ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Sentiment Analysis of Teacher Bullying Discourse in Indonesia Using Naive Bayes, Support Vector Machines, and Random Forest

Yenni Lestari Sitanggang[1], Ahmad Zamsuri[2], Yuvi Darmayunata[3], Guntoro[4]

[1,2,3,4]Faculty of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia

*Corresponding Author
Email: yennylestarisitanggang@gmail.com

Received: 01/12/2025 Revised: 15/12/2025 Accepted: 20/12/2025 Published: 29/12/2025

Abstract

Abstract

Bullying against teachers on social media has become an increasingly discussed issue, particularly on the X platform. This study aims to analyze public sentiment toward teacher bullying in Indonesia using a machine learning approach. Three classification methods are employed, namely Naive Bayes, Support Vector Machine (SVM), and Random Forest. The dataset consists of 3,351 tweets after preprocessing. Sentiment labels (positive, negative, and neutral) are assigned using a rule based approach, relying on a predefined keyword list. Feature extraction is conducted using TF-IDF, and classification performance is evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show that Naive Bayes achieves an accuracy of 71.83 percent, Support Vector Machine 86.89 percent, and Random Forest 87.48 percent. Based on these findings, Random Forest demonstrates the best performance for classifying sentiment on the issue of teacher bullying on X

Keywords: Sentiment Analysis, X Platform, Teacher Bullying, Naive Bayes, Support Vector Machine, Random Forest, Machine Learning

Introduction

Bullying in educational settings, both in Indonesia and globally, remains a serious and persistent problem. In Indonesia, bullying has increasingly been discussed in public discourse, including bullying directed at teachers through face to face interactions as well as through digital channels. Although moral education is formally promoted, values such as tolerance and empathy do not always translate into everyday practice, which can normalise intimidation toward individuals perceived as vulnerable. Empirical work also points to multiple contributing factors, including family conditions, school environment, peer groups, mass media exposure, and students' limited empathy (Andryawan et al., 2023).

The rapid development of technology and social media platforms such as X (formerly Twitter) further accelerates the circulation of public opinions from diverse perspectives, supported by features such as trending topics. These

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

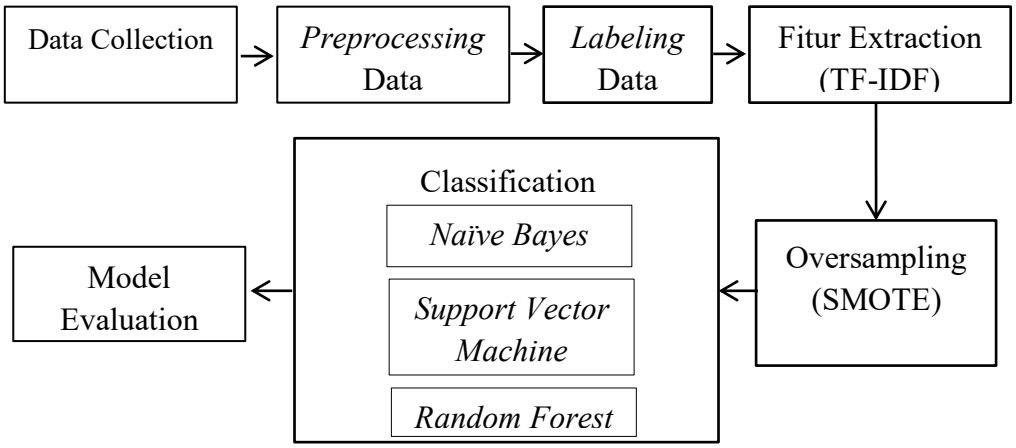
opinions may be positive, negative, or neutral. Sentiment analysis is therefore widely used as an effective approach to automatically identify the direction of public opinion in large scale text data (Raisa & Riza, 2023).

