Indonesian Journal of Data and Science

Volume 7 Issue 2
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i2.300

Research Article

Sentiment Classification of TikTok Comments on The Free Nutritious Meal Program Using Multinomial Naïve Bayes

Salom Sefanya Onibala [1]; Rosmasari[2,*]; Kezia Arum Sary [3]
[1] Universitas Mulawarman, Kalimantan Timur, Samarinda 75123, Indonesia, lomeyonibala@gmail.com
[2] Universitas Mulawarman, Kalimantan Timur, Samarinda 75123, Indonesia, rosmasari@unmul.ac.id
[3] Universitas Mulawarman, Kalimantan Timur, Samarinda 75123, Indonesia,kezia.arumsary@fisip.unmul.ac.id

Correspondence should be addressed to Rosmasari; rosmasari@unmul.ac.id

Received 10 March 2026; Accepted 04 July 2026; Published 31 July 2026

Abstract:
Introduction: TikTok has become an important platform for public discussion of government programs, including Indonesia's Free Nutritious Meal (MBG) Program, while the large volume of comments makes manual sentiment analysis inefficient. This study evaluates Multinomial Naïve Bayes for classifying public sentiment toward the MBG Program using TikTok comments. **Method**: A total of 1,048 comments were collected from the official National Nutrition Agency TikTok account using Apify, with 1,028 comments remaining after preprocessing. Sentiment pseudo-labels were generated using IndoRoBERTa, producing 333 positive and 695 negative comments. Undersampling balanced the dataset to 666 comments, followed by an 80:20 stratified train–test split. TF-IDF feature extraction and Multinomial Naïve Bayes classification were integrated in a Scikit-learn Pipeline, while GridSearchCV with 5-fold cross-validation optimized model parameters. **Results and Discussion**: The optimized model achieved 80.60% accuracy, 81.04% precision, 80.60% recall, and an 80.53% F1-score on 134 test comments. Negative sentiment dominated the original post-preprocessing dataset at 67.61%. However, these results primarily indicate agreement with IndoRoBERTa-generated pseudo-labels rather than human-verified sentiment labels. **Conclusion**: The TF-IDF–Multinomial Naïve Bayes pipeline provides a useful baseline for MBG-related TikTok sentiment classification, but manual label validation and comparison with alternative classifiers are required to establish stronger reliability and generalizability.

Keywords: Sentiment Classification, Multinomial Naive Bayes, IndoRoBERTa, Free Nutritious Meal Program, TikTok, TF-IDF

Dataset link:
https://drive.google.com/drive/folders/1yFHIoL5phPIMHXWD2b3jWkoAJKCfvx1c?usp=drive_link

1. Introduction

The rapid growth of social media has transformed the way people express their opinions on various issues, including government policies [1]. One of the fastest-growing social media platforms is TikTok, a short-video application developed by ByteDance that enables users to create, edit, and share interactive content [2]. TikTok has become one of the most engaging social media platforms, making it a popular medium for expressing public opinions on social issues and government policies. According to the We Are Social report, Indonesia had approximately 126.83 million TikTok users as of January 2024, ranking second among countries with the largest number of TikTok users worldwide [3].

Through the TikTok comment section, users can express support, criticism, and suggestions regarding government policies. The large volume of comments generated on social media provides a valuable source of data for understanding public opinion [4], [5]. However, manually analyzing such a large amount of textual data is inefficient and time-consuming. Therefore, sentiment analysis is required to automatically classify public opinions into

predefined sentiment categories [6]. Sentiment analysis is an application of text classification techniques that aims to categorize textual documents according to the sentiment expressed through the words contained in the documents [7].

Classification is one of the machine learning techniques used to predict the class of data based on patterns learned from labeled datasets. Generally, the classification process consists of two stages: the training phase, where a classification model is developed, and the testing phase, where the model is evaluated using unseen data [8], [9]. In sentiment analysis, the classification process assigns each user comment to a positive or negative sentiment category.

One of the government programs that has attracted considerable public attention is the Free Nutritious Meal (MBG) Program. Officially launched in 2025, the program aims to improve the nutritional status of the population, particularly school children and pregnant women, while supporting national efforts to reduce stunting [10], [11]. Despite its positive objectives, the implementation of the program has encountered several challenges, including food distribution, food quality, infrastructure readiness, and food poisoning incidents, which have generated diverse public responses on social media. Consequently, the number of comments discussing the MBG Program has continued to increase, highlighting the need for an efficient and accurate sentiment classification method [12], [13].

One of the most widely used algorithms for text classification is the Multinomial Naïve Bayes algorithm. This algorithm is an extension of the standard Naïve Bayes classifier that employs a probabilistic approach while considering the frequency of each word appearing in a document [14]. Unlike the standard Naïve Bayes algorithm, Multinomial Naïve Bayes utilizes word-frequency distributions, making it particularly suitable for textual data represented using Term Frequency–Inverse Document Frequency (TF-IDF) weighting. The algorithm also assumes that each word is conditionally independent of other words in determining the document class [15], [16]. These characteristics make Multinomial Naïve Bayes an efficient and widely adopted algorithm for sentiment analysis research.

