# Analisis Sentimen Pengguna Seputar Kendaraan Listrik Di Twitter Dengan Penerapan Algoritma *Naïve Bayes*, *KNN*, dan *Decision Tree* untuk Klasifikasi

Teguh Prasetyo[1], Arya Adhyaksa Waskita[2], Taswanda Taryo[3]

[1,2,3] Program Studi Magister Teknik Informatika, Program Pascasarjana, Universitas Pamulang
Jl. Raya Puspitek, Buaran, Kec. Pamulang. Kota Tangerang Selatan, Banten
Teguhprasetyo261211@gmail.com,

Diterima : 01 Februari 2025
Disetujui : 14 Februari 2025

***Abstract*— Penelitian ini bertujuan untuk menganalisis sentimen pengguna mobil listrik menggunakan tiga algoritma klasifikasi machine learning: Naïve Bayes, Decision Tree, dan K-Nearest Neighbor. Data penelitian diambil dari media sosial Twitter dengan jumlah data bersih sebanyak 1869 tweet. Proses analisis meliputi ekstraksi teks, preprocessing, serta penggunaan teknik word embedding TF-IDF dan Word2Vec. Hasil penelitian menunjukkan bahwa sentimen netral mendominasi sebesar 53,85%, diikuti oleh sentimen positif 35,85% dan sentimen negatif 10,30%. Dari model yang diuji, Decision Tree dengan embedding TF-IDF memiliki performa terbaik dengan akurasi 93,75%, sementara performa terendah ditunjukkan oleh Naïve Bayes dengan Word2Vec yang hanya mencapai 63,50%. KNN dengan TF-IDF memiliki akurasi 89,25%, lebih tinggi dibandingkan KNN dengan Word2Vec yang mencapai 87,75%. Sementara itu, Naïve Bayes dengan TF-IDF memiliki akurasi 81,50%, lebih tinggi daripada Naïve Bayes dengan Word2Vec yang hanya mencapai 63,50%. Berdasarkan analisis data perbandingan hasil model, Decision Tree dengan TF-IDF berhasil mengklasifikasikan 126 data sebagai positif, 1419 data sebagai netral, dan 55 data sebagai negatif. Selain itu, evaluasi kompleksitas waktu menunjukkan bahwa KNN dengan Word2Vec memiliki waktu training tercepat (0,003 detik), sementara Decision Tree dengan Word2Vec memiliki waktu prediksi tercepat (0,001 detik). Naïve Bayes dengan TF-IDF memiliki waktu training paling lama (0,496 detik), namun Decision Tree dengan TF-IDF memberikan keseimbangan optimal antara waktu training dan prediksi. Penelitian ini memberikan wawasan bagi industri otomotif dan pemerintah dalam memahami persepsi masyarakat terhadap mobil listrik. Hasil penelitian juga menyoroti efektivitas berbagai algoritma machine learning dan word embedding dalam analisis sentimen.***

***Keywords* — Analisis Sentimen, Kendaraan Listrik, Twitter, KNN, Decision Tree, Naïve Bayes**

## I. PENDAHULUAN

Kendaraan listrik semakin menjadi perhatian utama dalam industri otomotif dalam beberapa tahun terakhir. Jenis kendaraan ini mengandalkan motor listrik sebagai sumber tenaga utama untuk menggerakkan roda. Sebagai alternatif dari kendaraan berbahan bakar fosil, mobil listrik menawarkan berbagai keunggulan, seperti lebih ramah lingkungan, efisiensi energi yang lebih tinggi, serta biaya operasional yang lebih ekonomis. Sementara itu, energi fosil termasuk sumber daya yang tidak dapat diperbarui dan diperkirakan akan mengalami kelangkaan di masa depan [1].

Menurut data dari **www.dataindonesia.com**, penjualan mobil listrik di Indonesia mengalami lonjakan signifikan sebesar 383,46% dari tahun 2021 ke 2022. Tren ini menunjukkan bahwa kendaraan listrik semakin menarik perhatian masyarakat. Meskipun demikian, kehadiran mobil listrik masih menimbulkan perdebatan. Sebagian masyarakat mempertimbangkan keuntungan dan

kekurangan dibandingkan dengan kendaraan konvensional. Beberapa faktor yang menjadi perhatian utama adalah harga mobil listrik yang relatif lebih tinggi, keterbatasan infrastruktur stasiun pengisian daya di Indonesia, serta jarak tempuh dan kecepatan yang masih belum sebanding dengan kendaraan berbahan bakar fosil [2].

Mobil listrik menawarkan berbagai keunggulan potensial, salah satunya adalah kemampuannya dalam mengurangi emisi gas rumah kaca. Namun, kehadirannya masih memicu perdebatan di kalangan masyarakat, terutama di media sosial seperti Twitter. Sebagian orang berpendapat bahwa mobil listrik dapat menjadi solusi efektif untuk mengatasi defisit migas. Sementara itu, pihak lain menekankan bahwa transisi ke kendaraan listrik memerlukan persiapan yang matang, terutama dalam hal pengembangan infrastruktur pendukung.

Respon masyarakat Indonesia terhadap mobil listrik menjadi salah satu indikator bagi peneliti dalam menganalisis sentimen publik di media sosial Twitter. Pemahaman terhadap persepsi masyarakat mengenai mobil listrik sangat penting bagi produsen dan pemerintah dalam merumuskan strategi pemasaran serta kebijakan yang tepat. Dengan menganalisis sentimen positif dan negatif, produsen dapat mengevaluasi serta meningkatkan kualitas dan daya tarik mobil listrik agar lebih sesuai dengan kebutuhan konsumen.

