Journal of Information Systems and Informatics
Vol. 7, No. 1, March 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i1.1039 Published By DRPM-UBD

Sentiment Analysis of 'Free Lunch for Children' Policy on Social Media X Using Random Forest Algorithm

Anies[1], Muhammad Ikhsan[2]

[1,2]Computer Science, Faculty of Science and Technology, State Islamic University of North Sumatera, Indonesia
Email: [1]anies1904@gmail.com, [2]mhd.ikhsan@uinsu.ac.id

Abstract

The concept of a welfare state emphasizes the main role of the government in providing protection and improving welfare such as health and education to its people. The free lunch program proposed by Prabowo Subianto and Gibran Rakabuming Raka aims to improve the nutritional quality of school children while driving the national economy. The public's reaction to Prabowo Subianto's work program plan, on free school lunch program and nutritional support for Indonesian students, is very diverse in perspective on X. The Random Forest algorithm proved to be quite effective in classifying public sentiment regarding the policy of "Free Lunch for Children." With an overall accuracy of 73%, the model was able to categorize public opinion into positive, negative, and neutral categories. To improve the performance of the model, upsampling was performed to balance the classes in the dataset as well as hyperparameter tuning. After applying these techniques, the accuracy of the model increased significantly to 80%.

Keywords: Sentiment Analysis, Random Forest Algorithm, Social Policy

1. INTRODUCTION

The concept of a welfare state emphasizes the main role of the government in providing protection and improving welfare such as health and education to its people. Data compiled by the Ministry of Health in 2020 highlights the serious problems that continue to afflict most children in Indonesia, especially in the Jakarta and East Nusa Tenggara (NTT) regions, namely stunting and failure to thrive. These two diseases pose a significant threat to the physical development and well-being of children in this country. Stunting is of particular concern because it has a long-term impact on children's lives, both from a health and education perspective [1]. Stunting and failure to thrive are complex problems that are influenced by various factors, including access to adequate nutrition, poor hygiene, and inappropriate nutrition practices [2].

Data on the high number of stunting in the Jakarta and East Nusa Tenggara (NTT) regions can be seen as follows In Jakarta, the stunting rate reached 14.8% in 2022. Meanwhile, NTT has the highest stunting rate nationally in 2021 with a prevalence of 37.8%. Based on BPS data, in 2019, the percentage of stunting in NTT was at

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

30%, then decreased to 24.5% in 2020, 20.9% in 2021, 17.7% in 2022, and 15.2% in 2023. In the last few years. In recent years, NTT has experienced a significant decrease in stunting rates, with an average decrease of 3.4% over the past five years [3].

Addressing this issue requires comprehensive and collective action from the government, the community and the private sector. In addition, more effective prevention and intervention efforts are needed, including education on healthy eating habits and the development of nutrition programs for children [4]. The free meal program, which is set to begin in 2029, aims to reduce stunting across Indonesia by providing nutritious meals to children in their growing years. The program is expected to increase access to adequate and quality nutrition, especially in areas with high stunting rates such as Jakarta and East Nusa Tenggara (NTT).

This research focuses on the period from when the program was announced until February at the last presidential debate, to analyze its impact on reducing stunting. The research includes evaluating the implementation of the program in various regions, monitoring changes in stunting statistics, and identifying challenges and successes [5]. The results of this study are expected to provide a comprehensive picture of the effectiveness of the free meal program in reducing stunting, as well as provide recommendations for improvement and development of future nutrition policies [6].

Public reaction to the announcement of a free meal program set for 2029 to reduce stunting in Indonesia has been mixed. Many people welcomed the initiative, considering it an important step in addressing nutrition, which remains a major challenge, especially in areas with high stunting prevalence such as Jakarta and East Nusa Tenggara (NTT). Parents, health teams, and non-profit organizations praised the government's efforts in improving access to nutritious food for children. However, there are also concerns and skepticism emerging from various parties. Some are concerned about the effectiveness of the program's implementation, especially in remote areas with inadequate infrastructure [7].

