Current Science Research Bulletin

Volume 02 (05), May 2025, pp. 139-145

CrossRef DOI: [10.55677/csrb/02-V02I05Y2025](https://doi.org/10.55677/csrb/02-V02I05Y2025)

Available on: https://csrbjournal.org/index.php/csrb

e-ISSN: 3050-7103

p-ISSN: 3050-709X

Implementation of Random Forest Algorithm in Classifying Public Sentiment Towards Free Nutritious Meal Program

Ingka Amaliya Nur Sabrina[1], Triastuti Wuryandari[2], Rahmila Dapa*[3]

[1,2,3] Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro

ABSTRACT: The free nutritious meal program in Indonesia has garnered public attention and reactions on social media, especially on *X*. This study aims to analyze public sentiment towards the program using the Random Forest algorithm. The data were collected from *X* and labeled with positive (2371 tweets) and negative (432 tweets) using the InSet Lexicon. The optimal Random Forest model was determined through hyperparameter tuning using the GridSearchCV technique. The results of the study showed that Random Forest with parameters *max_features* = $sqrt(1980) \approx 44$, *n_estimators* = 100, *max_depth* = 40, *min_sample_split* = 2, and *min_sample_leaf* = 2 gave the best performance with accuracy of 87.54% and AUC score 0.8723. Based on the results, the Random Forest method proved to be effective in classifying public opinion on *X* regarding this program. The wordcloud visualization shows that the word "jepang" appears most frequently in positively labeled tweets, while the word "program" is more dominant in negatively labeled tweets. The results can inform government policy evaluations.

KEYWORDS: Free Nutritious Meal Program, Sentiment Classification, Random Forest, GridSearchCV

1. INTRODUCTION

Indonesia is a democratic country that guarantees freedom of speech, which is explicitly written in the Undang-Undang Dasar 1945 Pasal 28E Ayat (3). Public involvement in policy-making is crucial in Indonesia, especially regarding the widely discussed "Program Makan Bergizi Gratis" or Free Nutritious Meal Program on social media. Indonesia has approximately 24.85 million users on X, making it the fourth largest user base in the world. As a platform, X enables the public in Indonesia to share both supportive and critical views regarding the program's execution. Sentiment analysis is necessary to systematically understand public perception by categorizing opinions as either positive or negative.

Random Forest algorithm is a machine learning method used for sentiment analysis. According to Afdhal *et al*. (2022), this algorithm is widely chosen for sentiment classification because it is effective in overfitting and data containing noise and outliers. A study by Fitri *et al*. (2020) found that Random Forest achieved the highest accuracy of 97.16% and AUC value of 0.996 when compared to algorithm Naïve Bayes, Random Forest, and Support Vector Machine for analyzing sentiment in the Ruangguru application. This study aims to develop an effective Random Forest classification model to analyze public sentiment towards the "Program Makan Bergizi Gratis" based on tweets user X, providing a more objective analysis of public sentiment regarding the policy.

2. LITERATURE REVIEW

According to Mailo (2021), sentiment analysis is used to identify and understand how people express their emotions, both positive and negative sentiments. The main goal of sentiment analysis is to categorize the text based on the emotions it contains, that is, whether the text expresses positive or negative sentiment. Text mining helps in the process of obtaining information in documents or large data sources by automatically identifying patterns and trends that support sentiment classification.

Topics trending on X have attracted the attention of news media and sparked broader public discussion. In the pre-processing stage, irrelevant text is cleaned to prepare it for sentiment analysis. This process includes case folding, data cleaning (remove numbers, punctuation marks, emojis, and URL), and word normalization.

Text data labeling utilizes the Lexicon Based method which is done automatically. This method utilizes the Inset Lexicon Dictionary which contains 3609 positive words and 6609 negative words, the sentiment value or polarity score owned by each word in the text (Koto and Rahmaningtyas, 2017).

Corresponding Author: Rahmila Dapa

$$Polarity\ Score \begin{cases} Positive = positive\ score - negative\ score \geq 0 \\ Negative = positive\ score - negative\ score < 0 \end{cases} \quad (1)$$

Feature selection is the process of selecting the most relevant features for machine learning models (Büyükkeçeci and Okur, 2023). In this study, feature selection includes several stages. Tokenizing is the process of dividing a sentence into smaller units in the form of word collections, then stopword removal is the process of cleaning text document from irrelevant words, and stemming is a process of removing word endings or word affixes to create new basic words.