Previous studies have applied various classification algorithms for sentiment analysis on social media. Dina et al. [6] implemented Naïve Bayes on TikTok comments and reported an accuracy of 71.55%. Rachman et al. [17] compared Naïve Bayes and Support Vector Machine (SVM) and reported higher performance from SVM, although Naïve Bayes required lower computational complexity. In studies related to the MBG Program, Putra et al. [18] reported an accuracy of 88% using SVM, while other studies have investigated Naïve Bayes and alternative approaches. These findings indicate that model selection can influence sentiment-classification performance.

Although previous studies have examined MBG-related sentiment and social-media sentiment classification, research specifically focused on comments from the official National Nutrition Agency TikTok account and using IndoRoBERTa-based automatic labeling followed by a TF-IDF–Multinomial Naïve Bayes pipeline remains limited. The contribution of this study is therefore positioned primarily as a case study and methodological evaluation rather than as a claim of algorithmic novelty. The study examines public responses to the MBG Program in a TikTok-specific context, documents the complete preprocessing and balancing workflow, and evaluates how well a conventional Multinomial Naïve Bayes classifier can reproduce sentiment categories generated by an Indonesian transformer-based sentiment model. Because the labels are automatically generated, the study treats them as pseudo-labels produced under a weak-supervision setting rather than as fully verified ground truth. Accordingly, the findings provide an initial view of public response and a baseline classification workflow, while recognizing that human validation and comparison with stronger classifiers are needed to establish broader validity.

2. Method

Research Procedure

This study was conducted through a series of systematic stages to achieve its primary objective, namely classifying public sentiment toward the implementation of the Free Nutritious Meal (MBG) Program on the TikTok social media platform. The research procedure consisted of six main stages: problem identification, literature review, data collection, dataset preparation, process design, and model evaluation. Each stage was sequentially executed, with the output of one stage serving as the input for the subsequent stage, thereby ensuring a structured and comprehensive research workflow. The overall research procedure is illustrated in **Figure 1**.

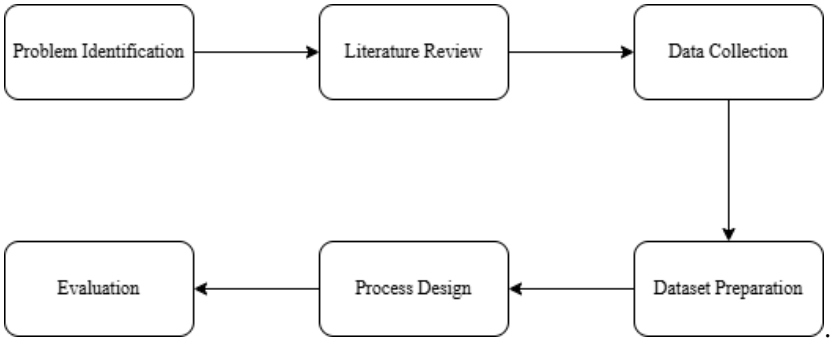

Figure 1. Research Procedure

The problem identification stage was conducted to identify the issues arising from the large number of TikTok user comments regarding the implementation of the Free Nutritious Meal (MBG) Program. Subsequently, a literature review was carried out to establish the theoretical foundation and determine the most appropriate research methodology. The comment data were collected through a web scraping process using the Apify platform. After the data collection stage, attribute selection was performed by retaining only the comment text, followed by automatic sentiment labeling into positive and negative categories using the IndoRoBERTa model.

The prepared dataset was then processed through several stages, including text preprocessing, data balancing, train–test splitting, TF-IDF feature extraction, and sentiment classification using the Multinomial Naïve Bayes algorithm. Finally, the classification model was evaluated using a confusion matrix and four performance metrics, namely accuracy, precision, recall, and F1-score, to assess its effectiveness in classifying sentiment from TikTok comments [20], [21].

Data Collection

The research data were obtained through a scraping process using the Apify platform to collect comments from the official TikTok account @badangizinasional.ri related to the Free Nutritious Meal (MBG) Program. The data collection process began by selecting relevant TikTok video URLs discussing the MBG Program. These URLs were then entered into the Apify platform to automatically extract user comments from each video [22].

The collected comments were then automatically labeled using the IndoRoBERTa sentiment-classification model. The model used in this study was w11wo/indonesian-roberta-base-sentiment-classifier, an Indonesian RoBERTa model fine-tuned for sentiment classification. The model output was mapped into two sentiment categories used in this study, namely positive and negative. The resulting labels were treated as pseudo-labels because they were generated automatically and were not used as human-verified ground truth. This distinction is important because errors or biases in the IndoRoBERTa predictions may propagate into the subsequent Multinomial Naïve Bayes training and evaluation. A manual validation of a representative sample of the labels was not conducted in the original experiment; therefore, the reliability of the pseudo-labels remains a limitation of this study and should be addressed in future research through independent human annotation and inter-annotator agreement.

