
Sentiment Analysis of the Free Nutritious Meal (MBG) Program on TikTok Using Naïve Bayes Classifier

Sukma Ajeng Griyandari^{1*}, Fahrur Rozi², Yayak Kartika Sari³

^{1,2,3}Bhinneka PGRI University, Faculty of Science and Technology, Information Technology Education Study Program, Mayor Sujadi Street 103B Tulungagung, East Java, Indonesia

Article History

Submitted: 13-07-2026

Revised: 14-08-2026

Published: 26-09-2026

DOI: 10.32664/j-intech.v14i03.2458

Keywords

Free Nutritious Meal Program; Naive Bayes Classifier; Sentiment Analysis; TikTok; TF-IDF

*Corresponding Author:

sukmagriyandari@gmail.com

Abstract

The Free Nutritious Meal (MBG) Program is an Indonesian government initiative to improve nutritional quality and support human resource development. As a public policy, it has generated diverse public responses on social media, particularly TikTok. The novelty of this study lies in utilizing TikTok comments, which remain underexplored in previous MBG sentiment studies, and employing expert-validated manual sentiment labeling to improve annotation reliability. This study applies the Naïve Bayes Classifier to analyze public sentiment toward the MBG Program using 3,190 TikTok comments, of which 1,294 were retained after relevance filtering, data cleaning, and manual labeling validated by two Indonesian language experts. Text preprocessing included cleaning, case folding, tokenization, normalization, stopword removal, and stemming, followed by TF-IDF feature weighting. The model was evaluated using stratified 10-fold cross-validation. It achieved 72.80% accuracy, 69.81% weighted precision, 73.25% weighted recall, and 71.49% F1-score. Negative sentiment dominated public discussion of the MBG Program. These findings demonstrate that combining expert-validated manual labeling, TF-IDF, and the Naïve Bayes Classifier provides a reliable approach for sentiment analysis of Indonesian social media text and offers useful insights for evaluating the implementation of the MBG Program.

1. Introduction

The rapid growth of social media has changed the way people express opinions and respond to public policies. This phenomenon has been driven by the continuous increase in internet usage in Indonesia, as reported by the Indonesian Internet Service Providers Association [1]. Among various social media platforms, TikTok has become one of the most widely used applications in Indonesia, allowing users to share short videos and interact through comments. Indonesia has the largest TikTok user population in the world, with approximately 157 million users [2]. These comment sections contain valuable information that reflects public perceptions, opinions, and reactions toward current issues. Therefore, social media data have become an important source

for sentiment analysis, enabling researchers to understand public responses more efficiently than traditional survey methods.

One of the Indonesian government programs that has attracted considerable public attention is the Free Nutritious Meals Program (Makan Bergizi Gratis or MBG). The program aims to improve nutritional intake among school children and other target groups while supporting human resource development [3]. Although the policy is expected to provide long-term social benefits, its implementation has generated mixed public responses. Discussions on social media reveal both supportive and critical opinions regarding food quality, distribution, implementation, and the overall effectiveness of the program [4]. Understanding these opinions is important because public sentiment may provide valuable feedback for evaluating government policies.

The enormous volume of user-generated comments makes manual analysis impractical and time-consuming. Text mining techniques, particularly sentiment analysis, provide an effective solution for automatically classifying opinions into positive, negative, and neutral sentiments [5], [6]. Among various machine learning algorithms, Naïve Bayes is widely used because of its computational efficiency, simplicity, and competitive performance in text classification tasks [7], [8]. Several previous studies have shown that Naïve Bayes combined with TF-IDF weighting can achieve satisfactory results for sentiment analysis across different social media platforms [9], [10], [11].

Several previous studies have investigated public sentiment toward the MBG program using social media data [12], [13], [14]. Most of these studies analyzed comments from platform X (formerly Twitter), while research focusing on TikTok remains limited despite the platform's large user base in Indonesia [2], [12], [14]. Furthermore, TikTok comments tend to contain informal language, abbreviations, slang expressions, and emojis, making sentiment classification more challenging [15]. This gap highlights the need for further investigation using TikTok data and appropriate preprocessing techniques to improve text representation before classification.

To address this research gap, this study analyzes public sentiment toward the Free Nutritious Meals (MBG) Program using TikTok comments and the Naïve Bayes Classifier algorithm. The preprocessing stage was implemented using Python integrated into RapidMiner to improve text normalization before feature extraction, followed by TF-IDF feature weighting and model evaluation using 10-fold cross-validation. The findings provide a clearer understanding of public opinion regarding the MBG program and demonstrate the applicability of the Naïve Bayes approach for sentiment analysis of Indonesian social media text.

Accordingly, the contributions of this study are as follows: (1) a curated dataset of 1,294 TikTok comments on the MBG Program with sentiment labels validated by two Indonesian language specialists; (2) an integrated preprocessing pipeline combining Python and RapidMiner Altair AI Studio for Indonesian text normalization, including informal language and slang; and (3) empirical evidence of the performance of the Naïve Bayes Classifier with TF-IDF on TikTok-sourced Indonesian social media text using stratified 10-fold cross-validation.

2. Research Method

This study employed a quantitative approach combined with text mining techniques to analyze public sentiment toward the Free Nutritious Meals (Makan Bergizi Gratis or MBG) Program based on TikTok comments. Text mining enables the extraction of meaningful patterns from large collections of unstructured textual data, while quantitative analysis is used to evaluate the classification model through measurable performance metrics [16], [17]. Sentiment classification was performed using the Naïve Bayes Classifier because of its computational efficiency and strong performance in text classification tasks [18]. The research procedure consisted of data collection, data selection, manual sentiment labeling, text preprocessing, TF-IDF feature weighting, sentiment classification using the Naïve Bayes algorithm, and model evaluation through 10-fold cross-validation. The preprocessing stage was implemented using Python integrated into RapidMiner Altair AI Studio to automate text normalization before feature extraction and classification. This integrated workflow ensured that all stages of data processing and model evaluation were conducted consistently within a single experimental environment. The overall research framework is illustrated in Figure 1.

