



Sentiment Analysis of the 2025 Budget Cut Policy using the SVM Method

Tomas Adi Tadu^{1*}, Fajar Hariadi²

^{1,2}Informatics Engineering Study Program, Wira Wacana Sumba Christian University
tomasaditadu@gmail.com^{1*}, fajar@unkriswina.ac.id²

Abstract

Budget cuts are a government policy to reduce the allocation of funds to various sectors or specific programs to align with state spending priorities. President Prabowo Subianto's efforts to streamline the budget remain controversial. The budget cuts are outlined in Presidential Instruction Number 1 of 2025. In the 2025 Fiscal Year, the budget cuts reached IDR 256.1 trillion. This study aims to analyze the Indonesian public's response to the 2025 budget efficiency policy via Twitter. The data used in this analysis are tweets discussing the policy. The data will then be classified into three sentiment categories: Negative, Neutral, and Positive. The results of Sentiment Analysis using the Support Vector Machine (SVM) method show that the model with 799 training data and 200 test data achieved an accuracy of 60.00%. The Negative class has a recall of 0.65, a precision of 0.66, and an f1-score of 0.65. The Neutral class has a precision of 0.45, a recall of 0.60, and an f1-score of 0.51, while the Positive class has a recall of 0.54 and an f1-score of 0.61. The macro averages of precision, recall, and f1-score are 0.60, 0.60, and 0.59, respectively, indicating that the performance has not yet reached its maximum. From the sentiment distribution graph, Negative sentiment dominates with 42.54% (425 tweets), followed by Positive 32.43% (324 tweets), and Neutral 25.03% (250 tweets).

Keywords: Budget Cuts, Sentiment Analysis, Twitter, Classification, Support Vector Machine

1. Introduction

Advances in internet technology have brought about changes in various aspects of human life, including education, commerce, government, and social communications. In this modern era, the internet and social media have grown rapidly, enabling the dissemination of a wide range of information and opinions from many people through social media.

Social media has become an integral part of everyday life. Today, many people have a regular habit of accessing social media, either to see updates from other users or to share their own posts. In fact, one of the first things people do after waking up is access social media. Furthermore, social media also allows people to access the latest news and express their opinions [1].

Twitter is one of the world's most popular social media platforms. Its popularity in Indonesia is evident from the 24.69 million users in the country. On Twitter, users can share their daily activities, photos, and comments on posts, as well as share others'. Posts frequently shared and discussed on Twitter are called Trending Topics. One of the most frequently discussed topics on Twitter in February 2025 was budget cuts.

Budget cuts are a government policy of reducing the allocation of funds to various sectors or specific programs to align with state spending priorities. President Prabowo Subianto's efforts to streamline the budget have been widely debated. These budget cuts are outlined in Presidential Instruction Number 1 of 2025. Furthermore, the policy was confirmed through the Minister of Finance's Letter Number S-37/MK.02/2025, signed by Minister of Finance Sri Mulyani Indrawati, and applied to operational and non-operational spending across all Ministries/Institutions. In the 2025 Fiscal Year, the amount was Rp 256.1 trillion. This policy is debated, because the policy aimed at streamlining budget utilization is considered inconsistent with the political policies in the formation of the government cabinet.

In his political policies, President Prabowo attempted to accommodate all political interests. As a result, the cabinet structure he formed became "bloated," with a larger number of officials than in the previous administration. The Red and White Cabinet he formed included 7 Coordinating Ministers, 41 Ministers, 55 Deputy Ministers, and 5 Ministerial-level Officials. This has recently become increasingly controversial. The Presidential Instruction stated that not all state institutions would be affected by the efficiency policy. It turns out that 17 Ministries/Institutions were not affected by the 2025 budget cuts [2].

This study aims to analyze the Indonesian public's response to the 2025 budget efficiency policy via Twitter. The data used in this analysis consists of tweets discussing the policy. The data will then be classified into three sentiment categories: Negative, Neutral, and Positive.

The advantage of Negative is that it allows the SVM model to detect opinions that are complaints, criticisms, or dissatisfaction. The advantage of Neutral is that it can be used for unclear texts that tend to be positive or negative, such as informative or descriptive texts without opinions. Meanwhile, Positive SVM can be trained to recognize text patterns containing optimistic, satisfied, or supportive sentiments.

