Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2391~2399
e-ISSN: 2548-6861

Comparative Analysis of Machine Learning and IndoBERT Models for Sentiment Analysis of YouTube Comments on the Free Nutritious Meals Program

Irvan Theo Shandy [1]*, Ucta Pradema Sanjaya [2]**
* Teknik Informatika, Universitas Ngudi Waluyo
** * Teknik Informatika, Universitas Ngudi Waluyo
irvantshandy@gmail.com [1], uctapradema@unw.ac.id [2]

Article Info

Article history:

Received 2026-04-13
Revised 2026-05-07
Accepted 2026-05-25

Keyword:

Free Nutritious Meals,
Sentiment Analysis,
YouTube,
Random Forest,
Support Vector Machine.

ABSTRACT

The Free Nutritious Meals Program has become one of the most widely discussed public policies in Indonesia and has generated various public responses on social media, particularly YouTube. Public comments on YouTube can be utilized as a valuable data source to understand public sentiment toward the implementation of the program. Therefore, this study aims to analyze and compare the performance of several classification algorithms in sentiment analysis of YouTube comments related to the Free Nutritious Meals Program. The dataset used in this study was obtained through a crawling process on one of Raymond Chin's YouTube videos discussing the MBG program. A total of 903 comments were collected, and after the preprocessing stage, 401 comments were selected for further analysis. The preprocessing steps included cleaning, normalization, tokenization, stopword removal, and stemming. Furthermore, the text data were transformed using the TF-IDF weighting method. This study compared several classification algorithms, namely Random Forest, Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, IndoBERT, and LightGBM. Model evaluation was conducted using confusion matrix analysis and performance metrics consisting of accuracy, precision, recall, and F1-score. The experimental results show that the Random Forest algorithm achieved the best performance with an accuracy of 0.9672, precision of 0.9683, recall of 0.9672, and F1-score of 0.9620. However, the confusion matrix analysis indicates that the model tends to be biased toward the positive sentiment class due to the imbalance in sentiment distribution within the dataset. In addition, the relatively small dataset and the use of comments from a single YouTube source may affect the generalization of the model results. Based on these findings, Random Forest can be considered the most effective algorithm for sentiment classification in this study. The results of this research are expected to provide insights into public perceptions regarding the MBG program and serve as evaluation material for policymakers in improving the implementation of public nutrition programs in Indonesia.

I. INTRODUCTION

The Free Nutritious Meals Program, initiated by President Prabowo Subianto and Vice President Gibran Rakabuming Raka, reflects the government's commitment to addressing the long-standing nutrition challenges facing Indonesia. This initiative is also aligned with the Indonesia Emas 2045 vision. To date, nutrition issues particularly among children remain a serious challenge that requires government attention. Nutritional problems, ranging from malnutrition to severe malnutrition, have serious impacts on children's growth and development, as well as their cognitive development. The Indonesian government has implemented the Free Nutritious

Meals program, which aims to address nutritional issues and improve the quality of human resources by combating stunting through the provision of nutritious meals for schoolchildren, toddlers, pregnant women, and breastfeeding mothers [1].

Similar programs have been implemented in various countries, but their success largely depends on public acceptance and sentiment toward these initiatives. Currently, free nutritious meals are a hot topic and have sparked a wide range of reactions on social media, particularly YouTube. With a massive user base of approximately 2.53 billion, this platform has become one of the primary channels for the public to express their opinions, criticisms, and sentiments regarding this issue. The comment sections on news videos or discussions about free nutritious meals contain thousands of opinions with a range of emotions, both positive and negative; this platform has great potential as a data source for understanding public opinion [2].

However, the high volume of public opinion data on social media has not been matched by comprehensive and systematic analysis, making it difficult for policymakers to understand public perceptions objectively and based on data. Therefore, a computational approach is needed that can process large amounts of data efficiently and accurately.

Sentiment analysis is used to identify the views expressed in textual data using a specific method. Sentiment analysis is a branch of Natural Language Processing (NLP) that aims to categorize public opinion into categories such as positive, negative, or neutral. This approach can provide a more objective and efficient understanding of public sentiment through text. Natural Language Processing (NLP) is also known as a branch of artificial intelligence (AI) that plays a role in understanding, processing, and analyzing human language. Previous studies have shown that sentiment analysis can provide a highly accurate picture of shifts in public opinion regarding a government program [3].

Sentiment analysis involves several stages, such as data preprocessing which includes cleaning, normalization, tokenization, stopword removal, and stemming as well as Term Frequency-Inverse Document Frequency (TF-IDF). These stages aim to convert textual data into numerical values so that it can be recognized when processed by machine learning algorithms within the system. This process is crucial because it improves data quality and enhances the model's performance in sentiment analysis. Various previous studies have shown that the preprocessing and feature weighting stages play a very important role in improving the accuracy of sentiment analysis algorithms [4].

