Vol. 8, No. 1, February 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Comparative Performance Analysis of Random Forest and Logistic Regression for Sentiment Classification of the Makan Bergizi Gratis Program on Platform X

Slamet Endro Prianto[1], Berlilana[2], Rujianto Eko Saputro[3]

[1]Magister of Computer Science, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

[2,3]Department of Computer Science, Universitas Amikom Purwokerto, Indonesia

Received:
November 3, 2025
Revised:
December 3, 2025
Accepted:
December 21, 2025
Published:
February 12, 2026

Corresponding Author:
Author Name*:
Slamet Endro Prianto
Email*:
slametendrop@gmail.com

DOI:
10.63158/journalisi.v8i1.1371

Abstract: The "Makan Bergizi Gratis" (MBG) program is a strategic government initiative designed to combat stunting that has triggered extensive public discourse regarding its implementation. This study aims to analyze public sentiment towards the MBG program and compare the classification performance of Random Forest (RF) and Logistic Regression (LR) algorithms. The research utilized a dataset of 1,002 tweets collected from Platform X (Twitter) between January and November 2025, processed through comprehensive preprocessing and TF-IDF feature extraction. A critical exploratory finding revealed a severe class imbalance, with Positive sentiment dominating at 95.3% of the dataset. Performance evaluation indicates that the RF algorithm marginally outperformed LR, achieving an accuracy of 77% and a weighted average F1-score of 0.68, whereas LR recorded an accuracy of 76% with a weighted F1-score of 0.67. However, further analysis demonstrates that both models faced significant challenges in minority class identification, recording a Precision and F1-score of 0.00 for Negative sentiment. This study provides empirical evidence that while RF exhibits superior robustness in handling complex data patterns, the high degree of data imbalance remains a critical hurdle for model accountability in public policy sentiment analysis.

Keywords: Sentiment Analysis, Random Forest, Logistic Regression, Makan Bergizi Gratis (MBG), Class Imbalance

1. INTRODUCTION

Malnutrition and stunting have emerged as key global and national problems because of their enormous influence on human resource quality and national development sustainability [1]. In Indonesia, this issue is a key priority for the "Indonesia Emas 2045" vision, which will result in large-scale nutritional intervention measures [2]. The Makan Bergizi Gratis (MBG) initiative, previously known as the Free Lunch initiative, has emerged as a key tool for addressing these requirements. However, since its launch, the initiative has sparked widespread public debate on social media sites such as X (Twitter), mainly about its large budget and operational issues [3]. Malnutrition and stunting have become critical global and national concern due to their significant impact on the quality of human resources and the sustainability of a country's development.

Debates about the program's effectiveness, budget ($10,000–$15,000 per serving), and food safety issues (such as cases of poisoning and menus deemed unfit for consumption) have sparked a massive wave of digital conversations. Analyzing public mood, whether good, negative, or neutral, is critical for stakeholders to maintain program accountability and efficacy [4] [5]. Machine Learning (ML) techniques are the most efficient solution for objectively processing large amounts of unstructured opinion data [6]. Classic algorithms like as Random Forest (RF) and Logistic Regression (LR) continue to be popular solutions for classifying social and political issues because to their competitive performance and ease of implementation [7].

Despite the increased interest in this area, current research on MBG sentiment analysis has limits [3]. Prior research has mostly concentrated on single-algorithm implementations, such as Naïve Bayes or Support Vector Machine (SVM), without undertaking thorough performance comparisons [8] [5] [9]. Furthermore, past research frequently misses the difficulty of significant data imbalance in government policy discussions, in which one sentiment class frequently dominates the dataset [5] [10] [11]. This gap results in a lack of empirical information about which categorization model is the most resilient for this specific area.

This work intends to close that gap by conducting a comparative evaluation of RF and LR algorithms under high data imbalance. This study provides new insights into model

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

dependability for high-noise social media data by contrasting RF's ensemble learning architecture with LR's linear probabilistic method. The findings are meant to help the government identify the best algorithm for categorizing public mood, allowing it to properly map popular acceptability of the MBG program.

2. METHODS

2.1 Data Collection

This study processes public opinion on the MBG using ML-based Sentiment Analysis techniques [12]. Data collection was conducted on the social media platform X (formerly Twitter) using the Twitter API v2 via the Python tweepy library, allowing for real-time capture of public discussions. The data crawling process began on November 27, 2025, and the program ran from January 2025 to November 2025. Exceeding the initial target of collecting 600–1,000 datasets, a total of 1,002 raw tweets were successfully collected. For example, "MBG", "MakanBergiziGratis", "MakanSiangGratis", "Prabowo", "Gibran", "Gizi", and "AnggaranMBG" are some hashtags and keywords relevant to key political figures and policies.

Several filtering criteria were used to verify that the data was high quality and genuine. To collect only organic public opinion, bot-like retweets and duplicate material were deleted [13]. Furthermore, in accordance with ethical concerns and data protection regulations, all personal user identifiers (usernames and user IDs) were anonymized immediately upon collection, and the data was utilized purely for academic research purposes.