However, the method used in this research utilizes the *Support Vector Machine* (SVM) as a classification technique. SVMs are capable of producing effective classification models, even when trained with relatively small amounts of data. Furthermore, this method facilitates finding the optimal hyperplane to distinguish between two classes in the input space. The classification process in SVM is carried out by finding a line or hyperplane that maximally separates the two data groups.

The classification process using SVM is carried out by finding the hyperplane with the largest separation margin between two data groups, resulting in a more generalized model capable of handling new data well. In other words, this method works by identifying the most optimal separation boundary, so that each new data can be accurately classified into the appropriate category. SVM's reliability in handling small data sizes and its ability to separate classes with maximum margin make it one of the most widely used algorithms in various classification applications.

2. Literary Studies

2.1 Twitter

Twitter is one of several emerging communication platforms created to allow users to express their opinions, aspirations, criticisms, and exchange information about current topics, without the constraints of time and space. This allows for direct communication; posting opinions on Twitter is known as tweeting.

This makes Twitter a valuable source of text data that can be mined and utilized for various research purposes in the field of information technology. Twitter has prepared structured text (XML and a list of categories) that always contains username information, timestamps, text, retweets, favorites, and other information. This can be seen from the metadata implementation process in each tweet. Meanwhile, unstructured content is the part that does not have a specific structure, in the form of text seen in text form or called a tweet [3].

2.2 Budget Cuts

Budget cuts are a government policy of reducing funding allocations to various sectors or specific programs to align with state spending priorities. President Prabowo Subianto's efforts to streamline the budget remain highly debated. These budget cuts are outlined in Presidential Instruction Number 1 of 2025. The policy was further confirmed through the Minister of Finance's Letter Number S-37/MK.02/2025, signed by Minister of Finance Sri Mulyani Indrawati, and will be applied to operational and non-operational spending across all Ministries/Institutions. In the 2025 Fiscal Year, the amount will be Rp 256.1 trillion [2].

2.3 Text Mining

Text mining and data mining are often considered similar disciplines because they both use the same algorithms. However, there are fundamental differences between the two: data mining focuses on structured data, while text mining deals with text that has special characteristics, is unstructured, and requires preprocessing. Furthermore, text mining is closely related to Natural Language Processing (NLP) [4].

2.4 Sentiment Analysis

Sentiment analysis is a branch of data mining that aims to process and extract information from textual data, particularly those containing opinions, evaluations, attitudes, emotions, assessments, and individual sentiments toward a product, figure, organization, or issue. In this study, the process of determining sentiment labels for responses was carried out by counting the number of positive and negative words contained within them. If the number of positive words is more dominant, then the response is categorized as positive sentiment. Conversely, if the number of negative words is greater, then the response is classified as negative sentiment [5].

Sentiment analysis aims to identify and determine the polarity of text or opinions within a document, whether they are positive or negative. In this study, Twitter was chosen as the data source because it is one of the most widely used platforms by internet users today, providing a variety of opinions that can be analyzed [6].

2.5 Lexicon

The lexicon-based labeling method in sentiment analysis is an approach that utilizes an Indonesian sentiment lexicon to classify opinions into positive, negative, or neutral categories. This lexicon contains words that have been scored based on sentiment polarity. The labeling process is performed automatically by matching words in the text to entries in the dictionary. In the context of local journals, this method is often used as the initial stage for building training data for machine learning models. For example, research [7] combine this method with Support Vector Machine to improve classification accuracy, while [8] utilized the InSet Lexicon in analyzing local app user reviews. While this method is efficient and does not require manual annotation, its main drawback is its reliance on the completeness and relevance of the vocabulary in the lexicon.

2.6 Classification

According to [9] Classification is the process of finding a model or function capable of describing and distinguishing between different classes of data. The primary goal of this process is to enable the model to predict the class of objects whose class is unknown. The model is built based on the analysis of data that already has class labels.

2.7 Support Vector Machine

Support Vector Machine (SVM) is a superior method compared to other techniques due to its ability to process high-dimensional data. This capability allows SVM to achieve optimal accuracy in the classification process, making it frequently used in various applications requiring complex data analysis [10].

