



Analisis Sentimen Penggunaan Kendaraan Listrik terhadap Lingkungan di Indonesia dengan Pendekatan Machine Learning

Muhammad Hafiz Albab¹, Arlita Dwina Firlana Sari², Salma Nabila Asrizal³, Robert Kurniawan⁴

^{1,2,3}Program Studi Statistika, Politeknik Statistika STIS Jakarta, Indonesia
⁴Program Studi Komputasi Statistik, Politeknik Statistika STIS Jakarta, Indonesia

¹212112214@stis.ac.id

²212111923@stis.ac.id

³212112346@stis.ac.id

⁴robertk@stis.ac.id

Corresponding author email: 212112214@stis.ac.id

Abstract: This study analyzes Indonesian public sentiment towards electric vehicles and their environmental impact using data from 1010 Twitter's tweets. The classification methods include SVM with kernels (linear, polynomial, sigmoid, and RBF), Naive Bayes, and Random Forest. Additionally, to overcome imbalance data, SMOTE method was applied. Sentiment analysis results indicate that 62.64% were positive and 37.36% were negative, with Naive Bayes identified as the best classification method, achieving an accuracy rate of 84%. These findings provide important insights for the government to support the adoption of electric vehicles in Indonesia, while emphasizing the use of renewable energy sources. This research also offers a basis for public decision-making regarding the purchase of electric vehicles by considering their contribution to air pollution in Indonesia.

Keywords: Electric vehicles, sentiment, machine learning, environment .

Abstrak: Penelitian ini menganalisis sentimen masyarakat Indonesia terhadap kendaraan listrik dan dampaknya terhadap lingkungan menggunakan data dari Twitter sebanyak 1010 data. Metode klasifikasi yang digunakan meliputi SVM dengan kernel (linear, polinomial, sigmoid, dan RBF), Naive Bayes, dan Random Forest. Selain itu, untuk mengatasi ketidakseimbangan data, diterapkan metode SMOTE. Hasil analisis sentimen menunjukkan bahwa terdapat 62,64% sentimen positif dan 37,36% sentimen negatif dengan metode klasifikasi terbaik adalah Naive Bayes yang memiliki tingkat akurasi 84%. Temuan ini memberikan wawasan penting bagi pemerintah untuk dapat mendukung penggunaan kendaraan listrik di Indonesia dengan tetap memperhatikan penggunaan sumber energi terbarukan. Penelitian ini juga memberikan dasar pengambilan keputusan bagi masyarakat terkait pembelian kendaraan listrik dengan mempertimbangkan kontribusinya terhadap polusi udara di Indonesia.

Kata kunci: Kendaraan listrik, sentimen, machine learning, lingkungan

I. PENDAHULUAN

Polusi udara merupakan permasalahan utama lingkungan karena berdampak pada kesehatan [1]. Penyumbang utama polusi udara salah satunya adalah sektor transportasi, terutama transportasi darat [2]. Menurut Climate Transparency Report 2020, sektor transportasi menyumbang 27% polusi udara. Pada gambar 1 terlihat bahwa tahun 2022, jumlah kendaraan bermotor di Indonesia meningkat mencapai lebih dari 148 juta unit [3]. Angka ini terus meningkat sejak tahun 2015.



Gambar 1. Perkembangan jumlah kendaraan bermotor tahun 2015-2022 (juta)



Peningkatan jumlah kendaraan bermotor menyebabkan emisi jelaga dari bensin dan solar menghasilkan semakin banyak partikel berbahaya yang terhirup sehingga berdampak buruk pada lingkungan dan kesehatan manusia [4]. Partikel yang dihasilkan oleh kendaraan bermotor seperti PM_{2.5} dan PM₁₀ memiliki dampak yang serius. Partikel-partikel ini dapat masuk ke saluran pernapasan, menyebabkan masalah pernapasan, iritasi mata, dan meningkatkan risiko penyakit jangka panjang seperti penyakit jantung dan paru-paru, bahkan hingga kematian. Dalam jangka panjang, memburuknya polusi udara dikhawatirkan dapat menghambat pencapaian target pembangunan jangka panjang Indonesia, yaitu Indonesia Emas 2045 tentang kesehatan [2].

Kekhawatiran atas partikel yang dihasilkan oleh kendaraan bermotor menyebabkan minat terhadap elektrifikasi dalam mobilitas meningkat pesat [5]. Kendaraan listrik merupakan salah satu solusi untuk mengurangi emisi [6]. Kendaraan listrik bertenaga baterai dianggap sebagai teknologi yang ramah lingkungan karena tidak menghasilkan emisi saat digunakan. Penggunaan kendaraan listrik merupakan salah satu upaya yang dapat dilakukan untuk mengurangi penggunaan bahan bakar untuk transportasi dan menggantinya dengan listrik [7]. Ini bisa memberikan kontribusi signifikan dalam mengurangi polusi udara lokal dan kemungkinan mengurangi emisi dalam sektor transportasi. Selain itu, dengan beralih menggunakan kendaraan listrik maka dapat meningkatkan keamanan pasokan energi dengan mengurangi ketergantungan pada impor bahan bakar fosil.

Hal ini menimbulkan persepsi baru masyarakat atas penggunaan kendaraan listrik. Namun, di sisi lain, bahan dasar pembuatan listrik di Indonesia masih menggunakan batu bara. Penggunaan batu bara yang berlebihan akan menimbulkan kekhawatiran terutama mengenai emisi gas rumah kaca [8]. Beberapa alasan penggunaan tersebut memunculkan berbagai persepsi masyarakat atas maraknya penggunaan kendaraan listrik di Indonesia. Adapun penelitian mengenai sentimen penggunaan kendaraan listrik telah dilakukan oleh beberapa peneliti.

