An Evaluation of SMOTE Effectiveness in Handling Class Imbalance in Public Opinion Data on the MBG Program

Nur Ghaniaviyanto Ramadhan [a,1,*], Azka Khoirunnisa [b,2]

[a, b] Department of Data Science, School of Computing, Telkom University, Bandung, Indonesia
[1] nuruer@telkomuniversity.ac.id*; [2] khoirunnisaazka@telkomuniversity.ac.id;
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Ensemble Learning
MBG Program
Imbalanced Data
Sentiment Analysis
SMOTE

ABSTRACT

The "Makan Bergizi Gratis" (MBG) Program is one of the strategic policies of the Government of Indonesia that reaps various opinions from the public, especially through social media. This study aims to classify public sentiment towards the MBG program with an ensemble learning-based machine learning approach, as well as evaluate the effectiveness of the SMOTE algorithm in dealing with class imbalance in opinion data. The dataset was collected from platform X (formerly Twitter) for the January–April 2025 period, totaling 4,374 tweets with label distributions: 1,783 positive, 1,634 negative, and 957 neutral. The preprocessing process includes data cleansing, normalization, stemming, and vectorization with TF-IDF. Five ensemble algorithms were used, namely Random Forest, AdaBoost, Bagging, Stacking, and Voting, tested in two scenarios: with and without the implementation of SMOTE. The results of the experiments showed that Random Forest provided the best and most consistent performance, with the F1-score increasing from 72.03% to 72.66% after the implementation of SMOTE. However, not all models benefit from SMOTE, such as Voting which experienced a drop in F1-score. These findings suggest that SMOTE is effective in increasing the sensitivity of the model to minority classes, but its success depends on the characteristics of the algorithm used. This study suggests the selective selection of balancing methods as well as the development of a more adaptive approach to handle unstructured opinion data.

1. Introduction

The MBG Program is one of the strategic initiatives launched by the Indonesian government to increase the nutritional intake of elementary school children through the provision of free healthy food [1]. Since its launch, this policy has provoked various reactions from the public, especially through social media platforms such as X (formerly Twitter). Social media is a rich and dynamic source of data to automatically analyze public opinion using the Natural Language Processing (NLP) approach [2].

Several previous studies have addressed sentiment analysis in the context of public policy and social issues. For example, Rahmatullah et al. [2] uses a Naïve Bayes algorithm to analyze sentiment towards the MBG program, but has not taken into account the problem of class imbalance in the data.

Research by Kedas et al. [3] dan Obiedat et al. [4] evaluate the use of SMOTE algorithms in sentiment classification tasks, particularly on deep learning and hybrid approaches, and show that synthetic oversampling techniques can improve minority class detection. However, these studies have not examined in depth the performance of SMOTE when combined with the ensemble learning model, especially on real and unstructured public opinion data, such as in the context of government policy evaluation.

Following up on these problems, various techniques have been developed to deal with data imbalances, one of which is the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE works by generating synthetic samples from minority classes through interpolation between data, so that the class distribution becomes more balanced [5]. This technique has been widely used in areas such as disease prediction, fraud detection, and risk evaluation, and has been proven to improve model performance on unbalanced data [6]. However, the application of SMOTE in the context of text-based public opinion data, especially in sentiment analysis on public policies such as the MBG program, is still very limited and rarely evaluated systematically.

Furthermore, it should be highlighted that the effectiveness of SMOTE is also strongly influenced by the classification algorithm used. Some ensemble learning approaches such as Random Forest, AdaBoost, Bagging, Stacking, and Voting are known to perform well in various classification tasks [7]. However, until now, there have not been many studies that have in-depth examined the effectiveness of SMOTE when combined with these models in social opinion data imbalance scenarios. In addition, no studies have been found that explicitly analyze the impact of SMOTE implementation on the trade-off between increased recall and the possibility of a decrease in precision or F1-score, especially in unstructured public opinion data.

