Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

Analysis of Public Opinion on TikTok Regarding the MBG Program Controversy Using the Support Vector Machine Algorithm

Fera Febrianti, Sahrul Ramadhan*, Irfan

Fakultas Teknik dan Ilmu Komputer, Program Studi Ilmu Komputer, Universitas Muhammadiyah Bima, Bima, Indonesia
Email: [1,*]febriantifera1802@email.com, [2]sahrulramadhanbin@gmail.com, [3]irfan1802@gmail.com
Email Penulis Korespondensi: febriantifera1802@email.com
Submitted: **07/05/2026**; Accepted: **29/06/2026**; Published**: 30/06/2026**

Abstract−This study examines the analysis of public opinion regarding the controversy surrounding the Free Nutritional Meal Program (MBG), a strategic policy of the Indonesian government aimed at reducing stunting rates and improving child nutrition. Despite its important social objectives, the program has sparked various public reactions concerning budget transparency, equitable distribution of aid, and food security. TikTok, as a social media platform with high levels of interaction, has become a primary platform for the public to express opinions on public policy. However, its use in sentiment analysis research remains relatively limited compared to other platforms such as Twitter and Instagram. This study aims to analyze public perceptions of the MBG program using a combination of the Support Vector Machine (SVM) algorithm and Word2vec. Research data was obtained through the collection of 2,381 TikTok comments, followed by preprocessing steps such as data cleaning, tokenization, slang normalization, stop-word removal, and stemming. After the data selection process, 2,376 comments were used in the lexicon-based sentiment labeling and classification process using SVM. The test results show that the SVM model achieved an accuracy of 80% before class imbalance handling, whereas after applying class imbalance handling techniques, the accuracy increased to 83%, with a weighted precision of 0.84, a recall of 0.83, and an F1 score of 0.83. This improvement indicates that processing the data in the database enhances the model's ability to recognize all sentiment classes more evenly, particularly positive sentiment, which previously had a smaller dataset. The sentiment analysis results show that the majority of opinions are dominated by neutral and negative sentiments, reflecting public concerns regarding the program's implementation effectiveness, budget management transparency, and equitable distribution. These findings suggest that public opinion on social media can be leveraged as a real-time source for evaluating government policies to help the government develop public communication strategies that are more transparent, responsive, and targeted toward the implementation of the MBG Program.

Keywords: Free Nutrious Meal; Sentiment Analysis; Support Vector Machine; Word2Vec; Tiktok;

1. INTRODUCTION

The Indonesian government has developed a nutritional intervention program known as the Free Nutritious Meals Program (MBG) to ensure that certain groups, particularly pregnant women, school-age children, and other vulnerable groups, have access to healthy, safe, and nutritious food [1]. This program was launched in response to the high prevalence of stunting in Indonesia a condition of impaired growth in children caused by prolonged inadequate nutrient intake. Based on 2025 data, the prevalence of stunting in Indonesia still stands at 19.8%, affecting approximately 4,482,340 children under five [2]. This high figure indicates that nutrition remains a major challenge in public health development, necessitating policies capable of sustainably improving the quality of the population's nutritional intake.

In early 2025, the Indonesian government officially launched the Free Nutritious Meals Program (MBG) as part of a national strategy to reduce stunting rates and improve the population's nutritional status. This program adopts the concept of providing nutritious meals in school settings, which has been successfully implemented in several countries [3]. The implementation of school lunch programs in countries such as the United States and Japan has demonstrated various benefits, including improved child health, enhanced concentration during learning, improved academic performance, and a reduced risk of hunger among school-age children [4]. The success of similar programs in these various countries serves as evidence that nutritional interventions through the provision of nutritious meals in schools can have a significant positive impact on the quality of human resources. Therefore, the MBG Program is expected to serve as a preventive measure to improve the quality of Indonesia's youth by ensuring that nutritional needs are met from an early age.

The implementation of the MBG Program in Indonesia has sparked various debates among the public, particularly regarding program management, transparency in budget allocation, the effectiveness of aid distribution, and several reports of suspected food poisoning. Discussions regarding this program have not only taken place directly in public spaces but have also spread across various social media platforms such as X, Instagram, YouTube, and TikTok [5], [6], [7], [8]. Among these platforms, TikTok has become one of the social media platforms with the highest user engagement in Indonesia and is widely used by the public to express opinions, criticism, or support for government policies. In addition to serving as a short-video entertainment platform, TikTok has evolved into a digital communication space that plays a crucial role in shaping public opinion on social issues and government policies [9]. With a user base dominated by people of working age, particularly the younger generation, TikTok actively fosters rapid and dynamic public discussions through its comment features and user interactions. Although TikTok's user demographics do not yet represent all segments of society, the high level of interaction and the speed of information dissemination make this platform relevant for analyzing public perceptions of government policies, including the MBG Program.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

In sentiment analysis research on government policies, various methods have been applied to understand public opinion in digital media. Previous studies have generally used data from platforms such as Twitter, Instagram, digital news portals, and online discussion forums as the primary sources for analysis. Several studies have also implemented text mining and Natural Language Processing (NLP) techniques using classification methods such as Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) to identify public sentiment toward public policy issues [5], [6], [7]. However, most research still focuses on text-based platforms like Twitter and has not yet widely utilized TikTok as a data source. In fact, TikTok has distinct interaction characteristics compared to other social media platforms, such as a high volume of comments, rapid information dissemination through audiovisual content, and a dominant user base of working-age individuals who actively voice opinions on social issues and public policy. These characteristics make TikTok a dynamic digital discussion space in shaping public perception of government policies. On the other hand, comments on TikTok which tend to be brief, informal, and heavily laden with slang also present unique challenges in sentiment analysis. Therefore, research utilizing comment data from TikTok is crucial for gaining a more contextual and comprehensive understanding of public responses to the MBG Program, while also expanding the scope of sentiment analysis on social media platforms that are relatively underutilized in public policy studies.

To address the limitations of previous research, this study employs a Natural Language Processing (NLP) approach to analyze unstructured TikTok user comments. The preprocessing steps include data cleaning, tokenization, slang normalization, stopword removal, and stemming, with the aim of improving text quality prior to classification [10]. Data was collected via web scraping from the TikTok platform, specifically comments related to the MBG Program. Text feature representation was performed using the Word2Vec method, which converts words into numerical vectors based on semantic relationships between words. Word2Vec was chosen for its ability to better represent word context compared to word-frequency-based methods such as TF-IDF [11]. This is particularly important given that TikTok comments tend to be brief, informal, use slang, and exhibit high linguistic contextual variation. For data labeling, this study employs a lexicon-based labeling approach that automatically categorizes sentiment into positive, negative, and neutral [12]. In the sentiment classification stage, the Support Vector Machine (SVM) algorithm was selected due to its stable performance on high-dimensional text data and its effectiveness on small to medium-sized datasets [13].

Unlike previous studies that used YouTube comments and the Multinomial Naïve Bayes algorithm with TF-IDF weighting, this study utilizes TikTok user comments as the primary data source to analyze public sentiment toward the Free Nutritious Meals Program (MBG). The choice of TikTok is based on the high level of user interaction and the more dynamic public discourse on this short-form video platform. Additionally, this study employs a combination of Word2Vec and Support Vector Machines (SVM) to represent semantic relationships between words and enhance classification performance on high-dimensional text data. The focus of this research is on analyzing public perceptions of the MBG program controversy through a Natural Language Processing (NLP) approach, aiming to provide a more contextual representation of sentiment compared to word-frequency-based methods such as TF-IDF [6].

Unlike previous studies that used data from the X (Twitter) platform and compared the Naïve Bayes and Logistic Regression algorithms with the SMOTE approach to address data imbalance, this study focuses more on analyzing the sentiment of TikTok user comments regarding the controversy surrounding the Free Nutritious Meals Program (MBG). This research combines the Word2Vec and Support Vector Machine (SVM) methods to represent semantic relationships between words and perform sentiment classification on unstructured comment text data. Additionally, a lexicon-based automatic labeling approach is also applied in the sentiment labeling process. The research focuses not only on improving the accuracy of the classification model but also on analyzing public perceptions regarding budget transparency, aid distribution, and the effectiveness of the MBG program's implementation. The use of TikTok as a data source is a significant contribution to this study, given that the platform is still relatively underutilized in sentiment analysis related to public policy compared to other social media platforms [14].

