

Penerapan Algoritma *Naïve Bayes* dalam Analisis Sentimen terhadap Mobil Listrik

Application of Naïve Bayes Algorithm in Sentiment Analysis of Electric Cars

Muhammad Febri Yulianto^{1*}, Fida Maisa Hana², Agung Prihandono³

^{1,2,3}*Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi,
Universitas Muhammadiyah Kudus*

*corr-author: 32021110002@std.umku.ac.id

ABSTRAK

Dalam era kemajuan teknologi yang pesat, mobil listrik menjadi inovasi utama dalam industri otomotif. Namun, analisis sentimen menunjukkan adanya tantangan, terutama kurangnya pemahaman masyarakat tentang mobil listrik, yang berdampak pada persepsi negatif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap mobil listrik di Indonesia melalui konten YouTube, dengan menggunakan algoritma Naive Bayes. Metode Knowledge Discovery in Databases (KDD) digunakan untuk mengumpulkan dan menganalisis opini masyarakat. Evaluasi model menggunakan Confusion Matrix menunjukkan peningkatan signifikan setelah penerapan teknik SMOTE Upsampling, meningkatkan akurasi dari 50,70% menjadi 78,16%. Analisis kelas menunjukkan peningkatan pada true positive, true negative, dan true neutral. Meskipun algoritma Naive Bayes menghasilkan akurasi 78,16%, presisi 71,43%, dan recall 57,97%, penelitian ini memiliki keterbatasan dan disarankan untuk membandingkan algoritma klasifikasi lain, memperluas dataset, serta menguji kombinasi algoritma yang lebih luas. Rekomendasi juga mencakup penerapan hasil penelitian ke dalam mesin klasifikasi dengan pengujian lebih lanjut. Studi ini memberikan wawasan untuk meningkatkan pemahaman dan penerimaan mobil listrik di masyarakat Indonesia.

Kata-kata kunci: Analisis sentimen; *Naive Bayes*; mobil listrik; Youtube

ABSTRACT

In an era of rapid technological progress, electric cars have become a major innovation in the automotive industry. However, sentiment analysis shows that there are challenges, especially the public's lack of understanding about electric cars, which has an impact on negative perceptions. This research aims to analyze public sentiment towards electric cars in Indonesia through YouTube content, using the Naive Bayes algorithm. The Knowledge Discovery in Databases (KDD) method is used to collect and analyze public opinion. Model evaluation using the Confusion Matrix shows a significant improvement after applying the SMOTE Upsampling technique, increasing accuracy from 50.70% to 78.16%. Class analysis shows an increase in true positive, true negative, and true neutral. Although the Naive Bayes algorithm produces 78.16% accuracy, 71.43% precision, and 57.97% recall, this research has limitations and it is recommended to compare other classification engine with further testing. This study provides insights to increase understanding and acceptance of electric cars in Indonesian society.

Keywords: *Sentiment analysis; Naive Bayes; electric cars; YouTube*

PENDAHULUAN

Di era modern yang sarat dengan perkembangan teknologi, dampaknya terasa jelas di berbagai bidang kehidupan, termasuk industri otomotif. Salah satu inovasi terbaru yang menonjol adalah mobil listrik, kendaraan yang mengandalkan motor listrik sebagai penggerak utama dan baterai sebagai sumber dayanya. Mobil listrik memiliki keunggulan berupa efisiensi energi dan dampak positif terhadap lingkungan. Namun, adopsi mobil listrik di Indonesia masih terbatas. Oleh karena itu, diperlukan analisis sentimen untuk memahami pandangan masyarakat terhadap mobil listrik, sebagai upaya untuk mendukung pengembangan dan peningkatan penggunaannya di Indonesia (Karimah et al., 2024).