A key methodological challenge in social media sentiment classification is class imbalance, where the number of instances across sentiment categories is uneven and can bias model learning. Oversampling techniques such as SMOTE are commonly applied to rebalance the training data so that classifiers can learn more optimally from minority classes (Adi et al., 2024).

Prior comparative studies have reported mixed findings, indicating that model performance is highly dependent on topic, language, and dataset characteristics. For example, Random Forest has been reported as performing strongly for depression related sentiment in Twitter text (Fachriza & Munawar, 2023), while other contexts show competitive results for Support Vector Machine (Zamsuri & Wileks, 2024) and evidence that Naive Bayes can remain robust for certain economic and industrial topics (Susanto & Dzulkarnain, 2023). Building on this motivation, the present study aims to classify public opinion on X regarding teacher bullying in Indonesia using Naive Bayes, Support Vector Machine, and Random Forest, and to compare their performance to identify the most effective method. The findings are expected to inform government efforts to strengthen teacher protection policies and improve educational quality.

Materials and Methods

This section explains the research workflow, data source, study setting, evaluation strategy, and the algorithms used. Figure 1 summarizes the end to end pipeline, beginning with data collection and ending with model evaluation.



Gambar 1 Research workflow

Data Collection
Data were collected by scraping posts from the X platform. The target data consisted of public opinions and discussions related to teacher bullying in Indonesia. Data acquisition was implemented in Python using Google Colab, yielding 3,355 tweets.


ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Data Pre processing

Pre processing was conducted to clean and standardize raw tweets so that downstream feature extraction and classification could be performed reliably. Because raw social media text often contains punctuation, numerals, links, and other non linguistic artifacts, multiple cleaning steps were applied. After pre processing, 3,351 tweets remained.

Steps included:

a. **Cleaning**
   Removal of punctuation, URLs, user mentions, hashtags, numerals, emojis, and special characters.
b. **Case folding**
   Conversion of all characters to lowercase to reduce variance caused by capitalization.
c. **Normalization**
   Conversion of non standard forms, slang, and informal spellings into standard Indonesian forms aligned with KBBI conventions.
d. **Tokenization**
   Segmentation of each tweet into individual tokens. Tokens separated by whitespace were stored for subsequent weighting and modeling.
e. **Stopword removal**
   Removal of high frequency functional words that contribute limited sentiment signal, such as dan, di, ke, yang, and similar forms, to reduce noise and improve efficiency.
f. **Stemming**
   Reduction of tokens to their base forms to consolidate morphological variants that convey equivalent meaning.

Data Labeling

The dataset was labeled into three sentiment categories: positive, negative, and neutral. Labeling was performed automatically using a rule based procedure that relied on predefined keywords. Positive tweets contained supportive expressions, negative tweets contained insults or condemnation, and neutral tweets primarily conveyed information without a clear evaluative stance. The labeled dataset was then split into 80 percent training data and 20 percent testing data so the models could learn patterns before being evaluated on unseen instances.

Feature Extraction using TF IDF

Feature extraction converts text into numerical representations suitable for machine learning. This study used TF IDF weighting to emphasize informative terms while down weighting ubiquitous words across the corpus.

Class Balancing using SMOTE

SMOTE (Synthetic Minority Over sampling Technique) was used to address class imbalance when one sentiment category is under represented relative to others. Importantly, oversampling was applied only to the training data to prevent data leakage and overly optimistic evaluation.

Model Implementation (Naive Bayes, Support Vector Machine, Random Forest)

1. **Naive Bayes**
   Naive Bayes is a probabilistic classifier that assumes conditional independence among features. It is widely used for text classification due to its computational efficiency and strong baseline performance.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

2. **Support Vector Machine**
   Support Vector Machine is a supervised method that separates classes by constructing an optimal decision boundary (hyperplane) in a high dimensional space. The method seeks a maximum margin separation to improve generalization.
3. **Random Forest**
   Random Forest is an ensemble learning method that aggregates multiple decision trees to improve predictive accuracy and reduce overfitting, producing more stable predictions than a single tree.

