



Research article

Analisis Sentimen Masyarakat terhadap Mobil Listrik di Indonesia Menggunakan Algoritma Naïve Bayes

Sentiment Analysis of Public Perception Towards Electric Cars in Indonesia Using the Naïve Bayes Algorithm

M. Syahrani¹, Muhammad Alief Indra Putra², Amrullah Assalami³

^{1,2,3} Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri Kota Tembilahan, Indragiri Hilir, Riau, Indonesia
email: ^{1,*} first@university.ac.id, ² second@gmail.com, ³ third@yahoo.com, ⁴ fourth@university.edu.my

* Correspondence

ARTICLE INFO

Article history:

Received mm dd, yyyy

Revised mm dd, yyyy

Accepted mm dd, yyyy

Available online mm dd, yyyy

Keywords:

Analisis Sentimen

Mobil Listrik

Naïve Bayes

TF-IDF

Media Sosial

Google Colab

ABSTRACT

Perkembangan mobil listrik di Indonesia menunjukkan peningkatan yang signifikan seiring dengan upaya pemerintah dalam mengurangi emisi karbon dan mendorong penggunaan energi terbarukan. Persepsi dan opini masyarakat terhadap mobil listrik menjadi faktor penting yang memengaruhi tingkat minat untuk memiliki kendaraan ramah lingkungan ini. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap mobil listrik di Indonesia menggunakan data ulasan dan komentar dari media sosial yaitu Aplikasi X (Twitter). Data dikumpulkan melalui teknik web scraping dan disimpan dalam format CSV untuk diproses lebih lanjut menggunakan Google Colab. Tahap preprocessing dilakukan untuk membersihkan data teks, termasuk penghapusan tanda baca, angka, simbol, *stopwords* bahasa Indonesia, *tokenisasi*, dan *stemming*. Selanjutnya, teks diubah menjadi representasi numerik menggunakan metode TF-IDF vectorization agar dapat diproses oleh model klasifikasi *Naïve Bayes*. Model kemudian dilatih dan diuji untuk mengklasifikasikan sentimen menjadi tiga kategori: positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa mayoritas komentar masyarakat terhadap mobil listrik bersifat negatif, dengan persentase mencapai sekitar 66,5%, sementara komentar positif sebesar 33,5%. Analisis lebih lanjut mengungkapkan bahwa sentimen negatif sebagian besar terkait dengan harga kendaraan yang relatif tinggi dan keterbatasan infrastruktur pengisian daya. Model *Naïve Bayes* dengan *CountVectorizer* terbukti efektif, dengan akurasi mencapai 78%, sehingga dapat digunakan sebagai alat analisis sentimen yang andal.

The development of electric cars in Indonesia shows significant improvement alongside government efforts to reduce carbon emissions and promote renewable energy use. Public perception and opinions of electric cars are important factors influencing interest in owning these environmentally friendly vehicles. This study aims to analyze public sentiment toward electric cars in Indonesia using review data and comments from the social media platform X (Twitter). Data was collected through web scraping techniques and stored in CSV format for further processing using Google Colab. Preprocessing was performed to clean the text data, including removal of punctuation, numbers, symbols, Indonesian stopwords, tokenization, and stemming. The text was then converted into numerical representation using TF-IDF vectorization so it could be processed by the Naïve Bayes classification model. The model was then trained and tested to classify sentiment into three categories: positive, negative, and neutral. The results show that the majority of public comments about electric cars were negative, with a percentage reaching approximately 66.5%, while positive comments accounted for 33.5%. Further analysis revealed that negative sentiment was largely related to the relatively high vehicle prices and limited charging infrastructure. The Naïve Bayes model with CountVectorizer proved effective, achieving 78% accuracy, thus making it a reliable tool for sentiment analysis.

