Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Sentiment Analysis of Free Nutritious Meal Programs Using Naïve Bayes on Platforms X and TikTok

Fadila Ullul Azmie[1*✉], Yudie Irawan[1], R.Rhoedy Setiawan[1]
[1]Faculty of Engineering, Universitas Muria Kudus, Kudus, Indonesia
**Corresponding Author: 202253161@std.umk.ac,id*

Article Information

Article history:

No. 1112
Rec. February 04, 2026
Rev. February 25, 2026
Acc. February 28, 2026
Pub. March 07, 2026
Page. 1208 – 1227

Keywords:

- *Sentimen Analysis*
- *Multinominal Naive Bayes*
- *MBG*
- *Media Social*
- *SMOTE*

ABSTRACT

This study analyzes public sentiment toward the Free Nutritious Meal Program (MBG) using the Naïve Bayes algorithm on data from X (Twitter) and TikTok. A total of 5,173 entries were collected via web scraping and processed through cleaning, normalization, tokenization, stopword removal, and stemming. To address class imbalance, SMOTE was applied, while evaluation employed accuracy, precision, recall, F1-score, and AUC-ROC metrics. Results show that without SMOTE, the model was biased toward the majority class, particularly on TikTok, achieving high precision but very low recall. With SMOTE, recall improved substantially and a better balance between precision and recall was achieved. On Twitter, the more moderate class distribution yielded stable performance, though SMOTE further enhanced sensitivity to positive sentiment. Word cloud analysis highlights differences in sentiment patterns: TikTok leaned strongly toward negative sentiment with dominant words such as "racun" (poison), "korupsi" (corruption), and "dapur" (kitchen), while Twitter displayed a more balanced discourse with positive terms like "gizi" (nutrition), "gratis" (free), and "program." These differences suggest TikTok tends to be a space of criticism, whereas Twitter more often reflects support for the program's benefits. Such findings underscore the importance of cross-platform analysis to comprehensively capture public perceptions.

How to Cite:

Azmie, F. U., Irawan, Y., & Setiawan, R. (2026). Sentiment Analysis of Free Nutritious Meal Programs Using Naïve Bayes on Platforms X and TikTok. Jurnal Teknologi Informasi dan Pendidikan, 19(1) 1208-1227. https://doi.org/10.24036/jtip.v19i1.1112

1. INTRODUCTION

Social media has become a primary platform for the public to express opinions and discuss various public issues, including government policies. With 212.9 million internet users in Indonesia approximately 77% of the total population platforms such as X (formerly Twitter) and TikTok serve as critical barometers for measuring public responses and sentiment toward programs or policies [1]. The Free Nutritious Meal Program, a government initiative aimed at improving public health and nutrition, particularly among children and pregnant women, has garnered widespread public attention [2]. The vast volume of textual data generated from public discussions across these social media platforms necessitates systematic analysis to comprehensively and objectively understand public perception [3].

Sentiment analysis is a technique within text mining that aims to identify and extract opinions from text and classify them into positive or negative categories [4]. Multinomial Naive Bayes is a probability-based classification algorithm that works by calculating the occurrence probability of each word in a document based on Bayes' theorem, assuming feature independence [5], [6]. This algorithm has been widely applied in various sentiment analysis studies and has proven to yield satisfactory results, particularly on social media data characterized by informal and diverse linguistic features [7]. Previous research on sentiment analysis of the Free Nutritious Meal Program employed the Random Forest method on platform X and demonstrated reasonably good performance in classifying public sentiment [2]. However, comparative studies of SVM and Naive Bayes algorithms in application review sentiment classification indicate that both algorithms exhibit different characteristics and performance depending on the dataset used [9]; therefore, further exploration of Multinomial Naive Bayes performance in the context of government policy programs with different dataset characteristics remains necessary.

The application of sentiment analysis to Indonesian social media data faces several significant challenges. First, the use of informal language, abbreviations, slang, and emoticons requires comprehensive preprocessing stages [10]. Second, imbalanced sentiment class distribution can affect classification algorithm performance, where models tend to be biased toward the majority class and produce low recall for the minority class [11]. Third, differing discussion characteristics across social media platforms (TikTok vs Twitter) generate distinct word patterns and sentiment distributions, which may affect model generalization capability. These challenges have not been comprehensively explored in previous research, particularly in the context of evaluating government policy programs in Indonesia.

To address these challenges, this study proposes a sentiment analysis approach using the Multinomial Naive Bayes algorithm with evaluation across three different dataset scenarios: the TikTok dataset, the Twitter dataset, and a combined dataset from both platforms. Each scenario is tested under two conditions: without balancing techniques and with the application of SMOTE to address class imbalance issues. Dataset characteristic analysis is conducted using word cloud visualization to identify frequently occurring words on each platform, such as "mbg", "program", "makan", "gizi", "gratis", "anak", and "enak", which reflect the context of public discourse on the Free Nutritious Meal Program. Comprehensive model performance evaluation is conducted using accuracy, precision, recall, F1-score, and AUC-ROC metrics to measure the model's capability in handling imbalanced data. The research implementation follows the CRISP-DM methodology to ensure a systematic and structured analysis process [8].

The main contributions of this study are: (1) a comparative analysis of Multinomial Naive Bayes performance on datasets with different imbalance levels and characteristics from two popular social media platforms in Indonesia, (2) an in-depth evaluation of SMOTE's impact on the precision-recall trade-off in the context of government policy program sentiment classification, and (3) identification of keyword patterns using word clouds that reflect the focus of public discourse on each platform. Practically, the results of this study can be utilized by the government to understand cross-platform public feedback, identify program improvement areas, and formulate more effective communication strategies. Academically, this study contributes to the development of Indonesian-language sentiment analysis literature, particularly in the context of public policy with imbalanced and multi-platform data. Based on this background, this study aims to analyze public sentiment toward the Free Nutritious Meal Program using the Multinomial Naive Bayes algorithm with performance comparisons across various dataset conditions to generate optimal implementation recommendations.