Model Evaluation
Model performance was evaluated using accuracy, precision, recall, F1 score, and the confusion matrix, following standard classification evaluation practice. Let TP denote true positives, TN true negatives, FP false positives, and FN false negatives.

1. **Accuracy**
   Accuracy is the proportion of correct predictions among all predictions.

   $$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

2. **Precision**
   Precision measures how many predicted positives are correct.

   $$\text{Precision} = TP / (TP + FP)$$

3. **Recall**
   Recall measures how many actual positives are correctly identified.

   $$\text{Recall} = TP / (TP + FN)$$

4. **F1 score**
   F1 score is the harmonic mean of precision and recall, providing a balanced metric under class imbalance.

   $$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Results and Discussion

Dataset Overview and Sentiment Distribution
Tweets related to teacher bullying in Indonesia were collected from the X platform using Python in Google Colab. The scraping process produced 3,355 tweets. After data preprocessing, which removed noise and standardized the text, 3,351 clean tweets remained for analysis. Sentiment labels were assigned automatically using a rule based approach grounded in predefined keywords, resulting in three classes: positive, negative, and neutral.

The labeled distribution indicates a clear class imbalance. Of the 3,351 tweets, 1,456 were labeled as negative, making this the majority class. Neutral tweets accounted for 1,393 instances, while positive tweets totaled 502.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

This distribution suggests that supportive discourse toward teachers was substantially less frequent than negative or neutral commentary. The sentiment distribution is illustrated in Figure 2.

