VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 8 No. 1 (2026), 83-95

DOI: 10.35580/variansiunm512

APPLICATION OF SVM FOR SENTIMENT ANALYSIS REGARDING THE EFFICIENCY OF APBN AND APBD IN 2025

Nabilah Nursya'bani, Ruliana*, Muhammad Kasim Aidid

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: *Budget Efficiency*, *Sentiment Analysis*, SVM.

Abstract:

The policy on expenditure efficiency in the 2025 APBN and APBD has triggered diverse public responses on social media, necessitating sentiment analysis to identify emerging opinion trends. The analysis employs the Support Vector Machine (SVM) method, a margin-based classification algorithm that constructs an optimal separation between classes through the identification of the best hyperplane, where optimality is achieved when the separating margin is maximized. This study aims to identify sentiment patterns and classify public opinion regarding the budget efficiency policy to provide a measurable quantitative overview beyond subjective assessment. Data were collected from the X platform during the period 15 January–25 March 2025 using the keyword "efisiensi anggaran." The results indicate that negative sentiment dominates at 53%, while positive sentiment accounts for 47%. The SVM model achieved an accuracy of 99%, indicating strong performance in classifying sentiment related to the 2025 budget efficiency policy.

1. Introduction

Social media has become a primary space for the public to express opinions on various public issues, including government policies. The platform X (Twitter) enables users to share their views rapidly and in real time, generating large volumes of opinion data. One policy that has attracted public attention is Presidential Instruction No. 1 of 2025 concerning the efficiency of expenditure in the implementation of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year. This policy aims to improve the effectiveness of public budget management, yet it has also generated diverse public responses on social media. The large volume of unstructured opinion data therefore requires a computational approach to systematically identify patterns of public sentiment.

One approach that can be used to analyze opinions in textual data is Text Mining, which focuses on extracting meaningful information from large collections of unstructured documents. This technique enables the identification of patterns, topics, and opinion tendencies contained in textual data (Anwar, 2022). In its development, text mining is often integrated with machine learning methods to improve the automated classification of text data. One of the algorithms widely used for document classification is Support Vector Machine (SVM), which works by constructing an optimal hyperplane that separates data into different classes with the maximum margin, thereby providing strong generalization capability in classification tasks (Aisah et al., 2024). In text analysis, feature representation is commonly performed using Term Frequency–Inverse Document Frequency (TF-IDF), which assigns weights to words based on their level of importance within a document relative to the entire corpus.

* Corresponding author.
E-mail address: ruliana.t@unm.ac.id

Previous studies have demonstrated that Support Vector Machine (SVM) exhibits strong performance in sentiment analysis of social media data. TikTok sentiment analysis conducted by Indriyani et al. (2023) reported that SVM achieved an accuracy of 84%. Similarly, a study by Kelvin et al. (2022) found that SVM reached an accuracy of 92.13% in sentiment analysis related to the COVID-19 pandemic on X (Twitter). Other findings also indicate that SVM provides better classification performance compared to algorithms such as Naïve Bayes and K-Nearest Neighbor in social media sentiment analysis (Pamungkas & Kharisudin, 2021).

However, studies examining public sentiment toward the efficiency policy of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year on social media remain limited. In addition, the characteristics of social media data, which are largely unstructured, dynamic, and linguistically diverse, pose significant challenges for sentiment classification. Therefore, further research is required to apply effective classification methods in order to systematically identify public sentiment trends toward this policy.

Based on this context, this study aims to apply the Support Vector Machine (SVM) method to sentiment analysis of posts on X (Twitter) related to the efficiency policy of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year. This research seeks to identify the overall tendency of public sentiment and evaluate the performance of SVM in classifying public opinions within social media text data. The novelty of this study lies in the application of SVM-based sentiment analysis to examine public responses to a relatively recent government budget efficiency policy, thereby providing empirical insights into public perceptions of the policy.

2. Literature Review

2.1. Text mining

Text Mining refers to the process of exploring unstructured or semi-structured textual data to identify patterns, topics, or previously unknown information (Fathiarahma et al., 2023). The primary objective of text mining is to evaluate information and support decision-making processes, particularly when dealing with large volumes of textual data (Husada & Paramita, 2021). This method is widely applied to address various text processing tasks, including classification, information retrieval, information extraction, and data grouping or clustering (Ridwansyah, 2022). Several stages are commonly involved in the text mining process (Sabrila et al., 2021):

a. *Preprocessing text*

Preprocessing text is an initial stage aimed at cleaning and standardizing textual data so that it can be effectively used in the model evaluation process (Maulaya & Junadhi, 2022). This stage is essential in Text Mining to improve data quality and enhance the performance of subsequent analysis. The following preprocessing steps are applied in this study:

1) *Case Folding* : the process of converting all characters in a document into lowercase letters to ensure text consistency (Puspitasari & Sutabri, 2024).
2) Normalisasi : the process of transforming non-standard or slang words into their standard linguistic forms so that they correspond to their original meanings (Puspitasari & Sutabri, 2024).
3) *Tokenization* : the process of splitting sentences into smaller units in the form of individual words or tokens (Puspitasari & Sutabri, 2024).
4) *Stopword Removal* : the process of removing commonly used words that do not carry significant meaning in the analysis, such as “yang”, “di”, and “dengan”, leaving only meaningful terms (Puspitasari & Sutabri, 2024).
5) *Stemming* : the process of reducing words to their root forms by removing prefixes, suffixes, or other affixes (Puspitasari & Sutabri, 2024).

b. *Term Frequency-Inverse Document Frequency* (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) is used to assign weights to extracted terms by considering the frequency of a word in a particular document relative to its occurrence across the entire document corpus (Fitri & Putri, 2022). The weighting is calculated as follows:

$$TF - IDF_{t,d} = tf_{t,d} \times log\left(\frac{N}{df_t}\right)$$

Notation:
$tf_{t,d}$ = frequency term t in document d
df_t = number of documents containing term t
N = total number of documents

2.2. Classification

In Machine Learning, classification refers to a method used to develop models capable of categorizing data into predefined classes based on their characteristics, allowing each data instance to be assigned to the most appropriate class according to similarity patterns (Amrozi et al., 2022). To evaluate the performance of a classification model, a Confusion Matrix is commonly employed. This method compares the predicted labels generated by the model with the actual labels in the dataset, enabling a clear assessment of model accuracy and the types of classification errors that occur. The comparison results are then used to calculate evaluation metrics such as accuracy, precision, recall, and F1-score (Husada & Paramita, 2021). The evaluation measurement using a confusion matrix is presented in Table 1.

Table 1. Confusion matrix (Salam et al., 2023)

Actual Data	**Prediction Data**	
	Positive	**Negative**
Positive	TP	FN
Negative	FP	TN

The values obtained from the Confusion Matrix are evaluated using several performance metrics (Salam et al., 2023):

a. Accuracy measures how correctly the model classifies data according to the actual conditions. It is calculated as:
$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$
b. Precision indicates how accurately the model identifies positive instances. It is calculated using the following formula:
$$Presisi = \frac{TP}{TP + FP} \times 100\%$$
c. *Recall* measures the model's ability to identify all actual positive instances, calculated as:
$$Recall = \frac{TP}{TP + FN} \times 100\%$$
d. *F1-score* represents the harmonic mean of precision and recall and is used to evaluate the overall performance of the classification model:
$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \times 100\%$$

Notation:

TP (*True Positive*) = positive data correctly classified as positive
TN (*True Negative*) = negative data correctly classified as negative
FP (*False Positive*) = negative data incorrectly classified as positive
FN (*False Negative*) = positive data incorrectly classified as negative

2.3. Support vector machine

Support Vector Machine (SVM) is a supervised learning method used for data classification by constructing an optimal hyperplane that separates different classes. The data points closest to the hyperplane are referred to as support vectors, and the margin between classes is maximized to improve classification accuracy (Sitorus et al., 2021).

a. Linear Support Vector Machine

Linear SVM separates two classes by constructing an optimal hyperplane that maximizes the margin determined by the support vectors (Suh et al., 2021; Yassin et al., 2021). The decision function can be expressed as (Pratiwi & Setyawan, 2021):

$$f(x) = \boldsymbol{w}^T x + b$$

The prediction for new data is calculated as:

$$f(\boldsymbol{x}_t) = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x}_t + b$$

Notation:

$\boldsymbol{w}^T$ = weight vector
x = input variabl
b = bias
α_i = Lagrange multiplier of the *i-th* data point
y_i = class label of the *i-th* support vector
x_i = *i-th* support vector
x_t = new data vector
n = number of support vectors

b. Nonlinear Support Vector Machine

In cases where data are not linearly separable, nonlinear SVM maps the data into a higher-dimensional space so that the classes can be separated linearly by a hyperplane (Rahman et al., 2021). This transformation is performed using kernel functions, which implicitly project the original feature space into a higher-dimensional feature space, enabling the model to capture nonlinear patterns effectively (Fadilah dkk., 2020; Yassin dkk., 2021).

c. Kernel Tricks

Kernel functions are used to map input data into a higher-dimensional space (kernel space), allowing linear separation in that transformed space (Nugroho et al., 2011). Several kernel functions commonly used in SVM are presented below.

