

NAÏVE BAYES AND SUPPORT VECTOR MACHINE BASED ON OPTIMIZATION FOR PUBLIC SENTIMENT ANALYSIS POST-2024 ELECTION

Fari Katul Fikriah^{1*}, Amelia Devi Putri Ariyanto²

^{1,2}Information System and Technology Study Program, Widya Husada Semarang University
Email: ¹farichatulfikriyah45@gmail.com, ²ameliadev@gmail.com

(Received: 20 June 2025, Revised: 02 July 2025, Accepted: 24 July 2025)

Abstract

The 2024 election has sparked an explosion of public opinion across various digital platforms, but the complexity and large volume of data make it difficult for policymakers to understand public sentiment promptly. Therefore, an accurate and efficient sentiment analysis method is needed to automatically classify public opinion. This study aims to analyze and compare the performance of the Naïve Bayes algorithm and an optimized Support Vector Machine (SVM) in classifying post-election public sentiment. The research method includes collecting 10,000 text data entries from various data sources, conducting text preprocessing, extracting features using the TF-IDF method, applying both algorithms with parameter tuning, and generating their performance using accuracy, precision, recall, and F1 score metrics. The results show that the optimized SVM algorithm delivers superior performance, achieving 88.24% accuracy, compared to 82.35% for Naïve Bayes. These findings indicate that SVM is more effective in handling complex public opinion sentiment classification, thus serving as a valuable reference for post-election policymaking.

Keywords: *Sentiment Analysis, Naïve Bayes, Support Vector Machine, TF-IDF, 2024 Election*

This is an open access article under the [CC BY](#) license.



**Corresponding Author: Fari Katul Fikriah*

1. INTRODUCTION

The 2024 General Election in Indonesia is an important moment that will not only determine the direction of national policy for the next five years but also trigger massive public opinion dynamics, especially in the digital space. Social media is the main medium for people to voice their opinions, hopes, and disappointments regarding the election process and results. Indonesia is a democratic country, with a system of government in which political power is held collectively by the people, and the democratic system is characterized by periodic general elections. [1]. To achieve a democratic nation, the general election mechanism is regarded as a means to realize this goal in alignment with Pancasila and the 1945 Constitution of the Republic of Indonesia. [2][3].

In practice, public enthusiasm for voting and voicing opinions through digital media is an important indicator of political participation. However, the emergence of various accompanying narratives, from support to criticism to disinformation, makes it difficult for policymakers to fully understand public aspirations. The emerging challenge is how to process

the explosion of public opinion data quickly and accurately to present a true picture of the public's collective emotions after the election. Such knowledge is crucial in the context of maintaining social stability and improving the quality of data-driven democracy.

In this context, sentiment analysis is one of the relevant approaches to comprehensively and data-drivenly capture public perception and emotions. The main challenge lies in processing large volumes of text data efficiently and accurately to gain a deep understanding of public responses after the election. [4]. Machine learning-based sentiment analysis techniques have proven effective in classifying public opinion into positive, negative, or neutral categories. Among the most commonly used methods are Naïve Bayes and Support Vector Machines (SVM) [5] [6]. Naïve Bayes is referred to as one of the commonly used classification algorithms, and it can be said that this algorithm is also the simplest. Naïve Bayes can estimate the probability of a class based on the distribution of words in a document. Naïve Bayes is often used by researchers because it has several advantages, including high accuracy, speed, and efficiency [7]. Meanwhile, the Support Vector

Machine (SVM) algorithm is a type of classification method that utilizes a kernel to map data into a higher-dimensional space, allowing non-linearly separable data to be classified [8] [9].

Although both algorithms have been widely used in previous research, comparative studies focused on the context of elections in Indonesia—particularly those applying optimization techniques to improve classification performance—are still very limited. The main questions to be addressed in this study are: which algorithm is more capable of accurately capturing the nuances of public sentiment in the post-2024 election context, and how parameter optimization techniques can improve the performance of both algorithms in terms of accuracy, precision, and processing efficiency. Furthermore, how feature extraction strategies can affect the final performance of the model is a crucial issue that also needs to be addressed. According to Putri Nurrahmah Wear et al [10]. Explaining the results of the Emilu often give rise to pros and cons in the eyes of the public, in their research using the Naive Bayes method, with the results obtained, namely with the Naive Bayes model, the number of negative comments was assessed as more than positive comments with the highest level of accuracy on Tiktok, namely 68% and the highest accuracy on X, namely 83%, so it can be concluded that public sentiment towards the 2024 Presidential Election Results on social media Tiktok and X contains more negative comments than positive comments.

