



# Citayam Fashion Week dalam Twitter.

Amri Muhaimin<sup>1</sup>, Muhammad Rafka<sup>2</sup>, Ryan Badai Alamsyah<sup>3</sup>

<sup>1,2,3</sup> Universitas Pembangunan Nasional "Veteran" Jawa Timur Program Study Sains Data

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

<sup>2</sup> [penulis.ketiga@universitas.ac.id](mailto:penulis.ketiga@universitas.ac.id)

<sup>3</sup> [penulis.ketiga@universitas.ac.id](mailto:penulis.ketiga@universitas.ac.id)

Corresponding author email: [amri.muhammad.stat@upnjatim.ac.id](mailto:amri.muhammad.stat@upnjatim.ac.id)

**Abstract:** *Citayam Fashion Week is a trending phenomenon in Indonesia. There have been a lot of comments from netizens on Twitter. By using Scrapy, the response text data is obtained and processed. Sentiment analysis is carried out to find out what sentiments fulfill this phenomenon. After that, it is visualized using Wordcloud for each sentiment that is formed. Finally, topic modeling was conducted to find out the dominant words in tweet responses that said the key to Citayam Fashion Week.*

**Keywords:** *phenomenon, model, sentiment, topic, wordcloud*

**Abstrak:** Citayam Fashion Week merupakan fenomena yang sedang trending di Indonesia. Banyak sekali tanggapan-tanggapan netizen yang dilontarkan di sosial media Twitter. Dengan menggunakan Scrapy, data teks tanggapan tersebut didapatkan dan diolah. Analisis sentimen dilakukan untuk mengetahui sentimen apa yang memenuhi fenomena tersebut. Setelah itu divisualisasikan menggunakan Wordcloud untuk masing-masing sentimen yang terbentuk. Terakhir dilakukan pemodelan topik untuk mengetahui kata-kata yang dominan dalam tanggapan tweet yang berkata kunci Citayam Fashion Week.

**Kata kunci:** fenomena, model, sentimen, topik, wordcloud

## I. PENDAHULUAN

Analisis Teks telah populer dan sering digunakan untuk menganalisis sebuah fenomena. Karena dengan menggunakan analisis teks, kita secara langsung melakukan analisis terhadap berdasarkan perorangan terhadap suatu fenomena. Tujuan umum untuk melakukan analisis teks adalah salah satunya mengetahui sentimen suatu fenomena tersebut. Beberapa alat yang biasa digunakan dalam analisis teks diantaranya adalah *Wordcloud*, Analisis Sentimen, dan Pemodelan Topik. Fenomena yang dianalisa adalah *Citayam Fashion Weeks*. Fenomena *Citayam Fashion Weeks* sangat populer pada akhir-akhir ini. Fenomena tersebut menjadi ramai karena banyaknya masyarakat dengan mayoritas remaja memenuhi jalanan Dukuh Atas Jakarta Pusat. Setelah menjadi ramai, banyak kalangan yang memberikan komentar melalui media sosial seperti Twitter, Instagram, dan Facebook. Melalui komentar-komentar tersebut teks analisis dapat dilakukan.

Riset terkait analisis teks sudah banyak dilakukan, salah satunya adalah analisis teks terkait covid-19 berdasarkan data twitter oleh [1]. Model yang digunakan adalah model BERT, dan data yang digunakan berasal dari komentar dari twitter. Akurasi yang didapatkan sebesar 99% untuk mengklasifikasikan komentar tersebut masuk kedalam sentimen positif atau negatif. Kemudian analisis teks juga dilakukan oleh [2], yaitu menganalisa terkait unggahan di Facebook dengan topik lingkungan. Model yang digunakan adalah *Social Cognitive Theory* (SCT). Hasil yang didapatkan berupa *Wordcloud* dan juga diagram hubungan antar kata.

Penelitian ini bertujuan untuk mengetahui sentimen dan juga pemodelan topik yang berhubungan dengan fenomena *Citayam Fashion Weeks*. Data yang digunakan berasal dari twitter dengan kata kunci #citayamfashionweeks atau *Citayam Fashion Weeks*. Kemudian dari data tersebut akan dikelola terkait analisis sentimen, pemodelan topik, dan pembuatan *wordcloud*. Sentimen analisis menggunakan pendekatan perhitungan kata yang terkandung dalam kamus positif atau negatif. Kemudian pembentukan topik menggunakan metode *Latent Dirichlet Analysis* untuk mendapatkan berapa topik yang terbentuk berdasarkan komentar terkait fenomena *Citayam Fashion Weeks* di twitter.