1. Introduction

Perkembangan teknologi kendaraan listrik di Indonesia terus mengalami peningkatan seiring dengan upaya pemerintah dalam mengurangi emisi karbon dan ketergantungan pada bahan bakar fosil. Kendaraan listrik

dianggap sebagai salah satu solusi untuk mendukung mobilitas yang ramah lingkungan dan mengurangi dampak negatif terhadap kualitas udara[1]. Di sisi lain, penerimaan masyarakat terhadap mobil listrik masih menghadapi sejumlah tantangan signifikan. Beberapa faktor yang memengaruhi persepsi masyarakat antara lain harga kendaraan yang relatif tinggi, keterbatasan infrastruktur pengisian daya, serta kurangnya pemahaman tentang manfaat jangka panjang penggunaan kendaraan listrik[2]. Kondisi ini menyebabkan munculnya opini yang beragam di masyarakat, mulai dari antusiasme terhadap inovasi teknologi hingga kekhawatiran dan ketidakpuasan terkait biaya, efisiensi, dan kemudahan penggunaan.

Komentar masyarakat yang tersebar di media sosial dan platform daring memberikan gambaran awal mengenai persepsi publik terhadap mobil listrik. Namun, komentar tersebut bersifat tidak terstruktur dan beragam, sehingga sulit dianalisis secara manual. Oleh karena itu, diperlukan metode yang mampu melakukan analisis sentimen secara sistematis dan kuantitatif, sehingga dapat memetakan opini masyarakat menjadi kategori yang jelas, seperti positif, negatif, atau netral. Analisis sentimen berbasis teks memungkinkan peneliti untuk mengidentifikasi proporsi opini positif dan negatif, mengetahui kata atau tema yang dominan dalam komentar, serta menilai faktor-faktor yang paling memengaruhi persepsi masyarakat terhadap mobil listrik[3].

Berdasarkan latar belakang tersebut, penelitian ini difokuskan pada beberapa permasalahan utama, yaitu bagaimana persepsi masyarakat Indonesia terhadap mobil listrik dapat diidentifikasi melalui komentar di media sosial X (Twitter), bagaimana distribusi sentimen positif dan negatif dalam komentar tersebut, dan sejauh mana algoritma *Naïve Bayes* efektif dalam mengklasifikasikan komentar masyarakat[4]. Dengan adanya penelitian ini, diharapkan dapat diketahui tingkat dominasi sentimen negatif maupun positif, serta faktor-faktor yang paling sering dibahas oleh masyarakat terkait mobil listrik.

Tujuan dari penelitian ini adalah untuk menganalisis sentimen masyarakat terhadap mobil listrik di Indonesia, mengetahui proporsi komentar positif dan negatif, serta menguji efektivitas algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen[5][6]. Penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai persepsi publik, sehingga produsen, pemerintah, dan pemangku kepentingan lainnya dapat membuat keputusan berbasis data dalam pengembangan produk, penyusunan strategi pemasaran, dan implementasi kebijakan publik yang mendukung penerimaan masyarakat terhadap kendaraan listrik[4].

Manfaat penelitian ini bersifat praktis dan akademis. Bagi produsen mobil listrik, hasil analisis sentimen dapat menjadi masukan untuk merancang strategi pemasaran, inovasi produk, dan peningkatan kualitas layanan yang sesuai dengan kebutuhan masyarakat. Bagi pemerintah dan pemangku kebijakan, penelitian ini dapat menjadi referensi dalam merancang kebijakan insentif, pembangunan infrastruktur pengisian daya, serta program edukasi untuk meningkatkan kesadaran dan penerimaan masyarakat terhadap mobil listrik. Selain itu, dari sisi ilmiah, penelitian ini memberikan kontribusi terhadap pengembangan metode analisis sentimen, khususnya untuk studi persepsi publik terhadap teknologi baru, sehingga dapat menjadi dasar bagi penelitian lanjutan di bidang analisis opini masyarakat dan pengambilan keputusan berbasis data.

Dengan pendekatan ini, penelitian tidak hanya menyajikan data kuantitatif terkait persepsi masyarakat, tetapi juga memberikan insight mengenai masalah dan peluang yang muncul dalam adopsi kendaraan listrik di Indonesia. Hal ini memungkinkan berbagai pihak untuk memahami opini publik secara lebih mendalam, melakukan evaluasi terhadap strategi yang sudah diterapkan, serta menyusun langkah-langkah yang lebih tepat sasaran untuk mendorong transformasi menuju transportasi yang lebih ramah lingkungan.