2.2 Data Preprocessing

The data preprocessing stage goes thru a number of rigorous procedures to ensure data quality. This procedure includes:

1) Cleaning: This step involves removing URLs, mentions, hashtags, numbers, and emoticons, which are considered as noise in the text data, to improve the quality of analysis [14]..
2) Case Folding: All text is converted to lowercase to standardize the input, ensuring that words like "Makan" and "makan" are treated as equivalent, thus enhancing consistency across the dataset [15].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3) Tokenization & Normalization: The text is split into individual tokens (words or phrases), and non-standard slang terms are converted into their formal, standard Indonesian equivalents using a custom colloquial dictionary (kamus alay). For example, "yg" is normalized to "yang" [16].
4) Stopword Removal & Stemming: Commonly occurring words that do not contribute to sentiment analysis, such as "dan" (and) or "itu" (that), are removed. Additionally, words are reduced to their root forms through stemming. In this study, we used the Sastrawi library, which is optimized for Indonesian morphological analysis, ensuring effective and accurate stemming [17].

2.3 Data Processing

The TF-IDF (Term Frequency-Inverse Document Frequency) method is used to analyze the text into Positive and Negative categories and convert it into numerical vectors. While deep learning embeddings (e.g., Word2Vec) offer semantic depth, TF-IDF was selected for this study due to its computational efficiency, interpretability, and proven effectiveness for classical Machine Learning models on sparse datasets. Words that appear too frequently in all documents (such as "*dan*" and "*yang*") have their values reduced (normalized) so they don't dominate important keywords (like "*gizi*" or "*mahal*"). Figure 1 shows the stages of analysis performed.

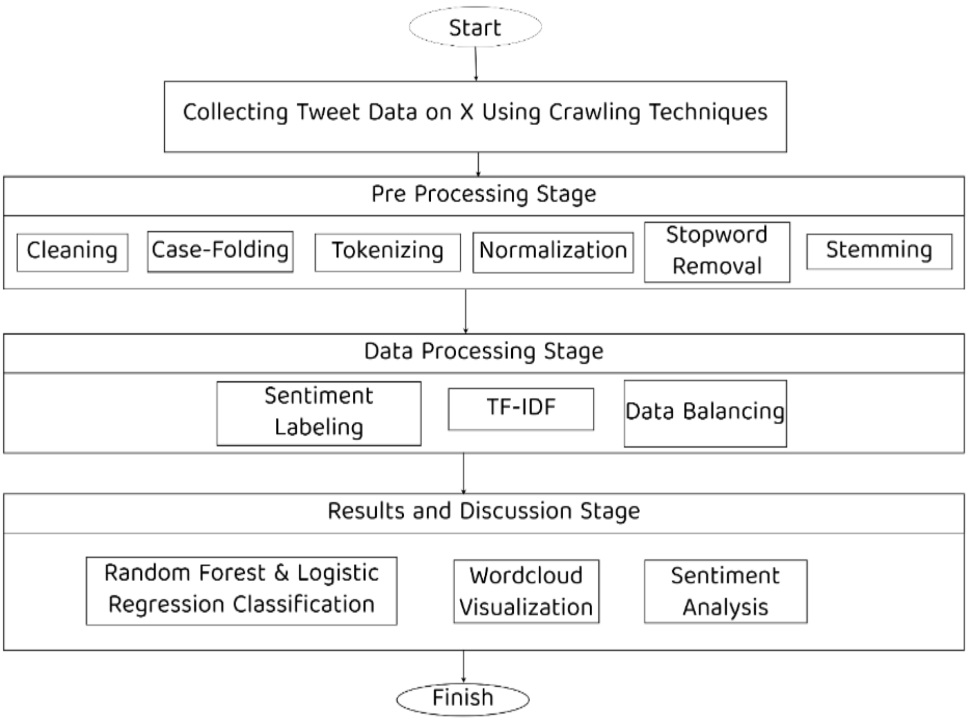

Figure 1. Research Processing Stages

2.4 Classification Algorithms and Hyperparameters

The study uses two widely recognized machine learning algorithms: Random Forest (RF) and Logistic Regression (LR), to classify the sentiment of public opinions.

1) The Random Forest (RF) algorithm

RF is an ensemble learning method for classification and regression that uses a number of decision trees trained on randomly selected data samples. Each tree has a different structure [18]. The RF algorithm produces more accurate and consistent predictions because it minimizes the risk of overfitting and improves model robustness, making it popular in many data analysis applications [19].

2) The Logistic Regression (LR) algorithm

LR is a classification method used to predict the likelihood that data will belong to one of two available classes. [20]. The following formula can be used to describe the basic equation of a LR model with one independent variable as shown in Equation 1 and 2.

$$P(Y = 1) = \frac{1}{1+e\ 2\text{-}(\beta0+\beta1x)} \tag{1}$$

$$P(Y = 0) = \frac{1}{1+e\ 2\text{-}(\beta0+\beta1x)} \tag{2}$$

To ensure the reproducibility of the study, the model parameter configurations are detailed in Table 1. It should be noted that in this experiment, the class_weight parameter is set to the default (None) and resampling techniques (e.g., SMOTE) or cost-sensitive learning were intentionally excluded from the experimental design. This decision was made to establish a baseline diagnostic of how standard ML algorithms behave under 'natural' social media data conditions. By retaining the original distribution, this research aims to empirically demonstrate the accuracy paradox where models achieve high accuracy solely by overfitting to the majority class thereby proving the necessity of advanced handling techniques in future operational implementations.