Therefore, there is a research gap that needs to be filled through a comprehensive study of the combination of balancing techniques such as SMOTE and various ensemble models on public sentiment data. This gap is the basis of this research. Although SMOTE has been proven to be effective in various domains, its performance when combined with various ensemble algorithms such as Random Forest, AdaBoost, bagging, stacking, and voting on unbalanced data in the realm of public opinion is still not widely researched. In addition, there have not been many studies that have systematically evaluated the potential trade-off between increased recall and decreased accuracy or risk of overfitting due to synthetic data, especially in informal language data such as tweets.

Although the SMOTE algorithm itself is well established, the novelty of this study lies in its systematic evaluation across multiple ensemble learning models on unstructured and real-world public opinion data in the context of a national policy (MBG). Additionally, this study highlights the inconsistency of SMOTE's effectiveness across models**,** particularly exposing its trade-offs between recall and potential overfitting an aspect that has not been widely addressed in prior works, especially in Indonesian language sentiment analysis.

As a form of contribution to this gap, this study proposes a comprehensive evaluation of the SMOTE algorithm combined with five ensemble models in classifying public sentiment towards the MBG program. The novelty of this study lies in: (1) the application of SMOTE to unstructured public opinion data on ongoing national policies; (2) thorough comparative analysis between ensemble models; and (3) evaluation of the trade-off effects that occur on various classification metrics. This research contributes to the distribution of knowledge by providing empirical evidence that the effectiveness of SMOTE is selective depending on the algorithm used, so that it can be a guide in the study of unbalanced data-driven sentiment analysis, especially in the field of public policy.

2. Method

Fig. 1 is the flow of the process of proposing this research which begins with crawling tweet data through platform X where then the data is cleaned, stemming, and TF-IDF. This study also applies the SMOTE algorithm for handling unbalanced data on the tweet data obtained.

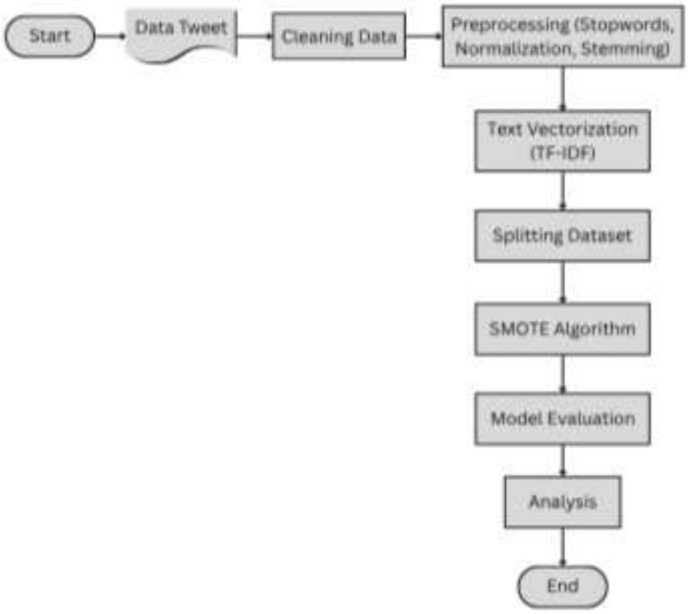

Fig. 1. Proposed Research Diagram

2.1. Dataset

The dataset used in this study was obtained through a data crawling process from social media platform X, with a data collection time range from January to April 2025. The crawling process is focused on tweets that contain keywords related to the MBG program, both in the context of support, criticism, and general discussion.

From the results of the crawling, 4374 tweets were collected that had gone through the initial verification and screening stage. The tweets were then manually labeled sentiment into three categories: positive, negative, and neutral. The distribution of data based on sentiment labels is as follows: Positive : 1783 tweets, Negative : 1634 tweets, and Neutral : 957 tweets This distribution shows a class imbalance**,** especially between positive and neutral classes, which has the potential to affect the performance of the classification model. Therefore, data balancing approaches such as SMOTE are implemented in the next stages to address these issues. Table 1 is a sample of tweet data and the labels used.