This study aims to explore and compare the performance of Naïve Bayes and SVM in analyzing public sentiment after the 2024 election, using a parameter optimization approach to achieve more optimal classification results. The optimization technique applied in this study focuses on improving classification accuracy through optimal parameter selection and enhancements in the feature extraction process. The benefits of this study include providing a scientific and practical foundation for policymakers, academics, and information technology practitioners to better understand public sentiment patterns, particularly in a dynamic socio-political context such as the post-election period. Additionally, this study is expected to enrich the literature and advance the development of machine learning-based sentiment analysis models through the integration of effective optimization techniques.

2. RESEARCH METHOD

This study uses a quantitative approach with an experimental method to compare the performance of the Naive Bayes and Support Vector Machine (SVM) algorithms in analyzing public sentiment after the 2024 election. The main data consists of public opinion posts on social media, especially from the Twitter/X platform, which are collected using web scraping techniques based on certain keywords relevant to election topics, such as candidate names,

political parties, and popular issues that emerged during and after the election. The steps are as follows:

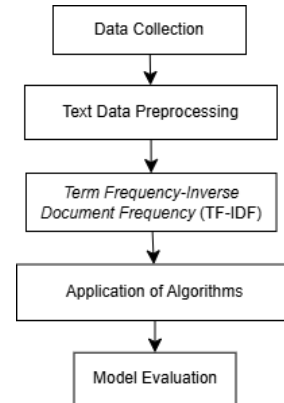


Figure 1. Research Steps

2.1 Data Collection

In this study, data were collected from social media using the Twitter/X API (Application Programming Interface) and crawling tools. The data collection period was determined to be from one week before to one month after the announcement of the official results of the 2024 election. Data were collected through a data crawling method using the Python programming language.

2.2 Text Data Preprocessing

Data obtained through crawling still contains a lot of noise. Therefore, preprocessing is carried out, which includes:

- Data cleaning: removing URL, emojis, numbers, punctuation, and special characters.
- Tokenization: breaking sentences into words.
- Stopword removal: eliminating common words that do not carry high informational value.

2.3 Feature Extraction TF-IDF

In this step, text features are converted into numeric form using the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which represents the importance weight of a word in a document. The processed text is transformed into a numerical representation by calculating the weight of each word based on its frequency in the document and how common the word is across the entire document collection.

2.4 Application Algorithms

This study uses two algorithms, namely Naïve Bayes and Support Vector Machine (SVM). The Naive Bayes algorithm uses a combination of values from the dataset and frequency to determine the probability [11]. This algorithm applies Bayes' theorem, which assumes that no attribute of the dataset depends on the value of the class variable. The Naive Bayes algorithm is formulated as follows:

$$P(x) = \frac{P(H)P(H)}{P(x)} \quad (1)$$

The naive Bayes algorithm, as shown in the image above, starts with the input dataset and then divides it into two parts: training data and testing data. training data is processed by calculating the probability of the testing data to produce a decision [12] [4]. The Support Vector Machine (SVM) algorithm is a classification method that can handle both linear and non-linear data. SVM aims to divide the dataset into classes to find the optimal hyperplane to separate classes in the data with a maximum margin [6].

2.5 Model Evaluation

During the model evaluation phase, this research seeks to assess how well the optimized Naïve Bayes and Support Vector Machine (SVM) algorithms perform in categorizing public sentiment following the 2024 Election. The assessment is carried out by comparing the model's predictions with the actual labels in the test dataset. The evaluation utilizes various performance metrics, including accuracy, precision, recall, and the F1-score. Accuracy reflects the proportion of correctly classified instances. Precision and recall evaluate how relevant and complete the classification outcomes are, while the F1-score, as the harmonic mean of precision and recall, provides a balanced measurement, especially when dealing with imbalanced data classes [13].

In addition to these metrics, the confusion matrix is examined to illustrate the detailed distribution of true and false predictions for each sentiment category (positive, negative, and neutral). The evaluation findings are further analyzed to determine how much the applied optimization enhances the performance of the Naïve Bayes and SVM algorithms compared to their unoptimized versions. Consequently, this evaluation stage forms the basis for concluding the model's effectiveness in accurately identifying public sentiment, thereby supporting decision-making and sentiment analysis in the post-election context.

3. RESULT AND DISCUSSION

4.1 Data Collection

The data used in this study were collected from Twitter (now X), discussion forums, and online news comment sections discussing the political situation after the 2024 Election. A total of 10,000 raw text entries were collected. After undergoing cleaning and validation, 8,500 entries were deemed suitable for analysis, with the distribution of sentiment classes shown in Table 1.