The free lunch program proposed by Prabowo Subianto and Gibran Rakabuming Raka aims to improve the nutritional quality of school children while driving the national economy [8]. Through this program, students in need will receive nutritionally balanced meals, with the hope of improving the quality of human resources to create the golden generation in 2045. In addition, the provision of nutritious food is also expected to boost students' learning activities which will ultimately contribute to economic growth [9]. The program not only focuses on health and education aspects, but also opens up new business opportunities and increases community participation in economic development. Partnerships with the private sector are an important element in the implementation of this program,

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

where the large funding required must be managed transparently. With an estimated cost of Rp15,000 per child, the program also requires effective logistics management to ensure that the food distribution can reach various areas, including remote ones. In addition to providing one nutritious meal a day, the program also has an educational component that teaches students and parents about the importance of nutritious food and healthy eating habits.

The 2024 President, Prabowo Subianto, proposed a free lunch program and nutritional support for students of various educational levels, related to the concept of a welfare state, as a measure to improve the socioeconomic welfare of the community [10]. One of the most important steps is to improve access to quality nutrition and appropriate health services, especially in areas where stunting rates are high. In addition, more effective prevention and intervention efforts are needed, including education on healthy eating habits and the development of nutrition programs for children. Improving people's nutrition and reducing the prevalence of stunting is one of the Ministry of Health's priority programs for the 2020-2024 period [11]. The public's reaction to Prabowo Subianto's work program plan on free school lunches and nutritional support for Indonesian students was very diverse in X perspectives.

The Random Forest method is a machine learning method for classification and regression by combining multiple decision trees into an ensemble, where each tree provides independent predictions and is combined by voting or averaging [8]. Its advantages include handling large datasets with many features, overfitting tolerance, and handling unbalanced data and also provides an estimate of the importance of features in prediction, thus providing additional insights [12].

2. METHODS

The research methodology used in this study follows a structured, multi-stage process illustrated in Figure 1, consisting of four key phases: Planning, Data Collection, Data Analysis, and Evaluation. These stages guide the systematic application of the Random Forest algorithm for sentiment analysis related to the "Free Lunch for Children" policy. Each stage is described in detail as follow.

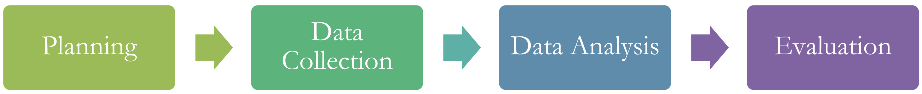

Figure 1. Research Framework

3.1. Planning

The initial stage began with identifying real-world problems, particularly focusing on public discourse around stunting and the "Free Lunch for Children" policy. The

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

primary research objective was formulated to explore public sentiment on this issue using machine learning techniques—specifically, the Random Forest algorithm. A thorough literature review was conducted to understand the theoretical foundation and identify gaps in previous studies, supporting the need for a data-driven approach to sentiment analysis in public policy evaluation.

3.2. Data Collection

Data was collected from the social media platform X (formerly Twitter) using a specialized scraping tool called Tweet-Harvest. This tool was selected due to recent changes in X's API access, which rendered other scraping tools like snscrape ineffective [13]. Tweet-Harvest enables automated retrieval of tweets, replies, threads, and comments based on specified keywords—in this case, the phrase "free lunch prabowo". The dataset includes rich metadata such as: tweet ID, tweet text, number of likes, number of retweets, timestamps, and user profile indicators This metadata provides contextual depth and supports multi-dimensional analysis beyond just sentiment classification [14].

3.3. Data Analysis

Once collected, the data was pre-processed to ensure consistency, remove noise, and handle missing values. Common preprocessing tasks included: lowercasing all text, removing stop words and punctuation, tokenization, lemmatization, and removing non-relevant or duplicated tweets. The cleaned data was then fed into the Random Forest algorithm, a widely used ensemble learning method known for its robustness and accuracy. This model works by constructing multiple decision trees and aggregating their predictions to determine the final class—typically using the mode of classes in classification tasks. To fine-tune the model and avoid overfitting, two critical hyperparameters were adjusted:

1) Number of Estimators (N Estimators): Defines the number of trees in the forest, typically ranging from 10 to 100. The optimal value is crucial for balancing bias and variance [15].
2) Maximum Depth (Max Depth): Determines how deep each tree can grow. Shallower trees reduce complexity and overfitting, while deeper trees may capture more detailed patterns [16].