Technology is currently advancing at a rapid pace, and sentiment analysis is being utilized in various machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbor (KNN). Each algorithm has different strengths and levels of accuracy in analyzing and interpreting text data. Therefore, research comparing several algorithms is essential to determine the best algorithm for analyzing public sentiment. The sentiment analysis approach using machine learning can identify public opinion and express emotions that are not always apparent in conventional surveys. This can help manage the massive volume of YouTube data that grows daily [5].

However, most previous studies have focused on using only one or two algorithms without conducting a comprehensive comparison between classical machine learning methods and transformer-based models. Furthermore, research on sentiment analysis regarding the Free Nutritious Meal Program remains limited, particularly studies that utilize data from YouTube as the primary source. This indicates a research gap that warrants further investigation [6] [7]. On the other hand, the use of data from social media platforms like YouTube also presents its own challenges, such as potential data bias, imbalanced class distribution, and limitations in comprehensively representing public opinion. Therefore, an analysis is needed that focuses not only on the model's accuracy but also on the interpretation of results and the validity of the data used [8].

This study compares several algorithms namely Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, and IndoBERT in classifying public sentiment toward the Free Nutritious Meal Program based on YouTube comment data. This is important given the massive volume of social media data. Therefore, this study aims to compare algorithms capable of accurately classifying public sentiment regarding the Free Nutritious Meal Program based on YouTube comment data [6] [7].

The primary contribution of this study is to conduct a comparative analysis between classical machine learning algorithms and transformer-based models in the context of public policy sentiment analysis, specifically regarding the Free Nutritious Meals Program. Additionally, this study provides initial insights into public perceptions of the program based on YouTube comment data, which is expected to serve as evaluation material for policymakers and a reference for future research.

II. METHOD

This study employs sentiment analysis to classify public opinion regarding the Free Nutritious Meals Program using data obtained from YouTube user comments. The research methodology is outlined below. This study employs a quantitative approach using experimental methods to compare several classification algorithms for sentiment analysis. In addition, this study also integrates a Natural Language Processing (NLP) approach to systematically process text data.

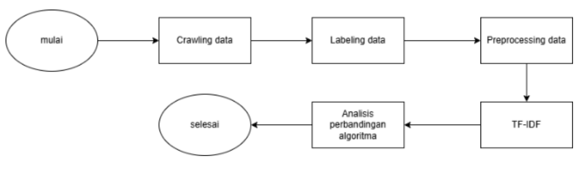

Figure 1. Research Flowchart

Figure 1 illustrates the stages of the research methodology used in the sentiment analysis of the Free Nutritious Meals Program (MBG). The research process began with the collection of data in the form of YouTube comments using crawling techniques. Next, the data was manually labeled to classify the comments into positive or negative categories. After that, the data underwent preprocessing, followed by a comparison of several classification algorithms to determine which algorithm had the highest accuracy rate.

A. Data Crawling

The data used in this study consists of comments from one of Raymond Chin's videos discussing a free nutritious meal program, which was uploaded to YouTube; the data was obtained through a data crawling process. This data collection process was carried out using Google Colab to facilitate the structured retrieval and processing of large amounts of data [8].

The crawling process was conducted using the YouTube Data API via a Python library (such as google-api-python-client) running on Google Colab. This process involved retrieving comments based on specific video IDs and filtering them based on relevance and text completeness.

Data collection is conducted within a specific timeframe (November 2025) and focuses solely on a single video as the primary data source. This may introduce data source bias as it does not represent the full spectrum of public opinion on the YouTube platform. Therefore, this limitation must be taken into account when interpreting the research results.

Data collection took place in November 2025, yielding 913 comments, which were then saved in Excel format (.xlsx) to facilitate processing in the next stage. Following this process, 401 comments were deemed suitable for use in the subsequent analysis phase. The reduction of data from 913 to 401 comments was achieved through data cleaning steps, such as removing duplicates, empty comments, and comments irrelevant to the research topic.

B. Data Labeling

At this stage, the data obtained from the crawling process is then labeled or categorized based on the sentiment contained in the comment text. Labeling is performed by dividing the comments into two main categories: positive sentiment and negative sentiment. This labeling stage is crucial in sentiment analysis because it serves as the primary foundation for training the machine learning models used [9].

The labeling process was performed manually by the researchers based on the context of the comments to ensure label accuracy [5]. To improve the reliability of the annotations, the labeling process should ideally involve more than one annotator and include a measure of inter-annotator agreement (e.g., using Cohen's Kappa); however, in this study, labeling was performed by a single annotator, which constitutes one of the study's limitations [6].

The labeling process was performed manually by researchers, taking into account the overall context of the sentences. Each comment is classified into a positive or negative category based on the opinion expressed. To improve labeling consistency, simple annotation guidelines defining the criteria for each sentiment class were used. However, this study did not involve more than one annotator, so annotation reliability measures such as Cohen's Kappa have not been calculated, which is one of the study's limitations [2].