Di Indonesia, penggunaan mobil listrik masih tergolong sedikit. Dikarenakan ada beberapa sebab, seperti harga mobil tergolong masih relatif tinggi, infrastruktur yang belum memadai (seperti stasiun pengisian daya listrik), dan pengetahuan masyarakat tentang mobil listrik yang masih relatif kurang. Untuk meningkatkan penggunaan mobil listrik di Indonesia, Perlu ada usaha meningkatkan pemahaman masyarakat mengenai keuntungan menggunakan kendaraan listrik. Beberapa cara yang diambil adalah dengan melakukan analisis sentimen terhadap penggunaan mobil listrik di media sosial. Oleh karena itu, diperlukan analisis sentimen mengenai mobil listrik [3].

Perlu diketahui bahwa analisis sentimen merupakan salah satu cabang dari pemrosesan bahasa alami (*natural language processing*) yang digunakan untuk melacak suasana hati masyarakat mengenai suatu produk atau isu yang beredar. Analisis sentimen disebut juga opinion mining [4]. Oleh sebab itu, Analisis sentimen, atau yang juga dikenal sebagai opinion mining, adalah proses untuk memahami, mengekstrak, dan memproses data teks secara otomatis guna mengidentifikasi sentimen yang terkandung dalam sebuah kalimat opini. Pengaruh besar dan manfaat yang ditawarkan oleh analisis sentimen mendorong perkembangan pesat dalam penelitian dan penerapan berbasis analisis ini. [5]

Selain masalah-masalah yang dijelaskan di atas, terdapat pula masalah seperti baterai mobil listrik yang lebih mahal daripada harga mobilnya, stasiun pengisian daya (charging station) yang belum banyak tersedia di Indonesia dan hanya terdapat di daerah-daerah kota besar, serta harga mobil listrik yang masih relatif mahal. Hal ini membuat penulis perlu melakukan analisis sentimen terhadap mobil listrik.

## II. TINJAUAN LITERATURE

Menurut tinjauan Pustaka yang telah peneliti pelajari, penelitian tentang Analisis Sentimen yang diambil oleh Eva Nurhazizah dkk tentang "Analisis Sentimen dan Jaringan Sosial pada Penyebaran Informasi Vaksinasi di Twitter" [6] menggunakan metode Social Network Analysis (SNA). Namun, penelitian ini tidak menguraikan secara rinci perangkat yang digunakan dalam analisisnya.

Studi lain yang dilakukan oleh Teguh Anysor Lorosae dkk tentang Analisis Sentimen berjudul "Analisis Sentimen Berdasarkan Opini Masyarakat pada Twitter Menggunakan Naïve Bayes" Diketahui bahwa peneliti menerapkan metode Naïve Bayes. Namun, tidak dijelaskan perangkat apa yang digunakan dalam penelitian tersebut. Selain itu, penelitian ini hanya menggunakan satu metode yang diterapkan oleh peneliti. [7].

Dalam studi lain yang dilakukan oleh Muhammad Nanda Fahriza dkk tentang Analisis Sentimen yang berjudul "Analisis Sentimen pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer GPT Menggunakan Metode Klasifikasi K-Nearest Neighbor (KNN)"

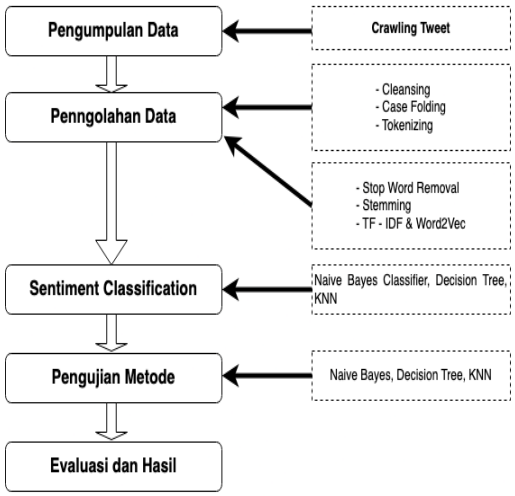
Diketahui bahwa peneliti menerapkan metode KNN. Namun, penelitian ini tidak menjelaskan secara spesifik perangkat yang digunakan. Selain itu, hanya satu metode yang diterapkan dalam penelitian ini. [8].

Dengan demikian, dalam studi ini peneliti menggunakan tiga metode dalam analisis sentimen untuk membandingkan performa dari model yang peneliti pilih. Metode tersebut adalah metode *KNN*, *Decision Tree*, dan *Naïve Bayes*. Studi ini dilakukan untuk mengkomparasi ketiga metode tersebut dengan mengintegrasikan Media Sosail Twitter sebagai platform untuk peneliti melakukan pengambilan data guna mengetahui opini masyarakat terhadap penggunaan mobil listrik dengan algoritma klasifikasi *machine learning*. Selain itu, tujuannya adalah untuk mengetahui berapa banyak jumlah distribusi sentiment seperti netral, positif dan negative terhadap opini masyarakat terhadap pengguna kendaraan listrik dengan cara pengumpulan data dari Twitter dan menganalisis data tersebut dengan Bahasa pemrograman *Python* dan google colab [9].

## III. METODE PENELITIAN

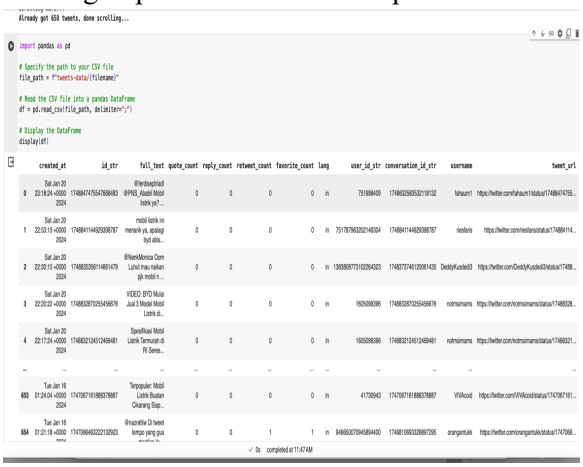
Studi yang dilakukan dengan pendekatan kuantitatif dengan jenis deskriptif. Studi ini menerapkan tiga algoritma klasifikasi, seperti *Naïve Bayes*, *KNN*, dan *Decision Tree*. Data yang digunakan dalam studi ini adalah data yang diambil dari media sosial Twitter mengenai mobil listrik. Alur proses penelitian ini ditampilkan pada Gambar 1. Proses dimulai dari pengumpulan data, kemudian dilanjutkan dengan pengolahan data (*preprocessing*) meliputi *cleansing*, *case folding*, *tokenizing*, *stop word removal*, *stemming* dan *embedding word* (*tf-idf* dan *word2vec*) [10] selanjutnya menggunakan klasifikasi *Naïve Bayes Classifier*, *Decision Tree* dan kemudian melakukan pengujian metode dan mengevaluasi kinerja model.

