Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Indonesian Public Sentiment Toward Electric Vehicles: Analysis of Social Media Data

Ni Wayan Sumartini Saraswati[1)*], I Wayan Dharma Suryawan[2)], I Dewa Made Krishna Muku[3)], I Kadek Agus Bisena[4)], Dewa Ayu Kadek Pramita[5)]
[1,2,3,4,5)]Institut Bisnis dan Teknologi Indonesia, Indonesia
[1)]sumartini.saraswati@instiki.ac.id, [2)] wayan.dharma@instiki.ac.id, [3)]dewamuku@instiki.ac.id, [4)]agus.bisena@instiki.ac.id, [5)]pramita.wayu@instiki.ac.id

Submitted :Aug 6, 2025 | **Accepted** : Aug 15, 2025 | **Published** : Aug 16, 2025

Abstract: The development of electric vehicles (EVs) in Indonesia has progressed significantly, supported by government subsidies for Battery-Based Electric Motor Vehicles. These subsidies have sparked mixed public reactions that some support them due to environmental benefits and pollution reduction, while others oppose them for various reasons. Social media platform X serves as a valuable source for gauging public opinion, though analyzing such data manually can be complex. To address this, sentiment analysis particularly using the Support Vector Machine (SVM) method offers an efficient solution. This study analyzes 23,031 Indonesian-language tweets from social media platform X, collected between October 2023 and July 2024, using SVM for sentiment classification. The best-performing model, with parameter C = 0.5 and without stemming, achieved an accuracy of 84.98%. The findings suggest that Indonesians generally view electric vehicles positively, with more favorable sentiments than negative ones. This study offers implications across methodological, industrial, and policy domains. Word cloud analysis further supports this, highlighting public support in areas such as pricing, infrastructure, and environmental impact. However, the study also identifies key concerns, including issues around subsidies, taxes, vehicle durability, battery types, and import regulations. Overall, the research provides meaningful insights into the diverse perspectives of Indonesian citizens regarding EVs, helping to inform future policy and development strategies.

Keywords: EVs, Indonesia, Sentiment Analysis, Support Vector Machine, Social Media X,

INTRODUCTION

Electric vehicles (EVs) have been in development since the 20th century, with adoption in Indonesia starting around 2010. Their usage has steadily increased over time. In April 2023, the Indonesian government introduced a subsidy policy for purchasing Battery-Based Electric Motor Vehicles (BEVs), including electric cars and buses. This policy aims to promote the use of electric vehicles to help reduce pollution and emissions (Kementrian Keuangan Republik Indonesia, 2023). By 2024, electric vehicles had gained popularity in Indonesia and were widely used in daily life. According to the Association of Indonesian Automotive Industries (GAIKINDO), EV sales in Indonesia have grown substantially since the introduction of the 2023 government subsidy policy, rising from over 10,000 units in 2022 to more than 43,000 units by 2024. 2022 (GAIKINDO, 2023, 2025) .The need for EVs is propelled by the urgent need to address environmental challenges and transform transportation towards sustainability (Omkar & Kumar, 2024). EVs help reduce greenhouse gas emissions, as they run on battery-powered electricity and produce 0% emissions, unlike conventional vehicles that rely on fossil fuels. Additionally, Indonesia has strong potential to grow its EVs industry, particularly in battery production, due to having the world's largest nickel reserves that a key material in battery manufacturing. The KBLBB policy in Indonesia received mixed public responses. While many supports electric vehicle adoption due to their environmental benefits, concerns remain. Key issues include inadequate infrastructure for charging and the high cost of EVs compared to conventional vehicles, which makes people hesitant to switch. Improved infrastructure and more affordable pricing are needed to boost adoption. International research has demonstrated that social media based sentiment analysis is effective in understanding public attitudes toward emerging transportation technologies. A study on autonomous vehicles (AVs) analyzed more than 950,000 tweets using web scraping, sentiment analysis, and statistical modeling to identify factors influencing public attitudes, such as safety concerns, technology

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

features, and policy issues (Wang et al., 2022). The findings provided valuable insights for technology development and policy formulation.

Building on these insights, examining public sentiment toward EVs in Indonesia can provide similarly valuable guidance for both policymakers and industry stakeholders. Public opinion about EVs is widely shared on social media platforms like Facebook, X (formerly Twitter), and Instagram, offering valuable insights for manufacturers and policymakers. These platforms help identify what people appreciate or criticize, especially regarding infrastructure, pricing, and policy. This information is crucial for shaping better policies and business strategies. However, manually analyzing large volumes of public opinion is challenging, making efficient and accurate methods essential. Since social media responses are text-based, sentiment analysis is a suitable approach (Habbat et al., 2021). As part of natural language processing, sentiment analysis identifies the mood of textual content, labeling it as positive, negative, or neutral (Jim et al., 2024). One effective method for sentiment analysis is Support Vector Machine (SVM), a machine learning technique that works by finding an optimal hyperplane by maximizing the distance between classes (Styawati et al., 2021). SVM has proven to be effective in predicting sentiment (Xu et al., 2024), making it a powerful tool for analyzing electric vehicle-related discussions on social media. Compared to Naïve Bayes, SVM generally achieves higher accuracy when feature independence assumptions are unrealistic, which is often the case in linguistic data. For instance, (Pavitha et al., 2022)compared Naïve Bayes and SVM for movie review sentiment analysis and found that while both algorithms performed well, SVM achieved slightly higher accuracy, indicating its stronger ability to handle diverse and complex textual expressions. Previous studies have successfully applied the SVM method for sentiment analysis in various contexts, such as movies or entertainment (Pavitha et al., 2022), tourism (Saraswati et al., 2023; Saraswati, Muku, et al., 2024; Saraswati, Putra, et al., 2024a), mobile applications (Faisal et al., 2024), food delivery services (Rismanah et al., 2024) and weather changes (Kolo & Supatman, 2024), achieving accuracy rates between 70% and 84.37%. However, no prior study has analyzed public sentiment toward EV adoption in Indonesia using a combination of SVM and the InSet Lexicon. This study addresses that gap by conducting sentiment analysis on social media X posts about EVs in Indonesia. It compares preprocessing techniques (with and without stemming) and tests different values of the SVM complexity parameter (C) to find the optimal model configuration. The study evaluates model performance using accuracy, precision, recall, and F1-score. In addition to classifying sentiments as positive or negative, it extracts key insights from each sentiment polarity to better understand public views. These insights aim to support policymakers and the automotive industry in shaping more effective policies and strategies by identifying public preferences, concerns, and perceptions surrounding electric vehicles.

