CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 1003-1016
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Sentiment Analysis of Indonesia's Free Nutritious Meal Program on X Posts Using SVM and Random Forest

Ferdy Aliansyah Hasyim, Talenta Parfaibya Mahenindra, Lilis Sriwahyuni, Alika Azka Shapira, Wigawijayanti, Nadhifa Zahra Ghaisani, Mirlan Sujana, Sri Nurdiati, and Mohamad Khoirun Najib*

School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

Abstract

The Free Nutritious Meal (Makan Bergizi Gratis/MBG) program aims to address stunting in Indonesia, yet its implementation has sparked extensive public debate. This study evaluates public perception on social media X by comparing Support Vector Machine (SVM) and Random Forest algorithms for sentiment classification. A dataset of 5,988 tweets was collected via time-based sampling by randomly selecting three days per month from January to October 2025, with labels generated using a lexicon-based approach and verified against a manual validation subset. To address class imbalance, this research evaluates SMOTE and Class Weighting strategies, integrated with 5-fold Stratified Cross-Validation for statistical stability. The analysis reveals that positive sentiment (47.62%) and negative sentiment (39.81%) dominate the discourse, reflecting a polarized public response. Methodologically, SVM with SMOTE achieved the superior performance, yielding a stable mean Macro F1-score of 66.83%. However, the models exhibit limitations in distinguishing neutral factual reports from polarized opinions due to feature overlap in high-dimensional TF-IDF representations. These findings offer a preliminary mapping of policy acceptance and provide methodological insights into the effectiveness of hyperplane-based optimization over tree-based ensembles for Indonesian-language public policy discourse.

Keywords: Indonesian Text Classification; MBG Program; Random Forest; Sentiment Analysis; Support Vector Machine

1. Introduction

Child nutrition remains a significant challenge in Indonesia, with direct implications for physical growth, cognitive development, and the formation of long-term human capital [1]. According to national nutrition survey reports, the prevalence of stunting in Indonesia reached 21.6% in 2022, a figure that still requires serious and multisectoral intervention [2]. In response, the Indonesian government launched the Free Nutritious Meal Program (Makan Bergizi Gratis/MBG) to improve nutritional intake, particularly among school-aged children from vulnerable households [3]. This program is designed to ensure the availability of affordable and nutritious food while supporting the achievement of national stunting reduction targets. The implementation of the MBG Program has been accompanied by various controversies and public debates. While some stakeholders have welcomed the policy as a progressive and strategic initiative, others have

---

*Corresponding author. E-mail: mkhoirun@apps.ipb.ac.id

Submitted: January 25, 2026 Reviewed: March 14, 2026 Accepted: April 18, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.40717

raised concerns regarding program effectiveness, food quality and safety, distribution mechanisms, budget transparency, and the long-term sustainability of the policy. These differing perspectives reflect the complex dynamics of public opinion surrounding the MBG Program.

The social media platform X has emerged as a digital public sphere in which citizens express aspirations and criticisms of government policies. Although such opinions are disseminated in an unstructured manner, they contain valuable information when analyzed using appropriate methods. Sentiment analysis provides an effective approach for extracting subjective information from textual data and classifying it into emotional polarities such as positive, negative, or neutral [4]. Previous studies have attempted to analyze this discourse; however, significant research gaps remain. [5] achieved high overall accuracy but struggled with minority sentiment classes. Other studies utilized SMOTE on a smaller dataset, but the analysis was limited to a short temporal window, which potentially overlooks the evolving nature of public concerns [6]. Another comprehensive study compared multiple algorithms but relied on general datasets that lack the domain-specific linguistic nuances of Indonesian public policy discussions [7].

This study evaluates the effectiveness of Support Vector Machine (SVM) and Random Forest by comparing two strategies for handling class imbalance: SMOTE and Class Weighting. This comparison is critical for capturing nuanced public feedback that is often overshadowed in imbalanced datasets. Utilizing 5,988 tweets collected via time-based sampling from January to October 2025, this research provides a longitudinal perspective to mitigate biases from short-term event spikes. Furthermore, the study incorporates a 5-fold Stratified Cross-Validation to ensure statistical stability and a detailed error analysis to identify recurring misclassification patterns in Indonesian-language policy discourse. While opinions on platform X do not necessarily represent the entire Indonesian population, particularly rural families with limited digital presence, this study offers a preliminary mapping of digital discourse and methodological insights into effective configurations for policy-related sentiment classification.

2. Methods

This section explains the methodological framework applied in this study to analyze public sentiment toward the Free Nutritious Meal Program on platform X. It begins with the data source and collection procedure, followed by the classification methods used, namely Support Vector Machine and Random Forest. The section then describes the hyperparameter tuning strategy and the evaluation metrics employed to assess model performance in a consistent and reliable manner.

2.1. Data Source and Collection

The data in this study were collected from the social media platform X using Twitter Harvest v2.6.1 via Google Colab. The collection period spanned from January to October 2025, strictly filtering for Indonesian-language tweets containing the keyword "Makan Bergizi Gratis." To manage data volume and prevent the analysis from being skewed by short-term emotional surges, time-based sampling was applied by selecting three random days per month. This systematic approach successfully captured the sustained baseline of public opinion over the ten-month period. Through this process, a total of 7,452 tweets were collected. Table 1 presents a summary of the dataset, which includes two main attributes: "Created At", indicating the date and time the tweet was posted, and "Full Text", representing the complete content of the tweet.