TF-IDF weighting is obtained by multiplying the Term Frequency (TF) and Invers Document Frequency (IDF) values to determine and calculate the importance of a particular word in a text document. TF is the total appearance of the word on the document. IDF is used to determine how important a word is in a set of documents. The TF and IDF values are multiplied to get the TF-IDF (Intan and Defeng, 2006).

$$W_{i,j} = \frac{n_{i,j}}{\sum T_j}\left(ln\left(\frac{N}{df_{(i)}}\right) + 1\right) \quad (2)$$

Descriptions:

$W_{i,j}$: TF-IDF weight on *term i* and document *j*

$n_{i,j}$: Number of times *term-i* appears in document-*j*

$\sum T_j$: Total occurrence of the *term* in document-*j*

N : Total number of documents

$df_{(i)}$: Number of documents containing *term i*

The random forest algorithm was developed from the CART method by utilizing random feature selection techniques and bootstrap incorporation (bagging) (Breiman, 2001). In the process, each tree in the random forest uses a different subset of features to reduce the correlation between trees, resulting in a model that is more robust and resistant to overfitting and outliers. Classification is performed by aggregating predictions from all trees through majority voting. The random forest classification algorithm can work with the following steps (Breiman and Cutler, 2003).

1. Draw a random sample of data of size m using the bootstrap resampling technique with returns.
2. Bootstrap samples are used to train the classification tree model for each node node.
   a. Randomly select as many features as d.
   b. The splitting of feature nodes is done by determining the optimal value based on the objective function. In classification tree building, this process involves dividing the data at each parent node into two child nodes, according to the following rules (Breiman *et al.*, 1984).
      i. Each splitting process is determined by the value obtained from one predictor variable.
      ii. For continuous predictor variables, all candidate separation nodes are determined by calculating the midpoint between two sorted values of variable Xr. Mathematically, the calculation can be seen as follows.

$$S_r = \frac{X_r + X_{r+1}}{2} \quad (3)$$

      where S_r is the *r-th* split point and X_r is the *r-th* observation.

      iii. Splitting node used in Random Forest Classification is by calculating impurity in a node using Gini index, the calculation can be seen as follows.

$$Gini\ Index = 1 - \sum_{a=1}^{b} \pi_a^2 \quad (4)$$

      Descriptions:

      π_a : Probability of the *a-th* class where *a* is a positive or negative label.

      b : The number of existing classes

      iv. Calculate the best split from the threshold value by minimizing the gini split. The gini split calculation can be seen as follows.

$$Gini\ Split = \sum_{a=1}^{b} \left(\frac{p_a}{p}\right) \times Gini\ Index \quad (5)$$

      Descriptions:

      p_a : Number of samples on the *a-th* branch after *split*

      p : Number of samples before *split*

3. Repeat steps 1 and 2 T times until a collection of trees forming a forest is formed. At the end of the leaf node (terminal node), the average observations in that area are calculated. Simplifying the model can be done like pruning a tree because there are

many possible subtrees. The model simplification is based on certain criteria until a simpler, more efficient, and still accurate model is obtained (Breiman *et al*., 1984). According to De'ath and Fabricius (2000), the homogeneity of a node is measured using the impurity level, i.e., the impurity value used will be zero when the node is fully homogeneous.

4. The prediction of sentiment label y is determined based on the majority vote of all decision trees, which combines the predictions from all decision tree.

Hyperparameter tuning is essential for finding the optimal combination of hyperparameters, as each one significantly impacts model performance. In this study, the hyperparameter tuning used is GridSearchCV to explore hyperparameter combinations and calculate cross-validation scores. It utilizes K-Fold Cross Validation, and the training data is divided into *k* sections, where the model is evaluated *k* times, with each part being used alternately as testing data. The average accuracy from these iterations helps identify the best hyperparameter combination. The following is an application of K-Fold Cross Validation (Ogundunmade *et al*., 2022):

1. Data is divided randomly into *k* subsets.
2. On each subset of the data generated:
   a. A subset is selected as testing data for each iteration, while the other subset is used for training data.
   b. The model is trained using training data then the performance of the model is evaluated using testing data., which can be calculated using the following Equation (6) (Mardiana *et al*., 2022):

$$Accuracy = \frac{\sum correctly\ classified\ testing\ data}{\sum total\ testing\ data} \tag{6}$$

3. Repeat steps 1 and 2 up to K times, or until the model has been trained and tested on all subsets.
4. The model's overall prediction error is evaluated by averaging the prediction errors of each tree.

