Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Sentiment Analysis of TikTok User Comments on The Free Nutritious Meal Program Using Support Vector Machine

Lina Nur Afifah[1], Sri Rahayu[2], Purwadi[3]*

[1,2]Informatic, Faculty of Computer Science, Amikom Purwokerto University
[3]Master of Computer Science, Faculty of Computer Science, Amikom Purwokerto University
linanurafifah14@gmail.com[1], srirahayu.23sa11a117@gmail.com[2],
purwadi@amikompurwokerto.ac.id[3]*

Abstract

This study aims to analyze user sentiment when leaving comments on TikTok about the Free Nutritious Food Program (MBG) to understand how the public views the program. Comment data was obtained through online collection and then divided into three groups: positive, negative, and neutral. Before further processing, the data went through a text cleaning and stemming stage to reduce word variation. The data was then represented using the TF-IDF method before being classified with a Support Vector Machine algorithm. The evaluation results showed that using stemming provided more accurate results than without using stemming, thereby improving the model's ability to recognize sentiments contained in comments using informal language. Additional analysis using word clouds, n-grams, and topic modeling provided an overview of words and issues frequently appearing in public discussions regarding the program.

***Keywords**: Sentiment Analysis; User Comments; Support Vector Machines; Social Media; TikTok*

1. Introduction

The Free Nutritious Food Program (MBG) is a government policy aimed at improving the nutritional quality of the community, especially for more vulnerable groups such as students and low-income families. This program is expected to be an important step in reducing the risk of malnutrition and helping improve food security at the national level. However, the implementation of this program has generated various responses from the public, particularly regarding the distribution of funds, the effectiveness of providing benefits, and the readiness of the necessary supporting facilities. Previous research has shown that the MBG issue often triggers polarization of public opinion on social media, particularly on platform X, where the public widely debates the program's effectiveness, budget burden, and government policy priorities [1].

In the age of social media, public perceptions of government policies are increasingly expressed through digital platforms like TikTok. Indonesia has a significant TikTok user base, making the platform a crucial public discussion platform for various national policy issues. Studies using TikTok comments as objects of sentiment analysis have also been conducted previously, For example, research on the school graduation ban policy in West Java showed that TikTok is an effective source of public opinion data, with the majority of comments successfully classified using the Naive Bayes algorithm [2].

Sentiment analysis is an important way to numerically understand public opinion. Several studies have shown that comments from TikTok and other social media users can be processed to determine whether they feel positive, negative, or neutral. One study on the TikTok app found that the SVM algorithm was able to produce up to 84% accuracy for comment classification, higher than the Naïve Bayes method on a dataset of 2,000 comments [3]. Meanwhile, another study showed that Naïve Bayes and LSTM methods can be used to analyze student comments, with Naïve Bayes achieving 83.69% accuracy on comments from the FTK online service at Ganesha University of Education [4].

These findings align with other studies that confirm the effectiveness of the Naïve Bayes algorithm in analyzing text-based public opinion, including a study of public sentiment about the KIP Kuliah Program on the platform X, which yielded an accuracy of 84.99% [5]. Thus, machine learning methods such as SVM, Naïve Bayes, and LSTM have proven relevant and widely used in social media based sentiment analysis. However, some previous studies have been limited to specific platforms such as the Google Play Store, Twitter/X, or internal surveys, making them less able to capture the characteristics of comments on TikTok, which are typically more relaxed, emotional, and volatile. In addition, some studies only divide sentiment into two categories, namely positive and negative, even though comments on social media are often unclear and require additional categories such as neutral or unknown. Multidimensional research on the previous

MBG Program even showed that negative sentiment dominated public discourse on platform X, and concluded that opinions on MBG were influenced by political, economic, health, and social factors [1].

Besides platform limitations and the variety of sentiment labels, another issue arises from the more complex preprocessing of TikTok data. Comments on TikTok often contain informal language, abbreviations, emojis, and informal sentence structures, requiring comprehensive preprocessing such as case folding, tokenizing, stopword removal, and stemming for effective classification. This is also confirmed by other studies highlighting that social media data requires a much more intensive data cleaning process before it can be modeled [2][1].

The research gap is also evident in the limited number of studies directly analyzing public sentiment toward the MBG Program through TikTok comments. Discussions about government programs are growing rapidly on TikTok and can directly reflect public opinion. Based on this gap, this study was conducted to analyze the sentiment of TikTok user comments related to the MBG Program using the Support Vector Machine (SVM) method and TF-IDF weighting. The selection of SVM is supported by previous research that demonstrated high performance in processing short and informal texts, including in the TikTok app sentiment analysis with an accuracy of up to 84% [3]. Furthermore, this study also added neutral and unknown sentiment categories to allow for more accurate interpretation of unclear comments. With this approach, the study is expected to provide a more comprehensive picture of public views on the MBG Program and play a role in evaluating policies based on public input.