Table 1. Experimental Hyperparameter Configuration

Algorithm	Parameter	Config Value	Description
RF	n_estimators	100	Number of decision trees in the forest.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Algorithm	Parameter	Config Value	Description
	criterion	'gini'	Function to measure the quality of a split.
	max_features	'sqrt'	Maximum number of features considered for splitting a node.
	random_state	42	Seed used by the random number generator for reproducibility.
LR	solver	'lbfgs'	Optimization algorithm suitable for small-to-medium datasets.
	max_iter	100	Maximum number of iterations taken for the solvers to converge.
	C	1.0	Inverse of regularization strength (standard value).
General	test_size	0.2	Proportion of the dataset included in the test split (20%).
	class_weight	None	No weights assigned to classes (intentionally set to test baseline performance on imbalanced data).

2.5. Model Evaluation and Baseline Considerations

To ensure the reproducibility of the study, hyperparameter configurations were fixed, and resampling techniques (e.g., SMOTE) or cost-sensitive learning were intentionally excluded from the experimental design. This decision was made to establish a baseline diagnostic of how standard machine learning algorithms perform under the conditions of real-world, imbalanced social media data. Retaining the natural class distribution allows the study to empirically demonstrate the accuracy paradox, where models may achieve high accuracy simply by overfitting to the majority class. This highlights the importance of advanced techniques for handling imbalanced data in future operational implementations.

3. RESULTS AND DISCUSSION

3.1 Data Characteristics and Sentiment Labeling

The data used in this study was sourced from the social media platform Twitter (X) and collected using web crawling techniques [24]. A total of 1,002 tweets were successfully collected manually on November 27, 2025. To ensure ground truth validity, sentiment labelling was performed manually by three independent annotators, a method chosen to

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

better capture sarcasm and context compared to automated lexicons. The class distribution reveals a severe imbalance, with Positive sentiment dominating (95.3%) and Negative sentiment comprising only 4.7%. Table 2 illustrates the labelling criteria used.

Table 2. Labeling Criteria

Sample Data	Full Text (Indonesia)	Sentiment
0	Program ini sangat membantu rakyat!	Positive
1	Anggaran MBG terlalu besar dan tidak efisien.	Negative
...	...	...
1001	Prabowo dan Gibran memang pro rakyat.	Positive
1002	Kualitas gizinya perlu diperhatikan lebih lanjut.	Negative

3.2 Data Preprocessing

Before entering the main data processing stage, the data preprocessing stage is an important step that aims to collect and analyze raw data from platforms such as Twitter and convert it into high-quality, structured data. This process begins with text cleaning, which means removing "noise" elements that are not relevant to sentiment analysis [21]. To avoid interference with the classification process, elements such as URLs, usernames (mentions), hashtags, numbers, emoticons, and excessive punctuation are removed at this stage [13].

Once the data is free of distracting characters, case folding is used to convert the text format. In this process, every letter in the document is converted to lowercase [17]. To ensure data consistency, this step is crucial to ensure that the system treats words that begin with capital letters (such as "Not satisfied") and lowercase letters (such as "not satisfied") as a single entity. To improve data readiness, this research methodology also applies tokenization, normalization, stop word removal, and stemming or lemmatization [22].

Table 3. Sample of Data Preprocessing Results (Indonesia Text)

cleaned_text	case_folded_text	Normalized	final_text
Program ini sangat membantu rakyat	program ini sangat membantu rakyat	program ini sangat membantu rakyat	program bantu rakyat
Anggaran MBG terlalu besar dan tidak efisien	anggaran mbg terlalu besar dan tidak efisien	anggaran mbg terlalu besar dan tidak efisien	anggar mbg efisien

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

cleaned_text	case_folded_text	Normalized	final_text
Lumayanlah tidak terlalu bagus tidak terlalu buruk	lumayanlah tidak terlalu bagus tidak terlalu buruk	lumayanlah tidak terlalu bagus tidak terlalu buruk	lumayan bagus buruk
Prabowo dan Gibran memang pro rakyat	prabowo dan gibran memang pro rakyat	prabowo dan gibran memang pro rakyat	prabowo gibran pro rakyat
Kualitas gizinya perlu diperhatikan lebih lanjut	kualitas gizinya perlu diperhatikan lebih lanjut	kualitas gizinya perlu diperhatikan lebih lanjut	kualitas gizi perhati