2.2. Support Vector Machine

Support Vector Machine (SVM) is a machine learning model widely utilized for sentiment classification due to its robust theoretical foundations in statistical learning theory. SVM works by finding the optimal separating plane (hyperplane) to distinguish between two groups of data. Let $x = \{x_1, x_2, \ldots, x_n\} \in R^n$ denote a dataset. This hyperplane acts as a boundary line that

Table 1: Research dataset

Created at	Full text
Wed Jan 01 03:46:37 +0000 2025	Makan Bergizi gratis nantinya susu diganti rumput eeehh daun kelor. Tidak apa2. Ok2 sajahh. Yg penting penguasa pejabat dan aparat Sipil/militer kasih contoh dulu tiap hari makan daun kelor
Wed Jan 01 03:56:08 +0000 2025	makan siang gratis belum matang. program mencla mencle. diubah ke sarapan bergizi gratis cek tempo 31 mei 24
⋮	⋮
Wed Oct 29 22:25:53 +0000 2025	Padahal setiap kali pemerintah menggelontorkan program besar seperti MBG (Makan Bergizi Gratis) yang dijual ke publik adalah narasi mulia: membantu anak miskin memperbaiki gizi mencerdaskan generasi. -@barengwarga
Wed Oct 29 23:42:31 +0000 2025	Program Makan Bergizi Gratis merupakan wujud nyata kepedulian Pemerintah terhadap masa depan anak bangsa #DariMBGUntukIndonesiaHebat https://t.co/SzIjqWjio9

separates the two classes and is represented by Eq. (1):

$$w^T x + b = 0 \tag{1}$$

where w denotes the weight vector, x represents the input feature vector, and b is the bias value [8]. The optimal hyperplane separating the two classes is obtained by maximizing the margin, defined as the distance between the hyperplane and the nearest data points from each class. These nearest data points are referred to as support vectors. Support Vector Machine (SVM) classifies data by selecting the hyperplane with the largest margin.

To handle data that is not linearly separable, the mapping function Φ is used to map the data to a higher-dimensional space. Since Φ is generally not explicitly known, the kernel function K is defined, which implicitly transforms through Eq. (2):

$$K\left(u,v\right) = \Phi^T(u) \cdot \Phi(v) \tag{2}$$

with $u, v \in R^n$ and $\Phi : R^n \to R^m, n < m$. Common kernel functions include the Linear, Polynomial, and Radial Basis Function (RBF) kernels, the latter of which is utilized in this study due to its effectiveness in capturing non-linear linguistic patterns [9].

Although SVM is fundamentally a binary algorithm, this study classifies three sentiment categories: positive, neutral, and negative. To handle this, the default One-vs-One (OvO) strategy in *Scikit-Learn* was utilized. As described by [10], OvO decomposes a k-class problem into $k(k-1)/2$ independent binary classifiers, training one for each class pair (c_i, c_j). During prediction, the final sentiment label is determined through majority voting among these classifiers. This approach effectively manages the dataset's multi-polarity by establishing tailored decision boundaries for each specific pair of sentiments.

2.3. Random Forest

The Random Forest algorithm was also used to classify sentiment in this study. This algorithm works by building several independent classification models, then combining the prediction results of each model to produce a final decision [11–13]. As a development of the decision tree technique, Random Forest uses a bagging approach, which is the process of bootstrap sampling from training data to form multiple subsets. Next, the decision tree model is trained repeatedly on each subset [14]. An illustration of how the Random Forest algorithm works is shown in Fig. 1.

During the construction of the ensemble's decision trees, the recursive partitioning of data at each node was determined exclusively using the Gini impurity criterion. Gini impurity evaluates the quality of a split by measuring the probability of incorrectly classifying a randomly chosen

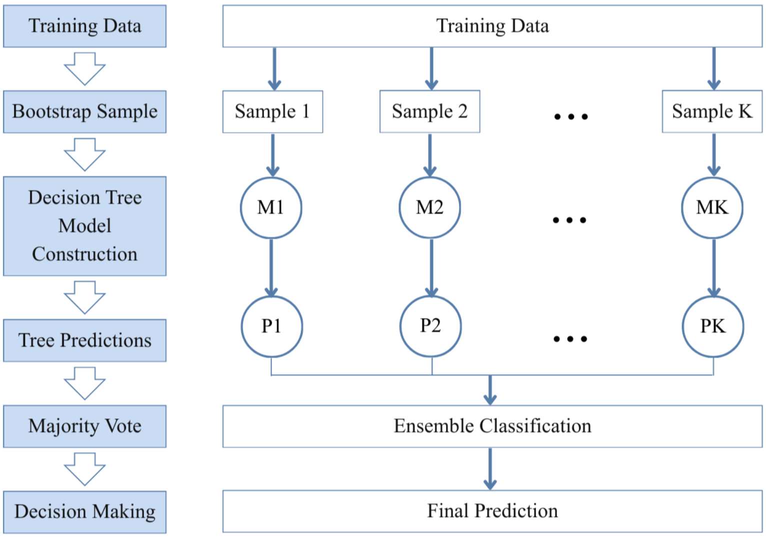

Fig. 1: Random Forest algorithm

element if it were labeled according to the class distribution within that specific node. This criterion was selected because it provides a computationally efficient evaluation for sparse, high-dimensional text data, and it serves as the standard default implementation in the Scikit-Learn framework [15].