Through this labeling process, the system can learn specific patterns from the text data, enabling it to automatically classify new data. Additionally, this stage serves to identify public opinion trends regarding the Free Nutritious Meals Program.

C. Data Preprocesing

At this stage, the data that has undergone the labeling process is further processed through the preprocessing stage. This stage aims to clean out invalid data and prepare the data so that it is more structured and ready for use in analysis using machine learning algorithms. Preprocessing plays a crucial role in text analysis because raw data generally still contains noise or irrelevant information [10].

It should be noted that these preprocessing steps are applied differently depending on the type of algorithm used. For classical machine learning algorithms (such as Random Forest, SVM, and Naïve Bayes), all preprocessing steps are performed. However, for transformer-based models such as IndoBERT, steps like stemming and stopword removal are not applied because these models already incorporate tokenization and contextual language representation mechanisms [3].

It should be noted that the preprocessing steps used in this study were generally applied to all algorithms, including transformer-based models such as IndoBERT. However, conceptually, transformer models like IndoBERT generally do not require stemming or stopword removal because they have built-in tokenization mechanisms capable of understanding more complex linguistic contexts. Therefore, applying the same preprocessing to all models may affect the performance of certain models and is a factor that needs to be considered in the analysis of results [11].

The preprocessing stage in this study consists of several main steps, namely cleaning, normalization, tokenization, stopword removal, and stemming, as shown in Figure 2 below.

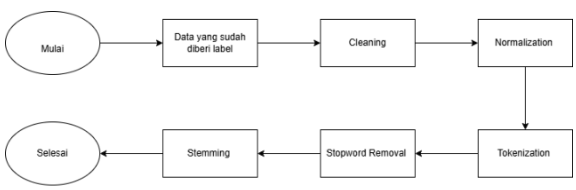

Figure 2. Data Preprocessing

1) Cleaning

The first step in preprocessing is data cleaning, which involves removing irrelevant elements such as symbols, excess punctuation, links (URLs), numbers, user mentions, hashtags, and emojis that have no significant meaning for the analysis. Additionally, duplicate comments are removed to maintain the quality of the dataset [9]. After the cleaning process is complete, the initial dataset of 913 comments is reduced to 401 clean comments.

2) Normalization

The normalization stage in the data preprocessing process involves converting non-standard words, abbreviations, and regionally specific conversational language into a more standardized format to prevent errors during machine learning analysis. This stage is crucial because YouTube comment sections frequently use informal language and abbreviated terms [8] [9].

3) Tokenization

Tokenization is the process of breaking text down into smaller units called tokens or words. Tokenization aims to facilitate the feature extraction process, where each token is treated as a potential feature in weight calculations using the TF-IDF method [9] [10].

4) Stopword Removal

The next step is stopword removal, which is performed to eliminate common words that are not relevant to the analysis process; if not removed, they will interfere with the analysis and affect the classification results. Examples include the words “apa,” “di,” or “untuk” [10].

5) Stemming

The final step is stemming, the process of removing affixes and converting words to their base forms. The stemming process is performed using the Sastrawi library, a popular tool in Indonesian research for stemming and stopword removal in local languages. Using this tool has been shown to improve the accuracy of the classification system [10].

D. TF-IDF

Data that has undergone preprocessing must then be converted from text format into a numerical representation so that it can be more easily analyzed using various algorithms such as Random Forest, Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, IndoBERT, and LightGBM. One method used for this transformation is Term Frequency-Inverse Document Frequency (TF-IDF), which assigns weights to each word ina document. TF-IDF measures a word’s importance by considering how frequently it appears within a document (term frequency) and its relative frequency across all documents (inverse document frequency). With this approach, words that play a significant role can be distinguished from less relevant words [9].

In this study, the TF-IDF method was used specifically for classical machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Gradient Boosting, XGBoost, Multinomial Naïve Bayes, and LightGBM. Meanwhile, for transformer-based models such as IndoBERT, text representation is performed using the model’s built-in embeddings without going through the TF-IDF process, as these models are already capable of capturing semantic context directly from the text [5].

Each unique word is then assigned a weight based on the TF-IDF calculation, which is mathematically formulated to describe the level of a word’s importance in sentiment analysis, as follows.

$$TF - IDF(t,d) = TF(t,d) \times IDF(t)$$

Where:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

Definitions:

- TF (t,d) : term frequency in document d
- Df (t) : number of documents containing term t
- N : total number of documents

E. Algorithm Comparison Analysis

This study conducted an analysis focused on comparing several algorithms to determine which algorithm has the highest accuracy rate in analyzing sentiment in YouTube comments. The evaluation was carried out to assess the performance of each algorithm so that the most effective algorithm for producing accurate results could be identified. The focus of this study is on several algorithms, including Random Forest, Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, IndoBERT, and LightGBM, each of which possesses unique characteristics and distinct advantages in text data processing. Through a thorough and structured evaluation, this research aims to provide a comprehensive understanding of the extent to which various methods can measure and analyze sentiment from user comments on the YouTube platform.