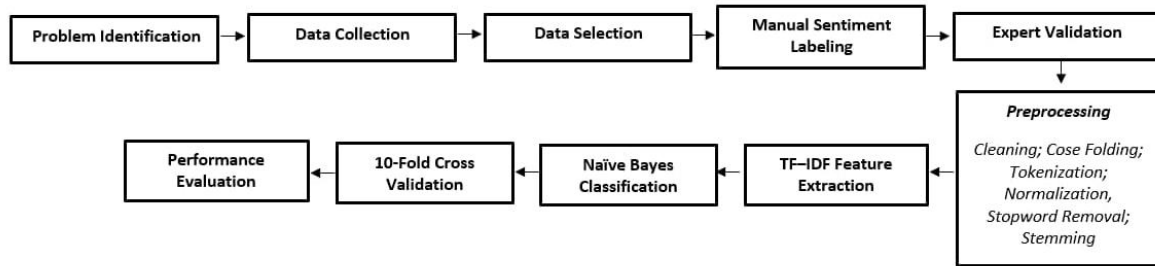


Figure 1. Research Flowchart

2.1 Data Collection and Data Selection

The data used in this study were secondary data collected from comments on a TikTok video discussing the Free Nutritious Meals (Makan Bergizi Gratis or MBG) Program uploaded by the @bbcnewsindonesia account. Comments were collected using the Apify TikTok Comments Scraper, which automatically extracts publicly available comments from TikTok videos. The collected comments reflected public responses during the implementation of the MBG program. Initially, 3,190 comments were successfully collected and stored in spreadsheet format as the raw dataset.

This study has a single-video sampling limitation because all comments were collected from one TikTok video uploaded by the @bbcnewsindonesia account. Therefore, the characteristics of the commenters and the editorial framing of the selected video may influence the comments, and the findings may not be generalizable to MBG sentiment across TikTok as a whole.

Since the raw dataset contained various types of irrelevant and noisy data, a manual data selection process was conducted using a purposive sampling approach. Comments were retained only if they contained opinions related to the MBG program. Comments consisting solely of emojis, advertisements, spam, duplicated content, unrelated discussions, or meaningless text were removed. In addition, excessively long comments containing multiple opinions were simplified through a cut-off process by preserving the dominant opinion without altering the original sentiment. This procedure was intended to improve data consistency and reduce ambiguity during the sentiment classification process. For excessively long comments, the text was shortened by retaining the core statement that best represented the overall meaning of the comment, while unnecessary supporting details were removed. In operational terms, the dominant opinion was identified by examining the overall context and the main emphasis of each comment. When a comment contained multiple opinions or ambiguous sentiment expressions, the researcher retained the statement that most clearly represented the commenter's primary stance and removed supporting statements that were not essential to conveying that stance. The cut-off decision was made during manual data selection and was discussed with the two Indonesian language experts during the labeling process. For example, a comment stating "MBG ini bagus kok, tapi saya kecewa pemerintah lambat" was simplified to "Saya kecewa pemerintah lambat" because the latter represented the main emphasis of the commenter's overall stance.

After the filtering and selection stages, 1,294 valid comments remained and were used as the final dataset for subsequent analysis. The selected dataset was then prepared for subsequent text mining analysis following the general Knowledge Discovery in Databases (KDD) process [19]. This procedure was applied only when a comment was too lengthy, contained multiple viewpoints, or included ambiguous sentiment expressions that could complicate the subsequent classification process. The purpose of the cut-off was not to change the original opinion, but to retain the statement that most clearly represented the commenter's overall stance.

An example of this procedure is presented in Table 1.

Table 1. Example of Comment Cut-off

Original Comment	Simplified Comment
Dengan adanya MBG: (1) Anak anak yg kurang mampu bisa merasakan makanan gratis dan orang tua tidak perlu keluar uang ekstra buat anaknya makan disekolah, (2) Bahan makanan yang diperlukan dibeli ke pasar pasar dan membuat ekonomi jadi berputar. UMKM jadi lebih makmur karena ada nya MBG. (3) Membuka lapangan pekerjaan dan sudah 300rb lebih lapangan pekerjaan karena adanya MBG. Dan targetnya sampai 2jt lapangan pekerjaan dari program MBG. Masih bingung dimananya yg bilang MBG enggk tepat dan buang huanng anggaran MBG ini bagus kok, tapi saya kecewa	Dengan adanya MBG, anak- yang kurang mampu dapat menikmati makanan gratis dan perekonomian masyarakat juga ikut berkembang.
pemerintah lambat	Saya kecewa pemerintah lambat

2.2 Manual Sentiment Labeling

After the data selection stage, each comment was manually assigned to one of three sentiment categories: positive, negative, or neutral. The labeling process considered the context, opinion orientation, and emotional expression contained in each comment. Positive comments expressed support or approval of the Free Nutritious Meals (MBG) Program, whereas negative comments reflected criticism or disagreement. Neutral comments contained factual statements or information without expressing clear positive or negative opinions.

To improve labeling reliability and reduce subjectivity, the manual sentiment labeling process involved two independent validators consisting of an Indonesian language lecturer and an Indonesian language teacher. Both validators independently reviewed each comment according to the predefined sentiment labeling criteria. Inter-annotator agreement was measured using Cohen's Kappa, a statistic used to quantify the level of agreement between two independent annotators while accounting for agreement that may occur by chance [20]. The agreement was calculated from the independent labels assigned by the two validators before the consensus discussion. The two validators achieved a Cohen's Kappa value of 0.859 across 150 labeled comments, indicating a high level of agreement. Disagreements were subsequently discussed and resolved through consensus to establish the final sentiment labels. Any disagreements were discussed until a consensus was reached before the final dataset was established.

Manual labeling was performed before the machine learning process to provide reliable class labels for supervised classification. This procedure is commonly applied to improve the quality of sentiment analysis datasets before model training [21]. Manual labeling was selected because it provides higher precision and labeling accuracy than fully automated annotation methods [22]. Furthermore, manual annotation plays an important role in developing reliable classification models [23]. The expert validation process ensured that the sentiment labels were based on linguistic interpretation rather than the subjective judgment of a single researcher, thereby improving the validity and credibility of the dataset. This approach helps reduce labeling errors caused by ambiguous expressions, slang, or informal language that frequently appear in TikTok comments. An example of the manual sentiment labeling process is presented in Table 2.