3. Research Method

The data used in this study is qualitative data derived from user posts or tweets on the social media platform Twitter. These tweets contain various opinions, views, or statements that can be analyzed to understand specific sentiments or patterns in digital communication. Twitter was chosen as the data source because of its dynamic nature, allowing users to express opinions openly in short text form.

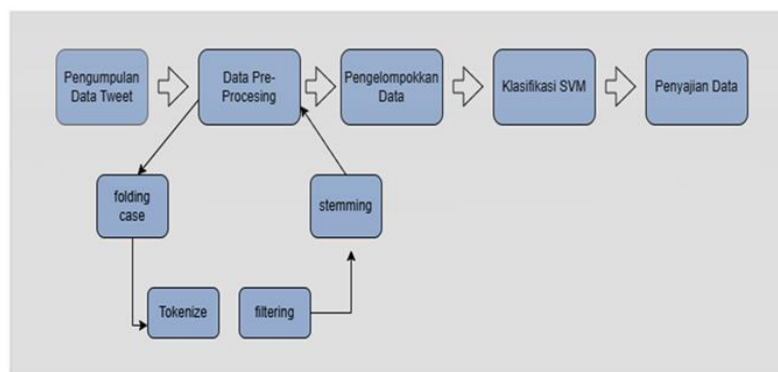


Figure 1: Research Flow

3.1. Data Collection

In this first process, data and information that can be used in this research are collected. The crawling data source itself is social media X. The data taken are tweets related to using the keyword "Budget Cuts" posted by social media users X, from February 2025 to March 2025 with a total of 1000 data, and crawled using the Google Colab platform.

3.2. Data Pre-Processing

This process is carried out using the Python programming language on the Google Colab platform, using the sncrape library for data collection. Before analyzing the data using the Support Vector Machine (SVM) method, a preprocessing stage is first performed. At this stage, tweets that have been successfully collected through the crawling process are further processed to remove irrelevant data. After that, a series of steps are carried out, such as case folding (converting letters into a uniform format), tokenization (breaking text into words), filtering (filtering out unnecessary words), and stemming (converting words to their basic form).

The following is an explanation of the Pre-processing stage:

- a. Transform Case or Folding Case Stage
The Transform Case or Folding Case stage is the process of converting all letters in the dataset to lowercase and removing all punctuation marks contained in the data. This step aims to align the text format so as to facilitate the further analysis process.
- b. Tokenization or Token Filter Stage
The Tokenization or Filter Tokens stage is the process of separating text in a dataset into individual words based on detected spaces. This step aims to simplify analysis by identifying each word as a separate unit.
- c. Filtering Stage
The filtering stage is the process of removing irrelevant or meaningless words from the dataset. This step is carried out to ensure that only significant words are used in the analysis.
- d. Stemming Stage
The stemming stage is the process of converting each word in the dataset to its base form. This step aims to simplify words by removing affixes so that analysis can be performed more effectively.

3.3. Data Grouping

At this stage, the obtained data will be categorized or labeled into three groups: Negative, Neutral, and Positive, using the lexicon method. The data will then be presented in tabular form to facilitate analysis. An example of the labeling process is shown below. After labeling, the dataset will be divided into 1,000 data points into training and test data, with a ratio of 70% training data and 30% test data.

- a. Training Data is data that will be used as a tool for the model knowledge base.
- b. Test Data is data that will be used to test a model that has been trained using training data.

3.4. SVM Classification

Before the classification process, a TF-IDF search is performed to determine the vector representation value of the data document. A Support Vector Machine (SVM) determines the position of the hyperplane using support vectors and margin values to achieve optimal separation. In this study, the data used as input is represented in vector form, obtained through a word weighting process. Next, the SVM model will go through a training phase, where the system learns from the training data to generate patterns or parameters that can be used in the testing process. This testing phase aims to classify new tweets and assign appropriate sentiment labels.

3.5. Data Presentation

After obtaining the percentage results from the classification process, the data will be presented in graphs or bar charts. This presentation aims to facilitate visualization of the percentage differences between the Negative, Neutral, and Positive categories, making them easier to analyze and understand.

4. Results And Discussion

4.1. Data Collection

Data collection process, data obtained from the results of crawling tweet X related to the topic of Budget Cuts from February 2025 to March 2025. The data obtained amounted to 1000 data, the tweet crawling process used python and node.js code using the opensource tweet harvest tool. The tweet data that was successfully collected can be seen in Table 1.