Table 1. Sentiment Data Distribution

Sentiment	Amount of Data	Percentage
positive	3400	40%
negative	3060	36%
netral	2040	24%

4.2 Text Data Processing and TF-IDF Results

In the text preprocessing stage, data cleaning is performed by removing punctuation, eliminating

stopwords, and normalizing words. After that, feature extraction is conducted using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The vectorization results show that the words with the highest weights are related to political issues, candidate names, policies, and public opinions. As an illustration, Table 2 presents the 10 words with the highest TF-IDF scores.

Table 2. TF-IDF Score

Text	TF-IDF Score
demokrasi	0.083
protes	0.072
pemimpin	0.072
netralisasi	0.067
kebijakan	0.066
korupsi	0.070

4.3 Application Algorithm

After the data collection, text preprocessing, and feature extraction using the TF-IDF method, the next stage is the application of classification algorithms to analyze public sentiment. This study employs two popular algorithms, namely Naïve Bayes and Support Vector Machine (SVM), both of which are optimized to enhance classification performance.

a. Naive Bayes

Naïve Bayes is applied as the baseline algorithm because of its strong ability to handle text data with high computational speed. The Naïve Bayes model works by calculating the probability that a piece of text belongs to one of the sentiment categories (positive, negative, or neutral) based on the distribution of words computed from the training data. At this stage, parameter tuning is performed to select the Naïve Bayes variant that best matches the characteristics of the data [14]. After testing several options, the Multinomial Naïve Bayes algorithm was chosen because it proved most effective in handling TF-IDF data. The model was trained on 80% of the dataset with a balanced distribution among the classes. The training process was fast, with an average computation time of less than 5 seconds for the entire training data.

The results show that Naïve Bayes is able to recognize common word patterns that indicate sentiment, especially in texts with explicit words such as *bagus*, *rusak*, *adil*, or *curang*. However, for ambiguous or sarcastic texts, the model tends to show decreased accuracy due to the limitation of the assumption of word independence.

b. Support Vector Machine (SVM)

Unlike Naïve Bayes, the SVM algorithm is applied using a supervised learning approach that builds an optimal hyperplane to separate data into different classes. In this study, the application of SVM focuses on the linear kernel, as experimental results indicate that the linear kernel is more stable than the RBF kernel for high-dimensional data such as TF-IDF vectors.

Additionally, optimization of the C and gamma parameters is performed to minimize classification errors and avoid overfitting. The parameter tuning

process uses the grid search cross-validation method with 5-fold cross-validation on the training data. The tuning results show that a C value of 1.0 and the linear kernel produce the best accuracy during validation. The SVM training time is relatively longer than that of Naïve Bayes, taking approximately 12 seconds for the entire training data. However, SVM is capable of distinguishing more nuanced sentiment patterns, especially in sentences with complex word structures and contextual meanings [15].

4. 4 Model Evaluation

Based on the test results, SVM demonstrated significant superiority in terms of accuracy, precision, recall, and F1-score. This is due to SVM's ability to build optimal separation margins even when the data have high-dimensional features. To ensure that the classification results from the Naïve Bayes and Support Vector Machine (SVM) models are not overfitting and have good generalization, a validation process was performed using the K-Fold Cross Validation technique. In this study, a K value of 10 (10-Fold Cross Validation) was used, which is a common practice in classification model validation to obtain more stable performance estimates. The K-Fold Cross Validation process divides the dataset into several equal parts. Each fold is used once as both test data and training data. The average of the results from each iteration is calculated to determine the final model performance.

The following table shows the average accuracy, precision, recall, and F1-score of each model after validation. Table 3 shows the evaluation metrics that show the performance differences between the two algorithms:

Metric	Naïve Bayes (%)	SVM (%)
Accuracy	82.35	88.24
precision	81.40	87.5
Recall	80.75	88
F1-Score	81.07	87.75

From the results above, it can be concluded that the SVM model with parameter optimization successfully provides better sentiment classification performance than Naïve Bayes, although it requires more computation time. This result is consistent with the findings of several previous studies, which state that SVM has advantages in text classification for data with complex public opinion contexts, such as post-election political sentiment.

4. CONCLUSION

The 2024 Election has generated various public opinions and sentiments that are widely disseminated through different digital media. Sentiment analysis is a crucial approach for capturing public perceptions quickly and accurately, thereby enabling policymakers to understand social dynamics following the election.

This study aims to compare the performance of the Naïve Bayes algorithm and the Support Vector

Machine (SVM) with parameter optimization in classifying public sentiment. The methods employed include data collection from online platforms, text preprocessing for data cleaning and normalization, feature extraction using TF-IDF, application of both algorithms with parameter tuning, and model performance evaluation using accuracy, precision, recall, and F1-score metrics.