Table 2 Example of Manual Sentiment Labeling.

No.	Comments	Sentiment
1	real ny di lapangan byk yg mubazir bocah2 gk mau makan	Negatif
2	mbg padahal sangat berguna untuk anak 2 knp pada membenci program ini?	Positif
3	program yang bikin negri bangkrut	Negatif
4	skrg aku bisa krj jd SPPG.. mashaAllah trimakasih BP. Prabowo	Positif
5	tolong ya kalo mau kasih mbg harus di cek dulu sebelum di kasih	Negatif
6	Terima kasih pak Prabowo, kami UMKM sdh merasakannya karena dapat menciptakan lapangan kerja	Positif
7	aku udh dapat	Netral
8	lebih baik di berikan langsung uang tunai ke orangtua, jadi membangkitkan daya beli ortu.. daripada mbg di korup oknum yg mengurus jalannya mbg	Negatif
9	muadzir pak! sayang ga di makan, anak berbeda2 selera mkn nya.apalagi anak TK,SD mknnya susah.ksi uangnya aja biar di masak sama ibu nya.	Negatif
10	banyak yang keracunan nih wo	Negatif

The labeled dataset was subsequently used as the input for the preprocessing stage, feature extraction using TF-IDF, and sentiment classification with the Naïve Bayes algorithm.

2.3 Text Preprocessing and Classification

Text preprocessing is an essential stage in text mining that aims to improve text quality by removing noise and transforming raw text into structured data suitable for machine learning algorithms [19]. Previous studies have also shown that preprocessing techniques, including the Sastrawi stemming algorithm, effectively normalize Indonesian text and improve feature representation for sentiment classification [24]. After manual sentiment labeling, the dataset was processed automatically using Python integrated into RapidMiner. The preprocessing stage was performed to improve text quality before feature extraction and sentiment classification. This stage consisted of several sequential processes, including cleaning, case folding, tokenization, stopword removal, and stemming. After tokenization, normalization was performed to convert informal Indonesian words, abbreviations, and slang into their standard forms using a custom normalization dictionary compiled for this study. The dictionary was implemented as a lookup table in the Python preprocessing script. For example, informal forms such as “gk”, “yg”, and “udh” were converted into “tidak”, “yang”, and “sudah”, respectively. This normalization step was applied before stopword removal to improve consistency in the representation of Indonesian social media text.

These processes aimed to remove noise, standardize text format, and convert the comments into a consistent representation suitable for text mining. After preprocessing, the cleaned text was transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting method. The resulting feature vectors were then classified into positive, negative, and neutral sentiment categories using the Naïve Bayes Classifier algorithm. In this study, Naïve Bayes was used as the sole classifier to maintain a focused evaluation of the preprocessing and TF-IDF feature extraction pipeline. Although comparisons with other classifiers such as Support Vector Machine or transformer-based models were feasible, comparative evaluation was outside the scope of this study. The use of alternative classifiers is therefore left for future research. The overall preprocessing and classification workflow is illustrated in Figure 2.

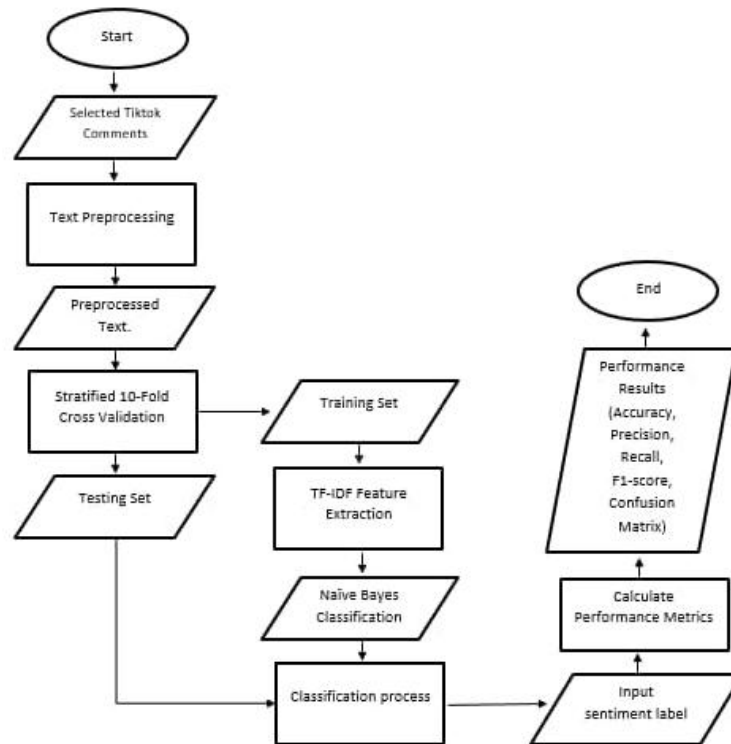


Figure 2. Text Preprocessing and Classification Workflow

- Cleaning, process removed URLs, punctuation marks, numbers, emojis, special characters, and other irrelevant symbols that did not contribute to sentiment analysis [19]. This step was performed to reduce noise and ensure that only relevant textual information was retained for subsequent processing.
- Case Folding, converted all characters into lowercase to ensure text consistency and eliminate differences between uppercase and lowercase words [19]. This process prevented identical words written in different letter cases from being treated as separate terms during feature extraction.
- Tokenization, divided each comment into individual words (tokens), enabling the text to be processed during feature extraction [19]. This step transformed each comment from a continuous text string into a sequence of individual terms.
- Normalization, converted informal Indonesian words, abbreviations, and slang into more standardized forms using a custom normalization dictionary compiled for this study. For example, “gk”, “yg”, and “udh” were converted into “tidak”, “yang”, and “sudah”, respectively. This step was applied before stopword removal to improve consistency in the representation of informal TikTok language.
- Stopword Removal, eliminated common Indonesian words that carried little semantic meaning, such as conjunctions and prepositions, while retaining important words related to sentiment [19]. his step reduced the number of less informative terms and helped focus the feature representation on more relevant words.
- Stemming, transformed inflected words into their root forms using the Indonesian stemming algorithm to reduce vocabulary variation [19]. Stemming also improve text consistency before classification. Similar findings were reported by Nafi' et al., who demonstrated that Sastrawi stemming contributed to better sentiment classification performance on Indonesian e-commerce reviews [24].

