The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 1, February 2025
https://doi.org/10.33022/ijcs.v14i1.4649

Sentiment Analysis of Public Responses Regarding the Use of Electric Cars in Indonesia with *Support Vector Machine* and *Random Forest* Methods

Yessica Anglila Seraphina[1], Putu Harry Gunawan[2]
yessicaanglilas@student.telkomuniversity.ac.id, phgunawan@telkomuniversity.ac.id[2]
[1,2] Informatics School Of Computing, Telkom University, Bandung, Indonesia

Article Information

Received : 15 Jan 2025
Revised : 30 Jan 2025
Accepted : 7 Feb 2025

Keywords

SVM, Random Forest, Sentiment Analysis, Electric Car, Social Media

Abstract

The diminishing use of fossil fuels has encouraged the search for alternative energy sources, one of which is the electric car. However, public acceptance of electric cars in Indonesia is not widely understood. This study aims to analyze public sentiment towards electric cars based on data from X social media. The dataset used consists of 3,450 data, which is analyzed using two machine learning methods, namely Support Vector Machine (SVM) and Random Forest. The research was conducted in three scenarios: SVM kernel comparison, Random Forest performance evaluation with various numbers of n-estimators (1, 10, 100), and performance comparison between the two methods. The experimental results show that Random Forest with 100 n-estimators produces the highest accuracy of 90.72% and F1-Score of 87.54%, while SVM with RBF kernel produces 89.35% accuracy and F1-Score of 85.15%. The performance difference of 1.37% shows that Random Forest is more effective in this sentiment analysis.

A. Introduction

The industrial revolution in the 1900s caused the world's energy sources to undergo several changes. One of the most influential changes is the change in the use of biomass such as firewood to fossils such as coal, petroleum and coal gas to meet energy needs. The increasing use of fossil energy causes greenhouse gas emissions that have an impact on climate instability and an increase in earth and sea water temperatures[1]. Statistics show an increase in fossil energy consumption worldwide. Increased consumption of fossil fuels increases carbon emissions in the atmosphere which impacts the environment as a cause of global warming[2].

This proves that fossil energy has become a basic need of Indonesian society. Therefore, the energy transition from fossil fuels to the use of alternative energy sources requires support by the world community. This shift has a major game-changing impact with the aim of slowing climate change and developing resources. Thus, researchers, designers, car manufacturers, and transportation agencies need to develop alternative energy resources, such as electric power, for vehicles as fossil fuels deplete [1]. The acceptance of electric vehicles in a country is greatly influenced by the public's view of the technology [3].

Positive views play an important role in encouraging the use of electric vehicles, but there are also many negative views regarding their implementation. With many variations in public opinion, sentiment analysis is a relevant technique for mapping public perceptions of electric vehicles. This research is to analyze the sentiment of public response to the use of electric cars in Indonesia. The data used for sentiment analysis in this study were obtained from the X social media platform. Social media platform X is one of the social media platforms that is widely used by the public to convey opinions, both in the form of positive responses and negative responses, one of which is related to electric cars.

This research implements two machine learning methods, namely Support Vector Machine (SVM) and Random Forest. The selection of these two methods is based on the results of previous studies which prove that both methods are able to produce a high level of accuracy in sentiment analysis. Research specifically regarding the evaluation of Indonesian people's perceptions of electric cars has not been found through this approach. Therefore, this research is expected to contribute to understanding public acceptance of electric cars in Indonesia.

SVM has been used in several studies, including studies that discuss the use of SVM and Random Forest methods for sentiment analysis of user reviews of KAI Access on Google Playstore[4], sentiment analysis of online transportation using Word2Vec text embedding model feature extraction and Support Vector Machine (SVM) algorithm[5], sentiment analysis of the use of the Shopee application using the Support Vector Machine (SVM) algorithm[6], and sentiment analysis related to the transfer of Indonesia's capital city using the Support Vector Machine (SVM) algorithm[7]. In these four studies, SVM obtained accuracy results of 97% on KAI Access, 87% on the Gojek application, 82% on the Grab application, 98% on the Shopee application, and 96.68% on sentiment analysis of moving the capital city of Indonesia. Seeing from previous research, the SVM method shows quite high accuracy results. Therefore, in this study, the SVM method is used because it has the ability to work well on high-dimensional data, can handle complex and non-linear data using kernels, and is able to produce an optimal separation hyperplane so that the resulting accuracy is higher than other methods[8].

