan yang lebih spesifik terhadap data yang diperoleh di tahap sebelumnya. Hal ini diharapkan akan memberikan ide akan parameter yang akan digunakan pada tahap ekstraksi fitur serta mengetahui informasi dari data yang diperoleh sehingga dapat melakukan analisis/pemodelan yang tepat. EVD yang dilakukan dalam penelitian ini adalah membuat plot *waveform*, plot *spectrum*, mencari nilai dominan dari *fast fourier transform* (FFT), dan plot *spectrogram* pada tiap kelas yang ada. *Pie chart* pada Gambar 2 dibuat untuk menunjukkan jumlah data dari masing-masing label.



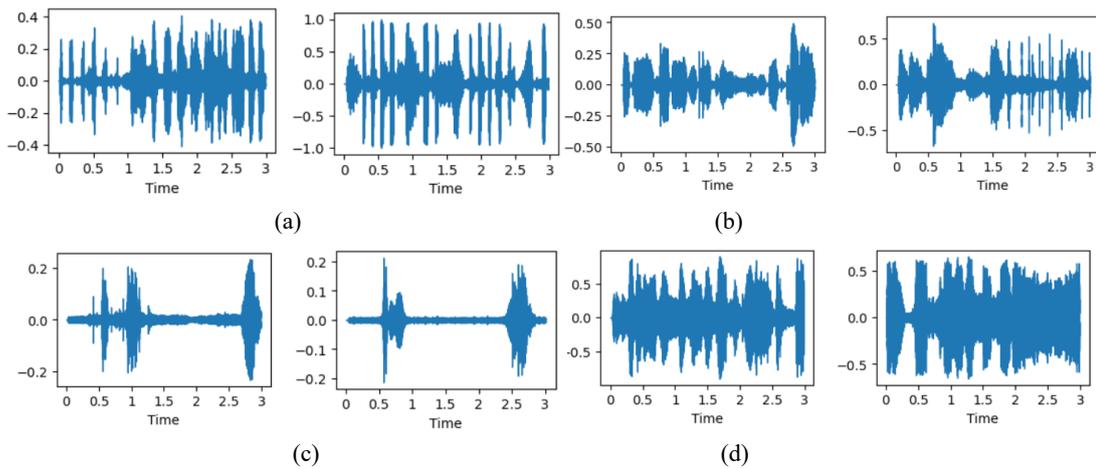
**Gambar 2.** *Pie Chart* Label Emosi

Dapat dilihat bahwa label angry merupakan label yang mempunyai jumlah data paling banyak di antara label yang lainnya, yakni sebesar 40%. Sedangkan untuk label happy merupakan label yang mempunyai jumlah data paling sedikit, yakni hanya sebesar 13% saja. Disebabkan distribusi jumlah data yang tidak merata maka dapat dikatakan bahwa terjadi *imbalanced* data dalam penelitian ini. Adapun rincian jumlah data dapat dilihat pada Tabel 1.

**Tabel 1.** Jumlah Data pada Tiap Kelas Emosi

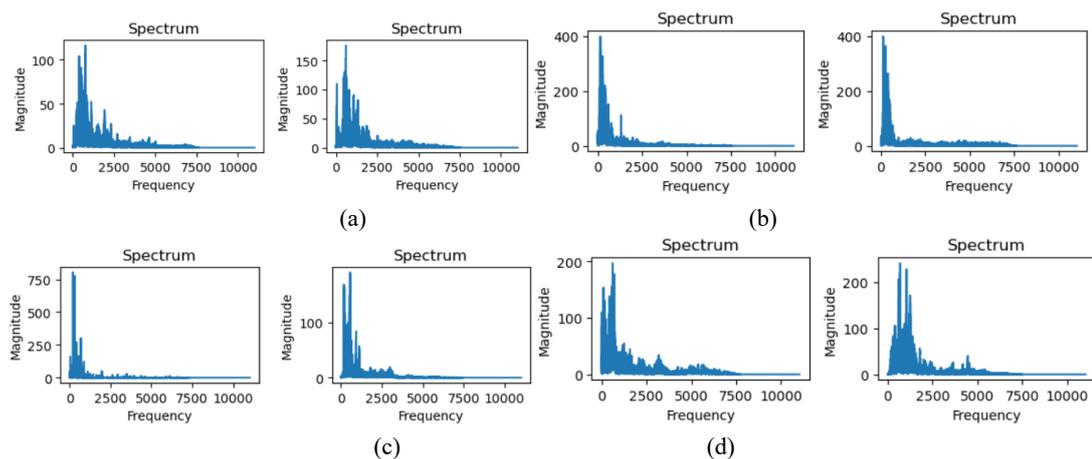
Happy	Sad	Neutral	Angry	Jumlah Data
427	897	624	1.302	3.250

Selanjutnya dibuat *waveform* pada Gambar 3 untuk setiap emosi sehingga dapat memberikan informasi terkait amplitudo dari sinyal audio. Hasilnya adalah pola yang dibentuk oleh keempat *waveform* cukup berbeda. *Waveform happy* mempunyai pola dengan amplitudo yang seringkali naik dalam beberapa waktu. *Waveform neutral* mempunyai pola yang cenderung stabil. *Waveform sad* mempunyai pola yang cukup stabil, yakni kenaikan amplitudo hanya di awal dan akhir, dan amplitudo yang dibentuk cenderung mendekati garis 0. Adapun pola dengan amplitudo tinggi dicurigai akibat adanya *noise* tangisan pada sampel suara *sad*. *Waveform angry* mempunyai pola dengan amplitudo yang lebih sering naik dan kenaikan tersebut cukup konstan jika dibandingkan dengan *waveform happy*.



**Gambar 3.** *Waveform* (a) *happy*; (b) *neutral*; (c) *sad*; (d) *angry*

Audio *angry* mempunyai pola amplitudo tinggi yang lebih konstan di setiap waktunya jika dibandingkan dengan *waveform* emosi lainnya disebabkan ketika emosi yang dirasakan adalah marah, manusia cenderung mengeluarkan suara yang lebih keras dibandingkan suara biasanya. Begitu pun dengan audio *happy* yang jika dilihat amplitudo yang terlihat cenderung lebih banyak yang naik karena ketika emosi yang dirasakan adalah senang, manusia cenderung sedikit menaikkan suara bicara atau tertawa. Ketika manusia merasakan emosi *neutral*, suara yang diberikan biasa saja dalam artian tidak terlalu kecil dan tidak terlalu besar, sedangkan ketika manusia merasakan emosi *sad*, suara yang diberikan cenderung memberikan amplitudo yang lebih kecil dibandingkan yang lainnya disebabkan rata-rata manusia ketika merasakan sedih cenderung menangis atau diam. Selanjutnya dibuat *spectrum* pada Gambar 4 sebagai representasi visual dari sinyal audio dalam domain frekuensi sehingga dapat mengidentifikasi frekuensi dari sinyal audio berdasarkan empat jenis emosi manusia. Hasilnya adalah *spectrum happy* dan *angry* mempunyai magnitudo dominan yang cenderung lebih banyak pada rentang frekuensi 0 – 2500 jika dibandingkan dengan *spectrum neutral* dan *sad*.



**Gambar 4.** *Spectrum* (a) *happy*; (b) *neutral*; (c) *sad*; (d) *angry*

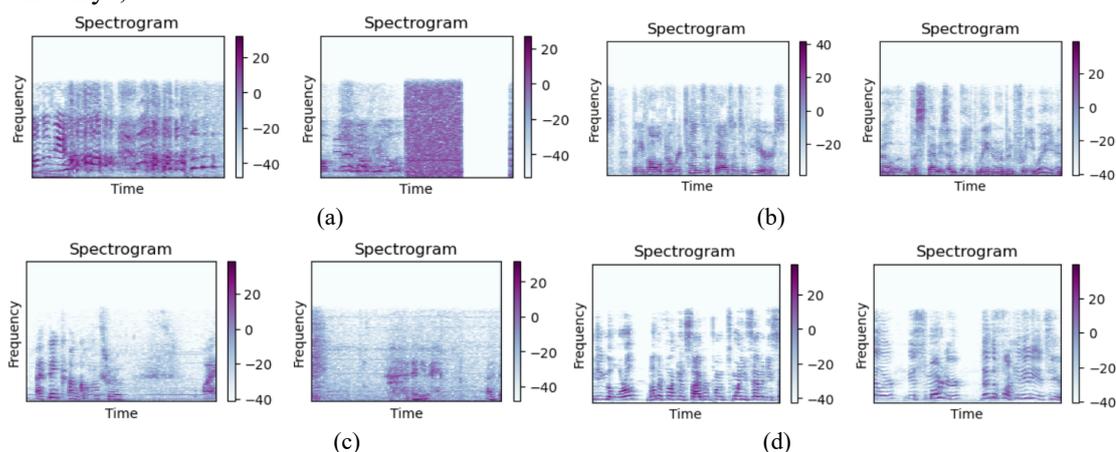
Untuk dapat menginterpretasikannya dengan jelas, peneliti mencari nilai rata-rata dominan frekuensi dari keseluruhan data tiap kelas. Hasil dari rata-rata tersebut akan dibagi dengan tiga yang merupakan durasi tiap audio yang kemudian akan menghasilkan rata-rata nilai Hz dari tiap kelas. Adapun nilai rata-rata dominan frekuensi dan rata-rata nilai Hz tiap kelas dapat dilihat pada Tabel 2.

**Tabel 2.** Nilai Rata-Rata Dominan Frekuensi dalam Hz

Happy	Sad	Neutral	Angry
19305,95	17237,37	16548,24	17640,45

Berdasarkan Tabel 2, label *happy* menghasilkan nilai 19305,95 Hz, *sad* menghasilkan nilai 17237,37 Hz, *neutral* menghasilkan nilai 16548,24 Hz, dan *angry* menghasilkan nilai 17640 Hz. Hal ini berarti dominan frekuensi yang paling besar dihasilkan oleh label *happy*, disusul dengan label *angry*, *sad*, dan *neutral*. Dengan demikian, dapat disimpulkan bahwa rata-rata dominan frekuensi yang dihasilkan label *happy* ternyata lebih besar dibandingkan rata-rata dominan frekuensi kelas yang lain.

Lalu, pembuatan *spectrogram* pada Gambar 5 dilakukan untuk menampilkan informasi spektrum frekuensi sinyal dalam bentuk warna sehingga lebih mudah dalam mengidentifikasi frekuensi dan waktu dari sinyal audio serta lebih mudah dalam menemukan pola dalam sinyal audio. Hasilnya adalah dapat dilihat bahwa *spectrogram happy* terlihat ada jarak antar *sample*, tetapi ada juga *spectrogram* yang mempunyai suara paling keras di tengah-tengah daripada suara sebelum dan sesudahnya. Pada label *sad* terdapat jarak yang seragam per *sample* diantara puncak-puncak frekuensi. Jarak tersebut dapat dijelaskan oleh pembicara yang mengambil waktunya untuk bernapas atau menghirup udara dari hidung. Frekuensi yang rata-rata rendah juga menjadi pola dari emosi sedih. Pada label *neutral* distribusi suara yang terdengar terlihat cukup merata dibandingkan dengan label lainnya. Hal ini berarti dalam kelas *neutral*, manusia cenderung berbicara tanpa henti dibandingkan ketika merasa senang, sedih, dan marah. *Spectrogram* label *angry* juga terdapat jarak, diasumsikan jarak tersebut adalah ketika manusia menggunakan waktu tersebut untuk bernapas yang mungkin dilakukan untuk mengontrol emosinya;



**Gambar 5.** *Spectrogram* (a) *happy*; (b) *neutral*; (c) *sad*; (d) *angry*

4. Ekstraksi fitur adalah tahap terpenting yang dilakukan pada penelitian ini disebabkan hasil akhir model akan ditentukan oleh seberapa baik data yang masuk ke dalam sistem. Ada banyak parameter yang diubah untuk mendapatkan fitur terbaik di antara kelas. Parameter tersebut ditentukan berdasarkan hasil dari tahap sebelumnya. Peneliti melakukan *label encoding* terhadap variabel-variabel yang digunakan. Hasil dari *label encoding* tersebut disimpan dalam format *.npy*. Lalu, peneliti melakukan ekstraksi fitur menggunakan metode MFCC dari data yang telah dilakukan standarisasi dan *balancing*.
5. Pembuatan model dengan arsitektur *deep learning* dilakukan setelah diperoleh data hasil ekstraksi fitur pada tahap sebelumnya. Model dibuat dengan metode RNN Bi-directional LSTM berdasarkan penelitian-penelitian sebelumnya yang menghasilkan model yang mendekati sempurna dengan menggunakan metode tersebut. Bidirectional LSTM merupakan tipe arsitektur *Recurrent Neural Network* (RNN) yang terdiri dari dua lapisan LSTM yang berbeda, yaitu lapisan LSTM maju (*forward LSTM layer*) dan lapisan LSTM mundur (*backward LSTM layer*). Arsitektur ini memungkinkan data masukan diolah dalam arah maju dan mundur, sehingga jaringan dapat menangkap karakteristik dan pola data yang sebelumnya diabaikan

oleh jaringan LSTM tradisional, pengolahan arah mundur pada jaringan Bi-LSTM memiliki keunggulan tersendiri, karena memungkinkan jaringan untuk belajar dari konteks data masa depan dan masa lalu. Dengan menangkap informasi dari kedua arah, jaringan dapat memodelkan ketergantungan dalam data masukan dengan efektif dan meningkatkan akurasi [12];

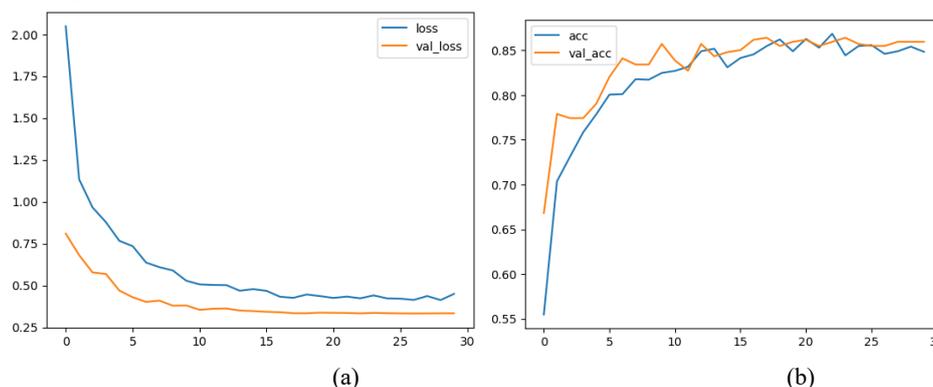
6. Tahap akhir adalah melakukan evaluasi berdasarkan akurasinya dan dilakukan prediksi langsung terhadap model yang telah dibuat. Prediksi dilakukan dengan merekam suara secara langsung dan disimpan dalam bentuk .mp3 yang setelahnya akan dimasukkan ke dalam model untuk dilakukan prediksi kelas emosi terhadap suara tersebut.

### III. HASIL DAN PEMBAHASAN

Pada paper ini, jaringan Bi-LSTM digunakan untuk mempelajari fitur emosi dari klip audio mentah yang sudah diekstraksi fitur dengan MFCC serta membagi data menjadi dua set secara acak, yaitu set pelatihan dengan 80% data dan set pengujian dengan 20% data. Meskipun jaringan dalam *deep learning* dianggap sebagai "black box" *approach*, yakni cara kerjanya tidak begitu jelas, jaringan *deep learning* ini biasanya digunakan untuk menemukan algoritma yang dapat melakukan prediksi. Jaringan Bi-LSTM yang dirancang juga digunakan untuk kekuatan prediktifnya daripada kekuatan penjelas yang lemah. Untuk mengurangi kemungkinan atau jumlah *overfitting* pada penelitian, beberapa teknik diperkenalkan. *Overfitting* adalah salah satu alasan prediksi buruk untuk data sampel yang tidak terlatih. Ketika *overfitting* terjadi, model yang terlalu dipaksakan hanya mengingat data pelatihan daripada belajar untuk memprediksi dengan lebih baik. Ada banyak alasan terjadinya *overfitting*. Jika jaringan *deep learning* sangat kompleks, maka *overfitting* akan terjadi. Jika jaringan *deep learning* *overtrained*, *overfitting* juga akan timbul. Ketika derajat kebebasan model yang diadopsi dalam pelatihan jaringan terlalu banyak, maka kondisi *overfitting* akan ada. Oleh karena itu, dilakukan teknik normalisasi *batch*, *early stopping*, dan pemilihan model untuk mengatasi *overfitting*.

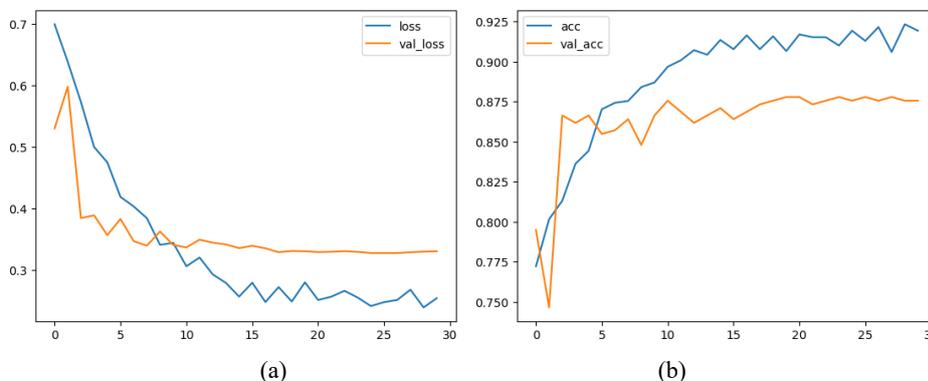
Dalam paper ini, hanya model-model terbaik yang dipilih untuk dicatat sebagai hasil yang tepat dan dapat diprediksi. Ketika akurasi validasi tidak lagi meningkat selama pelatihan model, itu menunjukkan bahwa model memiliki performa prediksi yang lebih superior (lihat Gambar 6). Dari gambar tersebut, dapat dilihat bahwa ketika akurasi validasi mencapai maksimum, akurasi pelatihan tidak mencapai maksimum. Ketika akurasi validasi menurun sementara akurasi pelatihan tetap meningkat, itu menunjukkan situasi *overfitting*. Oleh karena itu, pelatihan akan dihentikan dengan metode *early stopping*.

Pada grafik *loss* dan *accuracy* Gambar 6 dapat dilihat bahwa nilai *loss* dan *validation loss* dari model CNN berada di kisaran 0,5, sedangkan untuk nilai *accuracy* dan *validation accuracy* di kisaran 0,85. Disebabkan tidak adanya *gap* yang jauh antara garis kuning dan garis biru maka disimpulkan bahwa model yang dihasilkan tidak terjadi *overfit*.



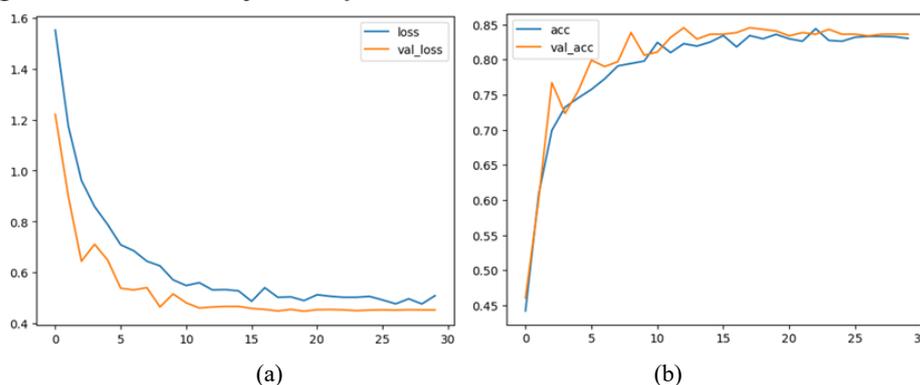
Gambar 6. Grafik (a) *loss*; (b) *accuracy* Model CNN

Pada grafik *loss* dan *accuracy* Gambar 7 dapat dilihat bahwa nilai *loss* dan *validation loss* dari model Bi-LSTM berada di kisaran 0,3, sedangkan untuk nilai *accuracy* dan *validation accuracy* dikisaran 0,88. Disebabkan tidak adanya *gap* yang jauh antara garis kuning dan garis biru maka disimpulkan bahwa model yang dihasilkan tidak terjadi *overfit*.



**Gambar 7.** Grafik (a) *loss*; (b) *accuracy* Model Bi-LSTM

Pada grafik *loss* dan *accuracy* Gambar 8 dapat dilihat bahwa nilai *loss* dan *validation loss* dari model LSTM berada di kisaran 0,5, sedangkan untuk nilai *accuracy* dan *validation accuracy* di kisaran 0,83. Disebabkan tidak adanya *gap* yang jauh antara garis kuning dan garis biru maka disimpulkan bahwa model yang dihasilkan tidak terjadi *overfit*.



**Gambar 8.** Grafik (a) *loss*; (b) *accuracy* LSTM

Tabel 3 menunjukkan hasil perbandingan dari beberapa metode yang dirancang dalam penelitian. Dari tabel ini, dapat dilihat bahwa jaringan Bi-LSTM menunjukkan keunggulan dalam kinerja secara keseluruhan. Rata-rata nilai *F1-Score* dan nilai ROC AUC yang dicapai dengan mempelajari fitur dalam bentuk MFCC lebih tinggi daripada jaringan lainnya. Dari gambar 6 dan gambar 8, juga dapat dilihat bahwa jaringan Bi-LSTM mencapai akurasi validasi tertinggi dengan jumlah *epoch* yang lebih sedikit dibandingkan dengan jaringan LSTM. Dengan kata lain, jaringan Bi-LSTM konvergen lebih cepat dibandingkan dengan jaringan LSTM dan CNN. Berdasarkan hasil tersebut, paper ini mengusulkan metode Bi-LSTM sebagai metode yang paling baik dalam memprediksi SER.

**Tabel 3.** Perbandingan Model

Model	<i>F1-Score</i>	ROC AUC
Bi-LSTM	0,878	0,979
LSTM	0,836	0,971
CNN	0,873	0,977

#### IV. KESIMPULAN

Berdasarkan hasil analisis dari setiap sampel suara didapatkan kesimpulan bahwa fitur penentu yang secara pasti dapat menjelaskan emosi seseorang melalui suara belum dapat ditentukan secara pasti.



Hal ini terjadi karena sifat natural manusia yang memiliki gaya berbicaranya sendiri dan juga intonasi serta pronounsiasi kata yang dapat mengganggu penentuan fitur. Akan tetapi, ada pola pada jeda, frekuensi suara, dan amplitudo yang dapat cukup menjelaskan emosi dari pembicara.

Dalam pembuatan model dengan *deep learning*, yaitu menggunakan metode Bi-LSTM yang dibandingkan dengan metode LSTM, dan CNN diperoleh hasil bahwa jaringan Bi-LSTM menunjukkan keunggulan dalam kinerja secara keseluruhan. Rata-rata nilai *F1-Score* dan nilai ROC AUC yang dicapai dengan mempelajari fitur dalam bentuk MFCC lebih tinggi daripada jaringan lainnya. Dengan demikian, jaringan Bi-LSTM konvergen lebih cepat dibandingkan dengan jaringan LSTM dan CNN. Oleh karena itu, metode Bi-LSTM yang diusulkan dalam paper ini dianggap sebagai metode yang paling baik dalam memprediksi SER. Pada penelitian selanjutnya dapat menggunakan data yang lebih heterogen dan melakukan *preprocessing* untuk pemilihan datanya sehingga tidak terlalu banyak *noise* yang ada. Pemilihan data yang heterogen sangat penting karena nyatanya dalam mengungkapkan emosi yang dirasakan, manusia mempunyai nada dan keras suara yang berbeda-beda.

## REFERENSI

1. Azhari, “IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK DALAM DETEKSI EMOSI MANUSIA BERDASARKAN EKSPRESI WAJAH,” Jul. 08, 2021. <http://eprosiding.ars.ac.id/index.php/pti/article/view/198>.
2. Rere, LM Rasdi. "Studi Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network." *Prosiding SeNTIK 3.1* (2019): 71-78.
3. S. Grover and A. Verma, “Design for emotion detection of punjabi text using hybrid approach,” in *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016, 2017*, vol. 2.
4. M. P. Skenduli, M. Biba, C. Loglisci, M. Ceci, and D. Malerba, “User-Emotion Detection Through Sentence-Based Classification Using Deep Learning: A Case-Study with Microblogs in Albanian Marjana,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11177 LNAI, Springer International Publishing, pp. v–vi, 2018.
5. A. A. Sundawa, A. G. Putrada, dan N. A. Suwastika, “Implementasi dan Analisis Simulasi Deteksi Emosi Melalui Pengenalan Suara Menggunakan Mel-Frequency Cepstrum Coefficient dan Hidden Markov Model Berbasis IOT.”
6. B. Schuller, G. Rigoll, dan M. Lang, “SPEECH EMOTION RECOGNITION COMBINING ACOUSTIC FEATURES AND LINGUISTIC INFORMATION IN A HYBRID SUPPORT VECTOR MACHINE-BELIEF NETWORK ARCHITECTURE.”
7. P. Juslin dan K. Scherer, “Speech emotion analysis,” *Scholarpedia*, vol. 3, no. 10, hlm. 4240, 2008, doi: 10.4249/scholarpedia.4240.Author 1, A.; Author 2, B. *Book Title*, 3rd ed.; Publisher: Publisher Location, Country, 2008; pp. 154–196.
8. K. Wang, N. An, B. N. Li, Y. Zhang, dan L. Li, “Speech emotion recognition using Fourier parameters,” *IEEE Trans Affect Comput*, vol. 6, no. 1, hlm. 69–75, Jan 2015, doi: 10.1109/TAFFC.2015.2392101.
9. F. KASYIDI, R. ILYAS, dan N. M. ANNISA, “Peningkatan Kemampuan Pengenalan Emosi Melalui Suara dalam Bahasa Indonesia,” *MIND Journal*, vol. 6, no. 2, hlm. 194–204, Des 2021, doi: 10.26760/mindjournal.v6i2.194-204.
10. A. Graves, “Supervised Sequence Labelling,” 2012, hlm. 5–13. doi: 10.1007/978-3-642-24797-2\_2.
11. F. W. Wibowo dan Institute of Electrical and Electronics Engineers, *2018 International Conference on Information and Communications Technology (ICOIACT) : 6-7 March 2018*.
12. Yildirim Ö (2018) A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med* 96:189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016>.



# Perspektif Opini Pelanggan pada Aplikasi MyPertamina menggunakan Metode Latent Dirichlet Allocation (LDA)

Ilham Ahmad Kamil<sup>1</sup>, Anisyaul Fitria<sup>2</sup>, Riri Dwi Setyawati Jatiningtyas<sup>3</sup>, Elfira Rahma Putri<sup>4</sup>, Mutiara Afifah<sup>5</sup>, Maryamah Maryamah<sup>6</sup>

<sup>1, 2, 3, 4, 5, 6</sup> *Teknologi Sains Data, Universitas Airlangga*

<sup>1</sup>[ilham.ahmad.kamil-2020@ftmm.unair.ac.id](mailto:ilham.ahmad.kamil-2020@ftmm.unair.ac.id)

<sup>2</sup>[anisyaul.fitria-2020@ftmm.unair.ac.id](mailto:anisyaul.fitria-2020@ftmm.unair.ac.id)

<sup>3</sup>[riri.dwi.setyawati.jatiningtyas-2020@ftmm.unair.ac.id](mailto:riri.dwi.setyawati.jatiningtyas-2020@ftmm.unair.ac.id)

<sup>4</sup>[elfira.rahma.putri-2020@ftmm.unair.ac.id](mailto:elfira.rahma.putri-2020@ftmm.unair.ac.id)

<sup>5</sup>[mutiara.afifah-2020@ftmm.unair.ac.id](mailto:mutiara.afifah-2020@ftmm.unair.ac.id)

<sup>6</sup>[maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

Corresponding author email: [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

**Abstract:** *Fluctuations in world oil prices are one of the major factors affecting the country's economic growth, including Indonesia. The government issued a fuel subsidy program, but the program was used by the wrong target and caused the government budget to swell. The solution is to provide programs to the right targets, the government created the MyPertamina application. MyPertamina's application reviews need to be further analyzed and researched to find out the customer's perspective on the application and help the government improve its work program. This study proposes the LDA (Latent Dirichlet Allocation) topic modeling method to analyze user reviews of the MyPertamina application. The methodology used in this research is starting with a literature study, data collection, data preprocessing, sentiment analysis, and topic modeling using LDA. The results of the analysis show that there are several negative reviews with most reviews being positive. This indicates that the MyPertamina application is a good program and right on target in distributing subsidized fuel. Based on the results of the analysis, there are some suggestions about improving application performance in the registration process and adding payment options through e-wallets other than LinkAja or m-banking. The conclusion of the perspective user, the MyPertamina application is effective in controlling subsidized fuel purchases and several user suggestions need to improve regarding registration and payment.*

**Keywords:** *MyPertamina, Opinion Perspective, User reviews, Latent Dirichlet Allocation, Topic Modelling*

**Abstrak:** Fluktuasi harga minyak dunia menjadi salah satu faktor besar yang mempengaruhi pertumbuhan ekonomi negara, termasuk Indonesia. Pemerintah mengeluarkan program subsidi BBM (Bahan Bakar Minyak) namun program dimanfaatkan oleh target yang kurang tepat sasaran dan menyebabkan membengkaknya anggaran pemerintah. Solusi memberikan program kepada target yang tepat, pemerintah membuat aplikasi MyPertamina. Ulasan aplikasi MyPertamina perlu dianalisa dan diteliti lebih lanjut untuk mengetahui perspektif opini pelanggan aplikasi dan membantu pemerintah meningkatkan program kerjanya. Penelitian ini mengusulkan metode *topic modelling* LDA (*Latent Dirichlet Allocation*) untuk menganalisis ulasan pengguna aplikasi MyPertamina. Metodologi yang digunakan dalam penelitian ini yaitu dimulai dengan studi literatur, pengumpulan data, data preprocessing, analisis sentimen, pemodelan topik menggunakan LDA. Hasil analisis menunjukkan bahwa meskipun terdapat beberapa ulasan negatif dengan mayoritas ulasan positif. Hal ini mengindikasikan bahwa aplikasi MyPertamina sudah bagus dan tepat sasaran dalam penyaluran BBM subsidi. Berdasarkan hasil analisa, beberapa saran juga disampaikan untuk meningkatkan performa aplikasi, seperti membuat pendaftaran yang lebih fleksibel dan menambahkan opsi pembayaran melalui e-wallet lain selain LinkAja atau m-banking. Opini pengguna terhadap aplikasi MyPertamina dapat dikatakan cukup efektif dalam mengontrol pembelian BBM subsidi dan beberapa saran dari opini pengguna menuliskan perlu adanya perbaikan terkait pendaftaran dan pembayaran.

**Kata kunci:** *MyPertamina, Perspektif Opini, Opini Pelanggan, Latent Dirichlet Allocation, Topic Modelling*



## I. PENDAHULUAN

Perekonomian internasional, khususnya Indonesia, terkena dampak signifikan akibat pandemi Coronavirus Disease 2019 (COVID-19) [1]. Fluktuasi harga minyak dunia menjadi salah satu faktor besar yang mempengaruhi pertumbuhan ekonomi negara, baik negara pengimpor maupun negara pengekspor minyak seperti Indonesia. Harga minyak bumi mengalami peningkatan tahun 2021 dari US\$ 70an per barel akan melonjak menjadi US \$100an pada tahun 2022 sehingga berdampak dalam ekonomi Indonesia. Dampak tersebut salah satunya terjadi peningkatan anggaran subsidi Bahan Bakar Minyak (BBM) dari APBN negara yang signifikan [2]. Padahal kebutuhan Bahan Bakar Minyak (BBM) di Indonesia sangat besar dengan jumlah penduduk yang besar.

Kebijakan yang diterapkan oleh pemerintah dalam mengendalikan segala permasalahan BBM subsidi adalah dengan membentuk aplikasi bernama MyPertamina yang dikelola oleh PT. Pertamina (Persero). MyPertamina merupakan program *loyalty* dan *e-payment* yang disediakan oleh PT Pertamina guna memberikan *user experiences* untuk seluruh pelanggan Pertamina. Dalam layanan *e-money*, telah terdaftar dan diawasi oleh Bank Indonesia. MyPertamina berfungsi sebagai *cashless payment* (sistem pembayaran non-tunai), sarana pelanggan untuk mendapatkan point dan reward serta pengaplikasian *e-voucher* yang dapat digunakan di berbagai merchant yang telah berkerjasama dengan PT Pertamina melalui MyPertamina Loyalty Program [3]. Tujuan lain dari aplikasi MyPertamina ini adalah untuk menjaga kuota subsidi BBM agar tidak melebihi target serta memastikan bahwa subsidi BBM diterima oleh masyarakat yang benar-benar membutuhkan atau tepat sasaran [4]. Cara untuk mengetahui layak atau tidaknya konsumen tersebut menerima bantuan subsidi BBM adalah dengan mendaftarkan kendaraan melalui aplikasi tersebut dengan memasukkan nama lengkap, nomor telepon yang aktif, Kartu Tanda Penduduk (KTP), nomor polisi, kapasitas mesin, dan foto kendaraan pribadi.

Melalui permasalahan tersebut selanjutnya, akan dilakukan *Data Mining* agar dapat mengetahui penyimpangan, design, maupun interaksi atau keterkaitan dari kumpulan data [5] dan menemukan tren dan pola tersembunyi yang tidak muncul dalam analisis kueri sederhana [6]. *Data Mining* yang termasuk di dalam penelitian ini yaitu *text mining*. *Text Mining* merupakan pengambilan atau penambangan data berupa data teks merupakan suatu proses ekstraksi baik berupa pola maupun informasi yang berasal dari sebuah data dalam bentuk teks [7]. *Text mining* sendiri terdiri atas *information retrieval*, *machine learning*, statistika, dan *linguistik komputasi* [8]. *Text mining* dapat digunakan dapat berbagai bidang seperti bisnis, komunikasi, sosial, dan lain-lain. Selanjutnya, dilakukan analisis menggunakan teknik *topic modelling* yang secara otomatis mampu menemukan topik dalam sekumpulan dokumen dengan melakukan pengindeksan laten semantic [9]. *Topic modelling* yang digunakan dalam penelitian ini adalah LDA (*Latent Dirichlet Allocation*) yang digunakan untuk mengetahui perspektif opini pelanggan dari aplikasi MyPertamina di kalangan masyarakat dari ulasan di dalam Google Playstore. Penggunaan LDA dalam penelitian ini dikarenakan LDA (*Latent Dirichlet Allocation*) merupakan suatu model probabilistik yang mampu memperkirakan probabilitas munculnya sebuah topik di dalam dokumen berdasarkan document term matrix. *Topic modelling* ini mampu memproses data yang berukuran besar secara efisien karena dalam pemrosesan datanya digunakan pemrograman Bayesian sehingga dapat mengolah dataset yang sangat besar dalam waktu yang wajar [10]. Selain itu, LDA menganggap sebuah dokumen sebagai sesuatu yang tersusun atas data acak serta mengandung topik tersembunyi di dalamnya, dan tiap topik dikarakterisasi melalui distribusi kata [11]. Dengan memanfaatkan ulasan pengguna aplikasi MyPertamina yang berupa rating dan komentar, dapat digunakan LDA untuk mendapatkan dan menganalisis topik-topik yang terbentuk sehingga dapat diketahui perspektif opini pelanggan dari aplikasi MyPertamina berdasarkan konsumen.

Oleh karena itu, pada penelitian ini peneliti mengusulkan penggunaan *topic modeling* LDA untuk mengetahui Perspektif Opini Pelanggan pada Aplikasi MyPertamina dengan cara mengolah data ulasan pengguna aplikasi MyPertamina dan mengolahnya dengan LDA untuk mendapatkan topik-topik penting yang ada. Hasil pemodelan topik ini dapat digunakan untuk mengetahui perspektif opini

pelanggan dari aplikasi MyPertamina sebagai penyalur subsidi BBM yang dapat digunakan sebagai pertimbangan dalam melakukan pengambilan keputusan bagi pemerintah maupun peneliti selanjutnya.

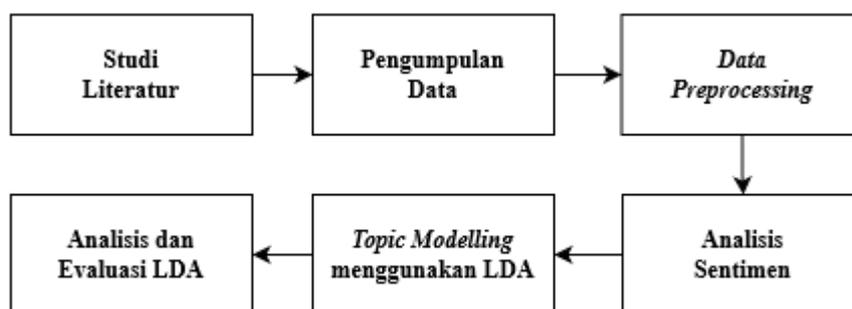
## II. METODE PENELITIAN

Dalam metode penelitian ini, akan dibahas mengenai sumber data dan metodologi yang digunakan dalam melakukan penelitian. Sumber data dalam penelitian menjelaskan mengenai subjek dari mana data penelitian tersebut diperoleh serta informasi mengenai cara pengambilan data dan pengolahannya. Lalu, metodologi penelitian merupakan sebuah pengetahuan untuk mendapatkan pengertian yang lebih mendalam terkait sistematisasi dan langkah-langkah yang dilakukan dalam penelitian [17].

### 2.1 Sumber Data

Pada penelitian ini kelompok kami menggunakan data *scraping*. Data *scraping* adalah data yang diperoleh dengan cara mengekstrak data dari aplikasi *enterprise*, *website*, *database* atau lainnya. Sumber data yang digunakan pada penelitian ini didapatkan dari data *scraping review* (ulasan) dari aplikasi MyPertamina di *Google Play Store* dengan mendapatkan 10.000 data ulasan dari aplikasi MyPertamina rating 1-5 diambil secara seimbang yaitu 2000 ulasan per rating.

### 2.2 Metodologi Penelitian



**Gambar 1.** Metodologi Penelitian

Alur penelitian yang dilakukan yaitu terdiri dari beberapa alur yang diawali dengan studi literatur hingga melakukan hasil dan visualisasi LDA. Berikut merupakan alur dari penelitian yang dilakukan antara lain :

#### 2.2.1 Studi Literatur

Studi literatur dilakukan dengan mencari informasi melalui studi buku, jurnal, internet, dan media yang berhubungan dengan topik modeling LDA dan informasi yang didapat akan digunakan sebagai acuan teori dan pembahasan penelitian.

#### 2.2.2 Pengumpulan Data

Pengumpulan data, peneliti melakukan pengumpulan sumber data dalam bentuk data primer. Data primer merupakan sumber data penelitian yang diperoleh peneliti secara langsung seperti melalui survei, eksperimen, dan sebagainya. Pada penelitian ini, didapatkan data primer dengan dilakukan *scraping* pada *Google Playstore* untuk mendapatkan ulasan pada aplikasi MyPertamina sebanyak 10.000 ulasan terbaru dengan rating 1-5 diambil secara seimbang yaitu 2000 ulasan per rating. Variabel data yang akan digunakan dalam penelitian ini yaitu *content* yang merupakan ulasan dari para pengguna aplikasi MyPertamina. Data yang diperoleh dilakukan proses *scraping* data dengan menggunakan Python. Pada analisis ini pengumpulan data dilakukan dengan *scraping* ulasan melalui Google Play. Untuk cara *scraping*nya sendiri yang harus dilakukan pertama adalah dengan mengimport seluruh library yang dibutuhkan yaitu *package google\_play\_scraper* sebagai API untuk *scraping* data dari Google Play, *pandas*, dan *numpy*. Selanjutnya, memilih aplikasi spesifik yang ingin *discraping* beserta jumlah data yang akan diambil dan juga filternya seperti waktu *publish* ulasan yang

terbaru dan juga rating dari ulasannya. Setelah itu, data yang didapatkan disimpan ke dalam bentuk format CSV untuk diolah lebih lanjut.

### 2.2.3 Data Preprocessing

Tahap *preprocessing* data terdiri dari *cleaning data*, *tokenize*, *stopword*, *remove blank*, dan *stemming*. Tahapan pertama yang dilakukan adalah *cleaning data* dimana proses tersebut merupakan proses pembersihan data dilakukan dengan menjadikan setiap katanya menjadi *lowercase*. Lalu, menghapus karakter unicode, URL, sebutan atau alias, tagar, tanda baca, dan nomor dengan bantuan regex. Hal ini ditujukan untuk memudahkan proses analisis selanjutnya. Kedua *tokenize*, setelah dilakukan *cleaning data* dilanjutkan dengan tokenisasi dengan bantuan *library word\_tokenize* dari *nlk.tokenize*. Tokenisasi berguna untuk mengubah susunan kalimat-kalimat menjadi bentuk kata. Ketiga *stopword*, tahapan *preprocessing* yang digunakan untuk menghapus kata-kata yang sering muncul dalam korpus seperti *ada*, *adalah*, *di*, *ke*. Kata-kata tersebut dihapuskan dari dataset karena seringkali tidak berguna saat dilakukan analisis lebih lanjut. Untuk melakukan *preprocessing stopwords* dapat digunakan *library stopwords* dari *nlk.corpus*. Keempat *remove blank*, tahapan yang digunakan untuk menghapus kolom yang kosong setelah *stopword* agar tidak ada kolom yang kosong sehingga seluruh data penting untuk dilakukan tahapan selanjutnya. Kelima *stemming*, Tahap *stemming* merupakan tahap terakhir pada *preprocessing*. *Stemming* berfungsi untuk mengembalikan semua kata pada korpus menjadi bentuk dasarnya kembali seperti *menangis* diubah menjadi *tangis*. Proses ini dapat dilakukan dengan bantuan *library StemmerFactory* dari *package Sastrawi*.

	content	content_clean	content_token	content_stopword	clean_data
0	Mekanisasi PERTANIAN, menambah lahan pertanian...	mekanisasi pertanian menambah lahan pertanian ...	[mekanisasi, pertanian, menambah, lahan, perta...	[mekanisasi, pertanian, menambah, lahan, perta...	mekanisasi tani tambah lahan tani bantu mesin ...
1	Baru terdaftar baru 2 kali di gunakan ko sudh ...	baru terdaftar baru kali di gunakan ko sudh di...	[baru, terdaftar, baru, kali, di, gunakan, ko,...	[terdaftar, kali, ko, sudh, suruh, update]	daftar kali ko sudh suruh update
2	Kurang efisien	kurang efisien	[kurang, efisien]	[efisien]	efisien
3	Aplikasi sampah lemot	sampah lemot	[sampah, lemot]	[sampah, lemot]	sampah lot
4	Aplikasi yg seharusnya digunakan di seluruh In...	yg seharusnya digunakan di seluruh indonesia ...	[yg, seharusnya, digunakan, di, seluruh, indon...	[indonesia, metode, pembayaran, via, linkajg, ...	indonesia metode bayar via linkajg monopoli bi...

**Gambar 2.** Hasil *PreProcessing* Data

### 2.2.4 Analisis Sentimen

Pada tahap ini dilakukan proses analisis sentimen dari data ulasan yang didapatkan melalui aplikasi MyPertamina. Analisis sentimen digunakan untuk mengidentifikasi opini *user* terhadap layanan dan fitur yang didapatkan melalui aplikasi MyPertamina [12]. *Output* dari analisis sentimen ini berupa klasifikasi ulasan negatif dan positif yang divisualisasikan ke dalam bentuk *bar chart* dan *WordCloud*. *WordCloud* merupakan sebuah sistem visualisasi data yang menampilkan kata-kata dengan penekanan pada frekuensi kemunculan kata yang terkait dengan data yang digunakan [13]. *WordCloud* melakukan visualisasi grafis yang dihasilkan menggunakan *web-oriented tools* [14]. Semakin sering kata tersebut muncul, maka ukuran font sebuah kata akan semakin besar. Melalui visualisasi *WordCloud*, proses penentuan topik yang paling sering dibicarakan atau sekedar kata yang paling sering muncul dapat diketahui [15]. Melalui *bar chart* dapat diketahui informasi terkait sebaran jumlah dari kata-kata yang sering digunakan dalam memberikan ulasan negatif maupun positif. Lalu, dengan menggunakan *WordCloud* akan terlihat kata-kata yang paling umum digunakan dalam memberikan ulasan positif maupun negatif.

### 2.2.5 Topic Modelling menggunakan LDA

Pemodelan LDA merupakan metode statistika yang menerapkan dasar-dasar dalam pencarian topik dalam kumpulan dokumen-dokumen dan dapat mengatasi permasalahan *overfitting* pada *Probabilistic Latent Semantic Analysis (PLSA)*. *Topic Modelling* termasuk dalam *Natural Language Processing (NLP)* yang bekerja untuk mempelajari topik dokumen,

menganalisis, dan mengelompokkannya menurut pola dan tema secara umum yang muncul pada seluruh dokumen [16]. *Topic Modelling* juga mencari makna tersembunyi yang ada di dalamnya agar dapat menemukan informasi yang berguna [11]. LDA digunakan untuk melabeli otomatis pada 10.000 data ulasan aplikasi MyPertamina. Tahapan dalam pemodelan LDA berdasarkan persamaan (1) adalah sebagai berikut.

- Melakukan pemberian label topik setiap kata (1 sampai i) pada setiap ulasan 1 sampai j mengikuti pola distribusi multinomial (Dirichlet distribution). Secara probabilitas, untuk setiap kata (i), ada 1 sampai (i) kata untuk suatu topik (k).
- Untuk setiap kata (i) dalam ulasan (j) dan topik (k), hitung proporsi kata-kata dalam dokumen yang dimasukkan pada topik tersebut.
- Melakukan proses resampling dengan memberi topik baru pada kata berdasarkan hasil perkalian  $p(k | j)$  dengan  $p(i | k)$ .
- Melakukan proses pemberian topik baru ke kata sampai iterasi yang telah ditentukan, sampai pada kondisi tidak ada perubahan topik pada kata (steady state).

Dengan menggunakan rumus persamaan 1 berikut :

$$\rho(\theta|\alpha) = \frac{r(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k r(\alpha_i)} \prod_{i=1}^k \theta_i^{\alpha_i - 1} \quad (1)$$

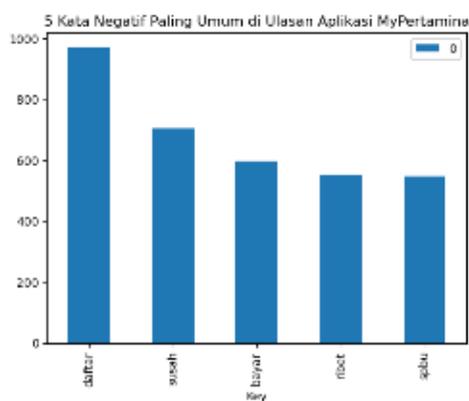
### 2.2.6 Analisis dan Evaluasi

Pada tahap analisis dan evaluasi, yaitu analisis dengan mencatat konsep hasil pengelompokan mata pelajaran yang dibentuk. Skor dari kelompok yang diperoleh kemudian dianalisis untuk memberi makna pada data. Pada tahap evaluasi penelitian ini yaitu mencermati persebaran topik yang ada berdasarkan kluster masing-masing topik.

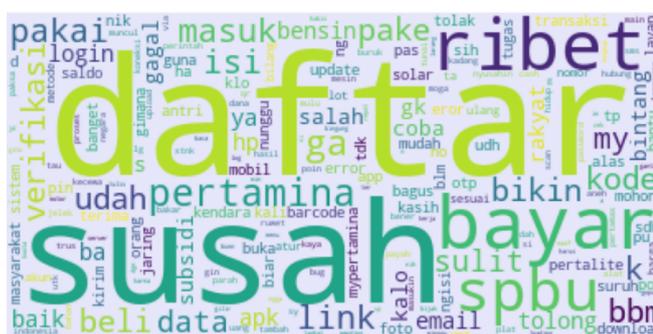
## III. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dipaparkan terkait hasil dari penelitian yang dilakukan disertai dengan analisis dan pembahasan dari hasil tersebut. Pertama, akan dipaparkan hasil analisis sentiment dalam bentuk visualisasi *bar chart* dan *WorldCloud*. Selanjutnya, dilakukan pemamparan hasil *topic modelling* yang dilakukan menggunakan metode LDA serta analisis dan evaluasi dari hasil yang didapatkan.

### 3.1 Most Common Word dan WordCloud



Gambar 3a.

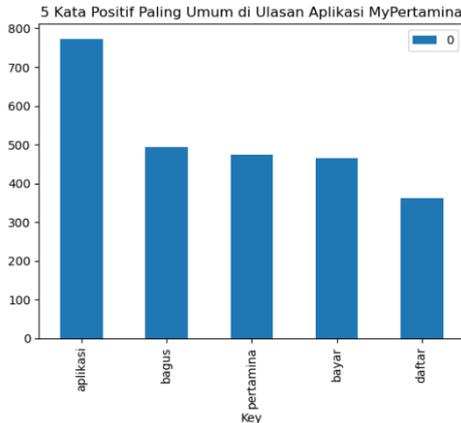


Gambar 3b.

Gambar 3 merupakan Kata Negatif : (3a) *Bar Chart* Kata Negatif (3b) *WordCloud* Kata Negatif

Berdasarkan Gambar 3. (a) *Bar Chart* dan (b) *WordCloud* 5 kata negatif ulasan MyPertamina dapat diketahui bahwa kata negatif yang didapatkan dari rating 1-2 yang paling sering muncul yaitu : daftar, susah, bayar, ribet, dan spbu. Kata daftar disebutkan 970 kali, susah 700 kali, bayar 600 kali, ribet 580

kali, dan ribet 570 kali. Dimana rata-rata memberikan ulasan negati bahwa aplikasi MyPertamina sulit dibagikan pendaftaran dan pembayaran yang ribet.

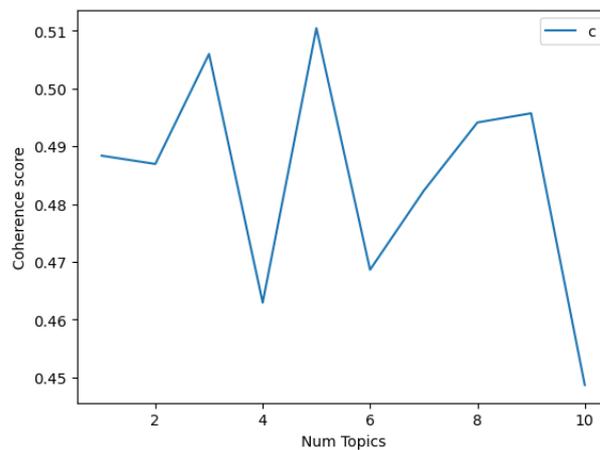

**Gambar 4a.**

**Gambar 4b.**
**Gambar 4.** Kata Positif : (4a) Bar Chart Kata Positif (4b) WordCloud Kata Positif

Berdasarkan Gambar 4. (a) Bar Chart dan (b) WordCloud 5 kata positif dapat diketahui bahwa kata positif yang didapatkan dari rating 3-5 yang sering muncul yaitu: aplikasi, bagus, pertamina, bayar dan daftar. Kata aplikasi disebutkan 780 kali, bagus 490 kali, pertamina 480 kali, bayar 475 kali, dan daftar 400 kali. Dimana rata-rata memberikan ulasan positif bahwa aplikasi MyPertamina sudah bagus aplikasi, proses pembayaran dan daftarnya.

### 3.2 Pemodelan Topik LDA

Pemodelan topik LDA pada penelitian ini digunakan untuk melihat topic mana yang sering dibahas dalam penelitian yang di uji. Biasanya dapat diketahui nilai koherensinya, sehingga dapat disimpulkan bahwa topic yang memiliki nilai koherensi tertinggi merupakan topic yang sering muncul pada ulasan aplikasi MyPertamina.


**Gambar 5.** Grafik garis nilai koherensi

Pada Gambar 5, didapatkan grafik garis nilai koherensi yang mana menggambarkan dari tabel 1 dimana setiap topik yang dihasilkan menunjukkan bahwa nilai kohorensi setiap topik mempunyai nilai yang berbeda. Rincian nilai koherensi dapat dilihat dari rincian nilai koherensi untuk setiap kata yang membentuk term-term topik.

**Tabel 1.** Nilai Koherence pada setiap topik

Topik	Nilai Koherence
T1	0.488362
T2	0.486934
T3	0.505951
T4	0.462964
T5	0.51042
T6	0.468642
T7	0.482297
T8	0.494109
T9	0.495702
T10	0.448719

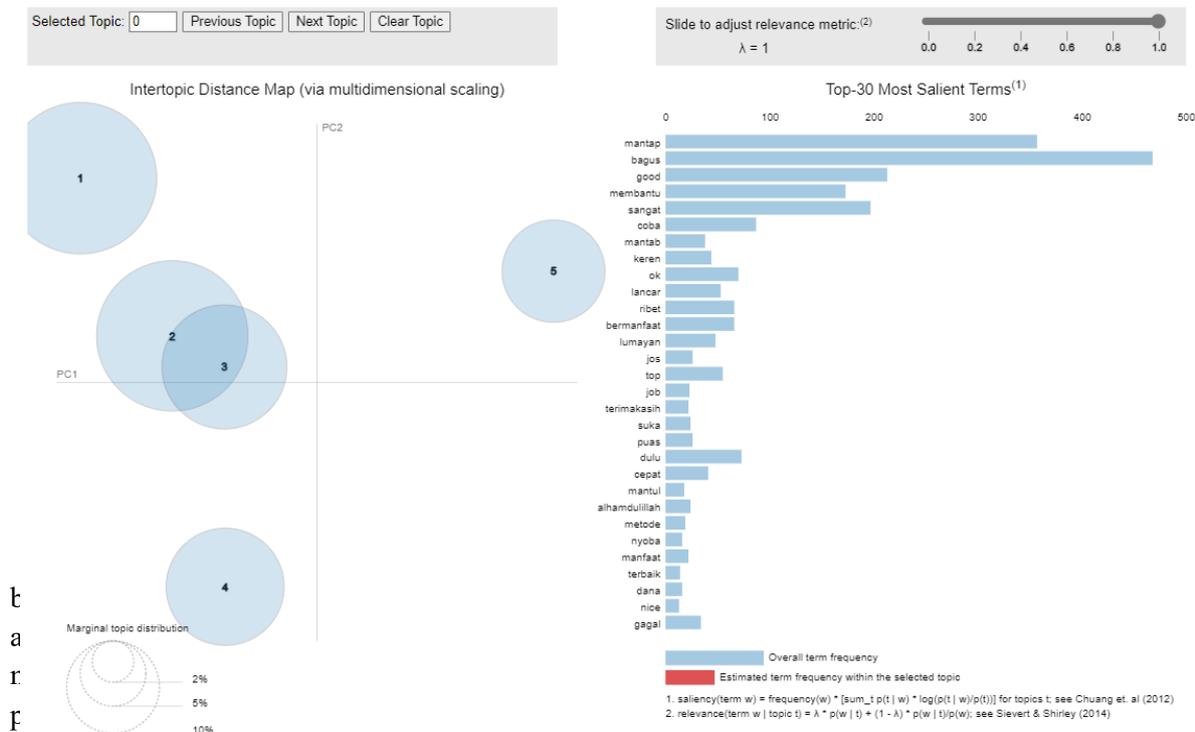
Berdasarkan Tabel 1. tersebut dapat diketahui bahwa topik yang memiliki nilai koherensi tertinggi adalah topik ke-5 dengan nilai 0.51402. Nilai koherensi yang tinggi menunjukkan topik tersebut mempunyai mempunyai peluang paling tinggi untuk muncul dalam ulasan aplikasi MyPertamina. Topik dengan nilai koherensi tertinggi tersebut akan dipilih untuk dilakukan pemodelan topik dipilih 10 topik karena grafik nilai koherensi setelah 10 mengalami penurunan nilai koherence yang cukup signifikan, serta didapatkan bahwa nilai coherence tertinggi 10 topik memiliki nilai yang tidak jauh berbeda. Berikut adalah hasil dari pemodelan topiknya.

**Tabel 2.** Hasil Nilai Koherence pada setiap topik

Topik	Nilai Koherence
T1	0.068*”good” + 0.012*”mantab” + 0.011*”ribet” + 0.008*”jos” +0.006*”terimakasih” + 0.068*”pembayaran” + 0.004*”mantul” + 0.004*”sangat” + 0.004*”metode” + 0.004*”dana”
T2	0.008*”lancar” + 0.007*”daftar” + 0.006*”saya” + 0.006*”bisa” +0.006*”susah” + 0.006*”di” + 0.006*”belum” + 0.005*”nya” + 0.005*”cepat” + 0.004*”bbm” + 0.008*”lancar” + 0.0057*”daftar”
T3	0.006*”saya” + 0.006*”bisa” + 0.006*”susah” + 0.006*”di” +0.006*”belum” + 0.005*”nya” + 0.005*”cepat” + 0.004*”bbm”
T4	0.064*”bagus” + 0.024*”membantu” + 0.023*”sangat” + 0.010*”ok” +0.009*”bermanfaat” + 0.008*”top” + 0.007*”pertamina” + 0.007*”semoga” + 0.006*”bisa”
T5	0.086*”mantap” + 0.010*”keren” + 0.006*”pertamina” + 0.005*”job” +0.005*”suka” + 0.005*”di” + 0.004*”nya” + 0.004*”bisa” + 0.004*”spbu + 0.003*”terbaik”

Tabel 2. Merupakan hasil pemodelan topik yang dapat dilihat dengan mencari topik yang dominan atau banyak muncul dengan TF-IDF dengan probabilitas kemunculan topik katanya yang dominan

dalam ulasan MyPertamina. Seperti dalam topik 1 yang memiliki isi goog, mantab, ribet, jos, terimakasih, pembayaran, mantul, sangat, metode, dan dan. Semakin tinggi nilai bobot maka mencerminkan semakin penting kata tersebut sehingga topik 1 good merupakan bobot paling penting begitu pula dengan topik lainnya.



tersebut. Kata mantap dan bagus merupakan kata yang paling penting.

**Tabel 3.** Kemunculan term-term topik

Topik	Term
Topik 1	Bantu, lumayan, bbm, beli, subsidi, bensin, barcode, pom, pake
Topik 2	Bayar, manfaat, top, link, mypertamina, linkaja, eror, mudah, metode, baik
Topik 3	Good, ribet, keren, download, banget, cepat, pertalite, guna, daftar
Topik 4	moga pertamina mudah layan mantab spbu sulit pakai transaksi mypertamina
Topik 5	Bagus, mantap, coba, daftar, susah, masuk, update, bintang, kode, pertamina

Berdasarkan implementasi topic modelling LDA yang dibuat terdapat 5 topik yang diklusterisasi. Peneliti menganalisis hasil pemodelan LDA 5 topik yang terbentuk dibagi menjadi 2 topik yang dominan diteliti yaitu tentang topik umum dan Ulasan dapat dilihat pada Tabel 4. Pada Tabel 4 topik umum yang diteliti dari ulasan aplikasi MyPertamina yang paling banyak membahas terkait pendaftara, pembayaran, dan aplikasi dengan ulasan secara keseluruhan membicarakan aplikasinya lumayan, mudah, ribet, sulit, cepat, susah, mantap, bantu, dan manfaat.

**Tabel 4.** Hasil Topik Dominan

Topik Umum	Ulasan
------------	--------



	n
Pendaftaran, Pembayaran, Aplikasi.	Lumayan, Mudah, Ribet, Sulit, Cepat, Susah, Mantap, Bantu, Manfaat.

**IV. KESIMPULAN**

Penelitian ini dilakukan untuk melihat perspektif opini pelanggan pada aplikasi MyPertamina. Hasil yang didapatkan bahwa Most Common Word pada ulasan negatif didapatkan bahwa aplikasi MyPertamina sulit dibagian daftarnya, serta sulit juga dibagian pembayaran, sedangkan pada ulasan positif didapatkan bahwa aplikasi MyPertamina sudah bagus aplikasi, proses pembayaran dan daftarnya. Wordcloud didapatkan kesimpulan bahwa sebagian besar masyarakat yang memberikan ulasan negatif mengalami kesulitan melakukan pendaftaran dan pembayaran pada aplikasi MyPertamina sedangkan yang memberikan ulasan positif beranggapan bahwa aplikasi MyPertamina merupakan aplikasi yang bagus dan memudahkan mereka dalam melakukan pembayaran pada SPBU.

Pada pemodelan LDA didapatkan nilai koherence antara 10 topik memiliki nilai yang tidak jauh berbeda, dengan nilai koherence tertinggi pada topic ke-5 dengan nilai 0.51402, sehingga digunakan num\_topik sebanyak 5 untuk membedakan topik-topik yang sering dibahas di ulasan aplikasi dan isi term dari topic ke-5 yaitu Bagus, mantap, coba, daftar, susah, masuk, update, bintang, kode, pertamina. Berdasarkan hasil visualisasi pyLDAvis kata yang paling banya adalah bagus dan mantap yang berarti aplikasi ini banyak memiliki ulasan positif. Sehingga dapat disimpulkan bahwa aplikasi MyPertamina memiliki ulasan yang cukup baik serta membantu dalam mempermudah penyaluran BBM Subsidi. Akan tetapi, terdapat beberapa ulasan yang mungkin perlu ditingkatkan dari aplikasi MyPertamina dari segi pendaftaran yang susah atau ribet dan pembayaran yang hanya memakai satu e-wallet saja yaitu LinkAja. Penelitian yang dilakukan diharapkan dapat menjadi salah satu saran untuk memperbaiki aplikasi MyPertamina menjadi lebih baik dengan memberikan solusi di beberapa pembayaran serta mempermudah dalam proses penggunaan dan pembayaran.

**UCAPAN TERIMA KASIH**

Ucapan terima kasih disampaikan kepada mata kuliah bahasa Indonesia yang diselenggarakan oleh UPN “Veteran” Jawa Timur sehingga artikel ini bisa ditulis dengan baik dan benar.

**REFERENSI**

- [1] S. Olivia, J. Gibson, and R. Nasrudin, “Indonesia in the time of covid-19,” *Bulletin of Indonesian Economic Studies*, vol. 56, no. 2, pp. 143–174, 2020. doi:10.1080/00074918.2020.1798581.
- [2] World Bank Group, “Indonesia Economic Prospects (IEP), June 2022: Financial deepening for stronger growth and Sustainable Recovery,” World Bank, <https://www.worldbank.org/en/country/indonesia/publication/indonesia-economic-prospects-iep-june-2022-financial-deepening-for-stronger-growth-and-sustainable-recovery> (accessed May 12, 2023).
- [3] MyPertamina. About Us. Retrieved December 10, 2022, from <https://mypertamina.id/about-us>.
- [4] Mustamu, D. D., & Putri, Y. R. Pengaruh Promosi Melalui Aplikasi Mypertamina Terhadap Keputusan Pembelian Bahan Bakar Pertamina Di Masyarakat Kota Bandung. *eProceedings of Management*, 6(2), 2019.
- [5] D. S. Pranav, D. Punj, T. Dubey, and P. Chawla, “Data Mining in cloud computing,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021. doi:10.1109/iccmc51019.2021.9418489
- [6] A. M. Hemeida et al., “Implementation of nature-inspired optimization algorithms in some data mining tasks,” *Ain Shams Engineering Journal*, vol. 11, no. 2, pp. 309–318, 2020. doi:10.1016/j.asej.2019.10.003
- [7] Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naïve Bayes Classifier. *Prosiding SNST Fakultas Teknik*, 1(1), 2016. doi:10.36499/psnst.v1i1.1506.
- [8] Nugraha, F. A., Harani, N. H., & Habibi, R. Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning. *Kreatif*, 2020.



- [9] M. O. Ajinaja, A. O. Adetunmbi, C. C. Ugwu, and O. S. Popoola, “Semantic similarity measure for topic modeling using latent Dirichlet allocation and collapsed Gibbs Sampling,” *Iran Journal of Computer Science*, vol. 6, no. 1, pp. 81–94, 2022. doi:10.1007/s42044-022-00124-7
- [10] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, no. 4/5, pp. 993–1022, 2003.
- [11] Smatana, Miroslav, and Peter Butka, "TopicAE: a topic modeling autoencoder." *Acta Polytechnica Hungarica*, vol. 16, no. 4, pp. 67-86, 2019.
- [12] A. Ligthart, C. Catal, and B. Tekinerdogan, “Systematic reviews in sentiment analysis: A Tertiary study,” *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4997–5053, 2021. doi:10.1007/s10462-021-09973-3
- [13] Qeis, M. I. Aplikasi wordcloud sebagai alat bantu analisis wacana. In *International Conference on Language, Culture, and Society-ICLCS LIPI*, 2015.
- [14] K. Padmanandam, S. P. Bheri, L. Vegesna, and K. Sruthi, “A speech recognized dynamic word cloud visualization for text summarization,” *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021. doi:10.1109/iciict50816.2021.9358693.
- [15] [1] R. Atenstaedt, “Word cloud analysis of the bjgp,” *British Journal of General Practice*, vol. 62, no. 596, pp. 148–148, 2012. doi:10.3399/bjgp12x630142
- [16] J. Lande, A. Pillay, and R. Chandra, *Deep learning for COVID-19 topic modelling via Twitter: Alpha, Delta and Omicron*, 2023.
- [17] Syahrums, S., & Salim, S., “Metodologi Penelitian Kuantitatif”, 2014.



# Sistem Rekomendasi Pencarian Indekos di Surabaya Menggunakan Random Forest

Theresa Agnes Virnauli Sinaga<sup>1</sup>, Lucia Bellanie Debra<sup>2</sup>, Jasmine Taj Ariva<sup>3</sup>, Devi Rizky Aditya<sup>4</sup>, Vivia Faustine Gunawan<sup>5</sup>, Maryamah Maryamah<sup>6</sup>

<sup>1,2,3,4,5</sup> *Teknologi Sains Data, Universitas Airlangga*

<sup>1</sup> [theresa.agnes.virnauli-2021@ftmm.unair.ac.id](mailto:theresa.agnes.virnauli-2021@ftmm.unair.ac.id)

<sup>2</sup> [lucia.bellanie.debra-2021@ftmm.unair.ac.id](mailto:lucia.bellanie.debra-2021@ftmm.unair.ac.id)

<sup>3</sup> [jasmine.taj.ariva-2021@ftmm.unair.ac.id](mailto:jasmine.taj.ariva-2021@ftmm.unair.ac.id)

<sup>4</sup> [devi.rizky.aditya-2021@ftmm.unair.ac.id](mailto:devi.rizky.aditya-2021@ftmm.unair.ac.id)

<sup>5</sup> [vivia.faustine.gunawan-2021@ftmm.unair.ac.id](mailto:vivia.faustine.gunawan-2021@ftmm.unair.ac.id)

<sup>6</sup> [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

Corresponding author email: [maryamah@ftmm.unair.ac.id](mailto:maryamah@ftmm.unair.ac.id)

**Abstract:** Looking for a boarding house is common for college students as a temporary place to live when they are studying outside their area of residence. Obtaining information regarding boarding prices that meet the criteria desired by students is quite difficult. In this paper, we proposed a recommendation system for finding boarding houses according to the criteria desired by students using the random forest method. This system can help students get boarding prices that match the boarding criteria they want, especially in the Surabaya area. The research method starts with data collection, preprocessing, and model training using the random forest. Based on the experimental results using the Decision Tree and Support Vector Machine (SVM) comparison method, the proposed method has the highest accuracy rate with a value of 78.55% and an error rate of 305887.80°. This recommendation system for prediction boarding house can help and make it easier for students to find boarding houses that match the criteria they want.

**Keywords:** Recommendation systems, boarding house, Machine Learning, random forest

**Abstrak:** Mencari tempat kos merupakan hal yang umum bagi mahasiswa sebagai tempat tinggal sementara ketika mereka menempuh masa pendidikan di luar daerah tempat tinggal mereka. Mendapatkan informasi terkait harga kos yang sesuai kriteria yang diinginkan oleh mahasiswa adalah hal yang cukup sulit. Oleh karena itu, kami mengusulkan sistem rekomendasi pencarian rumah kos sesuai kriteria yang diinginkan mahasiswa menggunakan metode *random forest*. Sistem ini dapat membantu mahasiswa mendapatkan harga kos yang sesuai dengan kriteria kos yang mereka inginkan khususnya di wilayah Surabaya. Metode penelitian dimulai dari pengumpulan data, preprocessing, pembuatan model menggunakan algoritma *random forest*. Berdasarkan hasil eksperimen dengan metode perbandingan *Decision Tree* dan *Support Vector Machine* (SVM), metode yang diusulkan memiliki tingkat akurasi paling tinggi dengan nilai sebesar 78.55% dan tingkat error sebesar 305887.80°. Prediksi harga ini dapat membantu dan mempermudah mahasiswa dalam pencarian kos yang sesuai dengan kriteria kos yang mereka inginkan.

**Kata kunci:** Sistem rekomendasi, Pencarian kos, *Machine Learning*, *random forest*

## I. PENDAHULUAN

Menurut Badan Pusat Statistik (BPS) Provinsi Jawa Timur Kota Surabaya pada tahun 2020, jumlah perguruan tinggi terhitung sebanyak 76 dengan total mahasiswa sejumlah 257.630 orang [1]. Berdasarkan data yang diperoleh dari survei pribadi peneliti terhadap responden yang berstatus sebagai mahasiswa di wilayah Surabaya, menunjukkan bahwa dari 54 responden hanya satu responden yang tidak memilih kos sebagai opsi tempat tinggal selama menempuh pendidikan. Dengan menyesuaikan kebutuhan mahasiswa yang membutuhkan tempat tinggal, keberadaan kos tentu memudahkan para mahasiswa khususnya dari luar daerah tempat mereka menuntut ilmu di lembaga pendidikan terkait. Kos atau *boarding house* didefinisikan sebagai tempat tinggal sementara yang disewakan oleh pemilik atau pengelola dengan harga terjangkau kepada mahasiswa atau pekerja. Kos umumnya terdiri dari beberapa kamar yang disewakan secara individu dengan dilengkapi fasilitas dasar seperti kamar mandi,



tempat tidur, lemari, meja, dan kursi. Kos umumnya juga menyediakan fasilitas tambahan seperti dapur bersama, ruang tamu, atau akses internet [2].

Ada banyak opsi hunian yang dapat dipertimbangkan oleh mahasiswa diantaranya rumah kontrakan, kos, rumah susun, dan masih banyak lagi yang disesuaikan dengan kenyamanan serta kemampuan dari calon mahasiswa. Namun, berdasarkan survei peneliti terhadap mahasiswa Surabaya menunjukkan bahwa mereka cenderung memilih kos sebagai hunian sementara. Hal ini didasari oleh beberapa hal diantaranya jarak kos ke kampus yang dekat, fasilitas yang memadai berupa ada tidaknya kamar mandi dalam, ada tidaknya biaya administrasi tambahan, luas kamar, ketersediaan kasur, dan ketersediaan luas parkir.

Dalam membantu mahasiswa dalam melakukan pemilihan kos terdapat beberapa *expert system berupa recommendation system* melalui penelitian yang telah dikembangkan. Salah satunya yaitu penelitian mengenai *web* sebagai sistem rekomendasi kos di Bali menggunakan metode *naive bayes* yang ditambahkan fitur *google map*. *Naive bayes* digunakan dalam pengembangan sistem rekomendasi ini karena dinilai memiliki akurasi yang tinggi [3]. Hasil diaplikasikan pada sistem dengan variabel input harga, jarak, kelengkapan fasilitas, yang kemudian menampilkan *output* rekomendasi kos berupa gambar kos, nama kos, dan harga. Selain itu, terdapat pula algoritma *Simple Additive Weighting (SAW)* pada penelitian rekomendasi tempat kos di Pringsewu dengan diterapkannya pembuatan *DSS (Decision Support System)*. Metode tersebut berjalan berdasarkan penambahan bobot dari rating kinerja tiap alternatif pada atribut, serta menggunakan *decision matrix* normalisasi ke skala yang dapat dibandingkan dengan peringkat alternatif [1]. Hasil pengaplikasian disesuaikan dengan kriteria yang diinginkan sebagai *input*, dimana *output* berupa daftar kos yang layak.

Meskipun *recommendation system* kos telah dikembangkan, tetapi masih terdapat hal di dalamnya yang perlu ditingkatkan. Pertama, sistem rekomendasi belum memberikan prediksi harga terbaik berdasarkan fasilitas kos yang ditawarkan. Sehingga mahasiswa masih mengalami kesulitan dalam memperkirakan budget yang sesuai terhadap kriteria kos idaman yang mereka tetapkan. Mahasiswa butuh pemahaman pola harga kos dengan faktor penentunya agar dapat mencari tempat tinggal, berupa kos-kosan yang menarik sesuai kriteria yang dibutuhkan. Sehingga, diperlukannya sistem prediksi berbasis *machine learning* yang dapat membantu para mahasiswa memilih kos beserta fasilitas yang diperoleh, dimana penentuan tersebut berdasarkan informasi yang diberikan [4]. Selain itu, belum terdapat *recommendation system* yang secara khusus dikembangkan untuk mahasiswa di wilayah Surabaya.

Dari beberapa permasalahan tersebut, kami mengusulkan sistem rekomendasi pencarian rumah kos sesuai kriteria yang diinginkan mahasiswa menggunakan metode *machine learning random forest*. Sistem ini dapat membantu mahasiswa mendapatkan harga kos yang sesuai dengan kriteria kos yang mereka inginkan khususnya di wilayah Surabaya. Metode penelitian dimulai dari pengumpulan data, preprocessing, pembuatan model menggunakan algoritma random forest. Data yang digunakan adalah kombinasi data penelitian pribadi oleh peneliti dan *scraping* data dari website mamikos. Dengan penggunaan metode *machine learning* mampu menghasilkan akurasi yang lebih akurat dibandingkan beberapa algoritma lain.

## II. METODE PENELITIAN

Metode penelitian yang digunakan dalam penulisan karya ilmiah ini dilandasi oleh beberapa referensi dari penelitian sebelumnya. Penelitian yang dikaji sebagai tinjauan pustaka adalah penelitian yang meneliti biaya sewa tempat tinggal dan aplikasi *expert system* berbasis *machine learning* sebagai pengambilan keputusan. *Expert system* merupakan sebuah sistem komputer untuk mengolah informasi yang ada menjadi sebuah penentu atau acuan dalam penentuan keputusan [5]. Penggunaan *expert system* dapat digunakan pada beberapa bidang seperti kesehatan untuk menganalisa diagnosis, pertanian untuk mengklasifikasikan hama, dan properti untuk menentukan harga sewa. Berdasarkan penelitian



independen yang dilakukan untuk memprediksi harga sewa rumah di Tanzania, Uganda dan Malawi pada tahun 2021 menggunakan *machine learning* memberikan hasil yang lebih baik dibandingkan menggunakan metode regresi Ordinary Least Square (OLS). Machine learning memperhitungkan dimensional struktur data dan korelasi antar variabel [6].

## 2.1 Teknik Pengambilan Data

Data yang digunakan pada penulisan karya ilmiah ini diambil dari dua teknik, yaitu data yang didapatkan melalui *web scraping* pada laman ‘mamikos.com’ sebagai data sekunder dan data yang didapatkan melalui survei sebagai data primer. Survei dilakukan dalam lingkup Universitas Airlangga, dimana mahasiswa angkatan 2021 prodi Teknologi Sains Data menjadi responden dari kuesioner yang dibagikan.

## 2.2 Data Pre-Processing

Data yang didapatkan dari kedua teknik yang telah disebutkan sebelumnya kemudian dibersihkan dan digabungkan. Namun, karena sumber data yang berbeda sehingga perlu dikonversi satuannya untuk menyelaraskan kedua *dataset*. Setelah data diselaraskan, data kemudian dijadikan satu untuk menjadi data bersih. Data yang telah digabungkan tersebut kemudian dilakukan *categorical encoding variable*, dimana teknik yang digunakan adalah *dummy encoding* atau mengubah variabel kategorik menjadi variabel *binary* atau dikenal sebagai variabel *dummy* [7]. Teknik ini dilakukan terhadap seluruh variabel independen agar dapat dimasukkan ke dalam model pada langkah selanjutnya.

## 2.3 Klasifikasi

Data yang telah dibersihkan kemudian dilatih menggunakan teknik *random forest classifier* guna memberikan *feedback* kepada pengguna apakah harga dan fasilitas yang ditawarkan kos sudah optimal atau diluar harga optimal. Random Forest merupakan sebuah algoritma machine learning yang termasuk dalam ensemble learning. Algoritma ini membangun beberapa (lebih dari dua) pohon keputusan secara acak, dengan setiap pohon keputusan menggunakan sampel acak dari data pelatihan dan subset acak dari atribut. Random Forest memiliki beberapa kelebihan antara lain dalam menangani data kompleks dan besar, toleransi terhadap overfitting, kemampuan mengatasi variabel tidak penting, dan memberikan perkiraan keakuratan model yang dihasilkan. Random Forest menghitung rata-rata dari hasil pada berbagai *decision tree* yang diaplikasikan pada berbagai subset set data untuk meningkatkan akurasi prediksi [4]. Data ini kemudian dibentuk menjadi bentuk model yang nantinya akan menerima nilai yang diberikan pengguna dan program akan mengembalikan dalam bentuk harga rentang untuk spesifikasi kos yang dimaksud. Random forest menggunakan rumus entropi sebagai penentu tingkat ketidakmurnian atribut dan nilai *information gain* seperti pada Persamaan 1 dan 2.

$$Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Pada Persamaan (1),  $Y$  merupakan variabel acak dari suatu dataset,  $c$  merupakan nilai kelas atau label yang mungkin diambil oleh variabel acak  $Y$ , lalu pada  $p(c|Y)$  merupakan probabilitas dari kemunculan kelas  $c$  pada variabel acak  $Y$ . Selanjutnya, dilakukan penjumlahan untuk semua nilai  $c$  yang mungkin terhadap nilai  $p(c|Y)$  yang dikalikan dengan  $\log_2$  dari nilai  $p(c|Y)$  dan dikalikan lagi dengan negatif satu.

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in values} \left| \frac{Y_v}{Y_a} \right| Entropy(Y_v) \quad (2)$$

Pada Persamaan (2),  $values(a)$  merupakan semua nilai yang mungkin dalam himpunan kasus  $a$ . Kemudian  $Y_v$  adalah subkelas dari  $Y$  dengan kelas  $v$  yang berhubungan dengan kelas  $a$ , sedangkan  $Y_a$  adalah semua nilai yang sesuai dengan  $a$ .

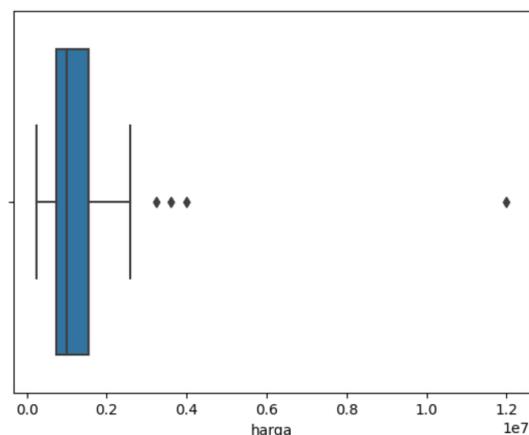
### III. HASIL DAN PEMBAHASAN

Pada penelitian ini, data yang diperoleh akan diuji menggunakan tiga metode, yaitu *decision tree*, *random forest*, dan *super vector machine*. Untuk menguji metode mana yang paling baik, penelitian ini menggunakan skor uji akurasi dan *mean average error* (MAE). Metode yang paling baik adalah metode yang memiliki skor akurasi tertinggi dan nilai MAE terkecil. Sebelum membangun model, akan dilakukan analisis statistika deskriptif dari dataset dengan variabel-variabel bernama “harga”, “leb-ih\_satu\_orang”, “km\_dalam”, “wifi”, “listrik”, “ac”, “tipe”, yang mana ditampilkan hasil perhitungan rata-rata, standar deviasi, nilai minimum, dan nilai maksimum setelah data yang bersifat kategorik diubah menjadi *dummy variable*. Hasil tampilan analisis statistika deskriptif tersebut dapat pada gambar 1. Selanjutnya, mengecek nilai kosong atau NA guna untuk memastikan bahwasanya di dalam dataset tidak mengandung nilai yang kosong, dimana variabel dengan NA akan langsung dihapus. Dilakukannya modifikasi langsung pada dataset yang telah berbentuk data frame dengan menghapus baris duplikat.

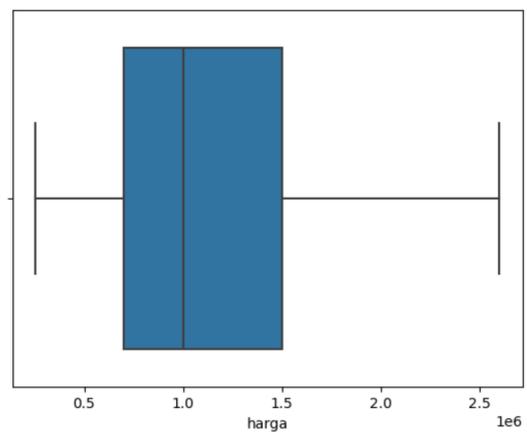
	harga	lebih_satu_orang	km_dalam	wifi	listrik	ac	tipe
count	1.010000e+02	101.000000	101.000000	101.000000	101.000000	101.000000	101.000000
mean	1.317495e+06	0.495050	0.485149	0.693069	0.554455	0.554455	1.297030
std	1.273151e+06	0.502469	0.502272	0.463521	0.499505	0.499505	0.806778
min	2.550000e+05	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.500000e+05	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
50%	1.000000e+06	0.000000	0.000000	1.000000	1.000000	1.000000	2.000000
75%	1.550000e+06	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000
max	1.200000e+07	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000

**Gambar 1.** Statistika Deskriptif Data

Selanjutnya untuk menentukan *outlier* yang perlu dihapus, dibuat visualisasi menggunakan *boxplot* untuk melihat harga kos yang *outlier*. Setelah nilai *outlier* diketahui, *outlier* kemudian dihapus menggunakan *lower bound* dan *upper bound* yang didapatkan dari nilai kuartil dan Interquartile Range (IQR). Perbedaan *boxplot* sebelum dan sesudah pembersihan dapat dilihat pada Gambar 2a. dan Gambar 2b. Setelah data dibersihkan, variabel x atau variabel prediktor dan variabel y atau variabel respon ditentukan, dimana variabel ‘harga’ menjadi variabel respon dan variabel lainnya menjadi variabel prediktor.



**Gambar 2a.** Variabel ‘Harga’ dengan outlier



**Gambar 2b.** Variabel ‘Harga’ setelah outlier dihapus

Perbandingan metode yang diusulkan, dibandingkan dengan dua metode *machine learning* lain yaitu *Decision Tree* dan *Support Vector Machine*. Nilai perbandingan dari ketiga metode

menggunakan hasil akurasi dan nilai *average error* pada Tabel 1. Nilai akurasi digunakan untuk melihat seberapa besar keakuratan model yang dihasilkan oleh algoritma. Pada tabel terlihat bahwa ketiga metode memiliki besaran nilai yang berbeda pada nilai akurasi Nilai akurasi tertinggi terdapat pada algoritma *Random Forest* sebesar 78,55 %. Hal ini menunjukkan bahwa keputusan yang dihasilkan oleh *Decision Tree* memberikan keakuratan yang tinggi. Setelah itu nilai akurasi tertinggi selanjutnya dimiliki oleh *Decision Tree* sebesar 78.50% dan *Super Vector Machine* sebesar 74.43%. Kemudian terdapat nilai *average error* yaitu nilai *error* yang mungkin dihasilkan dari perhitungan. Untuk itu, diperlukan nilai *error* terkecil yang dimiliki oleh *Random Forest* sebesar 305887.80°. Berdasarkan kedua indikator tersebut untuk menentukan algoritma yang dapat digunakan sebagai pengambilan keputusan model dipilih dengan nilai akurasi tertinggi dan *average error* terkecil yang dimiliki oleh *Random Forest* dengan selisih sebesar 0.05 lebih tinggi dibandingkan *Decision Tree*. Metode *Random Forest* memiliki hasil pengukuran lebih baik dalam peningkatan akurasi karena pemilihan pembangkitan simpul anak untuk setiap node dilakukan secara acak dan diakumulasikan hasil klasifikasi [7]. Melihat metode lain sebagai pembanding yaitu *Decision Tree* dan *Support Vector Machine* memiliki alur dan perhitungan yang berbeda dibandingkan dengan *Random Forest*.

**Tabel 1.** Nilai Hasil Perbandingan hasil

Algoritma	Nilai Akurasi (%)	Average Error(°)
Decision Tree	78.50	311400.48
Random Forest	78.55	305887.80
Super Vector Machine	74.43	315600.00

*Decision Tree* menggunakan struktur pohon untuk memodelkan keputusan berdasarkan atribut data yang dipunya. Adapun Setiap simpul mewakili keputusan atau tes pada atribut, dan cabang-cabang menggambarkan hasil keputusan. Pohon keputusan dapat digunakan untuk klasifikasi dan regresi, dan membantu interpretasi hubungan antara atribut dan hasil keputusan. SVM mengambil garis pemisah optimal (hyperplane) dengan margin maksimal tiap kelas data yang berbeda. SVM mampu menangani data linier maupun non-linier dengan transformasi non-linier. Kelebihan SVM adalah dalam mengoptimalkan fungsi margin maksimal, namun kurang efektif digunakan apabila dataset dalam jumlah besar.