This study holds significant value both academically and practically in the development of sentiment analysis regarding public policy on social media. From an academic perspective, this study contributes by using TikTok comments as a data source for public policy sentiment analysis, a source that has so far been underutilized compared to platforms such as Twitter and Instagram. Additionally, this study combines the Word2Vec and Support Vector Machine (SVM) methods to represent semantic relationships between words and improve classification accuracy on unstructured social media comment data. The use of a lexicon-based labeling approach in labeling public opinion data related to the MBG Program also serves as an additional contribution to the development of sentiment analysis methods based on Natural Language Processing (NLP). From a practical perspective, the results of this study are expected to provide insights into public perceptions, criticisms, and acceptance levels regarding the implementation of the Free Nutritious Meals (MBG) Program, particularly concerning budget transparency, distribution effectiveness, and on-the-ground implementation. These findings can serve as evaluation material for the government in designing more transparent, responsive, and targeted public communication policies and strategies [15].

Given the various limitations and gaps in previous research, a more comprehensive sentiment analysis of public opinion regarding the Free Nutritious Meals Program (MBG) is needed, utilizing comment data from the TikTok platform. This study aims to examine public sentiment toward the controversies surrounding the MBG Program using a Natural Language Processing (NLP) approach. The analysis was conducted by combining a lexicon-based labeling

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

method for automatic sentiment labeling, text feature representation using Word2Vec, and a Support Vector Machine (SVM) algorithm for sentiment classification. This combination of methods is expected to capture the semantic relationships between words in a more contextual manner and improve classification accuracy on social media comment data, which tends to be brief, informal, and unstructured. In addition to focusing on the performance of the classification model, this study also aims to uncover public perceptions regarding the implementation of the MBG Program, particularly concerning budget transparency, the effectiveness of aid distribution, and program implementation on the ground. The research results are expected to contribute to the development of social media-based sentiment analysis studies and serve as evaluation material for the government in designing public communication strategies that are more effective, transparent, and responsive to public aspirations.

2. RESEARCH METHODOLOGY

2.1 Research Stage

The purpose of This study evaluates the community opinion on Free Nutritious Meals (FNM) Program by analyzing comments posted by TikTok users. A systematic outline of the research process—from data collection to the evaluation of classification models illustrated in Figure 1.

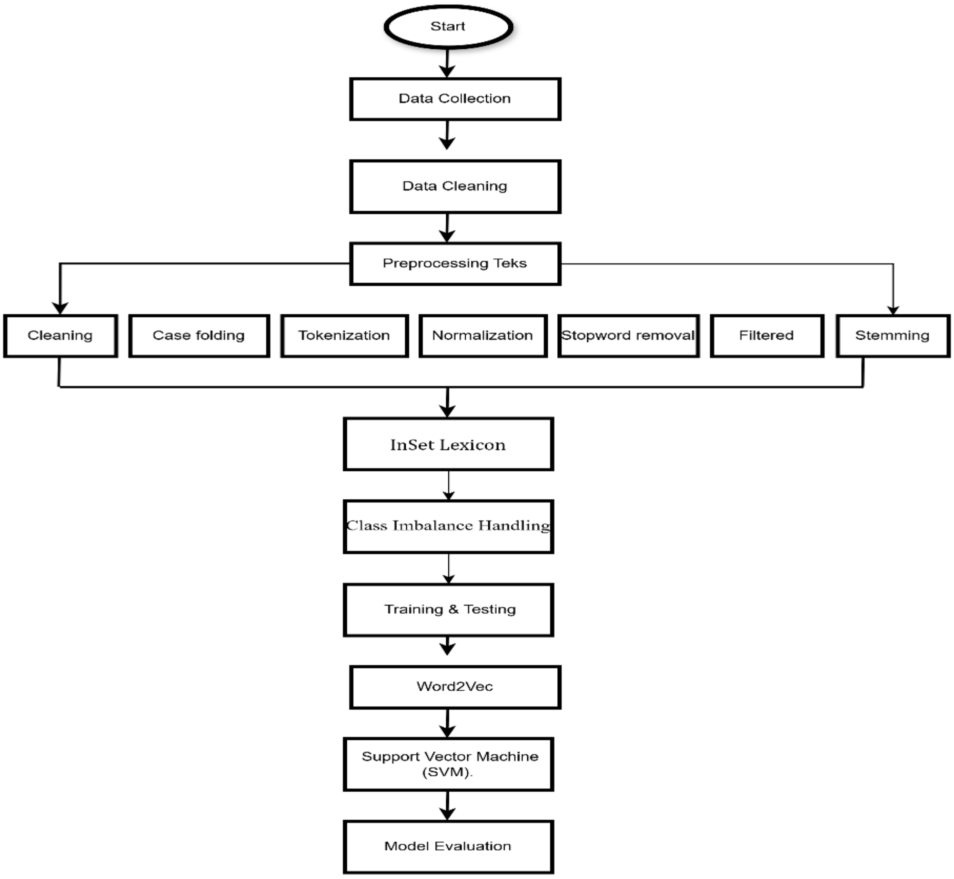

Figure 1. Research flowchart Source: Author's documentation.

2.2 Data Collection

Collecting data process includes the collection of key information sources for analysis. Data is obtained from TikTok platform using Apify's web scraping technique. This technique captured comments related to MBG and collected 2,381 random comments that reflected the public's spontaneous and unstructured perspectives, From a TikTok web scraping dataset comprising attributes such as comment, dig_count, reply_count, username, and url, this study uses only the comment column as the object of analysis [16].

2.3 Data Cleaning

Preprocessing removes irrelevant text data elements, such as hashtags, mentions, emojis, numbers, punctuation marks, special characters, duplicate data, empty text, metadata generated during data collection, and stopwords. Additional filtering and stemming processes are also applied to ensure that only relevant and meaningful base words for sentiment analysis are used.After this stage, the dataset was reduced to 2,376 comments ready for [17].

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

2.4 Data Filtering

Next, a filtering process was conducted to ensure that the data used was truly relevant to the research topic. At this stage, words with little emotional content, such as “uid,” “avatar,” “cid,” and “createTimeISO,” were removed because they did not contribute to the sentiment analysis [18].

2.5 Preprocessing Teks

Before sentiment analysis was performed, text preprocessing was conducted to improve the quality and consistency of the text data. This process is crucial for reducing noise and transforming disorganized comments on TikTok into a cleaner, more easily understandable format. Case folding, tokenization, normalization, stopword removal, and stemming are some of the preprocessing steps used in this study.All characters are converted to lowercase using the case folding method to eliminate inconsistencies caused by differences in capitalization. Afterward, each sentence is split into individual word tokens for tokenization. Additionally, normalization is used to standardize informal words, abbreviations, and slang commonly found in TikTok comments into standard Indonesian. Finally, the stemming process converts words into their root forms to make the text more readable and enhance the effectiveness of sentiment classification[19].

2.6 Data Slang

To address the use of slang and informal language that frequently appears in TikTok comments, this study employs a slang normalization technique using a manually compiled slang dictionary. During the pre-processing stage, each word/token is checked against the dictionary and automatically converted into standard Indonesian [20]. This approach aims to reduce linguistic variation caused by abbreviations, non-standard spellings, exaggerated expressions, and internet slang commonly used in social media communication. For example, terms like “gk,” “ga,” and “nggak” are normalized to “tidak,” while expressions like “mantappp” and “bagusss” are standardized to “mantap” and “bagus.” Additionally, TikTok-specific terms like “fix,” “real,” and “relate” are converted into forms with clearer semantic meanings. By standardizing these variations, the preprocessing step can improve text consistency as well as the effectiveness of feature extraction and sentiment classification.

2.7 Lexicon-Based Labeling

The sentiment labeling process in this study utilizes the InSet Lexicon, which was specifically developed for sentiment analysis in the Indonesian language. Automatic sentiment labeling is performed using a lexicon-based approach, which involves matching the words in each comment against a dictionary of positive and negative sentiments already available in the lexicon. Sentiment polarity is determined based on the total sentiment score of each word in the comment. Comments with a positive total score are classified as positive sentiment, comments with a negative total score fall into the negative sentiment category, while comments with a score of zero are categorized as neutral sentiment. In this study, the lexicon-based approach is used as an automatic annotation method to efficiently assign initial sentiment labels to large amounts of text data, prior to further classification using a Support Vector Machine (SVM) model [21], [22]. Although lexicon-based labeling may not yet be fully capable of capturing the sarcasm and contextual nuances found in social media language, this method remains a practical approach for large-scale sentiment annotation, particularly in research with limited resources.

