International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 07, Issue 03, August 2026
e-ISSN: 2745-9659
https://ijcis.net/index.php/ijcis/index

Public Sentiment Analysis of Indonesia's Free Nutritious Meals (MBG) Program Using Lexicon-Based Labeling, SMOTE and Machine Learning

1st Kharisma Wiati Gusti, 2nd Anni Alvionita Simanjuntak
[1] Informatics, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jakarta
[2] Law, Faculty of Law, Universitas Pembangunan Nasional "Veteran" Jakarta. Jakarta, Indonesia
[1] kharismawiatigusti@upnvj.ac.id, [2] annialvionita@upnvj.ac.id

Abstract—The Free Nutritious Meals (MBG) Program is a strategic policy of the Indonesian government aimed at improving national nutrition and has elicited diverse public reactions on social media. This study analyses public sentiment toward the MBG program using 11,003 cleaned public comments. Because the dataset lacked sentiment labels, a lexicon-based labelling approach using the Indonesian InSet Lexicon was applied to automatically classify comments as positive, negative, or neutral. The labelled data were then divided into 80% training data and 20% testing data using stratified sampling, and TF-IDF was used for feature extraction. Because the resulting class distribution was imbalanced (61.9% negative, 20.3% positive, and 17.8% neutral), the Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training data. Four classification algorithms (Naive Bayes, Support Vector Machine, Logistic Regression, and Random Forest) were trained and compared with and without SMOTE. The results show that SMOTE improved the macro-average F1-score across all models, particularly Naive Bayes (from 0.42 to 0.67). At the same time, SVM remained the best-performing model, achieving 84.83% accuracy and a macro F1-score of 0.81 after SMOTE, with a slight accuracy trade-off for substantially better recall in the neutral and positive minority classes. These findings indicate that combining lexicon-based labelling with SMOTE and classical machine learning provides a more balanced and reliable tool for monitoring public sentiment toward government policy programs such as MBG.

Keywords: sentiment analysis; Free Nutritious Meals; lexicon-based labeling; SMOTE; machine learning

I. INTRODUCTION

Indonesia's Free Nutritious Meals (MBG) Program is one of the Indonesian government's flagship policies, launched on January 6, 2025 [1] to address stunting and malnutrition among schoolchildren, toddlers, and pregnant and breastfeeding women [2]. As a large-scale policy under intense public scrutiny, the MBG program has produced comprehensive discussion on social media, ranging from appreciation of its potential health benefits to criticism concerning food poisoning incidents, food quality, the large budget allocation perceived as not yet matched by transparency and effective oversight, and concerns about the program's funding sources [3]. Understanding public sentiment toward this program is important for policymakers because it provides large-scale, real-time feedback that can complement formal program evaluations. Previous studies of MBG sentiment analysis algorithms such as Naive Bayes, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and IndoBERTweet to data collected from platforms such as X and YouTube, with varying accuracy depending on the methods and datasets used. Several studies have reported inconsistent findings. Studies on Platform X tend to find a predominance of positive sentiment, such as [4], which recorded 76.2% positive sentiment using Naive Bayes and ADASYN, and [5], which recorded 68.8% positive sentiment using Support Vector Machine (SVM). The LSTM approach in [6] produced results tending toward positive and neutral sentiment. However, the precision, recall, and F1-score for the negative class were all 0 due to class imbalance. In contrast, research on

YouTube comments [7] and research on Platform X using a different data-collection period [8] found a predominance of negative sentiment: 80.78% using SVM in the former and a negative majority using Naive Bayes with a self-training approach in the latter. These differences are thought to be influenced by the platform, data-collection period, search keywords, and labelling method used.

From a technical perspective, these studies have also consistently encountered imbalanced data across sentiment classes, which can cause models to be biased toward the majority class. Various handling techniques have been applied, including Adaptive Synthetic Sampling (ADASYN) [4], random undersampling [5], the class_weight="balanced" parameter [7], and a self-training approach for automatically labeling unlabeled data [8], with model accuracy ranging from 78% to 97.4%. These variations in results and approaches demonstrate the need for further research to more comprehensively confirm public sentiment trends toward MBG while evaluating the effectiveness of data-balancing techniques and classification algorithms best suited to Indonesian-language social media data [5][7].