Table 1. Sample Dataset

Dataset	Label
Ini mbg programnya buat anak sekolah yg katanya biar fokus belajar sekaligus cegah stunting tp pendidikan sm kesehatan cuma jd prioritas pendukung itu gmn ceritanya wkwkwkwkwk	Negative
Pak mbg bisa dihapus aja kok kalo nyatanya ga bagus. Gausah gengsi wkwk	Negative
@BaseAnakFK Kok bisa sih pemikirannya makan makan mulu sampe pendidikan dikesampingkan. Sejauh ini MBG juga kek gk ada gizi gizinya.. cuma sekeda makan isi perut. Udahhh.	Negative
MARI DUKUNG PROGRAM PEMERINTAH Makan Bergizi Gratis membantu anak-anak di Papua untuk tumbuh cerdas dan sehat. #MakanBergizi #GiziIndonesia #Makangratis #Gizi #MBG https://t.co/Ntkt1G2USL	Positive
@DPP_PKB @sahabat_arzeti Perlu lebih variatif. Tp yg penting negara wajib menjamin gizi seimbang di tiap porsi MBG	Positive
2. Kita disini untuk menjalankan program yang sudah dicanangkan oleh Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka yaitu program Makan Bergizi Gratis (MBG). ...	Positive
Program MBG merupakan inisiatif nasional yang bertujuan memastikan terpenuhinya kebutuhan gizi anak-anak sekolah. Selengkapnya: https://t.co/B06fe2GZSo https://t.co/EPDaHV0565	Netral

Dataset	Label
WAKIL Ketua DPR Sufmi Dasco Ahmad merespons adanya usulan masyarakat berkontribusi dalam program makan bergizi gratis (MBG) salah satunya melalui zakat. https://t.co/wZs4upxHMp	Neutral
Ketua Harian DPP Partai Gerindra Sufmi Dasco Ahmad menjelaskan penggunaan dana pribadi Presiden Prabowo Subianto untuk program Makan Bergizi Gratis (MBG) saat ini tak masalah. Sebab masih pada tahap uji coba. #presidenprabowo #prabowosubianto #indonesiamaju #indonesiaemas https://t.co/iFiLj8ov2w	Neutral

2.2. Cleaning Data

The data cleaning stage is carried out to ensure that the data used in the model analysis and training process is of good quality and free from noise that can affect the classification results [8], [9]. The data cleansing process in this study includes the following steps:

a) Duplicate tweet removal**,** which is tweets that have identical text and come from the same or different accounts. It is important to avoid bias due to the repetition of information [10].

b) Removal of empty or null tweets**,** i.e. entries that have no text content after the crawling process. This type of tweet does not provide information that can be used for sentiment analysis [11].

By doing this cleaning stage, the amount of data used in the next process becomes more representative.

2.3. Preprocessing

The preprocessing stage is carried out to convert the raw text from the crawled results into a cleaner format that is ready to be used in machine learning modeling [12], [13]. This process begins with converting HTML characters into their original symbols, such as changing & to &, so that the text can be read as it should. Furthermore, URLs are removed from text, especially words that begin with "http", because the information has no relevance to sentiment analysis. Mention of accounts such as @username is also removed to avoid bias against specific entities. In addition, punctuation marks such as periods, commas, and exclamation points are omitted to simplify sentence structure. Hashtags that are commonly used as topic markers, such as #MBG, are also removed because they do not directly contribute to the context of the sentiment.

After basic cleaning, a follow-up process is carried out in the form of removing stopwords [14], [15]. This process is done for common words in Indonesian such as "yang", "dan", "di", which have no important meaning in the analysis. Then word normalization is carried out, which is changing non-standard words or slang to standard forms, for example "gk" to "tidak" and "sm" to "sama" [16], [17]. The final stage is stemming, which is the process of changing words into their basic form, for example “makanannya” to “makan”, which is done using NLP libraries such as Sastrawi [18]–[20]. This entire preprocessing process aims to produce a simpler, consistent, and informative representation of text, so as to improve performance in the vectorization and sentiment classification stages [21].

2.4. Text Vectorization

After the preprocessing stage is completed, the next step is to perform a numerical representation of the text so that it can be processed by a machine learning algorithm. In this study, the method used for text representation is TF-IDF (Term Frequency–Inverse Document Frequency), which is one of the most popular vectorization techniques in text analysis [22].