To determine the best feature splits within the trees, the Gini Index was used as the impurity measure instead of alternatives like Information Gain or Gain Ratio. The Gini Index helps evaluate the "purity" of splits and is mathematically defined as shown in Equation 1.

$$\text{Gini Indeks} = 1 - \sum_{i=1}^{n} \left(p_i\right)^2 \tag{1}$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Where pi is the probability of class i at a particular node. For binary classification problems, such as sentiment (positive vs. negative), $p+$and $p-$represent the probabilities of positive and negative classes respectively.

3.4. Evaluation

In the final phase, the model's performance was assessed using standard evaluation metrics such as: accuracy, precision, recall, and f1-Score. These metrics offer a balanced view of how well the model handles both false positives and false negatives in sentiment classification [17]. The results were then interpreted in light of policy implications, particularly focusing on public sentiment trends around the "Free Lunch" initiative. The analysis provided data-driven insights that can support policymakers in evaluating the policy's reception and refining communication strategies.

3. RESULTS AND DISCUSSION

3.1. Data Analysis

The analysis phase of this research begins with the acquisition of public opinion data from the social media platform X (formerly Twitter) using the Tweet-Harvest tool. This tool was selected due to recent API restrictions on third-party scrapers like snscrape, which rendered traditional methods of tweet extraction obsolete [13]. Tweet-Harvest enables the real-time collection of tweet content, including metadata such as usernames, timestamps, tweet texts, retweets, likes, and replies, based on the specified keyword: "free lunch prabowo." This keyword was chosen to capture public responses to the "Free Lunch for Children" policy associated with Prabowo Subianto. A sample of the raw tweet data retrieved is shown in Table 1, where we see a range of expressions some critical, some supportive, and others neutral reflecting public sentiment toward the initiative.

Table 1. Data Sample (Indonesia)

Username	Tweets
freethink32	@geloraco @KPK_RI omon² terus persis @prabowo dr makan siang gratis mjd makan pagi. banyakan dramanya. tdk diketahui alamatnya tdk berwenang...bla...bla.... publik pingin aksi kongkret bknlah omongan. kalau bener tdk takut dgn presiden incapable @jokowi bapaknya @kaesangp ketum @psi_id
Sgrdamai	@rizieqdivist Mungkin ide program makan siang gratis pak Prabowo berawal dari sini
...	…
AlvaroGilan9	Prabowo - Gibran fokus pada kesehatan! Pembagian makan siang dan susu gratis tindak lanjut program gizi untuk anak-anak dan ibu hamil. dekade08

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Username	Tweets
bebyhasibuan_	Program makan siang gratis buat menurunkan angka stunting (kurang gizi) yang selama ini cukup tinggi dialami anak-anak Indonesia terutama untuk menghadapi bonus demografi 2035Kalau internet gratis? Bisa jamin gak dibuat nonton ... (Isi aja sendiri)

These tweets represent unstructured, raw textual data, which must be preprocessed before it can be used for sentiment classification and machine learning analysis. The preprocessing phase consisted of eight stages: text cleaning, case folding, tokenization, stopword removal, normalization, stemming, feature transformation (TF-IDF), and sentiment labeling. The purpose of these preprocessing steps is to reduce textual noise, normalize language structure, and convert raw tweets into a format that can be effectively used by classification algorithms. An example of this preprocessing workflow is provided in Table 2.