2.8 Class Imbalance Handling

There was a notable class imbalance in the dataset utilized in this investigation, where positive sentiment comments represented less than 10% of the overall dataset compared to negative and neutral sentiments. This imbalance could affect the model's ability to accurately classify minority groups class data. To address this issue, this support vector machine model applied this class_weight = balanced parameter, which automatically adjusted class weights during the training process according to the proportion of each sentiment class. This approach helped improve classification performance for minority sentiment classes while maintaining overall model stability [23].

2.9 Word2Vec

The Word2Vec method is used to convert text into numerical vector representations, thereby improving the quality of data representation. This method is more effective at capturing context than frequency-based methods because this system relies on the semantic relationships between words [24]. Combining word vectors into a single document vector yields the final representation for each comment In this study, the Word2Vec model was trained using the CBOW architecture, which has a vector size of 200, a window size of 5, a minimum word frequency (min_count) of 0, and 10 training intervals.

$$V_d = \frac{1}{n} \sum_i^n = 1 \quad V_i \tag{1}$$

where (V_d) represents the document vector, (V_i) represents the vector representation of each word token, and (n) indicates the number of words in document [24].

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

2.10 Training and Testing

The preprocessed dataset was divided into 30% test data and 70% training data. The training data was used to build a classification model, while the test data was used to evaluate the model's performance [25].

2.11 Classification Using SVM

The sentiment classification process used the The model used is the Support Vector Machine (SVM) algorithm, which employs a Radial Basis Function (RBF) kernel. model applied a regularization parameter of C = 5 and gamma = scale. In addition, the class_weight= balanced parameter was used to address the class imbalance problem, particularly because positive sentiment comments represented a smaller proportion of the dataset. The hyperparameter values determined using experimental testing to obtain the best classification performance on the dataset[25]. where the y parameter determines the strength of a data point relative to other data points. This kernel is effective for handling data with non-linearly separable distributions [15], [26], [27].

Table 1. Kernel Formula in SVM

Kernel	Equation
Linear	$K(xi, x) = x$
Polinomina	$K(xi, x) = (y.x + r)p, y > 0$
RBF	$K(x_i, x) = \exp(-\gamma\|x_i - x\|^2), \quad \gamma > 0$
Sigmoid	$K(x_i, x) = \tanh(\gamma(x_i \cdot x) + r)$

Where the y parameter determines the strength of the relationship between two data points This kernel is effective for handling data with non-linearly separable distributions [26].

$$f(x) = w.x + b \tag{2}$$

where X is feature vector generated from the Word2Vec representation, W is the weight vector, and b is the bias [28].

2.12 Model Evaluation

The primary metrics used to the metrics used to evaluate model performance are accuracy, precision, recall, and F1 score. The calculations are performed using the following formulas:

a. Accuracy, Overall, the prediction was correct.

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

b. Precision, Prediction accuracy per class [28].

$$\frac{TP}{TP+FP} \tag{4}$$

c. Recall, Detection coverage per class [29].

$$\frac{TP}{TP+FN} \tag{5}$$

d. F1-Score,used to evaluate how well precision and recall [29].

$$F1 = 2\,x\ = \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

3. RESULTS AND DISCUSSION

The results of a sentiment analysis conducted on TikTok user comments regarding the Free Nutritious Meals Program (MBG) are presented in this section using Natural Language Processing (NLP) methods. Data preprocessing, lexicon-based sentiment labeling, Word2Vec feature extraction, and classification using the Support Vector Machine (SVM) algorithm are all part of this analysis. To demonstrate the contribution of this research, model performance was evaluated using F1 score, accuracy, precision, and recall.

3.1 Data Colection

Data collection took place successfully using Apify and the TikTok Comments Scraper tool. Thousands of public comments regarding the MBG program were extracted from TikTok through this process for further analysis, as seen in Table 2.

Table 2. Data collection

	Column	Non-Null Count	Dtype
0	Text	2381 non-null	object

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

	Column	Non-Null Count	Dtype
1	diggCount	2381 non-null	int64
2	replyCommentTotal	2381 non-null	int64
3	createTimeISO	2381 non-null	object
4	uniqueId	2381 non-null	object
5	videoWebUrl	2381 non-null	object
6	Uid	2381 non-null	int64
7	cid	2381 non-null	int64
8	avatarThumbnail	2381 non-null	object

Data collection was successfully conducted using Apify and the TikTok Comments Scraper tool. A total of 2,381 public comments related to the Free Nutritious Food Program (MBG) were extracted from TikTok and prepared for further analysis. The detailed characteristics of the collected dataset are presented in Table 2. Table 2 presents the structure of the dataset obtained through the scraping process, including variables such as text, diggCount, replyCommentTotal, createTimeISO, uniqueId, videoWebUrl, uid, cid, and avatarThumbnail. Among these variables, only the text variable was selected as the primary input for sentiment analysis because it contains users' opinions, responses, and public perceptions regarding the MBG program. The remaining variables were used only as supporting metadata for identifying user interactions and comment information and were excluded from the sentiment classification process. As shown in Table 2, all 2,381 records were successfully collected without missing values in the selected variables, indicating that the dataset was sufficiently complete for further preprocessing and sentiment analysis stages. This dataset provides valuable insights into public perceptions of the MBG program, particularly concerning policy implementation, food quality, budget transparency, and social impact.

3.2 Data Cleaning

Prior to additional analysis, the data cleaning step was carried out to enhance the quality of the gathered dataset. Symbols, numerals, URLs, duplicate letters, and unrelated text components were eliminated from TikTok comments as part of this procedure, as seen in Table 3.

Table 3. Ouput cleaning data

	text	cleaning
1	"hentikan program mbg, kaji ulang semua progra...	hentikan program mbg kaji ulang semua program ...
2	"Hentikan program mbg,""0"",""0"",""2026-01-1...	Hentikan program mbg ocrumble
3	"anggran habis sampe triliunan banyak yg korup...	anggran habis sampe triliunan banyak korupsi ...
4	"bubarkan aja pak program ga bermutu ini,""1""...	bubarkan aja pak program bermutu ini halimp...
5	"sebelum ribuan,""1"",""0"",""2025-10-25T13:07...	sebelum ribuan mie

The cleaning stage is used to identify unnecessary elements in the TikTok content, such as text, URLs, numbers, symbols, duplicate characters, and irrelevant metadata obtained during the data collection process. As shown in Table 3, the cleaning process successfully transformed messy and unstructured text into text data that is more accurate and better suited for analysis, because TikTok comments often contain informal writing styles and unclear text, this process is important because comments containing additional symbols, timestamps, and irrelevant characters are simplified into more readable text while preserving the main semantic meaning.

3.3 Preprocessing Teks

A case folding procedure was used to normalize all text data into lowercase letters following the data cleaning step. The goal of this procedure was to improve the efficacy of the text preprocessing stage by minimizing inconsistencies brought on by variations in letter capitalization, explained in table 4.

Table 4. Output Case Folding

	cleaning	case_folding
1	hentikan program mbg kaji ulang semua program ...	hentikan program mbg kaji ulang semua program ...
2	Hentikan program mbg ocrumble	hentikan program mbg ocrumble
3	anggran habis sampe triliunan banyak korupsi ...	anggran habis sampe triliunan banyak korupsi ...
4	bubarkan aja pak program GAK bermutu ini	bubarkan aja pak program gak bermutu ini
5	sebelum ribuan mie	sebelum ribuan mie

To reduce inconsistencies caused by variations in capitalization, a process called case folding is used to standardize all text to lowercase. As a result, words with uppercase letters or mixed case, such as “GAK” and “Hentikan,” are converted to lowercase, as shown in Table 4. This process ensures that words with the same meaning are treated as identical tokens during analysis. The case conversion step reduces textual variation. It also improves the

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

consistency of the text data and supports more effective feature extraction and sentiment classification, can be seen in Figure 2.

Figure 2. Word Cloud Case Folding

Findings from the word cloud following the case conversion indicate that conversations about the MBG program are dominated by terms related to food, nutrition, government programs, and public issues. This suggests that the case conversion step improves text quality, facilitates feature extraction, and enhances the effectiveness of subsequent sentiment analysis processes.