Therefore, the separation process at each node calculates the Gini coefficient and subsequently determines the Gini decrease (or Gini gain) to find the most optimal partition. The formulas for calculating the Gini coefficient and Gini decrease are shown in Eq. (3)–Eq. (4):

$$Entropy(D) = -\sum_{i=1}^{k} p_i \log_2(p_i) \tag{3}$$

$$IG = Entropy(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} \times Entropy(D_i) \tag{4}$$

where D denotes the dataset, k is the number of partitions of D, p_i represents the proportion of partition D_i relative to D, IG is the information gain, $|D_i|$ is the number of observations in partition i, and $|D|$ is the total number of observations in the dataset.

Furthermore, the Gini coefficient is applied in each separation process. The separation process will consider the Gini coefficient to then calculate the Gini decrease. The formulas for calculating the Gini coefficient are shown in Eqs. (5)–(6)

$$gini(D) = 1 - \sum_{i=1}^{k} (p_i^2) \tag{5}$$

$$GD = gini(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} \times gini(D_i) \tag{6}$$

where D denotes the dataset, k is the number of partitions of D, p_i represents the proportion of partition D_i relative to D, GD is the Gini decrease, $|D_i|$ is the number of observations in partition i, and $|D|$ is the total number of observations in the dataset [16].

2.4. Hyperparameter Tuning

Hyperparameters are parameters selected before the model training process to determine how the model learns from data, so their selection greatly determines the quality of the final model results [17]. In this study, hyperparameter tuning is conducted using 5-fold Stratified Cross-Validation. This scheme divides the training data into five folds, where each fold is used as a validation set once while the remaining four folds serve as the training set, ensuring a robust evaluation and preventing overfitting. To ensure a fair and fully reproducible comparison, the exact same data splits (folds) are maintained across all algorithms and tuning methods by utilizing a fixed random state in the cross-validation generator. To optimize performance, the tuning process uses the macro-averaged F1-score as the evaluation metric. This metric is chosen to ensure that the model achieves balanced performance across all sentiment classes (positive, neutral, and negative), especially given the class imbalance in the dataset.

There are several techniques for automatically determining the best parameter combination in model training, such as GridSearchCV and RandomizedSearchCV [15]. For SVM, GridSearchCV is employed to exhaustively test multiple predefined parameter combinations. In contrast, the Random Forest model utilizes RandomizedSearchCV, which randomly selects 50 different parameter combinations for computational efficiency. These 50 iterations are chosen to provide a broad search across the parameter space without the high cost of an exhaustive grid search. To further mitigate the effects of an imbalanced dataset, a balanced class-weight parameter is explicitly included in the hyperparameter configurations for both models. The specific parameter ranges used in this study are detailed in Table 2.

Table 2: Parameters used in model training

Algorithm	Parameter	Value
SVM	Kernel	linear, RBF
	C	0.1, 1, 10, 100
	gamma	0.01, 0.1, 1
Random Forest	n_estimators	10, 50, 100, 200
	max_depth	10, 20, 30, 40, none
	min_samples_split	2, 5, 10, 20
	min_samples_leaf	1, 2, 4, 8

2.5. Evaluation Metrics

Next is the stage of evaluating and comparing the performance of both algorithms using several evaluation metrics, namely accuracy, precision, recall, and F1-score. Since this study addresses a multi-class classification problem with three sentiment classes (negative, neutral, and positive), the evaluation relies on a 3×3 confusion matrix. Let C_{ij} represent the number of instances from actual class i that are predicted as class j. For any specific class k, the True Positive (TP_k) is the number of correctly classified samples ($TP_k = C_{kk}$). The False Negative (FN_k) is the sum of all samples from actual class k that were incorrectly classified into other classes ($FN_k = \sum_{j \neq k} C_{kj}$). Conversely, the False Positive (FP_k) represents the sum of samples from other classes that were incorrectly classified as class k ($FP_k = \sum_{i \neq k} C_{ik}$) [18].

Accuracy in is used to measure the ratio of correct predictions in general, both positive and negative, to the total number of data samples. Accuracy is obtained by dividing the number of correct predictions by the total number of data samples, as represented by Eq. (5) [19].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

However, given that accuracy is prone to bias in unbalanced data, the measurement is reinforced with the F1-Score, which is the harmonic mean of precision and recall. On the other hand, precision measures the probability that samples classified as "Positive" (TP+FP) by the model are actually positive (TP), which can be seen in Eq. (8). Meanwhile, recall measures how many of the total samples that are originally positive (TP+FN) are successfully classified as positive class (TP) by the model, which is represented by Eq. (9) [20].

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Thus, the F1-score provides a single value that balances both, through Eq. (10).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Model performance was evaluated using Accuracy, Precision, Recall, and F1-score. To address class imbalance, Macro-averaging was applied to the latter three metrics. By calculating the arithmetic mean of the metrics independently for each class, this approach assigns equal weight to all categories regardless of sample size [19]. This ensures that the model's ability to classify minority segments, particularly neutral sentiment, is fairly represented in the overall evaluation.

3. Results and Discussion

This section presents the empirical findings of the study and discusses their implications in relation to the proposed classification framework. The discussion begins with the preprocessing and sentiment labeling stages, which form the basis for understanding the structure of the dataset before further exploration, feature extraction, and model evaluation are carried out.

