

Analisis Sentimen Kendaraan Listrik Menggunakan Perbandingan Naïve Bayes dan Support Vector Machine

Electric Vehicle Sentiment Analysis Using a Comparison of Naïve Bayes and Support Vector Machine

Faizah Dian Herawati^{*1}, Frederik Samuel Papilaya²

¹Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana

² Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana

Email: ¹682022605@student.uksw.edu, ² samuel.papilaya@uksw.edu

*Penulis Koresponden

Received: 17 Januari 2026

Accepted: 28 Januari 2026

Published: 02 Februari 2026



This work is licensed under a Creative Commons Attribution 4.0 International License.
Copyright (c) 2026 JUSTINDO

ABSTRAK

Perkembangan kendaraan listrik di Indonesia memicu berbagai opini dari masyarakat yang sering dibagikan melalui media sosial, terutama X. Opini tersebut perlu dianalisis untuk bisa memahami bagaimana masyarakat memandang kebijakan serta pelaksanaan kendaraan yang ramah lingkungan. Penelitian ini bertujuan untuk membandingkan dua jenis algoritma pengklasifikasian teks, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam melihat opini masyarakat terhadap kendaraan listrik dengan menggunakan pendekatan *Term Frequency–Inverse Document Frequency* (TF-IDF). Data yang digunakan adalah *tweet* dalam bahasa Indonesia yang dikumpulkan melalui proses *crawling*, lalu melewati beberapa tahap pra-proses seperti *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Setelah itu, data diberi label sentimen menjadi tiga kategori, yaitu positif, negatif, dan netral, sebelum diolah menggunakan algoritma klasifikasi. Untuk mengevaluasi kemampuan model, digunakan *confusion matrix* yang menunjukkan performa algoritma berdasarkan nilai akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki hasil yang lebih baik dengan akurasi sebesar 92%, sedangkan SVM mencapai akurasi sebesar 76%. Karena itu, algoritma *Naïve Bayes* dianggap lebih cocok untuk menganalisis sentimen dari *tweet* terkait kendaraan listrik di Indonesia.

Kata kunci: Analisis Sentimen, Kendaraan Listrik, Naïve Bayes, Support Vector Machine, TF-IDF

ABSTRACT

The development of electric vehicles in Indonesia has sparked various opinions from the public, which are often shared on social media, especially X. These opinions need to be analyzed to understand how the public views the policies and implementation of environmentally friendly vehicles. This study aims to examine public sentiment toward electric vehicles by comparing two types of text classification algorithms, namely *Naïve Bayes* and *Support Vector Machine* (SVM), using the *Term Frequency–Inverse Document Frequency* (TF-IDF) approach. The data used is Indonesian-language tweets collected through a crawling process, which then undergoes several pre-processing stages such as cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. After that, the data was labeled for sentiment into three categories: positive, negative, and neutral, before being processed using a classification algorithm. To evaluate the model's performance, a confusion matrix was used, which shows the algorithm's performance based on accuracy, precision, recall, and F1-score values. The research results show that the *Naïve Bayes* algorithm has better results with an accuracy of 92%, while SVM achieves an accuracy of 76%. Therefore, the *Naïve Bayes* algorithm is considered more suitable for analyzing the sentiment of tweets related to electric vehicles in Indonesia.

Keywords: Sentiment Analyst, Electric Vehicles, Naïve Bayes, Support Vector Machine, TF-IDF

1. Pendahuluan

Di Indonesia, sebagai negara dengan jumlah penduduk yang tinggi, membuat kebutuhan masyarakat kian meningkat. Termasuk dalam bidang teknologi otomotif sehingga mendorong kehadiran kendaraan listrik yang dianggap sebagai solusi terhadap permasalahan harga bahan bakar yang terus meningkat. Kendaraan listrik dianggap sebagai solusi yang lebih efisien dan lebih baik untuk lingkungan. Dengan munculnya tren kendaraan listrik, yang muncul sebagai alternatif bagi kendaraan konvensional berbahan bakar fosil. Karena kendaraan listrik cenderung lebih hemat energi, ramah lingkungan, dan memiliki biaya operasional yang lebih rendah. Tren ini tidak hanya terjadi karena kebijakan pemerintah dan industri otomotif, tapi juga muncul dalam pembicaraan opini di media sosial dan platform digital, di mana pengguna berbagi pendapat serta pengalaman mereka mengenai kendaraan listrik (Tarigan and Yusupa, 2024); (Kusuma et al., 2023).