### A. Desain Penelitian




Gambar 1. Flowchart Analisis Sentimen Pengguna Mobil Listrik

### B. Pengumpulan Data dan Persiapan Data



Gambar 2. Dataset Hasil Crawling

Data yang dikumpulkan sebanyak 2000 tweet selama beberapa kali tahapan pengumpulan dan disimpan kedalam file Csv (*Command Separated Values*) dengan format .csv. Data yang diambil dari twitter dengan *keyword* ‘‘mobil listrik’’ menggunakan API Twitter (X) [11] dan Tidak ada batasan dalam jumlah data yang dikumpulkan, namun dengan 2.000 data diharapkan dapat merepresentasikan opini masyarakat secara umum. Data yang disimpan mencakup tanggal pembuatan tweet, ID, serta pengguna Twitter yang mempostingnya. Data yang telah dikumpulkan dalam format CSV dapat dilihat pada Gambar 2.

C. Preprocessing Data

| | full_text | tweet_english | Score | sentimen |
|---|---|---|---|---|
| 0 | mobil listrik | electric car | 0.000 | Netral |
| 1 | mobil listrik byd darat indonesia | Indonesian land byd electric car | 0.000 | Netral |
| 2 | celaka kereta | wretched train | 0.000 | Netral |
| 3 | nikel beli murah | nickel buy cheap | 0.400 | Positif |
| 4 | mobil listrik tesla china kembali ada ganggu s... | Tesla China Electric Car Again Disrupts the au... | 0.000 | Netral |
| ... | ... | ... | ... | ... |
| 1864 | mobil listrik masuk brand brand bagus | Electric cars enter a good brand brand | 0.700 | Positif |
| 1865 | jiahnh cuman mobil listrik | jiahnh only electric cars | 0.000 | Netral |
| 1866 | harga mobil listrik byd | the price of an electric car byd | 0.000 | Netral |
| 1867 | mobil listrik byd aspal indonesia | Indonesian Asphalt Electric Car | 0.000 | Netral |
| 1868 | bener banget efektif nih kasi mobil jabat diki... | Really really effective, giving a little a lit... | 0.075 | Positif |

1869 rows × 4 columns

Gambar 3. Hasil Dataset Yang Sudah Dicleaning

*Preprocessing* merupakan tahap pengolahan data sebelum data tersebut diproses lebih lanjut. Dalam praktiknya, banyak dataset yang masih belum bersih, misalnya akibat kesalahan sistem saat pencatatan, yang dapat menyebabkan munculnya data duplikat. Data yang belum diolah atau data tidak bersih kategorinya seperti format data yang tidak beraturan, adanya data kosong, tipe data yang berbeda-beda, adanya atribut yang tidak penting, dan lain sebagainya. Semakin bersih pra proses yang dilakukan, maka kemungkinan besar hasil data tersebut semakin akurat. [12]. Berikut adalah langkah preprocessing yang dilakukan ;

1. *Cleaning*

Tahap pembersihan *(cleaning)* bertujuan untuk menghilangkan atau mengurangi kata maupun kalimat yang tidak relevan dalam data tweet, seperti tanda baca, karakter unicode, dan elemen lainnya. Proses cleaning ini terdiri dari lima tahapan yang akan dijalankan oleh sistem guna memperoleh hasil yang optimal, di antaranya: Membersihkan tanda baca; Membersihkan angka; Membersihkan *link , hashtag*; Membersihkan kelebihan spasi, *URLs,Hashtags, Mentions, Reservedwords (RT,FAV), Emojis, Smileys* [13] .

2. *Case Folding*

Pada tahap ini, seluruh huruf dalam teks akan dikonversi menjadi huruf kecil (*lowercase*). Adapun proses case folding meliputi langkah-langkah berikut: memeriksa setiap karakter dari awal hingga akhir teks, lalu mengubah huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*) jika ditemukan. [14].

3. *Tokenizing*

Tahap *tokenizing* berfungsi untuk memisahkan kata-kata dalam tweet menjadi unit-unit yang lebih kecil. Proses ini menggunakan spasi sebagai pemisah antar kata. Dalam sebuah tweet, terdapat sejumlah kata yang terhubung dan dipisahkan oleh spasi. Agar teks lebih mudah diproses, setiap kata dalam kalimat harus dipisahkan. Jika suatu karakter bukan tanda pemisah seperti titik (.), koma (,), atau spasi, maka karakter tersebut akan digabungkan dengan karakter berikutnya. [15] .

4. *Stopwords Removal*

Pada tahap ini, kumpulan tweet yang telah melewati proses tokenizing akan diproses lebih lanjut dengan menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis sentimen. Setiap kata dalam tweet akan diperiksa, dan jika kata tersebut termasuk dalam kategori kata sambung, kata depan, kata ganti, atau kata yang tidak relevan, maka kata tersebut akan dihapus. [16].