Table 2. SVM Kernel Functions

Kernel	Rumus
Linear	$K(\boldsymbol{x}_i; \boldsymbol{x}_j) = \boldsymbol{x}_i{}^T \boldsymbol{x}_j$
Polynomial	$K(\boldsymbol{x}_i; \boldsymbol{x}_j) = (\boldsymbol{x}_i{}^T \cdot \boldsymbol{x}_j + c)^d$
RBF (Gaussion)	$K(\boldsymbol{x}_i; \boldsymbol{x}_j) = exp\left(-\gamma \lvert\lvert \boldsymbol{x}_i - \boldsymbol{x}_j \rvert\rvert^2\right)$
Sigmoid	$K(\boldsymbol{x}_i; \boldsymbol{x}_j) = \tanh(k\ \boldsymbol{x}_i{}^T \boldsymbol{x}_j + r)$

Notation:

$K(x_i; x_j)$ = kernel function
$\boldsymbol{x}_i{}^T \boldsymbol{x}_j$ = *dot product of input vectors*
c = constant
d = polynomial degree
γ = $\frac{1}{2\sigma^2}$: gamma parameter
tanh = hyperbolic tangent function
r = bias constant

d. Expenditure Efficiency of the 2025 State Budget (APBN) and Regional Budget (APBD)

The Indonesian government under the leadership of Prabowo Subianto and Gibran Rakabuming Raka implemented a policy on the efficiency of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year through Presidential Instruction No. 1 of 2025. This policy targets savings of approximately IDR 306.69 trillion by reducing various expenditure components to improve the effectiveness of public financial management (Rizky et al., 2025). The policy includes reductions in spending across 16 major categories, such as office stationery, official travel, ceremonial activities, and building or vehicle rentals, with the objective of minimizing inefficiencies and enhancing the effective use of national and regional budgets.

3. Research Methodology

This study employs a quantitative method with descriptive and explanatory approaches to analyze public sentiment toward the efficiency policy of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year on X (Twitter). The descriptive approach is used to illustrate the distribution of sentiments expressed in public discussions, while the explanatory approach evaluates the performance of the Support Vector Machine (SVM) method in classifying sentiment. This method is selected because SVM is capable of handling high-dimensional textual data and has demonstrated strong performance in sentiment classification tasks.

The study utilizes primary data obtained from posts on X (Twitter) containing the keyword "efisiensi anggaran." Data were collected from 15 January 2025 to 25 March 2025 using the platform's authentication token to enable systematic data extraction. The collected data were then filtered based on topic relevance, posting period, and the use of the Indonesian language to ensure suitability for sentiment analysis. The overall data analysis process is illustrated in Figure 1.

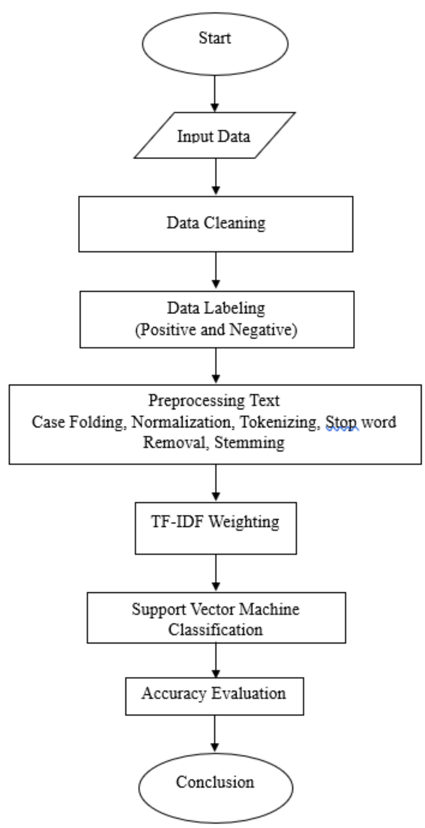

Figure 1. Flowchart

4. Result and Discussion

4.1. Descriptive Analysis

Descriptive statistics were used to illustrate user activity on X (Twitter) in discussing the efficiency policy of the State Budget (APBN) and Regional Budget (APBD) for the 2025 fiscal year. A total of 3,934 tweets were collected using the keyword *"efisiensi anggaran"* during the period 15 January to 25 March 2025, with an average of approximately 50 tweets per day. After the cleaning process, which included the removal of irrelevant content, duplicate data, symbols, URLs, and non-Indonesian language text, the dataset was reduced to 3,745 tweets suitable for analysis.

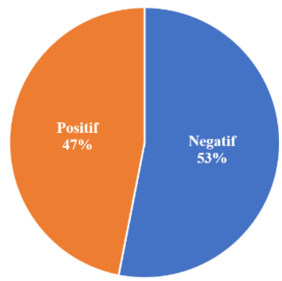

Figure 2. Sentiment Distribution of Budget Efficiency 2025

The distribution of public sentiment toward the 2025 budget efficiency policy is presented in Figure 2. Based on the figure, the pie chart illustrates that positive sentiment accounts for 47% (1,757 tweets), while negative sentiment represents 53% (1,988 tweets). This distribution indicates that public responses to the 2025 budget efficiency policy tend to be dominated by negative sentiment. Following the labeling stage, the dataset underwent a preprocessing phase to transform initially unstructured textual data into a structured format suitable for further analysis.