Opini masyarakat terhadap kehadiran teknologi otomotif berupa kendaraan listrik masih terus menjadi perbincangan hingga saat ini. Opini pengguna media sosial yang terdapat pada aplikasi X menjadi objek analisis untuk melihat berbagai opini mengenai kesan terhadap keberadaan kendaraan listrik, apakah positif atau negatif. Sebagian besar opini masyarakat pengguna X menghasilkan data teks dengan jumlah besar dan tidak terstruktur. Data data opini tersebut sulit untuk dianalisis dengan metode manual, sehingga memerlukan suatu pendekatan berbasis *machine learning* untuk klasifikasi sentimen secara efektif (Hendrawan and Kusmiyati, 2024). Analisis sentimen adalah bagian dari *text mining* yang bisa digunakan untuk mengklasifikasikan opini seseorang menjadi tiga kategori, yaitu positif, negatif, atau netral, berdasarkan bahasa dan situasi dalam teks (Lende et al., 2024).

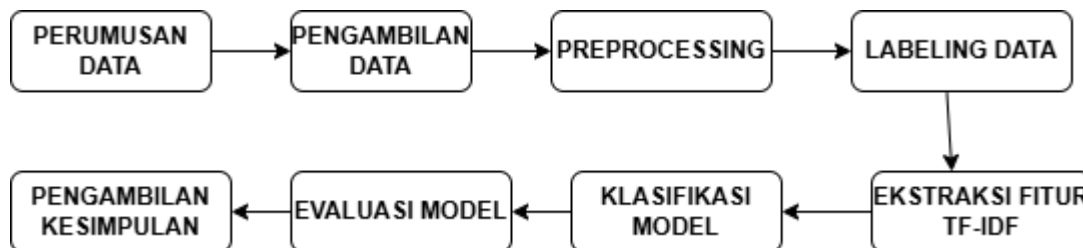
Sebelum melakukan analisis sentimen masyarakat pengguna X terhadap kendaraan listrik, diperlukan pengambilan data pada X. Data diperoleh dengan *crawling* data menggunakan *auth_token* dengan memanfaatkan *google collaboratory* untuk pengelolaan data dan implementasi algoritma klasifikasi. Kemudian dilakukan teks praproses dengan menggunakan metode *text mining* untuk mengolah data tidak terstruktur pada hasil *crawling* atau *scraping* agar data menjadi teks struktur dan siap untuk dianalisis menggunakan algoritma *text classification*. Melalui data tersebut maka akan disimpulkan tentang opini masyarakat pengguna X terhadap kendaraan listrik.

Beberapa penelitian terdahulu telah mencoba menganalisis sentimen masyarakat terhadap kendaraan listrik di X. Misalnya, penelitian oleh Agustian et al., (2022) menerapkan algoritma *Naive Bayes* dan *preprocessing* sederhana seperti pembersihan teks dan penerjemahan ke bahasa Inggris. Hasilnya menunjukkan akurasi 80%, dengan presisi 82% dan recal 80% namun penelitian ini belum membandingkan dengan algoritma lain, serta jumlah informasi yang digunakan relatif sedikit, yaitu hanya 1.589 tweet. Di sisi lain, Prawinata Dian Agus et al. (2024) mengungkap pendekatan *deep learning* dengan menggunakan LSTM dan fitur Word2Vec (CBOW & Skip-Gram). Dataset-nya lebih besar, yakni 30.000 tweet, dan hasilnya cukup baik dengan akurasi tertinggi mencapai 85,9%. Meskipun hasil pendekatan lebih akurat, penelitian hanya menggunakan satu algoritma tanpa membandingkan performa. Sementara itu, (Merdiansyah et al., 2024) memanfaatkan model berbasis transformer, yaitu IndoBERT, yang dilatih khusus untuk Bahasa Indonesia, dengan akurasi mencapai 98–99%. Namun dengan kebutuhan komputasi yang cukup tinggi, pendekatan ini cukup berat dari sisi komputasi, dan belum ada pembandingan dengan metode yang lebih ringan seperti TF-IDF + Naive Bayes dan SVM, yang lebih realistis untuk implementasi di skala kecil atau pemula.

Berdasarkan dari tiga penelitian terdahulu yang sudah disebutkan, maka penelitian ini menghadirkan pembaruan dengan membandingkan dua algoritma *text classification*, yaitu *Naive Bayes* dan SVM, serta menggunakan TF-IDF sebagai pendekatan pada data X terbaru yang berkaitan dengan *tweets* kendaraan listrik di Indonesia. Penelitian ini bertujuan untuk mengelompokkan komentar masyarakat menjadi tiga kelas yakni positif, negatif dan netral. Penggunaan data kontekstual terkini memungkinkan analisis yang lebih akurat terhadap opini publik, khususnya dalam merespons kebijakan pemerintah terkait kendaraan ramah lingkungan. Selain klasifikasi, penelitian ini juga menganalisis kecenderungan opini berdasarkan tema atau waktu tertentu, yang bertujuan memberikan pemahaman mendalam bagi pengambil kebijakan dan pelaku industri otomotif nasional.