#### IV. KESIMPULAN

Sistem rekomendasi harga kos dengan mengaplikasikan *machine learning random forest* memiliki tingkat akurasi lebih tinggi dibandingkan metode lainnya dengan akurasi sebesar 78.55% dan nilai *error* sebesar 305887.80°. Metode pembanding yang digunakan *Decision Tree* dan *Support Vector Machine*. Sistem rekomendasi yang dibangun akan membantu menyelesaikan permasalahan kebanyakan penyewa kos, dalam hal ini khususnya mahasiswa mendapatkan kos sesuai dengan kriteria yang diinginkan. Penggunaan akurasi dan nilai *average error* digunakan karena kedua indikator tersebut dinilai cukup untuk dijadikan sebagai standar pengukuran untuk mengakurasi hasil harga. Penelitian selanjutnya yang akan dilakukan adalah penambahan data kos yang lebih luas dan penerapan metode deep learning. Implementasikan lebih lanjut pada berbagai *platform*, salah satunya adalah dengan melakukan *deployment* ke dalam website juga akan dilakukan agar usulan dapat bermanfaat dan digunakan untuk keperluan yang lebih luas.

**UCAPAN TERIMA KASIH**

Ucapan terima kasih disampaikan kepada mata kuliah bahasa Indonesia yang diselenggarakan oleh UPN “Veteran” Jawa Timur sehingga artikel ini bisa ditulis dengan baik dan benar.

**REFERENSI**

- [1] BPS. (2021). Jumlah Perguruan Tinggi, Mahasiswa, dan Tenaga Pendidik (Negeri dan Swasta) di Bawah Kementerian Riset, Teknologi dan Pendidikan Tinggi Menurut Kabupaten/Kota, 2019 dan 2020. [Online]. Diakses pada 14 Mei 2023 melalui <https://jatim.bps.go.id/statictable/2021/09/06/2218/jumlah-perguruan-tinggi-mahasiswa-dan-tenaga-pendidik-negeri-dan-swasta-di-bawah-kementerian-ri-set-teknologi-dan-pendidikan-tinggi-menurut-kabupaten-kota-2019-dan-2020.html>.
- [2] H. Mukhlis, B. Ayshwarya, P. T. Nguyen, dan W. Hashim, "Boarding House Selection using SAW Method," July 2020.
- [3] S. Suryana, "Sistem Rekomendasi Tempat Kos Mahasiswa Baru dengan Metode Naïve Bayes Berbasis Web," *Jurnal Sistem Informasi Dan Komputer Terapan Indonesia (JSIKTI)*, vol. 3, no. 3, pp. 22-31, 2021. [Online]. Available: <https://doi.org/10.33173/jsikti.107>.
- [4] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, dan G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," *Procedia Computer Science*, vol. 199, pp. 806-813, 2021. [Online]. Tersedia: <https://doi.org/10.1016/j.procs.2022.01.100>.
- [5] W. S. Pramana, M. Sudarma, dan I. N. S. Kumara, "Expert system and classical probability for setting up hotel's dynamic price level: A case of four-star hotel in Bali," *International Journal of Electrical and Electronic Engineering and Telecommunications*, vol. 9, no. 2, pp. 124-131, 2020. [Online]. Tersedia: <https://doi.org/10.18178/IJEETC.9.2.124-131>.
- [6] W. T. Embaye, Y. A. Zereyesus, dan B. Chen, "Predicting the rental value of houses in household surveys in Tanzania, Uganda and Malawi: Evaluations of hedonic pricing and machine learning approaches," *PLoS ONE*, vol. 16, no. 2, pp. 1-20, February 2021. [Online]. Tersedia: <https://doi.org/10.1371/journal.pone.0244953>.
- [7] M. K. Dahouda dan I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381-114391, 2021. [Online]. Tersedia: <https://doi.org/10.1109/ACCESS.2021.3104357>.
- [8] Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>
- [9] V. W. Siburian dan I. E. Mulyana, *Prediksi Harga Ponsel Menggunakan Metode Random Forest, Prosiding Annual Research Seminar 2018*, vol. 4, no. 1, hal. 144-147, 2018.
- [10] Haifeng Wang and Dejin Hu, "Comparison of SVM and LS-SVM for Regression," 2005 International Conference on Neural Networks and Brain, Beijing, 2005, pp. 279-283, doi: 10.1109/ICNNB.2005.1614615.



# Optimasi Pengolahan Data Facemask di Teachable Machine melalui Analisis Nilai Epoch dan Learning Rate

Dimas Dzaky Daniswara<sup>1</sup>, Ikbar Athallah Taufik<sup>2</sup>, Prismahardi Aji Riyantoko<sup>3</sup>

<sup>1,2,3</sup> Program Studi Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

<sup>1</sup>[20083010006@student.upnjatim.ac.id](mailto:20083010006@student.upnjatim.ac.id)

<sup>2</sup>[20083010027@student.upnjatim.ac.id](mailto:20083010027@student.upnjatim.ac.id)

<sup>3</sup>[prismahardi.aji.ds@upnjatim.ac.id](mailto:prismahardi.aji.ds@upnjatim.ac.id)

Corresponding author email: [20083010006@student.upnjatim.ac.id](mailto:20083010006@student.upnjatim.ac.id)

**Abstract:** This study discusses the optimization of facemask data processing using Teachable Machine through the analysis of epoch and learning rate values. The purpose of this research is to improve the accuracy of recognizing the usage of masks on facial images using transfer learning techniques on Teachable Machine. The image data used in this study is the Face Mask Detection dataset obtained from Kaggle, which has been labeled with mask usage and non-usage categories. In the data processing stage, image augmentation techniques were employed to expand the sample size and avoid overfitting. Furthermore, model training was conducted with various values of epoch and learning rate to determine the optimal parameter combination for improving recognition accuracy. The results show that using 150 epochs and a learning rate of 0.001, a recognition accuracy of 99% was achieved on the dataset used. It is expected that the results of this study can contribute to the development of automatic mask usage detection applications.

**Keywords:** Teachable Machine, Machine Learning, Image Classification

**Abstrak:** Penelitian ini membahas tentang optimasi pengolahan data facemask menggunakan Teachable Machine melalui analisis nilai epoch dan learning rate. Tujuan penelitian ini adalah untuk meningkatkan akurasi pengenalan penggunaan masker pada gambar wajah menggunakan teknik transfer learning pada Teachable Machine. Data gambar yang digunakan dalam penelitian ini adalah dataset Face Mask Detection yang diperoleh dari sumber internet yaitu Kaggle dan telah dilabeli dengan kategori penggunaan masker dan tidak. Pada tahap pengolahan data, dilakukan teknik augmentasi gambar untuk memperluas jumlah sampel data dan menghindari overfitting. Selanjutnya, dilakukan pelatihan model dengan variasi nilai epoch dan learning rate untuk menentukan kombinasi parameter yang optimal untuk meningkatkan akurasi pengenalan. Hasil penelitian menunjukkan bahwa dengan menggunakan epoch sebanyak 150 dan learning rate sebesar 0.001, diperoleh akurasi pengenalan sekitar 96-100% pada dataset yang digunakan. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan aplikasi deteksi penggunaan masker secara otomatis.

**Kata kunci:** Teachable Machine, Pembelajaran Mesin, Klasifikasi Gambar

## I. PENDAHULUAN

Penggunaan masker menjadi salah satu cara yang efektif dalam mencegah penyebaran virus dan penyakit menular [1]. Oleh karena itu, pemerintah dan lembaga kesehatan di berbagai negara telah menerapkan kebijakan wajib menggunakan masker di tempat-tempat umum. Namun, masih terdapat individu yang tidak mematuhi kebijakan tersebut, yang dapat mengancam kesehatan dan keselamatan masyarakat. Pemanfaatan Machine Learning (Pembelajaran Mesin) dalam bidang pendeteksian objek telah mengalami perkembangan yang pesat dalam beberapa tahun terakhir. Kemajuan dalam algoritma, teknologi komputer, dan ketersediaan data telah memungkinkan pengembangan sistem pendeteksian objek yang semakin canggih dan akurat. Machine Learning dalam pendeteksian objek melibatkan penggunaan algoritma untuk melatih model komputasional yang mampu mengenali dan membedakan objek dalam gambar atau video. Model ini belajar secara mandiri dari data latihan dan dapat diterapkan untuk mendeteksi objek dalam konteks yang berbeda [2].

Untuk mengatasi hal tersebut, teknologi deteksi penggunaan masker dapat digunakan untuk mengidentifikasi individu yang tidak memakai masker di tempat-tempat umum. Salah satu teknologi yang dapat digunakan adalah Teachable Machine, yaitu platform pembelajaran mesin sederhana yang dikembangkan oleh Google [3]. Teachable Machine menggunakan pendekatan transfer learning yang

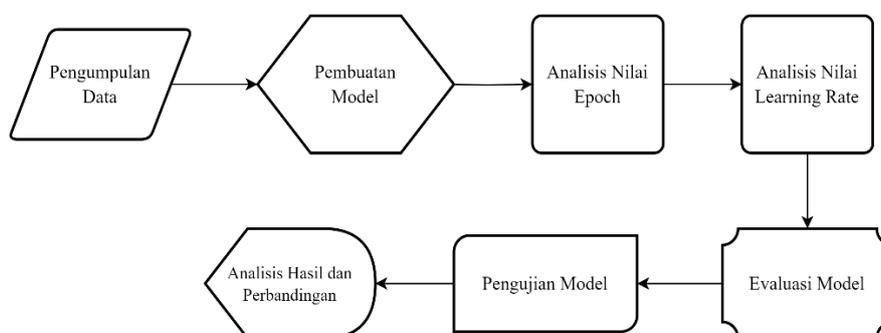
melibatkan penggunaan jaringan saraf yang telah dilatih sebelumnya. Transfer learning adalah metode dalam machine learning yang memanfaatkan kembali model yang telah dilatih sebelumnya sebagai titik awal untuk memodelkan tugas-tugas baru. Dalam transfer learning, model yang telah dilatih untuk satu tugas dapat digunakan kembali untuk tugas yang terkait sebagai langkah awal yang mengoptimalkan waktu dan sumber daya yang diperlukan dalam memodelkan tugas baru [4].

Penerapan Teachable Machine juga memberikan tingkat keakuratan, ketepatan, dan sensitivitas yang tinggi, dengan kisaran antara 97% hingga 100%. Teachable Machine yang dikembangkan oleh Google mampu mencapai tingkat akurasi uji hingga 100% melalui pembelajaran mesin. Namun, keakuratan ini dapat dipengaruhi oleh berbagai faktor, seperti kondisi pencahayaan di sekitar objek gambar, yang dapat mengakibatkan penurunan keakuratan [5]. Dalam penggunaannya, proses pelatihan model machine learning di Teachable Machine melibatkan beberapa faktor seperti jumlah epoch dan learning rate [6]. Penelitian lainnya pernah menganalisa mengenai Learning rate dan Batch size [7]. Dalam penelitian tersebut, membandingkan akurasi hasil antara beberapa learning rate dan ukuran batch size. Tujuannya adalah untuk mengevaluasi pengaruh learning rate dan ukuran batch size terhadap akurasi dan loss dalam proses deep learning, khususnya pada tahap pelatihan dan validasi.

Penelitian ini bertujuan untuk mengoptimalkan pengolahan data *facemask* di *Teachable Machine* melalui analisis nilai *epoch* dan *learning rate* [8]. Salah satu aspek utama dari *teachable machine* adalah bahwa ia sepenuhnya berbasis antarmuka pengguna grafis (GUI). Keunggulan utamanya terletak pada fleksibilitasnya yang memungkinkan pengguna untuk mengambil contoh secara langsung atau menggunakan file sebagai contoh. Model yang dibuat di dalam *teachable machine* merupakan model *tensorflow.js* yang dapat diakses di berbagai platform yang menggunakan *JavaScript*. Selain itu, *teachable machine* juga mendukung akses ke alat-alat seperti *Glitch*, *node.js*, dan lainnya. Selain itu, kemampuannya untuk mengekspor model ke berbagai format memungkinkan penggunaannya dalam platform seperti *Arduino*, *coral*, dan lain sebagainya [9]. Metode yang digunakan adalah eksperimen dengan memvariasikan nilai *epoch* dan *learning rate* pada proses pelatihan model deteksi penggunaan masker menggunakan *Teachable Machine* [10]. Selain itu, penelitian ini juga menguji performa model pada dataset *facemask* untuk mengukur kemampuan model dalam melakukan deteksi penggunaan masker pada dataset yang digunakan.

## II. METODE PENELITIAN

Beberapa langkah penelitian akan dilakukan sesuai dengan urutan yang ditunjukkan dalam diagram alur Gambar 1 berikut:



Alur penelitian yang akan dilakukan dijabarkan pada tahapan berikut:

### 1. Pengumpulan Data

Data gambar yang digunakan dalam penelitian ini adalah dataset Face Mask Detection yang diperoleh dari sumber internet yaitu Kaggle [11]. Data gambar dikategorikan

ke dalam dua kelompok yaitu gambar orang dengan menggunakan masker dan tanpa menggunakan masker.

**Tabel 1.** Dataset *Facemask*

No.	Jenis	Jumlah Dataset	Contoh Gambar
1	Menggunakan masker	2.994	
2	Tanpa menggunakan masker	2.994	

## 2. Pembuatan Model Menggunakan *Teachable Machine*

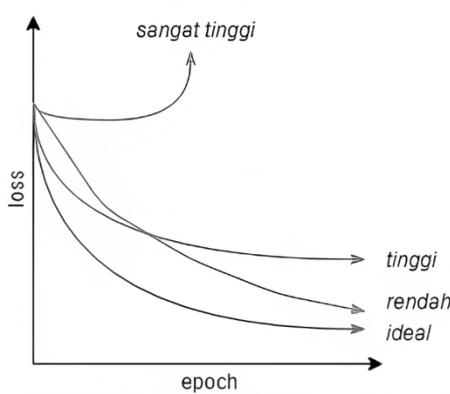
Pada tahap ini, data gambar yang telah dikumpulkan akan diolah menggunakan *Teachable Machine*. *Teachable Machine* adalah aplikasi web yang digunakan untuk membuat model machine learning dengan mudah tanpa memerlukan keterampilan pemrograman yang mendalam [3]. Model machine learning yang dibuat akan dilatih menggunakan data gambar yang telah dikategorikan sebelumnya.

## 3. Analisis Nilai *Epoch*

Pada tahap ini, dilakukan analisis terhadap nilai epoch yang digunakan dalam proses pelatihan model machine learning. *Epoch* adalah jumlah kali iterasi yang dilakukan untuk melatih model machine learning [12]. Nilai epoch yang berbeda akan mempengaruhi akurasi model machine learning yang dihasilkan. Dalam penelitian ini, dilakukan percobaan dengan menggunakan nilai *epoch* yang berbeda untuk menemukan nilai yang optimal yang dapat meningkatkan akurasi model.

## 4. Analisis Nilai *Learning Rate*

Pada tahap ini, dilakukan analisis terhadap nilai *learning rate* yang digunakan dalam proses pelatihan model *machine learning*. *Learning rate* adalah faktor yang mengatur seberapa besar bobot yang akan disesuaikan pada setiap iterasi pelatihan [13]. Nilai *learning rate* yang berbeda akan mempengaruhi kecepatan konvergensi model *machine learning*. Dalam penelitian ini, dilakukan percobaan dengan menggunakan nilai *learning rate* yang berbeda untuk menemukan nilai yang optimal yang dapat meningkatkan akurasi model.



## 5. Evaluasi Model

Setelah dilakukan pelatihan model dengan berbagai nilai *epoch* dan *learning rate*, dilakukan evaluasi terhadap model yang dihasilkan. Evaluasi dilakukan dengan menggunakan

data gambar yang tidak digunakan dalam proses pelatihan. Hasil evaluasi digunakan untuk menentukan nilai *epoch* dan *learning rate* yang optimal.

#### 6. Pengujian Model

Pada tahap ini, model machine learning yang telah dilatih dan dievaluasi akan diuji dengan menggunakan data gambar yang belum pernah dilihat sebelumnya. Pengujian dilakukan untuk mengetahui akurasi dan kinerja model machine learning yang telah dibuat.

#### 7. Analisis Hasil dan Perbandingan

Hasil pengujian dan analisis nilai *epoch* dan *learning rate* akan digunakan untuk membuat kesimpulan dan perbandingan dengan penelitian sebelumnya. Perbandingan dilakukan untuk mengetahui perbedaan dan kelebihan dari penelitian ini dengan penelitian sebelumnya dalam deteksi penggunaan masker menggunakan *Teachable Machine*.

### III. HASIL DAN PEMBAHASAN

Hasil pembahasan ini akan membahas tentang optimasi pengolahan data facemask di *Teachable Machine* melalui analisis nilai *epoch* dan *learning rate*. Dalam pengolahan data di *Teachable Machine*, nilai *epoch* dan *learning rate* sangat penting untuk menentukan kualitas dan akurasi model machine learning yang dihasilkan.

Pada penelitian ini, dilakukan beberapa eksperimen dengan mengubah nilai *epoch* dan *learning rate* untuk mengoptimalkan pengolahan data facemask di *Teachable Machine*. Eksperimen dilakukan dengan menggunakan dataset yang terdiri dari gambar orang yang menggunakan masker dan orang yang tidak menggunakan masker.

**Tabel 2.** Hasil modeling with mask

<i>Learning Rate</i>	<i>Epoch</i>			
	50	100	150	200
0,01	0,97	1	0,99	0,97
0,001	0,97	0,97	0,99	1
0,0001	0,99	0,96	1	1
0,00001	1	1	0,99	0,99
0,000001	0,97	0,99	1	0,97

Hasil pada tabel didapat dari pemrosesan model menggunakan *teachable machine*, pada tabel hasil pemrosesan klasifikasi dengan foto menggunakan masker. Masing-masing *learning rate* mulai dari 0,01 sampai 0,000001 diuji dengan *epoch* mulai dari 50 ke *epoch* 200 sehingga, diperoleh sebuah nilai akurasi yang cukup tinggi, beberapa hasil menunjukkan nilai akurasi mendekati 1 dan ada yang menunjukkan nilai akurasi 1. Nilai akurasi terendah berada di *learning rate* 0,0001 dengan *epoch* 100 yaitu sebesar 0,96.

**Tabel 3.** Hasil modeling without mask

<i>Learning Rate</i>	<i>Epoch</i>			
	50	100	150	200
0,01	1	0,97	0,99	1
0,001	0,97	1	0,99	0,99
0,0001	0,99	1	0,99	0,96
0,00001	0,95	0,99	1	0,99
0,000001	0,95	0,89	0,99	0,99

Hasil tabel selanjutnya juga didapat dari pemrosesan model menggunakan *teachable machine* yang sama seperti sebelumnya. Pada tabel terlihat masing-masing *learning rate* yaitu mulai dari 0,01 sampai dengan 0,000001, diuji dengan nilai *epoch* dari 50 sampai 200 menghasilkan, sebuah nilai



akurasi yang berbeda-beda, namun sedikit berbeda dari nilai tabel sebelumnya, nilai tabel 2. Memiliki hasil akurasi yang lebih menurun, seperti pada nilai akurasi di *learning rate* 0,000001 dengan *epoch* 100 yang mana hasil akurasinya sebesar 0,89 menunjukkan penurunan nilai yang cukup besar.

Hasil eksperimen menunjukkan bahwa nilai *epoch* dan *learning rate* mempengaruhi kualitas dan akurasi model yang dihasilkan. Nilai *epoch* yang terlalu rendah dapat mengakibatkan *underfitting*, sedangkan nilai *epoch* yang terlalu tinggi dapat mengakibatkan *overfitting*. Nilai *learning rate* yang terlalu rendah dapat memperlambat proses konvergensi, sedangkan nilai *learning rate* yang terlalu tinggi dapat menyebabkan model tidak stabil.

Berdasarkan hasil eksperimen, ditemukan bahwa nilai *epoch* yang optimal adalah 150 *epoch* dan nilai *learning rate* yang optimal adalah 0,001. Dengan menggunakan nilai *epoch* dan *learning rate* yang optimal, akurasi model dapat mencapai sekitar 96-100%, yang merupakan hasil yang cukup baik untuk deteksi penggunaan masker.

Dalam kesimpulannya, optimasi pengolahan data *facemask* di *Teachable Machine* melalui analisis nilai *epoch* dan *learning rate* dapat meningkatkan kualitas dan akurasi model *machine learning*. Nilai *epoch* dan *learning rate* yang optimal dapat membantu mencegah *overfitting*, *underfitting*, serta mempercepat proses konvergensi. Dengan menggunakan teknik optimasi ini, deteksi penggunaan masker dapat dilakukan dengan akurasi yang tinggi, sehingga dapat membantu dalam memerangi penyebaran virus Covid-19.

#### IV. KESIMPULAN

Optimasi nilai *epoch* dan *learning rate* pada pengolahan data *facemask* di *Teachable Machine* dapat meningkatkan akurasi model deteksi penggunaan masker. Pada penelitian ini, ditemukan bahwa nilai *epoch* dan *learning rate* yang optimal untuk mencapai akurasi tertinggi adalah 150 *epoch* dan *learning rate* sebesar 0,001.

Dalam mengoptimalkan nilai *epoch* dan *learning rate*, perlu dilakukan evaluasi secara berkala pada setiap iterasi pelatihan untuk mengetahui performa model pada setiap nilai *epoch* dan *learning rate*. Hal ini dapat dilakukan dengan memonitor metrik evaluasi seperti akurasi, presisi, *recall*, dan F1-score. Selain itu, penambahan data latih yang lebih banyak juga dapat membantu meningkatkan performa model.

Meskipun hasil penelitian menunjukkan peningkatan akurasi dengan menggunakan teknik optimasi nilai *epoch* dan *learning rate*, masih terdapat beberapa faktor yang dapat mempengaruhi performa model, seperti kualitas data latih dan kesesuaian algoritma yang digunakan. Oleh karena itu, penelitian lebih lanjut dapat dilakukan untuk meningkatkan performa model deteksi penggunaan masker dengan menggunakan metode-metode yang lebih canggih dan akurat.

#### UCAPAN TERIMA KASIH

#### REFERENSI

1. W. World Health Organization, “Anjuran mengenai penggunaan masker dalam konteks COVID-19,” *World Heal. Organ.*, no. April, pp. 1–17, 2020, [Online]. Available: [https://www.who.int/docs/default-source/searo/indonesia/covid19/anjuran-mengenai-penggunaan-masker-dalam-konteks-covid-19-june-20.pdf?sfvrsn=d1327a85\\_2](https://www.who.int/docs/default-source/searo/indonesia/covid19/anjuran-mengenai-penggunaan-masker-dalam-konteks-covid-19-june-20.pdf?sfvrsn=d1327a85_2)
2. M. Fasounaki, E. B. Yüce, S. Öncül, and G. Ince, “CNN-based Text-independent Automatic Speaker Identification Using Short Utterances,” *Proc. - 6th Int. Conf. Comput. Sci. Eng. UBMK 2021*, vol. 01, pp. 413–418, 2021, doi: 10.1109/UBMK52708.2021.9559031.
3. E. A. U. Malahina, R. P. Hadjon, and F. Y. Bisilisin, “Teachable Machine: Real-Time Attendance of Students Based on Open Source System,” *IJICS (International J. Informatics Comput. Sci.)*, vol. 6, no. 3, p. 140, 2022, doi: 10.30865/ijics.v6i3.4928.



4. D. Immanuel Salintohe, I. Alwiah Musdar, T. Informatika, and S. Kharisma Makassar, “Implementasi Machine Learning Untuk Mengidentifikasi Tanaman Hias Pada Aplikasi Tierra,” *Jtriste*, vol. 9, no. 1, pp. 1–15, 2022.
5. D. Agustian, P. P. G. P. Pertama, P. N. Crisnapati, and P. D. Novayanti, “Implementation of Machine Learning Using Google’s Teachable Machine Based on Android,” in *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, Oct. 2021, pp. 1–7. doi: 10.1109/ICORIS52787.2021.9649528.
6. F. N. Fajri, K. Malik, and G. Q. O. Pratamasunu, “Metode Pengumpulan Data Pada Deteksi Pakaian Hijab Syar’I Berdasarkan Citra Digital Menggunakan Teachable machine Learning,” *Justek J. Sains dan Teknol.*, vol. 5, no. 2, p. 194, 2022, doi: 10.31764/justek.v5i2.11614.
7. N. Rochmawati, H. B. Hidayati, Y. Yamasari, H. P. A. Tjahyaningtjas, W. Yustanti, and A. Prihanto, “Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 44–48, 2021, doi: 10.26740/jieet.v5n2.p44-48.
8. Z. Fitri, “Analisis Error dan Epoch dengan Pengembangan Adaptive Learning Rate dan Parameter Momentum pada Metode Backpropagation,” *J. Infomedia*, vol. 3, no. 2, Dec. 2018, doi: 10.30811/jim.v3i2.680.
9. M. P. Mathew and T. Y. Mahesh, “Object Detection Based on Teachable Machine,” *J. VLSI Des. Signal Process.*, vol. 7, no. 2, pp. 20–26, 2021, doi: 10.46610/jovdsp.2021.v07i02.003.
10. R. D. Nurfitia and G. Ariyanto, “Implementasi Deep Learning berbasis Tensorflow untuk Pengenalan Sidik Jari,” *Emit. J. Tek. Elektro*, vol. 18, no. 1, pp. 22–27, 2018, doi: 10.23917/emitor.v18i01.6236.
11. V. KUMAR, “Face Mask Detection,” *kaggle*, 2021. <https://www.kaggle.com/andrewmvd/face-mask-detection>  
<https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset> (accessed May 07, 2023)
12. A. P. Sugiarto and S. Sriyanto, “Model Pembelajaran dalam Jaringan Saraf Konvolusional untuk Klasifikasi Teks (Abstrak Penelitian Ilmu Komputer),” *Pros. Semin. Nas. ...*, pp. 95–106, 2020, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2469%0Ahttps://jurnal.darmajaya.ac.id/index.php/PSND/article/viewFile/2469/1152>
13. Sahrul, S. Hadinisa, M. Koyimatu, A. Irawan, and H. Nugroho, “Analisis Learning Rate pada Metode Transfer Learning untuk Sistem Pendeteksi Api,” *Semin. Nas. Microwave, Antena dan Propagasi*, pp. 1–4, 2018.



# Peramalan Lonjakan Kasus Harian Covid-19 Di Indonesia Dengan Model Arima

Aviolla Terza Damaliana<sup>1</sup>, Amri Muhaimin<sup>2</sup>, Prismahardi Aji Riyantoko<sup>3</sup>,

<sup>1,2,3</sup>, Program Studi Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

<sup>1</sup>[aviolla.terza.sada@upnjatim.ac.id](mailto:aviolla.terza.sada@upnjatim.ac.id)

<sup>2</sup>[amri.muhammad.stat@upnjatim.ac.id](mailto:amri.muhammad.stat@upnjatim.ac.id)

<sup>3</sup>[prismahardi.aji.ds@upnjatim.ac.id](mailto:prismahardi.aji.ds@upnjatim.ac.id)

Corresponding author email: [aviolla.terza.sada@upnjatim.ac.id](mailto:aviolla.terza.sada@upnjatim.ac.id)

**Abstract:** The World Health Organization (WHO) determined that Corona Virus Disease 2019 or also known as Covid-19 as a pandemic on March 11 2020. The Indonesian government announced the first case of Covid-19 in Indonesia on March 2 2020. There were 2 daily spikes in cases Covid-19 in Indonesia, namely during the first spike (14 May 2021 – 15 July 2021) and the second spike (26 December 2021 – 16 February 2022). Using the ARIMA method, the ARIMA model with the lowest MSE value was obtained, namely ARIMA (2,1,2) for the first spike with an MSE value of 2720874 and ARIMA (2,1,3) for the second spike with an MSE value of 13357211.

**Keywords:** Forecasting, Covid-19, ARIMA

**Abstrak:** Organisasi Kesehatan Dunia (WHO) menetapkan bahwa Corona Virus Disease 2019 atau disebut juga dengan Covid-19 sebagai pandemi pada tanggal 11 Maret 2020. Pemerintah Indonesia mengumumkan kasus Covid-19 di Indonesia pertama kali pada 2 Maret 2020. Terjadi 2 lonjakan kasus harian Covid-19 di Indonesia yaitu pada rentang waktu lonjakan pertama (14 Mei 2021 – 15 Juli 2021) dan lonjakan kedua (26 Desember 2021 – 16 Februari 2022). Menggunakan metode ARIMA diperoleh model ARIMA dengan nilai MSE terendah yaitu ARIMA (2,1,2) untuk lonjakan pertama dengan nilai MSE 2720874 dan ARIMA (2,1,3) untuk lonjakan kedua dengan nilai MSE 13357211.

**Kata kunci:** Peramalan, Covid-19, ARIMA

## I. PENDAHULUAN

Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia (WHO) menetapkan bahwa Corona Virus Disease 2019 atau disebut juga dengan Covid-19 sebagai pandemi. Covid-19 pertama kali diketahui di Kota Wuhan, Hubei, Tiongkok pada tanggal 31 Desember 2019. WHO menyampaikan bahwa kasus harian Covid-19 di Dunia hingga 10 Mei 2023 tercatat total 765.903.278, hingga menyebabkan 6.927.378 pasien meninggal dunia [1]. Indonesia termasuk negara dengan kasus harian Covid-19 tertinggi ke-20 di Dunia menurut catatan WHO pada 8 Mei 2023.

Total kasus Covid-19 di Indonesia hingga 8 Mei 2023 tercatat 6.791.121 kasus. Sampai saat ini pemerintah Indonesia berupaya menekan penularan tersebut dengan menekan kasus harian Covid-19. Salah satu cara yang dapat membantu pemerintah Indonesia dalam menekan kasus tersebut dengan menggunakan metode peramalan pada lonjakan kasus harian Covid-19 di Indonesia yang sebelumnya telah terjadi. Terdapat 2 lonjakan kasus harian Covid-19 tertinggi di Indonesia yaitu pada tanggal 15 Juli 2021 sejumlah 56.757 kasus harian Covid-19 dan tanggal 16 Februari 2022 sejumlah 64.718 kasus harian Covid-19 [2].

Salah satu metode peramalan yang dapat digunakan meramalkan lonjakan kasus harian Covid-19 tertinggi di Indonesia adalah metode ARIMA (Autoregressive Integrated Moving Average). Metode ARIMA adalah metode yang mudah dan cepat dengan menggunakan pola data yang ada, selain itu metode ARIMA mempunyai tingkat kedekatan yang lebih tinggi dan mempunyai nilai kesalahan yang kecil karena proses perhitungan yang secara bertahap [3]. Adapun penelitian terdahulu yang menggunakan metode ARIMA yaitu untuk meramalkan penyebaran Covid-19 di Arab Saudi diperoleh model ARIMA (2,1,1) sangat sesuai dengan nilai penyebaran sebenarnya [4]. Metode penelitian serupa juga digunakan untuk meramalkan penyebaran Covid-19 di 5 negara yaitu, AS, Brazil, India, Rusia dan Spanyol. Diperoleh peramalan Rusia dan Spanyol telah mencapai titik infleksi dalam penyebaran Covid-19 dengan model masing-masing ARIMA (3,0,0) dan ARIMA

(4,2,4). Sementara peramalan penyebaran Covid-19 AS, Brazil, dan India masih mengalami kurva eksponensial dengan model masing-masing secara berurutan yaitu ARIMA (1,2,1), ARIMA (3,1,2), dan ARIMA (4,2,4) (Sahai, 2020). Adapula penelitian terdahulu yang menggunakan metode yang sama dengan sehingga didapatkan model terbaik dari prediksi total kasus pasien positif Covid-19 di Kabupaten Sidoarjo yaitu ARIMA (2,2,1) dengan nilai MSE sebesar 1540,51. Sedangkan model terbaik untuk data total kasus pasien sembuh Covid-19 di Kabupaten Sidoarjo yaitu ARIMA (3,1,2) dengan nilai MSE sebesar 526,81 [5]. Dari berbagai penelitian serupa tersebut, penting dilakukan penelitian pada lonjakan kasus Covid-19 di Indonesia dengan tujuan menentukan model dalam meramalkan kedua lonjakan kasus harian Covid-19 di Indonesia menggunakan model ARIMA.

## II. METODE PENELITIAN

Penelitian ini menggunakan metode ARIMA, dimana metode ARIMA memiliki ketepatan yang sangat baik untuk peramalan jangka pendek, sedangkan kurang baik untuk peramalan jangka panjang. Karena memiliki kemungkinan cenderung flat (konstan) untuk periode yang cukup panjang [6]. Model ARIMA adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari runtun waktu (time series) secara statistik berhubungan satu sama lain (dependent). Oleh karena ARIMA merupakan salah satu jenis pemodelan yang menjelaskan runtun waktu tertentu berdasarkan nilai masa lalunya (lag) dan kesalahan perkiraan lag, sehingga persamaannya dapat digunakan untuk meramalkan nilai masa depan. Sehingga notasi modelnya berbentuk ARIMA ( $p,d,q$ ) [7], sebagai berikut.

1.  $p$  adalah orde untuk proses *autoregressive* (AR)
2.  $d$  adalah orde yang menyatakan banyaknya proses diferensi yang dilakukan pada data *time series* yang belum stasioner
3.  $q$  adalah orde untuk proses *moving average* (MA).

Pemodelan ARIMA hanya cocok untuk data runtun waktu yang stasioner. Sebuah data runtun waktu yang stasioner, memiliki mean, varians, dan koefisien autokorelasi (ACF & PACF) yang konstan terhadap waktu. Fungsi AutoKorelasi (ACF) dan Fungsi AutoKorelasi Parsial (PACF) ini merupakan sebuah cara menggambarkan bagaimana sebuah pengamatan berhubungan satu sama lainnya. Dimana karakteristik pembeda utama dari ACF dan PACF teoritis untuk proses yang stasioner sebagaimana tabel dibawah ini [8].

**Tabel 1.** Plot ACF dan plot PACF model ARIMA yang stasioner

Model	ACF	PACF
AR ( $p$ )	Turun secara eksponensial ( <i>dies down</i> )	Terpotong setelah lag $p$ ( <i>cut off after lag p</i> )
MA ( $q$ )	Terpotong setelah lag $q$ ( <i>cut off after lag q</i> )	Turun secara eksponensial ( <i>dies down</i> )
ARMA ( $p,q$ )	Turun secara eksponensial ( <i>dies down</i> )	Turun secara eksponensial ( <i>dies down</i> )

Data runtun waktu dengan varians yang tidak stasioner, sering kali akan menjadi stasioner jika dilakukan transformasi. Salah satunya dengan transformasi Box-Cox yang melibatkan proses estimasi dari data untuk mendapatkan transformasi yang sesuai [9]. Data runtun waktu dengan mean yang stasioner tidak lagi diperlukan diferensi maka umumnya mean tidak akan sama dengan nol. Apabila data runtun waktu tidak stasioner harus didiferensi ( $d > 0$ ). Dilakukan transformasi terlebih dahulu sebelum melakukan diferensi terhadap data. Jika melakukan diferensi terlebih dahulu, akan mengalami masalah karena runtun waktu hasil diferensi ini memiliki nilai negatif, dan transformasi sebuah bilangan negatif merupakan sesuatu yang tak-terdefiniskan [9].

Salah satu cara dalam pemilihan model terbaik dapat dilakukan menggunakan model ARIMA dengan nilai MSE terkecil yang dihasilkan dari persamaan sebagai berikut [10].

$$MSE = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2 (1)$$

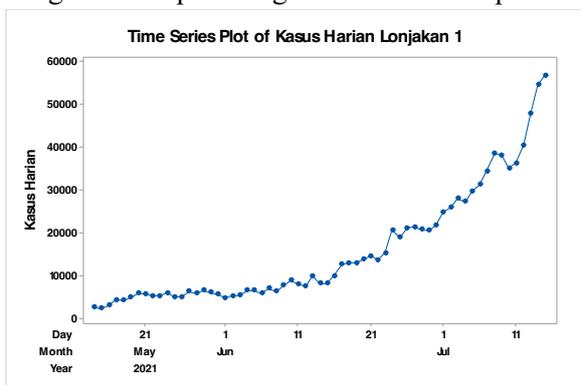
Dimana  $n$  = jumlah data,  $X_t$  = data pada periode ke  $t$ ,  $\hat{X}_t$  = data peramalan pada periode ke  $t$

Penelitian ini menggunakan data sekunder dari World Health Organization (WHO) pada laman website <https://covid19.who.int/data>. Data berupa data kuantitatif dari jumlah kasus harian Covid-19 di Indonesia dengan rentang waktu lonjakan pertama (14 Mei 2021 – 15 Juli 2021) sebanyak 63 data dan rentang waktu lonjakan kedua (26 Desember 2021 – 16 Februari 2022) sebanyak 53 data. Adapun langkah-langkah pengolahan data yang dilakukan sebagai berikut.

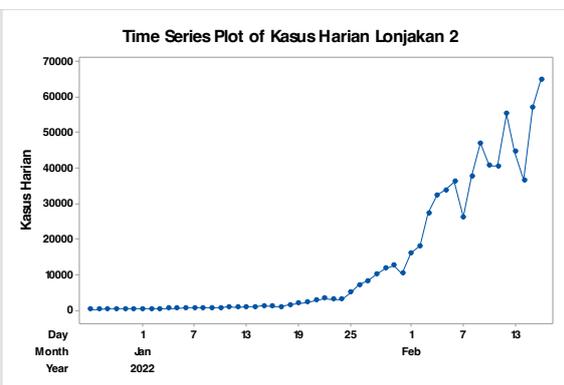
1. Identifikasi data menggunakan plot data,
2. Menganalisis stasioneritas dalam varians, apabila tidak memenuhi stasioneritas dalam varians maka dilakukan Transformasi,
3. Menganalisis stasioneritas dalam mean, apabila tidak memenuhi stasioneritas dalam mean maka dilakukan Diferensi,
4. Menentukan model ARIMA terbaik.
5. Melakukan prediksi untuk 7 hari kedepan.

### III. HASIL DAN PEMBAHASAN

Diperoleh plot data kasus harian Covid-19 di Indonesia pada lonjakan 1 (Gambar 1a) dan lonjakan 2 (Gambar 1b) mengalami kenaikan kasus. Dimana terjadi kenaikan kasus harian lonjakan 2 lebih tinggi dari kasus harian lonjakan 1. Sehingga mengakibatkan kebutuhan sarana dan prasarana kesehatan menjadi mengalami kenaikan, dan dapat berakibat fatal kepada pasien yang seharusnya mendapat penanganan secepat mungkin di sarana dan prasarana kesehatan tersebut.

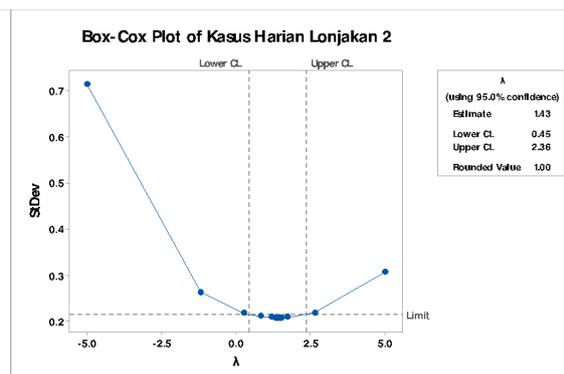
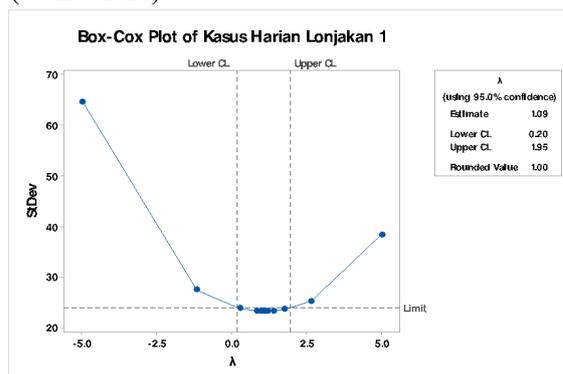


**Gambar 1a.** Plot Data Lonjakan 1



**Gambar 1b.** Plot Data Lonjakan 2

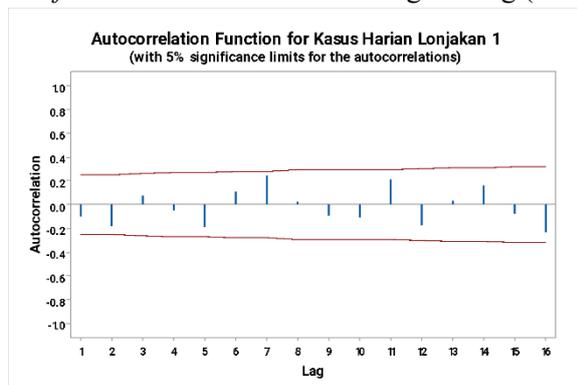
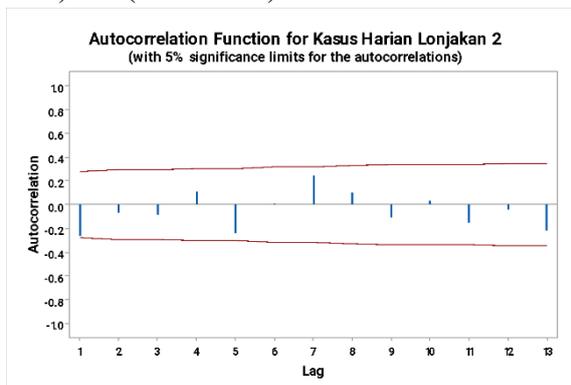
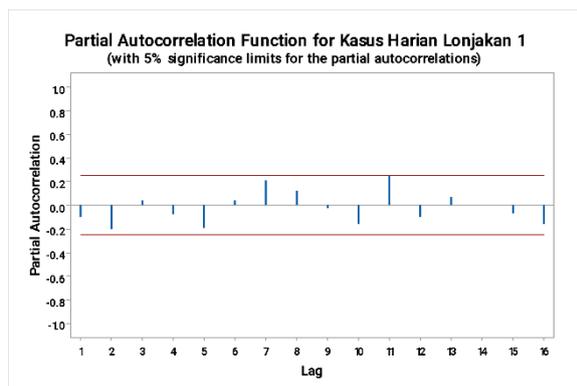
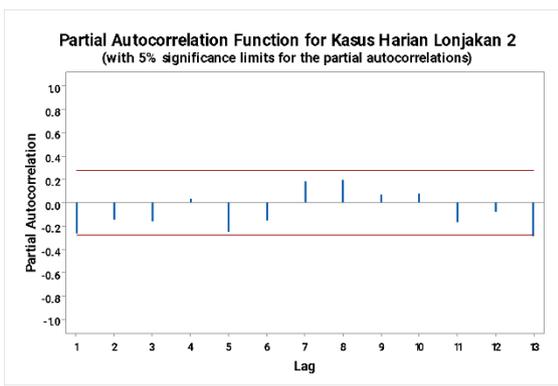
Pada data kasus harian Covid-19 lonjakan 1 (Gambar 1a) dilakukan transformasi Box-Cox sebanyak 2 kali supaya memenuhi stasioner dalam varians. Sedangkan, pada data kasus harian Covid-19 lonjakan 2 (Gambar 1b) dilakukan transformasi Box-Cox sebanyak 1 kali. Data yang memenuhi stasioner dalam varians dapat ditunjukkan oleh nilai Rounded Value = 1, dimana pada penelitian ini data telah memenuhi stasioner dalam varians dibuktikan pada nilai Rounded Value = 1 (Gambar 2a) dan (Gambar 2b).



**Gambar 2a.** Plot Box Cox Lonjakan 1

**Gambar 2b.** Plot Box Cox Lonjakan 2

Data dinyatakan memenuhi stasioner dalam mean apabila kurang dari 3 lag pertama yang keluar dari *Confidence Intervals* pada ACF dan PACF. Pada data kasus harian Covid-19 pada lonjakan 1 dan lonjakan 2 di Indonesia telah memenuhi stasioner dalam mean dengan masing-masing dilakukan diferensi 1 kali. Dapat dilihat dari (Gambar 3a) dan (Gambar 3b), dimana tidak terdapat lag yang keluar dari *Confidence Intervals* pada ACF. Dan pada PACF tidak terdapat lag yang keluar dari *Confidence Intervals* untuk masing-masing (Gambar 4a) dan (Gambar 4b).


**Gambar 3a.** ACF Lonjakan 1

**Gambar 3b.** ACF Lonjakan 2

**Gambar 4a.** PACF Lonjakan 1

**Gambar 4b.** PACF Lonjakan 2

Selanjutnya ACF akan digunakan untuk menentukan parameter AR dan PACF digunakan untuk menentukan parameter MA. Setelah data memenuhi stasioner dalam varians dan mean, maka dapat dilakukan pemilihan model terbaik pada data tersebut. Diperoleh nilai MSE model ARIMA untuk masing-masing lonjakan dengan mengacu pada “(1)” sehingga diperoleh sebagai berikut.

**Tabel 2.** MSE ARIMA Lonjakan 1

Model ARIMA	Nilai MSE
(1,1,0)	3158205
(0,1,1)	3158001
(2,1,1)	3140236
(2,1,2)	2720874

**Tabel 3.** MSE ARIMA Lonjakan 2

Model ARIMA	Nilai MSE
(1,1,0)	26226153
(0,1,1)	25028499
(2,1,2)	14273431
(2,1,3)	13357211

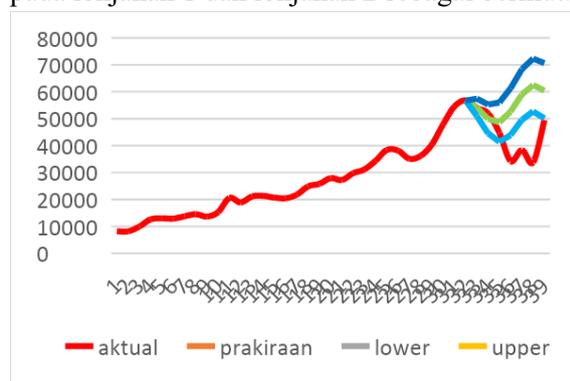
Dari kedua tabel diatas diperoleh model terbaik untuk kasus harian Covid-19 di Indonesia pada lonjakan 1 dan lonjakan 2 masing-masing adalah ARIMA (2,1,2) dan ARIMA (2,1,3). Karena pada model tersebut yang memiliki nilai MSE terkecil yaitu 2720874 dan 13357211. Setelah mendapatkan model terbaik dapat diperoleh nilai peramalan untuk kasus harian Covid-19 dalam jangka waktu tertentu. Sehingga diperoleh peramalan untuk 7 hari kedepan sebagai berikut.

**Tabel 4.** Peramalan Lonjakan 1

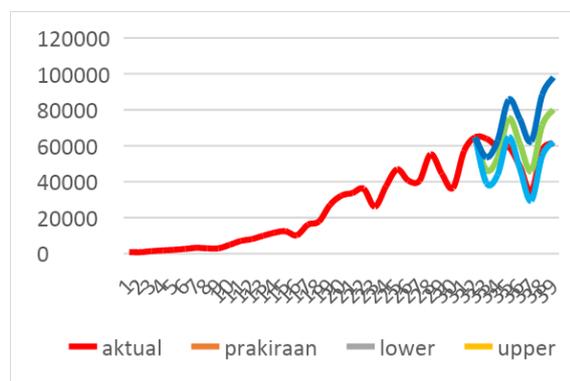
**Tabel 5.** Peramalan Lonjakan 2

Tanggal	Peramalan	Tanggal	Peramalan
16 Juli 2021	54174	17 Februari 2022	46382
17 Juli 2021	49905	18 Februari 2022	53313
18 Juli 2021	48695	19 Februari 2022	59384
19 Juli 2021	52632	20 Februari 2022	48484
20 Juli 2021	58974	21 Februari 2022	34418
21 Juli 2021	62498	22 Februari 2022	57491
22 Juli 2021	60376	23 Februari 2022	61488

Adapun dapat dibandingkan nilai peramalan selama 7 hari kedepan dengan nilai aktual kasus harian Covid-19 pada tanggal yang sama menggunakan grafik. Sehingga didapatkan grafik permalan pada lonjakan 1 dan lonjakan 2 sebagai berikut.



**Gambar 5a.** Peramalan Lonjakan 1



**Gambar 5b.** Peramalan Lonjakan 2

Dari gambar tersebut menunjukkan bahwa kasus harian Covid-19 pada lonjakan 1 (Gambar 5a) dan lonjakan 2 (Gambar 5b) apabila sesuai dengan peramalan maka akan terjadi kenaikan kembali pada jangka waktu 7 hari kedepan. Akan tetapi secara aktual lonjakan tertinggi sudah terlewati dan tidak terdapat kenaikan kasus harian Covid-19 kembali pada 7 hari kedepan.

#### IV. KESIMPULAN

Model ARIMA (2,1,2) diperoleh berdasarkan data kasus harian Covid-19 di Indonesia dengan rentang waktu lonjakan pertama (14 Mei 2021 – 15 Juli 2021). Dan model ARIMA (2,1,3) diperoleh berdasarkan data kasus harian Covid-19 di Indonesia dengan rentang waktu lonjakan kedua (26 Desember 2021 – 16 Februari 2022). Kebijakan pemerintah Indonesia memperketat protokol kesehatan dengan merubah PSBB menjadi PPKM di Indonesia sejak awal 2021 adalah sangat tepat. Adapun program dari pemerintah Indonesia yang mungkin berpengaruh terhadap penurunan lonjakan kasus harian Covid-19 adalah vaksinasi Covid-19 yang dimulai tanggal 13 Januari 2021 dan sosialisasi 5M dilingkungan masyarakat. Sebagai masukan untuk penelitian selanjutnya dapat mengkaji faktor-faktor yang mempengaruhi turunnya lonjakan kasus harian Covid-19 di Indonesia. Ataupun penelitian serupa bisa dilakukan menggunakan metode yang berbeda.

#### UCAPAN TERIMA KASIH

#### REFERENSI

1. World Health Organization. 2023. Global Situation – 10 May 2023. Available online: <https://covid19.who.int> (updated 2023 May 10).
2. Covid19. 2023. Covid 19 – 10 May 2023. Available online: [Peta Sebaran | Covid19.go.id](https://peta.sebaran.covid19.go.id) (updated 2023 May 10).



3. Mardiyah, I., Utami, W. D., Novitasari, D. C. R., Hafiyusholeh, M., & Sulistiyawati, D. (2021). Analisis prediksi jumlah penduduk di kota Pasuruan menggunakan metode arima. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 15(3), 525-534. [Google Scholar](#)
4. Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of infection and public health*, 13(7), 914-919. [Google Scholar](#)
5. Ainiyah, L., & Bansori, M. (2021). PREDIKSI JUMLAH KASUS COVID-19 MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)(STUDI KASUS KABUPATEN SIDOARJO) FORECASTING COVID-19 CASES USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) METHOD (CASE FOR SIDOARJO REGENCY. *vol, 10*, 62-68. [Google Scholar](#)
6. Hendrawan, B. (2012). Penerapan Model ARIMA dalam memprediksi IHS. *Jurnal Integrasi*, 4(2), 205-211. [Google Scholar](#)
7. Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1847979018808673. [Google Scholar](#)
8. Yunita, T. (2020). Peramalan Jumlah Penggunaan Kuota Internet Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA). *Journal of Mathematics: Theory and Applications*, 16-22. [Google Scholar](#)
9. Soelaeman, I., *Analisis Runtun Waktu (Edisi 1)*. Tangerang Selatan: Universitas Terbuka, 2016.
10. Rachmawati, A. K. (2021). Peramalan Penyebaran Jumlah Kasus Covid19 Provinsi Jawa Tengah dengan Metode ARIMA. *Zeta-Math Journal*, 6(1), 11-16. [Google Scholar](#)



# Pengaruh Nilai Tukar Rupiah terhadap Dolar Amerika Serikat dengan Jumlah Wisatawan Asia Tenggara

Vannesa Nathania<sup>1</sup>, Amirah Rizky Ramadhanti<sup>2</sup>, Trimono<sup>3</sup>,

<sup>1,2,3</sup>. Program Studi Sains Data, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur

<sup>1</sup>[22083010044@student.upnjatim.ac.id](mailto:22083010044@student.upnjatim.ac.id)

<sup>3</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

Corresponding author email: [22083010036@student.upnjatim.ac.id](mailto:22083010036@student.upnjatim.ac.id)

**Abstract:** Regression analysis is a statistical method used to calculate the estimated relationship between variables that aims to draw an understanding of the relationship between one dependent variable and one or more independent variables. The purpose of this study is to determine the effect of the rupiah exchange rate against the US dollar on the number of tourists from Southeast Asia. The method used in this research is a data-based analysis method with a regression analysis approach. The result of this study is that when the exchange rate increases, the number of tourists will decrease.

**Keywords:** Regression Analysis, Exchange Rates, Tourist.

**Abstrak:** Analisis regresi merupakan suatu metode statistik yang digunakan untuk menghitung perkiraan hubungan antara variabel-variabel yang bertujuan untuk menarik pemahaman mengenai hubungan antara satu variabel dependen dan satu atau lebih variabel independen. Tujuan dari penelitian ini yakni untuk mengetahui pengaruh nilai tukar rupiah terhadap Dolar Amerika Serikat dengan jumlah wisatawan dari Asia Tenggara. Metode yang digunakan dalam penelitian ini adalah metode analisis berdasarkan data dengan pendekatan analisis regresi. Hasil penelitian ini adalah ketika jumlah kurs mengalami kenaikan, jumlah wisatawan akan mengalami penurunan.

**Kata kunci:** Analisis Regresi, Kurs, Wisatawan.

## I. PENDAHULUAN

Indonesia merupakan suatu negara yang memiliki keanekaragaman budaya dan tempat-tempat yang menarik daya wisata. Indonesia merupakan wilayah yang terletak pada garis khatulistiwa sehingga termasuk negara tropis. Wisata di Indonesia didominasi oleh pantai, gunung, dan air terjun sehingga menarik wisatawan luar negeri untuk berkunjung. Selain itu, keunikan Indonesia terletak pada budaya, contohnya seperti tenun tradisional, tari daerah, serta makanan khas yang mampu menjadi daya tarik tersendiri bagi wisatawan luar. Dengan melihat banyaknya potensi yang bisa didapatkan karena keunikan Indonesia, maka sektor pariwisata harus terus ditingkatkan dengan memenuhi dan mengetahui dengan dipertimbangkan oleh pengaruh tertentu.

Selain berkunjung untuk berlibur, tidak jarang wisatawan luar mempelajari dan ikut melestarikan budaya di Indonesia. Dalam melakukan perencanaan liburan, terdapat beberapa tahapan yang perlu dilakukan oleh calon pengunjung, antara lain mematkan paspor berlaku setidaknya enam bulan, melakukan pembayaran VOA (*Visa On Arrival*), dan menyiapkan dokumen pendukung lainnya [1–3]. Paspor merupakan dokumen resmi yang dikeluarkan oleh pejabat berwenang dari suatu negara yang memuat identitas pemegangnya dan berlaku untuk melakukan perjalanan antar negara. Sedangkan, VOA adalah visa yang bisa dilakukan untuk beberapa aktivitas di negara tujuan. Aktivitas yang bisa dilakukan, antara lain kunjungan wisata, kunjungan pejabat pemerintah, pembicaraan bisnis dan kunjungan pertemuan, transit, dan pembelian barang kunjungan.

Saat melakukan transaksi, para wisatawan diwajibkan menggunakan rupiah sebagai mata uang yang berlaku. Ada beberapa sumber yang menyatakan bahwa barang dan makanan di Indonesia ramah di kantong menurut wisatawan. Hal ini menciptakan pemikiran bahwa nilai tukar rupiah memiliki pengaruh terhadap Jumlah Wisatawan di Indonesia. menukarkan uang negaranya menjadi uang rupiah. Untuk itu, pada artikel kali ini, kurs dianggap menjadi salah satu pertimbangan wisatawan luar negeri untuk berkunjung di Indonesia. Nilai Tukar Rupiah.



Nilai Tukar Rupiah merupakan nilai yang mewakili jumlah mata uang rupiah yang dijumlahkan ke nilai mata uang negara lain. Salah satu patokan yang digunakan adalah nilai tukar rupiah terhadap Dolar Amerika Serikat. Dalam perekonomian Indonesia, nilai tukar rupiah terhadap Dolar Amerika Serikat merupakan indikator penting. Menurut Badan Pusat Statistik, rata-rata nilai tukar rupiah terhadap Dolar Amerika Serikat dari tahun 2006 hingga 2019 mengalami fluktuasi. Terdapat beberapa faktor yang mempengaruhinya, antara lain tingkat inflasi, kebijakan pemerintah, perbedaan tingkat suku bunga, neraca perdagangan, serta ekspektasi nilai tukar di masa yang akan datang.

Berdasarkan penelitian yang dilakukan oleh Anggita Permata Yakup dan Tri Haryanto pada tahun 2019, dinyatakan bahwa pertumbuhan ekonomi memiliki pengaruh positif terhadap kunjungan wisatawan mancanegara [3]. Metode yang dilakukan dalam penelitian tersebut yakni menggunakan model persamaan simultan. Pengertian dari model persamaan simultan yakni merupakan suatu pemodelan yang digunakan untuk menganalisis keterkaitan antara variabel-variabel dalam suatu sistem.

Analisis regresi merupakan suatu metode statistik yang digunakan untuk menghitung perkiraan hubungan antara variabel-variabel dengan melakukan teknik-teknik pemodelan dan analisis terhadap beberapa variabel. Tujuan utama dari analisis regresi yakni menarik pemahaman mengenai hubungan antara satu variabel dependen dan satu atau lebih variabel independen (prediktor). Penghitungan perkiraan koefisien dilakukan dalam proses analisis regresi guna menggambarkan tingkat perubahan yang diharapkan dalam variabel dependen ketika variabel independen berubah (Amstrong, 2012) [4].

Adapun tujuan dari penelitian ini adalah mengetahui ada atau tidaknya pengaruh antara nilai tukar rupiah terhadap dolar Amerika dengan jumlah kunjungan wisatawan Asia Tenggara ke Indonesia. Jika diketahui memiliki pengaruh maka, dapat menentukan seberapa besar pengaruhnya.

## II. METODE PENELITIAN

Jenis penelitian yang digunakan yakni metode analisis berdasarkan data dengan pendekatan analisis regresi. Penelitian dengan menggunakan analisis regresi pada penelitian ini bertujuan untuk mengetahui pengaruh kurs rupiah dengan jumlah wisatawan mancanegara dengan periode 2006 hingga 2019.

### III.1. *Populasi dan Sampel*

Populasi merujuk pada sekumpulan objek atau subjek yang memiliki ciri khas tertentu yang telah ditentukan oleh peneliti untuk dianalisis dan dilakukan penarikan kesimpulan [5]. Menurut Hadari Nawawi, populasi dapat diartikan sebagai keseluruhan objek penelitian yang terdiri dari makhluk hidup, benda mati, gejala, nilai tes, serta peristiwa-peristiwa sebagai sumber data yang memiliki ciri khas tertentu di dalam suatu penelitian [6]. Dari kedua pendapat tersebut, dapat disimpulkan bahwa populasi adalah nilai tes pada sumber data yang memiliki nilai dan dapat diteliti.

Populasi yang digunakan dalam penelitian ini adalah nilai kurs tengah mata uang asing terhadap rupiah di Bank Indonesia, serta jumlah kunjungan wisatawan mancanegara ke Indonesia menurut kebangsaan. Sampel penelitian ini adalah data nilai kurs tengah mata uang Dolar Amerika Serikat pada rentang waktu 2006 sampai 2019 dan jumlah kunjungan wisatawan dari Asia Tenggara pada rentang waktu 2006 hingga 2019. Data yang digunakan bersumber pada Badan Pusat Statistik (BPS) [7-8]

### III.2. *Teknik Pengumpulan Data*

Teknik pengumpulan data dalam penelitian ini menggunakan teknik dokumentasi, yaitu pengumpulan data dari dokumen atau sumber tertulis. Jurnal, artikel, buku, dan basis data yang dapat ditemukan di internet. Literatur yang digunakan ada pada 3 lingkup yaitu:

- Pengaruh Nilai Tukar Rupiah (NTR) terhadap wisatawan [9];
- Faktor-faktor yang berpengaruh terhadap banyaknya jumlah wisatawan [10];

- Model Regresi untuk melakukan perhitungan [11]

### III.3. Variabel dan Pengukuran

Variabel bebas atau variabel independen dapat disebut sebagai variabel pemengaruh timbulnya variabel dependen atau variabel terikat. Sementara itu, variabel bebas atau variabel dependen disebut sebagai variabel yang dipengaruhi oleh variabel bebas [12]. Secara sederhana, variabel independen adalah sebab, sedangkan variabel dependen adalah akibat. Dalam penelitian ini, variabel independen yang digunakan sebagai sebab adalah kurs rupiah, sedangkan variabel dependen yang digunakan sebagai akibat adalah jumlah wisatawan.

### III.4. Teknik Analisis Data

Teknik Analisis Data yang digunakan pada penelitian ini adalah analisis regresi yang bertujuan untuk mengetahui pengaruh terhadap satu atau lebih variabel. Analisis regresi menarik pemahaman mengenai hubungan antara satu variabel dependen dan satu atau lebih variabel independen (prediktor). Penghitungan perkiraan koefisien dilakukan dalam proses analisis regresi guna menggambarkan tingkat perubahan yang diharapkan dalam variabel dependen ketika variabel independen berubah.

## III. HASIL DAN PEMBAHASAN

### IV.1. Statistika Deskriptif

Rata-rata jumlah wisatawan luar negeri yang berkunjung di Indonesia pada rentang waktu 2006 sampai dengan 2019 adalah 3.562.133. Jumlah wisatawan tertinggi ada pada angka 6.157.190 pada tahun 2019 dan nilai terendah ada pada tahun 2006 dengan jumlah 2.171.551 orang. Dari data nilai maksimal dan minimal dapat diketahui bahwa adanya peningkatan yang pesat dalam rentang waktu 13 tahun. Sedangkan rata-rata kurs rupiah ke Dolar Amerika Serikat adalah Rp11.450,57 per satu dolar. Nilai kurs tertinggi ada pada tahun 2018 dengan nilai Rp14.481, sedangkan nilai terendah kurs ada pada tahun 2010 dengan nilai Rp8.991. Berdasarkan data kurs dapat diketahui bahwa nilai kurs mengalami fluktuasi atau turun-naiknya perubahan harga.

**Tabel 1.** Statistika Deskriptif

Variabel	Max	Min	Mean
Jumlah_Wisatawan	6.157.190	2.171.551	3.562.133
Nilai Kurs	Rp14.481	Rp8.991	Rp11.450,57

### IV.2. Asumsi Regresi

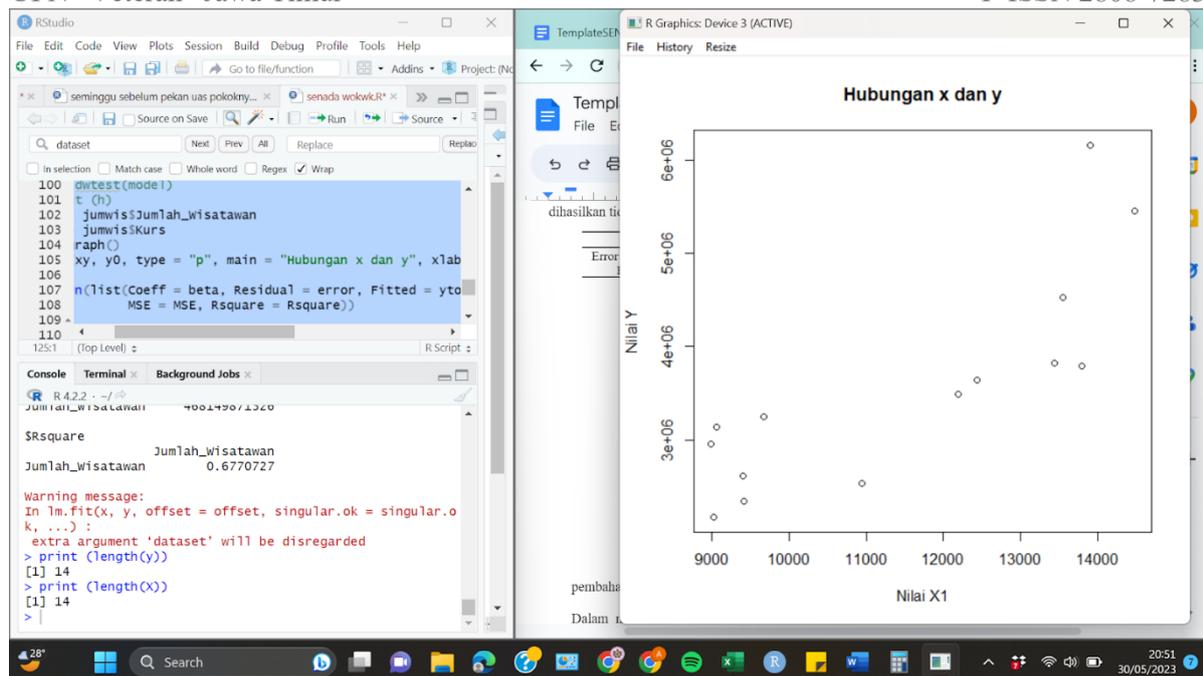
Dalam melakukan analisis regresi ada asumsi-asumsi yang harus dipenuhi; nilai error berdistribusi normal, hubungan antara x dan y linier, varians dari error homogen (homoskedalitas), dan nilai error tidak berkorelasi (non-autokorelasi). Jika salah satu asumsi tidak terpenuhi maka model yang dihasilkan tidak valid.

Asumsi-asumsi yang harus dipenuhi dalam analisis regresi yakni sebagai berikut.

- a. Nilai *error* berdistribusi normal
- b. Hubungan antara x dan y linear
- c. Homoskedalitas

**Tabel 2.** Hasil Asumsi Regresi

Asumsi	Alpha	P-value	Keputusan
Error berdistribusi normal	0.05	0.6709	Berdistribusi normal
Homoskedalitas	0.05	0.1407	Homoskedalitas



**Gambar 1.** Hubungan linear antara x dan y

Dalam kasus asumsi analisis regresi, normalitas dari nilai *error* merupakan asumsi penting karena mampu mempengaruhi validitas inferensi statistik, estimasi parameter yang optimal, konsekuensi statistik regresi, serta hasil uji hipotesis yang akurat. Distribusi normal dari *error* dapat diketahui bilamana memenuhi  $H_0$  dan menolak  $H_1$ .  $H_0$  yang dimaksud adalah sampel data berasal dari populasi yang berdistribusi normal, sedangkan  $H_1$  merupakan sampel data tidak berasal dari populasi yang berdistribusi normal. Nilai *alpha* yang digunakan yakni sebesar 0.05.

Dalam kasus ini, untuk dapat mengetahui korelasi antara x sebagai nilai kurs dan y sebagai jumlah\_wisatawan perlu menggunakan *plot*. *Output* yang diberikan memberikan gambaran bahwa nilai dari x dan y linear. Namun, terdapat beberapa nilai yang menjadi *outlier*.

Hasil dari uji asumsi normalitas nilai *error* menunjukkan bahwa nilai *p-value* lebih besar dari nilai *alpha*. Oleh karena itu,  $H_0$  gagal ditolak sehingga sampel data berasal dari populasi yang berdistribusi normal.

Homoskedastisitas adalah variansi error yang memiliki nilai homogen. Ketentuan variansi dari error dapat diketahui bilamana memenuhi  $H_0$  dan menolak  $H_1$ . Hipotesis nol adalah variansi bersifat homoskedastisitas, sedangkan hipotesis lainnya adalah variansi bersifat heteroskedastisitas. Nilai *alpha* yang ditetapkan dalam uji asumsi ini sebesar 0.05.

Hasil dari uji asumsi homoskedastisitas ini menunjukkan bahwa nilai *p-value* lebih dari nilai *alpha* sehingga hipotesis nol gagal ditolak, di mana variansi bersifat homoskedastisitas.

#### IV.3. Tahapan Analisis Regresi

Dalam melakukan analisis regresi, ada beberapa tahapan atau proses yang harus dilakukan untuk menghasilkan model yang sesuai. Beberapa tahapannya antara lain:

1. Menghitung nilai beta;

Nilai beta yang dimaksud merupakan estimasi untuk hubungan antara variabel independen (X) dan variabel dependen (y). Dalam menghitung nilai beta, terdapat beberapa langkah yang perlu dilakukan. Langkah-langkah tersebut adalah sebagai berikut.

- a. Melakukan transpose pada matriks X yang merupakan variabel independen.
- b. Mengalikan transpose matriks X dengan matriks X itu sendiri
- c. Melakukan invers matriks dari langkah kedua
- d. Menghitung perkalian invers matriks dari langkah ketiga dengan transpose X



- e. Melakukan perkalian langkah keempat dengan vektor  $y$  yang merupakan variabel dependen

Langkah-langkah di atas dapat dituliskan dengan rumus berikut.

$$\beta = solve(t(X) \% \times \%X) \% \times \%t(X) \% \times \%y$$

- 2. Menghitung nilai sigma (MSE);

MSE merupakan estimasi dari varian residual yang mampu menggambarkan seberapa baik model regresi memiliki kecocokan terhadap data. Dalam menghitung nilai sigma (MSE) dalam analisis regresi, terdapat beberapa langkah yang dapat dijabarkan sebagai berikut.

- a. Melakukan perkalian antara matriks  $X$  dengan invers dari perkalian matriks transpose  $X$ . Kemudian, hasil tersebut dikalikan dengan transpose matriks  $X$ . Matriks hasil tersebut diinisialkan sebagai  $H$ .

$$H = X \% \times \%solve(t(X) \% \times \%X) \% \times \%t(X)$$

- b. Melakukan pembuatan matriks identitas berukuran sebanyak  $n$  yang merupakan jumlah observasi dalam variabel dependen  $y$ . Pada matriks identitas, diagonal utama bernilai 1 dan elemen lainnya bernilai 0. Matriks hasil tersebut diinisialkan sebagai  $I$ .

$$I = diag(rep(1, n))$$

- c. Melakukan perhitungan jumlah kuadrat residual (SEE) dengan cara melakukan pengurangan antara matriks  $I$  dengan matriks  $H$ , kemudian mengalikan matriks transpose  $y$  dengan hasil pengurangan tersebut. Terakhir, hasil perkalian tersebut dikalikan dengan vektor  $y$ .

$$SSE = t(y) \% \times \%(I - H) \% \times \%y$$

- d. Menghitung derajat bebas (db) dengan melakukan pengurangan antara jumlah observasi ( $n$ ) dengan jumlah variabel independen ( $p$ ) dan 1.

$$db = n - (p + 1)$$

- e. Menghitung *Mean Squared Error* (MSE) dengan melakukan pembagian antara SSE dengan derajat bebas (db).

$$MSE = \frac{SSE}{db}$$

- 3. Menghitung *R-Square*;

*R-Square* merupakan suatu ukuran yang mampu menggambarkan seberapa baik kecocokan regresi dengan data. Nilai *R-Square* berada di antara 0 dan 1. Semakin mendekati 1, maka semakin baik model regresi dalam menjelaskan data. Hasil dari perhitungan SSE juga digunakan dalam rumus ini. Langkah-langkah dalam menghitung *R-Square* yakni sebagai berikut.

- a. Membuat matriks  $J$  berukuran  $n \times n$  yang semua elemennya bernilai 1.

$$J = matrix(rep(1, n \times n), ncol = n)$$

- b. Menghitung nilai SST dengan cara melakukan perkalian matriks antara transpose vektor  $y$ , hasil pengurangan antara  $I$  dengan perkalian antara  $(1/n)$  dengan  $J$ , dan vektor  $y$ .

$$SST = t(y) \% \times \% \left( I - \left( \frac{1}{n} \right) \times J \right) \% \times \%y$$

- c. Menghitung nilai *R-Square* dengan melakukan operasi pengurangan 1 dengan hasil dari pembagian antara SSE dan SST.

$$Rsquare = 1 - \left( \frac{SSE}{SST} \right)$$

- 4. Menghitung nilai prediksi dan *error*;

Perhitungan ini dapat membantu pengevaluasian seberapa baik model regresi dalam memprediksi nilai variabel dependen berdasarkan variabel independen, serta memahami



seberapa besar kecocokan model dengan data yang diteliti. Langkah-langkah dalam menghitung nilai prediksi dan *error* adalah sebagai berikut.

- a. Menghitung nilai prediksi ( $y$  hat) dengan melakukan perkalian matriks  $X$  dengan vektor beta yang telah dihitung sebelumnya. Hasil dari nilai prediksi tersebut berupa vektor yang memiliki nilai prediksi untuk variabel dependen.

$$\hat{Y} = X\% \times \% \beta$$

- b. Menghitung *error* dengan melakukan pengurangan vektor nilai aktual ( $y$ ) dari nilai prediksi ( $y$  hat). Hasil perhitungan tersebut berupa nilai *error* antara nilai aktual dan nilai prediksi.

$$error = y - \hat{Y}$$

- c. Melakukan penggabungan nilai aktual ( $y$ ), nilai prediksi ( $y$  hat), dan *error* menjadi satu matriks menggunakan fungsi *cbind*. Hasilnya akan terdapat tiga kolom, yaitu kolom ‘ $y$ ’ untuk nilai aktual, kolom ‘ $y$  hat’ untuk nilai prediksi, dan kolom ‘*error*’ untuk nilai *error*.

- d. Memberikan masing-masing kolom sesuai nilainya.

## 5. Uji kecocokan model (uji serentak);

Pengujian ini digunakan untuk mengevaluasi signifikansi model regresi dalam menjelaskan variabilitas dalam data. Langkah-langkah dalam melakukan uji kecocokan model adalah sebagai berikut.

- a. Melakukan perhitungan *Sum of Squares Regression* (SSR) dengan mengurangi SSE dari SST. Hasil dari SSR menggambarkan variabilitas yang dijelaskan oleh model regresi.

$$SSR = SST - SSE$$

- b. Menghitung *Mean Square Regression* (MSR) dengan melakukan pembagian antara SSR dengan jumlah variabel independen ( $p$ ). Hasil dari penghitungan MSR berupa estimasi dari varian yang dijelaskan oleh model regresi per variabel independen.

$$MSR = \frac{SSR}{p}$$

- c. Menghitung nilai statistik F ( $F_{hit}$ ) dengan membagi MSR dengan MSE. Hasil dari statistik F digunakan untuk membandingkan variabilitas yang dijelaskan oleh model regresi dengan variabilitas yang tidak dapat dijelaskan oleh model (*error*).

$$F_{hit} = \frac{MSR}{MSE}$$

- d. Menghitung *p-value* atau probabilitas untuk nilai statistik F. Hasil tersebut diinisialisasikan sebagai a.

$$a = 1 - pf(F_{hit}, p, db)$$

- e. Melakukan penggabungan nilai statistik F dan a menjadi satu matriks menggunakan fungsi *cbind*. Matriks hasil ini akan terdiri dari dua kolom, yakni kolom ‘ $F_{hit}$ ’ untuk nilai statistik F ( $F_{hit}$ ) dan kolom ‘*p-value*’ untuk nilai a. Matriks tersebut diinisialisasikan dengan  $F_{stat}$ .

- f. Melakukan pemberian nama pada kolom matriks F-stat agar mudah dikenali.

## 6. Uji Parsial Parameter Regresi;

Pengujian ini digunakan untuk mengevaluasi signifikansi dari masing-masing koefisien regresi (parameter). Statistik T digunakan dalam pengujian ini. Langkah-langkah yang harus dilakukan untuk melakukan uji parameter regresi adalah sebagai berikut.

- a. Menghitung *Standard Error of beta* ( $SE_{beta}$ ) dengan melakukan penghitungan akar kuadrat dari diagonal matriks invers dari hasil perkalian transpose matriks  $X$  dengan  $X$ . Kemudian, hasilnya dikalikan dengan vektor MSE.

$$SE\_beta = \sqrt{\text{diag}(\text{solve}(t(X)\% \times \%X)) \times \text{as.vector}(MSE)}$$

- b. Menghitung statistik T ( $t\_hit$ ) dengan melakukan pembagian setiap beta dengan nilai  $SE\_beta$ .

$$t\_hit = \left(\frac{1}{SE\_beta}\right) \times \beta$$

- c. Melakukan perhitungan nilai  $sig\_t$  yang merupakan  $p$ -value untuk nilai  $t\_hit$ . Hasil dari perhitungan tersebut mengindikasikan signifikansi statistik dari koefisien regresi individu. Dalam rumus ini, ‘pt’ berperan sebagai fungsi dalam R yang menghitung probabilitas distribusi t.

$$sig\_t = 2 \times (1 - pt(abs(t\_hit), n - p - 1))$$

- d. Melakukan penggabungan nilai beta,  $SE\_beta$ ,  $t\_hit$ , dan  $sig\_t$  menjadi satu matriks menggunakan fungsi *cbind*. Matriks tersebut diinisialkan sebagai  $T\_stat$  dan terdiri dari empat kolom, yakni kolom untuk nilai koefisien regresi, nilai  $SE\_beta$ , nilai  $t\_hit$ , dan nilai  $sig\_t$ .

- e. Memberi nama kolom pada matriks  $T\_stat$  sesuai dengan nilai di dalamnya supaya mudah dikenali.

#### 7. Interval Konfidensi.

Perhitungan ini digunakan untuk mengestimasi parameter populasi dengan tingkat kepercayaan tertentu. Langkah-langkah yang harus dilakukan yakni sebagai berikut.

- a. Memilih tingkat signifikansi  $alpha$ . Dalam penelitian ini, tingkat signifikansi yang digunakan sebesar 0,05.

- b. Menghitung nilai  $t_{min}$  yang merupakan kuantil distribusi T dengan derajat kebebasan ( $n - p - 1$ ) pada tingkat  $alpha/2$ . Kuantil ini digunakan sebagai batas bawah interval kepercayaan.

$$t_{min} = qt\left(\frac{\alpha}{2}, n - p - 1\right)$$

- c. Menghitung nilai  $t_{max}$  yang merupakan kuantil distribusi T dengan derajat kebebasan ( $n - p - 1$ ) pada tingkat  $1 - alpha/2$ . Kuantil ini digunakan sebagai batas atas interval kepercayaan.

$$t_{max} = qt\left(1 - \frac{\alpha}{2}, n - p - 1\right)$$

- d. Melakukan penjumlahan antara beta dan hasil perkalian antara  $SE\_beta$  dengan  $t_{min}$ . Hasil dari perhitungan ini menunjukkan batas bawah interval kepercayaan untuk koefisien regresi beta dan diinisialkan dengan  $beta\_min$ .

$$\beta_{min} = \beta + SE_{\beta} \times t_{min}$$

- e. Melakukan penjumlahan antara beta dan hasil perkalian antara  $SE\_beta$  dengan  $t_{max}$ . Hasil dari perhitungan ini menunjukkan batas atas interval kepercayaan untuk koefisien regresi beta dan diinisialkan dengan  $beta\_max$ .

$$\beta_{max} = \beta + SE_{\beta} \times t_{max}$$

- f. Melakukan penggabungan nilai  $beta\_min$  dan  $beta\_max$  menjadi satu matriks menggunakan fungsi *cbind*. Matriks ini akan diinisialkan sebagai  $T\_stat$  serta terdiri dari dua kolom, yakni kolom untuk batas bawah interval kepercayaan dan batas atas interval kepercayaan.

- g. Memberikan nama pada setiap kolom matriks  $T\_stat$  sesuai dengan nilai di dalamnya supaya mudah dikenali.

**Tabel 3.** Hasil Perhitungan Analisis Regresi

Tahapan	Output
---------	--------

Nilai MSE	468.149.871.326
Nilai R-Square	0.6770727
Uji Kecocokan Model	F_hitung = 25.16006 p_value = 0.00030
Uji Parsial Parameter Regresi	p_value b0 = 0.166 p_value x = 0.0003 t_hitung b0 = -1.472 t_hitung x = 5.015
Koefisien Regresi	b0 = -1513951.315 kurs = 443.304

Nilai koefisien merupakan estimasi koefisien regresi untuk masing-masing variabel dalam model regresi. Dalam praktikum ini, diketahui bahwa variabel yang ada yakni  $b_0$  dan  $kurs$ . Besarnya pengaruh setiap variabel terhadap variabel terikat ditunjukkan oleh nilai estimasi. Semakin besar nilai estimasi suatu variabel, semakin besar pengaruh variabel tersebut terhadap variabel terikat. Pada output di atas, terlihat koefisien regresi untuk variabel  $kurs$  adalah 443,304. Hal tersebut menandakan bahwa apabila variabel  $kurs$  meningkat satu satuan, variabel dependen *Jumlah\_Wisatawan* akan meningkat sebesar 443,304 satuan. Sehingga, dapat dirumuskan sebagai berikut:

$$Y = -1513951,315 + 443,304X + e$$

MSE atau *Mean Squared Error* merupakan suatu ukuran kesalahan prediksi pada model regresi. Pada output di atas, terlihat bahwa nilai MSE sebesar 468.149.871.326, menandakan besar rata-rata kuadrat selisih antara nilai prediksi yang diberikan model regresi dengan nilai sebenarnya. Semakin kecil nilai MSE, semakin baik kualitas prediksi nilai sebenarnya oleh model regresi.

Nilai *R-Square* merupakan penentu seberapa baik model regresi linier mampu menjelaskan variasi data yang diamati. Nilai ini berkisar antara 0 hingga 1, di mana model regresi akan semakin baik dalam menjelaskan variasi data apabila nilai *R-Square* tinggi. Nilai *R-Square* pada output di atas sebesar 0.6770727 yang artinya  $kurs$  berpengaruh terhadap  $y$  sebesar 67%. Dalam konteks ini, dapat disimpulkan bahwa jumlah wisatawan dipengaruhi oleh nilai kurs sebesar 67%. Kemudian, banyaknya penjualan produk dipengaruhi oleh biaya pengiklanan. Kemudian, 33% sisanya dipengaruhi oleh variabel lain yang tidak dijelaskan dalam penelitian ini.

Hasil regresi nilai aktual dari variabel dependen (*Jumlah\_Wisatawan*) dan nilai prediksi dari variabel dependen ( $\hat{y}$ ) dihasilkan dari model regresi untuk setiap observasi dalam data. Selisih antara nilai aktual dan nilai prediksi ditampilkan pada kolom ketiga (*Error*). Contohnya, baris pertama dengan kolom pertama berupa nilai aktual variabel dependen ( $y$ ) sebesar 2171551 dan kolom kedua berupa nilai prediksi ( $\hat{y}$ ) dari model regresi untuk observasi sebesar 2484651 akan menghasilkan -313099.64 pada kolom ketiga (*Error*) yang merupakan selisih antara keduanya.

Uji hipotesis kecocokan model regresi yang telah dilakukan memperoleh nilai  $F_{hit}$  sebesar 25.16006 dan nilai  $p$ -value sebesar 0.00030. Dapat dilihat bahwa nilai signifikansi lebih kecil dari 0.05, di mana hal tersebut menandakan bahwa terdapat perbedaan signifikan antara model regresi yang diuji dengan data yang digunakan untuk uji tersebut. Oleh karena itu, dapat disimpulkan bahwa model regresi yang diuji cocok untuk menjelaskan hubungan antara variabel dependen dan independen pada data yang diuji.

Uji hipotesis parameter model regresi berguna sebagai pengevaluasian signifikansi koefisien regresi dari setiap variabel independen dalam model regresi. Diketahui pada output di atas terdapat 2 variabel independen dalam model regresi, yakni konstanta ( $b_0$ ) dan  $kurs$ .

- Beta merupakan koefisien regresi yang menunjukkan perubahan rata-rata variabel dependen akibat perubahan satu unit pada variabel independen tertentu;



- SE<sub>beta</sub> merupakan standar error dari koefisien regresi yang mengukur seberapa jauh rata-rata sampel dari koefisien regresi sebenarnya;
- t<sub>hit</sub> merupakan nilai statistik uji t yang menghitung seberapa jauh koefisien regresi sampel berbeda dari nol dalam satuan standar *error*;
- *p-value* menunjukkan seberapa besar kemungkinan hasil yang diamati disebabkan oleh kebetulan semata.

Dari *output* di atas, dapat disimpulkan bahwa masing-masing variabel independen berpengaruh signifikan negatif terhadap variabel dependen dalam model regresi. Sehingga dapat disimpulkan bahwa, setiap kenaikan kurs akan menyebabkan penurunan jumlah wisatawan.

Rentang nilai untuk masing-masing koefisien regresi (beta) ditunjukkan oleh interval konfidensi beta dengan tingkat kepercayaan tertentu. Pada *output* di atas, terdapat kolom *lower* dan *upper* yang berarti batas bawah dan batas atas dari rentang interval kepercayaan untuk masing-masing koefisien regresi.

Contohnya, pada interval konfidensi beta yang diperoleh, koefisien regresi konstanta ( $b_0$ ) memiliki rentang interval kepercayaan antara -1513951.315 dan  $1.02e+6$  dengan tingkat kepercayaan 95%. Oleh karena itu, dapat dikatakan jika sampel yang sama dari populasi ditarik secara berulang dan diperoleh interval kepercayaan dari masing-masing sampel, nilai parameter populasi yang sebenarnya untuk  $b_0$  akan ditutup sekitar 95% oleh interval kepercayaan tersebut

#### IV. KESIMPULAN

Model regresi linier dapat menjelaskan 67% variansi nilai *Jumlah\_Wisatawan* dengan *kurs*, sehingga 33% variabel pengaruh tidak dapat dijelaskan oleh model ini. Dilihat dari nilai probabilitas *F-statistic* dapat diketahui bahwa  $0.003 < 0.05$  sehingga variabel berpengaruh secara signifikan terhadap *Jumlah\_Wisatawan*. Model regresi yang diperoleh adalah  $-1513951.315 + 443.304X + e$  sehingga setiap kenaikan satu satuan nilai  $x$  maka akan berpengaruh secara negatif pada  $y$ . Dapat diinterpretasikan ketika jumlah *kurs* mengalami kenaikan maka akan menyebabkan penurunan terhadap *Jumlah\_Wisatawan*.