Dengan menganalisis sentimen masyarakat, penelitian ini diharapkan dapat memberikan informasi yang akurat mengenai opini publik terhadap mobil listrik, sehingga dapat menjadi acuan bagi produsen, pemerintah, dan pemangku kepentingan dalam merumuskan strategi pemasaran, kebijakan, dan program edukasi yang tepat.

2. Research Methods

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis data teks untuk memahami sentimen masyarakat terhadap mobil listrik di Indonesia. Metodologi penelitian mencakup beberapa tahap, mulai dari pengumpulan data hingga evaluasi model klasifikasi.

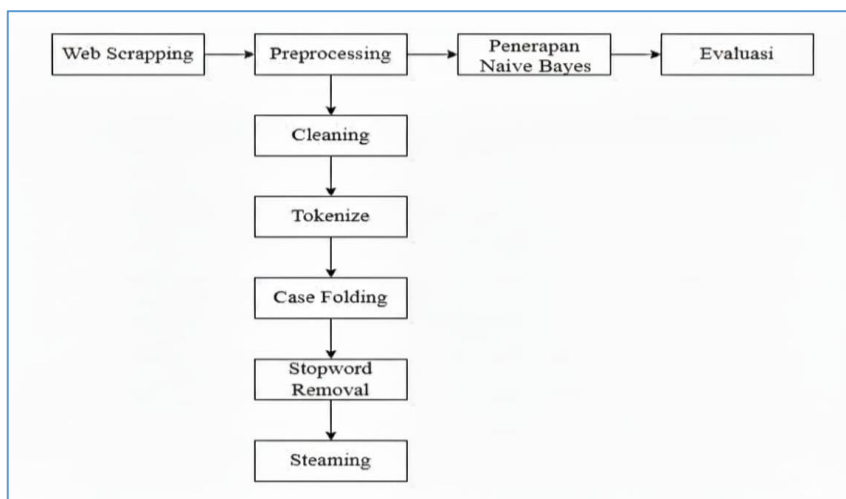
2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari media sosial, seperti Twitter, Instagram, atau forum publik, dengan topik seputar mobil listrik. Data dikumpulkan menggunakan teknik web scraping atau API platform yang bersangkutan, dan disimpan dalam format CSV.

Variabel data: komentar, ulasan, atau tweet masyarakat mengenai mobil listrik.

Kriteria data: komentar dalam bahasa Indonesia dan relevan dengan topik mobil listrik.

Jumlah sampel: sebanyak 1517 komentar, yang kemudian akan dianalisis untuk sentimen positif, negatif, atau netral.



Gambar 1. Alur Penelitian

2. 2. Preprocessing Data

Setelah data diperoleh, tahap *Preprocessing* data dilakukan untuk memastikan bahwa teks yang digunakan dalam penelitian berada dalam format yang bersih dan siap untuk diproses lebih lanjut. Berikut ini adalah yang dilakukan pada tahap *Preprocessing*.

2. 2. 1. Cleaning

Proses ini melakukan serangkaian langkah pembersihan data dengan menghapus seluruh data yang terdeteksi duplikat (penilaian user yang dilakukan dua kali). Tahap ini dimaksudkan untuk menghilangkan komponen analisis sentimen yang tidak penting.

2. 2. 2. Tokenize

Pada titik ini, proses pemisahan teks menjadi bagian-bagian kecil yang dikenal sebagai token dilakukan. Proses ini juga menghapus tanda baca yang tidak diperlukan misalnya “penyajian makanan cepat dan sangat bersih” akan di pisah menjadi token seperti [“ penyajian ”, “makanan”, “cepat”, “dan”, “sangat”, “bagus”].

2. 2. 3. Case Folding

Pada tahapan ini transformasi huruf dilakukan dengan mengubah seluruh karakter menjadi huruf kecil. Tujuannya agar tidak ada perbedaan pengolahan antara kata yang di tulis dengan huruf besar maupun kecil. Contoh: “Tong Tji Tea House harganya sangat terjangkau” akan diubah menjadi “tong tji tea house harganya sangat terjangkau”

2. 2. 4. Stop word Removal

Proses ini dikenal sebagai penghilangan stop words, dimana kata-kata yang dianggap kurang bermakna dalam konteks analisis akan dihilangkan untuk meningkatkan kualitas dan akurasi hasil analisis sebagainya. Contoh: “Tong Tji Tea house adalah salah satu restoran yang bersih di Amplas” setelah stop word beberapa kata akan dihapus menjadi “tong tji tea house restoran bersih di amplas”.