5. *Stemming*

Setiap kata dalam tweet seringkali memiliki berbagai bentuk morfologi. Oleh karena itu, setiap kata akan diubah menjadi bentuk dasarnya (stem) yang sesuai. Kata-kata ini akan direduksi dengan menghilangkan awalan atau akhiran yang ada. Proses stemming dilakukan dengan langkah-langkah berikut: pertama, kata yang digunakan adalah kata yang telah melalui tahap *stopword removal*; kemudian, setiap kata dalam tweet akan diperiksa dari awal hingga akhir; jika ditemukan kata yang mengandung imbuhan, maka imbuhan tersebut akan dihapus. [17].

D. Data Labelling Data

Data Labelling hasil *crawling* serta sudah melewati tahapan pengelolaan dataset yang dilaksanakan dengan bantuan python karena pada Rapid Miner harus menggunakan manual untuk pelabelan data, maka peneliti berinisiatif memakai *library python VaderSentiment* [18] seperti melihat *polarity, subjectivity* yang dipunyai oleh teks tweet yang sudah digunakan. *VaderSentiment* merupakan sebuah *library* yang disediakan oleh *Python* buat pemrosesan data dibidang *Natural Language Processing* yang bisa memberikan tag kata, ekstraksi kata,

penerjemahan kata serta *sentiment analysis*. Hasil dari objek VaderSentiment dapat digunakan dalam proses pembelajaran bahasa alami. Namun, karena saat ini VaderSentiment hanya mendukung bahasa Inggris, pada penelitian ini data berbahasa Indonesia terlebih dahulu diterjemahkan ke dalam bahasa Inggris. Penentuan kelas sentimen (positif, netral, dan negatif) didasarkan pada nilai polaritas. Nilai polaritas dalam analisis sentimen berada pada rentang antara 1 hingga -1, yang menunjukkan kelas sentimen data. Teks tweet dengan nilai polaritas mendekati 1 menunjukkan sentimen positif, nilai polaritas mendekati -1 menunjukkan sentimen negatif, dan nilai polaritas yang berada di sekitar 0 menunjukkan sentimen netral.

E. Pembobotan Kata (*Word Embedding*)

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari teknik TF-IDF adalah untuk menilai sejauh mana suatu kata (*term*) penting dalam sebuah dokumen, berdasarkan konteks koleksi dokumen yang lebih besar.[19].

Dalam metode TF-IDF, nilai TF (Term Frequency) dan IDF (Inverse Document Frequency) dikombinasikan untuk menghasilkan bobot kata (term weight) bagi setiap kata dalam sebuah dokumen. Bobot ini menggambarkan seberapa penting suatu kata dalam dokumen tersebut jika dibandingkan dengan koleksi dokumen yang lebih besar. Rumus untuk metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah sebagai berikut:

$$tf = 0{,}5 + 0{,}5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf_t = \log\left(\frac{D}{dft}\right) \quad (2)$$

$$W_{d.t} = tf_{d.t} \times idf_{d.t} \quad (3)$$

Metode Word2Vec bertujuan untuk mengidentifikasi hubungan tersembunyi antar kata. Setiap kata diwakili oleh distribusi bobot pada elemennya. Metode ini memiliki dua jenis model arsitektur, yaitu Continuous Bag-of-Words (CBOW) dan Skip-Gram. Model CBOW berfokus pada memprediksi kata target berdasarkan konteks kata, sementara model Skip-Gram bertujuan untuk memprediksi kemungkinan kata-kata yang bisa menjadi konteks dari kata target. [20].

F. Klasifikasi

Setelah tahap pembobotan kata, langkah berikutnya adalah melakukan klasifikasi menggunakan algoritma. Algoritma yang digunakan dalam proses ini meliputi Naïve Bayes, KNN, dan Decision Tree.

1. *Naïve Bayes* dipergunakan untuk melakukan klasifikasi sentiment tweet dalam tiga kategori : positf, netral dan negatif. Model yang dibangun berdasarkan data yang melewati proses latih, pemberian *labelling* sentimen dan di beri pembobotan kata. *Naive Bayes* menggunakan Pengklasifikasian probabilistik sederhana adalah metode yang menghitung sejumlah probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang ada. Proses ini digunakan untuk menentukan kemungkinan setiap kelas berdasarkan data yang diberikan [21].
2. *K-Nearest Neighbors* bertujuan untuk mengklasifikasikan sentimen teks (cuitan di Twitter) ke dalam kategori positif, negatif, atau netral. Cara kerjanya membandingkan cuitan baru dengan cuitan-cuitan lain yang sudah ada di dataset. Kemudian mencari "tetangga terdekat" dari cuitan baru tersebut berdasarkan kemiripan karakteristik (misalnya, kata-kata yang digunakan) dan Mengklasifikasikan sentimen cuitan baru berdasarkan sentimen mayoritas dari "tetangga terdekat"-nya. [22].
3. *Decision Tree* juga untuk mengklasifikasikan sentimen cuitan. Cara kerjanya adalah ,embangun struktur pohon keputusan berdasarkan fitur-fitur dari data training (misalnya, keberadaan kata-kata tertentu, panjang cuitan, dll.).Setiap cabang pohon merepresentasikan sebuah keputusan berdasarkan fitur tertentu. Sentimen cuitan ditentukan berdasarkan jalur yang dilalui cuitan tersebut dalam pohon keputusan [23].