3.1. Data Preprocessing and Sentiment Labeling

The collected data are raw data that contains unnecessary information. The data preprocessing stage aims to improve data quality and consistency so that machine learning models can work optimally. Data preprocessing begins with removing irrelevant information, such as punctuation marks, numbers, symbols, and converting all capital letters to lowercase letters. Then, data with identical content is deleted to prevent bias due to repeated data. Next, each text in the data is broken down into separate words to remove stopwords or common words that have no specific weight or meaning, such as conjunctions. Finally, each word with an affix is returned to its base form.

Table 3: Example of preprocessed data results

No	Before preprocessed	After preprocessed
1	Wah program bagus nih untuk mengatasi masalah gizi buruk di Indonesia. Semoga pelaksanaannya berjalan lancar dan bisa menjangkau anak-anak yang membutuhkan. Kita doakan saja yang terbaik! https://t.co/iFPCrLKbSq	program bagus nih atas gizi buruk indonesia moga laksana jalan lancar jangkau anak anak butuh doa baik
2	makan siang gratis belum matang. program mencla mencle. diubah ke sarapan bergizi gratis cek tempo 31 mei 24	makan siang gratis matang program mencla mencle ubah sarapan gizi gratis cek tempo mei
3	Makan bergizi gratis tingkatkan kualitas sumber daya manusia Indonesia #DukungMBG https://t.co/SWcX250IRz	makan gizi gratis tingkat kualitas sumber daya manusia indonesia

Table 3 presents examples of the results of the data preprocessing steps performed. All uppercase letters were converted to lowercase, and punctuation marks, numbers, hashtags, and links were removed. In addition, common words such as "wah," "di," and "kita" were eliminated. Words with affixes, such as "semoga," "pelaksanaannya," and "berjalan," were also reduced to their base forms. The raw data consisted of 7,452 observations; however, after removing duplicate texts, the number of observations was reduced to 5,988 tweets.

Sentiment labeling used a semi-automatic approach, combining a lexicon method with manual validation. First, an online lexicon assigned a polarity score to each text, classifying it as positive (>0), negative (<0), or neutral (=0). To avoid bias from the program's name, the words "makan", "gizi", and "gratis" were given a weight of 0. Next, to check the quality of these automatic results, we manually validated 500 randomly selected tweets. In this manual check, only 53 tweets (10.6%) needed label corrections. This shows a consistency rate of 89.4% between the lexicon and human validation. Because of this high accuracy, the lexicon approach is proven to be reliable for this dataset. Therefore, the automatic labels, combined with the 500 manually corrected ones, are valid to be used as the ground truth for training the machine learning models.

Table 4: Example of sentiment labeling results on data

No	Text	Lexicon Label	Validation Label
1	Bersama Kita Wujudkan Makan Bergizi Gratis! Perubahan besar dimulai dari langkah kecil. Melalui program MBG kita gotong royong menghadirkan makanan sehat bagi pelajar dan masyarakat yang membutuhkan. Setiap piring bergizi adalah wujud nyata kepedulian kita.	Positive	Positive
2	Tujuan utama Program Makan Bergizi Gratis (MBG) adalah : untukmeningkatkan kualitas sumber daya manusia Indonesia melalui perbaikan gizi terutama pada anak-anak sekolah ibu hamil dan menyusui serta untuk menekan angka	Negative	Neutral
3	Stop program makan bergizi gratis kembalikan anggaran ke masing2 K/L... biar normal lagi	Positive	Negative

Table 4 provides illustrative examples of the sentiment labeling results. Each data instance was initially assigned a polarity score based on the frequency and weight of words with positive or negative connotations found within the lexicon. Subsequently, automated sentiment labels were determined according to these polarity scores. To ensure the integrity of the training data, these automated labels underwent a manual validation process to correct any contextual misinterpretations by the lexicon, particularly in cases of neutral or ambiguous statements.

3.2. Data Exploration

3.2.1. Sentiment Distribution

Data exploration was conducted to provide an initial understanding of the sentiment characteristics within the collected tweets. Following the semi-automatic labeling process, a clear imbalance in the distribution of sentiment classes was observed. The dataset is predominantly composed of positive and negative categories, which appear in significantly higher proportions compared to neutral sentiment. Out of the 5,988 analyzed tweets, the majority expressed positive sentiment with 2,852 tweets, followed by 2,313 tweets expressing negative sentiment, and 823 tweets classified as neutral. The distribution patterns of these three categories are visually presented in the following pie chart.

Based on Fig. 2, the majority of the analyzed tweets exhibited a positive sentiment toward the MBG Program, accounting for 47.63% of the total dataset. Negative sentiment followed as the second-largest category, representing 38.63%, while neutral sentiment accounted for the remaining 13.74%. These results indicate that while public opinion on X predominantly reflects support for the program, the substantial proportion of negative sentiment suggests a significant level of criticism and policy-related skepticism. The smaller neutral segment highlights that a

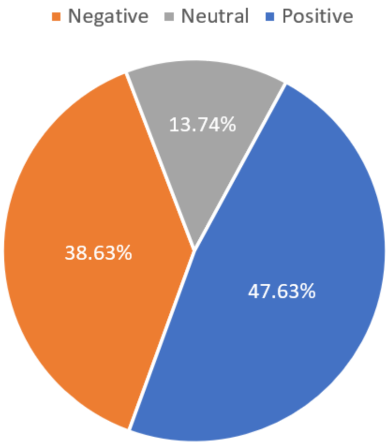

Fig. 2: Percentage of labeling results

minor portion of users focused on purely informative monitoring or news dissemination without expressing a clear stance.