2. RESEARCH METHOD

The research was conducted through several stages as illustrated in Figure 1. Each step in the process was systematically designed to ensure a structured research flow aligned with the established objectives.

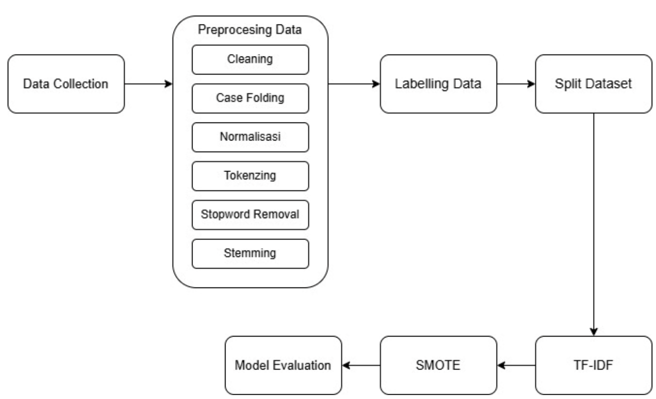


Figure 1. Research Flow

2.1. Data Collecting

Data Data collection was conducted separately from X (Twitter) and TikTok platforms using web scraping methods. The selection of both platforms was based on differences in user characteristics and communication formats [1]. For platform X, data were collected using the tweet-harvest tool on Google Colab with Auth Token, based on the keyword "MBG" from August 2024 to December 2025, yielding 1,861 tweets [12]. TikTok data were obtained using the Apify API and TikTok Scraper based on the same keyword "MBG" collected within the same period of August 2024 to December 2025, yielding 3,312 comments [13]. Both keywords were selected because "MBG" is the official acronym and the most widely used search term associated with the Free Nutritious Meal Program across both platforms. The pandas library was used to manage and store data in CSV format. The total collected data amounted to 5,173 unique Indonesian-language contents [2].

2.2. Preprocesing Data

The preprocessing stage was conducted to clean and standardize text data [10]. The process consists of six sequential sub-stages: cleaning to remove URLs, mentions, hashtags, emojis, and punctuation; case folding to convert text to lowercase [14]; normalization to transform slang into standard forms according to KBBI [10]; tokenizing using NLTK to break text into individual tokens; stopword removal to eliminate insignificant words [15]; and stemming using Sastrawi to convert affixed words into base words [16].

Table 1. Preprocessing Stages and Examples

Stage	Before	After
Cleaning	Penyebab Siswa Sukoharjo Keracunan MBG Diduga Akibat Ayam Tak Matang https://t.co/9wfua7U7XZ	Penyebab Siswa Sukoharjo Keracunan MBG Diduga Akibat Ayam Tak Matang
Case Folding	Penyebab Siswa Sukoharjo Keracunan MBG Diduga Akibat Ayam Tak Matang	penyebab siswa sukoharjo keracunan mbg diduga akibat ayam tak matang
Normalization	penyebab siswa sukoharjo keracunan mbg diduga akibat ayam tak matang	penyebab siswa sukoharjo keracunan mbg diduga akibat ayam tidak matang
Tokenzing	penyebab siswa sukoharjo keracunan mbg diduga akibat ayam tidak matang	['penyebab', 'siswa', 'sukoharjo', 'keracunan', 'mbg', 'diduga', 'akibat', 'ayam', 'tidak', 'matang']
Stopword Removal	['penyebab', 'siswa', 'sukoharjo', 'keracunan', 'mbg', 'diduga', 'akibat', 'ayam', 'tidak', 'matang']	['penyebab', 'siswa', 'sukoharjo', 'keracunan', 'mbg', 'diduga', 'akibat', 'ayam', 'matang']
Stemming	['penyebab', 'siswa', 'sukoharjo', 'keracunan', 'mbg', 'diduga', 'akibat', 'ayam', 'matang']	sebab siswa sukoharjo racun mbg duga akibat ayam tidak matang

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

2.3. Labelling Data

The preprocessed data then entered the sentiment labeling phase using the Indonesian sentiment lexicon-based approach (InSet Lexicon) to automatically assign initial labels [17]. Each word in the text was assigned a sentiment score based on the lexicon, and these scores were summed to obtain the total document score. Documents with a total score ≥ 0 were labeled as positive sentiment, while scores < 0 were labeled as negative sentiment [18].

2.4. Split Data

The labeled dataset was divided into three subsets: training set, validation set, and testing set using the stratified splitting method with an 80:10:10 ratio [20]. Data splitting was conducted in two stages to ensure proper stratification across each subset. In the first stage, the dataset was split into 80% training data and 20% temporary data. In the second stage, the 20% temporary data was divided into two equal parts: 10% for validation data and 10% for testing data [9].

2.5. TF-IDF

After data splitting, feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) to convert text into numerical representations [22]. TF-IDF calculates word weights based on term frequency in documents and inverse document frequency across the corpus using the formula $\text{TF-IDF} = \text{TF}(t,d) \times \log(N/df)$ [23]. Implementation used scikit-learn's TfidfVectorizer with parameters max_features=5000, ngram_range=(1,2), and min_df=2 [9]. The process was applied only to training data (fit_transform) then transformed to test data to prevent data leakage, producing a sparse matrix input for Multinomial Naive Bayes.

2.6. SMOTE

To address class imbalance issues that can cause model bias toward the majority class, the Synthetic Minority Over-sampling Technique was applied to the training data [21]. SMOTE works by generating synthetic minority class samples through linear interpolation between existing minority samples and their k-nearest neighbors [11]. The SMOTE algorithm randomly selects a sample from the minority class, identifies its k nearest neighbors in the feature space, and creates new synthetic samples along the line connecting the sample with its neighbors.

The implementation used the imbalanced-learn library with parameters n_neighbors=5 and random_state=42 for reproducibility [21]. SMOTE was applied only to

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

the training data after the TF-IDF process to maintain evaluation validity, while the test data retained the original distribution without SMOTE.