III. RESULT AND DISCUSSION

This section presents the results of the analysis process and a discussion of the findings. The results presented are the outcome of the data processing stages, ranging from preprocessing to the application of the classification methods used in this study. Furthermore, these results were analyzed using machine learning to determine the performance of each algorithm in classifying sentiment data. This study aims to classify public comments regarding the Free Nutritious Meals Program into two sentiment categories, namely positive and negative, using several algorithms that were compared.

In addition to presenting the classification results, this study also conducts a more in-depth analysis of the model’s performance, data distribution, and potential biases that arise during the classification process. This analysis is important to ensure that the results obtained are not only numerically high but also scientifically valid and can be interpreted accurately.

A. Data Crawling

This study uses data from text-based YouTube comments. Data was collected via web crawling using Google Colab. A total of 913 comment records were saved in an Excel file (.xlsx).

The collected comments represent public reactions and responses, ranging from criticism and suggestions to support for the free nutritious meals program. In this study, YouTube serves as a platform for public discussion regarding the Free Nutritious Meals program. Examples of the crawled data are shown in Table 1.

TABLE I
DATA CRAWLING RESULTS

Text
Nyatanya pasar2 tradisional makin sepi sejak adanya mbg ini, dan yang bilang dan ngerasain ya masyarakat menengah ke bawah 😥
Program yg bikin bahan pokok dipasar jadi melonjak tinggi dan bikin emak2 pusing ngatur keuangan karena bahan pokok sudah semahal itu 😥

B. Data Labeling

After the comment data crawling process, the next step is manual data labeling. The purpose of this manual labeling is to assign a sentiment class or category to each collected comment. In this study, comments were classified into two sentiment categories—positive and negative—taking into account the context of the discussion regarding the Free Nutritious Food (MBG) program. A total of 913 comments were selected, resulting in 401 comments.

The distribution of sentiment classes in the dataset is also an important aspect to consider, as it can affect the performance of the classification model. Based on the data labeling results, the number of comments with positive sentiment dominates over those with negative sentiment. This data imbalance can cause the model to tend to predict the majority class, resulting in high accuracy scores, but it may not necessarily be able to classify the minority class optimally [1]. Therefore, the interpretation of the model evaluation results must take into account the data distribution conditions used in this study. Table 2 is an example of data labeling.

TABLE II
DATA LABELING RESULT

Label	Text
Negatif	Nyatanya pasar2 tradisional makin sepi sejak adanya mbg ini, dan yang bilang dan ngerasain ya masyarakat menengah ke bawah 😥
Positif	Program yg bikin bahan pokok dipasar jadi melonjak tinggi dan bikin emak2 pusing ngatur keuangan karena bahan pokok sudah semahal itu 😥

The following visualization compares the number of positive and negative sentiments shown in Figure 1, indicating that the proportion of positive sentiments is significantly higher.

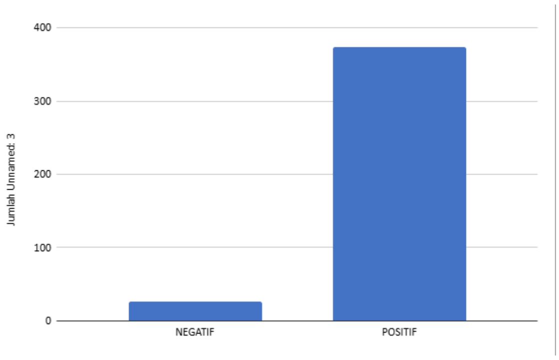

Figure 3. Comparison of positive and negative sentiment

C. Data Preprocessing

Preprocessing is a crucial step in text data analysis because the raw data is obtained from social media platforms. This process transforms the raw data into cleaner data that can be used for further analysis. The preprocessing steps can be carried out as follows.

1) Cleaning

The first step in preprocessing is data cleaning. During this step, irrelevant elements are removed, such as symbols, excess punctuation, links (URLs), numbers, user mentions, hashtags, and emojis that do not contribute significant meaning to the analysis. Additionally, duplicate comments are removed to ensure the quality of the dataset is maintained.

TABLE III
CLEANING

Text
nyatanya pasar tradisional makin sepi sejak adanya mbg ini dan yang bilang dan ngerasain ya masyarakat menengah ke bawah
program yg bikin bahan pokok dipasar jadi melonjak tinggi dan bikin emak pusing ngatur keuangan karena bahan pokok sudah semahal itu

Based on the table above, it is clear that the text has undergone a cleaning process to remove unnecessary elements, resulting in cleaner sentences that are ready for processing in the next stage. This stage is important so that the analysis can focus more on the main information in the text.