The overall data collection procedure is illustrated in **Figure 2**.

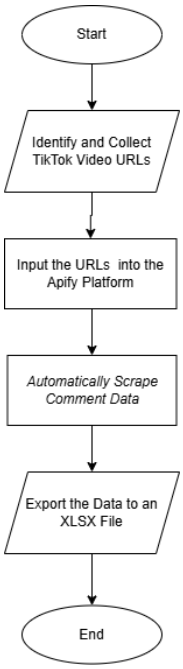

Figure 1. Scraping Data

Process Design

The process design stage began with text preprocessing to reduce noise and standardize Indonesian social-media text. The preprocessing pipeline consisted of case folding, cleaning, tokenization, word normalization, stopword removal, and stemming. Word normalization used a normalization dictionary to map informal, abbreviated, or non-standard forms into their standardized equivalents. Stopword removal used an Indonesian stopword list to remove high-frequency function words that were considered less informative for classification. Stemming was applied to reduce inflected Indonesian words to their root forms. Because the exact normalization dictionary and stopword list can influence the resulting vocabulary, these resources should be reported or made available with the dataset in a reproducible version of the study. The preprocessing stage was followed by data balancing using undersampling to address the class imbalance.

The preprocessed text data were transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF assigns a weight to each term based on its frequency within a document and its distribution across the entire document collection, thereby emphasizing informative terms while reducing the influence of commonly occurring words [28]. Mathematically, Term Frequency (TF) is defined as the number of occurrences of term k in document j, as expressed in Equation (1).

$$TF\left(t_k, d_j\right) = f(t_k, d_j) \tag{1}$$

Where $f(t_k, d_j)$ represents the frequency of occurrence of *term* t_k in document d_j.

Subsequently, Inverse Document Frequency (IDF) was used to measure the uniqueness of a term across the document collection [29]. Terms that appear in fewer documents receive higher IDF values, whereas terms that occur frequently across many documents receive lower values. The IDF value was calculated using Equation (2).

$$IDF\left(t_k\right) = \log\frac{D}{df(t)} \tag{2}$$

Where D denotes the total number of documents in the corpus, and $df(t_k)$ represents the number of documents containing *term* t_k.

The final TF-IDF weight was obtained by multiplying the TF and IDF values, as expressed in Equation (3).

$$TF\ IDF\ (\ t_k\ ,d_j) = TF(t_k, d_j) * \ IDF\ (\ t_k\ ) \tag{3}$$

The resulting TF-IDF weights were then used as the numerical representation of the documents before the classification process using the Multinomial Naïve Bayes algorithm. Mathematically, the Naïve Bayes method is formulated based on Equation (4).

$$P(Y|X) = \frac{P(X|Y) \cdot \ P(Y)}{P(X)} \tag{4}$$

At Equation (4), P(Y|X) represents the posterior probability, which is the probability of class Y given the observed data X. The term P(X|Y) denotes the probability of observing data X given class Y, P(Y) represents the prior probability of class Y, while P(X) denotes the probability of observing data X [30]. In the classification process, the prior probability of each class is first calculated, followed by the likelihood of each feature for the corresponding class. To prevent zero-probability problems, Laplace smoothing is applied. Finally, the posterior probability is computed to determine the most appropriate sentiment class for each testing instance [31].

The TF-IDF feature extraction and Multinomial Naïve Bayes classification processes were integrated using the Scikit-learn Pipeline, enabling both feature transformation and classification to be performed automatically within a single workflow. To obtain the optimal model, GridSearchCV with 5-fold cross-validation was employed for hyperparameter optimization [32]. The evaluated hyperparameters included ngram_range, min_df, max_df, sublinear_tf, and alpha. The best-performing parameter combination was subsequently selected as the final model for the testing phase.

Testing

The evaluation stage was conducted to measure the performance of the proposed model in classifying sentiment from TikTok comments. The trained model was used to predict the sentiment labels of the testing dataset, and the predicted results were then compared with the actual labels using a confusion matrix [33]. Based on the confusion matrix, the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were obtained and subsequently used to calculate the evaluation metrics, namely accuracy, precision, recall, and F1-score [34].

The accuracy metric measures the overall proportion of correctly classified data and is calculated using Equation (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{5}$$

Precision measures the proportion of correctly predicted positive instances among all instances predicted as positive and is calculated using Equation (6).

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

Recall measures the model's ability to correctly identify all actual positive instances and is calculated using Equation (7).