Table 3. Example of Text Preprocessing

Original Comment	Cleaning	Case Folding	Tokenization	Stopword Removal	Stemming
Pemborosan anggaran pak @PRABOWO SUBIANTO lebih baik siapkan anggaran cairkan ke orang tua anak-anak wajib bawa nasi dari rumah senilai 15k sesuai yang anak-anak mereka sukai. Jangan bapak buat proyek-proyek yang banyak dikorupsi!!! 	Pemborosan anggaran pak PRABOWO SUBIANTO lebih baik siapkan anggaran cairkan ke orang tua anak wajib bawa nasi dari rumah senilai sesuai yang anak mereka sukai. Jangan bapak buat proyek yang banyak dikorupsi	pemborosan anggaran pak prabowo subianto lebih baik siapkan anggaran cairkan ke orang tua anak wajib bawa nasi dari rumah senilai sesuai yang anak mereka sukai jangan bapak buat proyek yang banyak dikorupsi	[pemborosan, anggaran, pak, prabowo, subianto, lebih, baik, siapkan, anggaran, cairkan, orang, tua, anak, wajib, bawa, nasi, rumah, senilai, sesuai, anak, mereka, sukai, jangan, bapak, buat, proyek, banyak, dikorupsi]	[pemborosan, anggaran, prabowo, subianto, siapkan, anggaran, cairkan, orang, tua, anak, wajib, bawa, nasi, rumah, senilai, sesuai, anak, sukai, jangan, bapak, buat, proyek, korupsi]	[boros, anggar, prabowo, subianto, siap, anggar, cair, orang, tua, anak, wajib, bawa, nasi, rumah, nilai, sesuai, anak, suka, jangan, bapak, buat, proyek, korupsi]

2.4 TF-IDF Feature Weighting

After the preprocessing stage, the textual data were transformed into numerical feature vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) weighting scheme. TF-IDF is a widely used feature extraction technique in text mining because it assigns higher weights to words that frequently occur in a document but rarely appear across the entire document collection, allowing important terms to contribute more effectively to the classification process [10]. The TF-IDF weighting consists of three sequential calculations: Term Frequency (TF), Inverse Document Frequency (IDF), and the final TF-IDF weight.

The Term Frequency (TF) of a term is calculated using (1),

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Where $f_{t,d}$ denotes the frequency of term t in document d , $\sum_k f_{k,d}$ represents the total number of terms contained in document d .

The Inverse Document Frequency (IDF) is calculated using (2),

$$IDF(t) = \log \frac{N}{df_t} \quad (2)$$

Where N is the total number of documents and df_t is the number of documents containing term t .

Finally, the TF-IDF weight is obtained by multiplying TF and IDF as shown in (3),

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

The resulting TF-IDF vectors represented the numerical features of each TikTok comment and were subsequently used during the sentiment classification stage.

2.5 Naïve Bayes Classification

The sentiment classification process was carried out using the Naïve Bayes Classifier, a probabilistic machine learning algorithm that applies Bayes' theorem under the assumption that features are conditionally independent given a class label [18]. This algorithm is widely used in text classification because of its simplicity, computational efficiency, and competitive performance in sentiment analysis tasks.

For a document represented by a set of features X and a sentiment class H , the posterior probability is calculated using (4),

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (4)$$

where $P(H|X)$ is the posterior probability that document X belongs to class H , $P(H)$ is the prior probability of class H , $P(X|H)$ is the likelihood of observing features X given class H , and $P(X)$ is the overall probability of observing features X .

The prior probability of each sentiment class is computed using (5),

$$P(H) = \frac{\text{Number of documents in class } H}{\text{Total number of documents}} \quad (5)$$

To avoid zero probabilities for unseen terms, Laplace smoothing was applied when calculating the conditional probability of a term, as shown in (6),

$$P(x_i|H) = \frac{nk+1}{nH+V} \quad (6)$$

where nk is the frequency of term x_i in class H , nH is the total number of terms in class H , and V is the number of unique terms in the vocabulary.

Each TikTok comment represented by TF-IDF features was classified into one of three sentiment categories: positive, negative, or neutral. The final class label was determined by selecting the class with the highest posterior probability.

2.6 Model Evaluation

The performance of the proposed sentiment classification model was evaluated using 10-fold cross-validation, which is one of the most widely used validation techniques in machine learning because it provides a reliable estimation of model performance while reducing evaluation bias [25], [11]. In this method, the dataset was divided into ten approximately equal subsets. During each iteration, nine subsets were used as training data and the remaining subset was used for testing. This process was repeated ten times so that every subset served once as the testing set.

The prediction results obtained from each fold were summarized using a multiclass confusion matrix. A confusion matrix is a tabular evaluation method that compares the actual class labels with the predicted class labels, allowing the classification performance of each sentiment category to be analyzed [9]. Since this study classified comments into three sentiment classes (positive, negative, and neutral), a multiclass confusion matrix was employed.

The classification results from each fold were evaluated using a multiclass confusion matrix. Four performance metrics were calculated, namely accuracy, precision, recall, and F1-score, to assess the effectiveness of the Naïve Bayes classifier in predicting positive, negative, and neutral sentiments [26], [9].

The overall classification accuracy is calculated using Equation (7),

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \times 100\% \quad (7)$$

where TP denotes True Positive, TN denotes True Negative, FP denotes False Positive, and FN denotes False Negative.

Precision is calculated using Equation (8),

$$Precision = \frac{TP}{(TP+FP)} \quad (8)$$

Recall is calculated using Equation (9),

$$Recall = \frac{TP}{(TP+FN)} \quad (9)$$

The F1-score, which represents the harmonic mean of precision and recall, is calculated using Equation (10),

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

The values of accuracy, precision, recall, and F1-score obtained from each fold were then averaged to provide the overall performance of the proposed sentiment classification model.