#### UCAPAN TERIMA KASIH

Penulis ingin menyampaikan rasa terima kasih setulus-tulusnya kepada semua pihak yang telah membantu dan berkontribusi dalam penelitian ini. Terima kasih kepada Tim Senada yang telah memberikan tempat bertumbuh dengan kesempatan untuk menampung penulisan artikel. Rasa terima kasih juga disampaikan kepada pak Trimono selaku dosen pengampu mata kuliah Statistika Komputasi atas ilmu yang dapat diimplementasikan pada penelitian ini.

Penulis juga ingin mengucapkan terima kasih kepada staf-staf yang berada di Gedung Kuliah Bersama yang telah memberikan kesempatan yang baik memberikan tempat yang nyaman dan sesuai kebutuhan penulis. Selanjutnya, penulis berterima kasih pada keluarga dan kolega yang memberikan dukungan penuh berupa pengertian dan juga semangat yang diberikan. Rasa terima kasih ini penulis sampaikan sejujur-jujurnya dengan kerendahan hati. Semoga, hasil dari penelitian yang telah dibuat memberikan dampak yang baik pada perkembangan pariwisata di Indonesia

#### REFERENSI

1. Apa Saja Persyaratan Perjalanan WNA untuk Memasuki Wilayah Indonesia? Ini Jawabannya. Available online: <https://www.imigrasi.go.id/id/2022/06/21/apa-saja-persyaratan-perjalanan-wna-untuk-memasuki-wilayah-indonesia-ini-jawabannya/> (30/05/2023)
2. Visa on Arrival (VOA). Available online: <https://www.imigrasi.go.id/en/visa-kunjungan-saat-kedatangan/> (30/05/2023)



3. Yakup, A. P., & Haryanto, T. (2019). Pengaruh pariwisata terhadap pertumbuhan ekonomi di Indonesia. *Bina Ekonomi*, 23(2), 39-47.
4. Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of forecasting*, 28(3), 689-694.
5. Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
6. Nawawi, Hadari. (2012). Metode Penelitian Bidang Sosial. Yogyakarta: Gajah Mada University Press.
7. Kurs Tengah Beberapa Mata Uang Asing Terhadap Rupiah di bank Indonesia dan Harga Emas di Jakarta (rupiah), 2000-2019. Available online: <https://www.bps.go.id/linkTableDinamis/view/id/952> (30/05/2023).
8. Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia Menurut Kebangsaan (Kunjungan). Available online: <https://www.bps.go.id/indicator/16/1821/2/jumlah-kunjungan-wisatawan-mancanegara-ke-indonesia-menurut-kebangsaan.html> (30/05/2023).
9. Faidzin, Nur. (2017). Pengaruh Kurs Rupiah Per Dollar Amerika terhadap Jumlah Wisatawan Mancanegara (*Inbound*) dan Jumlah Devisa Pariwisata di Indonesia Tahun 2006: MI – 2015: M12.
10. Prabowo, Destrian. Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia Tahun 2013-2017.
11. Agus Tri basuki. (2015). Analisis Regresi dalam Penelitian Ekonomi dan Bisnis. Yogyakarta: Rajawali Pers.
12. Sugiyono. (2017). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.



# Evaluasi Kualitas Pelayanan Kesehatan di Kota Surabaya: Studi Kasus pada Puskesmas dan Rumah Sakit

Irma Octavia Chaniago<sup>1</sup>, Halim Arif Cahyono<sup>2</sup>, Fikrie Hartanta Sembiring<sup>3</sup>, Regita Putri Permata,

<sup>1,2,3</sup>Prodi Sains Data, Institut Teknologi Telkom Surabaya

<sup>1</sup>[irma.octavia.21@student.ds.itelkom-sby.ac.id](mailto:irma.octavia.21@student.ds.itelkom-sby.ac.id)

<sup>2</sup>[halim.arif.21@student.ds.itelkom-sby.ac.id](mailto:halim.arif.21@student.ds.itelkom-sby.ac.id)

<sup>3</sup>[fikrie.hartanta.21@student.ds.itelkom-sby.ac.id](mailto:fikrie.hartanta.21@student.ds.itelkom-sby.ac.id)

Corresponding author email: [irma.octavia.21@student.ds.itelkom-sby.ac.id](mailto:irma.octavia.21@student.ds.itelkom-sby.ac.id)

**Abstract:** *This study aims to evaluate the quality of health services in the city of Surabaya, with a focus on health centers and hospitals, based on data obtained from the Central Bureau of Statistics (BPS). This study uses a quantitative data analytics approach to analyze the relationship between the variables of the number of health facilities, the number of health workers, the percentage of illness, and the percentage of people's reasons for health facilities. The data used in this study includes information on the number of health facilities, including puskesmas and hospitals, as well as the number of health workers available in the city of Surabaya. In addition, the data also includes the percentage of illness rates which describes the prevalence of disease in the community, as well as the percentage of people's reasons for health facilities which reflect the perceptions and preferences of the population regarding health services. Data analysis was performed using statistical methods to identify patterns and relationships between the observed variables, namely health workers, disease levels, and community preferences. The results showed a significant relationship between the number of health facilities and the number of health workers with the quality of health services. The more health facilities and health workers available, the higher the quality of service tends to be. The evaluation shows that the quality of health services in the city of Surabaya must be further improved. Although health facilities and health personnel are adequate, there is evidence that indicates deficiencies in several aspects of service quality.*

**Keywords:** *quality, facilities, variables, health, surabaya*

**Abstrak:** Penelitian ini bertujuan untuk melakukan evaluasi kualitas pelayanan kesehatan di Kota Surabaya, dengan fokus pada puskesmas dan rumah sakit, berdasarkan data yang diperoleh dari Badan Pusat Statistik (BPS). Penelitian ini menggunakan pendekatan kuantitatif *data analytics* untuk menganalisis hubungan antara variabel jumlah fasilitas kesehatan, jumlah tenaga kesehatan, persentase angka sakit, dan persentase alasan masyarakat terhadap fasilitas kesehatan. Data yang digunakan dalam penelitian ini mencakup informasi tentang jumlah fasilitas kesehatan, termasuk puskesmas dan rumah sakit, serta jumlah tenaga kesehatan yang tersedia di Kota Surabaya. Selain itu, data juga mencakup persentase angka sakit yang menggambarkan prevalensi penyakit di masyarakat, serta persentase alasan masyarakat terhadap fasilitas kesehatan yang mencerminkan persepsi dan preferensi penduduk terkait pelayanan kesehatan. Analisis data dilakukan menggunakan metode statistik untuk mengidentifikasi pola dan hubungan antara variabel yang diamati, yakni tenaga kesehatan, tingkat penyakit, dan preferensi masyarakat. Hasil penelitian menunjukkan hubungan yang signifikan antara jumlah fasilitas kesehatan dan jumlah tenaga kesehatan dengan kualitas pelayanan kesehatan. Semakin banyak fasilitas kesehatan dan tenaga kesehatan yang tersedia, maka cenderung meningkat pula kualitas pelayanan. Evaluasi menunjukkan bahwa kualitas pelayanan kesehatan di Kota Surabaya harus lebih ditingkatkan. Meskipun fasilitas kesehatan dan tenaga kesehatan cukup memadai, namun terdapat bukti yang menunjukkan adanya kekurangan dalam beberapa aspek kualitas pelayanan.

**Kata kunci:** kualitas, fasilitas, variabel, kesehatan, surabaya

## I. PENDAHULUAN

Pelayanan kesehatan yang berkualitas menjadi salah satu faktor krusial dalam meningkatkan kesejahteraan masyarakat. Evaluasi kualitas pelayanan kesehatan menjadi penting guna memastikan bahwa setiap individu mendapatkan akses terhadap layanan yang memadai, aman, dan efektif. Mutu pelayanan di fasilitas kesehatan perlu ditingkatkan, salah satu kunci utamanya adalah pelayanan yang diberikan oleh tenaga kesehatan [1]. Kota Surabaya, sebagai ibu kota di Provinsi Jawa Timur,



memiliki peran yang signifikan dalam menyediakan pelayanan kesehatan yang berkualitas kepada penduduknya.

Fasilitas kesehatan merupakan sarana pelayanan kesehatan tingkat pertama yang berperan sebagai tonggak pelayanan awal, yaitu menjadi gerbang pelayanan kesehatan tingkat lanjut terutama di era jaminan kesehatan nasional seperti saat ini [2]. Studi kasus mengenai evaluasi kualitas pelayanan kesehatan pada fasilitas kesehatan khususnya puskesmas dan rumah sakit di Kota Surabaya menjadi relevan untuk mengetahui sejauh mana tingkat kualitas pelayanan yang saat ini tersedia berdasarkan jumlah tenaga kesehatan.

Evaluasi merupakan bagian penting dari proses manajemen karena evaluasi memberikan umpan balik terhadap pelaksanaan program atau kegiatan. Menurut American Public Health Association, evaluasi adalah proses penentuan nilai atau tingkat keberhasilan dalam mencapai tujuan tertentu. Menurut WHO, pemantauan dan evaluasi dapat dilakukan dan harus berkelanjutan [3]. Evaluasi merupakan alat pengambilan keputusan yang penting pada tingkat perumusan kebijakan dan pelaksanaan program. Evaluasi ini dapat memberikan wawasan mendalam tentang kekuatan dan kelemahan sistem pelayanan kesehatan di kota ini, serta menyoroti area yang memerlukan perbaikan.

Penelitian ini bertujuan untuk melakukan evaluasi kualitas pelayanan kesehatan di Kota Surabaya dengan fokus pada puskesmas dan rumah sakit. Dalam penelitian ini, akan dilakukan analisis kuantitatif berdasarkan data yang tersedia dari berbagai sumber, termasuk Badan Pusat Statistik (BPS) dan data internal dari puskesmas dan rumah sakit terkait. Terkait banyaknya akses data dari berbagai sumber website, menyulitkan pekerja dalam mendapatkan dan menganalisis data dengan baik. *Microsoft Excel* mempermudah pekerjaan dan menggabungkan data dari beberapa sumber website (*Join table*), khususnya di bidang pelayanan kesehatan, dan kualitas Kesehatan.

Hasil-hasil studi sebelumnya menunjukkan bahwa aplikasi Microsoft Excel sangat membantu para pengusaha kecil dan menengah dalam mengelola transaksi keuangan mereka. Studi yang dilakukan oleh Achadiyah (2019), menunjukkan bahwa aplikasi Microsoft Excel dapat digunakan untuk otomatisasi kegiatan akuntansi. Lebih lanjut, Achadiyah menemukan para pemilik usaha sangat terbantu dalam mengelola keuangan mereka. Selain itu, Microsoft Excel mampu melakukan perhitungan-perhitungan numerik baik dengan operasi-operasi aritmetika biasa maupun dengan fungsi-fungsi matematika, termasuk fungsi-fungsi statistika sederhana (statistiks deskriptif). Dengan demikian, MS Excel dapat digunakan untuk analisis data statistiks. Akan tetapi, untuk keperluan analisis data statistik yang lebih rumit, pemakaian fungsi-fungsi matematika dasar dan statistika (*built-in*) memerlukan perhitungan manual yang cukup panjang., pemilik usaha sangat terbantu dalam proses pengambilan keputusan (Achadiyah, 2019) [4]. Adapun hasil penelitian sebelumnya oleh mahasiswa Kesehatan Masyarakat mengenai Kualitas Pelayanan Kesehatan di Puskesmas Ciampea Kabupaten Bogor Tahun 2019 dengan menerapkan analisis multivariat dengan menggunakan alat bantu statistik excel sebagai analisis data [5].

Hasil penelitian ini akan menjadi sumber informasi penting untuk pemangku kebijakan dan praktisi kesehatan dalam meningkatkan kualitas pelayanan kesehatan di tingkat regional. Pentingnya evaluasi kualitas pelayanan kesehatan di Kota Surabaya tidak hanya berkontribusi pada peningkatan kesehatan masyarakat secara keseluruhan, tetapi juga berpotensi untuk meningkatkan keberlanjutan sistem pelayanan kesehatan di masa depan. Dengan mengevaluasi kualitas pelayanan saat ini, akan tercipta landasan yang kuat untuk mengidentifikasi kebutuhan dan merumuskan langkah-langkah strategis yang diperlukan dalam meningkatkan aksesibilitas, efisiensi, dan kepuasan pasien.

Kesimpulannya, evaluasi kualitas pelayanan kesehatan di Kota Surabaya melalui studi kasus pada puskesmas dan rumah sakit merupakan langkah penting dalam memahami kondisi dan tantangan yang dihadapi dalam penyediaan pelayanan kesehatan. Penelitian ini diharapkan dapat memberikan masukan berharga untuk perbaikan sistem pelayanan kesehatan di kota ini.

## II. METODE PENELITIAN

### 2.1. Metode

#### 3.1.1 Statistika Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu himpunan data sehingga memberikan informasi yang berguna [6]. Contoh statistika deskriptif yang sering muncul adalah, tabel, diagram, grafik, dan besaran-besaran lain di majalah dan koran-koran. Dengan statistika deskriptif, kumpulan data yang diperoleh akan tersaji dengan ringkas dan rapi serta dapat memberikan informasi inti dari kumpulan data yang ada.

Metode penelitian ini menerapkan metode statistika deskriptif untuk menganalisis data yang diperoleh dari Badan Pusat Statistik (BPS) mengenai evaluasi kualitas pelayanan kesehatan di Kota Surabaya. Data tersebut meliputi variabel jumlah fasilitas kesehatan, jumlah tenaga kesehatan, persentase angka sakit, dan persentase alasan masyarakat terhadap fasilitas kesehatan. Metode statistika deskriptif ini membantu dalam memberikan pemahaman yang komprehensif tentang situasi pelayanan kesehatan di Kota Surabaya berdasarkan data yang tersedia dari BPS.

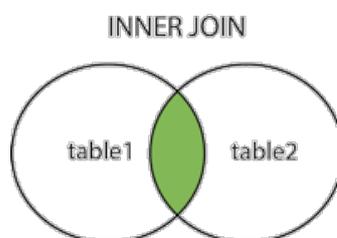
Metode penelitian yang digunakan merupakan metode *web scraping* dengan membangun fungsi implementasi dari *text mining* untuk dapat menganalisis komentar yang berbentuk ujaran kebencian pada media sosial youtube yang berada di salah satu konten Surya Citra Televisi (SCTV) yang mengarah terhadap Ketua DPR RI Puan Maharani. Rancangan ini juga dapat dilihat pada gambar 1 yang mengilustrasikan seluruh metode pada penelitian yang terdiri dari beberapa proses, yaitu *scraping data*, *cleaning data*, *removing unnecessary data*, dan hasil akhir dari pengolahan data.

**Tabel 1.** Tabel statistika deskriptif fasilitas kesehatan dan tenaga kesehatan

Statistika deskriptif	Rata	Median	Standar Deviasi	Varians	Kurtosis	Skewness	Range	Min	Maks
Fasilitas Kesehatan	5,87	5	2,59	6,71	0,16	0,77	10	2	12
Tenaga Kesehatan	26,41	25	12,51	156,71	-0,28	0,76	42	11	53

#### 2.1.2 Join Tablee

Metode penelitian ini menerapkan metode *inner join*. *Inner join* digunakan untuk menghasilkan baris data dengan cara menggabungkan 2 buah tabel atau lebih menggunakan pasangan data yang match pada masing-masing tabel [7]. Dengan adanya metode tersebut, penulis lebih mudah mengintegrasikan data dari berbagai sumber yang relevan dalam evaluasi kualitas pelayanan kesehatan di Kota Surabaya. Data yang digunakan mencakup informasi tentang jumlah fasilitas kesehatan dan jumlah tenaga kesehatan.



**Gambar 1.** Gambar *inner join*

Dalam metode ini, tabel-tabel yang berbeda akan digabungkan berdasarkan kunci atau variabel yang sama untuk menyatukan informasi yang saling terkait. Melalui penggabungan tabel, akan tercipta satu set data yang lengkap dan terintegrasi yang memungkinkan analisis yang lebih komprehensif. Metode ini memungkinkan peneliti untuk mengidentifikasi hubungan dan pola yang lebih kompleks antara variabel-variabel yang diamati, serta mendapatkan wawasan yang lebih



mendalam tentang kualitas pelayanan kesehatan di Kota Surabaya berdasarkan data yang tersedia dari berbagai sumber yang relevan.

**Tabel 2.** Tabel fasilitas kesehatan

No	Kecamatan	Jenis Faskes	Penyelenggara Faskes	Nama Faskes
1	Asemrowo	Puskesmas	Pemerintah Kota	Puskesmas Asemrowo
2	Asemrowo	Puskesmas	Pemerintah Kota	Pustu Tambak Langon
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
181	Wonokromo	Puskesmas	Pemerintah Kota	Puskesmas Wonokromo
182	Wonokromo	Puskesmas	Pemerintah Kota	Pustu Sawunggaling

**Tabel 3.** Tabel tenaga kesehatan

Id	Kecamatan	Dokter	Perawat	Total Tenaga Kesehatan
1	Asemrowo	6	5	11
2	Benowo	8	9	17
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
30	Wonocolo	21	17	38
31	Wonokromo	24	27	51

**Tabel 4.** Tabel persentase angka sakit di provinsi Jawa Timur Tahun 2020

Kabupaten/Kota Se Jawa Timur	Tahun	Angka Kesakitan Menurut Kabupaten/Kota (Persen)
Kabupaten Pacitan	2020	21,13
Kabupaten Ponorogo	2020	31,24
:	:	:
:	:	:
:	:	:
Kota Surabaya	2020	19,32
Kota Batu	2020	27,81

**Tabel 5.** Tabel persentase angka sakit di kota Surabaya

Alasan	Proporsi (%)
Mengobati sendiri	89,66%
mendampingi	5,60%
Merasa tidak perlu	2,10%
Lainnya	2,63%

**Tabel 6.** Tabel pivot fasilitas kesehatan

Id	Kecamatan	Jumlah Fasilitas Kesehatan
1	Asemrowo	2
2	Benowo	5
:	:	:
:	:	:
:	:	:

30	Wonocolo	7
31	Wonokromo	12

Pada metode ini data yang dilakukan penggabungan (*join*) yakni tabel fasilitas kesehatan dan tenaga kesehatan berdasarkan id kecamatan yang dijadikan sebagai *primary key*.

**Tabel 7.** Tabel join fasilitas kesehatan dan tenaga kesehatan

<b>Id</b>	<b>Kecamatan</b>	<b>Jumlah Fasilitas Kesehatan</b>	<b>Total Tenaga Kesehatan</b>
1	Asemrowo	2	11
2	Benowo	5	17
:	:	:	:
:	:	:	:
:	:	:	:
30	Wonocolo	7	38
31	Wonokromo	12	51

### 3.1.3 Analisis Regresi

Regresi linier adalah salah satu dari jenis analisis peramalan atau prediksi yang sering digunakan pada data berskala kuantitatif (interval atau rasio) [8].

Persamaan regresi dituliskan dalam bentuk yang paling sederhana yaitu satu peubah bebas (X) dengan satu peubah tak bebas (Y) mempunyai persamaan:

$$Y_i = a_0 + bx \quad (1)$$

Jika data hasil pengamatan Y terjadi karena akibat variabel-variabel bebas  $x_1, x_2, \dots, x_k$ , maka hal ini menggunakan analisis regresi berganda. Model regresi linear ganda Y atas  $x_1, x_2, \dots, x_k$  akan ditaksir oleh:  $\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$ . Dengan  $a_0, a_1, \dots, a_k$  merupakan koefisien-koefisien yang harus ditentukan berdasarkan data hasil pengamatan. Persamaan tersebut berisikan  $k + 1$  buah koefisien, sehingga apabila koefisien-koefisien  $a_0, a_1, \dots, a_k$  ditentukan dengan metode kuadrat terkecil, maka diperlukan penyelesaian sistem persamaan yang terdiri atas  $k + 1$  buah persamaan. Koefisien-koefisien tersebut disebut koefisien regresi parsial, karena hanya memberikan gambaran parsial apa yang terjadi pada Y untuk perubahan X yang berhubungan dengan koefisien dimaksud. Pengujian signifikansi koefisien-koefisien regresi secara parsial akan digunakan statistik uji berdistribusi t (student), sedangkan secara menyeluruh (simultan) akan menggunakan statistik uji berdistribusi F (fisher) dengan kritikal error  $\alpha = 0,05$  atau derajat kepercayaan  $1 - \alpha = 0,95$  [9].

Metode analisis regresi digunakan dalam penelitian ini untuk mengidentifikasi hubungan antara variabel independen dan variabel dependen dalam konteks evaluasi kualitas pelayanan kesehatan di Kota Surabaya. Dalam penelitian ini, variabel independen meliputi jumlah fasilitas kesehatan, jumlah tenaga kesehatan, persentase angka sakit, dan persentase alasan masyarakat terhadap fasilitas kesehatan.

### 3.1.4 Hubungan Fasilitas Kesehatan dan Tenaga Kesehatan

Dalam penelitian ini, dilakukan analisis untuk mengevaluasi hubungan antara fasilitas kesehatan dan tenaga kesehatan dalam konteks evaluasi kualitas pelayanan kesehatan. Tujuan dari analisis ini adalah untuk memahami sejauh mana jumlah fasilitas kesehatan dan jumlah tenaga kesehatan berperan dalam meningkatkan kualitas pelayanan kesehatan di Kota Surabaya. Penambahan fasilitas kesehatan dapat meningkatkan aksesibilitas pelayanan kesehatan bagi masyarakat, sedangkan peningkatan jumlah tenaga kesehatan dapat meningkatkan kapasitas dan ketersediaan layanan kesehatan. Keberadaan fasilitas kesehatan yang memadai dan tenaga kesehatan yang mencukupi menjadi faktor krusial dalam memberikan pelayanan yang memadai, efektif, dan aman bagi pasien.

## 2.2. Data

Data yang digunakan dalam penelitian ini berasal dari Badan Pusat Statistik (BPS) Kota Surabaya. Data ini mencakup informasi tentang jumlah fasilitas kesehatan, jumlah tenaga kesehatan, persentase angka sakit, dan persentase alasan masyarakat terhadap fasilitas kesehatan. Terdapat dua data yang telah diolah sebelumnya dengan menggunakan metode *join* yakni dari data yang terkait fasilitas kesehatan dan tenaga kesehatan.

**Tabel 8.** Tabel join fasilitas kesehatan dan tenaga kesehatan

<b>Id</b>	<b>Kecamatan</b>	<b>Jumlah Fasilitas Kesehatan</b>	<b>Total Tenaga Kesehatan</b>
1	Asemrowo	2	11
2	Benowo	5	17
:	:	:	:
:	:	:	:
:	:	:	:
30	Wonocolo	7	38
31	Wonokromo	12	51

**Tabel 9.** Tabel persentase angka sakit di provinsi Jawa Timur

<b>Kabupaten/Kota Se Jawa Timur</b>	<b>Tahun</b>	<b>Angka Kesakitan Menurut Kabupaten/Kota (Persen)</b>
Kabupaten Pacitan	2020	21,13
Kabupaten Ponorogo	2020	31,24
:	:	:
:	:	:
:	:	:
Kota Surabaya	2020	19,32
Kota Batu	2020	27,81

**Tabel 10.** Tabel persentase angka sakit di kota Surabaya

<b>Alasan</b>	<b>Proporsi (%)</b>
Mengobati sendiri	89,66%
mendampingi	5,60%
Merasa tidak perlu	2,10%
Lainnya	2,63%

Variabel penelitian yang akan diamati meliputi jumlah fasilitas kesehatan dan jumlah tenaga kesehatan dalam satu tabel, serta persentase angka sakit dan persentase alasan masyarakat terhadap fasilitas kesehatan yang meliputi masing-masing tabel.

**Tabel 11.** Tabel variabel

<b>Variabel</b>	<b>Tabel</b>	<b>Keterangan</b>
Kecamatan	1	Nama kecamatan di Kota Surabaya
Total Tenaga Kesehatan	1	Jumlah tenaga kesehatan (dokter, perawat)
Jumlah Fasilitas Kesehatan	1	Jumlah fasilitas kesehatan (puskesmas, rumah sakit)
Kabupaten/Kota Se Jawa Timur	2	Nama kabupaten/kota di provinsi Jawa Timur
Tahun	2	Waktu data terambil
Angka Kesakitan Menurut Kabupaten/Kota (%)		Proporsi angka sakit masyarakat

Alasan	3	Alasan masyarakat tidak memilih untuk berkunjung ke fasilitas kesehatan
Proporsi (%)	3	Proporsi alasan masyarakat

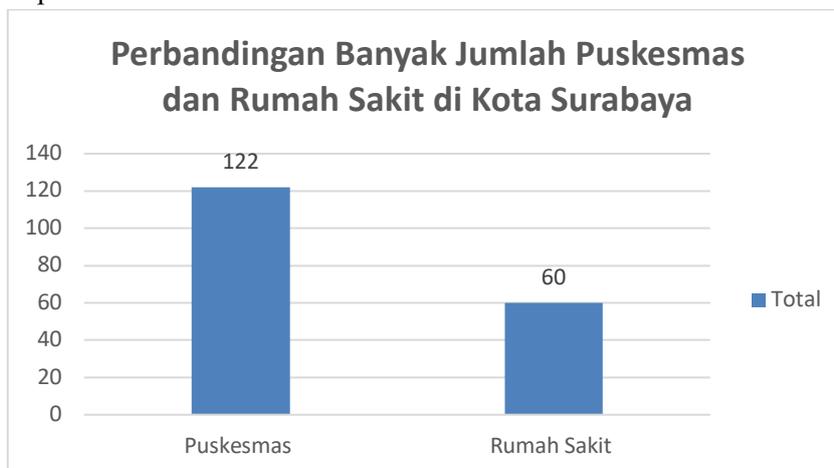
### 2.3. Tahap Analisis

Pada tahap ini pertama dilakukan pengumpulan data tentang jumlah fasilitas kesehatan, jumlah tenaga kesehatan, proporsi angka sakit berdasarkan persentase, dan proporsi alasan masyarakat berdasarkan persentase. Metode yang digunakan yakni metode statistika deskriptif pada data tenaga kesehatan untuk mengetahui kecenderungan arah pada grafik histogram dengan bantuan excel sebagai analisis multivariat. Kemudian, melakukan *join table* dengan menggabungkan data yang berbeda yakni tabel fasilitas kesehatan dan tenaga kesehatan untuk mencakup informasi lebih berdasarkan Id Kecamatan sebagai kunci atau variabel yang sama untuk menyatukan informasi yang saling terkait. Setelah dilakukan analisis, selanjutnya membuat visualisasi *dashboard* dari informasi yang dibutuhkan seperti membandingkan jumlah fasilitas kesehatan yang terdiri dari puskesmas dan rumah sakit, membandingkan jumlah tenaga kesehatan yang terdiri dari dokter dan perawat, serta visualisasi dari proporsi angka sakit dan alasan masyarakat. Pada *dashboard* juga ditampilkan hasil perhitungan dari analisis korelasi dan analisis regresi dari fasilitas kesehatan terhadap tenaga kesehatan.

## III. HASIL DAN PEMBAHASAN

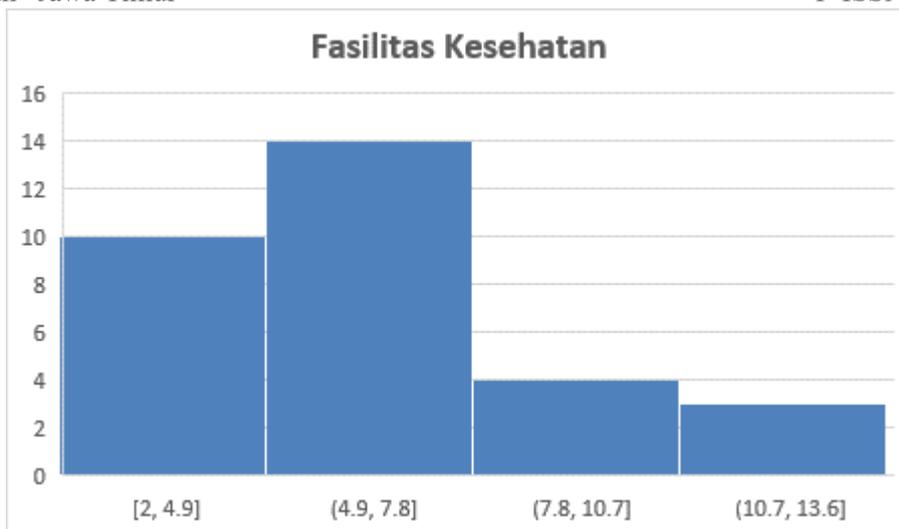
### 4.1 Deskripsi Data:

Jumlah Fasilitas Kesehatan: Berdasarkan data BPS, Kota Surabaya memiliki total 182 fasilitas kesehatan yang diantaranya terdapat 122 puskesmas dan 60 rumah sakit yang menyediakan pelayanan kesehatan kepada penduduk.



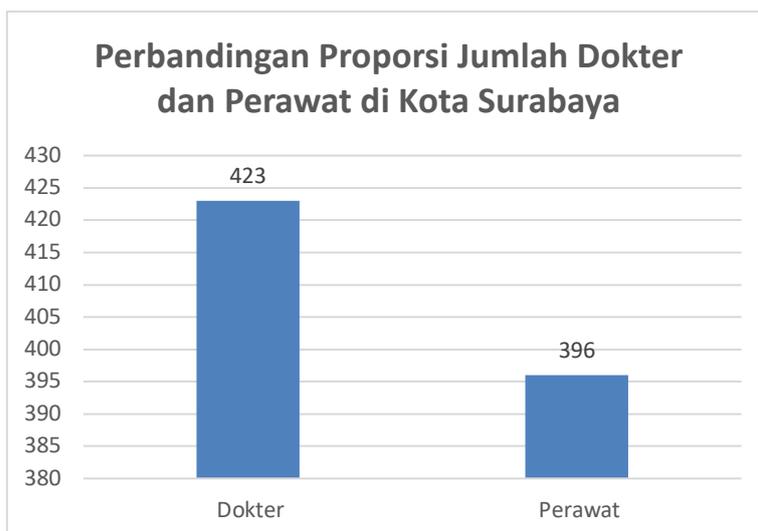
**Gambar 2.** Grafik perbandingan banyak jumlah puskesmas dan rumah sakit di Kota Surabaya

Setelah dilakukan analisis, diketahui grafik jumlah fasilitas kesehatan di kota Surabaya yaitu skewness positif yang berarti sebagian besar jumlah fasilitas kesehatan di Kota Surabaya berada di bawah rata-rata, tetapi ada sejumlah kecil kecamatan dengan jumlah fasilitas kesehatan yang sangat banyak.



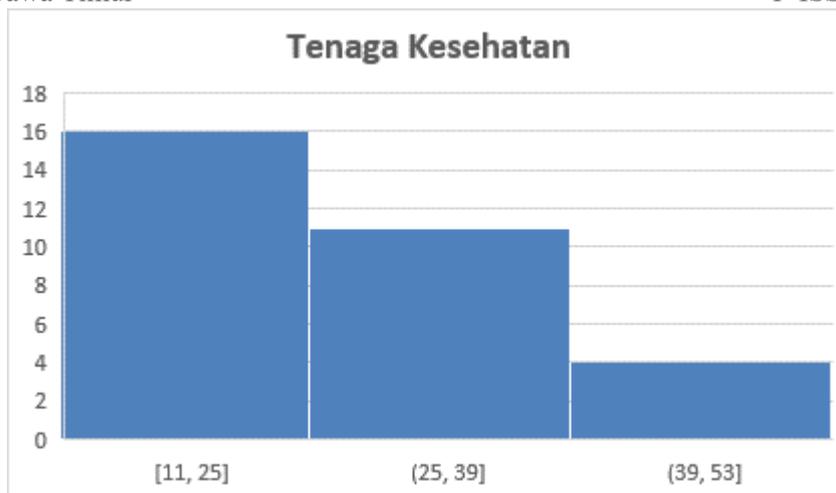
**Gambar 3.** Grafik histogram fasilitas kesehatan

Jumlah Tenaga Kesehatan: Data menunjukkan bahwa terdapat 819 tenaga kesehatan yang diantaranya terdiri dari 423 dokter dan 396 perawat yang bekerja di puskesmas dan rumah sakit di Kota Surabaya.



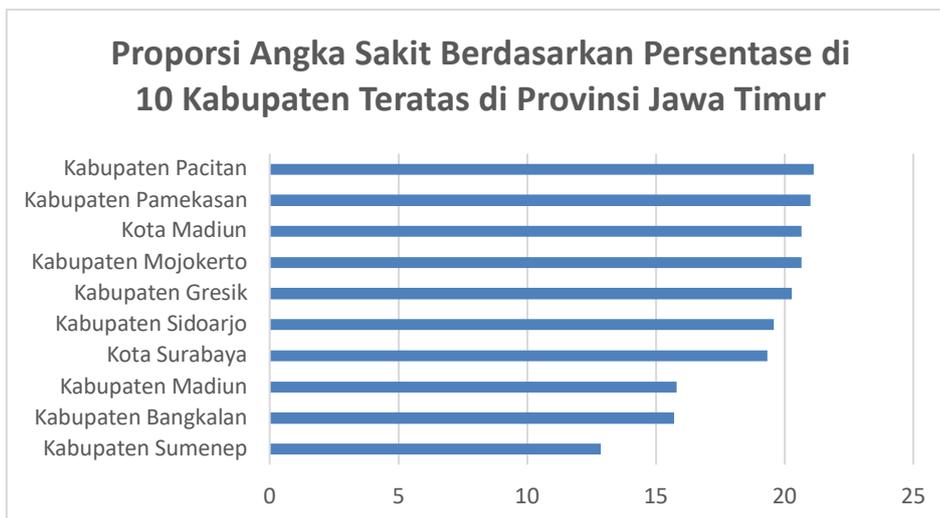
**Gambar 4.** Grafik perbandingan proporsi jumlah dokter dan perawat

Setelah dilakukan analisis, diketahui grafik jumlah tenaga kesehatan di kota Surabaya yaitu skewness positif yang berarti sebagian besar jumlah tenaga kesehatan di kota Surabaya berada di bawah rata-rata, tetapi ada sejumlah kecil kecamatan dengan jumlah tenaga kesehatan yang sangat banyak.



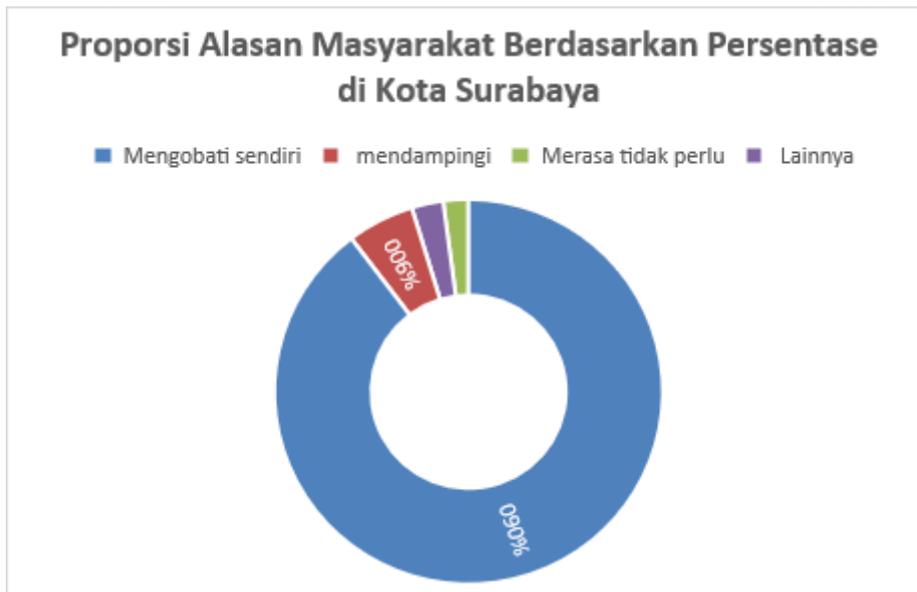
**Gambar 5.** Grafik histogram tenaga kesehatan

Persentase Angka Sakit: Berdasarkan data BPS, persentase angka sakit di Kota Surabaya mencapai 19% di provinsi Jawa Timur, menunjukkan adanya beban penyakit yang signifikan pada masyarakat.



**Gambar 6.** Grafik perbandingan proporsi angka sakit berdasarkan persentase di provinsi Jawa Timur

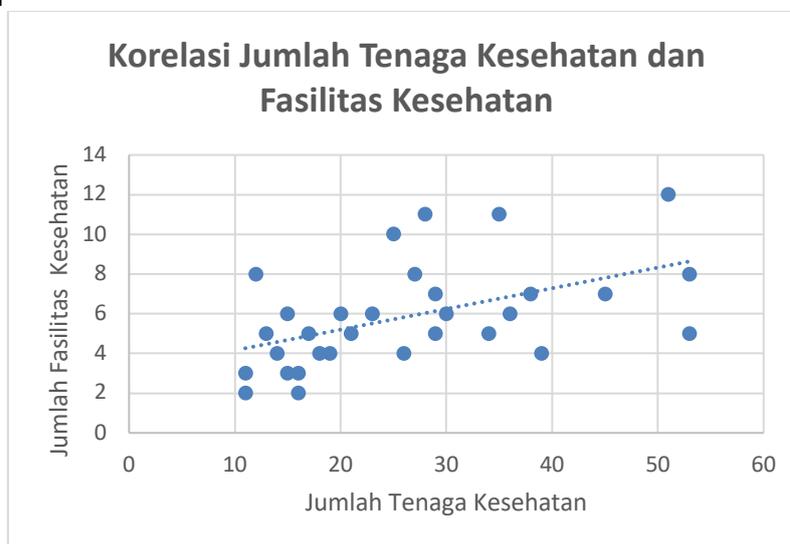
Persentase Alasan Masyarakat terhadap Fasilitas Kesehatan: Data menunjukkan bahwa 89,66% masyarakat lebih memilih untuk mengobati sendiri, 5,6% masyarakat sudah didampingi, 2,1% masyarakat merasa tidak perlu untuk berkunjung ke fasilitas kesehatan, dan 2,63% alasan lainnya.



**Gambar 7.** Visualisasi perbandingan proporsi alasan masyarakat berdasarkan persentase di Kota Surabaya

#### 4.2 Analisis Korelasi Fasilitas Kesehatan terhadap Tenaga Kesehatan:

Hubungan antara Jumlah Fasilitas Kesehatan dan Jumlah Tenaga Kesehatan: Hasil analisis korelasi menunjukkan bahwa terdapat hubungan positif yang signifikan antara fasilitas kesehatan dan tenaga kesehatan dalam meningkatkan kualitas pelayanan kesehatan. Pada kondisi nyata yang terjadi di kota Surabaya tahun 2020, pemerintah juga harus menghitung kemampuan dan kapasitas tenaga kesehatan ketika menambah ruang-ruang perawatan. Mengingat sejumlah dokter pun mulai bertumbuhan [10].



**Gambar 8.** Visualisasi korelasi jumlah tenaga kesehatan dan fasilitas kesehatan

**Tabel 12.** Tabel analisis korelasi

	Tekes (y)	Faskes (x)
Tekes (y)	1	
Faskes (x)	0,503	1

Dari hasil tabel 12 menunjukkan bahwa korelasi positif kuat dengan nilai 0,5 yang artinya semakin banyak fasilitas kesehatan yang tersedia, maka semakin besar jumlah tenaga kesehatan yang tersedia, maka kemungkinan terciptanya pelayanan kesehatan yang berkualitas juga semakin tinggi.

#### 4.3 Analisis Regresi Fasilitas Kesehatan terhadap Tenaga Kesehatan: rumus tabel

Pengaruh Jumlah Fasilitas Kesehatan terhadap Jumlah Tenaga Kesehatan terhadap Persentase Angka Sakit: Analisis regresi menunjukkan bahwa jumlah fasilitas kesehatan dan jumlah tenaga kesehatan secara signifikan mempengaruhi persentase angka sakit di Provinsi Jawa Timur. Penambahan fasilitas kesehatan dan tenaga kesehatan yang memadai berpotensi mengurangi angka sakit di masyarakat.

**Tabel 13.** Tabel anova

	df	SS	MS	F	Significance F
Regression	1	51,001	51,001	9,828	0,0039
Residual	29	150,482	5,189		
Total	30	201,483			

Dari hasil tabel 14 menunjukkan bahwa nilai signifikan F sebesar 0,0039 yang mana kurang dari 0,05 sehingga dapat diputuskan tolak  $H_0$ , maka dapat disimpulkan bahwa secara serentak variabel dalam penelitian berpengaruh terhadap fasilitas kesehatan.

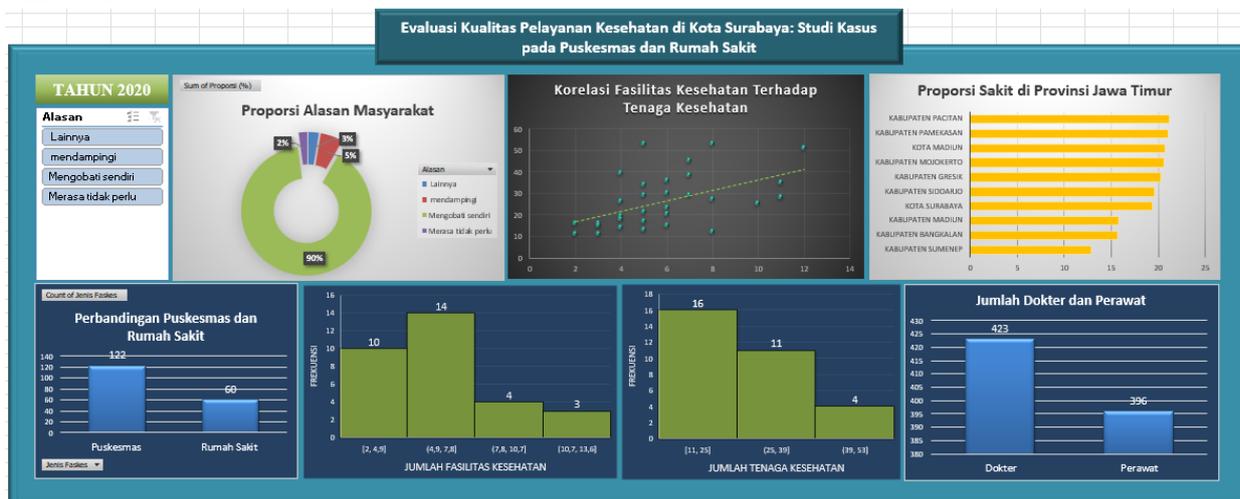
**Tabel 14.** Tabel *coefficient beta*

	Coefficients	Standard Error	t Stat	P-value
Intercept	3,119	0,968	3,221	0,003
Tenaga Kesehatan	0,104	0,033	3,135	0,003

Tabel 14 menunjukkan bahwa tenaga kesehatan berpengaruh positif terhadap fasilitas kesehatan ditunjukkan dengan nilai p-value sebesar 0,003 kurang dari 0,05. Hal ini menunjukkan bahwa dengan meningkatnya tenaga kesehatan sebanyak 100 orang, maka fasilitas kesehatan dapat bertambah 10 sarana ataupun prasarana di kecamatan yang ada di Surabaya tanpa dipengaruhi variabel yang lainnya.

#### 4.4 Dashboard

Berikut disajikan dashboard evaluasi kualitas Kesehatan di Kota Surabaya. Pemantauan data historis akan lebih mudah apabila menggunakan visualisasi data dan analisis menggunakan bantuan *dashboard*.



Gambar 9. Dashboard

## IV. KESIMPULAN

Berdasarkan penelitian kuantitatif yang dilakukan menggunakan data dari BPS tentang Evaluasi Kualitas Pelayanan Kesehatan di Kota Surabaya dengan studi kasus pada Puskesmas dan Rumah Sakit, diperoleh beberapa kesimpulan penting sebagai berikut:

1. Evaluasi menunjukkan bahwa kualitas pelayanan kesehatan di Kota Surabaya masih perlu ditingkatkan. Meskipun terdapat jumlah fasilitas kesehatan dan tenaga kesehatan yang



memadai, terdapat temuan yang mengindikasikan adanya kekurangan dalam beberapa aspek kualitas pelayanan.

2. Analisis menunjukkan adanya hubungan yang signifikan antara jumlah fasilitas kesehatan dan jumlah tenaga kesehatan dengan kualitas pelayanan kesehatan. Semakin banyak fasilitas kesehatan dan tenaga kesehatan yang tersedia, maka cenderung meningkat pula kualitas pelayanan.
3. Persentase angka sakit dan alasan masyarakat terhadap fasilitas kesehatan juga mempengaruhi kualitas pelayanan kesehatan. Temuan ini menunjukkan bahwa faktor-faktor seperti tingkat penyakit di masyarakat dan alasan-alasan individu untuk tidak memanfaatkan fasilitas kesehatan berdampak pada kualitas pelayanan.

Dalam rangka meningkatkan kualitas pelayanan kesehatan di Kota Surabaya, rekomendasi perbaikan yang dapat dilakukan antara lain adalah peningkatan jumlah fasilitas kesehatan dan tenaga kesehatan sesuai dengan kebutuhan, meningkatkan sosialisasi dan edukasi mengenai pentingnya pemanfaatan fasilitas kesehatan kepada masyarakat, serta memperkuat sistem pemantauan dan evaluasi kualitas pelayanan secara berkelanjutan.

## UCAPAN TERIMA KASIH

## REFERENSI

1. D. Ayatulloh, N. Nursalam, and N. D. Kurniawati, “The Effect of Knowledge Management in Healthcare Services: A Systematic Review,” *Jurnal Pendidikan Keperawatan Indonesia*, vol. 7, no. 1, 2021.
2. R. N. Karimah and A. P. Wicaksono, “Prototype Sistem Informasi Pelayanan Bayi Baru Lahir pada Fasilitas Kesehatan Primer,” *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, 2018, doi: 10.23917/khif.v4i1.5330.
3. kesmas, “Evaluasi Bidang Kesehatan,” <http://www.indonesian-publichealth.com/evaluasi-bidang-kesehatan/>, Mar. 22, 2014.
4. W. Weli, R. Pambudi, E. Meilia, and D. Joshua, “Optimizing the Use of Microsoft Excel for Employees’ Administrative Tasks at Le Premier Hotel Kota Deltamas, Cikarang,” *MITRA: Jurnal Pemberdayaan Masyarakat*, vol. 5, no. 1, 2021, doi: 10.25170/mitra.v5i1.1347.
5. D. Erdini, E. Dwimawati, and I. Chotimah, “EVALUASI PROGRAM TB PARU DI PUSKESMAS CIAMPEA KABUPATEN BOGOR TAHUN 2019,” *PROMOTOR*, vol. 3, no. 5, 2020, doi: 10.32832/pro.v3i5.4202.
6. Saddam Hussein, “Statistika Deskriptif: Pengertian, Fungsi, Perbedaan Statistika Deskriptif dan Inferensial, dan Contoh Perhitungannya,” <https://geospasialis.com/statistika-deskriptif/>, Jun. 10, 2020.
7. MUJIHARTO PANGA, “SINTAX SQL, INNER JOIN, LEFT JOIN, RIGHT JOIN, AND FULL JOIN,” [https://mahasiswa.ung.ac.id/921411144/home/2012/12/6/syntax\\_sql\\_inner\\_join\\_left\\_join\\_right\\_join\\_and\\_full\\_join.html](https://mahasiswa.ung.ac.id/921411144/home/2012/12/6/syntax_sql_inner_join_left_join_right_join_and_full_join.html), Dec. 06, 2012.
8. Anwar Hidayat, “Pengertian Analisis Regresi Korelasi Dan Cara Hitung ,” <https://www.statistikian.com/2012/08/analisis-regresi-korelasi.html>, May 21, 2020.
9. H. Basri, “PEMODELAN REGRESI BERGANDA UNTUK DATA DALAM STUDI KECERDASAN EMOSIONAL.”



10. BBC NEWS INDONESIA, “Covid: ‘Fungsi fasilitas kesehatan kolaps dan nakes kelelahan’, apakah cukup dijawab dengan penambahan tempat tidur perawatan?,” <https://www.bbc.com/indonesia/indonesia-57878074>, Jul. 2020



# Analisis Perubahan Tingkat Pengangguran di Kabupaten/Kota Kalimantan Barat Tahun 2010-2018

Arindra Harris Abdillah<sup>1</sup>, Muchamad Risqi<sup>2</sup>, Trimono, S.Si., M.Si<sup>3</sup>,

<sup>1,2,3</sup> Program Studi Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

<sup>1</sup>[22083010023@student.upnjatim.ac.id](mailto:22083010023@student.upnjatim.ac.id)

<sup>3</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

Corresponding author email: [22083010029@student.upnjatim.ac.id](mailto:22083010029@student.upnjatim.ac.id)

**Abstract:** This article discusses the changes in the unemployment rate in West Kalimantan's districts and cities from 2010 to 2018. The analysis aims to study the variations in the unemployment rate during this period, utilizing data from each area and applying statistical approaches to identify influencing factors. The results show significant fluctuations in the unemployment rate across the region. Some areas experienced notable decreases, while others saw increases or insignificant changes.

Potential factors influencing the unemployment rate include economic growth and labor policies. The findings have important implications for local governments and stakeholders in formulating strategies to reduce unemployment. Collaboration between the government, private sector, and educational institutions is crucial to improve job opportunities and strengthen workforce training programs. In conclusion, this analysis provides insights into the changes in the unemployment rate in West Kalimantan's districts and cities from 2010 to 2018 and the factors that contribute to these changes. The article serves as a basis for effective decision-making in addressing unemployment challenges, guiding policy formulation and interventions in the region.

**Keywords:** district/city, unemployment, open, level

**Abstrak:** Artikel ini membahas tentang perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat Tahun 2010-2018. Analisis ini bertujuan untuk mempelajari perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat dari tahun 2010 hingga 2018. Penelitian ini menggunakan data tingkat pengangguran di setiap Kabupaten/Kota sebagai sumber informasi dan menerapkan pendekatan statistik serta analisis untuk mengidentifikasi faktor-faktor yang mempengaruhi perubahan tersebut. Hasil analisis menunjukkan variasi yang signifikan dalam tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode penelitian. Beberapa daerah mengalami penurunan tingkat pengangguran yang signifikan, sementara daerah lain mengalami kenaikan atau perubahan yang tidak signifikan.

Faktor-faktor yang berpotensi mempengaruhi perubahan tingkat pengangguran antara lain pertumbuhan ekonomi dan kebijakan ketenagakerjaan. Hasil penelitian ini memiliki implikasi penting bagi pemerintah daerah dan stakeholder terkait dalam merumuskan kebijakan untuk mengurangi tingkat pengangguran di Kabupaten/Kota Kalimantan Barat. Diperlukan kerjasama antara pemerintah, sektor swasta, dan lembaga pendidikan dalam meningkatkan peluang kerja dan menguatkan program pelatihan tenaga kerja guna mengatasi masalah pengangguran. Kesimpulannya, analisis ini memberikan wawasan yang lebih baik tentang perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode 2010-2018 serta faktor-faktor yang berperan dalam perubahan tersebut. Artikel ini diharapkan dapat menjadi dasar untuk pengambilan keputusan yang lebih efektif dalam menghadapi tantangan pengangguran di wilayah tersebut.

**Kata kunci:** media sosial, ujaran kebencian, komentar youtube

## I. PENDAHULUAN

Tingkat pengangguran adalah salah satu ukuran penting untuk mengevaluasi situasi ketenagakerjaan suatu daerah. Kabupaten/Kota Kalimantan Barat, yang terletak di Kalimantan, Indonesia, memiliki potensi ekonomi yang sangat menjanjikan. Dalam rangka memahami dinamika ketenagakerjaan dan perekonomian di wilayah ini, penting untuk menganalisis perubahan tingkat pengangguran dari tahun 2010 hingga 2018. Penelitian ini bertujuan untuk melakukan analisis terhadap perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama delapan tahun terakhir. Data yang digunakan dalam penelitian ini diperoleh dari dataset Tingkat Pengangguran Terbuka (TPT) menurut Kabupaten/Kota yang disediakan oleh Pemerintah Provinsi Kalimantan Barat melalui situs [data.kalbarprov.go.id](http://data.kalbarprov.go.id).

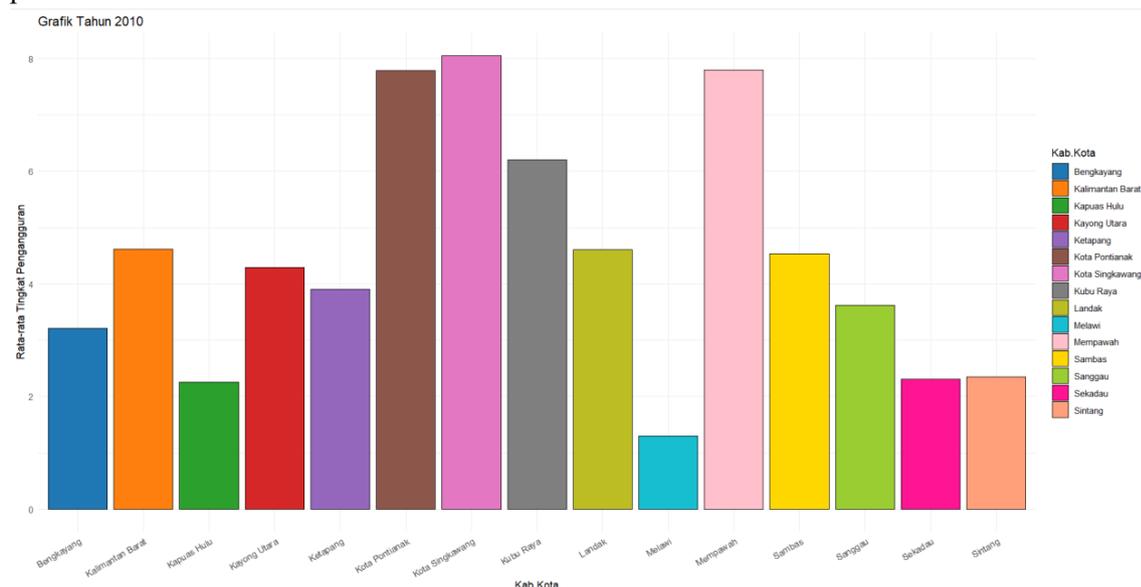
Melalui analisis ini, diharapkan dapat diidentifikasi persentase dan pola perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode yang diteliti. Selain itu, penelitian ini juga akan mengidentifikasi urutan daerah yang mengalami tingkat pengangguran dari terendah hingga tertinggi. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang kondisi ketenagakerjaan di Kabupaten/Kota Kalimantan Barat dari tahun 2010 hingga 2018. Temuan-temuan ini akan menjadi dasar untuk merumuskan kebijakan dan strategi yang efektif dalam mengatasi masalah pengangguran, meningkatkan peluang kerja, dan mendorong pertumbuhan ekonomi yang berkelanjutan di wilayah ini.

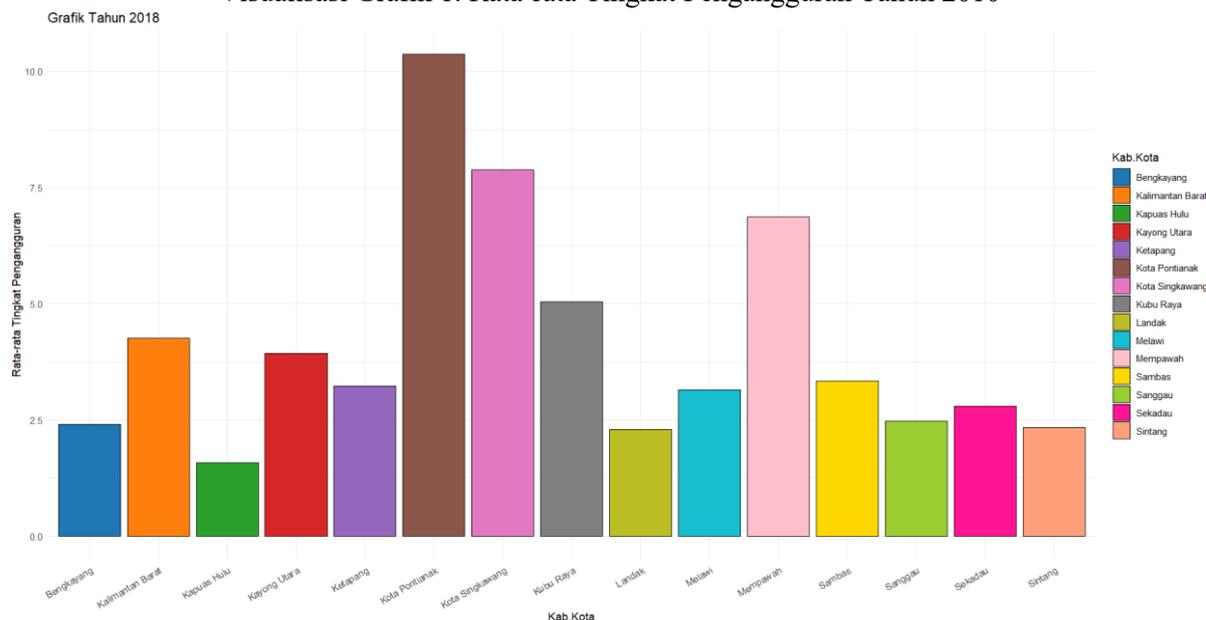
Penelitian ini memiliki relevansi yang penting bagi pemerintah daerah, kalangan akademisi, dan pemangku kepentingan terkait, dalam upaya meningkatkan kualitas hidup masyarakat, mengurangi tingkat pengangguran, serta memperkuat pembangunan ekonomi di Kabupaten/Kota Kalimantan Barat. Dengan pemahaman yang lebih mendalam tentang perubahan tingkat pengangguran dan faktor-faktor yang mempengaruhinya, diharapkan dapat diambil langkah-langkah yang efektif dalam menghadapi tantangan dan memanfaatkan peluang di sektor ketenagakerjaan wilayah ini.

## II. KAJIAN PUSTAKA

Dokumen Kajian mengenai tingkat pengangguran dan perubahan tingkat pengangguran telah menjadi isu yang penting dalam bidang ekonomi dan ketenagakerjaan. Banyak penelitian sebelumnya telah dilakukan untuk menganalisis faktor-faktor yang memengaruhi tingkat pengangguran dan dampaknya terhadap perekonomian suatu daerah. Sebagai contoh, di Indonesia, juga telah banyak penelitian tentang tingkat pengangguran. Sebagai contoh, penelitian yang dilakukan oleh Pratama dan Yusuf pada tahun 2016 menganalisis faktor-faktor yang memengaruhi tingkat pengangguran di Provinsi Jawa Barat. Hasil penelitian mereka menunjukkan bahwa pertumbuhan ekonomi, investasi, dan tingkat pendidikan memiliki hubungan yang signifikan dengan tingkat pengangguran di wilayah tersebut. Namun, dalam konteks Kabupaten/Kota Kalimantan Barat, penelitian tentang perubahan tingkat pengangguran masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan pengetahuan ini dengan menganalisis perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode tahun 2010-2018.

Berikut ini visualisasi rata-rata tingkat pengangguran tahun 2010 dan 2018 dari data yang kami peroleh:



**Visualisasi Grafik 1. Rata-rata Tingkat Pengangguran Tahun 2010****Visualisasi Grafik 2. Rata-rata Tingkat Pengangguran Tahun 2018**

Data yang digunakan dalam penelitian ini diperoleh dari Pemerintah Provinsi Kalimantan Barat melalui platform [data.kalbarprov.go.id](http://data.kalbarprov.go.id). Dataset ini menyediakan informasi tentang tingkat pengangguran terbuka (TPT) berdasarkan Kabupaten/Kota di Kalimantan Barat selama periode yang diteliti. Dalam penelitian ini, metode analisis statistik deskriptif, analisis korelasi dan teknik analisis regresi akan digunakan untuk mengidentifikasi tren dan pola perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat. Selain itu, penelitian ini juga akan mengidentifikasi faktor-faktor yang berkontribusi terhadap perubahan tingkat pengangguran, seperti urutan daerah yang mengalami tingkat pengangguran dari terendah hingga tertinggi, persentase dan pola perubahan tingkat pengangguran.

Melalui ringkasan pustaka ini, dapat disimpulkan bahwa penelitian sebelumnya telah memberikan wawasan tentang faktor-faktor yang mempengaruhi tingkat pengangguran dan metode analisis yang dapat digunakan untuk menganalisis perubahan tingkat pengangguran. Meskipun demikian, penelitian yang berfokus pada Kabupaten/Kota Kalimantan Barat masih terbatas, sehingga penelitian ini diharapkan dapat memberikan kontribusi yang berharga dalam pemahaman kita tentang perubahan tingkat pengangguran di wilayah ini.

### III. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis terhadap perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode tahun 2010-2018. Metode penelitian yang digunakan meliputi pengumpulan data, pengolahan data, analisis statistik deskriptif, analisis korelasi antara tahun 2010- 2018, analisis regresi linear, dan analisis perubahan tahun 2010-2018. Data yang digunakan diperoleh dari dataset Tingkat Pengangguran Terbuka (TPT) menurut Kabupaten/Kota yang disediakan oleh Pemerintah Provinsi Kalimantan Barat melalui laman resmi [data.kalbarprov.go.id](http://data.kalbarprov.go.id).

Berikut adalah langkah-langkah yang dijalankan dalam metode penelitian ini:

Pengumpulan Data:

- Data tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode 2010-2018 diunduh dari dataset yang disediakan.
- Data tersebut mencakup variabel Kabupaten/Kota, tahun 2010, tahun 2018, dan tingkat pengangguran pada tahun-tahun tersebut.

**Pengolahan Data:**

- Data yang diunduh kemudian diproses dan dibersihkan agar siap digunakan dalam analisis.
- Langkah-langkah pengolahan data meliputi pemilihan variabel yang relevan, penanganan missing data, dan penyesuaian format data.

**Analisis Statistik Deskriptif:**

- Data yang telah diproses akan dianalisis menggunakan metode statistik deskriptif untuk mengidentifikasi perubahan tingkat pengangguran.
- Analisis ini meliputi perhitungan rata-rata, max,min,mean,median , dan visualisasi grafik yang memperlihatkan perubahan tingkat pengangguran dari tahun 2010 (data lama) dan 2018 (data baru).

**Tabel 1.** Analisis Statistik Deskriptif Tahun 2010

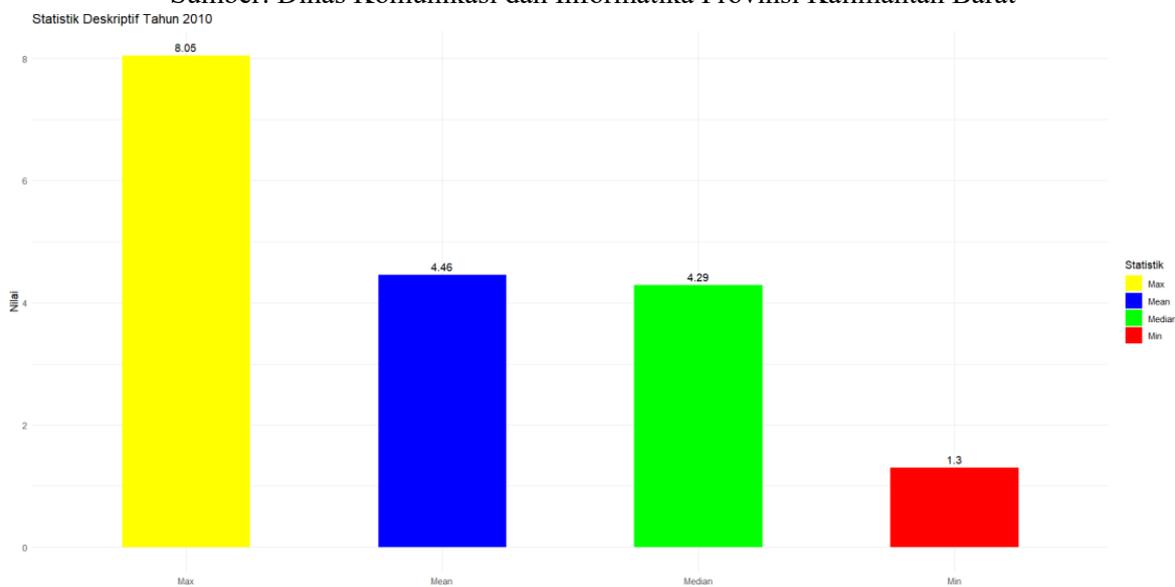
Tahun 2010			
Mean	Median	Min	Max
4.13	3.23	1.58	10.37

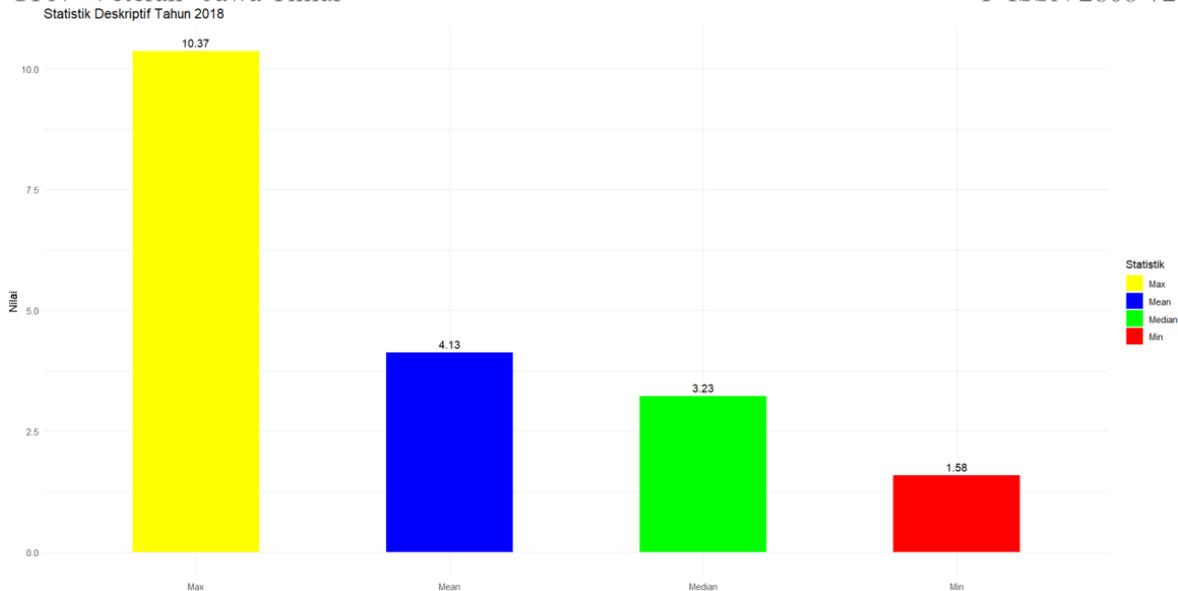
Sumber: Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat

**Tabel 2.** Analisis Statistik Deskriptif Tahun 2018

Tahun 2018			
Mean	Median	Min	Max
4.13	3.23	1.58	10.37

Sumber: Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat

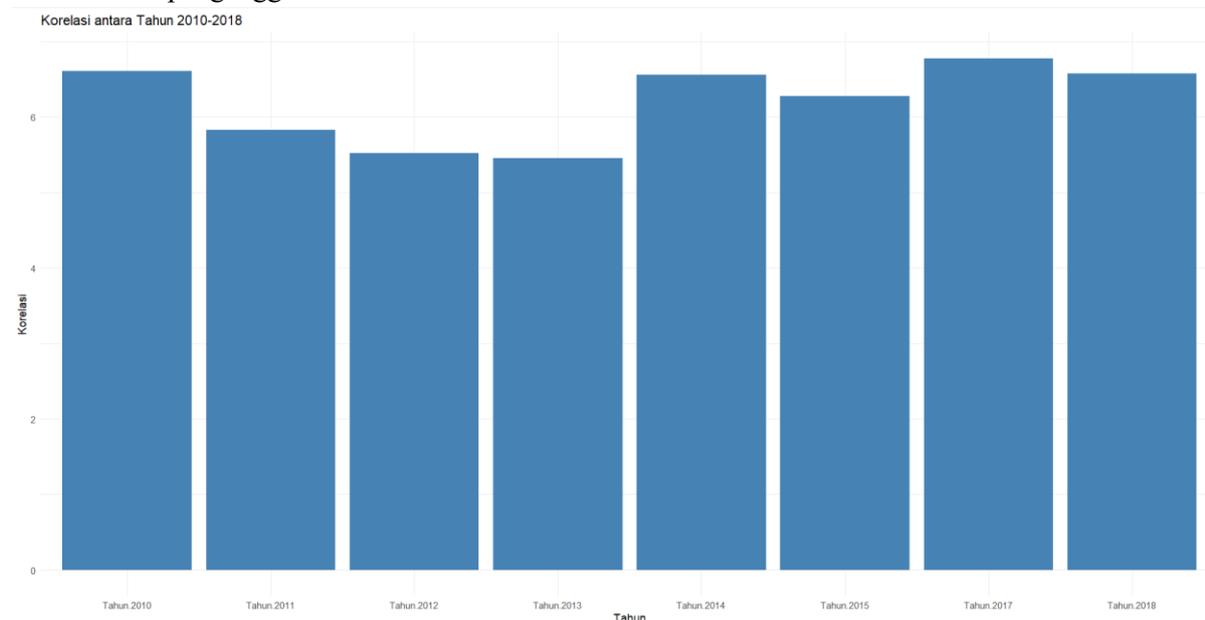
**Visualisasi Grafik 3.** Statistik Deskriptif Tahun 2010



Visualisasi Grafik 4. Statistik Deskriptif Tahun 2018

**Analisis Korelasi :**

- Data yang telah diproses akan di lanjutkan analisis menggunakan metode korelasi untuk mengidentifikasi adanya hubungan atau keterkaitan antara variabel tingkat pengangguran pada tahun- tahun yang berbeda.
- Analisis korelasi ini dapat membantu dalam mengidentifikasi pola perubahan tingkat pengangguran dan visualisasi grafik yang memperlihatkan perubahan tingkat pengangguran dari tahun ke tahun.



Visualisasi Grafik 5. Analisis Korelasi dari tahun 2010 - tahun 2018

Metode penelitian yang dilakukan dalam artikel ini diharapkan dapat memberikan pemahaman yang komprehensif tentang perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat selama periode tahun 2010-2018.

**IV. HASIL DAN PEMBAHASAN**

Hasil dan Pembahasan penelitian ini menunjukkan perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat dari tahun 2010 hingga 2018. Berikut adalah beberapa temuan utama:



- Tahun 2010-2018: Tingkat pengangguran rata-rata -11.33%.

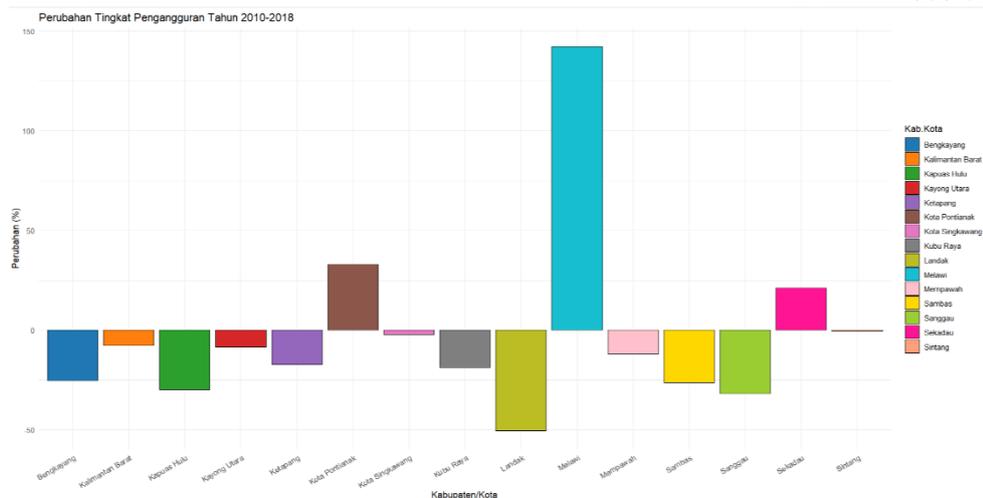
**Pola Perubahan Tingkat Pengangguran:**

- Berdasarkan dari analisis perubahan tahun 2010-2018, kita membuat rank<sub>2018</sub> untuk mengetahui Kab.kota yang memiliki tingkat pengangguran dari terkecil hingga tertinggi. Pada Kolom Perubahan<sub>Tahun\_10\_18</sub> digunakan untuk mengetahui perubahan dari tahun awal yaitu 2010 hingga tahun terakhir 2018, lalu pada kolom persentase<sub>bulat</sub> menunjukkan hasil pembulatan dari kolom Perubahan<sub>Tahun\_10\_18</sub>.

**Tabel 3.** Analisis Perubahan Tahun 2010-2018

<b>Analisis s Perubahan Tahun 2010-2018</b>						
No.	Kab.Kota	Tahun.2 010	Tahun.2 018	Rank_ 2018	Perubahan_Tahun_10 _18	Persentase_B ulat
Kalimantan						
1	Barat	4.62	4.26	11	-77.922.078	-8%
2	Sambas	4.53	3.34	9	-262.693.157	-27%
Bengkayan						
3	g	3.21	2.40	4	-252.336.449	-26%
4	Landak	4.61	2.29	2	-503.253.796	-51%
5	Mempawah	7.80	6.87	13	-119.230.769	-12%
6	Sanggau	3.62	2.47	5	-317.679.558	-32%
7	Ketapang	3.90	3.23	8	-171.794.872	-18%
8	Sintang	2.35	2.34	3	-0.4255319	-1%
Kapas						
9	Hulu	2.25	1.58	1	-297.777.778	-30%
10	Sekadau	2.31	2.80	6	212.121.212	21%
11	Melawi	1.30	3.15	7	1.423.076.923	142%
Kayong						
12	Utara	4.29	3.93	10	-83.916.084	-9%
13	Kubu Raya	6.20	5.04	12	-187.096.774	-19%
Kota						
14	Pontianak	7.79	10.37	15	331.193.838	33%
Kota						
15	Singkawang	8.05	7.88	14	-21.118.012	-3%

Sumber: Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat



- Grafik Visualisasi menunjukkan hasil akhir presentase perubahan tingkat pengangguran dari tahun 2010 hingga 2018. Semakin presentase ke atas/tinggi maka semakin sedikit pengangguran di kota/kabupaten tersebut yang berarti perubahan menjadi lebih baik dan jika semakin kebawah/rendah grafik yang ditunjukkan maka semakin banyak pengangguran atau tidak ada perubahan sama sekali.

Visualisasi Grafik 6. Analisis perubahan tingkat pengangguran tahun 2010-2018

Urutan Daerah dengan Tingkat Pengangguran Tertinggi hingga Terendah:

- Daerah Kabupaten Landak memiliki tingkat pengangguran tertinggi pada tahun 2010-2018 dengan presentase -51%
- Daerah Kota Pontianak mengalami peningkatan tingkat pengangguran terendah yang signifikan dari tahun 2010 hingga 2018 dengan presentase 33%
- Daerah Kabupaten Malawi memiliki tingkat pengangguran terendah pada tahun 2010-2018 dengan presentase 142%

Tabel presentasi tinggi dan rendah tingkat pengangguran tahun 2010 vs 2018:

**Tabel 4.** Persentase Tinggi dan Rendah

Persentase Tinggi dan Rendah		
Perubahan_Persentase_Tahun_10_18	Tahun.2010_vs_2018	Persentase_Bulat
-77.922.078	Tinggi	-8%
-262.693.157	Tinggi	-27%
-252.336.449	Tinggi	-26%
-503.253.796	Tinggi	-51%
-119.230.769	Tinggi	-12%
-317.679.558	Tinggi	-32%
-171.794.872	Tinggi	-18%
-0.4255319	Tinggi	-1%
-297.777.778	Tinggi	-30%
212.121.212	Rendah	21%
1.423.076.923	Rendah	142%
-83.916.084	Tinggi	-9%
-187.096.774	Tinggi	-19%
331.193.838	Rendah	33%
-21.118.012	Tinggi	-3%



Sumber: Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat

Berdasarkan data diatas dibandingkan antara tahun 2010 dan 2018 dan hasil analisis perubahan tahun 2010- 2018. Sehingga data yang dihasilkan mempunyai kesamaan, dapat dilihat pada kolom ke-2 bahwa faktor tinggi rendahnya bergantung terhadap kolom ke-1.

Faktor-faktor yang Berkontribusi terhadap Perubahan Tingkat Pengangguran:

- Faktor rendahnya pengangguran di daerah tertentu adalah pertumbuhan ekonomi, investasi, tingkat pendidikan dan kebijakan ketenagakerjaan yang memiliki hubungan yang signifikan dengan perubahan tingkat pengangguran di Kabupaten/Kota Kalimantan Barat.

Faktor tingginya pengangguran di daerah tertentu adalah Keadaan krisis ekonomi, seperti resesi atau inflasi yang tinggi, dapat mengakibatkan penutupan usaha, pemotongan anggaran, dan pengurangan tenaga kerja. Hal ini dapat menyebabkan peningkatan tingkat pengangguran di daerah tersebut.

## V. KESIMPULAN

Penelitian ini memberikan pemahaman yang lebih dalam tentang bagaimana tingkat pengangguran di Kabupaten/Kota Kalimantan Barat yang mengalami perubahan dari tahun 2010 hingga 2018. Dalam periode tersebut, terlihat fluktuasi tingkat pengangguran yang menunjukkan adanya ketidakstabilan ekonomi di wilayah tersebut. Faktor-faktor ekonomi seperti pertumbuhan ekonomi, investasi, dan tingkat pendidikan memiliki pengaruh signifikan terhadap perubahan tingkat pengangguran.

Penelitian ini juga memberikan sumbangan berharga dalam mengisi kesenjangan pengetahuan yang terbatas mengenai perubahan tingkat pengangguran di wilayah ini. Temuan ini dapat menjadi referensi penting bagi kalangan akademisi dan pihak-pihak yang berkepentingan dalam memahami dan mengatasi masalah pengangguran di Kabupaten/Kota Kalimantan Barat.

## UCAPAN TERIMA KASIH

## REFERENSI

1. Pemerintah Provinsi Kalimantan Barat. (2018). Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota 2010-2018. Diakses dari : <https://data.kalbarprov.go.id/dataset/tingkat-pengangguran-terbuka-tpt-menurut-kab-kota-2010-2018>
2. BPS. (2018). Tingkat Pengangguran Terbuka Menurut Propinsi 1986-2018. Retrieved from <https://www.bps.go.id/statictable/2014/09/15/981/tingkat-pengangguran-terbuka-tpt-menurut-provinsi-1986---2018.html>
3. Sejati, D. P. (2020). PENGANGGURAN SERTA DAMPAKNYA TERHADAP PERTUMBUHAN DAN PEMBANGUNAN EKONOMI. AKSELERASI: Jurnal Ilmiah Nasional, 2(3), 98 - 105.
4. Suhandi, S., Wiguna, W. ., & Quraysin, I. . (2021). DINAMIKA PERMASALAHAN KETENAGAKERJAAN DAN PENGANGGURAN DI INDONESIA. Jurnal Valuasi: Jurnal Ilmiah Ilmu Manajemen Dan Kewirausahaan, 1(1), 268-283
6. Salsabila, N. A., & Pramukty, R. . (2023). Pengaruh Investasi, Indeks Pembangunan Manusia, dan Tingkat Pengangguran Terbuka Terhadap Ketimpangan Distribusi



**SENADA 2023**

Seminar Nasional Sains Data 2023 (SENADA 2023)

UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

Pendapatan. EKOMA : Jurnal Ekonomi, Manajemen, Akuntansi, 2(2), 436–445.

<https://doi.org/10.56799/ekoma.v2i2.1586>



# Analisis Pengaruh COVID-19 Terhadap Jumlah Pengangguran di Jawa Timur Menggunakan Metode Uji Mean

Larasati Romadhani Yunita Putri<sup>1</sup>, Mirechelin Kristanaya<sup>2</sup>, Trimono<sup>3</sup>

*1, 2, 3 Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran Jawa Timur"*

<sup>1</sup>[22083010053@studentupnjatim.ac.id](mailto:22083010053@studentupnjatim.ac.id)

<sup>2</sup>[22083010032@studentupnjatim.ac.id](mailto:22083010032@studentupnjatim.ac.id)

<sup>3</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

Corresponding author email: [22083010053@studentupnjatim.ac.id](mailto:22083010053@studentupnjatim.ac.id)

**Abstract:** *The unemployment rate in Indonesia is one of the social issues affected by the COVID-19 pandemic. This study aims to analyze the impact of the Covid-19 pandemic on increasing the number of unemployed by comparing the unemployment rate in 2019 and 2021. The data used comes from the Central Bureau of Statistics using the two population means test as a method to test significant differences between the number of unemployed before and during a pandemic. The results of this study indicate a significant increase in the number of unemployed in 2021 compared to 2019 as a result of the COVID-19 pandemic. This research is important because it provides a better understanding of the effects of the Covid-19 pandemic on the unemployment rate in East Java, Indonesia, which has drastically changed the employment landscape.*

**Keywords:** *Unemployment, Means Test, COVID-19 Pandemic, Increase in Unemployment*

**Abstrak:** Tingkat pengangguran di Indonesia menjadi salah satu isu sosial yang terdampak karena adanya pandemi COVID-19. Penelitian ini bertujuan untuk menganalisis dampak pandemi Covid-19 terhadap peningkatan jumlah pengangguran dengan membandingkan tingkat pengangguran di tahun 2019 dan 2021. Data yang digunakan bersumber dari Badan Pusat Statistik dengan menggunakan uji mean 2 populasi sebagai metode untuk menguji perbedaan signifikan antara jumlah pengangguran sebelum dan selama pandemi. Hasil penelitian ini menunjukkan adanya peningkatan signifikan dalam jumlah pengangguran di tahun 2021 dibandingkan dengan tahun 2019 sebagai dampak dari pandemi COVID-19. Penelitian ini penting karena memberikan pemahaman yang lebih baik tentang efek pandemic Covid-19 terhadap tingkat pengangguran di Jawa Timur, Indonesia, yang telah mengubah peta ketenagakerjaan secara drastis.

**Kata kunci:** Pengangguran, Uji Mean, Pandemi COVID-19, Peningkatan Pengangguran

## I. PENDAHULUAN

Indonesia menjadi negara yang memiliki jumlah penduduk terbanyak keempat di dunia. Berdasarkan data Worldometers, pada April 2019, Indonesia tercatat memiliki jumlah penduduk sebanyak 269 juta jiwa (3,49 % dari total populasi dunia). Peningkatan jumlah penduduk dari tahun ke tahun ini menimbulkan berbagai permasalahan kependudukan di Indonesia. Salah satu permasalahan tersebut adalah pengangguran. Merujuk pada Badan Pusat Statistik Indonesia disebutkan bahwa tingkat pengangguran terbuka (TPT) Agustus 2020 sebesar 7,07 persen meningkat 1,84 persen poin dibandingkan dengan Agustus 2019. Maknanya, jumlah tingkat pengangguran terbuka di Agustus 2019 sebesar 5,23 persen, jika dibandingkan dengan Agustus 2020 hal ini mengalami kenaikan jumlah pengangguran sebesar 1,84 persen. Tingkat pengangguran yang tinggi ini memiliki dampak negatif yang luas, baik terhadap individu maupun bagi perekonomian secara keseluruhan.

Peningkatan jumlah penduduk tidak menjadi satu-satunya penyebab pengangguran. Beberapa faktor bisa menjadi penyebab terjadinya pengangguran, mulai dari jumlah lapangan pekerjaan yang terbatas, adanya PHK (Pemutusan Hubungan Kerja), dan terjadinya situasi pandemi yang melanda dunia sejak tahun 2020. Selain itu, pengurangan biaya perusahaan dengan diberlakukannya kebijakan PHK juga dapat meningkatkan jumlah pengangguran. Salah satu faktor paling signifikan yang memengaruhi tingkat pengangguran adalah pandemi.



Indonesia merupakan salah satu negara yang terdampak parah dengan adanya pandemi Covid-19. Mengutip dari detik.com, kasus pertama di Indonesia terjadi pada WNI (Warga Negara Indonesia) yang merupakan seorang ibu dan anak yang tinggal di Depok, Jawa Barat. Dari kasus pertama tersebut, berdasar pada Gugus Tugas Percepatan Penanganan COVID-19, 2020 hingga per tanggal 28 Mei 2020, tercatat 31.024 kasus COVID-19 yang telah menyebar di 34 provinsi di Indonesia. Kementerian Kesehatan Republik Indonesia menyebutkan, Hingga 24 November 2021, Pemerintah Republik Indonesia telah melaporkan 4.254.443 kasus terkonfirmasi positif COVID-19 dan ada 143.766 kematian (CFR: 3,4%) terkait COVID-19 yang dilaporkan dan 4.102.700 pasien telah sembuh dari penyakit tersebut.

Beberapa penelitian sebelumnya telah mengungkapkan dampak signifikan dari pandemi ini terhadap pasar tenaga kerja. Penelitian pada jurnal Ilmu Sosial dan Pendidikan (JISIP) Vol. 5 No. 3 Juli 2021 menunjukkan bahwa diberlakukannya PSBB (Pembatasan Sosial Berskala Besar) berdampak pada jumlah pengangguran yang terjadi di Indonesia terus meningkat seiring adanya pandemi COVID-19. Pada jurnal tersebut juga menunjukkan bahwa ada sebanyak 212.394 pekerja yang di-PHK dan sebanyak 1.205.191 pekerja yang dirumahkan. Dari sektor nonformal, tercatat sekitar 282 ribu orang yang tidak memiliki penghasilan. Pada penelitian oleh Rusman Universitas Muhammadiyah Mamuju Sulawesi Barat yang merujuk pada Badan Pusat Statistik menyebutkan bahwa 21.32 juta (turun 7.80 juta) penduduk usia kerja di Indonesia terdampak akibat pandemi Covid-19. Dalam periode Agustus 2021 tahun ini 1.82 juta orang kehilangan pekerjaan akibat dari pandemi. Dalam penelitian-penelitian yang telah disebutkan di atas membuktikan bagaimana pandemi COVID-19 berpengaruh terhadap jumlah pengangguran di Indonesia.