2. Research Method

This study used a quantitative approach with text mining methods to analyze sentiment from TikTok user comments regarding the Free Nutritious Food Program (MBG). The research process was carried out regularly and followed the general stages of sentiment analysis, namely data collection, data labeling, preprocessing, feature extraction, classification, model assessment, word cloud, n-gram analysis, topic modeling, and word frequency analysis.

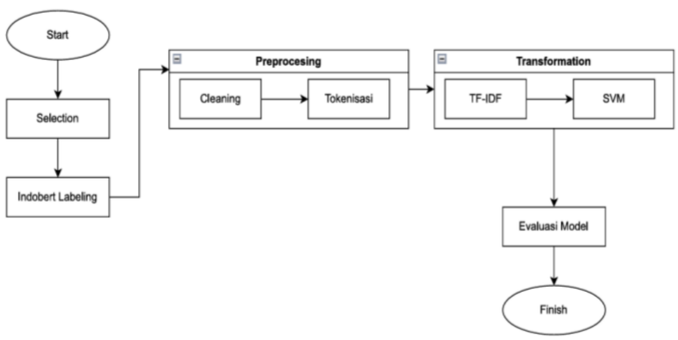

Fig. 1: Research Method

2.1. Data Collection

The data for this study were obtained through web scraping techniques to collect 400 TikTok comments discussing the MBG Program. Data collection was time-limited to align with the discussion period surrounding the MBG policy.

2.2. Data Labeling

The data was manually labeled into three sentiment categories: positive, negative, and neutral, following previous research methods aimed at maintaining class consistency through a manual labeling process [5][2]. The labeling results showed 180 negative comments, 119 positive comments, and 95 neutral comments. The dominance of negative comments indicates that most TikTok users provided critical responses to the MBG Program, while positive and neutral comments were fewer, indicating a difference of opinion but tended to express more dissatisfaction.

sentiment	count
negative	180
positive	119
neutral	95

dtype: int64

Fig. 2: Dataset Labeling

2.3. Pre-Processing

In the preprocessing stage, data cleaning was carried out through case folding, tokenizing, stopword removal, and stemming to ensure the text was in a format ready for processing. After the cleaning process was completed, a missing values check was performed on all columns to ensure there were no empty data that could interfere with the analysis process. The results showed that almost all columns had no missing values, except for several columns such as pinnedByAuthor (205), repliesToId (151), and replyCommentTotal (243). These columns did not affect the analysis process because they were not used as features in sentiment modeling. The main columns analyzed, namely text and sentiment, were free of empty values, so the data was considered valid and ready to enter the feature extraction stage.

```
=== Missing value setelah cleaning ===
videoWebUrl          0
submittedVideoUrl    0
input                0
cid                  0
createTime           0
createTimeISO        0
text                 0
diggCount            0
likedByAuthor        0
pinnedByAuthor     205
repliesToId        151
replyCommentTotal  243
uid                  0
uniqueId             0
avatarThumbnail      0
mentions             0
detailedMentions     0
raw_label            0
sentiment            0
dtype: int64
```

Fig. 3: Clean Missing Value Results

2.4. Feature Extraction (TF-IDF)

Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method to convert words into numbers indicating their weight. The use of TF-IDF is based on research from KIP Kuliah, which demonstrated the method's effectiveness in accurately representing text [5], as well as research on TikTok comment analysis that used TF-IDF to determine word weight before entering it into the classification [2].

2.5. Classification Using Support Vector Machines (SVM)

Sentiment classification was performed using the Support Vector Machine (SVM) algorithm. This algorithm was selected based on previous research findings, such as the KIP Kuliah sentiment study, which demonstrated an accuracy rate of up to 84.99% [5], and the TikTok study, which found SVM to be more effective in processing short, informal texts [2].

2.6. Model Assessment

Model assessment was conducted using a confusion matrix to calculate accuracy, precision, recall, and F1-score. This assessment method was used continuously in the KIP Kuliah research [5], TikTok comment research [2], and multidimensional MBG research on the X platform [1] to comprehensively assess the model's classification ability.

2.7. Word Cloud

A word cloud is a visual representation that shows the most frequently occurring words in a dataset. The size of a word reflects its frequency; the larger the word, the more frequently it appears. Word clouds are used to identify common topics, dominant words, and a preliminary overview of public opinion without manually reviewing the entire data.