Tantangan yang muncul dalam analisis sentimen terhadap mobil listrik meliputi kurangnya informasi dan pemahaman masyarakat, yang menyebabkan pandangan terhadap teknologi ini kurang positif. Selain itu, keterbatasan infrastruktur dan harga yang tinggi juga menjadi penghalang. Analisis sentimen di media sosial, terutama YouTube, dinilai sebagai pendekatan yang tepat untuk menilai pandangan masyarakat terhadap mobil listrik (Karimah et al., 2024). YouTube kini menjadi salah satu platform paling populer untuk menonton video. Oleh karena itu, memahami apakah pandangan pengguna internet terhadap YouTube umumnya positif, negatif, atau netral menjadi cukup sulit (Karimah et al., 2024).

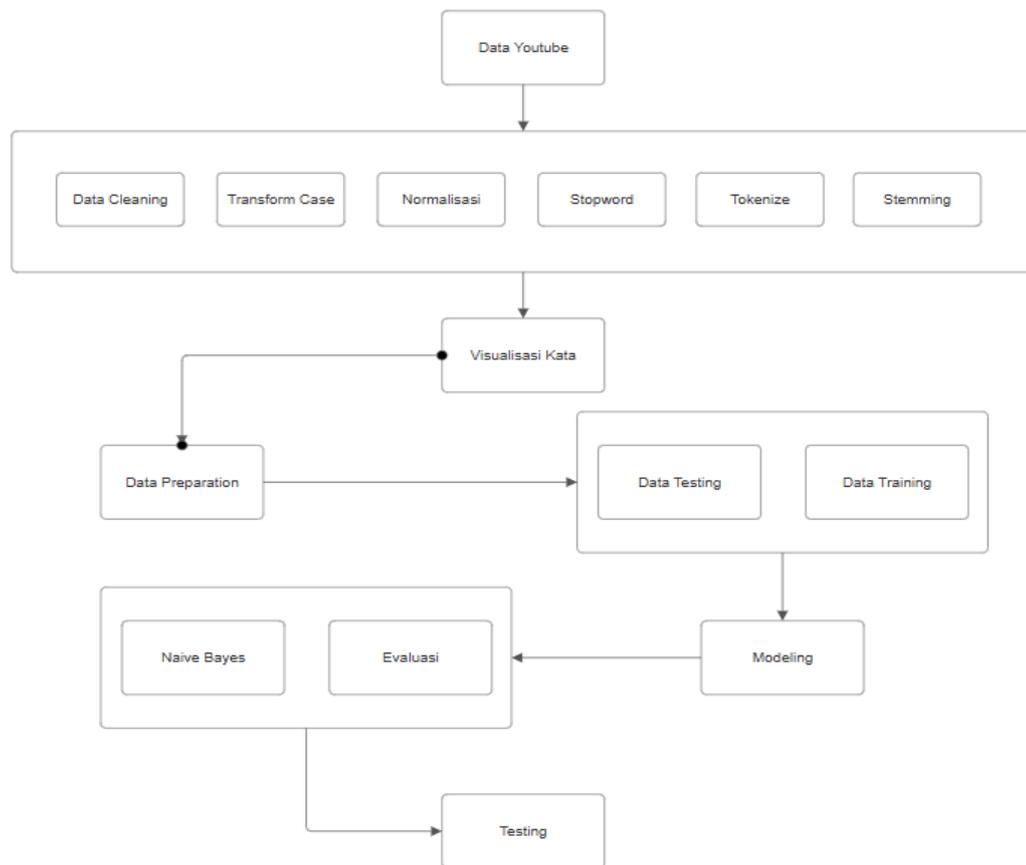
Kendaraan listrik juga menawarkan kinerja yang memadai, sehingga mampu bersaing dengan kendaraan konvensional dalam hal kecepatan, jarak tempuh, dan kapasitas angkut. Meskipun saat ini kendaraan listrik belum menjadi pilihan utama bagi mayoritas konsumen di dunia, banyak produsen otomotif terkemuka telah berinvestasi dalam teknologi ini dan memperkenalkan model-model yang semakin canggih dan terjangkau. Dengan kemajuan teknologi yang terus berkembang dan permintaan yang meningkat, kendaraan listrik diharapkan menjadi alternatif yang lebih populer dan berkelanjutan bagi kendaraan konvensional di masa depan (Alfarizi & Fitriani, 2023).