2. Metode Penelitian

Dalam penelitian ini, algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* digunakan sebagai metode klasifikasi sentimen. Algoritma *Naïve Bayes* dipilih karena kemampuannya dalam mengolah data teks berdimensi tinggi. SVM digunakan sebagai pembanding karena kemampuannya dalam memisahkan kelas data secara optimal. Tahapan metode penelitian dapat dilihat pada Gambar 1. berikut:



Gambar 1. Tahapan Penelitian

2.1. Perumusan data

Perumusan data penelitian merupakan tahap awal yang bertujuan untuk menentukan karakteristik data yang digunakan dalam penelitian ini. Data penelitian dirumuskan sebagai *tweet* berbahasa Indonesia yang memuat opini masyarakat terkait kendaraan listrik di Indonesia. Setiap *tweet* ditetapkan sebagai satu unit analisis dan akan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Selain itu, ditetapkan kriteria data yang digunakan, meliputi relevansi isi *tweet* terhadap topik penelitian. Perumusan data ini menjadi dasar dalam analisis sentimen.

2.2. Pengambilan data

Berdasarkan perumusan data yang telah ditetapkan, pengumpulan data dilakukan untuk memperoleh data mentah yang sesuai dengan kriteria penelitian dengan *scraping* dari *media social X* yang menyediakan data *tweets* opini pengguna terkait kendaraan listrik. Data yang dikumpulkan berupa teks yang berisi keluhan, pendapat, pengalaman, maupun penilaian terhadap kehadiran kendaraan listrik. Pengambilan data menggunakan kata kunci yang berkaitan dengan kendaraan listrik, seperti, "mobil listrik", "kendaraan ramah lingkungan", "kendaraan listrik", dan "motor listrik". Data dibatasi pada *tweets* berbahasa Indonesia dengan rentang waktu tertentu agar sesuai dengan tujuan penelitian. Data yang didapatkan masih berupa data mentah dan harus diproses tahap selanjutnya.

2.3. Preprocessing

Setelah selesai dalam tahap pengumpulan data dari X data tersebut akan diproses sebelum diklasifikasikan. Tahapan sebelum dilakukan klasifikasi model adalah *preprocessing*. *Preprocessing* data adalah tahap yang wajib dilakukan setelah melakukan *scraping* data teks dari X. Tahapan *preprocessing* yang dilakukan meliputi, *cleaning*, *case folding*, normalisasi kata, *tokenize*, *stopword removal*, dan *stemming*. Pada tahap *cleaning*, semua bagian teks yang tidak penting akan dibuang, seperti URL, tanda baca, angka, emoji, tagar (#), *mention* (@), atau karakter lain yang tidak diperlukan. Tujuan dilakukan proses *cleaning* adalah agar data yang di ambil fokus pada kata yang memiliki makna, dan kata yang tidak relevan dengan topik bisa di kurangi sejak awal sehingga hasil analisis bisa maksimal. Proses *cleaning* perlu dilakukan karena dataset hasil dari *scraping* banyak mengandung kata-kata yang tidak relevan untuk analisis sentimen (Ritonga and Sriani, 2025). Tahapan selanjutnya adalah tahapan *case folding*, setelah dataset dilakukan tahap *cleaning* data, semua teks dilakukan tahap *case folding*, untuk mengubah seluruh data teks menjadi huruf kecil atau *lowercase* agar kata "Mobil Listrik", mobil listrik", dan "MOBIL LISTRIK" terbaca sebagai kata yang memiliki arti yang sama oleh komputer (Ritonga and Sriani, 2025). Tahapan selanjutnya adalah tahapan normalisasi kata, normalisasi kata adalah proses memperbaiki kata yang tidak sesuai standar, seperti kesalahan menulis, kata kasual, atau singkatan, menjadi bentuk bahasa yang benar dan terstandar dengan bantuan kamus slang. Karena dataset berasal dari pengguna media sosial, pasti terdapat kalimat yang menggunakan kata tidak baku, slang, atau singkatan seperti "gokil" atau

"gw". Tahap normalisasi ini sangat penting terutama untuk data dari media sosial, karena banyak ditemukan penggunaan bahasa yang tidak formal dan istilah slang yang bisa mengganggu hasil analisis (Rifaldi et al., 2023). Tahapan selanjutnya adalah tahapan tokenisasi, yaitu memisahkan kalimat panjang menjadi beberapa kata yang disebut dengan token (Ridwansyah, 2022). Misalnya pada kalimat "mobil listrik hemat bensin" menjadi ["mobil", "listrik", "hemat", "bensin"]. Tahap tokenisasi merupakan tahap penting dalam *preprocessing* karena komputer tidak bisa menganalisis kalimat secara utuh, token lebih mudah untuk dianalisis dan dimasukkan kedalam model klasifikasi *machine learning*. Tahapan selanjutnya adalah tahapan *stopword removal* yaitu tahap penghapusan *stopword* merupakan proses menghilangkan kata atau token yang tidak berpengaruh terhadap analisis sentimen, seperti konjungsi. Tujuan dari penghapusan *stopword* adalah meningkatkan akurasi model dalam memprediksi sentimen dengan menghilangkan kata yang tidak penting dari teks, sehingga model hanya fokus pada kata-kata yang memiliki makna dan berpengaruh. Dengan mengurangi jumlah kata yang diproses, jumlah fitur berkurang dan proses komputasi menjadi lebih cepat dan efisien (Handaya and Wahyu, 2024). Tahapan selanjutnya adalah tahapan *Stemming*, setelah teks sudah melalui tahapan *cleaning*, normalisasi, dan dipecah menjadi token, setiap token biasanya masih dalam bentuk turunan kata. *Stemming* menghapus kata imbuhan seperti kata "menyukai" kedalam bentuk dasarnya yaitu "suka". Tahapan ini membantu mengurangi variasi kata yang memiliki arti sama (Setiawan and Nastiti, 2024). Seluruh tahapan *preprocessing* dilakukan untuk menghasilkan data teks terstruktur sehingga proses klasifikasi sentimen menjadi maksimal.