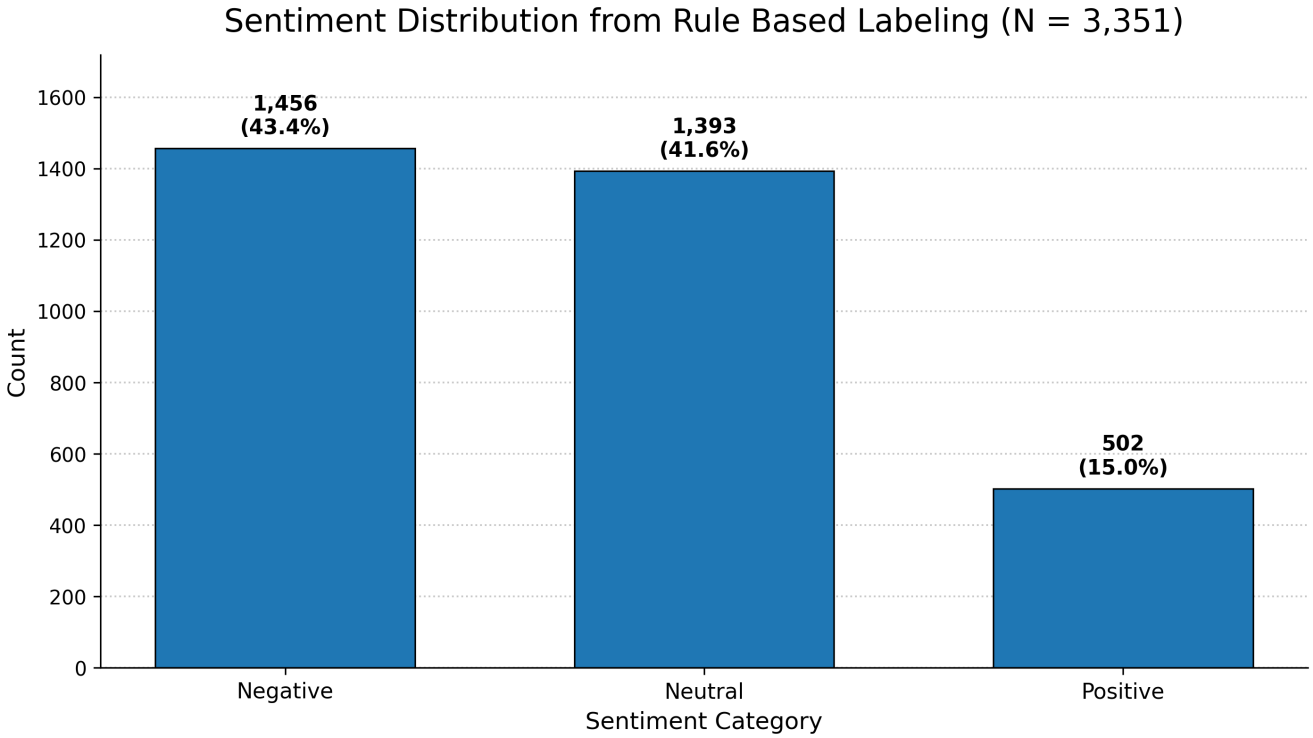


Figure 2. Sentiment label distribution (automatic rule based labeling).

Train Test Split, Feature Representation, and Class Balancing

Following scraping, preprocessing, and labeling, the dataset was split using an 80:20 ratio, where 80 percent of the data were used for training and 20 percent were reserved for testing. This split is commonly used to support robust generalization assessment on unseen data.

Text features were represented using TF IDF, which assigns weights to terms based on their frequency in a tweet and their rarity across the corpus. This weighting emphasizes discriminative terms and reduces the influence of ubiquitous words, supporting more informative classification.

Because the positive class was a minority, SMOTE was applied to the training subset only. Applying SMOTE exclusively to training data prevents leakage into the evaluation set. After SMOTE, the training data became balanced with 1,164 instances per sentiment class. This balancing strategy reduces the tendency of models to over predict the majority class and supports improved learning for minority sentiments.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