3.3 TF-IDF Implementation

The feature extraction phase involves using the Term Frequency-Inverse Document Frequency (TF-IDF) method to transform the text dataset into a numerical vector representation. This process is very important because machine learning algorithms can only process data in a mathematical format [13]. Figure 5 shows a sample weighting matrix that has been created. This matrix has documents (tweets) on each row (indicated by rows 0 to 4) and word (token) weights on each column (such as *adu*, *bangsa*, program, and indonesia). The numerical values that appear, such as 0.353553 or 0.420669, indicate the importance or relevance of the word in the specific document and the overall data corpus, as shown in Figure 4.

```
        adu    bangsa     domba   dukung      emas   gampang  generasi  \
0  0.353553  0.000000  0.353553  0.00000  0.000000  0.353553  0.000000
1  0.000000  0.447214  0.000000  0.00000  0.000000  0.000000  0.000000
2  0.000000  0.000000  0.000000  0.00000  0.000000  0.000000  0.000000
3  0.000000  0.000000  0.000000  0.57735  0.000000  0.000000  0.000000
4  0.000000  0.000000  0.000000  0.00000  0.420669  0.000000  0.420669

     goblok  indonesia      kebo  ...  presiden   program    tangan   tentara  \
0  0.353553   0.000000  0.353553  ...  0.000000  0.000000  0.000000  0.000000
1  0.000000   0.000000  0.000000  ...  0.447214  0.000000  0.447214  0.000000
2  0.000000   0.000000  0.000000  ...  0.000000  0.000000  0.000000  0.362651
3  0.000000   0.000000  0.000000  ...  0.000000  0.000000  0.000000  0.000000
4  0.000000   0.420669  0.000000  ...  0.000000  0.420669  0.000000  0.000000

     ternak   tolong      urus     usut     wujud        ya
0  0.353553  0.00000  0.000000  0.00000  0.000000  0.000000
1  0.000000  0.00000  0.000000  0.00000  0.000000  0.447214
2  0.000000  0.00000  0.362651  0.00000  0.000000  0.000000
3  0.000000  0.57735  0.000000  0.57735  0.000000  0.000000
4  0.000000  0.00000  0.000000  0.00000  0.420669  0.000000
```

Figure 4. TF-IDF Data

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.4 Classification Model

3.4.1 Performance Evaluation

The RF algorithm achieved an overall accuracy of 77% (0.77), but the results reveal a severe performance bias stemming from data imbalance. The model demonstrated excellent performance on the majority class, Positive (2), with a near-perfect Recall of 0.99 and a strong F1-score of 0.86 (150 True Positives out of 151 support points). In stark contrast, the model failed completely to predict the Negative (0) class, recording a Recall, Precision, and F1-score of 0.00 (from 7 support points). The Neutral (1) class also performed very poorly, with a Recall of 0.07 and an F1-score of 0.13 (from 42 support points). Consequently, the low weighted F1-score of 0.68 indicates that despite the high accuracy, the model is unacceptable for effective multi-class sentiment analysis as it essentially ignores the minority classes [17], [13], [23].

The evaluation results of the LR model show an overall accuracy of 76%, but in-depth analysis reveals critical failures in handling the minority class due to severe data imbalance. This model shows a complete inability to identify Negative tweets (0), with Precision, Recall, and F1-score all at 0.00, and only successfully identifies 2 out of 42 Neutral (1) data points. Conversely, the model is highly dominant in the Positive class (2), with a Recall of 0.99, where 150 out of 151 Positive data points are correctly predicted. This phenomenon confirms that LR has a strong bias toward the majority class, as reflected by the very low macro-average F1-score (0.32), making the model invalid for effectively distinguishing sentiment in this dataset [24], [25], [26]. Figure 5 to 7 are performance evaluation of RF and LR.

	precision	recall	f1-score
0	0.00	0.00	0.00
1	0.75	0.07	0.13
2	0.77	0.99	0.86
accuracy			0.77
macro avg	0.51	0.35	0.33
weighted avg	0.74	0.77	0.68

Random Forest

	precision	recall	f1-score
0	0.00	0.00	0.00
1	0.50	0.05	0.09
2	0.77	0.99	0.86
accuracy			0.76
macro avg	0.42	0.35	0.32
weighted avg	0.68	0.76	0.67