The classification results are evaluated by using a confusion matrix and AUC-ROC curves serve to assess prediction accuracy and overall model performance. Confusion matrix compares correctly and incorrectly classified data. (Normawati and Prayogi, 2021). Meanwhile, AUC (Area Under the Curve) evaluates performance in binary classification. The accuracy is further illustrated through the confusion matrix table.

Table 1. Confusion Matrix Table

Actual Classification	Prediction Classification	
	Positive Class	Negative Class
Positive Class	True Positive (TP)	False Negative (FN)
Negative Class	False Positive (FP)	True Negative (TN)

There are parameters to evaluate classification performance including accuracy, precision, recall, and f1-score.

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

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

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

$$F1 - Score = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{10}$$

AUC (Area Under Curve) is a measure that shows how much area is under the ROC (Receiver Operating Characteristic) curve to evaluate a model's ability to distinguish between positive and negative classes. The ROC curve shows the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds (Stern, 2021). The AUC is calculated by summing the trapezoidal areas between ROC points and is categorized into five interpretation levels as in the following table (Rahmi *et al*., 2023).

Table 2. Interpretation of AUC Values

AUC Value	Descriptions
0,9 - 1	Very high accuracy
0,8 – 0,9	High accuracy
0,7 – 0,8	Medium accuracy

Corresponding Author: Rahmila Dapa

0,6 – 0,7	Weak accuracy
0,5 – 0,6	False accuracy

3. DATA AND METHODOLOGY

This study is based on data taken from X in the form of 3015 Indonesian language tweets obtained with the keyword "makan bergizi gratis". The data collected are tweets posted when the program began to be implemented more widely in the regions, namely January 13-14, 2025. The period was chosen because it coincided with the trial implementation of the free nutritious meal program in several cities in Indonesia. The independent variable is the number of terms contained in public tweets on social media X, while the dependent variable in this study is the sentiment label, namely positive and negative. The following are the stages of analysis applied in this study:

1. Data Retrieval.
2. Data Pre-Processing (case folding, data cleaning and word normalization).
3. Data Labelling
4. Feature Selection (tokenizing, stopword removal, and stemming).
5. TF-IDF Weighting
6. Splitting Data is done by trial and error on the proportion of training data and testing data of 90:10, 80:20, 75:25, and 70:30.
7. Hyperparameter Tuning (GridSearchCV and K-Fold Cross Validation).
8. Random Forest Classification
   Perform classification with random forest on training data.
   a. Randomly draw *m* data samples from the bootstrap dataset.
   b. Train the classification tree model using bootstrap samples for each node.
   c. Repeat steps a and b T times until a collection of trees forming a forest is formed.
   d. Predict the sentiment class *y* in random forest classification based on the majority vote of all trees.
9. Model Evaluation with Confusion Matrix and AUC values.
10. Wordcloud Visualization

4. RESULTS AND DISCUSSION

This study consists of 2803 Indonesian tweets collected through scraping using Python at Google Collaboratory. The initial dataset consisted of 3,015 tweets containing the keyword "makan bergizi gratis”. Data pre-processing, including the removal of duplicates and incomplete tweets, resulted in 2,803 tweets ready for analysis. Data pre-processing produces a form of data that is ready to be processed before going into further classification analysis process.

In this study, the results of tweet classification in the form of positive and negative sentiments are automatically processed based on the InSet Lexicon dictionary. There are 2803 tweets about the free nutritious meal program, 2371 tweets have positive sentiment labels and 432 tweets have negative labels. This shows that most people support the program. The feature selection process is carried out to reduce the dimensional space of the document vector and to improve the quality of analysis.