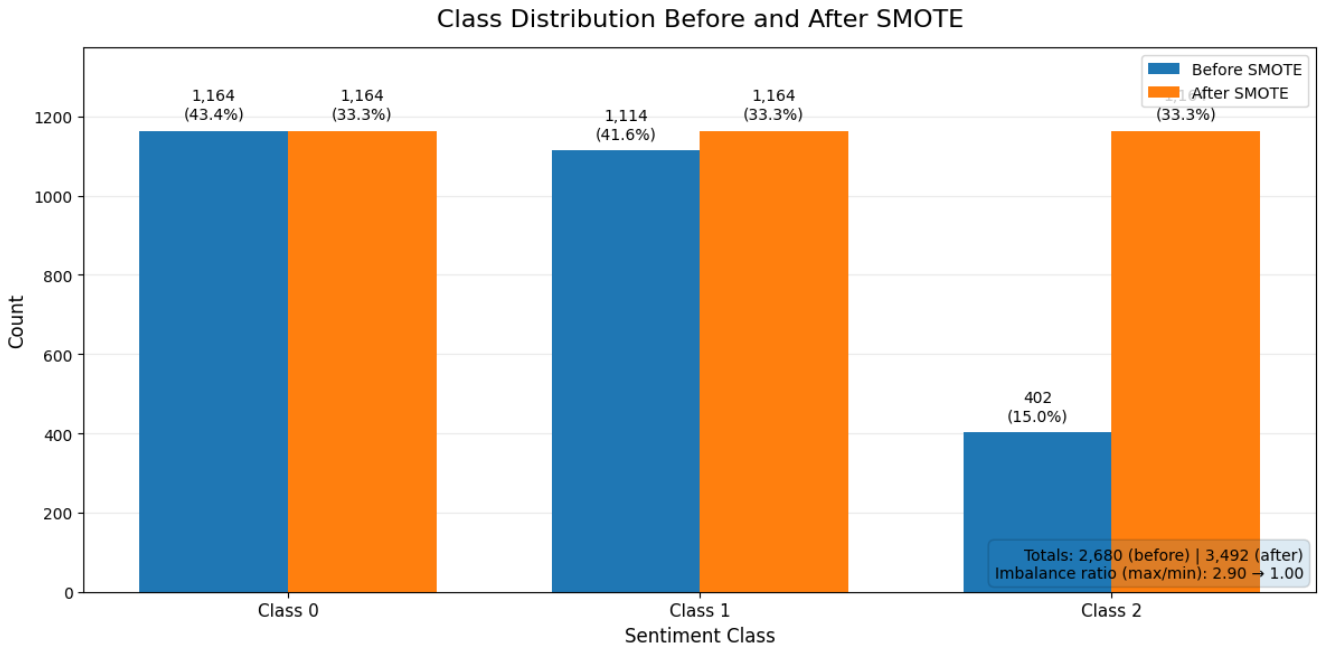


Figure 3. Training data distribution after SMOTE balancing.

Classification Performance

Model performance was evaluated on the test set (n = 671) using accuracy, precision, recall, and F1 score. Tables 1 to 3 summarize the classification results for Naive Bayes, Support Vector Machine, and Random Forest.

Table 1. Model evaluation: Naive Bayes

Sentiment	Precision	Recall	F1 score	Support
Negative	0.78	0.74	0.76	292
Neutral	0.81	0.72	0.77	279
Positive	0.44	0.63	0.52	100
Accuracy			0.72	671
Macro avg	0.68	0.70	0.68	671
Weighted avg	0.74	0.72	0.73	671

Naive Bayes achieved an accuracy of 0.72 on the test set. Performance was comparatively strong for the negative and neutral classes, while the positive class remained challenging. The low precision for the positive class indicates that many tweets predicted as positive were actually neutral or negative, which is plausible given lexical overlap and the limited size of positive examples.

Table 2. Model evaluation: Support Vector Machine

Sentiment	Precision	Recall	F1 score	Support
Negative	0.91	0.86	0.89	292
Neutral	0.84	0.95	0.89	279
Positive	0.82	0.68	0.74	100

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Accuracy			0.87	671
Macro avg	0.86	0.83	0.84	671
Weighted avg	0.87	0.87	0.87	671

Support Vector Machine achieved an accuracy of 0.87, with consistently high performance for negative and neutral tweets. The neutral class recall of 0.95 indicates that the model successfully captured most neutral instances. Although positive performance improved substantially compared with Naive Bayes, recall for positive tweets remained lower than for the other classes, suggesting that some positive content was still mapped to neutral or negative categories.

Table 3. Model evaluation: Random Forest

Sentiment	Precision	Recall	F1 score	Support
Negative	0.93	0.86	0.89	292
Neutral	0.81	0.99	0.89	279
Positive	1.00	0.60	0.75	100
Accuracy			0.87	671
Macro avg	0.91	0.82	0.84	671
Weighted avg	0.89	0.87	0.87	671

Random Forest achieved the highest accuracy among the evaluated models at 0.87. The model produced very strong results for negative and neutral classes, including a neutral recall of 0.99. For the positive class, the model achieved perfect precision but only moderate recall, indicating that when the model predicted positive it was almost always correct, yet it missed a meaningful portion of truly positive tweets. This pattern suggests a conservative decision boundary for the positive class.

Confusion Matrix Analysis

The confusion matrix provides a more granular view of model behavior across sentiment classes. For Naive Bayes, 217 of 292 negative tweets were correctly classified, while 26 were misclassified as neutral and 49 as positive. In the neutral class, 202 of 279 tweets were correctly classified, with misclassifications split between negative (46) and positive (31). For the positive class, only 63 of 100 tweets were correctly classified, confirming that the model struggled most with minority positive sentiment.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