3. Result and Discussions

3.1 Dataset Characteristics

Initially, 3,190 comments were collected from a TikTok video discussing the Free Nutritious Meals (Makan Bergizi Gratis or MBG) Program. After the data selection process, which removed duplicate comments, advertisements, emojis, unrelated discussions, meaningless text, and comments containing multiple unrelated opinions, 1,294 valid comments remained for further analysis. These comments were manually labeled into three sentiment categories, namely positive, negative, and neutral. The final dataset consisted of 730 negative comments (56.41%), 339 positive comments (26.20%), and 225 neutral comments (17.39%). This distribution indicates a class imbalance, with the neutral class representing the smallest proportion of the dataset.

The final dataset showed a moderate class imbalance, with the neutral class representing 17.39% of the data. Although stratified 10-fold cross-validation was applied to maintain the class proportions across the training and testing folds, it does not directly address class imbalance during model training. In this study, the original class distribution was retained without applying oversampling or class-weighting techniques. Therefore, the model performance, particularly for the minority neutral class, should be interpreted in light of the class imbalance. Future studies may consider applying techniques such as SMOTE or class-weighting to further evaluate their effects on minority-class classification performance.

Figure 3 presents the distribution of sentiment labels in the final dataset. The results show that negative sentiment dominates the dataset, followed by positive sentiment and neutral sentiment. This distribution indicates that public responses toward the MBG program on TikTok were generally inclined toward criticism rather than support. In addition, the unequal number of instances across sentiment classes reflects the presence of class imbalance, which may influence the classification performance, particularly for the minority class. Therefore, the use of stratified 10-fold cross-validation was considered appropriate to preserve the proportion of each sentiment class during model evaluation.

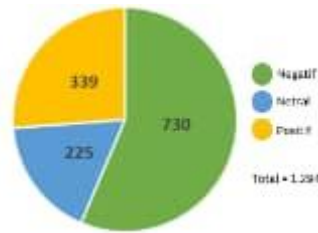


Figure 3. Distribution of Sentiment Labels in The Final Dataset

3.2 Text Preprocessing Result

After the preprocessing stage, all TikTok comments were transformed into a cleaner and more consistent textual representation suitable for sentiment classification. The preprocessing successfully removed irrelevant elements such as URLs, punctuation marks, emojis, numbers, and special characters, while normalizing informal Indonesian words and reducing inflected words to their root forms through stemming. As a result, the vocabulary became more standardized, allowing the TF-IDF weighting method to extract more representative textual features for classification.

To provide an overview of the processed textual data, a word cloud was generated from the preprocessed comments, as presented in Figure 4. The visualization highlights the most frequently occurring terms in the dataset after preprocessing. Words closely related to the MBG program, such as *mbg*, *anak*, *program*, *makan*, and *gratis*, appear prominently because they occur frequently in users' comments.



Figure 4. Word Cloud of Preprocessed TikTok Comments

The word cloud also reveals the presence of several opinion-related terms associated with both supportive and critical responses toward the MBG program. However, negative expressions appear more frequently than positive ones, which is consistent with the sentiment distribution presented in Figure 3. These findings indicate that discussions surrounding the MBG program on TikTok were largely dominated by critical opinions, although positive and neutral responses were also observed. Overall, the preprocessing stage produced a structured textual representation that effectively captured the dominant topics and sentiment-related terms before the classification process.

3.3 Model Implementation

The sentiment classification model was implemented using RapidMiner Altair AI Studio. After the manual labeling process, the TikTok comments were imported into RapidMiner and processed automatically using the Execute Python operator. The Python script performed text normalization, including lowercase conversion, removal of punctuation, numbers, emojis, and special characters, normalization of informal Indonesian words, stopword removal, and stemming using the Sastrawi library.

The cleaned comments were then converted into text attributes using the Nominal to Text operator so that they could be processed by the text mining module. Feature extraction was performed through the Process Documents from Data operator, which tokenized the comments and transformed them into numerical representations using TF-IDF weighting.

Sentiment classification was carried out using the Naïve Bayes Classifier combined with 10-fold cross-validation. The cross-validation process divided the dataset into ten subsets while preserving the proportion of sentiment classes through stratified sampling. In each iteration, nine subsets were used for training and one subset was used for testing, allowing every comment to be evaluated once as testing data.

Figure 5 presents the overall workflow of the implemented sentiment classification model.

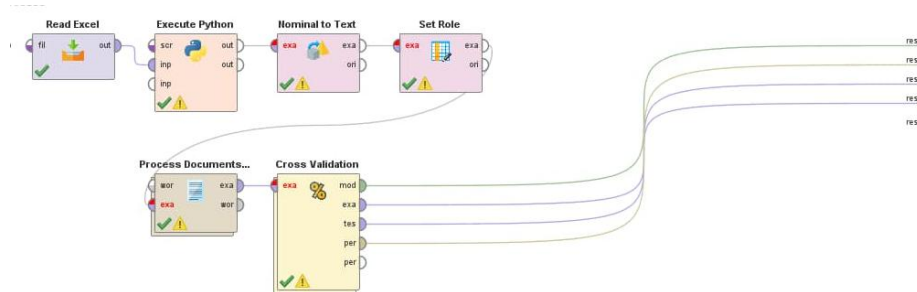


Figure 5. Workflow of The Sentiment Classification Model in RapidMiner.

As shown in Figure 5, the workflow consists of data import, automated preprocessing through Python, feature extraction using TF-IDF, and sentiment classification using the Naïve Bayes algorithm with stratified 10-fold cross-validation. This integrated workflow ensures that all preprocessing, feature extraction, model training, and evaluation stages are performed consistently within a single environment.

3.4 Classification Performance

The performance of the proposed sentiment classification model was evaluated using the Performance (Classification) operator in RapidMiner Altair AI Studio after applying the Naïve Bayes classifier with stratified 10-fold cross-validation. RapidMiner automatically generated a multiclass confusion matrix and calculated the overall classification accuracy. In addition, weighted precision and weighted recall were reported for the multiclass classification task, while the weighted F1-score was calculated manually based on the weighted precision and recall values.

Figure 6 presents the confusion matrix obtained from RapidMiner Altair AI Studio. The classifier correctly predicted 257 positive comments, 163 neutral comments, and 522 negative comments, while the remaining comments were incorrectly assigned to other sentiment categories. The larger number of correctly classified negative comments is partly influenced by the higher proportion of negative instances in the dataset, as presented in Figure 3. In addition, several misclassifications occurred between the positive and neutral classes, indicating that informal TikTok comments often contain similar lexical characteristics that make sentiment discrimination more challenging.