LITERATURE REVIEW

A study that implemented SVM for both sentiment analysis and a recommendation system is presented by (Pavitha et al., 2022) in the context of movie reviews. This study proposes a movie recommendation system that utilizes Cosine Similarity to suggest movies based on genre, cast, overview, and ratings, while also incorporating sentiment analysis of user reviews to assess public opinion more effectively. To classify sentiments as positive or negative, the system employs two supervised machine learning algorithms Naïve Bayes and SVM, where SVM slightly outperforms NB in accuracy (98.63% vs. 97.33%). In the context of mobile applications, (Faisal et al., 2024) conducted sentiment analysis on public perceptions of the digital service PLN Mobile. The study utilized the SVM algorithm to classify user reviews, which is known for its effectiveness in handling textual data. The model evaluation results showed high precision for the negative class (90.91%) and very high recall for the positive class (99.03%), indicating that most positive reviews were accurately identified, although detection of negative reviews remains a weakness (recall 32.61%). The findings of this study not only reflect the dynamic public perception of the application but also reveal widespread misunderstandings regarding electrical service procedures and a lack of awareness about users' rights and responsibilities.

In the rapidly evolving e-commerce landscape, platforms like ShopeeFood have transformed consumer behavior by emphasizing convenience, efficiency, and a diverse range of food choices. To better understand customer satisfaction and concerns, (Rismanah et al., 2024) conducted a sentiment analysis of ShopeeFood customer reviews collected from Twitter using a Knowledge Discovery in Database (KDD) approach. The data, consisting of 2,246 customer reviews, underwent a comprehensive preprocessing stage including cleaning, case folding, tokenization, and stopword removal, followed by sentiment classification using the SVM algorithm. The evaluation results yielded an accuracy of 80.31%, precision of 73.22%, and recall of 95.58%, indicating the model's strong ability to identify positive sentiments, though with some trade-off in precision. Understanding public sentiment toward weather changes is becoming increasingly important due to its wide-ranging impact across various sectors of human life. Study by (Kolo & Supatman, 2024), Twitter data was utilized to analyze public opinions on weather changes The sentiment classification was performed using the SVM algorithm, which is well-regarded for text classification tasks. The model achieved an accuracy of 70%, with performance metrics showing precision and recall of 39% and an F1-score of 37%, particularly favoring the identification of positive sentiment.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In tourism context, reviews left by tourists on platforms such as TripAdvisor serve not only as decision-making tools for future visitors but also as valuable data for tourism stakeholders. In this context, the study by (Saraswati et al., 2023) conducted sentiment analysis on large-scale review data to explore the public image of Bali's tourist attractions. Using the VADER Lexicon method combined with term frequency, bigrams, and topic-based trigrams, the analysis revealed predominantly positive sentiment, emphasizing Bali's beautiful beaches, stunning temples, and friendly locals. However, negative sentiments were also detected, highlighting issues such as plastic waste, unclean tourist spots, and aggressive street hawkers, which could harm Bali's tourism image. The study demonstrates how sentiment analysis can extract meaningful insights from user-generated content and guide tourism stakeholders in enhancing destination image and tourist satisfaction. Study by (Saraswati, Putra, et al., 2024b) has proposed a hybrid approach combining lexicon-based methods with active learning (AL), specifically the LeALSVM method, which integrates VADER with AL-SVM to enhance sentiment analysis performance on large-scale tourism data. By leveraging VADER to guide the query process in AL, the system requires only a small amount of labeled data while maintaining high classification accuracy. Experimental results demonstrated that LeALSVM outperforms standalone methods (VADER, SVM, and lexicon-SVM) and achieves results closer to expert-labeled ground truth, offering an efficient and effective solution for automated sentiment analysis in big data tourism applications. In other aspects of electric vehicles, it is done by (Batmetan & Hariguna, 2024). In this research, SVM combined with TF-IDF is applied to analyze over 1,000 online comments regarding electric vehicle (EV) incentives, which include tax deductions and financial rewards. The model achieved 78.1% accuracy, with strong performance in classifying positive and neutral sentiments but significantly lower recall for negative sentiments. Table 1 presents a summary of the previous studies.

Table 1. Summary of the Previous Studies

Author	Domain	Methodology	Key Findings	Limitations
(Pavitha et al., 2022)	Movie recommendation and review sentiment	Cosine Similarity, Naïve Bayes, SVM	SVM: 98.63%, NB: 97.33%	Generalization limited to the film domain
(Faisal et al., 2024)	PLN Mobile app	SVM	Negative precision: 90.91%, Positive recall: 99.03%	Low negative recall (32.61%), indicating difficulty detecting minority sentiments
(Rismanah et al., 2024)	ShopeeFood (Twitter)	KDD, SVM	Accuracy: 80.31%, Precision: 73.22%, Recall: 95.58%	Not comparing SVM performance with other algorithms directly
(Kolo & Supatman, 2024)	Weather change perception (Twitter)	SVM	Accuracy: 70%, Precision: 39%, Recall: 37%	Low performance across all metrics, especially for negative sentiment
(Saraswati et al., 2023)	Bali tourism	VADER + n-gram	Predominantly positive sentiment	Data is limited to TripAdvisor
(Saraswati, Putra, et al., 2024b)	Tourism (big data)	LeALSVM	Outperforms standalone methods, close to expert-labeled ground truth	Tested only on tourism big data, so generalization to other domains is unproven.
(Batmetan & Hariguna, 2024)	Electric vehicle incentives	SVM with TF-IDF	Accuracy: 78.1%, Positive precision: 80%, Positive recall: 94.6%	Low negative recall (10.6%), needs improvement in balanced sentiment classification

Previous studies have shown that SVM achieves high accuracy (70–98%) in sentiment analysis across domains such as movie reviews, tourism, mobile apps, delivery services, environmental issues, and EVs. However, many are limited by narrow domain focus, lack of preprocessing variation, or absence of hybrid approaches combining lexicon-based and machine learning methods. To date, no research has examined public sentiment toward EV adoption in Indonesia using SVM with a lexicon-based method (InSet Lexicon) while evaluating multiple preprocessing techniques and model parameters. This study addresses that gap by analyzing social media opinions and offering strategic insights for policymakers and the automotive industry.