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

Furthermore, the F1-score is the harmonic mean of precision and recall, providing a more balanced evaluation of the classification performance, particularly for datasets with imbalanced class distributions [35]. Its calculation is presented in Equation (8).

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \times 100\% \quad (8)$$

These four-evaluation metrics were used to assess the performance of the Multinomial Naïve Bayes model in classifying TikTok comments related to the Free Nutritious Meal (MBG) Program, thereby providing a comprehensive evaluation of the model in terms of accuracy, precision, recall, and overall classification balance.

3. Result and Discussion

Dataset Preparation

A total of 1,048 comments were collected through a web scraping process using the Apify platform from the official TikTok account @badangizinasional.ri. The scraped dataset contained several attributes, including text, diggCount, replyCommentTotal, createTimeISO, uniqueId, videoWebUrl, uid, cid, and avatarThumbnail. However, this study only utilized the text attribute because it contains the comment content used as the object of sentiment analysis, whereas the remaining attributes were excluded as they were not relevant to the sentiment classification process. An example of the attribute selection results is presented in **Table 1**.

Table 1. Attribute Selection Results

No.	*Text*
1.	*makasih min mbg nya enak banget*
2.	*agar segera di tindak lanjuti, dapur makan bergizi gratis yg dibangun diatas tanah kodim 0827/Sumenep setiap hari digunakan untuk latihan sepeda motor cross yang pastinya sangat berdebu, hal ini sangat tidak higienis dan dapat dipastikan dapurnya akan tercemar oleh debu akibat latihan cross tersebut, lagian dapur MBG kok dibangun di lapangan cross*
3.	"*saya setuju* *yang utama anak itu makan bergizi dulu* *demikian motorik anak akan cerdas*."
4.	*Maju Terus BGN*
⋮	⋮
1045.	*di SDN cadasngampar cirebon*
1046.	*kayaknya di provinsi kaka ada yg belum buat akun payroll. atau udah semua?*
1047.	*iya bang mau makan apa*
1048.	*Saya juga pusing ini bu hehe*

Subsequently, each comment was automatically assigned a sentiment label using the IndoRoBERTa model and classified into two sentiment categories, namely positive and negative. The labeling results indicate that the dataset is dominated by negative sentiment, as presented in **Table 2**.

Table 2. Sentiment Labeling Results

Sentiment	Number of Comments	Percentage
Positive	336	32.06%
Negative	712	67.94%
Total	**1,048**	**100%**

The next stage was text preprocessing, which was performed to improve the quality of the dataset before the classification process [36]. The preprocessing steps included case folding, cleaning, tokenization, word normalization, stopword removal, and stemming. Examples of the preprocessing results are presented in **Table 3**.

Table 3. Preprocessing Results

No.	Comment	Preprocessing Result
1.	*Pak SD dkt rmh saya blm dpt makan bergizi..*	*sd dkt rmh belum makan gizi*
2.	*pak prabowo,, program bapak sangat bermanfaat sekali pak,, kami liat ini menangisss,,,*	*prabowo program manfaat liat menang*
3.	*semoga dengan program ini bisa menjadi negara lebih baik dan maju*	*moga program bisa negara baik maju*
4.	*sujud syukur hamba kalo di respon*	*sujud syukur hamba respon*
5.	*silahkan simak video YouTube saya berjudul CARUT MARUT PROGRAM MBG PRESIDEN PRABOWO*	*silah simak video youtube judul carut marut program makan gizi gratis presiden prabowo*

After the preprocessing stage, the dataset consisted of 1,028 comments. Since the class distribution remained imbalanced, an undersampling technique was applied to balance the dataset, resulting in 333 positive comments and 333 negative comments. A summary of the number of comments at each stage of the data preparation process is presented in **Table 4**.

Table 4. Number of Comments at Each Stage of Data Preparation

Stage	Positive	Negative	Total Comments
Sentiment Labeling	336	712	1,048
After Preprocessing	333	695	1,028
After Undersampling	333	333	666

The balanced dataset obtained through undersampling was subsequently used for TF-IDF feature extraction, Multinomial Naïve Bayes model training, and classification performance evaluation.

Model Development

The undersampled dataset was then divided into 80% training data and 20% testing data using the stratified splitting method to preserve the proportion of each sentiment class. The distribution of the training and testing datasets is presented in **Table 5**.

Table 5. Splitting Data

Dataset	Positive	Negative	Total Comments
Training (80%)	266	266	532
Testing (20%)	67	67	134
Total	**333**	**333**	**666**

The text data were subsequently converted into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The TF-IDF feature extraction and Multinomial Naïve Bayes classification processes were integrated using the Scikit-learn Pipeline, allowing feature transformation and classification to be performed automatically within a single workflow. To obtain the optimal model, hyperparameter optimization was

carried out using GridSearchCV with 5-fold cross-validation. The evaluated parameters included alpha, ngram_range, min_df, max_df, and sublinear_tf.