4.2. Sentiment Classification using Support vector machine

a. Data Crawling using X auth token

Data collection was conducted through a crawling process using an authentication token from X (Twitter) with the assistance of Google Colab. The authentication token was obtained by logging into an X account and extracting the *auth_token* value from browser cookies via developer tools, which was then used to authorize data requests. Users subsequently defined the search keyword, number of tweets, and time range to systematically extract the data. Through this process, 3,934 tweets from the period 15 January to 25 March 2025 were collected and stored in CSV format.

b. Data Cleaning

The data cleaning stage was conducted to remove irrelevant elements within the tweets, including special characters, URLs, punctuation marks, numbers, mentions, and hashtags. After this process, the dataset used for analysis decreased from 3,934 tweets to 3,745 tweets, ensuring that the remaining data were more suitable for subsequent text processing and sentiment classification.

Table 3. Cleaning Result

Full_Text	Cleaning
Sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg https://t.co/Qja55bQnPr	Sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg

c. Data Labeling

The sentiment labeling process was conducted using a text processing system in Google Colab to categorize tweets based on the polarity of words appearing in each sentence. Through this process, tweets were classified into positive and negative sentiment categories.

Table 4. Data Labeling Result

Tweet	Sentiment
Sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg	Positive
Efisiensi anggaran yang diterapkan juga berdampak pada kampus di bawah naungan Kemenkes Dampaknya sangat terasa padahal mahasiswa telah membayar UKT secara penuh Sebenarnya pemerintah ingin fokus mengalokasikan anggaran ke sektor apa	Negative

d. Preprocessing Text

The text preprocessing stage consisted of several steps, including (1) case folding, (2) normalization, (3) tokenization, (4) stopword removal, and (5) stemming. These steps were applied to transform raw and unstructured text data into a structured format suitable for further analysis. The results of the preprocessing stage are presented below:

1) Case Folding

Table 5. Case Folding Result

Tweet	Case Folding
Sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg	sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg

2) Normalization

Table 6. Normalization Result

Case Folding	Normalization
sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gile bro ini gua lulus kerja apaan nih inpo loker bg	sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gila bro ini saya lulus kerja apaan ini info loker bang

3) Tokenization

Table 7. Tokenization Result

Normalization	Tokenization
sudut pandang saya sebagai mahasiswa teknik sipil terhadap efisiensi anggaran yang terjadi di tahun ini gila bro ini saya lulus kerja apaan ini info loker bang	['sudut', 'pandang', 'saya', 'sebagai', 'mahasiswa', 'teknik', 'sipil', 'terhadap', 'efisiensi', 'anggaran', 'yang', 'terjadi', 'di', 'tahun', 'ini', 'gila', 'bro', 'ini', 'saya', 'lulus', 'kerja', 'apaan', 'ini', 'info', 'loker', 'bang']

4) Stopword Removal

Table 8. Stopword Removal Result

Tokenization	Stopword Removal
['sudut', 'pandang', 'saya', 'sebagai', 'mahasiswa', 'teknik', 'sipil', 'terhadap', 'efisiensi', 'anggaran', 'yang', 'terjadi', 'di', 'tahun', 'ini', 'gila', 'bro', 'ini', 'saya', 'lulus', 'kerja', 'apaan', 'ini', 'info', 'loker', 'bang']	sudut pandang mahasiswa teknik sipil efisiensi anggaran terjadi tahun gila lulus kerja info loker

5) Stemming

Table 9. Stemming Result

Stopword Removal	Stemming
sudut pandang mahasiswa teknik sipil efisiensi anggaran terjadi tahun gila lulus kerja info loker	sudut pandang mahasiswa teknik sipil efisiensi anggar jadi tahun gila lulus kerja info loker

e. TF-IDF Weighting

The preprocessed text data were subsequently transformed using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting scheme. This method assigns weights to terms based on their frequency in a document relative to their occurrence across the entire corpus. The TF-IDF weighting results for several selected terms are presented in Table 10:

Table 10. TF-IDF Result

Term	TF			DF	IDF	TF-IDF		
	d1	d2	d3			d1	d2	d3
anggar	0,07	0,09	0,1	3700	0,012088	0,000863	0,001099	0,001209
bawah	0,07	0,05	0,0	57	4,185126	0,2929588	0,190233	0,0
benar	0,0	0,05	0,0	33	4,731669	0,0	0,215076	0,0
dampak	0,0	0,09	0,0	234	2,772856	0,0	0,252078	0,0
dasar	0,0	0,0	0,0	28	4,895972	0,0	0	0,0
dwifungsi	0,0	0,0	0,0	19	5,283738	0,0	0,0	0,0
efisiensi	0,07	0,05	0,1	3681	0,17237	0,001231	0,000784	0,001724
gila	0,07	0,0	0,0	20	5,232445	0,373746	0	0
mahasiswa	0,07	0,05	0,0	33	4,731669	0,337976	0,215076	0,0

f. Kernel Testing

At this stage, several kernel functions of the Support Vector Machine (SVM) were evaluated to determine the most suitable kernel for classifying tweets related to the budget efficiency policy. Four kernels were tested: linear, polynomial, radial basis function (RBF), and sigmoid. Each kernel was trained using TF-IDF features and evaluated using accuracy, precision, recall, and F1-score metrics to compare classification performance.