METHOD

The sentiment analysis conducted in this study went through several stages which are depicted in the flowchart in Fig. 1. It showed that the study will conduct two experiments, namely using stemming and without

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

stemming. It is going through the stemming process to find out whether stemming can provide better results in the sentiment analysis model.

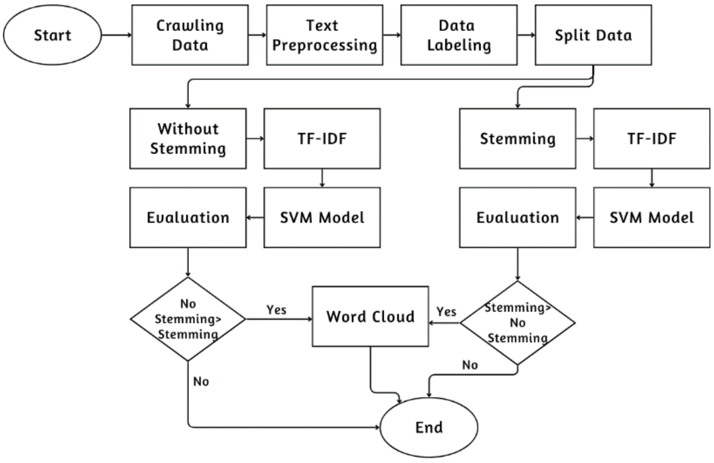


Fig. 1 Sentiment Analysis Flowchart

Data Collection

The data collection process was carried out using the crawling technique using Tweet Harvest created by Helmi Satria (Helmi, 2022). This tool can be used to retrieve tweets from social media X by requiring a token from account X. Tweet data was taken from the time span October 2023 - July 2024. This study gathers public opinions expressed in the Indonesian language with the aim of capturing insights that reflect the broader perceptions and attitudes of Indonesian society. It uses several keywords for the crawling process as shown in Table 2. The data collected was 23,031 data.

Table 2 Keywords Used to Crawling Data

No	Keywords
1	mobil listrik (electric car)
2	subsidi kendaraan listrik (electric vehicle subsidy)
3	kendaraan listrik (electric vehicle)

Pre-processing

The data that has been obtained cannot be analyzed because it is not well structured, there are duplications, and missing values. The data also still contains symbols and words that are not needed. The preprocessing phase reduces noise and enhances feature quality, helping to improve text classification performance (Occhipinti et al., 2022). The cleaning stage removes unnecessary elements such as punctuation, symbols, links, duplicates, and missing values, resulting in a refined dataset of 22,421 rows. Case folding standardizes text by converting all characters to lowercase (Isnan et al., 2023). Tokenization separates words using delimiters, enabling individual word analysis (Alomari & Ahmad, 2024; Postiglione, 2024). Normalization converts text to its standard form by correcting abbreviations, misspellings, and uncommon or out-of-vocabulary words (Meel & Vishwakarma, 2021), enhancing model accuracy. A normalization dictionary (Putri, 2021) specific to Indonesian is used for this purpose. Finally, stopword removal filters out common words like "dan", "yang", and "adalah" that do not provide important meaning (Mohd Nafis & Awang, 2021), allowing the model to focus on more meaningful terms for better performance.

Data Labeling

The data labeling process involves assigning sentiment labels to each tweet by matching root words with entries in a sentiment lexicon. Sentiment scores are calculated for each tweet based on the presence of words from the InSet Lexicon, which contains 3,609 positive and 6,609 negative words. Tweets with a sentiment score of ≥ 0 are labeled as positive, while those with a score < 0 are labeled as negative (Ependi et al., 2023). The InSet Lexicon (Koto & Rahmaningtyas, 2017), serves as the primary reference for this classification. As a result of this process, 18,535 tweets were classified as positive and 3,886 as negative, indicating an imbalanced dataset. Neutral sentiments are excluded from this study due to the binary nature of the InSet Lexicon, which only supports positive and negative classifications. The analysis therefore focuses exclusively on these two sentiment categories to provide a more explicit understanding of public opinion regarding electric vehicle public perception in Indonesia.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Split Data

Following the labeling process using the InSet Lexicon, the dataset is partitioned into training and testing subsets to evaluate the performance of the SVM model. The training data, comprising 20% of the overall dataset, is used to train the model, enabling it to learn patterns associated with sentiment classification. The remaining 80% serves as test data to assess the model's ability to classify unseen data accurately. The decision to use a smaller proportion of training data is based on the consideration of the performance of the lexicon method in sentiment analysis is generally yield moderate, several studies report an accuracy of 60% to 70% (Hendrawati et al., 2024). This suggests that even with limited training data, sufficient model performance can be achieved when SVM combined with lexicon-based labeling, thus optimizing resource efficiency without significantly compromising accuracy. The specific distribution of training and testing data is detailed in Table 3.

Table 3. Number of Training Data and Test Data

Class	Number of Tweet	Train Data (20%)	Test Data (80%)
Positive	18.535	3.707	14.828
Negative	3.886	777	3.109
Total	22.421	4.484	17.937

Stemming

The stemming stage changes words in the dataset into basic forms by removing word endings (Dinata et al., 2020). The purpose of the stemming process is to facilitate the identification of relationships between words that have the same meaning. For Indonesian language, this study conducted stemming using the Sastrawi library. Sastrawi reduces affixed words in Indonesian to their root form according to standard dictionary rules (Paskahningrum et al., 2023).

Weighting

The word weighting stage uses the TF-IDF method. The TF-IDF process calculates the weight by combining Term Frequency (TF) with Inverse Document Frequency (IDF) (Chamira, 2022). TF is how often words appear in a document (Putu et al., 2025). IDF reduces the weight of words that often appear in many documents and increases the weight of words that rarely appear. The results of the weighting are obtained from the multiplication of TF and IDF is shown in (1).

$$TFIDF(t,d,D) = TF(t,d) \cdot IDF(t,D) \quad (1)$$

Where t is a word (term), d is a document in the document set D, and D is a document set (corpus).