Selengkapnya dijelaskan pada bagian selanjutnya yaitu Metode Penelitian, Langkah Analisis, Analisis dan Pembahasan, terakhir kesimpulan.

## II. METODE PENELITIAN

Bagian ini membahas tentang metode-metode yang digunakan dalam menyelesaikan penelitian terkait. Terdapat empat metode yang digunakan dalam penelitian ini, Analisis Teks, Analisis Sentimen, Wordcloud, dan Pemodelan Topik.

### II.1. Analisis Teks

Analisis teks adalah istilah metode penelitian yang digunakan untuk menggambarkan, menafsirkan, dan memahami teks. Berbagai jenis informasi dapat diturunkan dari sebuah teks, dari makna harfiahnya hingga subteks, simbolisme, asumsi, dan nilai yang diungkapkannya. Metode yang digunakan untuk melakukan analisis teks tergantung pada bidang dan tujuan penelitian. Umumnya bertujuan untuk menghubungkan teks dengan konteks sosial, politik, budaya atau seni yang lebih luas [3].

### II.2. Analisis Sentimen

Analisis sentimen adalah jenis pengolahan dan pemrosesan bahasa alami yang mengkategorikan teks ke dalam berbagai tingkatan teks yang berupa dokumen, paragraf, kalimat, klausa, frasa, atau daftar frasa untuk mendapatkan informasi sentimen yang terkandung di dalamnya. Analisis sentimen digunakan untuk menemukan opini atau kecenderungan suatu opini dari beberapa review atau ulasan positif, negatif atau netral. Analisis sentimen sangat berguna untuk memahami minat dan kecenderungan seseorang dalam mengekspresikan diri di Internet [4].

### II.3. Wordcloud

Dalam kamus Cambridge dijelaskan bahwa *Wordcloud* adalah sebuah visualisasi yang menunjukkan daftar kata-kata yang digunakan dalam sebuah teks, secara umum semakin besar frekuensi kata tersebut dalam teks semakin besar pula ukuran kata tersebut di dalam visualisasi. *Wordcloud* digunakan untuk membantu evaluator dengan analisis tekstual eksplorasi dengan mengidentifikasi frekuensi kata-kata yang sering muncul dalam serangkaian wawancara, dokumen, maupun teks lainnya. Visualisasi itu juga dapat digunakan untuk mengkomunikasikan poin-poin atau tema yang paling menonjol dalam tahap pelaporan [5].

### II.4. Pemodelan Topik

Pemodelan topik atau *topic modeling* adalah suatu metode clustering yang terdapat dalam *unsupervised learning*. Dalam clustering terdapat tiga tipe, yaitu *soft fuzzy clustering*, *hard clustering*, dan *hierarchical clustering*. Untuk pemodelan topik termasuk dalam *soft/fuzzy clustering* yang mana setiap objek dapat memiliki lebih dari satu *cluster* dengan tingkat tertentu [6]. Salah satu teknik yang digunakan dalam pemodelan teknik adalah metode *Latent Dirichlet Analysis* (LDA). Metode LDA digunakan untuk melakukan pengelompokan, meringkas, menghubungkan, ataupun memproses suatu data. Ide dasar dari metode LDA ialah menyiratkan bahwa dokumen direpresentasikan sebagai campuran acak atas topik laten, di mana setiap topik didefinisikan oleh distribusi kata [7].

## III. LANGKAH ANALISIS

Penelitian ini menggunakan data teks yang diambil dari twitter menggunakan bantuan Scrapy. Data yang didapatkan selanjutnya diolah untuk dianalisa dan dihasilkan Wordcloud, Pemodelan Topik, dan

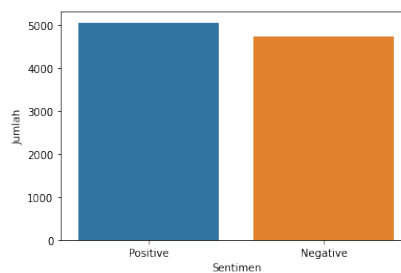
persentasi hasil sentimen. Scrapy data menggunakan kata kunci Citayam Fashion Week, dengan jumlah kurang lebih 10000 tweet yang diperoleh.