The optimization results indicated that the best parameter combination consisted of alpha = 0.5, max_df = 0.8, min_df = 3, ngram_range = (1,3), and sublinear_tf = False, achieving a cross-validation accuracy of 78.95%, as presented in Table 6.

Table 6. Best Hyperparameter Configuration Obtained Using GridSearchCV

Parameter	Best Value
Alpha	0.5
Max_df	0.8
Min_df	3
Ngram_range	(1, 3)
Sublinear_tf	False
Validation Accuracy	**78.95%**

The model with the best hyperparameter configuration was subsequently evaluated using the testing dataset. Its classification performance was assessed based on the accuracy, precision, recall, F1-score, and confusion matrix metrics.

Classification Performance

The best model obtained from the hyperparameter optimization process was then used to predict the sentiment labels of 134 testing comments. The model performance was evaluated using the accuracy, precision, recall, F1-score, and confusion matrix metrics. The overall performance evaluation results of the Multinomial Naïve Bayes model are presented in Table 7.

Table 7. Classification Performance of The Multinomial Naïve Bayes Model

Evaluation Metric	Value (%)
Accuracy	80.60
Precision	81.04
Recall	80.60
F1-score	80.53

In addition to the overall performance evaluation, the classification performance for each sentiment class was analyzed using the classification report. The evaluation results for each class are presented in Table 8.

Table 8. Classification Performance for Each Sentiment Class

Sentiment Class	*Precision*	*Recall*	*F1-Score*
Positive	0.85	0.75	0.79
Negative	0.77	0.87	0.82

The results presented in Table 8 indicate that the proposed model achieved satisfactory performance in classifying both sentiment classes. The negative class obtained a higher recall value (0.87), indicating that most negative comments were correctly identified by the model. In contrast, the positive class achieved a higher precision value (0.85), indicating that predictions classified as positive were generally more accurate.

The visualization of the confusion matrix is presented in Figure 3. As shown in the confusion matrix, the model correctly classified 108 out of 134 testing comments, consisting of 58 negative and 50 positive comments. Meanwhile, 26 comments were misclassified, indicating that some positive and negative comments shared similar textual characteristics, making them more difficult to distinguish. Nevertheless, the model achieved an accuracy of 80.60%,

demonstrating that the Multinomial Naïve Bayes algorithm is effective in classifying TikTok comments related to the Free Nutritious Meals (MBG) Program.

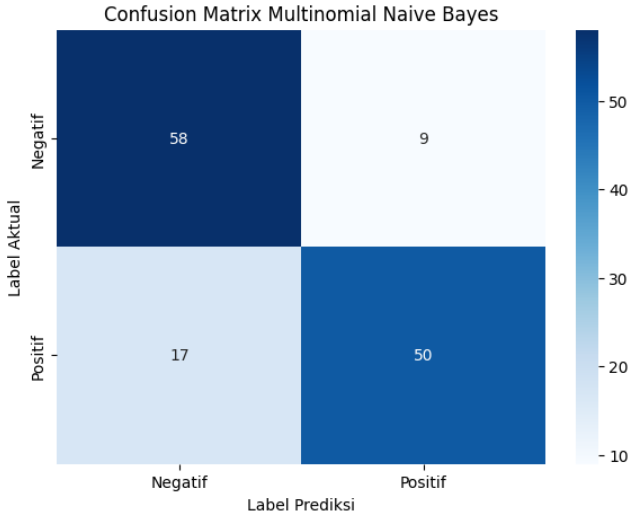

Figure 2. Visualisasi Confusion Matrix

Discussion

The results of this study demonstrate that the Multinomial Naïve Bayes algorithm effectively classified the sentiment of TikTok comments related to the Free Nutritious Meals (MBG) Program, achieving an accuracy of 80.60%. This result indicates that the combination of TF-IDF feature weighting and the Multinomial Naïve Bayes algorithm was able to effectively represent textual characteristics and distinguish between positive and negative comments. Furthermore, the implementation of a Scikit-learn Pipeline enabled feature extraction and classification to be performed automatically within a single workflow, thereby reducing the possibility of errors during the modeling process.

The sentiment distribution revealed that negative comments dominated the post-preprocessing dataset, accounting for 695 comments (67.61%), while 333 comments (32.39%) were classified as positive by the IndoRoBERTa pseudo-labeling process. This distribution should not be interpreted as a directly verified measure of public sentiment because the labels were generated automatically and were not manually validated. Nevertheless, within the pseudo-labeled dataset, negative sentiment was more prevalent and may indicate concerns expressed by users about aspects of MBG implementation. Based on the content of the comments examined in this study, recurring concerns included food quality, distribution equity, facility readiness, and implementation problems. Positive comments generally expressed support, appreciation, or expectations that the program would continue to improve. These observations should be interpreted as thematic indications from the collected comments rather than as a formal topic model or aspect-based sentiment analysis.