3.4 Normalisasi

In order to transform informal, shortened, and non-standard words that are frequently found in TikTok comments into their standard versions, the normalization stage was completed. By guaranteeing that words with similar meanings are rendered consistently, this procedure aids to increase text consistency and raises the caliber of sentiment analysis.

Table 5. shows the output of the normalization stage.

	Case folding	normalisasi
1	hentikan program mbg kaji ulang semua program ...	hentikan program makan bergizi gratis kaji ula...
2	hentikan program mbg ocrumble	hentikan program makan bergizi gratis ocrumble
3	anggran habis sampe triliunan banyak korupsi ...	anggaran habis sampe triliunan banyak korupsi ...
4	bubarkan aja pak program tidak bermutu ini	bubarkan aja pak program tidak bermutu ini
5	sebelum ribuan mie	sebelum ribuan mie

The normalization process is used to convert informal words, abbreviations, and non-standard terms commonly found in TikTok comments into standard Indonesian. Table 5 shows that terms such as “mbg” are normalized to “makan bergizi gratis” (free nutritious meals), and terms such as ‘anggran’ are corrected to “anggaran” (budget). This method improves the consistency of the text.

3.5 Tokenizing

The tokenization stage was performed by splitting each normalized sentence into individual word units (tokens). This process allows the text data to be represented as a sequence of words that can be further processed in Natural Language Processing (NLP). For example, the sentence “hentikan program makan bergizi gratis” was transformed into tokens such as [hentikan, program, makan, bergizi, gratis]

Table 6. shows the tokenization output.

	normalisasi	tokenizing
1	hentikan program makan bergizi gratis kaji ula...	[hentikan, program, makan, bergizi, gratis, ka...
2	hentikan program makan bergizi gratis ocrumble	[hentikan, program, makan, bergizi, gratis, oc...
3	anggaran habis sampe triliunan banyak korupsi ...	[anggaran, habis, sampe, triliunan, banyak, ko...
4	bubarkan aja pak program bermutu ini halimpadang	[bubarkan, aja, pak, program, bermutu, ini, ha...
5	sebelum ribuan mie	[sebelum, ribuan, mie]

To complete the tokenization stage, each normalized sentence is broken down into individual words or tokens. For example, “stop the free nutritious meal program” is converted to “stop, program, meal, nutritious, free,” as shown in Table 6. This process enables the representation of textual data in a structured format that can be processed computationally through natural language processing (NLP). Tokenization is crucial because it simplifies text analysis by focusing on word-level representations. It also allows for further preprocessing steps, such as feature extraction, sentiment classification, and stopword removal, to be performed more efficiently.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

3.6 Stopword

In order to exclude unusual words significantly add plays a major role in sentiment analysis, the stopword elimination stage was carried out. This stage aids in cutting out superfluous words and keeping only those that are crucial to the subject under study.

Table 7. Stopword Display Output

	tokenizing	stopword
1	[hentikan, program, makan, bergizi, gratis, ka...	[hentikan, program, makan, bergizi, gratis, ka...
2	[hentikan, program, makan, bergizi, gratis, oc...	[hentikan, program, makan, bergizi, gratis,oc…
3	[anggaran, habis, sampe, triliunan, banyak, ko...	[anggaran, habis, sampe, triliunan, korupsi, t...
4	[bubarkan, aja, pak, program, tidak, bermutu, ini, ha...	[bubarkan, pak, program, tidak, bermutu,]
5	[sebelum, ribuan, mie]	[sebelum, ribuan, mie]

Common words that are irrelevant to sentiment analysis and lack significant semantic value are removed through the stopword removal process. Table 7 shows that important terms related to sentiment remain, but uninformative words such as “just,” “this,” and other frequently occurring function words are removed from the tokenized text. This process reduces noise in the dataset and allows the model to focus on more meaningful words, which are better able to capture the context and sentiment of the comments. As a result, classification performance and feature extraction quality can be improved.

3.7 Filtered

The filtering stage was performed to remove words that were considered irrelevant or unrelated to the sentiment analysis process. This step aims to retain only meaningful words that contribute to identifying public sentiment toward the MBG program.

Table 8. Display of the word cloud results and filtered output

	stopword	filtered
1	[hentikan, program, makan, bergizi, gratis, ka...	[hentikan, program, makan, bergizi, gratis, ka...
2	[hentikan, program, makan, bergizi, gratis, oc...	[hentikan, program, makan, bergizi, gratis]
3	[anggaran, habis, sampe, triliunan, korupsi, t...	[anggaran, habis, sampe, triliunan, korupsi, t...
4	[bubarkan, pak, program, tidak, bermutu, halimpadang]	[bubarkan, program, tidak, bermutu]
5	[sebelum, ribuan, mie]	[sebelum, ribuan, mie]

After the stopword removal process, a filtering step is performed to eliminate irrelevant tokens, uncommon words, and terms that do not contribute significantly to sentiment analysis. Table 8 shows that some tokens deemed meaningless or irrelevant to the research context were removed, resulting in cleaner and more focused textual data. Words such as usernames, vague terms, or location tokens are removed, while words related to emotions are retained. By reducing noise and ensuring that the remaining tokens better reflect the discussion topics and emotions expressed in TikTok comments, this process helps improve the quality of the dataset.

In addition to stopword removal, the filtering process was further applied to identify and remove words that were not relevant to the sentence analysis. At this stage, only words containing meaningful and relevant information were retained for further processing. Tokens such as “baik”, “cocok”, “tepat”, and “target” were preserved because they contribute to sentiment interpretation. In addition, the word frequency visualization showed that terms such as “makan”, “gratis”, “bergizi”, “anak”, and “pak” appeared most frequently in the dataset.

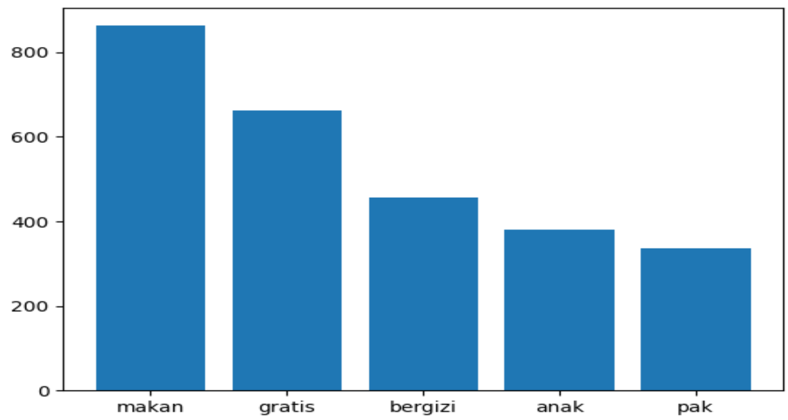

Figure 3. word frequency diagram

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

3.8 Stemming

The stemming stage was performed to transform each word into its root form in order to reduce variations of words with similar meanings. This process helps improve consistency in text representation and enhances the effectiveness of sentiment classification.

```
...                                              filtered  \
 1  [hentikan, program, makan, bergizi, gratis, ka...
 2         [hentikan, program, makan, bergizi, gratis]
 3  [anggaran, habis, sampe, triliunan, korupsi, t...
 4                       [bubarkan, program, bermutu]
 5                              [sebelum, ribuan, mie]

                                                 stemming
 1  [henti, program, makan, gizi, gratis, kaji, ul...
 2                [henti, program, makan, gizi, gratis]
 3  [anggar, habis, sampe, triliun, korupsi, tolol...
 4                              [bubar, program, mutu]
 5                                   [belum, ribu, mie]
```

Figure 4. Output Result Stemming

This process aims to reduce variations in words that are different but have the same meaning forms, thereby making the text representation more consistent. Based on the results obtained, words such as “hentikan” were changed to “henti,” “bergizi” to “gizi,” “anggaran” to “anggar,” and “bubarkan” to “bubar.” These changes demonstrate that the stemming process successfully removes affixes such as prefixes and suffixes without altering the core meaning of the words.

3.9 Lexicon-Based Labeling

Sentiment labeling was performed using a lexicon-based approach by assigning scores to each word based on a predefined sentiment dictionary.