2.8. N-gram analysis

N-gram analysis is a technique that examines the sequential occurrence of words in text. N-grams can be bigrams (two consecutive words) or trigrams (three consecutive words). This analysis provides a richer understanding of context than unigrams, helping to uncover frequently used phrase patterns in comments.

2.9. Topic Modelling Latent Dirichlet Allocation (LDA)

Topic Modeling (LDA - Latent Dirichlet Allocation) is a method for identifying hidden themes or topics within a collection of comments. LDA works by grouping documents based on word distribution, resulting in several topics frequently discussed by the public. This technique is useful for uncovering key issues related to the MBG program, such as perceived benefits, criticisms of implementation, economic concerns, and evaluations of policy effectiveness.

2.10. Word Frequency

Word frequency and dominant issue analysis were used to identify the most frequently occurring words or phrases in the dataset. This analysis helps identify trends in public opinion and supports understanding of the dominant topics discussed. The results can be used to inform more informed policy recommendations for the government and relevant parties.

3. Result and Discussion

This chapter discusses the results of sentiment analysis of TikTok user comments regarding the Free Nutritional Meal Program (MBG) using the text mining method and the Support Vector Machine (SVM) algorithm.

3.1. TikTok Comment Sentiment Distribution

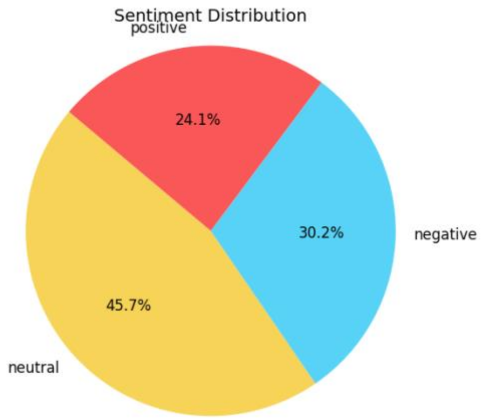

Fig. 4: Sentiment Distribution

Table 1: Sentiment Count

Sentiment	Count
Neutral	180
Negative	119
Positive	95

An initial analysis was conducted by examining the sentiment of all collected comments. From the visualization, the neutral sentiment category was the most common, with approximately 180 comments. This indicates that most TikTok users provided informative, impartial comments or did not explicitly express support or opposition to the MBG program. Negative sentiment accounted for 119 comments, while positive sentiment accounted for 95. These findings indicate that public perception of the MBG program is diverse, but not dominated by extreme sentiment. The large number of neutral comments could indicate that the public is still understanding the policy or is simply providing information without including strong emotional opinions.

3.2. Model Performance Evaluation

Table 2: Classification Report without Stemming

	Precision	Recall	F1-Score	Support
negative	0.50	0.45	0.48	22
neutral	0.59	0.76	0.67	38
Positive	0.40	0.21	0.28	19
accuracy			0.54	79
macro avg	0.50	0.48	0.47	79
weighted avg	0.52	0.54	0.52	79

Table 3: Classification Report with Stemming

	Precision	Recall	F1-Score	Support
negative	0.65	0.50	0.56	22
neutral	0.63	0.87	0.73	38
Positive	0.50	0.26	0.34	19
Accuracy			0.62	79
macro avg	0.59	0.54	0.55	79
weighted avg	0.61	0.62	0.59	79

An evaluation was conducted to determine the model's accuracy in classifying sentiment before and after using stemming. The model without stemming achieved an accuracy of 54.43%. This indicates that the variety of word forms makes it difficult for the model to recognize sentiment patterns. The F1-score for the positive class was also low, at 0.28, indicating the model's difficulty understanding diverse and ambiguous positive expressions.

After stemming was applied, accuracy increased significantly to 62.03%. This improvement occurred because stemming successfully simplified word variations, resulting in a more consistent feature distribution. The F1-score for the neutral class increased to 0.73, indicating that the model excelled at classifying informative or descriptive comments. Although the positive class remained difficult to predict, the improved performance demonstrated that stemming had a positive impact on classification quality.

Overall, the evaluation results confirmed that preprocessing is crucial for improving SVM performance in sentiment analysis, especially when dealing with unstructured social media text. Although model performance has improved, accuracy can still be improved by adding more data, balancing classes, or using more sophisticated feature representation techniques such as embedding or transformer-based models.