Table 11. Kernel Testing

Kernel	Parameters	Value
Linear	*Complexity* (C)	[0,5; 0,75; 1]
Polinomial	*Complexity* (C)	[0,5; 0,75; 1]
	Degree (d)	[1; 2; 3]
	Gamma (γ)	[0,5; 0,75; 1]
RBF	*Complexity* (C)	[0,5; 0,75; 1]
	Gamma (γ)	[0,5; 0,75; 1]
Sigmoid	*Complexity* (C)	[0,5; 0,75; 1]
	Gamma (γ)	[0,5; 0,75; 1]

g. Kernel Testing Result

For the linear kernel, the complexity parameter (C) was tested using values of 0.5, 0.75, and 1. The results showed a consistent increase in performance as the value of C increased, with the best performance obtained at C = 1, achieving an accuracy of 96.4% and an F1-score of 96.1%.

For the polynomial kernel, experiments were conducted on the parameters C, gamma, and degree to obtain the optimal combination. Testing C values of 0.5, 0.75, and 1 indicated that model performance improved with increasing C, with the best result at C = 1 (accuracy 86.1% and F1-score 83.3%). Further testing of gamma values (0.5, 0.75, and 1) showed the highest performance at gamma = 1, achieving an accuracy of 96.4% and an F1-score of 96.1%. The degree parameter was tested using values of 1, 2, and 3, where the highest accuracy of 99.4% and F1-score of 99.3% were obtained at degree = 2. Therefore, the optimal parameter combination for the polynomial kernel was C = 1, gamma = 1, and degree = 2.

For the RBF kernel, the parameters C and gamma were evaluated to obtain optimal performance. Testing C values of 0.5, 0.75, and 1 showed improved performance with increasing C, with the best result at C = 1, achieving an accuracy of 89.5% and an F1-score of 87.73%. Further testing of gamma values (0.5, 0.75, and 1) showed the highest performance at gamma = 1. The combination of C = 1 and gamma = 1 produced an accuracy of 98.91% and an F1-score of 98.82%, indicating the optimal parameter configuration for the RBF kernel.

For the sigmoid kernel, experiments were also conducted on the parameters C and gamma. Testing C values of 0.5, 0.75, and 1 showed improved performance with higher C values, with the best result obtained at C = 1, producing an accuracy of 85.93% and an F1-score of 83.03%. Further testing of gamma values (0.5, 0.75, and 1) showed the highest performance at gamma = 0.5. The combination of C = 1 and gamma = 0.5 produced an accuracy of 91.2% and an F1-score of 90.14%, representing the optimal parameter configuration for the sigmoid kernel.

Based on the kernel testing results, the best-performing model for each kernel function was identified. A summary of the experimental results is presented in Table 12:

Table 12. Performance Comparison of SVM Kernels

Kernel	Parameters	Accuracy	Precision	Recall	*F1-score*
Linear	C =1	96,4%	97,5%	94,7%	96,1%
Polinomial	C =1; Gamma=1 Degree=2	99,4%	99,7%	98,9%	99,3%
RBF	C =1; Gamma=1	98,9%	99,5%	98,1%	98,8%
Sigmoid	C =1; Gamma = 0,5	91,2%	95,2%	85,6%	90,1%

Based on Table 11, the polynomial kernel achieved the best overall performance compared with the other kernels. This is indicated by an accuracy of 99.4%, precision of 99.7%, recall of 98.1%, and an F1-score of 98.8%, which are the highest values among all tested kernels. The selection of the polynomial kernel with parameters C = 1, gamma = 1, and degree = 2 indicates that the sentiment data exhibit nonlinear patterns.

h. Confusion matrix

The following confusion matrix shows the SVM model using a polynomial kernel with parameters C = 1, Gamma = 1, and Degree = 2:

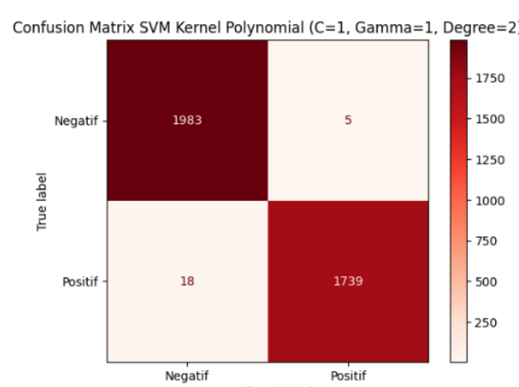

Figure 3. Confusion matrix

As shown in Figure 3, a total of 1,983 negative-labeled data points were correctly classified, while 5 negative data points were incorrectly predicted as positive. For the positive class, 1,739 data points were correctly predicted as positive, whereas 18 positive data points were misclassified as negative. The results of the model performance evaluation are presented in Table 13.