Table 9. Output clas sentimen

sentiment	
Negatif	1377
Netral	769
Positif	230
Name: count, dtype: int64	

The results of the lexicon-based labeling process show the distribution of sentiment classes in the dataset Table 9. With 1,377 comments, negative sentiment dominates the dataset, followed by neutral sentiment with 769 comments and positive sentiment with 230 comments. This analysis shows that the majority of TikTok users hold negative views about the MBG program, particularly regarding transparency, budget management, food quality, and policy implementation. Positive comments express support for and appreciation of the program, while neutral comments are typically informative or discussion-oriented, without strong emotional undertones. In addition, the imbalance in sentiment distribution indicates that the number of positive samples in the dataset is smaller than expected, which could affect how classification is performed for the minority class.

Table 10. Score Sentimen

	stemming	score	sentiment
1	[henti, program, makan, gizi, gratis, kaji, ul...	-6	Negative
2	[henti, program, makan, gizi, gratis]	-3	Negative
3	[anggaran, habis, sampe, triliun, korupsi, tolol...	-7	Negative
4	[bubar, program, tidak, mutu]	-3	Negative
6	[gila, banget, program, tidak, jelas, jadi, pr...	-7	Negative
7	[gak, perlu, penting, makan, gizi, gratis, nya...	-1	Negative
8	[lapar, urus, perintah, bang, kasih, tiap, bul...	-4	Negative
9	[sabar, presidan, harus, lebih, lihat, lagi, program]	0	Neutral
10	[biasa, racun, bahaya, jika, buat, tidak, benar]	-8	Negative

Examples of sentiment analysis results generated using the InSet Lexicon method are presented in Table 10. The polarity scores of the words in the text, following preprocessing and stemming, were used to assign a sentiment score to each comment. Comments with a negative total score are considered negative sentiment, comments with a positive total score are considered positive sentiment, and comments with a zero total score are considered neutral sentiment. The table shows that comments containing words such as “corruption,” “poison,” “danger,” and “stupid” received very negative scores, indicating that the public is critical of the MBG program. Comments containing more

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

neutral words, such as “suitable” and “target,” received a score of zero. These results indicate that the lexicon-based labeling process is capable of capturing the polarity of public opinion reflected in TikTok comments.

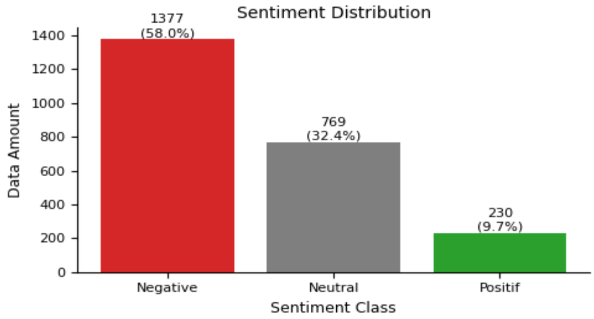

Figure 5. Sentiment class distribution

Figure 5 shows the sentiment class distribution derived from TikTok comments regarding the MBG program. The results indicate that negative sentiment dominates the dataset with 1,377 comments (58.0%), followed by neutral sentiment with 769 comments (32.4%), and positive sentiment, which represents the smallest proportion of the dataset with 230 comments (9.7%). This distribution indicates that public responses to the MBG program on TikTok tend to be more negative than positive, particularly regarding transparency, budget management, food quality, and policy implementation. Positive comments express appreciation and support for the program, while neutral comments typically contain informative discussions or opinions without strong emotional bias. Additionally, the imbalance in sentiment distribution indicates that the number of positive samples in the dataset is smaller, which may impact the model’s ability to classify, particularly for minority sentiment classes.

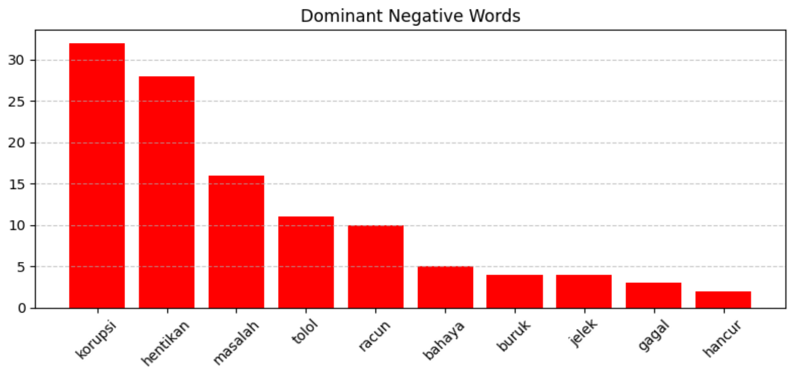

Figure 6. Output display for the negative class

Figure 6 shows the words most frequently used in negative comments about the MBG program. Terms such as “corruption,” “stop,” “problem,” “stupid,” and “poison” are among the most frequently used, indicating that many users are expressing criticism, distrust, and concern about how the program is being implemented. Words such as “danger,” “bad,” “failure,” and “ruin” further indicate a negative view of food security, policy management, and government performance. The prevalence of these negative expressions suggests that public criticism on TikTok is largely focused on transparency, budget management, and the quality of the MBG program. Furthermore, these findings support the sentiment distribution results, which show that negative sentiment constitutes the largest proportion of comments found in the dataset.

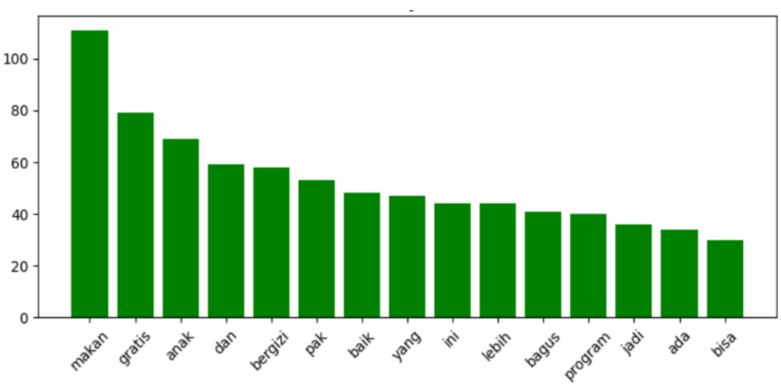

Figure 7. Output display for the positive class

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

Figure 7 shows the words that appear most frequently in positive comments about the MBG program. Words such as “eat,” “free,” “children,” “nutritious,” and ‘good’ indicate that supportive comments are typically associated with the benefits of providing free nutritious meals to children. Other frequently used words, such as “great,” “more,” and “can,” reflect positive views regarding the program’s objectives and its potential social impact. These findings suggest that users who express positive attitudes tend to value programs that improve children’s nutrition and support community well-being. While the number of positive comments is smaller than that of negative sentiment, the nature of the positive comments indicates that some TikTok users view the MBG program as beneficial and useful for the community.

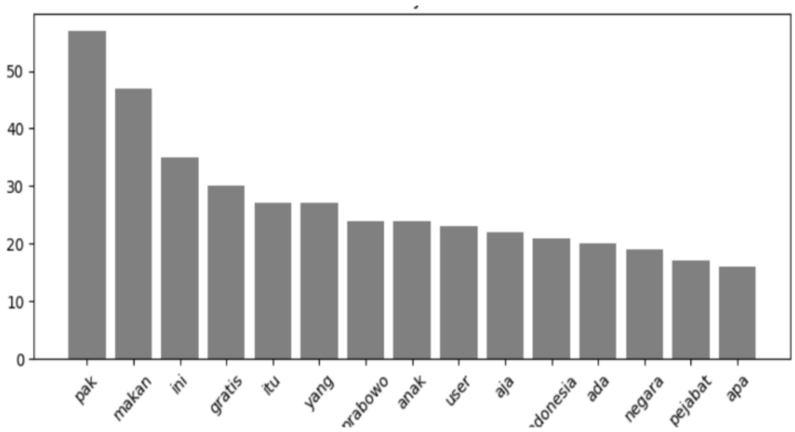

Figure 8. Output of the Neutral class

Figure 8 shows the words that appear most frequently in comments related to the MBG program with neutral sentiment. Terms such as “sir,” “food,” “free,” “Prabowo,” “children,” and “country” indicate that neutral comments largely focus on general conversation, informational statements, and public discussion regarding the program's implementation. Neutral sentiment differs from negative comments in that they do not explicitly criticize or endorse. Instead, they focus more on discussions about government policies, public figures, and the overall objectives of the free meal program. The results show that neutral comments are essentially descriptive and informative, reflecting public interest in the MBG program without significant emotional bias.