SVM Model

The classification process is carried out using SVM. This study employed an SVM because of several factors. Text data contain a large number of features, yet SVM performs well under such conditions as it is capable of handling high-dimensional data. Additionally, SVM has better generalization capabilities than other machine learning methods (Zhang et al., 2025). SVM is used to find the optimal hyperplane by maximizing the distance between classes (Pratiwi et al., 2021). The working concept of the SVM algorithm is shown in Fig. 2.

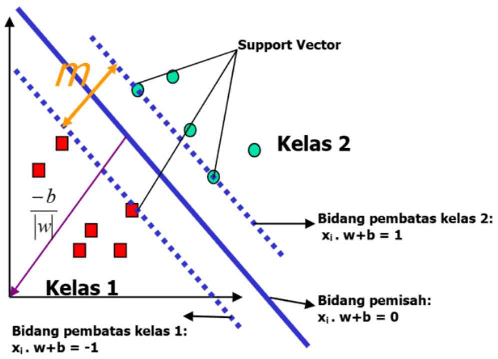


Fig. 2 Hyperplane SVM (10)

In order to obtain the most optimal hyperplane line in separating data into two classes, the hyperplane margin calculation is used and the maximum point is searched (Fahlevvi, 2022). To obtain the hyperplane on SVM, we can use (2).

*name of corresponding author

DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$(w.x_i)+b=0 \qquad (2)$$

SVM has several kernels, such as Linear, Polynomial, RBF, and Sigmoid (Rabbani et al., 2023), with this study using the help of the Linear kernel. Several other parameters used are the complexity parameter (complexity value) with values of 0.5, 1, 10, 100, 1000, gamma = 1 (learning rate value), degree = 1.1.

Model Evaluation

Model evaluation was conducted using the test data separated from the training set. This approach ensures that the model is tested on data it has not seen during training, providing a realistic assessment of its classification performance. Therefore, cross-validation was not applied in this study. The model was evaluated using a confusion matrix to measure accuracy, precision, recall, and F1-score. Evaluation was performed for two classes, positive and negative, resulting in a 2×2 confusion matrix. The confusion matrix displays the actual and predicted data, as shown in Fig. 3.

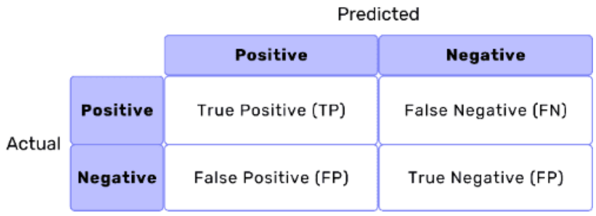


Fig. 3 Confusion Matrix

TP or True Positive is the number of correct predictions in the positive class, FP or False Positive is the number of incorrect predictions in the positive class, TN or True Negative is the number of correct predictions in the negative class, FN or False Negative is the number of incorrect predictions in the negative class. The average search for accuracy, precision, and recall values displayed in percentages using formula (3) – (6):

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+FN+TN}\times 100\% \qquad (3)$$

$$\text{Precision}=\frac{TP}{TP+FP}\times 100\% \qquad (4)$$

$$\text{Recall}=\frac{TP}{TP+FN}\times 100\% \qquad (5)$$

$$\text{F1-score} = 2\times\frac{\text{Precision}\times \text{recall}}{\text{Precision}+\text{recall}}\times 100 \qquad (6)$$

After testing, the results of positive and negative sentiments are visualized in the form of a word cloud. It displayed the frequency of occurrence of words in the document (Tupari et al., 2023). The more often a word appears in a document, the larger the size of the word in the word cloud.

Testing Scenario

This study uses two testing scenarios. The imbalance class data from InSet Lexicon will implement two pre-processing. The first scenario is an SVM model without using stemming and the second scenario is using stemming. The more complete scenarios are shown on the Table 4.

Table 4 Testing Scenario

Scenario	Preprocessing	C Parameter
1	Without Stemming	0.5
		1
		10
		100
		1000
2	With Stemming	0.5
		1
		10
		100
		1000

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Experimental Setup

All experiments in this study were implemented in Python 3.10 using the Scikit-learn library version 1.3.2 for the SVM model, the Sastrawi library version 1.0.1 for Indonesian stemming, and the NLTK library version 3.8.1 for tokenization and stopword removal. Pandas 2.1.4 and NumPy 1.26.2 were used for data manipulation, while Matplotlib 3.8.2 and WordCloud 1.9.3 were employed for visualization. This study were conducted on a desktop PC running Microsoft Windows 10 Pro (Version 10.0.19045), equipped with a 13th Gen Intel® Core™ i5-13400 processor (2.50 GHz, 10 cores, 16 threads), 32 GB RAM, and a 64-bit system architecture.

RESULT

The sentiment analysis performance for each parameter size of the SVM model are shown in Table 5. Based on the results of the model evaluation, the parameter C=0.5 gives the best results in the second scenario, namely the model without using stemming achieves an accuracy of 84.98%. If each parameter size is compared between the model using stemming and without stemming, the model without stemming also gives better evaluation results.

Table 5. Evaluation Model Results

Scenario	Pre-processing	C Parameter	Accuracy	Precision	Recall	F1-Score
1	With stemming	0.5	84.16%	84.00%	84,16%	84,07%
		1	83.46%	84,44%	83,46%	83,89%
		10	81.69%	84,44%	81,69%	82,73%
		100	81.69%	84,44%	81,69%	82,73%
		1000	81.69%	84,44%	81,69%	82,73%
2	Without stemming	**0.5**	**84.98%**	**85,15%**	**84,98%**	**85,06%**
		1	84.44%	85,61%	84,44%	84,93%
		10	82.85%	85,32%	82,85%	83,78%
		100	83.21%	85,39%	83,21%	84,04%
		1000	83.21%	85,39%	83,21%	84,04%

The complexity parameter (C) in SVM controls the trade-off between maximizing the margin and minimizing misclassification on the training data. A larger C value narrows the margin and penalizes errors more strictly, potentially leading to overfitting. Conversely, a smaller C value allows a wider margin and is more tolerant of errors but may result in lower accuracy. Experimental results show that higher C values (e.g., 100 and 1000) do not significantly improve performance and yield lower evaluation results than smaller values. Among the tested values, C = 0.5 provides the best balance between model complexity and accuracy, making it the optimal choice for sentiment classification and further analysis of term frequency. As for the confusion matrix, it is shown in Fig. 4.