G. Evaluasi

Evaluasi dilakukan guna mengukur kinerja ketiga model yang digunakan dalam studi ini yaitu *Naïve Bayes*, *K-Nearest Neighbors* dan *Decision Tree* dalam mengklasifikasikan sentimen

menggunakan *Confusion Matrix*. *Confusion matrix* adalah sebuah matriks yang digunakan untuk mengevaluasi hasil dari model klasifikasi, dengan menunjukkan jumlah data uji yang benar maupun yang salah. Dengan adanya matriks ini, kita dapat menilai seberapa baik kinerja model klasifikasi tersebut. [24]. Tabel 1 menunjukkan visualisasi dari *confusion matrix*, yang terdiri dari 4 komponen utama yaitu :

Tabel 1. Tabel *Confusion Matrix*

| Kelas Aktual | Kelas Prediksi | |
|---|---|---|
| | Positif | Negatif |
| Positif | Benar Positif | Salah Negatif |
| Negatif | Salah Positif | Benar Negatif |

Akurasi adalah rasio antara jumlah data tweet yang berhasil terdeteksi dalam pengujian. Nilai akurasi menggambarkan seberapa dekat hasil prediksi sistem dengan prediksi yang dilakukan oleh manusia. [25] Berikut rumusnya:

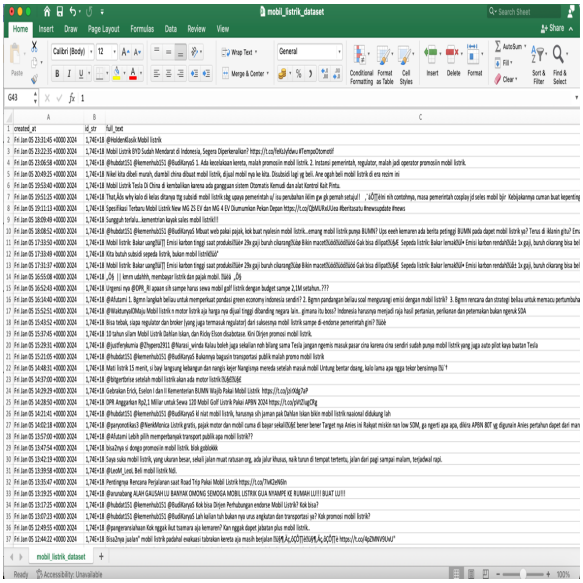
$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

## IV. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dijelaskan mengenai step by step dilakukan di penelitian ini. Penelitian ini dilakukan menggunakan Bahasa pemrograman python dan implementasikan menggunakan platform google colab. Proses dari pengumpulan data, tahap preprocessing, implementasi model, hingga evaluasi model yang digunakan menggunakan python secara struktur menggunakan library yang python miliki.

### A. Pengumpulan Data

Data yang dikumpulkan dari X dengan keyword ‘mobil listrik’ disertai rentang pengambilan data tweets yaitu dari periode 01 Januari 2024 – 31 Januari 2024. Data yang dikumpulkan berjumlah 2000 data. Data yang berhasil terkumpul dalam format ‘csv’ dalam gambar 4. Berikut :



Gambar 4. Hasil pengumpulan data tweet dalam format csv

### B. *Preprocessing Data*

Setelah data dikumpulkan, tahap berikutnya adalah *preprocessing*. Proses pertama yang dilakukan adalah pembersihan dataset (*cleaning*), yang bertujuan untuk mengurangi atau menghapus elemen-elemen dalam tweet yang tidak diperlukan, seperti tanda baca, *unicode*, dan sebagainya. Selanjutnya, dilakukan *case folding*, dimana seluruh huruf diubah menjadi huruf kecil (*lowercase*). Tahap berikutnya adalah *tokenizing*, yang berfungsi untuk memisahkan kata-kata yang terdapat dalam tweet. Pada tahap *stopword removal*, tweet yang sudah melalui proses *tokenizing* akan melanjutkan ke penghapusan *stopwords*. Kata-kata yang tidak relevan dengan analisis sentimen, seperti kata sambung, kata depan, atau kata ganti, akan dihapus. Terakhir, pada tahap stemming, setiap kata yang muncul dalam tweet yang memiliki variasi morfologi akan direduksi menjadi bentuk dasarnya (*stemmed word*). Semua tahapan preprocessing ini akan ditampilkan dalam tabel 2 hingga tabel 5.

Tabel 2. Proses *Case Folding*

| No | Sebelum | Sesudah |
|---|---|---|
| 1. | Mobil listrik Bakar uang Emisi karbon tinggi saat produksi gaji buruh cikarang Bikin macet Gak bisa dilipat Sepeda listrik Bakar lemak | mobil listrik bakar uang emisi karbon tinggi saat produksi gaji buruh cikarang bikin macet gak bisa dilipat sepeda listrik |

| | | |
|---|---|---|
| | Emisi karbon rendah gaji | bakar lemak emisi karbon rendah gaji |
| 2. | ALAH GAUSAH LU BANYAK OMONG SEMOGA MOBIL LISTRIK GUA NYAMPE KE RUMAH LU BUAT LU | alah gausah lu banyak omong semoga mobil listrik gua nyampe ke rumah lu buat lu |

Tabel 3. Proses *Tokenizing*

| No | Sebelum | Sesudah |
|---|---|---|
| 1. | dah jadi kayak norwegia pak sini banyak charger ev gratis karena pasti bagian kebijakan banyak bengkel konversi ev atpm atpm berlomba masukin mobil listrik innova reborn pasti dah listrik | ‘sudah’ , ‘jadi, ‘norwegia’, ‘charger’, gratis pasti bagian ‘bijak’, ‘bengkel’, ‘konversi, ‘lomba’, ‘masuk’, ‘mobil’ ‘innova’, ‘pasti’, ‘listrik’ |
| 2. | kedua predikat tesla perusahaan jagoannya elon sebagai raja mobil listrik pun disusul oleh perusahaan mobil asal tiongkok byd | ‘dua’, ‘predikat’, ‘tesla’ ‘perusahaan’, ‘jagoan’, ‘elon’, ‘sebagai’, ‘raja’, ‘mobil’ ‘listrik’, ‘oleh’, ‘asal’, ‘tiongkok’, ‘byd’ |

Tabel 4. Proses *Stopwords*

| No | Sebelum | Sesudah |
|---|---|---|
| 1. | Malah dipake buat belanja mobil dinas baru Mau ngirit pak Investasi ke trafo yg bisa kirim signal kalau jaringan listrik bermasalah | pakai, belanja, mobil, dinas, irit, investasi, trafo, kirim, signal, jaringan, listrik |

Tabel 5. Proses *Stemming*

| Kata Imbuhan | Kata Dasar |
|---|---|
| mendarat | darat |
| dibeli | beli |
| dikembalikan | kembali |
| contohnya | contoh |
| dilipat | lipat |
| bagusin | bagus |
| memperbanyak | banyak |

## C. Pelabelan Data

Setelah melewati tahap *preprocessing*, kemudian dilakukan pelabelan data. Dari proses pelabelan data didapatkan 1869 data tweet adalah 670 tweet masuk ke dalam kelas positif, 999 tweet kelas netral dan 200 negatif. Hasil dari pelabelan dapat dilihat pada gambar 5 sebagai berikut.