2.4. Labeling data

Setelah melakukan tahap *preprocessing*, proses selanjutnya adalah proses *labeling* sentimen, di mana *tweet* yang telah dikumpulkan akan diberi label. Setiap *tweet* diklasifikasikan ke dalam salah satu dari tiga kategori sentimen, yaitu positif, negatif, atau netral. Proses pelabelan sentimen menggunakan sebuah metode, yaitu *lexicon based*. *Lexicon based* adalah metode pelabelan dataset sentimen dengan bahasa Indonesia menggunakan kamus *lexicon*. Dalam kamus *lexicon* terdapat daftar kata negatif dan positif yang digunakan untuk acuan perhitungan bobot dari sebuah kata. Kamus *labeling lexicon* tersebut bernama *lexicon InSet* (Asri et al., 2025); (Musfiroh, 2021)

2.5. Ekstraksi fitur TF-IDF

Setelah data sentimen sudah melalui tahap *labeling*, langkah berikutnya adalah ekstraksi fitur, yaitu mengubah teks menjadi bentuk numerik. Dalam penelitian ini, pendekatan yang digunakan adalah TF-IDF (*Term Frequency - Inverse Document Frequency*). TF-IDF digunakan untuk mengukur seberapa penting suatu kata dalam dokumen dengan mempertimbangkan frekuensi kata dan distribusinya. TF-IDF membantu membuat representasi numerik yang lebih baik dari teks untuk model klasifikasi sehingga hasil prediksi bisa lebih akurat (Lestari and Hutagalung, 2025).

2.6. Klasifikasi Model

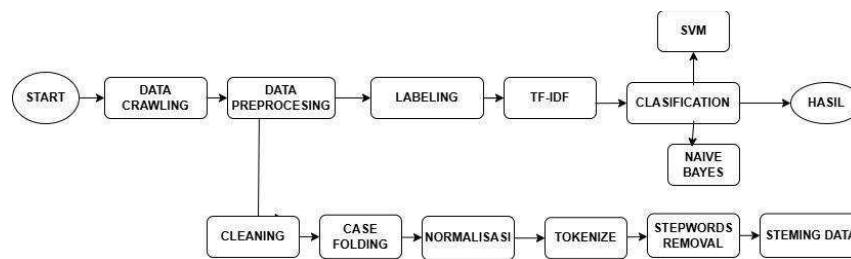
Penelitian ini menggunakan dua jenis algoritma klasifikasi model yaitu *Naïve bayes* dan *support vector machine (SVM)* dengan *python* sebagai bahasa pemrograman. *Library sklearn* juga digunakan untuk membuat model klasifikasi. Kedua algoritma klasifikasi tersebut dijelaskan sebagai berikut:

2.6.1 Klasifikasi *Naïve Bayes*

Naïve bayes merupakan salah satu algoritma klasifikasi sederhana berbasis probabilitas yang sering digunakan dalam melakukan analisis sentimen, terutama pada data teks. Dalam proses perhitungannya, *Naïve bayes* mengasumsikan bahwa setiap fitur bersifat *independent* setelah diketahui kelasnya, sehingga perhitungannya lebih sederhana dan efisien. Dengan asumsi tersebut, *Naïve Bayes* mampu menangani dataset dalam skala besar terutama ketika dikombinasikan dengan teknik ekstraksi fitur seperti TF-IDF pada data berbasis teks (Permataning Tyas et al., 2022).