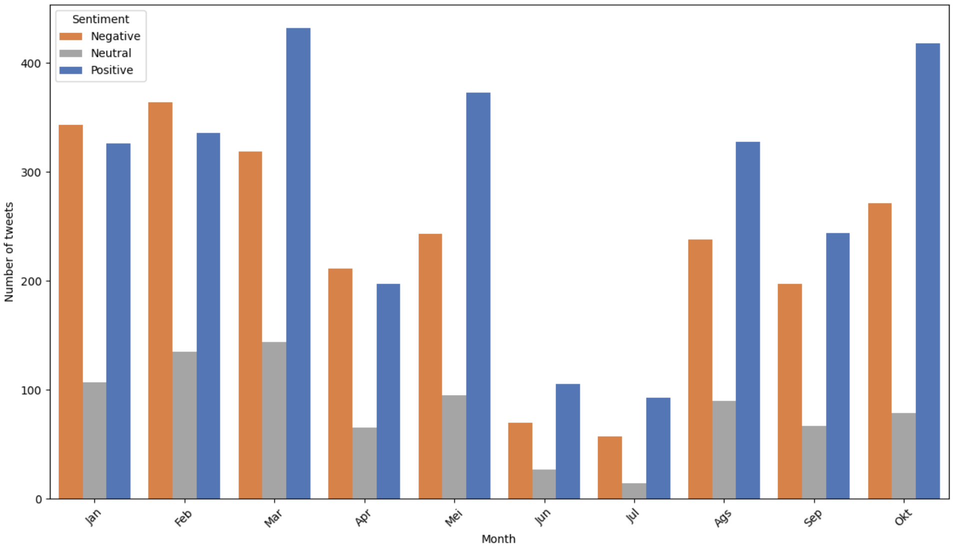

Fig. 3: Comparison of the number of tweets per sentiment category each month

The temporal distribution of sentiment from January to October 2025, as illustrated in Fig. 3, shows dynamic fluctuations throughout the study period. In early January and February, negative sentiment slightly outweighed positive sentiment, indicating initial public skepticism during the program's initial announcement phase. However, a significant shift occurred in March, with positive sentiment surging to its highest point in the first half of the year, followed by a significant decline in discussion volume during the mid-year period (June and July). Sentiment increased significantly again from August to October, with positive sentiment reaching a second peak, indicating that the sampling method successfully captured underlying sentiment and significant shifts in public discourse as the program approached its implementation phase.

3.2.2. Word Cloud Based on Sentiment

Word clouds were employed to identify dominant keywords and linguistic patterns within each sentiment category. Unlike unigram analysis, the bigram approach used in this study (as shown in Fig. 4) captures the contextual relationship between adjacent words, providing deeper insights into the specific themes of public discourse.

(a) **(b)** **(c)**

Fig. 4: Word cloud of the three sentiment categories: (a) negative, (b) neutral, (c) positive.

Fig. 4a shows that negative sentiment is characterized by clusters of bigrams reflecting systemic anxiety and safety risks. Dominant bigrams such as "orang tua" (parents) and "bahan baku" combined with "racun" and "generasi emas" suggest a high level of concern regarding food safety and its potential impact on children's health. Furthermore, the presence of "badan nasional" and "bgn" (National Nutrition Agency) in a negative context, along with terms like "butuh" (need) and "lanjut", indicates public skepticism and demands for better management and transparency in the program's execution to avoid mass health incidents.

Neutral sentiment, as presented in Fig. 4b, is dominated by institutional and operational terminology, reflecting an informational discourse. The prominence of bigrams like "babinsa", "koramil", and "laksana giat" suggests that neutral tweets are primarily composed of field monitoring reports and official activity updates related to local military units. The mentions of "badan nasional", "rp triliun" (trillions of rupiah), and "ekonomi lokal" suggest discussions focused on the logistical aspects, budget allocation, and the broader economic framework of the program, confirming that the neutral category serves as a channel for information dissemination regarding the program's roll-out.

Finally, Fig. 4c shows that positive sentiment is anchored in the themes of national progress and student welfare. Dominant bigrams, such as "sehat cerdas" (healthy and smart), "penuh" (full), and "terima manfaat" (receiving benefits), reflect a strong public appreciation for the program's role in improving nutrition. The frequent appearance of "presiden subianto" and "lapang kerja" (job opportunities) alongside "dukung penuh" indicates that positive sentiment is closely tied to political trust and the hope for improved human resource quality. These patterns suggest that supporters view the MBG Program as a strategic investment in the nation's future intelligence and economic growth."

3.3. Feature Extraction

Feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) to transform textual data into a numerical representation in the form of an $n \times m$ matrix, where n denotes the number of data instances (tweets), and m represents the number of unique terms. Each term is assigned a weight that reflects its importance and frequency within the dataset. Table 5 presents the results of TF-IDF feature extraction in the form of a matrix with dimensions $5,988 \times 7,301$.

Table 5: Results of TF-IDF application

Index	**aaa**	**aah**	**abad**	**abadi**	$\cdots$	**zulkarman**	**zulkifli**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
767	0.0	0.0	0.0	0.2	0.0	0.0	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6047	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Based on Table 5, a total of 7,301 unique terms were identified as features in the TF-IDF matrix. Each cell in the matrix contains the TF-IDF weight of a specific term in a particular document, where a value of 0.0 indicates that the term does not appear in the document, while positive values indicate the relevance of the term to the document and the overall corpus.

