INNOVATIVE: Journal Of Social Science Research
Volume 4 Nomor 4 Tahun 2024 Page 837-845
E-ISSN 2807-4238 and P-ISSN 2807-4246
Website: https://j-innovative.org/index.php/Innovative

Sentiment Analysis Of Electric Vehicles On Twitter Using Machine Learning

Gusti Ayu Sri Adinda Prameswari[1✉], Muhammad Rakib[2], Valentino Aris[3]
Bisnis Digital, Universitas Negeri Makassar
Email: adindaprmswrii@gmail.com[1✉]

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap kendaraan listrik menggunakan teknik pembelajaran mesin. Data dikumpulkan dari Twitter dengan menggunakan kata kunci terkait kendaraan listrik. Data yang diperoleh diolah melalui beberapa tahap seperti pembersihan data, tokenisasi, dan stemming. Selanjutnya, data dianalisis menggunakan tiga algoritma pembelajaran mesin: Support Vector Machine (SVM), Naive Bayes Classifier (NBC), dan Random Forest. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan akurasi terbaik dalam mengklasifikasikan sentimen pengguna. Berdasarkan analisis diagram tulang ikan, ditemukan bahwa sentimen negatif pengguna terutama disebabkan oleh delapan faktor utama: baterai, infrastruktur pengisian daya, biaya awal, jangkauan, dukungan kebijakan, kinerja, variasi model, dan keandalan. Temuan ini diharapkan dapat memberikan wawasan bagi produsen kendaraan listrik untuk meningkatkan kualitas produk dan kepuasan pelanggan.

Kata Kunci: *Analisis Sentimen, Kendaraan Listrik, Machine Learning, Support Vector Machine, Naive Bayes, Random Forest*

Abstract

This research aims to analyze Twitter user sentiment towards electric vehicles using machine learning techniques. Data were collected from Twitter using keywords related to electric vehicles. The obtained data were processed through several stages such as data cleaning, tokenization, and stemming. Subsequently, the data were analyzed using three machine learning algorithms: Support Vector Machine (SVM), Naive Bayes Classifier (NBC), and Random Forest. The results showed that the SVM algorithm provided the best accuracy in classifying user sentiment. Based on fishbone diagram analysis, it was found that negative user sentiment was primarily caused by eight main factors: battery, charging infrastructure, initial cost, range, policy support, performance, model variety, and reliability. These findings are expected to provide insights for electric vehicle manufacturers to improve product quality and customer satisfaction.

Keyword: *Sentiment Analysis, Electric Vehicles, Machine Learning, Support Vector Machine, Naive Bayes, Random Forest*

PENDAHULUAN

Climate change has become a serious issue and a major concern for many countries. The impact of this change is closely related to the continuously increasing carbon emissions in Indonesia (Cakrawati Sudjoko, 2021). Uncontrolled use of fossil energy poses significant threats, including energy crises and environmental pollution contributing to climate change (Kene et al., 2021). In 2023, Indonesia was recorded as the country with the second largest carbon emissions from land-use change, after Brazil (Annur, 2023). Furthermore, data show that emissions from two-wheeled motor vehicles in Indonesia reach 300 million kilograms per day (Ismoyo, 2023).

The use of environmentally friendly alternative energy is expected to reduce carbon emission problems in Indonesia (Kwee, 2023). Electric vehicles, as one of the alternative solutions, can help address carbon emission issues, especially in urban areas. The development of electric cars has become a priority program for the Indonesian government. Various efforts are being made, including providing subsidies for the purchase of electric cars to increase their usage (Aziz et al., 2020). President Joko Widodo even stated that in the future, only electric cars would be allowed to operate in the new capital city (Fauzi, 2020).

The use of electric vehicles, such as electric cars and motorcycles, has significant potential to reduce pollutant emissions such as CO, NOx, HC, SO2, and PM (Resosudarmo et al., 2009). Electric vehicles produce much less air pollution compared to conventional vehicles using fossil fuels. This is crucial considering the increasing air pollution in major cities in Indonesia, such as Jakarta, which in August 2019 was recorded as the city with the highest air pollution levels in the world (Rahayu, 2019).

Sentiment analysis towards electric vehicles also becomes an important focus in this

research. Public sentiment towards this innovation can influence the adoption and development of electric vehicle technology (Riyadi et al., 2021). For instance, a study conducted by Ananda (Ananda et al., 2023) used the Support Vector Machine method to classify public sentiment tweets about electric vehicles. This analysis provides insights into the perceptions and potential barriers faced in the implementation of electric vehicles in Indonesia.