Setelah diperoleh, selanjutnya data tersebut diproses berbasis teks. Seperti menghilangkan tanda baca, merubah ke dalam huruf kecil, menghilangkan angka, menghilangkan stopwords, melakukan stemming, menghilangkan kata kunci Citayam Fashion Week, melakukan tokenisasi, dan merubah kedalam bentuk tabular dengan nilai Term Frequency – Inverse Document Frequency (TF-IDF). Setelah itu dilakukan analisis sentimen dengan mengklasifikasikan teks yang belum terklasifikasi secara manual dengan menggunakan model Naïve Bayes.

Pada bagian analisis sentimen, langkah awal kita melakukan klasifikasi secara manual pada 50 teks secara acak. Selanjutnya dari hasil klasifikasi teks tersebut, dilakukan pembuatan fitur sebanyak 50 fitur yang akan digunakan dalam klasifikasi. Namun, 50 fitur tersebut direduksi menggunakan Principal Component Analysis (PCA) menjadi kedalam dua dimensi saja, yaitu PC<sub>1</sub> dan PC<sub>2</sub>. Setelah itu baru dijalankan algoritma Naïve Bayes classifier. Setelah sentimen analisis, dilakukan analisis Wordcloud berdasarkan kelas sentimen yang telah diprediksi dengan model naïve bayes. Selanjutnya dilakukan topik pemodelan dengan menggunakan algoritma Latent Dirichlet Allocation (LDA). Topik yang dibentuk menggunakan eksperimen untuk mengoptimasi jumlah topik yang terbentuk.

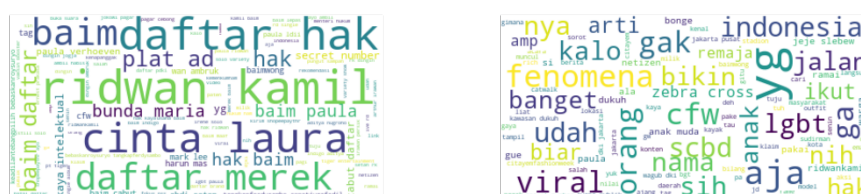
#### IV. HASIL DAN PEMBAHASAN

Data yang diolah terdapat beberapa nilai yang hilang sehingga diperlukan penghapusan data yang hilang. Penurunan dimensi dibutuhkan karena data merupakan data sparse, untuk mendapatkan hasil klasifikasi yang optimal, maka dilakukan PCA menjadi dua variabel atau dua komponen saja. Selanjutnya didapatkan hasil diagram bar sebagai berikut



Gambar 1. Diagram Batang Sentimen Tweet

Dari diagram batang tersebut ternyata tweet Citayam Fashion Week tidak menunjukkan perbedaan sentimen yang signifikan. Tetapi sentimen positif lebih kental dibandingkan sentimen negatif. Selanjutnya dibangkitkan Wordcloud berdasarkan kelas sentimen tersebut dengan hasil berikut.



Gambar 2. Wordcloud Sentimen Negatif (kanan) dan Positif (kiri)

Berdasarkan Wordcloud tersebut, pada bagian negatif dan positif sedikit mempunyai pola yang berbeda. Wordcloud negatif mempunyai kata-kata LGBT, sedangkan di positif tidak ada kata LGBT.



Wordcloud positif didominasi oleh nama-nama orang terkenal seperti Baim Wong, Ridwan Kamil, dan Cinta Laura. Sedangkan Wordcloud negatif didominasi oleh kata-kata seperti scbd, viral, lgbt, fenomena, dan seterusnya.

Selanjutnya dilakukan pemodelan topik menggunakan LDA. Penentuan topik yang terbentuk berdasarkan eksperimen dengan nilai coherence. Semakin besar nilai, maka jumlah topik yang terbentuk semakin optimal. Berdasarkan hasil analisis coherence didapatkan nilai optimal adalah 12 topik. Sehingga akan dibangkitkan atau dibentuk topik sebanyak 12. Berikut adalah 12 topik yang terbentuk dengan menggunakan 5 kata saja didalam modelnya.