Table 2. Pre-Processing Data Sample (Indonesia)

Full text	Text Clean	Stopword Remover	Stemming
@geloraco @KPK_RI omon² terus persis @prabowo dr makan siang gratis mjd makan pagi. banyakan dramanya. tdk diketahui alamatnya tdk berwenang...bla...bla.... publik pingin aksi kongkret bknlah omongan. kalau bener tdk takut dgn presiden incapable @jokowi bapaknya @kaesangp ketum @psi_id	RI omon terus persis dr makan siang gratis mjd makan pagi banyakan dramanya tdk diketahui alamatnya tdk berwenangblabla publik pingin aksi kongkret bknlah omongan kalau bener tdk takut dgn presiden incapable bapaknya ketum id	['bicara', 'persis', 'dari', 'makan', 'siang', 'gratis', 'menjadi', 'makan', 'pagi', 'banyakan', 'dramanya', 'tidak', 'alamatnya', 'tidak', 'berwenang', 'publik', 'pingin', 'aksi', 'kongkret', 'bknlah', 'omongan', 'bener', 'tidak', 'takut', 'dengan', 'presiden', 'incapable', 'bapaknya', 'ketum']	bicara persis dari makan siang gratis jadi makan pagi banyak drama tidak alamat tidak wenang publik pingin aksi kongkret bknlah omong bener tidak takut dengan presiden incapable bapak tum
@rizieqdivist Mungkin ide program makan siang gratis pak Prabowo berawal dari sini	Mungkin ide program makan siang gratis pak		ide program makan siang gratis Prabowo

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Full text	**Text Clean**	**Stopword Remover**	**Stemming**
	Prabowo berawal dari sini		
...	...	...	...
Prabowo - Gibran fokus pada kesehatan! Pembagian makan siang dan susu gratis tindak lanjut program gizi untuk anak-anak dan ibu hamil. dekade08	Prabowo Gibran fokus pada kesehatan Pembagian makan siang dan susu gratis tindak lanjut program gizi untuk anakanak dan ibu hamil decade		subianto gibran fokus sehat bagi makan siang susu gratis tindak program gizi anakanak hamil decade
Program makan siang gratis buat menurunkan angka stunting (kurang gizi) yang selama ini cukup tinggi dialami anak-anak Indonesia terutama untuk menghadapi bonus demografi 2035Kalau internet gratis? Bisa jamin gak dibuat nonton ... (Isi aja sendiri)	Program makan siang gratis buat menurunkan angka stunting kurang gizi yang selama ini cukup tinggi dialami anakanak Indonesia terutama untuk menghadapi bonus demografi Kalau internet gratis Bisa jamin gak dibuat nonton Isi aja sendiri		program makan siang gratis turun angka stunting gizi alami anakanak indonesia hadap bonus demografi internet gratis jamin tidak nonton isi saja

Once cleaned and structured, the data underwent manual sentiment labeling. Each tweet was analyzed for its emotional tone and categorized into three sentiment classes: Positive, Neutral, or Negative. This labeling is essential for supervised learning algorithms like Random Forest, as it provides the ground truth required to train and validate the model. Examples of labeled data are presented in Table 3.

Table 3. Labelling Data (Indonesia)

Preprocessing Results	**Label**
bicara persis dari makan siang gratis jadi makan pagi banyak drama tidak alamat tidak wenang publik pingin aksi kongkret bknlah omong bener tidak takut dengan presiden incapable bapak tum	Negative
ide program makan siang gratis Prabowo	Neutral
...	...
prabowo gibran fokus sehat bagi makan siang susu gratis tindak program gizi anakanak hamil decade	Positive
program makan siang gratis turun angka stunting gizi alami anakanak indonesia hadap bonus demografi internet gratis jamin tidak nonton isi saja	Positive

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Following the labeling process, the next step was featuring extraction using Term Frequency–Inverse Document Frequency (TF-IDF). This technique quantifies the importance of each word in a document relative to the entire corpus, allowing the algorithm to better identify which terms contribute most to classification decisions [17]. TF-IDF is computed as the product of Term Frequency (TF) and Inverse Document Frequency (IDF) for each word in each document. This converts the raw text into a structured numerical format, where each word is assigned, a weight based on its relevance.