The results of the pre-processing and feature selection process show that there are 1980 words that appear in the entire dataset. An example of manual calculation of TF-IDF weighting is as follows.

$$W_{anggota,tweet1} = \frac{n_{i,j}}{\sum T_j} \times \left(ln \left(\frac{N}{df_{(i)}} \right) + 1 \right)$$

$$W_{anggota,tweet1} = \frac{1}{31} \times \left(ln \left(\frac{2803}{21} \right) + 1 \right) = 0{,}190127$$

Table 3. TF-IDF Weighting Results

Tweet Number	abang	⋯	anggota	⋯	program	⋯	zamroni
1	0	⋯	0.190127	⋯	0.073151	⋯	0
2	0	⋯	0	⋯	0.226768	⋯	0
3	0	⋯	0	⋯	0.412305	⋯	0
4	0	⋯	0	⋯	0.125982	⋯	0
5	0	⋯	0	⋯	0.107985	⋯	0
6	0	⋯	0	⋯	0	⋯	0
7	0	⋯	0.392928	⋯	0	⋯	0

Corresponding Author: Rahmila Dapa

8	0	⋯	0	⋯	0.251964	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
2803	0	⋯	0	⋯	0	⋯	0

The division of training data and testing data is carried out on word data that has passed the weighting process. Data division is done by trial and error with the proportion of training data and testing data 90:10, 80:20, 75:25, and 70:30. This study uses GridSearchCV and K-Fold Cross Validation to determine the best parameters in the random forest model. The parameter values used in the hyperparameter tuning process can be seen as follows.

1. *max_features* = sqrt, log2
2. *n_estimators* = 50, 100, 200
3. *max_depth* = 10, 20, 40
4. *min_sample_split* = 2, 10
5. *min_sample_leaf* = 2, 4

The following results of the comparison of accuracy values in each data splitting performed are shown in Table 4.

Table 4. Measures of Model Goodness for All Proportions

Proportion Training:Testing		Model Goodness Measure	
		Accuracy	AUC Value
90:10	Training Data	0.8913	0.9476
	Testing Data	0.8754	0.8723
80:20	Training Data	0.8907	0.9455
	Testing Data	0.8377	0.8431
75:25	Training Data	0.8691	0.8561
	Testing Data	0.8516	0.8000
70:30	Training Data	0.8532	0.8433
	Testing Data	0.8180	0.7620

Based on the hyperparameter tuning results in Table 4, it is obtained that splitting data with a proportion of 90:10 has the best accuracy value among other proportions, so that proportion will be used for future analysis. Based on the hyperparameter tuning output using GridSearchCV on the random forest classification model have the best parameters namely the value of max_features = sqrt, n_estimators = 100, max_depth = 40, min_sample_split = 2, and min_sample_leaf = 2. The following Table 5 is the score of the k-fold cross validation accuracy value at 90:10 data splitting.

Table 5. K-Fold Cross Validation Score

Fold-	Evaluation (accuracy)
1	0.85770
2	0.85375
3	0.84920
4	0.84127
5	0.85317
6	0.84127
7	0.86507
8	0.82539
9	0.81746
10	0.87301
Average	0.84773

The K-Fold Cross Validation accuracy value obtained is 0.84773, indicating that the model is able to make accurate predictions. Based on Table 4, the results show that the model has a high accuracy value with the AUC value of the training testing data of 0.9476 and 0.8723. This relatively high AUC value indicates that the model is able to predict public opinion sentiment with high accuracy, both on trained data and on new data.

Corresponding Author: Rahmila Dapa

Table 6. Random Forest Classification Confusion Matrix

Actual Training	Training Prediction		Actual Testing	Testing Prediction	
	Positive	Negative		Positive	Negative
Positive	1981	152	Positive	219	19
Negative	122	267	Negative	16	27

Based on Table 6, the evaluation results on the testing data using the following precision, recall, and f1-score values can be obtained.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} = \frac{(219+27)}{(219+16+27+19)} = 0.8754$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{219}{(219+16)} = 0.9319$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{219}{(219+19)} = 0.9201$$

$$F1-Score = \frac{2 \times Precision \times Recall}{(Precision+Recall)} = \frac{2 \times 0.9319 \times 0.9201}{(0.9319+0.9201)} = 0.9260$$

Based on Table 4, the model with a 90:10 data split shows strong performance, with training and testing accuracy of 89.13% and 87.54%. The recall, precision, and f1-score values for the testing data are 92.01%, 93.19%, and 92.6%, indicating the model's strong ability to accurately classify the two classes. The training-testing AUC values reached 0.9476 and 0.8723, further confirming the model's effectiveness. The small accuracy gap between training and testing data indicates good generalization ability without overfitting, which means the hyperparameter configuration was effective applied can be considered effective for capturing patterns in the data optimally.