2.7. Evaluation Models

Evaluation employed six primary metrics [26]: Confusion Matrix for visualizing the distribution of correct and incorrect predictions [27]; Accuracy for overall correctness; Precision for the accuracy of positive predictions; Recall for minority class detection capability [28]; F1-Score as the harmonic mean of precision and recall [29]; and AUC-ROC for model discriminative ability [30]. The results from six experiments were compared to analyze the impact of dataset characteristics, the effectiveness of SMOTE, the precision-recall trade-off, and optimal configuration.

3. RESULTS AND DISCUSSION

This section presents the results of the performance evaluation of the Multinomial Naive Bayes algorithm in sentiment classification across three different scenarios. The first scenario used the TikTok dataset, the second scenario used the Twitter dataset, and the third scenario combined both datasets. In the first and second scenarios, performance comparisons were conducted with and without SMOTE to address class imbalance, while the third scenario used the combined dataset without SMOTE. The evaluation employed accuracy, precision, recall, F1-score, confusion matrix, and ROC curve metrics.

3.1. Scenario 1: Performance of Multinomial Naive Bayes on the TikTok Dataset

3.1.1. TikTok Dataset Without SMOTE

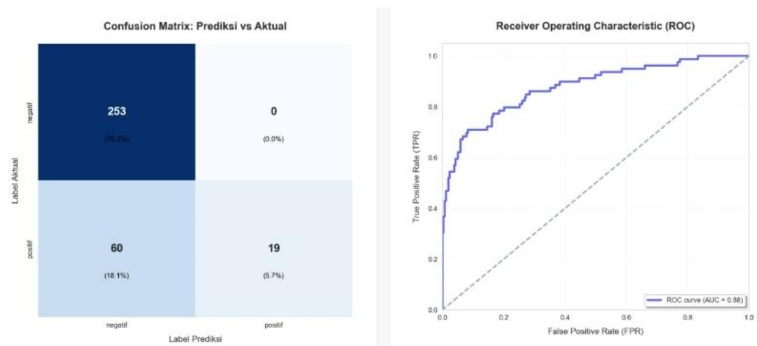
The TikTok dataset consists of 3,312 processed text entries that have undergone complete preprocessing and TF-IDF feature extraction (as shown in the dashboard overview). The data exhibits a significant class imbalance with 788 positive reviews (23.8% of total data) and 2,524 negative reviews (76.2% of total data), as illustrated in the sentiment distribution charts. This imbalanced distribution presents a notable challenge for the Multinomial Naive Bayes classifier, potentially leading to bias toward the majority class.

Figure 2. performance matrix visualization

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

Based on the Model Performance, Multinomial Naive Bayes without SMOTE achieved an accuracy of 82.01%, which is relatively high but requires further analysis using other metrics. The precision value reached 1.0000 or 100%, indicating that all positive predictions made by the model were genuinely positive with no false positive errors whatsoever. This perfect precision score indicates that the model is highly conservative in predicting the positive class.

However, the recall value was only 25.32%, very low compared to precision. This low recall indicates that Multinomial Naive Bayes only managed to detect approximately one-quarter of all actual positive reviews, missing the vast majority of other positive reviews (high false negatives). The F1-score of 40.40% reflects the extreme imbalance between precision and recall, confirming that the model exhibits significant bias. The visualization of performance metrics in Figure 6 shows a clear disparity between the very high precision and the very low recall.



(a) (b)

Figure 3. Result (a) confussion matrix visualization, (b) ROC Curve visualization

The model successfully identified 253 true negatives, demonstrating strong capability in recognizing negative sentiment. However, for the positive class, there were only 19 true positives out of the total positive reviews available. More critically, there were 60 false negatives, where positive reviews were incorrectly predicted as negative. The low number of true positives and high number of false negatives explain the very low recall value of 24.05%.

Notably, there were no false positives whatsoever (0 cases), which explains the perfect 100% precision score. This pattern indicates that when faced with imbalanced data, Multinomial Naive Bayes tends to be highly conservative and only predicts the positive class when it has extremely high certainty. The algorithm prefers to "play it safe" by predicting the majority class (negative) to avoid errors, resulting in many positive reviews going undetected.

The ROC Curve shows an AUC of 0.88, which is in the good category. The curve shows a more gradual increase compared to the ideal curve, indicating that the model

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

requires a fairly substantial threshold increase to achieve a high true positive rate, with the consequent increase in false positive rate.

Figure 4. World Cloud

The word clouds reveal linguistic patterns for each sentiment class. The positive sentiment shows keywords like "baik", "sekolah", "anak", "enak", "syukur", and "alhamdulillah", reflecting gratitude and appreciation for the Free Nutritious Meal Program. The negative sentiment displays keywords such as "racun", "korupsi", "untung", and "uang", reflecting public concerns about corruption and financial mismanagement in the program's implementation.

3.1.2. TikTok Dataset With SMOTE

The application of SMOTE to the TikTok dataset transformed the class distribution from extremely imbalanced to perfectly balanced. The training dataset, which previously had 788 positive comments (23.8%) and 2,524 negative comments (76.2%), was balanced to 2,524 positive (50%) and 2,524 negative (50%) through the synthesis of 1,736 new positive samples. The total dataset increased from 3,312 to 5,048 entries, while the testing dataset maintained its original distribution for objective evaluation. This 50:50 balance creates optimal learning conditions where the model is no longer biased toward the majority class and is forced to learn the distinguishing features between the two sentiments equally.