Berkaitan dengan masalah tersebut penelitian ini mencoba untuk mengkaji menggunakan analisis numerik terkait pengaruh COVID-19 terhadap peningkatan pengangguran khususnya di Provinsi Jawa Timur. Metode yang digunakan adalah Uji Mean Dua Populasi yaitu uji yang menggunakan dua sampel atau lebih sebagai objek penelitiannya. Uji ini memiliki tujuan untuk melihat ada tidaknya perbedaan yang signifikan terhadap kedua sampel populasi dari data yang kita uji.

## II. METODE PENELITIAN

Pada Indonesian Journal of Business Analytics (IJBA) mengatakan bahwa pertumbuhan kasus Covid-19 mempengaruhi pertumbuhan ekonomi nasional bahkan Indonesia masuk ke dalam resesi. Dalam jurnal ini juga mengatakan bahwa peningkatan pengangguran yang terjadi merupakan imbas dari pemberlakuan PSBB sebagai bentuk antisipasi pemerintah agar tidak terjadi penularan yang semakin meluas. Selain itu, jurnal ini menunjukkan meskipun pertumbuhan angka pengangguran menurun dari tahun 2020 ke tahun 2021 sebesar 34,41% atau 9,10 juta penduduk usia kerja harus terdampak covid 19, turun 10,02 juta penduduk, tetap pemerintah harus berhati-hati karena pandemi covid 19 ini belum usai bahkan pertengahan tahun 2021 pertumbuhan kasus semakin meningkat bahkan di kenal dengan gelombang ke 2. Ini mungkin akan berdampak pula pada pertumbuhan angka pengangguran seperti yang terjadi di tahun 2020 saat awal virus covid 19 masuk ke Indonesia, jika pemerintah tidak waspada.

Pada penelitian ini, kami akan membandingkan secara numerik menggunakan uji mean dua populasi untuk mengetahui apakah ada perbedaan secara signifikan pada rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 dan rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021. Perlu dicatat bahwa tahun 2019 sebelum terjadinya pandemi COVID-19, sementara tahun 2021 merupakan tahun di mana pandemi tersebut telah berdampak signifikan pada perekonomian dan tingkat pengangguran. Oleh karena itu, analisis ini bertujuan untuk mengetahui apakah pandemi COVID-19 mempengaruhi kondisi ekonomi dan tingkat pengangguran khususnya di Jawa Timur yang membuat adanya perbedaan secara signifikan antara kedua tahun tersebut.



Untuk melakukan uji mean dua populasi terdapat beberapa prosedur:

### 2.1. Pengumpulan Data

Pada penelitian ini data yang kami gunakan berasal dari Badan Pusat Statistik Provinsi Jawa Timur (BPS Provinsi Jawa Timur). Kami mengambil dua kolom sampel yaitu Jumlah Pengangguran Angkatan Kerja di Jawa Timur pada tahun 2019 dan Jumlah Pengangguran Angkatan Kerja di Jawa Timur pada tahun 2021. Kolom yang kita ambil merupakan nilai total dari kolom yang pernah bekerja dan kolom yang tidak pernah bekerja.

### 2.2. Analisis Deskriptif

Tahap ini bertujuan untuk mengetahui penyebaran dan pemusatan data yang akan kita teliti, serta kita juga bisa mengetahui ringkasan dari data yang kita pakai. Analisis deskriptif meliputi:

#### 2.2.1. Nilai Minimum

Nilai minimum memberikan informasi tentang batas bawah atau ekstrem terendah dari data yang sedang diteliti.

#### 2.2.2. Quartil

Tahap ini akan memberikan informasi tentang titik yang membagi data dalam urutan terurut menjadi empat bagian yang sama.

#### 2.2.3. Median

Median akan memberikan informasi mengenai nilai tengah pada data yang sedang kita teliti.

#### 2.2.4. Mean

Mean akan memberikan informasi mengenai rata-rata dari data yang sedang kita teliti.

#### 2.2.5. Nilai Maximum

Nilai minimum memberikan informasi tentang batas atas atau ekstrem tertinggi dari data yang sedang diteliti.

#### 2.2.6. Grafik / Plot

Tahap ini akan menyajikan informasi berupa tampilan visual dari data yang kita teliti.

### 2.3. Menentukan tingkat signifikansi atau $\alpha$

Tingkat signifikansi diperlukan sebagai pembanding terhadap p-value pada tiap pengujian data. Pada penelitian ini kami menggunakan tingkat signifikansi sebesar 0.05 atau sebesar 5%.

### 2.4. Uji Normalitas Data

Untuk melakukan uji mean dua populasi, data dari kedua populasi yang akan dibandingkan harus berdistribusi secara normal sehingga data tersebut perlu untuk diuji normalitasnya. Hal ini dapat diuji menggunakan metode statistik seperti uji Kolmogorov-Smirnov atau Shapiro-Wilk.

Pada penelitian ini kami menggunakan uji Kolmogorov-Smirnov. Di mana ketika p-value yang didapatkan dari uji normalitas ini lebih besar dari  $\alpha$ /tingkat signifikansinya maka data yang di uji terdistribusi secara normal. Dan sebaliknya ketika p-value lebih kecil dari  $\alpha$  maka data yang di uji tidak berdistribusi normal.

### 2.5. Uji Homogenitas Varians Kedua Populasi

Syarat kedua untuk melakukan uji mean dua populasi adalah varians populasinya harus homogen, maka tahap ini adalah tahap yang harus dilakukan. Ketika nilai p-value yang didapat pada tahap ini lebih besar dari atau sama dengan tingkat signifikansi yang telah ditetapkan, maka data pada populasi tersebut memiliki varians yang homogen sehingga memenuhi syarat untuk bisa di uji mean dua populasi.

#### 2.5.1. Transformasi Data

Data yang kami gunakan pada penelitian ini perlu untuk tahap dilakukan transformasi data karena ketika kedua populasi tidak dilakukan transformasi data, akan menghasilkan p-value yang kurang dari tingkat signifikansi sehingga menyebabkan data pada penelitian ini tidak memenuhi syarat homogenitas varians pada kedua populasinya.

## 2.6. Menentukan Hipotesis

Pada penelitian ini kami menggunakan uji di sisi kanan dengan hipotesis yang bertujuan untuk membandingkan populasi jumlah pengangguran di Jawa Timur pada tahun 2019 dengan populasi jumlah pengangguran di Jawa Timur pada tahun 2021. Hipotesis nol ( $H_0$ ) menyatakan bahwa rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 kurang dari atau sama dengan rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021. Sebaliknya, hipotesis alternatif ( $H_1$ ) menyatakan bahwa rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 lebih besar daripada rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021. Dengan keterangan sebagai berikut:

- $H_0 : \mu_1 \leq \mu_2$

Rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 kurang dari atau sama dengan rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021.

- $H_1 : \mu_1 > \mu_2$

Rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 lebih besar daripada rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021.

## 2.7. Statistika Uji

### 2.7.1. Mencari Z-Hitung

Tahap ini diperlukan untuk mengambil keputusan atau membuat inferensi tentang populasi berdasarkan sampel data yang diteliti. Berdasar pada data yang kita pakai, statistik uji yang paling cocok untuk digunakan pada data kami adalah menggunakan uji-Z karena ukuran sampel pada data ini memiliki jumlah yang lebih dari 30 dan dengan data yang berdistribusi normal.

Pada uji-Z kita akan mencari nilai Z-Hitung dan Z-Tabel. Untuk mencari nilai Z-Hitung, dapat menggunakan rumus:

$$Z - \text{Hitung} = \frac{(\text{mean1} - \text{mean2})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Dengan mean1 dan mean2 adalah rata-rata tiap sampel,  $s_1$  dan  $s_2$  adalah simpangan baku tiap sampel, dan  $n_1$  dan  $n_2$  adalah ukuran tiap sampel.

### 2.7.2. Mencari Z-tabel

Pada penelitian ini kami mencari Z-tabel pada RStudio menggunakan sintaks `qnorm(p, mean=0, sd=1)` dengan  $p$  merupakan peluang yang didapat dari  $1 - \alpha$  di mana  $\alpha$  adalah tingkat signifikansi yang telah ditetapkan sebelumnya.

## 2.8. Menentukan Kriteria Uji

Pada tahap ini kita akan mengetahui apakah data yang diuji akan gagal tolak  $H_0$  atau menolak  $H_0$ , yang berdasar pada nilai Z-Hitung dan Z-Tabel yang telah kita cari dengan kriteria sebagai berikut:

- Gagal tolak  $H_0$  jika  $Z\text{-Hitung} \leq Z\text{-Tabel}$
- Menolak  $H_0$  jika  $Z\text{-Hitung} > Z\text{-Tabel}$

## 2.9. Interpretasikan Hasil

Tahap ini memuat penjelasan dari hasil yang didapat pada penelitian ini.

## III. HASIL DAN PEMBAHASAN

### 3.1. Tabel Data

**Tabel 1.** Data Jumlah Pengangguran di Jawa Timur pada Tahun 2019 dan 2021

Kabupaten / Kota	Jumlah Pengangguran 2019	Jumlah Pengangguran 2021
Pacitan	3389	7527



Ponorogo	17922	22931
Trenggalek	14004	14568
Tulungagung	19201	29630
Blitar	20606	24072
Kediri	32104	44706
Malang	54106	77267
Lumajang	15181	19439
Jember	48278	73017
Banyuwangi	37054	50659
Bondowoso	13797	20835
Situbondo	11148	14787
Probolinggo	24079	31063
Pasuruan	46653	53865
Sidoarjo	54464	130979
Mojokerto	22159	35033
Jombang	30286	50063
Nganjuk	17681	26982
Madiun	13866	18846
Magetan	11299	14812
Ngawi	17790	21216
Bojonegoro	26068	35057
Tuban	17536	32766
Lamongan	25838	33476
Gresik	36390	57912
Bangkalan	26745	41796
Sampang	13312	17966
Pamekasan	10590	14197
Sumenep	14187	15630
Kediri	6157	9971
Blitar	3684	5264
Malang	27664	46542
Probolinggo	5083	8543
Pasuruan	5191	6977
Mojokerto	1715	4768
Madiun	3776	7859
Surabaya	91912	152273
Batu	2839	8101

### 3.2. Tahap Analisis Deskriptif

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1715	11186	17736	22204	27434	91912	4768	14623	23502	33721	43979	152273

**Gambar 1.** Analisis Deskriptif  
Jumlah Pengangguran di Jawa Timur  
pada Tahun 2019

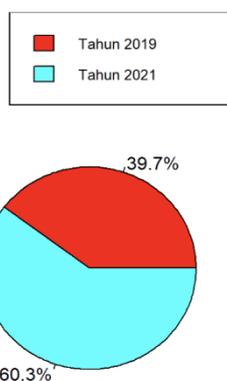
**Gambar 2.** Analisis Deskriptif  
Jumlah Pengangguran di Jawa Timur  
pada Tahun 2021

Rata-rata (mean) jumlah pengangguran dari kedua populasi yang diteliti menunjukkan perbedaan antara tahun 2019 dan tahun 2021. Pada tahun 2019, rata-rata jumlah pengangguran sebesar 22,204 sedangkan pada tahun 2021 rata-rata jumlah pengangguran meningkat menjadi 33,721. Selain itu, terdapat perbedaan yang cukup besar antara jumlah pengangguran terendah pada tahun 2019, yaitu sebanyak 1,715, dengan jumlah pengangguran terendah pada tahun 2021, yaitu sebanyak 4,768.

Selanjutnya, jumlah pengangguran tertinggi pada tahun 2019 mencapai 91,912, sedangkan pada tahun 2021 jumlah pengangguran mencapai 152,273. Perbedaan ini menunjukkan peningkatan yang signifikan dalam jumlah pengangguran dari tahun ke tahun.

Selain itu, terdapat perbedaan nilai tengah (median) jumlah pengangguran antara tahun 2019 dan tahun 2021. Pada tahun 2019, nilai tengah pengangguran adalah 17,736, sedangkan pada tahun 2021 nilai tengahnya meningkat menjadi 23,502. Berikut presentase jumlah pengangguran di Jawa Timur pada tahun 2019 dan 2021.

**Persentase Jumlah Pengangguran  
di Jawa Timur pada Tahun 2019 dan 2021**



**Gambar 3. Pie Chart**

Jumlah Pengangguran di Jawa Timur pada Tahun 2019 dan 2021

Pada gambar 3 disajikan dalam bentuk *pie chart* dengan presentase jumlah pengangguran di Jawa Timur pada tahun 2019 sebesar 39.7% sedangkan pada tahun 2021 sebesar 60.3%.

### 3.3. Tahap Uji Normalitas Data

Tahapan selanjutnya adalah menguji normalitas kedua populasi yang kita teliti. Dari uji Kolmogorof-Smirnov yang kita gunakan didapat nilai p-value yang akan dibandingkan dengan tingkat signifikansi/alpha yang telah kita tetapkan yaitu 0.05. Berikut hasil uji normalitas pada kedua sampel populasi.

Exact two-sample Kolmogorov-Smirnov test

```
data: x1 and x2
D = 0.21053, p-value = 0.3727
alternative hypothesis: two-sided
```

**Gambar 4. Uji Normalitas Dua Sampel Populasi**  
Jumlah Pengangguran di Jawa Timur pada Tahun 2019 dan 2021

Berdasar pada gambar 4, uji normalitas ini memiliki p-value sebesar 0.3727 yang mana nilai p-value ini lebih besar dari pada tingkat signifikansi yang telah kita tetapkan. Maka dapat

disimpulkan bahwa kedua populasi yang kami teliti berdistribusi normal dan memenuhi syarat normalitas pada kedua populasinya.

### 3.4. Tahap Uji Homogenitas Varians Data

#### 3.4.1. Transformasi Data

Kami mengakarkan tiap sampel populasinya agar bisa menghasilkan nilai p-value yang lebih dari tingkat signifikansi yang telah ditetapkan.

#### 3.4.2. Menguji Homogenitas Varians Data

Setelah melakukan transformasi data, maka kita bisa masuk pada tahap ini. Data pada uji ini, menghasilkan p-value sebesar 0.131 yang mana nilai ini lebih besar dari pada tingkat signifikansi yang telah ditetapkan sehingga dapat disimpulkan bahwa kedua populasi pada data ini memenuhi syarat homogenitas varians nya dengan hasil uji sebagai berikut:

```
F test to compare two variances
data: a and b
F = 0.60501, num df = 37, denom df = 37,
p-value = 0.131
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.314437 1.164112
sample estimates:
ratio of variances
 0.6050124
```

**Gambar 5.** Uji F untuk mengetahui homogenitas kedua populasi

### 3.5. Statistik Uji

#### 3.5.1. Tahap Mencari Z-Hitung

Setelah menguji homogenitas varians nya, selanjutnya mencari nilai Z-Hitung sample kedua populasi dari data yang kita teliti. Nilai Z-Hitung dari kedua populasi pada penelitian ini memiliki nilai sebesar -1.933314. Nilai ini akan dibandingkan pada z-tabel yang akan kita cari pada tahap selanjutnya.

#### 3.5.2. Tahap Mencari Z-Tabel

Setelah itu, nilai Z-Tabel yang diperoleh dari data kedua populasi pada penelitian ini sebesar 2.326348. Nilai ini didapat dari sintaks pemrograman Bahasa R yang telah kami sebut sebelumnya. Dimana nilai Z-Tabel ini nantinya akan dibandingkan dengan nilai Z-Hitung.

### 3.6. Keputusan/Interpretasi Hasil

Berdasar pada nilai Z-Hitung dan Z-Tabel yang dihasilkan pada tahap sebelumnya, kita tahu bahwa nilai Z-Hitung lebih kecil daripada nilai Z-Tabel, sehingga kondisi ini memenuhi pada kriteria uji pada poin A yang telah kita tentukan sebelumnya. Kriteria uji pada poin A tersebut menunjukkan jika nilai Z-Hitung kurang dari atau sama dengan nilai Z-Tabel maka akan gagal menolak  $H_0$  yang berarti kondisi kedua sampel populasi pada penelitian ini memenuhi kondisi  $H_0$ .

Maka keputusan yang bisa kita ambil yaitu rata-rata jumlah pengangguran di Jawa Timur pada tahun 2019 kurang dari atau sama dengan rata-rata jumlah pengangguran di Jawa Timur pada tahun 2021. Hal ini membuktikan bahwa pandemi COVID-19 membawa pengaruh yang signifikan pada peningkatan jumlah pengangguran pada tahun 2019 (sebelum terjadi pandemi) dan pada tahun 2021 (setelah terjadi pandemi).

## IV. KESIMPULAN

Dengan menggunakan metode Uji Mean Dua Populasi, terbukti bahwa pandemi Covid-19 memberikan pengaruh signifikan terhadap kenaikan jumlah pengangguran di Jawa Timur. Perbandingan rata-rata antara data pengangguran pada tahun 2019 dan data pengangguran pada tahun 2021 di Jawa Timur menunjukkan adanya perbedaan yang signifikan. Hal ini mengindikasikan bahwa dampak pandemi Covid-19 secara nyata telah berkontribusi terhadap peningkatan jumlah pengangguran di wilayah tersebut.

**UCAPAN TERIMA KASIH**

Ucapan terima kasih disampaikan kepada mata kuliah bahasa Indonesia yang diselenggarakan oleh UPN “Veteran” Jawa Timur sehingga artikel ini bisa ditulis dengan baik dan benar.

**REFERENSI**

1. Anam, K., & Yahya, M. S. (2021). Inovasi Guru Dalam Pembelajaran Masa Pandemi covid-19. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 5(3), 2090.
2. Nurrahman, A. (2020, March 1). Upaya Pemerintah Dalam Mengatasi Permasalahan Pengangguran di Indonesia. Repository IPDN. (<http://eprints.ipdn.ac.id/6283/> )
3. Jurnal Kependudukan Indonesia. (n.d.). Dampak Pandemi Covid-19 Dalam Perspektif Kependudukan. Retrieved from (<https://ejurnal.kependudukan.lipi.go.id/index.php/jki/article/download/571/pdf> )
4. Sabiila, S. I. (n.d.). Kasus Corona pertama di Indonesia, Ini Kilas Balik Usai 2 Tahun Berlalu. Detiknews. Retrieved from (<https://news.detik.com/berita/d-5964691/kasus-corona-pertama-di-indonesia-ini-kilas-balik-usai-2-tahun-berlalu#:~:text=2> )
5. Rokom. (2021, January 19). 3 provinsi Ini Paling tinggi Tingkat Penularan covid-19. Sehat Negeriku. Retrieved from (<https://sehatnegeriku.kemkes.go.id/baca/umum/20200610/4434057/3-provinsi-paling-tinggi-tingkat-penularan-covid-19/>)
6. Teguh Ali Fikri, Y. (2021). Analisis Peningkatan Angka Pengangguran Akibat Dampak Pandemi covid 19 di Indonesia. *Indonesian Journal of Business Analytics*, 1(2), 107–116.
7. Trimono, T., Riyantoko, P. A., & Agista, F. (n.d.). *Model Arma-Garch dan ensemble Arma-Garch Untuk Prediksi value-at-risk pada portofolio saham*. PROSIDING SEMINAR NASIONAL SAINS DATA. Available online: (<https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/52> )
8. Rusman. Dampak Pandemi COVID-19 Terhadap Angka Pengangguran Di Indonesia. (ISSN Cetak: 2654-5306).
9. Rpubs. Available online : <https://rpubs.com/> (27/05/2023).



# Analisis Dampak Covid-19 terhadap Nilai Ekspor Provinsi Jawa Timur dengan Uji Mean

Affa Lelira Ibrahim<sup>1</sup>, Deannisa Syafira Putri<sup>2</sup>, Trimono<sup>3</sup>

<sup>1, 2, 3</sup>Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[22083010065@student.upnjatim.ac.id](mailto:22083010065@student.upnjatim.ac.id)

<sup>2</sup>[22083010062@student.upnjatim.ac.id](mailto:22083010062@student.upnjatim.ac.id)

Corresponding author email: [trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

**Abstract:** *The export sector is a sector that plays an important role in the economy in Indonesia. One of the factors in the decline in the value of exports in Indonesia occurred due to the Covid-19 pandemic. The purpose of this study is to compare the average export value in 2020 with the average export value in 2021 in East Java province. The results of this study indicate that the Covid-19 pandemic has had an impact on the export sector in East Java. This research was conducted using the mean 2 population test analysis method with data obtained from the Central Badan Pusat Statistik (BPS).*

**Keywords:** *Export, Covid-19, East Java*

**Abstrak:** Sektor Ekspor merupakan salah satu sektor yang berperan penting dalam perekonomian di Indonesia. Salah satu faktor dari penurunan nilai ekspor di Indonesia ini terjadi karena adanya pandemi Covid-19. Tujuan dari penelitian ini adalah untuk membandingkan nilai rata-rata ekspor tahun 2020 dengan nilai rata-rata ekspor tahun 2021 di provinsi Jawa Timur. Hasil dari penelitian ini menunjukkan bahwa pandemi Covid-19 memberikan dampak pada sektor Ekspor di Jawa Timur. Penelitian ini dilakukan dengan menggunakan metode analisis uji mean 2 populasi dengan data yang didapatkan dari Badan Pusat Statistik (BPS).

**Kata kunci:** Ekspor, Covid-19, Jawa Timur

## I. PENDAHULUAN

Pada akhir tahun 2019, Virus SARS-CoV-2 pertama kali terdeteksi di China yang membuat seluruh negara terdampak pandemi Covid-19 dan menyebabkan lebih dari 178 juta kasus dikonfirmasi dan 3,9 juta kematian[1]. Virus ini juga memasuki negara Indonesia tepatnya di bulan Maret 2020. Dalam meminimalisir penyebaran virus Covid-19, pemerintah membatasi segala kegiatan dengan dilakukannya *lockdown*. Akibat diberlakukannya *lockdown* ini, beberapa sektor perdagangan di Indonesia terkena imbasnya, salah satu contohnya adalah sektor perdagangan ekspor.

Sektor perdagangan ekspor di sebuah negara memainkan peran penting dalam meningkatkan pertumbuhan ekonomi. Hasil dari sektor perdagangan ekspor ini dapat menyuplai anggaran negara melalui pendapatan dan mata uang asing yang dapat digunakan untuk memperbaiki infrastruktur dan menciptakan iklim investasi yang menarik (Mohsen, 2015). Peran sektor perdagangan ekspor juga dinilai sangat penting dalam mendorong suatu negara untuk dapat berinovasi dalam produksi maupun penggunaan teknologi yang canggih. Selain itu, peran sektor perdagangan ekspor juga berpengaruh bagi perusahaan karena kegiatan ekspor mendorong motivasi perusahaan untuk mengadopsi praktik terbaik yang dilakukan dalam kancah internasional dan penerapan inovasi teknologi terdepan yang mendorong efisiensi dan meningkatkan kualitas produk sehingga pada akhirnya menghasilkan daya saing ekspor (Bbaale et al., 2019)

Pada Provinsi Jawa Timur, nilai ekspor sendiri mengalami surplus dan defisit yang disebabkan oleh beberapa faktor, seperti faktor harga ekspor, GDP per kapita, kurs nilai tukar riil, migas dan non migas, dsb[2]. Berdasarkan Badan Pusat Statistik (BPS) Jawa Timur nilai ekspor pada tahun 2017 hingga 2018 meningkat sebesar 1,29%[3]. Sedangkan, pada tahun 2018 hingga 2019 juga meningkat sebesar 2,74%[4]. Dikarenakan pada tahun 2020 hingga 2021 Indonesia dilanda pandemi Covid-19 yang menyebabkan keterbatasannya kegiatan ekspor di provinsi Jawa Timur. Sehingga masalah tersebut menjadi latar belakang tujuan dilakukannya penelitian ini, yaitu untuk membuktikan apakah pandemi Covid-19 memengaruhi nilai ekspor di Provinsi Jawa Timur berdasarkan perbandingan rata-rata nilai FOB tahun 2020 dan 2021 dengan menggunakan analisis uji mean 2 populasi.



## II. KAJIAN PUSTAKA

### Penelitian Terdahulu yang Relevan

Pada dasarnya, penelitian mengenai analisis ujaran kebencian pada kolom komentar youtube sudah dilakukan pada peneliti terdahulu. Salah satunya adalah penelitian tentang “Ujaran Kebencian dalam Kolom Komentar Youtube pada Tahun Politik Pemilihan Presiden 2019” pada tahun 2020 oleh Fitri Jamilah dan Primasari Wahyuni. Yang berisi mengenai ujaran kebencian yang terjadi pada pemilihan presiden tahun 2019 di channel YouTube [4].

## III. METODE PENELITIAN

### III.1. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, dimana data didapatkan atau diambil dari *website* Badan Pusat Statistik (BPS). Data ini berisi informasi mengenai nilai *Free on Board* (FOB) ekspor pada tahun 2020 dan 2021 di Provinsi Jawa Timur. Berikut adalah data yang akan digunakan dalam penelitian ini.

**Tabel 1.** nilai *Free on Board* (FOB) ekspor tahun 2020 dan 2021

2020	2021
5.288.050.602	2.941.437.995
1.382.401.517	2.088.634.060
1.168.105.357	1.553.410.680
858.352.824	910.503.123
753.444.271	1.024.472.058
770.885.076	764.261.526
1.275.803.477	1.718.090.471
946.239.028	1.550.263.778
437.201.657	542.201.822
595.088.894	746.443.301
585.242.649	562.401.772
744.903.880	761.423.680
512.776.326	503.687.573
566.347.987	857.630.300
415.954.150	371.871.419
457.453.120	541.706.355
266.931.410	340.830.147
124.253.138	185.907.591
349.749.049	305.690.444
219.312.530	222.389.142
241.346.988	168.106.619
134.265.106	91.829.671
82.612.261	102.728.447
215.952.220	203.280.796
129.716.956	156.988.839
104.142.992	255.270.754
161.658.525	185.116.880
103.382.149	146.159.273
95.234.014	93.665.136
744.903.880	761.423.680
1.173.178.804	1.268.916.471

### III.2. Analisis Data

Pada data ini, dilakukan analisis data dengan menggunakan salah satu teknik pengujian hipotesis dalam ilmu statistika yaitu uji mean 2 populasi. Uji mean 2 populasi adalah metode yang digunakan untuk menguji kesamaan rata-rata dari 2 populasi yang bersifat independen, maka dari itu uji ini juga disebut uji rata-rata 2 sampel independen (bebas). Independen disini maksudnya yaitu sampel atau populasi yang digunakan tidak saling memengaruhi atau tidak saling berhubungan. Dalam uji mean 2 populasi ini, terdapat asumsi yang harus dipenuhi, yaitu data harus berdistribusi normal dan variansi kedua populasi harus sama[5]. Dengan sampel yang diambil adalah nilai FOB ekspor di tahun 2020 dan 2021. Pengujian ini nantinya akan dihitung dengan menggunakan *software Rstudio*. Hasil dari pengujian mean 2 populasi ini akan membuktikan apakah pandemi Covid-19 memengaruhi nilai ekspor di Provinsi Jawa Timur berdasarkan perbandingan rata - rata nilai FOB tahun 2020 dan 2021.

Dalam perhitungan uji mean 2 populasi ini, uji statistik yang akan digunakan adalah uji Z. Uji Z adalah salah satu uji statistika dimana hipotesisnya didekati dengan distribusi normal dengan varians populasi yang telah diketahui bernilai sama. Uji Z dapat digunakan untuk menguji data dengan jumlah sampel yang besar, yaitu lebih besar dari 30[6]. Berikut adalah rumus untuk menghitung uji Z.

$$IV. \quad Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} [7]$$

Berdasarkan penjelasan di atas dapat disimpulkan bahwa dalam perhitungan uji hipotesis mean 2 populasi, akan dicari terlebih dahulu asumsi apakah data yang dianalisis berdistribusi normal dan memiliki nilai varians yang sama. Kedua asumsi ini harus terpenuhi terlebih dahulu agar nantinya hasil dari pengujian mean 2 populasi dapat menampilkan hasil yang akurat. Berikut adalah penjelasan untuk uji asumsi yang dicari.

#### III.2.1. Uji Distribusi Normal

Uji asumsi berdistribusi normal atau juga disebut uji normalitas merupakan sebuah uji hipotesis yang dilakukan untuk menilai sebaran data pada sebuah kelompok data atau variabel. Uji normalitas ini berisi proses statistik yang digunakan untuk menguji apakah suatu sampel atau populasi mengikuti distribusi normal atau tidak. Data yang berdistribusi normal maksudnya adalah data yang memiliki pola persebaran yang simetris dan mengikuti kurva normal. Dalam pengujian hipotesis statistik, uji normalitas ini banyak digunakan. Hal ini dikarenakan jika data yang diuji tidak berdistribusi normal, maka analisis yang didasarkan pada asumsi normalitas mungkin tidak akurat[8-9]. Karena dalam analisis ini *software* yang digunakan untuk menghitung uji hipotesis adalah *Rstudio*, maka cara untuk menentukan uji normalitas ini menggunakan *function* bawaan dari *Rstudio* sendiri, yaitu Kolmogorov-Smirnov test atau *ks.test*.

#### III.2.2. Uji Homogenitas Varians

Uji homogenitas adalah uji asumsi yang digunakan untuk menentukan apakah varians dari 2 atau lebih sebuah data sama bersifat sama atau tidak. Pengujian ini menggunakan pendekatan distribusi F. Dalam *software Rstudio*, uji homogenitas varians biasanya dicari dengan menggunakan *function* *var.test*[10].

## V. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah untuk membandingkan rata-rata nilai FOB ekspor di tahun 2020 dengan rata-rata nilai FOB ekspor di tahun 2021. Metode ini akan menghasilkan sebuah gambaran tentang seberapa besar pengaruh Covid-19 terhadap nilai ekspor di Provinsi Jawa Timur. Berikut adalah gambar hasil analisis hipotesis uji mean 2 populasi.

<pre>Langkah Awal : Menentukan Uji Normalitas Data Pertama 1. Hipotesis : H0: Data berdistribusi normal H1: Data tidak berdistribusi normal 2. Diketahui : Tingkat signifikansi = alpha = 0.05 Mean = 22831.13 standar Deviasi = 12577.51 3. Statistika Uji : Nilai p-value = 0.6030086 4. Kriteria Uji : H0 ditolak jika p-value &lt; 0.05 H0 gagal ditolak jika p-value &gt; 0.05 5. Kesimpulan : Data berdistribusi Normal  Data Kedua 1. Hipotesis : H0: Data berdistribusi normal H1: Data tidak berdistribusi normal 2. Diketahui : Tingkat signifikansi = alpha = 0.05 Mean = 24023.65 standar Deviasi = 11598.19 3. Statistika Uji : Nilai p-value = 0.8401508 4. Kriteria Uji : H0 ditolak jika p-value &lt; 0.05 H0 gagal ditolak jika p-value &gt; 0.05 5. Kesimpulan : Data berdistribusi Normal</pre>	<pre>Langkah kedua : Uji Homogenitas varians 1. Hipotesis : H0: Kedua data memiliki varians sama H1: Kedua data tidak memiliki varians sama 2. Statistika Uji : p-value uji = 0.659886 3. Kriteria Uji : H0 ditolak jika p-value &lt; 0.05 H0 gagal ditolak jika p-value &gt; 0.05 4. Kesimpulan : Kedua data memiliki varians sama  Langkah ketiga : Menentukan Uji Mean 1. Hipotesis : H0: mu1 &lt;= mu2 H1: mu1 &gt; mu2 2. Diketahui : Tingkat signifikansi = alpha = 0.05 Mean1 = mu1 = 22831.13 Mean2 = mu2 = 24023.65 standar Deviasi1 = 12577.51 standar Deviasi2 = 11598.19 n1 = 31 n2 = 31 3. Statistika Uji : Nilai zhit = -0.0002746756 Nilai ztabel = 1.644854 4. Kriteria Uji : H0 ditolak jika zhit &gt; z_tabel H0 gagal ditolak jika zhit &lt; z_tabel 5. Kesimpulan : mu1 &lt;= mu2</pre>
--	---

**Gambar 1 :** (1a) gambar hasil uji normalitas (1b) gambar hasil uji homogenitas varians dan uji mean

Berdasarkan gambar di atas, dapat dilihat bahwa kedua data yang digunakan memiliki nilai mean sebesar 22831,13 dan 24023,65, dengan nilai  $\alpha$  yang digunakan sebesar 0,05. Standar deviasi pada data pertama sebesar 12577,51 dan pada data kedua sebesar 11598,19. Dalam asumsi uji normalitas, data pertama diasumsikan berdistribusi normal karena memiliki nilai  $p$ -value sebesar 0,6030086, dimana  $p$ -value tersebut lebih besar dari nilai  $\alpha$  sehingga memenuhi kriteria uji bahwa data tersebut berdistribusi normal. Hal ini juga berlaku untuk data kedua, dimana data kedua juga diasumsikan berdistribusi normal dikarenakan memiliki nilai  $p$ -value yang lebih besar dari nilai  $\alpha$ , yaitu sebesar 0,8401508 sehingga memenuhi kriteria uji bahwa data tersebut berdistribusi normal.

Untuk uji asumsi homogenitas varians sendiri, juga menunjukkan bahwa kedua data memiliki nilai varians yang sama. Hal ini dibuktikan dengan pemenuhan kriteria uji, dimana nilai  $p$ -value dari hasil uji varians lebih besar dari nilai  $\alpha$ , yaitu sebesar 0,659886. Maka dari itu, dapat ditarik kesimpulan bahwa kedua data memiliki nilai varians yang sama.

Dikarenakan semua uji asumsi telah terpenuhi maka data siap untuk dianalisis dengan menggunakan uji mean 2 populasi. Dalam uji mean 2 populasi, hasil dari perhitungan  $Z$ -hitung sebesar -0,0002746756 dan  $Z$ -tabel sebesar 1,644854. Maka dari itu, nilai dari  $Z$ -tabel lebih besar dari  $Z$ -hitung sehingga memenuhi kriteria uji bahwa rata-rata data pertama lebih kecil daripada rata-rata data kedua.

Berdasarkan hasil dari pengujian di atas, dapat disimpulkan bahwa nilai rata-rata FOB ekspor tahun 2020 lebih kecil daripada nilai rata-rata FOB ekspor tahun 2021. Rendahnya nilai rata-rata FOB ekspor di tahun 2020 menunjukkan bahwa pandemi Covid-19 berpengaruh terhadap nilai ekspor di Provinsi Jawa Timur, dimana pada masa itu setiap negara memiliki kebijakan pembatasan sektor ekspor dan impor untuk mengantisipasi terjadinya penularan Covid-19. Sedangkan, pada tahun 2021 nilai rata-rata FOB ekspor menunjukkan kenaikan. Hal ini dikarenakan pada tahun 2021 kebijakan *new normal* telah diberlakukan. Pemberlakuan masa *new normal* ini menjadi masa transisi setelah pandemi Covid-19, dimana kebijakan pemerintah akan kembali seperti semula. Oleh karena itu, sektor ekspor mulai bangkit kembali sehingga nilai rata-rata FOB ekspor di tahun 2021 meningkat.

**VI. KESIMPULAN**

Pandemi Covid-19 yang melanda dunia menyebabkan aktivitas ekspor antar negara menjadi terhambat. Salah satu daerah di Indonesia yang terdampak pandemi ini yaitu provinsi Jawa Timur. Berdasarkan hasil dan pembahasan sebelumnya dapat dilihat bahwa nilai rata - rata FOB ekspor di provinsi Jawa Timur tahun 2020 lebih kecil daripada tahun 2021. Rendahnya nilai ekspor di tahun 2020 ini disebabkan karena melonjaknya pandemi Covid-19 dan didukung pula dengan kebijakan pemerintah dalam membatasi ekspor untuk mengantisipasi penyebaran Covid-19. Sedangkan, pada tahun 2021, nilai rata-rata FOB ekspor mulai meningkat karena kebijakan *new normal* sudah diberlakukan. Kebijakan ini membuka kembali sektor ekspor secara lebih luas sehingga sektor ekspor mulai kembali optimal.

**UCAPAN TERIMA KASIH****REFERENSI**

1. Yip, B., & Perasso, V. (2021). *Asal Covid-19: Apakah Kita Perlu Tahu dari mana asal virus Corona Ini?*. BBC
2. D. Lubis, A. D. L. (2010). *Analisis Faktor Yang Mempengaruhi Kinerja Ekspor Indonesia*.
3. Kementerian Keuangan Direktorat Jenderal Perbendaharaan (2018). *Kajian Fiskal Regional Provinsi Jawa Timur Tahun 2018*.
4. Badan Pusat Statistik (BPS) Jawa Timur. (2018). *Nilai Ekspor Jawa Timur Menurut Kategori Migas dan Non Migas Bulanan (US\$), 2018-2020*. BPS provinsi Jawa Timur. ( <https://jatim.bps.go.id/indicator/8/44/2/nilai-ekspor-jawa-timur-menurut-kategori-migas-dan-non-migas-bulanan.html> )
5. Nurfitriani, I. (2022, August 23). *ANALISIS DATA MENGGUNAKAN UJI RATA-RATA DUA POPULASI INDEPENDEN DAN DEPENDEN DI RStudio*. Analisis data menggunakan uji rata-rata dua populasi Independen Dan Dependen di Rstudio. ([https://rstudio-pubs-static.s3.amazonaws.com/934402\\_e90fd1f55b114b2f96708593203c3389.html](https://rstudio-pubs-static.s3.amazonaws.com/934402_e90fd1f55b114b2f96708593203c3389.html) )
6. Hatta, H. (2010, November 5). *Uji Z*. STATISTIKA. (<https://hatta2stat.wordpress.com/2009/10/01/uji-z/#:~:text=Uji%20Z%20adalah%20salah%20satu%20uji%20statistika%20yang,samp el%2030%20atau%20lebih%20dianggap%20sampel%20berukuran%20besar>)
7. Long, T. J. (2022). *Uji Hipotesis Rata-Rata Dua Populasi*. Uji hipotesis rata-rata dua populasi. (<https://jagostat.com/metode-statistika-2/uji-hipotesis-rata-rata-dua-populasi>)
8. Hidayat, A. (2021, August 16). *Uji Normalitas Dan Metode Perhitungan (Penjelasan Lengkap)*. Uji Statistik. (<https://www.statistikian.com/2013/01/uji-normalitas.html> )
9. Vierda, A. (2023, February 14).  $\sqrt$  *Uji Normalitas: Pengertian, rumus, metode contoh SPSS*. Wiki Statistika. (<https://wikistatistika.com/uji-normalitas/>)
10. Hidayat, A. (2017, March 1). *Penjelasan Lengkap Uji homogenitas*. Uji Statistik. (<https://www.statistikian.com/2013/01/uji-homogenitas.html>)



# Analisis Perbandingan Tingkat Kemiskinan Di Jawa Timur Selama Dan Sesudah Krisis Pangan

Shafira Amanda Putri<sup>1</sup>, Nabilah Selayanti<sup>2</sup>, Penulis ketiga<sup>3</sup>

<sup>1,3</sup> Afiliasi pertama (jurusan Sains Data, UPN “Veteran” Jawa Timur)

<sup>1</sup> [22083010008@student.upnjatim.ac.id](mailto:22083010008@student.upnjatim.ac.id)

<sup>3</sup> [22083010013@student.upnjatim.ac.id](mailto:22083010013@student.upnjatim.ac.id)

<sup>2</sup> Afiliasi kedua (jurusan Sains Data, UPN “Veteran” Jawa Timur)

Corresponding author email: [penulis.kedua@universitas.ac.id](mailto:penulis.kedua@universitas.ac.id)

**Abstract:** Poverty poses a complex issue, particularly during times of economic downturns resulting from a food crisis. The aim of this research is to analyze the poverty levels in East Java during and after the food crisis. This study employs a quantitative comparative research design that seeks to compare the poverty levels before and after the food crisis. The secondary data utilized in this research includes the number of poverty cases in 2008 and 2009 in East Java. The analysis method employed is the dependent samples t-test, which allows for the comparison of poverty levels during and after the food crisis. The findings of this study reveal a significant difference in poverty levels in East Java during and after the food crisis.

**Keywords:** poverty, food crisis, secondary data, dependent samples t-test.

**Abstrak:** Kemiskinan menjadi masalah yang cukup rumit terutama pada masa keterpurukan ekonomi akibat krisis pangan. Tujuan penelitian ini untuk menganalisis tingkat kemiskinan di Jawa Timur selama dan sesudah krisis pangan. Jenis penelitian yang digunakan dalam penelitian ini adalah menggunakan penelitian kuantitatif komperatif yang bertujuan untuk membandingkan tingkat kemiskinan selama dan sesudah krisis pangan. Data sekunder yang digunakan dalam penelitian ini yang terkumpul meliputi jumlah tingkat kemiskinan tahun 2008 dan jumlah kemiskinan tahun 2009 di Jawa Timur. Metode analisis yang digunakan adalah uji mean 2 populasi dependen karena dengan metode tersebut dapat mengetahui perbandingan kemiskinan selama dan sesudah krisis pangan. Berdasarkan hasil penelitian menunjukkan adanya perbandingan yang signifikan dalam tingkat kemiskinan di Jawa Timur selama dan sesudah krisis pangan.

**Kata kunci:** kemiskinan, krisis pangan, data sekunder, uji mean 2 populasi.

## I. PENDAHULUAN

Kemiskinan merupakan permasalahan utama untuk negara di dunia, baik di negara maju maupun di negara berkembang. Kemiskinan bukan suatu masalah yang mudah pada negara berkembang bahkan tergolong sebagai masalah yang rumit. Indonesia termasuk negara berkembang yang memiliki masalah cukup rumit pada kemiskinan. Ketidakmampuan dalam segi ekonomi, segi material, dan fisik untuk mencukupi kebutuhan dasar pangan dan non pangan yang menggunakan pengeluaran sebagai tolak ukur kemiskinan.

Jawa Timur merupakan provinsi Indonesia yang terletak pada bagian timur pulau jawa dan memiliki ibu kota Surabaya. Luas dari Provinsi ini 48.033 kilometer persegi. Jawa Timur mempunyai posisi yang strategis di bidang industri salah satunya yaitu pertanian. Pada sektor pertanian Jawa Timur memiliki iklim yang mendukung serta mempunyai lahan yang luas sehingga menghasilkan seperti beras, jagung, dan buah-buahan yang cukup tinggi. Meskipun dengan lokasi strategis di bidang industri, banyak masyarakat di kabupaten Jawa Timur yang tergolong ke dalam masyarakat miskin. Karena kenyataannya petani di negara berkembang terjadi penurunan harga komoditas, dengan konsekuensi penurunan pendapatan dan keberlanjutan usaha tani.

Pada tahun 2008 terjadi peristiwa krisis pangan termasuk di Jawa Timur. Harga minyak terjadi penurunan dan diikuti dengan krisis finansial dengan penurunan pertumbuhan ekonomi global yang telah mendorong penurunan harga pangan, sehingga terjadilah peristiwa baru yaitu krisis pangan. Kenaikan harga dipicu dengan terjadinya krisis pangan dari krisis energi yang terjadi sehingga harga minyak dan bionergi juga mengalami kenaikan. Kenaikan harga tersebut berdampak pada ketahanan



pangan dan kemiskinan. Krisis finansial juga memberikan dampak pada kebutuhan dan investasi untuk infrastruktur pembangunan.

Untuk itu diperlukan suatu analisis tentang tingkat kemiskinan di Provinsi Jawa Timur selama krisis pangan dan sesudah krisis pangan dengan tujuan untuk membandingkan tingkat kemiskinan. Maka akan dilakukan penelitian apakah ada perbedaan pada tingkat kemiskinan ketika peristiwa krisis pangan pada tahun 2008 dengan sesudah peristiwa krisis pangan tahun 2009.

## II. METODE PENELITIAN

### 2.1. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini yaitu penelitian kuantitatif komparatif. Makna dari data “kuantitatif” yaitu data tersebut berupa informasi yang dinyatakan dengan bilangan dan dapat diukur atau dihitung secara langsung. Peneliti menggunakan jenis penelitian tersebut karena tujuan penelitian ini bersifat membandingkan keberadaan satu variabel pada sampel atau waktu yang berbeda.

### 2.2. Sumber data

Dalam penelitian ini data yang digunakan menggunakan data sekunder. Data sekunder yaitu data atau informasi yang diperoleh secara tidak langsung karena diperoleh melalui perantara seperti melalui orang lain atau lewat dokumen. Data sekunder ini bersumber dari laman resmi Badan Pusat Statistika tahun 2015

Data yang dianalisis dalam penelitian ini yaitu data dari laman resmi Badan Pusat Statistika tahun 2015 dengan menggunakan data per-Kabupaten/Kota di provinsi Jawa Timur dengan mengambil data pada tahun 2008-2009. Sampel yang digunakan dalam penelitian ini meliputi sampel kemiskinan tahun 2008 (selama krisis pangan) dan kemiskinan tahun 2009 (sesudah krisis pangan).

### 2.3. Metode Analisis

Dalam menganalisis perbandingan tingkat kemiskinan pada tahun 2008 (selama krisis pangan) dan tahun 2009 (sesudah krisis pangan). Metode analisis yang digunakan yaitu uji mean 2 populasi dependen. Karena dengan metode tersebut dapat mengetahui perbandingan kemiskinan selama dan sesudah krisis pangan. Untuk melakukan uji mean 2 populasi dependen menggunakan langkah-langkah seperti berikut :

#### 1. Analisis Deskriptif

Analisis deskriptif merupakan kumpulan data-data dalam metode penelitian dengan cara data tersebut disusun, diolah dan dianalisis yang dapat memberikan gambaran mengenai masalah yang ada. Pada penelitian ini menggunakan analisis deskriptif untuk menampilkan hasil dari data penduduk kemiskinan 2008 (selama krisis pangan) dan penduduk kemiskinan 2008 (sesudah krisis pangan). Dalam analisis deskriptif ini peneliti menyajikan pada kedua kelompok dalam bentuk tabel yang terdiri dari jumlah data, mean, median, varian, standar deviasi.

#### 2. Uji Normalitas

Uji normalitas dilakukan untuk memeriksa apakah data dalam setiap populasi berdistribusi normal atau tidak. Data bisa dikatakan berdistribusi normal ketika nilai  $p\text{-value} \geq 0,05$  (alpha). Uji normalitas ini dapat memastikan bahwa asumsi ini terpenuhi sebelum menerapkan metode statistik parametrik. Jika data berdistribusi normal, maka dapat menggunakan metode parametrik untuk mengambil hasil statistik yang akurat dan mengambil kesimpulan.

#### 3. Uji Homogenitas

Untuk mengetahui homogenitas pada data dua kelompok dengan dua variabel dependen maka dilakukan melalui uji Homogenitas. Uji ini digunakan untuk memeriksa apakah varian antara Kabupaten/Kota Jawa Timur pada tahun 2008 (selama krisis pangan) dan tahun 2009 (sesudah krisis pangan) adalah homogen. Uji homogenitas dalam konteks uji mean 2 populasi penting karena asumsi dasar dari beberapa metode statistik yaitu kedua populasi tersebut memiliki varians yang sama.

#### 4. Uji mean dua populasi

Dalam uji hipotesis dilakukan untuk mengetahui perbandingan rata-rata antara dua populasi yang saling bergantung. Uji ini digunakan untuk mengetahui apakah terdapat perbedaan yang signifikan antara Kabupaten/Kota Jawa Timur pada tahun 2008 (selama krisis pangan) dan tahun 2009 (sesudah krisis pangan). Dalam melakukan uji ini, terdapat hipotesis terlebih dahulu, yaitu:

- $H_0 = \mu_1 \geq \mu_2$  (rata - rata penduduk miskin 2008 lebih banyak dari rata rata penduduk miskin 2009)
- $H_1 = \mu_1 < \mu_2$  (rata - rata penduduk miskin 2008 lebih kecil daripada rata - rata penduduk miskin 2009)

### III. HASIL DAN PEMBAHASAN

#### 3.1. Analisis Deskriptif

Analisis deskriptif dalam penelitian ini digunakan untuk melihat gambaran tentang karakteristik dan perubahan tingkat kemiskinan antara tahun 2008 dan 2009 di Kabupaten/Kota Jawa Timur. Untuk lebih jelasnya dapat dilihat sebagai berikut :

**Tabel 1.** Hasil Analisis Deskriptif

Deskripsi	Penduduk Miskin 2008	Penduduk Miskin 2009
Jumlah data	38	38
Mean	172.35	154.2342
Median	160.8	146.4
Varian	11042.31	9116.319
Standar deviasi	105.0824	95.47941

Berdasarkan Tabel.1 dapat diketahui bahwa rata-rata *penduduk miskin 2008* yaitu sebesar 172.35 dengan nilai tengah 11042.31 dan variasi 1042.31 serta standar deviasi sebesar 105.0824. Sedangkan untuk *penduduk miskin 2009*, rata-ratanya sebesar 154.2342 dengan nilai tengah 146.4 dan variasi 9116.319 serta standar deviasi nya sebesar 95.47941. Berdasarkan hasil analisis deskriptif tersebut dapat diketahui bahwa tingkat kemiskinan di Kabupaten/Kota Jawa Timur pada tahun 2009 cenderung lebih rendah daripada tahun 2008. Hal ini mengalami penurunan penduduk miskin dari tahun 2008 ke tahun 2009, serta perubahan dalam sebaran data dengan variasi yang lebih rendah pada tahun 2009.

#### 3.2. Uji Normalitas

Uji normalitas digunakan untuk mengetahui apakah populasi pada data tersebut berdistribusi normal atau tidak. Pada uji normalitas ini menggunakan uji *Kolmogorov Smirnov* hasilnya akan dijelaskan pada tabel dibawah ini :

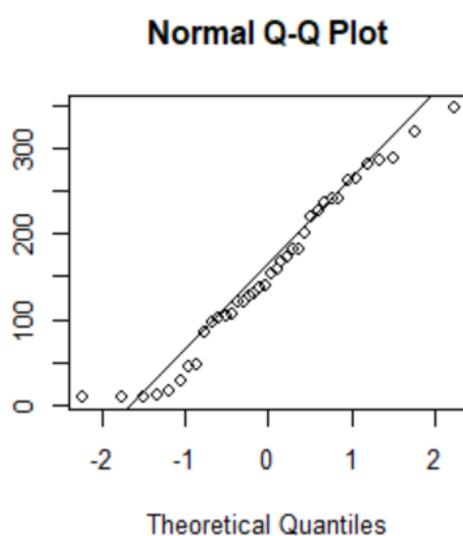
**Tabel 2.** Hasil Uji Normalitas

Deskripsi	Penduduk Miskin 2008	Penduduk Miskin 2009
-----------	----------------------	----------------------

Alpha	0.05	0.05
P-value	0.9589028	0.9538778
Kesimpulan	Data berdistribusi normal	Data berdistribusi normal

Berdasarkan hasil dari uji normalitas Tabel.1 menggunakan uji *Kolmogorov Smirnov* dapat diketahui bahwa p-value penduduk miskin 2008 (selama krisis pangan) yaitu sebesar 0.9589028 dan penduduk miskin 2009 (sesudah krisis pangan) yaitu sebesar 0.9538778. Karena nilai tingkat signifikan lebih dari 0,05 maka pada dua kelompok sampel tersebut dinyatakan berdistribusi normal.

Dari hasil Tabel.1 diatas, peneliti dapat membuat Q-Q Plot untuk masing-masing kelompok data penduduk miskin 2008 dan penduduk miskin 2009. Hasil dari plot bisa dilihat dibawah ini :



**Gambar 1.** Q-Q plot uji normalitas

Berdasarkan Gambar.1 normalitas pada data tersebut dapat diketahui dengan melihat persebaran data (titik) pada sumbu diagonal dari grafik. Pada grafik diatas menunjukkan bahwa data menyebar di sekitar garis diagonal dan mengikuti garis diagonal, sehingga dapat disimpulkan bahwa data tersebut dapat memenuhi asumsi normalitas.

### 3.3. Uji Homogenitas Varians

Salah satu syarat untuk melakukan uji mean 2 populasi yaitu melakukan uji homogenitas varians. Peneliti menggunakan uji F untuk mengetahui variasi data dalam ‘Penduduk Kemiskinan 2008’ harus serupa dengan variasi data dalam populasi ‘Penduduk Kemiskinan 2009’.

**Tabel 3.** Hasil Uji Homogenitas Varians

	<b>Penduduk Miskin 2008 dan 2009</b>
F	1.2113
num df	37
denom df	37
p-value	0.5629

Berdasarkan Tabel.3 diatas, dari hasil uji homogenitas varian terdapat dua kelompok data yang dibandingkan, yaitu penduduk miskin 2008 dan penduduk miskin 2009. Dapat diketahui dari hasil diatas bahwa nilai F diperoleh sebesar 1.2113 dengan derajat kebebasan numerator (num df) dan derajat kebebasan denominator (denom df) masing-masing yaitu 37 serta p-value sebesar 0.5629.

Dalam hal ini menunjukkan bahwa nilai signifikansi pada masing-masing kelompok lebih dari 0,05. Maka dapat disimpulkan bahwa antara kelompok penduduk miskin 2008 (selama krisis pangan) dan kelompok miskin 2009 (sesudah krisis pangan) bersifat homogen, sehingga dapat dilanjutkan dengan melakukan uji mean 2 populasi.

### 3.4. Uji Mean 2 Populasi

Uji mean 2 populasi ini bertujuan untuk mengetahui apakah ada perbandingan yang signifikan ketika penduduk miskin 2008 (selama krisis pangan) dan penduduk miskin 2009 (sesudah krisis pangan). Uji mean 2 populasi dilakukan dengan menggunakan statistika uji z.

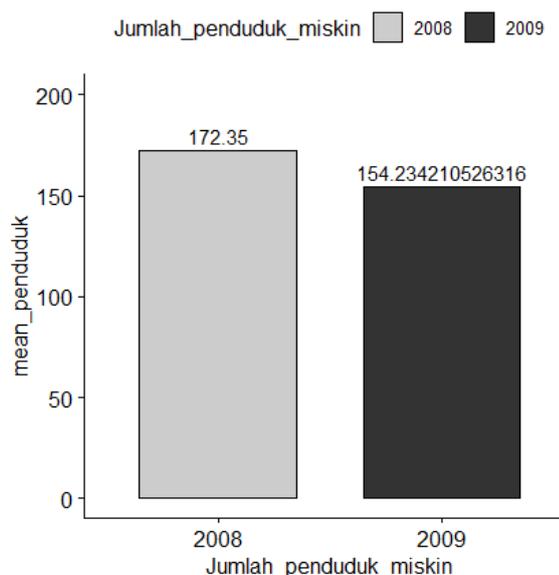
**Tabel 4.** Hasil Uji Mean 2 Populasi

Penduduk Miskin 2008 dan 2009		
Zhitung	-343.3985	
Ztabel	1.959964	
Keputusan	tolak $H_0$	
Kesimpulan	$\mu_1 \geq \mu_2$	

Berdasarkan Tabel.4 dari pengujian tersebut dapat diketahui bahwa nilai dari  $Z_{hit}$  sebesar -343.3985 sedangkan  $Z_{tab}$  sebesar 1.959964 dengan kriteria uji yaitu tolak  $H_0$ . Karena pada hasil tersebut nilai  $Z_{hitung} < Z_{tabel}$  dapat diambil keputusan yaitu tolak  $H_0$ .

Maka dapat disimpulkan bahwa  $\mu_1 \geq \mu_2$  yaitu *rata - rata penduduk miskin 2008* (selama krisis pangan) lebih banyak dari *rata rata penduduk miskin 2009* (sesudah krisis pangan).

Hal tersebut dapat disimpulkan berdasarkan hasil pedoman pengambilan keputusan berdasarkan nilai probabilitas bahwa gagal tolak  $H_0$  dan  $H_1$  diterima yang artinya bahwa terdapat perbandingan yang signifikan antara rata - rata jumlah penduduk miskin 2008 (selama krisis pangan) lebih banyak dari rata rata penduduk miskin 2009 (sesudah krisis pangan).



**Gambar 2.** bar chart uji mean 2 populasi

Berdasarkan Gambar.2 dapat diketahui bahwa rata-rata jumlah penduduk miskin pada tahun 2008 (selama krisis pangan) lebih tinggi daripada jumlah penduduk miskin 2009 (sesudah krisis pangan). Berdasarkan hasil analisis deskriptif juga diketahui bahwa



mean dari kelompok penduduk miskin 2008 yaitu sebesar 172.35 dan mean dari kelompok penduduk miskin 2009 sebesar 154.2342.

Hal ini menunjukkan bahwa terdapat perbandingan yang signifikan pada saat peristiwa krisis pangan di Provinsi Jawa Timur yaitu pada tahun 2008 (selama krisis pangan) dan tahun 2009 (sesudah krisis pangan).

#### **IV. KESIMPULAN**

Template ini adalah versi pertama untuk SENADA. Format ini seharusnya menjadi acuan dalam mengajukan submission SENADA.

#### **UCAPAN TERIMA KASIH**

Ucapan terima kasih disampaikan kepada mata kuliah bahasa Indonesia yang diselenggarakan oleh UPN “Veteran” Jawa Timur sehingga artikel ini bisa ditulis dengan baik dan benar.

#### **REFERENSI**

- 1.



## Analisis Keterkaitan Jumlah Rumah Ibadah berdasarkan Penganutnya

Ajeng Puspa Wardani<sup>1</sup>, Maulidya Prastita Syah<sup>2</sup>, Trimono<sup>3</sup>

1, 2, 3 Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran Jawa Timur"

<sup>1</sup>[22083010040@studentupnjatim.ac.id](mailto:22083010040@studentupnjatim.ac.id)

<sup>2</sup>[22083010039@studentupnjatim.ac.id](mailto:22083010039@studentupnjatim.ac.id)

<sup>3</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

Corresponding author email: [22083010039@studentupnjatim.ac.id](mailto:22083010039@studentupnjatim.ac.id)

**Abstract:** This study aims to analyze the relationship between the number of houses of worship and the number of religious adherents in Indonesia. The data used was obtained from sources of religious institutions. This study uses two methods, namely descriptive analysis to provide an overview and simple linear regression analysis to determine whether there is a relationship between the number of houses of worship and the number of religious adherents. The independent variable in this study is the number of houses of worship, while the dependent variable is the number of religious adherents. The results show that the correlation between the number of houses of worship and the number of religious adherents is not very significant. Other factors influence this correlation, such as traditions, places of worship that are not recorded, and so on. The results of this study can be used as a reference for evaluating the government and religious institutions in planning the construction of houses of worship and religious policies.

**Keywords:** Number of Houses of Worship, Religious Adherents, Descriptive Analysis, Regression Analysis, Influence Factors.

**Abstrak:** Penelitian ini bertujuan untuk menganalisis keterkaitan antara jumlah rumah ibadah dengan jumlah penganut agama di Indonesia. Data yang digunakan diperoleh dari sumber lembaga keagamaan. Penelitian ini menggunakan dua metode yaitu analisis deskriptif untuk memberikan gambaran secara umum dan analisis regresi linier sederhana untuk menentukan apakah terdapat hubungan antara jumlah rumah ibadah dengan jumlah penganut agama. Variabel independen dalam penelitian ini adalah jumlah rumah ibadah, sedangkan variabel dependen adalah jumlah penganut agama. Hasil menunjukkan bahwa korelasi antara jumlah rumah ibadah dan jumlah penganut agama tidak terlalu signifikan. Adanya faktor-faktor lain yang mempengaruhi korelasi tersebut seperti tradisi, rumah ibadah yang tidak terdata dan lain-lain. Hasil penelitian ini dapat menjadi acuan evaluasi bagi pemerintah dan lembaga keagamaan dalam perencanaan pembangunan rumah ibadah dan kebijakan agama.

**Kata Kunci:** Jumlah Rumah Ibadah, Penganut Agama, Analisis Deskriptif, Analisis Regresi, Faktor-Faktor Pengaruh.

### I. PENDAHULUAN

Indonesia adalah negara majemuk yang kaya akan keragaman suku, bahasa, dan kepercayaan. Dengan lebih dari 300 suku, 1.340 etnik suku, dan 829 bahasa yang dituturkan, serta lebih dari 100 kepercayaan, Indonesia merupakan bukti nyata suatu kemajemukan. Meskipun demikian, dalam dasar hukum Indonesia, hanya enam aliran agama yang diakui secara resmi, yaitu agama Islam, Kristen, Katolik, Hindu, Buddha, dan Konghucu. Hal ini dapat menjadi bukti bahwa kemajemukan negara Indonesia terjalin dengan utuh dan berdaulat.

Pemerintah memiliki kewajiban untuk mengatur tentang aspek-aspek dalam rangka menciptakan keseimbangan antar umat beragama di Indonesia. Agar terciptanya harmoni pemerintah salah satunya melakukan penetapan terhadap lembaga keagamaan, seperti masjid, gereja, pura, dan vihara. Penetapan tersebut dilakukan agar terselenggaranya aktivitas Ibadah yang teratur. Selain itu, penetapan tempat ibadah juga membantu dalam pemetaan kehadiran agama di suatu wilayah.

Penelitian ini bertujuan untuk menganalisis keterkaitan antara jumlah rumah ibadah dengan jumlah penganut agama di suatu daerah di Indonesia. data yang digunakan meliputi informasi tentang jumlah rumah ibadah yang ada di Indonesia serta data tentang jumlah penganut agama di Indonesia. Metode analisis yang digunakan dalam penelitian ini melibatkan dua metode yaitu analisis deskriptif dan analisis regresi. Analisis deskriptif akan memberikan gambaran umum tentang distribusi jumlah rumah ibadah dan penganut agama di daerah yang diteliti. Sementara itu, analisis regresi akan digunakan untuk memahami keterkaitan atau korelasi antara jumlah rumah ibadah dan penganut agama. Metode ini akan



membantu mengidentifikasi apakah ada hubungan positif atau negatif antara kedua variabel tersebut. Analisis regresi juga akan memberikan gambaran tentang seberapa kuat hubungan tersebut dan apakah hubungan tersebut signifikan.

Penelitian ini memiliki implikasi penting dalam pemahaman tentang distribusi agama di suatu daerah dan kebutuhan masyarakat terkait tempat ibadah. Hasil penelitian ini dapat menjadi dasar untuk merencanakan kebijakan dan pembangunan yang memperhatikan kebutuhan agama dalam masyarakat. Dengan demikian, analisis keterkaitan antara jumlah rumah ibadah dan penganut agama ini memiliki nilai penting dalam konteks menghormati dan mempromosikan kebebasan beragama serta menciptakan harmoni antar agama di Indonesia.

## II. METODE PENELITIAN

### 2.1. Tinjauan Pustaka

#### 2.1.1. TOLERANSI UMAT BERAGAMA DAN PENDIRIAN RUMAH IBADAH ( Studi kasus di Gayo Lues)

Indonesia merupakan negara yang multikultural, yaitu negara yang memiliki beraneka ragam suku, budaya dan juga agama. Masyarakat Indonesia juga merupakan masyarakat yang majemuk, yaitu suatu kondisi dimasyarakat yang terdiri dari berbagai perbedaan strata, ekonomi, ras, suku, bangsa, agama dan budaya. Seringkali perbedaan dan keragaman ini rentan akan konflik, terutama konflik yang diakibatkan oleh sikap intoleransi. Konflik yang timbul akibat sikap toleransi ini akan merusak nilai-nilai kerukunan yang ada di Indonesia terutama kerukunan antar umat beragama. Tujuan dari penelitian ini adalah untuk mengetahui bagaimana penerapan nilai-nilai toleransi umat beragama dan pendirian rumah ibadah yang ada di Kabupaten Gayo Lues. Dalam penelitian ini penulis menggunakan metode kualitatif (*field research*) yang berbasis pada penelitian lapangan dan berfokus pada pengamatan yang mendalam, metode ini digunakan untuk meneliti pada kondisi obyek yang alamiah dan menggunakan sistem pengumpulan data dengan tahapan observasi, wawancara dan kemudian dokumentasi. Berdasarkan dari hasil penelitian menunjukkan bahwa nilai-nilai toleransi yang diterapkan masyarakat Gayo Lues berupa rasa hormat, menghargai dan menerima perbedaan, kemudian minimnya kasus atau konflik yang terjadi di Gayo Lues menunjukkan bahwa masyarakat Gayo Lues sudah menerapkan nilai-nilai toleransi. Sejauh ini belum pernah terjadi kasus atau konflik antar umat beragama di Gayo lues.

#### 2.1.2. KEKRISTENAN INDONESIA DI ERA POSTMODERN DITILIK DARI JUMLAH PENGANUT DAN JUMLAH RUMAH IBADAH

Seperti diketahui kehidupan di dunia manusia dipengaruhi oleh perkembangan zaman, dimana zaman terus mengalami perubahan, dari zaman kuno, abad pertengahan, pra modern menuju ke modern dan sekarang memasuki era *postmodern*. Era *postmodern* juga diikuti oleh pemikiran filosofis dari *postmodernisme*. Pemikiran filosofis tersebut telah jauh mempengaruhi kehidupan banyak manusia, yang juga telah menyentuh dalam aspek iman Kristen (Made Nopen Supriadi, dalam Tinjauan Teologi Terhadap *Postmodernisme* dan Implikasinya bagi Iman Kristen, 7 Maret 2020). Melihat fenomena yang cukup menarik tersebut, melalui Jurnal ini penulis tertarik untuk mengulas masalah ini yang tertuang dalam “Kekristenan Indonesia di era *Postmodern* ditilik dari Jumlah Penganut dan Jumlah Rumah Ibadah. Tujuan penulisan ini dimaksudkan untuk menjabarkan secara jelas sejauhmana perkembangan jumlah umat Kristen di Indonesia dan juga perkembangan jumlah rumah ibadahnya di era *postmodern*. Melalui kedua indikasi tersebut yaitu jumlah penganut dan jumlah rumah



ibadah, maka akan diketahui dan terlihat jelas sejauh mana kondisi kekristenan di Indonesia. Sampai sekarang jumlah rumah ibadah umat Islam dalam hal ini Masjid dan Musholla tetap memiliki porsi yang paling besar yang diperkirakan mencapai 781,2 ribu buah pada tahun 2016 kemudian naik terus dan mencapai 815,8 ribu buah pada tahun 2020. Dari uraian tersebut di atas maka dapat disimpulkan bahwa ke kristenan di Indonesia walaupun banyak menghadapi tantangan di era *postmodern* ini, namun ke kristenan di Indonesia tetap terus berkembang. Hal ini terlihat jelas dari perkembangan jumlah umat Kristen yang setiap tahunnya terus mengalami pertumbuhan. Perkembangan ini juga terlihat jelas dari pertumbuhan jumlah rumah ibadah yang ada dalam hal ini gereja. Indikasi lainnya yang juga menunjukkan bahwa ke kristenan di Indonesia tampak terus berkembang juga bisa dilihat dari berita-berita yang ada di mas media, di mana belakangan ini banyak orang-orang yang bertobat dan mengalami penjamahan dari Tuhan Yesus.

## 2.2. Data yang digunakan

Data yang kami gunakan untuk memenuhi penelitian dalam pembuatan artikel ini meliputi:

2.2.1. Data jumlah rumah ibadah pada tahun 2021. Data ini kami ambil dari situs Satu data - Kementerian agama RI (<https://satudata.kemenag.go.id/dataset/detail/jumlah-rumah-ibadah>) yang memaparkan data mengenai jumlah rumah ibadah, dengan penjelasan variabel data yang meliputi : Provinsi, masjid, gereja\_kristen, gereja\_katolik, pura, vihara, klenteng, jumlah, dan tahun.

2.2.2. Data jumlah penduduk menurut agama pada tahun 2021. Data ini kami ambil dari situs Satu data - Kementerian agama RI (<https://satudata.kemenag.go.id/dataset/detail/jumlah-penduduk-menurut-agama>) yang memaparkan data mengenai jumlah penduduk menurut agama, dengan penjelasan variabel data meliputi: provinsi, islam, kristen, katolik, hindu, buddha, khonghucu, tahun, dan variabel lainnya berupa jumlah penduduk penganut kepercayaan per provinsi.

## 2.3. Metode Analisis

### 2.3.1. Proses Pengambilan Data.

Dalam proses pengambilan data kami menggunakan metode *Import Data*. *Import data* merupakan salah satu metode pengambilan data yang fokus memindahkan data dari sumbernya ke dalam format yang dapat digunakan dan dimanfaatkan dalam alat analisis yang sedang digunakan. Data yang kami *import* berformat excel kemudian disimpan dalam *file* dengan ekstensi .xls atau .xlsx. Dengan metode *import*, prosedur pengambilan data menjadi lebih efisien sehingga dapat menghemat waktu.

### 2.3.2. Proses Analisis data.

Dalam proses analisis, kami menggunakan dua metode yaitu analisis deskriptif dan analisis regresi dua variabel. Analisis deskriptif adalah suatu metode statistik yang bertujuan untuk memberikan pemahaman yang lebih baik tentang karakteristik, pola, dan distribusi data yang diamati. Dengan analisis deskriptif dapat memberikan gambaran awal data sebelum melanjutkan ke metode analisis yang lebih lanjut. Gambaran awal yang diberikan dapat berupa statistik seperti ukuran pemusatan (rata-rata, median, modus) dan ukuran penyebaran (jangkauan, simpangan baku, dan kuartil). Dengan gambaran awal ini diharapkan dapat pemahaman yang lebih komprehensif terkait data yang akan di analisis. Analisis deskriptif juga memungkinkan untuk membuat hipotesis awal sebagai pembanding hasil penelitian di akhir.



Metode analisis regresi dua variabel yang juga dikenal sebagai regresi linear sederhana bertujuan untuk mendapatkan hasil hubungan matematis antara variabel bebas dan variabel terikat, menentukan besarnya perubahan variabel terikat jika variabel bebas mengalami kenaikan, dan nilai dari variabel terikat jika variabel bebas sama dengan nol. Prinsip utama dari persamaan regresi sederhana ini adalah kedua variabel yang digunakan harus memiliki hubungan kausalitas baik secara teoritis maupun secara logis berdasarkan penelitian terdahulu. Artinya data yang digunakan harus memiliki alasan yang kuat bahwa variabel bebas akan memengaruhi variabel terikat. Dalam pemodelan analisis regresi penting untuk mencari data yang berhubungan secara teoritis guna memberikan kerangka kerja yang baik untuk memahami hubungan kedua variabel dan memberikan dasar yang kuat untuk melakukan analisis lanjutan.

### 2.3.3. Bahasa Pemrograman

Dalam penelitian ini, kami menggunakan bahasa pemrograman R dalam lingkungan *software* R studio. Bahasa pemrograman R menyediakan banyak fungsi dan *package* yang mendukung analisis statistik seperti deskriptif, regresi, dan uji hipotesis. Selain itu, R dapat mendukung pembuatan dalam visualisasi data yang informatif seperti visualisasi grafik batang, grafik garis, grafik titik, dan lain lainnya. Dengan menggunakan bahasa pemrograman R kami dapat melakukan analisis dan visualisasi data dengan mudah dan jelas.

## III. HASIL DAN PEMBAHASAN

Data masukan yang digunakan dalam analisis regresi dua variabel sebagai berikut:

**Tabel 1.** Data Jumlah Penduduk Menurut Agama

Provinsi	Islam	Kristen	Katolik	Hindu	Buddha	Khonghucu
Aceh	5.271.485	63.486	5.511	95	7.045	1
Sumatera Utara	10.124.754	4.085.380	654.764	16.091	355.447	770
Sumatera Barat	5.470.109	83.794	46.773	95	3.415	5
Riau	5.726.306	638.119	69.867	767	136.542	2.204
Jambi	3.425.742	119.044	20.94	488	34.293	791
Sumatera Selatan	8.324.093	82.875	49.954	41.121	67.588	97
Bengkulu	1.989.720	32.935	8.038	4.138	2.082	11
Lampung	8.531.111	122.692	77.755	125.1	24.507	129
Kep. Bangka Belitung	1.316.560	30.755	19.058	1.225	63.778	29.378
Kepulauan Riau	1.631.245	248.235	52.191	927	146.626	3.254



DKI Jakarta	9.442.139	963.715	437.996	20.413	395.365	1.698
Jawa Barat	46.923.543	865.382	299.89	17.082	98.753	12.111
Jawa Tengah	36.296.971	594.871	347.772	14.618	51.276	1.344
DI Yogyakarta	3.415.882	89.454	165.191	3.421	3.07	72
Jawa Timur	39.925.335	681.474	275.735	104.987	71.198	2.065
Banten	11.410.170	314.182	145.206	8.566	142.9	2.267
Bali	430.924	69.602	34.852	3.714.068	29.023	560
Nusa Tenggara Barat	5.260.683	13.539	10.021	130.966	16.911	44
Nusa Tenggara Timur	518.916	1.987.688	2.941.807	5.803	378	24
Kalimantan Barat	3.290.832	633.174	1.210.295	2.788	313.504	15.095
Kalimantan Tengah	1.970.660	443.05	87.325	151.445	2.87	178
Kalimantan Selatan	3.996.958	55.005	22.523	23.823	12.185	202
Kalimantan Timur	3.365.718	288.806	170.453	8.552	15.648	345
Kalimantan Utara	511.874	138.437	43.064	337	4.122	148
Sulawesi Utara	845.194	1.672.955	118.108	15.793	3.869	428
Sulawesi Tengah	2.408.937	496.912	27.523	110.87	4.222	27
Sulawesi Selatan	8.263.575	692.908	152.282	63.499	20.636	76
Sulawesi Tenggara	2.565.710	44.795	16.215	50.814	1.528	15
Gorontalo	1.176.947	17.517	1.114	4.115	945	5
Sulawesi Barat	1.208.735	192.483	15.652	19.51	419	6
Maluku	991.664	743.412	129.179	6.601	366	63
Maluku Utara	986.743	330.052	6.749	114	134	127



Papua	627.581	2.995.059	675.154	3.139	2.068	74
Papua Barat	438.841	619.802	89.441	1.189	895	21

**Tabel 2. Jumlah Rumah Ibadah**

Provinsi	Masjid	Gereja Kristen	Gereja Katolik	Pura	Vihara	Klenteng
Aceh	4,267	188	19	1	11	0
Sumatera Utara	10,532	15,798	2,156	36	353	60
Sumatera Barat	5,307	340	134	1	8	0
Riau	6,853	3,012	243	5	308	67
Jambi	4,244	497	72	3	55	12
Sumatera Selatan	9,304	609	201	25	182	5
Bengkulu	3,097	173	41	19	10	0
Lampung	11,948	868	355	337	182	1
Kep. Bangka Belitung	1,011	215	31	13	225	47
Kepulauan Riau	1,907	398	71	5	205	12
DKI Jakarta	3,419	1,291	73	23	331	8
Jawa Barat	58,421	2,295	109	29	187	42
Jawa Tengah	50,23	2,876	379	101	541	30
DI Yogyakarta	8,127	297	199	27	24	1
Jawa Timur	48,17	2,749	287	400	162	23
Banten	8,716	935	35	13	116	10
Bali	249	486	48	7,397	50	19
Nusa Tenggara Barat	5,117	93	10	138	54	1



Nusa Tenggara Timur	901	6,781	3,432	31	3	0
Kalimantan Barat	4,252	3,471	2,758	18	965	209
Kalimantan Tengah	2,262	2,111	189	232	12	2
Kalimantan Selatan	2,765	319	84	106	26	1
Kalimantan Timur	3,13	1,616	271	48	28	2
Kalimantan Utara	661	593	133	3	10	1
Sulawesi Utara	1,022	5,548	294	34	33	6
Sulawesi Tengah	3,794	2,262	168	73	20	0
Sulawesi Selatan	14,046	2,743	424	59	39	2
Sulawesi Tenggara	3,468	316	60	187	18	0
Gorontalo	2,567	198	10	5	4	0
Sulawesi Barat	2,659	1,144	101	58	3	0
Maluku	1,305	2,559	297	7	8	0
Maluku Utara	1,097	1,109	70	1	1	1
Papua	395	6,342	832	21	19	0
Papua Barat	388	2,001	163	9	6	0

Kedua tabel tersebut menjadi acuan dalam penelitian kali ini. Dengan tabel penduduk berdasarkan agama sebagai variabel bebas dan tabel rumah ibadah sebagai variabel terikat. Secara umum persamaan yang digunakan dalam perhitungan ini adalah sebagai berikut:

$$Y = a + bX$$

Keterangan:

Y = Variabel dependen (variabel bebas)

X = Variabel independen (variabel terikat)

a = Koefisien *intercept* (nilai Y saat X=0)

b = Koefisien *slope* (kecenderungan nilai Y)

Untuk nilai koefisien *intercept* (a) dan koefisien *slope* (b) sendiri didapatkan melalui perhitungan data X dan Y yang di-input-kan, persamaanya sebagai berikut:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$
$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Sebelum melakukan analisis regresi kami melakukan analisis deskriptif guna mengetahui gambaran data secara singkat. Dari *output* yang dihasilkan dapat dilihat kolom provinsi memiliki panjang 34 dengan data bertipe *character*. Kemudian untuk kolom masjid, gereja Kristen, gereja Katolik, pura, vihara, dan klenteng terdapat informasi mengenai nilai terendah dan nilai tertinggi, kuartil satu sampai tiga, dan rata-rata. Untuk *output* keseluruhan sebagai berikut:

provinsi	masjid	gereja_kristen	gereja_katolik	pura
Length:34	Min. : 1.011	Min. : 1.109	Min. : 2.156	Min. : 1.000
Class :character	1st Qu.: 2.848	1st Qu.: 2.605	1st Qu.: 42.750	1st Qu.: 7.099
Mode :character	Median : 4.692	Median : 11.290	Median :105.000	Median : 24.000
	Mean : 84.619	Mean :193.844	Mean :159.157	Mean : 61.041
	3rd Qu.: 13.521	3rd Qu.:318.250	3rd Qu.:232.500	3rd Qu.: 58.750
	Max. :901.000	Max. :935.000	Max. :832.000	Max. :400.000
vihara	klenteng			
Min. : 1.00	Min. : 0.00			
1st Qu.: 10.25	1st Qu.: 0.00			
Median : 30.50	Median : 1.50			
Mean :123.50	Mean : 16.53			
3rd Qu.:182.00	3rd Qu.: 12.00			
Max. :965.00	Max. :209.00			

**Gambar 1.** Analisis Deskriptif

Dengan menggunakan analisis regresi kami memperoleh hasil hubungan antara agama dan tempat ibadahnya serta memprediksi nilai tempah ibadahnya (Y) berdasarkan nilai agama (X). Berikut pembahasan *output* tiap agama:

#### 1. Islam

```
> regresi(islam1, masjid1)
Intercept (a): 103933.1
Slope (b): -0.002758204
Nilai R-squared: 0.0230454
```

**Gambar 2.** Islam & Masjid

Dari analisis regresi dua variabel antara agama Islam dengan masjid didapati nilai *a* atau koefisien *intercept* bernilai 103933.1, artinya nilai *Y* yang diperkirakan adalah 103933.1 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar -0.002758204. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 2.3% variasi data dalam variabel ‘masjid’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 103933.1 - 0.002758204X$ .

## 2. Kristen

```
> regresi(kristen1, gereja_kristen1)
Intercept (a): 263896.3
Slope (b): -0.1164588
Nilai R-squared: 0.1532065
```

**Gambar 3.** Kristen & Gereja Kristen

Dari analisis regresi dua variabel antara agama Kristen dengan gereja Kristen didapati nilai *a* atau koefisien *intercept* bernilai 263896.3, artinya nilai *Y* yang diperkirakan adalah 263896.3 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar -0.1164588. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 15.3% variasi data dalam variabel ‘gereja kristen’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 263896.3 - 0.1164588X$ .

## 3. Katolik

```
> regresi(katolik1, gereja_katolik1)
Intercept (a): 163102
Slope (b): -0.015913
Nilai R-squared: 0.002587382
```

**Gambar 4.** Katolik & Gereja Katolik

Dari analisis regresi dua variabel antara agama Katolik dengan gereja Katolik didapati nilai *a* atau koefisien *intercept* bernilai 163102, artinya nilai *Y* yang diperkirakan adalah 163102 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar -0.015913. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 0.025% variasi data dalam variabel ‘gereja katolik’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 163102 - 0.015913X$ .

## 4. Hindu

```
> regresi(hindu1, pura1)
Intercept (a): 61972.05
Slope (b): -0.00677419
Nilai R-squared: 0.002035653
```

**Gambar 5.** Hindu & Pura

Dari analisis regresi dua variabel antara agama Hindu dengan pura didapati nilai *a* atau koefisien *intercept* bernilai 61972.05, artinya nilai *Y* yang diperkirakan adalah 61972.05 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar -0.00677419. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 0.02% variasi data dalam variabel ‘pura’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 61972.05 - 0.00677419X$ .

## 5. Buddha

```
> regresi(buddha1, vihara1)
Intercept (a): 113.7101
Slope (b): 6.441368e-05
Nilai R-squared: 0.005985417
```

**Gambar 6.** Buddha & Vihara

Dari analisis regresi dua variabel antara agama Buddha dengan vihara didapati nilai *a* atau koefisien *intercept* bernilai 113.7101 artinya nilai *Y* yang diperkirakan adalah 113.7101 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar 6.441368e-05. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang sangat lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 0.05% variasi data dalam variabel ‘vihara’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 113.7101 - 0.000064X$ .

## 6. Khonghucu

```
> regresi(khonghucu1, klenteng1)
Intercept (a): 16.02908
Slope (b): 3.966804e-06
Nilai R-squared: 0.0004696839
```

**Gambar 7.** Khonghucu & Klenteng

Dari analisis regresi dua variabel antara agama Khonghucu dengan klenteng didapati nilai *a* atau koefisien *intercept* bernilai 16.02908 artinya nilai *Y* yang diperkirakan adalah 16.02908 ketika  $X=0$ . Koefisien *slope* atau kecenderungan nilai *Y* jika terjadi penambahan adalah sebesar 3.966804e-06. Rendahnya nilai koefisien *slope* menunjukkan hubungan yang sangat lemah antara kedua variabel ini. Kemudian nilai *r* kuadrat pada *output* di atas menunjukkan bahwa sekitar 0.004% variasi data dalam variabel ‘klenteng’ dapat dijelaskan dengan model regresi ini. Maka persamaan untuk analisis regresi ini adalah  $Y = 16.02908 - 0.0000039X$ .



Dari pembahasan *output* di atas dapat dilihat bahwa jumlah penganut suatu agama tidak berpengaruh secara signifikan terhadap peningkatan jumlah rumah ibadah. Kami menyimpulkan hal ini dipengaruhi beberapa hal seperti pelaksanaan ibadah yang dapat dilaksanakan di luar tempat ibadah itu sendiri. Contohnya seorang umat muslim dapat melakukan ibadah di rumahnya masing-masing, selain itu pada data yang kami dapatkan hanya terdapat jumlah masjid dan belum tercantum jumlah mushollanya. Kemudian tempat ibadah juga ada yang dibangun karena faktor sosial dan budaya, seperti tradisi keagamaan dalam masyarakat, tingkat kereligiusan masyarakat dalam beragama, dan lain sebagainya juga memengaruhi jumlah tempat ibadah di suatu daerah. Faktor distribusi geografi penganut agama juga dapat memengaruhi jumlah tempat ibadah di daerah tersebut. Contoh kasus jika di suatu daerah memiliki wilayah yang cukup luas dan distribusi penganut agamanya beragam maka besar kemungkinan untuk dibangun berbagai tempat ibadah di daerah tersebut.

#### IV. KESIMPULAN

Dalam penelitian ini dilakukan analisis deskriptif dan analisis regresi satu variabel untuk mengkaji keterkaitan antara jumlah penganut agama dan jumlah tempat ibadah. Melalui penelitian ini penggunaan analisis deskriptif difokuskan untuk melihat variasi jumlah penganut dan tempat ibadah pada beberapa provinsi. Hasil yang didapatkan menunjukkan bahwa terdapat variasi yang signifikan untuk tiap provinsi. Kemudian untuk analisis regresi difokuskan untuk melihat keterkaitan antara jumlah penganut dengan jumlah tempat ibadah pada suatu provinsi. Hasil analisis mengindikasikan bahwa tingkat keterkaitan atau hubungan antara jumlah penganut dengan jumlah tempat ibadah cukup rendah. Kami menyimpulkan bahwa beberapa faktor seperti sosial, budaya, maupun kepentingan penganut itu sendiri memberikan pengaruh yang signifikan terhadap jumlah tempat ibadah. Dari penelitian ini kami ingin memberikan pandangan terkait konteks yang diujikan bahwa hubungan antara penganut sebuah agama dan tempat ibadahnya tidak hanya bergantung pada faktor kuantitas penganut itu sendiri melainkan juga ada faktor eksternal yang memengaruhi.

#### UCAPAN TERIMA KASIH

Terima kasih kami ucapkan untuk Tim SENADA yang telah meluangkan waktunya untuk menyiapkan acara dan *template* artikel yang luar biasa ini. Terima Kasih juga kami sampaikan kepada bapak Trimono selaku dosen pembimbing dalam pengerjaan artikel ini. Tak lupa juga kami sampaikan terima kasih kepada pihak-pihak terkait yang membantu pembuatan artikel ini.