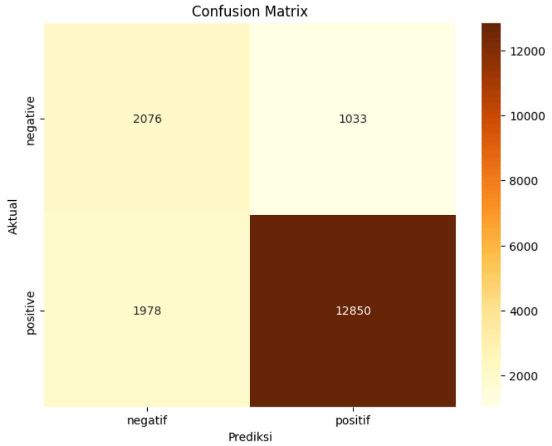


Fig. 4 Confusion Matrix

Evaluation results indicate that the model without stemming outperforms the model with stemming. This suggests that avoiding stemming allows the model to better recognize patterns and maintain stability. Stemming may reduce performance by removing important lexical information, particularly in imbalanced datasets where minority class features are already limited. Without stemming, the data preserves a greater variety of unique words, maintaining a high- dimensional feature space that aligns well with SVM's capabilities. However, the evaluation metrics should be interpreted with caution. The dataset's class imbalance may cause the model to favor the majority class, meaning that minority opinions, such as negative sentiment toward EV adoption, might be underrepresented.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179

e-ISSN : 2541-2019
p-ISSN : 2541-044X

At this stage, the model was tested under several data split schemes: 20:80, 50:50, 80:20, 100, and the lexicon method. The evaluation used 100 manually labeled data points as the ground truth (y_test). A significance test was then conducted to compare the classification results across these schemes. The purpose of this test was to ensure that the observed performance differences, such as accuracy, precision, and F1-score, were statistically significant. The results are presented in Table 6. Based on the evaluation, the without stemming model with a 20% training and 80% testing data split consistent with the approach proposed in this study achieved the highest accuracy compared to the other data split ratios.

Table 6. Performance Evaluation Against Ground Truth

Ratio	Evaluation	Accuracy	Precision	Recall	F1-Score
	Lexicon	70,37%	73,33%	73,33%	73,33%
20%:80%	Stemming	84,00%	88,00%	81,48%	84,62%
	Without Stemming	**86,00%**	**81,25%**	**96,30%**	**88,14%**
50%:50%	Stemming	80,00%	80,36%	83,33%	81,82%
	Without Stemming	85,00%	85,45%	87,04%	86,24%
80%:20%	Stemming	80,00%	80,36%	83,33%	81,82%
	Without Stemming	80,00%	81,48%	81,48%	81,48%
100%	Stemming	76,00%	74,19%	85,19%	79,31%
	Without Stemming	75,00%	73,02%	85,19%	78,63%

The prediction results are further visualized through unigram, bigram, and trigram word clouds, with unigrams presented in Fig. 5.The unigram analysis reveals frequently occurring words such as “mobil,” “listrik,” “kendaraan,” and “subsidi.” The term “subsidi” appears in both positive and negative sentiments, indicating mixed public opinions on the subsidy policy, some support it, while others oppose it. Similarly, the term “baterai” reflects differing views on battery usage, with positive sentiment associating it with industrial growth (e.g., “nikel,” “pabrik,” “produsen”), and negative sentiment suggesting potential concerns or drawbacks. Prominent electric vehicle brands like “Hyundai,” “BYD,” “Tesla,” and “Wuling” appear in positive sentiment, indicating their strong presence in the market. Infrastructure-related terms such as “SPKLU,” “pengisian,” and “ekosistem” suggest growing public appreciation for the expanding EVs support network. Positive sentiment words like “harga,” “murah,” and “beli” reflect perceptions of affordability and purchasing interest. Additionally, “PLN” is noted as a key stakeholder in supporting the electric vehicle ecosystem.

(a) (b)

Fig. 5 Unigram (a) sentiment positive (b) sentiment negative

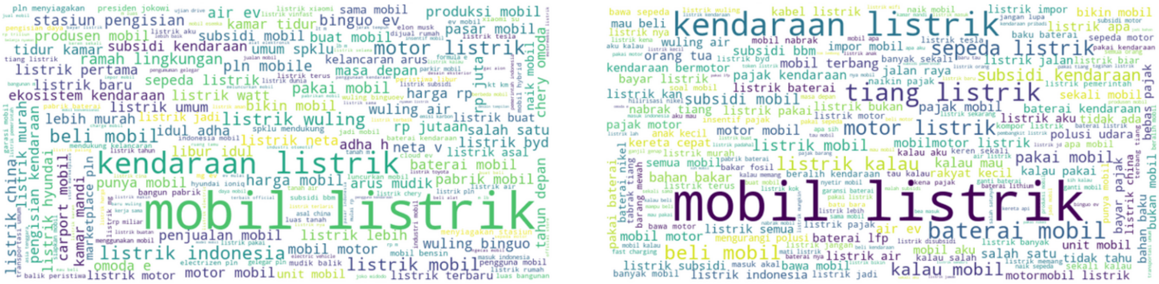
Negative sentiment terms such as “pajak” and “bayar” suggest that taxes and payment obligations related to electric vehicles are still perceived as burdensome by some members of the public. The recurring presence of “subsidi” in negative sentiment also highlights opposition to the subsidy policy. Bigram analysis in Fig. 6 provides deeper insight, showing that phrases like “mobil listrik” and “kendaraan listrik” appear in both sentiments, reflecting divergent public views on electric vehicles. In positive sentiment, bigrams such as “lebih murah,” “beli mobil,” and “harga mobil” reinforce perceptions of affordability. Phrases like “pln menyiagakan,” “spklu mendukung,” and “ekosistem kendaraan” indicate public recognition of Indonesia’s growing readiness to support electric mobility infrastructure. Additionally, terms such as “bangun pabrik” and “pabrik mobil” reflect support for domestic electric vehicle production. Bigrams like “pasar mobil,” “produsen mobil,” and references to brands such as “listrik Wuling,” “listrik BYD,” and “listrik Hyundai” further demonstrate the growing role of major manufacturers in the Indonesian EV market. The presence of the phrase “ramah lingkungan” underscores strong public support for the nvironmental benefits of electric vehicles.