Penelitian oleh Suryani terkait sentimen terhadap mobil listrik mengumpulkan 680 data. Dari data tersebut, 20% atau 139 data menyatakan sentimen positif, 71% atau 485 data menyatakan sentimen netral, dan 9% atau 59 data menyatakan sentimen negatif. Penelitian tersebut menggunakan algoritma Naive Bayes Classifier dan Support Vector Machine (SVM) dan menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dibandingkan Naive Bayes [6]. Penelitian lain yakni dari Alhari tentang analisis sentimen terkait kendaraan listrik di Indonesia menggunakan metode SVM. Dari penelitian tersebut, diperoleh akurasi sebesar 87%. Hasil analisis sentimen yang diperoleh sebesar 79% memberikan sentimen positif, sedangkan 21% memberikan sentimen negatif dari total 1502 opini Tweet yang terakumulasi tentang kehadiran teknologi mobil listrik di Indonesia [9]. Selain itu, penelitian lain oleh Rahat terkait analisis sentimen ulasan maskapai penerbangan menggunakan metode SVM dan Naive Bayes memperoleh akurasi terbaik pada metode SVM yaitu sebesar 82%. Penelitian tersebut menggunakan 10.000 data tweet [10].

Berdasarkan penelitian-penelitian terdahulu, kebaruan penelitian ini adalah analisis sentimen mobil listrik yang lebih difokuskan dampaknya untuk lingkungan. Penelitian ini membandingkan tiga metode klasifikasi yaitu Support Vector Machine dengan Kernel linear, Polinomial, Sigmoid, dan RBF, serta metode lain seperti Naive Bayes dan Random Forest. Selain itu, pada penelitian ini dilakukan metode SMOTE untuk mengatasi jumlah data imbalance. Untuk mengetahui apakah kendaraan listrik dapat menjadi alternatif transportasi yang menarik bagi pengguna kendaraan pribadi, mengurangi emisi dan polusi udara, serta mendukung peralihan ke energi hijau, diperlukan analisis sentimen tentang kendaraan listrik dan efeknya terhadap lingkungan.



Penelitian ini bertujuan untuk mengetahui berbagai opini masyarakat tentang keberadaan kendaraan listrik, baik opini positif ataupun negatif dan membandingkan metode klasifikasi seperti SVM, Naive Bayes, dan Random Forest. Dalam penelitian ini diharapkan dapat menjadi dasar pengambilan keputusan bagi masyarakat untuk membeli mobil listrik mengingat dampaknya dari sisi lingkungan yang pada akhirnya dapat mengurangi atau meningkatkan polusi udara di Indonesia, khususnya di kota-kota besar. Selain itu, dalam konteks perencanaan pembangunan diharapkan penelitian ini dapat memberikan gambaran bagi pemerintah, khususnya Kementerian PPN/Bappenas, dalam merancang Rencana Pembangunan Jangka Panjang Nasional (RPJPN) untuk periode 2025-2045 terkait kesehatan masyarakat jika penggunaan mobil listrik terus meningkat di Indonesia. Penelitian ini juga membuka wawasan penting bagi industri kendaraan listrik dalam mengembangkan kebijakan yang mendukung penggunaan energi hijau dan perbaikan kualitas udara di Indonesia.

II. TINJAUAN PUSTAKA

2.1. Text Mining

Text mining adalah proses menemukan informasi dalam jumlah teks yang besar dan secara otomatis mengidentifikasi pola serta hubungan menarik dalam data teks [11]. Dalam *text mining*, sumber data utama untuk menemukan pola menarik tersebut merupakan data teks yang tidak terstruktur [12].

2.2. Data Crawling

Data crawling merupakan proses sistem untuk mengumpulkan dan mengunduh data dari berbagai halaman web yang biasanya digunakan oleh mesin pencari untuk mengumpulkan, mengindeks, dan memungkinkan pengguna mencari informasi di internet berdasarkan *query* yang diinginkan [13]. Dalam analisis sentimen, *data crawling* berperan penting dalam mengumpulkan opini dan ulasan dari berbagai platform seperti media sosial, situs berita, dan ulasan konsumen.

2.3. Analisis Sentimen

Perkembangan aplikasi berbasis internet, seperti media sosial, telah menghasilkan banyak komentar dan ulasan yang mengekspresikan pandangan dan opini para pengguna internet. Untuk dapat menganalisis opini tersebut, digunakan analisis sentimen, yang juga dikenal sebagai *opinion analysis* atau *opinion mining*. Analisis sentimen adalah proses mengumpulkan dan menganalisis opini, pikiran, dan kesan orang mengenai berbagai topik, produk, subjek, dan layanan [14].

2.4. Machine Learning

Machine learning mempelajari bagaimana sebuah program dapat meningkatkan kinerjanya berdasarkan data [15]. Contoh penerapan seperti mengenali pola-pola kompleks dan membuat keputusan berdasarkan data biasa dilakukan dengan klasifikasi dan clustering. Menurut penelitian Han, penerapan *machine learning* yang berkaitan dengan *data mining*, antara lain *supervised learning* yang biasanya berkaitan dengan klasifikasi, *unsupervised learning* yang biasanya berkaitan dengan *clustering*, *semi-supervised learning*, dan *active learning* [15].