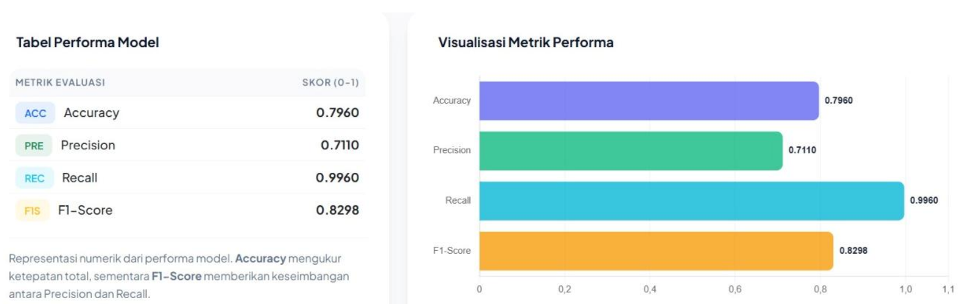
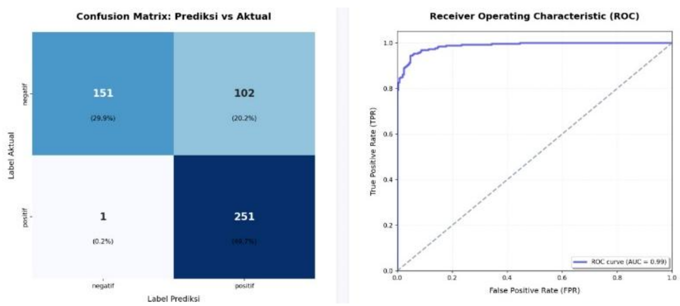


Figure 5. performance matrix visualization

The confusion matrix reveals a dramatic transformation in classification patterns. True positives skyrocketed from 40 to 499 cases (a 1,147% increase), false negatives

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

plummeted from 118 to just 6 (a 94.9% reduction). On the negative side, 191 false positives emerged (from 0 previously), explaining the drop in precision to 72.32%. This pattern indicates the model has transformed from extremely conservative to appropriately balanced, showing a preference for detecting positives despite the consequence of false positives.



(a) (b)

Figure 6. Result (a) confussion matrix visualization, (b) ROC Curve visualization

The ROC Curve shows a substantial improvement in the model's discrimination ability, with AUC increasing from 0.84 to 0.98 (a 16.7% increase). An AUC value of 0.98 places the model in the "excellent" or "near-perfect classifier" category. The curve that rapidly rises toward the top-left corner indicates the model achieves a very high true positive rate with a low false positive rate across various thresholds, proving that SMOTE has improved the model's fundamental ability to distinguish between the two classes at a probabilistic level.

Figure 7. World Cloud

The word cloud visualization reveals linguistic characteristics consistent with the original data but with a more balanced distribution. The left panel (positive sentiment) features dominant words such as "*mbg*", "*anak*" (child), "*makan*" (eat), "*enak*" (delicious), "*sehat*" (healthy), and "*baik*" (good). The right panel (negative sentiment) displays dominant words like "*racun*" (poison), "*korupsi*" (corruption), "*sekolah*" (school), "*mbg*", "*makan*" (eat), "*buat*" (for/make), and "*anak*" (child). The overwhelming dominance of the word "*racun*" (poison) is the most significant finding, indicating food safety concerns are the biggest issue. The vocabulary overlap in both sentiments proves that classification cannot be done based on specific words alone but must consider word combinations and context within sentences.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

3.2. Scenario 2: Performance of Multinomial Naive Bayes on the Twitter Dataset

3.2.1. Twitter Dataset Without SMOTE

The Twitter dataset consists of 1,861 entries with a more moderate class imbalance compared to TikTok. The distribution shows 636 positive reviews (34.2%) and 1,225 negative reviews (65.8%), presenting a ratio of approximately 2:1 between negative and positive sentiments. This relatively balanced distribution, with the minority class representing more than one-third of the data, creates more favorable learning conditions for the Multinomial Naive Bayes algorithm compared to the extreme imbalance observed in TikTok (76.2% vs 23.8%). The larger proportion of positive samples allows the model to learn positive sentiment patterns more effectively without requiring synthetic oversampling techniques.

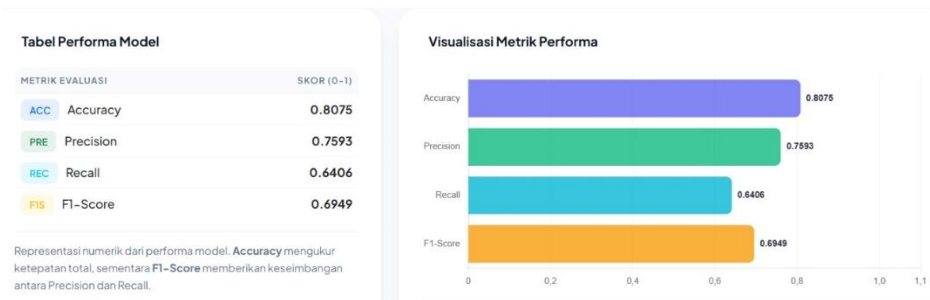
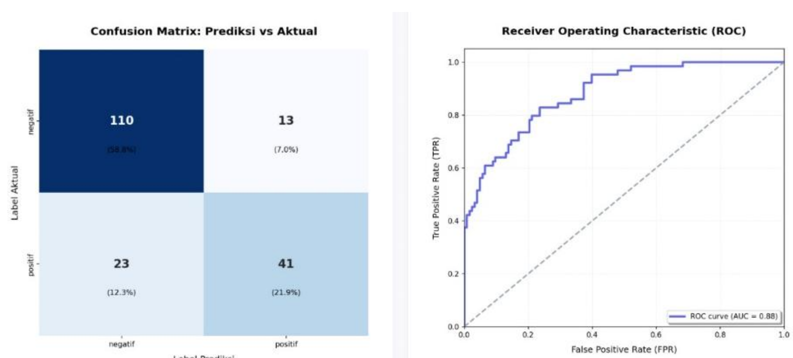


Figure 8. performance matrix visualization



(a) (b)

Figure 9. Result (a) confussion matrix visualization, (b) ROC Curve visualization

The confusion matrix reveals a more balanced classification pattern with 225 true negatives and 93 true positives, demonstrating solid capability in recognizing both sentiment classes. The model generates 34 false positives (negative reviews incorrectly predicted as positive) and 21 false negatives (positive reviews incorrectly predicted as negative). The relatively balanced ratio of false positives to false negatives (34:21 or approximately 1.6:1) indicates the model does not exhibit extreme bias toward either class,

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

contrasting sharply with TikTok's highly conservative pattern of zero false positives but 118 false negatives. This balanced error distribution confirms that the moderate class imbalance (34.2% vs 65.8%) allows Multinomial Naive Bayes to estimate posterior probabilities more proportionally for both classes.