Logistic Regression

Figure 5. Training Performance

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

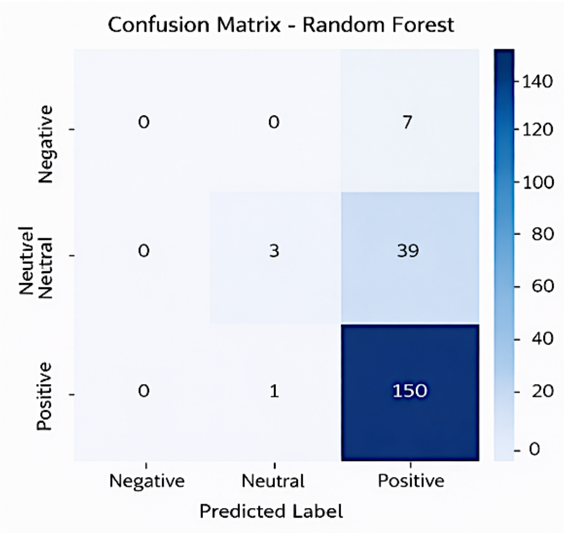

Figure 6. RF Algorithm Prediction Label

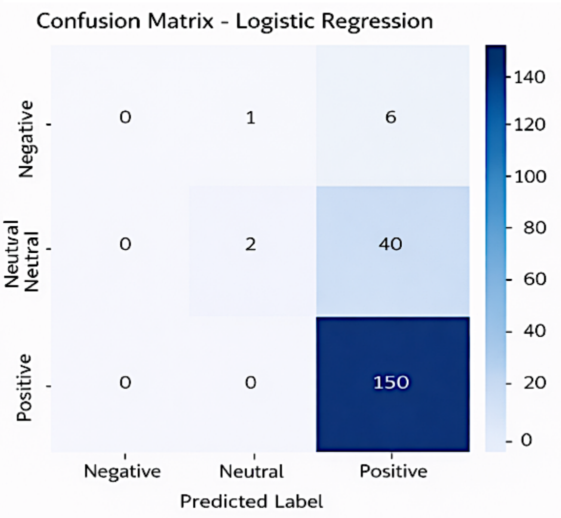

Figure 7. LR Algorithm Prediction Label

3.4.2 Comparative Evaluation Algorithm

At this stage, a comparative test was conducted to evaluate the performance of the RF and LR algorithms in classifying sentiment using the Testing Set. Based on the data presented in the table, RF demonstrated marginally superior performance with an accuracy rate of 0.77, outperforming LR, which achieved 0.76 [27] [28]. The comparison reveals critical failures shared by both models. Both algorithms exhibit an extreme bias toward the majority Positive class, recording a near-perfect Recall of 0.99 and a strong F-1 score of 0.86 for Positive sentiment (RM: Precision 0.77; LR: Precision 0.77) [29]. In stark contrast, both RF and LR failed completely to recognize the Negative class, recording 0.00 for Precision, Recall, and F-1 score [30]. Both models also performed poorly on the Neutral class (RM Recall 0.07, LR Recall 0.05 [28]). This phenomenon indicates that the high overall accuracy of both models is unreliable and is primarily driven by the ability to correctly predict the majority (Positive) class, confirming that both algorithms are severely impacted by data imbalance [28]. Therefore, referencing the Macro F1-score and Balanced Accuracy is essential to expose the models' lack of generalization.

Table 2. Algorithm Performance Comparison

	Accuracy	Precision			Recall			F-1		
		Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
RM	0.77	0.77	0.75	0.00	0.99	0.07	0.00	0.86	0.13	0.00
LR	0.76	0.77	0.50	0.00	0.99	0.05	0.00	0.86	0.09	0.00

The comparison graph visualization shows the superiority of the RF algorithm with an accuracy of 77%, outperforming LR which recorded 76%, as shown in Figure 8. This performance difference proves that the ensemble learning architecture in RF is more effective in capturing complex linguistic patterns and noise in social media data compared to the linear approach of LR, making it a more reliable model for sentiment classification of the MBG program [19] [13] [31]. The comparison graph (Figure 7) highlights RF's slight superiority, attributed to its ensemble architecture which handles noise better than LR's linear approach, though both remain compromised by the imbalance.

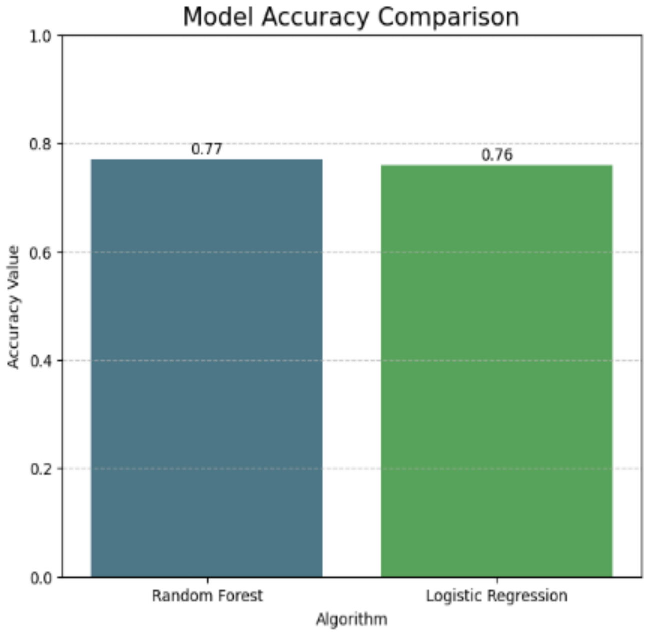

Figure 8. RF and LR of Algorithm Accuracy

3.5 Wordcloud Preprocessing

The Word Cloud visualization of the MBG program reveals a sharp contrast in public sentiment, with Negative sentiment driven by practical concerns regarding implementation and budget efficiency. Dominant words such as "program," "healthy," "nutrition," "economy," and "price" reflect doubts about nutritional quality, cost management, and potential market disruptions, thus questioning the promised health and well-being benefits. On the other hand, Positive sentiment focuses on political support and emotional expectations regarding program outcomes, characterized by words such as "free," "lunch," "Prabowo," "Gibran," "people," and "children," highlighting enthusiasm for promises of basic needs fulfilment and trust in the implementing political figures.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Therefore, this visualization effectively separates the discourse between operational concerns (Negative) and idealistic expectations and political support for the goal of societal well-being (Positive), as shown in Figure 9 and 10 [5], [17], [32].