Wordcloud visualization is used to show words that are often used in positive and negative labels. The following are the results of the wordcloud based on public opinion towards the free nutritious meal program.

Figure 1. Positive and Negative Labels Wordcloud

Based on the positive wordcloud, the word that appears most often in positive sentiment tweets is "Japan", this word is most likely related to support for the free nutritious meal program. Then the negative wordcloud shows the word that appears most often in negative sentiment tweets is “program”, most likely this word is related to criticism of the program.

5. CONCLUSIONS

Based on the analysis and discussion of the implementation of the random forest algorithm in the classification of public sentiment towards the free nutritious meal program on X social media. The results of data labeling or sentiment scoring based on the InSet Lexicon dictionary obtained positive labels as many as 2371 tweets and negative labels as many as 432 tweets. This shows that the majority of tweets express positive sentiment toward the program. Classification of public sentiment using random forest algorithm with the proportion of training data and testing data of 90:10 produces optimal parameters, namely max_features = $sqrt(1980) \approx 44$, n_estimators = 100, max_depth = 40, min_sample_split = 2, and min_sample_leaf = 2 obtained from the classification model produces training-testing data accuracy values of 89.13% and 87.54%. The recall, precision, and f1-score values for testing data are 92.01%, 93.19%, and 92.6%. The AUC value of training-testing data is 0.9476 and 0.8723. Based on the results of wordcloud visualization, the one that appears most in positively labeled tweets is "Japan" 1729 times. This indicates that the word was frequently associated with supportive opinions for the free nutritious meal program. Meanwhile, the word that appeared the most in negatively labeled tweets was "program" 145 times. This word most likely appeared in contexts containing criticism of the program.

REFERENCES

1. Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Penerapan algoritma Random Forest untuk analisis sentimen komentar di YouTube tentang Islamofobia. *Jurnal Nasional Komputasi dan Teknologi Informasi, 5*(1).
2. Breiman, L., & Cutler, A. (2003). *Manual on setting up, using, and understanding Random Forest V4.0*.
3. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
4. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
5. Buyukkececi, M., & Okur, M. C. (2023). A comprehensive review of feature selection and feature selection stability in machine learning. *Journal of Science, 36*(4), 1506–1520.
6. De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology, 81*(11), 3178–3192.
7. Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis sentimen terhadap aplikasi Ruangguru menggunakan algoritma Naive Bayes, Random Forest dan Support Vector Machine. *Transformatika, 18*(1), 71–80.
8. Intan, R., & Defeng, A. (2006). Hard: Subject-based search engine menggunakan TF-IDF dan Jaccard's coefficient. *Jurnal Teknik Industri: Jurnal Keilmuan dan Aplikasi Teknik Industri, 8*(1), 61–72.
9. Koto, F., & Rahmaningtyas, G. Y. (2017). InSet Lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *IEEE Xplore*, 391–394.
10. Mailo, F. F., & Lazuardi, L. (2021). Analisis sentimen data Twitter menggunakan metode text mining tentang masalah obesitas di Indonesia. *Journal of Information Systems for Public Health, 6*(1), 28–36.
11. Mardiana, L., Kusnandar, D., & Satyahadewi, N. (2022). Analisis diskriminan K-fold cross validation untuk klasifikasi kualitas air di Kota Pontianak. *Buletin Ilmiah Matematika, Statistika dan Terapannya (Bimaster), 11*(1), 97–102.
12. Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier dan Confusion Matrix pada analisis sentimen berbasis teks pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI), 5*(2), 697–711.
13. Ogundunmade, T. P., Adepujo, A. A., & Allam, A. (2022). Stock price forecasting: Machine learning models with K-fold and repeated cross validation approaches. *Modern Economy and Management, 1*(1).
14. Rahmi, I. A., Afendi, F. M., & Kurnia, A. (2023). Metode AdaBoost dan Random Forest untuk prediksi peserta JKN-KIS yang menunggak. *Jambura Journal of Mathematics, 5*(1), 83–94.
15. Stern, R. H. (2021). Interpretation of the area under the ROC curve for risk prediction models. *arXiv*.

Corresponding Author: Rahmila Dapa