TF-IDF works by measuring how important a word is in a document relative to the entire corpus. The Term Frequency (TF) value represents the frequency with which a word appears in a single document, while the Inverse Document Frequency (IDF) measures how rarely the word appears in the entire document [12]. Mathematically, the TF-IDF gives high weight to words that often appear in one document but rarely appear in another, making them more effective in capturing words that are discriminatory for classification.

Using TF-IDF, each tweet in the dataset is converted into a fixed-dimensional numeric vector**,** where each dimension represents a term (word) from the corpus after preprocessing [23]. The results of this transformation are used as inputs for the training and testing process of the sentiment classification model. The use of TF-IDF in this study aims to maintain a balance between word frequency and its relevance in the overall document, while reducing the dominance of generic words that are less informative [24]. This method was also chosen because of its ability to produce more spars and efficient representations than methods such as Bag-of-Words. In formula (1) is a formula to produce the TF-IDF vector value.

$$TFIDF(t,d) = TF(t,d)\ x\ log\left(\frac{N}{DF(t)}\right) \tag{1}$$

Where TF(t,d) is the frequency of the term *t* in document *d*. N is the total number of documents. DF(t) is the sum of documents that contain the term *t*.

2.5. SMOTE Algorithm

One of the main challenges in sentiment classification is class imbalance, which is a condition in which the amount of data in one class (e.g. positive opinions) is much more than the other class (e.g. negative or neutral opinions) [25]. This imbalance can cause classification models to tend to favor the majority class, thus reducing the accuracy of identifying minority classes that are often more analytically important [26]. To overcome these problems, this study applies SMOTE to training data. SMOTE is an oversampling technique that works by creating a new synthetic sample of a minority class, not just by copying existing data, but by interpolating between adjacent minority data points in the feature space [27], [28].

Technically, for each minority sample, the SMOTE algorithm selects k-nearest neighbors (generally k=5), then randomly selects one of its neighbors and generates a new data point between them based on Euclidean distances [29]. This process is repeated until the number of minority samples is equal to the number of majority samples, so that the dataset becomes balanced. In the context of this study, SMOTE was applied after the vectorization process (TF-IDF) and before the model training stage. This ensures that only the training data is synthetically expanded, while the test data remains pure to maintain the validity of the model performance evaluation [30]. The implementation of SMOTE is expected to improve the model ability to detect minority opinions.

In this study, the SMOTE algorithm was chosen as the oversampling technique due to its well-established performance and wide adoption in handling class imbalance, especially in early benchmark studies involving text classification and structured tabular data. While advanced techniques such as ADASYN and Borderline-SMOTE offer adaptive or edge-focused sampling, SMOTE was selected for its simplicity, interpretability, and reproducibility as a baseline method. This also allows a more controlled evaluation of the effects of oversampling across ensemble models. Formula (2) is how to produce synthetic samples using SMOTE.

$$X_{new} = x_i + \delta\,.\,(x_{nn} - x_i) \tag{2}$$

Where x_i is a genuine minority sample, x_{nn} is one of the closest neighbors of x_i, and δ is a random number between 0 and 1.

2.6. Model Evaluation

To evaluate the performance of the model in classifying public opinion of the MBG program, this study uses four main evaluation metrics, namely Accuracy, Precision, Recall, and F1-Score. These four metrics were chosen because they were able to provide a comprehensive picture of the model's capabilities, especially in the context of unbalanced data [5].

All models in this study were evaluated using test data that were not affected by the SMOTE oversampling process to maintain the objectivity of performance measurement. By comparing the results of the trained model with and without SMOTE, this evaluation aims to find out the extent to which the application of SMOTE impacts the model's ability to detect minority opinions, as well as how consistent it is across different types of ensemble models.

The use of questionnaires is generally needed in exploratory research or surveys, where the opinions, perceptions, or subjective assessments of respondents are the main data in answering the problem formulation. However, in the context of this study, the approach used is quantitative-computational based on experiments and objective model performance measurement.