Table 13. Polynomial Kernel SVM Model Evaluation

		Precision	Recall	F1-score	Support
Negative		99,1%	99,7%	99,4%	1988
Positive		99,7%	98,9%	99,3%	1757
Accuracy	99%				3745
Macro avg		0,99	0,99	0,99	3745
Weigthed avg		0,99	0,99	0,99	3745

Table 12 shows that the model with C = 1, Gamma = 1, and Degree = 2 achieves an accuracy of 99%. For the negative class, the model obtained precision of 99.1% and recall of 99.7%, while for the positive class it achieved precision of 99.7% and recall of 98.9%. The F1-score of 99% indicates that the model has very good and balanced performance in classifying sentiments.

4.3. Discussion

In general, public sentiment toward the 2025 APBN and APBD expenditure efficiency policy on the X platform shows high public attention, with negative sentiment dominating at 53% compared to positive sentiment at 47%. To identify the most frequently occurring words in each sentiment category, a word cloud visualization was generated based on positive and negative labels.

Figure 4. Positive and Negative Sentiment Wordcloud

ased on the positive sentiment wordcloud, public discussions are dominated by issues related to leadership and governance, indicated by the frequent appearance of words such as Prabowo Subianto, "presiden", "negara", and "pemerintah." In addition, words such as "program", "instruksi", and "prioritas" indicate a focus on policy implementation. Economic-related terms such as "harga", "dampak", "triliun", and "bantuan" suggest that positive sentiment is often associated with economic aspects and public welfare.

Based on the negative sentiment wordcloud, public discussions are largely dominated by criticism of government policies, indicated by the appearance of words such as "negara", Prabowo Subianto, "presiden", and "pemerintah", Terms such as "potong", "gaji", "harga", "hutang" and "mahal" indicate that negative sentiment is mainly related to economic issues, public spending, and concerns about budget management and its impact on society..

The application of the support vector machine (SVM) method in sentiment analysis of the 2025 APBN and APBD expenditure efficiency policy demonstrates very strong performance, achieving 99% accuracy with balanced precision, recall, and F1-score values. The confusion matrix shows that most data points in both classes were correctly classified with a very low error rate, indicating that the model is effective and stable in distinguishing between positive and negative sentiments.

According to the model performance classification proposed by Sang et al. (2021) in the study titled *"Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine,"* an accuracy value within the range of 91%–100% is categorized as very good. Therefore, the 99% accuracy obtained by the SVM model in this study indicates that the model has excellent performance in classification tasks.

5. Conclusion

Based on the sentiment classification results of the 2025 APBN and APBD expenditure efficiency policy on the X platform using the Support Vector Machine method, a total of 3,934 tweets were collected, with 3,745 data points considered suitable for analysis. The results indicate a high level of public attention, with negative sentiment dominating (53%) compared to positive sentiment (47%). The SVM model achieved optimal performance with 99% accuracy, and precision, recall, and F1-score values ranging from 98% to 99%, indicating that the model is effective in distinguishing between positive and negative sentiments.

References

Aisah, I. S., Irawan, B., & Suprapti, T. (2024). Algoritma *Support vector machine* (SVM) untuk Analisis Sentimen Ulasan Aplikasi Al Qur'an Digital. *Jurnal Mahasiswa Teknik Informatika*, *7*(6), 3759–3765. https://doi.org/10.36040/jati.v7i6.8263

Amrozi, Y., Yuliati, D., Susilo, A., Novianto, N., & Ramadhan, R. (2022). Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *11*(3), 394–399. https://doi.org/10.32736/sisfokom.v11i3.1502

Anwar, K. (2022). Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *2*(4), 148–155. https://doi.org/10.30865/klik.v2i4.315

Bei, F., & Sudin, S. (2021). Analisis Sentimen Aplikasi Tiket Online di Play Store Menggunakan Metode *Support vector machine* (SVM). *Seminar Nasional Sistem Informasi Dan Manajemen Informatika (SISMATIK)*, *01*(01), 91–97. https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/13

Fadilah, W. R. U., Agfiannisa, D., & Azhar, Y. (2020). Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode *Support vector machine*. *Fountain of Informatics Journal*, *5*(2), 45. https://doi.org/10.21111/fij.v5i2.4449