Naive Bayes, yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes, memprediksi probabilitas di masa depan dengan mendasarkan perhitungannya pada data atau pengalaman masa lalu (Karimah et al., 2024). Naive Bayes adalah salah satu metode yang dapat digunakan untuk analisis sentimen, dan secara teori, metode ini menunjukkan konsistensi data serta akurasi dalam perhitungan klasifikasi. Naive Bayes umumnya diterapkan pada teknik klasifikasi, terutama dalam konteks platform seperti Youtube. Keunggulan utama klasifikasi Naive Bayes adalah kemampuannya menghasilkan hipotesis yang kuat terkait suatu kondisi atau peristiwa. Perhitungan probabilitas kelompok dalam metode ini didasarkan pada pendekatan algoritma Bayesian yang menggunakan persamaan matematis sebagai dasar (Caswadi et al., 2023). Metode klasifikasi Naive Bayes digunakan untuk membuat keputusan dengan memprediksi suatu kasus berdasarkan hasil klasifikasi yang telah diperoleh (Ningsih et al., 2024).

Dengan demikian, penelitian ini bertujuan untuk mengidentifikasi dan menganalisis sentimen masyarakat terhadap mobil listrik melalui konten yang tersedia di platform YouTube. Melalui pendekatan ini, diharapkan dapat terungkap berbagai perspektif dan opini yang berkembang di masyarakat mengenai mobil listrik. Analisis ini juga akan memberikan gambaran tentang bagaimana teknologi ini dipersepsikan, serta faktor-faktor yang memengaruhi pandangan tersebut. Dengan menggunakan algoritma Naive Bayes, penelitian ini berupaya untuk mengklasifikasikan sentimen secara akurat, sehingga dapat memberikan dasar yang kuat untuk memahami dinamika penerimaan mobil listrik di Indonesia.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menilai tingkat akurasi tertinggi serta mengevaluasi sentimen pengguna Youtube, apakah bersifat positif atau negatif, yang dapat mempengaruhi persepsi masyarakat terhadap Mobil Listrik. Bab ini menjelaskan metode yang digunakan dalam penelitian untuk mencapai tujuan yang telah dirumuskan. Metodologi penelitian dirancang sebagai panduan sistematis dalam mengumpulkan, mengolah, dan menganalisis data terkait mobil listrik. Pendekatan ini diharapkan dapat memberikan jawaban yang valid dan terpercaya terhadap rumusan masalah penelitian. Dalam penelitian ini menggunakan algoritma *Naive Bayes*, dimana dilakukan beberapa langkah seperti yang dijelaskan pada Gambar 1.



Gambar 1. Tahapan penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode analisis deskriptif. Pendekatan kuantitatif dipilih karena mampu mengolah data numerik untuk memberikan gambaran objektif tentang tren, pola, serta hubungan antar variabel dalam data terkait mobil listrik. Algoritma *Naive Bayes* dipilih karena dianggap salah satu yang terbaik dalam data mining dan sangat efektif untuk proses klasifikasi data.

Evaluasi model menggunakan *Confusion Matrix* menunjukkan peningkatan signifikan setelah penerapan teknik *SMOTE Upsampling*, meningkatkan akurasi dari 50,70% menjadi 78,16%. Analisis kelas menunjukkan peningkatan pada *true positive*, *true negative*, dan *true neutral*. Meskipun algoritma *Naive Bayes* menghasilkan akurasi 78,16%, presisi 71,43%, dan recall 57,97%, penelitian ini memiliki keterbatasan dan disarankan untuk membandingkan algoritma klasifikasi lain, memperluas dataset, serta

menguji kombinasi algoritma yang lebih luas. Rekomendasi juga mencakup penerapan hasil penelitian ke dalam mesin klasifikasi dengan pengujian lebih lanjut. Studi ini memberikan wawasan untuk meningkatkan pemahaman dan penerimaan mobil listrik di masyarakat Indonesia.