Table 4. TF-IDF

Term	Term Frequency				DF	IDF	TF-IDF			
	K1	K2	K3	K4			K1	K2	K3	K4
Aksi	1	0	0	0	1	1.38629	1.38629	0	0	0
alamat	1	0	0	0	1	1.38629	1.38629	0	0	0
Alami	0	0	0	1	1	1.38629	0	0	0	1.38629
anakanak	0	0	1	1	2	0.69314	0	0	0.69314	0.69314
Angka	0	0	0	1	1	1.38629	0	0	0	1.38629
…	…	…	…	…	…	…	…	…	…	…
Takut	1	0	0	0	1	1.38629	1.38629	0	0	0
Tidak	3	0	0	1	4	0.69314	2.07944	0	0	0.69314
Tindak	0	0	1	0	1	1.38629	0	0	1.38629	0
Tum	1	0	0	0	1	1.38629	1.38629	0	0	0
Turun	0	0	0	1	1	1.38629	0	0	0	1.38629

After getting the TF-IDF representation of the documents, the next step is to build the decision tree. This step starts by determining the best feature to perform the first split [18]. In this example, we use the Gini Index (as default from the Python Scikit-learn library) to select the best feature. In Random Forest, we need to create multiple decision trees. Each tree is built with a random subset of the terms (features) in the dataset. Suppose we build 3 decision trees for this analysis; each tree will have a different subset of terms, as shown in Equation 2 to 4.

$$\text{Gini}(t)=1-\sum_{i=1}^{c} p_i^2 \quad (2)$$

$$\text{Gini}(\text{lunch})=1-\left(p_{\text{negative}}^2+p_{\text{neutral}}^2+p_{\text{positive}}^2\right) \quad (3)$$

$$\text{Gini}(\text{lunch})=1-\left(\left(\frac{1}{4}\right)^2+\left(\frac{1}{4}\right)^2+\left(\frac{2}{4}\right)^2\right)=1-\frac{6}{16}=0.625 \quad (4)$$

The Random Forest algorithm constructed three decision trees using different random subsets. From the decision tree calculation, the predicted sentiments for each document are as follows: K1: Negative, K2: Neutral, K3: Positive, and K4: Positive. After having several decision trees, the next step is voting to determine the final prediction of the Random Forest model. For each document, the majority

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

prediction is calculated from all trees. The majority voting result will be the final Random Forest prediction as in Table 5.

Table 5. Majority Voting

Document	1st Tree	2nd Tree	3rd Tree	Majority Voting
K1	Negative	Negative	Negative	Negative
K2	Positive	Neutral	Neutral	Neutral
K3	Positive	Positive	Positive	Positive
K4	Positive	Positive	Positive	Positive

For each document, the majority prediction is calculated from all trees. The majority voting result will be the final Random Forest prediction as in the following table. Thus, the final prediction from Random Forest for this dataset is Negative for K1, Neutral for K2, and Positive for K3 and K4. These results are obtained based on the voting of the three trees that have been built previously.

3.2. Model Evaluation

This study utilizes the Random Forest model to classify sentiment based on tweet data that has undergone extensive preprocessing and vectorization using the TF-IDF (Term Frequency–Inverse Document Frequency) technique. Random Forest is an ensemble learning method that combines the predictions of multiple decision trees to produce more accurate and stable results. Its inherent capability to reduce overfitting and handle high-dimensional data makes it an ideal choice for text classification tasks such as sentiment analysis.

As illustrated in Figure 3, the Random Forest classifier used in this research was configured with the parameter n_estimators=100, indicating that the model builds 100 decision trees. During the training phase, each tree is built using a bootstrapped sample of the training data and a random subset of TF-IDF features. This randomness introduces diversity among the trees and improves the generalization performance of the overall model. Once trained, the ensemble of decision trees is used to predict the sentiment category—positive, negative, or neutral—of tweets in the test dataset.

```
# RandomForest Classifier
nb = RandomForestClassifier(n_estimators=100, random_state=42)
nb.fit(tfidf_train, y_train)
y_pred = nb.predict(tfidf_test)
```

Figure 3. Random Forest Classifier

Following prediction, the model's performance was evaluated using four standard classification metrics: accuracy, precision, recall, and F1-score [19]. Accuracy

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

provides a general measure of how many tweets were correctly classified, while precision assesses how many of the predicted positive sentiments were actually correct. Recall measures the model's ability to identify all relevant instances of a particular sentiment, and the F1-score offers a balanced assessment by combining both precision and recall. These metrics collectively provide a comprehensive understanding of how well the model performs across all sentiment categories.