#### REFERENSI

1. Keragaman Indonesia - Indonesia.go.id. Available online: (<https://www.indonesia.go.id/ragam/budaya/kebudayaan/keragaman-indonesia>). (accessed on 25 May 2023)
2. Suku Bangsa - Indonesia.go.id. Available online: (<https://indonesia.go.id/profil/suku-bangsa/kebudayaan/suku-bangsa>). (accessed on 25 May 2023)
3. Mengenal 7 Kepercayaan di Indonesia yang Ada Sejak Ratusan Tahun Lalu. Available online: (<https://kumparan.com/berita-heboh/mengenal-7-kepercayaan-di-indonesia-yang-ada-sejak-ratusan-tahun-lalu-1sT4jfEWrkM>). (accessed on 25 May 2023)
4. 6 Agama yang Diakui di Indonesia - Gramedia Literasi. Available online: (<https://www.gramedia.com/literasi/agama-yang-diakui-di-indonesia/>). (accessed on 25 May 2023)



5. Lismeiliani, Rini. Toleransi Umat Beragama dan Pendirian Rumah Ibadah ( Studi kasus di Gayo Lues), Universitas Islam Negeri Ar-Raniry Darussalam, Banda Aceh, 25 November 2022
6. Wiyono, Harry. Kekristenan Indonesia di Era Postmodern Ditilik dari Jumlah Penganut dan Jumlah Rumah Ibadah, Sekolah Tinggi Teologi IKAT Jakarta, Jakarta, 2022.
7. Satu Data - Kementerian Agama RI. Available online: (<https://satudata.kemenag.go.id/dataset/detail/jumlah-penduduk-menurut-agama>) (accessed on 26 May 2023)
8. Satu Data - Kementerian Agama RI. Available online: (<https://satudata.kemenag.go.id/dataset/detail/jumlah-rumah-ibadah>). (accessed on 26 May 2023)
9. Pratomo, D.S.; Astuti, E.Z. Analisa Regresi I dan Korelasi antara Pengunjung dan Pembeli terhadap Nominal Pembelian di Indomaret Kedungmungdu Semarang dengan Metode Kuadrat Terkecil.07–11.
10. Ghozali, I. (2016) Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23. Edisi 8. Semarang: Badan Penerbit Universitas Diponegoro.
11. Sekaran, U. & Bougie, R.J., (2016). Research Methods for Business: A skill Building Approach. 7<sup>th</sup> Edition, John Wiley & Sons Inc. New York, US.
12. Suhandi, Nazori. Putri, Efri Ayu kartika. Agnisa, Sari.(2018). Analisis Pengaruh Jumlah Penduduk terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linear di Kota Palembang. JURNAL ILMIAH INFORMATIKA GLOBAL.9(2).79-81
13. Contoh dan Cara Penulisan Daftar Pustaka Menurut APA Style. Available online: (<https://www.brainacademy.id/blog/menyusun-daftar-pustaka#:~:text=Untuk%20menulis%20daftar%20pustaka%20dari,terbit%20%E2%80%93%20judul%20buku%20%E2%80%93%20penerbit>). (accessed on 26 May 2023)



# Statistika Deskriptif Pada Analisis Ketimpangan Kemiskinan (Studi Kasus Data Kemiskinan Wilayah Perkotaan dan Pedesaan Seluruh Provinsi di Indonesia periode tahun 2021-2022)

Muhammad Nashif Farid<sup>1</sup>, Mohammad Sufa Ammar Habibi<sup>2</sup>, Trimono<sup>3</sup>

<sup>1, 2, 3</sup>Program Studi Sains Data, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur

<sup>1</sup>[22083010024@student.upnjatim.ac.id](mailto:22083010024@student.upnjatim.ac.id)

<sup>2</sup>[22083010014@student.upnjatim.ac.id](mailto:22083010014@student.upnjatim.ac.id)

<sup>3</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id)

**Abstract:** *The trend of poverty in Indonesia is a problem that requires special attention in its mitigation. This study discusses the inequality of poverty between urban and rural areas in Indonesia, based on data from the Central Statistics Agency (Badan Pusat Statistik or BPS) for the period 2021-2022. The research employs descriptive analysis and Pearson correlation test as the methods. The results of the descriptive analysis indicate a significant difference between the regions. Furthermore, the results of the correlation test also demonstrate that the majority of poverty cases in Indonesia are concentrated in rural areas. Therefore, further research is needed*  
**Keyword :** *Poverty, Inequality*

**Abstrak:** Trend masalah kemiskinan di Indonesia merupakan permasalahan yang perlu mempunyai perhatian khusus dalam penanggulangannya. Penelitian ini membahas ketimpangan kemiskinan antara wilayah perkotaan dengan pedesaan di Indonesia yang bersumber dari Badan Pusat Statistik (BPS) periode tahun 2021-2022. Penelitian ini menggunakan metode analisis deskriptif dan uji korelasi pearson. Hasil uji analisis deskriptif menunjukkan bahwa terdapat perbedaan yang cukup signifikan antar wilayah. Kemudian hasil uji korelasi juga menunjukkan bahwa angka kemiskinan di Indonesia sebagian besar terdapat dalam rentang wilayah pedesaan. Oleh karenanya diperlukan penelitian lebih lanjut dalam pengambilan keputusan oleh pemerintahan dengan memperhatikan faktor dan implikasi kemiskinan yang terjadi di Indonesia.

**Kata kunci :** Kemiskinan, Ketimpangan, Analisis Deskriptif

## I. PENDAHULUAN

Ketimpangan kemiskinan merupakan salah satu isu sosial yang kompleks dan mempengaruhi banyak negara, termasuk Indonesia. Di tengah perkembangan ekonomi dan modernisasi, ketimpangan dalam distribusi pendapatan dan kemiskinan masih menjadi tantangan serius yang dihadapi oleh masyarakat Indonesia. Fenomena ini melibatkan ketidakadilan dalam akses terhadap sumber daya dan peluang ekonomi, yang menghasilkan disparitas yang signifikan antara kelompok masyarakat yang kaya dan miskin.

Indonesia, sebagai negara kepulauan dengan beragam budaya, etnis, dan geografi, menghadapi tantangan yang kompleks dalam mengatasi ketimpangan kemiskinan. Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa tingkat kemiskinan di Indonesia masih cukup tinggi, meskipun ada kemajuan yang telah dicapai dalam beberapa tahun terakhir. Ketimpangan ini dapat dipengaruhi oleh berbagai faktor, termasuk tingkat pengangguran, akses terhadap pendidikan dan pelayanan kesehatan, serta perbedaan dalam distribusi infrastruktur dan sumber daya.

Pada artikel ini, penulis akan menganalisis tingkat kemiskinan di Indonesia dengan menggunakan data yang diperoleh dari BPS. Kami akan fokus pada perbandingan antara kemiskinan di wilayah perkotaan dan pedesaan, serta melihat perbedaan kemiskinan antar provinsi di Indonesia. Melalui analisis ini, diharapkan dapat memperoleh pemahaman yang lebih baik tentang ketimpangan kemiskinan di Indonesia dan merumuskan strategi dan kebijakan yang lebih efektif dalam mengurangi kesenjangan sosial dan ekonomi.

**Tabel 1.** Persentase Kemiskinan Perkotaan dan Pedesaan di Indonesia Tahun 2021 - 2022

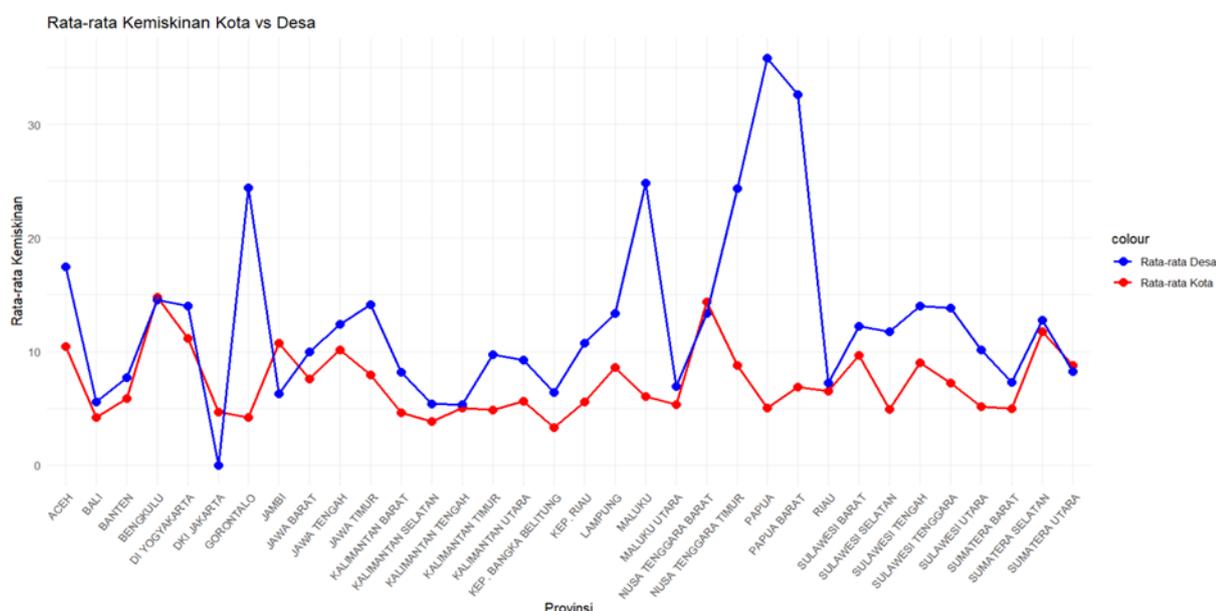
No	Provinsi	Rata-rata Kota	Rata-rata Desa	Rata-rata Keseluruhan
1	ACEH	10,425	17,4375	13,93125
2	SUMATERA UTARA	8,805	8,26	8,5325
3	SUMATERA BARAT	4,995	7,3	6,1475
4	RIAU	6,5175	7,2125	6,865
5	JAMBI	10,7525	6,3025	8,5275
6	SUMATERA SELATAN	11,7375	12,7525	12,245
7	BENGKULU	14,81	14,5725	14,69125
8	LAMPUNG	8,61	13,365	10,9875
9	KEP. BANGKA BELITUNG	3,34	6,3975	4,86875
10	KEP. RIAU	5,5575	10,715	8,13625
11	DKI JAKARTA	4,6725	0,0000	4,6725
12	JAWA BARAT	7,5975	9,9625	8,78
13	JAWA TENGAH	10,17	12,4	11,285
14	DI YOGYAKARTA	11,1575	14,02	12,58875
15	JAWA TIMUR	7,965	14,1075	11,03625
16	BANTEN	5,8975	7,74	6,81875
17	BALI	4,2	5,5425	4,87125
18	NUSA TENGGARA BARAT	14,385	13,3475	13,86625
19	NUSA TENGGARA TIMUR	8,7525	24,3675	16,56
20	KALIMANTAN BARAT	4,6175	8,1875	6,4025
21	KALIMANTAN TENGAH	5,05	5,3225	5,18625
22	KALIMANTAN SELATAN	3,8425	5,3675	4,605
23	KALIMANTAN TIMUR	4,88	9,7125	7,29625
24	KALIMANTAN UTARA	5,6025	9,2575	7,43
25	SULAWESI UTARA	5,1575	10,1525	7,655
26	SULAWESI TENGAH	9,0325	14,025	11,52875
27	SULAWESI SELATAN	4,9275	11,76	8,34375
28	SULAWESI TENGGARA	7,2425	13,85	10,54625
29	GORONTALO	4,1875	24,4475	14,3175
30	SULAWESI BARAT	9,6575	12,225	10,94125
31	MALUKU	6,035	24,835	15,435
32	MALUKU UTARA	5,3275	6,925	6,12625
33	PAPUA BARAT	6,885	32,61	19,7475
34	PAPUA	5,04	35,82	20,43
Rata-rata Keseluruhan		7,29	12,66	9,97

Tabel 1 menampilkan persentase rata rata kemiskinan di Indonesia dan berbagai provinsi yang ada di Indonesia dalam kurun waktu 2021 – 2022. Indonesia memiliki persentase rata rata kemiskinan sebesar 9,972537 persen (sekitar 7,289191 persen dari perkotaan dan 12,65588 pedesaan). Provinsi Papua menjadi penyumbang persentase tertinggi kemiskinan di pedesaan, dengan persentase mencapai 35,82 persen. Sementara itu, provinsi Bengkulu merupakan penyumbang persentase tertinggi kemiskinan di perkotaan, dengan persentase sebesar 14,81 persen. Hal ini menunjukkan bahwa provinsi

Papua dan Bengkulu memiliki tantangan yang perlu ditangani dalam mengurangi tingkat kemiskinan, baik di pedesaan maupun di perkotaan.

Secara khusus, provinsi Papua memiliki persentase rata-rata kemiskinan terbesar, yaitu sebesar 20,43 persen. Hal ini terdiri dari persentase kemiskinan di pedesaan sebesar 35,82 persen dan persentase kemiskinan di perkotaan sebesar 5,04 persen. Di sisi lain, provinsi DKI Jakarta memiliki persentase rata-rata kemiskinan terkecil, yaitu sebesar 2,33625 persen. Hal ini terdiri dari persentase kemiskinan di perkotaan sebesar 4,6725 persen, sementara tidak ada persentase kemiskinan di pedesaan.

Perbedaan yang signifikan antara tingkat kemiskinan di pedesaan dan perkotaan di berbagai provinsi mengindikasikan adanya disparitas sosial dan ekonomi yang perlu mendapat perhatian serius. Dalam konteks ini, analisis lebih lanjut diperlukan untuk memahami faktor-faktor yang menyebabkan perbedaan ini. Faktor-faktor seperti akses terhadap pekerjaan, pendidikan, layanan kesehatan, serta infrastruktur dasar perlu dieksplorasi agar langkah-langkah kebijakan yang tepat dapat dirumuskan untuk mengurangi tingkat kemiskinan dan mengatasi ketimpangan antara penduduk desa dan kota di Indonesia.



**Gambar 1.** Lineplot Perbandingan Kemiskinan Daerah Perkotaan dan Pedesaan 2021-2022

Gambar 1 memperlihatkan ketimpangan antara penduduk desa dan kota yang terjadi di berbagai provinsi di Indonesia, tercatat di provinsi Papua persentase rata-rata kemiskinan di pedesaan sangat tinggi hingga mencapai 35,82 persen. Sedangkan, di daerah perkotaan hanya menginjak angka 5,04 persen, artinya terjadi ketimpangan sekitar 30 persen antara penduduk pedesaan dan perkotaan di Papua. Kemudian pada provinsi Gorontalo, tercatat persentase kemiskinan penduduk pedesaan sebesar 24,45 persen. Sedangkan, di daerah perkotaan hanya menginjak 4,18 persen, artinya terjadi ketimpangan sekitar 20 persen antara penduduk pedesaan dan perkotaan di Gorontalo. Begitu pula di provinsi Maluku, tercatat persentase rata-rata kemiskinan penduduk pedesaan di Maluku mencapai 24,8 persen yang sangat jauh jika dibandingkan dengan persentase penduduk di perkotaan yang hanya menginjak angka 6 persen, artinya terjadi ketimpangan sekitar 18 persen antara penduduk perkotaan dan pedesaan di Maluku. Perbedaan tersebut mengindikasikan bahwa tingkat kemiskinan di pedesaan jauh lebih tinggi dibandingkan dengan perkotaan.

Sedangkan, tercatat ada beberapa provinsi di Indonesia yang mengalami ketimpangan cukup rendah. Tercatat di provinsi Bengkulu rata-rata kemiskinan di daerah pedesaan 14,81 persen dan di daerah perkotaan sebesar 14,57 persen, artinya hanya terjadi ketimpangan rata-rata kemiskinan sebesar 0,23 persen antara penduduk perkotaan dan pedesaan. Kemudian provinsi Riau dengan rata-rata



kemiskinan di pedesaan sebesar 6,51 dan di perkotaan 7,21 persen, dengan artian bahwa hanya terjadi ketimpangan rata-rata kemiskinan sebesar 0,69 persen di provinsi Riau.

Perbedaan ini mengindikasikan bahwa tingkat kemiskinan di pedesaan cenderung lebih tinggi dibandingkan dengan perkotaan di sebagian besar provinsi. Ketimpangan ini menunjukkan adanya disparitas sosial dan ekonomi yang perlu ditangani oleh pemerintah dan pemangku kepentingan lainnya.

Dalam konteks ini, analisis lebih lanjut diperlukan untuk memahami faktor-faktor yang berkontribusi pada tingkat kemiskinan yang tinggi di pedesaan. Hal ini meliputi aspek ekonomi, sosial, dan infrastruktur yang mempengaruhi akses terhadap pekerjaan, pendidikan, layanan kesehatan, serta fasilitas dasar lainnya. Dengan memahami faktor-faktor tersebut, langkah-langkah kebijakan yang tepat dapat dirumuskan untuk mengurangi tingkat kemiskinan di pedesaan dan meningkatkan kualitas hidup penduduk secara keseluruhan.

## II. KAJIAN PUSTAKA

### 2.1 Kemiskinan

Kemiskinan adalah keadaan saat ketidakmampuan untuk memenuhi kebutuhan dasar seperti makanan, pakaian, tempat berlindung, pendidikan, dan kesehatan. Supriatna (1997:90) menyatakan bahwa kemiskinan adalah situasi yang serba terbatas yang terjadi bukan atas kehendak orang yang bersangkutan. Suatu penduduk dikatakan miskin bila ditandai oleh rendahnya tingkat pendidikan, produktivitas kerja, pendapatan, kesehatan dan gizi serta kesejahteraan hidupnya, yang menunjukkan lingkaran ketidakberdayaan. Kelompok penduduk miskin yang berada di masyarakat pedesaan dan perkotaan, umumnya berprofesi sebagai buruh tani, petani gurem, pedagang kecil, nelayan, pengrajin kecil, buruh, pedagang kaki lima, pedagang asongan, pemulung, gelandangan dan pengemis, dan pengangguran. Kelompok miskin ini akan menimbulkan problema yang terus berlanjut bagi kemiskinan kultural dan struktural, bila tidak ditangani secara serius, terutama untuk generasi berikutnya.

### 2.2 Ketimpangan

Ketimpangan merupakan suatu kondisi dimana terdapat perbedaan yang cukup signifikan antara dua hal atau lebih. Konteks ketimpangan ini dapat dilibatkan dengan berbagai permasalahan yang sering terjadi di ruang lingkup kehidupan seperti sosial, politik, ekonomi hingga pendidikan. Ketimpangan dalam masalah kemiskinan dapat terjadi disebabkan beberapa faktor seperti halnya ketimpangan pembangunan dan pendidikan. Ketimpangan pembangunan ekonomi antar wilayah menurut Sjafrizal (2012) merupakan fenomena umum yang terjadi dalam proses pembangunan ekonomi suatu daerah. Ketimpangan ini pada awalnya disebabkan oleh adanya perbedaan kandungan demografi yang terdapat pada masing-masing wilayah. Akibat dari perbedaan ini, kemampuan suatu daerah untuk meningkatkan pertumbuhan ekonomi dan mendorong proses pembangunan juga menjadi berbeda.

## III. METODE PENELITIAN

Penelitian ini membahas mengenai ketimpangan rata-rata kemiskinan di Indonesia. Jenis pada penelitian ini menggunakan penelitian kuantitatif. Data yang digunakan dalam penelitian ini adalah data sekunder dengan jenis data survey. Variabel yang diselidiki pada studi ini, ialah variabel Rata-rata Desa yang dilihat dari variabel Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Pedesaan 2021 Semester 1 (Maret), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Pedesaan 2021 Semester 2 (September), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Pedesaan 2022 Semester 1 (Maret), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Pedesaan 2022 Semester 2 (September), dan variabel Rata-rata Kota yang dilihat dari variabel Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Perkotaan 2021 Semester 1 (Maret), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Perkotaan 2021 Semester 2 (September), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Perkotaan 2022 Semester 1 (Maret), Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah Perkotaan 2022 Semester 2

Teknik analisis dalam penelitian ini menggunakan analisis deskriptif, analisis uji korelasi Pearson, dan uji T. Analisis deskriptif berdasarkan data sekunder, jurnal, artikel, dan hasil-hasil penelitian yang berhubungan dengan permasalahan. Kemudian analisis uji korelasi akan digunakan untuk mengevaluasi hubungan antara variabel Rata-rata Desa dan Rata-rata Kota dengan Rata-rata keseluruhan. Uji korelasi Pearson dapat mengungkapkan sejauh mana korelasi antara variabel tersebut, apakah ada hubungan positif, negatif, atau tidak ada hubungan sama sekali. Kemudian uji T digunakan untuk membandingkan Rata-rata Keseluruhan dengan Rata-rata Desa dan Rata-rata Kota. Hal ini bertujuan untuk menentukan apakah ada perbedaan signifikan antara tingkat kemiskinan di daerah pedesaan dan perkotaan. Uji T akan memberikan informasi apakah perbedaan yang diamati antara dua kelompok tersebut bersifat signifikan secara statistik ataukah hanya kebetulan.

#### IV. HASIL DAN PEMBAHASAN

##### 4.1 Analisis Deskriptif

Dalam tahap awal yang perlu dilakukan adalah melakukan analisis deskriptif untuk mengetahui beberapa informasi statistik yang ada dalam hasil olah data.

**Tabel 2.** Hasil Analisis Deskriptif Perhitungan Statistika Sederhana

X	Mean	Std. Deviasi	Q1	Median	Q3
Rata-rata Kota	7,29	2,98	5	6,28	9,03
Rata-rata Desa	12,66	7,75	7,3	11,24	14,03
Rata-rata Keseluruhan	9,97	4,35	6,82	8,66	12,59

Berdasarkan hasil tabel di atas menampilkan variabel rata-rata desa dan rata-rata kota. Dimana hasil analisis deskriptif perbedaan mean (rata-rata) antara variabel rata-rata desa dan rata-rata kota cukup besar yaitu 5,37 persen, artinya rata-rata kemiskinan kota dan rata-rata kemiskinan desa terdapat ketimpangan cukup besar. Kemudian dari hasil analisis deskriptif juga ditampilkan Std. deviasi (standar deviasi) dari variabel rata-rata Kota, rata-rata Desa, dan rata-rata keseluruhan relatif rendah yaitu 2,98, 7,75 dan 4,35 persen. Maka dapat disimpulkan bahwa tingkat kemiskinan yang relatif seragam atau homogen di antara unit analisis. Kemudian dari hasil analisis deskriptif juga ditampilkan kuartil 1, median, dan kuartil 3. Hal ini berguna untuk mengelompokkan data berdasarkan hasil kuartil.

**Tabel 3.** Pemetaan Wilayah dengan Indikasi Kemiskinan

No	Provinsi	Rata-rata Keseluruhan	Kategori
1	ACEH	13,93125	Tinggi
2	SUMATERA UTARA	8,5325	Sedang
3	SUMATERA BARAT	6,1475	Rendah
4	RIAU	6,865	Sedang
5	JAMBI	8,5275	Sedang
6	SUMATERA SELATAN	12,245	Sedang
7	BENGKULU	14,69125	Tinggi
8	LAMPUNG	10,9875	Sedang
9	KEP. BANGKA BELITUNG	4,86875	Rendah
10	KEP. RIAU	8,13625	Sedang



11	DKI JAKARTA	4,6725	Rendah
12	JAWA BARAT	8,78	Sedang
13	JAWA TENGAH	11,285	Sedang
14	DI YOGYAKARTA	12,58875	Sedang
15	JAWA TIMUR	11,03625	Sedang
16	BANTEN	6,81875	Rendah
17	BALI	4,87125	Rendah
18	NUSA TENGGARA BARAT	13,86625	Tinggi
19	NUSA TENGGARA TIMUR	16,56	Tinggi
20	KALIMANTAN BARAT	6,4025	Rendah
21	KALIMANTAN TENGAH	5,18625	Rendah
22	KALIMANTAN SELATAN	4,605	Rendah
23	KALIMANTAN TIMUR	7,29625	Sedang
24	KALIMANTAN UTARA	7,43	Sedang
25	SULAWESI UTARA	7,655	Sedang
26	SULAWESI TENGAH	11,52875	Sedang
27	SULAWESI SELATAN	8,34375	Sedang
28	SULAWESI TENGGARA	10,54625	Sedang
29	GORONTALO	14,3175	Tinggi
30	SULAWESI BARAT	10,94125	Sedang
31	MALUKU	15,435	Tinggi
32	MALUKU UTARA	6,12625	Rendah
33	PAPUA BARAT	19,7475	Tinggi
34	PAPUA	20,43	Tinggi

Tabel 3 menunjukkan hasil setelah dilakukan pengelompokkan sesuai kuartil dari analisis deskriptif. Data yang masuk dalam kuartil 1 akan termasuk pada kelompok rendah dan yang berada diantara kuartil 1 dan kuartil 3 akan masuk dalam kelompok sedang, kemudian kuartil 3 akan masuk dalam kelompok tinggi.

#### 4.2 Analisis uji korelasi

Setelah mendapatkan informasi statistik dari analisis deskriptif, tahap selanjutnya yaitu mencari nilai korelasi antara variable rata-rata kota dan rata-rata desa terhadap rata-rata keseluruhan, untuk mengetahui nilai korelasi atau hubungan antar variabel tersebut. kemudian kita cari nilai korelasi menggunakan analisis uji korelasi untuk mengetahui nilai korelasi antar variabel.

**Tabel 4.** Uji Korelasi Pearson

X	Rata-rata Keseluruhan
Rata-rata Kota	0,4707585
Rata-rata Desa	0,9407781

Tabel 4 menunjukkan nilai korelasi antara variabel Rata-rata Kota dan Rata-rata Keseluruhan, ditemukan bahwa korelasi antara kedua variabel tersebut cukup rendah, dengan nilai sekitar 0,47. Hal ini menunjukkan adanya hubungan yang lemah antara rata-rata kemiskinan di desa dengan rata-rata



kemiskinan secara keseluruhan. Selanjutnya, pada Tabel 4 juga terlihat nilai korelasi antara variabel Rata-rata Desa dan Rata-rata Keseluruhan, terdapat korelasi yang cukup tinggi antara kedua variabel tersebut, dengan nilai sekitar 0,94. Hal ini menunjukkan adanya hubungan yang kuat antara rata-rata kemiskinan di kota dengan rata-rata kemiskinan secara keseluruhan. Hasil di atas menunjukkan bahwa nilai rata-rata kemiskinan desa merupakan sumbang nilai terbesar terhadap kemiskinan yang terjadi di Indonesia.

## V. KESIMPULAN

Trend masalah kemiskinan di Indonesia dalam kurun waktu 2 tahun terakhir berdasarkan Data dari Badan Pusat Statistik menunjukkan bahwa terdapat perbedaan yang signifikan antara Wilayah pedesaan dengan perkotaan. Hal ini menunjukkan bahwa terjadi ketimpangan kemiskinan di Indonesia. Dari adanya masalah tersebut menjadikan langkah keputusan yang harus diambil dalam mengatasi permasalahan juga berbeda-beda yang perlu diadakan analisis statistik lagi secara mendalam dengan memperhatikan variabel penyebab atau faktor ketimpangan kemiskinan di Indonesia.

Berdasarkan hasil proses hitung analisis deskriptif juga menunjukkan perbedaan yang cukup signifikan antara persentase kemiskinan di pedesaan dengan perkotaan. Dengan adanya analisis deskriptif juga dapat disimpulkan beberapa wilayah provinsi yang termasuk dalam kluster kemiskinan rendah, sedang dan tinggi. Dal tersebut juga memudahkan untuk klasifikasi fokus penanganan permasalahan utama.

Berdasarkan hasil tersebut, maka sudah waktunya bagi pemerintah untuk melakukan *crash program* untuk menyelesaikan permasalahan ketimpangan kemiskinan antar wilayah di Indonesia. Hal tersebut perlu dilakukan dalam menekan angka kemiskinan di Indonesia dengan penelitian lanjutan mengenai faktor dan implikasi kemiskinan yang terjadi ditengah masyarakat.

Namun juga perlu diingat bahwa, apapun yang diprogramkan oleh pemerintah khususnya dalam penanggulangan kemiskinan tidak akan terlaksana dan mencapai apa yang diharapkan, jika tidak didukung oleh masyarakat sebagai sasaran implementasi setiap kebijakan pembangunan dan kemasyarakatan

## UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada mata kuliah bahasa Indonesia yang diselenggarakan oleh UPN "Veteran" Jawa Timur sehingga artikel ini bisa ditulis dengan baik dan benar.

## REFERENSI

1. Badan Pusat Statistik, (2022) Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah 2021-2022. Jakarta : BPS
2. Kadji, Yulianto. "Kemiskinan dan Konsep teoritisnya." Guru Besar Kebijakan Publik Fakultas Ekonmi Dan Bisnis UNG (2012): 1-7.
3. Kemiskinan (7 Februari 2023) Dalam Wikipedia, diakses pada tanggal 29 Mei 2023 dari <https://id.wikipedia.org/wiki/Kemiskinan>.
4. Ramdani, Martiyan. "Determinan kemiskinan di Indonesia tahun 1982-2012." Economics Development Analysis Journal 4.1 (2015): 58-64.
5. Supriatna, Tjahya. Birokrasi, pemberdayaan, dan pengentasan kemiskinan. Humaniora Utama Press, 1997.
6. Sjafrizal, Ekonomi Wilayah dan Perkotaan, Jakarta: Rajawali Press, 2012, hlm.108-110.



# Analisis Kesamaan Rata-rata Indeks Harga Konsumen (IHK) Menurut Kelompok dan Sub Kelompok pada Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau dengan Kelompok dan Sub Kelompok 9 Pendidikan

Melya Vebryanti<sup>1</sup>, Wahyu Melinda Permanasari<sup>2</sup>, Trimono<sup>3</sup>,

<sup>1,2,3</sup>Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup> [22083010063@student.upnjatim.ac.id](mailto:22083010063@student.upnjatim.ac.id)

<sup>3</sup> [trimono.stat@unjatim.ac.id](mailto:trimono.stat@unjatim.ac.id)

Corresponding author email: [22083010043@student.upnjatim.ac.id](mailto:22083010043@student.upnjatim.ac.id)

**Abstract:** *The Consumer Price Index (CPI) is one of the indicators used to measure the inflation rate in a country or region. According to the Central Statistics Agency (BPS), the CPI is used to calculate the average price changes over a specific period of time for a set of goods and services used by the population and households. The CPI typically provides information about price trends for a group of goods or services consumed by households at a specific point in time. An increase in the CPI for the food sub-group will have a significant impact on food prices. This will ultimately affect the cost of living. Education is an important foundation in building a society with high knowledge. Next, a comparison of the averages between the CPI for the food sub-group and the education sub-group is conducted. To achieve the best results, the Two-Sample Mean Test method is used. In the results section, it can be concluded that the Consumer Price Index (CPI) for the food and education sectors has different averages, indicating that the inflation rates in these two sectors are different.*

**Keywords:** *Consumer Price Index, Food Sub-Group, Education Sub-Group, Mean Test Two Populations*

**Abstrak:** Indeks Harga Konsumen (IHK) adalah salah satu indikator untuk mengukur tingkat inflasi dalam suatu negara atau wilayah. Menurut Badan Pusat Statistik (BPS), Indeks Harga Konsumen (IHK) digunakan untuk menghitung perubahan harga rata-rata selama periode waktu tertentu dari kumpulan barang dan jasa yang digunakan oleh penduduk dan rumah tangga selama periode waktu tertentu. IHK biasanya memberikan informasi tentang tren harga untuk sekelompok barang atau jasa yang dikonsumsi oleh rumah tangga pada suatu titik waktu tertentu. Kenaikan IHK pada sub kelompok makanan akan berdampak pada kenaikan harga-harga makanan secara signifikan. Ini nantinya akan berpengaruh pada kenaikan biaya hidup. Lalu pendidikan merupakan pondasi penting dalam membangun masyarakat yang memiliki pengetahuan yang tinggi. Selanjutnya dilakukan perbandingan rata-rata antara IHK pada sub kelompok makanan dan sub kelompok pendidikan. Untuk menghasilkan hasil yang terbaik, maka menggunakan metode Uji Mean Dua Populasi. Pada bagian hasil, dapat disimpulkan bahwa Indeks Harga Konsumen (IHK) pada sektor makanan dan pendidikan memiliki rata-rata yang berbeda yang berarti tingkat inflasi pada kedua sektor tersebut berbeda.

**Kata kunci:** Indeks Harga Konsumen, Sub Kelompok Makanan, Sub Kelompok Pendidikan, Uji Mean Dua Populasi

## I. PENDAHULUAN

Indeks Harga Konsumen (IHK) adalah salah satu indikator untuk mengukur tingkat inflasi dalam suatu negara atau wilayah. Menurut Badan Pusat Statistik (BPS), Indeks Harga Konsumen (IHK) digunakan untuk menghitung perubahan harga rata-rata selama periode waktu tertentu dari kumpulan barang dan jasa yang digunakan oleh penduduk dan rumah tangga selama periode waktu tertentu. IHK biasanya memberikan informasi tentang tren harga untuk sekelompok barang atau jasa yang dikonsumsi oleh rumah tangga pada suatu titik waktu tertentu.

Indeks Harga Konsumen (IHK) merupakan salah satu indikator ekonomi yang digunakan untuk mengukur tingkat perubahan harga (inflasi/deflasi) di tingkat konsumen. Dengan adanya perubahan pola konsumsi masyarakat, maka mulai Januari 2020, pengukuran inflasi di Indonesia menggunakan IHK tahun dasar 2018=100. Beberapa perubahan mendasar dalam penghitungan IHK (2018=100) dibandingkan (IHK 2012=100), khususnya dari sisi cakupan, klasifikasi pengelompokan



komoditas, metodologi penghitungan IHK, paket komoditas, dan diagram timbang. Perubahan tersebut didasarkan pada Survei Biaya Hidup (SBH) yang dilaksanakan oleh BPS selama tahun 2018, sebagai salah satu bahan dasar utama dalam penghitungan IHK. [1]

Di Indonesia terdapat Indeks Harga Konsumen (IHK), perkembangan IHK dapat memperlihatkan tingkat harga suatu barang dan jasa yang dibeli masyarakat. IHK bermanfaat untuk mengetahui tingkat kenaikan pendapatan, harga, juga dapat dijadikan sebagai indikator ekonomi dan tolak ukur besarnya biaya produksi. [2] IHK sendiri dibagi menjadi sembilan kelompok dan sub kelompok. Kelompok dan sub kelompok IHK mengacu pada klasifikasi kelompok barang dan jasa yang digunakan untuk menghitung IHK suatu negara atau wilayah.

IHK kelompok dan sub kelompok 1 makanan, minuman, dan tembakau, ini mengacu pada barang pokok maupun jasa yang dibutuhkan dalam kehidupan sehari-hari. Subkelompok makanan dan minuman mencakup harga bahan makanan, minuman, dan produk pertanian. Subkelompok tembakau mencakup harga rokok dan produk tembakau lainnya. Perubahan harga pada kelompok ini dapat memiliki dampak yang signifikan terhadap daya beli masyarakat, karena konsumsi barang-barang tersebut merupakan kebutuhan utama. Sedangkan IHK kelompok dan sub kelompok 9 pendidikan mencakup biaya pendidikan, buku, dan perlengkapan sekolah.

Makanan sebagai kebutuhan pokok, memiliki peran penting dalam memenuhi kebutuhan dasar manusia. Kenaikan IHK pada kelompok ini akan berdampak pada kenaikan harga-harga makanan secara signifikan. Ini nantinya juga akan berpengaruh pada kenaikan biaya hidup. Lalu pendidikan sebagai kebutuhan pengembangan pada diri manusia, memiliki peran penting dalam memenuhi kebutuhan dasar. Pendidikan merupakan pondasi penting dalam membangun masyarakat yang memiliki pengetahuan yang tinggi.

Pada penelitian ini menggunakan metode statistika Uji Mean Dua Populasi untuk memprediksi dan melihat rata-rata IHK Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau dengan rata-rata IHK Kelompok dan Sub Kelompok 9 Pendidikan. Data yang digunakan dan akan dilakukan perbandingan adalah data yang dikumpulkan oleh Badan Pusat Statistik (BPS) yaitu data IHK Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau Tahun 2022 di 90 kota di Indonesia dan data IHK Kelompok dan Sub Kelompok 9 Pendidikan Tahun 2022 di 90 kota di Indonesia. Penelitian ini bertujuan untuk melihat apakah rata-rata IHK kedua kelompok dan sub kelompok tersebut sama.

## **II. METODE PENELITIAN**

### **3.1 Variabel yang digunakan**

Variabel yang digunakan dalam penelitian ini adalah data IHK Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau Tahun 2022 di 90 kota di Indonesia dan data IHK Kelompok dan Sub Kelompok 9 Pendidikan Tahun 2022 di 90 kota di Indonesia. Dan yang akan dianalisis yaitu subkelompok makanan dan sub kelompok pendidikan.

### **3.2 Metode Pengumpulan Data**

Metode pengumpulan data yang dilakukan dalam penyusunan artikel ini adalah Metode Literatur. Metode literatur adalah mengumpulkan, memilih, dan menganalisis beberapa sumber bacaan yang berkaitan dengan permasalahan yang berhubungan dalam penyusunan artikel. Dengan metode ini, penulis mengumpulkan dan memilih sumber bacaan (buku-buku) yang berkaitan dengan kegiatan tersebut. Metode ini digunakan untuk mendapatkan data tentang Indeks Harga Konsumen (IHK) pada dua sub kelompok yaitu makanan dan pendidikan. Data tersebut merupakan data sekunder karena diperoleh dari Badan Pusat Statistik (BPS).

### **3.3 Metode Analisis Data**

Metode yang digunakan untuk menganalisis data Indeks Harga Konsumen pada artikel ini, penulis menggunakan metode Uji Mean Dua Populasi. Uji Mean Dua Populasi merupakan pengujian

untuk melihat ada atau tidaknya perbedaan rata-rata dari dua sampel yang dibandingkan. Pada metode ini terdapat 3 step yang diujikan pada data yang akan dibandingkan :

### 1. Uji Normalitas

Step ini merupakan pengujian pada data untuk melihat apakah data tersebut berdistribusi normal atau tidak. Hipotesis pada Uji Normalitas ini antara lain :

- H<sub>0</sub>: Data Berdistribusi Normal
- H<sub>1</sub>: Data Tidak Berdistribusi Normal

### 2. Uji Variansi dua Populasi

Misalkan terdapat dua populasi, yaitu X<sub>1</sub> dengan rata-rata populasi  $\mu_1$  dan variansi populasi  $\sigma_1^2$ , dan populasi X<sub>2</sub> dengan rata-rata populasi  $\mu_2$  dan variansi populasi  $\sigma_2^2$ . Dari kedua populasi diambil sampel masing-masing sebanyak n<sub>1</sub> dan n<sub>2</sub>. Dari kedua populasi tersebut dapat dilakukan uji untuk variansi keduanya sama atau berbeda. Asumsi yang digunakan yaitu:

- Data sampel adalah data kontinu.
- Masing-masing populasi berdistribusi Normal

Step ini merupakan pengujian pada kedua data yang dibandingkan untuk melihat apakah kedua data tersebut memiliki variansi populasi yang sama atau tidak.

- Hipotesis pada Uji Variansi dua Populasi ini antara lain :
  - H<sub>0</sub> : Variansi populasi 1 sama dengan variansi populasi 2
  - H<sub>1</sub> : Variansi populasi 1 tidak sama dengan variansi populasi 2

#### 2) Tingkat Signifikansi

Tingkat signifikansi yang umumnya digunakan adalah  $\alpha = 0,05$

#### 3) Statistik Uji

Statistik uji menggunakan nilai F hitung yang diperoleh melalui persamaan berikut :

$$F_{hitung} = \frac{S_1^2}{S_2^2}$$

#### 4) Kaidah Pengambilan Keputusan

H<sub>0</sub> ditolak jika F<sub>hitung</sub> > F<sub>tabel</sub> atau F<sub>hitung</sub> < -F<sub>tabel</sub>

F<sub>tabel</sub> adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas  $v_1 = n_1 - 1$  dan  $v_2 = n_2 - 1$ .

### 3. Uji Hipotesis Perbandingan Rata-Rata Dua Populasi, dengan Sampel yang Diambil Saling Bebas.

Hal penting yang perlu dipahami pada uji ini adalah antara populasi 1 dan populasi 2, ukuran sampel yang diambil tidak harus sama. Misalkan akan diuji perbandingan rata-rata IHK pada kota-kota di Indonesia. Populasi yang diamati adalah IHK pada sektor makanan dan sektor pendidikan yang keduanya saling bebas. Asumsi yang digunakan adalah:

- Sampel berasal dari populasi yang berdistribusi Normal/setiap populasi berdistribusi Normal
- Data sampel adalah data kontinu

Step ini merupakan pengujian pada kedua data yang dibandingkan untuk menentukan apakah kedua data tersebut memiliki rata-rata populasi yang sama atau tidak.

#### 1) Hipotesis pada Uji Mean dua Populasi ini antara lain :

- H<sub>0</sub> : Rata-rata populasi 1 sama dengan rata-rata populasi 2
- H<sub>1</sub> : Rata-rata populasi 1 tidak sama dengan rata-rata populasi 2

#### 2) Tingkat Signifikansi

$\alpha = 0,05$

#### 3) Statistik Uji

Statistik uji t digunakan saat: nilai  $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui

**Subkasus 1** ( $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui, dan terdapat informasi bahwa  $\sigma_1^2 = \sigma_2^2$ )

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{dengan } S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

**Subkasus 2** ( $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui, dan terdapat informasi bahwa  $\sigma_1^2 \neq \sigma_2^2$ )

$$t_{\text{hitung}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

4) Kaidah Pengambilan Keputusan untuk statistik uji T

**Subkasus 1** :  $H_0$  ditolak jika  $t_{\text{hitung}} > t_{\text{tabel}}$  atau  $t_{\text{hitung}} < -t_{\text{tabel}}$

untuk  $t_{\text{tabel}}$  adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas ( $n_1 + n_2 - 2$ )

**Subkasus 2** :  $H_0$  ditolak jika  $t_{\text{hitung}} > t_{\text{tabel}}$  atau  $t_{\text{hitung}} < -t_{\text{tabel}}$

untuk  $t_{\text{tabel}}$  adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas  $v$ , dengan  $v$  didefinisikan sebagai berikut :

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2}} - 2$$

Dalam penelitian ini data yang digunakan adalah data sekunder dengan jenis data yang digunakan adalah data pada tahun 2022. Sumber data diperoleh dari Badan Pusat Statistika (BPS) dan data tersebut meliputi data IHK (Indeks Harga Konsumen). Data IHK akan diramalkan menggunakan data masa lalu yang akan dianalisis untuk memperkirakan suatu nilai pada masa yang akan datang.

### III. HASIL DAN PEMBAHASAN

```
> data_rata
  Bulan Rata.SK1.Makanan Rata.SK9.Pendidikan
1 Januari           111.9110           107.7379
2 Februari          110.3897           107.8106
3 Maret             112.2807           107.8107
4 April             114.2647           107.8317
5 Mei               115.9179           107.8524
6 Juni              118.3862           107.9162
7 Juli              119.9352           108.6231
8 Agustus           116.9399           109.4638
9 September         115.8922           109.9864
10 Oktober          114.1669           110.2014
11 November         113.9823           110.2183
12 Desember         116.3880           110.2268
```

**Gambar 1.** Data Rata-rata

Gambar 1. adalah data yang digunakan dalam Uji Mean Dua Populasi. Data ini merupakan data yang berhasil diolah dengan cara mencari rata-rata sub kelompok yang dibutuhkan dari data aslinya. Berdasarkan yang terlihat pada gambar tersebut bahwa terdapat rata-rata perbulan sektor makanan dan rata-rata perbulan sektor pendidikan.

```
- Uji Normalitas Data -
Menentukan apakah data yang akan di uji mean berdistribusi normal atau tidak

Data : data_rata$Rata.SK1.Makanan

1. Hipotesis :
H0 : Data berdistribusi Normal
H1 : Data tidak berdistribusi Normal

2. Tingkat Signifikansi = 0.05

3. Statistika uji :
Nilai p-value : 0.9847946

4. Kriteria uji :
H0 ditolak jika p-value < 0.05

5. Kesimpulan : data_rata$Rata.SK1.Makanan yaitu Data berdistribusi Normal
```

**Gambar 2.** Uji Normalitas Data Sektor Makanan

```
- Uji Normalitas Data -  
Menentukan apakah data yang akan di uji mean berdistribusi normal atau tidak  
  
Data : data_rata$Rata.SK9.Pendidikan  
  
1. Hipotesis :  
H0 : Data berdistribusi Normal  
H1 : Data tidak berdistribusi Normal  
  
2. Tingkat Signifikansi = 0.05  
  
3. Statistika uji :  
Nilai p-value : 0.2231659  
  
4. Kriteria uji :  
H0 ditolak jika p-value < 0.05  
  
5. Kesimpulan : data_rata$Rata.SK9.Pendidikan yaitu Data berdistribusi Normal
```

**Gambar 3.** Uji Normalitas Data Sektor Pendidikan

Berdasarkan hasil uji normalitas, pada Gambar 2 dan Gambar 3 dapat dilihat bahwa distribusi data normal dan memenuhi asumsi normalitas. Berdasarkan yang terlihat pada kedua gambar tersebut, dapat disimpulkan bahwa kedua data tersebut berdistribusi normal.

```
- Uji Kesamaan Variansi Dua Populasi -  
Menentukan apakah populasi 1 dan populasi 2 yang akan di uji mean terdapat kesamaan variansi atau tidak  
  
Populasi 1 : data_rata$Rata.SK1.Makanan  
Populasi 2 : data_rata$Rata.SK9.Pendidikan  
  
1. Hipotesis :  
H0 : Variansi populasi 1 sama dengan variansi populasi 2  
H1 : Variansi populasi 1 tidak sama dengan variansi populasi 2  
  
2. Tingkat Signifikansi = 0.05  
  
3. Statistika uji :  
Statistik Uji F : 6.150005  
Nilai p-value: 0.002743939  
  
4. Kriteria uji :  
H0 ditolak jika p-value < 0.05  
  
5. Kesimpulan :  
Terdapat perbedaan yang signifikan antara variansi populasi 1 dan populasi 2
```

**Gambar 4.** Uji Kesamaan Variansi Dua Populasi

Hasil uji kesamaan variansi dengan tingkat signifikansi sebesar 5% menunjukkan bahwa di antara populasi 1 dan populasi 2 terdapat perbedaan yang signifikan antara variansi populasi 1 dan populasi 2. Dikarenakan nilai p-value kurang dari tingkat signifikansi. Hal ini menunjukkan bahwa kedua populasi memiliki tingkat dispersi yang berbeda secara signifikan atau dapat disimpulkan bahwa terdapat perbedaan antara kedua variansi populasi tersebut.

```
- Uji Mean 2 Populasi Saling Bebas -  
  
1. Hipotesis :  
H0 : Rata - rata populasi 1 sama dengan rata - rata populasi 2  
H1 : Rata - rata populasi 1 tidak sama dengan rata - rata populasi 2  
  
2. Tingkat Signifikansi = 0.05  
  
3. Statistika uji :  
Nilai |T-hitung| = 7.261982  
  
4. Kriteria uji :  
H0 ditolak jika |T-hitung| > -2.129992  
  
5. Kesimpulan :  
Terdapat perbedaan yang signifikan antara rata - rata populasi 1 dan rata - rata populasi 2
```

**Gambar 5.** Uji Mean Dua Populasi Saling Bebas

Berdasarkan yang terlihat pada gambar 5. kriteria uji pada uji mean tersebut adalah H0 di tolak jika nilai  $t_{hitung}$  lebih besar dari nilai  $t_{tabel}$ . Oleh karena itu hasil uji mean menunjukkan kesimpulan bahwa terdapat perbedaan yang signifikan antara rata-rata populasi 1 dan rata-rata populasi 2.



#### **IV. KESIMPULAN**

Dapat disimpulkan bahwa Indeks Harga Konsumen (IHK) pada sektor makanan dan pendidikan memiliki rata-rata yang berbeda yang berarti tingkat inflasi pada kedua sektor tersebut berbeda. Tingkat inflasi yang berbeda ini menggambarkan bahwa inflasi pada setiap sektor berbeda yang dapat diartikan tidak meratanya kenaikan atau penurunan inflasi antar kelompok dan sub kelompok secara signifikan. Misalnya, suatu instansi dapat menggunakan hasil perbandingan ini untuk menentukan kebijakan moneter atau kebijakan fiskal yang lebih tepat sesuai dengan kebutuhan kelompok-kelompok yang berbeda.

#### **UCAPAN TERIMA KASIH**

Terima kasih kami ucapkan kepada Badan Pusat Statistik (BPS) yang telah menyediakan dataset mengenai Indeks Harga Konsumen (IHK) ini, sehingga kami bisa melakukan analisis terhadap dataset tersebut untuk membandingkan rata-rata sub kelompok.

#### **REFERENSI**

1. Nabila Aulia Putri Ganessa; Sheilta Alphenia; Aisya Putri Zanuazizqi;dkk, ANALISIS FAKTOR YANG MEMPENGARUHI INDEKS HARGA KONSUMEN, *Jurnal Khazanah Agustus 2021, Volume 13 No. 1*, 14-23.
2. Farah Yuni Lestari; Moh. Yamin Darsyah, Peramalan Indeks Harga Konsumen di Indonesia Menggunakan Metode *Moving average* dan *Holt Exponential Smoothing*, *Prosiding Seminar Nasional Mahasiswa Unimus 2018. Vol. 1*, 400-404
3. Amir Salim; Fadilla, Pengaruh Inflasi Terhadap Pertumbuhan Ekonomi Indonesia, *Ekonomika Sharia: Jurnal Pemikiran dan Pengembangan Ekonomi Syariah Agustus 2021, Volume 7 Nomor 1*, 17-28
4. Inflasi. Available online: <https://www.bps.go.id/subject/3/inflasi.html#subjekViewTab2> (accessed on 31 May 2023)



# Peramalan Penumpang Pelabuhan Terbesar Indonesia Efek Pandemi COVID-19: Pendekatan Pemodelan *Seasonal Autoregressive Integrated Moving Average* dan Intervensi

Regita Putri Permata<sup>1</sup>, Rifdatun Ni'mah<sup>2</sup>, Irma Octavia Chaniago<sup>3</sup>

<sup>1,2,3</sup>Sains Data – Institut Teknologi Telkom Surabaya

<sup>2</sup>[rifdatun@ittelkom-sby.ac.id](mailto:rifdatun@ittelkom-sby.ac.id)

<sup>3</sup>[irma.octavia.21@student.ds.ittelkom-sby.ac.id](mailto:irma.octavia.21@student.ds.ittelkom-sby.ac.id)

Corresponding author email: [regitapermata@ittelkom-sby.ac.id](mailto:regitapermata@ittelkom-sby.ac.id)

**Abstract:** Forecasting of the movement of port passengers which tends to fluctuate and is uncertain is needed to increase the accuracy of port development. Modeling and forecasting the number of port passengers requires methods that can capture the effect of policies on simplifying human movement during the COVID-19 pandemic, one of which is an intervention model. The SARIMA forecasting model was built from data on the monthly number of passengers at the main domestic ports from January 2006 to February 2023. The results showed that there was a drastic decrease in the four ports in May 2020 due to social modifications. Intervention analysis is able to determine the size and effect of the Covid-19 pandemic on the number of port passengers. The SARIMA Intervention Model is the application of appropriate intervention analysis to explain the dynamics and impacts of interruptions and time series changes in more detail and precisely with seasonal patterns in the data.

**Keywords:** Intervention Analysis, Covid-19 Pandemic, SARIMA.

**Abstrak:** Peramalan terhadap pergerakan penumpang pelabuhan yang cenderung fluktuatif dan tidak menentu diperlukan untuk meningkatkan ketepatan pengembangan pelabuhan. Pemodelan dan peramalan jumlah penumpang pelabuhan membutuhkan metode yang dapat menangkap efek kebijakan pembatasan pergerakan manusia selama masa pandemi COVID-19, salah satunya model intervensi. Model peramalan SARIMA dibangun dari data jumlah penumpang secara bulanan di pelabuhan utama dalam negeri mulai periode Januari 2006 hingga Februari 2023. Pembentukan fungsi intervensi ditetapkan pada Maret 2020 sesuai pemberlakuan kebijakan pembatasan pergerakan. Hasil penelitian menunjukkan bahwa terjadi penurunan drastis pada keempat pelabuhan pada bulan Mei 2020 akibat pembatasan sosial. Analisis intervensi mampu mengetahui besar dan juga efek pandemi Covid-19 terhadap jumlah penumpang pelabuhan. Model Intervensi SARIMA merupakan penerapan analisis intervensi yang tepat untuk menjelaskan dinamika dan dampak interupsi dan perubahan deret waktu secara lebih rinci dan tepat dengan pola musiman pada data.

**Kata kunci:** Analisis Intervensi, Pandemi Covid-19, SARIMA.

## I. PENDAHULUAN

Arus penumpang di pelabuhan cenderung fluktuatif dan tidak menentu. Pengetahuan akan keadaan arus penumpang diperlukan agar pengembangan Pelabuhan yang dilakukan tepat dan berguna. Peramalan terhadap arus penumpang Pelabuhan dapat dilakukan untuk mengetahui bagaimana arus penumpang [1]. Penelitian peramalan arus penumpang pelabuhan telah banyak dilakukan. Pratiwi *et al.* [2] meramal arus penumpang pelabuhan Tanjung Priok dengan menggunakan metode ARIMA Box-Jenkins dan metode variasi kalender ARIMAX. Peramalan arus penumpang pelayaran menggunakan ARIMA Box Jenkins juga diterapkan oleh Farri dan Irhamah [3] untuk Rute Surabaya-Jayapura di Pelabuhan Tanjung Perak. Putri dan Sofro [4] meramalkan jumlah keberangkatan penumpang pelayaran dalam negeri di pelabuhan Tanjung Perak menggunakan metode ARIMA dan SARIMA.

Desember 2019, muncul virus Covid-19 yang menyebabkan pandemi dan meningkat dengan kecepatan drastis hingga mencakup hampir seluruh dunia. Indonesia mengkonfirmasi adanya virus COVID-19 telah menyebar di Indonesia per 1 Maret 2020. Covid-19 menyerang manusia meskipun memiliki tingkat kematian yang rendah dibandingkan dengan sindrom pernapasan akut parah (SARS) dan sindrom pernapasan Timur Tengah (MERS). Virus ini memiliki transmisi yang tinggi menuntut adanya protokol dalam pembatasan pergerakan manusia. Protokol tersebut berimbas pada arus penumpang di pelabuhan yang dibatasi oleh pemerintah. Akibat adanya peristiwa tersebut jumlah penumpang pelabuhan selama pandemi Covid-19 turun secara drastis bahkan pernah tanpa adanya



penumpang sama sekali, padahal sebelumnya berpola musiman naik setiap bulan-bulan libur panjang. Selama Januari–Oktober 2020, jumlah penumpang angkutan laut mencapai 11,7 juta orang atau turun sebesar -40,28% dibanding periode yang sama tahun 2019. Penurunan jumlah penumpang terjadi di semua pelabuhan utama, yaitu Pelabuhan Tanjung Priok sebesar -76,34%, Belawan sebesar -73,36%, Makassar sebesar -67,49%, dan Tanjung Perak sebesar -58,14% [5]. Pemodelan dan peramalan jumlah penumpang pelabuhan membutuhkan metode yang dapat menangkap efek dari kejadian tersebut, salah satunya model intervensi.

Analisis intervensi telah digunakan dalam beberapa penelitian, yaitu Model Analisis Intervensi ARIMA Peramalan Permintaan Pariwisata Selama Pandemi Covid-19 dengan metode ARIMAX dan Pemodelan Intervensi. Hasil penelitian tersebut menyatakan bahwa model intervensi memberikan prakiraan lebih akurat dibandingkan ARIMAX serta mampu memberikan besar pengaruh intervensi terhadap jumlah pengunjung internasional yang berkunjung ke Indonesia [6]. Dalam kasus lain, penerapan fungsi step pada intervensi dan pembelajaran mesin ekstrem untuk peramalan penumpang bandara di Sorong. Peneliti mengusulkan beberapa metode untuk meramalkan jumlah penumpang bandara di Sorong salah satunya model ARIMA intervensi. Hasil penelitian menunjukkan bahwa intervensi ARIMA dapat menurunkan nilai MAPE sebesar 45% [7]. Model analisis intervensi sangat tepat digunakan untuk pola data yang memiliki kecenderungan tiba-tiba naik ataupun tiba-tiba turun. Penelitian dengan intervensi juga dilakukan untuk meramalkan jumlah penumpang domestik di Bandara Sultan Hasanuddin. Hasil penelitiannya menunjukkan bahwa intervensi digunakan untuk memodelkan data dengan peristiwa tertentu dalam waktu singkat, seperti kecelakaan, bencana alam, dan promosi [8]. Peramalan dengan model intervensi dalam model SARIMA pernah dilakukan oleh Christie *et al.* [9] untuk memprediksi jumlah pengunjung objek Wisata Londa akibat efek pandemi Covid-19. Penelitian dengan menerapkan analisis intervensi dalam model SARIMA juga dilakukan oleh Widianingsih *et al.* [10] untuk memprediksi data laju inflasi Kota Tasikmalaya akibat adanya adanya kenaikan harga minyak mentah dunia.

Peneliti mengawali pendekatan model deret waktu dengan SARIMA dan akan dilanjutkan dengan penerapan intervensi pada SARIMA untuk menangkap efek perubahan drastis yang diakibatkan oleh pandemi Covid-19. Penelitian ini bertujuan untuk meramalkan arus penumpang pada pelabuhan utama dalam negeri dengan menggunakan model SARIMA dan mempertimbangkan model intervensi akibat efek pandemi Covid-19.

## II. METODE PENELITIAN

### 3.1 Tinjauan Pustaka

#### 1. ARIMA

ARIMA( $p, d, q$ ) adalah model deret waktu nonstasioneritas yang diturunkan dari ARMA( $p, q$ ) sedemikian rupa sehingga membutuhkan  $d$  kali pembeda agar stasioner. Bentuk umum ARIMA ( $p, d, q$ ) diberikan sebagai berikut [11].

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t. \quad (1)$$

Model ARIMA perkalian dengan periode musiman  $s$  dilambangkan dengan ARIMA ( $p, d, q$ )( $P, D, Q$ )<sup>s</sup>. Bentuk umum model ARIMA musiman diberikan sebagai berikut.

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D \dot{Y}_t = \theta_q(B)\Theta_Q(B^s)a_t, \quad (2)$$

dimana  $\phi_p(B)$  dinotasikan sebagai operator AR,  $\Phi_p(B^s)$  dinotasikan sebagai operator seasonal AR,  $\theta_q(B)$  dinotasikan sebagai operator MA,  $\Theta_Q(B^s)$  dinotasikan sebagai operator seasonal MA, dan  $a_t$  sebagai residual *white noise* dengan rata-rata nul dan varians konstan  $\sigma_a^2$ .

## 2. Analisis Intervensi

Model deret waktu paling populer yang banyak digunakan dalam peramalan data deret waktu adalah model Autoregressive Integrated Moving Average, juga dikenal sebagai model ARIMA. Dalam penerapannya, model ini membutuhkan pemenuhan asumsi stasioneritas pada mean (mean) dan varians dari deret waktu.

Model intervensi adalah model analisis data deret waktu yang awalnya banyak digunakan untuk membahas peristiwa eksternal yang dianggap terkait dengan variabel yang diamati. Untuk prosesnya ikuti ARIMA  $(p,d,q)(P,D,Q)^S$  model, bentuk persamaan matematis dapat ditulis [11]:

$$\phi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\theta_Q(B^S)a_t, \quad (3)$$

atau

$$Y_t = \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t, \quad (4)$$

dengan,

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

$B$  adalah *backshift*, sebagai contohnya:  $B^k Y_t = Y_{t-k}$ .

Jika didefinisikan untuk  $N_t = \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t$ , maka persamaan (4) dapat dituliskan  $Y_t = N_t$ .

Persamaan model dari (4), untuk  $d = 0$  dapat diartikan bahwa perubahan  $Y_t$  adalah sebagai  $a_t$ . Jika diperhatikan ada pengaruh dari beberapa kejadian intervensi  $X_t$  untuk deret waktu, maka kita dapat menulis model umum sebagai berikut

$$Y_t = f(X_t) + N_t, \quad (5)$$

dengan  $Y_t$  adalah variabel respon pada saat  $t$ ,  $X_t$  adalah variabel intervensi dan  $N_t$  adalah model *noise* yang mengikuti ARIMA  $(p,d,q)$ .

Secara umum ada dua macam variabel intervensi, yaitu fungsi *step* (*step function*) dan fungsi *pulse* (*pulse function*). *Step function* adalah suatu bentuk intervensi yang terjadinya dalam kurun waktu yang panjang. Secara matematik, bentuk intervensi *step function* ini biasanya dinotasikan sebagai berikut

$$X_t = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \quad (6)$$

dimana  $T$  adalah waktu mulainya terjadi intervensi.

Sedangkan *pulse function* adalah suatu bentuk intervensi yang terjadinya hanya dalam suatu waktu tertentu. Secara matematik, bentuk intervensi *pulse function* ini biasanya dinotasikan sebagai berikut

$$X_t = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases} \quad (7)$$

dimana  $T$  adalah waktu terjadinya intervensi.

### 3. Ukuran Akurasi Peramalan

Kami menggunakan *Root Mean Square Error of Prediction* (RMSEP) dan *Mean Average Percentage Error* (MAPE) sebagai kriteria pemilihan model. Rumus RMSEP diberikan masing-masing sebagai berikut [11].

$$RMSEP = \sqrt{\frac{\sum_{l=1}^L (Y_{n+l} - \hat{Y}_n(l))^2}{L}} \quad (8)$$

$$MAPE = \frac{\sum_{l=1}^L |Y_{n+l} - \hat{Y}_n(l)|}{Y_{n+1}} \times 100\% \quad (9)$$

dimana  $Y_{n+1}$  menunjukkan nilai sebenarnya,  $\hat{Y}_n(l)$  menunjukkan nilai perkiraan, dan  $L$  menunjukkan horizon perkiraan.

### 3.2 Data

Data yang digunakan pada penelitian ini adalah data jumlah penumpang secara bulanan di Pelabuhan terbesar Indonesia periode Januari 2006 sampai dengan Februari 2023 yang diperoleh dari *website* resmi BPS Indonesia, <https://www.bps.go.id/>. Variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

**Tabel 1. Variabel Penelitian**

No	$Y_t$	Keterangan
1	$Y_1$	Jumlah penumpang di Pelabuhan Belawan
2	$Y_2$	Jumlah penumpang di Pelabuhan Tanjung Priok
3	$Y_3$	Jumlah penumpang di Pelabuhan Tanjung Perak
4	$Y_4$	Jumlah penumpang di Pelabuhan Makassar

Kejadian intervensi menggunakan tanggal kali pertama ditetapkan pemerintah Kasus COVID-19 di Indonesia pada 2 Maret 2020, sehingga intervensi penelitian ini pada Maret 2020.

### 3.3 Langkah Analisis

Langkah-langkah analisis data yang dilakukan dengan menggunakan SARIMA dan analisis intervensi adalah sebagai berikut.

- Melakukan pembagian data menjadi dua kelompok berdasarkan penetapan waktu terjadinya intervensi.
  - Kelompok 1: Data sebelum terjadi intervensi (Jan 2006- Feb 2020)
  - Kelompok 2: Data setelah terjadi intervensi hingga data berakhir (April 2020- Feb 2023)
- Membangun model sebelum terjadi intervensi  
Membangun model ARIMA untuk data sebelum analisis intervensi ( $N_t$ ) untuk estimasi dan diagnosis parameter intervensi, yaitu  $p$  dan  $q$ .
- Peramalan data pada periode intervensi berbasis ARIMA untuk data sebelum intervensi.
- Pembentukan fungsi intervensi
  - Menghitung fungsi respon untuk standarisasi residual model ARIMA sebelum dilakukan intervensi.
  - Identifikasi order b,r,s dari grafik residual yang telah didapatkan.
  - Mengidentifikasi model intervensi menggunakan order intervensi.
  - Melakukan penafsiran dan pengujian signifikansi parameter model intervensi.

- e. Melakukan peramalan menggunakan model dengan kesalahan minimum.

### III. HASIL DAN PEMBAHASAN

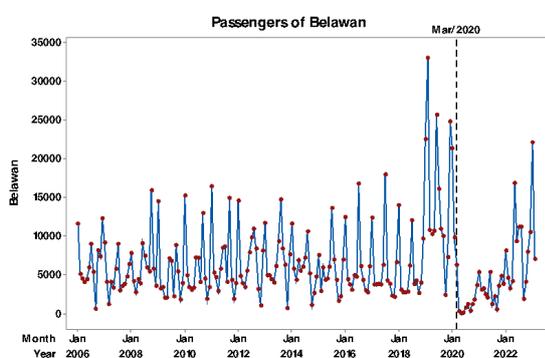
#### 4.1 Karakteristik Data

Tabel 2 menunjukkan karakteristik data berdasarkan statistika deskriptif. Penelitian ini menganalisis dari empat Pelabuhan terbesar di Indonesia yaitu Belawan, Tanjung Priok, Tanjung Perak, dan Makassar.

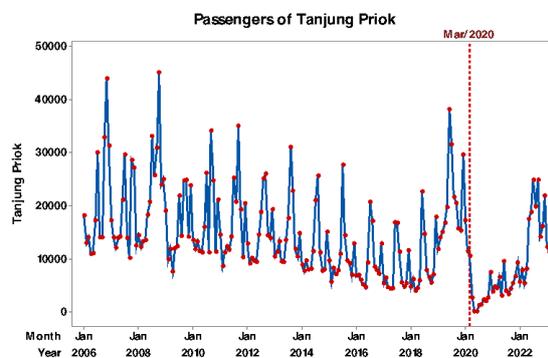
**Tabel 2.** Statistika Deskriptif Jumlah Penumpang

Statistika Deskriptif	Jumlah Penumpang			
	Belawan	Tanjung Priok	Tanjung Perak	Makassar
Rata-rata	6430,383	13930,621	33671,231	34301,995
Deviasi Standar	5014,927	8396,428	21784,003	15783,780
Minimum	0	30	2810	118
Maksimum	33003	45064	132579	117195

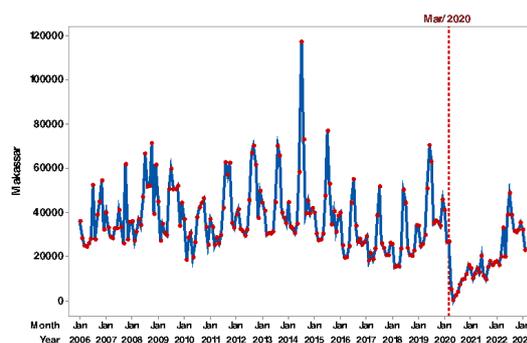
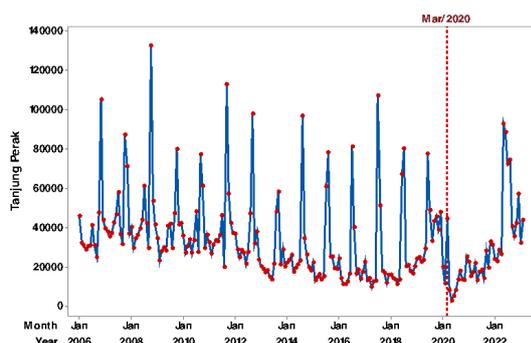
Hasil eksplorasi data jumlah penumpang di empat Pelabuhan terbesar Indonesia menunjukkan bahwa rata-rata jumlah penumpang Makassar lebih besar dari tiga Pelabuhan lainnya dengan mencapai 34302 orang selama kurang lebih 16 tahun. Titik terendah jumlah penumpang nul di Pelabuhan Belawan terjadi pada kasus COVID-19 tinggi di Indonesia yaitu pada Mei 2020. Pada bulan yang sama kondisi jumlah penumpang terendah juga dialami oleh Pelabuhan Tanjung Priok, Tanjung Perak, dan Makassar. Standar deviasi yang beragam dari Pelabuhan tersebut menunjukkan angka tersebut pun beragam.



**Gambar 1a.**



**Gambar 1b.**

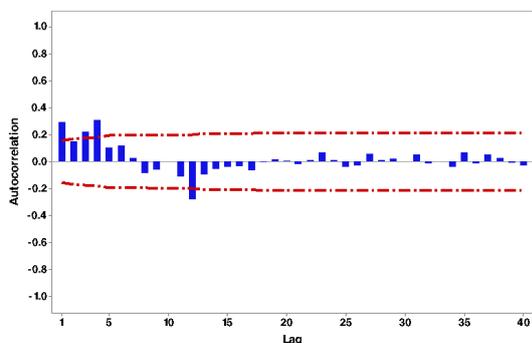
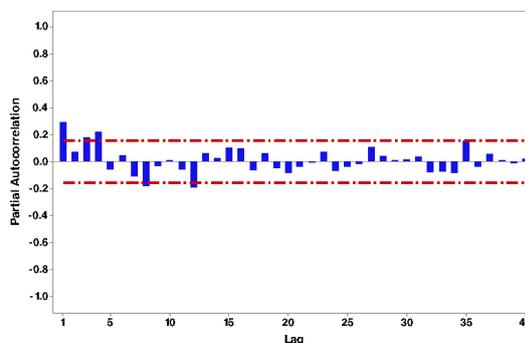


**Gambar 1c.**
**Gambar 1d.**
**Gambar 1.** Pola data Jumlah Penumpang: (1a) Belawan (1b) Tanjung Priok (1c) Tanjung Perak (1d) Makassar

Jumlah penumpang dari empat pelabuhan menunjukkan bahwa data memiliki pola tren, dan pola musiman serta adanya kecenderungan polanya sama setiap tahun. Imbas Covid-19 mengakibatkan jumlah penumpang turun drastis bahkan sebelum bulan Maret 2020. Selama liburan musim, pola musiman yang jelas, seperti pada bulan Juli dan Desember, jumlah penumpang akan meningkat, kemudian menurun di luar musim liburan.

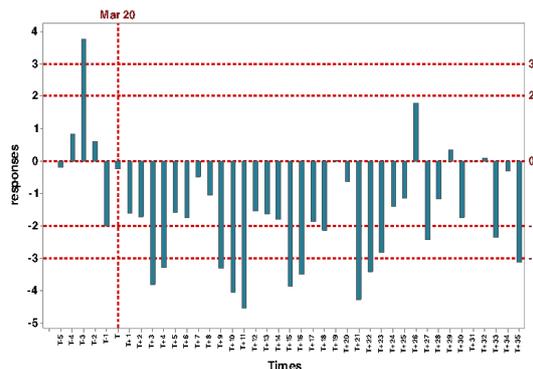
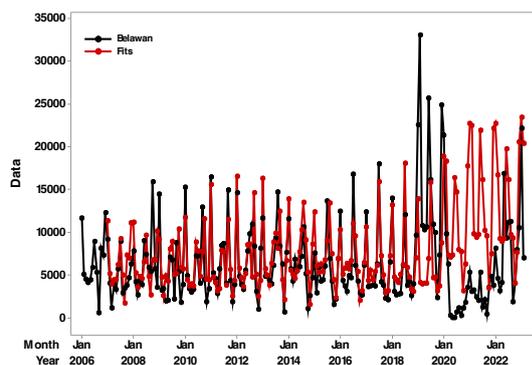
#### 4.2 Pemodelan "Jumlah Penumpang Belawan"

Untuk membangun model ARIMA, data jumlah penumpang Belawan dipotong sebelum tanggal intervensi Maret 2020, sehingga diperoleh 170 data. Untuk membuat deret waktu stabil dan stasioner, digunakan uji ADF yang membuktikan bahwa deret waktu berisi akar unit yang mengambil satu perbedaan, yaitu  $d = 12$ . Sehingga grafik stasioner ACF dan PACF disajikan pada Gambar 2.


**Gambar 2a.**

**Gambar 2b.**
**Gambar 2.** Plot Korelasi Residual Jumlah Penumpang Belawan: (2a) *Autocorrelation Function* (2b) *Partial Autocorrelation Function*

Hasil dari Gambar 2 dapat diduga sementara melalui plot ACF dan PACF sehingga dugaan model SARIMA  $(1,1,0)^{12}$  atau SARIMA  $(0,1,1)^{12}$ , hasil error minimum dengan model SARIMA  $(1,1,0)^{12}$ . Bagian ini menyajikan hasil model intervensi dengan mengilustrasikan dampak dari fungsi *step*. Secara matematis, jenis intervensi fungsi *step* dari indikator intervensi deterministik (dummy) ditulis sebagai:

$$S_t = \begin{cases} 0, & t < 172 \\ 1, & t \geq 172 \end{cases}$$



**Gambar 3a.**
**Gambar 3b.**
**Gambar 3.** Pemodelan ARIMA data Jumlah Penumpang Belawan: (3a) Peramalan model SARIMA (1,1,0)<sup>12</sup> (3b) Nilai Respon Residual

Menurut Gambar 3b, dapat dilihat bahwa nilai respon pada waktu mulai dari T+3 memiliki selang kepercayaan lebih dari -3 sampai 3. Jadi kemungkinan  $b = 3$ , dan orde  $s$  mulai dari  $T > 3$  yang mana nilainya adalah 1,6,7,8,12,13,17,18,19 serta  $r=0$ . Estimasi parameter dan uji signifikansi menunjukkan bahwa urutan model ini menghasilkan parameter yang signifikan dengan  $s=1,6,8$ . Tabel 3 menunjukkan model intervensi output.

**Tabel 3.** Hasil Estimasi Parameter Model Intervensi “Jumlah Penumpang Belawan”

Parameter	Estimasi	t Value	P-value	Lag	Tipe
AR 1	-0,26029	-3,17	0,0015	12	Signifikan
$\omega_0$	-20216	-4,29	0,0001	0	Signifikan
$\omega_1$	-12589,3	-2,51	0,0122	1	Signifikan
$\omega_6$	9572,3	2,55	0,0107	6	Signifikan
$\omega_8$	-19806,8	-5,78	0,0001	8	Signifikan

Tabel 3 menunjukkan bahwa parameter signifikan sehingga urutan intervensi adalah  $b=3$ ,  $s=(1,6,8)$ , dan  $r=0$ . Hal ini menunjukkan bahwa efek pengumuman kasus pertama COVID-19 terhadap penurunan jumlah penumpang terdeteksi pada T+3 atau Juni 2020 sebesar -20.216 orang namun sebenarnya dampak hingga 0 penumpang terjadi pada Mei 2020. Beberapa peristiwa dampak COVID-19 terhadap penurunan jumlah penumpang Belawan pasca intervensi ditunjukkan pada Tabel 4. Namun pada Mei 2022 atau T+26 terdapat efek kenaikan penumpang yaitu dari 4.092 penumpang pada bulan April 2022 menjadi 16.782 penumpang. Bulan Mei 2022 merupakan masa arus mudik-balik 2022 dimana potensi lonjakan penumpang untuk menggunakan kapal laut sebagai moda transportasi. Masa mudik tahun 2022 berbeda dengan tahun 2020 dan 2021 karena masyarakat diperbolehkan mudik pada tahun 2022 dengan adanya syarat perjalanan yang harus dipenuhi yaitu vaksin lengkap, mendapatkan booster atau vaksin dosis ketiga serta tetap menerapkan protokol kesehatan yang ketat.

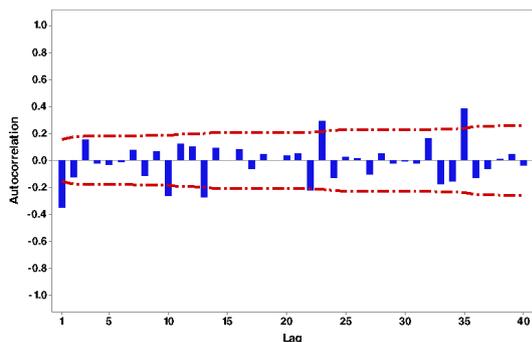
**Tabel 4.** Peristiwa Efek Covid-19 pada Pelabuhan Belawan

Waktu (t)	Bulan	Peristiwa
T+3	June, 2020	3 bulan setelah dikonfirmasi Pandemi Covid-19 di Indonesia
T+4	Juli, 2020	Kota Medan (lokasi pelabuhan Belawan) masuk ke dalam zona merah yaitu wilayah dengan resiko penyebaran Covid-19 tinggi
T+9	Des, 2020	Kota Medan menjadi satu-satunya wilayah yang masuk kategori risiko tinggi penyebaran virus Corona atau zona merah di Sumatera Utara.
T+11	Feb, 2021	Bulan Januari-Februari merupakan bulan dimana terjadinya musiman jumlah penumpang meningkat karena liburan. Namun pada Februari 2021 terdapat jenis Covid-19 varian Beta sehingga PPKM mikro se-Jawa-Bali ditingkatkan

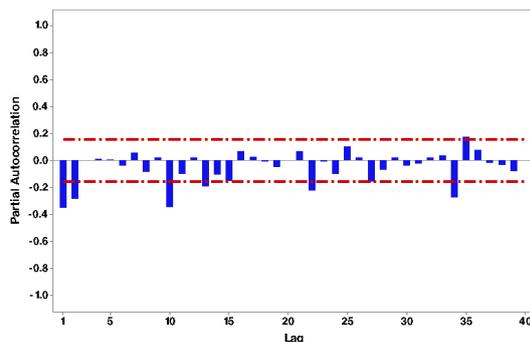
Untuk menghindari kasus Covid-19 yang semakin tinggi, pada tahun 2021 pelabuhan Belawan selalu menerapkan protocol kesehatan dengan ketat, serta adanya vaksin gratis [12]. Jadi model intervensi untuk jumlah penumpang Belawan ditunjukkan pada persamaan 10.

### 4.3 Pemodelan “Jumlah Penumpang Tanjung Priok”

Langkah pertama pemodelan dengan ARIMA adalah mendeteksi stasioneritas data. Pola data jumlah penumpang Tanjung Priok tidak stasioner sehingga distasionerkan dengan  $d=1$  dan  $D=1$  pada order musiman.



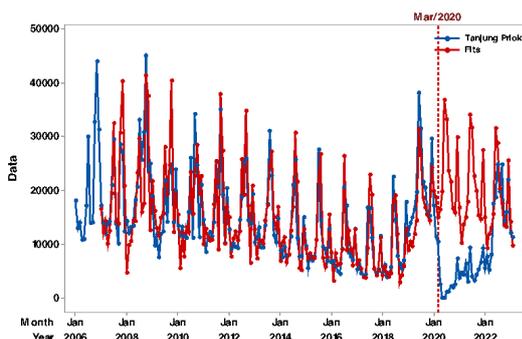
Gambar 4a.



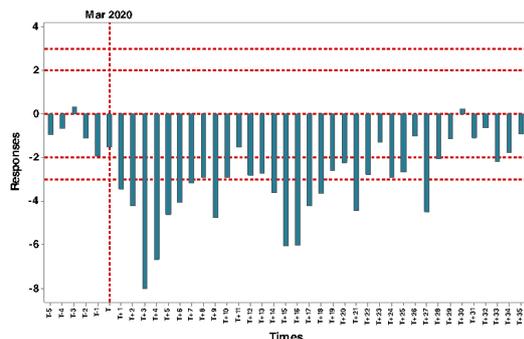
Gambar 4b.

**Gambar 4.** Plot Korelasi Residual Jumlah Penumpang Tanjung Priok: (4a) *Autocorrelation Function* (4b) *Partial Autocorrelation Function*

Gambar (4a) dan (4b) menunjukkan bahwa lag ACF diluar batas interval kepercayaan adalah lag 1,10,13,24, sedangkan lag PACF yang berada diluar batas interval adalah lag 1,2,10,13,22. Lag PACF menandakan ada komponen musiman, maka dugaan sementara model adalah SARIMA (0,0,2) (0,1,0)<sup>12</sup> atau SARIMA ([1,2,3,13,23],0,0) (0,1,0)<sup>12</sup>. Model yang dipilih adalah SARIMA (2,0,0)(3,1,0)<sup>12</sup> dengan nilai MSE minimum.



Gambar 5a.



Gambar 5b.

**Gambar 5.** Pemodelan ARIMA data Jumlah Penumpang Tanjung Priok: (5a) Peramalan model SARIMA ([1,2,3,13,23],0,0) (0,1,0)<sup>12</sup> (5b) Nilai Respon Residual

Gambar 5b dapat dilihat bahwa nilai respon setelah T+1 lebih besar dari interval kepercayaan. Pengaruh COVID-19 terhadap jumlah penumpang Tanjung Priok menurun hingga T+16. Artinya pasca kejadian COVID-19 sangat membuat jumlah penumpang Tanjung Priok turun drastis selama kurang lebih 16 bulan. Sehingga kemungkinan order  $b=2$ ,  $s = (1,2,3,4,6,8,13,14,15,16,17)$  dan untuk order  $s$  signifikan adalah  $s=(3,6,8,15)$ , dan  $r=0$ . Untuk memudahkan interpretasi set  $r = 0$ . T+1 atau April 2020, jumlah penumpang Tanjung Priok turun pasca intervensi. Estimasi parameter dan uji

signifikansi menunjukkan bahwa urutan model ini menghasilkan parameter yang signifikan dapat dilihat dari Tabel 5.

**Tabel 5.** Hasil Estimasi Parameter Model Intervensi “Jumlah Penumpang Tanjung Priok”

Parameter	Estimasi	t Value	P-value	Lag	Tipe
AR 1	-0,487	-6,77	0,0001	1	Signifikan
AR 2	-0,328	-4,5	0,0001	2	Signifikan
AR 10	-0,294	-3,9	0,0001	10	Signifikan
AR 13	-0,225	-3,02	0,0025	13	Signifikan
AR 22	-0,145	-1,96	0,0504	22	Signifikan
$\omega_0$	-10717,5	-6,1	0,0001	0	Signifikan
$\omega_3$	-17601,3	-6,45	0,0001	3	Signifikan
$\omega_6$	9477,3	2,91	0,0036	6	Signifikan
$\omega_8$	-5945,8	-2,22	0,0267	8	Signifikan
$\omega_{15}$	3576,7	2,66	0,0079	15	Signifikan

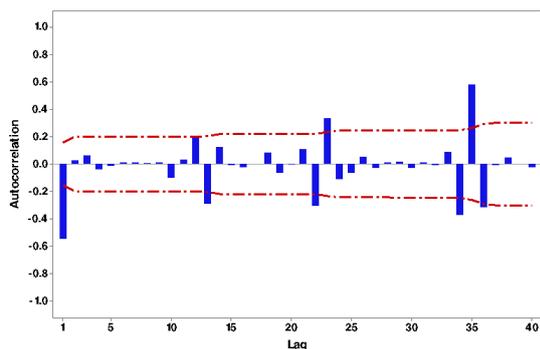
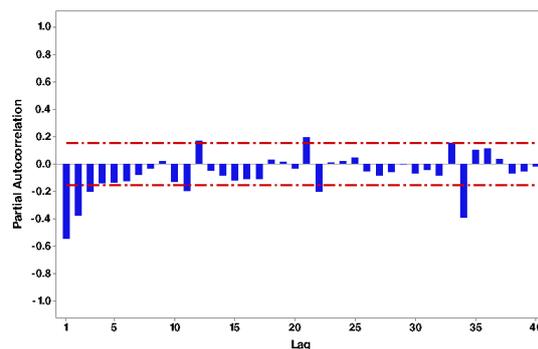
Tabel 5 menunjukkan bahwa parameter signifikan sehingga urutan intervensi adalah  $b=1$ ,  $s = (3,6,8,15)$ , dan  $r=0$  serta model SARIMA([1,2,10,13,22],1,0)(0,1,0)<sup>12</sup>. Hal ini menunjukkan bahwa efek pengumuman kasus pertama COVID-19 terhadap penurunan jumlah penumpang terdeteksi pada T+1 atau April 2020 sebesar -10717,5 orang. Beberapa peristiwa dampak COVID-19 terhadap penurunan jumlah penumpang Tanjung Priok pasca intervensi ditunjukkan pada Tabel 6.

**Tabel 6.** Peristiwa Efek Covid-19 pada Pelabuhan Tanjung Priok

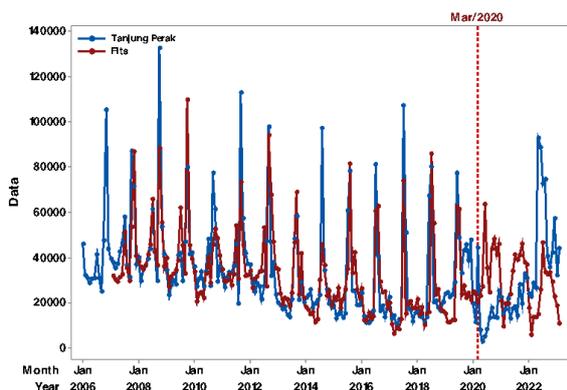
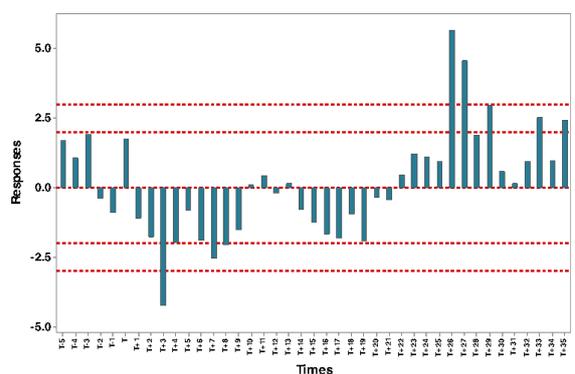
Waktu (t)	Bulan	Peristiwa
T+1	April 2020	Satu bulan setelah dikonfirmasi pandemi Covid-19 di Indonesia
T+4	Juli 2020	Lonjakan kasus Covid-19 pada masa PSBB transisi 2
T+7	Oktober 2020	Kasus lonjakan Covid-19 masih tinggi sehingga masih ada peraturan pembatasan sosial
T+16	Juli 2021	Peningkatan kasus Covid-19 akibat varian Delta, dengan kecepatan penularan 6x lebih cepat. pemerintah saat itu menetapkan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) Darurat di berbagai kabupaten/kota di Jawa-Bali.

#### 4.4 Pemodelan “Jumlah Penumpang Tanjung Perak”

Pemodelan pada jumlah penumpang Tanjung Perak perlu distasionerkan dengan  $d=1$ ,  $D=1$  untuk order nonmusiman dan musiman.