3.3. Confusion Matrix Analysis

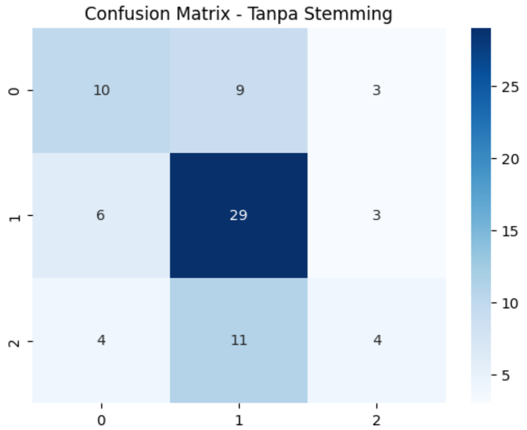

Fig. 5: Confusion Matrix without Stemming

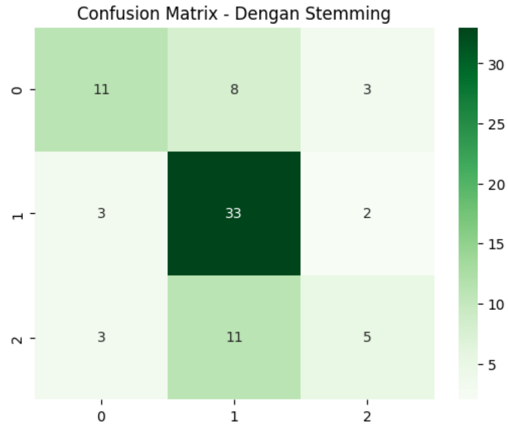

Fig. 6: Classification Matrix with Stemming

The results of the Confusion Matrix without stemming show that the model experienced several classification errors, particularly in the negative and positive classes. For example, some comments were actually negative but were classified as neutral or positive. Furthermore, the neutral class appeared to be more easily guessed correctly, although some errors still occurred. Overall, the model without stemming still performed poorly, primarily due to the variety of words in TikTok comments, which are often informal, use slang, and don't always follow standard Indonesian language structure.

Confusion Matrix Results with Stemming: After stemming, classification performance improved. The number of correct predictions in the neutral class increased from 29 to 33, while errors in the negative and positive classes decreased. These improvements demonstrate that stemming can reduce word form variation, making it easier for the model to recognize sentiment patterns. For example, words like "makan," "makan," "dimakan," or "ngemakan" can be transformed into the same base form, thereby improving feature consistency. Overall, stemming improves accuracy, precision, and recall, making it suitable for use in sentiment analysis of non-standard TikTok comments.

3.4. Analysis of Dominant Issues in TikTok Comments

In addition to sentiment analysis, this study also conducted issue extraction to determine the most frequently discussed topics among the public regarding the MBG program. Based on word frequency calculations and issue groupings, three main issues were identified.

Table 4: Issue Frequency Table

Issue	Comments
Budget	24
Food Quality	91
Abuse	10

Food quality is a key issue. Many TikTok users commented on the condition of the food, such as whether it was safe to consume, its taste, its nutritional content, and its hygiene. This demonstrates that the public is highly concerned about the quality of government assistance, as it directly impacts the health of recipients. Budgetary issues arose as the public questioned the transparency of the program's use of funds. Meanwhile, issues of misuse arose from public reports of possible improper distribution, inappropriate use of funds, or individuals exploiting the program improperly. These findings suggest that improving food quality and distribution oversight are two key aspects that the government needs to address to ensure the MBG program is more effective and well-received by the public.