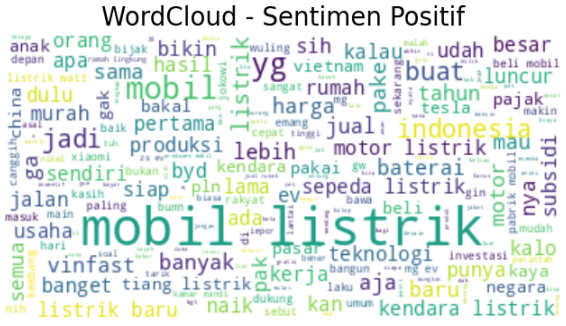
| | full_text | tweet_english | Score | sentimen |
|---|---|---|---|---|
| 0 | mobil listrik | electric car | 0.000 | Netral |
| 1 | mobil listrik byd darat indonesia | Indonesian land byd electric car | 0.000 | Netral |
| 2 | celaka kereta | wretched train | 0.000 | Netral |
| 3 | nikel beli murah | nickel buy cheap | 0.400 | Positif |
| 4 | mobil listrik tesla china kembali ada ganggu s... | Tesla China Electric Car Again Disrupts the au... | 0.000 | Netral |
| ... | ... | ... | ... | ... |
| 1864 | mobil listrik masuk brand brand bagus | Electric cars enter a good brand brand | 0.700 | Positif |
| 1865 | jiahnh cuman mobil listrik | jiahnh only electric cars | 0.000 | Netral |
| 1866 | harga mobil listrik byd | the price of an electric car byd | 0.000 | Netral |
| 1867 | mobil listrik byd aspal indonesia | Indonesian Asphalt Electric Car | 0.000 | Netral |
| 1868 | bener banget efektif nih kasi mobil jabat diki... | Really really effective, giving a little a lit... | 0.075 | Positif |

1869 rows x 4 columns

Gambar 5. Hasil pelabelan data

## D. *Word Embbedding (TF-IDF dan Word2Vec)*

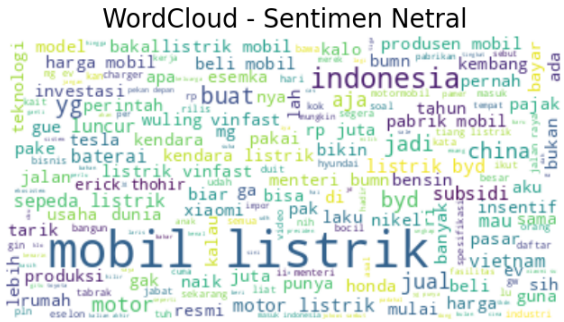
Setelah melewati tahap pelabelan data, selanjutnya adalah menerapkan fitur *word embbeding* atau *Mengonversi* teks menjadi representasi numerik dengan menyoroti kata-kata paling penting berdasarkan frekuensi kemunculannya di dalam dokumen dan seberapa jarang kata tersebut muncul di dokumen lainnya. Proses ini menggunakan dua *word embedding* yaitu *TF-IDF* dan *Word2vec.* Untuk mempermudah pemahaman, hasil dari proses TF-IDF dan Word2vec ini akan divisualisasikan dalam bentuk word cloud, di mana kata-kata dengan skor tertinggi akan ditampilkan lebih besar, sehingga memudahkan identifikasi kata kunci penting secara visual.




Gambar 6. Visualisasi *Wordcloud* Sentimen Positif

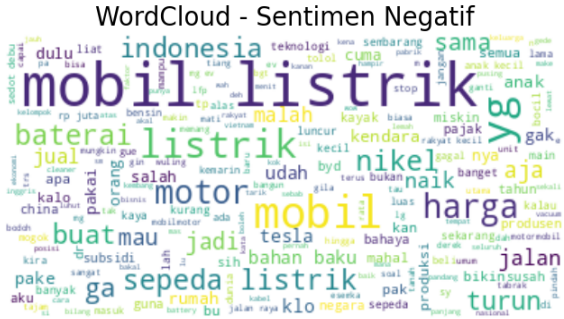
Pada gambar 6. Ulasan positif diidentifikasi berdasarkan frekuensi kata yang muncul dalam ulasan. Berikut ini adalah hasil visualisasi dari ulasan positif yang diperoleh melalui ekstraksi informasi dari ulasan-ulasan yang ditulis oleh pengunjung. Dari informasi yang diperoleh, dapat

diketahui bahwa pada kelas sentimen positif, kata-kata yang paling sering muncul adalah *mobil, listik, baterai, sepeda, motor, byd, indonesia, jual*.