The ROC Curve demonstrates excellent discriminative ability with an AUC of 0.90, placing the model in the "excellent" category and significantly outperforming TikTok without SMOTE (0.84). The AUC value of 0.90 indicates the model has a 90% probability of assigning a higher score to a randomly chosen positive instance compared to a randomly chosen negative instance. The curve's shape, rising steeply toward the top-left corner, confirms the model achieves high true positive rates while maintaining controlled false positive rates across various classification thresholds. This superior AUC demonstrates that the more balanced class distribution (34.2% positive) provides conducive conditions for the algorithm to learn characteristic patterns of both sentiments effectively.

Figure 10. World Cloud

Word cloud shows the difference in sentiment on Twitter. Positive sentiment uses words like "*gizi*", "*gratis*", "*makan*", "*program*", "*presiden*", and "*prabowo*", indicating public appreciation for the free meal program and political support. Negative sentiment uses words like "*racun*", "*siswa*", "*sekolah*", "*program*", and "*mbg*", reflecting public concerns about food safety and its impact on students. The presence of similar words in both groups shows that sentence context is more important than single keywords in determining sentiment.

3.2.2. Twitter Dataset With SMOTE

After applying SMOTE to the training data, the class distribution became perfectly balanced with 1,225 positive reviews (50.0%) and 1,225 negative reviews (50.0%), as shown in Figure 10. The test data retained the original distribution for objective evaluation of the model's generalization ability.

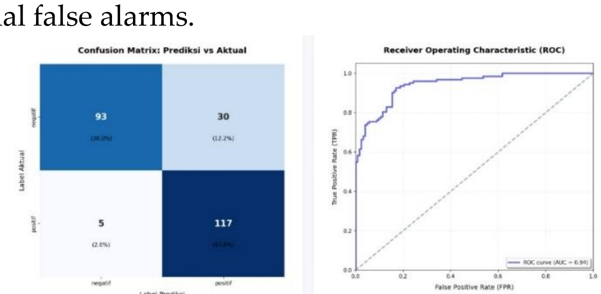
P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Jurnal Teknologi Informasi dan Pendidikan
Volume 19, No. 1, March 2026
https://doi.org/10.24036/jtip.v19i1.1112

Figure 11. performance matrix visualization

Based on the Model Performance Table in Figure 11, Multinomial Naive Bayes with SMOTE achieved an accuracy of 79.60%, which represents the overall correctness of predictions across both sentiment classes. The precision of 71.10% indicates the proportion of correctly identified positive sentiments among all instances predicted as positive, showing the model's reliability when it predicts a positive classification. The recall value of 99.60% is particularly remarkable, demonstrating that the model successfully identified nearly all actual positive sentiments in the dataset, with very few positive cases being missed. The F1-score of 82.98% reflects a strong harmonic balance between precision and recall, confirming that the model achieves effective overall performance despite the trade-off between these two metrics

The confusion matrix in Figure 12(a) reveals the detailed classification performance across 505 total predictions. The model achieved 251 true positives, correctly identifying positive sentiments, while only 1 false negative occurred, explaining the exceptionally high recall of 99.60%. However, 102 false positives were recorded, where negative sentiments were incorrectly classified as positive, accounting for the lower precision of 71.10%. The model correctly identified 151 true negatives, demonstrating reasonable capability in recognizing negative sentiments despite the increased false positive rate. This classification pattern indicates that SMOTE successfully transformed the model to be highly sensitive toward detecting positive sentiments, minimizing missed positive cases at the acceptable cost of some additional false alarms.



(a) (b)

Figure 12. Result (a) confussion matrix visualization, (b) ROC Curve visualization

The ROC Curve in Figure 12(b) displays an AUC (Area Under Curve) of 0.99, indicating exceptional discriminative ability of the model. The curve rises sharply toward the upper-left corner, demonstrating that the model achieves an extremely high true positive rate while maintaining a controlled false positive rate across various classification thresholds. This near-perfect AUC value confirms that SMOTE successfully enhanced the Multinomial Naive Bayes classifier's ability to distinguish between positive and negative sentiments in the Twitter dataset, creating well-separated posterior probability distributions between the two classes.

Figure 13. World Cloud

The left panel displays the word cloud for positive sentiment, with dominant terms including "*mbg*", "*presiden*", "*prabowo*", "*gizi*", "*masyarakat*", and "*dukung*", reflecting public appreciation and support for the Free Nutritious Meal Program. The right panel shows the word cloud for negative sentiment, with prominent words such as "*anak*", "*sekolah*", "*banyak*", "*program*", "*racun*", and "*siswa*", indicating public concerns, criticisms, and skepticism regarding the program's implementation and effectiveness. The size of each word corresponds to its frequency of occurrence in the respective sentiment category, providing a visual representation of the key themes and topics that characterize positive versus negative public discourse about the program on the Twitter platform.