In addition to SVM, the Random Forest method has also been used in several studies, including sentiment analysis of Gojek and Grab users on Twitter social media using Random Fores[9], sentiment analysis of Dana application reviews using the Random Forest metho[10], and sentiment analysis of Yogyakarta Tentrem Hotel reviews using the Random Forest Classifier algorithm[11]. In these studies, Random Forest obtained accuracy results of 84% on the Dana application, 76.25% on Grab and Gojek application users on social media, and 90% on Tentrem Yogyakarta Hotel reviews. Seeing these results, Random Forest shows good performance because this algorithm is able to handle data that has many variables (high-dimensional data) by building a large number of decision trees and voting to get the final result[12]. Another advantage is its ability to prevent overfitting on large datasets, resulting in a more stable and accurate model[13]. Therefore, in this research, the Random Forest method is used as a comparison to see its performance against the SVM method.

B. Research Method

The system for analyzing the sentiment of public response to electric cars in Indonesia with the Support Vector Machine (SVM) and Random Forest methods on the X platform used in this study is shown in Figure 1. The first stage before getting the results of the comparison of the two methods is data crawling, where data is taken from the X platform using the keywords "mobil listrik", "ioniq5", "AirEV", "BYD Seal". The next stage after the data is collected is data labeling, which is the process of labeling the data with a value of 1 for positive sentiment and 0 for negative sentiment.The next stage after the data labeling process is complete is preprocessing to clean and prepare the data. The data that has been processed is then divided into training data and test data. Furthermore, the data is processed using the SVM and Random Forest methods. The results of the two methods are then compared to determine the method that provides the highest accuracy in analyzing sentiment towards the use of electric cars in Indonesia.

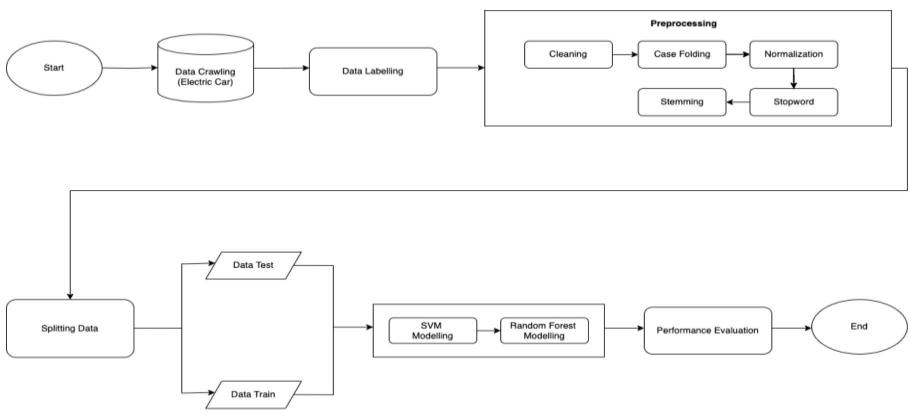

Figure 1. Constructed System

a. Dataset

This research uses Indonesian language datasets which are public responses to the use of electric cars in Indonesia obtained from X social media platforms.

Data collection is done through retrieving information from various sources called the crawling process method such as blogs, social media, and other websites. Labeling is done manually after the data is collected, then grouping the data into two classes, namely positive opinions and negative opinions. Table 1 is an example of data labeling results

Table 1. Dataset

Label	Sentence
1	Wuling Air EV nih mobil paling favorit gw sih. Gampang dibawa gw juga masih jomblo jadi cocok buat sendiri. Klo kalian suka sama Air EV gak guys?
0	lah geotermal kurang green apelagi dah? secara literal tinggal ngerebus aer doang dari gas kentut bumi jadi energi lebih kotor tambang nikel buat tesla ama mobil jelek ioniq

b. Data Pre-processing

The data preprocessing flowchart shown in Figure 2 in this study is conducted to ensure that the data is optimally prepared so that it can be efficiently used in the analysis and development of machine learning models. The data preprocessing stages applied include cleaning, letter folding, stopword elimination, stemming, and tokenization.