2.5. Hubungan Kendaraan Listrik dengan Lingkungan

Penelitian oleh Hu menggunakan model COPERT IV untuk menganalisis emisi kendaraan menemukan hasil bahwa adopsi *electric vehicle* atau kendaraan bermotor (EV) memiliki potensi besar untuk mengurangi emisi kendaraan bermotor secara signifikan dan memperbaiki kualitas udara perkotaan di Shanghai [16]. Penelitian mengenai penilaian siklus hidup (*Life Cycle Assessment* atau LCA) dari EVs dan kendaraan dengan mesin pembakaran internal (*Internal Combustion Engine Vehicles* atau ICEVs) di Hong Kong yang dilakukan oleh Shafique memberikan hasil bahwa EV dengan

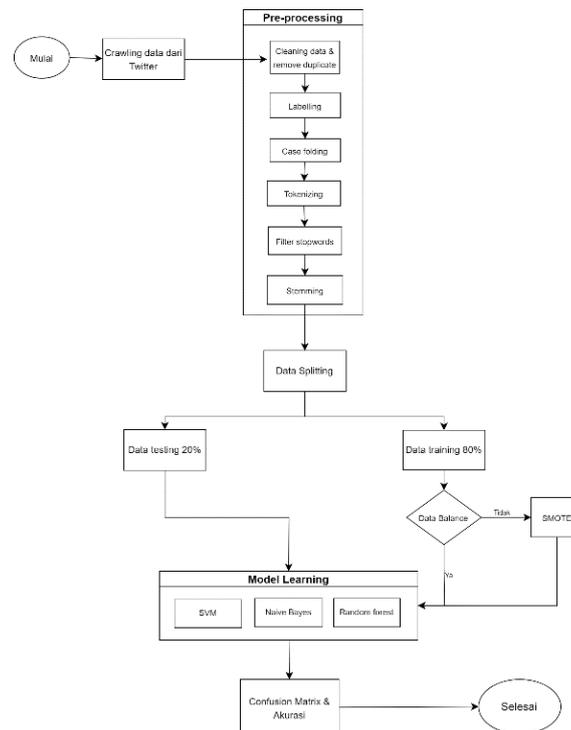
campuran energi tahun 2050 merupakan pilihan yang optimal dengan dampak lingkungan paling sedikit dan kendaraan ICEV berbahan bakar bensin memiliki dampak lingkungan terbesar [17]. Namun, pada penelitian Tang, meskipun *Battery Electric Vehicles* (BEVs) memiliki potensi untuk mengurangi emisi CO₂ secara signifikan dibandingkan dengan *Internal Combustion Engine Vehicles* (ICEVs), efektivitas pengurangan emisi CO₂ dari BEVs sangat bergantung pada campuran pembangkit listrik regional karena di daerah dengan penetrasi energi terbarukan yang tinggi, BEVs menunjukkan kinerja yang lebih baik dalam pengurangan emisi karbon tetapi di daerah dengan dominasi pembangkit listrik tenaga batu bara, BEVs dapat menghasilkan lebih banyak emisi CO₂ dibandingkan ICEVs [18].

III. METODE PENELITIAN

Penelitian ini menggunakan analisis sentimen dengan metode klasifikasi yang dilakukan adalah algoritma *machine learning*, yaitu SVM, Naive Bayes, dan Random Forest. Dalam penelitian ini, digunakan platform media sosial Twitter sebagai sumber untuk memperoleh *dataset* yang berhubungan dengan topik yang diteliti. Data yang diambil merupakan data dari bulan Januari 2023 sampai Mei 2024 dan menghasilkan hasil crawling sebanyak 1410 cuitan.

3.1. Tahapan Penelitian

Gambar 2 mengilustrasikan kerangka analisis yang dilakukan dalam penelitian ini.



Gambar 2. Diagram Alur Tahapan Penelitian

Tahapan-tahapan yang digunakan pada penelitian ini sesuai dengan Gambar 2 dan akan dipaparkan lebih lanjut sebagai berikut.

1. Crawling

Pengambilan data dari Twitter menggunakan metode *crawling* dengan alat Google Collab. Kemudian *tweet* atau cuitan para pengguna Twitter yang diperoleh merupakan pendapat masyarakat mengenai kendaraan listrik. Proses pengambilan data di penelitian ini terhubung ke akun twitter



dengan *query* "Kendaraan listrik, Polusi", "Kendaraan listrik, Lingkungan". Jenis pencarian yang digunakan adalah yang terbaru atau populer.

2. *Cleaning Data dan Remove Duplicate*

Tahap ini merupakan tahapan untuk menghilangkan data-data dan juga karakter tertentu, seperti tanda baca dan simbol, yang tidak sesuai dengan konteks yang ingin kita analisis. Setelah dilakukan tahapan ini, data 1410 cuitan berkurang menjadi 1010 cuitan.

3. *Labelling*

Pelabelan data merupakan suatu tahapan untuk menentukan respon *tweet* di kumpulan data. Opini atau sentimen di kumpulan data terbagi menjadi tiga, yaitu opini positif, netral, dan negatif. Namun, opini netral berarti tidak mengandung keduanya, baik negatif atau positif sehingga opini netral ini tidak digunakan dalam tahapan setelah *pre-processing*.

4. *Case Folding, Tokenization, Filter Stopwords, Stemming, dan Splitting Data*

Case Folding merupakan proses mengubah gaya huruf pada suatu kalimat menjadi huruf kecil. Tokenization merupakan tahap pemecahan suatu kalimat menjadi satuan-satuan kata. Filter Stopwords merupakan langkah untuk menghapus kata-kata yang dianggap tidak memiliki makna dalam dokumen *tweet* yang digunakan dalam penelitian ini. Stemming digunakan untuk menemukan kata dasar dengan mengeliminasi imbuhan seperti awalan, sisipan, akhiran, dan kombinasi awalan dan akhiran dari kata. Dataset yang telah melalui tahapan *pre-processing* akan dibagi (*split*) menjadi dua bagian, yaitu 80% untuk data training dan 20% untuk data testing.

3.2. *Klasifikasi Data*

Pada penelitian ini, klasifikasi data dilakukan dengan pendekatan Machine Learning, yaitu SVM, Naive Bayes, dan Random Forest yang dipaparkan lebih lanjut sebagai berikut.