HASIL DAN PEMBAHASAN

Data dikumpulkan dengan mengimpor file "mobil_listrik.csv" ke dalam perangkat lunak pengolahan data, seperti Python atau Microsoft Excel (Gambar 2). Selanjutnya, data diolah untuk memverifikasi kelengkapan, konsistensi, dan akurasi informasi yang digunakan dalam analisis. Setelah data berhasil dikumpulkan, dilakukan proses pengolahan atau *preprocessing*. Tahap ini mencakup aktivitas membangun data dan diikuti dengan proses pembersihan data agar data tersebut siap untuk dikelola ke tahap berikutnya. Adapun langkah-langkah dalam *preprocessing* data meliputi:

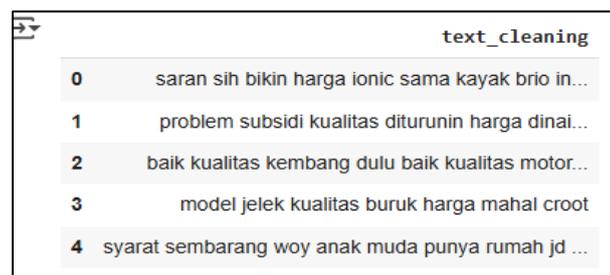


	id_komentar	nama_akun	tanggal	text_cleaning	sentimen
0	UgzblI5eyrly3-gdUUJ4AaABAg	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	UgzEDUIV3OTrV943p8p4AaABAg	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	UgwwJqu6JMF4EH2CsVV4AaABAg	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	UgyYicCMR1rKwuOj2Y14AaABAg	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
4	UgxKAcLuAwZQOK6es-x4AaABAg	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya ruma...	negatif

Gambar 2. Contoh dataset analisis sentimen mobil listrik

1. Data Cleaning

Data Cleaning adalah langkah penting dalam analisis sentimen mobil listrik untuk memastikan data yang digunakan bersih, relevan, dan siap dianalisis. Proses ini mencakup penghapusan karakter khusus, duplikasi, dan kata-kata tidak bermakna (*stopwords*) yang dapat mengganggu hasil analisis (Gambar 3). Selain itu, normalisasi teks seperti mengubah kata-kata tidak baku menjadi bentuk standar juga dilakukan agar model analisis sentimen dapat mengenali makna dengan lebih akurat. Dengan data yang telah dibersihkan, analisis sentimen dapat memberikan wawasan lebih mendalam tentang bagaimana masyarakat menanggapi mobil listrik, termasuk aspek harga, kualitas, subsidi, dan inovasi teknologi. Hasil dari analisis ini dapat membantu produsen dan pemerintah dalam merancang strategi yang lebih baik untuk pengembangan dan adopsi kendaraan listrik di masa depan.



	text_cleaning
0	saran sih bikin harga ionic sama kayak brio in...
1	problem subsidi kualitas diturunin harga dinai...
2	baik kualitas kembang dulu baik kualitas motor...
3	model jelek kualitas buruk harga mahal croot
4	syarat sembarang woy anak muda punya rumah jd ...

Gambar 3. Hasil proses data cleaning

2. Transform Cases

Pada tahap ini dilakukan konversi huruf kapital dalam dataset menjadi huruf kecil. Langkah ini bertujuan untuk menyamakan format teks pada model klasifikasi sehingga dapat mencegah kesalahan selama proses *tokenize*.