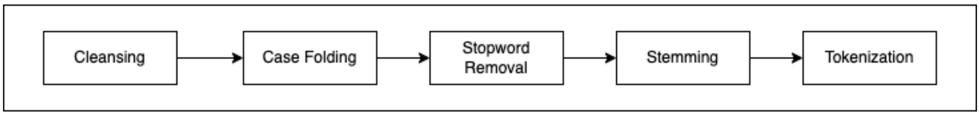

Figure 2. Data Preprocessing Flowchart

c. Support Vector Machine

Support Vector Machine (SVM) is a learning method that has many advantages, making it very popular. SVM is supported by a strong theoretical foundation that is able to perform classification with a higher level of accuracy than most other algorithms[6]. The outermost data object in SVM that is closest to the hyperplane is called the support vector, this support vector is taken into account by SVM to be able to find the most optimal hyperplane while other objects are not taken into account [14]. In this study, SVM is used to classify datasets that have been processed in the pre-processing stage. The data used as input has been manually labeled into two classes, namely positive opinions with the label "1" and negative opinions with the label "0". In this research, SVM acts as a text classification method by utilizing the concept of hyperplane. and in this research also compares the level of accuracy produced by various kernels in SVM.

d. Random Forest

In this study, Random Forest used a large number of decision trees during the training process to improve classification accuracy. The number of trees used in this study are 1, 10, and 100 trees. Random Forest is an ensemble method in machine learning that makes decision trees the basis for classification. One important aspect of Random Forest is the use of bootstrap sampling techniques to build prediction trees. Each decision tree makes predictions by randomly selecting predictors, and the prediction results from all will be combined using the majority vote method for classification or averaging for regression[15].

Figure 3. Random Forest Process

To get the final result, a decision tree is created by determining the root node and ending with several leaf nodes. Figure 3 is an illustration of the decision tree.

e. Confusion Matrix

Confusion matrix tables are used for classification model performance assessment by comparing the predicted results of the model with the actual labels in the data set. Through this evaluation, metrics such as accuracy, F1 Score, precision, and recall are obtained which provide a more in-depth picture of the model's performance. There are four possible outcomes in classification evaluation, if positive data is successfully predicted as positive, it is called true positive (TP), if positive data is predicted as negative, it is called false negative (FN). If the data is negative and correctly predicted as negative, it is called true negative (TN), but if the negative data is predicted as positive, it is called false positive (FP).

Based on the true positive (TP), true negative (TN), false positive (FP), false negative (FN) values, the recall, accuracy, and precision values can be obtained with the following equations:

Equation (1) shows the formula of Recall. Recall is metric that can measure ability model toidentify all class instances in the dataset.

$$Recall = \frac{TP}{TP+ FN} \quad (1)$$

Equation (2) shows the formula of Precision. Precision is a metric that measures the accuracy of the information to correctly identify instances of positive classes.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Equation (3) shows the formula of F1-Score. F1-Score is the average of the precision and recall.

$$F1 = 2 \times \frac{(precision \times recall)}{(precision + recall)} \quad (3)$$

Equation (4) shows the formula of Accuracy. Accuracy in the confusion matrix is used to measure how good a classification model is.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (4)$$

C. Result and Discussion

a. Model Evaluation

The test scenarios in this study were carried out with two classification models, namely SVM and Random Forest. The first scenario aims to determine the highest level of accuracy by comparing various types of kernels owned by SVM. The second scenario compares the performance of the Random Forest model with variations in the number of n-estimators. The data used in the test is obtained from the results of pre-processing on a dataset that has been processed in advance. The third scenario compares the results of both methods.