Gambar 7. Visulisasi *Wordcloud* Sentimen Netral

Pada Gambar 7. Ulasan netral diidentifikasi berdasarkan frekuensi kata yang muncul dalam ulasan. Berikut ini adalah hasil visualisasi ulasan netral yang diperoleh melalui ekstraksi informasi dari ulasan-ulasan yang ditulis oleh pengunjung. Dari informasi yang didapat, dapat diketahui bahwa pada kelas sentimen netral, kata-kata yang paling sering muncul adalah *mobil, listrik, indonesia, buat, jadi, china.*




Gambar 8. Visulisasi *Wordcloud* Sentimen Negatif

Pada Gambar 8. Ulasan negatif diidentifikasi berdasarkan frekuensi kata yang muncul dalam ulasan. Berikut ini adalah hasil visualisasi ulasan negatif yang diperoleh dari ekstraksi informasi yang didapatkan dari ulasan-ulasan yang ditulis oleh pengunjung. Informasi yang diperoleh menunjukkan bahwa pada kelas sentimen negatif, kata-kata yang paling sering muncul adalah. *mobil, listrik, sembarang, susah, jalan, turun, turun, sepeda.*

E. Klasifikasi

Penelitian ini memanfaatkan dataset dari tweet. Dataset tersebut menggunakan 3 Algoritma klasifikasi yaitu *Naïve Bayes*, *KNN* dan *Decision Tree*. Pembuatan model dibuat menggunakan skenari pembagian *data training* dan *data testing* dalam rasio 20 : 80. Proses ini dijalankan dengan *library* pada bahasa python 3 yang bernama *scikit-learn* untuk proses klasifikasi.

F. Evaluasi

Setelah pembuatan model, Langkah selanjutnya adalah evaluasi model yang sudah dibuat. Dengan melakukan evaluasi hasilnya menggambarkan efisiensi setiap model dalam hal waktu yang dibutuhkan untuk pelatihan dan prediksi. ditunjukkan pada tabel 6.

Tabel 6. Waktu Training dan Prediksi

| Model | Waktu Training | Waktu Prediksi |
|---|---|---|
| NB + TF-IDF | 0,496 detik | 0,008 detik |
| KNN + TF-IDF | 0,217 detik | 0,145 detik |
| DT + TF-IDF | 0,314 detik | 0,005 detik |
| NB + Word2Vec | 0,007 detik | 0,001 detik |
| **KNN + Word2Vec** | **0,003 detik** | **0,184 detik** |
| DT + Word2Vec | 0.394 detik | 0,001 detik |

Kesimpulan dari hasil evaluasi menunjukkan bahwa K-Nearest Neighbors + Word2Vec memiliki waktu pelatihan tercepat (0,003 detik), namun waktu prediksinya relatif lebih lambat (0,184 detik). Sementara itu, Decision Tree + Word2Vec mencatatkan waktu prediksi tercepat (0,001 detik). Naive Bayes + TF-IDF, di sisi lain, memerlukan waktu pelatihan paling lama (0,496 detik). Adapun Decision Tree + TF-IDF memberikan keseimbangan yang baik antara waktu pelatihan dan prediksi, dengan waktu prediksi tercepat di antara model berbasis TF-IDF (0,005 detik). Pemilihan model yang tepat sangat bergantung pada prioritas yang diinginkan, apakah lebih mengutamakan kecepatan pelatihan atau kecepatan prediksi.

Setelah melakukan training, selanjutnya adalah akurasi, Akurasi mengukur proporsi dari prediksi yang benar di bandingkan dengan total jumlah prediksi yang dibuat. Ini dihitung dengan

membagi jumlah prediksi benar dengan jumlah total prediksi pada tabel 7.

Tabel 7. Hasil Akurasi

| *Model* | *Accuracy* |
|---|---|
| NB + TF-IDF | 81,50% |
| **DT + TF-IDF** | **93,75%** |
| KNN + TF-IDF | 89,25% |
| NB + Word2Vec | 63,50% |
| DT + Word2Vec | 82,50% |
| KNN + Word2Vec | 87,75% |

Berdasarkan kesimpulan yang terdapat pada Tabel 7, model Decision Tree + TF-IDF (DT + TF-IDF) menunjukkan performa terbaik dengan akurasi 93,75%, yang mengindikasikan bahwa Decision Tree sangat efektif ketika digunakan dengan fitur TF-IDF dalam klasifikasi teks. K-Nearest Neighbors + TF-IDF (KNN + TF-IDF) juga mencatatkan akurasi tinggi sebesar 89,25%, menunjukkan bahwa KNN cukup efektif dengan TF-IDF. Meskipun Naïve Bayes + TF-IDF (NB + TF-IDF) memiliki akurasi lebih rendah di 81,50%, namun masih lebih baik dibandingkan metode berbasis Word2Vec. Secara umum, Word2Vec menghasilkan akurasi yang lebih rendah dibandingkan TF-IDF, dengan model Naïve Bayes + Word2Vec (NB + Word2Vec) mencatatkan performa terendah di 63,50%, yang menunjukkan ketidaksesuaian Naïve Bayes dengan fitur Word2Vec dalam kasus ini. KNN + Word2Vec mencapai akurasi terbaik di antara model berbasis Word2Vec (87,75%), meskipun tetap kalah dibandingkan dengan KNN + TF-IDF. Sementara itu, DT + Word2Vec mencatatkan akurasi yang cukup baik (82,50%), tetapi masih lebih rendah dibandingkan kombinasi TF-IDF. Secara keseluruhan, dapat disimpulkan bahwa TF-IDF lebih efektif dibandingkan Word2Vec dalam representasi teks untuk model yang digunakan. Decision Tree dengan TF-IDF merupakan kombinasi terbaik dalam eksperimen ini, sementara KNN menunjukkan performa yang cukup konsisten baik dengan TF-IDF maupun Word2Vec.