$$\begin{aligned}t_0 &= 0.096_{dingin} + 0.074_{tag} + 0.064_{senin} \\t_1 &= 0.188_{hak} + 0.165_{daftar} + 0.031_{confess} \\t_2 &= 0.028_{ikuti} + 0.026_{paten} + 0.025_{gelar} \\t_3 &= 0.100_{trend} + 0.044_{rebut} + 0.041_{outfit} \\t_4 &= 0.206_{stadion} + 0.065_{forever} + 0.027_{cipta} \\t_5 &= 0.102_{banget} + 0.050_{milik} + 0.044_{brand} \\t_6 &= 0.198_{merek} + 0.040_{ridwan_kamil} + 0.039_{bgt} \\t_7 &= 0.059_{fenomena} + 0.038_{bonge} + 0.020_{via_youtube} \\t_8 &= 0.329_{baim} + 0.051_{paki} + 0.014_{pribadi} \\t_9 &= 0.067_{baimwong} + 0.053_{shopee} + 0.037_{scbd} \\t_{10} &= 0.101_{viral} + 0.100_{apem} + 0.035_{video} \\t_{11} &= 0.059_{link} + 0.047_{amp} + 0.042_{pake}\end{aligned}$$

Model yang terbentuk berdasarkan kata yang berkontribusi besar dalam membentuk topik tersebut. Misalnya, topik  $t_0$ , dibentuk dengan kata dingin, tag, dan senin yang merupakan kata yang berkontribusi besar dalam membentuk topik tersebut. Jika kita lihat, topik  $t_8$ , mempunyai kontribusi kata baim dengan nilai yang besar dibandingkan kata-kata yang lain di topik yang lain. Berarti topik 8 kemungkinan membahas tentang baim. Pemodelan topik dengan data tweet dan kata kunci Citayam Fashion Week didominasi oleh subjek yang berperan dalam mengembangkan fenomena tersebut.

## V. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah, sentimen antara positif dan negatif dalam twitter untuk fenomena Citayam Fashion Week adalah berimbang. Jumlah perbedaan tidak begitu signifikan. Pola Wordcloud yang terbentuk antar sentimen berbeda signifikan. Sentimen positif didominasi oleh subjek atau tokoh-tokoh yang terkenal dikalangan masyarakat. Sedangkan sentimen negatif didominasi oleh kata-kata yang memang mempunyai konotasi negatif seperti lgbt. Pembentukan model topik yang optimal didapatkan 12 topik dihitung dari topik 0. Pemodelan pada masing-masing kata tersebut tidak terdapat kata yang overlapping. Artinya LDA mampu memodelkan topik dengan baik. Kata-kata yang membentuk model didominasi oleh tokoh dan juga kata-kata asing yang berhubungan dengan milenial.

## UCAPAN TERIMA KASIH

Terimakasih kepada anak dan istri saya yang telah memberikan support. Beserta dua mahasiswa saya yang menyusun artikel ini.

## REFERENSI

1. Theocharopoulos, Panagiotis & Tsoukala, Anastasia & Georgakopoulos, Spiros & Tasoulis, Sotiris & Plagianakos, Vassilis. (2022). Text Analysis of COVID-19 Tweets. 10.1007/978-3-031-08223-8\_42.
2. Demirel, Apollo. (2022). Voluntary Simplicity: An Exploration Through Text Analysis. International Journal of Consumer Studies. 46. 10.1111/ijcs.12644.



3. Jacqueline C.K. Lam, Lawrence Y.L. Cheung, Shanshan Wang, Victor O.K. Li, Stakeholder concerns of air pollution in Hong Kong and policy implications: A big-data computational text analysis approach, *Environmental Science & Policy*, Volume 101, 2019, Pages 374-382, ISSN 1462-9011, doi: 10.1016/j.envsci.2019.07.007.
4. P. Y. Saputra, D. H. Subhi, and F. Z. A. Winatama, “Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, 2019, doi: 10.33795/jip.v5i4.259.
5. M. G. Pradana, “Penggunaan fitur wordcloud dan document term matrix dalam text mining,” *J. Ilm. Inform.*, vol. 8, no. 1, pp. 38–43, 2020.
6. B. W. Arianto and G. Anuraga, “Topic Modeling for Twitter Users Regarding the ‘Ruanggguru’ Application,” *J. ILMU DASAR*, vol. 21, no. 2, p. 149, 2020, doi: 10.19184/jid.v21i2.17112.
7. A. Daud, J. Li, L. Zhou, and F. Muhammad, “Knowledge discovery through directed probabilistic topic models: A survey,” *Front. Comput. Sci. China*, vol. 4, no. 2, pp. 280–301, 2010, doi: 10.1007/s11704-009-0062-y.