Fathiarahma, A., Voutama, A., Ridwan, T., & Heryana, N. (2023). Analisis *Text mining* Klasifikasi Kegiatan Keluarga Menggunakan Orange dengan Metode Naive Bayes. *Jurnal Teknologi Terpadu*, *9*(1), 35–41. https://doi.org/10.54914/jtt.v9i1.606

Fitri, D. A., & Putri, A. (2022). Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma *Support vector machine*. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(3), 472–478. https://doi.org/10.37859/coscitech.v3i3.4260

Ginting, E., Putri, S. E., & Sinaga, Z. Y. (2024). Analisis Keterkaitan Pemerintah Pusat dalam Peningkatan APBN dan APBD Tahun Anggaran 2023. *Jurnal Ilmu Hukum Dan Sosial*, *2*(1), 91–99. https://doi.org/10.51903/hakim.v2i1.1545

Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen pada Maskapai Penerbangan di *Platform* Twitter Menggunakan Algoritma *Support vector machine* (SVM). *Jurnal Teknika*, *10*(1), 18–26. https://doi.org/10.34148/teknika.v10i1.311

Husnaningtyas, N., & Dewayanto, T. (2023). Financial Fraud Detection and *Machine learning* Algorithm (Unsupervised Learning): Systematic Literature Review. *Jurnal Riset Akuntansi Dan Bisnis Airlangga*, *8*(2), 1521–1542. https://doi.org/10.20473/jraba.v8i2.49927

Indriani, A. (2020). Analisa Perbandingan Metode Naïve Bayes Classifier dan K-Nearest Neighbor terhadap Klasifikasi Data. *Sebatik*, *24*(1), 1–7. https://doi.org/10.46984/sebatik.v24i1.909

Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Naïve Bayes dan *Support vector machine*. *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, *10*(2), 176–184.

https://doi.org/10.37373/tekno.v10i2.419

Instruksi Presiden (Inpres) No. 1 Tahun 2025 Tentang Efisiensi Belanja dalam Pelaksanaan Anggaran Pendapatan dan Belanja Negara dan Anggaran Pendapatan dan Belanja Daerah Tahun Anggaran 2025. (2025). jdih.setneg.go.id

Jun, Z. (2021). The Development and Application of *Support vector machine*. *Journal of Physics: Conference Series*, *1748*(5), 0–6. https://doi.org/10.1088/1742-6596/1748/5/052006

Kelvin, Banjarnahor, J., Indra, E., & Sinurat, S. H. (2022). Analisis Perbandingan Sentimen Corona Virus Disease-2019 (COVID19) pada Twitter Menggunakan Metode Logistic Regression dan *Support vector machine* (SVM). *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *5*(2), 47–52. https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365

Larasati, R. C., Dewi, C., & Christanto, H. J. (2024). Analisis Sentimen Produk Kecantikan Jenis Moisturizer di Twitter Menggunakan Algoritma Super Vector Machine. *Jurnal Tekinkom*, *7*(1), 124–134. https://doi.org/10.37600/tekinkom.v7i1.1243

Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode *Support vector machine* (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). *Prosiding Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, *2*(2), 64–74. https://conference.upnvj.ac.id/index.php/senamika/article/view/1630

Maulaya, A. K., & Junadhi. (2022). Analisis Sentimen Menggunakan *Support vector machine* Masyarakat Indonesia Di Twitter Terkait Bjorka. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(3), 495–500. https://doi.org/10.37859/coscitech.v3i3.4358

Namira Nur Az-Zahra, R., Rizqi Fitrialdi, M., Nurjanah, E., Darmawan, M., & Firmansyah, R. (2021). Analisis Sentimen Media Sosial Tiktok dengan Metode Supervised Learning pada Algoritma *Machine learning*. *Jurnal Buffer Informatika*, *7*(1), 18–25. https://journal.uniku.ac.id/index.php/buffer

Nugroho, A. S., Witarto, A. B., & Handoko, D. (2011). *Support vector machine*: Teori dan Aplikasinya dalam Bioinformatika. *Proceedings of the 2011 Chinese Control and Decision Conference, CCDC 2011*, 842–847. https://doi.org/10.1109/CCDC.2011.5968300

Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada *Machine learning*: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, *4*(1), 61–72. https://doi.org/10.54082/jiki.168

Nurkholifah, M., Jasmarizal, Umar, Y., & Rahmaddeni. (2023). Analisa Performa Algoritma *Machine learning* Dalam Prediksi Penyakit Liver. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *4*(1), 164–172. https://doi.org/10.35870/jimik.v4i1.149

Pamungkas, F. S., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia terhadap Pandemi Covid-19 pada Media Sosial Twitter. *Prosiding Seminar Nasional Matematika*, *4*, 1–7. https://journal.unnes.ac.id/sju/prisma/article/view/45038