The presence of terms like “mengurangi polusi” and “polusi udara” in negative sentiment suggests public skepticism about the effectiveness of electric vehicles in reducing pollution, particularly due to reliance on fossil

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

fuel-based power plants, as indicated by "bahan bakar" and "bakar fosil." Concerns about financial burdens are reflected in phrases such as "pajak mobil" and "pajak kendaraan." The term "beli mobil" implies lingering hesitation toward purchasing electric vehicles. Additionally, the frequent appearance of "baterai mobil" in negative sentiment points to issues such as battery lifespan, replacement costs, and recycling challenges. The contrasting terms "baterai kendaraan," "baterai LFP," and "baterai nikel" highlight public debate over battery types and their advantages or drawbacks. Lastly, terms like "impor mobil" and "listrik China" indicate concerns about dependency on imported electric vehicles and limited domestic alternatives.



(a) (b)

Fig. 6 Bigram (a) sentiment positive, (b) sentiment negative

The trigram analysis provides deeper insights into public sentiment toward electric vehicle adoption in Fig. 7. In positive sentiment, trigrams such as "beli mobil listrik" and "pakai mobil listrik" suggest public support for using and purchasing EVs. This is further reinforced by mentions of improved infrastructure, such as "pengisian kendaraan listrik," "stasiun pengisian kendaraan," and "listrik umum SPKLU," indicating that facility development supports EV adoption. Conversely, negative sentiment is reflected in the trigram "subsidi mobil listrik," suggesting disapproval of the subsidy policy. The presence of the term "kalau mobil listrik" indicates skepticism or sarcasm, reflecting uncertainty or criticism toward electric vehicle implementation.

The analysis reveals that in positive sentiment, terms like "mobil listrik Wuling" and "mobil listrik China" highlight the growing popularity and perceived quality of these brands in Indonesia. Their frequent mention suggests strong public interest and acceptance. Conversely, in negative sentiment, terms such as "impor mobil listrik" and "mobil listrik impor" reflect concerns about dependence on imported electric vehicles, indicating a public preference for increased domestic production. Additionally, terms like "baterai mobil listrik," "baterai kendaraan listrik," and "pajak mobil listrik" express doubts and concerns regarding battery issues and tax burdens, highlighting ongoing hesitations toward full electric vehicle adoption

(a) (b)

Fig. 7 Trigram (a) sentiment positive, (b) sentiment negative

DISCUSSIONS

The wordcloud analysis reveals that EVs adoption in Indonesia is driven by positive factors such as lower prices, improved charging infrastructure, and perceived environmental benefits. Industrial support is also growing through the establishment of battery factories and EVs manufacturers. However, challenges persist, including controversial over subsidy policies, tax burdens, battery-related concerns (cost and durability), and reliance on imported vehicles, particularly from foreign manufacturers. Additionally, skepticism remains about the environmental impact of EVs, especially due to the continued use of fossil fuels by PLN. Although there is enthusiasm for buying EVs, there are still doubts from the public regarding prices, battery reliability, and available infrastructure.

The findings of this study have several implications from the perspectives of the methodology, industry, and policy. From the methodological perspective, the results indicate that the best-performing model is without

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

stemming and with a C parameter of 0.5. This suggests that the model retains important features without stemming modifications in the context of Indonesian. The optimal parameter values can serve as a reference for future research on other sentiment classification tasks. From an industrial perspective, aspects such as pricing, charging infrastructure, and subsidy policies can provide valuable insights for industry stakeholders to enhance competitiveness through more competitive pricing strategies and a stronger supporting ecosystem such as expanding charging stations and improving battery availability. From a public policy perspective, insights from word cloud analysis can inform policymakers in refining subsidy schemes to ensure more targeted implementation. Additionally, these insights can serve as an evaluation of concerns regarding electricity generation that still relies on fossil fuels, highlighting the need for a transition to renewable energy for greater environmental benefits.

Compared to other studies, the research conducted by (Kolo & Supatman, 2024) analyzed public opinion sentiment related to weather changes using stemming in the preprocessing stage, achieving an accuracy of 70%, recall of 39%, and precision of 39%. In contrast, (Rismanah et al., 2024) analyzed customer sentiment toward ShopeeFood without applying stemming, and obtained an accuracy of 80.31%, recall of 95.58%, and precision of 73.22%. These findings align with the present study, indicating that the absence of stemming in the preprocessing stage tends to yield higher evaluation results compared to when stemming is applied. Likewise, when compared to the earlier study on electric vehicle incentives (Batmetan & Hariguna, 2024), which reported an accuracy of 78.1%, the present study achieved a notably higher accuracy of 84.98% under the optimal configuration (without stemming, C = 0.5). This represents a performance improvement of nearly 7 percentage points, underscoring the effectiveness of the chosen preprocessing approach and parameter tuning in enhancing the SVM model's classification capability.

This study has several limitations that may affect the generalizability and validity of the findings. Imbalance in class distribution may influence evaluation metrics. The imbalance in class distribution could bias the model toward the majority class. Also, the model only explores the impact of stemming and does not account for other preprocessing techniques such as lemmatization or stopword removal in isolation. Lastly, the study focuses solely on SVM with a linear kernel and a limited set of C values (0.5, 1, 10, 100, 1000). This methodological choice may not capture the potential performance of SVM with non-linear kernels, such as polynomial or radial basis function (RBF) kernels, which can better model complex, non-linear relationships in text features. Moreover, hyperparameter optimization was minimal, so the reported results may not reflect the optimal performance achievable by alternative methods or more extensive parameter tuning.