Table 1: *Crawling Results*

No.	<i>Tweet</i>
1.	@budimandjatmiko For number one, I don't agree....bud get cuts everywhere, layoffs are going crazy, it's hard to find work.
2.	@ARSIPAJA From various efficiency cuts... only the DPR/DPRD budget is safe
3.	@dinamo_udi81001 @Jatosint @prabowo @EmmanuelMacron @Elysee The layoffs of many honorary employees at RRI/TVRI and ministries occurred as a result of budget cuts, which is not part of the diplomacy program.
...	...
999.	To fund priority programs, President Prabowo has implemented \$19 billion in budget cuts, allocating more funds to social programs and economic development. https://t.co/6ANryKF2XB
1000.	President Prabowo has directed \$19 billion in budget cuts to be redirected to free meal programs and economic stimulus. https://t.co/iobqhJ711z

4.2. Data Pre-Processing

In this stage, the data will then be prepared before being used in the classification process using the SVM method. The pre-processing processes used in this study are Transform Case or Folding Case, Tokenization or Filter Tokens, filtering, and *Stemming*.

- a) *Folding Case*: This process converts all letters in the dataset to lowercase and removes all punctuation. The results of implementing Folding Case on the dataset can be seen in Table 2.

Table 2: *Folding Case Results*

No.	<i>After Folding Case</i>
1.	for number one, i don't agree with budget cuts everywhere, layoffs are going crazy, it's hard to find work

2. Of the various efficiency cuts, only the DPRD budget is safe.
3. The layoff of many honorary employees at RRI and the ministry occurred as a result of the budget cuts, which were not part of the diplomatic program.
- ...
999. To fund priority programs, President Prabowo cut the budget by billions and allocated more funds for social programs and economic development.
1000. President Prabowo directed billions of rupiah in budget cuts to be diverted to free meal programs and economic stimulus.

- b) *Tokenization or Filter Tokens*: The process of separating text in a dataset into individual words based on detected spaces. The results of implementing Tokenization or Filter Tokens on the dataset can be seen in Table 3.

Table 3: Tokenization Results

No.	After Tokenization
1.	['for', 'number', 'one', 'no', 'agree to cuts', 'budget', 'where', 'where', 'layoff', 'crazy', 'sck', 'difficult', 'looking for', 'work']
2.	['from', 'various', 'pruning', 'efficiency', 'only', 'budget', 'dprdprd', 'only', 'the', 'safe']
3.	['phk', 'many', 'employees', 'honorary', 'in', 'rritvri', 'and', 'ministry', 'happens', 'as', 'impact', 'cuts', 'budget', 'that', 'not', 'part', 'of', 'program', 'diplomacy']
...	...
999.	['to', 'fund', 'program', 'priority', 'president', 'prabowo', 'do', 'cut', 'budget', 'amount', 'billion', 'allocate', 'fund', 'more', 'big', 'share', 'program', 'social', 'amp', 'development', 'economy']
1000.	['president', 'prabowo', 'direct', 'cut', 'budget', 'billion', 'for', 'diverted', 'to', 'program', 'meal', 'free', 'amp', 'stimulus', 'economy']

- c) *Filtering*: The process of removing irrelevant or meaningless words from a dataset. The results of implementing tokenization or filtering on the dataset can be seen in Table 4.

Table 4: Filtering Results

No.	After Filtering
1.	['number', 'agree to cuts', 'budget', 'where', 'layoff', 'crazy', 'sck', 'difficult', 'looking for', 'work']
2.	['pruning', 'efficiency', 'budget', 'dprdprd', 'safe']
3.	['phk', 'employee', 'honorary', 'rritvri', 'ministry', 'impact', 'cuts', 'budget', 'program', 'diplomacy']
...	...
999.	['fund', 'program', 'priority', 'president', 'prabowo', 'cut', 'budget', 'billion', 'allocate', 'fund', 'program', 'social', 'amp', 'development', 'economy']
1000.	['president', 'prabowo', 'direct', 'cut', 'budget', 'billion', 'diverted', 'program', 'meal', 'free', 'amp', 'stimulus', 'economy']

- d) *Stemming*: The process of converting each word in a dataset to its base form. This step aims to simplify words by removing affixes so that analysis can be performed more effectively. The results of implementing Tokenization or Filtering on the dataset can be seen in Table 5.