3. Normalisasi

Normalisasi merupakan proses membersihkan dan menyederhanakan teks sebelum dianalisis guna meningkatkan akurasi hasil. Dalam opini masyarakat mengenai mobil listrik, sering ditemukan variasi kata, slang, ejaan tidak baku, atau bahkan emotikon yang dapat mengganggu analisis sentimen. Misalnya, kata “mahal bgt” dapat dinormalisasi menjadi “mahal banget” agar lebih mudah dipahami oleh model analisis. Dengan normalisasi, data menjadi lebih seragam, memungkinkan model *machine learning* atau analisis berbasis leksikon untuk menginterpretasikan sentimen secara lebih akurat (Gambar 4). Hal ini penting untuk memahami persepsi publik terhadap mobil listrik, baik dalam aspek harga, kualitas, efisiensi, maupun kebijakan terkait.

```
[ ] # normalisasi
norm = {'dgn' : 'dengan', 'gue' : 'saya', 'ngaco' : 'sembarangan', 'kwalitas' : 'kualitas', 'blom' : 'belum', 'aing' : 'aku', 'tehnolog' : 'teknologi'}

def normalisasi(str_text):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text

data['text_cleaning'] = data['text_cleaning'].apply(lambda x: normalisasi(x))
```

Gambar 4. Hasil normalisasi

4. Tokenizing

Proses ini dimaksudkan untuk memecah teks ulasan atau komentar menjadi unit kata atau frasa yang lebih kecil agar lebih mudah dianalisis. Dalam konteks mobil listrik, tokenizing membantu dalam memahami opini pengguna terhadap berbagai aspek, seperti performa baterai, harga, fitur teknologi, dan kebijakan pemerintah (Gambar 5). Dengan tokenizing, kata-kata yang memiliki makna serupa dapat diidentifikasi, sehingga model analisis sentimen dapat mengklasifikasikan apakah sebuah ulasan bersifat positif, negatif, atau netral.

```
[ ] tokenized = data['text_cleaning'].apply(lambda x:x.split())
tokenized
```

	text_cleaning
0	[saran, sih, bikin, harga, ionic, sama, kayak, ...]
1	[problem, subsidi, kualitas, diturunin, harga, ...]
2	[baik, kualitas, kembang, dulu, baik, kualitas, ...]
3	[model, jelek, kualitas, buruk, harga, mahal, ...]
4	[syarat, sembarangan, woy, anak, muda, punya, ...]
...	...
1512	[apa, kabar, padahal, negri, luar, biasa, negara]
1513	[antar, anak, sekolah, antar, bantu, pasar, kl...]
1514	[esemka, bangga, solo]
1515	[cerdas, orang, hny, dn, pasar, jln, x, da, gi...]
1516	[niat, beli, ev, murah, malah, ikut, dinaikin, ...]

1514 rows x 1 columns
dtype: object

Gambar 5. Hasil tokenizing

5. Filtering

Proses penting untuk memastikan data yang digunakan bersih dan relevan. Dalam analisis sentimen, data yang dikumpulkan dari berbagai sumber seperti media sosial, ulasan pelanggan, dan forum diskusi sering kali mengandung noise, seperti kata-kata tidak relevan, duplikasi, atau bahkan informasi yang bias. Oleh karena itu, proses filtering

dilakukan dengan beberapa teknik, seperti menghilangkan stopwords, normalisasi teks, serta penghapusan karakter khusus atau emoji yang tidak berkontribusi terhadap pemahaman sentimen.

6. Stopword

Langkah ini dilakukan untuk menghapus kata-kata yang tidak relevan dalam analisis sentimen, seperti kata sambung dan kata keterangan (Gambar 6).

```
[ ] # stopwords
import Sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary
more_stop_word = []

stop_words = StopWordRemoverFactory().get_stop_words()
new_array = ArrayDictionary(stop_words)
stop_words_remover_new = StopWordRemover(new_array)

def stopword(str_text):
    str_text = stop_words_remover_new.remove(str_text)
    return str_text

data['text_cleaning'] = data['text_cleaning'].apply(lambda x: stopword(x))
data.head()
```

	nama_akun	tanggal	text_cleaning	sentimen
0	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio ...	positif
1	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinai...	negatif
2	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor...	positif
3	yp office	2023-08-04 08:29:54+00:00	model jelek kualitas buruk harga mahal croot	negatif
4	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat sembarangan woy anak muda punya rum...	negatif