2) Normalization

The normalization stage aims to convert non-standard words, abbreviations, and local colloquialisms into more standard forms. This is important to avoid errors in the machine learning analysis process, given that YouTube comments generally use informal language. Examples of normalization results are shown in Table IV below.

TABLE IV
NORMALIZATION

Text
nyatanya pasar tradisional makin sepi sejak adanya mbg ini dan yang bilang dan merasakan ya masyarakat menengah ke bawah
program yang membuat bahan pokok di pasar jadi melonjak tinggi dan membuat emak pusing mengatur keuangan karena bahan pokok sudah semahal itu

Based on Table IV, it can be seen that the comment text becomes more organized with more consistent word forms. This process helps reduce unnecessary word variation, so that words with the same meaning can be recognized as a single unit by the system. With more structured data, subsequent stages such as tokenization and TF-IDF weighting can run more optimally, thereby improving sentiment classification accuracy.

3) Tokenization

Tokenization is the process of breaking text down into smaller units called tokens or words. Tokenization aims to facilitate the feature extraction process, where each token is treated as a potential feature in weight calculations using the TF-IDF method.

TABLE V
TOKENIZING

Text
'nyatanya', 'pasar', 'tradisional', 'makin', 'sepi', 'sejak', 'adanya', 'mbg', 'ini', 'dan', 'yang', 'bilang', 'dan', 'merasakan', 'ya', 'masyarakat', 'menengah', 'ke', 'bawah'
'program', 'yang', 'membuat', 'bahan', 'pokok', 'di', 'pasar', 'jadi', 'melonjak', 'tinggi', 'dan', 'membuat', 'emak', 'pusing', 'mengatur', 'keuangan', 'karena', 'bahan', 'pokok', 'sudah', 'semahal', 'itu'

As shown in Table V, each sentence has been broken down into individual words. This process allows the system to identify each word as a feature to be used in subsequent analysis stages. Through tokenization, complex sentence structures are converted into a simpler form, thereby facilitating the process of word weighting and sentiment classification.

4) Stopword Removal

The next step is stopword removal, which is performed to eliminate common words that are not relevant to the analysis process; if not removed, they will interfere with the analysis and affect the classification results. Examples include the words “apa,” “di,” or “untuk”.

TABLE VI
STOPWORD REMOVAL

Text
'nyatanya', 'pasar', 'tradisional', 'makin', 'sepi', 'sejak', 'adanya', 'mbg', 'bilang', 'merasakan', 'masyarakat', 'menengah', 'bawah'
'program', 'membuat', 'bahan', 'pokok', 'pasar', 'melonjak', 'tinggi', 'membuat', 'emak', 'pusing', 'mengatur', 'keuangan', 'bahan', 'pokok', 'semahal'

As shown in Table VI, common words that lack significant meaning have been removed from the text. This process results in a more relevant and informative set of words. By removing stopwords, the data becomes more concise and focused, thereby improving the quality of the analysis and helping the algorithm identify sentiment patterns more accurately.

5) Stemming

One of the key text preprocessing techniques in natural language processing. This technique is used to simplify text, thereby improving the efficiency and accuracy of machine learning algorithms.

TABLE VII
STEMMING

Text
Nyatanya pasar pasar tradisional makin sepi sejak adanya mbg ini, dan yang bilang dan ngerasain ya masyarakat menengah ke bawah
Program yg bikin bahan pokok dipasar jadi melonjak tinggi dan bikin emak emak pusing ngatur keuangan karena bahan pokok sudah semahal itu

As shown in Table VII, the words in the sentences have been converted to their base forms. This process aims to reduce morphological variation so that words with similar meanings can be recognized as the same entity. Thus, stemming can improve data processing efficiency and help enhance the model’s accuracy in the sentiment classification process.

D. TF-IDF

TF-IDF is used to measure how important a word is within a document relative to the entire corpus. The higher a word’s TF-IDF value, the greater its influence or weight in representing the document’s content. TF-IDF is a commonly used method in text processing because it assigns weights to each word based on its level of importance within a document relative to the entire corpus.

E. Confusion Matrix

The performance of the classification model in this study was evaluated using a confusion matrix, which serves to illustrate the model’s ability to accurately classify sentiment data. The confusion matrix reveals the number of correct and incorrect predictions based on a comparison between the actual labels and the model’s predicted labels [11] [12].

Figure 4 shows the confusion matrix generated by the model. Based on the figure, the true negative (TN) value is 0, meaning that the model did not correctly predict any negative sentiment data as negative. Furthermore, the false positive (FP) value of 5 indicates that there were 5 data points with an actual negative label that were predicted as positive by the model. This indicates an error in the model's classification of negative sentiment. On the other hand, the false negative (FN) value of 0 indicates that no positive sentiment data was incorrectly predicted as negative. Meanwhile, the true positive (TP) value of 76 indicates that 76 data points with an actual positive label were correctly predicted by the model.