2. 2. 5. Steaming

Tahapan terakhir pada *Preprocessing* data adalah steaming yang berfungsi untuk mengembalikan kata yang berimbuhan ke bentuk dasarnya. Langkah ini sangat penting karena dalam tahapan pemrosesan teks dapat menyederhanakan variasi kata dan meningkatkan kinerja algoritma klasifikasi. Contoh: "Makanan di restoran ini sangat lezat dan pelayanannya memuaskan, saya sudah merekomendasikanya ke teman-teman". Akan diubah menjadi “makan”, “di”, “restoran”, “ini”, “sangat”, “enak”, “dan”, “layanan”, “puas”, “saya”, “sudah”, “rekomendasi”, “kepada”, “teman”.

2. 3. Representasi Data

Mengubah data teks dan numerik menjadi format yang dapat diproses oleh Naïve Bayes. Menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah kata menjadi vektor numerik. TF-IDF memberi bobot lebih tinggi pada kata-kata yang penting dan jarang muncul, sehingga model lebih peka terhadap kata-kata kunci dalam komentar mahasiswa.

Secara matematis, TF-IDF dihitung dengan persamaan (1) berikut:

$$TF - IDF (t, d) = TF (t, d) \times IDF(t)$$

Di mana:

TF (Term Frequency) mengukur seberapa sering kata t muncul dalam dokumen d. Dihitung dengan persamaan (2) berikut:

$$TF (t, d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{total kata dalam dokumen } d}$$

IDF (Inverse Document Frequency) mengukur seberapa unik suatu kata dalam keseluruhan korpus dokumen. Dihitung dengan persamaan (3) berikut:

$$IDF(t) = \log \frac{N}{DF(t)} + 1$$

Di mana N adalah jumlah total dokumen dalam korpus, dan DF(t) adalah jumlah dokumen yang mengandung kata t. Semakin jarang sebuah kata muncul di seluruh dokumen, semakin besar nilai IDF-nya, sehingga kata tersebut dianggap lebih bermakna.

2. 4. Pembagian Data

Metode Support Vector Machine digunakan untuk mengklasifikasikan teks ke dalam tiga kategori sentimen: positif, netral, dan negatif. SVM bekerja dengan cara mencari hyperplane terbaik yang memisahkan kelas-kelas data dalam ruang vektor berdimensi tinggi.

Rumus dasar fungsi keputusan SVM adalah:

$$f(x) = \text{sign}(w \cdot x + b)$$

di mana:

- x adalah vektor fitur hasil TF-IDF,
- w adalah bobot dari masing-masing fitur,
- b adalah bias,
- dan $\text{sign}()$ menentukan kelas hasil prediksi.

2. 5. Training Data

1. Akurasi (Accuracy)

Mengukur proporsi prediksi yang benar terhadap total data uji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi (Precision)

Mengukur tingkat ketepatan model dalam memprediksi kelas tertentu.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall (Sensitivity)

Mengukur kemampuan model dalam mendeteksi kelas yang benar.

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score

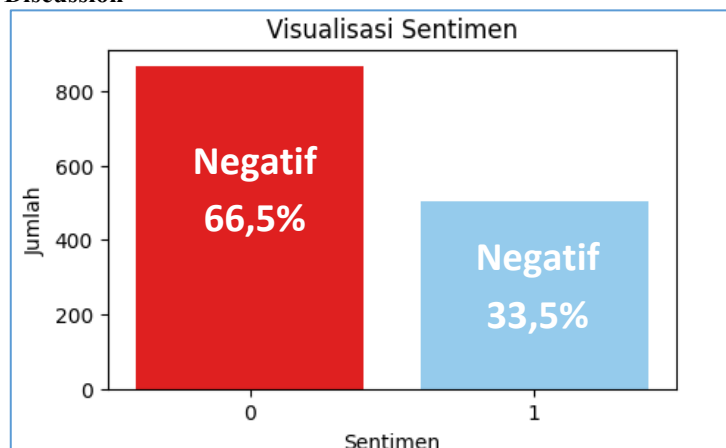
Menggabungkan presisi dan recall menjadi satu ukuran performa.

$$F1 = 2X = \frac{Precision \times Recall}{Precision + Recall}$$

2. 6. Visualisasi

Visualisasi pada penelitian ini menggunakan *word cloud*. *Word cloud* menampilkan kata-kata yang menjelaskan setiap kelompok klasifikasi.