2.6.2 Klasifikasi Support Vector Machine SVM

Dalam proses klasifikasi, SVM memiliki kemampuan dalam menangani data yang berdimensi tinggi seperti hasil dari pembobotan dengan TF-IDF. SVM bekerja dengan mencari *hyperplane* antara kelas positif, negative, dan netral, sehingga setiap contoh dapat sentiment dan mengukur akurasinya, sehingga menghasilkan klasifikasi yang efektif (Azzahra and Mailoa, 2025). Tahapan klasifikasi dengan *Naïve Bayes* dan SVM terdapat pada Gambar 2. berikut:



Gambar 2. Tahapan klasifikasi *Naïve Bayes* dan SVM

2.7. Evaluasi model

Evaluasi model dilakukan untuk mengetahui bagaimana algoritma *Naïve Bayes dan Support Vector Machine* dalam mengklasifikasikan sentimen dari data uji. Pada tahap evaluasi model menggunakan *confusion matrix* untuk membandingkan hasil prediksi model dengan label sentimen sebenarnya, sehingga dapat menghitung beberapa nilai seperti akurasi, *precision*, *recall*, dan *F1-score*. Akurasi menunjukkan seberapa tepat model dalam memberikan prediksi secara keseluruhan, sedangkan *precision* menggambarkan kemampuan model dalam memprediksi sentimen secara benar, dan *recall* menggambarkan seberapa baik model dalam menangkap semua kelas sentimen. *F1-score* merupakan kombinasi dari *precision* dan *recall* yang digunakan untuk menilai keseimbangan antara keduanya. Hasil dari semua matriks evaluasi tersebut kemudian dibandingkan untuk mengetahui algoritma mana yang lebih baik dalam analisis sentimen terkait kendaraan listrik berdasarkan data yang digunakan.

3. Hasil dan Pembahasan

3.1. Perumusan data

Penelitian ini berdasarkan pada data yang telah ditetapkan, dengan fokus pada *tweet* berbahasa Indonesia yang mengandung opini masyarakat terkait kendaraan listrik. Data yang digunakan mencakup topik seperti kendaraan listrik, mobil listrik, motor listrik, serta kendaraan ramah lingkungan. Dari data mentah hasil pengumpulan, dilakukan proses seleksi dan *preprocessing* untuk memastikan kesesuaian data dengan tujuan penelitian. Proses ini menghasilkan data yang digunakan sebagai dasar dalam memprediksi perilaku ekonomi. Setiap *tweet* yang telah diproses akan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Data hasil perumusan ini kemudian digunakan pada tahap analisis sentimen dengan metode TF-IDF serta algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*.

3.2. Pengambilan data

Proses pengambilan data menggunakan informasi dari data tweets dengan *crawling* data menggunakan *auth_token* dari X, dan memanfaatkan *platform google colab*. Kata kunci yang digunakan untuk pengambilan data adalah menggunakan kata kunci “kendaraan listrik, mobil listrik, motor listrik, dan kendaraan ramah lingkungan” dengan jumlah data yang berhasil dikumpulkan berjumlah 3466 tweet terbaru tahun 2023-2025 dengan tweet berbahasa Indonesia. Hasil scraping disimpan dalam file dengan format csv dan masih berupa data mentah, sehingga harus memerlukan tahap *preprocessing* sebelum dilakukan analisis. Contoh hasil *crawling* dapat dilihat pada Tabel 1:

Tabel 1. Hasil *Scraping*

Hasil Scarping
@unpad Jadi kendaraan listrik tetap punya potensi besar. Tapi harus jalan bareng energi bersih dan reformasi sistem transportasi. Kalau enggak ya cuma solusi semu. #NetZero #EnergiBersih #RisetUnpad
Apakah Kawan KABAR salah satu pengguna kendaraan listrik (EV)? Bagikan pengalamanmu yang paling berkesan saat mengendarai EV dengan reply post ini! #BicaraSebenTAR #KoalisiIndonesiaBebasTAR #KABAR
Seiring dengan melonjaknya permintaan kendaraan listrik Indonesia telah menjadi pemimpin dalam produksi nikel menempatkan Halmahera Morowali dan Konawe Utara di jantung rantai pasokan. FYI Petrokimia Gresik juga sudah operasikan 201 kendaraan listrik di lingkungan pabrik. Aksinya bukan main-main! #UjiEmisiGratisPG #HariLingkunganHidup #PetrokimiaGresik #SolusiAgroindustri

Hasil Scarping

Kendaraan listrik ini punya kapasitas luas tanpa emisi dan biaya operasional yang jauh lebih hemat dibanding mobil niaga konvensional. Cocok banget buat kamu yang pengen bisnis berkelanjutan.

3.3. Preprocessing

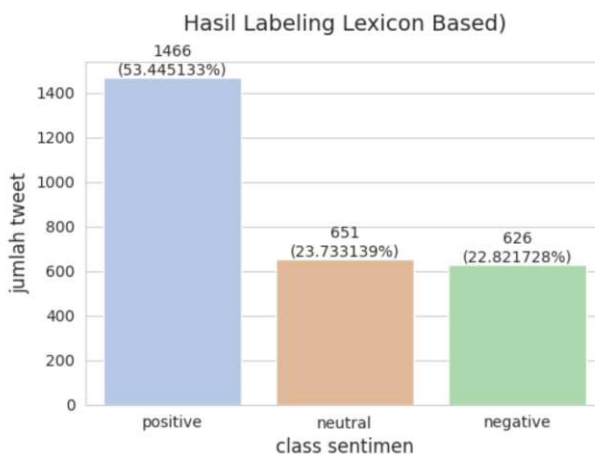
Setelah dataset hasil *scraping* digabungkan, dataset akan melakukan tahap preprocessing, karena data hasil *scraping* merupakan data yang tidak terstruktur. Dalam *preprocessing* juga terdapat proses penghapusan data duplikat dan penghapusan data yang kosong. Hasil *preprocessing* dapat dilihat pada Tabel 2:

Tabel 2. Hasil *Preprocessing*

Tahapan	Hasil
Data Tweet	@unpad Jadi kendaraan listrik tetap punya potensi besar. Tapi harus jalan bareng energi bersih dan reformasi sistem transportasi. Kalau enggak ya cuma solusi semu. #NetZero #EnergiBersih #RisetUnpad
Cleaning	Jadi kendaraan listrik tetap punya potensi besar. Tapi harus jalan bareng energi bersih dan reformasi sistem transportasi. Kalau enggak ya cuma solusi semu.
Case Folding	jadi kendaraan listrik tetap punya potensi besar. tetapi harus berjalan bareng energi bersih dan reformasi sistem transportasi. kalau enggak ya cuma solusi semu.
Normalisasi kata	jadi kendaraan listrik tetap punya potensi besar. tapi harus jalan bareng energi bersih dan reformasi sistem transportasi. kalau tidak hanya solusi semu.
Tokenize	[jadi, kendaraan, listrik, tetap, punya, potensi, besar, tapi, harus, jalan, bareng, energi, bersih, dan, reformasi, sistem, transportasi, kalau, tidak, hanya, solusi, semu]
Stopword Removal	[kendaraan, listrik, potensi, besar, jalan, bareng, energi, bersih, reformasi, sistem, transportasi, tidak, hanya, solusi, semu]
Stemming	kendaraan listrik potensi besar jalan bareng energi bersih reformasi sistem transportasi tidak hanya solusi semu

3.4. Labeling data

Setelah dilakukan *preprocessing*, kemudian dilakukan pelabelan dataset dengan menggunakan *lexicon based*. Hasil pelabelan dapat dilihat pada Gambar 3:



Gambar 3. Labeling data

Berdasarkan Gambar 3. Hasil *labeling* data sentiment menggunakan *lexicon based* bahwa dari 2743 tweet yang dianalisis, mayoritas data termasuk dalam sentimen positif dengan jumlah 1466 tweet atau sekitar 53,45%. Hal ini menunjukkan bahwa sebagian besar pengguna cenderung memberikan opini yang bersifat positif terhadap kendaraan listrik. Selanjutnya, sentiment netral berjumlah 651 tweet atau 23,73%, yang menunjukkan terdapat opini yang tidak menunjukkan kecenderungan sentiment atau emosi tertentu. Sementara itu, sentimen negatif memiliki jumlah 626 tweet atau sekitar 22,82%, yang menunjukkan bahwa adanya kritik atau opini kurang positif terhadap kendaraan listrik. Secara keseluruhan, distribusi sentimen ini menunjukkan bahwa persepsi publik terhadap kendaraan listrik cenderung positif, meskipun masih terdapat proporsi sentimen netral dan negatif yang cukup signifikan dan perlu diperhatikan dalam analisis lanjutan.

3.5. Pembobotan TF-IDF

Tahap selanjutnya adalah tahap pembobotan kata menggunakan pendekatan TF-IDF. Pendekatan ini dilakukan dengan menggunakan *Count-Vectorizer*. Pendekatan *Count-Vectorizer* digunakan untuk menentukan intensitas kata yang muncul dalam setiap tweet. Data dibagi menjadi 80% data pelatihan dan 20% data uji sebelum pembobotan kata. Pemodelan dilakukan menggunakan data pelatihan berdasarkan komentar positif dan negatif. Pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata dari setiap teks yang dipresentasikan dalam bentuk vector. Hasil ekstraksi fitur terdiri dari 4502 data sebagai data latih dan 1125 data sebagai data uji.

3.6. Klasifikasi model

3.6.1 Klasifikasi model *Naive Bayes*

Hasil klasifikasi sentimen dengan model *Naive Bayes* menunjukkan persentase *fi-score*(kelas tertinggi, 94%) dengan presentasi *recall* tertinggi pada kelas positif yaitu 98%. Persentase presisi tertinggi diperoleh dari kelas negatif dengan nilai 95%. Secara keseluruhan klasifikasi dengan model *Naive Bayes* memiliki akurasi keseluruhan sebesar 72%. Gambar 4. Menunjukkan hasil klasifikasi model menggunakan *Naive Bayes*.

```

classification_report:
              precision    recall  f1-score   support

   negative         0.95         0.91         0.93         519
    neutral         0.94         0.82         0.87         528
   positive         0.91         0.98         0.94        1147

 accuracy                   0.92         2194
 macro avg         0.93         0.90         0.92         2194
 weighted avg         0.93         0.92         0.92         2194
    
```