Pratiwi, N., & Setyawan, Y. (2021). Analisis Akurasi dari Perbedaan Fungsi Kernel dan Cost pada *Support vector machine* Studi Kasus Klasifikasi Curah Hujan di Jakarta. *Journal of Fundamental Mathematics and Applications (JFMA)*, *4*(2), 203–212. https://doi.org/10.14710/jfma.v4i2.11691

Purnajaya, A. R., & Pernando, Y. (2023). Analisa Sentimen Informasi Hoaks Pasca Pandemi Covid-19 dengan *Text mining*. *Journal of Computer System and Informatics (JoSYC)*, *4*(3), 460–469. https://doi.org/10.47065/josyc.v4i3.3358

Puspitasari, D., & Sutabri, T. (2024). Analisis Sentimen Berdasarkan pada Twitter (X) terhadap Layanan Indihome Menggunakan Algoritma *Support vector machine* (SVM). *Jurnal Manajemen Informatika Dan Bisnis Digital (JUMINTAL)*, *3*(2), 58–71. https://doi.org/10.55123/jumintal.v3i2.4449

Putra, K. T., Hariyadi, M. A., & Crysdian, C. (2023). Perbandingan Feature Extraction TF-IDF dan BOW Untuk Analisis Sentimen Berbasis SVM. *Jurnal Cahaya Mandalika*, *3*(2), 1449–1463. https://doi.org/https://doi.org/10.36312/jcm.v3i2

Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan *Support vector machine*. *Jurnal Rekayasa Sistem Dan Teknologi Informasi*, *5*(1), 17–23. https://doi.org/10.29207/resti.v5i1.2700

Reviantika, F., Azhar, Y., & Marthasari, G. I. (2021). Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. *Jurnal Sistem Cerdas*, *3*(4), 387–392. https://doi.org/10.37396/jsc.v4i3.166

Ridwansyah, T. (2022). Implementasi *Text mining* terhadap Analisis Sentimen Masyarakat Dunia di Twitter terhadap Kota Medan Menggunakan K-Fold Cross Validation dan Naïve Bayes Classifier. *Jurnal KLIK: Kajian Ilmiah*

Informatika Dan Komputer, *2*(5), 178–185. https://doi.org/10.30865/klik.v2i5.362

Rizky, G. M., Tanzil, V. J., Rafi, A., Pasara, D., & Tiasono, E. J. (2025). Dampak Efisiensi Anggaran pada ATR / BPN Kota Tangerang : Optimalisasi Layanan Pertanahan dan Iklim Investasi. *Jurnal Sosial Dan Humaniora*, *9*(2), 326–342. https://doi.org/https://doi.org/10.37817/ikraith-humaniora.v9i2

Sabrila, T. S., Sari, V. R., & Minarno, A. E. (2021). Analisis Sentimen pada Tweet tentang Penanganan Covid-19 Menggunakan Word Embedding pada Algoritma *Support vector machine* dan K-Nearest Neighbor. *Fountain of Informatics Journal*, *6*(2), 69. https://doi.org/10.21111/fij.v6i2.5536

Salam, R. R., Jamil, M. F., Ibrahim, Y., Rahmaddeni, R., Soni, S., & Herianto, H. (2023). Analisis Sentimen terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan *Support vector machine*. *MALCOM: Indonesian Journal of Machine learning and Computer Science*, *3*(1), 27–35. https://doi.org/10.57152/malcom.v3i1.590

Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan *Support vector machine*. *E-Proceeding of Engineering*, *8*(5),8954–8963. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15900

Santoso, P., Abijono, H., & Anggreini, N. L. (2021). Algoritma Supervised Learning dan Unsupervised Learning dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, *4*(2), 315–318. https://doi.org/10.33379/gtech.v4i2.635

Sitorus, Y. W., Sukarno, P., & Mandala, S. (2021). Analisis Deteksi Malware Android Menggunakan Metode *Support vector machine* & Random Forest. *E-Proceeding of Engineering*, *8*(6), 12500–12518. https://openlibrarypublications.telkomuniversity.ac.id/ind

Suh, Y. S., Shin, S. K., Baang, D., Seo, S. M., & Lee, J. B. (2021). A Brief Review of Linear *Support vector machine* for *Machine learning* Programming. *Transactions of the Korean Nuclear Society Virtual Spring Meeting*, *3*, 13–15. https://doi.org/10.1007/978-3-319-41063-0_4

Yassin, B., Mohamed, C., & Al-Amrani, Y. (2021). A Nonlinear *Support vector machine* Analysis Using Kernel Functions for Nature and Medicine. *E3S Web of Conferences*, *319*, 1–5. https://doi.org/10.1051/e3sconf/202131901103

Yul, H. (2025). *Efisiensi Anggaran dalam Evaluasi Pembangunan: Analisis Fungsi Organisasi BPKAD NTB, Opini Publik di Media Sosial X dan Rekomendasi Kebijakan*. *8*(1), 109–124. https://doi.org/http://dx.doi.org/10.35914/jemma.v8i1.3168