3.4. Sentiment Classification Model

Sentiment classification was conducted using Support Vector Machine (SVM) and Random Forest algorithms. The dataset was split into a 90% training set and a 10% testing set. During tuning, a 5-fold Stratified Cross-Validation was applied only to the training set. Two main strategies were applied to address class imbalance: Synthetic Minority Oversampling Technique (SMOTE) and Class Weighting, which were integrated into a pipeline. This ensures that the balancing process is applied only to the training folds during each iteration, leaving the validation folds unchanged.

Table 6: Results of SMOTE application

Category	**Negative**	**Neutral**	**Positive**	**Total**
Before SMOTE	2,081	741	2,567	5,389
After SMOTE	2,567	2,567	2,567	7,701

Table 6 presents the results of applying SMOTE to the training data. Through SMOTE, the number of negative and neutral sentiment samples was synthetically increased to match the positive sentiment class, resulting in 2,567 samples for each category. This balancing process reduces algorithmic bias toward the majority class during training and enhances the model's ability to predict previously underrepresented sentiment classes.

Based on the cross-validation tuning, the SVM model demonstrated high stability. The RBF kernel was consistently selected, with optimal parameters at C = 10 and gamma = 0.1 across all scenarios. During this phase, the SVM model with SMOTE achieved the highest cross-validation Macro F1-score of 67.83%. Following this, the best-performing SVM configurations were retrained on the entire 90% training set and evaluated on the unseen 10% test set. The final results are presented in Table 7.

Table 7: Evaluation of the SVM models (Macro-Averaged)

Category	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Baseline	81.30	77.25	72.28	73.73
SMOTE	81.97	77.15	75.48	76.18
Class Weighting	80.63	75.06	76.05	75.50

According to Table 7, the SVM with SMOTE achieved the best overall performance, reaching a final test-set Macro F1-score of 76.18%. The 2.45% increase over the baseline demonstrates that SMOTE effectively strengthened the model's ability to identify the Neutral class. Furthermore, the difference between the cross-validation score (67.83%) and the final test score (76.18%) is expected, as the final model was trained on the entire 90% training data, giving it more information to learn from.

Hyperparameter tuning for the Random Forest model showed significant sensitivity to data distribution across scenarios. The Baseline reached its optimum with 50 trees, min_samples_split = 10, min_samples_leaf = 2, and unrestricted depth. In contrast, both the SMOTE and Class Weighting scenarios converged on the exact same optimal configurations: an ensemble of 100 trees, min_samples_split = 2, min_samples_leaf = 1, and a max_depth = 40. These variations indicate that as the minority class volume or weight increases, the tree-based model requires a larger forest and finer split conditions to capture the balanced patterns, while simultaneously constraining the depth to prevent overfitting.

Table 8: Evaluation of the Random Forest models (Macro-Averaged)

Category	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Baseline	70.78	72.96	61.02	63.27
SMOTE	67.11	62.27	64.61	62.87
Class Weighting	65.94	61.16	62.82	61.78

As presented in Table 8, the Random Forest model achieved its highest Macro F1-Score of 63.27% in the baseline scenario, experiencing a slight performance decline after the application of SMOTE (62.87%) and Class Weighting (61.78%). This decline suggests that in a highly dimensional and sparse TF-IDF feature space, the synthetic samples generated by SMOTE may introduce noise and overlapping boundaries between classes. Tree-based recursive partitioning algorithms, such as Random Forest, are highly sensitive to such noise, often struggling to find optimal splits and subsequently failing to generalize well on the minority class. However, despite these improvements, Random Forest's overall performance remained significantly lower than SVM's. This disparity is primarily due to the high-dimensional and sparse nature of the TF-IDF features, which are mathematically more difficult for tree-based recursive partitioning to navigate compared to the hyperplane-based approach of SVM.

Among all evaluated models, the SVM model with SMOTE achieves the highest accuracy, precision, recall, and Macro F1-score, and can therefore be considered the best performing model. To ensure stability, a 5-fold Stratified Cross-Validation was performed, yielding a mean Macro F1-score of 66.83% with a low standard deviation of 2.5%. This consistency confirms the model's robust generalization and ensures that the results are not biased by a specific data split.

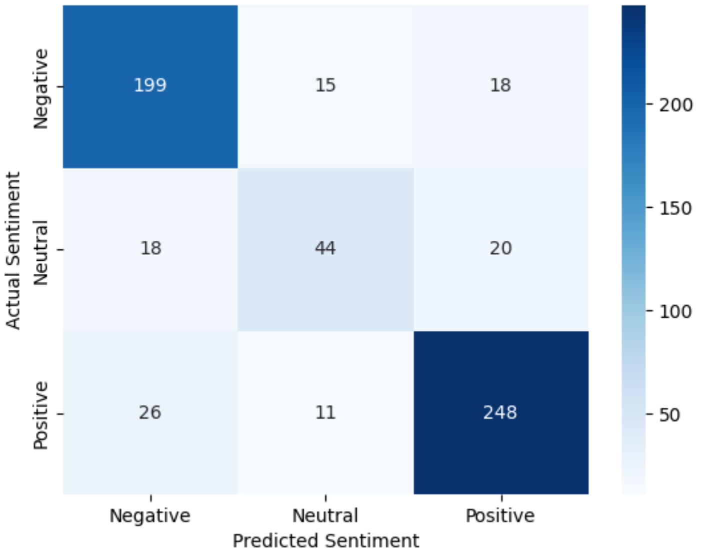

Fig. 5: Confusion matrix diagram of the SVM model with SMOTE.