The obtained accuracy is lower than the 88% accuracy reported in the cited SVM-based MBG study [18]. However, direct comparison should be made cautiously because the studies differ in data sources, dataset sizes, labeling procedures, preprocessing, balancing strategies, feature representations, and experimental settings. In the present study, the target labels were generated automatically by IndoRoBERTa, whereas the external studies may use different labeling procedures. Therefore, the 80.60% accuracy reported here primarily measures how well Multinomial Naïve Bayes reproduces the IndoRoBERTa-generated pseudo-labels under the selected experimental configuration. It should not be interpreted as definitive evidence that the model captures human-verified public sentiment.

The confusion matrix showed that 108 of 134 testing comments were classified correctly and 26 were misclassified. These errors are plausible in TikTok comments because social-media language often contains slang,

abbreviations, spelling variations, sarcasm, implicit meaning, code-mixing, and mixed positive-negative expressions. A comment may also mention a positive aspect of the MBG Program while simultaneously criticizing its implementation, making a binary sentiment decision difficult. The preprocessing pipeline can reduce lexical variation but may also remove contextual cues that are useful for interpreting sarcasm or mixed sentiment. Because the original experiment did not retain a systematic manual error-analysis annotation for all 26 misclassified comments, specific causes cannot be quantified reliably. Future work should manually inspect these errors and group them into categories such as sarcasm, ambiguous sentiment, mixed sentiment, informal language, and preprocessing-induced information loss.

This study contributes a case-specific analysis of TikTok comments related to the MBG Program collected from the official National Nutrition Agency account and evaluates a conventional TF-IDF–Multinomial Naïve Bayes pipeline on IndoRoBERTa-generated pseudo-labels. The contribution is therefore centered on the dataset context, the documented processing workflow, and the empirical evaluation of a lightweight classifier rather than on proposing a new classification algorithm. The findings provide an initial computational description of public responses in the collected TikTok data, while the limitations of automatic labeling and undersampling mean that the results should be considered exploratory rather than a definitive measurement for policy evaluation.

4. Conclusion

This study developed a sentiment classification pipeline for TikTok comments concerning the Free Nutritious Meals (MBG) Program using text preprocessing, TF-IDF feature extraction, and Multinomial Naïve Bayes classification. The data were collected from the official National Nutrition Agency TikTok account using Apify. Of 1,048 collected comments, 1,028 remained after preprocessing. Automatic IndoRoBERTa-based pseudo-labeling produced 333 positive and 695 negative comments, after which undersampling generated a balanced dataset of 666 comments. The balanced data were split using an 80:20 stratified train–test split and classified using a TF-IDF–Multinomial Naïve Bayes pipeline optimized through GridSearchCV with 5-fold cross-validation.

The experimental results showed that the model achieved 80.60% accuracy, 81.04% precision, 80.60% recall, and 80.53% F1-score on the testing set. These values indicate reasonable agreement between Multinomial Naïve Bayes predictions and the IndoRoBERTa-generated pseudo-labels under the selected experimental configuration. The original post-preprocessing pseudo-label distribution was dominated by the negative class (67.61%), but this distribution was altered by undersampling before model training. Therefore, the observed negative sentiment proportion should be interpreted only as a characteristic of the automatically labeled dataset and not as a definitive estimate of verified public opinion. The study also cannot establish that Multinomial Naïve Bayes is the best classifier for this task because no controlled baseline comparison with SVM, Logistic Regression, Random Forest, or other classifiers was performed.

Several limitations should be considered. First, the sentiment labels were generated automatically using IndoRoBERTa and were not manually validated, so label noise and model bias may affect the classification results. Second, undersampling removed 362 negative comments from the post-preprocessing dataset, potentially reducing the diversity of negative expressions available for model training. Third, the study used comments from a specific TikTok account and therefore may not represent all Indonesian public opinion about the MBG Program. Fourth, the absence of a controlled comparison with alternative classifiers limits conclusions about the relative suitability of Multinomial Naïve Bayes. Future research should conduct manual validation of a representative sample of pseudo-labels, report human–model agreement, use larger and more diverse TikTok datasets, compare SVM, Logistic Regression, Random Forest, and transformer-based classifiers under the same data split, and investigate topic or aspect-based sentiment analysis to identify the specific implementation issues underlying positive and negative responses.

References:

[1] W. F. Satrya, R. Aprilliyani, and E. H. Yossy, “Sentiment analysis of Indonesian police chief using multi-level ensemble model,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 620–629, 2022, doi: https://doi.org/10.1016/j.procs.2022.12.177.

[2] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, no. August 2022, 2023, doi: https://doi.org/10.1016/j.eswa.2023.119862.

[3] Z. Yao, J. Yang, J. Liu, M. Keith, and C. H. Guan, "Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19," *Cities*, vol. 116, no. July 2020, p. 103273, 2021, doi: https://doi.org/10.1016/j.cities.2021.103273.