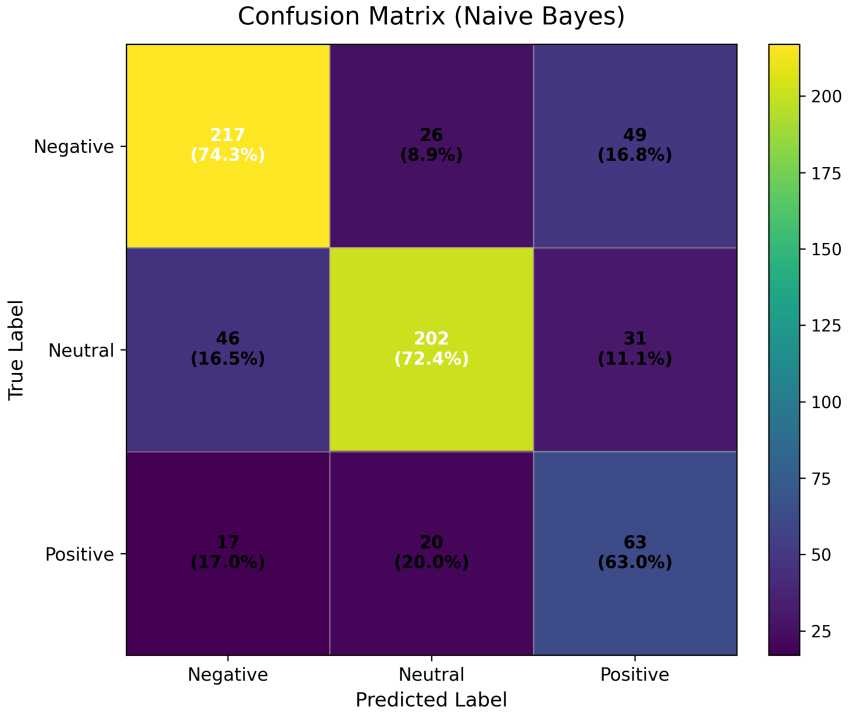


Figure 4. Confusion matrix: Naive Bayes

Support Vector Machine showed a more stable pattern. In the negative class, 251 of 292 tweets were correctly classified, while 31 were assigned to neutral and 10 to positive. In the neutral class, 264 of 279 were correctly classified, with relatively few confusions into negative (10) or positive (5). In the positive class, 68 of 100 were correctly classified, with the remaining errors split into negative (14) and neutral (18).

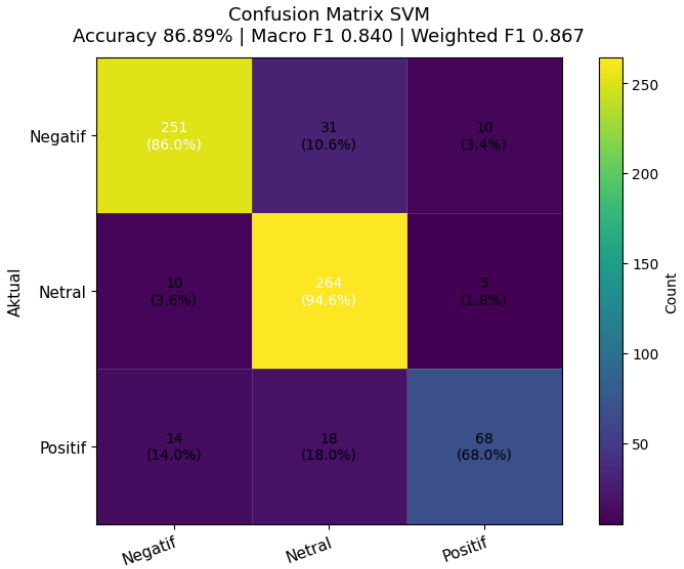


Figure 5. Confusion matrix: Support Vector Machine

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Random Forest delivered particularly strong recognition of neutral sentiment. For negative tweets, 250 of 292 were correctly classified and 42 were misclassified as neutral. For neutral tweets, 277 of 279 were correctly classified, with only 2 misclassified as negative. For positive tweets, 60 of 100 were correctly classified, while 16 were misclassified as negative and 24 as neutral. The absence of negative tweets misclassified as positive suggests that the model maintained strict separation between negative and positive signals.