By understanding public sentiment and perceptions, the government and stakeholders can design more effective strategies to promote the use of electric vehicles. For example, Fitriana et al (Fitriana et al., 2021) in their research on sentiment analysis of opinions on the Covid-19 vaccine on Twitter, showed that sentiment analysis approaches can provide valuable insights into public opinion, which can then be used to inform policies and communication strategies. By analyzing online reviews, social media posts, and other forms of customer feedback, businesses can identify common themes, both positive and negative, related to their products and services (Rakib et al., 2024). Thus, the development of electric vehicles not only requires technological innovation and supportive policies but also a deep understanding of public sentiment and perceptions to ensure broader and sustainable adoption.

RESEARCH METHODS

This research uses a quantitative approach supported by sentiment analysis techniques. A quantitative approach was chosen because it allows for the collection of extensive data and generalization of findings, as well as facilitating the use of statistical methods to test research hypotheses (Santoso et al., 2022). Sentiment analysis techniques are used to identify and classify sentiment in collected tweets, enabling a systematic evaluation of user views on electric vehicles (Syakuro, 2017).

Data were collected from the social media platform Twitter using a Python library called tweet-harvest. The data collection process began by determining relevant keywords for the research, namely “Kendaraan Listrik”, “Motor Listrik”, “Mobil Listrik”, “Wuling EV”, and “Hyundai Ioniq”. Tweets containing these keywords were downloaded and stored in the form of a dataset. The collected data reflect real-time public opinions and encompass various perspectives on Twitter.

This research involves several main stages::

a. Data Collection: Done by downloading tweets from Twitter containing relevant keywords.
b. Data Preprocessing: Involves removing links, hashtags, usernames (@mentions),

punctuation, numbers, and other special characters. The review text is then tokenized and stemmed to convert words to their base forms.

c. Data Analysis: The processed data are labeled using a lexicon dictionary method to identify positive or negative sentiment. Data analysis is conducted with machine learning algorithms, namely Support Vector Machine (SVM), Naïve Bayes, and Random Forest. The models are trained with training data and evaluated using confusion matrices to determine the best model.

This research method enables researchers to accurately and efficiently identify and interpret public sentiment towards electric vehicles, providing valuable insights for manufacturers, policymakers, and other stakeholders.

RESULT AND DISCUSSION

The data was collected from the social media platform Twitter using five main keywords: "kendaraan listrik", "Mobil listrik", "Motor listrik", "Wuling EV", and "Hyundai Ioniq". The data collection process resulted in 3,608 data columns. Data labeling was carried out using the lexicon dictionary method with Python. This method involves the use of a list of words or phrases that have already been assigned sentiment values. Each token in the text is compared to this lexicon dictionary, and the corresponding sentiment value is assigned. The results showed 2,036 positive sentiments and 1,176 negative sentiments.

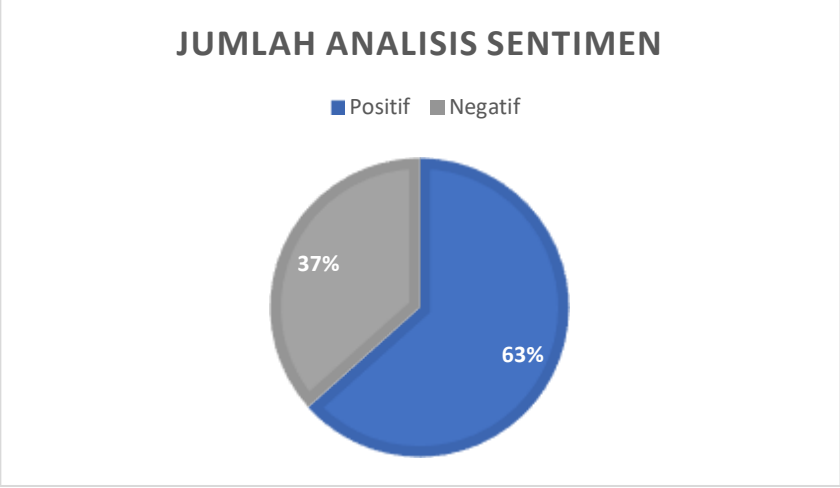

Figure 1. Sentiment Class Distribution

The classification process was carried out by splitting the data into training and testing sets. Three trials were conducted for each algorithm: the first trial used a 90:10 data split, the second trial used an 80:20 data split, and the third trial used a 70:30 data split.