3.2.1. Support Vector Machine (SVM)

SVM merupakan metode klasifikasi *supervised learning* yang sering digunakan dalam analisis sentimen. SVM memiliki keakuratan yang tinggi, karena kemampuan mereka untuk memodelkan batas-batas keputusan non-linier yang kompleks [15]. Tujuan SVM adalah untuk menemukan *hyperplane* yang optimal. Menurut Wang, metode SVM untuk analisis sentimen akan memiliki kinerja yang lebih baik pada dataset dengan dua kelas, yaitu positif dan negatif [20].

Pada metode SVM, klasifikasi dapat dilakukan secara linear dan non-linear. Klasifikasi non-linear perlu menggunakan fungsi kernel yang memproyeksikan titik dataset ke dimensi ruang yang lebih tinggi [21]. Beberapa fungsi kernel, antara lain Linear, Polinomial, Radial Basis Function (RBF), Sigmoid. SVM dengan kernel linear hanya memiliki satu parameter yang dapat dioptimalkan pemilihannya, yaitu parameter Cost (C) [22]. Nilai parameter Cost dapat mengontrol kesalahan training, kesalahan testing, jumlah support vector, dan margin SVM [23]. Kemudian, Kernel RBF dan Kernel Sigmoid memiliki parameter Cost (C) dan Gamma (γ). Parameter Gamma (γ) pada Kernel RBF digunakan untuk mengatur seberapa luas atau sempit pengaruh dari titik data individu pada ruang fitur dan pada Kernel Sigmoid Gamma (γ) adalah parameter skala yang mengatur data input [24]. Sedangkan pada Kernel Polinomial, pemilihan degree menjadi sangat penting. Pada penelitian Tharwat, degree yang terlalu tinggi dapat menyebabkan model overfits dan error rate yang tinggi sedangkan degree yang terlalu rendah dapat menyebabkan tingginya bias [23].

Dalam analisis sentimen, SVM dapat menghasilkan akurasi yang baik karena memiliki kemampuan yang baik dalam menangani dataset yang kompleks dan memisahkan data yang tidak linear. Menurut penelitian Kanakaraddi yang membandingkan beberapa algoritma machine



learning, yaitu SVM, Naive Bayes, Max Entropy, LSTM, CNN, dan Random Forest, diperoleh hasil bahwa SVM menghasilkan akurasi paling tinggi, yaitu sebesar 79,9% [25].

3.2.2. Naïve Bayes

Naive Bayes merupakan metode probabilistik untuk mengklasifikasikan teks. Algoritma ini didasarkan pada Teorema Bayes dengan asumsi independensi antar data. Teorema ini digunakan untuk menghitung probabilitas posterior dari kelas berdasarkan bukti yang diberikan. Jika $P\{A|B\}$ adalah probabilitas peristiwa A dengan syarat peristiwa B, Gorunescu mendefinisikan rumus Bayes yang disederhanakan dengan [19]:

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

Dimana,

$P\{B|A\}$ = probabilitas posterior dari kelas B yang diberikan fitur A.

$P\{A|B\}$ = probabilitas likelihood dari fitur A yang diberikan kelas B

$P\{B\}$ = probabilitas prior dari kelas B.

$P\{A\}$ = probabilitas total dari fitur A.

Naive Bayes sering digunakan untuk analisis sentimen karena Naive Bayes merupakan metode yang sederhana dan dapat menghasilkan efisiensi yang tinggi. Berdasarkan penelitian Nayak, yang melakukan analisis sentimen tentang review film pada aplikasi Twitter, metode Naive Bayes mendapatkan akurasi paling baik, yaitu sebesar 89% dibandingkan dengan SVM sebesar 88% dan Random Forest sebesar 85% [26].

3.2.3. Random Forest

Random Forest merupakan metode ensemble learning yang menggunakan kombinasi dari banyak Decision Tree untuk meningkatkan akurasi klasifikasi. Setiap Decision Tree di dalam Random Forest dibuat dengan memilih atribut secara acak di setiap node sehingga setiap tree bergantung pada nilai acak yang dihasilkan secara independen untuk semua Decision Tree. Ketika melakukan klasifikasi, setiap decision tree memiliki masing-masing prediksi dan hasil yang paling baik akan dipilih oleh pohon-pohon tersebut untuk menjadi hasil akhir [15].

3.3. Evaluasi Model

Untuk mengevaluasi model klasifikasi yang dihasilkan, salah satu alat yang dapat digunakan adalah confusion matrix yang dapat dilihat pada Tabel 1. Elemen-elemen dalam tabel confusion matrix terdiri dari True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN).

Tabel 1. Confusion Matrix

		Predicted Class		Total
		Yes	No	
Actual Class	Yes	TP	FN	P
	No	FP	TN	N
Total		P'	N'	P+N

TP dan TN merupakan ukuran yang melihat data yang diklasifikasikan dengan benar, sedangkan FP dan FN merupakan ukuran saat diklasifikasikan dengan salah. Kemudian, ukuran evaluasi yang umum digunakan adalah dengan menghitung akurasi, presisi, recall, dan F1-Score yang diperoleh melalui formula berikut [15] :

$$\text{Akurasi} = \frac{TP+TN}{P+N} \quad (2)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (3)$$



$$\text{Recall} = \frac{TP}{P} \tag{4}$$

$$\text{F1-score} = \frac{2 \times \text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \tag{5}$$

3.4. Synthetic Minority Over- Sampling Technique (SMOTE)

Langkah penggunaan SMOTE bertujuan untuk meningkatkan nilai akurasi hasil klasifikasi algoritma Naive Bayes, Support Vector Machine (SVM), dan Random Forest karena SMOTE akan menyeimbangkan data training yang sebelumnya *imbalance*. Algoritma SMOTE dilakukan dengan menambahkan data baru yang dihasilkan secara artifisial pada kelas minoritas sehingga jumlah data di kedua kelas menjadi lebih seimbang [27]. Dalam *supervised classification*, *imbalance* data terjadi ketika probabilitas priori antar dua kelas berbeda secara signifikan sehingga terdapat kelas minoritas dengan jumlah kejadian yang relatif lebih sedikit dibandingkan kelas mayoritas [28].