This research does not rely on the opinions of experts or respondents as the basis for decision-making, but on the empirical evaluation of machine learning models through evaluation metrics such as accuracy, precision, recall, and F1-score. The entire testing process was carried out on datasets that had been systematically processed, with different treatment in the aspect of balancing data (with and without SMOTE), and the results were measured between ensemble models.

Thus, the validity of the decision in this study does not come from individual subjectivity, but from the results of reproducible data-based experiments**.** Therefore, the use of questionnaires as an additional data collection instrument is considered irrelevant and unnecessary**,** as it will not add evidential value to the main purpose of the study, which is to evaluate the effectiveness of SMOTE in combination with ensemble algorithms against unbalanced public opinion data.

3. Results and Discussion

This section presents a comparative analysis of the performance of five *ensemble learning* models with and without the application of the SMOTE algorithm. The selection of five ensemble models—Random Forest, AdaBoost, Bagging, Stacking, and Voting—was based on the popularity and diversity of ensemble approaches in the current literature. Random Forest and Bagging represent bootstrap aggregating-based ensemble techniques, AdaBoost reflects an iterative boosting approach, Stacking represents a combination of heterogeneous meta-learner-based models, while Voting is used to evaluate the individual contributions of models in majority-based ensemble schemes. By selecting models from these different categories, this study seeks to comprehensively evaluate the effects of SMOTE within the framework of ensemble learning.

The evaluation was carried out using the Accuracy, Precision, Recall, and F1-Score metrics which represent the performance of the model in classifying public opinion on unbalanced data. The aim of this experiment was to find out the extent to which SMOTE can improve the model's ability to recognize minority classes more accurately. Full results are presented in Table 2.

Table 2. Result of Experiment

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Without SMOTE	Random Forest	76.63	74.76	71.93	72.03
	AdaBoost	69.96	66.1	64.36	63.51
	Bagging	75.02	72.5	70.95	71.05
	Stacking	74.68	71.48	70.63	70.62
	Voting	76.29	74.21	71.86	71.97
With SMOTE	Random Forest	76.63	74.5	72.53	72.66
	AdaBoost	67.66	65.13	64.58	64.38
	Bagging	74	71.2	70	70
	Stacking	73	70	70	71
	Voting	72.84	76	66	65

In conditions without the implementation of SMOTE, the Random Forest model obtained the best results overall, with an Accuracy of 76.63% and an F1-Score of 72.03%. This shows that Random Forest is able to handle class imbalances relatively well even without the oversampling technique. The Voting Model also shows a competitive performance, with an Accuracy of 76.29% and an F1-Score of 71.97%, slightly below Random Forest. Meanwhile, AdaBoost was the lowest-performing model, particularly on F1-Score (63.51%), which indicates that it is less effective in detecting minority classes.

After the implementation of SMOTE on the training data, Random Forest remained the best model with the F1-Score increasing to 72.66%, although the Accuracy remained unchanged (76.63%). This improvement reflects that SMOTE has successfully helped the model to better recognize minority classes, demonstrated by the increase in Recall from 71.93% to 72.53%. However, the implementation of SMOTE did not have a positive impact on all models. AdaBoost saw a decline in all metrics, with the F1-Score reaching just 64.38%. Similarly, Voting saw a significant decline, with F1-Score dropping from 71.97% to 65%, and Accuracy dropping to 72.84%, although Precision actually increased to 76%. This shows that the Voting model is becoming more "confident" in predicting, but more misclassified in the minority class (Recall drops to 66%).

Random Forest shows the most stable and optimal performance, both with and without SMOTE. SMOTE has been shown to be effective in improving recall and F1-score in certain models, but inconsistently improving all metrics across models. Models such as Voting and AdaBoost tend to experience performance degradation after SMOTE, likely due to overfitting the synthetic data generated.