Table 1. Comparison of Classification Accuracy of SVM, NBC, and Random Forest Algorithms

Experiment	Accuracy		
	SVM	NBC	*Random Forest*
Comparison 90 : 10	73%	70%	72%
Comparison 80 : 20	74%	70%	73%
Comparison70 : 30	74%	67%	72%
Avarage	73,6%	69%	71,3%

The table above shows the accuracy results of three machine learning algorithms, namely Support Vector Machine (SVM), Naive Bayes Classifier (NBC), and Random Forest, tested with three data split scenarios: 90:10, 80:20, and 70:30. SVM showed stable accuracy with a highest of 74% in the 80:20 and 70:30 splits, and an average of 73.6%, demonstrating good performance in classifying electric vehicle sentiment data. The NBC model exhibited greater accuracy variation, peaking at 70% in the 90:10 and 80:20 splits, but dropping to 67% in the 70:30 split, with an average of 69%. The Random Forest model demonstrated stable and good accuracy across all experiments, peaking at 73% in the 80:20 split, with an average of 71.3%, showing competitive performance compared to SVM.

Figure 2. Visualization of Positive Reviews Using Wordcloud

The image above is a wordcloud that shows word associations from tweets with positive sentiment regarding electric vehicles, helping to understand the topics and keywords in positive discussions on Twitter. Larger words indicate higher frequency, such as "kendara" (vehicle) and "mobil" (car), showing a focus on vehicle types. Brand names like "Hyundai" and "Wuling", especially the "Ioniq" model, are frequently mentioned, indicating appreciation for innovation and performance. Words like "listrik" (electric),

"motor", and "EV" stand out, indicating positive sentiment towards electric technology. The word "harga" (price) appears prominently, indicating discussions about competitive pricing or incentives. Words like "ramah" (friendly), "subsidi" (subsidy), "pemerintah" (government), and "inovatif" (innovative) highlight that environmental sustainability, government support, and technological innovation are key factors in positive sentiment. Overall, this wordcloud shows that the public values energy efficiency, technological innovation, and government support for environmentally friendly vehicles.

Figure 3. Visualization of Negative Reviews Using Wordcloud

The image above is a wordcloud that shows word associations from tweets with negative sentiment regarding electric vehicles, helping to understand the topics and keywords in negative discussions on Twitter. Larger words indicate higher frequency, such as "listrik" (electric) and "motor" (motor), indicating complaints related to the performance and reliability of electric technology. Brand names like "Hyundai" and "Wuling", especially "Ioniq", are frequently mentioned in a negative context due to technical and operational issues. Other prominent keywords include "baterai" (battery), "harga" (price), and "pajak" (tax), indicating issues with battery longevity, high initial costs, and insufficient tax policies. Words like "jalan" (road) and "infrastruktur" (infrastructure) highlight user frustration with the lack of adequate charging infrastructure. Overall, this wordcloud illustrates various issues faced by electric vehicle users, indicating the need for infrastructure improvements, cost reductions, and more supportive policies to address negative sentiment.

Based on data analysis, several main categories were identified, including battery issues, limited charging infrastructure, high initial costs, limited range, insufficient policy support, inadequate vehicle performance, limited model variety, and reliability issues. Each category encompasses specific causes that contribute to user dissatisfaction and

complaints. With this visualization, we can more clearly see cause-and-effect relationships, which can be used to design more effective solutions to address these problems.