IV. HASIL DAN PEMBAHASAN

Tahapan awal sebelum pembentukan model, yaitu melakukan preprocessing (cleaning dan case folding data, melakukan labelling data menurut sentimen, tokenization, filter stopwords, dan stemming). Berikut beberapa contoh hasil proses cleaning dan case folding data.

Tabel 2. Contoh Proses Cleaning dan Case Folding Data

No	Data Asli	Cleaning dan Case Folding Data
1	@IDFreeport Mobil listrik memang menjadi pilihan yang ramah lingkungan dengan kontribusi besar dalam mengurangi emisi karbon.	mobil listrik menjadi pilihan yang ramah lingkungan dengan kontribusi besar dalam mengurangi emisi karbon
2	Menggali pertanyaan krusial tentang keberlanjutan mobil listrik dalam konteks tantangan lingkungan global. Apakah mobil listrik menjadi solusi efektif dalam menghadapi krisis lingkungan? #userstory https://t.co/nSm0kytyzN	menggali pertanyaan krusial tentang keberlanjutan mobil listrik dalam konteks tantangan lingkungan global apakah mobil listrik menjadi solusi efektif dalam menghadapi krisis lingkungan
3	@rsyarief @JeblukAkun Perlu diketahui kalau yg kelihatan bersih itu justru yg palÃ-ng mencemari. Emg bensin oktan tinggi kek pertamax atau mobil listrik dsb itu ramah lingkungan tpi proses pembuatannya jauh lebih merusak lingkungan.	perlu diketahui kalau yang kelihatan bersih itu justru yang paling mencemari dan memang benar bensin oktan tinggi seperti pertamax atau mobil listrik dan sebagainya itu proses pembuatannya jauh lebih merusak lingkungan

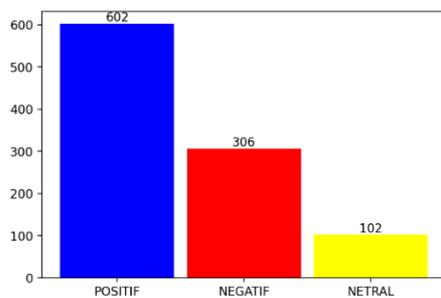
Dari 1410 cuitan data yang didapatkan, ditemukan sekitar 300 cuitan data yang mengalami *duplicate* dan terdapat beberapa cuitan yang tidak sesuai dengan konteks lingkungan, seperti promosi, ujaran kebencian, bahkan terkadang sama sekali tidak membahas tentang kendaraan listrik, sehingga data yang digunakan hanya 1010 cuitan. Setelah itu, berdasarkan tabel 2 di atas, dilakukan cleaning dan case folding data dengan menghilangkan beberapa kata atau simbol yang tidak penting, yaitu mention, tagar, titik, tautan, dan tanda baca serta mengganti beberapa huruf kapital menjadi huruf non kapital. Tahapan selanjutnya, yaitu melakukan labelling data menurut 3 sentimen dengan sentimen positif dikodekan menjadi 2, sentimen negatif dikodekan menjadi 1, dan sentimen netral dikodekan menjadi 0 seperti tabel 3 di bawah ini.

Tabel 3. Contoh Pengklasifikasian Labelling Data Menurut Sentimen

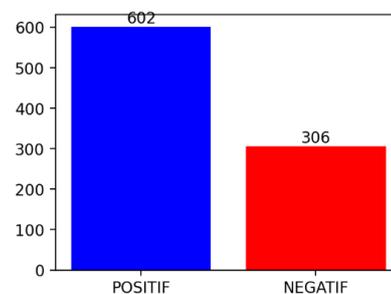
No	Cleaning Data	Labelling Sentimen	Kode
1	mobil listrik menjadi pilihan yang ramah lingkungan dengan kontribusi besar dalam mengurangi emisi karbon	Positif	2

2	menggal pertanyaan krusial tentang keberlanjutan mobil listrik dalam konteks tantangan lingkungan global apakah mobil listrik menjadi solusi efektif dalam menghadapi krisis lingkungan	Netral	1
3	perlu diketahui kalau yang kelihatan bersih itu justru yang paling mencemari dan memang benar bensin oktan tinggi seperti pertamax atau mobil listrik dan sebagainya itu proses pembuatannya jauh lebih merusak lingkungan	Negatif	0

Berdasarkan tabel 3 di atas, kata yang mengandung “ramah lingkungan” atau “mengurangi emisi karbon” dikategorikan menjadi sentimen positif. Untuk kata yang masih mengandung pertanyaan, seperti “apakah” dalam menanggapi isu lingkungan dikategorikan menjadi sentimen netral. Selanjutnya, kata yang mengandung “paling mencemari” atau “merusak lingkungan” dalam proses pembuatannya ataupun dalam proses pengisian menggunakan listrik karena di Indonesia masih menggunakan batubara dikategorikan menjadi sentimen negatif.



Gambar 3a.



Gambar 3b.