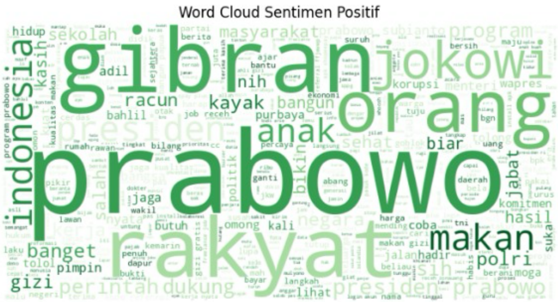

Figure 9. Wordcloud Positive

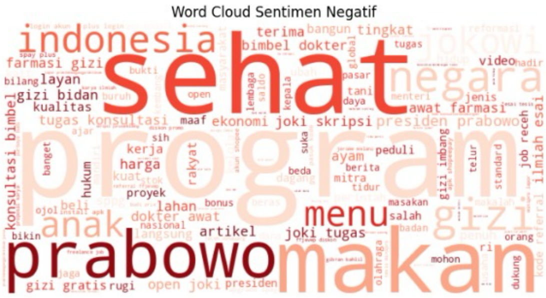

Figure 10. Wordcloud Negative

3.6 Discussion

A critical finding in this study is the "accuracy paradox," where a seemingly high accuracy rate of 77% masks the model's complete failure to detect Negative sentiment. This phenomenon is particularly concerning given the highly imbalanced nature of the dataset, with Positive sentiment dominating the majority class. The model's ability to predict the Negative class was non-existent, as reflected by the Recall, Precision, and F1-score all being 0.00 for the Negative sentiment. This performance bias highlights the limitations of traditional evaluation metrics, such as accuracy, in the context of imbalanced data, where the model's high performance on the majority class overshadows its inability to generalize to minority classes.

The low Macro-Average F1-score of 0.33 further reflects the poor performance on the minority class and underscores the necessity of implementing resampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or Random Undersampling in future sentiment analysis projects. These techniques are vital to balance the representation of minority classes and mitigate the bias towards the majority class in machine learning models. Without these interventions, models risk high accuracy scores while providing misleading results that fail to capture the true sentiment across all categories.

The comparison between Logistic Regression (LR) and Random Forest (RF) algorithms highlights the limitations of linear decision boundaries when applied to high-dimensional,

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

noisy data. Logistic Regression, with its linear decision boundary, struggled significantly with the text data's complexity, failing to detect Negative sentiment entirely. On the other hand, Random Forest—an ensemble method that builds multiple decision trees and averages their outputs—showed marginally better performance. RF's ability to capture non-linear relationships between features allowed it to slightly overcome the limitations of LR, offering improved robustness in detecting Positive sentiment. However, without class weighting or advanced resampling techniques, even RF could not entirely mitigate the negative impacts of the 95% imbalance between classes.

While both models demonstrated dominance in predicting Positive sentiment (with nearly perfect Recall values), their failure to detect Negative sentiment suggests that class imbalance heavily influences the models' ability to generalize. This confirms that more sophisticated strategies are necessary to develop truly effective models for sentiment analysis in imbalanced datasets, especially in the context of public opinion about governmental programs.

To address the issues identified in this study, future research should explore the following advanced techniques:

1) XGBoost: This algorithm utilizes boosting to iteratively correct errors made by previous models, which has shown to be highly effective in handling imbalanced datasets. Boosting can help improve the model's ability to predict minority classes by focusing learning efforts on difficult-to-classify samples.
2) Support Vector Machine (SVM): By using non-linear kernels, particularly the Radial Basis Function (RBF) kernel, SVM can handle the high-dimensional, sparse feature space typical in text data more effectively than linear models like LR. This method may offer better generalization and performance on minority sentiment classes.
3) Deep Learning (IndoBERT): Traditional methods like TF-IDF do not capture the rich semantic context of words, particularly slang and nuanced sentiment. Deep learning models such as IndoBERT (a pre-trained BERT model for Indonesian language) are capable of understanding the deeper meaning and context of phrases, which is essential for detecting subtle negative sentiments. IndoBERT's ability to capture contextual information would likely outperform TF-IDF in tasks where semantic understanding is crucial.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4) Resampling and Cost-Sensitive Learning: To prevent model bias toward the majority class, integrating resampling techniques or cost-sensitive learning could help to adjust for class imbalance. This would ensure that the model is equally sensitive to all sentiment classes, particularly in contexts like governmental program analysis, where understanding public criticism (Negative sentiment) is just as crucial as identifying support (Positive sentiment).

The findings from this study offer several important insights for future governmental sentiment analysis initiatives. Given the widespread public debate over programs such as Makan Bergizi Gratis (MBG), accurately classifying public sentiment can provide invaluable feedback to policymakers and help ensure that the program's operations align with public expectations. However, relying solely on traditional machine learning models without addressing data imbalance issues would likely lead to misleading conclusions about public opinion, especially if models are evaluated using accuracy alone.

Policymakers should therefore prioritize the development of more robust sentiment analysis models, incorporating the techniques outlined above, to better capture the nuances of public sentiment and guide future decision-making. Further research into more advanced algorithms and data handling techniques will be essential to improve the effectiveness of sentiment analysis systems in this area.