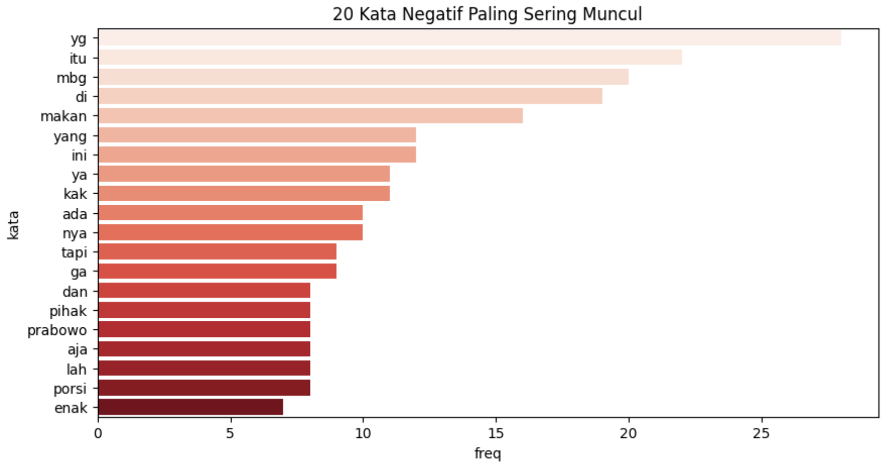

Fig. 7: Most Frequently Appearing Negative Words

To better understand conversation patterns, a word frequency analysis was conducted. Frequently occurring words included "yg," "kak," "menu," "makan," "gizi," "porsi," and "amanah." This pattern indicates that public discussion focused heavily on menu quality, adequate portion sizes, and assessments of program organizers. Analysis of bigrams such as "nutritionist," "menunya," "makan gizi," and "sekolah anak" further confirmed that food quality was a central topic of discussion. Meanwhile, trigrams such as "elementary school, junior high school class," and "junior high school class," indicated a focus on program implementation at various levels of education. Topic modeling using the Latent Dirichlet Allocation (LDA) method produced three main groups, namely (1) topics related to the menu and portion size of food, (2) topics discussing the taste and quality of the food provided, and (3) topics related to aspects of responsibility, nutritional adequacy, and consistency in implementing the program. These three topics are in accordance with the main issues that have been previously identified. Word sentiment analysis also shows that the number of positive words is greater, namely 58 words, while negative words are very few, which indicates that the community tends to have a positive or neutral perception of the MBG program.

Overall, the combination of sentiment analysis, confusion matrix, word frequency, and topic modeling provides a comprehensive picture of public perception of the MBG program. These results confirm that food quality and transparency in program implementation are two key indicators used by the public to assess its success.

4. Conclusion

Sentiment analysis of TikTok comments about the Free Nutritious Meals (MBG) program shows that public opinion is diverse, with the majority having neutral sentiment, followed by negative and positive sentiment. By applying stemming techniques, model accuracy increased from 54% to 62%, and precision and recall values increased, especially in the neutral sentiment category. This indicates that variations in language, such as slang, abbreviations, and non-standard language, affect the model's ability to accurately recognize sentiment. Furthermore, three main issues emerged in user comments: food quality, program budget utilization, and the risk of misuse. This suggests that public evaluation focuses not only on the program's objectives but also on how it is implemented. Based on the results of this study, it is recommended to improve food quality standards, increase transparency in budget utilization, and implement technology-based or community-participatory distribution monitoring mechanisms to make the MBG program more effective.

References

[1] A. Putriyekti, D. Sulistiawati, N. Khairani, and R. R. Atmaja, "ANALISIS MULTIDIMENSIONAL SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X," vol. 2, no. 1, pp. 1004–1024, 2025.

[2] K. Setiawan and E. Alafia, "Analisis Sentimen Komentar TikTok terhadap Kebijakan Larangan Wisuda Sekolah oleh Gubernur Jawa Barat Menggunakan Algoritma Naive Bayes," vol. 6, no. 3, pp. 1834–1841, 2025.

[3] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine Tiktok application sentiment analysis using naïve bayes algorithm and support vector machine," vol. 10, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.

[4] N. Kadek, T. Ari, I. G. A. Gunadi, and I. M. G. Sunarya, "Sentiment Analysis of Online Services at the Engineering and Vocational Faculty of Ganesha Education University Using Naïve Bayes and LSTM Algorithms Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM," vol. 4, no. July, pp. 1120–1129, 2024.

[5] D. Pramudita, Y. Akbar, and T. Wahyudi, "Sentiment Analysis of the Indonesian Smart College Card Program on Social Media X Using the Naive Bayes Algorithm Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah Pada Media Sosial X Menggunakan Algoritma Naive Bayes," vol. 4, no. October, pp. 1420–1430, 2024.

[6] D. F. M. Dina, T. Haryanti, M. A. Haq, "Analisis Sentimen Terhadap Komentar Pada Media Sosial TikTok yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naïve Bayes," vol. 7, no. 1, 2025.

[7] M. M. Laia, F. N. Hasan, A. Y. Kuntoro, "Analisis Sentimen Program Makan Gratis Pada Platform X Menggunakan Algoritma Naïve Bayes,"2025.

[8] R. N. Rahmawaty, D. Indramaya, A. Pambudi, "Penerapan Metode Naïve Bayes dan Cosine Similarity Dalam Analisis Sentimen Terhadap Platform Film Ilegal di Media Sosial X (Twiter)," vol. 3, no. 1, 2024.

[9] F. Fatkhurrohman, B. I. Nugroho, N. Fadillah, "Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM,", vol. 4, no. 3, pp. 3906-3917, 2025.