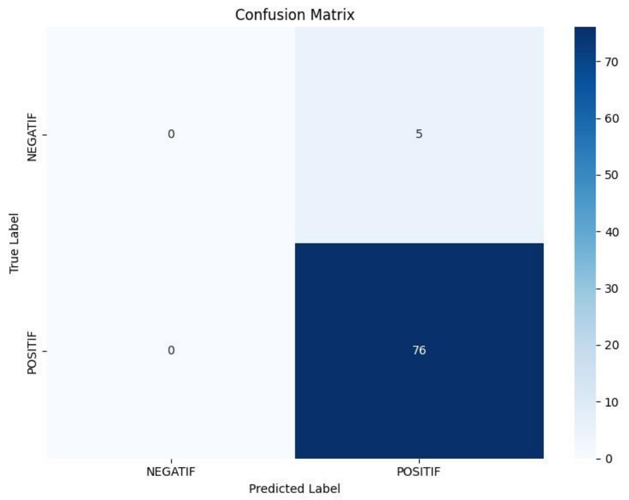

Figure 4. Confusion Matrix

Based on the confusion matrix results above, it can be concluded that the model performs very well in recognizing positive sentiment, as indicated by the high true positive rate and the absence of false negatives. However, the model is still unable to identify negative sentiment effectively, as none of the negative data was classified correctly. This indicates that the model is biased toward the positive class and requires improvements, such as data balancing or enhanced training methods, to achieve more accurate classification performance.

Although the model achieves a fairly high accuracy rate, the confusion matrix results indicate that the model's ability to distinguish between the two sentiment classes remains unbalanced. A True Negative (TN) value of 0 indicates that the model is not yet capable of properly identifying negative sentiment data. This suggests a tendency for the model to focus more on the positive class, which is likely influenced by an unbalanced data distribution and a relatively small dataset size [2]. These results show that using the accuracy metric alone is insufficient to fully describe the model's performance. Therefore, additional evaluation metrics such as precision, recall, and F1-score are needed to analyze the model's ability to classify data more comprehensively and objectively [3].

F. Word Cloud

The word cloud visualization in this study is used to show the distribution and frequency of words in the analyzed text data. Through this display, frequently occurring words appear larger, making it easier to identify the main topics, opinion trends, and sentiment patterns present in the data.

Figure 5 displays three types of word clouds: one for the overall data, one for negative sentiment, and one for positive sentiment. In the word cloud for the overall data, words such as "food," "program," "children," "nutrition," and "school" appear most prominently. This indicates that the main discussion focuses on nutrition programs, particularly for school-aged children. Furthermore, in the negative sentiment word cloud, words such as "food," "nutrition," "free," "quiet market," "business," and "quiet" stand out quite prominently. This indicates that negative sentiment is more closely related to the program's impact on economic conditions or specific businesses, as well as the presence of less positive views regarding its implementation. Meanwhile, in the positive sentiment word cloud, words such as "food," "program," "nutrition," "school," "children," and "love" appear more dominant. This suggests that positive sentiment tends to highlight the program's benefits, particularly in improving children's nutrition and supporting activities within the school environment.

Figure 5. Word Cloud

Based on this visualization, it can be concluded that the main topic in the data centers on programs providing nutritious meals to schoolchildren. Additionally, there is a difference in emphasis between positive and negative sentiment, with positive sentiment highlighting the benefits of the program, while negative sentiment tends to focus on criticism or the impacts felt by the community.

In addition to showing the frequency of word occurrences, word cloud visualizations also provide an initial overview of public opinion trends regarding the program under study. Negative sentiment is generally associated with economic issues, such as rising prices of basic commodities and their impact on traditional market activities. Conversely, positive sentiment more often highlights the program's benefits in improving children's nutritional quality and supporting educational activities in schools. These results indicate that public perception of the Free Nutritious Meals Program is influenced by social and economic factors directly experienced by the public.

G. Comparative Analysis of Algorithms

This study compares the performance of several classification algorithms to identify the most optimal algorithm for sentiment analysis of YouTube comments related to the Free Nutritious Food Program (MBG). The algorithms compared include Random Forest, Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, IndoBERT, and LightGBM. All algorithms were tested using the same dataset to ensure the comparison results were objective and fair. The performance of each algorithm was evaluated using several evaluation metrics, namely accuracy, precision, recall, and F1-score. These metrics were selected because they provide a comprehensive overview of the model's performance in classifying sentiment. The results of the performance comparison of each classification algorithm used in this study are presented in Figure 4.