Based on the findings of this study, several suggestions are proposed for future research, such as the implementation of Aspect-Based Sentiment Analysis (ABSA), which involves separating sentiments based on specific aspects such as price and infrastructure. This approach provides a deeper understanding of the specific factors that most influence public opinion on electric vehicles. Furthermore, future research could explore hybrid approaches that combine lexicon-based and machine learning methods, or leverage deep learning architectures such as LSTM, BiLSTM, or BERT. These methods have shown strong performance in capturing semantic and contextual nuances in text, which may enhance classification accuracy, especially in handling the complexities of the Indonesian language and domain-specific vocabulary.

CONCLUSION

Sentiment analysis related to EVs using the SVM model with data without stemming, imbalance data, and parameter 0.5 gave the best accuracy results reaching 84.98%. This is because without stemming, the diversity of important features for classification can be maintained. The novelty of this research lies in the finding that avoiding stemming in Indonesian text preserves important lexical features, improving model performance, especially in domains with diverse vocabulary like EV adoption. This insight provides a methodological reference for future sentiment analysis studies in Indonesian. Meanwhile, insight analysis based on unigrams, bigrams, and trigrams shows that the adoption of EVs in Indonesia has received support from the public, especially in terms of low prices, infrastructure that has begun to develop, and has better environmental benefits compared to conventional cars. However, there are several challenges related to the adoption of EVs, which are depicted from negative public sentiment. Several challenges and concerns from the public include controversial subsidy policies, taxes that are considered burdensome, and car battery issues related to costs, durability, and battery types. In addition, dependence on imports is also a concern related to the dominance of foreign manufacturers in the Indonesian market. So, even though the public is enthusiastic about adopting EVs, some people still have doubts about the reliability of the battery and the readiness of its supporting infrastructure. These results can inform policymakers in refining subsidy schemes for more targeted implementation and guide industry stakeholders in enhancing competitiveness through pricing strategies, infrastructure expansion, and local production. For future work, the findings highlight the potential of extending this research using hybrid sentiment analysis approaches or deep learning models to capture more nuanced linguistic features and context. Such advancements could improve model robustness and support more precise, data-driven recommendations for both industry stakeholders and policymakers.

ACKNOWLEDGMENT

We thank INSTIKI Research and Development Program in 2025 for the support, both morally and financially, for the successful implementation of this research.

REFERENCES

Alomari, D., & Ahmad, I. (2024). Exploring Character Trigrams for Robust Arabic Text Classification: A Comparative Analysis in the Face of Vocabulary Expansion and Misspelled Words. *IEEE Access*, *12*(March), 57103–57116. https://doi.org/10.1109/ACCESS.2024.3390048

Batmetan, J. R., & Hariguna, T. (2024). Sentiment Unleashed: Electric Vehicle Incentives Under the Lens of Support Vector Machine and TF-IDF Analysis. *Journal of Applied Data Sciences*, *5*(1), 122–132. https://doi.org/10.47738/jads.v5i1.162

Chamira, S. (2022). Implementasi Metode Text Mining Frequency-Invers Document Frequency (Tf-Idf) Untuk Monitoring Diskusi Online. *Journal of Informatics, Electrical and Electronics Engineering*, *1*(3), 97–102. https://doi.org/10.47065/jieee.v1i3.353

Dinata, R. K., Safwandi, S., Hasdyna, N., & Mahendra, R. (2020). Kombinasi Algoritma Brute Force dan Stemming pada Sistem Pencarian Mashdar. *CESS (Journal of Computer Engineering, System and Science)*, *5*(2), 273. https://doi.org/10.24114/cess.v5i2.17989

Ependi, U., Aliya, S., & Wibowo, A. (2023). Sentiment Analysis of Covid-19 Handling in Indonesia Based on Lexicon Weighting. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *12*(1), 76–82. https://doi.org/10.32736/sisfokom.v12i1.1615

Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Dan Komunikasi Pemerintahan*, *4*(1), 1–13. https://doi.org/10.33701/jtkp.v4i1.2701

Faisal, H., Febriandirza, A., & Hasan, F. N. (2024). Analisis Sentimen Terkait Ulasan Pada Aplikasi PLN Mobile Menggunakan Metode Support Vector Machine. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, *5*(1), 303–312.

GAIKINDO. (2023). *GAIKINDO: Permintaan Mobil Listrik Ada, tapi tak Besar*. GAIKINDO. https://www.gaikindo.or.id/gaikindo-permintaan-mobil-listrik-ada-tapi-tak-besar/

GAIKINDO. (2025). *Jumlah Total Whole Sales Khusus Mobil BEV dan PHEV Berdasar Merek di Indonesia sepanjang Januari – Desember 2024*. GAIKINDO. https://www.gaikindo.or.id/jumlah-total-whole-sales-khusus-mobil-bev-dan-phev-berdasar-merek-di-indonesia-sepanjang-januari-desember-2024/

Habbat, N., Anoun, H., & Hassouni, L. (2021). A Novel Hybrid Network for Arabic Sentiment Analysis using fine-tuned AraBERT model. *International Journal on Electrical Engineering and Informatics*, *13*(4), 801–812. https://doi.org/10.15676/ijeei.2021.13.4.3

Helmi, S. (2022). *Tweet Harvest*. Github. https://github.com/helmisatria/tweet-harvest

Hendrawati, T., Ginantra, N. L. W. S. R., & Saiman, C. M. (2024). Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine dan Lexicon Based. *Tematik*, *11*(1), 56–64. https://doi.org/10.38204/tematik.v11i1.1890

Isnan, M., Elwirehardja, G. N., & Pardamean, B. (2023). Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model. *Procedia Computer Science*, *227*, 168–175. https://doi.org/https://doi.org/10.1016/j.procs.2023.10.514

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*(January), 100059. https://doi.org/10.1016/j.nlp.2024.100059

Kementrian Keuangan Republik Indonesia. (2023). *Siaran Pers: Akselerasi Transformasi Ekonomi, Pemerintah Luncurkan Insentif Pembelian KBLBB Roda Empat dan Bus Per 1 April 2023*. Kementrian Keuangan Republik Indonesia. https://www.kemenkeu.go.id/informasi-publik/publikasi/siaran-pers/Pemerintah-Luncurkan-Insentif-Pembelian-KBLBB