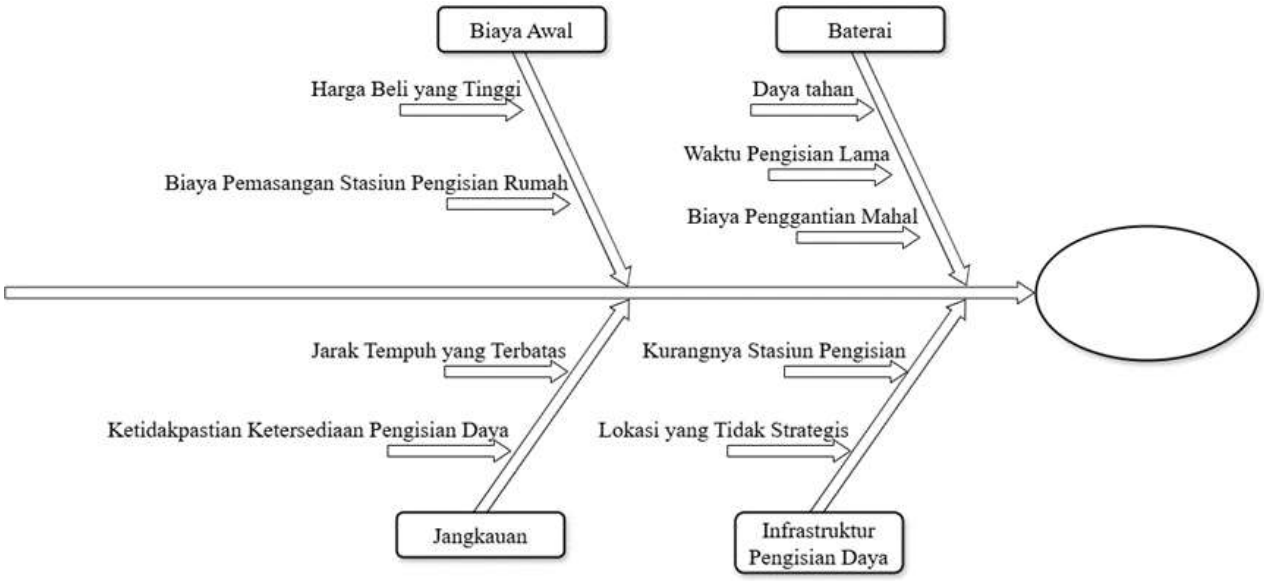

Figure 4. Fishbone Diagram Based on Negative Feedback (1)

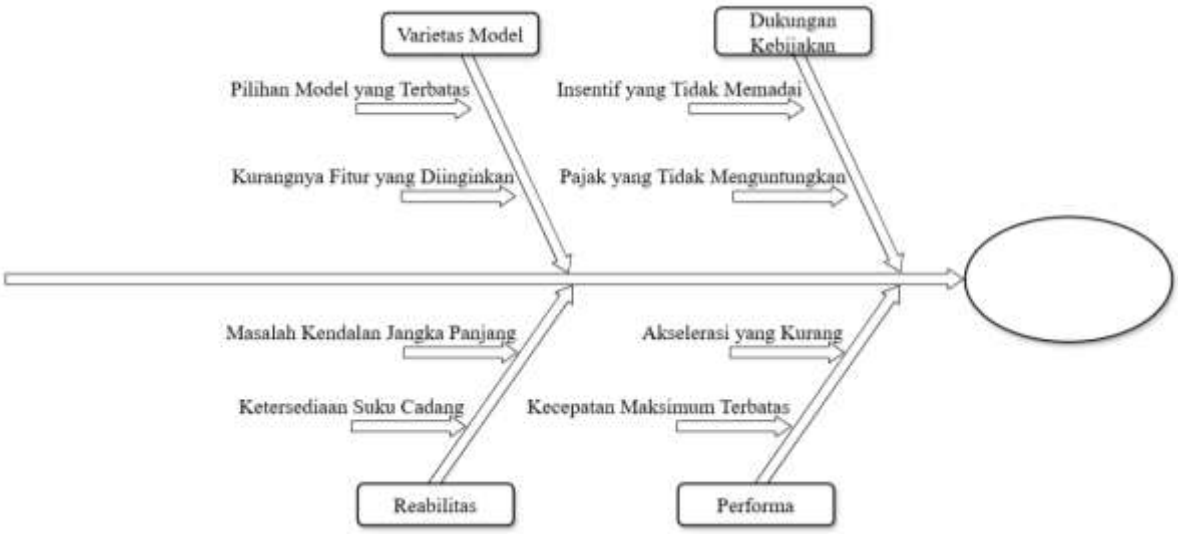

Figure 5. Fishbone Diagram Based on Negative Feedback (2)

CONCLUSION

Based on the analysis conducted, several conclusions were obtained as follows:

1. This research successfully collected 3608 Twitter review data using Python with the assistance of the tweet-harvest library. After labeling, the data were divided into two sentiment classes: positive and negative. Topic analysis of the reviews indicates that in the positive sentiment class, the frequently discussed topics include technological aspects. On the other hand, in the negative sentiment class, the dominant topics focus

on technical issues and performance of electric vehicles.