**Gambar 6a.**

**Gambar 6b.**
**Gambar 6.** Plot Korelasi Residual Jumlah Penumpang Tanjung Perak: (6a) *Autocorrelation Function* (6b) *Partial Autocorrelation Function*

Gambar 6 menunjukkan hasil *plot* ACF dan PACF yang sudah stasioner, didapatkan pola ACF ada pola musiman yang tertangkap yaitu pada lag 13, 22, 23, 34, 35 dan 36. Sedangkan pola lag pada PACF *diesdown* untuk pola nonmusiman, dan menangkap pola musiman pada lag 11,12, 21,22, dan 34. Dugaan model SARIMA untuk jumlah penumpang di Tanjung Perak adalah SARIMA  $(0,1,[1,13,22,23,34,35,36]) (0,1,0)^{12}$  atau SARIMA  $([1,2,3,13,22,23,34,35,36],1,0)(0,1,0)^{12}$ . Nilai MS minimum pada model SARIMA  $(0,1,[1,13,22,23,34,35,36])(0,1,0)^{12}$ .


**Gambar 7a.**

**Gambar 7b.**
**Gambar 7.** Pemodelan ARIMA data Jumlah Penumpang Tanjung Perak: (7a) Peramalan model SARIMA (7b) Nilai Respon Residual

Gambar 7a menunjukkan bahwa *fit* yang diprediksi dengan SARIMA tidak mampu menangkap beberapa kejadian setelah intervensi. Gambar 7b menunjukkan bahwa intervensi efek Covid-19 pada penurunan jumlah penumpang di Tanjung Perak terjadi pada T+3 atau pada bulan Juni 2020. Kemungkinan order yang tepat untuk kasus ini adalah  $b=3$ ,  $s=23$  dan 24 menggunakan interval -3 sampai 3, sedangkan  $r=0$ . Model SARIMA signifikan pada lag 1, 23, 34, dan 35 sehingga menggunakan model SARIMA  $(0,1,[1,23,34,35])(0,1,0)^{12}$ . Jumlah penumpang biasanya meningkat pada bulan Juni atau Juli. Dalam kasus ini, jumlah penumpang Bulan Juni 2020 terjadi peningkatan sebesar 79% dari kondisi paling buruk di Bulan Mei 2020 sebesar 2.810 menjadi 5.048 penumpang.

**Tabel 7.** Hasil Estimasi Parameter Model Intervensi “Jumlah Penumpang Tanjung Perak”

Parameter	Estimasi	t Value	P-value	Lag	Tipe
MA 1	0,596	6,82	0,0001	1	Signifikan
MA 23	-0,201	-2,05	0,0408	23	Signifikan

Parameter	Estimasi	t Value	P-value	Lag	Tipe
MA 34	0,325	3,03	0,0024	34	Signifikan
MA 35	-0,562	-4,48	0,0001	35	Signifikan
$\omega_0$	1435	1,1	0,2724	0	Signifikan
$\omega_{23}$	-82498,8	-6,09	0,0001	23	Signifikan
$\omega_{24}$	92913,6	6,47	0,0001	24	Signifikan

Pada Tabel 7 menunjukkan bahwa parameter signifikan order intervensi adalah  $b=3$ ,  $s = (23 \text{ dan } 24)$ , dan  $r=0$ . Untuk model SARIMA([1,23,24],1,0)(0,1,0)<sup>12</sup>. Beberapa peristiwa dampak COVID-19 terhadap penurunan jumlah penumpang Tanjung Perak pasca intervensi ditunjukkan pada Tabel 8.

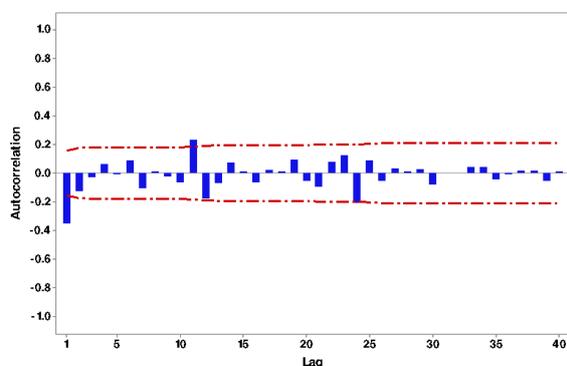
**Tabel 8.** Peristiwa Efek Covid-19 pada Pelabuhan Tanjung Perak

Waktu (t)	Bulan	Peristiwa
T+3	Juni 2020	3 bulan setelah dikonfirmasi adanya Covid-19. Adanya pola peningkatan pertama setelah terjadinya penurunan drastis di bulan Mei 2020.
T+26	Mei 2022	Efek positif melonjaknya jumlah penumpang
T+27	Juni 2022	Efek positif melonjak jumlah penumpang

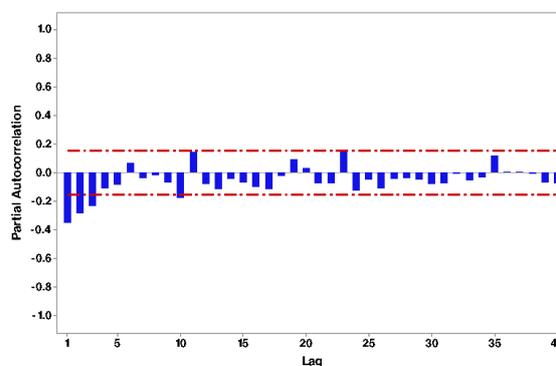
Mei dan Juni 2020 dalam prediksi harusnya meningkat jumlah penumpang karena efek musiman. Namun adanya fakta efek pandemi COVID-19 jumlah penumpang di Tanjung Perak turun drastis. Peningkatan jumlah penumpang mulai terasa imbasnya setelah T+23 atau Februari 2022 sampai Juni 2022. Namun efek positif sangat terasa pada Mei dan Juni 2022. Menurut Badan Pusat Statistik (BPS), jumlah penumpang kapal pesiar yang berangkat dari pelabuhan Tanjung Perak di Surabaya, Jawa Timur, meningkat 609,89 *year-on-year* menjadi 92.790 di bulan Mei. Hal ini disebabkan liburan Idul Fitri, serta perjalanan penumpang meningkat. Pemerintah mengizinkan masyarakat mudik untuk Idul Fitri tahun 2022.

#### 4.5 Pemodelan “Jumlah Penumpang Pelabuhan Makassar”

Pemodelan pada jumlah penumpang Pelabuhan Makassar dengan melakukan stasioneritas data dengan *differencing* nonmusiman dan musiman  $d=1$  dan  $D=1$ . Berikut ditampilkan plot korelasi residual pada Gambar 8.



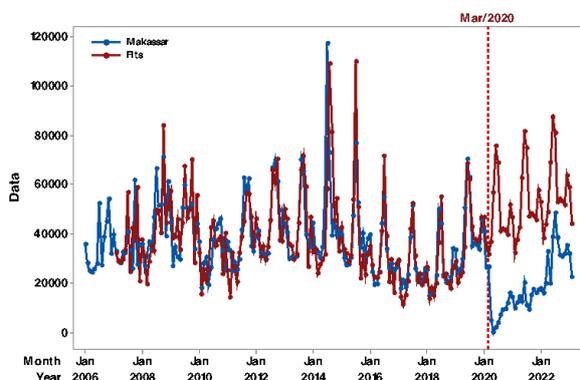
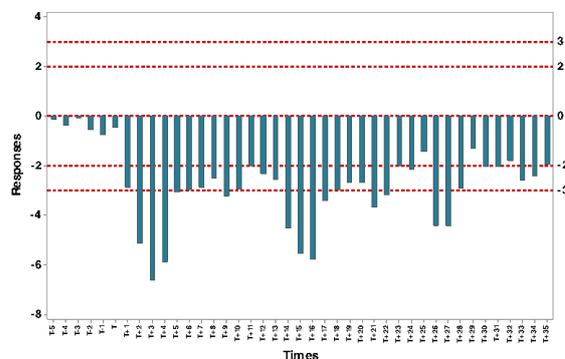
**Gambar 8a.**



**Gambar 8b.**

Gambar 8. Plot Korelasi Residual Jumlah Penumpang Makassar: (8a) *Autocorrelation Function* (8b) *Partial Autocorrelation Function*

Menurut Gambar 8 dapat diduga model ARIMA sementara berdasarkan *lag* signifikan dari ACF dan PACF adalah Model SARIMA (3,1,0) (0,1,0)<sup>12</sup> atau SARIMA(0,1,1) (0,1,1)<sup>12</sup>, model yang digunakan dengan MSE minimum adalah SARIMA (0,1,1) (0,1,1)<sup>12</sup>.


**Gambar 9a.**

**Gambar 9b.**

**Gambar 9.** Pemodelan ARIMA data Jumlah Penumpang Makassar: (9a) Peramalan model SARIMA (9b) Nilai Respon Residual

Menurut Gambar (9a), justifikasi model SARIMA yang digunakan tidak mampu memprediksi data aktual penumpang pelabuhan Makassar setelah dikonfirmasi terjadi Covid-19. Gambar (9b) menunjukkan bahwa nilai respon pada waktu mulai dari  $T+2$  dengan selang kepercayaan lebih dari  $-3$  sampai  $3$ . Jadi kemungkinan nilai respon order  $b = 2$ , dan orde  $s$  mulai dari  $T > 2$  yang mana nilainya adalah 1,2,12,13,14,15,19,20,24,25 serta  $r=0$ . Setelah dilakukan estimasi parameter model intervensi dengan nilai respon  $s$  signifikan adalah 1 dan 13.

**Tabel 9.** Hasil Estimasi Parameter Model Intervensi “Jumlah Penumpang Makassar”

Parameter	Estimasi	t Value	P-value	Lag	Tipe
MA 1	0,6288	10,48	0,0001	1	Signifikan
MA 12	0,3892	5,33	0,0001	12	Signifikan
$\omega_0$	-42885,7	-5,01	0,0001	0	Signifikan
$\omega_1$	-47531,8	-5,2	0,0001	1	Signifikan
$\omega_{13}$	4637,3	2,54	0,0110	13	Signifikan

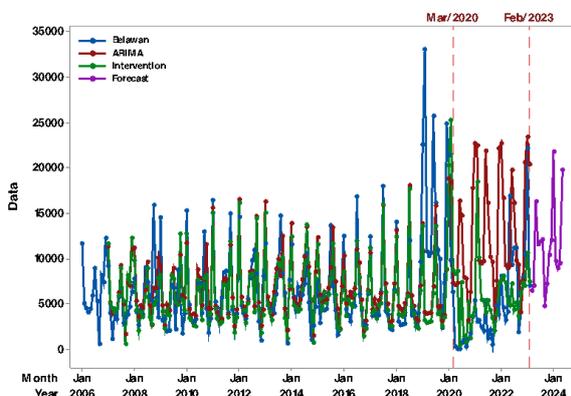
Estimasi parameter pada tabel 9 menunjukkan signifikan. Dari Tabel 9 terlihat bahwa setelah kejadian intervensi dua bulan, COVID-19 berdampak  $-42.885,7$  pada jumlah penumpang pelabuhan Makassar. Beberapa kejadian dampak COVID-19 terhadap penurunan jumlah penumpang setelah kejadian intervensi ditunjukkan pada Tabel 10.

**Tabel 10.** Peristiwa Efek Covid-19 pada Pelabuhan Makassar

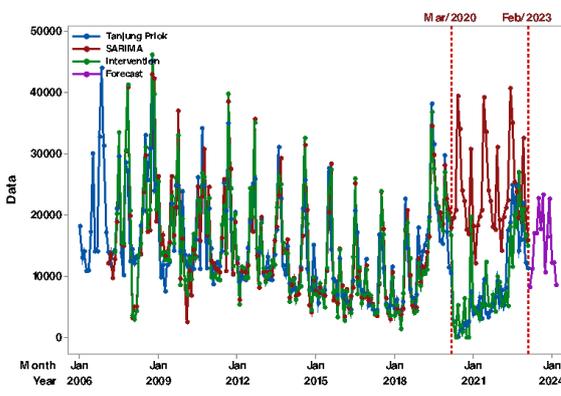
Waktu (t)	Bulan	Peristiwa
T+2	Mei 2020	2 bulan setelah dikonfirmasi adanya Covid-19. Adanya penurunan drastis di bulan Mei 2020. Imbas <i>social distancing</i>
T+3	Juni 2020	Adanya pola peningkatan pertama setelah terjadinya penurunan drastis di bulan Mei 2020.
T+16	Juli 2021	Pemerintah saat itu menetapkan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) Darurat di berbagai kabupaten/kota di Jawa-Bali.

#### 4.6 Peramalan Model Intervensi

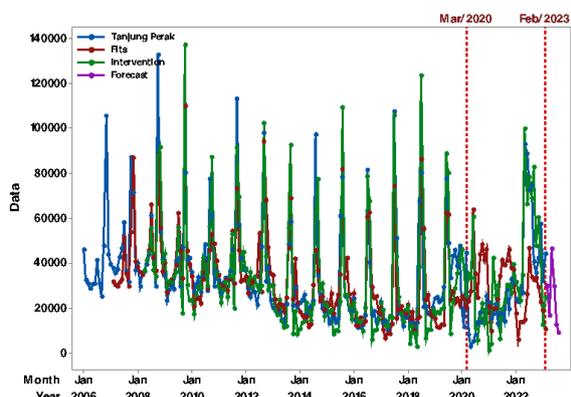
Pemodelan intervensi yang telah dilakukan dapat menghitung hasil peramalan jumlah penumpang selama 12 bulan mendatang disajikan pada Gambar 10.



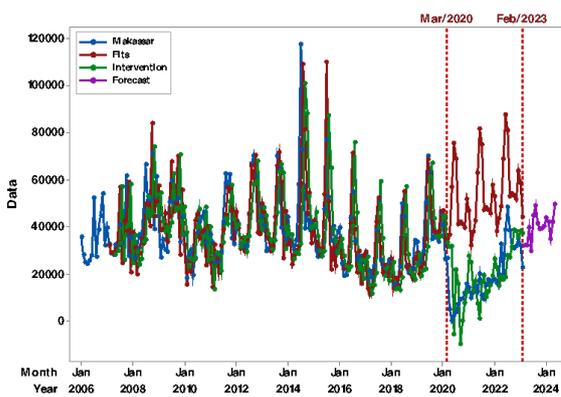
Gambar 10a.



Gambar 10b.



Gambar 10c.



Gambar 10d.

Gambar 10. Peramalan Jumlah Penumpang;(10a) Belawan (10b) Tanjung Priok (10c) Tanjung Perak (10d) Makassar

#### IV. KESIMPULAN

Berdasarkan hasil analisis mengenai penelitian jumlah penumpang Pelabuhan terbesar di Indonesia menunjukkan bahwa pandemic Covid-19 sangat berdampak signifikan pada pola dan tren jumlah penumpang yang menyebabkan penurunan drastis dari Maret hingga Juni 2020. Efek intervensi dari pandemic Indonesia terhadap jumlah penumpang kapal diterapkan pada bulan Maret 2020. Adanya efek tak terduga yang terjadi pada bulan Mei dan Juni 2022 untuk ke-empat pelabuhan utama tersebut. Hal positif ini disebabkan oleh liburan Idul Fitri, serta perjalanan penumpang meningkat. Pemerintah mengizinkan masyarakat mudik untuk Idul Fitri tahun ini. Hasil penelitian ini menunjukkan bahwa model Intervensi SARIMA merupakan penerapan analisis intervensi yang tepat untuk menjelaskan dinamika dan dampak interupsi dan perubahan deret waktu secara lebih rinci dan tepat. Saran untuk penelitian selanjutnya adalah menggunakan deteksi outlier, sehingga outlier dapat diidentifikasi dan model intervensi dapat terdistribusi secara normal.

#### UCAPAN TERIMA KASIH

**REFERENSI**

1. B. P. Dewani, S. Suhartono and M. S. Akbar, "Peramalan Jumlah Penumpang dan Barang di Bandar Udara Internasional Juanda dan Pelabuhan Tanjung Perak Menggunakan Model Hybrid ARIMAX dan Deep Learning Neural Networks," *Inferensi*, vol. 2, no. 1, pp. 1-11, 2019.
2. A. Pratiwi, D. Safitri and B. Warsito, "Peramalan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Priok dengan Metode Arima Box-Jenkins dan Metode Variasi Kalender Arimax," *Statistika*, vol. 6, no. 1, pp. 56-64, Mei 2018.
3. E. A. Farri and I. Irhamah, "Peramalan Penumpang Angkutan Laut Rute Surabaya-Jayapura di PT Pelayaran Nasional Indonesia (Persero) Cabang Surabaya," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, pp. 169-176, 2020.
4. S. Putri and A. Sofro, "Peralaman Jumlah Keberangkatan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Perak Menggunakan Metode Arima dan Sarima," *MATHunesa Jurnal Ilmiah Matematika*, vol. 10, no. 1, pp. 61-67, 2022.
5. Office of Chief Economist, "Daily Economic and Market Review," Bank Mandiri, Jakarta, 2020.
6. F. Rianda and H. Usman, "Forecasting Tourism Demand during the COVID-19 Pandemic: ARIMAX and Intervention Modelling Approaches," *Journal of Mathematics and Its Applications*, vol. 17, no. 1, p. 0285–0294, 2023.
7. N. Faizin, A. Fauzan and A. H. Primandari, "Implementation of the Step Function Intervention and Extreme Learning Machine for Forecasting the Passenger's Airport in Sorong," *Journal of Mathematics and Its Applications*, vol. 17, no. 1, p. 0535–0544, 2023.
8. A. F. Sustrisno, R. and I. Setiawan, "Intervention Model Analysis the Number of Domestic Passengers at Sultan Hasanuddin Airports," *Journal of Statistics*, vol. 1 nomor 1, pp. 41-49, 2021.
9. G. Christie, D. Hatidja and R. Tumilaar, "Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application of the SARIMA Method in the Step Function Intervention to Predict the Number of Visitors at Londa Tourism Object)," *JURNAL ILMIAH SAINS*, vol. 22, no. 2, pp. 96-103, 2022.
10. P. Widianingsih, G. Darmawan and N. Sunengsih, "Analisis Intervensi dalam Model SARIMA untuk Memprediksi Laju Inflasi di Kota Tasikmalaya," *Formosa Journal of Science and Technology*, vol. 1, no. 4, pp. 293-304, 2022.
11. W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*. 2nd edition, USA: Pearson Education Inc., 2006.
12. Dinas Kominfo Kota Medan, "Wali Kota Medan Tinjau Pelaksanaan Vaksinasi Massal di Polresta Pelabuhan Belawan," Pemerintah Kota Medan, 5 Juli 2021. [Online]. Available: <https://pkk.pemkomedan.go.id/artikel-21268-wali-kota-medan-tinjau-pelaksanaan-vaksinasi-massal-di-polresta-pelabuhan-belawan.html>. [Accessed 29 Mei 2023].
13. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, Second Edition, United States: Pearson Education, Inc, 2006.



# Analisis Sentimen Persepsi Publik Terhadap UPN "Veteran" Jawa Timur Menggunakan Metode SVM, Naïve Bayes, dan Multinomial Logistic Regression

Rahmatul Amanillah<sup>1</sup>, Sahat Renaldi. S<sup>2</sup>, Prismahardi Aji Riyantoko<sup>3</sup>

<sup>1,2,3</sup> Program Studi Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

<sup>1</sup> [20083010002@student.upnjatim.ac.id](mailto:20083010002@student.upnjatim.ac.id)

<sup>2</sup> [20083010026@student.upnjatim.ac.id](mailto:20083010026@student.upnjatim.ac.id)

<sup>3</sup> [prismahardi.aji.ds@upnjatim.ac.id](mailto:prismahardi.aji.ds@upnjatim.ac.id)

Corresponding author email: [20083010002@student.upnjatim.ac.id](mailto:20083010002@student.upnjatim.ac.id)

**Abstract:** The development of information technology has influenced various aspects of human life. Through the social media platform Twitter, users can express their opinions or thoughts regarding real-world events. These opinions can be utilized to gain deeper insights, one of which is through sentiment analysis. This study conducts a public sentiment analysis on UPN "Veteran" East Java. The researcher utilizes data obtained through web crawling using Twint. The researcher employs feature extraction methods such as Bag of Words (BoW) and Term Inverse - Document Term Frequency (TF-IDF), as well as algorithms SVM, Naive Bayes, and Multinomial Logistic Regression to build sentiment classification models. Based on the results of the analysis, which includes sentiment labeling using Python TextBlob, it is found that 12.9% of the data is classified as negative sentiment, 55.2% as neutral sentiment, and 31.9% as positive sentiment. Additionally, it is observed that there is an increasing trend in the number of tweets related to UPN "Veteran" East Java in February 2023. The evaluation results of the classification models indicate that Logistic Regression with BoW achieves the best performance, with an accuracy of 0.75, precision of 0.74, and recall of 0.60. The BoW feature extraction method appears to work better across all models tested in this study.

**Keywords:** UPNVJT, Sentiment Analysis, Feature Extraction, Classification

**Abstrak:** Perkembangan teknologi informasi mempengaruhi berbagai lini kehidupan manusia. Melalui media sosial Twitter pengguna dapat mengutarakan opini/pendapatnya terkait kejadian di dunia nyata. Opini/pendapat tersebut dapat dimanfaatkan untuk menggali *insight* lebih, salah satunya adalah melalui analisis sentimen. Penelitian ini melakukan analisis sentimen publik terhadap UPN "Veteran" Jawa Timur. Peneliti menggunakan data yang diambil dengan Teknik *crawling* dengan Twint. Peneliti menggunakan metode ekstraksi fitur *Back of Word* (BOW) dan *Term Invers - Document Term Frequency* (TF-IDF) dan algoritma SVM, *Naive Bayes*, dan *Multinomial Logistic Regression* untuk membangun model klasifikasi sentimen. Berdasarkan hasil dari proses analisis pada data yang telah dilakukan pelabelan sentimen dengan *Python TextBlob*, 12,9% dikelompokkan dalam sentimen negatif, 55,2% sentimen netral, dan 31,9% sentimen positif. Selain itu, didapatkan informasi bahwa terjadi peningkatan trend jumlah *tweet* terkait UPN "Veteran" Jawa Timur bulan Februari 2023. Hasil uji dari model klasifikasi menunjukkan bahwa *Logistic Regression* dengan BoW menghasilkan kinerja terbaik akurasi yang didapatkan sebesar 0.75, presisi sebesar 0.74 dan *recall* sebesar 0.60. Metode ekstraksi fitur BoW tampak bekerja lebih baik pada semua model yang diuji dalam penelitian ini.

**Kata kunci:** UPNVJT, Analisis Sentimen, Ekstraksi Fitur, Klasifikasi

## I. PENDAHULUAN

Pengaruh teknologi dalam kehidupan manusia sekarang sangat cepat. Salah satunya dapat dilihat dari berbagai bentuk aktivitas kehidupan yang dilakukan melalui media sosial. *Twitter* adalah salah satu media sosial yang diminati sekarang ini. *Twitter* kerap menjadi pilihan penggunanya untuk mengutarakan opini/pendapat terkait hal-hal yang terjadi di dunia nyata. Melalui fitur *trending topic* dan *keyword search*, pengguna dapat mengetahui topik yang sedang menjadi pembicaraan dan juga dapat mengetahui opini/pendapat pengguna lainnya terkait hal tersebut. Berbicara mengenai opini/pendapat pengguna media sosial terkait suatu topik tertentu merupakan hal yang menarik untuk diteliti. Data teks terkait opini/pendapat tersebut bisa dimanfaatkan untuk menggali *insight* yang lebih, salah satunya adalah melalui analisis sentimen.

Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini [1]. Analisis sentimen digunakan untuk menilai opini dan kecenderungan suatu data terhadap suatu topik yang dikelompokkan pada kelas tertentu. Media sosial seperti *Twitter*



memberikan kesempatan bagi para peneliti untuk mencari *insight* dari data *tweet* pengguna sesuai kebutuhan penelitiannya.

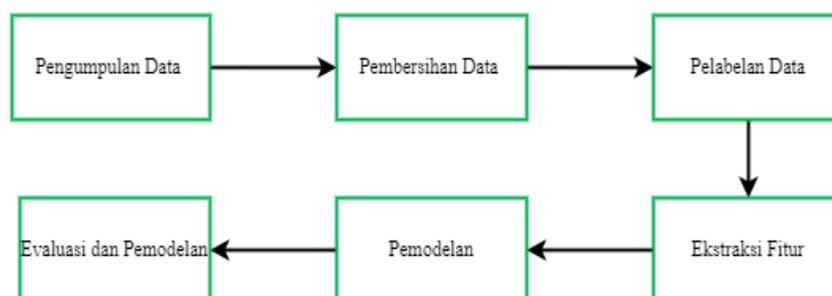
Dalam mengembangkan sistem untuk analisis sentimen, ada banyak metode yang bisa dilakukan. Penelitian yang dilakukan Arif, dkk. yaitu melakukan analisis sentimen terhadap maskapai penerbangan menggunakan metode *Naive Bayes* dengan seleksi fitur *information gain* menghasilkan rata-rata akurasi sebesar 80% [2]. Primandani, dkk. juga melakukan analisis sentimen yang berkaitan dengan adanya wacana pemindahan ibukota. Berdasarkan hasil pengujian terhadap *tweet* sentimen pemindahan ibukota dari media sosial *Twitter* sebanyak 1236 *tweet* (404 positif dan 832 negatif) menggunakan metode SVM diperoleh akurasi sebesar 96.68% [3]. Erna, dkk. dalam penelitiannya terkait analisis sentimen nasabah pada layanan perbankan didapatkan hasil bahwa model terbaik untuk mengklasifikasikan sentimen layanan pada BRI adalah *SMOTE-SVM* dengan Kernal RBF dengan akurasi mencapai 99.7% pada data *training* dan 99.2% pada data *testing*, sedangkan untuk Bank Mandiri model terbaik adalah dengan *SMOTE-Naive Bayes* yang berhasil mencapai nilai AUC sebesar 99.6% [4]. Pada tahun 2020, Imamah, dkk. melakukan penelitian sentimen analisis terhadap kasus Covid-19 menggunakan 355.384 data *tweet*. Dengan metode *Logistic Regression* dan ekstraksi fitur TF-IDF didapatkan akurasi sebesar 94.71% [5].

Pada penelitian ini, peneliti melakukan analisis sentimen publik terhadap UPN "Veteran" Jawa Timur di media sosial *Twitter*. UPN "Veteran" Jawa Timur merupakan perguruan tinggi negeri terakreditasi A di Jawa Timur. Untuk meningkatkan kualitas dan mutu universitas, UPN "Veteran" Jawa Timur perlu mengetahui opini publik terkait reputasi dan pelayanan serta pengabdian yang dilakukan selama ini. Opini publik dapat menjadi tolak ukur keberhasilan pelayanan yang ada di UPN "Veteran" Jawa Timur. Bila terdapat lebih banyak sentimen negatif, maka UPN "Veteran" Jawa Timur harus berbenah untuk meningkatkan reputasinya di masa depan sehingga tidak mempengaruhi tingkat minat masyarakat yang terhadap universitas. Peneliti mengambil data dari media sosial *Twitter* dengan *keyword search* "upnjatim" dan "upnvjt" yang terhitung sejak tanggal 1 Oktober 2022 sampai 25 Maret 2023. Total terdapat 909 baris dan 36 kolom data yang didapatkan dari proses *crawling*. Data mentah ini kemudian akan melewati proses pembersihan sebelum masuk dalam tahap pemodelan. Penelitian ini dilakukan untuk membangun model analisis sentimen untuk memberikan gambaran dan masukan terhadap UPN "Veteran" Jawa Timur supaya dapat meningkatkan kualitas dan mutu dengan menggunakan analisis sentimen sebagai tolak ukur penilaian terhadap institusi tersebut.

Berdasarkan tinjauan literatur dari penelitian sebelumnya, metode SVM, *Naive Bayes* dan *Multinomial Logistic Regression* memiliki kemampuan yang baik dalam mengklasifikasi sentimen. Oleh karena itu, peneliti mengusulkan metode SVM, *Naive Bayes* dan *Multinomial Logistic Regression* untuk diterapkan pada *tweet* terkait *keyword search* yang diambil dalam penelitian ini. Peneliti juga melakukan ekstraksi fitur dengan metode *Back of Word* (BOW) dan *Term Frequency - Invers Document Frequency* (TFIDF) serta membandingkan kinerja dari ketiga model.

## II. METODE PENELITIAN

Metodologi pada suatu penelitian merujuk pada alur atau langkah-langkah sistematis untuk mencapai tujuan penelitian sesuai dengan yang direncanakan. Metodologi penelitian dirancang dengan alur yang terstruktur dan ilmiah. Dalam penelitian ini, alur penelitian dimulai dari proses pengumpulan data yang akan digunakan, pembersihan data, pengelompokan sentimen, ekstraksi fitur untuk memperkecil dimensi data, pemodelan, sampai dengan evaluasi dan validasi model. Peneliti menggunakan *Confusion Matrix* dan *k-Fold cross validation* untuk mengukur kebaikan model.

**Gambar 1.** Alur Penelitian

### 3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan dari hasil proses *crawling* data *Twitter* menggunakan *Twint*. *Keyword* yang digunakan untuk mengambil data adalah "upn jatim" dan "upnvjt". Penelitian ini mengambil data *tweet* yang terhitung sejak tanggal 1 Oktober 2022 sampai 25 Maret 2023. Total terdapat 909 baris dan 36 kolom data mentah yang dihasilkan dari proses *crawling*.

### 3.2 Pembersihan Data

Data yang diperoleh dari proses *crawling* merupakan data mentah yang masih mengandung unsur-unsur yang tidak diperlukan. Pembersihan dilakukan untuk mempersiapkan data sebelum melakukan pemodelan. Pada data sebanyak 909 baris dan 36 kolom akan dilakukan pembersihan mulai dari penghapusan kolom yang tidak diperlukan, tautan (*URLs*), *mention*, konversi tagar (*hashtag*) menjadi kata, penghapusan simbol dan angka, serta pola-pola lainnya. Proses dilanjutkan dengan *case folding* atau mengubah setiap kata ke dalam huruf kecil, *stopword* atau pembuangan *term* yang tidak memiliki arti/tidak relevan, *stemming* atau penghapusan imbuhan suatu kata, penghapusan *emoticon*, *tokenizing* atau pemotongan *string* input berdasarkan kata yang menyusunnya, konversi slang/kata-kata informal menjadi formal, penghapusan tanda baca, baris yang kosong, dan data duplikat.

### 3.3 Pengelompokan Data

Data *tweet* dikategorikan dalam 3 label berdasarkan sentimen yaitu label -1 (negatif), 0 (netral), dan 1 (positif). Pengelompokan ini dilakukan dengan menggunakan *TextBlob*. *TextBlob* adalah sebuah pustaka di Python untuk memproses data teks [6]. Analisis sentimen menggunakan *TextBlob* hanya tersedia dalam bahasa Inggris, sehingga peneliti menerjemahkan data dari hasil pra-pemrosesan ke dalam bahasa Inggris sebelum dilakukan analisis sentimen.

### 3.4 Ekstaksi Fitur

Setelah dilakukan proses pembersihan data, data *tweet* kemudian diekstraksi untuk memperkecil dimensi data. Ekstraksi fitur merupakan proses penting pada klasifikasi teks untuk mengubah format tekstual yang tidak terstruktur menjadi terstruktur sehingga dapat diproses oleh algoritma *machine learning* untuk mengklasifikasikan ke *class* yang telah ditentukan [7]. Dalam penelitian ini, peneliti menggunakan metode ekstraksi fitur *Back of Word* (BoW) dan *Term Frequency - Invers Document Frequency* (TF-IDF). Metode ekstraksi fitur dengan *Back of Word* (BoW) menggambarkan jumlah kemunculan suatu kata di dalam dokumen. Gambar 2 merepresentasikan bagaimana *Back of Word* (BoW) bekerja mengekstraksi sebuah fitur.

Dataset	
Doc1	upn kampus bela negara
Doc2	fakultas kedokteran upn jatim
Doc3	kampus upn bersih

**Vocabulary (sorted) :**  
[“bela”, “bersih”, “fakultas”, “jatim”, “kampus”, “kedokteran”, “negara”, “upn”]

**Bag of Word (BoW) :**

	0	1	2	3	4	5	6	7
Doc1	1	0	0	0	1	0	1	1
Doc2	0	0	1	1	0	1	0	1
Doc3	0	1	0	0	1	0	0	1

**Gambar 2.** Ilustrasi Cara Kerja Metode Back of Word (BoW)

Sedangkan, metode *Term Frequency - Invers Document Frequency* (TF-IDF) adalah suatu proses untuk memberikan bobot relasi suatu kata (*term*) dengan dokumen. Proses TF-IDF mengkombinasikan dua skema perhitungan bobot, yakni frekuensi kemunculan suatu kata di dalam suatu dokumen dan *inverse* frekuensi dokumen yang memuat kata tersebut. Metode TF-IDF menentukan frekuensi relatif kata-kata dalam dokumen tertentu melalui *invers* proporsi kata di seluruh korpus dokumen. *Invers Document Frequency* adalah matrik untuk menentukan seberapa jarang suatu kata didasarkan pada suatu dokumen [8]. Berikut adalah persamaan untuk menghitung nilai IDF:

$$IDF_t = \log \log 10 \left( \frac{D}{df_t} \right) \quad (1)$$

$$W_{d,t} = tf_{d,t} * IDF_t \quad (2)$$

Mengacu pada (1) nilai D merupakan jumlah dokumen yang berisi *term* (t) dan nilai dfi merupakan jumlah kemunculan kata terhadap D. Adapun formula yang digunakan untuk menghitung bobot (w) masing-masing dokumen ditunjukkan pada persamaan (2) dengan  $W_{d,t}$  berarti bobot dokumen ke-d pada kata ke-t, dan  $tf_{d,t}$  adalah frekuensi kata.

### 3.5 Pemodelan

Dalam penelitian ini, peneliti menggunakan metode klasifikasi SVM (*Support Vector Machine*), *Naive Bayes*, dan *Multinomial Logistic Regression*. Metode pengklasifikasian dengan SVM dilakukan pada pengaturan *kernel* RBF (*Radial Basis Function*). Peneliti menggunakan jenis *kernel* ini karena berdasarkan studi literatur yang telah dilakukan, RBF *kernel* memiliki kinerja yang baik untuk kasus *multiclass classification*. Penelitian yang dilakukan oleh Harun, dkk. yang membandingkan *kernel* SVM dalam proses klasifikasi *multiclass Human Development Index* (HDI). Hasil klasifikasi *Human Development Index* (HDI) dengan menggunakan *kernel* RBF merupakan *kernel* terbaik untuk mengatasi masalah HDI [9].

Untuk melakukan proses pemodelan diperlukan hasil dari data *tweet* yang telah diekstraksi fitur. Peneliti membandingkan hasil antara dua metode ekstraksi fitur yang digunakan. Persentase jumlah data yang digunakan sebagai data *training* adalah sebesar 80%. Setelah dilakukan *training process*, maka akan dilakukan pengujian terhadap data *testing*.

### 3.6 Evaluasi dan Validasi

Evaluasi dan validasi adalah langkah untuk mengetahui performa atau kinerja dari model yang telah dibangun. Dalam kasus ini digunakan *Confusion Matrix* dan *k-Fold cross validation* untuk

mengukur kebaikan model. Setelah dilakukan pengujian pada data *testing*, maka akan dilakukan perbandingan hasil klasifikasi model dengan hasil klasifikasi yang sebelumnya sudah didefinisikan. Peneliti juga membandingkan hasilnya dengan data *tweet* yang telah disintesis dengan SMOTE (*Synthetic Minority Oversampling Technique*). Model terbaik berdasarkan hasil pada *confusion matrix* kemudian akan dilakukan validasi dengan *k-Fold cross validation* untuk bisa lebih detail melihat kemampuan model. *K-Fold cross validation* bekerja dengan membagi data ke dalam sejumlah nilai dan melakukan iterasi sebanyak *k*.

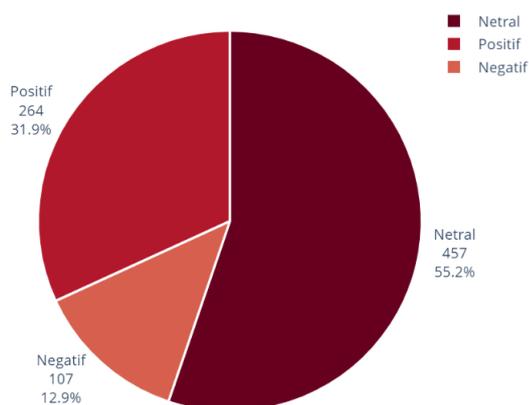
### III. HASIL DAN PEMBAHASAN

#### 4.1 Pembersihan Data

Peneliti menggunakan modul NLTK yang tersedia di *python* sebagai salah satu alat untuk melakukan pembersihan terhadap data *tweet*. Data awal yang didapatkan dari proses *crawling* terdiri dari 909 baris dan 36 kolom direduksi menjadi 828 baris dan 4 kolom. Adapun kolom yang digunakan untuk proses analisis berikutnya adalah kolom *date* (tanggal *tweet* dikirim), *tweet* (data text), *tweet\_eng* (data text dalam bahasa Inggris), dan *sentiment* (label).

#### 4.2 Sentimen

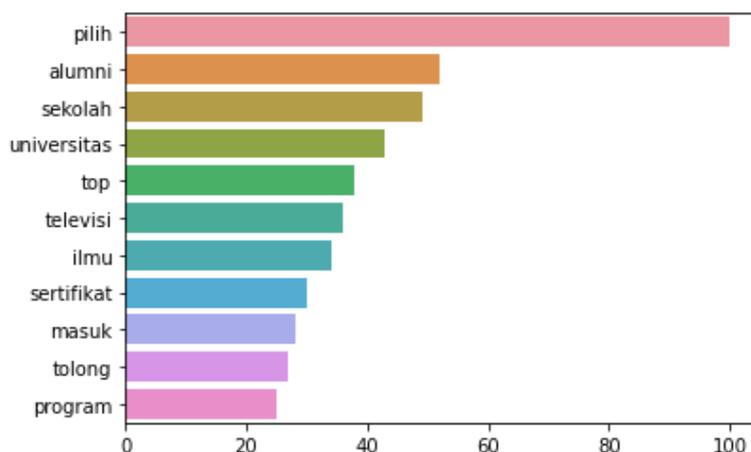
Proses pengelompokan sentimen dilakukan secara otomatis dengan bantuan *package* TextBlob pada data kolom *tweet\_eng*. Dari total 828 *tweet*, terdapat 107 *tweet* yang masuk dalam sentimen negatif, 457 *tweet* netral, dan sisanya sebesar 164 merupakan *tweet* positif.



Gambar 3. Sentimen

#### 4.3 Explorasi Data Analisis (EDA)

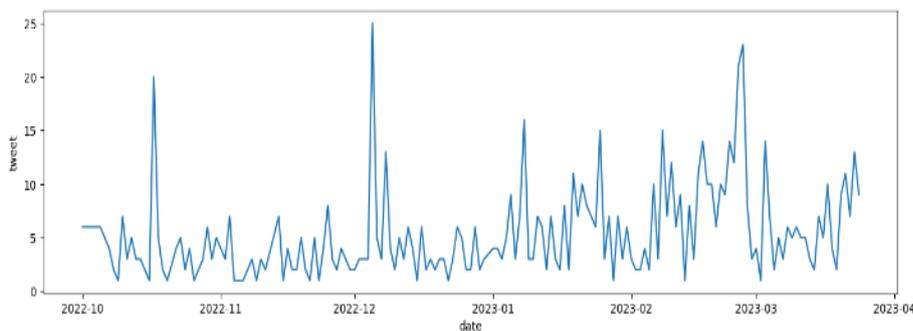
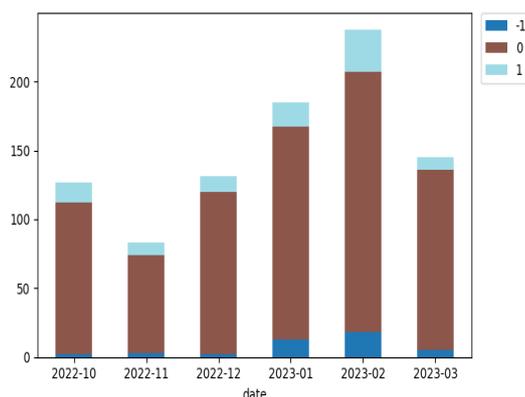
Peneliti melakukan *exploratory data analysis* untuk mendapatkan *insight* tambahan dari data yang telah dibersihkan. Peneliti menggali informasi terkait kata-kata yang sering muncul pada sentimen positif, melihat sentimen negatif dengan visualisasi *wordcloud*, *time series* analisis banyaknya *tweet* terkait UPN "Veteran" Jawa Timur dari hari ke hari, serta melihat banyaknya sentimen negatif, netral, dan positif di setiap bulannya dengan visualisasi *stacked bar chart*.

**Gambar 4.** Kata yang Sering Muncul terkait Sentimen Positif

Pada Gambar 4 dapat dilihat bahwa kata ”pilih”, ”alumni”, ”sekolah”, ”universitas”, ”top”, merupakan lima kata teratas yang paling sering muncul pada *tweet* sentimen positif. Kata-kata tersebut merepresentasikan ketertarikan terhadap UPN ”Veteran” Jawa Timur. Kata ”universitas” dan ”top” mengindikasikan tanggapan public bahwa UPN “Veteran” Jawa Timur merupakan salah satu universitas yang dikenal baik di kalangan masyarakat. Dilanjutkan dengan adanya kata “pilih”, “alumni”, dan “sekolah” memberikan gambaran bahwa UPN “Veteran” Jawa Timur merupakan salah satu pilihan bagi banyak siswa dari berbagai sekolah untuk melanjutkan pendidikan.

**Gambar 5.** Wordcloud Sentimen Positif

Dari *wordcloud* yang ditampilkan pada Gambar 5 memuat kata “ketat”, “akreditasi”, “snbp” yang sebelumnya tidak diketahui. Kata "ketat" dan “snbp” dapat mengindikasikan bahwa ada aturan atau kebijakan yang ketat dalam hal seleksi untuk bisa diterima di UPN "Veteran" Jawa. Selain itu, tampak juga kata “akreditasi” yang mungkin menjadi salah satu fokus utama yang dipertimbangkan banyak orang dalam memilih universitas.

**Gambar 6.** Jumlah Tweet per Hari**Gambar 7.** Jumlah Tweet per Bulan

Pada Gambar 6 dan Gambar 7 dapat dilihat bahwa terjadi peningkatan jumlah tweet pada periode bulan Februari 2023. Peneliti mengindikasikan peningkatan ini berkaitan dengan fenomena yang terjadi pada bulan ini yaitu penerimaan mahasiswa baru melalui jalur SNBP (Seleksi Nasional Berbasis Prestasi).

#### 4.4 Ekstraksi Fitur

Setelah sebelumnya telah melalui tahap preprocessing, *tweet* yang tersisa kemudian dilakukan ekstraksi fitur untuk memperkecil dimensi data. Proses ekstraksi dilakukan dengan memberi pembobotan pada setiap vektor kata untuk setiap kalimat. Metode ekstraksi fitur yang digunakan adalah *Back of Word* (BoW) dan *Term Frequency - Inverse Document Frequency* (TF-IDF).

**Tabel 1.** Tweet Mentah dan Tweet Bersih

Ekstraksi Fitur	Train (80%)	Test (20%)
Data awal	662	166
BoWs	8167	1881
TF-IDF	8167	1881

#### 4.5 Pemodelan

Penelitian ini mengimplementasikan algoritma SVM, *Naive Bayes*, dan *Logistic Regression* pada sentimen di media sosial *Twitter* terkait UPN "Veteran" Jawa Timur. Perbandingan data *training* dan data *testing* adalah 8 : 2. Metode pengklasifikasian dengan SVM dilakukan pada pengaturan kernel RBF (*Radial Basis Function*) dengan *random state* 42, serta pada *Logistic Regression* dengan pengaturan *multiclass multinomial* dengan *random state* yang sama. Selanjutnya dilakukan ekstraksi fitur dengan menjalankan model pada data *testing* untuk melihat nilai akurasi, presisi, dan *recall*.

Selain itu, peneliti juga membandingkan hasil pemodelan terhadap data yang disintesis dengan SMOTE. *Synthetic Minority Over sampling Technique* (SMOTE) merupakan salah satu metode yang dapat diterapkan untuk menangani ketidakseimbangan sentimen/label. Peneliti mengimplementasikan SMOTE pada data *training*.

**Tabel 2.** Akurasi, Presisi, *Recall* tanpa SMOTE

Metode	BOW			TF-IDF		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
SVM	0.66	0.44	0.44	0.67	0.68	0.46
Naive Bayes	0.72	0.64	0.56	0.62	0.46	0.39
<b>Multinomial Logistic Regression</b>	<b>0.75</b>	<b>0.74</b>	<b>0.60</b>	0.70	0.84	0.50

**Tabel 3.** Akurasi, Presisi, *Recall* dengan SMOTE

Metode	BOW			TF-IDF		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
SVM	0.59	0.49	0.48	0.70	0.79	0.52
Naive Bayes	0.55	0.52	0.53	0.58	0.51	0.53
<b>Multinomial Logistic Regression</b>	0.58	0.54	0.59	0.73	0.69	0.61

Hasil pengujian data tanpa SMOTE pada tabel 2 terlihat bahwa metode *Logistic Regression* bekerja lebih baik saat dikombinasikan dengan BoW. Nilai akurasi yang didapatkan sebesar 0.75, presisi sebesar 0.74 dan *recall* sebesar 0.60. Hasil tersebut merupakan hasil yang paling tinggi jika dibandingkan dengan metode *Support Vector Machine* (SVM) dan *Naive Bayes* yang dikombinasikan dengan ekstraksi fitur BoW maupun TF-IDF. Namun, antara metode ekstraksi fitur BoW dengan TF-IDF, metode BoW cenderung memberikan hasil yang lebih baik.

Pada data yang disintesis dengan SMOTE, hasil akurasi, presisi, *recall* tidak lebih baik dibandingkan tanpa SMOTE. Data yang diekstraksi dengan TF-IDF dan disintesis dengan SMOTE, menghasilkan skor yang lebih tinggi dibandingkan jika data hasil SMOTE tersebut diekstraksi dengan BoW di setiap model yang diuji.

#### 4.6 Evaluasi dan Validasi

Pemodelan data *tweet* terkait topik UPN “Veteran” Jawa Timur dengan metode *Logistic Regression* dan BoW menunjukkan performa paling baik dibandingkan dengan yang lain, Untuk memastikan lebih detail terkait kebaikan model *Logistic Regression* dengan kombinasi BoW ini, maka akan dilakukan validasi dengan *k-Fold cross validation*. Pada tahap validasi dengan *k-Fold cross validation*, data *training* dibagi secara random ke dalam k bagian dengan perbandingan yang sama. Proses validasi pada penelitian ini menggunakan nilai  $k = 5$ .

**Tabel 4.** Nilai Akurasi, Presisi, *Recall*, dengan Validasi k-Fold

k-Fold	Akurasi	Presisi	Recall
1	0.74	0.77	0.61
2	0.69	0.71	0.58
3	0.82	0.88	0.74
4	0.77	0.76	0.64
5	0.76	0.80	0.63



Pengujian dengan menggunakan *k-Fold cross validation*, nilai akurasi, presisi, dan *recall* tertinggi adalah pada saat nilai  $k$  adalah 3. Nilai akurasi saat  $k = 3$  adalah 0.82, presisi sebesar 0.88, dan *recall* 0.74.

#### IV. KESIMPULAN

Berdasarkan hasil analisis data, didapatkan 12.9% dikelompokkan dalam sentimen negatif, 55.2% sentimen netral, dan 31.9% sentimen positif. Melalui proses eksplorasi data analisis diketahui terdapat peningkatan jumlah *tweet* terkait UPN “Veteran” Jawa timur pada bulan Februari 2023. Hasil dari pengujian model klasifikasi sentiment dari beberapa model didapatkan metode *Logistic Regression* bekerja lebih baik saat dikombinasikan dengan BoW. Pengujian pada data tanpa SMOTE dengan *Logistic Regression* dan BoW menghasilkan akurasi sebesar 0.75, presisi sebesar 0.74 dan *recall* sebesar 0.60. Pengujian dengan menggunakan *k-Fold cross validation*, akurasi, presisi, dan *recall* tertinggi adalah pada saat nilai  $k$  adalah 3. Metode ekstraksi fitur BoW tampak bekerja lebih baik pada semua model yang diuji dalam penelitian ini.

#### UCAPAN TERIMA KASIH

#### REFERENSI

1. B. Brahimi, M. Touahria, dan A. Tari, “Improving sentiment analysis in Arabic: A combined approach,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, hal. 1242–1250, 2021, doi: 10.1016/j.jksuci.2019.07.011.
2. A. B. P. Negara, H. Muhandi, dan I. M. Putri, “Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, hal. 599, 2020, doi: 10.25126/jtiik.2020711947.
3. P. Arsi dan R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, hal. 147, 2021, doi: 10.25126/jtiik.0813944.
4. E. D. N. Sari dan I. Irhamah, “Analisis Sentimen Nasabah pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM),” *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.44565.
5. Imamah dan F. H. Rachman, “Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regresion,” *Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020*, hal. 238–242, 2020, doi: 10.1109/ITIS50118.2020.9320958.
6. D. Hazarika, G. Konwar, S. Deb, dan D. J. Bora, “Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing,” *Proc. Int. Conf. Res. Manag. Technovation 2020*, vol. 24, hal. 63–67, 2020, doi: 10.15439/2020km20.
7. M. R. Faisal dan D. T. Nugrahadi, “2517-5933-1-Pb,” vol. 8, no. 1, hal. 62–69, 2020.
8. A. S. Neogi, K. A. Garg, R. K. Mishra, dan Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, hal. 100019, 2021, doi: 10.1016/j.jjime.2021.100019.
9. H. Al Azies, D. Trishnanti, dan E. Mustikawati P.H, “Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI),” *IPTEK J. Proc. Ser.*, vol. 0, no. 6, hal. 53, 2019, doi: 10.12962/j23546026.y2019i6.6339.
10. Z. Drus dan H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” *Procedia Comput. Sci.*, vol. 161, hal. 707–714, 2019, doi:



10.1016/j.procs.2019.11.174.Arn, U. D. (2018). *Apa Itu Text Mining?*. Tersedia dari:

Www.Garudacyber.Co.Id. <https://garudacyber.co.id/artikel/1254-apa-itu-text-mining>



# One-Shot Learning Menggunakan Siamese Neural Network Untuk Pendeteksian Wajah

Muhammad Dzealfath<sup>1</sup>, Ardy Octavian Winarta<sup>2</sup>, Prizka Rismawati Arum<sup>3</sup>

<sup>1,2,3</sup> Program Studi Statistika, Universitas Muhammadiyah Semarang

<sup>1</sup> [dzealfath.aptx@gmail.com](mailto:dzealfath.aptx@gmail.com)

<sup>2</sup> [ardyocta01@gmail.com](mailto:ardyocta01@gmail.com)

<sup>3</sup> [prizka.rismawatiarum@unimus.ac.id](mailto:prizka.rismawatiarum@unimus.ac.id)

Corresponding author: [prizka.rismawatiarum@unimus.ac.id](mailto:prizka.rismawatiarum@unimus.ac.id)

**Abstract:** *One-Shot Learning using Siamese Neural Network is a method used for facial recognition, this method implements a deep learning algorithm which allows facial detection using one single image. This research aims to extract features from the images and calculate the dissimilarity of each compared images. This research uses Convolutional Neural Network (CNN) to perform feature extraction, where the images will be divided into two pairs and fed into two different CNN networks, then calculate the Euclidian Distance to measure similarity, calculate Contrastive Loss, and train Siamese Network to produce accurate results. The results of this research show that the algorithm can produce accurate results with high mean dissimilarity value for non-similar images and low mean dissimilarity value for similar images.*

**Keywords:** *Convolutional Neural Network (CNN), Face Detection, One-Shot Learning, Siamese Network*

**Abstrak:** *One-Shot Learning menggunakan Siamese Neural Network adalah metode yang digunakan untuk identifikasi wajah, metode ini mengimplementasikan algoritma deep learning yang memungkinkan pendeteksian wajah dengan menggunakan satu keping gambar. Penelitian ini bertujuan untuk mengekstrak fitur dari gambar dan menghitung ketidakmiripan dari gambar yang saling dibandingkan. Penelitian ini menggunakan Convolutional Neural Network (CNN) untuk melakukan ekstraksi fitur, dimana gambar akan dibagi menjadi dua pasangan dan dimasukkan kedalam dua jaringan CNN yang berbeda, kemudian menghitung Euclidian Distance untuk mengukur kesamaan, menghitung Contrastive Loss, dan melatih Siamese Network untuk menghasilkan hasil yang akurat. Hasil dari penelitian ini menunjukkan bahwa algoritma tersebut dapat menghasilkan hasil yang akurat dengan rata rata nilai ketidakmiripan yang tinggi untuk gambar yang tidak mirip dan rata-rata nilai ketidakmiripan yang rendah untuk gambar yang mirip.*

**Kata kunci:** *Convolutional Neural Network (CNN), Deteksi Wajah, One-Shot Learning, Siamese Network.*

## I. PENDAHULUAN

Masalah yang dihadapi dalam pengenalan wajah adalah metode konvensional yang membutuhkan data pelatihan yang besar dan model pelatihan yang memakan waktu. Selain itu, model yang dilatih menggunakan metode ini masih mengalami kesulitan dalam mengenali wajah baru atau asing. Ini karena model hanya dapat mengenali wajah yang sudah ada di data pelatihan dan tidak dapat mengenali wajah baru secara akurat. Oleh karena itu diperlukan metode baru dalam pengenalan wajah untuk mengatasi masalah tersebut [1][2]. Salah satu metode yang dapat digunakan adalah *one-shot learning* menggunakan *Siamese Neural Network*. Metode ini memanfaatkan arsitektur *Siamese Neural Network* yang memungkinkan model untuk membandingkan dua citra wajah secara langsung dan memberikan skor kemiripan di antara keduanya. Dalam pelatihan model, hanya diperlukan satu gambar untuk setiap wajah, sehingga meminimalkan jumlah data yang diperlukan. Dengan menggunakan metode ini, model dapat mengenali wajah baru dengan akurasi yang lebih tinggi dibandingkan dengan metode konvensional dalam pengenalan wajah [3].

Bentuk wajah memberikan informasi visual secara multi-dimensi, yang menyediakan berbagai informasi mengenai ciri-ciri seorang individu, seperti identitas, jenis kelamin, usia, ras, suasana hati, dan niat. Proses pengenalan wajah bisa dilakukan melalui gambar diam maupun urutan video yang berawal dari sebuah gambar diam. Wajah umumnya terdiri dari dua mata, mulut, hidung, dan berbagai fitur lain yang ditempatkan pada lokasi yang sama [4][5].

Disisi lain pembelajaran mesin juga telah berhasil digunakan untuk mencapai kinerja canggih dalam berbagai implementasi seperti pencarian web, deteksi spam, pembuatan teks, serta pengenalan ucapan dan gambar [6][7][8]. Namun, algoritma ini sering tidak bekerja dengan maksimal ketika



dipaksa untuk membuat prediksi tentang data yang hanya memiliki sedikit informasi yang dipelajari. Tujuan kami adalah ingin menggeneralisasi model ke kategori yang tidak biasa ini tanpa memerlukan pelatihan ulang ekstensif yang mungkin mahal atau tidak mungkin karena data yang terbatas atau dalam data yang pengambilannya dibatasi seperti pengambilan web. Salah satu tugas yang sangat menarik adalah klasifikasi ini memiliki batasan bahwa kita hanya boleh mengamati satu contoh dari setiap kelas sebelum membuat prediksi [9]. Teknik ini disebut *one-shot learning* dan merupakan fokus utama dari model kami yang disajikan dalam karya ini. Ini harus dibedakan dari pembelajaran *zero-shot*, di mana model tidak dapat melihat contoh apa pun dari kelas target [10].

Pengenalan wajah merupakan salah satu aplikasi penting dalam teknologi kecerdasan buatan yang dapat digunakan dalam berbagai bidang, seperti keamanan, manajemen sumber daya manusia, dan pengenalan identitas [11]. Oleh karena itu, penting untuk mengatasi permasalahan dalam pengenalan wajah, seperti membutuhkan data pelatihan yang besar dan model pelatihan yang memakan waktu serta kesulitan dalam mengenali wajah baru [12]. Dengan mengatasi masalah ini, kami dapat mengembangkan model yang lebih akurat dan efisien dalam mengenali wajah, yang dapat membantu meningkatkan keamanan dan efisiensi di berbagai aplikasi. Selain itu, pengenalan wajah juga dapat membantu pengembangan teknologi canggih dan kecerdasan buatan, seperti mobil otomatis, server robotik, dan sistem pengenalan emosi manusia. Secara keseluruhan, mengatasi masalah pengenalan wajah sangat penting untuk membuka peluang baru dalam aplikasi kecerdasan buatan dan dapat membantu meningkatkan keamanan dan efisiensi di berbagai bidang. Penelitian ini bertujuan untuk mengusulkan metode baru dalam pengenalan wajah menggunakan *one-shot learning* dengan arsitektur *Siamese Neural Network* untuk mengatasi keterbatasan metode konvensional.

## II. METODE PENELITIAN

Secara keseluruhan, penelitian tentang algoritme *one-shot learning* masih belum matang dan mendapat perhatian terbatas dari komunitas pembelajaran mesin. Namun demikian ada beberapa garis kunci penelitian yang mendahului makalah ini. Dalam [13], Li Fei-Fei et al. mengembangkan 1 kerangka kerja Bayesian variasional untuk klasifikasi gambar *one-shot learning* menggunakan premis bahwa kelas yang dipelajari sebelumnya dapat dimanfaatkan untuk membantu memperkirakan kelas yang akan datang ketika sangat sedikit data yang tersedia dari kelas tertentu. Informasi ini dapat berhasil dimasukkan ke dalam prior, yang diperbarui ketika data dari kelas baru diamati, dan kemudian digabungkan dengan kemungkinan untuk menghasilkan distribusi posterior kedalam kelas yang baru. Kemudian dalam [14], Xieyuanli, Chen et al. meneliti tentang bagaimana cara mengatasi masalah *overloop* dalam pendeteksian menggunakan lidar dengan mengembangkan model baru dari *OverlapNet* yang dimasukkan kedalam algoritma *monte-carlo* yang berdiri diatas jaringan *siamese network* dan menghasilkan model yang dapat mendeteksi secara dini *overloop* dan model dapat menggeneralisasi secara baik terhadap data lain yang diberikan. Sedangkan pada penelitian yang dilakukan oleh Hindy, Hanan et al. [15], menggunakan *siamese network* untuk mendeteksi pola baru dari serangan siber yang dimana peneliti menggunakan hanya satu model pelatihan pola serangan siber dan model tersebut dapat mendeteksi 3 pola serangan siber lain yang belum pernah dilihat sebelumnya.

Prosedur penelitian yang akan digunakan adalah mengumpulkan data latih dan uji dari subjek, kemudian menguji hipotesis dengan mengimplementasikan *One-Shot Learning* menggunakan *Siamese Neural Network* untuk pendeteksian wajah, dan akhirnya mengevaluasi akurasi dari pendeteksian wajah tersebut. Seperti yang diketahui bahwa penerapan dari *siamese network* adalah dengan memberikan gambar berpasangan guna membandingkan antara dua gambar dan dari gambar tersebut harus memiliki label seperti “sama” dan “tidak sama”. Setelah memiliki data berpasangan tersebut akan dibuat peta penyematan pada jaringan (*embedding mapping network*). Peta jaringan

yang telah tersedia akan dijadikan acuan untuk menentukan input jaringan pada *siamese network*. Hasil keluaran dari *siamese network* ini sendiri akan digunakan untuk menghitung jarak menggunakan *euclidian distance*, yang kemudian akan digunakan untuk menghitung *contrastive loss* yang diterapkan bersamaan dengan backpropagation neural network. Dan langkah terakhir adalah dengan melatih jaringan *contrastive loss* yang telah didapatkan.

### III. HASIL DAN PEMBAHASAN

Dari keseluruhan data yang ada kami akan mengambil beberapa gambar yang dimana kami mengambil sample dari gambar yang sama dan gambar yang berbeda, kemudian kami akan memberikan label pada masing-masing gambar seperti pada contoh dibawah agar dapat diolah dan dimasukkan kedalam *Convolutional Neural Network (CNN)*.

Input pair		Label
		<b>SAMA</b>
		<b>BERBEDA</b>
		<b>SAMA</b>
		<b>BERBEDA</b>

**Gambar 1.** Labeling data

Setelah kita mendapatkan data tersebut, kami akan membuat *siamese network* yang berdiri diatas CNN. Kami akan membagi gambar tersebut menjadi 2 pasangan, yang dimana masing-masing pasangan berisikan gambar dengan label yang sama dan berbeda. Dari pasangan gambar tersebut kami akan memasukan kedalam dua jaringan CNN yang berbeda dengan satu pasangan gambar ke dalam jaringan CNN A dan pasangan gambar lain kedalam jaringan CNN B. Perlu diingat bahwa peran CNN di sini hanya untuk mengekstrak fitur dan bukan untuk mengklasifikasikan. Seperti yang kita ketahui bahwa jaringan ini harus memiliki bobot dan arsitektur yang sama, jika Jaringan A kita adalah CNN dengan tiga layer maka Jaringan B kita juga harus CNN dengan tiga layer dan kita harus menggunakan set bobot yang sama untuk kedua jaringan ini. Jadi Jaringan A dan Jaringan B akan memberi kita penyematan untuk gambar masukan dan masing-masing. Untuk melakukan proses ekstrasi fitur ini yang dimana kami akan menggunakan dua lapisan konvolusi dengan aktivasi relu dan max pooling untuk melakukan proses ekstrasi fitur dan diakhiri dengan flat layer. Setelah mendapatkan hasil dari ekstrasi fitur yang dilakukan pada CNN, kami mendapatkan total 20.000 titik data dengan rincian 10.000 titik data untuk gambar yang sama dan 10.000 titik data untuk gambar yang berbeda. Dari data tersebut kemudian kita akan membagi data tersebut dengan proporsi 75:15 untuk masing masing data latih dan data uji, atau sekitar 15.000 untuk data latih dan 5.000 untuk data uji. Dan masing-masing akan dimasukkan kedalam jaringan A dan jaringan B sehingga menghasilkan vektor fitur yang baru. Setelah kami mendapatkan vektor fitur dari gambar tadi, kami akan menggunakan kedua vektor fitur tadi untuk mengukur kesamaan menggunakan *euclidian distance*

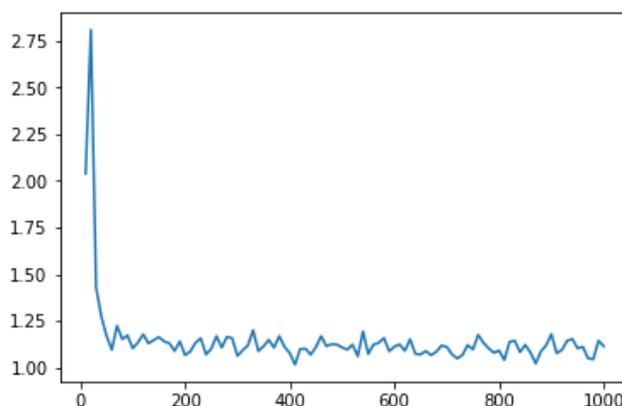
sebagai tolok ukur yang didapatkan dari jaringan A dan jaringan B yang akan digunakan sebagai masukan pada tahap *contrastive loss*, *euclidian distance* sendiri dapat dihitung dengan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Pada tahap *contrastive loss*, loss akan dihitung dan algoritma backpropagation akan dijalankan. *Contrastive loss* sendiri dapat dihitung dengan rumus:

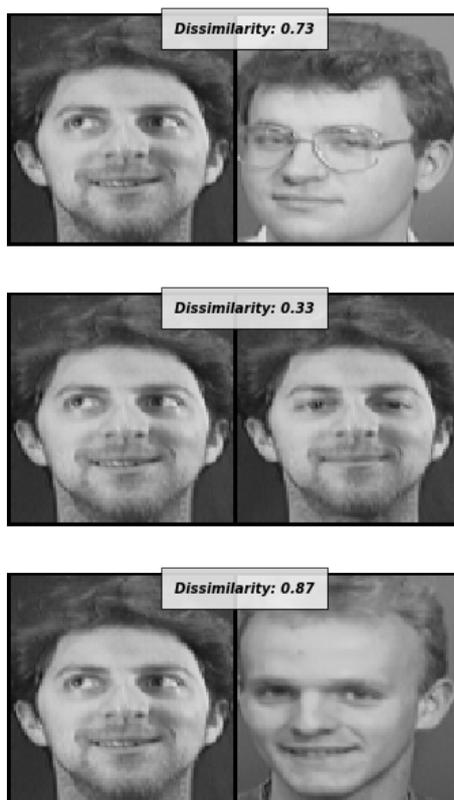
$$L_{\text{contrastive}}(x_0, x_1, y) = \frac{1}{2} y \| f(x_0) - f(x_1) \|^2 + \frac{1}{2} (1 - y) \{\max(0, m - \| f(x_0) - f(x_1) \|_2)\}^2 \quad (2)$$

Setelah kami mendapatkan hasil dari *contrastive loss*, kami melakukan proses pelatihan menggunakan model *siamese network*. Kami menggunakan *optimizer* adam dengan *learning rate* sebesar 0.0005, dengan total 100 epoch, dengan hasil dari training kami menghasilkan loss pada setiap titik data menurun seperti yang ditampilkan pada gambar dibawah



**Gambar 2.** Loss dari pelatihan

Dengan rata-rata loss yang dihasilkan adalah 1.00, kemudian kami mencoba menerapkan langsung pada data gambar yang ada yang kami pilih secara acak dan mencoba membandingkan antara kemudian kami dua menghitung gambar nilai ketidakmiripan dari dua gambar yang dibandingkan tersebut. Dari hasil tersebut kami mendapatkan hasil dengan rata-rata hasil yang akurat seperti sebagai berikut:



**Gambar 3.** Hasil akurasi dari *siamese network*

Angka yang berada diatas gambar menunjukkan nilai ketidakmiripan, yang berarti semakin tinggi angka tersebut maka gambar semakin tidak mirip dan semakin rendah angka tersebut maka gambar semakin mirip. Dari 3 gambar diatas hasil yang didapatkan sangat baik, dengan gambar pertama menunjukkan bahwa ada ketidak miripan yang tinggi, kemudian gambar kedua menunjukan bahwa gambar tersebut memiliki ketidakmiripan yang rendah atau dapat dikatakan gambar tersebut mirip, dan yang terakhir angka menunjukkan bahwa gambar memiliki nilai ketidak miripan yang tinggi. Hasil penelitian ini juga menunjukkan bahwa *siamese network* mencapai akurasi 98,5% pada dataset AT&T. Hasil ini mengungguli metode konvensional seperti Principal Component Analysis (PCA) dan Linear Discriminant Analysis (LDA), yang masing-masing mencapai akurasi 85,25% dan 92,75% [16]. Akurasi yang tinggi dari *siamese network* dalam pengenalan wajah menunjukkan keefektifan arsitektur jaringan dalam mempelajari representasi fitur yang menangkap kesamaan antar wajah.

#### IV. KESIMPULAN

Algoritma ini mampu mendeteksi perbedaan antara dua citra yang dibandingkan dengan baik walaupun dengan data pelatihan yang kecil. Kinerja model diukur dengan nilai *contrastive loss* dan mencapai akurasi yang baik. Namun, terdapat beberapa keterbatasan dalam pengembangan dan penerapan model dalam kehidupan sehari-hari, seperti model membutuhkan pembelajaran berulang setiap menerima data baru dan model membutuhkan gambar yang jelas untuk mengurangi resiko gambar tidak terdeteksi. Sebagai kesimpulan, penelitian kami menunjukkan keefektifan jaringan Siam dalam pengenalan wajah dan keunggulannya dibandingkan metode konvensional. Hasilnya menunjukkan bahwa *Siamese Network* dapat diterapkan dalam berbagai tugas pengenalan wajah, termasuk sistem keamanan, manajemen sumber daya manusia, dan pengenalan identitas. Penelitian di masa depan dapat mengeksplorasi penggunaan arsitektur jaringan yang lebih kompleks dan kumpulan data yang lebih besar untuk lebih meningkatkan kinerja *Siamese Network* dalam pengenalan wajah

**REFERENSI**

1. Z. Ding, Y. Guo, L. Zhang, and Y. Fu, “Generative One-Shot Face Recognition,” Sep. 2019, [Online]. Available: <http://arxiv.org/abs/1910.04860>
2. S. Omatu, A. Chekima, J. Dargham, and E. Moug, “Face Recognition State-of-the-art, Enablers, Challenges and Solutions: A Review,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 96–105, Apr. 2020, doi: 10.30534/ijatcse/2020/1691.22020.
3. G. R. Koch, “Siamese Neural Networks for One-Shot Image Recognition,” 2015.
4. S. Z. Li and A. K. Jain, “Handbook of Face Recognition (The Second Edition) Springer-Verlag,” 2011.
5. A. R. Dewi and I. Septadepi, “Face Recognition System Using Eigenface Method based on Facial Component Region,” 2012.
6. Y. Zhang, “Speech Recognition Using Deep Learning Algorithms,” 2013.
7. K. Kowsari *et al.*, “Text Classification Algorithms: A Survey,” *Information (Switzerland)*, vol. 10, Apr. 2019, doi: 10.3390/info10040150.
8. S. Annareddy and S. Tammina, “A Comparative Study of Deep Learning Methods for Spam Detection,” in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 66–72. doi: 10.1109/I-SMAC47947.2019.9032627.
9. B. M. Lake, R. Salakhutdinov, J. Gross, and J. B. Tenenbaum, “One shot learning of simple visual concepts,” 2011. [Online]. Available: <http://web.mit.edu/brenden/www/charactervideos.html>.
10. W. Wang *et al.*, “Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions,” Feb. 2021, [Online]. Available: <http://arxiv.org/abs/2102.12122>
11. C. Mingsung and L. Cai, *Research on the Application of Face Recognition System*. 2020. doi: 10.2991/assehr.k.200727.057.
12. R. Jafri and H. Arabnia, “A Survey of Face Recognition Techniques,” *JIPS*, vol. 5, pp. 41–68, Jun. 2009, doi: 10.3745/JIPS.2009.5.2.041.
13. L. Fei-Fei, R. Fergus, and P. Perona, “A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories,” 2006. [Online]. Available: [www.google.com](http://www.google.com)
14. X. Chen, T. Labe, A. Milioto, T. Röhling, J. Behley, and C. Stachniss, “OverlapNet: a siamese network for computing LiDAR scan similarity with applications to loop closing and localization,” *Auton Robots*, vol. 46, no. 1, pp. 61–81, Jan. 2022, doi: 10.1007/s10514-021-09999-0.
15. H. Hindy *et al.*, “Leveraging siamese networks for one-shot intrusion detection model,” *J Intell Inf Syst*, 2022, doi: 10.1007/s10844-022-00747-z.
16. A. Sahoolzadeh, B. Heidari, and C. Dehghani, “A new face recognition method using PCA, LDA and neural network,” *International Journal of Computer Science and Engineering*, vol. 2, Jan. 2008.



# Model Prediksi Kepadatan Lalu Lintas: Perbandingan Antara Algoritma Random Forest dan XGBoost

Fitri Indah Sari<sup>1</sup>, Ellexia Leonie Gunawan<sup>2</sup>, Chelsea Ayu Adhigiadany<sup>3</sup>, Angela Lisanthoni<sup>4</sup>

<sup>1,2,3,4</sup> Sains Data, UPN "Veteran" Jawa Timur

<sup>1</sup> [21083010025@student.upnjatim.ac.id](mailto:21083010025@student.upnjatim.ac.id)

<sup>2</sup> [21083010027@student.upnjatim.ac.id](mailto:21083010027@student.upnjatim.ac.id)

<sup>3</sup> [21083010028@student.upnjatim.ac.id](mailto:21083010028@student.upnjatim.ac.id)

<sup>4</sup> [21083010032@student.upnjatim.ac.id](mailto:21083010032@student.upnjatim.ac.id)

Corresponding author email: [21083010032@student.upnjatim.ac.id](mailto:21083010032@student.upnjatim.ac.id)

**Abstract:** Population in general will continue to grow so traffic congestion is a phenomenon that needs to be overcome. Therefore, it is necessary to predict accurate traffic density to develop the right strategy to reduce the level of congestion. So, the aim of this study is making a comparison between the Random Forest and XGBoost algorithms which are one of the forecasting models forming algorithms. The results will be compared based on three types of evaluation parameters including the coefficient of determination ( $R^2$ ), Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE). Based on the research, it proves that there is no significant difference in predicting traffic volume between the Random Forest and XGBoost algorithms. The two algorithms being compared have an accuracy of approximately 95%, however, XGBoost has the advantage that the time required for prediction is 532% faster than Random Forest. By considering accuracy and efficiency, the XGBoost algorithm is the right choice in building traffic forecasting models.

**Keywords:** Regression Supervised Learning, Traffic Forecasting, Random Forest, XGBoost

**Abstrak:** Populasi manusia secara umum akan terus bertambah sehingga kepadatan lalu lintas adalah fenomena yang perlu diatasi. Oleh sebab itu, diperlukan prediksi kepadatan lalu lintas yang akurat untuk menyusun strategi yang tepat guna mengurangi tingkat kemacetan. Sehingga tujuan dari penelitian ini adalah membandingkan algoritma *Random Forest* dan *XGBoost* yang merupakan salah satu algoritma pembentuk model prediksi. Hasil akan dibandingkan berdasarkan tiga jenis evaluasi parameter diantaranya koefisien determinasi ( $R^2$ ), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE). Berdasarkan penelitian yang dilakukan, menunjukkan bahwa tidak ada perbedaan signifikan dalam melakukan prediksi volume lalu lintas antara algoritma *Random Forest* dan *XGBoost*. Kedua algoritma yang dibandingkan memiliki hasil akurasi kurang lebih 95% namun, *XGBoost* memiliki kelebihan yakni waktu yang dibutuhkan dalam prediksi 532% lebih cepat dibanding *Random Forest*. Dengan mempertimbangkan akurasi dan efisiensi, algoritma *XGBoost* adalah pilihan yang tepat dalam membangun model prediksi kepadatan lalu lintas.

**Kata kunci:** Regresi Supervised Learning, Prediksi Kepadatan lalu lintas, Random Forest, XGBoost

## I. PENDAHULUAN

Kepadatan lalu lintas sudah menjadi permasalahan yang semakin umum di masa sekarang. Berdasarkan *The 2022 Revision of World Population Prospects* tercatat sebesar 8 miliar jiwa pada 15 November 2022 [1]. Dari data *2023 World Population by Country (Live)* tercatat bahwa populasi Indonesia pada 2022 mencapai 275 juta jiwa dengan luas daratan 1,9 juta  $km^2$ , itu berarti populasi masyarakat Indonesia mencapai 135 juta jiwa/ $km^2$  [2].

Kepadatan lalu lintas tentunya didasari oleh beberapa aspek salah satunya adalah cuaca [3]. Dengan berbagai permasalahan tersebut, adanya prediksi tentang kepadatan lalu lintas menjadi sangat penting. Dengan adanya prediksi kepadatan lalu lintas diharapkan dapat membantu dalam menghindari bahkan mengatasi peristiwa kemacetan. Prediksi yang akurat akan memberikan hasil yang bagus dan juga bermanfaat bagi pengguna jalan. Cara untuk memprediksi dengan nilai akurasi yang bagus tentunya menggunakan teknik machine learning seperti regresi linier, *random forest*, dan *XGBoost* [4]. Pada penelitian sebelumnya, untuk memprediksi kepadatan lalu lintas menggunakan metode *random forest algorithm* dibuktikan bahwa *random forest algorithm* merupakan algoritma terbaik untuk memprediksi lalu lintas berdasarkan data [5]. Kemudian metode-metode tersebut akan dievaluasi kinerjanya menggunakan matrik evaluasi seperti *Root Mean Square Error* (RMSE), koefisien determinasi ( $R^2$ ), dan juga *Mean Absolute Error* (MAE). Dari hasil evaluasi ini nantinya

akan membantu menentukan metode manakah yang paling efektif dalam memprediksi kepadatan lalu lintas.

Tujuan daripada pembuatan artikel ini yaitu untuk mencari model terbaik dalam melakukan *traffic forecasting* (prediksi kepadatan lalu lintas). Dalam perencanaan transportasi, mengetahui informasi arus lalu lintas sangat krusial. Oleh sebab itu, dibutuhkan model yang mampu memproyeksikan dengan tepat dan akurat volume lalu lintas di masa mendatang. Dalam artikel ini, akan dibandingkan dua model yaitu *random forest* dan *XGBoost* dalam rangka menemukan model optimal yang dapat digunakan dalam *traffic forecasting*. Kedua model ini telah banyak digunakan dalam berbagai aplikasi *machine learning* dan memiliki kelebihan masing-masing. Dengan membandingkan kedua model ini, diharapkan dapat ditemukan model terbaik yang dapat digunakan dalam perencanaan transportasi untuk memprediksi volume lalu lintas di masa depan.

## II. KAJIAN PUSTAKA

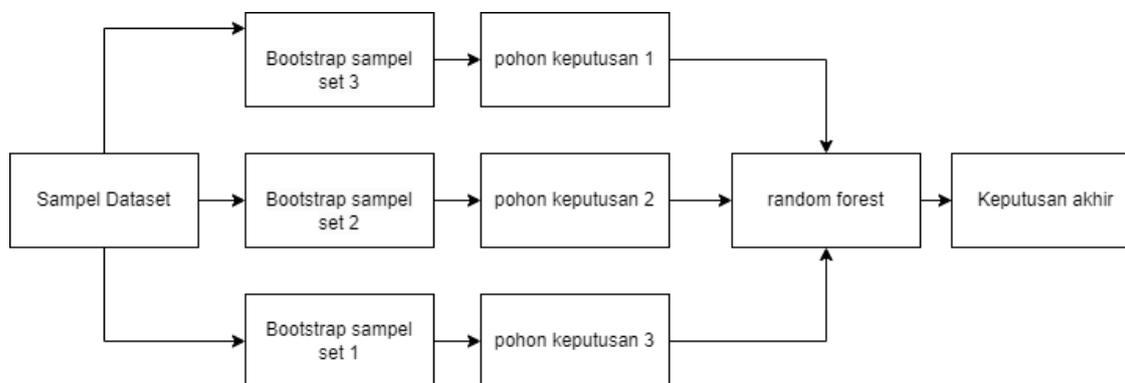
### 2.1. Traffic Forecasting

*Traffic forecasting* adalah salah satu jenis model *supervised machine learning* yang memiliki tujuan utama memprediksi kepadatan lalu lintas pada waktu dekat di masa depan berdasarkan data kepadatan lalu lintas sekarang dan data di masa lampau sehingga model akan mempelajari pola data input [6]. Prediksi *traffic forecasting* yang akurat sangat membantu pengambilan keputusan di masa depan agar dapat merencanakan strategi dengan tepat dan lebih efisien dalam mengurangi kepadatan lalu lintas, meningkatkan kualitas udara, serta mobilitas [7].

Traffic forecasting akan memprediksi volume kemacetan dimana volume kemacetan yang dimaksud adalah jumlah kendaraan yang melintas pada jalan yang diobservasi. Volume kemacetan dipengaruhi oleh beberapa faktor diantaranya cuaca dan hari libur. ketika hari libur dan suhu meningkat maka mobilitas akan ikut meningkat cukup signifikan; ketika curah hujan dan jumlah tingkat salju meningkat, maka mobilitas menurun cukup signifikan; serta, jika prosopir langit yang tertutup oleh awan menaik, maka mobilitas meningkat meskipun tidak secara signifikan [8]. Berdasarkan penelitian [9], kepadatan lalu lintas dipengaruhi oleh hari libur, suhu, curah hujan, jumlah tingkat salju, dan proporsi langit yang tertutup oleh awan sebesar 93.3%.

### 2.2. Random Forest Algorithm

Algoritma *Random Forest* dapat digunakan untuk melakukan regresi maupun klasifikasi tergantung tujuan analisis yang diinginkan. Jika digunakan untuk melakukan regresi, maka hasil yang ditampilkan adalah rata - rata dari pohon yang berbeda [10].

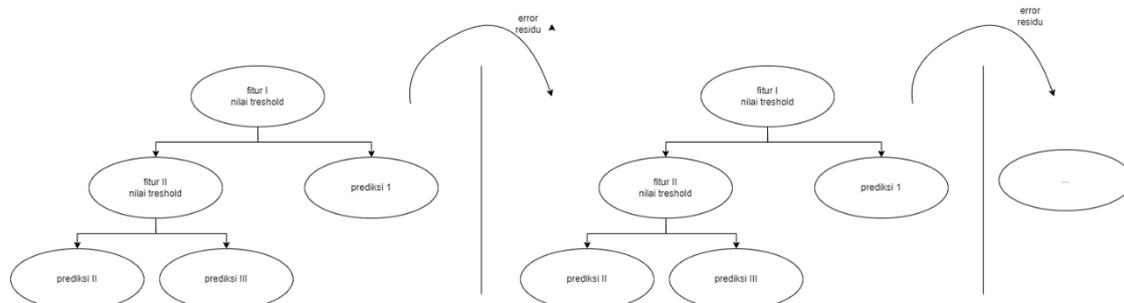


**Gambar 1.** Proses algoritma *Random Forest*

Gambar 1 menunjukkan ilustrasi algoritma *Random Forest* berjalan. Data akan diproses melalui metode bootstrap untuk menemukan k sampel data yang berbeda. Di setiap *bootstrap*, terdapat pohon keputusan yang dilakukan secara berulang untuk memecah data menjadi dua kelompok dengan

kriteria tertentu. Setiap pohon akan memberikan keputusan hingga didapatkan keputusan yang berbeda. Kemudian, diambil satu keputusan sebagai keputusan akhir berdasarkan suara voting terbanyak [11].

### 2.3. XGBoost Algorithm



**Gambar 2.** Proses algoritma XGBoost

*eXtreme Gradient Boosting* adalah kepanjangan dari XGBoost. Algoritma ini mengimplementasikan pengembangan dari pohon keputusan dan menerapkan teknik ansambel artinya model akan terus diperbarui untuk memperbaiki kesalahan pada model sebelumnya. Algoritma ini dibuat dalam rangka meningkatkan efisien waktu pemroses termasuk automasi menangani data hilang, melakukan pohon keputusan secara paralel, serta dapat melakukan training data terus-menerus untuk meningkatkan hasil akurasi [12]. Ilustrasi proses algoritma XGBoost ditampilkan pada gambar 2 [13]. Algoritma ini dirumuskan sebagai berikut:

$$obj(\theta) = \sum_i L(\hat{y}_i, y_i) + \sum_k L\Omega(f_k), \quad f_k \in F \quad (1)$$

L menyatakan perbedaan standar deviasi antara nilai prediksi  $\hat{y}_i$  dan nilai asli  $y_i$ .  $\Omega$  menyatakan fungsi regularisasi kompleksitas model untuk menghindari overfitting.  $f$  menyatakan fungsi dalam ruang fungsional  $F$  dan  $F$  menyatakan himpunan semua pohon yang dibuat [14].

### 2.4. Evaluasi Parameter

Untuk melakukan evaluasi terhadap model, terdapat tiga parameter yang akan dibandingkan yaitu koefisien determinasi ( $R^2$ ) yang digunakan untuk menafsir kelayakan model dengan rentang nilai  $0 \leq R^2 \leq 1$ . *Root Mean Square Error* (RMSE) yaitu selisih nilai prediksi dengan nilai sebenarnya dan akar dari *Mean Square Error* (MSE), serta *Mean Absolute Error* (MAE) yaitu rata-rata kesalahan mutlak antara nilai prediksi dengan nilai yang sebenarnya. Ketiga parameter dirumuskan sebagai berikut [14] [15]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f(\hat{y}_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

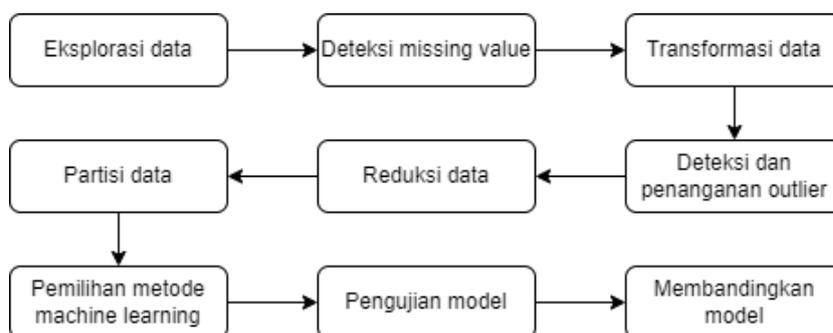
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

## III. METODE PENELITIAN

Dalam studi ini, dataset yang digunakan peneliti adalah Metro Interstate Traffic Volume, yang disediakan oleh repositori data UCI. Dataset ini terdiri 48204 baris dan 9 kolom, dimana data ini berisi informasi volume lalu lintas Interstate 94 Westbound untuk MN DoT ATR station 301, berkisar antara Minneapolis dan St Paul, MN dari tahun 2012-2018 dengan berbagai kondisi, seperti cuaca, hari libur, dan suhu. Sehingga pada penelitian ini, peneliti ingin memprediksi kepadatan lalu lintas berdasarkan beberapa kondisi informasi sebagai berikut:

- holiday yaitu hari libur nasional Amerika Serikat kategorikal ditambah hari libur regional, Minnesota State Fair
- temp yaitu numerik rata-rata suhu dalam kelvin
- rain\_1h yaitu numeric jumlah dalam mm hujan yang terjadi dalam satu jam
- snow\_1h yaitu jumlah numerik dalam mm dari salju yang terjadi dalam satu jam
- cloud\_all yaitu persentase numerik tutupan awan
- weather\_main yaitu kategori deskripsi tekstual singkat tentang cuaca saat ini
- weather\_description yaitu kategorikal deskripsi tekstual yang lebih panjang tentang cuaca saat ini
- date\_time yaitu tanggal, waktu, dan jam dari data yang dikumpulkan dalam waktu CST lokal
- traffic\_volume yaitu numeric hourly I-94 ATR 301 melaporkan volume lalu lintas arah barat



**Gambar 3.** Diagram alur pengolahan data

Gambar 3 menunjukkan alur metode dalam pengolahan data. Tahap *preprocessing* data merupakan langkah penting dalam penelitian yang melibatkan analisis data. Tahapan *preprocessing* data yang dilakukan oleh peneliti meliputi eksplorasi data, penanganan *missing value*, transformasi data, deteksi dan penanganan outlier, reduksi data, dan partisi data. Sedangkan algoritma yang digunakan peneliti adalah *Random Forest* dan *XGBoost*. *Random Forest* dan *XGBoost* merupakan algoritma machine learning yang populer dan sering digunakan karena kemampuannya dalam mengatasi overfitting, meningkatkan akurasi prediksi serta dapat menangani masalah regresi. Basis algoritma *Random Forest* dan *XGBoost* menggunakan *decision tree*, oleh sebab itu peneliti melakukan perbandingan keakuratan dan efisiensi terhadap dataset yang digunakan.