Based on these studies, this research aims to automatically label an unlabeled MBG comment dataset using a lexicon-based approach, apply the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance in the training data, compare the performance of four classical machine learning algorithms with and without SMOTE, and analyse the resulting sentiment distribution and its implications for public policy communication.

II. RESEARCH METHODS

This study uses a dataset 11,375 public comments related to the MBG program [9], which underwent text preprocessing (cleaning, case folding, normalization, tokenization, and stopword removal) before analysis.

After empty and duplicate entries were removed, 11,003 comments remained for further processing. Because the dataset had no sentiment labels, an automatic lexicon-based labeling technique was applied using the InSet lexicon, an Indonesian sentiment word list containing positive and negative polarity scores [10] [11]. The sentiment score of each comment was calculated as the sum of the polarity weights of its constituent words; comments with a positive total score were labeled positive, those with a negative total score were labeled negative, and those with a score of zero were labeled neutral.

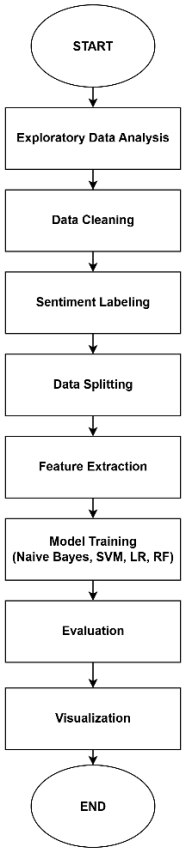

Figure 1. Research Stages

The labeled dataset was then divided into training (80%) and testing (20%) subsets using stratified random sampling to preserve the original class proportions. Text features were extracted using Term Frequency–Inverse Document Frequency (TF-IDF), with a maximum of 5,000 features. Because the training data were imbalanced (approximately 62% negative, 20% positive, and 18% neutral), the Synthetic Minority Over-sampling Technique (SMOTE) [12] was applied only to the training data using the implementation from the imbalanced-learn library [13] [14]. This generated synthetic samples for the minority classes until all classes were balanced; the testing data were left unchanged to maintain a realistic evaluation of real-world performance. Four supervised classification algorithms were then trained and compared, both with and without SMOTE: Multinomial Naive Bayes, Linear

Support Vector Machine (SVM), Logistic Regression, and Random Forest, all implemented using the scikit-learn library [13] [14]. Model performance was evaluated using accuracy and macro-average F1-score, and the confusion matrix of the best-performing model was further analyzed to understand its class-level behavior.

III. RESULTS AND ANALYSIS

This section presents the distribution of sentiment labels obtained through lexicon-based labeling, the effect of SMOTE on classification performance, and a more detailed analysis of the best-performing model.

3.1 Sentiment Label Distribution

Applying the InSet lexicon-based labeling procedure to the 11,003 cleaned comments produced the following sentiment-class distribution:

a. Negative: 6,813 comments (61.9%)
b. Positive: 2,232 comments (20.3%)
c. Neutral: 1,958 comments (17.8%)
d. Total comments analysed: 11,003

As shown in Figure 2, negative sentiment clearly dominates public discussion of the MBG program. This imbalance is consistent with previous research reporting that criticism related to food-safety incidents and budget transparency tends to generate more vocal reactions on social media than expressions of support. This finding directly motivated the use of SMOTE during model training, as explained in the following subsection.

3.2 Effect of SMOTE on Model Performance

SMOTE was applied only to the 80% training set (8,802 comments), synthetically balancing the negative, positive, and neutral classes before training; the 20% testing set (2,201 comments) retained its original imbalanced distribution to ensure a realistic evaluation.

Table 1. Testing Accuracy Before and After SMOTE

No	Model	Baseline Accuracy	SMOTE Accuracy
1	Naive Bayes	67,20%	73,74%
2	SVM	85,19%	84,83%
3	Logistic Regression	79,78%	81,37%
4	Random Forest	77,92%	77,33%

Table 1 compares the testing accuracy of each model before and after SMOTE. Naive Bayes benefited the most from SMOTE, with accuracy increasing from 67.20% to 73.74%, while SVM and Random Forest showed slight decreases in raw accuracy, from 85.19% to 84.83% and from 77.92% to 77.33%, respectively. However, when evaluated in Table 2 using macro-average F1-score, which assigns equal weight to all three classes, every model improved after SMOTE: Naive Bayes increased from 0.42 to 0.67, Logistic Regression from 0.71 to 0.77, Random Forest from 0.72 to 0.73, and SVM from 0.80 to 0.81. This indicates that SMOTE substantially improved each model's ability to correctly classify the neutral and positive minority classes, even in cases, even when overall accuracy decreased slightly, as illustrated in Figure 2.