PerformanceVector (Performance)				
accuracy: 72.00% +/- 4.54% (micro average: 72.00%)				
	true negativ	true neutr	true positf	class precision
pred. negativ	522	41	70	82.45%
pred. neutr	89	163	12	81.74%
pred. positf	119	21	257	54.74%
class recall	71.51%	72.44%	75.01%	

Figure 6. Confusion Matrix Generated by RapidMiner

The overall classification accuracy achieved by the proposed model was 72.80%, indicating that the classifier correctly predicted nearly three-quarters of the TikTok comments. Most classification errors occurred between the positive and neutral classes, suggesting that these two sentiment categories share similar linguistic characteristics in informal social media comments. In contrast, negative comments were identified more accurately because they generally contained more distinctive sentiment expressions.

Table 4. Examples of Positive-Neutral Misclassification.

No.	Comment	Actual Label	Predicted Label
1	"izin tanya pak angaran mbg serap dr angran rapbn arti hasil pajak rakyat juga."	Neutral	Positive
2	"sumbermalang kabupaten situbondo belum nyampe."	Neutral	Positive
3	"terima kasih pak kami sudah dapat"	Positive	Neutral

These misclassified comments share several characteristics that make positive and neutral sentiment difficult to distinguish [27]. The comments are relatively short, use informal Indonesian expressions, and provide limited contextual information. In particular, expressions such as "terima kasih" may indicate positive sentiment, while comments describing a situation or asking for information may contain words that are also associated with supportive expressions. Such contextual ambiguity can lead the classifier to assign a positive label to comments that are neutral in meaning, or a neutral label to short positive expressions.

To facilitate visual interpretation, the confusion matrix was further represented as a heatmap, as illustrated in Figure 7. Darker colors indicate higher prediction frequencies. The heatmap clearly shows that the largest concentration of correct predictions belongs to the negative sentiment class, whereas the positive and neutral classes exhibit more frequent misclassifications. This visualization provides a clearer overview of the classifier's performance across all sentiment categories.

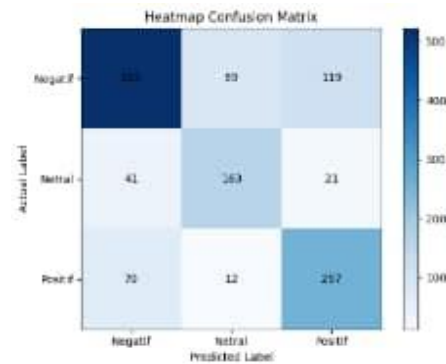


Figure 7. Heatmap Visualization of The Confusion Matrix

To evaluate the consistency of the proposed model, the performance obtained from each fold of the stratified 10-fold cross-validation is summarized in Table 5. Since RapidMiner reports precision and recall as weighted averages for multiclass classification, the weighted F1-score was calculated manually using the weighted precision and weighted recall values.

Table 5. Results of Stratified 10-fold Cross Validation

Fold	Accuracy	Recall	Precision	F1-score
1	0,730769	0,748648	0,721281	0,73471
2	0,713178	0,73389	0,681664	0,706814
3	0,697674	0,697084	0,666277	0,681332
4	0,643411	0,644124	0,609788	0,626486
5	0,767442	0,799941	0,752377	0,77543
6	0,751938	0,737882	0,715251	0,72639
7	0,684615	0,680447	0,650705	0,665244
8	0,784615	0,757048	0,735296	0,746014
9	0,723077	0,732444	0,693607	0,712497
10	0,782946	0,793141	0,755051	0,773627
Mean	0,727967	0,732465	0,69813	0,714854

The results demonstrate that the proposed model achieved relatively stable performance across the ten validation folds. The classification accuracy ranged from 64.34% to 78.46%, with the highest accuracy obtained in Fold 8 and the lowest in Fold 4. The relatively lower performance observed in Fold 4 may be influenced by the random partitioning of the dataset determined by the local random seed. Although stratified 10-fold cross-validation preserves the overall class proportions across folds, the local random seed determines how individual comments are assigned to each training and testing subset, which may lead to variation in fold-level performance. The highest recall (79.99%) and F1-score (77.54%) were achieved in Fold 5, whereas the highest precision (75.51%) was obtained in Fold 10. The relatively small variation among folds indicates that the Naïve Bayes classifier provides consistent performance despite differences in the training and testing partitions, demonstrating good generalization capability for sentiment classification of TikTok comments.

3.5 Discussion

The findings indicate that negative sentiment dominates public responses toward the Free Nutritious Meals (MBG) Program on TikTok. This pattern suggests that discussions on social media tend to focus more on criticism than appreciation of the program. Most negative comments were related to issues such as food quality, implementation, distribution, and reports of food poisoning, whereas positive comments mainly emphasized the perceived benefits of the program for children's nutrition and employment opportunities. Meanwhile, neutral comments generally consisted of questions or factual statements without expressing explicit opinions. These findings demonstrate that public opinion regarding the MBG program on TikTok remains predominantly critical.

From a policy perspective, the negative sentiment identified in this study highlights several aspects of MBG implementation that may require greater attention from program administrators. Negative comments related to food quality suggest the need to strengthen food quality and safety controls during preparation and before meals are distributed. Concerns regarding distribution indicate the importance of improving delivery coordination and ensuring that meals reach intended beneficiaries in a timely and equitable manner. In addition, reports of food poisoning appearing in public comments suggest the need for stricter monitoring of food handling, hygiene, and distribution procedures. These measures may help address public concerns while improving the implementation and communication of the MBG Program.

The proposed Naïve Bayes classifier achieved satisfactory performance in classifying Indonesian TikTok comments. This result is supported by the preprocessing stage, which normalized informal language through text cleaning, normalization, stopword removal, and stemming before TF-IDF feature extraction. Nevertheless, several comments were still misclassified, particularly those containing ambiguous expressions, informal

language, abbreviations, or sarcasm. These linguistic characteristics are common in social media and remain challenging for probabilistic classifiers such as Naïve Bayes.