The effectiveness of SMOTE in improving model performance is greatly influenced by the internal characteristics of each algorithm. Models such as Random Forest and Bagging based on sampling and random tree formation are better able to utilize the data from SMOTE synthesis. In contrast, models like AdaBoost that are sensitive to noise and small errors tend to experience performance degradation when synthetic data introduces unrepresentative variations. Similarly, in the Voting ensemble, the effect of aggregation between models can neutralize the positive impact of SMOTE if not all models get balanced benefits from the technique.

To further validate comparative claims related to the effectiveness of the use of SMOTE techniques in improving model performance, the Wilcoxon Signed-Rank Test was conducted on evaluation metrics (Accuracy, Precision, Recall, and F1-Score) on five ensemble models (Random Forest, AdaBoost, Bagging, Stacking, and Voting), by comparing the results of the model before and after the implementation of SMOTE.

The test results showed that the difference in model accuracy with and without SMOTE was close to a level of statistical significance (p-value = 0.067), indicating that SMOTE has the potential to have a real impact on improving model accuracy, although it is not yet statistically significant at a 95% confidence level. Meanwhile, for the Precision ($p = 0.625$), Recall ($p = 0.313$), and F1-Score ($p = 0.813$) metrics, no statistically significant differences were found. These findings suggest that while SMOTE does not necessarily result in statistically significant performance improvements on all metrics, it still provides practical improvements, particularly in terms of improving the model's ability to recognize minority classes in imbalanced data.

Discussion

This study aims to classify public sentiment towards the MBG program using a machine learning approach based on ensemble models, and to evaluate the influence of unbalanced data handling techniques, specifically the use of SMOTE. The dataset used shows an uneven class distribution between positive, negative, and neutral sentiments, which poses the risk of model bias toward the majority class and potentially underrepresents critical public feedback.

1. Effectiveness and Limitations of SMOTE in Sentiment Classification

The implementation of SMOTE on MBG sentiment data revealed improved model performance in several ensemble algorithms, particularly Random Forest and Stacking, which showed notable increases in both recall and F1-score. These findings indicate that SMOTE is effective in enhancing the model's sensitivity to minority classes such as negative or neutral opinions that would otherwise be overlooked in imbalanced settings. This is crucial in public policy sentiment analysis, where the identification of critical or negative feedback is as important as detecting supportive opinions, especially for evaluating program effectiveness and public trust.

However, the effectiveness of SMOTE is not uniform across all models. For instance, the Voting and AdaBoost models experienced a decline in F1-score after oversampling. Although an increase in precision was observed in the Voting model, this came at the cost of reduced recall, suggesting that the model became overconfident yet less capable of capturing minority class signals. This inconsistency highlights that SMOTE may lead to overfitting, particularly when used in complex or sensitive algorithms that are vulnerable to synthetic data variations. Models like Voting and Stacking may struggle with the artificial structure of interpolated samples, making their predictions less robust.

2. Repeated Testing and Cross-Validation

To ensure the reliability of the findings, all models were evaluated using 5-fold cross-validation**.** This approach provides a more consistent and generalized assessment of model performance across different data splits. Additionally, multiple trials were conducted using randomized seed initializations to test the stability of each model. The average scores reported in Table 2 reflect consistent patterns across repeated evaluations, not isolated or one-off outcomes. This reinforces the validity of the experimental design and confirms that the observed differences are reproducible.

3. Implications for Public Policy Evaluation

From a practical perspective, the findings stress the importance of adopting selective and data-aware balancing techniques in sentiment analysis on public policy. While SMOTE has proven useful, its implementation must be tailored to the characteristics of each ensemble model and the complexity of unstructured social media data. Based on the results, Random Forest combined with SMOTE is recommended for similar contexts due to its stable performance. Meanwhile, more caution is advised when applying SMOTE to models like Voting or AdaBoost. The use of advanced variations such as SMOTE-Tomek or Borderline-SMOTE could be explored in future work to mitigate the risks of overfitting and better simulate real-world minority patterns.

Although this study focuses on the analysis of public opinion on MBG policies, the ensemble learning and unbalanced data handling approaches used have broad relevance to other policy domains such as education, environment, and transportation. This technique can be applied to other public policy opinion data that also have unbalanced characteristics, as long as context adjustment and domain validation are carried out accordingly.