Gambar 4. Hasil klasifikasi naive bayes

3.6.2 Klasifikasi *Support Vector Machine*

Hasil klasifikasi sentimen dengan model SVM menunjukkan persentase *fi-score* (kelas tertinggi, 87%) dengan presentasi *recall* tertinggi pada kelas positif yaitu 88%. Presentasi presisi tertinggi diperoleh dari kelas positif dengan 85%. Secara keseluruhan klasifikasi dengan model SVM memiliki akurasi keseluruhan sebesar 76%. Gambar 5. Menunjukkan hasil klasifikasi model menggunakan *Naive Bayes*.

```

classification_report:
              precision    recall  f1-score   support

   negative         0.71         0.70         0.71         107
    neutral         0.52         0.47         0.50         123
   positive         0.85         0.88         0.87         319

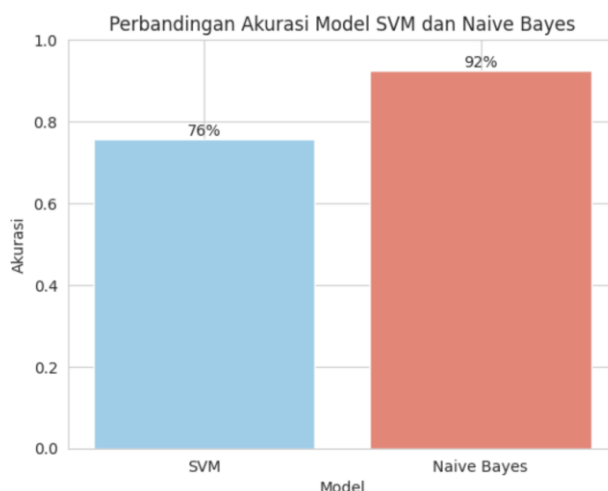
 accuracy                   0.76         549
 macro avg         0.69         0.69         0.69         549
 weighted avg         0.75         0.76         0.75         549
    
```

Gambar 5. Hasil klasifikasi SVM

3.7. Evaluasi dan visualisasi model

Hasil visualisasi perbandingan akurasi algoritma *naive bayes* dan SVM. Algoritma *naive bayes* mendapatkan akurasi sebesar 92%. Sedangkan SVM mendapatkan akurasi sebesar 76%. Dari hasil akurasi tersebut dapat dilihat bahwa algoritma *Naive Bayes* memiliki akurasi lebih tinggi dari algoritma SVM dalam mengklasifikasikan sentiment pada data teks X terkait kendaraan listrik. Sehingga algoritma *Naive Bayes* lebih efektif dalam menangani karakteristik data sentimen yang

memiliki sifat pendek, tidak terstruktur, serta memiliki distribusi kata yang tidak merata. Gambar 6. Menunjukkan hasil visualisasi perbandingan algoritma:



Gambar 6. Hasil visualisasi perbandingan algoritma

Selain itu, hasil pelabelan sentimen menunjukkan bahwa sentimen positif mendominasi dengan proporsi lebih dari 50%, sementara sentimen netral dan negatif relatif seimbang. Kondisi ini mendukung kinerja baik algoritma *Naive Bayes* yang dikenal stabil pada data dengan distribusi kelas yang tidak sepenuhnya seimbang, serta berbasis probabilistik sederhana menggunakan representasi fitur TF-IDF. Temuan ini sejalan dengan penelitian Agustian et al. (2022) yang menyatakan bahwa algoritma *Naive Bayes* mampu memberikan performa yang baik dalam analisis sentimen X, meskipun pendekatan yang digunakan tidak sepenuhnya sederhana. Oleh karena itu, hasil penelitian ini menegaskan bahwa metode *Naive Bayes* merupakan pendekatan yang efektif dan realistis dalam analisis sentimen pada data media sosial X. Dengan demikian, pendekatan ini sangat layak untuk dipakai dalam penelitian berskala kecil hingga menengah.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil menyajikan pembaruan dengan membandingkan dua algoritma, *Naive Bayes* dan *Support Vector Machine (SVM)*, dalam mengklasifikasikan sentimen masyarakat terhadap kendaraan listrik di Indonesia menggunakan pendekatan TF-IDF pada data X terbaru. Dari hasil evaluasi, terlihat bahwa algoritma *Naive Bayes* menunjukkan performa yang lebih baik dibandingkan SVM, dengan akurasi sebesar 92%. Nilai *precision*, *recall*, dan *F1-score* pada *Naive Bayes* relatif seimbang di ketiga kelas sentimen. Pada kelas negatif, *Naive Bayes* mendapatkan *precision* 0,95, *recall* 0,91, dan *F1-score* 0,93; pada kelas netral, *precision* 0,94, *recall* 0,82, dan *F1-score* 0,87; sedangkan pada kelas positif, *precision* 0,91, *recall* 0,98, dan *F1-score* 0,94. Sementara itu, algoritma SVM hanya mencapai akurasi 76% dan menunjukkan performa yang lebih rendah, terutama pada kelas netral. Temuan ini menunjukkan bahwa *Naive Bayes* lebih efektif dalam menangani data teks X pada penelitian ini, serta mampu memberikan gambaran yang lebih akurat mengenai opini publik terhadap kendaraan listrik, yang dapat menjadi pertimbangan bagi pengambil kebijakan serta pelaku industri otomotif di Indonesia.