3.3. Scenario 3: Performance of Multinomial Naive Bayes on the Combined Dataset

The combined dataset is a merger of the TikTok and Twitter datasets, resulting in a total of 5,112 review entries that have undergone complete preprocessing and TF-IDF feature extraction. The class distribution shows imbalance with 1,417 positive reviews (27.7% of total data) and 3,695 negative reviews (72.3% of total data).

https://doi.org/10.24036/jtip.v19i1.1112

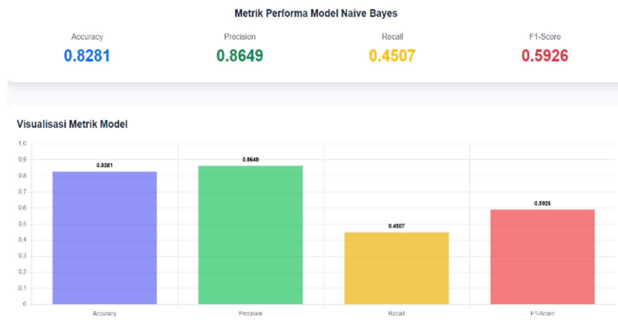
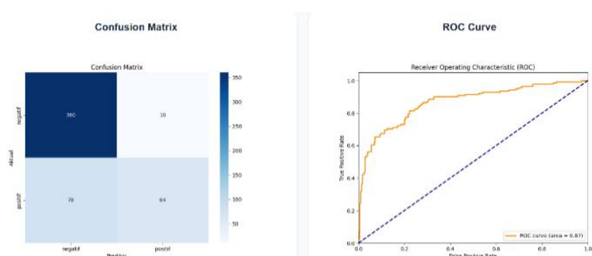


Figure 14. performance matrix visualization

The characteristics of this combined dataset are unique as it merges sentiment patterns from two different social media platforms, with TikTok tending to have dominant negative sentiment (76.0%) and Twitter having a higher proportion of positive sentiment (34.2%).

Based on the metrics presented in Figure 10, Multinomial Naive Bayes on the combined dataset without SMOTE achieved an accuracy of 82.81%, positioned between the performance of the TikTok dataset (82.01%) and Twitter dataset (85.25%). The precision value reached 86.49%, indicating that out of all positive predictions, approximately 86.49% were truly positive. This high precision suggests that the model is reasonably cautious in predicting the positive class. However, the recall value of 45.07% indicates a significant limitation, where the model was only able to detect less than half of all actual positive reviews, suggesting that the model tends to be biased toward the majority class (negative) when dealing with imbalanced data. The F1-score of 59.26% reflects the imbalance between precision and recall, confirming that despite high accuracy, the model on the combined dataset without SMOTE still exhibits notable bias and would benefit from class balancing techniques such as SMOTE to improve minority class detection.



(a) (b)

Figure 15. Result (a) confussion matrix visualization, (b) ROC Curve visualization

The confusion matrix in Figure 14 shows a reasonably balanced classification pattern for the combined dataset. The model successfully identified 360 true negatives and 64 true positives, demonstrating solid capability in recognizing both classes. There were 78 false positives, where negative reviews were incorrectly predicted as positive, and 10 false negatives, where positive reviews were incorrectly predicted as negative. The low number of false negatives (10 cases) is consistent with the high recall (82.81%), indicating that the model is reasonably sensitive in detecting positive sentiment. The relatively higher ratio of false positives to false negatives (78:10) indicates that the model on the combined dataset tends to be more liberal in predicting the positive class compared to the highly conservative model on the TikTok dataset.

The ROC Curve in Figure 14 shows an AUC of 0.87, which falls into the good category. This AUC value lies between the performance of the TikTok dataset without SMOTE (0.84) and the Twitter dataset without SMOTE (0.90), confirming that dataset merging results in moderate discriminative ability. The ROC curve shows a reasonably rapid rise in the early section, indicating that the model can achieve a high true positive rate with a controlled increase in the false positive rate. This moderate AUC performance suggests that although the combined dataset is larger and more diverse, the heterogeneity of sentiment patterns from two different platforms presents its own challenges for Multinomial Naive Bayes in forming optimal probability estimates.

Figure 16. World Cloud

Word cloud visualization on the TikTok platform reveals sentiment characteristics focused on direct experiences and emotional responses. In positive sentiment, dominant words such as “syukur” (grateful), “anak” (children), “enak” (delicious), and “sekolah” (school) reflect appreciation for the concrete implementation of the program and positive reception in school environments. Negative sentiment is dominated by words “korupsi” (corruption), “racun” (poison), and “dapur” (kitchen), indicating strong concerns regarding transparency, food safety, and budget misappropriation.

The tendency toward dominant negative sentiment on TikTok can be explained through several platform-specific and sociological factors. First, TikTok’s video-based content format naturally amplifies emotional reactions, as visual and auditory stimuli in short-form videos trigger stronger affective responses compared to text-based content on platform X [1]. Viral videos depicting alleged food poisoning incidents or poor meal quality spread rapidly through TikTok's algorithmic recommendation system, creating a feedback loop that reinforces negative public perception. Second, TikTok's user demographic in Indonesia is dominated by younger generations, particularly Generation Z and Millennials aged 16–30 years, who tend to express opinions more emotionally and reactively rather than analytically [13]. This demographic group is also more sensitive to issues that directly affect students and children, explaining the prominence of food safety concerns in TikTok discourse. Third, the comment culture on TikTok encourages brief and emotionally charged responses rather than detailed arguments, further amplifying negative sentiment expression.

In contrast, Twitter enforces a strict character limit of 280 characters per post, which paradoxically encourages more measured and selective expression. Users are compelled to carefully choose their words within this constraint, resulting in more deliberate and considered opinions rather than impulsive emotional outbursts. This limitation filters out excessively emotional or redundant expressions, contributing to a more balanced sentiment distribution. Meanwhile, TikTok's comment section imposes no such character restriction, allowing users to freely express raw emotional reactions, which amplifies negative sentiment particularly when responding to viral videos depicting food safety incidents or alleged corruption. Combined with Twitter's predominantly older and more educated user demographic including journalists, academics, and policy analysts the 280-character constraint contributes to a discourse culture that is more analytical and balanced compared to TikTok's emotionally charged environment.