Table 2. Macro F1 Before and After SMOTE

No	Model	F1-macro (Baseline)	F1-macro (SMOTE)
1	Naive Bayes	0.415410	0.675263
2	SVM	0.808074	0.806679
3	Logistic Regression	0.708128	0.772581
4	Random Forest	0.718766	0.733821

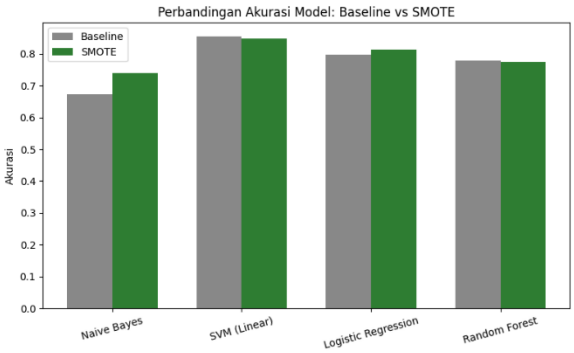

Figure 2. Comparison of model accuracy before (baseline) and after SMOTE.

3.3 Confusion Matrix Analysis of the Best Model

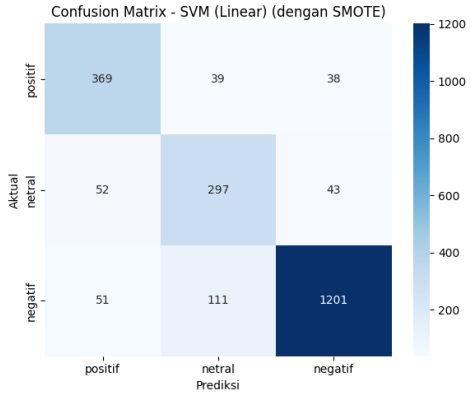

Figure 3. Confusion matrix of the best-performing model (Linear SVM) after SMOTE on the testing data.

Figure 3 presents the confusion matrix of the best-performing model after SMOTE, namely Linear SVM, which achieved 84.83% accuracy and a macro F1-score of 0.81 on the testing data. The model correctly classified 88% of negative comments, 83% of positive comments, and 76% of neutral comments, representing improved recall for the minority classes compared with the model before SMOTE, which correctly classified only 68% of neutral comments and 77% of positive comments. Most remaining misclassifications occurred between the neutral and negative classes, which is understandable given the linguistic similarity between mild criticism and genuinely neutral statements in informal Indonesian text.

3.4 Impact of SMOTE on Minority-Class Recall

In addition to aggregate accuracy and F1-score, class-level recall provides a clearer picture of SMOTE's benefits. Across the four models, the average recall for the neutral class increased from 58% before SMOTE to 74% after SMOTE, while the average recall for the positive class increased from 61% to 78%.

This improvement is particularly important for policy-monitoring applications, where failure to detect genuinely neutral or positive feedback can create a distorted and overly negative picture of public opinion, leading to misguided communication strategies.

3.5 Limitations

Despite the improvements achieved through SMOTE, several limitations of this study should be noted:

a. Synthetic Data Risk: SMOTE generates synthetic samples by interpolating among existing minority-class examples in the TF-IDF feature space, which can sometimes produce unrealistic feature combinations that do not correspond to natural language.
b. Lexicon Coverage: The InSet lexicon may not fully cover informal expressions, slang, or newly coined Indonesian words commonly used on social media, which can lead to labeling errors even before SMOTE is applied.
c. Generalizability: The dataset was collected from a specific period and source; therefore, the findings, including the effects of SMOTE, may not fully generalize to other platforms or later stages of program implementation.

3.6 Comment Sentiment Distribution

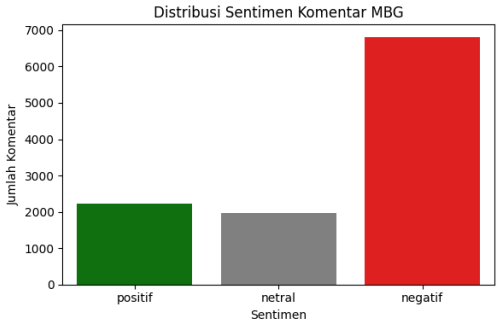

Figure 4. Distribution of sentiment labels in the analyzed public comments related to MBG.