Kolo, S. Y., & Supatman, S. (2024). Analisis Sentimen Terhadap Opini Masyarakat Terkait Perubahan Cuaca Di Indonesia Menggunakan Algoritma Support Vector Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1412–1416. https://doi.org/10.36040/jati.v8i2.8988

Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *2017 International Conference on Asian Language Processing (IALP)*, 391–394. https://doi.org/10.1109/IALP.2017.8300625

Meel, P., & Vishwakarma, D. K. (2021). HAN, image captioning, and forensics ensemble multimodal fake news detection. *Information Sciences*, *567*, 23–41. https://doi.org/10.1016/j.ins.2021.03.037

Mohd Nafis, N. S., & Awang, S. (2021). An Enhanced Hybrid Feature Selection Technique Using Term

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 3, July 2025
DOI : https://doi.org/10.33395/sinkron.v9i3.15179
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification. *IEEE Access*, *9*, 52177–52192. https://doi.org/10.1109/ACCESS.2021.3069001

Nur Rismanah, S., Astuti, R., & M. Basysyar, F. (2024). Penerapan Algoritma Support Vector Machine Dalam Menganalisis Sentimen Ulasan Pelanggan Shopeefood Berdasarkan Twitter. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 406–412. https://doi.org/10.36040/jati.v8i1.8401

Occhipinti, A., Rogers, L., & Angione, C. (2022). A pipeline and comparative study of 12 machine learning models for text classification. *Expert Systems with Applications*, *201*(March), 117193. https://doi.org/10.1016/j.eswa.2022.117193

Omkar, M., & Kumar, M. V. (2024). Design of Filter for Single-Phase Bidirectional Battery Charger for Electrical Vehicle Applications. *International Journal on Electrical Engineering and Informatics*, *16*(4), 544. https://doi.org/10.15676/ijeei.2024.16.4.3

Paskahningrum, Y. K., Utami, E., & Yaqin, A. (2023). A Systematic Literature Review of Stemming in Non-Formal Indonesian Language. *International Journal of Innovative Science and Research Technology*, *8*(1). https://doi.org/10.5281/zenodo.7547482

Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022). Movie recommendation and sentiment analysis using machine learning. *Global Transitions Proceedings*, *3*(1), 279–284. https://doi.org/10.1016/j.gltp.2022.03.012

Postiglione, A. (2024). Finite State Automata on Multi-Word Units for Efficient Text-Mining †. *Mathematics*, *12*(4). https://doi.org/10.3390/math12040506

Pratiwi, R. W., H, S. F., Dairoh, D., Af'idah, D. I., A, Q. R., & F, A. G. (2021). Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM). *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, *4*(1), 40–46. https://doi.org/10.20895/inista.v4i1.387

Putri, A. A. (2021). *Text Pre-Processing*. Github. https://doi.org/10.1007/978-3-030-85085-2_3

Putu, N., Widiantari, T., Made, E. I., Dwi, A., Kadek, N., & Rusjayanthi, D. (2025). *Food Recipe Recommendation System with Content-Based Filtering and Collaborative Filtering Methods*. *9*(3), 1167–1176.

Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 153–160. https://doi.org/10.57152/malcom.v3i2.897

Rismanah, S. N., Astuti, R., & M. Basysyar, F. (2024). Penerapan Algoritma Support Vector Machine Dalam Menganalisis Sentimen Ulasan Pelanggan Shopeefood Berdasarkan Twitter. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 406–412. https://doi.org/10.36040/jati.v8i1.8401

Saraswati, N. W. S., Muku, I. D. M. K., Suryawan, I. W. D., Pramita, D. A. K., & Bisena, I. K. A. (2024). Balinese Temple: The Image and Characteristics of Tourists based on Sentiment Analysis. *2024 IEEE International Symposium on Consumer Technology (ISCT)*, 19–24. https://doi.org/10.1109/ISCT62336.2024.10791104

Saraswati, N. W. S., Putra, I. K. G. D., Sudarma, M., & Sukarsa, I. M. (2023). The Image of Tourist Attraction in Bali Based on Big Data Analytics and Sentiment Analysis. *2023 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS)*, 82–87.

Saraswati, N. W. S., Putra, I. K. G. D., Sudarma, M., & Sukarsa, I. M. (2024a). Enhance Sentiment Analysis in Big Data Tourism Using Hybrid Lexicon and Active Learning Support Vector Machine. *Bulletin of Electrical Engineering and Informatics*, *99*(1), 1–12. https://doi.org/10.11591/eei.v13i5.7807

Saraswati, N. W. S., Putra, I. K. G. D., Sudarma, M., & Sukarsa, I. M. (2024b). Enhance Sentiment Analysis in Big Data Tourism Using Hybrid Lexicon and Active Learning Support Vector Machine. In *Bulletin of Electrical Engineering and Informatics*.

Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(5), 904–910. https://doi.org/10.29207/resti.v5i5.3380

Tupari, T., Abdullah, S., & Chairani, C. (2023). Visualisasi Data Analisa Sentimen RUU Omnibus Law Kesehatan Menggunakan KNN dengan Software RapidMiner. *Jurnal Informatika: Jurnal Pengembangan IT*, *8*(3), 261–268. https://doi.org/10.30591/jpit.v8i3.5641

Wang, S., Li, M., Yu, B., Bao, S., & Chen, Y. (2022). Investigating the Impacting Factors on the Public's Attitudes towards Autonomous Vehicles Using Sentiment Analysis from Social Media Data. *Sustainability (Switzerland)*, *14*(19). https://doi.org/10.3390/su141912186

Xu, Q. A., Jayne, C., & Chang, V. (2024). An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews. *Technological Forecasting and Social Change*, *202*(April 2023), 123326. https://doi.org/10.1016/j.techfore.2024.123326

Zhang, Z., Liu, X., Wang, Y., Li, E., & Zhang, Y. (2025). Stability Prediction Model of Transmission Tower Slope Based on ISCSO-SVM. *Electronics (Switzerland)*, *14*(1). https://doi.org/10.3390/electronics14010126

*name of corresponding author