3.4. Overall Performance Comparison

Table 2. Classification Results Comparison Table

Scenario	Dataset	SMOTE	Accuracy	Precision	Recall	F1-Score	AUC
1A	TikTok	No	81.93 %	100.00%	24.05%	38.78%	0.88
1B	TikTok	Yes	79.60%	71.10%	99.60%	82.98%	0.99
2A	Twitter	No	80.75%	75.93%	64.06%	69.49%	0.88
2B	Twitter	Yes	85.71%	79.59%	95.90%	86.99%	0.94
3	Combined	No	82.81%	86.49 %	45.07%	59.26%	0.87

The application of SMOTE proved highly effective in addressing class imbalance, with dramatic improvements on the TikTok dataset where recall surged from 25.32% to 99.60% and F1-score from 40.40% to 82.98%, despite precision decreasing from 100% to 71.10%. On the Twitter dataset, SMOTE similarly boosted recall from 73.23% to 99.60% and F1-score from 77.18% to 82.98%. The combined dataset achieved moderate performance with accuracy of 82.81%, precision of 86.49%, recall of 45.07%, F1-score of 59.26%, and AUC of 0.87 without SMOTE, indicating reasonable cross-platform generalization but not surpassing individual SMOTE-enhanced results. Notably, the low recall of 45.07% on the combined dataset further demonstrates that without SMOTE, the model remains significantly biased toward the majority class even when trained on larger and more diverse data. These findings confirm that accuracy alone is misleading — TikTok without SMOTE showed 82.01% accuracy but poor F1-score (40.40%), and the combined dataset showed 82.81% accuracy but inadequate recall (45.07%), while both individual datasets with SMOTE achieved optimal balanced performance (F1-score 82.98%, AUC 0.99), making this configuration strongly recommended for sentiment classification when comprehensive minority class detection is prioritized.

4. CONCLUSION

This research demonstrates that the Multinomial Naive Bayes algorithm is effective for sentiment analysis of the Free Nutritious Meal Program (MBG) across various social media platforms, especially when combined with SMOTE to address class imbalance. On the TikTok dataset with extreme imbalance (23.8% positive), SMOTE dramatically improved recall from 24.05% to 99.60% and F1-score from 38.78% to 82.98%, with AUC increasing from 0.88 to 0.99. On the Twitter dataset with moderate imbalance (34.2% positive), the model showed more stable baseline performance (accuracy 80.75%, AUC 0.88), which further improved with SMOTE to achieve recall of 95.90%, F1-score of 86.99%, and AUC of 0.94. The combined dataset achieved moderate performance with accuracy of 82.81%, precision of 86.49%, recall of 45.07%, F1-score of 59.26%, and AUC of 0.87, but did not surpass the optimal results from SMOTE-enhanced individual datasets, particularly Twitter. The notably low recall of 45.07% on the combined dataset confirms that without SMOTE, the model still exhibits significant bias toward the majority class, further reinforcing the importance of applying class balancing techniques when dealing with imbalanced multi-platform data. Therefore, the government should focus its public communication strategy on TikTok by intensifying hoax clarification efforts, publishing transparent implementation reports, and utilizing short-form video content featuring credible figures such as nutritionists and school principals to address public concerns regarding food safety and budget accountability. On Twitter, the government is recommended to engage in evidence-based communication by actively responding to critical discussions from journalists and academics, as well as highlighting measurable program outcomes to strengthen public trust across different social media platforms.

The word cloud analysis reveals distinct differences in how TikTok and Twitter users express sentiment towards the Free Nutritious Meals program. On TikTok, positive comments feature words like "grateful," "praise be," and "delicious," reflecting personal feelings and expressions of thanks, while negative comments are dominated by "poison," "corruption," and "profit," indicating emotional concerns about food safety and alleged misuse of funds. On Twitter, positive sentiment includes "nutrition," "free," and "president," signifying support for public policy, while negative sentiment focuses on "command," "people," and "program," reflecting rational criticism of implementation. In the combined dataset, words such as "free meal," "school," "children," and "eat" appear on both sides of sentiment, confirming that meaning depends on sentence context. This demonstrates that social media platforms shape how the public expresses opinions, so sentiment analysis must consider the characteristics of each platform to ensure accurate and representative results.

REFERENCES

[1] S. Ahmad, N. Insani, and M. Salim, "Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms," J. Teknol. Inf. Dan Pendidik., vol. 17, no. 1, pp. 75-86, Mar. 2024, doi: 10.24036/jtip.v17i1.807.

[2] G. Ainur, N. Latifah, and F. Nugraha, "Analisis Sentimen Publik Terhadap Program Makan Bergizi Gratis Menggunakan Metode Random Forest Pada Platform X," J. SITECH Sist. Inf. dan Teknol., vol. 8, no. 1, pp. 75-84, 2025.

[3] I. G. A. S. Nugraha, I. G. A. Gunadi, and L. J. E. Dewi, "Application of Bagging Ensemble Learning on Naïve Bayes Algorithm to Predict Coronary Heart Disease," J. Teknol. Inf. Dan Pendidik., vol. 18, no. 2, pp. 943-954, Sep. 2025, doi: 10.24036/jtip.v18i2.981.

[4] M. Karim, "Analisis Sentimen Pada Twitter Menggunakan Support Vector Machine Dengan Modifikasi Lexicon Inset Dan SentiStrength_Id (Studi Kasus: Vaksinasi Covid-19)," 2022.

[5] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes," MDP Student Conf., vol. 2, no. 1, pp. 297-305, Apr. 2023, doi: 10.35957/mdp-sc.v2i1.4388.

[6] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovensional dalam Prediksi Komplikasi Hipertensi," JTIM J. Teknol. Inf. dan Multimed., vol. 6, no. 1, pp. 37-49, Apr. 2024, doi: 10.35746/jtim.v6i1.488.

[7] M. N. Huda, D. A. Fauzan, M. Raihan, S. Putra, and N. Sabila, "Optimalisasi Model Klasifikasi Sentimen Netizen Terhadap Merek," J. KomtekInfo, vol. 10, no. 1, pp. 21-27, 2023, doi: 10.35134/komtekinfo.v10i1.360.