#### IV. HASIL DAN PEMBAHASAN

Penelitian ini melakukan analisa dengan dua algoritma, yaitu *Random Forest* dan *XGBoost*. Data dipisahkan menjadi dua bagian yaitu data training sebanyak 70% dan data testing sebanyak 30% dengan alat bantu Machine Learning bahasa pemrograman Python.

**Tabel 1.** Perbedaan volume lalu lintas data asli dan prediksi menggunakan *random forest* terhadap 25 data teratas

Indeks	Data Asli	Prediksi
0	6685	5776.19
1	1336	1195.81
2	6172	6092.79
3	2721	2766.86
4	973	936.28
5	1939	1878.79
6	393	343.77
7	5370	4568.07
8	5864	5401.22
9	6218	5894.28
Indeks	Data Asli	Prediksi

10	4058	3684.97
11	3486	3453.21
12	3047	3169.11
13	3619	3473.41
14	374	302.30
15	3113	3125.47
16	2620	3131.39
17	3538	3273.73
18	6599	6139.82
19	5875	5723.51
20	621	701.40
21	3660	3350.04
22	5883	5602.97
23	6516	6372.90
24	429	410.83

Tabel 1 merupakan perbedaan volume lalu lintas dari data asli dengan data prediksi pada 25 data teratas. Gambar 4 merupakan visualisasi perbedaan volume lalu lintas dari data asli dan data prediksi yang terdapat pada Tabel 1. Sumbu x pada grafik di Gambar 4 menunjukkan indeks dari data, sedangkan sumbu y menunjukkan volume lalu lintas. Kurva biru menunjukkan volume lalu lintas dari data asli sedangkan kurva oranye menunjukkan volume lalu lintas dari data prediksi menggunakan algoritma *Random Forest*.



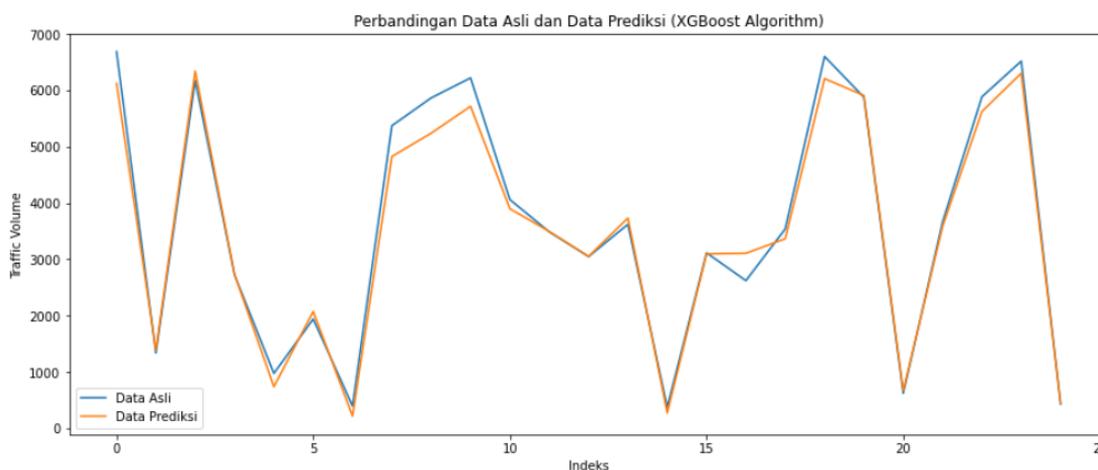
**Gambar 4.** Visualisasi grafik perbedaan volume lalu lintas data asli dan data prediksi dengan *random forest* terhadap 25 data teratas

**Tabel 2.** Perbedaan volume lalu lintas data asli dan prediksi menggunakan *XGBoost* terhadap 25 data teratas

Indeks	Data Asli	Prediksi
0	6685	6121.34
1	1336	1384.57
2	6172	6339.75
3	2721	2731.25
4	973	735.72
5	1939	2073.62
6	393	215.20
7	5370	4823.57
8	5864	5237.60
9	6218	5714.29
10	4058	3896.49
11	3486	3493.68

Indeks	Data Asli	Prediksi
12	3047	3050.46
13	3619	3730.57
14	374	271.96
15	3113	3096.56
16	2620	3105.85
17	3538	3362.94
18	6599	6206.52
19	5875	5906.25
20	621	666.80
21	3660	3580.76
22	5883	5622.24
23	6516	6300.70
24	429	446.37

Tabel 2 merupakan perbedaan volume lalu lintas dari data asli dengan data prediksi pada 25 data teratas. Gambar 5 merupakan visualisasi perbedaan volume lalu lintas dari data asli dan data prediksi yang terdapat pada Tabel 2. Sumbu x pada grafik di Gambar 5 menunjukkan indeks dari data, sedangkan sumbu y menunjukkan volume lalu lintas. Kurva biru menunjukkan volume lalu lintas dari data asli sedangkan kurva oranye menunjukkan volume lalu lintas dari data prediksi menggunakan algoritma *XGBoost*. Dari Gambar 2 dan 5 dapat disimpulkan bahwa hasil prediksi *XGBoost* lebih baik daripada *Random Forest* karena perbedaan antara data asli dan data prediksi yang lebih kecil sesuai dengan nilai RMSE.



**Gambar 5.** Visualisasi grafik perbedaan volume lalu lintas data asli dan data prediksi dengan *XGBoost* terhadap 25 data teratas

Dalam penelitian ini, evaluasi parameter dilakukan dengan menggunakan koefisien determinasi ( $R^2$ ), *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan waktu yang dibutuhkan saat pemrosesan setiap algoritma. Hasil pengujian evaluasi parameter antara dua algoritma yang digunakan, yaitu *Random Forest* dan *XGBoost* ditampilkan pada tabel 3

**Tabel 3.** Perbandingan evaluasi parameter dan waktu pemrosesan antara *Random Forest* dan *XGBoost*

Algoritma	R2	RMSE	MAE	Waktu pemrosesan (s)
Random Forest	95.53	420.09	226.87	19.91
XGBoost	95.92	401.02	236.38	3.15

Berdasarkan tabel 3, didapat nilai akurasi algoritma *Random Forest* sebesar 95.53% sedangkan untuk algoritma *XGBoost* sebesar 95.92%. Nilai RMSE algoritma *Random Forest* sebesar 420.09, sedangkan *XGBoost* sebesar 401.02 dan nilai MAE *Random Forest* sebesar 226.87, sedangkan



Nilai  $R^2$  *XGBoost* yang lebih tinggi, menunjukkan bahwa kelayakan model lebih bagus. Nilai RMSE *XGBoost* yang lebih rendah menunjukkan bahwa perbedaan nilai prediksi dengan nilai sebenarnya lebih kecil sehingga lebih bagus. Nilai MAE *Random Forest* yang lebih rendah menunjukkan bahwa rata - rata kesalahan mutlak antara nilai prediksi dengan nilai yang sebenarnya lebih rendah. Hasil  $R^2$ , RMSE, dan MAE dari kedua algoritma ini tidak berbeda secara signifikan. Namun, untuk perbedaan nilai waktu pemrosesan antara *Random Forest* dan *XGBoost* memiliki perbedaan yang signifikan. *XGBoost* memiliki waktu pemrosesan yang sangat cepat, yaitu 532% lebih cepat dibanding *Random Forest*.

## V. KESIMPULAN

Berdasarkan penelitian yang dilakukan dengan menggunakan algoritma *Random Forest* dan *XGBoost* tidak menunjukkan perbedaan yang signifikan mengenai prediksi data volume lalu lintas. Nilai  $R^2$  pada algoritma *XGBoost* yang lebih tinggi, menunjukkan bahwa kelayakan model lebih bagus. Nilai RMSE *XGBoost* yang lebih rendah menunjukkan bahwa perbedaan nilai prediksi dengan nilai sebenarnya lebih kecil sehingga lebih bagus. Sedangkan nilai MAE *Random Forest* yang lebih rendah menunjukkan bahwa rata - rata kesalahan mutlak antara nilai prediksi dengan nilai yang sebenarnya lebih rendah. Hasil  $R^2$ , RMSE, dan MAE dari kedua algoritma ini tidak berbeda secara signifikan. Namun, untuk perbedaan nilai waktu pemrosesan antara *Random Forest* dan *XGBoost* memiliki perbedaan yang sangat signifikan. *XGBoost* memiliki waktu pemrosesan yang sangat cepat, yaitu 532% lebih cepat dibanding *Random Forest*. Dari hasil perbandingan, algoritma *XGBoost* adalah pilihan yang tepat dalam membangun model prediksi kepadatan lalu lintas karena memiliki akurasi tinggi dan efisiensi dalam waktu.

## UCAPAN TERIMA KASIH

## REFERENSI

1. U. nations, "Data Portal Population Division," population.un.org, [Online]. Available: <https://population.un.org/dataportal/home>. [Accessed 1 April 2023].
2. w. p. review, "2023 World Population by Country (Live)," worldpopulationreview.com, [Online]. Available: <https://worldpopulationreview.com/>. [Accessed 1 April 2023].
3. M. Hudzaifah and A. A. Rismayadi, "PERAMALAN ARUS LALU LINTAS BERDASARKAN WAKTU TEMPUH DAN CUACA MENGGUNAKAN METODE TIME SERIES DECOMPOSITION," *JURNAL RESPONSIF*, vol. III, no. 2, 2021.
4. B. ALAOUI, D. BARI and Y. GHABBAR, "Surface Weather Parameters Forecasting Using Analog Ensemble the Main Airports of Morocco," *Journal of Meteorological Research*, vol. XXXVI, 2022.
5. A. C. N and D. H. V. Kumaraswamy, "Traffic Prediction using Random Forest Machine Learning Algorithms," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. XI, no. 4, pp. 55 - 59, 2022.
6. J. S. ANGARITA-ZAPATA, A. D. MASEGOSA and I. TRIGUERO, "A Taxonomy of Traffic Forecasting Regression Problems From a Supervised Learning Perspective," *IEEE Access*, vol. VII, 2019.
7. J. Liu, N. Wu, Y. Qiao and Z. Li, "A scientometric review of research on traffic forecasting in transportation," *IET Intell Transp Syst*, vol. XV, no. 1, pp. 1-16, 2021.
8. P. Lin, Y. He, M. Pei and R. Yang, "Data-driven spatial-temporal analysis of highway traffic volume considering weather and festival impacts," *Travel Behaviour and Society*, vol. XXIX, pp. 95 - 112, 2022.



9. H. J. Nascimento, "Statistical estimation of traffic volume for the Minneapolis-St Paul Metropolitan area," *International Journal of Advanced Engineering Research and Science (IJAERS)*, vol. IX, no. 11, pp. 392 - 397, 2022.
10. I. Alam, D. M. Farid and R. J. F. Rossetti, "The Prediction of Traffic Flow with Regression Analysis," in *Emerging Technologies in Data Mining and Information Security*, 2018.
11. Y. Liu and H. Wu, "Prediction of Road Traffic Congestion Based on Random Forest," in *10th International Symposium on Computational Intelligence and Design*, Hangzhou, China, 2017.
12. J. Brownlee, *XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn, Machine Learning Mastery*, 2016.
13. A. I. A. Osman, A. N. Ahmed, M. F. Chow, Y. F. Huang and A. El-Shafie, " Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Engineering Journal*, vol. XII, no. 2, pp. 1545 - 1556, 2021.
14. J. Luo, Z. Zhang, Y. Fu and F. Rao, "Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms," *Results in Physics*, vol. XXVII, 2021.
15. A. E. Putra and A. Juarna, "Prediksi Produksi Daging Sapi Nasional dengan Metode Regresi Linier dan Regresi Polinomial," *Jurnal Ilmiah KOMPUTASI*, vol. XX, no. 2, pp. 209 - 215, 2021.



# *Adaptive Synthetic Support Vector Machine Multiclass* untuk mengklasifikasikan *Imbalance data* pada Sentimen kenaikan Bahan Bakar Minyak

Ismatullah<sup>1</sup>, Fatkhurokhman Fauzi<sup>2</sup>, Indah Manfaati Nur<sup>3</sup>,

<sup>1,2,3</sup>Program Studi Statistika, Universitas Muhammadiyah Semarang

Corresponding author email: [fatkhurokhmanf@unimus.ac.id](mailto:fatkhurokhmanf@unimus.ac.id)

**Abstract:** The phenomenon of increasing fuel oil (BBM) has become a trending topic for people in Indonesia in September 2022. Indonesian people from various provinces have chosen their opinions regarding this phenomenon on social media, one of which is Twitter. Sentiment analysis is used to describe a person's opinion on social media about a phenomenon. In this study, we will look at public sentiment regarding the increase in fuel prices which is labeled into three categories, namely positive, neutral and negative. This study applies adaptive synthetics to overcome data imbalances caused by negative sentiment. The data used in this research is public opinion related to the increase in fuel prices in every province in Indonesia. Each province is limited to 100 opinions. The classification method applied to this research is the multiclass Support Vector Machine (SVM). The results obtained are that people in all provinces in Indonesia have a negative opinion regarding the increase in fuel prices. The results of the multiclass SVM classification show an average accuracy of 87.94%, with the highest accuracy of 95%.

**Keywords:** Adaptive Synthetic, BBM, Sentiment, SVM

**Abstrak:** Fenomena kenaikan Bahan Bakar Minyak (BBM) menjadi *trending* topik masyarakat di Indonesia pada bulan September 2022. Masyarakat Indonesia dari berbagai provinsi menyuarakan pendapatnya terkait fenomena tersebut di media sosial salah satunya *twitter*. Analisis sentimen digunakan untuk mengetahui gambaran pendapat seseorang di media sosial terhadap suatu fenomena. Pada penelitian ini akan melihat sentimen masyarakat terkait kenaikan BBM yang dilabelkan kedalam tiga kategori yaitu positif, netral, dan negatif. Penelitian ini menerapkan *adaptive synthetic* untuk menangani *imbalance data* yang disebabkan oleh sentimen negatif. Data yang digunakan pada penelitian ini adalah pendapat masyarakat terkait kenaikan BBM di setiap provinsi di Indonesia. Setiap provinsi dibatasi 100 pendapat. Metode klasifikasi yang diterapkan pada penelitian adalah Support Vector Machine (SVM) *multiclass*. Hasil yang didapat adalah masyarakat diseluruh provinsi di Indonesia berpendapat negatif terkait kenaikan BBM. Hasil klasifikasi SVM *multiclass* menunjukkan rata-rata akurasi sebesar 87.94%, dengan akurasi tertinggi sebesar 95%.

**Kata kunci:** Adaptive Synthetic, BBM, Sentimen, SVM

## I. PENDAHULUAN

Kenaikan harga Bahan Bakar Minyak (BBM) mengundang pro dan kontra di masyarakat Indonesia. Fenomena ini mengundang masyarakat berkomentar di media sosial tentang pandangannya terkait kenaikan BBM. Sebagai negara demokrasi Indonesia menjunjung tinggi kebebasan berpendapat. Pendapat tersebut dapat dijadikan sebagai acuan dalam evaluasi kebijakan pemerintah dalam kasus ini adalah kenaikan BBM.

*Twitter* menjadi salah satu *platform* media sosial yang digunakan masyarakat untuk menyuarakan pendapatnya [1]. Pendapat masyarakat yang terdapat di *twitter* dapat dianalisis, metode tersebut dinamakan analisis sentimen. Tujuan dari analisis sentimen yaitu mengklasifikasi polaritas teks pada dokumen, kalimat, dan fitur serta menentukan apakah komentar/pendapat tersebut bersentimen positif, netral, atau negatif[2]. Analisis sentimen tentang kenaikan harga BBM perlu dilakukan untuk mengetahui respon pendapat/pandangan masyarakat khususnya pengguna *twitter* mengenai kenaikan harga BBM yang ditetapkan oleh pemerintah pada tanggal 3 September 2022. Sentimen dari masyarakat tentang kenaikan harga BBM merupakan faktor yang penting bagi pemerintah untuk menetapkan suatu kebijakan atau mengevaluasi suatu kebijakan.

Klasifikasi teks atau kategorisasi teks merupakan proses yang secara otomatis menempatkan dokumen teks ke dalam suatu kategori berdasarkan isi dari teks tersebut. Klasifikasi teks berbasis statistik, klasifikasi teks berbasis koneksi, dan klasifikasi teks berbasis aturan adalah tiga kategori utama pendekatan klasifikasi teks telah digunakan dalam penelitian. Teknik klasifikasi berbasis statistik berkinerja lebih baik daripada yang lain [3].

Salah satu metode klasifikasi berbasis statistik yang dapat digunakan untuk mengklasifikasi dokumen yaitu *Support Vector Machine* (SVM). Penelitian SVM *multiclass* pernah dilakukan oleh Irmanda dan Astriratma [4] dengan mengklasifikasikan jenis pantun. Hasil akurasi yang didapat oleh SVM *multiclass* sebesar 81.91%. Hal ini menunjukkan bahwa metode SVM *multiclass* memiliki performansi yang sangat baik. Penelitian lainnya terkait metode SVM *multiclass*[5]–[7]

Sebagian besar studi menggunakan teknik klasifikasi standar yang mengasumsikan bahwa data *training* didistribusikan secara merata di semua kategori. Terlepas dari kenyataan bahwa ketidakseimbangan jumlah data *training* sering ditemui dalam praktiknya. Akan ada kelas minoritas dan kelas mayoritas di *dataset*. Kinerja pengklasifikasi teks seringkali mengalami penurunan ketika dihadapkan pada kondisi tersebut [8]. Untuk menyeimbangkan jumlah data *training* yang digunakan per estimasi, peneliti menggunakan metode “*Adaptive Synthetic*” untuk menyelesaikan permasalahan tersebut.

Penelitian ini bertujuan untuk mengklasifikasi *tweet* dari pengguna Twitter pada setiap provinsi di Indonesia terkait kenaikan harga BBM. *Dataset* dibagi menjadi *training set* dan *testing set* kemudian *testing set* diklasifikasikan berdasarkan model SVM *multiclass* pada *training set* yang telah dilakukan *preprocessing*. Hasil klasifikasi SVM *multiclass* akan dievaluasi berdasarkan nilai akurasi.

## II. METODE PENELITIAN

Pengumpulan data dalam penelitian ini menggunakan metode *scraping*. *Web scraping* adalah teknik untuk mengekstraksi informasi secara otomatis dari suatu *website* atau beberapa *website* [9]. Teknik ini menghasilkan konten yang relevan berdasarkan kueri dan mengubah format tidak terstruktur menjadi representasi terstruktur [10]. Pada tahap *scraping*, peneliti mengambil data *tweet* dari pengguna Twitter terkait kenaikan harga BBM. Di bawah ini adalah contoh data yang dikumpulkan oleh peneliti.

**Tabel 1.** Sampel Data Hasil *Scraping*

Username	Tweet	User Location
@SilviaPutrii9	Penyesuaian bbm bertujuan untuk mengurangi beban subsidi besar akibat harga minyak dunia yang terus naik, kebijakan pemerintah sudah benar#BantuanBBMUntukRakyat #BLTBBMTepatSasaran #HematCermatBBM	Pontianak, Kalimantan Barat
@inthesky014	BBM naik itu biar yang sering bawa motor ke masjid pada jalan kaki, supaya pahalanya lebih banyak, subhanallah pemerintah kita ini sangat memperdulikan iman rakyatnya	Cianjur, Indonesia
@Kentrिंगmanikk	@M45Broo_ Saya bisnis laundry dukung 1000% BBM naik. Bila perlu naikan terus, hilangkan subsidi BBM. Larang mobil pake BB fosil secepatnya. Subsidi bisa di pakai utk hal lain. Sudah tdk efisien negeri ini	Tanjung Emas, Indonesia

Jumlah data yang diambil pada setiap provinsi sebanyak 100 *tweet* dengan total *tweet* yang digunakan dalam penelitian ini sebanyak 3400. Data yang diambil adalah *tweet* dalam bahasa

Indonesia. Kategori sentimen sebanyak tiga kategori yaitu sentimen positif, sentimen netral, dan sentimen negatif.

Langkah pertama pada penelitian ini adalah *text preprocessing*. *Text preprocessing* dilakukan untuk mengelola data teks sebelum membuat model *machine learning* [11]. Misalnya, menghapus tagar, URL, *stopword*, tanda baca, *@username*, dan karakter duplikat dalam sebuah kata [12]. Selain itu, *case folding*, normalisasi kata, *cleansing*, pemfilteran/penghapusan kata berhenti, *stemming*, dan tokenisasi juga dilakukan dalam *preprocessing* teks [13].

a) Case Folding

Data diubah menjadi huruf kecil sehingga huruf besar dan huruf kecil dengan arti yang sama tidak diperlakukan berbeda [14].

b) Word Normalization

Normalisasi kata digunakan untuk mengubah kata yang tidak baku atau disingkat menjadi kata baku dalam Bahasa Indonesia.

c) Cleansing

*Cleansing* digunakan untuk membersihkan kata-kata yang tidak diperlukan seperti *hashtag* (#), alamat *website*, *username* (@username), angka, emoji dan *email*.

d) Filtering/Stopwords Removal

*Stopwords removal* adalah langkah utama dalam *preprocessing* teks di *Natural Language Processing* (NLP). Tahap ini menyaring kata-kata yang mengandung sedikit informasi atau kata yang tidak ada makna semantik dari teks yang diberikan [15].

e) Stemming

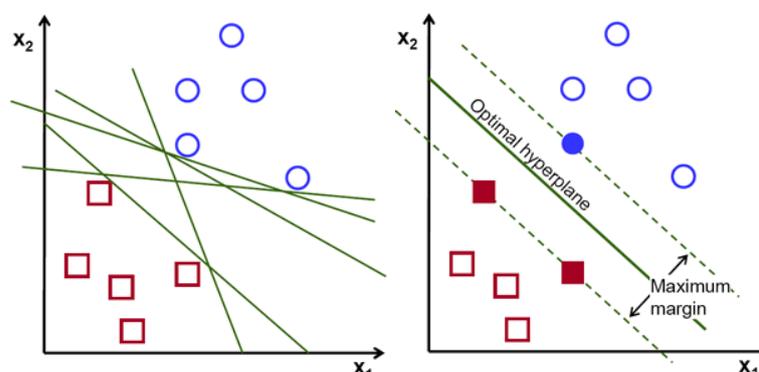
Langkah ini dilakukan untuk menemukan akar kata dengan menghapus awalan atau akhiran.

f) Tokenization

Tokenisasi mengacu pada proses mengubah teks apa pun menjadi serangkaian token, masing-masing berbeda dan tidak bergantung satu sama lain.

Tahap kedua adalah penerapan metode *Adaptive Synthetic* (ADASYN) *Imbalanced class data* adalah kondisi data yang tidak seimbang antar kelas data [16]. Kondisi data yang tidak seimbang merupakan masalah dalam klasifikasi karena *learning classifier* akan cenderung memprediksi kelas data mayoritas dibandingkan dengan kelas minoritas. Akibatnya, akurasi prediksi yang dihasilkan menguntungkan untuk sebagian besar kelas data *training*, sedangkan untuk kelas minoritas akan menghasilkan akurasi prediksi yang buruk (Chawla, 2003). Kami mengusulkan metode adaptif untuk memfasilitasi pembelajaran dengan *imbalanced class data*. Tujuannya di sini adalah dua, yaitu: mengurangi bias dan belajar secara adaptif. Algoritma yang diusulkan untuk masalah klasifikasi *multiclass* dijelaskan dalam algoritma ADASYN [17].

Tahap ketiga adalah klasifikasi dengan metode *Support Vector Machine* (SVM) *multiclass*. SVM merupakan metode klasifikasi yang memiliki tingkat akurasi yang baik. Tujuan dari algoritma SVM adalah untuk menemukan *hyperplane* dalam ruang berdimensi-N ( $N$  — jumlah fitur) yang secara jelas mengklasifikasikan titik data [18].



**Gambar 1.** *Support Vector Machine* (SVM)

Untuk memisahkan dua kelas titik data tersebut, ada banyak kemungkinan *hyperplane* yang bisa dipilih. Tujuan kita adalah menemukan bidang yang memiliki margin maksimum, yaitu jarak maksimum antara titik data dari kedua kelas. Memaksimalkan jarak margin memberikan penguatan sehingga titik data di masa mendatang dapat diklasifikasikan dengan lebih baik [19].

*Hyperplanes* adalah batas keputusan yang membantu mengklasifikasikan titik data. Poin data yang jatuh di kedua sisi hyperplane dapat dikaitkan dengan kelas yang berbeda. Juga, dimensi hyperplane bergantung pada jumlah fitur. Jika jumlah fitur input adalah 2, maka hyperplane hanya berupa garis. Jika jumlah fitur masukan adalah 3, maka *hyperplane* menjadi bidang dua dimensi. Sulit membayangkan ketika jumlah fitur melebihi 3[20]. Didalam algoritma SVM, dilihat batas maksimal antara poin data dan *hyperplane*. *Los function* membantu memaksimumkan batas adalah *hinge loss*[21]. Bidang pembatas pertama menjadi batas kelas pertama sedangkan bidang pembatas kedua adalah batas dari kelas kedua, sehingga diperoleh Persamaan 1.

$$\begin{aligned} xi . w + b &\geq +1 \text{ for } yi = +1 \\ xi . w + b &\leq -1 \text{ for } yi = -1 \end{aligned} \quad (1)$$

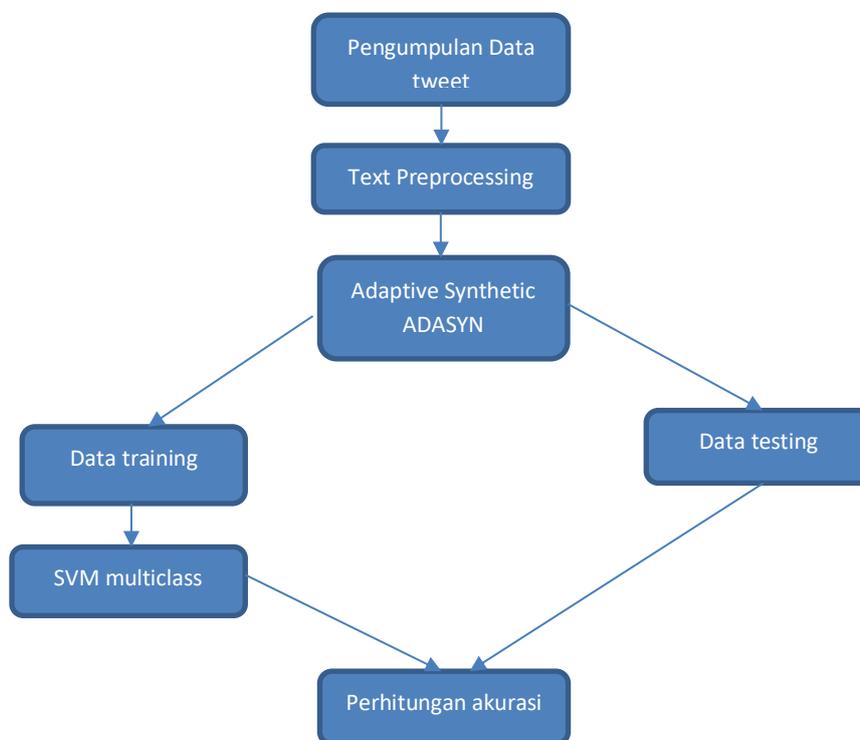
Keterangan:

w : Normal bidang

b : Posisi bidang relatif terhadap pusat koordinat

Secara umum, cara kerja dari SVM adalah menemukan jarak terjauh dari hyperplane dengan kedua kelas. Proses penentuan jarak terjauh dilakukan berulang kali hingga menemukan hyperplane terbaik. Untuk itulah diperlukan optimasi pada SVM untuk menemukan jarak maksimum hyperplane dengan kedua kelas tersebut. Dalam pembangunan SVM, terdapat dua bentuk optimasi yang digunakan untuk menemukan hyperplane. Bentuk optimasi pertama yaitu Primal Form SVM dan yang kedua adalah Dual Form SVM. Primal Form tidak dapat digunakan dalam penelitian ini karena tidak akan pernah memenuhi konstrain.

Langkah terakhir pada penelitian ini adalah menghitung akurasi pada setiap provinsi di Indonesia. Perhitungan akurasi digunakan untuk mengetahui seberapa baik metode ADASYN dan SVM dalam mengklasifikasi sentimen kenaikan BBM di seluruh provinsi di Indonesia. Berikut bagan penelitian disajikan pada Gambar 2.



**Gambar 2.** Diagram penelitian

### III. HASIL DAN PEMBAHASAN

Pada bagian ini akan diperlihatkan gambaran komentar masyarakat Indonesia terhadap kenaikan Bahan Bakar Minyak (BBM) melalui bentuk *wordcloud*. Sentimen masyarakat Indonesia terhadap kenaikan BBM dibagi menjadi tiga label yaitu positif (pro terhadap kenaikan BBM), netral (tidak

memihak), dan negative (kontra terhadap kenaikan BBM). Sentimen yang dihasilkan lebih banyak kontra terhadap kenaikan BBM, sehingga mengakibatkan *imbalance* data. Pada penelitian ini menerapkan metode *Adaptive Synthetic* (ADASYN) untuk mengatasi permasalahan tersebut. Hasil ADASYN akan diklasifikasikan menggunakan metode *Support Vector Machine* (SVM).

Pembahasan pertama pada penelitian dimulai dari visualisasi sentimen masyarakat Indonesia menggunakan *wordcloud*. Visualisasi dilakukan untuk setiap provinsi di Indonesia. Semakin besar ukuran kata pada *wordcloud* maka topik tersebut sering dibicarakan oleh masyarakat. Berikut merupakan sampel hasil *wordcloud* setiap provinsi.



**Gambar 3.** Sampel *Wordcloud* Sentimen Kenaikan Harga BBM (a) Positif Provinsi Aceh (b) Netral Provinsi Aceh (c) Negatif Provinsi Aceh (d) Positif Provinsi Kalimantan Timur (e) Netral Provinsi Kalimantan Timur (f) Negatif Kalimantan Timur (g) Positif Provinsi Jawa Tengah (h) Netral Provinsi Jawa Tengah (i) Negatif Provinsi Jawa Tengah (j) Positif Provinsi DKI Jakarta (k) Netral Provinsi DKI Jakarta (l) Negatif Provinsi DKI Jakarta.

Berdasarkan Gambar 3 terlihat bahwa sentimen positif di setiap provinsi frekuensinya hampir merata dan tidak ada kata yang menonjol mendukung kenaikan BBM. Kata-kata yang muncul pada sentimen positif lebih menekankan pada apa yang harus dilakukan oleh pemerintah seperti “subsidi” dan “sesuai harga”. Hal yang sama terjadi pada sentimen netral frekuensi kata terbanyak adalah “harga”, “turun”, “naik”. Sedangkan kata yang sering muncul pada sentimen negatif adalah “harga”, “demo”, “bakar”, dan “naik”. Kata tersebut menggambarkan luapan amarah dari masyarakat Indonesia terkait kenaikan BBM.

Untuk mengetahui lebih detail mengenai kata yang sering diucapkan oleh masyarakat terkait kenaikan BBM disajikan pada Tabel 2. Tabel 2 menyajikan frekuensi kemunculan kata terbanyak untuk setiap sentimen positif, netral, dan negatif di setiap provinsi akan digunakan untuk merepresentasikan kata-kata yang paling sering digunakan oleh masyarakat terkait dengan kenaikan



harga BBM. Kata-kata yang paling sering digunakan oleh masyarakat umum pada setiap kategori sentimen di setiap provinsi terdapat pada tabel di bawah ini.

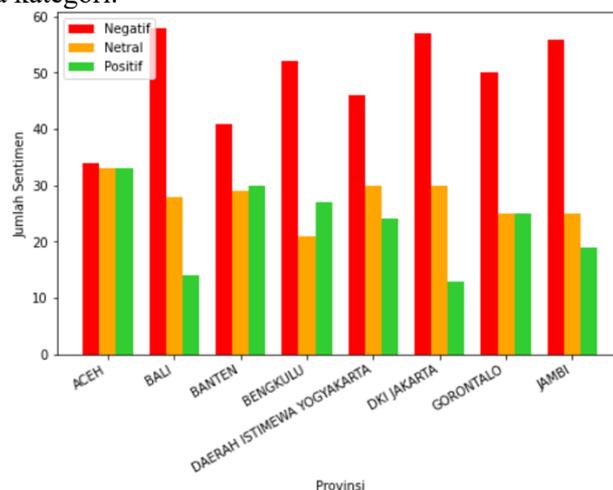
**Tabel 2.** Frekuensi Kata yang Sering Digunakan

No	Provinsi	Kata yang Sering Digunakan (%)		
		Negatif	Netral	Positif
1	Aceh	Harga (5.97 %)	Harga (9.71%)	Harga (14.72 %)
2	Bali	Harga (9.88 %)	Harga (9.42 %)	Harga (10.47 %)
3	Banten	Harga (13.64 %)	Harga (5.61 %)	Harga (13.92 %)
4	Bengkulu	Harga (11.11 %)	BLT (7.35%)	Harga (13.22 %)
5	DI Yogyakarta	Harga (6.71 %)	Harga (6.36 %)	Harga (14.16 %)
6	DKI Jakarta	Bakar (5.70 %)	Bakar (5.32 %)	Bakar (5.87 %)
7	Gorontalo	Harga (14.29 %)	Harga (11.97 %)	Harga (14.46 %)
8	Jambi	Harga (11.76 %)	Harga (10 %)	Harga (14.30 %)
9	Jawa Barat	Harga (7.19 %)	Harga (7.46 %)	Harga (11.32 %)
10	Jawa Tengah	Harga (8.01 %)	Harga (6.56 %)	Harga (12.88 %)
11	Jawa Timur	Harga (8.41 %)	Harga (7.41 %)	Harga (12.69 %)
12	Kalimantan Timur	Harga (10.26 %)	Sepakat (10.04 %)	Harga (12.62 %)
13	Kalimantan Selatan	Harga (10.29 %)	Harga (8.14 %)	Kalimantan (10.36 %)
14	Kalimantan Tengah	Naik (5.86 %)	Harga (5.02 %)	Harga (12.54 %)
15	Kalimantan Barat	Moga (14.29)	Salah (3.39 %)	Sesuai (9.77 %)
16	Kalimantan Utara	Harga (13.33 %)	Harga (11.11 %)	Harga (10.53 %)
17	Bangka Belitung	Harga (7.14 %)	Harga (8.22 %)	Pulau (13.61 %)
18	Riau	Blitar (9.21 %)	Bantu (10.00 %)	Masyarakat (23.08 %)
19	Lampung	Harga (9.42 %)	Harga (5.18 %)	Dukung (9.74 %)
20	Maluku	Naik (7.41 %)	Harga (15.00 %)	Harga (10.59 %)
21	Maluku Utara	Harga (8.47 %)	Utara (8.33 %)	Utara (13.24 %)
22	Nusa Tenggara Barat	Harga (10.17 %)	Harga (5.61 %)	Tenggara (5.39 %)
23	Nusa Tenggara Timur	Nusa (12.00 %)	Harga (5.99 %)	Dukung (9.11 %)
24	Papua	Turun (18.18 %)	Sa (7.53 %)	Masyarkat (5.84 %)
25	Papua Barat	Oktober (16.67 %)	Harga (9.93 %)	Perintah (7.59 %)
26	Riau	Harga (6.63 %)	Harga (5.40 %)	Harga (14.10 %)
27	Sulawesi Barat	Harga (11.70 %)	Harga (7.38 %)	Utara (8.56 %)
28	Sulawesi Selatan	Harga (10.22 %)	Harga (5.47 %)	Subsidi (5.00 %)
29	Sulawesi Tengah	Harga (14.29 %)	Harga (11.90 %)	Dukung (9.43 %)
30	Sulawesi Tenggara	Harga (10.28 %)	Bantu (5.84 %)	Subsidi (5.27 %)
31	Sulawesi Utara	Harga (10.34 %)	Harga (8.77 %)	Harga (15.53 %)
32	Sumatera Barat	Harga (11.59 %)	Harga (9.09 %)	Jambi (10.93 %)
33	Sumatera Selatan	Harga (8.96 %)	Bantu (4.08 %)	Dukung (8.23 %)
34	Sumatera Utara	Turun (15.38 %)	Harga (4.90 %)	Dukung (10.67 %)

Kata “harga” adalah kata yang paling sering digunakan pada kategori negatif, netral, dan positif. Pada kategori negatif, kata "harga" merupakan kata yang paling sering digunakan di 25 provinsi dari total 34 provinsi di Indonesia. Pada kategori netral, kata "harga" juga mendominasi kata yang paling banyak digunakan di 25 provinsi dari total 34 provinsi di Indonesia. Pada kategori positif, kata yang paling sering digunakan cukup beragam. Kata "harga" merupakan kata yang paling sering digunakan

di 16 provinsi, kata "dukung" merupakan kata yang sering digunakan di 5 provinsi, kata "subsidi" paling sering digunakan di 2 provinsi yaitu Sulawesi Utara dan Sulawesi Selatan.

Terdapat perbedaan frekuensi kata yang paling sering digunakan untuk setiap provinsi di Indonesia. Daerah yang cukup berbeda jika dibandingkan dengan daerah lain adalah provinsi DKI Jakarta, Kalimantan Timur, Kepulauan Riau, Maluku Utara, Nusa Tenggara Timur, Papua, Papua Barat, Sulawesi Tenggara, Sumatera Selatan, dan Sumatera Utara. Di daerah-daerah tersebut, kata yang paling sering digunakan pada setiap kategori cukup beragam dan tidak ada kata yang mendominasi untuk semua kategori.



**Gambar 4.** Jumlah Sentimen Per Label Per Provinsi

Rata-rata frekuensi sentimen negatif mendominasi setiap provinsi di Indonesia, hal ini menunjukkan bahwa penolakan kenaikan BBM terjadi di berbagai provinsi di Indonesia. Ketidakseimbangan label pada analisis klasifikasi menimbulkan masalah serius (Gambar 4). Apabila terdapat data baru masuk dan diprediksi berdasarkan model terbentuk, data tersebut akan diarahkan ke label yang paling dominan. Metode ADASYN membangkitkan data sintetik untuk menyeimbangkan frekuensi data. Hasil ADASYN akan diklasifikasikan menggunakan metode SVM. Berikut hasil klasifikasi menggunakan metode SVM *multiclass*.

**Tabel 3.** Akurasi Metode SVM *Multiclass* Per Provinsi

Provinsi	Akurasi	Provinsi	Akurasi
Aceh	0.70	Kepulauan Riau	0.90
Bali	0.90	Lampung	0.90
Bangka Belitung	0.90	Maluku	0.85
Banten	0.90	Maluku Utara	0.85
Bengkulu	0.80	NTB	0.90
DI Yogyakarta	0.95	NTT	0.85
DKI Jakarta	0.95	Papua	0.95
Gorontalo	0.85	Papua Barat	0.95
Jawa Barat	0.95	Riau	0.90
Jambi	0.90	Sulawesi Barat	0.90
Jawa Tengah	0.90	Sulawesi Selatan	0.85
Jawa Timur	0.90	Sulawesi Tengah	0.90
Kalimantan Barat	0.85	Sulawesi Tenggara	0.90
Kalimantan Selatan	0.90	Sulawesi Utara	0.75
Kalimantan Tengah	0.90	Sumatera Barat	0.90
Kalimantan Timur	0.80	Sumatera Selatan	0.85
Kalimantan Utara	0.85	Sumatera Utara	0.90

Berdasarkan Tabel 3 akurasi metode SVM *multiclass* terendah adalah provinsi Aceh sebesar 70%. Sedangkan Akurasi tertinggi terdapat pada provinsi DI Yogyakarta, DKI Jakarta, Jawa Barat, Papua, Papua Barat sebesar 95%. Rata-rata akurasi seluruh provinsi di Indonesia sebesar 87.94%. Hal ini



menunjukkan bahwa metode ADASYN SVM *multiclass* mampu memberikan akurasi yang baik untuk pengklasifikasian data tidak seimbang (*imbalance data*).

#### IV. KESIMPULAN

Sentimen negatif terhadap kenaikan BBM di seluruh provinsi di Indonesia mendominasi tweet. Hal ini dibuktikan dengan frekuensi sentimen negatif lebih banyak dibandingkan dengan netral dan positif. Provinsi dengan sentimen negatif terbanyak adalah DKI Jakarta. Penerapan metode ADASYN mengatasi ketidaksamaan frekuensi antara sentimen negatif, positif, dan netral. Hasil penyamaan ADASYN dilakukan klasifikasi dengan menggunakan metode SVM *multiclass* menghasilkan rata-rata akurasi 87.94%. Akurasi tertinggi adalah provinsi DKI Jakarta, DI Yogyakarta, Jawa Barat, Papua, dan Papua Barat dengan akurasi mencapai 95%.

#### UCAPAN TERIMA KASIH

#### REFERENSI

1. Z. Jianqiang and G. Xiaolin, “Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis,” *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.
2. D. Anjas Ramadhan and E. Budi Setiawan SSi, “ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE.”
3. A. Ridok and R. Latifah, “Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN,” in *Konferensi Nasional Sistem & Informatika*, 2015, pp. 1–6. [Online]. Available: [www.kompas.com](http://www.kompas.com)
4. H. N. Irmanda and R. Astriratma, “Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM),” *RESTI*, vol. 1, no. 3, pp. 915–922, 2017.
5. H. Guo and W. Wang, “An active learning-based SVM multi-class classification model,” *Pattern Recognit*, vol. 48, no. 5, pp. 1577–1597, 2015, doi: <https://doi.org/10.1016/j.patcog.2014.12.009>.
6. V. Blanco, A. Japón, and J. Puerto, “Optimal arrangements of hyperplanes for SVM-based multiclass classification,” *Adv Data Anal Classif*, vol. 14, no. 1, pp. 175–199, 2020, doi: 10.1007/s11634-019-00367-6.
7. C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans Neural Netw*, vol. 13, no. 2, pp. 415–425, 2002, doi: 10.1109/72.991427.
8. A. Ridok and R. Latifah, “Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN,” 2015. [Online]. Available: [www.kompas.com](http://www.kompas.com)
9. S. Chaudhari, A. R. V. G. Tekkur, P. G. L., and S. R. Karki, “Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef,” *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020.
10. T. Karthikeyan, K. Sekaran, D. Ranjith, V. Vinoth kumar, and J. M. Balajee, “Personalized content extraction and text classification using effective web scraping techniques,” *International Journal of Web Portals*, vol. 11, no. 2, pp. 41–52, Jul. 2019, doi: 10.4018/IJWP.2019070103.
11. A. Kurniasih and L. P. Manik, “On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 927–934, 2022, doi: 10.14569/IJACSA.2022.01306109.
12. A. F. Hidayatullah, S. Cahyaningtyas, and A. M. Hakim, “Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset,” *IOP Conf Ser Mater Sci Eng*, vol. 1077, no. 1, p. 012001, Feb. 2021, doi: 10.1088/1757-899x/1077/1/012001.
13. L. P. Manik *et al.*, “Aspect-Based Sentiment Analysis on Candidate Character Traits in Indonesian Presidential Election,” in *Proceeding - 2020 International Conference on Radar, Antenna, Microwave, Electronics and Telecommunications, ICRAMET 2020*, Institute of



- Electrical and Electronics Engineers Inc., Nov. 2020, pp. 224–228. doi: 10.1109/ICRAMET51080.2020.9298595.
14. M. Mayo, “A General Approach to Preprocessing Text Data,” Dec. 01, 2017.
  15. A. A. V. A. Jayaweera, Y. N. Senanayake, and P. S. Haddela, “Dynamic Stopword Removal for Sinhala Language,” 2019.
  16. R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, “Data mining with imbalanced class distributions: concepts and methods,” 2009.
  17. H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning,” 2008.
  18. F. Fauzi, “K-Nearset Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah Info Artikel,” *Jurnal MIPA*, vol. 40, no. 2, 2017, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/JM>
  19. M. Y. Darsyah, I. J. Suprayitno, F. Fuzi, B. W. Otok, and B. S. S. Ulama, “Smooth Support Vector Machine (SSVM) for classification of human development index,” in *Journal of Physics: Conference Series*, 2019. doi: 10.1088/1742-6596/1217/1/012114.
  20. Z. F. Hussain *et al.*, “A new model for iris data set classification based on linear support vector machine parameter’s optimization,” *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 1079–1084, 2020, doi: 10.11591/ijece.v10i1.pp1079-1084.
  21. D. Agustina, E. Putri, F. Fauzi, S. N. Alawiyah, and R. Wasono, “METODE SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI DATA EKSPRESI GEN MICROARRAY,” in *EDUSAINTEK 4*, 2020, pp. 1–10.



# Analisis dan Simulasi Teori Antrian pada Indomaret Darmo Surabaya

Thesion Marta Sianipar<sup>1</sup>, Halim Arif Cahyono<sup>2</sup>, Amalia Nur Alifah<sup>3</sup>

<sup>1, 2, 3</sup>Sains Data, Institut Teknologi Telkom Surabaya,

<sup>1</sup>[thesion.marta.21@student.ds.ittelkom-sby.ac.id](mailto:thesion.marta.21@student.ds.ittelkom-sby.ac.id)

<sup>2</sup>[halim.arif.21@student.ds.ittelkom-sby.ac.id](mailto:halim.arif.21@student.ds.ittelkom-sby.ac.id)

<sup>3</sup>[amaliaalfah@ittelkom-sby.ac.id](mailto:amaliaalfah@ittelkom-sby.ac.id)

Corresponding author email: [thesionmartasianipar@gmail.com](mailto:thesionmartasianipar@gmail.com)

**Abstract:** Service quality is one of the key factors to get a good rating from consumers. However, an increase in the number of consumers is sometimes not matched by an increase in the number of service providers. This is what causes the queue. Queuing theory provides several alternative mathematical models that can be used to determine some of the properties of a queuing system and optimize decision making. For this reason, queuing theory simulation is one of the important things to do so that the service system in one place can be optimal. In this research, analysis and simulation of queuing theory will be carried out at Indomaret Darmo Surabaya. The data used are primary data taken directly by researchers. In this study, the results of queuing theory analysis will be displayed in several forms of data visualization and determine the confidence intervals of service duration, service waiting time, and idle time. The results show that the service duration at Indomaret Darmo Surabaya can be trusted to be between 4.48 to 5.128 minutes. While the average waiting time can be trusted to be between 3,974 to 5,505 minutes. In addition, the results of the analysis show that the average, median, mode, and standard deviation of idle time are 0 (less than 1 minute). This shows that there is almost no free time for services at Indomaret Darmo Surabaya.

**Keywords:** idle time, queuing theory, service duration, waiting time

**Abstrak:** Kualitas pelayanan menjadi salah satu faktor kunci untuk mendapatkan penilaian yang baik dari konsumen. Namun, adanya peningkatan jumlah konsumen terkadang tidak dimbangi dengan peningkatan jumlah penyedia layanan. Hal inilah yang menyebabkan terjadinya antrian. Teori antrian memberikan beberapa alternatif model matematis yang dapat digunakan untuk menentukan beberapa sifat dari suatu sistem antrian dan mengoptimalkan pengambilan keputusan. Untuk itu, simulasi teori antrian merupakan salah satu hal yang penting untuk dilakukan agar sistem pelayanan pada suatu tempat dapat menjadi optimal. Pada penelitian ini akan dilakukan analisis dan simulasi teori antrian pada Indomaret Darmo Surabaya. Data yang digunakan adalah data primer yang diambil secara langsung oleh peneliti. Pada penelitian ini, hasil analisis teori antrian akan ditampilkan dalam beberapa bentuk visualisasi data serta menentukan interval kepercayaan dari durasi pelayanan, waktu tunggu pelayanan, serta idle time. Hasil menunjukkan bahwa durasi pelayanan di Indomaret Darmo Surabaya dapat dipercayai berada antara 4,48 hingga 5,128 menit. Sedangkan rata-rata waktu tunggu dapat dipercayai berada antara 3,974 hingga 5,505 menit. Selain itu, hasil analisis menunjukkan bahwa rata-rata, median, modus, dan standar deviasi dari idle time bernilai 0 (kurang dari 1 menit). Ini menunjukkan bahwa hampir tidak ada waktu kosong pada pelayanan di Indomaret Darmo Surabaya.

**Kata kunci:** durasi pelayanan, idle time, teori antrian, waktu tunggu

## I. PENDAHULUAN

Dalam era globalisasi, perkembangan ekonomi memunculkan ketergantungan dan persaingan antar wilayah. Perubahan bisnis yang semakin cepat mendorong perusahaan atau organisasi untuk merespons perubahan tersebut. Tantangan utama yang dihadapi perusahaan adalah bagaimana mereka dapat menarik dan mempertahankan pelanggan mereka [1]. Pertumbuhan populasi yang terus meningkat mendorong perusahaan jasa untuk mengubah sistem mereka menjadi lebih efektif guna memberikan layanan terbaik. Kualitas pelayanan menjadi salah satu faktor kunci untuk mendapatkan penilaian yang baik dari konsumen. Namun, adanya peningkatan jumlah konsumen terkadang tidak dimbangi dengan peningkatan jumlah penyedia layanan. Hal inilah yang menyebabkan terjadinya



antrian. Teori antrian memberikan beberapa alternatif model matematis yang dapat digunakan untuk menentukan beberapa sifat dari suatu sistem antrian dan mengoptimalkan pengambilan keputusan [2].

Untuk meningkatkan jumlah pelanggan, selain melakukan promosi, menciptakan produk baru yang lebih menarik, kepuasan konsumen dalam hal kemudahan dan kecepatan pelayanan juga harus diperhatikan [3]. Simulasi teori antrian telah banyak dilakukan oleh beberapa peneliti pada objek penelitian yang berbeda-beda, diantaranya adalah pada pelayanan teller bank, supermarket, hingga pom bensin [4-6]. Sedangkan penelitian ini akan melakukan analisis serta simulasi pada salah satu rantai ritel waralaba di Indonesia, yaitu Indomaret. Indomaret merupakan jaringan mini market yang mencakup kebutuhan pokok dan sehari-hari, dengan luas ritel kurang dari 200 m<sup>2</sup>. Pada Desember 2010, jumlah Indomaret di Indonesia adalah sebanyak 4.955 toko. Dari jumlah tersebut, 3.058 toko dimiliki secara pribadi dan 1.897 toko waralaba sisanya dimiliki oleh masyarakat dan berlokasi di kota Jabodetabek, Jawa Barat, Jawa Timur, Jawa Tengah, Yogyakarta, Bali, dan Lampung. Indomaret mudah ditemukan di kawasan perumahan, gedung perkantoran, serta ruang publik. Lebih dari 3.500 produk makanan dan non makanan tersedia dengan harga bersaing, mencakup hampir semua kebutuhan sehari-hari konsumen.

Data yang digunakan pada penelitian ini adalah data primer yang diambil secara langsung oleh peneliti di Indomaret Darmo Surabaya. Sistem antrian yang digunakan pada Indomaret Darmo adalah First in First Out (FIFO), yaitu antrian yang memprioritaskan pelayanan pada pengunjung yang datang pertama. Indomaret menggunakan sistem pelayanan multi channel-single phase, yaitu dengan beberapa fasilitas pelayanan dalam satu antrian tunggal atau satu fase pelayanan. Dalam penelitian ini, peneliti hanya fokus pada hitungan data antrian berdasarkan 1 kasir saja dari beberapa kasir yang ada. Seperti halnya sistem single channel-single phase, hanya ada satu server atau fasilitas pelayanan dan satu tahapan atau fase pelayanan.

Pada penelitian ini, hasil analisis teori antrian akan ditampilkan dalam beberapa bentuk visualisasi data serta menentukan interval kepercayaan pada beberapa aspek. Selama ini, pada penelitian terdahulu, belum ada penelitian yang membandingkan hasil analisis serta simulasi teori antrian antara data real time dengan data yang dibangkitkan secara acak. Sehingga pada penelitian ini, kami mencoba untuk menganalisis, mensimulasikan, serta membandingkan hasil simulasi antara data realtime dengan data bangkitan. Harapannya, perbandingan hasil analisis teori antrian antara data realtime dengan data hasil bangkitan adalah sama. Dengan hasil analisis serta simulasi pada penelitian ini, harapannya dapat memberikan evaluasi pada pelayanan, serta dapat meningkatkan kualitas pelayanan yang ada pada Indomaret Darmo Surabaya.

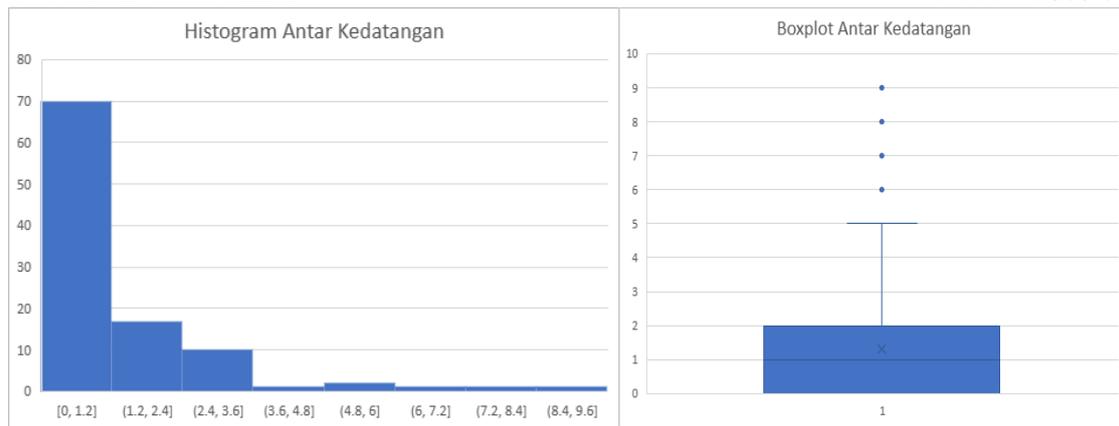
## II. METODE PENELITIAN

Penelitian ini adalah penelitian tindakan (*action research*). Data yang digunakan adalah data kuantitatif dan sumber data primer. Dalam penelitian ini, teknik pengumpulan data menggunakan metode observasi langsung, yaitu dengan mengamati waktu kedatangan, awal dan akhir pelayanan pelanggan. Studi antrian di Indomaret Darmo Surabaya dilakukan pada Jumat, 27 Januari 2023 pada pukul 14.00 WIB. Data yang diperoleh adalah sebanyak 104 data pengunjung.

### 2.1. Data Pengamatan

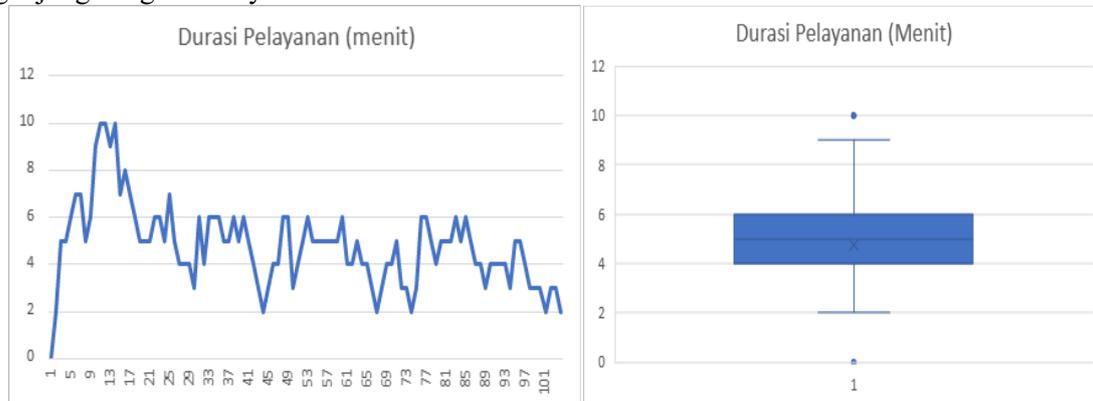
Pada penelitian ini, data yang diambil berupa data waktu kedatangan, awal pelayanan, serta akhir pelayanan. Waktu kedatangan merupakan waktu kedatangan konsumen, yaitu adalah waktu saat konsumen masuk ke indomaret. Sedangkan awal pelayanan adalah waktu dimana konsumen mulai dilayani, dan akhir pelayanan adalah waktu dimana konsumen selesai dilayani.

Setelah memperoleh data waktu kedatangan, awal pelayanan, serta akhir pelayanan, selanjutnya waktu antar kedatangan serta waktu durasi pelayanan dapat ditentukan. Waktu antar kedatangan merupakan jarak antara kedatangan suatu konsumen dengan konsumen sebelumnya. Sedangkan durasi pelayanan adalah total waktu yang dibutuhkan dalam melakukan pelayanan customer.

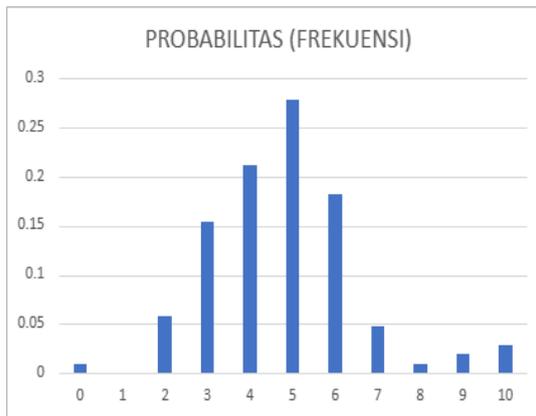
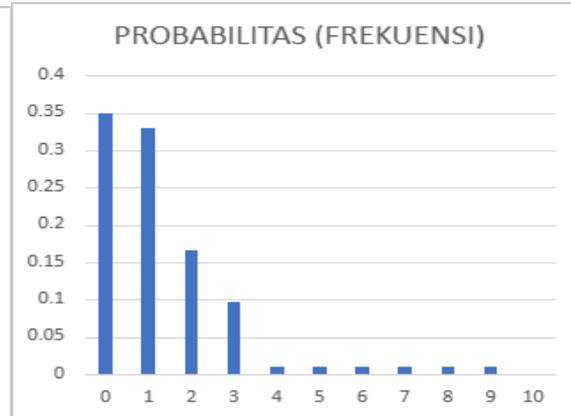
**Gambar 1a. Histogram****Gambar 1b. Boxplot****Gambar 1. Visualisasi Waktu Antar Kedatangan : (1a) Histogram (1b) Boxplot**

Dalam grafik histogram tersebut, interval waktu antar kedatangan pelanggan yang paling banyak terjadi adalah pada rentang waktu 0 hingga 1.2 menit. Analisis data juga menunjukkan bahwa distribusi waktu antar kedatangan tidak berdistribusi normal dan memiliki skewness positif. Dapat dilihat dari grafik boxplot antar kedatangan di atas, terdapat beberapa data yang teridentifikasi sebagai outlier, dengan rentang waktu antar kedatangan sekitar 6 menit hingga 9 menit. Dalam hal ini, outlier pada waktu antar kedatangan menunjukkan adanya beberapa pengunjung yang mengalami waktu antar kedatangan yang signifikan lebih lama dibandingkan dengan kebanyakan pengunjung lain.

Berdasarkan hasil statistika deskriptif, rata-rata jarak waktu antar kedatangan konsumen di Indomaret adalah 1,33 menit dengan median 1 menit. Modus jarak waktu antar kedatangan adalah 0 menit, yang menunjukkan bahwa nilai paling sering muncul dalam data adalah 0 menit. Hal ini menunjukkan bahwa selisih waktu kedatangan antar pengunjung paling banyak adalah kurang dari 0 menit, ini artinya indomaret tergolong ramai, dilihat dari selisih waktu kedatangan antara satu pengunjung dengan lainnya.

**Gambar 2a. Line Chart****Gambar 2b. Boxplot****Gambar 2. Visualisasi Durasi Pelayanan : (2a) Line Chart (2b) Boxplot**

Jika dilihat dari grafik line chart di atas, terlihat bahwa durasi pelayanan pelanggan Indomaret Darmo Surabaya paling sering berkisar antara 4 menit hingga 6 menit. Dari grafik boxplot durasi pelayanan, terdapat 2 titik outlier pada 0 menit dan 10 menit. Kuartil 1 (Q1) berada pada rentang waktu 4 menit, sedangkan kuartil 2 (Q2) berada pada rentang waktu 6 menit. Rentang durasi pelayanan paling sedikit 2 menit dan paling maksimal 9 menit. Berdasarkan hasil statistika deskriptif menunjukkan bahwa Rata-rata durasi pelayanan adalah 4.788 menit dengan median (nilai tengah) durasi pelayanan adalah 5 menit. Sedangkan modus (nilai paling sering muncul) dari durasi pelayanan adalah 5 menit.


**Gambar 3a.** Durasi Pelayanan

**Gambar 3b.** Waktu Antar Kedatangan

**Gambar 3.** Bar Chart Probabilitas (frekuensi): (3a) Durasi Pelayanan (3b) Waktu Antar Kedatangan Probabilitas (frekuensi) durasi pelayanan merupakan probabilitas munculnya durasi pelayanan. Grafik barchart di atas menggambarkan tingkat probabilitas dari setiap kategori durasi pelayanan. Kategori durasi pelayanan yang memiliki probabilitas tertinggi adalah 5 menit. Sedangkan probabilitas (frekuensi) waktu antar kedatangan merupakan probabilitas munculnya waktu antar kedatangan. Informasi ini berguna untuk memahami pola waktu antar kedatangan pelanggan di indomaret. Grafik barchart di atas menggambarkan probabilitas setiap kategori waktu antar kedatangan, dengan sumbu X sebagai kategori waktu. Kategori waktu dengan probabilitas tertinggi adalah 0 menit, dengan probabilitas sebesar 0,35 atau 35%.

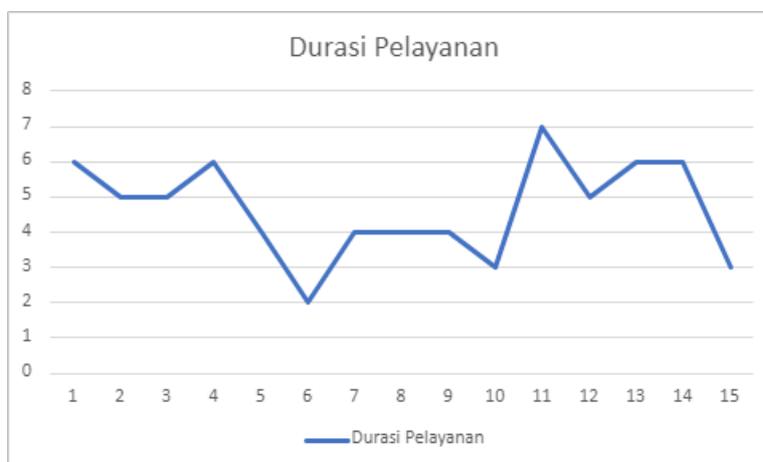
### 2.2. Data Acak

Selain mengambil data pengamatan atau real data, peneliti juga mencoba Menyusun data dengan membangkitkan bilangan acak berdasarkan tabel 1. Tabel 1 menentukan interval bilangan acak untuk waktu antar kedatangan dan juga durasi pelayanan berdasarkan kumulatif probabilitas yang diperoleh dari data pengamatan. Dengan demikian data acak dapat dibangkitkan.

**Tabel 1.** Interval Bilangan Acak

KEDATANGAN			PELAYANAN		
ANTAR KEDATANGAN	KUMULATIF PROBABILITAS	INTERVAL BILANGAN ACAK	DURASI PELAYANAN	KUMULATIF PROBABILITAS	INTERVAL BILANGAN ACAK
0	0.349514563	1-34	0	0.009615385	1
1	0.67961165	35-67	2	0.067307692	2-6
2	0.844660194	68-84	3	0.221153846	7-22
3	0.941747573	85-94	4	0.432692308	23-43
4	0.951456311	95	5	0.711538462	44-71
5	0.961165049	96	6	0.894230769	72-89
6	0.970873786	97	7	0.942307692	90-94
7	0.980582524	98	8	0.951923077	95
8	0.990291262	99	9	0.971153846	96-97

9	1	100	10	1	98-99
---	---	-----	----	---	-------



**Gambar 4.** Linechart Durasi Pelayanan dari Data Acak

Grafik line chart untuk durasi pelayanan pada data acak menunjukkan nilai maksimal sebesar 7 menit dan nilai minimal sebesar 2 menit. Ini menandakan bahwa durasi pelayanan terpendek adalah 2 menit dan durasi pelayanan terpanjang adalah 7 menit. Hasil ini menunjukkan hasil yang tidak jauh berbeda dari data pengamatan.

### III. HASIL DAN PEMBAHASAN

Pada bagian ini akan ditentukan waktu tunggu, waktu sistem, serta idle time dari hasil data pengamatan maupun data acak. Waktu tunggu merupakan selisih antara waktu kedatangan dengan awal pelayanan pengunjung, waktu sistem adalah total antara waktu tunggu dengan durasi pelayanan pengunjung. Sedangkan *idle time* adalah total waktu kekosongan (ketika tidak ada pelayanan dan tidak ada pengunjung yang hadir). Dengan demikian, penting untuk mengetahui waktu tunggu serta idle time. Waktu tunggu dapat menentukan seberapa baik sistem pelayanan, sedangkan idle time dapat menentukan seberapa ramai indomaret tersebut.

#### 3.1. Data Pengamatan

##### 3.1.1. Durasi Pelayanan

Berikut akan dihitung estimasi durasi pelayanan dengan selang kepercayaan 95%. Estimasi titik dari  $\mu$  adalah  $\bar{x} = 4,78$  dari  $n = 104$  dengan  $s = 1,74$  untuk selang kepercayaan 95% artinya  $(1-\alpha) \times 100\% = 95\%$  maka  $\alpha = 0,05$  karena  $\sigma$  tidak diketahui Dimana  $t_{\frac{0,05}{2}} = t_{0,025} = 1,983$  dengan derajat bebas  $v = n - 1 = 103$ , maka

$$\begin{aligned} \bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} &< \mu < \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \\ 4,78 - t_{\frac{0,05}{2}} \frac{1,74}{\sqrt{104}} &< \mu < 4,78 + t_{\frac{0,05}{2}} \frac{1,74}{\sqrt{104}} \\ 4,78 - 1,983 \frac{1,74}{\sqrt{104}} &< \mu < 4,78 + 1,983 \frac{1,74}{\sqrt{104}} \\ 4,48 &< \mu < 5,128 \end{aligned}$$

Berdasarkan analisis, estimasi selang kepercayaan dengan error 5% menunjukkan bahwa rata-rata durasi pelayanan dapat dipercayai berada antara 4,48 hingga 5,128 menit. Ini berarti bahwa jika data dikumpulkan dan dihitung ulang sebanyak banyak mungkin, 95% dari hasil akan berada dalam interval ini. Oleh karena itu, interval ini memberikan tingkat keyakinan tinggi pada perkiraan rata-rata durasi pelayanan.

##### 3.1.2. Waktu Tunggu

Berdasarkan statistika deskriptif dari data pengamatan, diperoleh bahwa rata-rata waktu tunggu adalah 9,53 menit dengan nilai tengah (median) dari waktu tunggu adalah 9 menit. Sedangkan modus (nilai yang paling sering muncul) dari waktu tunggu adalah 6 menit.

Berikut akan dihitung estimasi waktu tunggu dengan selang kepercayaan 95%. Estimasi titik dari  $\mu$  adalah  $\bar{x} = 4,74$  dari  $n = 104$  dengan  $s = 3,93$  untuk selang kepercayaan 95% artinya  $(1-\alpha) \times 100\%$

= 95% maka  $\alpha = 0,05$  karena  $\sigma$  tidak diketahui Dimana  $t_{0,05} = t_{0,025} = 1,983$  dengan derajat bebas  $v = n - 1 = 103$ , maka

$$\begin{aligned} \bar{x} - t_{\alpha} \frac{S}{\sqrt{n}} &< \mu < \bar{x} + t_{\alpha} \frac{S}{\sqrt{n}} \\ 4,74 - t_{0,05} \frac{3,93}{\sqrt{104}} &< \mu < 4,74 + t_{0,05} \frac{3,93}{\sqrt{104}} \\ 4,74 - 1,983 \frac{3,93}{\sqrt{104}} &< \mu < 4,74 + 1,983 \frac{3,93}{\sqrt{104}} \\ 3,974 &< \mu < 5,505 \end{aligned}$$

Interval kepercayaan untuk rata-rata waktu tunggu pada tingkat selang kepercayaan 95% menunjukkan bahwa rata-rata waktu tunggu dapat dipercayai berada antara 3,974 hingga 5,505 menit. Ini berarti bahwa jika data dikumpulkan dan dihitung ulang sebanyak banyak mungkin, 95% dari hasil akan berada dalam interval ini. Oleh karena itu, interval ini memberikan tingkat keyakinan tinggi pada perkiraan rata-rata waktu tunggu.

### 3.1.3. Waktu Sistem dan Idle Time

Berdasarkan statistika deskriptif dari data pengamatan, diperoleh rata-rata waktu sistem adalah 14,32 menit. Dengan nilai median waktu sistem adalah 13,5 menit. Sedangkan modus waktu sistem adalah 20 menit. Berdasarkan statistika deskriptif dari data pengamatan, diperoleh bahwa rata-rata, median, modus, dan standar deviasi semuanya bernilai 0. Ini menunjukkan bahwa tidak ada waktu kosong pada saat pelayanan. Range, minimum, dan maksimum juga bernilai 0, yang menunjukkan bahwa waktu kekosongan di indomaret kurang dari 1 menit. Hal ini menandakan bahwa indomaret tersebut cenderung ramai dan padat pengunjung.

### 3.2. Perbandingan Hasil Data Pengamatan dengan Data Acak

Perbandingan tingkat kepadatan pengunjung dari data real dan data acak yang diambil dari probabilitas data real bisa saja sama, tergantung bagaimana pembangkitan data acak dilakukan. Perbandingan antara real data dengan data yang dibangkitkan dengan bilangan acak memungkinkan untuk membandingkan kinerja sistem antrian sebenarnya dengan hipotesis yang dibangun. Data real menunjukkan bagaimana sistem antrian benar-benar beroperasi, sedangkan data yang dibangkitkan dengan bilangan acak memperkirakan bagaimana sistem antrian akan beroperasi jika berjalan secara acak. Jika data real dan data yang dibangkitkan dengan bilangan acak mirip, ini menunjukkan bahwa sistem antrian memiliki distribusi waktu yang baik dan tidak terlalu dipengaruhi oleh variasi eksternal. Jika terdapat perbedaan yang signifikan antara kedua data, ini menunjukkan bahwa sistem antrian memiliki masalah yang perlu diatasi.

Pada parameter idle time, ditemukan bahwa tidak ada waktu kosong pada saat pelayanan baik pada data real maupun pada data yang dibangkitkan dengan bilangan acak. Arti dari tidak ada waktu kosong pada parameter idle time adalah bahwa tidak ada suatu saat dimana pelayanan tidak melakukan pelayanan. Ini berarti bahwa pelayanan selalu dalam keadaan bekerja dan tidak pernah mengalami kekosongan waktu selama proses pelayanan. Oleh karena itu, data yang dibangkitkan dengan bilangan acak dapat digunakan sebagai alternatif yang memadai dalam melakukan simulasi atau analisis antrian yang merepresentasikan data real.

## IV. KESIMPULAN

Berdasarkan analisis, selang kepercayaan error 5% menunjukkan bahwa durasi pelayanan rata-rata yang dipercaya adalah antara 4,48 hingga 5,128 menit. Interval kepercayaan 95% untuk rata-rata waktu tunggu adalah  $3,974 < \mu < 5,505$  dan interval kepercayaan 95% untuk rata-rata waktu sistem adalah  $8,594 < \mu < 10,463$ . Hasil perbandingan tingkat kepadatan pengunjung antara data real dan data acak pada penelitian ini relatif sama. Hal ini menunjukkan bahwa distribusi waktu sistem antrian baik dan tidak terlalu dipengaruhi oleh variasi luar. Parameter idle time menunjukkan bahwa tidak ada waktu kosong pada data real maupun data acak. Ini berarti bahwa pelayanan selalu bekerja dan tidak pernah mengalami waktu kosong selama proses pelayanan.

Berdasarkan hasil analisis teori antrian pada Indomaret Darmo Surabaya, terdapat beberapa saran yang dapat diberikan. Berdasarkan hasil analisis rata-rata estimasi durasi pelayanan, dapat ditemukan tingkat error yang diterima oleh sistem. Jika error melebihi toleransi yang ditentukan, maka perlu



ditempuh tindakan untuk memperbaiki proses pelayanan dan meminimalkan durasi pelayanan. Oleh karena itu, durasi pelayanan di indomaret Darmo dapat lebih dioptimalkan lagi. Selain itu, hasil analisis interval kepercayaan untuk waktu tunggu dapat digunakan untuk menentukan bagaimana konsep ini mempengaruhi pengalaman pelanggan. Dari sini, dapat diambil tindakan untuk memperbaiki waktu tunggu untuk memberikan pengalaman yang lebih baik bagi pelanggan. Dengan mempertimbangkan saran di atas dan memahami hasil analisis, Indomaret Darmo Surabaya dapat melakukan tindakan untuk meningkatkan pelayanan dan memastikan pengalaman pelanggan yang memuaskan.

**UCAPAN TERIMA KASIH****REFERENSI**

1. Putri, W. S., & Ahmad, D. (2020). Penerapan Teori Antrian pada Pelayanan Teller Bank BNI Kantor Cabang Pembantu Air Tawar. *Journal of Mathematics UNP*, 5(1).
2. Dimiyati, T. T., & Dimiyati, A. (2009). *Operation Research: Model-model Pengambilan Keputusan*.
3. Puspita, Nia Sari. 2017. “Penerapan Teori Antrian pada Pelayanan Teller Bank X.
4. Farkhan, F., Hendikawati, P., & Arifudin, R. (2013). Aplikasi teori antrian dan simulasi pada pelayanan teller bank. *Unnes Journal of Mathematics*, 2(1).
5. Hardiyani, R. (2013). Analisis Penerapan Teori Antrian Pada Sistem Pembayaran Supermarket Di Golden Market Jember.
6. Pellondou, E. H., Fanggidae, R. P., & Nyoko, A. E. (2021). Analisis Teori Antrian Pada Jalur Sepeda Motor Stasiun Pengisian Bahan Bakar Umum (SPBU) Oebobo. *GLORY: Jurnal Ekonomi & Ilmu Sosial*, 2(1-Mar), 19-31.
7. Suprianto, Muhammad, & Indriani, A. (2019). Penerapan Metode Simulasi Kejadian Diskrit Pada Layanan Antar Pesan. 205.



# Regresi Data Panel untuk Memodelkan Tingkat Kemiskinan di Provinsi Papua

Dwi Rahmadini<sup>1</sup>, Ina Alfidayanti<sup>2</sup>, M. Al Haris<sup>3</sup>

<sup>1,2,3</sup>Afiliasi Program Studi Statistika, Universitas Muhammadiyah Semarang, Jawa Tengah

<sup>1</sup>[dwirahmadini21@gmail.com](mailto:dwirahmadini21@gmail.com)

<sup>2</sup>[inaalfida7@gmail.com](mailto:inaalfida7@gmail.com)

<sup>3</sup>[alharis@unimus.ac.id](mailto:alharis@unimus.ac.id)

Corresponding author email: [alharis@unimus.ac.id](mailto:alharis@unimus.ac.id)

**Abstract:** The purpose of this research was to model and determine the factors that influence the poverty rate in the Province of Papua from 2018 to 2020 by applying panel data analysis. The independent variables in this research were life expectancy, average length of schooling, open unemployment rate, and labor force participation rate. The estimation model used was the common effect model (CEM), fixed effect model (FEM) and random effect model (REM). The best model to determine the factors that influence the poverty rate in the Province of Papua was Random Effect Model and (REM) and the independent variable that has a significant effect on the poverty rate in Papua Province in 2018-2020 was the average length of school.

**Keywords:** The poverty rate, Papua Province, Panel Data Analysis

**Abstrak:** Tujuan dari penelitian ini adalah untuk memodelkan dan menentukan faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Papua dari tahun 2018 hingga 2020 dengan menerapkan analisis data panel. Variabel independen dalam penelitian ini adalah angka harapan hidup, rata-rata lama sekolah, tingkat pengangguran terbuka, dan tingkat partisipasi angkatan kerja. Model estimasi yang digunakan adalah *Common Effect Model* (CEM), *Fixed Effect Model* (FEM), dan *Random Effect Model* (REM). Model terbaik untuk menentukan faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Papua adalah *Random Effect Model* (REM). Berdasarkan model terbaik diketahui bahwa variabel independen yang memiliki pengaruh signifikan terhadap tingkat kemiskinan di Provinsi Papua pada tahun 2018-2020 adalah rata-rata lama sekolah.

**Kata kunci:** Analisis Data Panel, Angka Tingkat kemiskinan, Provinsi Papua

## I. PENDAHULUAN

Salah satu permasalahan yang tetap menjadi isu penting bagi negara berkembang adalah kemiskinan, seperti halnya di Indonesia [1], [2]. Menurut sumber informasi Badan Pusat Statistik (2008), kemiskinan dianggap sebagai masalah yang kompleks karena berkaitan erat dengan setiap aspek kehidupan, baik itu sosial, ekonomi, budaya, atau lain-lain [3]. Setiap negara memiliki definisi sendiri tentang apa itu miskin atau tidak miskin, karena kemiskinan itu termasuk kondisi yang bersifat relatif bagi masing-masing negara. Oleh karena itu, pengentasan kemiskinan harus dilakukan dengan baik, dengan melibatkan seluruh aspek kehidupan masyarakat. Upaya penanganan kemiskinan ditujukan untuk mencapai tujuan nasional yaitu mewujudkan masyarakat adil dan makmur [4], [5].

Pada September tahun 2015, Perserikatan Bangsa Bangsa (PBB) merilis *Sustainable Development Goals* (SDGs) yang merupakan tujuan untuk aksi global dari tahun 2016 hingga 2030 yang terdiri dari 17 tujuan dan 169 target. Disebutkan dalam SDGs bahwa tidak ada kemiskinan (*no poverty*) yang menjadi poin utama [6]. Artinya dunia telah sepakat untuk mengakhiri kemiskinan dalam bentuk apapun, termasuk di Indonesia. Angka kemiskinan terparah di Indonesia terjadi di kawasan timur Indonesia. Menurut World Bank tahun 2007, setelah mengendalikan karakteristik lain. Jumlah penduduk miskin di kawasan timur Indonesia masih tinggi, terutama di Provinsi Papua [7].

Papua merupakan salah satu provinsi dengan tingkat kemiskinan tertinggi secara nasional. Sebagian besar penduduk miskin Papua tinggal di pedesaan, bahkan lebih dari sepertiga penduduk pedesaan di Papua hidup dalam kemiskinan. Menurut Badan Pusat Statistik (BPS), jumlah penduduk yang hidup di bawah garis kemiskinan di Provinsi Papua pada September 2021 ada 944.490 jiwa atau

27,38% dari total penduduk. Dibandingkan dengan 33 provinsi lainnya, proporsi ini merupakan yang tertinggi. Dari tempat tinggal, terdapat 895.260 jiwa di pedesaan Papua hidup di bawah garis kemiskinan. Angka tersebut mencapai 36,5% dari total penduduk [3]. Ini berarti lebih dari sepertiga penduduk di pedesaan Papua hidup dalam kemiskinan. Sementara itu, jumlah penduduk miskin di daerah perkotaan Papua berjumlah 49.230 jiwa atau 4,94% dari total penduduk. Dari data tersebut menunjukkan ketimpangan penduduk miskin di pedesaan dan perkotaan Papua sangat jelas. Oleh karena itu, dibutuhkan kehati-hatian dalam memperbaiki kondisi masyarakat pedesaan di Provinsi Papua [7].

Tingginya angka kemiskinan tentunya tidak terlepas dari faktor-faktor yang mempengaruhinya. Salah satu metode yang paling tepat untuk menganalisis faktor-faktor yang mempengaruhi kemiskinan di Papua adalah regresi data panel [2], [4], [8]. Regresi data panel mengembangkan analisis regresi yaitu dengan menggabungkan data *cross-sectional* dan data deret waktu. Pada umumnya, regresi data panel berguna untuk mengamati data yang diteliti secara terus menerus selama periode waktu. Manfaat lain dari analisis regresi data panel yaitu lebih memaknai keseluruhan data deret waktu, dengan pertimbangan keragaman yang terjadi dalam unit *cross-sectional* [9].

Pada penelitian sebelumnya, diketahui beberapa faktor telah terbukti berpengaruh signifikan terhadap tingkat kemiskinan. Sebagai contoh, penelitian yang dilakukan oleh [10] pada tahun 2018 tentang “Analisis Regresi Data Panel Pada Faktor-faktor yang Mempengaruhi Tingkat Kemiskinan Provinsi Sulawesi Selatan Tahun 2011–2015”, memberikan kesimpulan bahwa rata-rata lama sekolah dan laju pertumbuhan penduduk berpengaruh signifikan terhadap tingkat kemiskinan di Provinsi Sulawesi Selatan. Sedangkan penelitian yang dilakukan oleh [11] pada tahun 2012 yaitu tentang “Analisis Data Panel Model Efek Acak pada Data Kemiskinan di Provinsi Sulawesi Selatan”, memberikan kesimpulan bahwa tingkat pengangguran terbuka dan partisipasi angkatan kerja berpengaruh signifikan terhadap tingkat kemiskinan di Provinsi Sulawesi Selatan.

Metode yang tepat untuk menganalisis faktor-faktor yang mempengaruhi kemiskinan di Papua adalah regresi data panel. Dalam penelitian ini, data yang digunakan ditabulasikan dalam struktur data panel. Data *cross-sectional* dalam studi kasus ini berasal dari data kabupaten/kota di Provinsi Papua. Sedangkan data deret waktu adalah data antar waktu, yaitu data hasil kajian dalam periode tertentu.

## II. METODE PENELITIAN

### 2.1. Jenis Data dan Sumber Data Penelitian

Penelitian ini menggunakan data sekunder periode 2018 hingga 2020 untuk setiap Kabupaten/Kota di Provinsi Papua. Data ini bersumber dari Badan Pusat Statistik (BPS) Provinsi Papua. Variabel dependen ( $Y_{it}$ ) yang dipakai yaitu tingkat kemiskinan dengan variabel independennya yaitu terdiri atas angka harapan hidup ( $X_{1it}$ ), rata-rata lama sekolah ( $X_{2it}$ ), tingkat pengangguran terbuka ( $X_{3it}$ ) dan tingkat partisipasi angkatan kerja ( $X_{4it}$ ).

### 2.2. Tahapan analisis data

Analisis data yang dilakukan untuk memodelkan Tingkat Kemiskinan Provinsi Papua adalah sebagai berikut:

#### 2.2.1. Mengestimasi model regresi data panel yang terdiri dari *Common Effect Model* (CEM), *Fixed Effect Model* (FEM), dan *Random Effect Model* (REM).

*Common Effect Model* (CEM) mengasumsikan bahwa koefisien *slope* dan intersep konstan antar individu dan waktu. menggunakan metode *Ordinary Least Squares* untuk melakukan estimasi parameter [12]. bentuk persamaan model adalah sebagai berikut:

$$Y_{it} = \alpha + \sum_{k=1}^K \beta_k X_{it} + u_{it};$$

dengan  $i = 1, 2, 3, \dots, N$ ,  $t = 2018, 2019, 2020, \dots, T$ ,  $\alpha$  adalah intersep,  $K$  adalah jumlah variabel independen,  $\beta$  adalah *slope* berukuran  $(K \times 1)$ , dan  $u$  adalah slope berukuran  $(N \times T) \times 1$ .

*Fixed Effect Model* (FEM) adalah model yang menduga bahwa intersep bersifat tidak konstan, sementara koefisien *slope* diasumsikan tetap antar unit *cross section*. Perbedaan intersep mengakibatkan adanya *dummy* variabel pada model [13]. FEM membentuk dua model yaitu *fixed individual effect* yang *dummy* variabelnya terdiri atas 29 kabupaten/kota, dengan bentuk persamaan model adalah sebagai berikut:

$$Y_{it} = \sum_{j=1}^N D_{jt} \alpha_j + \sum_{k=1}^K \beta_k X_{kit} + u_{it};$$

dengan  $D_{jt}$  bernilai 1 jika  $j = i$ . Sedangkan model *fixed time effect dummy* variabelnya terdiri dari tahun 2018; 2019; dan 2020. Bentuk persamaan model adalah:

$$Y_{it} = \sum_{s=1}^N D_{is} \alpha_s + \sum_{k=1}^K \beta_k X_{kit} + u_{it};$$

dengan  $s = 1, 2, 3, \dots, N$ ;  $t = 2018, 2019, 2020$  dengan  $D_{is}$  bernilai 1 jika  $s = t$ .