4. Conclusion

This study contributes not by proposing a new algorithm, but by providing empirical insight into how SMOTE behaves differently depending on the ensemble algorithm used**,** and by identifying when its application might be detrimental (e.g., in Voting and AdaBoost). Such findings are valuable for future researchers and practitioners working on imbalanced, unstructured opinion data in local language contexts.

This study also evaluates the effectiveness of the SMOTE algorithm in addressing class imbalances in public sentiment data related to the MBG program. Five ensemble models: Random Forest, AdaBoost, Bagging, Stacking, and Voting were tested under two conditions: with and without the implementation of SMOTE. The results showed that Random Forest was the most stable and consistent model to produce high performance, especially in F1-score and recall metrics, both before and after the implementation of SMOTE.

In terms of methodological processes, the findings show that SMOTE is effective in increasing the sensitivity of the model to minority classes, especially in noise-resistant models such as Random Forest. However, the impact is not uniform for all algorithms. Some models such as Voting and AdaBoost actually experienced a decline in performance after oversampling, which indicates the potential for overfitting synthetic data. This reinforces that the success of the balancing method is highly dependent on the characteristics of the algorithm used, so its application cannot be equalized.

To ensure the validity of the results, all experiments were conducted with a 5-fold cross-validation approach as well as repeated experiments *with* different random seeds. This process ensures that the results obtained are not accidental, but stable and reproducible across different data divisions.

Based on these results, it is recommended that the use of SMOTE be tailored to the specific behavior of each model, and not applied uniformly. Further research can be directed towards testing more adaptive SMOTE variants such as Borderline-SMOTE or SMOTE-Tomek, as well as the integration of feature selection techniques to improve model resistance to noise. In addition, expanding the coverage of data from various social media platforms and the application of aspect-based sentiment analysis can provide a more comprehensive insight into public opinion on government policies such as the MBG program.

Declarations

Author contribution. Author 1 contributed to the research conceptualization, methodology design, and overall manuscript writing. Author 2 contributed to the methodology and actively participated in the writing process.

Funding statement. The authors sincerely appreciate the valuable support and funding granted by Telkom University through its internal research program, as stated in contract number 186/LIT06/PPM-LIT/2025.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] A. Kiftiyah, F. A. Palestina, F. U. Abshar, and K. Rofiah, "Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial–Politik," *Pancasila: Jurnal Keindonesiaan*, vol. 5, no. 1, pp. 101–112, 2025.
[2] B. Rahmatullah, S. A. Saputra, P. Budiono, and D. P. Wigandi, "Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes," *JIfoTech*, vol. 5, no. 1, Mar. 2025.
[3] S. Kedas, A. Kumar, and P. K. Jain, "Dealing with class imbalance in sentiment analysis using deep learning and SMOTE," in *Advances in Data Computing, Communication and Security*, Singapore: Springer Nature Singapore, 2022, pp. 407–416.
[4] R. Obiedat *et al.*, "Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022.
[5] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, "Chronic diseases prediction using machine learning with data preprocessing handling: a critical review," *IEEE Access*, vol. 12, pp. 80698–80730, 2024.
[6] V. S. Spelmen and R. Porkodi, "A review on handling imbalanced data," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, 2018, pp. 1–11.
[7] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, p. e1249, Jul. 2018.
[8] P. Kashyap, A. Pareek, S. Mishra, Z. Khan, R. Garg, and H. K. Tripathy, "Sentiment polarity analysis of twitter data using machine learning models," in *Innovative Computing and Communications*, Singapore: Springer Nature Singapore, 2024, pp. 623–635.
[9] B. Bala and S. Behal, "A brief survey of data preprocessing in machine learning and deep learning techniques," in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, 2024, pp. 1755–1762.
[10] B. Menaouer, S. Fairouz, M. B. Meriem, S. Mohammed, and M. Nada, "A sentiment analysis of the Ukraine-Russia War tweets using knowledge graph convolutional networks," *Int. J. Inf. Technol.*, Jan. 2025.
[11] G. Taiwo, M. Saraee, and J. Fatai, "Crime prediction using twitter sentiments and crime data," *Informatica (Vilnius)*, vol. 48, Feb. 2024.
[12] G. Popoola, K.-K. Abdullah, G. S. Fuhnwi, and J. Agbaje, "Sentiment analysis of financial news data using TF-IDF and machine learning algorithms," in *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, Houston, TX, USA, 2024, pp. 1–6.
[13] A. S. Safitri, I. Wijayanto, and S. Hadiyoso, "Improving classification accuracy with preprocessing techniques for sentiment analysis," in *2024 International Conference on Data Science and Its Applications (ICoDSA)*, Kuta, Bali, Indonesia, 2024, vol. 7, pp. 487–490.
[14] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Comput. Math. Organ. Theory*, vol. 25, no. 3, pp. 319–335, Sep. 2019.
[15] D. J. Ladani and N. P. Desai, "Stopword Identification and Removal Techniques on TC and IR applications: A Survey," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020, pp. 466–472.
[16] H. Raza, M. Faizan, A. Hamza, A. Mushtaq, and N. Akhtar, "Scientific text sentiment analysis using machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 157–165, 2019.
[17] Student, The University of Texas at Arlington, Arlington, Texas, United States of America and S. S. A. Challapalli, "Sentiment analysis of the Twitter dataset for the prediction of sentiments," *Journal of Sensors, IoT & Health Sciences*, vol. 2, no. 4, pp. 1–15, Dec. 2024.
[18] A. Sabir, H. A. Ali, and M. A. Aljabery, "ChatGPT tweets sentiment analysis using machine learning and data classification," *Informatica (Vilnius)*, vol. 48, May 2024.