Another strength of this study lies in the manual sentiment labeling process, which was validated by two Indonesian language experts consisting of an Indonesian language lecturer and an Indonesian language teacher. Their involvement helped ensure that the assigned sentiment labels accurately reflected the contextual meaning of informal TikTok comments, including slang expressions and ambiguous language commonly found on TikTok. This expert validation reduced labeling subjectivity and improved the reliability of the training dataset, which contributed to the overall performance of the Naïve Bayes classifier.

The classification performance obtained in this study is comparable to previous studies. Amelia Br. Siregar et al. (2025) reported an accuracy of 65% using Naïve Bayes with TF-IDF for sentiment analysis of the Free Lunch Program on X (Twitter), where negative sentiment also dominated the discussion. Likewise, Cam et al. (2024) reported that Naïve Bayes achieved an accuracy of 68.4%, although it was outperformed by Support Vector Machine and Multilayer Perceptron. Qi and Shabrina (2023) also found that Support Vector Classifier produced higher accuracy than Multinomial Naïve Bayes in Twitter sentiment classification. In this study, the Naïve Bayes classifier achieved an accuracy of 72.80%. However, differences in accuracy across studies may be influenced by differences in the characteristics of the datasets and experimental settings, including the social media platform, dataset size, class distribution, sentiment categories, language characteristics, and preprocessing techniques. Therefore, the performance results should be considered within the context of the dataset and methodology used in each study.

This study has several limitations. The dataset was collected from comments on a single TikTok video discussing the MBG program, which may not fully represent public opinion across different social media platforms. In addition, manual sentiment labeling may still introduce subjective interpretation, particularly for sarcastic or ambiguous comments. Furthermore, this study evaluated only the Naïve Bayes classifier without comparing it with other machine learning algorithms. Therefore, future research is recommended to employ larger and more diverse datasets, investigate alternative classification algorithms, and explore advanced deep learning approaches to further improve sentiment classification performance.

4. Conclusions and Future Works

This study presented a sentiment analysis of TikTok comments regarding the Free Nutritious Meals (MBG) Program using the Naïve Bayes classifier with TF-IDF feature weighting and manually labeled data validated by two Indonesian language experts. From the initial collection of 3,190 comments, a total of 1,294 valid comments were obtained after the data selection and preprocessing stages. The experimental results showed that the proposed model achieved an accuracy of 72.80%, with a weighted precision of 69.81%, a weighted recall of 73.25%, and an F1-score of 71.49%. The findings also revealed that negative sentiment dominated public responses toward the MBG program, indicating that discussions on TikTok were largely characterized by criticism rather than supportive opinions. These findings demonstrate that combining expert-validated manual sentiment labeling, systematic text preprocessing, TF-IDF feature extraction, and the Naïve Bayes classifier provides a feasible approach for sentiment classification of Indonesian social media text.

Despite these results, this study has several limitations. The dataset was collected from comments on a single TikTok video, which may not fully represent public opinion across TikTok or other social media platforms. In addition, only the Naïve Bayes classifier was evaluated without comparison to other machine learning or deep learning approaches. Although inter-annotator agreement was assessed using Cohen's Kappa before the consensus discussion, the labeling process involved only two validators. Future research is therefore recommended to utilize larger and more diverse datasets, incorporate data from multiple social media platforms, compare Naïve Bayes with alternative classifiers such as Support Vector Machine, Logistic Regression, Random Forest, and transformer-based models, and involve additional annotators to further strengthen the reliability and generalizability of the sentiment classification results.

5. Declarations

Author Contributions

Sukma Ajeng Griyandari, as the first author, was primarily responsible for the conceptualization of the study, methodology, data collection, data preprocessing, data analysis, investigation, writing of the original draft, and manuscript preparation. Fahrur Rozi, as the second author, contributed through supervision, academic guidance, methodological guidance, critical review of the research, and review and editing of the manuscript. Yayak Kartika Sari, as the third author, contributed through academic review, critical evaluation of the research, feedback on the relevance and potential contribution of the study, and review and editing of the manuscript.

Funding Statement

This research received no external funding and was fully self-funded by the first author.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Data Availability Statement

The data analyzed in this study were collected from publicly available comments on a TikTok video posted by the @bbcnewsindonesia account. The comments were collected using the Apify TikTok Scraper. The dataset was derived from user-generated comments on a single publicly accessible video and is not publicly shared in its original form due to privacy and platform related considerations.

Use of Artificial Intelligence (AI)

AI tools were used to assist with language translation and language refinement during the preparation of this manuscript. AI tools were not used to generate research data, perform data analysis or sentiment classification, or make scholarly judgments. The authors remain fully responsible for the content, analysis, interpretation, and conclusions of the manuscript.

Ethical Clearance

This study analyzed publicly available comments from a TikTok video and did not involve direct interaction with or intervention on human participants. Therefore, formal ethical clearance was not required. The study was conducted in accordance with applicable ethical and privacy principles.

6. References

- [1] APJII, *SURVEI PENETRASI INTERNET DAN PERILAKU PENGGUNAAN INTERNET Sebagai perwakilan Pengurus APJII, kami dengan bangga mempersembahkan Profil Internet Indonesia 2025*. 2025.
- [2] Y. P. Galuh Putri Riyanto, "Indonesia Pengguna TikTok Terbesar di Dunia, Tembus 157 Juta Kalahkan AS," *kompas.com*. Accessed: Oct. 28, 2025. [Online]. Available: https://tekno.kompas.com/read/2024/10/25/15020057/indonesia-pengguna-tiktok-terbesar-di-dunia-tembus-157-juta-kalahkan-as?utm_source=chatgpt.com#goog_rewarded
- [3] N. N. Peni Cahyati, "A COMPREHENSIVE STUDY ON MBG (MAKAN BERGIZI GRATIS) IN THE PRABOWO-GIBRAN CABINET : EVALUATING THE PSYCHOLOGICAL AND HEALTH IMPACTS OF THE POLICY ON," *Jurnal Ilmu Psikologi dan Kesehatan (SIKONTAN)*, vol. 3, no. 4, pp. 177–186, 2025, doi: 10.47353/sikontan.v3i4.2783.
- [4] H. F. Maulana and F. Arroyan, "Narrative Fractures in Policy Communication : Framing and Public Sentiment on Indonesia ' s Free Nutritious Meals Program," in *Proceedings of Jogjakarta Communication Conference (JCC)*, 2025, pp. 309–321.
- [5] Y. Sergio, V. Putranta, B. Rahayudi, and W. Purnomo, "Analisis Sentimen Masyarakat terhadap Kebijakan Penghapusan Subsidi BBM pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier

- dengan Ekstraksi Fitur N-Gram TF-IDF,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 1501–1510, 2023.
- [6] M. R. Ghufroon, M. F. M. Arsyada, M. R. Lukman, Y. A. H. Putra, and N. A. Rakhmawati, “Analisis Sentimen Pengguna Twitter Terhadap Pemilu 2024 Berbasis Model XLM-T,” *J-INTECH (Journal of Information and Technology)*, vol. 11, no. 2, pp. 307–315, 2023, doi: 10.32664/j-intech.v11i2.1013.
 - [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.
 - [8] R. A. Nugroho, I. Cholissodin, and Indriati, “Implementasi Naïve Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 1, pp. 301–310, 2021.
 - [9] A. Akbar, M. Putra, and M. L. Jundillah, “Analisis Sentimen Pengguna Youtube Terhadap Uang Baru Tahun Emisi 2022 Menggunakan Metode Naïve Bayes Classifier,” *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 3, no. 1, pp. 17–27, 2024, doi: 10.30872/atasi.v3i1.1177.
 - [10] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes Pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH TF-IDF Weighting Using Naïve Bayes on Public Sentiment on The Issue of Rising BIPIH,” *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
 - [11] S. Khoerunnisa, D. F. Shiddiq, and D. Nurhayati, “Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 566–577, 2025, doi: 10.57152/malcom.v5i2.1852.
 - [12] R. Amelia Br Siregar, A. Ramadhan Manik, A. Syahri, M. Budi Akbar, A. Arnita, and F. Ramadhani, “Penerapan Naïve Bayes Untuk Analisis Sentimen Netizen Terhadap Program Makan Siang Gratis Pada Media Social X,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 5621–5628, 2025, doi: 10.36040/jati.v9i4.13870.
 - [13] B. Rahmatullah, S. A. Saputra, P. Budiono, and D. P. Wigandi, “Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes,” *Journal of Information Technology*, vol. 5, no. 1, pp. 259–264, 2025, doi: 10.46229/jifotech.v5i1.978.
 - [14] A. Putriyekti, D. Sulistiawati, N. Khairani, R. Rizky, and A. Kusuma, “Analisis Multidimensional Sentimen Masyarakat terhadap Program Makan Bergizi Gratis pada Media Sosial X,” *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 01 Februari, pp. 1004–1024, 2025.
 - [15] M. A. Dzakwan, M. Amani, M. Putra, and N. Rakhmawati, “Analisis Sentimen Komentar Pada Video Kampanye Tiktok Terhadap Preferensi Publik Dalam Pilkada Jakarta 2024 Dengan Metode Naïve Bayes Classifier,” *Etika Teknologi Informasi*, vol. 1, no. 1, 2024.
 - [16] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner Analysis Data Mining Netflix Data Using The Rapid Miner Application,” *Journal of Business and Audit Information Systems*, vol. 4, no. 1, pp. 13–21, 2021.
 - [17] A. Srirahayu and L. S. Pribadie, “Review Paper Data Mining Klasifikasi Data Mining,” *Jurnal Ilmiah Informatika Global*, vol. 14, no. 1, 2023, doi: 10.36982/jiig.v14i1.2981.
 - [18] A. K. Ginting, D. Purba, L. M. Sinaga, and M. Sagala, “Analisis Sentimen Masyarakat Twitter terhadap Emas Digital Menggunakan Algoritma Naïve Bayes,” *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, vol. 07, no. 01, pp. 8–14, 2025, doi: 10.54367/kakifikom.v7i1.4909.

- [19] E. T. Naldy and A. Andri, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," *Jurnal Nasional Ilmu Komputer*, vol. 2, no. 2, pp. 89–101, 2021, doi: 10.47747/jurnalnik.v2i2.525.
- [20] I. T. Julianto, M. Affan, and A. Sidqi, "Automatic Sentiment Annotation Using Grok AI for Opinion Mining in a University Learning Management System," *Journal of Intelligent Systems Technology and Informatics*, vol. 1, no. 3, pp. 72–77, 2025, doi: 10.64878/jistics.v1i3.42.
- [21] I. F. Rahman, A. N. Hasanah, N. Heryana, U. S. Karawang, T. Timur, and J. Barat, "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL (SIGNAL) DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 963–969, 2024, doi: 10.23960/jitet.v12i2.4073.
- [22] A. H. Pratama and M. Hayaty, "Performance of Lexical Resource and Manual Labeling on Long Short-Term Memory Model for Text Classification," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 1, pp. 74–84, 2023, doi: 10.26555/jiteki.v9i1.25375.
- [23] N. Hidayah and Dodiman, "Implementasi Algoritma Multinomial Naïve Bayes , TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," *Jurnal Akademik Pendidikan Matematika*, vol. 10, no. 1, pp. 8–15, 2024, doi: 10.55340/japm.v10i1.1491.
- [24] A. Nafi, A. Tri, J. Harjanta, B. A. Herlambang, and S. Fahmi, "Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna," *J-INTECH (Journal of Information and Technology)*, vol. 12, no. 2, pp. 330–339, 2024, doi: 10.32664/j-intech.v12i02.1450.
- [25] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [26] F. Zamzami, R. Hidayat, and R. Fathonah, "Penerapan Algoritma Naive Bayes Classifier untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan IKN," *Faktor Exacta*, vol. 17, no. 1, pp. 47–57, 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [27] R. Wati, I. P. Astuti, and G. A. Buntoro, "Analisis Sentimen terkait Penerapan E-Parking Ponorogo (Parkir- Go) dengan Metode Bidirectional Encoder from Transformers (BERT) Sentiment Analysis of Social Media Comments Related to the Implementation of Ponorogo E-Parking (Parkir-Go) Using the Bi," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 13, no. 2, pp. 220–227, 2025, doi: 10.26418/justin.v13i2.86993.