Figure 4 illustrates the sentiment distribution obtained through lexicon-based labeling, confirming the predominance of negative sentiment discussed earlier.

a. Public sentiment toward the MBG program is overwhelmingly negative, driven by concerns about safety and transparency.
b. SMOTE improves the macro F1-score of every model, particularly Naive Bayes.
c. Linear SVM remains the most reliable model, balancing accuracy and class-level recall after SMOTE.

International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 07, Issue 03, August 2026
e-ISSN: 2745-9659
https://ijcis.net/index.php/ijcis/index

3.7 Visualization

a. Positive Sentiment

The sentiment data are depicted using a word cloud. It is dominated by words such as "presiden" (president), "rakyat" (people), "negara" (country), "Indonesia," "mahasiswa" (student), "allah," "mantap" (excellent), "semoga" (hopefully), "dukung" (support), "sehat" (healthy), and "prabowo." This pattern indicates comments expressing support, prayers, and positive hopes for the program or government, for example, phrases such as "semoga presiden sehat" (may the president stay healthy), "mantap pak" (excellent, sir), and "allah lindungi" (may Allah protect), among others. Religious words (Allah, swt) and optimistic words (mantap, semoga, dukung) are distinctive markers of this class.

Figure 5. Positive Sentiment Word Cloud

b. Neutral Sentiment

The neutral-sentiment word cloud is dominated by the words "indonesia," "mahasiswa" (student), "rakyat" (people), "tiyo," "presiden" (president), "pemerintah" (government), "kampus" (campus), "ugm," and "bem." Its content is more descriptive or factual, mentioning institutions (UGM, BEM = Student Executive Board) and the parties being discussed, without words carrying strong emotional weight. Interestingly, the word "bodoh" (stupid) also appears fairly prominently. This may reflect a limitation of lexicon-based labeling: a sentence containing a negative word that is offset by another word with an opposing score can yield a total score of zero and thus fall into the neutral category, even though its actual meaning is sarcastic or negative. This pattern is relevant for discussion in the Limitations section of this article.

Figure 6. Neutral Sentiment Word Cloud

c. Negative Sentiment

The negative-sentiment word cloud is dominated by the words "rakyat" (people), "negara" (country), "Indonesia," "mahasiswa" (student), "korupsi" (corruption), "kritik" (criticism), "koruptor" (corruptor), "terror" (terror), "takut" (afraid), "pejabat" (official), and "Prabowo." This clearly indicates comments criticising corruption, fear, and disappointment with officials or the government, consistent with our finding that concerns about the transparency and integrity of the program or government dominate negative sentiment.

Interestingly, the word "tiyo" appears prominently in all three sentiment classes (positive, neutral, and negative), strongly indicating that "Tiyo" is the name of a person or figure (a student) discussed across sentiments rather than a word that inherently carries polarity. This is a common pattern in lexicon-based sentiment analysis: personal names are usually lexically neutral and can therefore appear in all classes depending on sentence context. The appearance of words such as "ugm," "bem," "kampus" (campus), and "mahasiswa" (student) in all classes also indicates that much of the discussion involves university students or campus activists.

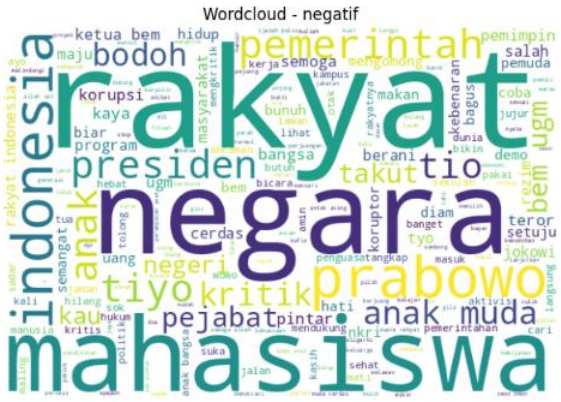

Figure 7. Negative Sentiment Word Cloud

3.8 Error Analysis

Manual examination of a sample of misclassified comments showed that most remaining errors, even after SMOTE, involved short comments with limited contextual information, for which lexicon-based scoring alone was insufficient to determine polarity reliably.