Table 11. topic modeling Negative

Topic Negative	Dominant word	Interprestasi
Topic Negatif 1	makan, gratis, gizi, program, racun, gak, tidak, juga, aku, uang	This topic highlights public concerns about the quality and safety of the free nutritious meal program. The use of words such as “racun,” “tidak,” and “uang” suggests that some social media users doubt the program’s effectiveness and associate it with potential waste of government funds.
Topic Negative 2	makan, jabat, uang, jangan, gak, rakyat, hari, belum, dulu, sudah	This topic reflects public criticism of government officials or agencies regarding the management of programs. The words “jabat,” “uang,” and “rakyat” indicate negative perceptions regarding the use of public funds, as well as the belief that the needs of other members of the public have not yet been fully met.
Topic Negative 3	makan, gratis, gizi, racun, program,	This topic reflects the public’s skepticism regarding the priority of implementing the free nutritious meal program. The use of the phrase “masih butuh” indicates that some members of the public believe there are other needs that are considered more urgent than the implementation of this program.
Topic Negative 4	masak, jam, banget, sppg, basi, perintah, rakyat, negara, daerah, program	This topic concerns technical issues related to food distribution and service. The terms “basi,” “masak,” and “SPPG” indicate complaints regarding food quality, processing procedures, and program distribution in several regions.
Topic Negative 5	makan, tidak, guru, racun, sehat, mampu, bukan, bagi, kurang, dapur	This topic reflects concerns about health standards and the program’s targeting. The words “guru,” “sehat,” and “mampu” suggest a discussion about who is eligible to receive program benefits and the quality of the food provided.

The results indicate that topic modeling provides a deeper understanding than word cloud visualization, as it is able to identify the key issues that trigger negative sentiment among the public.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

Table 12. Topic modeling Positive

Topic Positive	Dominant word	Interprestasi
Topic Positive 1	baik, sukses, manfaat, nak, sangat, bantu,	This topic highlights public support for the benefits of free nutritious meal programs for children. The words “sukses,” “manfaat,” and “bantu” indicate the hope that the program can help meet nutritional needs and support the future of the younger generation.
Topic Positive 2	sehat, bantu, manfaat, presiden, baik, kecil, ada, dana, kurang, bikin	This topic reflects the public’s appreciation for the government’s efforts to improve children’s health and well-being. The use of the words “sehat,” “presiden,” and “manfaat” indicates support for government policies that are seen as having a positive impact on society.
Topic Positive 3	bagus, bantu, baik, kerja, guru, jangan, nanti, kasih, cepat, mantap	This topic reflects a positive assessment of the program’s implementation within the school environment. The words “guru,” “baik,” and “mantap” indicate that the community believes the program can assist students’ learning activities and support the educational process.
Topic Positive 4	aku, bantu, buang, baik, gin, rakyat, layak, bagus, evaluasi, kalian	This topic indicates that the public supports the continuation of the program, provided that evaluations and improvements are taken into account. The words “rakyat,” “layak,” and “evaluasi” suggest that the public supports the program as long as it is implemented effectively and continuously improved.
Topic Positive 5	baik, tiap, sangat, kasih, senang, awas, penting, hari, bapak, terimakasih	This topic reflects the public’s satisfaction and appreciation for the free nutritious meal program. The words “senang,” “terimakasih,” and “penting” indicate that many people feel the program provides tangible benefits and is worth continuing.

Overall, the results of topic modeling on positive sentiment indicate that the public supports the free nutritious meals program because it is considered to:

Help meet children’s nutritional needs, Improve public health, Support the educational process in schools, Provide social benefits to the people, Be a beneficial program worthy of continuation. These results demonstrate that the topic modeling method provides a deeper understanding of the reasons behind the public’s positive sentiment compared to relying solely on word cloud visualizations.

Table 13. Topic modeling Neutral

Topic Neutral	Dominant word	Interprestasi
Topic Neutral 1	negara, kelola, apakah, pasti, program, ini	This topic reflects public discussions regarding the government’s management of programs. The use of the words “kelola,” “negara,” and “apakah” indicates a conversation that questions or discusses the mechanisms for implementing programs without directly expressing a positive or negative stance.
Topic Neutral 2	sehat, lanjut, perbaikin	This topic relates to discussions about food quality and the implementation of programs in schools or cafeterias. The words “sehat,” “lanjut,” and “perbaikin” indicate that the discussion revolves around food service, but they do not explicitly indicate support for or opposition to the program.
Topic Neutral 3	uang, buang, sia,	This topic reflects general conversations on social media discussing issues such as spoiled food, the use of money, and references to specific figures. However, the context of these discussions tends to consist of comments or informal conversations, so they cannot be clearly categorized as either positive or negative sentiment.
Topic Neutral 4	sri, lanjut, rakyat, presiden, mulyani, allah, jabat, tuju, bener, uang	This topic reflects public discussions regarding government policies and the actions of government officials in implementing programs. The terms “presiden,” “Sri Mulyani,” and “uang” indicate discussions about policy and budgetary issues without a clear bias toward either support or opposition.
Topic Neutral 5	jabat, guru, ganti, aman, betul, dek, salsa, kak, cerdas, aku	This topic reflects general conversations among social media users regarding the implementation of programs in educational settings. The words “guru,” “aman,” and “cerdas” indicate that the discussions are informative in nature or consist of casual conversations without strong emotional sentiment.

Overall, the topic modeling results for neutral sentiment indicate that public discussions mainly focused on government management of the program, the quality of school meals, general conversations on social media, program

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

policies and budgets, as well as informative discussions regarding program implementation. These findings suggest that neutral sentiment primarily reflects informational exchanges and public discussions without strong emotional tendencies, either positive or negative, toward the Free Nutritious Meals (MBG) program.

3.10 Split Data (Train-test)

The training dataset is used to train the classification model, while the test dataset is used to evaluate the model's performance. The ratio between the two datasets is 70:30. In sentiment classification.

Table 14. shows the code and results for distinguishing between test and training data.

Train:		Test:	
sentiment		sentiment	
Negative	964	Negative	413
Neutral	538	Neutral	231
Positive	161	Positive	69

The train-test splitting results show that the training dataset consisted of 964 negative comments, 538 neutral comments, and 161 positive comments. Meanwhile, the testing dataset contained 413 negative comments, 231 neutral comments, and 69 positive comments. These results indicate that negative sentiment dominated both the training and testing datasets, while positive sentiment represented the smallest portion of the data. This distribution reflects the imbalance of sentiment classes in the collected TikTok comments regarding the MBG program.

3.11 Word2Vec

After the preprocessing and sentiment labeling were completed, the Word2Vec method was used to represent the text data as numerical vectors.

Figure 9 shows the Word2Vec vector representations generated from tokenized comments on TikTok. The Continuous Bag of Words (CBOW) architecture was used to convert each word in the study into a 300-dimensional dense vector. The figure displays numerical values representing the semantic characteristics learned from the contextual relationships of words in the dataset. Word2Vec has the ability to identify semantic relationships and contextual similarities between words, which distinguishes it from conventional frequency-based techniques. For example, due to the context of the comments, words related to criticism, risk, or negative expressions may be placed closer together in the vector space. Hasilnya adalah vektor berdimensi 300 dengan nilai-nilai seperti 0,252431, -0,333062, -0,582866, dan seterusnya.

```
Data ke-0
============================================================
Representasi Teks (Token):
['sdh', 'jelas', 'resiko', 'masalah', 'racun', 'ribet', 'jaga', 'bersih', 'repot', 'pihak', 'blm',

Representasi Vektor (Word2Vec):
Dimensi: 300
Nilai awal: [ 0.25243104 -0.33306249 -0.58286599 -0.50659149  0.35087918]

 0.252431 -0.333062 -0.582866 -0.506591  0.350879  0.397111 -0.581309 -0.364867
-0.045649  0.243764 -0.280373  0.199828 -0.367278 -0.082067  0.602070  0.075374
-0.388879  0.245985  0.702462  0.069122  0.488037 -0.057892 -0.309251 -0.096654
-0.398070 -0.073617  0.411355 -0.321888  0.535100  0.437296 -0.072803  0.333704
-0.155028 -0.282489  0.187403 -0.466423 -0.261299  0.216460  0.281257  0.072204
 0.445153 -0.539263 -0.393761  0.063952 -0.508180 -0.439135  0.117161 -0.447205
-0.033098 -0.372924  0.590425 -0.511315  0.233576 -0.219026  0.578052 -0.287779
-0.228669 -0.654172 -0.149656 -0.502162  0.183833  0.282212 -0.016058 -0.364004
-0.607088 -0.731715  0.949587 -0.915422  0.355070  0.684078 -0.430152 -0.334325
-0.246429  0.138727  0.134165 -0.634685  0.318582 -0.493299  0.404181 -0.482761
 0.325423  0.353533  0.053126 -0.361913 -0.141358 -0.173009  0.278739 -0.553766
-0.455340  0.061065 -0.348872  0.311757 -0.345271 -0.506477 -0.419522 -0.282527
-0.120662  0.394061  0.050923 -0.294744  0.528474 -0.302769 -0.407318 -0.068544
-0.284574  0.322794  0.699791 -0.135594  0.290308 -0.297315  0.411666  0.482776
 0.503636 -0.363480 -0.049133 -0.088808  0.332093  0.461875 -0.444387  0.321324
-0.489604 -0.160757 -0.622796  0.426679  0.392274 -0.274099 -0.304785  0.493314
-0.240931 -0.324291 -0.512539 -0.369661  0.044592  0.289323 -0.378982 -0.267857
 0.503900  0.551430  0.173940  0.350944 -0.230446  0.296882  0.175143 -0.396036
 0.050094  0.805798  0.268070  0.068858 -0.508367  0.345914  0.018767  0.303424
```