Table 5: Stemming Results

No.	Stemming
1.	number agreed to cut the budget where the layoffs are crazy sck hard to find work
2.	reduce the efficiency of the DPRD's fencing safely

3.	layoffs of honorary Rritvri ministerial employees as a result of cutting diplomatic program budgets
...	...
999.	President Prabowo's priority program funds have been cut by billions in budget allocation for social programs and economic development.
1000.	President Prabowo aims to reduce the billion-dollar budget for the free meal and economic stimulus program

4.3. Data Grouping

In this process, labeling will be carried out on the dataset using the lexicon method, where each data will be determined to be included in the positive, negative or neutral sentiment category, the results of labeling using the lexicon method on the dataset can be seen in Table 6.

Table 6: Lexicon Labeling Results

No.	Tweet	Sentiment
1.	number agrees to budget cuts where crazy layoffs sck hard to find work	Neutral
2.	reduce the efficiency of the DPRD's fencing safely	Neutral
3.	layoffs of honorary Rritvri ministerial employees as a result of cutting diplomatic program budgets	Negative
...
999.	President Prabowo's priority program funds have been cut by billions in budget allocation for social programs and economic development.	Positive
1000.	President Prabowo aims to reduce the billion-dollar budget for the free meal and economic stimulus program	Positive

4.4. SVM Classification

In this section, the dataset that has gone through the pre-processing and grouping process will then be used to develop a model by finding a dividing line or hyperplane to separate each class, whether positive, negative, or neutral. The division of training data and test data is 79.8% test data and 20.02% training data, or from 1000 data, 799 data will be used as training data and 200 data will be used as test data. The results of the confusion matrix for 300 test data can be seen in Figure 2.

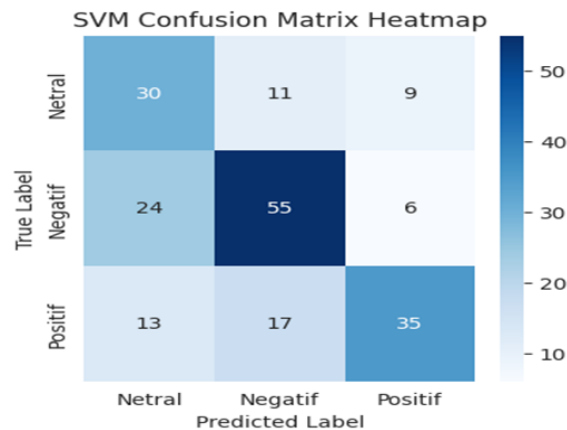


Figure 2: Confusion Matrix Results

The confusion matrix provides a comprehensive overview of the classification model's performance in classifying data. For data actually labeled Positive, there were 30 data points that were correctly classified by the model as Positive. However, the model predicted 11 data points that should have been Positive as Negative, and 9 others as Neutral. For data labeled Negative, the model correctly classified 55 data points as Negative, but there were still 24 data points classified as Positive and 6 others as Neutral. Meanwhile, for data labeled Neutral, the model only correctly classified 35 data points as Neutral, there were 17 data points incorrectly classified as Negative and 4 others as Positive. Of the total 799 test data points, 200 data points were correctly classified. Figure 3 shows the results of the precision, accuracy, recall, f1-value, and macro-average of the model:

b. Negative Sentiment Wordcloud

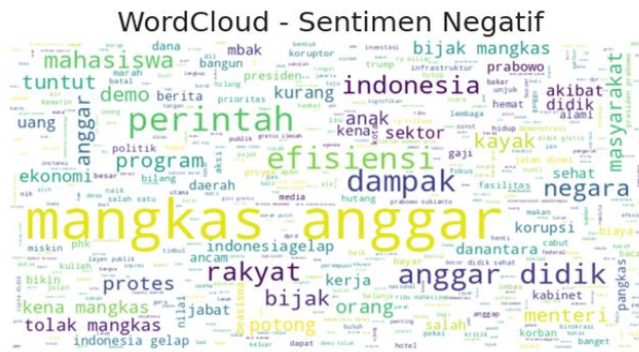


Figure 6: Negative Sentiment Wordcloud

c. Neutral Sentiment Wordcloud

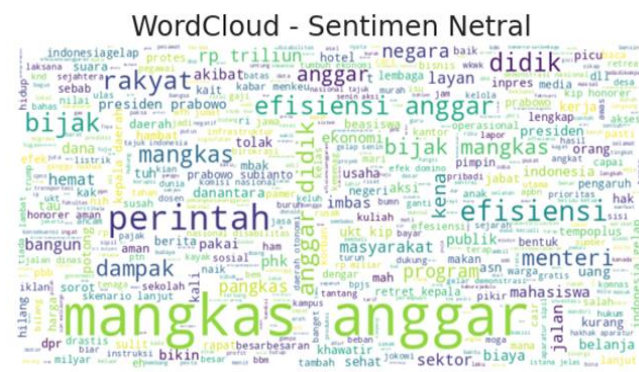


Figure 7: Neutral Sentiment Wordcloud

Figures 5, 6, and 7 above show the results of the word cloud visualization, which is used to illustrate the most frequently occurring words in each sentiment label. The size of the word in the word cloud represents its frequency of occurrence; the larger the word, the more frequently it appears.

5. Conclusion

The results of Sentiment Analysis using the Support Vector Machine (SVM) method show that the model with 799 training data and 200 test data achieved an accuracy of 60.00%. The Negative class has a recall of 0.65, a precision of 0.66, and an f1-score of 0.65. The Neutral class has a precision of 0.45, a recall of 0.60, and an f1-score of 0.51, while the Positive class has a recall of 0.54 and an f1-score of 0.61. The macro averages of precision, recall, and f1-score are 0.60, 0.60, and 0.59, respectively, indicating that the performance has not yet reached its maximum. From the sentiment distribution graph, Negative sentiment dominates with 42.54% (425 tweets), followed by Positive 32.43% (324 tweets), and Neutral 25.03% (250 tweets).

Reference

- [1] U. Para, F. Hariadi, and A. C. Talakua, "Sentiment Analysis of Capture Marriage Tradition in Sumba Region Using SVM Method," *J. Minfo Polgan*, vol. 12, no. 1, pp. 84–95, 2023, doi: 10.33395/jmp.v12i1.12320.
- [2] Rony Arianto Nugroho, "Budget Cuts." Kompas, Jakarta, 2025. [Online]. Available: <https://www.kompas.id/artikel/What-are-the-impacts-of-budget-cuts-when-bureaucratic-integrity-plummets?>
- [3] H. Eni Tri and S. Ari, "Sentiment Analysis of Public Response to Daily Covid-19 News on the Ministry of Health's Twitter Account," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 32–37, 2021, [Online]. Available: <http://repository.teknokrat.ac.id/3224/>
- [4] F. Fathonah and A. Herliana, "Application of Text Mining Sentiment Analysis Regarding the Covid-19 Vaccine Using the Naïve Bayes Method," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [5] F. S. Pamungkas and I. Kharisudin, "Sentiment Analysis with SVM, NAIVE BAYES and KNN for the Study of Indonesian Public Responses to the Covid-19 Pandemic on Twitter Social Media," *Pros. Semin. Nas. Mat.*, vol. 4, pp. 1–7, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/45038>
- [6] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [7] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 8, no. 1, pp. 59–71, 2022, doi: 10.23917/khif.v8i1.15213.
- [8] R. Damanhuri and V. A. Husein, "Sentiment Analysis of Indonesian-Language Access by KAI Application Reviews Using Word-Embedding and Classical Machine Learning," vol. 15, no. September, 2024, doi: 10.14710/jmasif.15.2.62383.
- [9] D. Bayusetyo, R. Santoso, and Tarno, "Classification of Prospective Blood Donors Using the Naive Bayes Classifier Method," *J. Gaussian*, vol. 6, no. 2, pp. 193–200, 2017, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/16948/16246>
- [10] P. Arsi and R. Waluyo, "Sentiment Analysis of the Discourse on Relocating Indonesia's Capital City Using the Support Vector Machine (SVM) Algorithm," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jiik.0813944.