The initial model achieved an accuracy of 73%, reflecting a fairly good ability to classify sentiment from social media text. However, upon further inspection, it became evident that the dataset was imbalanced, with certain sentiment categories being underrepresented. This imbalance negatively impacts classification models by causing them to disproportionately favor the majority class, which reduces the accuracy and fairness of predictions across all categories.

To address this issue, the study applied the Upsampling technique to the minority classes. Upsampling involves duplicating samples from underrepresented sentiment categories until their frequency is comparable to that of the majority class [20]. This process balances the training dataset and helps ensure that the model learns to recognize all sentiment classes equally. By exposing the classifier to a more balanced distribution of samples, the likelihood of biased predictions is significantly reduced.

	precision	recall	f1-score	support
negatif	0.76	0.81	0.79	257
netral	0.81	0.79	0.80	271
positif	0.85	0.81	0.83	243
accuracy			0.80	771
macro avg	0.81	0.80	0.81	771
weighted avg	0.81	0.80	0.80	771

Figure 4. Performance Evaluation

After implementing upsampling and retraining the model, the overall classification performance improved. As shown in Figure 4, the Random Forest model's accuracy increased to 80%, indicating a meaningful enhancement in predictive power. This result confirms the effectiveness of the upsampling approach in resolving class imbalance and producing more equitable and accurate predictions across all sentiment categories. The improvement in accuracy and other evaluation metrics demonstrates that Random Forest, combined with proper data balancing techniques, is highly capable of handling real-world sentiment analysis tasks in a social media context.

3.3. Discussion

The findings of this study reveal several important insights into the effectiveness of using the Random Forest algorithm for sentiment analysis of public opinion on

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

social media, specifically related to the "Free Lunch for Children" policy. The application of a structured methodology from data collection via Tweet-Harvest to preprocessing and TF-IDF vectorization established a solid foundation for the classification model. The successful execution of this process demonstrates that social media platforms like X (formerly Twitter) can serve as reliable, real-time sources for gauging public sentiment on socio-political issues. Initially, the model achieved an accuracy of 73%, indicating reasonable predictive capability. However, a closer examination of the dataset revealed an issue commonly encountered in sentiment classification tasks class imbalance. In this study, the number of positive, negative, and neutral tweets was not evenly distributed. This skew resulted in a model that leaned heavily toward predicting the dominant class, thereby affecting the overall quality and fairness of its predictions. Such imbalances are known to distort accuracy metrics and often fail to reflect true model performance across all classes.

To address this challenge, upsampling was introduced as a technique to synthetically balance the dataset by duplicating samples from the minority classes. This intervention had a marked effect on the model's behavior. After retraining the model on the balanced dataset, the accuracy improved to 80%, and the distribution of correct predictions across all classes became more consistent. This improvement highlights the importance of balancing techniques in machine learning, particularly when working with real-world, user-generated data that is naturally imbalanced. Beyond accuracy, improvements were also observed in precision, recall, and F1-score, suggesting the model's increased ability to correctly identify and differentiate between sentiment classes. This balanced performance is crucial in sentiment analysis applications, especially in policy-related contexts where both positive and negative feedback carry significant weight in influencing decision-making and public perception.