IV. DISCUSSION

A comparison of the baseline and SMOTE results shows that accuracy alone can be misleading when the data are imbalanced. For SVM and Random Forest, accuracy decreased slightly after SMOTE, whereas the macro F1-score increased. This means that model performance became more balanced across all classes, even though the number of correctly classified comments decreased slightly. In contrast, Naive Bayes improved on both metrics, indicating that this model was the most sensitive to the initial class imbalance.

Compared with related studies, the SVM accuracy obtained in this study was 84.83% after SMOTE. This result falls within the range commonly reported for classical machine learning approaches applied to MBG-related social media data (approximately 67–97%, depending on the algorithm and dataset), while transformer-based approaches such as IndoBERTweet have reported much higher accuracy (up to 97.7%) when combined with pseudo-labeling strategies. This gap indicates that although lexicon-based labeling combined with SMOTE and classical machine learning offers a practical, resource-efficient solution with better class balance, contextual language models remain a promising direction for future improvement. Key observations from this study include:

a. The predominance of negative sentiment (61.9%) reflects public concerns about food-poisoning incidents and budget-transparency issues surrounding the MBG program.
b. SMOTE improved the macro F1-score of every model, particularly Naive Bayes, from 0.42 to 0.67. This indicates that class imbalance concealed the model's poor performance on minority classes in the baseline results.
c. SVM remained the best-performing model after SMOTE, achieving 84.83% accuracy and a macro F1 Score of 0.81. This suggests that the sentiment classes in this dataset are largely linearly separable in the TF-IDF feature space.
d. SVM and Random Forest showed slight decreases in accuracy after SMOTE despite increases in macro F1-score, illustrating the common trade-off between accuracy and class balance that results from oversampling.
e. The average recall for the neutral class across all models increased from 58% to 74% after SMOTE, directly addressing the weakest aspect of the baseline models.
f. Lexicon-based labeling with InSet does not fully capture negation, sarcasm, and context-dependent sentiment, which can introduce noise that SMOTE amplifies by oversampling mislabeled minority examples.
g. Compared with related studies using Naive Bayes and SVM on similar MBG datasets, which reported accuracy ranging from approximately 67% to 97.4%, the results of this study fall within a comparable range even after accounting for the accuracy trade-off caused by SMOTE.
h. Studies using deep learning approaches such as IndoBERTweet has reported much higher accuracy (up to 97.67%), indicating room for improvement through contextual embeddings, potentially combined with class-balancing techniques such as SMOTE.
i. Classical machine learning combined with SMOTE remains a practical and resource-efficient option for sentiment-analysis tasks in policy monitoring, where balanced class performance is as important as overall accuracy.
j. Because SMOTE was applied only to the training data, the reported testing metrics reflect

performance on the original, imbalanced distribution of public comments rather than on an artificially balanced distribution.

k. These findings can inform government communication strategies by highlighting aspects of the MBG program (such as food safety and budget transparency) that require more targeted public messaging, while avoiding overreaction to seemingly uniform waves of negative sentiment.

V. CONCLUSION

This study applied a lexicon-based labeling approach using the InSet lexicon to automatically label 11,003 public comments related to the MBG program, followed by TF-IDF feature extraction, SMOTE-based class balancing of the training data, and a comparison of four classification algorithms. Linear SVM achieved the best overall performance, with 84.83% accuracy and a macro F1-score of 0.81 after SMOTE. SMOTE improved the macro-average F1-score of every model, most notably Naive Bayes, from 0.42 to 0.67, demonstrating that addressing class imbalance substantially improves minority-class detection despite slightly reducing raw accuracy in already strong models. Negative sentiment dominated public opinion (61.9%), reflecting concerns about food safety and program transparency. These results indicate that combining lexicon-based labeling, SMOTE, and classical machine learning is a practical, resource-efficient, and class-balanced approach for large-scale public sentiment monitoring and can serve as a decision-support tool for policymakers. Future research could explore manual validation of a subset of the labeled data, alternative resampling techniques such as ADASYN, aspect-based sentiment analysis, and transformer-based models such as IndoBERT to further improve classification accuracy and class balance.