The results indicate that the SVM algorithm with parameter optimization delivers better performance than Naïve Bayes. The SVM model achieved an accuracy of 88.24%, whereas Naïve Bayes attained an accuracy of 82.35%. Additionally, SVM outperformed Naïve Bayes in terms of precision, recall, and F1-score. This difference demonstrates that SVM is more effective in handling sentiment text data with complex opinion contexts, although it requires relatively more computation time. Therefore, it can be concluded that the application of an optimization-based SVM algorithm is more recommended for post-election public sentiment analysis, as it provides more accurate and reliable classification results to support strategic decision-making in the socio-political domain.

5. REFERENCE

- [1] L. Damayanti and K. M. Lhaksmana, "Sentiment Analysis of the 2024 Indonesia Presidential Election on Twitter," *Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, 2024, doi: 10.33395/v8i2.13379.
- [2] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 746–757, Aug. 2023, doi: 10.26555/jiteki.v9i3.26490.
- [3] B. Bansal and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 346–353. doi: 10.1016/j.procs.2018.08.183.
- [4] N. Mirantika, R. Trisudarmo, and T. S. Syamfithriani, "Implementation of Naïve Bayes Algorithm for Early Detection of Stunting Risk," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] S. A. Salleh *et al.*, "Support Vector Machine (SVM) and Object Based Classification in Earth Linear Features Extraction: A Comparison," *Revue Internationale de Géomatique*, vol. 33, no. 1, pp. 183–199, 2024, doi: 10.32604/rig.2024.050723.
- [6] G. Ayu, P. Febriyanti, and A. Baita, "Comparison of Support Vector Machine and Decision Tree Algorithm Performance with Undersampling Approach in Predicting Heart Disease Based on Lifestyle," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [7] T. A. Amini and K. Setiawan, "Application of the Naive Bayes Algorithm in Twitter Sentiment

- Analysis of 2024 Vice Presidential Candidate Gibran Rakabuming Raka using Rapidminer,” *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 4, no. 1, pp. 234–246, Apr. 2024, doi: 10.35870/ijsecs.v4i1.2236.
- [8] R. Putra, Y. Yusra, and M. Fikry, “Penerapan Metode SVM pada Klasifikasi Sentimen terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024,” *Jurnal Informatika Universitas Pamulang*, vol. 8, no. 2, pp. 145–152, Jun. 2023, doi: 10.32493/informatika.v8i2.30355.
- [9] Muthia Sakhdiah, Admi Salma, Dony Permana, and Dina Fitria, “Sentiment Analysis Using Support Vector Machine (SVM) of ChatGPT Application Users in Play Store,” *UNP Journal of Statistics and Data Science*, vol. 2, no. 2, pp. 151–158, May 2024, doi: 10.24036/ujsds/vol2-iss2/158.
- [10] P. Nurrahmah Wear, N. bin Idris, I. Rosita, U. Mulia Jl Letjen Zaini Azhar Maulani No, B. Selatan, and K. Timur, “Analisis Sentimen Terhadap Hasil Pilpres 2024 Pada Aplikasi Tiktok Dan X Menggunakan Metode Naïve Bayes Classifier.”
- [11] O. Peretz, M. Koren, and O. Koren, “Naive Bayes classifier – An ensemble procedure for recall and precision enrichment,” *Eng Appl Artif Intell*, vol. 136, Oct. 2024, doi: 10.1016/j.engappai.2024.108972.
- [12] A. Z. Arrayyan, H. Setiawan, and K. T. Putra, “Naive Bayes for Diabetes Prediction: Developing a Classification Model for Risk Identification in Specific Populations,” *Semesta Teknika*, vol. 27, no. 1, pp. 28–36, Apr. 2024, doi: 10.18196/st.v27i1.21008.
- [13] T. H. Tanjung and M. Furqan, “Classification of Heart Disease Using Support Vector Machine,” *sinkron*, vol. 8, no. 3, pp. 1803–1812, Jul. 2024, doi: 10.33395/sinkron.v8i3.13904.
- [14] G. Forda Nama, A. H. Shalihah, B. Wintoro, Y. Mulyani, and D. Despa, “Implementation of Naïve Bayes Classifier & Support Vector Machine Algorithm for Sentiment Classification using Twitter Data on Indonesian Presidential Candidates In 2024,” 2024. [Online]. Available: <https://www.jisem-journal.com/>
- [15] D. Srivastava, L. Bhambhu, and D. K. Srivastava, “Data classification using support vector machine,” 2010. [Online]. Available: www.jatit.org