The results also validate the choice of the Random Forest algorithm for this task. Its ensemble nature building multiple decision trees on varied data subsets and combining their outputs proves advantageous in managing textual complexity and reducing overfitting, especially when combined with feature extraction techniques like TF-IDF. The algorithm's robustness, ease of interpretability, and resilience to noise make it a suitable candidate for sentiment classification tasks, particularly in the highly dynamic and informal environment of social media platforms. Furthermore, the use of TF-IDF as a vectorization method contributed to the model's ability to emphasize contextually important terms in tweets, allowing the classifier to focus on the most informative words rather than frequent but insignificant terms. This increased the semantic richness of the input features and helped distinguish subtle differences between sentiment categories.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The successful application of this model in analyzing sentiments around the "Free Lunch" policy demonstrates the potential of machine learning as a decision-support tool for policymakers. By transforming subjective public opinion into quantifiable insights, such systems can inform better policy evaluation, communication strategies, and public engagement. However, it is also important to acknowledge the limitations of the current approach. Despite improvements, the model still depends on manual labeling and rule-based preprocessing, which may introduce bias or inconsistencies. Moreover, social media data is highly context-sensitive, and sarcasm, slang, or mixed sentiment can still confuse even the most robust models. As such, future studies could explore integrating deep learning approaches or contextual language models like BERT, which are capable of capturing more nuanced sentiment features.

The findings affirm that the integration of TF-IDF-based feature extraction, class balancing via upsampling, and Random Forest modeling forms an effective pipeline for sentiment classification in a public policy context. The improved evaluation metrics post-upsampling emphasize the value of addressing data distribution issues in building fair and accurate models. These results provide a practical pathway for using social media sentiment as a reliable metric for real-time policy monitoring and evaluation.

4. CONCLUSION

The results of this study demonstrate that the Random Forest algorithm is a reliable and effective method for classifying public sentiment related to the “Free Lunch for Children” policy. Initially, the model achieved a commendable accuracy of 73%, successfully categorizing tweets into positive, negative, and neutral sentiment classes. However, performance limitations due to class imbalance were evident, particularly in the neutral sentiment category. To address this issue, two key optimization strategies were implemented: upsampling of minority classes and hyperparameter tuning. These enhancements significantly improved the model's overall performance, with accuracy rising to 80%. Additionally, metrics such as precision and recall showed marked improvements, especially within the neutral class, indicating that the classifier became more equitable and accurate across all sentiment categories. This improvement confirms that combining data balancing techniques with model optimization not only enhances prediction accuracy but also leads to more balanced and consistent sentiment classification. Such enhancements enable the model to better capture the diversity and nuance of public opinion expressed on social media. Ultimately, these findings underscore the potential of Random Forest—when properly tuned and supported by preprocessing and balancing strategies—as a robust tool for sentiment analysis in the context of policy evaluation and public discourse monitoring.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

REFERENCES

[1] M. A. P. Ginting and S. Sriani, "Developing a Web-Based Application for Palm Seedling Eligibility Using C5.0 Algorithm and CART Algorithm," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 12, no. 1, pp. 97–108, Mar. 2024, doi: 10.33558/piksel.v12i1.8810.

[2] H. C. Morama, D. E. Ratnawati, dan I. Arwani, "Analisis sentimen berbasis aspek terhadap ulasan Hotel Tentrem Yogyakarta menggunakan algoritma Random Forest Classifier," *J. Pengemb. Teknol. Inf. Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022.

[3] R. I. Astuti and Suhana, "Pengaruh Kompetensi Dan Knowledge Sharing Terhadap OCB Dengan Mediasi Komitmen Organisasional (Studi Pada Pegawai Bpsdmd Provinsi Jawa Tengah," *J. Ilm. EDUNOMIKA*, vol. 7, no. 1, 2023.

[4] Amri, "Analisis Pemanfaatan Teknologi Informasi Dan Komunikasi Dalam Menunjang Terwujudnya Makassar Sebagai 'Smart City,'" *J. Komun. KAREBA*, vol. 5, no. 2, 2021.

[5] I. S. Tinendung and I. Zufria, "Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, p. 2014, Oct. 2023, doi: 10.30865/mib.v7i4.6908.

[6] C. P. Yanti, N. W. Eva Agustini, N. L. W. Sri Rahayu Ginantra, and D. A. Putri Wulandari, "Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 2, p. 756, Apr. 2023, doi: 10.30865/mib.v7i2.5900.

[7] S. Y. Pangestu, Y. Astuti, and L. D. Farida, "Algoritma Support Vector Machine dalam Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia," *J. Mantik Penusa*, vol. 03, no. 01, 2019.