[19] N. A Semary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, "Enhancing machine learning-based sentiment analysis through feature extraction techniques," *PLoS One*, vol. 19, no. 2, p. e0294968, Feb. 2024.
[20] N. Romadoni, A. M. Siregar, D. S. Kusumaningrum, and T. Rohana, "Classification Model of Public Sentiments About Electric Cars Using Machine Learning," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 303–314, 2024.
[21] N. G. Ramadhan and F. Adhinata, "Sentiment analysis on vaccine COVID-19 using word count and Gaussian Naïve Bayes," *Indones. J. Electr. Eng. Comput. Sci.*, Jun. 2022.
[22] K. Alemerien, A. Al-Ghareeb, and M. Z. Alksasbeh, "Sentiment analysis of online reviews: A Machine Learning based approach with TF-IDF vectorization," *J. Mob. Multimed.*, vol. 20, no. 5, pp. 1089–1116, Dec. 2024.
[23] Y. Terentyeva, "Sentiment Analysis, InSet Lexicon, SentiStrength Lexicon, Naive Bayes, Multinomial Naive Bayes, TF-IDF, Machine Learning," *International Journal of Open Information Technologies*, vol. 12, no. 7, pp. 32–37, 2024.
[24] E. Triana, A. I. Purnamasari, A. Bahtiar, and E. Tohidi, "Improved spam email detection performance based on Naïve Bayes approach TF-IDF Vectorizer with multi-metric optimization," *j. of artif. intell. and eng. appl.*, vol. 4, no. 3, pp. 1667–1672, Jun. 2025.
[25] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 1, Jun. 2024.
[26] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artif. Intell. Rev.*, vol. 57, no. 6, May 2024.
[27] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018.
[28] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, "Prediction of cardiovascular disease (CVD) in the upcoming year using tree-based ensemble model," in *12th International Conference on Information and Communication Technology (ICOICT)*, 2024, pp. 210–216.
[29] P. P. Putra, "Optimizing sentiment analysis on imbalanced hotel review data using SMOTE and ensemble machine learning techniques," *J. Appl. Data Sci.*, vol. 6, no. 2, pp. 921–935, May 2025.
[30] A. Adiwijaya and N. G. Ramadhan, "Analyzing risk factors and handling imbalanced data for predicting stroke risk using machine learning," *Int. J. Adv. Intell. Inform.*, vol. 11, no. 1, pp. 39–54, Feb. 2025.