Based on the Confusion Matrix in Fig. 5, the model demonstrates strong performance in recognizing sentiment categories, although the neutral sentiment remains the most challenging to distinguish. Despite theoretical concerns regarding synthetic data, these results suggest that balancing the dataset at the feature-space level via SMOTE significantly enhances the model's generalization capabilities for imbalanced social media data. This improvement is particularly evident in the model's increased ability to identify the Neutral class compared to the baseline results.

Table 9: Examples of misclassified neutral tweets

No	Text	Actual Label	Predicted Label
1	kapolsek teluk nibung monitoring kegiatan makan bergizi gratis oleh sppg yayasan al muslimun nusantara siswa menerima manfaat dukung generasi sehat amp kuat	Neutral	Positive
2	keterlibatan umkm pada pelaksanaan program makan bergizi gratis mbak masih minim salah satunya karena banyak umkm yang belum memiliki legalitas dan modal md	Neutral	Negative
3	makan bergizi gratis mbak di kecamatan kamal bangkalan jatim terkontaminasi ulat sayur namun pihak sppg menilai jenis ulat tersebut bisa dikonsumsi dan tinggi protein	Neutral	Positive

According to Table 9, neutral sentiment frequently suffers from misclassification due to feature overlap in factual discourse. Objective reports often contain polarized vocabulary that biases the model's decision boundary. For example, factual reports in Tweets 1 and 3 are misclassified as Positive because they contain high-weight tokens like "manfaat" (benefit) and "sehat" (healthy). Conversely, Tweet 2 is pulled toward the Negative class due to terms like "minim" (minimal) that the model strongly associates with criticism. In the high-dimensional TF-IDF feature space, these informational vectors fall into a "gray area" near the SVM hyperplane, making it difficult for the RBF kernel to distinguish objective reporting from explicit opinions.

4. Conclusion

This study demonstrates that public sentiment toward the Free Nutritious Meal (Makan Bergizi Gratis/MBG) program on platform X is characterized by a mix of strong support and significant implementation concerns. The analysis indicates that while positive sentiment appears more frequent, a substantial portion of the discourse highlights critical issues regarding food safety and delivery. To ensure the reliability of these findings, the automated labels were cross-validated against a manually annotated subset, confirming the general accuracy of the labeling process. Methodologically, the Support Vector Machine (SVM) with SMOTE proved more effective than Random Forest for this high-dimensional TF-IDF dataset, showing superior stability with a consistent 5-fold cross-validation mean Macro F1-score of 66.83%.

However, these findings should be interpreted with certain limitations in mind. The fact that 10.6% of the labels required manual correction highlights the difficulty of using lexicon-based approaches to capture sarcasm and complex context in public policy discussions. Furthermore, the model still struggles to consistently distinguish neutral factual reporting from polarized opinions due to feature overlap. Finally, since the data only comes from platform X, it may not fully represent the general public's perspective. Future research should consider collecting data from multiple platforms, expanding the manually annotated dataset, and exploring transformer-based models to improve the classification of complex sentiments.

CRediT Authorship Contribution Statement

Ferdy Aliansyah Hasyim: Conceptualization, Software, Formal analysis, Writing–Original Draft. **Talenta Parfaibya Mahenindra:** Conceptualization, Methodology, Writing–Original Draft. **Lilis Sriwahyuni:** Supervision, Validation, Writing–Review & Editing. **Alika Azka Shapira:** Conceptualization, Investigation, Writing–Original Draft. **Wigawijayanti:** Conceptualization, Visualization, Writing–Original Draft. **Nadhifa Zahra Ghaisani:** Conceptualization, Investigation, Writing–Original Draft. **Mirlan Sujana:** Conceptualization, Writing–Original

Draft. **Sri Nurdiati:** Supervision, Validation, Writing–Review & Editing. **Mohamad Khoirun Najib:** Supervision, Validation, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted technologies

Generative AI and AI-assisted technologies were used during the preparation of this manuscript under the supervision and responsibility of the authors. Google Gemini 3 was utilized for writing assistance and translation, while Grammarly was employed for grammar correction, tone adjustment, and improving readability. All AI-generated outputs were reviewed, verified, and approved by the responsible authors.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The data and code supporting the findings of this study are available from the corresponding author upon reasonable request and subject to confidentiality agreements.

References

[1] Indriya Laras Pramesthi, Luh Ade Ari Wiradnyani, Roselynne Anggraini, Judhiastuty Februhartanty, Wowon Widaryat, Bambang Hadi Waluyo, Agung Tri Wahyunto, Muchtaruddin Mansyur, and Umi Fahmida. "Evaluating the Impact of Indonesia's National School Feeding Program (ProGAS) on Children's Nutrition and Learning Environment: A Mixed-Methods Approach". In: *Nutrients* 17.22 (2025), p. 3575. DOI: 10.3390/nu17223575.