3. Results and Discussion



Gambar 2 Visualisasi Sentimen

Hasil analisis sentimen terhadap komentar masyarakat mengenai mobil listrik menunjukkan bahwa persepsi publik di Indonesia masih cenderung didominasi oleh pandangan negatif. Dari keseluruhan data yang dianalisis, sebesar 66,5% komentar dikategorikan sebagai sentimen negatif, sedangkan 33,5% sisanya merupakan sentimen positif. Proporsi dominan pada kategori negatif menggambarkan adanya ketidakpuasan dan keraguan yang signifikan terhadap adopsi kendaraan listrik dalam kehidupan sehari-hari. Sentimen negatif tersebut umumnya berkaitan dengan isu-isu utama seperti tingginya harga kendaraan listrik, terbatasnya infrastruktur pengisian daya, serta ketidakpastian mengenai ketahanan dan biaya pemeliharaan baterai. Temuan ini sejalan dengan berbagai laporan dan survei nasional yang menunjukkan bahwa faktor ekonomi dan ketersediaan fasilitas pendukung masih menjadi hambatan utama dalam implementasi kendaraan listrik di Indonesia.

Sebaliknya, kelompok sentimen positif mencerminkan adanya bagian masyarakat yang telah menunjukkan penerimaan terhadap inovasi mobil listrik. Komentar positif banyak menyoroti keunggulan seperti efisiensi energi, teknologi yang lebih modern, serta kontribusinya terhadap pengurangan emisi karbon dan pencemaran udara. Walaupun persentasenya lebih rendah, keberadaan sentimen positif ini menunjukkan adanya peluang penguatan penerimaan publik apabila hambatan-hambatan utama dapat ditangani melalui kebijakan dan strategi yang tepat.

Akurasi Model Naive Bayes					
0.7815533980582524					
Classification Report:					
	precision	recall	f1-score	support	
negatif	0.81	0.88	0.84	137	
positif	0.71	0.58	0.64	69	
accuracy			0.78	206	
macro avg	0.76	0.73	0.74	206	
weighted avg	0.78	0.78	0.78	206	

Dari sisi performa analisis, model *Multinomial Naïve Bayes* yang diterapkan dalam penelitian ini menunjukkan tingkat akurasi sebesar 78% pada pengujian data. Nilai akurasi tersebut mengindikasikan bahwa model mampu melakukan klasifikasi sentimen dengan cukup baik, meskipun masih terdapat ruang perbaikan terutama pada pengenalan sentimen positif yang jumlah datanya lebih sedikit. Konsistensi hasil evaluasi melalui precision, recall, dan f1-score menunjukkan bahwa model bekerja secara stabil dalam membedakan polaritas komentar. Dengan demikian, algoritma ini dapat dianggap layak digunakan untuk analisis sentimen berbahasa Indonesia, khususnya pada penelitian yang berkaitan dengan persepsi masyarakat terhadap isu-isu teknologi dan kebijakan publik.



Gambar 3 Visualisasi data kata terbanyak negatif

Pada *WordCloud* sentimen negatif, kata-kata yang paling sering muncul dalam komentar masyarakat terhadap mobil listrik divisualisasikan menggunakan peta warna merah (*colormap='Reds'*). Visualisasi ini menunjukkan bahwa kata-kata dominan berkaitan dengan berbagai kendala dan kekhawatiran masyarakat, antara lain “kendaraan”, “harga”, dan lainnya. Kemunculan kata-kata tersebut menegaskan bahwa isu utama yang memengaruhi persepsi negatif masyarakat terkait dengan harga kendaraan yang relatif tinggi, keterbatasan infrastruktur pengisian daya, serta ketidakpastian mengenai daya tahan dan biaya perawatan baterai. *WordCloud*