Model	Accuracy	F1-Score	Precision	Recall	Training Time (s)	Prediction Time (s)	Model Size (MB)
Random Forest	0.9672	0.9620	0.9683	0.9672	0.3316	0.0351	0.000
Gradient Boosting	0.9672	0.9620	0.9683	0.9672	0.6458	0.0012	0.000
SVM	0.9672	0.9620	0.9683	0.9672	0.0435	0.0019	0.000
XGBoost	0.9508	0.9475	0.9459	0.9508	0.4222	0.0108	0.000
Multinomial Naïve Bayes	0.9508	0.9367	0.9533	0.9508	0.0046	0.0005	0.000
IndoBERT	0.9344	0.9028	0.8732	0.9344	2045.5944	23.1804	421.752
LightGBM	0.9180	0.9222	0.9273	0.9180	0.0547	0.0043	0.000

Figure 6. Algorithm Comparison Results

As shown in Figure 4, the Random Forest algorithm demonstrates the best performance compared to other algorithms, with an accuracy of 0.9672, an F1-score of 0.9620, a precision of 0.9683, and a recall of 0.9672. These values indicate that Random Forest is highly capable of classifying sentiment data accurately and consistently.

Random Forest delivers the best performance because this algorithm works by building many decision trees, which are then combined to produce a more stable final prediction. On small datasets like the one used in this study, Random Forest tends to be more effective at reducing the risk of overfitting compared to some other algorithms. Additionally, Random Forest handles TF-IDF-processed text features well because it can automatically select important features during the tree-building process.

Additionally, the Gradient Boosting and Support Vector Machine (SVM) algorithms also demonstrate highly competitive performance, with accuracy, F1-score, precision, and recall values comparable to those of Random Forest. However, the primary difference among these three algorithms lies in computational efficiency, particularly regarding training time and prediction time.

Although algorithms such as XGBoost and SVM also demonstrate high performance, these two algorithms have different characteristics when processing text data. SVM is known for its strong classification capabilities on high-dimensional data, but its performance is heavily influenced by kernel parameters and data distribution. Meanwhile, XGBoost has powerful boosting capabilities, but on small datasets, its performance is not always superior to that of Random Forest.

Thus, although some algorithms have equally high accuracy levels, the selection of the best algorithm is not based solely on evaluation metrics but also considers computational time efficiency. In the context of this study, Random Forest can be considered the most optimal algorithm because it provides a balance between high classification performance and good process efficiency.

In transformer models such as IndoBERT, the performance achieved has not yet surpassed that of Random Forest. This may be due to the relatively small size of the dataset, as transformer models generally require larger amounts of data to learn linguistic context optimally. Additionally, the complexity of transformer models requires more computational resources than classical machine learning algorithms.

IV. CONCLUSION

This study aims to analyze and compare the performance of several classification algorithms in sentiment analysis of YouTube comments regarding the Free Nutritious Meals Program (MBG). Based on the research results, the sentiment analysis process—which begins with data crawling, manual labeling, and preprocessing steps including cleaning, normalization, tokenization, stopword removal, and stemming—has proven capable of producing data ready for optimal classification.

The application of the TF-IDF weighting method successfully represented the text data in numerical form so that it could be processed by machine learning algorithms. The results of the comparison of the Random Forest, Gradient Boosting, Support Vector Machine (SVM), XGBoost, Multinomial Naïve Bayes, IndoBERT, and LightGBM indicate that the Random Forest algorithm delivers the best performance with an accuracy of 0.9672, an F1-Score of 0.9620, a precision of 0.9683, and a recall of 0.9672.

The results of the evaluation using a confusion matrix indicate that the classification model is still unable to identify negative sentiment optimally. This is influenced by an imbalanced data distribution, in which the number of data points with positive sentiment is more dominant than those with negative sentiment in the research dataset. This imbalance can lead to bias in the classification process and increase the likelihood of overfitting due to the relatively small dataset size. Therefore, future research is advised to use a larger dataset with a more balanced class distribution and to apply validation techniques such as K-Fold Cross Validation so that model performance evaluation can be conducted more stably and accurately.

This study also shows that social media sentiment analysis can be used to describe public perceptions of government policies, particularly the Free Nutritious Meals Program. Positive sentiment generally reflects public support for the program's benefits in helping meet the nutritional needs of schoolchildren. On the other hand, negative sentiment is more often related to the economic impact and the program's implementation within the community. It is hoped that the findings of this study can serve as input for the government in conducting evaluations and formulating policies that are more effective and aligned with the needs of the community.

Nevertheless, this study still has limitations because the data used was derived solely from a single YouTube video by Raymond Chin. As a result, the findings do not fully represent the overall opinions of the Indonesian public or social media users in general. Therefore, future research is advised to utilize a more diverse range of data sources, such as X/Twitter, Instagram, TikTok, and online news portals, so that the sentiment analysis results obtained can be broader, more representative, and more comprehensive.

Thus, Random Forest can be concluded as the most optimal algorithm for classifying the sentiment of YouTube comments related to the Free Nutritious Meal Program (MBG). The results of this study are expected to serve as evaluation material for the government in understanding public response to the MBG program and as a reference for future research involving the expansion of data volume, sentiment classes, and a wider variety of methods.