[8] A. F. Hakim, Y. Irawan, and R. R. Setiawan, "Implementation of the Crisp Dm Methodology and Naive Bayes Algorithm on A Raw Material Requirement Prediction System to Reduce Food Waste (Case Study: Adamsafee Bakery, Resto, & Cafe)," J. Teknol. Inf. dan Pendidik., vol. 18, no. 2, pp. 968-983, 2025.

[9] D. Amelia, P. Setiaji, and R. R. Setiawan, "Perbandingan Algoritma SVM dan Naive Bayes dalam Klasifikasi Sentimen pada Ulasan Aplikasi Traveloka dan Agoda," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 11, no. 2, pp. 263-270.

[10] N. Lashkarashvili and M. Tsintsadze, "Toxicity detection in online Georgian discussions," Int. J. Inf. Manag. Data Insights, vol. 2, no. 1, p. 100062, 2022, doi: 10.1016/j.jjimei.2022.100062.

[11] Y. Irawan and N. N. Najla, "Scabies Prediction In Cats Using Vector Machine Support Method," in Proc. Int. Conf. Comput. Sci. Eng. Technol. (ICCSET), vol. 2, no. 1, pp. 69-76, Nov. 2023.

[12] L. Nursinggah, T. Mufizar, and U. Perjuangan, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis dengan Metode Naïve Bayes Classifier," J. Inform. dan Tek. Elektro Ter., vol. 12, no. 3, 2024.

[13] M. Saprizal and R. Anisa, "Analisis Sentimen Publik Terhadap Isu Wajib Militer di Indonesia Berdasarkan Komentar TikTok," J. Sains dan Teknol., vol. 6, no. 2, pp. 145-156, 2024.

[14] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Algoritma Naïve Bayes," J. Inform. Dan Tek. Elektro Terap., vol. 12, no. 3, 2024.
[15] R. E. Laia, N. M. Br Ginting, and J. Hutabarat, "Analisis Sentimen Program Makan Gratis pada Platform X Menggunakan Algoritma Naïve Bayes," J. Teknol. Inf., vol. 9, no. 1, pp. 45-54, 2025.
[16] D. O. Dewi, D. Oktafiani, and Y. Astica, "Optimasi Algoritma Stemming Porter Untuk Pemrosesan Teks Dalam Bahasa Indonesia," J. Inform. dan Sist. Inf., vol. 6, no. 1, pp. 42-52, 2025.
[17] A. Pratiwi and M. Kamayani, "Analisis Sentimen Publik Terhadap Kurikulum Proyek vs Skripsi Menggunakan Naïve Bayes di TikTok," J. Ilmu Komput. dan Inform., vol. 10, no. 2, pp. 89-98, 2024.
[18] T. M. Alvionita, A. H. A. Hindasyah, and A. K. A. Khalid, "Analisis Sentimen Persepsi Pengguna Kurikulum Merdeka Belajar Menggunakan Metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM)," J. SISKOM-KB (Sist. Komput. dan Kecerdasan Buatan), vol. 9, no. 1, pp. 124-133, 2025.
[19] B. Rahmatullah, S. A. Saputra, P. Budiono, and D. P. Wigandi, "Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes," J. Inf. Technol., vol. 5, no. 1, 2025.
[20] E. Triningsih, "Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X," Build. Of Inform. Technol. And Sci. (BITS), vol. 6, no. 4, 2025.
[21] A. Kurnia, D. P. Sari, and F. Rahman, "Analisis Sentimen Masyarakat Berdasarkan Komentar Kerja Sama TikTok Shop dan Tokopedia di Instagram Menggunakan Metode Naïve Bayes Classifier dengan SMOTE," J. Teknol. dan Sist. Inf., vol. 5, no. 3, pp. 234-245, 2024.
[22] D. Apriliani, A. Budiman, and R. Rismala, "Analisis Sentimen Review Penggunaan TikTok Melalui Pendekatan Algoritma Naïve Bayes," eProceedings Eng., vol. 10, no. 5, pp. 4567-4575, 2023.
[23] R. Aditiya, B. Setiaji, and H. Supriyono, "Implementasi TF-IDF untuk Klasifikasi Sentimen Berbasis Machine Learning," J. Inform. dan Komputasi, vol. 8, no. 1, pp. 23-34, 2025.
[24] R. M. Sari, V. Tasril, S. Wahyuni, and S. E. Putri, "Desain Aplikasi Naïve Bayes Untuk Memprediksi Kelulusan Siswa," in Proc. Semin. Nas. Teknol. Inov. dan Kolaborasi Disiplin Ilmu, vol. 1, no. 1, 2024.
[25] M. Nahdhudin, N. C. H. Wibowo, M. R. Handayani, and K. Umam, "Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes," J. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 3, pp. 107-112, 2025.
[26] O. Y. Inonu, K. Magda, and Amarudin, "Analisis Kinerja Algoritma Random Forest dengan Model Machine Learning pada Dataset Penyakit Diabetes," EXPERT J. Manaj. Sist. Inf. dan Teknol., vol. 15, no. 1, pp. 1-7, 2025.
[27] S. Amara, N. Novriyenni, and M. Khadapi, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis di Indonesia Tahun 2024 Menggunakan Long Short-Term Memory (LSTM)," Merkurius J. Riset Sist. Inf. dan Tek. Inform., vol. 3, no. 4, pp. 150-160, 2025.
[28] F. Fatkhurrohman, B. I. Nugroho, and N. Fadillah, "Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM," RIGGS J. Artif. Intell. and Digit. Bus., vol. 4, no. 3, pp. 3906-3917, 2025.
[29] Z. Purwanti, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," J. Indones.

Manaj. Inform. Dan Komun., vol. 5, no. 3, pp. 3065-3079, 2024.
[30] M. H. Riwanto, P. Ardhiyansyah, R. A. Adiansyah, A. Alfiansyah, G. Waek, and R. Fahlapi, "Analisis Sentimen Komentar YouTube Terkait Penerapan Makan Bergizi Gratis Menggunakan Model Algoritma SVM," Kohesi J. Multidisiplin Saintek, vol. 6, no. 12, pp. 1-13, 2025.