The first scenario is to find the highest level of accuracy by comparing 4 SVM kernels. Previously done is the separation of feature data into x and sentiment label analysis into y. By pre-processing text data into numerical representations carried out using CountVectorizer using n-gram (1, 2) to consider combinations of one word to two words. Based on the first scenario, the highest accuracy result is obtained by the RBF kernel which gets an accuracy result of 89.35%, F1-Score 85.15%. The following is a table of the first scenario test results.

Table 2. SVM Results

Kernel	Accuracy	F1-Score
Polynomial	89,20%	84,32%
RBF	89,35%	85,15%
Sigmoid	89,13%	84,01%
Linear	89,13%	84,01%

In the second scenario, a comparison of the number of n-estimators in the Random Forest algorithm, namely 1, 10, and 100. The test results presented in

Table 8 show that n-estimators of 100 produce the best performance with an accuracy of 90.72%, F1- Score of 87.54%.

Table 3. Random Forest Test Results

n-estimators	Accuracy	F1-Score
1	85,14%	84,97%
10	90,29%	88,18%
100	90,72%	87,54%

In the third scenario, a comparison of the results between the two SVM and *Random Forest* methods was carried out. The highest accuracy results are obtained by *Random Forest* with an accuracy of 90.72%, F1-Score 87.54%.

Table 4. Comparison of The Results of The Two Methods

Methods	Accuracy	F1-Score
SVM	89,35%	85,15%
Random Forest	90,72%	87,54%

The first scenario aims to determine the accuracy level by comparing the four SVM kernels. The test results show that the RBF kernel produces the best accuracy rate of 89.35% and an F1-Score of 85.15%. In other kernels, the accuracy results do not show much difference.The polynomial and RBF kernels experienced an increase in accuracy of 0.15%, while the sigmoid and linear kernels produced the same accuracy results. The second scenario compares the Random Forest algorithm with different numbers of n-estimators, namely 1, 10, and 100. The best accuracy results were obtained with n-estimators set to 100, achieving an accuracy of 90.72% and an F1-Score of 87.54%.

The most significant increase in accuracy was observed when increasing the number of n-estimators from 1 to 10, which amounted to 5.15%. However, from n-estimators 10 to 100, the increase in accuracy was only 0.43%, indicating that increasing the number of n-estimators beyond a certain point did not lead to a significant improvement. The third scenario conducted a performance comparison between the two methods: SVM and Random Forest. The test results show that Random Forest produces the highest accuracy, with an accuracy value of 90.72% and an F1-Score of 87.54%. Meanwhile, the SVM method produces an accuracy value of 89.35% and an F1-Score of 85.15%. The difference in accuracy between the two methods is 1.37%, with Random Forest showing better performance

b. Sample Testing

Based on the results of sample testing on six text samples, the Random Forest model is able to produce fairly good predictions. From a total of six texts, the model successfully predicted five texts correctly, while one text experienced a prediction error.

Table 5. Sample Testing Three Text Data Using Random Forest

Text	Label	Predictions
Mobil ini sangat efisien dan ramah lingkungan, saya sangat menyukainya	1	1
Kendaraan listrik ini memberikan pengalaman berkendara yang luar biasa	1	1
Sangat hemat biaya dan mudah untuk digunakan, benar-benar inovatif	1	1
antara bakar sendiri terus nyalahin masaa atau ya dibakar aja biar diganti ioniq5 biar si paling EV ramah lingkungan tai lincung	0	1
Tpi ioniq 5 designnya kureng suka sih :)	0	0
Batrei ioniq 5 sampe 400jt.. mau ga mau harus beli kalau ngedrop meningan beli mbl bahan bakar minyak belinya seperlu kita pake	0	0

The model successfully recognized positive text with an original label of one, and was also able to predict negative text with an original label of zero. However, in one of the negative texts in the Indonesian dataset, the model incorrectly predicted the text "antara bakar sendiri terus nyalahin masaa atau ya dibakar aja biar diganti ioniq5 biar si paling EV ramah lingkungan." The error occurred because this text was ambiguous, where the actual context was sarcasm.