[4] F. Guo, Z. Liu, Q. Lu, S. Ji, and C. Zhang, "Public Opinion About COVID-19 on a Microblog Platform in China: Topic Modeling and Multidimensional Sentiment Analysis of Social Media," *J. Med. Internet Res.*, vol. 26, no. 1, 2024, doi: https://doi.org/10.2196/47508.

[5] M. Isnan, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput. Sci.*, vol. 227, pp. 168–175, 2023, doi: https://doi.org/10.1016/j.procs.2023.

[6] D. F. M. Dina, T. Haryanti, and M. A. Haq, "Analisis Sentimen Terhadap Komentar Pada Media Sosial Tiktok Yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naive Bayes," *Comput. Insight J. Comput. Sci.*, vol. 7, no. 1, pp. 1–9, 2025, doi: https://doi.org/10.30651/comp_insight.v7i1.26327.

[7] R. Haque, N. Islam, M. Tasneem, and A. K. Das, "Multi-class sentiment classification on Bengali social media comments using machine learning," *Int. J. Cogn. Comput. Eng.*, vol. 4, no. December 2021, pp. 21–35, 2023, doi: https://doi.org/10.1016/j.ijcce.2023.01.001.

[8] N. Puspitasari, R. Rosmasari, F. W. Pratama, and H. Sulastri, "Quality Classification of Palm Oil Varieties Using Naive Bayes Classifier," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 11–23, 2022, doi: https://doi.org/10.31849/digitalzone.v13i1.9773.

[9] D. G. Pradana, M. L. Alghifari, M. F. Juna, and D. Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 55–60, 2022, doi: https://doi.org/10.56705/ijodas.v3i2.35.

[10] M. R. Ananda *et al.*, "Implementasi Support Vector Machine Dalam Analisis Sentimen Netizen Terhadap Program Makan Bergizi Gratis," *J. Mhs. Tek. Inform.*, vol. 10, no. 2, pp. 3594–3601, 2026.

[11] W. Maharani, H. Daud, N. Muhammad, and E. A. Kadir, "Leveraging Social Media Data for Forest Fires Sentiment Classification: A Data-Driven Method," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 3, pp. 392–407, 2024, doi: https://doi.org/10.20473/jisebi.10.3.392-407.

[12] W. N. Mergianti, N. Khudin, A. Afandi, N. A. Shofah, and D. C. R. Novitasari, "Analisis Sentimen Pengguna Aplikasi X Terhadap Pelaksanaan Makan Bergizi Gratis Menggunakan Metode Adaptive Neuro-Fuzzy Inference System," *Sains Data J. Stud. Mat. dan Teknol.*, vol. 1, pp. 24–33, 2026, doi: https://doi.org/10.10.52620/sainsdata.v4i1.339 ABSTRAK.

[13] Z. P. Tiararosa, I. S. Wijaya, and E. Rohaini, "Perbandingan SVM dan Naive Bayes pada Analisis Sentimen MBG di Platform TikTok dan Instagram," *J. Inform. dan Rekayasa Komput.*, vol. 6, no. April, pp. 1983–1993, 2026, doi: https://doi.org/10.33998/jakakom.v6i1.

[14] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes," *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: https://doi.org/10.15294/sji.v10i1.39952.

[15] D. Bor, B. S. Lee, and E. J. Oughton, "Quantifying polarization across political groups on key policy issues using sentiment analysis," pp. 1–31, 2021.

[16] R. W. Harwenda, M. D. Angelo, I. Budi, A. B. Santoso, and P. K. Putra, "Sentiment Analysis on Government Public Policies: A Systematic Literature Review," *Dinasti Int. J. Educ. Manag. Soc. Sci.*, vol. 6, no. 5, pp.

4192–4211, 2025, doi: https://doi.org/10.38035/dijemss.v6i5.4699.

[17] Herlawati, D. B. Srisulistiowati, S. C. Agustin, P. H. Syafina, N. Rachmatin, and S. Setiawati, "Metode Naïve Bayes dan Support Vector Machine untuk Mengolah Sentimen Ulasan dan Komentar di Platform Digital," *J. Students' Res. Comput. Sci.*, vol. 5, no. 2, pp. 197–212, 2024, doi: https://doi.org/10.31599/dby15h32.

[18] J. A. Zulqornain, Indriati, and P. P. Adikara, "Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorial Propotional Difference (CPD)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 7, pp. 2886–2890, 2021.

[19] J. A. Widians, U. Hairah, and T. Fadila, "Sistem Pakar Diagnosis Penyakit Mioma Uteri Menggunakan Teorema Bayes," *J. Rekayasa Teknol. Inf.*, vol. 8, no. 2, p. 174, 2024, doi: https://doi.org/10.30872/jurti.v8i2.1446.