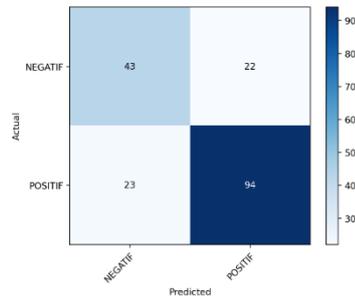
Gambar 3. Jumlah Pengklasifikasian Labelling data menurut Sentimen : (3a) sebelum menghilangkan netral (3b) setelah menghilangkan netral

Berdasarkan gambar 3a di atas, diperoleh pengklasifikasian data hasil *labelling* sentimen, yaitu dari 1010 cuitan sebanyak 306 sentimen negatif, 102 sentimen netral, dan 602 sentimen positif. Namun, penelitian ini hanya akan difokuskan pada sentimen positif dan negatif dalam melakukan klasifikasi lebih lanjut setelah sentimen netral dihilangkan, sehingga hanya terdapat 602 atau 66,3% sentimen positif dan 306 atau 33,7% sentimen negatif seperti yang terlihat pada gambar 3b dengan total data yang digunakan sebesar 908 cuitan. Tahap berikutnya melakukan preprocessing dengan cara tokenization, filter stopwords, dan stemming seperti tabel 4 dibawah ini.

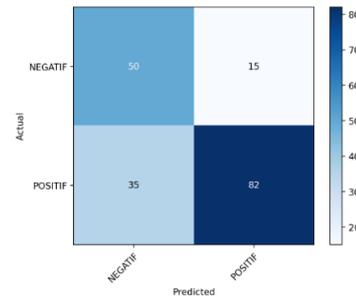
Tabel 4. Contoh Hasil Tahapan Preprocessing

No	Tahapan	Hasil yang Diperoleh
1	Cleaning data	perlu diketahui kalau yang kelihatan bersih itu justru yang paling mencemari dan memang benar bensin oktan tinggi seperti pertamax atau mobil listrik dan sebagainya itu proses pembuatannya jauh lebih merusak lingkungan
2	Tokenization	['perlu', 'diketahui', 'kalau', 'yang', 'kelihatan', 'bersih', 'itu', 'justru', 'yang', 'paling', 'mencemari', 'dan', 'memang', 'benar', 'bensin', 'oktan', 'tinggi', 'seperti', 'pertamax', 'atau', 'mobil', 'listrik', 'dan', 'sebagainya', 'itu', 'proses', 'pembuatannya', 'jauh', 'lebih', 'merusak', 'lingkungan']
3	Filter Stopwords	[kelihatan', 'bersih', 'mencemari', 'bensin', 'oktan', 'pertamax', 'mobil', 'listrik', 'proses', 'pembuatannya', 'merusak', 'lingkungan']
4	Stemming	lihat bersih cemar bensin oktan pertamax mobil listrik proses buat rusak lingkungan

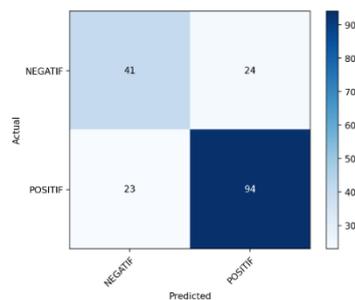
Berdasarkan tabel 4 di atas, Pada tahap tokenization, data dalam bentuk kalimat dikonversi dalam bentuk beberapa kata. Setelah itu, pada tahap Filter Stopwords dilakukan penghilangan kata yang dianggap tidak memiliki arti, seperti kata “yang”, “seperti”, “itu”, “sebagainya”, dan yang lainnya. Pada



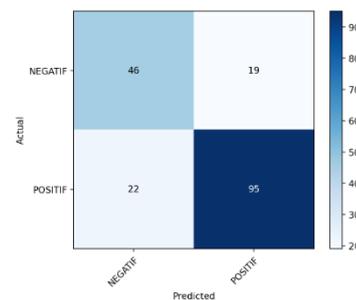
Gambar 6a.



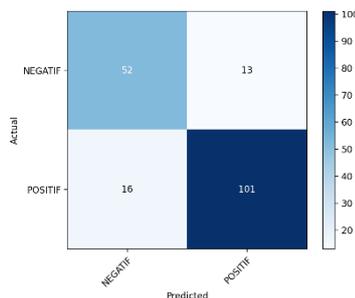
Gambar 6b.



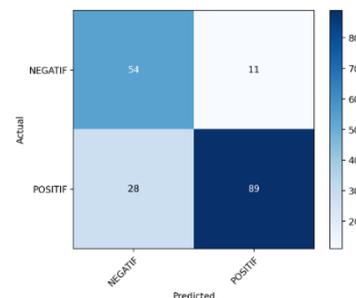
Gambar 6c.



Gambar 6d.



Gambar 6e.



Gambar 6f.

Gambar 6. Confusion Matrix dari masing-masing model : (6a) SVM Linear (6b) SVM Polinomial (6c) SVM Sigmoid (6d) SVM RBF (6e) Naive Bayes (6f) Random Forest

Berdasarkan gambar 6 di atas, kategori True Positive (TP) dengan jumlah terbanyak diperoleh oleh model Random Forest, yaitu sebesar 54. Artinya, data yang bersentimen negatif terprediksi sebagai sentimen negatif sebesar 54. Selanjutnya, kategori True Negative (TN) dengan jumlah terbanyak diperoleh oleh model Naive Bayes, yaitu sebesar 101. Artinya, data yang bersentimen positif terprediksi sebagai sentimen positif sebesar 101. Maka dapat disimpulkan bahwa model Naive Bayes merupakan model dengan prediksi data sentimen yang lebih baik dibandingkan kelima model lainnya karena hanya 29 data yang terprediksi salah, yaitu baik berupa False Negative (data bersentimen negatif diprediksi menjadi sentimen positif) maupun False Positive (data bersentimen positif diprediksi menjadi sentimen negatif). Oleh karena itu, disimpulkan bahwa berdasarkan model Naive Bayes, hasil prediksi data sentimen positif berjumlah 114 (62,64% dari total keseluruhan data testing) dan data sentimen negatif berjumlah 68 (37,36% dari total keseluruhan data testing).