D. Conclusion

This research aims to develop a sentiment analysis program to evaluate public responses regarding the use of electric cars in Indonesia. The analysis was conducted using SVM and Random Forest methods, with a dataset consisting of 3,450 tweets collected from platform "X". The dataset was processed using both algorithms with the aim of determining which algorithm has the best performance in classifying positive or negative sentiment. This research provides results that the highest accuracy of 90.72% with an F1-Score of 87.54% was produced by the Random Forest algorithm while an accuracy of 89.35% with an F1-Score of 85.15% was produced by the SVM algorithm. This study concludes that Random Forest outperforms SVM due to its ability to handle pre-processing variations more effectively. Random Forest is an ensemble method that works by building multiple decision trees and combining their predictions. This ensemble approach reduces overfitting and improves generalization ability, which makes Random Forest algorithm superior to SVM method which is more sensitive to data scaling and kernel selection.

E. References

[1] A. E. Setyono and B. F. T. Kiono, "Dari Energi Fosil Menuju Energi Terbarukan: Potret Kondisi Minyak dan Gas Bumi Indonesia Tahun 2020 – 2050," *Jurnal*

Energi Baru dan Terbarukan, vol. 2, no. 3, pp. 154–162, Oct. 2021, doi: 10.14710/jebt.2021.11157.
[2] H. Khan, I. Khan, and T. T. Binh, "The heterogeneity of renewable energy consumption, carbon emission and financial development in the globe: A panel quantile regression approach," *Energy Reports*, vol. 6, pp. 859–867, Nov. 2020, doi: 10.1016/j.egyr.2020.04.002.
[3] V. Tulus and P. Sidabutar, "Kajian pengembangan kendaraan listrik di Indonesia: prospek dan hambatannya," *Jurnal Paradigma Ekonomika*, vol. 15, no. 1, pp. 2085–1960, Jan. 2020.
[4] N. B. Sidauruk, N. Riza, R. Nuraini, and S. Fatonah, "Penggunaan Metode SVM DAN Random Forest Untuk Analisis Sentimen Ulasan Pengguna Terhadap KAI Access Di Google Playstore," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 3, pp. 1901–1906, Jun. 2023.
[5] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," vol. 4, no. 1, pp. 96–106, 2023, doi: 10.33365/jtsi.v4i1.2445.
[6] T. T. Thet, J. C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J Inf Sci*, vol. 36, no. 6, pp. 823–848, Dec. 2023, doi: 10.1177/0165551510388123.
[7] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
[8] K. X. Han, W. Chien, C. C. Chiu, and Y. T. Cheng, "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset," *Applied Sciences (Switzerland)*, vol. 10, no. 3, Feb. 2020, doi: 10.3390/app10031125.
[9] B. Irawan and O. Nurdiawan, "Analisis Sentimen Terhadap Pengguna Gojek Dan Grab Pada Media Sosial Twitter Menggunakan Random Forest," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 5, Oct. 2023.
[10] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," vol. 6, no. 9, pp. 4305–4313, Sep. 2022, [Online]. Available: http://j-ptiik.ub.ac.id
[11] H. Chyntia Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," vol. 6, no. 4, pp. 1702–1708, Apr. 2022, [Online]. Available: http://j-ptiik.ub.ac.id
[12] M. Chen, Z. Zhang, J. Shen, Z. Deng, J. He, and S. Huang, "A Quantitative Investment Model Based on Random Forest and Sentiment Analysis," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1575/1/012083.
[13] S. Khomsah, "Sentiment Analysis On YouTube Comments Using Word2Vec and Random Forest Sentimen Analisis pada Opini YouTube Menggunakan Word2Vec dan Random Forest," *Jurnal Informatika dan Teknologi Informasi*, vol. 18, no. 1, pp. 61–72, 2021, doi: 10.31515/telematika.v18i1.4493.
[14] N. Fitriyah, B. Warsito, D. Asih, and I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM),"

JURNAL GAUSSIAN, vol. 9, no. 3, pp. 376–390, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[15] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27–31, Mar. 2018, Accessed: Jan. 04, 2025. [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/IJAIDM/index