REFERENCES

[1] M. F. Saleh and R. Imanda, "Public Sentiment Analysis of the Free Meal Program : A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform," vol. 9, no. 1, pp. 131–139, 2025.

[2] F. Nutritious, M. Program, and L. R. Methods, "Research Horizon," vol. 0696, 2025.

[3] N. Putu, A. Noviyanti, I. G. Ngurah, and A. Cahyadi, "Penerapan Multinomial Naïve Bayes dan Chi-Square pada Analisis Sentimen Makan Bergizi Gratis," vol. 3, pp. 785–794, 2025.

[4] M. S. Dewi and A. H. Hasugian, "Mutiara Sintia Dewi, 2) Abdul Halim Hasugian," vol. 10, no. 2, pp. 911–922, 2025.

[5] E. Setiadi and D. A. Sulistyo, "Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning," vol. 15, no. 3, pp. 507–513, 2025.

[6] R. E. Saputro, "Literature Review : Komparasi Algoritma Machine Learning dalam Analisis Sentimen Makan Bergizi Gratis Literature Review : Comparison of Machine Learning Algorithms for Sentiment Analysis of Free Nutritious Meals," vol. 15, pp. 1110–1120, 2026.

[7] F. A. Hakim, I. Wisma, D. Prastya, and J. R. Budiani, "Sentiment Analysis of the Free Nutritious Meal Program (MBG) on Social Media X (Twitter) Using K-Nearest Neighbor and Artificial Neural Network," vol. 10, no. 1, 2026.

[8] W. Wahyuni, T. P. Lestari, M. Apriliana, and R. Gumelta, "Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments," vol. 9, no. 4, pp. 1964–1971, 2025.

[9] I. Hermansyah and M. S. Hasibuan, "Sentiment Analysis of Free Nutritious Meal Programme on Social Media X using Linear Regression and Random Forest Algorithms," vol. 13, no. 225, pp. 49–54, 2026, doi: 10.33558/piksel.v13i1.10633.

[10] T. A. Aziz and I. Ismayadi, "Analisis Sentimen Terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Logistic Regression dan SVM," vol. 24, no. 1, pp. 18–28, 2025.

[11] M. Ilham and B. Priambodo, "Analisis Sentimen Publik Terhadap Program Makan Siang Gratis Menggunakan BERT Neural Network Pada Platform X," vol. 6, no. 2, pp. 1039–1047, 2024.

[12] M. Haikal and A. Erfina, "JURNAL LOCUS : Penelitian & Pengabdian Analisis Sentimen Warganet terhadap Program Makan Bergizi Gratis (MBG) Menggunakan Model Indobert," vol. 5, no. 3, 2026.

[13] D. A. Vonega, A. Fadila, and D. E. Kurniawan, "Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024", JAIC, vol. 6, no. 2, pp. 129–135, Nov. 2022.

[14] J. Friadi, and D. E. Kurniawan, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam: Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine," Jurnal Sistem Informasi Bisnis, vol. 14, no. 4, pp. 403-407, Oct. 2024. https://doi.org/10.21456/vol14iss4pp403-407.

[15] N. N. Aprianti, N. M. M. R. Desmayani, L. G. B. Libraeni, I. G. A. Indrawan, and M. L. Radhitya, "Public Sentiment Analysis of the Free Nutritious Meals Program (MBG) on Social Media X Using the Naive Bayes Method", JAIC, vol. 9, no. 6, pp. 3929–3936, Dec. 2025.

[16] F. A. Hakim, I. W. D. Prastya, and J. R. Budiani, "Sentiment Analysis of the Free Nutritious Meal Program (MBG) on Social Media X (Twitter) Using K-Nearest Neighbor and Artificial Neural Network", JAIC, vol. 10, no. 1, pp. 865–876, Feb. 2026.

[17] A. P. Arunia, R. R. Sani, I. N. Dewi, and M. T. Sulistyono, "Exploring Public Opinion on the 'Makan Bergizi Gratis' Program on X: A Comparative Analysis of IndoBERT-Large and NusaBERT-Large Models", JAIC, vol. 10, no. 1, pp. 856–864, Feb. 2026.

[18] I. Nugroho and S. Mujiyono, "Analysis of Public Sentiment Towards the Free Nutritious Meals Program on TikTok Social Media Using the K-Nearest Neighbor Algorithm", JAIC, vol. 10, no. 1, pp. 975–983, Feb. 2026.

[19] N. P. Zakiyah, K. Umam, and A. A. Mahfudh, "Public Opinion on The MBG Program: Comparative Evaluation of InSet and VADER Lexicon Labeling Using SVM on Platform X", JAIC, vol. 9, no. 6, pp. 3937–3944, Dec. 2025.