Daftar Pustaka

- Agustian, A., Tukero, T., Nurapriani, F., 2022. Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter. *Jurnal TIK* 7, 243–249. <https://doi.org/10.51179/tika.v7i3.1550>
- Asri, Y., Kuswardani, D., Suliyanti, W.N., Manullang, Y.O., Ansyari, A.R., 2025. Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application.

- Indonesian Journal of Electrical Engineering and Computer Science 38, 677. <https://doi.org/10.11591/ijeecs.v38.i1.pp677-688>
- Azzahra, W.L., Mailoa, E., 2025. Analisis Sentimen terhadap RSUD Salatiga Menggunakan SVM dan TF-IDF. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi* 6, 478–489. <https://doi.org/10.35870/jimik.v6i1.1208>
- Handaya, K.K., Wahyu, S., 2024. Sentiment Analysis on Hate Speech Post 2024 Election for Elected President Using a Hybrid Model Machine Learning. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi* 10, 128–139. <https://doi.org/10.24014/coreit.v10i2.31927>
- Hendrawan, G.N., Kusmiyati, H., 2024. Evaluasi Performa Naive Bayes dan SVM dalam Analisis Sentimen Kendaraan Listrik di Media Sosial Twitter. *CESS (Journal of Computer Engineering, System and Science)* 9, 299–313.
- Kusuma, G.H., Permana, I., Salisah, F.N., Afdal, M., Jazman, M., Marsal, A., 2023. Pendekatan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kendaraan Listrik Pada Sosial Media X. *JUSINFO (Jurnal Sistem Informasi)* 9, 65–76. <https://doi.org/10.19109/jusifo.v9i2.21354>
- Lende, S., Pati, G.K., Adis, A., Stella, S., Sumba, M., 2024. Analisis Sentimen Komentar Pengunjung Air Terjun Waikelo Sawa Desa Tema Tana Wewewa Timur Kabupaten Sumba Barat Dengan Metode Naive Bayes Classifier. *Jurnal Ilmu Komputer dan Bisnis XV*, 42–50.
- Lestari, V.B., Hutagalung, C.A., 2025. Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. *J-KOMA Journal of Computer Science and Applications* 08, 22–2025.
- Merdiansyah, R., Siska, S., Ali Ridha, A., 2024. Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)* 7, 221–228. <https://doi.org/10.55338/jikomsi.v7i1.2895>
- Musfiroh, D., 2021. Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 1, 24–33.
- Permataning Tyas, S.M., Rintyarna, B.S., Suharso, W., 2022. The Impact of Feature Extraction to Naive Bayes Based Sentiment Analysis on Review Dataset of Indihome Services. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi* 13, 1–10. <https://doi.org/10.31849/digitalzone.v13i1.9158>
- Prawinata Dian Agus, Rahajoe Ani Dijah, Diyasa I Gede Susrama Mas, 2024. Analisis Sentimen Kendaraan Listrik Pada Twitter Menggunakan Metode Long Short Term Memory. *SABER: Jurnal Teknik Informatika, Sains dan Ilmu Komunikasi* 2, 300–313. <https://doi.org/10.59841/saber.v2i1.857>
- Ridwansyah, T., 2022. Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naive Bayes Classifier. *KLIK: Kajian Ilmiah Informatika dan Komputer* 2, 178–185.
- Rifaldi, D., Abdul Fadlil, Herman, 2023. TEKNIK PREPROCESSING PADA TEXT MINING MENGGUNAKAN DATA TWEET “MENTAL HEALTH.” *DECODE: Jurnal Pendidikan Teknologi Informasi* ISSN 3, 161–171.
- Ritonga, R.R., Sriani, S., 2025. Public Opinion Sentiment Analysis Towards Government Budget Efficiency Policy on Twitter (X) Using the Naive Bayes Classifier Algorithm. *Journal of Information Systems and Informatics* 7, 2496–2515. <https://doi.org/10.51519/journalisi.v7i3.1234>
- Setiawan, M.J., Nastiti, V.R.S., 2024. DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees. *Jurnal Sisfo Kom (Sistem Informasi dan Komputer)* 13, 337–345. <https://doi.org/10.32736/sisfokom.v13i3.2239>
- Tarigan, V.T., Yusupa, A., 2024. Perbandingan Algoritma Machine Learning dalam Analisis Sentimen Mobil Listrik di Indonesia pada Media Sosial Twitter/X. *Jurnal Informatika Polinema* 10, 479–490. <https://doi.org/10.33795/jip.v10i4.5130>