Performa masing-masing model yang digunakan untuk klasifikasi, yaitu SVM dengan kernel linear (SVM-L), SVM dengan kernel polinomial (SVM-P), SVM dengan kernel Radial Basis Function (SVM-RBF), SVM dengan kernel Sigmoid (SVM-S), Naïve Bayes (NB), dan Random Forest (RF) dapat dilihat pada tabel 5 berikut.



Tabel 5. Perbandingan Evaluasi Model SVM, Naive Bayes, dan Random Forest

Model	Ukuran Evaluasi			
	Akurasi	Presisi	Recall	F1-Score
SVM-L	75,27%	75,36%	75,27%	75,32%
SVM-P	72,53%	75,35%	72,53%	73,08%
SVM-S	74,18%	74,09%	74,18%	74,13%
SVM-RBF	77,47%	77,73%	77,47%	77,58%
NB	84,07%	84,27%	84,07%	84,14%
RF	78,57%	80,73%	78,57%	78,97%

Berdasarkan tabel 5 di atas, dari keenam model yang digunakan, dapat disimpulkan bahwa model Naïve Bayes (NB) merupakan model dengan prediksi data sentimen terbaik karena berdasarkan beberapa ukuran evaluasi, model yang dihasilkan dengan metode Naïve Bayes memiliki nilai yang paling besar daripada model dengan metode lainnya, yaitu dengan akurasi sebesar 84,07%, presisi sebesar 84,27%, recall sebesar 84,07%, dan F1-score sebesar 84,14%. Perlu diketahui bahwa, nilai akurasi yang lebih dari 80% hanya diperoleh oleh model Naïve Bayes (NB), sedangkan akurasi yang terkecil diperoleh oleh model SVM Polynomial (SVM-P), yaitu sebesar 72,53%.

Berdasarkan hasil yang didapatkan, sentimen positif mengenai efisiensi mobil listrik terhadap lingkungan tergolong tinggi jika dibandingkan sentimen negatif. Hal ini seperti penelitian oleh Alhari tentang sentiment terkait kendaraan listrik di Indonesia menggunakan metode SVM yang memperoleh 79% sentimen positif dan sisanya sentimen negatif [9]. Penelitian yang dilakukan oleh Qin tentang sentiment masyarakat China terhadap kendaraan listrik dengan 90,11% memiliki sentimen positif, 9,13% memiliki sentimen negatif, dan sisanya netral [29]. Selanjutnya, metode klasifikasi terbaik yang didapatkan pada penelitian ini, yaitu metode Naive Bayes dengan akurasi 84,07%. Penelitian berbeda mengenai topik review film pada aplikasi twitter oleh Nayak, metode Naive Bayes mendapatkan akurasi paling baik, yaitu sebesar 89% dibandingkan dengan SVM dan Random Forest [25].

V. KESIMPULAN

Berdasarkan metode terbaik yang diperoleh, hasil menunjukkan bahwa sebesar 62,64% sentimen positif dan 37,36% sentimen negatif dengan nilai akurasi sebesar 84,07%, presisi sebesar 84,27%, recall sebesar 84,07%, dan F1-score sebesar 84,14%. Hal ini berarti masyarakat Indonesia memiliki perspektif yang baik mengenai adopsi kendaraan listrik sebagai solusi untuk mengurangi emisi dan meningkatkan keberlanjutan lingkungan.

Dengan demikian, untuk dapat mengurangi polusi dari kendaraan berbahan bakar bensin yang kian mengkhawatirkan, pemerintah dapat mendukung penggunaan kendaraan listrik, baik sebagai kendaraan pribadi maupun kendaraan umum, seperti memberikan subsidi untuk pembelian kendaraan listrik dan membangun infrastruktur transportasi publik yang menggunakan listrik. Namun, mengingat bahan bakar utama kendaraan listrik masih dihasilkan dari batu bara yang juga dapat mencemari lingkungan, pemerintah diharapkan dapat mendorong penggunaan sumber energi terbarukan untuk menghasilkan listrik, seperti tenaga surya. Kemudian, diharapkan masyarakat dapat mengambil keputusan yang bijak terkait pembelian kendaraan listrik dan mengetahui kontribusinya terhadap polusi udara di Indonesia. Selain itu, untuk penelitian selanjutnya diharapkan dapat menggunakan data yang lebih banyak. Kemudian, dikarenakan metode yang digunakan pada penelitian ini hanya metode supervised learning sehingga penelitian selanjutnya dapat dilakukan dengan membandingkan metode deep learning, seperti metode Convolutional Neural Network (CNN) ataupun Long Short Term Memory (LSTM).



UCAPAN TERIMA KASIH

Terima kasih kepada semua pihak yang telah terlibat dalam penulisan penelitian ini.