Gambar 6. Hasil *stopword*

7. Stemming

Teknik untuk mengubah kata-kata menjadi bentuk dasarnya, sehingga kata-kata dengan makna serupa dapat dianalisis secara lebih efektif. Dalam konteks mobil listrik, opini pengguna sering kali mengandung variasi kata seperti "mahal," "kemahalan," atau "terlalu anmahal," (Gambar 7).

```
[ ] # stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(text_cleaning):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in text_cleaning:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

tokenized.to_csv('mobil_listrik_bersih1.csv', index=False)
data_clean = pd.read_csv('mobil_listrik_bersih1.csv', encoding='latin1')
```

Gambar 7. Hasil *stemming*

KESIMPULAN

Berdasarkan pengujian analisis sentimen dengan menggunakan algoritma Naive Bayes. Hasil Evaluasi model menggunakan Confusion Matrix menunjukkan peningkatan signifikan setelah penerapan teknik SMOTE Upsampling, meningkatkan akurasi dari 50,70% menjadi 78,16%. Analisis kelas menunjukkan peningkatan pada true positive, true negative, dan true neutral. Meskipun algoritma Naive Bayes menghasilkan akurasi 78,16%, presisi 71,43%, dan recall 57,97%, penelitian ini memiliki keterbatasan dan disarankan untuk membandingkan algoritma klasifikasi lain, memperluas dataset, serta menguji kombinasi algoritma yang lebih luas. Rekomendasi juga mencakup penerapan hasil penelitian ke dalam mesin klasifikasi dengan pengujian lebih lanjut. Studi ini memberikan wawasan untuk meningkatkan pemahaman dan penerimaan mobil listrik di masyarakat Indonesia.

DAFTAR PUSTAKA

- Alfarizi, S., & Fitriani, E. (2023). Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization. *Indonesian Journal on Software Engineering (IJSE)*, 9(1), 19–27. <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse>
- Caswadi, Dienwati, N., Dwilestari, G., Fathurrohman, & Tohidi, E. (2023). Penerapan Algoritma Naive Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Youtube Mengenai Intensif Mobil Listrik. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3851–3857.
- Fahrezy, D. C., & Utomo, F. S. (2024). Analisis Sentimen Opini Publik Tentang Kendaraan Listrik di Indonesia. 1(2), 127–139.
- Hadiwinata, L. N., Murtiningsih, B. S. E., & Berto, A. R. (2023). Analisis Teks Dan Jaringan Promosi Media Sosial Youtube Mobil Listrik Ioniq 5 Menggunakan Metode Sna. *Perspektif Komunikasi: Jurnal Ilmu Komunikasi Politik Dan Komunikasi Bisnis*, 7(1), 1. <https://doi.org/10.24853/pk.7.1.1-18>
- Huzna, A. N., Nurhayati, I., Saputri, A. E., Qomarul Huda, M., Informasi, S., Sains, F., Teknologi, D., Syarif, U., & Jakarta, H. (2024). Analisis Sentimen Terhadap Mobil Listrik Di Indonesia Pada Twitter: Penerapan Naive Bayes Classifier Untuk Memahami Opini Publik. *Just IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 14(2), 80–149. <https://jurnal.umj.ac.id/index.php/just-it/index>
- Karimah, A., Dwilestari, G., & Mulyawan. (2024). Analisis Sentimen Komentar Video Mobil Listrik di Platform Youtube dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–773.
- Ningsih, W., Alfianda, B., Rahmadden, R., & Wulandari, D. (2024). Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 556–562. <https://doi.org/10.57152/malcom.v4i2.1253>
- Rahmat Bachtiar, D., Rosyid, H. A., & Nabila, K. (2022). Pengaruh Perkembangan Kendaraan Listrik Terhadap Industri Otomotif Pada Era Society 5.0. *Jurnal Inovasi Teknik Dan Edukasi Teknologi*, 2(6), 277–281. <https://doi.org/10.17977/um068v1i62022p277-281>