4 CONCLUSION

This study analyzed public sentiment toward the Makan Bergizi Gratis (MBG) program by comparing Random Forest (RF) and Logistic Regression (LR) algorithms using 1,002 tweets from the X platform. RF slightly outperformed LR, achieving an accuracy of 77%, while LR recorded 76%. However, both models suffered from severe data imbalance, with Positive sentiment dominating 95.3% of the dataset, leading to a critical failure in predicting Negative sentiment (0), with Recall, Precision, and F1-score all at 0.00. Future research should prioritize data balancing techniques (e.g., SMOTE, Random Undersampling) and incorporate contextual embeddings like IndoBERT to better capture sarcasm and subtle negative sentiments. From a policy perspective, Negative sentiment—often tied to budget efficiency and food safety—should not be overlooked, as it provides crucial feedback for improving program accountability and implementation. RF's ensemble learning approach proved more effective than LR's linear method, capturing complex linguistic patterns in noisy data, making it a more reliable choice for sentiment classification in social media discussions. The visualizations also highlight the contrast between Negative concerns and Positive political support, offering valuable insights for stakeholders to evaluate program effectiveness.

REFERENCES

[1] H. Rahman, M. Rahmah, dan N. Saribulan, "Efforts to Address Stunting in Indonesia: Bibliometric and Content Analysis (Upaya Penanganan Stunting Di Indonesia Analisis Bibliometrik dan Analisis Konten)," *J. Ilmu Pemerint. Suara Khatulistiwa*, vol. 8, no. 01, hal. 44–59, 2024, doi: 10.33701/jipsk.v8i1.3184.

[2] A. D. Arieffiani and M. R. L. Ekowanti, "Evaluating local government policy innovations: A case study of Surabaya's efforts in combating stunting and enhancing public health services quality," *Jurnal Bina Praja*, vol. 16, no. 1, pp. 1-20, 2024, doi: 10.21787/jbp.16.2024.1-20.

[3] A. Kiftiyah, F. A. Palestina, F. U. Abshar, dan K. Rofiah, "Makan Bergizi Gratis (MBG) Program in the Perspective of Social Justice and Socio – Political Dynamics," *Jurnal Keindonesiaan*, vol. 05, no. 1, hal. 101–112, 2025, doi: 10.52738/pjk.v5i1.726.

[4] F. Fatkhurrohman, B. I. Nugroho, and N. Fadillah, " Sentiment Analysis of the Indonesian Government's Free Nutritional Meal Program Through Twitter Using the SVM Method (Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM)," RIGGS J. Artif. Intell. Digital Bus., vol. 4, no. 3, pp. 3906-3917, 2025, doi: 10.31004/riggs.v4i3.2533.

[5] E. Triningsih, M. Afdal, I. Permana, dan N. Evrilyan Rozanda, "Sentiment Analysis of the Free Nutritional Meal Program Using Machine Learning Algorithms on Social Media X (Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X)," *Technol. Sci.*, vol. 6, no. 4, hal. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[6] T. V. S. Krishna, T. S. R. Krishna, S. Kalime, C. V. M. Krishna, S. Neelima, dan R. R. Pbv, "A novel ensemble approach for Twitter sentiment classification with ML and LSTM algorithms for real-time tweets analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 34, no. 3, hal. 1904–1914, Jun 2024, doi: 10.11591/ijeecs.v34.i3.pp1904-1914.

[7] S. U. Hassan, J. Ahamed, dan K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustain. Oper. Comput.*, vol. 3, hal. 238–248, Jan 2022, doi: 10.1016/j.susoc.2022.03.001.

[8] A. Miftahusalam, A. Febby Nuraini, A. A. Khoirunisa, dan H. Pratiwi, "Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Twitter Sentiment Analysis of Public Opinion on the Elimination of Honorary Staff (Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer)," *in Seminar Nasional Official Statistics*, vol. 2022, no. 1, pp. 563-572, Nov. 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.

[9] R. Jader dan S. Aminifar, "Predictive Model for Diagnosis of Gestational Diabetes in the Kurdistan Region by a Combination of Clustering and Classification Algorithms: An Ensemble Approach," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9749579.

[10] Intan Permata dan Esther Sorta Mauli Nababan, "Application of Game Theory in Determining Optimum Marketing Strategy In Marketplace," *J. Ris. Rumpun Mat. dan Ilmu Pengetah. ALAM*, vol. 2, no. 2, hal. 65–71, Jul 2023, doi: 10.55606/jurrimipa.v2i2.1336.

[11] V. Agresia dan R. R. Suryono, "Comparison of SVM, Naïve Bayes, and Logistic Regression Algorithms for Sentiment Analysis of Fraud and Bots in Purchasing Concert Ticket," *J. Inovtek Polbeng*, vol. 10, no. 2, hal. 2025, 2025, doi: 10.35314/npyfdh47.

[12] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, dan M. Hamdi, "Predictive Analysis Of Heart Diseases With Machine Learning Approaches," *Malaysian J. Comput. Sci.*, vol. 2022, no. Special Issue 1, hal. 132–148, 2022, doi: 10.22452/mjcs.sp2022no1.10.