[8] D. Agung Prabowo and Sudianto, "Analisis Sentimen Sepak Bola Indonesia pada Twitter menggunakan K-Nearest Neighbors dan Random Forest," *JSAI J. Sci. Appl. Inform.*, vol. 6, no. 2, pp. 217–227, Jun. 2023, doi: 10.36085/jsai.v6i2.5337.

[9] A. Syakur, "Implementasi Metode Lexicon Base Untuk Analisis Sentimen Kebijakan Pemerintah Dalam Pencegahan Penyebaran Virus Corona Covid-19 Pada Twitter," *J. Ilm. Inform. Komput.*, vol. 26, no. 3, pp. 247–260, 2021, doi: 10.35760/ik.2021.v26i3.4720.

[10] D. Sabarudin dan P. Purwadi, "Ideologi kebangsaan, identitas politik, dan ekonomi calon presiden Prabowo Subianto dalam program BEST RESULTS FAST 2024–2029," *Univ. Kebangsaan Republik Indones.*, vol. 1, no. 1, pp. 30–43, 2023.

[11] D. Nurwahidah, G. Dwilestari, N. D. Nuris, dan R. Narasati, "Analisis sentimen data ulasan pengguna aplikasi Google Kelas pada Google Play

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Store menggunakan algoritma Naïve Bayes," *JATI (J. Mahasiswa Tek. Inform.)*, vol. 7, no. 6, pp. 3673–3678, 2023.

[12] A. Sagita, A. Faqih, G. Dwilestari, B. Siswoyo, dan D. Pratama, "Penerapan metode Random Forest dalam menganalisis sentimen pengguna aplikasi CapCut di Google Play Store," *JATI (J. Mahasiswa Tek. Inform.)*, vol. 7, no. 6, pp. 3307–3313, 2023.

[13] R. Chairunnisa, "Analisis Sentimen terhadap Karyawan Dirumahkan pada Media Sosial Twitter menggunakan Fitur N-Gram dan Pembobotan Augmented TF – IDF Probability dengan K-Nearest Neighbour," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 6, no. 4, pp. 1960–1965, 2022.

[14] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.

[15] S. N. Sari, M. R. Faisal, D. Kartini, I. Budiman, T. H. Saragih, and M. Muliadi, "Perbandingan Ekstraksi Fitur dengan Pembobotan Supervised dan Unsupervised pada Algoritma Random Forest untuk Pemantauan Laporan Penderita COVID-19 di Twitter," *J. Komputasi*, vol. 11, no. 1, pp. 33–42, Apr. 2023, doi: 10.23960/komputasi.v11i1.6650.

[16] A. M. A. Rahim, Inggrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier," *Indones. J. Comput. Sci.*, vol. 12, no. 5, Nov. 2023, doi: 10.33022/ijcs.v12i5.3413.

[17] A. Elhan, M. K. D. Hardhienata, Y. Herdiyeni, S. H. Wijaya, and J. Adisantoso, "Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme Random Forest dan BERT," *J. Ilmu Komput. Dan Agri-Inform.*, vol. 9, no. 2, pp. 199–211, Nov. 2022, doi: 10.29244/jika.9.2.199-211.

[18] A. H. Lubis, L. P. A. Lubis, and Sriani, "Sentiment analysis on twitter about the death penalty using the support vector machine method," *TEKNOSAINS J. Sains Teknol. Dan Inform.*, vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373.

[19] S. Ariqoh, M. A. Sunandar, and Y. Muhyidin, "Analisis Sentimen Pada Produk Cushion di Website Female Daily Menggunakan Metode Support Vector Machine (SVM)," *STORAGE J. Ilm. Tek. Dan Ilmu Komput.*, vol. 2, no. 3, pp. 137–142, Aug. 2023, doi: 10.55123/storage.v2i3.2345.

[20] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi Dan Teknol. Inf. JNKTI*, vol. 5, no. 3, pp. 423–431, Jun. 2022, doi: 10.32672/jnkti.v5i3.4382.