Figure 9. Dimensi Word2vec 300

```
Data ke-0
============================================================
Representasi Teks (Token):
['sdh', 'jelas', 'resiko', 'masalah', 'racun', 'ribet', 'jaga', 'bersih', 'repot', 'pihak', 'blm', 'tentu', 'saji', 'higinis', 'cukup', 'nila

Representasi Vektor (Word2Vec):
Dimensi: 200
Nilai awal: [ 0.00025149 -0.00143085  0.00114851  0.00431142  0.00588752]
```

Figure 10. Dimensi Word2vec 200

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

An example of Word2Vec vector representations generated from tokenized TikTok comments using a 200-dimensional vector space is shown in Figure 10. This study uses Word2Vec with the CBOW architecture to convert text data into dense numerical vectors capable of identifying semantic relationships between words. By using a 200-dimensional representation, the goal is to assess the impact of vector dimension on sentiment classification performance by comparing it with more complex representations, such as a 300-dimensional representation. Subsequently, a Support Vector Machine (SVM) classifier uses the generated vectors. The comparison results show differences in classification accuracy and efficiency, particularly for minority sentiment classes. This indicates that vector dimensions and class balancing strategies can influence how effective a sentiment classification model is

3.12 Support Vector Machine Algorithm

Once support vector machine model has been has studied at training information, the next step is test it on he data being tested.This evaluation process uses Metrics such as accuracy, precision, recall, and F1 score are used to measure how well a model performs in identify. all sentiment categories.

Table 15. SVM Classification Accuracy Results with Balance

	precision	recall	f1-score	support
Negative	0.93	0.78	0.85	413
Neutral	0.74	0.94	0.83	231
Positive	0.70	0.71	0.71	69
Accuracy			0.83	713
macro avg	0.79	0.81	0.79	713
weighted avg	0.84	0.83	0.83	713

As presented in Table 15, the Support Vector Machine (SVM) model achieved an overall accuracy of 83% on the test dataset consisting of 713 samples. Model performance was evaluated using precision, recall, and F1-score metrics for each sentiment class. For the negative class, the model achieved a precision of 0.93, a recall of 0.78, and an F1-score of 0.85, indicating strong performance in identifying negative sentiment despite the presence of some false negatives. As shown in Table 15, the model also demonstrated good performance in detecting the neutral class, achieving a precision of 0.74, a recall of 0.94, and an F1-score of 0.83. In contrast, the positive class showed relatively lower performance, which may be influenced by class imbalance in the dataset. Based on the evaluation results in Table 15, the model obtained an overall macro F1-score of 0.79 and a weighted F1-score of 0.83. Overall, these findings indicate that the SVM model performed consistently well, particularly for the dominant sentiment classes.

Table 16. SVM Classification Accuracy Results Without Balancing

	precision	recall	f1-score	support
Negative	0.89	0.75	0.82	413
Neutral	0.73	0.93	0.82	231
Positive	0.62	0.61	0.61	69
Accuracy			0.80	713
macro avg	0.75	0.76	0.75	713
weighted avg	0.81	0.80	0.80	713

As presented in Table 16, the Support Vector Machine (SVM) model without applying class balance weight achieved an overall accuracy of 80% on the test dataset consisting of 713 samples. The model's performance was evaluated using precision, recall, and F1-score metrics for each sentiment class. Based on the results shown in Table 16, for the negative class, the model achieved a precision of 0.89, a recall of 0.75, and an F1-score of 0.82, indicating that the model performed reasonably well in identifying negative sentiment. For the neutral class, the model demonstrated good performance with a precision of 0.73, a recall of 0.93, and an F1-score of 0.82, suggesting that the model was able to accurately classify most neutral sentiment data. In contrast, the positive class showed the lowest performance, with a precision of 0.62, a recall of 0.61, and an F1-score of 0.61. As shown in Table 16, this lower performance may be influenced by class imbalance due to the smaller amount of positive data compared to the other sentiment classes. Overall, the evaluation results in Table 16 indicate that the model achieved a macro average F1-score of 0.75 and a weighted average F1-score of 0.80. These findings suggest that the model was relatively stable in performing sentiment classification but still faced limitations in recognizing minority classes, particularly positive sentiment.

3.13 Discussion

Initially, the InSet Lexicon was used for sentiment labeling, specifically designed for Indonesian sentiment analysis. According to the lexicon, each token is assigned a polarity score. A negative score indicates negative sentiment, a positive score indicates positive sentiment, and a score of zero indicates neutral sentiment. Based on how the sentiment is distributed, this score indicates that the sentiment is distributed neutrally. the dataset contains more negative sentiment (1,377 comments, or 58.0%) than neutral sentiment (769 comments, or 32.4%) and positive sentiment (230 comments, or 9.7%). This indicates that public response to the MBG program on TikTok tends to be negative, particularly regarding transparency, food safety, and policy implementation.

After sentiment labeling, the Word2Vec method is used to convert the tokenized text into numerical vector representations. This is done using the CBOW (Continuous Word Embedding) architecture. Each word is represented by a 30-dimensional vector, and these vectors are used to average the document-level representations. This method captures semantic relationships and contextual meanings between words more effectively than traditional frequency-based methods.

In addition, this study conducts a comparative analysis to highlight its advantages over previous research. Previous studies typically relied on platforms such as Twitter, Instagram, or online news portals, which generally feature shorter, more structured text. In contrast, comments on TikTok are more dynamic, informal, and make extensive use of slang and expressions relevant to specific contexts. Although this makes classification more difficult, it allows for a more accurate representation of public opinion in real time.

Previous research has typically Metode konvensional untuk mengekstraksi fitur termasuk Term-FrequencyInverse-Document-Frequency (TF-IDF). in combination with algorithms such as Logistic Regression or Naive Bayes. These methods are not always effective, as they can only capture semantic relationships between words. Instead, this study uses Word2Vec to create dense vector representations, which enable the model to understand contextual relationships within social media text.

In addition, this study addresses the issue of class imbalance by using the parameter `class_weight = balanced` in the SVM model. This method differs from many previous studies that did not explicitly address imbalanced data; it improves model stability and enhances performance for the minority class specifically positive sentiment. Overall, the results show that the proposed method is effective at handling unstructured and complex social media data. The combination of TikTok as a data source and the use of Word2Vec as an SVM provides a more contextual and representative sentiment analysis than conventional methods, particularly in terms of capturing semantic relationships and public opinion in real time.