REFERENCES

[1] S. Deny, "Program Makan Bergizi Gratis Resmi Dimulai Hari Ini 6 Januari 2025," Liputan6.com. Accessed: Aug. 07, 2026. [Online]. Available: https://www.liputan6.com/bisnis/read/5864557/program-makan-bergizi-gratis-resmi-dimulai-hari-ini-6-januari-2025

[2] M. I. Wibisono and S. Santosa, "Implementasi Program Makan Bergizi Gratis dalam Meningkatkan Ketahanan Gizi Pelajar di Indonesia," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 11973–11981, Jan. 2026, doi: 10.31004/riggs.v4i4.5365.

[3] H. Fidelya Ardhana and U. Negeri Jakarta Rini Riris Setyowati, "Analisis Implementasi Dan Dampak Pada Makanan Bergizi Gratis (MBG): Studi Pendekatan Transdisipliner," *Jurnal Jendela Pendidikan*, vol. 6, 2026.

[4] M. E. Zainuri, "ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA PLATFORM X DENGAN MENGGUNAKAN METODE NAIVE," Universitas Dinamika, Surabaya, 2026. Accessed: Aug. 07, 2026. [Online]. Available: https://repository.dinamika.ac.id/id/eprint/8387/1/21410100027-2026-UNIVERSITASDINAMIKA.pdf

[5] M. A. Firmansyah, A. Saeppani, I. Fadil, U. Sebelas, and A. Sumedang, "Analisis Sentimen Publik Program Makan Bergizi Gratis Menggunakan Support Vector Machine," *Jurnal Pustaka Nusantara Multidisplin*, vol. 3, no. 4, 2025.

[6] M. S. Arum and A. Hendrawan, "Analisis Sentimen Komentar YouTube Terhadap Program Makan Bergizi Gratis (MBG) Menggunakan Long Short-Term Memory(LSTM)," *KETIK:JurnalInformatika*, pp. 121–129, 2026, Accessed: Aug. 07, 2026. [Online]. Available: https://jurnal.faatuatua.com/index.php/KETIK/article/view/627/237

[7] B. Andianto and H. Februariyanti, "Analisis Opini Publik Terhadap Program Makan Bergizi Gratis (MBG) Pada Media Sosial Youtube Menggunakan Algoritma Support Vector Machine," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 13,

no. 2, pp. 111–124, 2026, [Online]. Available: http://jurnal.mdp.ac.id

[8] S. J. Rain *et al.*, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X DENGAN NAÏVE BAYES," 2026.

[9] H. Hidayah, "Indonesia MBG Sentiment Analysis ," Kaggle. Accessed: Aug. 07, 2026. [Online]. Available: https://www.kaggle.com/datasets/neevvv/indonesia-mbg-sentiment-analysis](https://www.kaggle.com/datasets/neevvv/indonesia-mbg-sentiment-analysis

[10] A. Okta, K. Adi, A. Sanjaya, A. B. Setiawan, and P. Korespondens, "Penerapan Inset Lexicon untuk Analisis Sentimen Penonton Video JKT48 di YouTube 1*," 2025.

[11] D. Pratiwi, N. Khoerani, S. Sari, and P. Korespondensi, "ANALISIS SENTIMEN TERHADAP KEBIJAKAN SUBSIDIPEMBELIAN KENDARAAN BERTENAGA LISTRIK DI INDONESIA MENGGUNAKAN PENDEKATAN INSET LEXICON DAN METODE SUPPORT VECTOR MACHINE," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 12, no. 6, pp. 1303–1314, 2025.

[12] K. W. Gusti, "KLASIFIKASI BENCANA ALAM PADA TWITTER MENGGUNAKAN NAÏVE BAYES, SUPPORT VECTOR MACHINE DAN LOGISTIC REGRESSION," *Technologia : Jurnal Ilmiah*, vol. 14, no. 4, p. 349, Oct. 2023, doi: 10.31602/tji.v14i4.11614.

[13] F. Febriansyah Putra, S. Fakhri Wibowo, and J. R. Zebua, "Prediksi Kelulusan Mahasiswa Menggunakan Random Forest Dan Smote Untuk Mengatasi Imbalanced Data," *Jurnal Ilmiah Sains Teknologi dan Informasi*, vol. 4, pp. 156–165, 2026, doi: 10.59024/jiti.v4i3.2225.

[14] R. H. Tanjung, F. Damayanti, and A. Zaki, "Optimasi Random Forest Melalui Feature Engineering dan SMOTE untuk Klasifikasi Kesehatan Mental," *Bulletin of Information Technology (BIT)*, vol. 7, pp. 204–212, 2026.