REFERENSI

1. S. C. Anenberg, J. Miller, D. Henze, and R. A. Y. Minjares, “Pollution-related health impacts of transportation sector emission in 2010 and 2015,” *ICCT, Clim. Clean Air Coal.*, pp. 1–55, 2019.
2. J. Syahputri, E. B. Suarga, I. Rahman, T. N. Zahari, and D. A. Ramdani, *Dampak Polusi Udara dari Transportasi terhadap Kesehatan di Indonesia; Analisis dan Rekomendasi untuk Pengambil Kebijakan*. 2023.
3. BPS, “perkembangan jumlah kendaraan bermotor menurut jenis unit,” 2022. <https://www.bps.go.id/statistics-table/2/NTcjMg==/perkembangan-jumlah-kendaraan-bermotor-menurut-jenis--unit-.html>.
4. K. Arole *et al.*, “Impacts of particles released from vehicles on environment and health,” *Tribol. Int.*, vol. 184, no. March, p. 108417, 2023, doi: 10.1016/j.triboint.2023.108417.
5. A. Ajanovic and R. Haas, “Electric vehicles: solution or new problem?,” *Environ. Dev. Sustain.*, vol. 20, no. June, pp. 7–22, 2018, doi: 10.1007/s10668-018-0190-3.
6. S. Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, “Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm,” *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 1–9, 2023, doi: 10.57152/predatecs.v1i1.814.
7. A. Suwandono and L. Nadya, “Observing the Correlation of the Use of Electric Vehicles with Increasing the Degree of Public Health in Indonesia,” *Proc. Int. Conf. Democr. Natl. Resil. (ICDNR 2021)*, vol. 620, no. Icdnr, pp. 74–78, 2022, doi: 10.2991/assehr.k.211221.013.
8. D. H. D. Rocha, D. S. Siqueira, and R. J. Silva, “Effects of coal compositions on the environment and economic feasibility of coal generation technologies,” *Sustain. Energy Technol. Assessments*, vol. 47, no. February, 2021, doi: 10.1016/j.seta.2021.101500.
9. M. I. Alhari, O. N. Pratiwi, and M. Lubis, “Sentiment Analysis of The Public Perspective Electric Cars in Indonesia Using Support Vector Machine Algorithm,” *2022 Int. Conf. Sci. Inf. Technol. Smart Adm. ICSINTESA 2022*, no. February, pp. 155–160, 2022, doi: 10.1109/ICSINTESA56431.2022.10041604.
10. A. M. Rahat, A. Kahir, and A. K. M. Masum, “Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset,” *Proc. 2019 8th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2019*, pp. 266–270, 2020, doi: 10.1109/SMART46866.2019.9117512.
11. C. C. Aggarwal and C. X. Zhai, *Mining text data*, vol. 9781461432234. 2013.
12. R. Feldman and J. Sanger, *Advanced approaches in analyzing unstructured data*. 2018.
13. C. Olston and M. Najork, *Web crawling*, vol. 4, no. 3. 2010.
14. M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022.
15. J. Han, M. Kamber, and J. Pei, *Data mining concepts and techniques*. 2016.
16. X. Hu, N. Chen, N. Wu, and B. Yin, “The potential impacts of electric vehicles on urban air quality in Shanghai city,” *Sustain.*, vol. 13, no. 2, pp. 1–12, 2021, doi: 10.3390/su13020496.
17. M. Shafique, A. Azam, M. Rafiq, and X. Luo, “Life cycle assessment of electric vehicles and internal combustion engine vehicles: A case study of Hong Kong,” *Res. Transp. Econ.*, vol. 91, no. June, p. 101112, 2022, doi: 10.1016/j.retrec.2021.101112.
18. B. Tang, Y. Xu, and M. Wang, “Life Cycle Assessment of Battery Electric and Internal Combustion Engine Vehicles Considering the Impact of Electricity Generation Mix: A Case Study in China,” *Atmosphere (Basel)*, vol. 13, no. 2, 2022, doi: 10.3390/atmos13020252.
19. F. Gorunescu, *Data mining concepts, models and techniques*, vol. 11, no. 1. 2018.
20. C. K. Wang, *Sentiment Analysis Using Support Vector Machines, Neural Networks, and Random Forests*, no. Icaiai. Atlantis Press International BV, 2023.
21. A. N. Khobragade, M. M. Raghuwanshi, and L. Malik, “Evaluating Kernel Effect on Performance of SVM Classification using Satellite Images,” *Int. J. Sci. Eng. Res.*, vol. 7, no. 3, pp. 742–748, 2016.
22. I. Guyon, B. Boser, and V. Vapnik, “Automatic Capacity Tuning of Very Large VC-Dimension Classifiers,” *Adv. Neural Inf. Process. Syst.*, vol. 5, pp. 147–155, 1993, [Online]. Available: <http://www.clopinet.com/isabelle/Papers/autocapa.ps>.
23. A. Tharwat, “Parameter investigation of support vector machine classifier with kernel functions,” *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1269–1302, 2019, doi: 10.1007/s10115-019-01335-4.



24. C. A. A. Kaestner, “Support Vector Machines and Kernel Functions for Text Processing,” *Rev. Informática Teórica e Apl.*, vol. 20, no. 3, p. 130, 2013, doi: 10.22456/2175-2745.39702.
25. S. G. Kanakaraddi, A. K. Chikaraddi, K. C. Gull, and P. S. Hiremath, “Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms,” *Proc. 5th Int. Conf. Inven. Comput. Technol. ICICT 2020*, pp. 287–292, 2020, doi: 10.1109/ICICT48043.2020.9112546.
26. A. Nayak and S. Natarajan, “Comparative study of Naïve Bayes , Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter feeds,” *Int. J. Adv. Stud. Comput. Sci. Eng. IJASCSE*, vol. 5, no. 1, pp. 14–17, 2016.
27. C. Kaope and Y. Pristyanto, “The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 2, pp. 227–238, 2023, doi: 10.30812/matrik.v22i2.2515.
28. A. Artetxe, A. Beristain, and M. Graña, “Predictive models for hospital readmission risk: A systematic review of methods,” *Comput. Methods Programs Biomed.*, vol. 164, pp. 49–64, 2018, doi: 10.1016/j.cmpb.2018.06.006.
29. Q. Qin, Z. Zhou, J. Zhou, Z. Huang, X. Zeng, and B. Fan, “Sentiment and attention of the Chinese public toward electric vehicles: A big data analytics approach,” *Eng. Appl. Artif. Intell.*, vol. 127, no. PA, p. 107216, 2024, doi: 10.1016/j.engappai.2023.107216.