4. CONCLUSION

This study demonstrates that the combination of Support Vector Machine, Word2Vec feature extraction, and lexicon-based sentiment labeling is effective for classifying public sentiment toward the Free Nutritious Food Program (MBG) using TikTok comments as a data source. The experimental results show that the implementation of balance weight improved the model performance, increasing the accuracy from 80% to 83% and producing a more balanced classification across sentiment classes, particularly for the minority positive sentiment class. In both scenarios, negative sentiment was found to be the dominant public response, reflecting concerns related to transparency, policy implementation, food safety, and budget management, while positive sentiment indicated that some users recognized the benefits of the program in improving children’s nutrition and social welfare. These findings also confirm that TikTok can serve as a valuable platform for capturing real-time public opinion because of its dynamic, informal, and expressive communication patterns. However, several limitations remain, particularly the inability of lexicon-based labeling to effectively capture sarcasm, ambiguity, slang, irony, and contextual meaning frequently found in TikTok comments. In addition, class imbalance affected the model performance when balance weight was not applied, causing lower detection capability for minority sentiment classes. Ethical considerations regarding the collection and analysis of user-generated TikTok comments, including privacy and responsible data usage, should also be acknowledged. Therefore, future studies are recommended to explore transformer-based and contextual approaches such as BERT, IndoBERT, or hybrid deep learning models to better understand nuanced language patterns and improve sentiment classification performance on complex and imbalanced social media datasets.

REFERENCES

[1] N. Dwi Prandika, D. Natalia Tinambunan, N. Indah Sari, and L. Pratama, “Strategi Sekolah dalam Pengelolaan Program Makan Bergizi Gratis sebagai Upaya Peningkatan Disiplin dan Prestasi Peserta Didik,” *Jurnal Ilmiah Multidisiplin*, vol. 1, no. 2, p. 2025, 2025, doi: https://doi.org/10.66914/nzspvv17.

[2] M. Rifai Ar Rahman, Z. Abidin Dalimunthe, A. Pane, M. Febriana, P. Ayainas, and U. Hasanah, “Analisis Kualitatif Peran Program Makanan Bergizi dalam Pencegahan Stunting: Studi Kasus di Desa Kuala Indah, Batubara,” *Umi Hasanah Journal of Human And Education*, vol. 4, no. 5, p. 1025, 2024, doi: https://doi.org/10.31004/jh.v4i5.1627.

[3] W. Trisno Aji, “Makan Bergizi Gratis di Era Prabowo-Gibran: Solusi untuk Rakyat atau Beban Baru?,” *NAAFI: JURNAL ILMIAH MAHASISWA*, vol. 2, no. 2, 2025, doi: 10.62387/naafijurnalilmiahmahasiswa.v2i2.134.

[4] A. G. Auliawan and W. Harsiwi, “Kyushoku di Jepang Sebagai Referensi Program Makan Bergizi Gratis di Indonesia,” *KIRYOKU*, vol. 9, no. 1, pp. 184–197, Apr. 2025, doi: 10.14710/kiryoku.v9i1.184-197.

[5] E. Triningsih, M. Afdal, I. Permana, and N. Evrilyan Rozanda, “Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X,” *Technology and Science (BITS)*, vol. 6, no. 4, pp. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[6] L. Najib, A. A. Mahfudh, and S. Bakhri, “Analisis Sentimen Persepsi Publik Terhadap Program MBG Pada Komentar YouTube Menggunakan Naïve Bayes dan Resampling,” *Technology and Science (BITS)*, vol. 7, no. 4, 2026, doi: 10.47065/bits.v7i4.9400.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 409−424
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9880

[7] Wardianto, P. Muhamad Jakak, and M. Rohman, "Analisis Sentimen Public Program Makan Bergizi Gratis Platform Instagram Dengan Algoritma SVM," *SMARTICS Journal*, vol. 11, no. 1, pp. 14–20, Apr. 2025, doi: 10.21067/smartics.v11i1.11852.

[8] M. Fikri, D. A. Azhar, and Y. Pradana, "Discourse About the News of Makan Bergizi Gratis (MBG) Program on Tiktok," *Citizen : Jurnal Ilmiah Multidisiplin Indonesia*, vol. 5, no. 5, pp. 1398–1409, Oct. 2025, doi: 10.53866/jimi.v5i5.940.

[9] K. Setiawan *et al.*, "Analisis Sentimen Komentar TikTok terhadap Kebijakan Larangan Wisuda Sekolah oleh Gubernur Jawa Barat Menggunakan Algoritma Naive Bayes," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 6, no. 3, 2025, doi: 10.63447/jimik.v6i3.1615.

[10] M. Zahara, M. Rizka, inul Abdi, and H. Mahyar, "Analisis Sentimen Program Makan Siang Gratis pada TikTok dengan Pendekatan NLP Berbasis IndoBERT," *Jurnal Infomedia : Teknik Informatika, Multimedia, dan Jaringan*, vol. 10, no. 2, Dec. 2025, doi: http://dx.doi.org/10.30811/jim.v10i2.8032.

[11] A. Rahmatullah and Q. Annisa, "Application Of TF-IDF And Word2vec For Feature Extraction In Sentiment Analysis Of Free Nutritious Food Policies," *Journal of Computer Electronic and Telecommunication*, vol. 6, no. 2, Jan. 2026, doi: 10.52435/complete.v6i2.741.

[12] J. Eisenstein, "Unsupervised Learning for Lexicon-Based Classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.10965.

[13] P. Arsi and R. Waluyo, "Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma Support Vector Machine (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.

[14] R. Hidayat and D. J. Ratnaningsih, "Analisis Sentimen Program Makanan Bergizi Gratis Menggunakan Algoritma Random Forest dan Naive Bayes," *Journal of Computing and Informatics Research*, vol. 5, no. 1, pp. 395–400, 2025, doi: 10.47065/comforch.v5i1.2355.

[15] Wardianto, P. Muhamad Jakak, and M. Rohman, "Analisis Sentimen Public Program Makan Bergizi Gratis Platform Instagram Dengan Algoritma SVM," *SMARTICS Journal*, vol. 11, no. 1, pp. 14–20, Apr. 2025, doi: 10.21067/smartics.v11i1.11852.

[16] S. A. Mazhar, "Methods of Data Collection: A Fundamental Tool of Research," *Journal of Integrated Community Health*, vol. 10, no. 01, pp. 6–10, Jun. 2021, doi: 10.24321/2319.9113.202101.

[17] S. Xu *et al.*, "Data cleaning in the process industries," *Reviews in Chemical Engineering*, vol. 31, no. 5, Jan. 2015, doi: 10.1515/revce-2015-0022.

[18] D. Ma, X. Shang, N. M. Ridler, and W. Wu, "Assessing the Impact of Data Filtering Techniques on Material Characterization at Millimeter-Wave Frequencies," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–4, 2021, doi: 10.1109/TIM.2021.3067224.

[19] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," *Journal of Management Analytics*, vol. 7, no. 2, pp. 139–172, Apr. 2020, doi: 10.1080/23270012.2020.1756939.

[20] L. A. Syafaah and S. Haryanto, "Slang Semantic Analysis on TikTok Social Media Generation Z," *Proceeding ISETH (International Summit on Science, Technology, and Humanity)*, pp. 476–484, Jan. 2024, doi: 10.23917/iseth.3898.

[21] A. S. Rizkia, W. Wufron, and F. F. Roji, "Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, Jul. 2025, doi: 10.57152/malcom.v5i3.2151.

[22] D. P. Nguyen, D. T. V. Duc, N. T. M. Trang, V. Q. Ket, and N. A. Hoang, "Sentiment index as a predictor of CPI: A lexicon-based approach using economic news data in Vietnam," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 3, p. 100620, Sep. 2025, doi: 10.1016/j.joitmc.2025.100620.

[23] M. Carvalho, A. J. Pinho, and S. Brás, "Resampling approaches to handle class imbalance: a review from a data perspective," *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01119-4.

[24] D. Intan Af *et al.*, "Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen," vol. 6, no. 3, 2021, doi: https://doi.org/10.30591/jpit.v6i3.

[25] A. H. Anshor and T. N. Wiyatno, "Comparative Analysis of Random Forest and Support Vector Machine for Sundanese Dialect Classification Using Speech Recognition Features," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 14, no. 2, pp. 269–276, May 2025, doi: 10.32736/sisfokom.v14i2.2347.

[26] T. Abdul Aziz and I. Ismayadi, "Analisis Sentimen Terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Logistic Regression dan SVM," *In Search*, vol. 24, pp. 18–28, 2052, doi: https://doi.org/10.37278/insearch.v24i1.1238.

[27] M. J. Setiawan and V. R. S. Nastiti, "DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 3, pp. 337–345, Nov. 2024, doi: 10.32736/sisfokom.v13i3.2239.

[28] E. R. Kaburuan and N. R. Setiawan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 105–116, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1614.

[29] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 79–91. doi: 10.18653/v1/2020.eval4nlp-1.9.