## V. KESIMPULAN

Berdasarkan penelitian tentang analisis sentimen pengguna mobil listrik di Twitter dengan menggunakan metode Naïve Bayes, Decision Tree, dan K-Nearest Neighbor, ditemukan bahwa sentimen netral mendominasi dengan persentase 53,85%, diikuti sentimen positif 35,85% dan negatif 10,30%. Dari ketiga algoritma yang diuji, Decision Tree dengan TF-IDF mencatatkan akurasi terbaik sebesar 93,75%, sementara Naïve Bayes dengan Word2Vec mencatatkan akurasi terendah 63,50%. KNN dengan TF-IDF menghasilkan akurasi 89,25%, lebih baik dari KNN dengan Word2Vec yang 87,75%. Naïve Bayes dengan TF-IDF memperoleh akurasi 81,50%. Distribusi sentimen dari model terbaik, Decision Tree + TF-IDF, adalah 126 data positif, 1419 netral, dan 55 negatif. Dalam hal kompleksitas waktu, KNN + Word2Vec memiliki waktu training tercepat (0,003 detik), sementara Decision Tree + Word2Vec memiliki waktu prediksi tercepat (0,001 detik). Naïve Bayes + TF-IDF memerlukan waktu training paling lama (0,496 detik), sedangkan Decision Tree + TF-IDF menawarkan keseimbangan antara waktu training dan prediksi dengan waktu prediksi tercepat dalam kelompok TF-IDF (0,005 detik). Penggunaan TF-IDF lebih efektif dibandingkan Word2Vec dalam meningkatkan akurasi klasifikasi pada penelitian ini.

## DAFTAR PUSTAKA

[1] S. Alfarizi and E. Fitriani, "Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization," *Indonesian Journal on Software Engineering (IJSE)*, vol. 9, no. 1, pp. 19–27, 2023, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ijse

[2] NURUL AFIFAH, Dony Permana, Dodi Vionanda, and Dina Fitria, "Sentiment Analysis of Electric Cars Using Naive Bayes Classifier Method," *UNP Journal of Statistics and Data Science*, vol. 1, no. 4, pp. 289–296, Aug. 2023, doi: 10.24036/ujsds/vol1-iss4/68.

[3] P. G. Aryanti and I. Santoso, "ANALISIS SENTIMEN PADA TWITTER TERHADAP MOBIL LISTRIK MENGGUNAKAN ALGORITMA NAIVE BAYES." [Online]. Available: https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive

[4] A. Deviyanto, M. R. Didik Wahyudi, and T. Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, "PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *Jurnal*

*Informatika Sunan Kalijaga)*, vol. 3, no. 1, pp. 1–13, 2018, [Online]. Available: https://twitter.com/search?l=id&q=AHY%20since%3A2017-01-01%20until%3A2017-01-

[5] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," 2017. [Online]. Available: https://t.co/jrvaMsgBdH

[6] E. Nurhazizah, R. Nur Ichsan, and S. Widiyanesti, "24~35 Diterima Februari 11," *JURNAL SWABUMI*, vol. 10, no. 1, p. 2022, 2022.

[7] T. Ansyor Lorosae and B. Dwi Prakoso, "Seminar Nasional Teknologi Informasi dan Multimedia," *UNIVERSITAS AMIKOM Yogyakarta*, 2018.

[8] M. N. Fahriza and N. Riza, "ANALISIS SENTIMEN PADA ULASAN APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR(KNN) Sistematic Literature Review," 2023.

[9] Edwin Febrywinata, "Pengenalan Dan Klasifikasi Jenis Buah Menggunakan Metode CNN Secara Sederhana Dengan Menggunakan Google Colab," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 185–193, Jun. 2024, doi: 10.61132/merkurius.v2i4.162.

[10] F. Nur Rozi and D. Harini Sulistyawati, "KLASIFIKASI BERITA HOAX PILPRES MENGGUNAKAN METODE MODIFIED K-NEAREST NEIGHBOR DAN PEMBOBOTAN MENGGUNAKAN TF-IDF," 2019.

[11] R. Hutagaol and D. Yandra Niska, "SOSIAL MEDIA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER," 2023.

[12] A. Wijaya, C. Rozikin, and B. N. Sari, "Penerapan Text Mining Untuk Klasifikasi Judul Berita Hoax Vaksinasi COVID-19 Menggunakan Algoritma Support Vector Machine," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 16, pp. 11–20, 2022, doi: 10.5281/zenodo.7058890.

[13] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.

[14] M. Furqan, S. Mayang Sari, and P. Ilmu Komputer Fakultas Sains dan Teknologi, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia Sentiment Analysis using K-Nearest Neighbor towards the New Normal During the Covid-19 Period in Indonesia," 2022. [Online]. Available: www.tripadvisor.com

[15] H. Parasian Doloksaribu and Y. T. Samuel, "KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI," vol. 16, no. 1, 2022, doi: 10.47111/JTI.

[16] B. Laurensz, A. Sentimen, and E. Sediyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic)," 2021.

[17] I. Habib Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, 2023.

[18] Y. Yudi, U. Gultom, and R. Haroen, "Perancangan Sistem Informasi E-Commerce Pada Sales Auto 2000 Ciledug," *Jurnal Manajamen Informatika Jayakarta*, vol. 1, no. 2, p. 134, 2021, doi: 10.52362/jmijayakarta.v1i2.449.

[19] R. Rakhmat Sani, Y. Ayu Pratiwi, S. Winarno, E. Devi Udayanti, and dan Farrikh Al Zami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia," 2022.

[20] A. Fitri Niasita, P. P. Adikara, and S. Adinugroho, "Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan Automated Lexicon Word2Vec dan Naive-Bayes," 2019. [Online]. Available: http://j-ptiik.ub.ac.id

[21] N. Tri Romadloni, I. Santoso, S. Budilaksono, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE."

[22] R. Miya Juwita, E. Haerani, S. Kurnia Gusti, and dan Siti Ramadhani, "Klasifikasi Berita Menggunakan Metode K-Nearest Neighbor," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, 2022.

[23] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 646, Dec. 2021, doi: 10.32493/informatika.v5i4.7622.

[24] J. Minfo Polgan *et al.*, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier", [Online]. Available: www.tokopedia.com

[25] M. Syarifuddinn, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.