*Random Effect Model* (REM) digunakan untuk menangani permasalahan pada FEM. Penambahan *dummy variabel* pada model *fixed effect* akan menghalangi model aslinya, sehingga terdapat penambahan variabel error term pada model. Bentuk persamaan model adalah [14], [15]:

$$Y_{it} = \alpha_i + \sum_{k=1}^K \beta_k X_{kit} + w_{it};$$

dengan  $i = 1, 2, \dots, N$ ;  $t = 2018, 2019, 2020$ ,  $w_{it}$  merupakan gabungan galat yaitu  $u_{it}$  dan  $\varepsilon_i$ .

2.2.2. Melakukan uji Chow, uji Hausman dan uji Lagrange Multiplier untuk memilih model terbaik.

a. Uji *Chow*

Uji *Chow* digunakan untuk menentukan antara *Fixed Effect Model* atau *Common Effect Model* yang sebaiknya digunakan [12]. Uji Hipotesis uji *Chow* adalah sebagai berikut:

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_n = 0$  (*Common Effect Model*)

$H_1 : \text{minimal ada satu } \alpha_i \neq 0; i = 1, 2, 3, \dots, n$  (*Fixed Effect Model*)

Statistik uji yang digunakan merupakan uji F, yaitu:

$$F_{hitung} = \frac{(RSS_1 - RSS_2)/(N-1)}{RSS_2/(NT-N-K)}$$

dengan

$N$  : total individu (data cross section)

$T$  : total periode waktu (data time series)

$K$  : total variabel penjelas

$RSS_1$  : *residual sum of squares* hasil pendugaan *Common Effect Model*

$RSS_2$  : *residual sum of squares* hasil pendugaan *Fixed Effect Model*

b. Uji *Hausman*

Uji *Hausman* digunakan untuk menguji apakah ada hubungan antara kesalahan dalam model (kesalahan gabungan) pada satu atau lebih variabel penjelas dalam model (Baltagi, 2015).

$H_0$ : Korelasi  $(X_{it}, \varepsilon_{it}) = 0$  (Model efek acak)

$H_1$ : Korelasi  $(X_{it}, \varepsilon_{it}) \neq 0$  (Model efek tetap)

Menggunakan statistik uji *Chi-Square* berdasarkan kriteria *Wald*, yaitu:

$$W = (\hat{\beta}_{MET} - \hat{\beta}_{MEA})' [\text{var}(\hat{\beta}_{MET} - \hat{\beta}_{MEA})]^{-1} (\hat{\beta}_{MET} - \hat{\beta}_{MEA})$$

dengan:

$\hat{\beta}_{MET}$  : vektor estimasi kemiringan model efek tetap.

$\hat{\beta}_{MEA}$  : vektor estimasi kemiringan model efek acak.

c. Uji *Lagrange Multiplier*

Uji *Lagrange Multiplier* digunakan untuk menentukan apakah model efek acak (REM) lebih baik daripada model efek umum (CEM), digunakan uji *Lagrange Multiplier* (LM). Uji signifikansi model acak ini dikembangkan oleh *Breusch-Pagan* [16]. Pengujian ini didasarkan

pada nilai residu dari metode efek umum. Hipotesis Uji LM *Breusch-Pagan* adalah sebagai berikut:

$$H_0: \sigma_e^2 = 0 \text{ (Model efek umum)}$$

$$H_1: \text{minimal ada } \sigma_e^2 \neq 0 \text{ (Model efek acak)}$$

Statistik uji berdasarkan distribusi *Chi-Squares* yaitu dengan rumus sebagai berikut:

$$LM = \frac{NT}{2(T-1)} \left[ \frac{\sum_{i=1}^N [T\bar{e}]^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right]^2$$

2.2.3. Melakukan uji signifikansi parameter pada model regresi terpilih untuk data panel menggunakan uji-F (simultan) dan uji-T (parsial).

Uji-F digunakan untuk menentukan apakah model data panel signifikan dan layak dianalisis lebih lanjut atau tidak. Uji-F mengasumsikan bahwa gangguan  $\varepsilon_i$  mengikuti distribusi normal. Statistik uji yang digunakan diformulasikan sebagai berikut:

$$F_{hitung} = \frac{(SSE_P - SSE_{DV}) / (N - 1)}{(SSE_{DV}) / (NT - N - 1)}$$

Uji T digunakan untuk menguji signifikansi variabel bebas  $X_j$  terhadap variabel terikat Y (dengan asumsi variabel bebas lainnya konstan), berarti  $\beta_j = 0$ . Sehingga rumusnya adalah sebagai berikut:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, j = 0, 1, 2, \dots, k \text{ (k merupakan koefisien slope)}$$

saat menguji hipotesis koefisien regresi parsial ( $\beta_j$ ), digunakan statistik uji-t sebagai berikut:

$$t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$$

untuk menguji apakah  $\beta_j$  sama dengan 0 ( $H_0: \beta_j = 0$ ), maka nilai  $\beta_j$  pada persamaan diganti dengan nol sehingga persamaan uji-t menjadi:

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

persamaan tersebut mengikuti distribusi t dengan derajat kebebasan sebesar  $(n - k)$ .

2.2.4. Melakukan interpretasi model yang diperoleh.

### III. HASIL DAN PEMBAHASAN

#### 3.1. Pemodelan *Common Effect Model* (CEM)

Hasil estimasi parameter model CEM pada pemodelan faktor-faktor yang mempengaruhi kemiskinan di Papua disajikan pada Tabel 1.

**Tabel 1.** Estimasi parameter model CEM

Variabel	Parameter	Nilai Estimasi	Nilai t hitung	Nilai p
Intersep	$\alpha$	43,536	2,83	<b>0,006</b>
AHH	$\beta_1$	0,131	0,601	0,549
RLS	$\beta_2$	-2,236	-4,65	<b><math>1,2 \times 10^{-5}</math></b>
TPT	$\beta_3$	-0,610	-1,50	0,137
TPAK	$\beta_4$	-0,095	-1,42	0,160
Nilai F-hitung				27,727
Nilai p				$1,4 \times 10^{-14}$
Nilai $R^2$				0,575

#### 3.2. Pemodelan *Fixed Effect Model* (FEM)

Hasil estimasi parameter model FEM pada pemodelan faktor-faktor yang mempengaruhi kemiskinan di Papua disajikan pada Tabel 2.

**Tabel 2.** Estimasi parameter model FEM

Variabel	Parameter	Nilai Estimasi	Nilai t hitung	Nilai p
AHH	$\beta_1$	-1,824	-0,733	0,4667

RLS	$\beta_2$	-2,332	-0,551	0,5840
TPT	$\beta_3$	-0,469	-0,095	0,2781
TPAK	$\beta_4$	-0,015	0,334	0,7399
Nilai F-hitung				1,5724
Nilai $p$				0,1948
Nilai $R^2$				0,1043

### 3.3. Pemodelan *Random Effect Model* (REM)

Hasil estimasi parameter model REM pada pemodelan faktor-faktor yang mempengaruhi kemiskinan di Papua disajikan pada Tabel 3.

**Tabel 3.** Estimasi parameter model REM

Variabel	Parameter	Nilai Estimasi	Nilai t hitung	Nilai $p$
Intersep	$\alpha$	36,564	1,671	0,0947
AHH	$\beta_1$	0,098	0,285	0,7754
RLS	$\beta_2$	-1,954	-3,630	<b>0,0002</b>
TPT	$\beta_3$	-0,596	-1,624	0,1043
TPAK	$\beta_4$	0,002	0,037	0,1043
Nilai F-hitung				46,4
Nilai $p$				$2,02 \times 10^{-9}$
Nilai $R^2$				0,3614

### 3.4. Pemilihan model regresi data panel

#### a. Uji Chow

Uji *Chow* digunakan untuk menentukan antara *Fixed Effect Model* atau *Common Effect Model* yang sebaiknya digunakan [12]. Uji Hipotesis uji *Chow* adalah sebagai berikut:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_n = 0 \text{ (Common Effect Model)}$$

$$H_1: \text{minimal ada satu } \alpha_i \neq 0; i = 1, 2, 3, \dots, n \text{ (Fixed Effect Model)}$$

Berdasarkan hasil perhitungan diperoleh nilai  $p = 3,9 \times 10^{-12} < \alpha (0,05)$  yang menyebabkan  $H_0$  ditolak. Hasil tersebut dapat disimpulkan bahwa setidaknya ada satu kabupaten/kota yang memiliki perbedaan intersep yang berbeda, sehingga model yang sesuai untuk mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan di Papua yaitu model FEM.

#### b. Uji Hausman

Uji Hausman digunakan untuk membandingkan FEM dengan REM, dengan asumsi sebagai berikut:

$$H_0: \text{Korelasi } (X_{it}, \varepsilon_{it}) = 0 \text{ (Model REM terbaik)}$$

$$H_1: \text{Korelasi } (X_{it}, \varepsilon_{it}) \neq 0 \text{ (Model FEM terbaik)}$$

Berdasarkan hasil perhitungan, diperoleh nilai  $p = 0,2416 > \alpha (0,05)$  yang menyebabkan tidak cukup bukti menolak  $H_0$ , yang berarti bahwa model REM konsisten. Oleh karena itu dapat disimpulkan model REM adalah model yang lebih baik daripada model FEM. Karena hasil uji *chow* dan uji *hausman* menunjukkan hasil yang berbeda, maka akan dilanjutkan dengan uji *Lagrange Multiplier* (LM).

#### c. Uji *Lagrange Multiplier*

Pengujian ini dilakukan untuk mengetahui apakah model REM lebih baik dari model CEM, dengan asumsi sebagai berikut:

$$H_0: \sigma_e^2 = 0 \text{ (Model CEM terbaik)}$$

$$H_1: \sigma_e^2 \neq 0 \text{ (Model REM terbaik)}$$

Hasil perhitungan diperoleh nilai  $p = 4,49 \times 10^{-11} < \alpha (0,05)$  yang menyebabkan  $H_0$  ditolak. Hal tersebut dapat disimpulkan bahwa model REM merupakan model terbaik.

### 3.5. Pengujian Signifikansi Model

Penelitian ini menggunakan dua pengujian signifikansi, yaitu uji-F (simultan) dan Uji-T (parsial). Pada uji F dengan memperhatikan model yang terpilih yaitu model REM yang ditunjukkan pada Tabel 3, diperoleh nilai  $p = 2,03 \times 10^{-9} < \alpha (0,05)$  atau disimpulkan  $H_0$

ditolak. Hal tersebut menunjukkan bahwa terdapat variabel bebas yang berpengaruh terhadap variabel respon.

Pada pengujian menggunakan uji T (parsial) dengan model REM yang ditunjukkan pada Tabel 3, diketahui bahwa hanya variabel rata-rata lama sekolah (RLS) yang memiliki nilai  $\rho = 0,000 < \alpha (0,05)$  atau menolak  $H_0$ . Sehingga dapat disimpulkan hanya variabel rata-rata lama sekolah (RLS) yang berpengaruh signifikan terhadap tingkat kemiskinan di Provinsi Papua. Model REM yang terbentuk adalah sebagai berikut:

$$Y_{it} = \alpha_i + 0,0987X_{1it} - 1,9545X_{2it} - 0,5966X_{3it} + 0,0015X_{4it} + W_{it};$$

dengan  $i = 1,2,\dots,N$  dan  $t = 2018,2019,2020$ . Nilai intersep ( $\alpha_i$ ) setiap kabupaten/kota dapat dilihat pada Tabel 4.

**Tabel 4.** Nilai Intersep Setiap Kabupaten/Kota

Parameter	Nilai Intersep
$\alpha_{Asmat}$	-5.05351998
$\alpha_{Biak Numfor}$	6.56933430
$\alpha_{Boven Digoel}$	-2.80854620
$\alpha_{Deiyai}$	5.32532551
$\alpha_{Dogiyai}$	-2.83727796
$\alpha_{Jayapura}$	-4.46550734
$\alpha_{Jayawijaya}$	6.29689843
$\alpha_{Keerom}$	-7.84673629
$\alpha_{Kepulauan Yapen}$	4.19997522
$\alpha_{Kota Jayapura}$	-4.46550734
$\alpha_{Mambero Tengah}$	-0.05488321
$\alpha_{Mambero Raya}$	-0.53796941
$\alpha_{Mappi}$	-2.46696434
$\alpha_{Merauke}$	-13.11319851
$\alpha_{Mimika}$	-4.66850109
$\alpha_{Nduga}$	-1.81239131
$\alpha_{Paniai}$	2.53727815
$\alpha_{Pegunungan Bintang}$	-5.47078849
$\alpha_{Puncak}$	-1.03066952
$\alpha_{Puncak Jaya}$	-1.81239131
$\alpha_{Sarmi}$	0.75943187
$\alpha_{Talikara}$	-2.58594307
$\alpha_{Weropen}$	6.27627411
$\alpha_{Yakuhimo}$	4.14995733
$\alpha_{Yalimo}$	-2.98394100
$\alpha_{Intan Jaya}$	4.04867628
$\alpha_{Lanny Jaya}$	2.34797812
$\alpha_{Nabire}$	3.95826323
$\alpha_{Sapiro}$	13.72521378

Mengacu pada model efek acak (REM), diperoleh nilai koefisien determinasi sebesar 0,36139. Oleh karena itu, dapat disimpulkan bahwa keempat variabel tersebut hanya dapat menjelaskan tingkat kemiskinan sebesar 36,14% sedangkan sisanya dijelaskan oleh variabel lain yang tidak dijelaskan dalam model.

#### IV. KESIMPULAN

Berdasarkan pengujian yang telah dilakukan, dapat diketahui bahwa model terbaik untuk memodelkan faktor-faktor yang mempengaruhi kemiskinan di Papua yaitu *Random Effect Model*



(REM) dengan  $R^2$  sebesar 0,36139. Hal ini menunjukkan bahwa variabel bebas dalam model dapat menjelaskan 36,14% angka kemiskinan di Provinsi Papua tahun 2018-2020, sedangkan sisanya dijelaskan oleh variabel lain yang belum teruji. Faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Papua adalah rata-rata lama sekolah (RLS).

### UCAPAN TERIMA KASIH

### REFERENSI

1. E. N. Amaliah, Darnah, and Sifriyani, “Regresi Data Panel dengan Pendekatan Common Effect Model (CEM), Fixed Effect model (FEM) dan Random Effect Model (REM) (Studi Kasus: Persentase Penduduk Miskin Menurut Kabupaten/Kota di Kalimantan Timur Tahun 2015-2018),” *ESTIMASI J. Stat. Its Appl.*, vol. 1, no. 2, pp. 106–115, 2020, doi: 10.20956/ejsa.v1i2.10574.
2. Y. Hikmah, “Pemodelan Panel Spasial pada Data Kemiskinan di Provinsi Papua,” *Stat. J. Theor. Stat. Its Appl.*, vol. 17, no. 1, pp. 1–15, 2017, doi: 10.29313/jstat.v17i1.2318.
3. Badan Pusat Statistik, “PENDAPATAN NASIONAL (National Income of Indonesia),” 2020.
4. R. N. Tri Wahyuni and A. Damayanti, “Faktor-Faktor yang Menyebabkan Kemiskinan di Provinsi Papua: Analisis Spatial Heterogeneity,” *J. Ekon. dan Pembang. Indones.*, vol. 14, no. 2, pp. 128–144, 2014, doi: 10.21002/jepi.v14i2.441.
5. Y. S. Barlinti, “Kebijakan-Kebijakan Pemerintah dalam Penanggulangan Kemiskinan,” *Lex Jurnalica*, vol. 4, no. 3, pp. 158–167, 2007.
6. Kementerian PPN, “Pedoman Teknis Penyusunan Rencana Aksi - Tujuan Pembangunan Berkelanjutan/ Sustainable Development Goals (TPB/SDGs),” *Kementerian PPN*, p. 114, 2020.
7. V. B. Kusnandar, “Lebih dari Sepertiga Penduduk Perdesaan Papua Hidup Miskin pada September 2021,” *Databoks.Katadata.Co.Id*, no. September, p. 2021, 2022.
8. [8] Marlina and U. Usman, “Pengaruh PDRB dan Pengeluaran Pemerintah Provinsi Papua Terhadap Kemiskinan di Papua,” *J. Ekon. Reg. Unimal*, vol. 3, no. 2, pp. 15–22, 2020, doi: 10.29103/jeru.v3i2.3202.
9. A. Indrasetianingsih and T. K. Wasik, “Model Regresi Data Panel untuk Mengetahui Faktor yang Mempengaruhi Tingkat Kemiskinan di Pulau Madura,” *J. Gaussian*, vol. 9, no. 3, pp. 355–363, 2020, doi: 10.14710/j.gauss.v9i3.28925.
10. S. Royat, “Analisis Regresi Data panel pada Faktor-Faktor yang Mempengaruhi Tingkat kemiskinan Provinsi Sulawesi Selatan Tahun 2011-2015,” Universitas Islam Negeri Alauddin Makassar, 2027.
11. A. Anisa, N. Ilyas, and H. Hadijah, “Analisis Data Panel Model Efek Acak pada Data Kemiskinan di Provinsi Sulawesi Selatan,” *J. Mat. Stat. dan Komputasi*, vol. 8, no. 2, pp. 110–130, 2012.
12. B. H. Baltagi, *The Oxford handbook of panel data*. Oxford Handbooks, 2015.
13. D. N. Gujarati and D. C. Porter, *Basic Econometrics*, 4th ed. New York: McGraw-Hill/Irwin, 2009.
14. S. P. Rahayu and N. Pratiwi, “Aplikasi Metode Regresi Spasial Data Panel pada Penerimaan Daerah Sektor Pariwisata di Kabupaten/Kota di Daerah Istimewa Yogyakarta,” *J. Stat. Ind. dan Komputasi*, vol. 7, no. 1, pp. 13–23, 2022.
15. D. Venosia, Suliyanto, Sediono, and Nur Chamidah, “Pemodelan Persentase Kepesertaan BPJS Non Penerima Bantuan Iuran Dengan Pendekatan Regresi Data Panel,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 1, pp. 116–126, 2022, doi: 10.36456/jstat.vol15.no1.a4863.
16. A. Widarjono, *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*, 2nd ed. 2007.



# Analisis Kesamaan Rata-rata Indeks Harga Konsumen (IHK) Menurut Sub Kelompok Makanan dengan Sub Kelompok Pendidikan

Melya Vebryanti<sup>1</sup>, Wahyu Melinda Permanasari<sup>2</sup>, Trimono<sup>3</sup>

<sup>1, 2, 3</sup>Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>22083010063@student.upnjatim.ac.id

<sup>2</sup>22083010043@student.upnjatim.ac.id

<sup>3</sup>trimono.stat@upnjatim.ac.id

Corresponding author email: 22083010043@student.upnjatim.ac.id

**Abstract:** *The Consumer Price Index (CPI) is one of the indicators used to measure the level of inflation in a country or region. According to the Central Statistics Agency (BPS), the CPI is used to calculate the average price changes during a specific period of time for a set of goods and services used by the population and households. The CPI typically provides information about price trends for a group of goods or services consumed by households at a particular point in time. An increase in the CPI for the food subgroup will have a significant impact on food prices. This will subsequently affect the cost of living. Meanwhile, education is an important foundation for building a society with a high level of knowledge.*

*Next, a comparison of the average CPI between the food subgroup and the education subgroup is conducted. This comparison aims to determine whether there is social inequality between the two subgroups. The Mean Test Two Populations is used for the best results. In the results section, it can be concluded that the Consumer Price Index (CPI) for the food and education subgroups has different means, indicating that the inflation rates for the two subgroups are different, and there is indeed social inequality between them.*

**Keywords:** *The Consumer Price Index, food subgroup, Education subgroup, Mean Test Two Populations*

**Abstrak:** Indeks Harga Konsumen (IHK) adalah salah satu indikator untuk mengukur tingkat inflasi dalam suatu negara atau wilayah. Menurut Badan Pusat Statistik (BPS), Indeks Harga Konsumen (IHK) digunakan untuk menghitung perubahan harga rata-rata selama periode waktu tertentu dari kumpulan barang dan jasa yang digunakan oleh penduduk dan rumah tangga selama periode waktu tertentu. IHK biasanya memberikan informasi tentang tren harga untuk sekelompok barang atau jasa yang dikonsumsi oleh rumah tangga pada suatu titik waktu tertentu. Kenaikan IHK pada sub kelompok makanan akan berdampak pada kenaikan harga-harga makanan secara signifikan. Ini nantinya akan berpengaruh pada kenaikan biaya hidup. Lalu pendidikan merupakan pondasi penting dalam membangun masyarakat yang memiliki pengetahuan yang tinggi. Selanjutnya dilakukan perbandingan rata-rata antara IHK pada sub kelompok makanan dan sub kelompok pendidikan. Dilakukannya perbandingan tersebut untuk melihat ada atau tidaknya ketimpangan sosial antara dua sub kelompok tersebut. Untuk hasil yang terbaik, maka digunakan metode Uji Mean Dua Populasi. Pada bagian hasil, dapat disimpulkan bahwa Indeks Harga Konsumen (IHK) pada sub kelompok makanan dan pendidikan memiliki rata-rata yang berbeda yang berarti tingkat inflasi pada kedua sub kelompok tersebut berbeda dan memang terdapat ketimpangan sosial antara keduanya.

**Kata kunci:** Indeks Harga Konsumen, Sub Kelompok Makanan, Sub Kelompok Pendidikan, Uji Mean Dua Populasi

## I. PENDAHULUAN

Indeks Harga Konsumen (IHK) adalah salah satu indikator untuk mengukur tingkat inflasi dalam suatu negara atau wilayah. Menurut Badan Pusat Statistik (BPS), Indeks Harga Konsumen (IHK) digunakan untuk menghitung perubahan harga rata-rata selama periode waktu tertentu dari kumpulan barang dan jasa yang digunakan oleh penduduk dan rumah tangga selama periode waktu tertentu. IHK biasanya memberikan sebuah informasi tentang tren harga untuk sekelompok barang atau jasa yang dikonsumsi oleh masyarakat pada suatu titik waktu yang tertentu.[1]



Indeks Harga Konsumen (IHK) merupakan salah satu indikator ekonomi yang digunakan untuk mengukur tingkat perubahan harga (inflasi/deflasi) di tingkat konsumen. Dengan adanya perubahan pola konsumsi masyarakat, maka mulai Januari 2022, pengukuran inflasi di Indonesia menggunakan IHK tahun dasar 2018=100. Beberapa perubahan mendasar dalam penghitungan IHK (2018=100) dibandingkan (IHK 2012=100), khususnya dari sisi cakupan, klasifikasi pengelompokan komoditas, metodologi penghitungan IHK, paket komoditas, dan diagram timbang. Perubahan tersebut didasarkan pada Survei Biaya Hidup (SBH) yang dilaksanakan oleh BPS selama tahun 2018, sebagai salah satu bahan dasar utama dalam penghitungan IHK. [2]

Di Indonesia terdapat Indeks Harga Konsumen (IHK), perkembangan IHK dapat memperlihatkan tingkat harga suatu barang dan jasa yang dibeli masyarakat. IHK bermanfaat untuk mengetahui tingkat kenaikan pendapatan, harga, juga dapat dijadikan sebagai indikator ekonomi dan tolak ukur besarnya biaya produksi. [3] IHK sendiri dibagi menjadi sembilan kelompok dan sub kelompok. Kelompok dan sub kelompok IHK mengacu pada klasifikasi kelompok barang dan jasa yang digunakan untuk menghitung IHK suatu negara atau wilayah.

IHK kelompok dan sub kelompok 1 makanan, minuman, dan tembakau, ini mengacu pada barang pokok maupun jasa yang dibutuhkan dalam kehidupan sehari-hari oleh masyarakat. Subkelompok makanan dan minuman mencakup harga bahan makanan, minuman, dan produk pertanian. Subkelompok tembakau mencakup harga rokok dan produk tembakau lainnya. Perubahan harga pada kelompok ini dapat memiliki dampak yang signifikan terhadap daya beli masyarakat, karena konsumsi barang-barang tersebut merupakan kebutuhan utama. Sedangkan IHK kelompok dan sub kelompok 9 pendidikan mencakup biaya pendidikan, buku, dan perlengkapan sekolah.

Makanan merupakan kebutuhan pokok yang memiliki peran penting dalam memenuhi kebutuhan dasar manusia. Kenaikan IHK pada kelompok ini akan berdampak pada kenaikan harga-harga makanan secara signifikan. Ini nantinya juga akan berpengaruh pada kenaikan biaya hidup. Lalu pendidikan merupakan kebutuhan pengembangan pada diri manusia yang memiliki peran penting dalam memenuhi kebutuhan dasar. Pendidikan merupakan pondasi penting dalam membangun masyarakat yang memiliki pengetahuan yang tinggi.

Pada penelitian ini digunakan metode statistika Uji Mean Dua Populasi untuk memprediksi dan melihat rata-rata IHK Kelompok dan Sub Kelompok makanan dengan rata-rata IHK Kelompok dan Sub Kelompok Pendidikan. Data yang digunakan dan akan dilakukan perbandingan adalah data yang dikumpulkan oleh Badan Pusat Statistik (BPS) yaitu data IHK Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau Tahun 2022 di 90 kota di Indonesia dan data IHK Kelompok dan Sub Kelompok 9 Pendidikan Tahun 2022 di 90 kota di Indonesia. Penelitian ini bertujuan untuk melihat apakah rata-rata IHK kedua kelompok dan sub kelompok tersebut sama. Dilakukannya penelitian ini juga untuk melihat apakah terdapat perbedaan yang signifikan antara rata-rata pada kedua sub kelompok tersebut, untuk menyimpulkan bahwa adanya ketimpangan sosial yang terkait aksesibilitas, kualitas, atau biaya dari makanan dan pendidikan.

Misalnya, jika rata-rata IHK sub kelompok makanan lebih tinggi daripada rata-rata IHK sub kelompok pendidikan, maka dapat diindikasikan bahwa kenaikan harga makanan relatif lebih tinggi daripada kenaikan biaya pendidikan. Dalam konteks tersebut, kelompok masyarakat memiliki kesulitan dalam memenuhi kebutuhan makanan dapat mengalami ketimpangan sosial yang lebih tinggi dibandingkan dengan kelompok yang dapat mengakses makanan dengan harga yang lebih terjangkau. Demikian pula, jika rata-rata IHK sub kelompok pendidikan lebih tinggi daripada rata-rata IHK sub kelompok makanan, hal tersebut dapat menunjukkan bahwa biaya pendidikan meningkat dengan lebih cepat daripada kenaikan harga makanan. Dalam situasi tersebut, kelompok masyarakat



yang memiliki kesulitan dalam mengakses pendidikan berkualitas mungkin akan menghadapi ketimpangan sosial yang lebih besar dibandingkan kelompok yang lebih mampu secara finansial.

## II. METODE PENELITIAN

### 2.1 Variabel yang digunakan

Variabel yang digunakan dalam penelitian ini adalah data IHK Kelompok dan Sub Kelompok 1 Makanan, Minuman, dan Tembakau Tahun 2022 dari 90 kota di Indonesia dan data IHK Kelompok dan Sub Kelompok 9 Pendidikan Tahun 2022 dari 90 kota di Indonesia. Dan yang akan dianalisis yaitu sub kelompok makanan dan sub kelompok pendidikan.

### 2.2 Metode Pengumpulan Data

Dalam metode pengumpulan data yang dilakukan dalam penyusunan artikel ini adalah Metode Literatur. Metode ini digunakan untuk mendapatkan data tentang Indeks Harga Konsumen (IHK) pada dua sub kelompok yaitu makanan dan pendidikan. Data yang digunakan dalam penelitian tersebut merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS).

### 2.3 Metode Analisis Data

Metode yang digunakan untuk menganalisis data Indeks Harga Konsumen pada artikel ini, penulis menggunakan metode Uji Mean Dua Populasi. Uji Mean Dua Populasi merupakan pengujian untuk melihat ada atau tidaknya perbedaan rata-rata dari dua sampel yang dibandingkan. Pada metode ini terdapat 3 step yang diujikan pada data yang akan dibandingkan :

#### 1. Uji Normalitas

Step ini merupakan pengujian pada data untuk melihat apakah data tersebut berdistribusi normal atau tidak. Pada pengujian ini digunakan Uji Normalitas Kolmogorov Smirnov. Menurut tinjauan pustaka yang dilakukan, mengutip Website Belajar Matematika dan Statistika [4]. Asumsi yang digunakan yaitu :

1. Data diambil dari populasi yang merupakan data dengan distribusi kontinu
2. Data diukur secara kuantitatif
3. Data bersifat independen

#### 1) Hipotesis pada Uji Normalitas ini antara lain :

- a.  $H_0$  : Data Berdistribusi Normal
- b.  $H_1$  : Data Tidak Berdistribusi Normal

#### 2) Tingkat Signifikansi

- a. Tingkat signifikansi yang digunakan adalah  $\alpha = 0,05$

#### 3) Statistik Uji

Pada statistik uji ini digunakan fungsi  $ks.test()$  yang dimana akan menghasilkan nilai p-value. Pengujian tersebut digunakan untuk menunjukkan seberapa konsistenkah data tersebut dengan hipotesis nol.

#### 4) Kaidah pengambilan keputusan :

1. Jika nilai p-value lebih besar dari tingkat signifikansi maka data penelitian tersebut berdistribusi normal
2. Jika nilai p-value lebih kecil dari tingkat signifikansi maka data penelitian tersebut tidak berdistribusi normal

#### 2. Uji Variansi dua Populasi

Misalkan terdapat dua populasi, yaitu  $X_1$  dengan rata-rata populasi  $\mu_1$  dan variansi populasi  $\sigma_1^2$ , dan populasi  $X_2$  dengan rata-rata populasi  $\mu_2$  dan variansi populasi  $\sigma_2^2$ . Dari kedua populasi diambil sampel masing-masing sebanyak  $n_1$  dan  $n_2$ . Dari kedua populasi tersebut dapat dilakukan uji untuk variansi keduanya sama atau berbeda. Asumsi yang digunakan yaitu:

1. Data sampel adalah data kontinu.
2. Masing-masing populasi berdistribusi Normal

Step ini merupakan pengujian pada kedua data yang dibandingkan untuk melihat apakah kedua data tersebut memiliki variansi populasi yang sama atau tidak.

- 1) Hipotesis pada Uji Variansi dua Populasi ini antara lain :
  - a.  $H_0$  : Variansi populasi 1 sama dengan variansi populasi 2
  - b.  $H_1$  : Variansi populasi 1 tidak sama dengan variansi populasi 2
- 2) Tingkat Signifikansi  
Tingkat signifikansi yang digunakan adalah  $\alpha = 0,05$
- 3) Statistik Uji

Statistik uji menggunakan nilai F hitung yang diperoleh melalui persamaan berikut :

$$F_{hitung} = \frac{S_1^2}{S_2^2}$$

- 4) Kaidah Pengambilan Keputusan  
 $H_0$  ditolak jika  $F_{hitung} > F_{tabel}$  atau  $F_{hitung} < -F_{tabel}$   
 $F_{tabel}$  adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas  $v_1 = n_1 - 1$  dan  $v_2 = n_2 - 1$ .
3. Uji Hipotesis Perbandingan Rata-Rata Dua Populasi, dengan Sampel yang Diambil Saling Bebas

Hal penting yang perlu dipahami pada uji ini adalah antara populasi 1 dan populasi 2, ukuran sampel yang diambil tidak harus sama. Misalkan akan diuji perbandingan rata-rata IHK pada kota-kota di Indonesia. Populasi yang diamati adalah IHK pada sektor makanan dan sektor pendidikan yang keduanya saling bebas. Asumsi yang digunakan adalah:

1. Sampel berasal dari populasi yang berdistribusi Normal/setiap populasi berdistribusi Normal
2. Data sampel adalah data kontinu

Step ini merupakan pengujian pada kedua data yang dibandingkan untuk menentukan apakah kedua data tersebut memiliki rata-rata populasi yang sama atau tidak.

- 1) Hipotesis pada Uji Mean dua Populasi ini antara lain :
  - a.  $H_0$  : Rata-rata populasi 1 sama dengan rata-rata populasi 2
  - b.  $H_1$  : Rata-rata populasi 1 tidak sama dengan rata-rata populasi 2
- 2) Tingkat Signifikansi  
Tingkat signifikansi yang digunakan adalah  $\alpha = 0,05$
- 3) Statistik Uji

Statistik uji t digunakan saat: nilai  $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui

**Subkasus 1** ( $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui, dan terdapat informasi bahwa  $\sigma_1^2 = \sigma_2^2$ )

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{dengan } Sp = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

**Subkasus 2** ( $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui, dan terdapat informasi bahwa  $\sigma_1^2 \neq \sigma_2^2$ )

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- 4) Kaidah Pengambilan Keputusan untuk statistik uji T  
**Subkasus 1** :  $H_0$  ditolak jika  $t_{hitung} > t_{tabel}$  atau  $t_{hitung} < -t_{tabel}$   
untuk  $t_{tabel}$  adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas  $(n_1 + n_2 - 2)$   
**Subkasus 2** :  $H_0$  ditolak jika  $t_{hitung} > t_{tabel}$  atau  $t_{hitung} < -t_{tabel}$   
untuk  $t_{tabel}$  adalah kuantil ke  $-(\alpha/2)$  distribusi t berderajat bebas  $v$ , dengan  $v$  didefinisikan sebagai berikut :

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2}} - 2$$

Dalam penelitian ini data yang digunakan adalah data sekunder dengan jenis data yang digunakan adalah data pada tahun 2022. Sumber data diperoleh dari Badan Pusat Statistika (BPS) dan data tersebut meliputi data IHK (Indeks Harga Konsumen). Tujuan dari penelitian ini dimana membandingkan rata-rata sub kelompok pada Indeks Harga Konsumen (IHK) yaitu dapat memberikan wawasan tentang perbedaan tingkat inflasi atau perubahan harga antara sub kelompok yang berbeda dalam populasi. Jika terdapat perbedaan yang signifikan dalam perubahan harga antara sektor-sektor ini, hal ini dapat mengindikasikan ketimpangan dalam akses terhadap makanan atau pendidikan, yang dapat menjadi fokus analisis lebih lanjut untuk memahami faktor-faktor yang mempengaruhi ketimpangan tersebut.

### III. HASIL DAN PEMBAHASAN

#### 4.1 Analisis Statistik

Analisis Statistik pada pengujian dataset ini adalah mencari nilai rata-rata pada kedua populasi yang nanti nya akan dilakukan pengujian lanjutan pada data tersebut. Berikut adalah hasil dari perhitungan rata-rata pada kedua populasi tersebut

**Tabel 1.** Data Rata-rata

No.	Bulan	Rata-rata Sub Kelompok Makanan	Rata-rata Sub Kelompok Pendidikan
1.	Januari	111,9110	107,7379
2.	Februari	110,3897	107,8106
3.	Maret	112,2807	107,8107
4.	April	114,2647	107,8317
5.	Mei	115,9179	107,8524
6.	Juni	118,3862	107,9162
7.	Juli	119,9352	108,6231
8.	Agustus	116,9399	109,4638
9.	September	115,8922	109,9864
10.	Oktober	114,1669	110,2014
11.	November	113,9823	110,2183
12.	Desember	116,3880	110,2268

```
> data_rata
  Bulan Rata.SK1.Makanan Rata.SK9.Pendidikan
1 Januari      111.9110      107.7379
2 Februari     110.3897      107.8106
3 Maret        112.2807      107.8107
4 April        114.2647      107.8317
5 Mei          115.9179      107.8524
6 Juni         118.3862      107.9162
7 Juli         119.9352      108.6231
8 Agustus      116.9399      109.4638
9 September    115.8922      109.9864
10 Oktober     114.1669      110.2014
11 November    113.9823      110.2183
12 Desember    116.3880      110.2268
```

**Gambar 1.** Data Rata-rata

Gambar 1. adalah data yang digunakan dalam Uji Mean Dua Populasi. Data ini merupakan data yang berhasil diolah dengan cara mencari rata-rata sub kelompok yang dibutuhkan dari data aslinya. Berdasarkan yang terlihat pada gambar tersebut bahwa terdapat rata-rata perbulan sektor makanan dan rata-rata perbulan sektor pendidikan. Pada gambar tersebut dapat dilihat bahwa sub kelompok makanan memiliki rata-rata yang lebih besar pada setiap bulannya.

### 3.2 Uji Normalitas Data

Uji Normalitas adalah pengujian dimana untuk menentukan atau menganalisis suatu data apakah data tersebut berdistribusi normal atau tidak. Pengujian ini dilakukan untuk melihat apakah data yang diteliti merupakan data berdistribusi normal, agar pengujian lanjutan atau pada uji variansi dan uji mean dapat dilakukan.

```
- Uji Normalitas Data -  
Menentukan apakah data yang akan di uji mean berdistribusi normal atau tidak  
  
Data : data_rata$Rata.SK1.Makanan  
  
1. Hipotesis :  
H0 : Data berdistribusi Normal  
H1 : Data tidak berdistribusi Normal  
  
2. Tingkat Signifikansi = 0.05  
  
3. Statistika uji :  
Nilai p-value : 0.9847946  
  
4. Kriteria uji :  
H0 ditolak jika p-value < 0.05  
  
5. Kesimpulan : data_rata$Rata.SK1.Makanan yaitu Data berdistribusi Normal
```

**Gambar 2.** Uji Normalitas Data Sektor Makanan

```
- Uji Normalitas Data -  
Menentukan apakah data yang akan di uji mean berdistribusi normal atau tidak  
  
Data : data_rata$Rata.SK9.Pendidikan  
  
1. Hipotesis :  
H0 : Data berdistribusi Normal  
H1 : Data tidak berdistribusi Normal  
  
2. Tingkat Signifikansi = 0.05  
  
3. Statistika uji :  
Nilai p-value : 0.2231659  
  
4. Kriteria uji :  
H0 ditolak jika p-value < 0.05  
  
5. Kesimpulan : data_rata$Rata.SK9.Pendidikan yaitu Data berdistribusi Normal
```

**Gambar 3.** Uji Normalitas Data Sektor Pendidikan

Berdasarkan hasil uji normalitas, pada Gambar 2 dan Gambar 3 dapat dilihat bahwa kedua data tersebut berdistribusi normal dan memenuhi asumsi uji normalitas. Sehingga dapat dilakukan pengujian lanjutan karena data telah disimpulkan bahwa berdistribusi normal.

### 3.3 Uji Kesamaan Variansi Dua Populasi

Uji Kesamaan Variansi Dua Populasi ini dilakukan untuk menentukan apakah variansi dari kedua populasi berbeda secara signifikan atau tidak. Tujuan dari pengujian ini adalah untuk menguji hipotesis nol yang menyatakan bahwa kedua populasi memiliki variansi yang sama.

- Uji Kesamaan Variansi Dua Populasi -  
Menentukan apakah populasi 1 dan populasi 2 yang akan di uji mean terdapat kesamaan variansi atau tidak

Populasi 1 : data\_rata\$Rata.SK1.Makanan  
Populasi 2 : data\_rata\$Rata.SK9.Pendidikan

1. Hipotesis :  
H0 : Variansi populasi 1 sama dengan variansi populasi 2  
H1 : Variansi populasi 1 tidak sama dengan variansi populasi 2
2. Tingkat Signifikansi = 0.05
3. Statistika uji :  
Statistik Uji F : 6.150005  
Nilai p-value: 0.002743939
4. Kriteria uji :  
H0 ditolak jika p-value < 0.05
5. Kesimpulan :  
Terdapat perbedaan yang signifikan antara variansi populasi 1 dan populasi 2

**Gambar 4.** Uji Kesamaan Variansi Dua Populasi

Hasil uji kesamaan variansi dua populasi dengan tingkat signifikansi sebesar 0.05 (5%) menunjukkan bahwa terdapat perbedaan yang signifikan antara variansi populasi 1 dan populasi 2. Dikarenakan nilai p-value kurang dari tingkat signifikansi. Hal ini menunjukkan bahwa kedua populasi memiliki tingkat dispersi yang berbeda secara signifikan atau dapat disimpulkan bahwa terdapat perbedaan antara kedua variansi populasi tersebut, yang dimana hipotesis nol ditolak.

### 3.4 Uji Mean Dua Populasi Independen

Uji Mean Dua Populasi Independen ini dilakukan untuk membandingkan rata-rata dua populasi yang independen atau tidak terikat satu sama lain. Tujuan pengujian ini adalah untuk menentukan apakah terdapat perbedaan yang signifikan antara kedua populasi yang diujikan.

- Uji Mean 2 Populasi Saling Bebas -

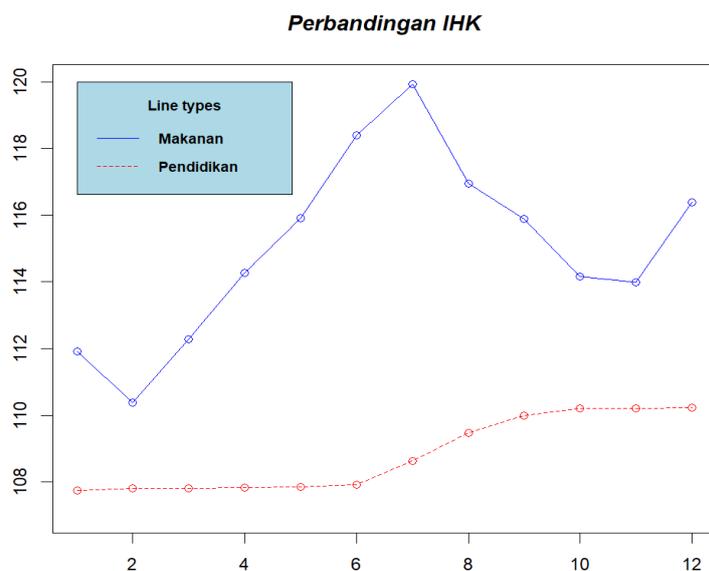
1. Hipotesis :  
H0 : Rata - rata populasi 1 sama dengan rata - rata populasi 2  
H1 : Rata - rata populasi 1 tidak sama dengan rata - rata populasi 2
2. Tingkat Signifikansi = 0.05
3. Statistika uji :  
Nilai |T-hitung| = 7.261982
4. Kriteria uji :  
H0 ditolak jika |T-hitung| > -2.129992
5. Kesimpulan :  
Terdapat perbedaan yang signifikan antara rata - rata populasi 1 dan rata - rata populasi 2

**Gambar 5.** Uji Mean Dua Populasi Saling Bebas

Berdasarkan yang terlihat pada gambar 5. kriteria uji pada uji mean tersebut adalah H0 ditolak jika nilai t hitung lebih besar dari nilai t tabel. Berdasarkan gambar tersebut terlihat bahwa nilai t hitung memang lebih besar dari nilai t tabel. Oleh karena itu hasil uji mean menunjukkan bahwa H0 ditolak dan dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara rata-rata populasi 1 dan rata-rata populasi 2.

### 3.5 Visualisasi Data

Visualisasi Data ini bertujuan untuk merepresentasikan rata-rata dari dua populasi tersebut dalam grafik agar memudahkan dalam memahami, menganalisis serta menginterpretasikan perbandingan kedua populasi tersebut.



**Gambar 6.** Grafik Perbandingan IHK sub kelompok makanan dan pendidikan

Berdasarkan gambar 6. garis berwarna biru merupakan rata-rata IHK dari sub kelompok makanan, sedangkan yang merah merupakan rata-rata IHK dari sub kelompok pendidikan, perbandingan tersebut merupakan rata-rata pada 12 bulan (satu tahun pada tahun 2022). Dari gambar diatas dapat disimpulkan bahwa rata-rata IHK sub kelompok makanan lebih tinggi daripada rata-rata IHK sub kelompok pendidikan, maka dapat diindikasikan bahwa kenaikan harga makanan relatif lebih tinggi daripada kenaikan biaya pendidikan. Dalam konteks tersebut, kelompok masyarakat memiliki kesulitan dalam memenuhi kebutuhan makanan dapat mengalami ketimpangan sosial yang lebih tinggi dibandingkan dengan kelompok yang dapat mengakses makanan dengan harga yang lebih terjangkau.

#### IV. KESIMPULAN

Tujuan dari penelitian ini, dimana membandingkan rata-rata sub kelompok pada Indeks Harga Konsumen (IHK) yaitu dapat memberikan wawasan tentang perbedaan tingkat inflasi atau perubahan harga antara sub kelompok yang berbeda dalam populasi. IHK yang berbeda ini menggambarkan bahwa inflasi pada setiap sektor berbeda yang dapat diartikan tidak meratanya kenaikan atau penurunan inflasi antar kelompok dan sub kelompok secara signifikan. Hal tersebut terjadi karena terdapat perbedaan yang signifikan dari rata-rata dua populasi yang diujikan. Jika terdapat perbedaan yang signifikan dalam perubahan harga antara sektor-sektor ini, hal ini dapat mengindikasikan ketimpangan sosial dalam akses terhadap makanan atau pendidikan, yang dapat menjadi fokus analisis lebih lanjut untuk memahami faktor-faktor yang mempengaruhi ketimpangan tersebut. Misalnya, suatu instansi dapat menggunakan hasil perbandingan ini untuk menentukan kebijakan moneter atau kebijakan fiskal yang lebih tepat sesuai dengan kebutuhan kelompok-kelompok yang berbeda. Dan dari hasil analisis yang telah dilakukan, membuktikan bahwa adanya ketidakmerataan inflasi pada sub kelompok makanan dan pendidikan yang juga mengindikasikan adanya ketimpangan sosial dalam akses terhadap pendidikan.

Dapat disimpulkan bahwa Indeks Harga Konsumen (IHK) pada sub kelompok makanan dan pendidikan memiliki rata-rata yang berbeda secara signifikan, yang berarti adanya ketimpangan sosial yang terkait aksesibilitas, kualitas, atau biaya dari makanan dan pendidikan. Dan dari hasil analisis yang telah dilakukan, didapat rata-rata IHK sub kelompok makanan lebih tinggi daripada rata-rata IHK sub kelompok pendidikan, maka dapat diindikasikan bahwa kenaikan harga makanan relatif lebih tinggi daripada kenaikan biaya pendidikan. Dalam konteks tersebut, kelompok masyarakat memiliki kesulitan dalam memenuhi kebutuhan makanan dapat mengalami ketimpangan sosial yang lebih



tinggi dibandingkan dengan kelompok yang dapat mengakses makanan dengan harga yang lebih terjangkau.

### UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Badan Pusat Statistik (BPS) yang telah menyediakan dataset mengenai Indeks Harga Konsumen (IHK) ini, sehingga kami bisa melakukan analisis terhadap dataset tersebut untuk membandingkan rata-rata sub kelompok yang kami gunakan pada artikel ini.

### REFERENSI

1. BAB I. Available online: <http://repository.unimus.ac.id/5991/3/BAB%20I.pdf> (accessed on 4 June 2023)
2. Penyusunan Inflasi Daerah Kabupaten Gresik Triwulan 3 Tahun 2021. Available online: <https://gresikkab.go.id/documents/1632796143-3.%20Laporan%20triwulanan%203.pdf> (accessed on 4 June 2023)
3. Nabila Aulia Putri Ganessa; Sheilta Alphenia; Aisya Putri Zanuvarizqi; dkk, ANALISIS FAKTOR YANG MEMPENGARUHI INDEKS HARGA KONSUMEN, *Jurnal Khazanah Agustus 2021, Volume 13 No. 1*, 14-23.
4. Uji Kolmogorov-Smirnov. Available online: <https://jagostat.com/metode-statistika-2/uji-kolmogorov-smirnov> (accessed on 4 June 2023)
5. Farah Yuni Lestari; Moh. Yamin Darsyah, Peramalan Indeks Harga Konsumen di Indonesia Menggunakan Metode *Moving average* dan *Holt Exponential Smoothing*, *Prosiding Seminar Nasional Mahasiswa Unimus 2018. Vol. 1*, 400-404
6. Amir Salim; Fadilla, Pengaruh Inflasi Terhadap Pertumbuhan Ekonomi Indonesia, *Ekonomica Sharia: Jurnal Pemikiran dan Pengembangan Ekonomi Syariah Agustus 2021, Volume 7 Nomor 1*, 17-28
7. Inflasi. Available online: <https://www.bps.go.id/subject/3/inflasi.html#subjekViewTab2> (accessed on 31 May 2023)
8. Fitri Zulaikah, ANALISIS FAKTOR INDEKS HARGA KONSUMEN (IHK) PADA SUB-SUB KELOMPOK PENGELUARAN YANG MEMPENGARUHI LAJU INFLASI KABUPATEN PATI TAHUN 2008, Tesis Gelar Ahli Madya, Universitas Negeri Semarang, Semarang, 2009.
9. Pengaruh Indeks Harga Konsumen (Ihk) Dan Inflasi Terhadap Suku Bunga Tahun 2008-2018 Indonesia. Available online: <https://ejournal.uksw.edu/ecodunamika/article/view/4042/1563> (accessed on 4 June 2023)



# Analisis Sentimen Brand Ambassador Artis Korea Selatan pada Produk Indonesia dengan *Lexicon*

Galuh Etha Pratiwi<sup>1</sup>, Tiani Wahyu Utami<sup>2</sup>, Rochdi Wasono<sup>3</sup>

<sup>1, 2, 3</sup>Prodi Statistika Universitas Muhammadiyah Semarang

Corresponding author email: [tianjutami@unimus.ac.id](mailto:tianjutami@unimus.ac.id)

**Abstract:** *The use of South Korean artists as brand ambassadors for Indonesian products is one way that e-commerce and local companies do to attract consumers. However, it reaped positive and negative responses from the public. The social media used to issue opinions regarding this matter is Twitter. This study conducts sentiment analysis from tweets on Twitter regarding this topic. The method used is Lexicon sentiment. The sentiment Lexicon works by first making a dictionary of opinion words and matching them with the sentiment words contained in the Lexicon dictionary. The Lexicon dictionary used is InSet Lexicon. Sentiment classification is divided into 2 namely, positive and negative. Data collection uses the Twitter API with the keyword &quot;BA Artis Korea&quot; as many as 366 data tweets. The research phase consists of data collection, text preprocessing, Lexicon sentiment, visualization, and testing. The results of the classification using the Lexicon method obtained 67.1% positive and 32.9% negative. It was found that the use of South Korean artists to become brand ambassadors for Indonesian products received positive sentiment. Accuracy, precision, and recall values obtained by looking at the confusion matrix table are 78% accuracy, 68% precision, and 61% recall.*

**Keywords:** *Brand Ambassador, Lexicon, Sentiment, Twitter*

**Abstrak:** Penggunaan artis Korea Selatan sebagai brand ambassador produk Indonesia merupakan salah satu cara yang dilakukan *e-commerce* dan perusahaan lokal untuk menarik konsumen. Namun, hal tersebut menuai tanggapan positif dan negatif dari masyarakat. Media sosial digunakan untuk mengeluarkan pendapat mengenai hal tersebut yaitu Twitter. Penelitian ini melakukan analisis sentimen dari cuitan pada Twitter mengenai topik tersebut. Metode yang digunakan *Lexicon sentiment*. *Lexicon sentiment* bekerja dengan membuat kamus kata opini terlebih dahulu dan mencocokkan dengan kata sentimen terdapat pada kamus *Lexicon*. Kamus *Lexicon* yang digunakan yaitu *InSet Lexicon*. Klasifikasi sentimen dibagi menjadi 2 yaitu, positif dan negatif. Pengumpulan data menggunakan Twitter API dengan kata kunci “BA Artis Korea” sebanyak 366 data tweet. Tahap penelitian terdiri dari pengumpulan data, *text preprocessing*, *Lexicon sentiment*, visualisasi, dan pengujian. Hasil dari klasifikasi dengan metode *Lexicon* diperoleh 67,1% positif dan 32,9% negatif. Diperoleh bahwa penggunaan artis Korea Selatan menjadi brand ambassador produk Indonesia mendapatkan sentimen positif. Nilai akurasi, presisi, dan *recall* yang didapatkan dengan melihat tabel confusion matrix yaitu akurasi 78%, presisi 68%, dan *recall* 61%.

**Kata kunci:** Brand Ambassador, *Lexicon*, Sentiment, Twitter

## I. PENDAHULUAN

Seiring dengan pesatnya perkembangan teknologi dan arus informasi memudahkan masyarakat Indonesia lebih terbuka dalam pengetahuan secara luas. Termasuk dengan kemajuan perkembangan internet dan media sosial yang mudah diakses. Hadirnya kemudahan di internet yang membuat internet tidak lagi menjadi sarana untuk mencari informasi. Internet juga bisa digunakan sebagai lahan bisnis yang lebih dikenal sebagai bisnis online atau *e-commerce*. Laporan Survei Status Literasi Digital di Indonesia berkolaborasi dengan Katadata Insight Center (KIC) dan Kementerian Komunikasi dan Informatika (KOMINFO) menyatakan bahwa 27,6% mayoritas masyarakat Indonesia berbelanja online minimal sebulan sekali [1]. *E-commerce* merupakan kegiatan usaha yang mengimpilasikan manufaktur, konsumen, perdagangan perantara, dan pelayanan jasa dengan menggunakan jaringan komputer atau dikenal dengan internet [2]. Dengan adanya *e-commerce* memudahkan pelaku bisnis untuk mengalihkan cara penjualan secara online atau media *e-commerce*.



Menurut Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) yang dirilis pada bulan November 2022 mencatat sejumlah alasan konsumen memilih untuk belanja kebutuhan secara online [3]. Alasan yang paling banyak yaitu harga yang jauh lebih murah dibandingkan membeli langsung di toko dengan persentase sebanyak 15,2%. Alasan kedua yang paling banyak yaitu dapat dilakukan dimana saja sehingga memudahkan konsumen yang ingin membeli suatu barang namun tidak memiliki waktu untuk datang langsung ke tempat. Semakin banyak *e-commerce* dan perusahaan lokal yang melayani pembelian online maka akan semakin ketat persaingan di antara satu sama lain. Berbagai cara dilakukan para pelaku bisnis online untuk menarik konsumen, dengan cara memberikan gratis ongkir untuk pengiriman, menawarkan promo-promo menarik, dan beberapa cara lainnya. Cara lain yang dilakukan pelaku usaha online adalah menggunakan brand ambassador yang tengah diminati oleh masyarakat Indonesia. Brand ambassador merupakan seseorang yang dipercayai untuk mewakili suatu merk dan menjadi wajah dari merk tersebut [4]. Di zaman pasar modern seperti sekarang, tanpa adanya brand ambassador akan susah untuk menjual produk. Untuk mengingat suatu produk, dibutuhkan pemilihan brand ambassador yang tepat agar suatu produk akan tetap berjalan dengan stabil di antara banyaknya produk dan *e-commerce* serupa. Dewasa ini, beberapa perusahaan lokal Indonesia termasuk *e-commerce* mulai berbondong-bondong menggunakan artis asal Korea Selatan sebagai brand ambassador. Produk industri hiburan asal Korea Selatan mulai banyak digemari di Indonesia, baik dalam bentuk musik, drama, makanan, dan budaya. Menurut laporan *Katadata Insight Center* (KIC) bersama *zigi.id* mengatakan bahwa sekitar 1000 responden yang disurvei, sebanyak 10,9% responden terkena gelombang korea atau disebut dengan korean wave pada tahun 2020. Gelombang tersebut terlihat sudah menguat di awal tahun 2010 hingga puncaknya di awal pandemi Covid-19 tahun 2020. Sebanyak 2,1% perbedaan dari tahun 2019 ke 2020 yang menjadi puncak digemarinya hiburan yang berasal dari Korea Selatan tersebut. Oleh karena itu, banyak perusahaan lokal dan *e-commerce* melihat sebuah kesempatan untuk menggandeng artis Korea Selatan sebagai brand ambassador yang akan menjadi wajah bagi perusahaan mereka.

Penggunaan artis korea sebagai brand ambassador menjadi trend di dalam pemasaran online juga dilakukan oleh beberapa *e-commerce*. Beberapa *e-commerce* yang sudah akrab dengan keseharian seperti Tokopedia, Shopee, Blibli, Lazada, dan lainnya pun mulai memakai artis asal Korea Selatan menjadi brand ambassador. Tokopedia yang menggandeng boygrup BTS, Shopee yang menggandeng girlgrup Blackpink, dan lainnya. Penggunaan artis Korea Selatan menjadi brand ambassador seakan menjadi trend di kalangan bisnis online. Target pemasaran kali ini adalah masyarakat yang menyukai artis-artis yang menjadi brand ambassador produk lokal dan *e-commerce* tersebut. Keterlibatan brand ambassador merupakan salah satu faktor kuat untuk mempengaruhi keputusan minat beli dari konsumen. Pemanfaatan artis Korea Selatan mewakili suatu merk produk atau *e-commerce* memiliki keuntungan dari sisi penjualan dikarenakan target pasar mereka yaitu penggemar artis tersebut yang dimana dikenal loyal dan bersedia untuk membeli [5].

Namun, penggunaan artis Korea Selatan sebagai brand ambassador menuai sentimen positif dan negatif. Twitter merupakan salah satu media sosial yang banyak digunakan oleh konsumen untuk mengeluarkan pendapat mengenai hal tersebut. Twitter memiliki kelebihan dibandingkan media sosial lainnya yaitu tampilan yang simpel, topik terupdate, dan cepat dalam menggambarkan sebuah opini pengguna. Fungsi pada Twitter yang digunakan untuk berbagi bahan menjadi sarana yang digunakan di berbagai aspek salah satunya menjadi media pelatihan berupa sentimen analisis, *text mining*, dan *artificial intelligence* [6]. Untuk mengetahui berbagai sentimen masyarakat Indonesia terhadap penggunaan artis dari Korea Selatan sebagai brand ambassador produk lokal maka diperlukan analisis sentimen terhadap pengguna Twitter.

Analisis sentimen adalah penelitian yang menerapkan teknik untuk mengumpulkan opini, emosi, dan sentimen dalam bentuk dan ciri yang dimiliki yang dituangkan dalam bentuk tekstual [7]. Terdapat 2 pendekatan untuk melakukan analisis sentimen, yang pertama yaitu machine learning



dengan melatih data latih pada dataset yang dilabelkan secara manual. Untuk dapat mengklasifikasikan suatu opini (negatif atau positif) perlu dilakukan identifikasi terlebih dahulu apakah suatu kalimat tersebut mengandung opini atau tidak. Pendekatan 2 yaitu berbasis leksikal (*Lexicon*) yang tidak membutuhkan data latih untuk mendapatkan polaritas dari suatu sentimen [8]. Data diproses dengan cara memberikan skor sentimen untuk setiap opini yang ada. Metode *Lexicon* bekerja dengan cara membuat kamus kata opini (*Lexicon*) terlebih dahulu. Metode ini akan mengklasifikasikan teks opini secara otomatis dengan melihat data latih berupa teks opini yang sudah diklasifikasikan secara manual sebelumnya [9]. *Lexicon* bekerja dengan cara memeriksa data yang sudah diproses dan bersih kemudian dibandingkan dengan kamus *Lexicon* yang sudah ada sebelumnya. Kamus *Lexicon* yang digunakan pada penelitian ini adalah InSet (Indonesian *Lexicon Sentiment*) yang dikembangkan oleh [10]. Kamus InSet mengandung kata-kata yang sudah ada bobot atau nilainya per setiap kata. Bobot tersebut nantinya akan digunakan untuk menilai emosi di dalam suatu kalimat. Bobot ini terbagi menjadi positif dan negatif yang akan dijumlahkan untuk menilai suatu sentimen itu mengarah ke positif atau negatif. Kamus InSet terdiri dari 3609 kata positif dan 6609 kata negatif.

Penelitian terdahulu dilakukan oleh [11] yang berjudul “Penerapan *Lexicon-Based* Untuk Analisis Pada Twitter Terhadap Isu Covid-19” didapatkan kesimpulan yaitu opini masyarakat untuk isu Covid-19 memperoleh sentimen positif dengan kesimpulan bahwa pengguna Twitter percaya bahwa Covid-19 nyata adanya masyarakat pengguna Twitter percaya bahwa covid itu nyata dibandingkan dengan masyarakat yang mempercayai isu konspirasi yang mengatakan bahwa covid tidak nyata. Penelitian sebelumnya juga pernah dilakukan oleh [12] yang berjudul “Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode *Lexicon Based*” dalam penelitian ini hasil akhir dari opini yang diperoleh dari Twitter dibagi menjadi 3 kategori yaitu positif, negatif, dan netral sehingga bisa disimpulkan sentimen masyarakat terhadap topik tersebut. Penelitian mengenai penggunaan artis Korea Selatan sebagai brand ambassador pernah dilakukan oleh [5] yang berjudul “Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal”, penelitian ini terfokus kepada brand ambassador produk kecantikan lokal dan menggunakan metode *support vector machine* (SVM) dengan hasil analisis menunjukkan kecenderungan masyarakat bersifat positif dengan adanya keputusan tersebut. Berdasarkan uraian tersebut, maka penelitian ini akan meneliti mengenai sentimen brand ambassador pada artis Korea Selatan pada produk Indonesia menggunakan metode *Lexicon sentiment*. Sehingga diperoleh gambaran umum sentimen pengguna Twitter, hasil klasifikasi dan visualisasi, dan tingkat akurasi, presisi, dan recall. Sumber kamus *Lexicon* yang digunakan pada penelitian ini yaitu InSet (Indonesian *Sentiment Lexicon*) yang dikembangkan oleh [10].

## II. METODE PENELITIAN

### 2.1. Sumber Data dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder yang merupakan data cuitan (tweet) pengguna Twitter. Data dikumpulkan dari hasil *crawling* yang menggunakan kata kunci yang sudah ditentukan yaitu “BA Artis Korea”. Proses *crawling* data menggunakan salah satu fitur dari Twitter yaitu Twitter API (*Application Programming Interface*). Twitter API adalah fitur dimana Twitter mengizinkan pihak ketiga untuk mengakses informasi dari Twitter yang terdapat pada website <https://developers.twitter.com/>. Data tweet dikumpulkan pada bulan Januari – Februari 2023 dengan jumlah data tweet sebanyak 366 tweet. Variabel penelitian yang digunakan yaitu tanggal tweet dibuat dan data tweet yang berisi cuitan dari pengguna Twitter.

### 2.2. Langkah Penelitian

Metode penelitian yang digunakan yaitu *Lexicon sentiment* dan menggunakan kamus InSet



*Lexicon*. Penelitian ini bertujuan untuk mendapatkan gambaran umum sentimen pengguna Twitter terhadap brand ambassador artis Korea Selatan pada produk Indonesia, mendapatkan hasil klasifikasi dan visualisasi menggunakan metode *Lexicon* sentiment terhadap brand ambassador artis Korea Selatan pada produk Indonesia, dan mendapatkan tingkat akurasi, presisi, dan recall metode *Lexicon* dalam pengklasifikasian sentimen pengguna Twitter terhadap brand ambassador artis Korea Selatan pada produk Indonesia. Adapun langkah-langkah sebagai berikut:

1. Proses *crawling* data tweet menggunakan Twitter API.
2. Text preprocessing yang bertujuan untuk memperbaiki data awal menjadi data yang sudah bisa digunakan. Tahap *text preprocessing* yang dilakukan, yaitu: *Case folding*, *Punctuation removal*, *Tokenizing*, *Stopword removal*, *Stemming*.
3. Klasifikasi menggunakan pendekatan *Lexicon* dengan menggunakan kamus InSet *Lexicon*:
  - *Polarity tagging*, menganalisis polaritas dalam kata dan membagi menjadi positif dan negatif.
  - *Word frequency*, menghitung banyaknya satu kata dalam satu kalimat sentimen.
  - *Word attitude*, mengubah polaritas kata positif dan negatif menjadi angka (nilai).
  - *Overall attitude*, menjumlahkan nilai per kata positif dan negative
  - *Sentiment score*, menghitung total nilai sentimen dalam satu kalimat. Tahap ini merupakan tahap terakhir dalam analisis *Lexicon*.
4. Skenario pengujian, dilakukan pengujian dengan menghitung presisi, *recall*, akurasi, dan *f-measure*.

### III. HASIL DAN PEMBAHASAN

Data *tweet* di *crawling* menggunakan *software RStudio* dalam rentang bulan Januari – Februari 2023 dengan kata kunci pencarian BA Artis Korea. Jumlah data yang berhasil di *crawling* sebanyak 366 *tweet*. *Text preprocessing* merupakan tahap awal yang dilakukan dalam analisis text mining yang bertujuan untuk memperbaiki data sesuai yang dibutuhkan dalam analisis selanjutnya. Analisis *text preprocessing* menggunakan *Google Colab*.

- *Case Folding*

Tahap ini dilakukan untuk proses penyamaan huruf yang terdapat dalam data dengan mengubah huruf kapital menjadi huruf tidak kapital.

- *Punctuation Removal*

Tahap *punctuation removal* dilakukan pembersihan terhadap data tweet yang didalamnya masih terdapat link, username, hashtag, dan lain-lain yang tidak digunakan dalam proses klasifikasi. Selain itu, dilakukan penghapusan angka dan tanda baca. Sehingga data tweet hanya berisi data yang bersih dan bisa digunakan untuk proses selanjutnya.

- *Tokenizing*

Tahap ini dilakukan untuk membagi kata-kata yang menyusun dokumen atau kalimat pada data tweet. Proses ini memudahkan untuk melakukan pembobotan pada setiap kata dan menghasilkan kata-kata yang sudah terpisah satu sama lain tidak berbentuk satu kalimat.

- *Stopword Removal*

Tahap *stopword removal* dilakukan untuk penghapusan kata yang tidak digunakan dalam analisis. Tabel di atas menunjukkan hasil dari tahapan *stopword removal*. Pada data *tweet* pertama dilakukan penghapusan pada kata “[‘sih’, ‘soalnya’, ‘kan’, ‘ada’, ‘jadi’, ‘gitu’, ‘pasti’]”.

- *Stemming*

Tahap *stemming* bertujuan untuk mengubah kata berimbuhan menjadi kata dasar. Imbuhan pada awal kata, imbuhan akhir kata, dan imbuhan pada tengah kata sehingga diperoleh kata dasar yang akan memudahkan proses analisis sentimen selanjutnya. Tabel di atas menunjukkan hasil dari proses stemming. Data tweet pertama dilakukan perubahan pada kata “[‘untungnya’,

‘meningkatkan’, ‘penjualan’, ‘idolanya’]” menjadi “[‘untung’, ‘tingkat’, ‘jual’, ‘idol’]”. Kata “untungnya” dan “idolanya” dilakukan penghapusan imbuhan pada akhir kata yaitu “- nya”. Kata “penjualan” dan “meningkatkan” dilakukan penghapusan imbuhan pada awal dan akhir kata sehingga didapatkan kata dasar yaitu “jual” dan “tingkat”.

### 3.1. *Lexicon Sentiment*

Setelah dilakukan proses text preprocessing yaitu *case folding*, *punctuation removal*, *tokenizing*, *stopword removal*, dan *stemming* maka diperoleh data *tweet* bersih yang akan digunakan untuk proses. Setelah dilakukan proses *text preprocessing* yang dilakukan untuk mempersiapkan data untuk bisa dilanjutkan dengan analisis selanjutnya yaitu *Lexicon sentiment*. *Lexicon sentiment* dibangun sesuai dengan kata-kata sentimen yang terdapat di dalam kamus yang digunakan. Pada analisis ini menggunakan kamus *Lexicon* yaitu Indonesia *Sentiment Lexicon (InSet)* dan *Google Colab*. Kamus pada *Lexicon sentiment* digunakan untuk menentukan polaritas sentimen dari suatu opini. Hasil dari perhitungan polaritas sentimen bisa dilihat pada Tabel di bawah ini:

**Tabel 1.** Hasil Skor Polaritas

Opini	<i>Clean Tweet</i>	Skor Polaritas	Polaritas
1	orang indonesia perasaan kaga seputih salju skinker lokal pake ba orang korea sehat	3	Positif
2	thats why produk skincare brand lokal gw review recommend always range harga item please skincare brand lokal stop bayar mahal pake ba artis korea harga skincare ikutan mahal marketnya geser	-1	Negatif

Pada Tabel 1 terdapat hasil akhir dari skor polaritas pada masing-masing terhadap setiap kalimat sentimen. Proses ini dilakukan dengan membandingkan data *tweet* yang sudah melalui tahap *text preprocessing* terlebih dahulu. Kemudian dibandingkan dengan kata-kata sentimen yang terdapat di dalam kamus *InSet Lexicon*. Kamus ini berisi kata-kata sentimen beserta bobot nilai yang berfungsi untuk menentukan nilai polaritas. Nilai polaritas diperoleh dengan menjumlahkan seluruh bobot kata yang terdapat dalam sebuah kalimat opini. Pada Tabel 1 diketahui skor polaritas dari setiap kata yang membentuk kalimat opini. Kalimat opini yang berbunyi “orang indonesia perasaan kaga seputih salju skinker lokal pake ba orang korea sehat” mengandung 1 kata sentimen yang terdapat dalam kamus *InSet Lexicon* yaitu “sehat”. Tahap-tahap yang dilakukan untuk menghitung skor polaritas diuraikan sebagai berikut:

- *Polarity tagging*: tahap ini menganalisis suatu kata yang terdapat dalam kalimat opini dan membagi dalam kelas positif dan negatif. Kata “sehat” dalam kalimat opini 1 terdapat pada kelas positif dan negatif sehingga kata “sehat” digolongkan sebagai kata positif dan negatif.
- *Word frequency*: tahap ini menghitung banyaknya satu kata dalam sebuah kalimat sentimen. Dalam opini 1 terdapat 1 kata sentimen yaitu “sehat”. Kata “sehat” pada kalimat opini 1 hanya ada satu. Jika dalam satu kalimat opini terdapat 2 kata sentimen maka akan dihitung menjadi dua dan seterusnya.
- *Word attitude*: tahap ini bertujuan untuk mengubah polaritas kata positif dan negatif menjadi angka atau nilai. Kata “sehat” yang terdapat dalam kelas positif dan negatif diubah menjadi nilai sesuai dalam kamus yang dipakai yaitu *InSet Lexicon*. Kata “sehat” mempunyai 2 nilai, yaitu (4) untuk kelas positif dan (-1) untuk kelas negatif.
- *Overall attitude*: tahap ini menghitung nilai keseluruhan dari suatu kalimat opini sehingga mendapatkan nilai polaritas untuk setiap kata sentimen. Kata “sehat” mempunyai nilai 4 kelas positif dan -1 kelas negatif. Maka total nilai kata “sehat”, yaitu:

$$\text{Overall attitude} = (1 \times 4) + (1 \times (-1)) = 4 + (-1) = 3$$

Kata “sehat” terdapat 1 kali dalam kalimat opini 1 maka *word frequency* = 1. *Word frequency* tersebut dikalikan dengan nilai sentimen pada kelas positif yaitu 4 dan nilai sentimen pada kelas negatif yaitu -1. Sehingga diperoleh total nilai polaritas untuk kata “sehat” adalah 3.

- *Sentiment score*: tahap ini menghitung total nilai sentimen dalam sebuah kalimat.

$$\text{Sentiment score} = 3 = 3$$

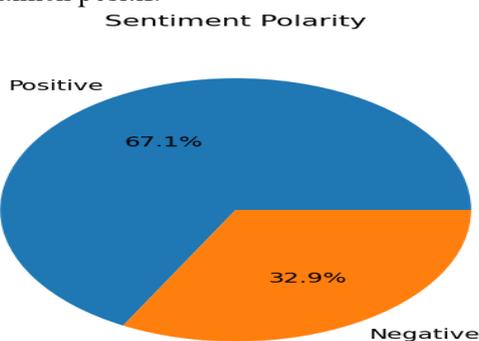
Karena dalam kalimat opini 1 hanya mengandung 1 sentimen yaitu “sehat” maka nilai polaritas total dalam satu kalimat opini adalah 3. Karena *sentiment score* > 0, maka disimpulkan

bahwa sentimen kalimat opini 1 adalah positif. Nilai polaritas kemudian diubah menjadi polaritas pada setiap kalimat sentimen yang sudah diproses dengan kamus InSet *Lexicon* untuk mengetahui polaritas sebuah kalimat.

Pada Tabel 1 menampilkan skor polaritas dan polaritas untuk kalimat opini yang telah diproses melalui *text preprocessing* sehingga didapatkan kalimat opini yang sudah bersih. Kalimat opini “orang indonesia perasaan kaga seputih salju skinker lokal pake ba orang korea sehat” mendapatkan skor polaritas 3 dan *sentiment score*  $> 0$  maka polaritasnya positif. Kalimat opini “thats why produk skincare brand lokal gw review recommend always range harga item please skincare brand lokal stop bayar mahal pake ba artis korea harga skincare ikutan mahal marketnya geser” mendapatkan skor polaritas -1 dan *sentiment score*  $< 0$  maka polaritasnya negatif.

### 3.2. Visualisasi Sentimen

Setelah diketahui nilai polaritas dan polaritas dari suatu data, dilanjutkan dengan visualisasi. Visualisasi dilakukan untuk mengetahui pembagian dari polaritas positif dan negatif dalam bentuk *piechart* dan *wordcloud* kata-kata yang sering muncul dalam kelas positif dan negatif. Pada gambar 4.2 di bawah ini diketahui sebanyak 67,1% mempunyai polaritas positif dan 32,9% mempunyai polaritas negatif. Dari 325 data *tweet* yang terkait dengan penggunaan artis Korea Selatan sebagai *brand ambassador* didapatkan hasil bahwa 67,1% memiliki sentimen positif terkait topik tersebut dan 32,9% memiliki sentimen negatif terkait dengan topik tersebut. Hal ini menyatakan bahwa *brand ambassador* yang menggunakan artis Korea Selatan pada produk Indonesia mendapatkan sentimen positif.



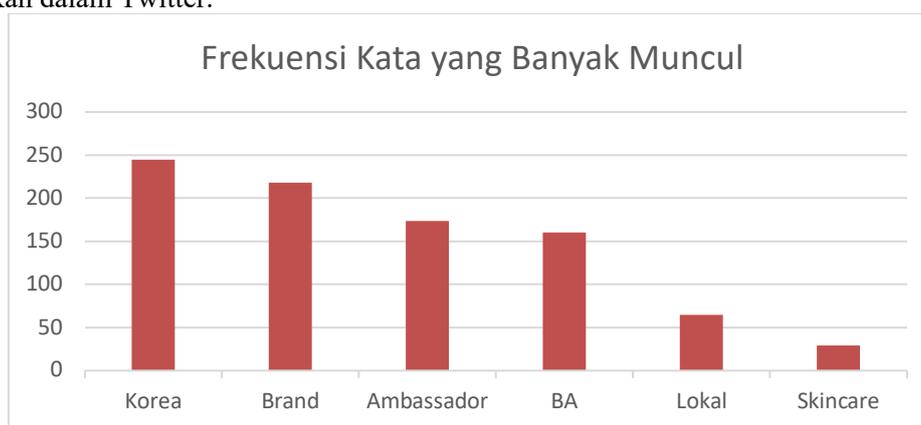
Gambar 1. Polaritas Sentimen



Gambar 2. Wordcloud Sentimen

Pada Gambar 2 merupakan visualisasi dari kata positif dan negatif yang paling banyak muncul pada data opini masyarakat mengenai penggunaan artis Korea Selatan sebagai *brand ambassador* produk Indonesia. *Wordcloud* untuk kata positif menunjukkan kata “*brand ambassador*” dan “*korea*” merupakan kata paling banyak digunakan. Hal ini dikarenakan kata kunci yang digunakan untuk *crawling* data *tweet* adalah BA artis Korea yang merupakan singkatan dari *brand ambassador*. Selain itu terdapat beberapa kata lainnya yaitu “*seru*”, “*suka*”, “*artis*”, dan “*lokal*”. Kata-kata tersebut dapat diartikan bahwa artis Korea Selatan yang menjadi *brand ambassador* disukai oleh konsumen yakni pengguna Twitter. Konsumen merasakan bahwa artis-artis tersebut menjadi lokal dikarenakan memakai produk Indonesia dan familiar dalam keseharian.

*Wordcloud* untuk kata negatif menampilkan bahwa kata “korea” yang paling banyak digunakan. Kata “korea” merupakan kata yang termasuk ke dalam kata kunci *crawling* data. Hal ini menyatakan bahwa kata Korea juga termasuk ke dalam *tweet* negatif. Kata lainnya yang muncul dalam *wordcloud* negatif adalah “beda”, “kulit”, “lokal”, dan “*denial*”. Kata-kata tersebut dapat diartikan bahwa konsumen merasa bahwa *brand ambassador* yang digunakan oleh produk Indonesia berbeda dalam warna kulit masyarakat Indonesia. Selain itu, kecantikan Korea dijadikan standar oleh beberapa produk kecantikan. Dengan memakai wajah artis tersebut, produk diiklankan dengan harapan bahwa konsumen akan membeli produk dan bisa menjadi seperti wajah dari *brand ambassador* yang diiklankan. Kata *denial* atau dalam bahasa Indonesia yaitu menyangkal/penyangkalan. Beberapa konsumen menyangkal dengan penggunaan artis Korea sebagai *brand ambassador* produk Indonesia dan menyampaikan opini negatif dalam *tweet* yang dituliskan dalam Twitter.



**Gambar 3.** Histogram kata yang paling banyak muncul

Gambar 3 menunjukkan frekuensi kata yang paling banyak muncul pada data *tweet* yang sudah di proses melalui *text preprocessing*. Kata pada data *tweet* yang paling banyak muncul adalah “korea” yang muncul sebanyak 245 kali. Kata “*brand*” muncul sebanyak 218 kali dan dilanjutkan dengan kata “*ambassador*” yang muncul sebanyak 174 kali. Oleh karena itu, bisa dikatakan bahwa data *tweet* yang sudah *dicrawling* memuat kata kunci yang sesuai dengan yang ditentukan yaitu “BA Artis Korea”. Sehingga data *tweet* sudah memenuhi informasi yang dibutuhkan pada penelitian yang menggunakan *brand ambassador* artis Korea Selatan pada produk Indonesia.

Nilai polaritas untuk sentimen positif yang paling tinggi yaitu 23. Nilai polaritas untuk sentimen negatif yang paling tinggi yaitu -14. Pada Tabel berikut ditampilkan sentimen positif teratas dan sentimen negatif teratas.

**Tabel 2.** Top Positif dan Negatif Sentimen

Data <i>Tweet</i>	Skor Polaritas	Polaritas
@arcus aorta skg skin care lokal aja brand ambassador nya artis korea ya bun..etapi iklan mie sama kopi/susu pun skg artis kepop. Kadang pengen tau bayarannya dibanding artis lokal apa ga beda jauh sampe perusahaannya mutusin pake artis korea/kepop tp ya meh ngopo bar kuwi...	23	Positif
@waesehoon itu fandom sebelah tuh bilang kalo idolnya 1st ga buat male naa karena dikira sh itu bukan ga tapi ha ada jg ba cmn dr korea grgr brand nya labil ngasih tittle buat sh yaudah dah tuh udah dikasih tau tp tetep denial bilang brandnya boong lah apa lah	-14	Negatif

### 3.3. Tabel Confusion Matrix

Tabel *confusion matrix* digunakan untuk mengevaluasi suatu klasifikasi dengan menghitung berdasarkan data penelitian untuk menghindari *false statment*. Tabel 10 di bawah ini merupakan tabel *confusion matrix* dari data *tweet* yang telah diklasifikasikan menggunakan *Lexicon InSet*. Pada tabel *confusion matrix* yang digunakan pada penelitian ini, data prediksi yaitu data dalam

format hasil klasifikasi yang dihasilkan oleh algoritma yang digunakan, data aktual dihasilkan dengan penghitungan manual dengan cara membandingkan dengan kamus InSet *Lexicon*.

**Tabel 3.** Hasil Tabel *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	66	30
Negatif	41	188

Pada Tabel 3 diperoleh hasil sebanyak 66 data *tweet* termasuk dalam *true positive* yaitu data aktual positif yang diprediksi benar positif, 30 data *tweet* termasuk dalam *false positive* yaitu data aktual positif yang diprediksi salah negatif, 41 data *tweet* termasuk dalam *false negative* yaitu data aktual negatif yang diprediksi salah positif, dan 188 data *tweet* termasuk dalam *true negative* yaitu data aktual negatif yang diprediksi benar negatif. Setelah mendapatkan hasil dari tabel *confusion matrix* maka dilanjutkan dengan menghitung nilai presisi, *recall*, akurasi, dan *f-measure*.

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{66}{66+30} = \frac{66}{96} = 0,68 = 68\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{66}{66+41} = \frac{66}{107} = 0,61 = 61\%$$

$$\text{Accurancy} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{66+188}{66+30+188+41} = \frac{254}{325} = 0,78 = 78\%$$

$$F - \text{Measure} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{(\text{Presisi} + \text{Recall})} = 2 \times \frac{0,68 \times 0,61}{(0,68 + 0,61)} = \frac{0,82}{1,29} = 0,63 = 63\%$$

Berdasarkan hasil perhitungan di atas diketahui nilai presisi, *recall*, akurasi, dan *f-measure* dari proses analisis sentimen *Lexicon* menggunakan kamus InSet *Lexicon*. Nilai presisi yang dihasilkan yaitu 68%, *recall* yaitu 61%, akurasi yaitu 78%, dan *f-measure* yaitu 63%.

#### IV. KESIMPULAN

Setelah melakukan tahapan analisis dan memperoleh hasil yang telah dibahas sebelumnya, ada beberapa kesimpulan yang diperoleh diantaranya yaitu:

1. Setelah dilakukan klasifikasi sentimen dengan menggunakan kamus InSet *Lexicon* diperoleh total nilai polaritas untuk sentimen positif yang paling tinggi adalah 23. Nilai polaritas untuk sentimen negatif yang paling tinggi adalah -14. *Wordcloud* untuk kata positif menunjukkan kata “*brand ambassador*” dan “*korea*” merupakan kata paling banyak digunakan. Hal ini dikarenakan kata kunci yang digunakan untuk *crawling* data *tweet* adalah BA artis Korea yang merupakan singkatan dari *brand ambassador*. Selain itu terdapat beberapa kata lainnya yaitu “*seru*”, “*suka*”, “*artis*”, dan “*lokal*”. Kata-kata tersebut dapat diartikan bahwa artis Korea Selatan yang menjadi *brand ambassador* disukai oleh konsumen yakni pengguna Twitter. Konsumen merasakan bahwa artis-artis tersebut menjadi lokal dikarenakan memakai produk Indonesia dan familiar dalam keseharian. *Wordcloud* untuk kata negatif menampilkan bahwa kata “*korea*” yang paling banyak digunakan. Kata “*korea*” merupakan kata yang termasuk ke dalam kata kunci *crawling* data. Hal ini menyatakan bahwa kata Korea juga termasuk ke dalam *tweet* negatif. Kata lainnya yang muncul dalam *wordcloud* negatif adalah “*beda*”, “*kulit*”, “*lokal*”, dan “*denial*”. Kata-kata tersebut dapat diartikan bahwa konsumen merasa bahwa *brand ambassador* yang digunakan oleh produk Indonesia berbeda dalam warna kulit masyarakat Indonesia. Selain itu, kecantikan Korea dijadikan standar oleh beberapa produk kecantikan. Dengan memakai wajah artis tersebut, produk diiklankan dengan harapan bahwa konsumen akan membeli produk dan bisa menjadi seperti wajah dari *brand ambassador* yang diiklankan. Kata *denial* atau dalam bahasa Indonesia yaitu menyangkal/penyangkalan. Beberapa konsumen menyangkal dengan penggunaan artis Korea



- sebagai *brand ambassador* produk Indonesia dan menyampaikan opini negatif dalam *tweet* yang dituliskan dalam Twitter.
2. Tingkat akurasi yang diperoleh dari hasil klasifikasi menggunakan kamus InSet *Lexicon* dengan yaitu 78%. Nilai presisi yaitu 68% dan nilai *recall* yaitu 61%.

**UCAPAN TERIMA KASIH****REFERENSI**

1. Yoshio, A. Mayoritas Masyarakat Indonesia Pernah Belanja Online. *Katada.Co.Id*, 2021. <https://databoks.katadata.co.id/datapublish/2022/04/27/mayoritas-masyarakat-indonesia-pernah-belanja-online>
2. Prasetyo, T., dan Barkatullah, A., *Bisnis E-commerce: Suatu Sistem Keamanan dan Hukum di Indonesia*. Pustaka Pelajar, 2005
3. Annur, C. M. Ragam Alasan Konsumen Pilih Berbelanja Online. *Asosiasi Penyelenggara Jasa Internet Indonesia (APJII)*, November, 2020.
4. <https://databoks.katadata.co.id/datapublish/2020/11/11/ragam-alasan-konsumen-pilh-berbelanja-online>
5. Poghosyan, A. Celebrity Endorsement as one of nowadays Major Ways to Influence Customer Buying Behaviour. *European Scientific Journal*, 2015; pp 30–38.
6. Slamet, R., Gata, W., Novtariany, A., Hilyati, K., & Jariyah, F. A. Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal. *INTECOMS: Journal of Information Technology and Computer Science*, 2022; 5(1), 145–153. <https://doi.org/10.31539/intecom.v5i1.393>
7. Hartanto. Text Mining dan Sentimen Analisis Twitter. *Jurnal Psikologi Ilmiah*, 2017; 9(1), 18–25.
8. Liu, B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 2012; 5, 1–167.
9. Medhat, W., Hassan, A., & Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 2014; 5(4), 1093–1113.
10. Azhar, Y. Metode *Lexicon*-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 2018; 6(3), 237.
11. <https://doi.org/10.23887/janapati.v6i3.11739>
12. Koto, F., & Rahmaningtyas, G. Y. Inset *Lexicon*: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017, 2018-Janua(December)*, 391–394. <https://doi.org/10.1109/IALP.2017.8300625>
13. Nooryuda Prasetya, Y., & Winarso, D. Penerapan *Lexicon* Based Untuk Analisis Sentimen Pada Twitter Terhadap Isu Covid-19. *Jurnal Fasilkom*, 2021; 11(2), 97–103.
14. Arief, R., & Imanuel, K. Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode *Lexicon* Based Universitas Gunadarma 1, 2 Jalan Margonda Raya No 100 Depok Jawa Barat 16424 Sur-el: rifiana@staff.gunadarma.ac.id 1, karel4404@gmail.com 2. *Jurnal Ilmiah MATRIK*, 2019; 21(3), 242–250