[13] I. Hermansyah dan M. S. Hasibuan, "Sentiment Analysis of Free Nutritious Meal Programme on Social Media X using Linear Regression and Random Forest Algorithms," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 13, no. 1, hal. 55–68, 2025, doi: 10.33558/piksel.v13i1.10633.

[14] S. A. H. Bahtiar, C. K. Dewa, dan A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *J. Inf. Syst. Informatics*, vol. 5, no. 3, hal. 915–927, Agu 2023, doi: 10.51519/journalisi.v5i3.539.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[15] A. A. Rohman, A. G. Trisnapradika, dan K. Kunci, "Perbandingan Algoritma NBC, SVM, Logistic Regression untuk Analisis Sentimen Terhadap Wacana KaburAjaDulu di Media Sosial X," *Technol. Sci.*, vol. 7, no. 1, hal. 169–178, 2025, doi: 10.47065/bits.v7i1.7261.

[16] Tundo dan D. N. Rachmawati, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, 2024, doi: 10.35870/jimik.v5i3.978.

[17] M. Azhari dan P. Parjito, "Analisis Sentimen Opini Publik Program Makan Siang Gratis dengan Random Forest Pada Media," *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, hal. 1932–1942, 2024, doi: 10.47065/bits.v6i3.6423.

[18] H. Junianto, R. E. Saputro, B. A. Kusuma, D. Intan, dan S. Saputra, "Comparison Of Logistic Regression And Random Forest In Sentiment Analysis Of Disdukcapil Application Reviews," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 6, hal. 1539–1547, 2024, doi: 10.52436/1.jutif.2024.5.6.1802.

[19] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis," *KDIR*, vol. 1, pp. 525-531, 2019.

[20] Z. S. Munmun, S. Akter, dan C. R. Parvez, "Machine Learning-Based Classification of Coronary Heart Disease: A Comparative Analysis of Logistic Regression, Random Forest, and Support Vector Machine Models," *OALib*, vol. 12, no. 03, hal. 1–12, 2025, doi: 10.4236/oalib.1113054.

[21] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, dan M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, hal. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[22] M. E. Kabir, "Topic and sentiment analysis of responses to Muslim clerics' misinformation correction about COVID-19 vaccine: Comparison of three machine learning models," *Online Media Glob. Commun.*, vol. 1, no. 3, hal. 497–523, 2022, doi: 10.1515/omgc-2022-0042.

[23] S. Das, K. Bhattacharyya, dan S. Sarkar, "Performance Analysis of Logistic Regression, Naive Bayes, KNN, Decision Tree, Random Forest and SVM on Hate Speech Detection from Twitter," *Int. Res. J. Innov. Eng. Technol.*, vol. 07, no. 03, hal. 07–03, 2023, doi: 10.47001/irjiet/2023.703004.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[24] S. Bhushan, M. Alshehri, I. Keshta, A. K. Chakraverti, J. Rajpurohit, dan A. Abugabah, "An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition," *Electron.*, vol. 11, no. 6, Mar 2022.

[25] M. G. Hussain, B. Sultana, M. Rahman, dan M. R. Hasan, "Comparison analysis of Bangla news articles classification using support vector machine and logistic regression," *Telkomnika (Telecommunication Comput. Electron. Control*, vol. 21, no. 3, hal. 584–591, 2023, doi: 10.12928/TELKOMNIKA.v21i3.23416.

[26] M. Alfyando, Anggraeny Fetty Tri, dan SihanantoAndreas Nugroho, "Perbandingan Algoritma Random Forest dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store," *J. Sist. Inf. dan Ilmu Komput.*, vol. 2, no. 1, hal. 77–86, 2023, doi: 10.59581/jusiik-widyakarya.v2i1.2262.

[27] G. A. Saputri dan D. Alita, "Analisis Sentimen Twitter Terhadap Pemindahan Ibu Kota Negara Menggunakan Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, hal. 213–223, Des 2024, doi: 10.30591/jpit.v9i3.6612.

[28] R. Saputra dan F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, hal. 411–419, Jul 2024, doi: 10.47233/jteksis.v6i3.1378.

[29] Asro, Sudaryono, dan A. Sulaiman, "Analisis Sentimen tentang Transformasi Program Makan Siang menjadi Makan Bergizi Gratis menggunakan Logistik Regression pada laman Youtube," *J. ICT Inf. Commun. Technol.*, vol. 24, no. 1, hal. 87–94, 2024, doi: 10.36054/jict-ikmi.v24i1.257.

[30] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, dan 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, doi: 10.23960/jitet.v12i3.4902.

[31] M. W. Arif, " Sentiment Analysis of the Free Nutritional Meal Policy on Social Media Using Python TextBlob-Based Natural Language Processing in Indonesia (Analisis Sentimen Kebijakan Makan Bergizi Gratis di Media Sosial Menggunakan Natural Language Processing Berbasis Python TextBlob di Indonesia)," *J. Pendidik. dan Teknol. Indonesia (JPTI)*, vol. 5, no. 9, pp. 2463-2471, 2025., doi: 10.52436/1.jpti.931.

[32] Y. Al Amrani, M. Lazaar, dan K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Comput. Sci.*, vol. 127, hal. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.