[2] Prima Maharani Putri, Aqilla Shafa Shafira, and Gembong Satria Mahardhika. "Stunting Reduction Strategy in Indonesia: Maternal Knowledge Aspects". In: *The Indonesian Journal of Public Health* 19.2 (2024), pp. 329–343. DOI: 10.20473/ijph.v19i2.2024.329-343.

[3] Firre An Suprapto, Editha Praditya, Reffi Marizka Dewi, and Wignyo Adiyoso. "A Policy Implementation Review of the Free Nutritious Meal (MBG) Program". In: *The Journal of Indonesia Sustainable Development Planning* 6.2 (2025), pp. 297–312. DOI: 10.46456/jisdep.v6i2.798.

[4] Bing Liu. *Sentiment Analysis and Opinion Mining.* Vol. 5. Synthesis Lectures on Human Language Technologies 1. San Rafael, CA: Morgan & Claypool Publishers, 2012. DOI: 10.1007/978-3-031-02145-9.

[5] Fatkhurrohman Fatkhurrohman, Bangkit Indarmawan Nugroho, and Nurul Fadillah. "Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM". In: *RIGGS: Journal of Artificial Intelligence and Digital Business* 4.3 (2025), pp. 3906–3917. DOI: 10.31004/riggs.v4i3.2533.

[6] Elsa Triningsih, M. Afdal, Inggih Permana, and Nesdi Evrilyan Rozanda. "Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X". In: *Building of Informatics, Technology and Science (BITS)* 6.4 (2025), pp. 2240–2250. DOI: 10.47065/bits.v6i4.6534.

[7] Musriatun Napiah, Sujiliani Heristian, Mugi Raharjo, and Rachmat Adi Purnama. "Analyzing Public Sentiment Toward Makanan Bergizi Gratis Program Using Machine Learning". In: *Computer Science (CO-SCIENCE)* 6.1 (2026), pp. 30–38. DOI: `10.31294/co-science.v6i1.10445`.

[8] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez. "A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends". In: *Neurocomputing* 408 (2020), pp. 189–215. DOI: `10.1016/j.neucom.2019.10.118`.

[9] Yulia Restiani and Joko Purwadi. "Support Vector Machine for Classification: A Mathematical and Scientific Approach in Data Analysis". In: *Jurnal Penelitian Pendidikan IPA* 10.11 (2024), pp. 9896–9903. DOI: `10.29303/jppipa.v10i11.8122`.

[10] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. `http://probml.github.io/book1`.

[11] Sunwoo Han, Brian D. Williamson, and Youyi Fong. "Improving Random Forest predictions in small datasets from two-phase sampling designs". In: *BMC Medical Informatics and Decision Making* 21.1 (2021), p. 322. DOI: `10.1186/s12911-021-01688-3`.

[12] Ghevira Chairunisa, Mohamad Khoirun Najib, Sri Nurdiati, Salsabila F Imni, Wardah Sanjaya, Rizka D Andriani, Renda SP Putri, Dhea Ekaputri, et al. "Life Expectancy Prediction Using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions". In: *Jurnal Sintak* 2.2 (2024), pp. 71–82. DOI: `10.62375/jsintak.v2i2.249`.

[13] EH Nugrahani, S Nurdiati, F Bukhari, MK Najib, DM Sebastian, and PAN Fallahi. "Sensitivity and feature importance of climate factors for predicting fire hotspots using machine learning methods". In: *IAES International Journal of Artificial Intelligence* 13.2 (2024), pp. 2210–2223. DOI: `10.11591/ijai.v13.i2.pp2212-2225`.

[14] Gi-Wook Cha, Hyeun-Jun Moon, and Young-Chan Kim. "Comparison of Random Forest and Gradient Boosting Machine Models for Predicting Demolition Waste Based on Small Datasets and Categorical Variables". In: *International Journal of Environmental Research and Public Health* 18.16 (2021), p. 8530. DOI: `10.3390/ijerph18168530`.

[15] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.

[16] Robert G. Gallager. "Claude E. Shannon: A retrospective on his life, work, and impact". In: *IEEE Transactions on Information Theory* 47.7 (2001), pp. 2681–2695. DOI: `10.1109/18.959253`.

[17] Mulia Kevin Suryadi, Rudy Herteno, Setyo Wahyu Saputro, Mohammad Reza Faisal, and Radityo Adi Nugroho. "Comparative Study of Various Hyperparameter Tuning on Random Forest Classification with SMOTE and Feature Selection Using Genetic Algorithm in Software Defect Prediction". In: *Journal of Electronics, Electromedical Engineering, and Medical Informatics* 6.2 (2024), pp. 137–147. DOI: `10.35882/jeeemi.v6i2.375`.

[18] Alaa Tharwat. "Classification Assessment Methods". In: *Applied Computing and Informatics* 17.1 (2021), pp. 168–192. DOI: `10.1016/j.aci.2018.08.003`.

[19] Margherita Grandini, Enrico Bagli, and Giorgio Visani. "Metrics for Multi-class Classification: An Overview". In: *arXiv preprint* (2020). DOI: `10.48550/arXiv.2008.05756`. arXiv: `2008.05756 [stat.ML]`.

[20] Hongwon Yun. "Prediction Model of Algal Blooms Using Logistic Regression and Confusion Matrix". In: *International Journal of Electrical and Computer Engineering (IJECE)* 11.3 (2021), pp. 2407–2413. DOI: `10.11591/ijece.v11i3.pp2407-2413`.