2. In the data processing, which involved training and testing data splits of 90:10, 80:20, and 70:30, the classification algorithms tested included Support Vector Machine (SVM), Naïve Bayes Classifier (NBC), and Random Forest. From the test results, the Support Vector Machine (SVM) algorithm demonstrated the best performance with the highest accuracy among the three algorithms. This indicates that SVM effectively captures patterns from the review data compared to NBC and Random Forest in the context of this research.
3. Based on the fishbone diagram analysis, 8 main dimensions were identified in the negative sentiment reviews. These issues encompass various aspects including Battery, Charging Infrastructure, Initial Cost, Range, Policy Support, Performance, Model Variety, and Reliability. This analysis provides deep insights into the areas that electric vehicle manufacturers need to improve to enhance customer satisfaction and reduce negative feedback in the future.

REFERENCES

Ananda, N., Fikry, M., Handayani, L., & Iskandar, I. (2023). *Klasifikasi Sentimen Tweet Masyarakat Terhadap Kendaraan Listrik Menggunakan Support Vector Machine*. *8*(4), 568–577. https://doi.org/10.32493/informatika.v8i4.36754

Annur, C. M. (2023). *Indonesia Penghasil Emisi Karbon Terbesar Kedua Dunia dari Sektor Alih Fungsi Lahan*. Databoks. https://databoks.katadata.co.id/datapublish/2023/12/05/indonesia-penghasil-emisi-karbon-terbesar-kedua-dunia-dari-sektor-alih-fungsi-lahan

Aziz, M., Marcellino, Y., Rizki, I. A., Ikhwanuddin, S. A., & Simatupang, J. W. (2020). Studi Analisis Perkembangan Teknologi Dan Dukungan Pemerintah Indonesia Terkait Mobil Listrik. *TESLA: Jurnal Teknik Elektro*, *22*(1), 45. https://doi.org/10.24912/tesla.v22i1.7898

Cakrawati Sudjoko. (2021). Strategi Pemanfaatan Kendaraan Listrik Berkelanjutan Sebagai Solusi Untuk Mengurangi Emisi Karbon. *Jurnal Paradigma: Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, *2*(2), 54–68.

Fauzi, Y. (2020). *Arahan Jokowi, Ibu Kota Baru Eksklusif Mobil Listrik-Otonom*. CNN INDONESIA. https://www.cnnindonesia.com/otomotif/20200116181124-579-466028/arahan-jokowi-ibu-kota-baru-eksklusif-mobil-listrik-otonom

Fitriana, F., Utami, E., & Al Fatta, H. (2021). Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, *5*(1), 19–25.

https://doi.org/10.31603/komtika.v5i1.5185

Ismoyo, B. (2023). *ESDM: Populasi Motor 120 Juta Unit, Sembur Emisi 300 Juta Kg per Hari*. https://www.cnnindonesia.com/otomotif/20230729214512-595-979474/esdm-populasi-motor-120-juta-unit-sembur-emisi-300-juta-kg-per-hari

Kene, R., Olwal, T., & van Wyk, B. J. (2021). Sustainable electric vehicle transportation. *Sustainability (Switzerland)*, *13*(22). https://doi.org/10.3390/su132212379

Kwee, E. Y. (2023). *Grid optimization and demand side mangement for electric vehicles penetration in remote area*.

Rahayu, L. S. (2019). *AirVisual: Paling Berpolusi, Udara Jakarta Pagi Ini Sangat Tidak Sehat!* Detiknews. https://news.detik.com/berita/d-4667683/airvisual-paling-berpolusi-udara-jakarta-pagi-ini-sangat-tidak-sehat

Rakib, M., Azis, M., Najib, M., & Isma, A. (2024). Impact of Digital Literacy, Business Innovation, Competitive Advantage on the Existence of SMEs: A quantitative study in Makassar City, Indonesia. *Quality - Access to Success*, *25*(198), 277–283. https://doi.org/10.47750/QAS/25.198.30

Resosudarmo, B., Nurdianto, D. A., & Yusuf, A. (2009). Greenhouse Gas Emission in Indonesia: The Significance of Fossil Fuel Combustion. *Regional Development, Energy and the Environment in Indonesia*, *July*, 146–159. https://www.researchgate.net/publication/342716834

Riyadi, A. F., Rahman, F. R., Nofa Pratama, M. A., Khafidli, M. K., & Patria, H. (2021). Assessment of Social Sentiment Toward Electric Vehicle Technology: Empirical Evidence in Indonesia. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, *11*(2), 141.

Santoso, A., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization. *Journal of Practical Computer Science*, *2*(1), 24–31. https://doi.org/10.37366/jpcs.v2i1.1084

Syakuro, A. (2017). Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro. *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*, 1–89. http://etheses.uin-malang.ac.id/11706/