[20] S. K. Q. Asit, C. M. Nahumury, J. Feninlambir, W. ode N. Hasiyati, and F. S. Tomia, "Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan Dfd Dan Flowchart Pada Bengkel Latansa Walding," *J. Tagalaya Pengabdi. Kpd. Masy.*, vol. 1, no. 2, pp. 193–200, 2024, doi: https://doi.org/10.71315/jtpkm.v1i2.30.

[21] N. Salsabilah, Irawati, and L. N. Hayati, "Performance Analysis of Convolutional Neural Networks and Naive Bayes Methods for Disease Classification in Tomato Plant Leaves," *Indones. J. Data Sci.*, vol. 6, no. 3, pp. 445–453, 2025, doi: https://doi.org/10.56705/ijodas.v6i3.255.

[22] C. Hill, F. Irshaidat, M. Johnson, and J. Fresneda, "An analytical assessment of sentiment analysis trends and methods through systematic review and topic modeling," *Decis. Anal. J.*, vol. 17, no. October, p. 100644, 2025, doi: https://doi.org/10.1016/j.dajour.2025.100644.

[23] N. P. Yusal, A. Erfina, and C. Warman, "Sarcasm and Irony Detection in Lazada App Reviews Using IndoBERT," vol. 6, no. 3, pp. 618–625, 2025.

[24] C. Shaw, P. LaCasse, and L. Champagne, "Exploring emotion classification of indonesian tweets using large scale transfer learning via IndoBERT," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, pp. 1–12, 2025, doi: https://doi.org/10.1007/s13278-025-01439-6.

[25] M. A. Imbiri, L. Y. Baisa, and J. J. A. Limbong, "Sentiment Classification and Influential Actor Detection on Twitter (Case Study : The Raja Ampat Mining Conflict)," vol. 7, no. 1, pp. 12–28, 2026.

[26] N. N. A. S. Wahyuni, I. G. I. Sudipa, N. N. A. J. Sastaparamitha, A. G. Willdahlia, and I. G. A. A. M. Aristamy, "Public Response on X to the Revocation of Indonesia's 3-Kg LPG Retail Ban: A Support Vector Machine Study," *Indones. J. Data Sci.*, vol. 6, no. 3, pp. 412–425, 2025, doi: https://doi.org/10.56705/ijodas.v6i3.349.

[27] N. Raveendhran and N. Krishnan, "A novel hybrid SMOTE oversampling approach for balancing class distribution on social media text," *Bull. Electr. Eng. Informatics*, vol. 14, no. 1, pp. 638–646, 2025, doi: https://doi.org/10.11591/eei.v14i1.8380.

[28] P. Guleria, J. Frnda, and P. N. Srinivasu, "NLP based text classification using TF-IDF enabled fine-tuned long short-term memory: An empirical analysis," *Array*, vol. 27, no. August, p. 100467, 2025, doi: https://doi.org/10.1016/j.array.2025.100467.

[29] L. Zhang, "Features extraction based on Naive Bayes algorithm and TF-IDF for news classification," *PLoS One*, vol. 20, no. 7 July, pp. 1–17, 2025, doi: https://doi.org/10.1371/journal.pone.0327347.

[30] N. Kewsuwun and S. Kajornkasirat, "A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2829–2838, 2022, doi: https://doi.org/https://doi.org/10.1016/j.rineng.2024.102148

[31] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia*

Comput. Sci., vol. 161, pp. 765–772, 2021, doi: https://doi.org/10.1016/j.procs.2019.11.181.

[32] M. A. K. Raiaan *et al.*, "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," *Decis. Anal. J.*, vol. 11, no. April, p. 100470, 2024, doi: https://doi.org/10.1016/j.dajour.2024.100470.

[33] Y. N. Fuadah, M. A. Pramudito, and K. M. Lim, "An Optimal Approach for Heart Sound Classification Using Grid Search in Hyperparameter Optimization of Machine Learning," *Bioengineering*, vol. 10, no. 1, 2023, doi: https://doi.org/10.3390/bioengineering10010045.

[34] S. I. G. Iwan, M. Habibi, E. S. Jullev Atmadji, and I. Arfiani, "Predictive Modeling of Air Quality Levels Using Decision Tree Classification: Insights from Environmental and Demographic Factors," *Indones. J. Data Sci.*, vol. 5, no. 3, pp. 251–258, 2024, doi: https://doi.org/10.56705/ijodas.v5i3.201.

[35] C. Dewi, R. C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam J. Comput. Sci.*, vol. 10, no. 4, pp. 485–498, 2023, doi: https://doi.org/10.1142/S2196888823500100.

[36] M. Yusrizal and T. B. Sasongko, "Analisis Sentimen Masyarakat Terhadap Presiden dan Calon Presiden Terpilih 2024 Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1673, 2024, doi: https://doi.org/10.30865/mib.v8i3.7882.