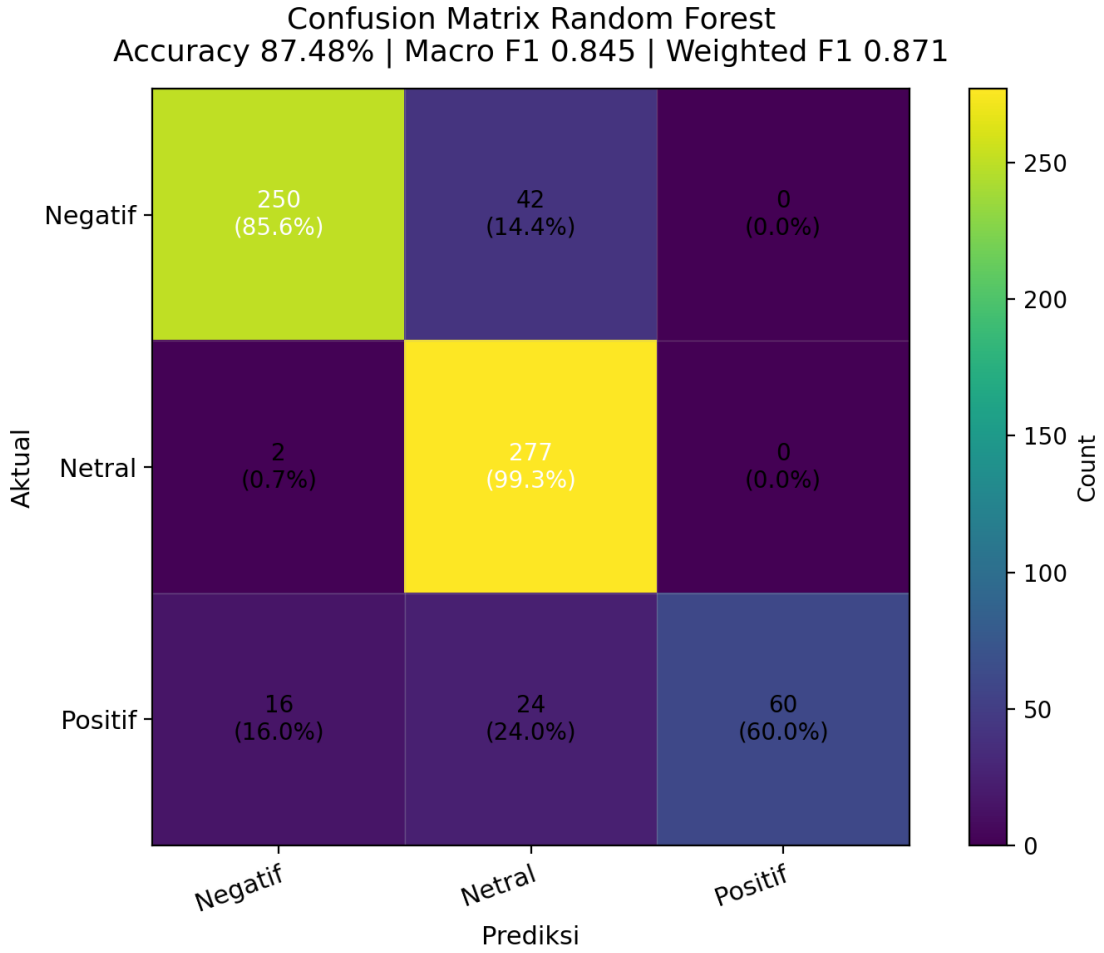


Figure 6. Confusion matrix: Random Forest

Comparative Summary and Implications

A direct comparison of overall accuracy shows that Naive Bayes achieved 0.72, Support Vector Machine achieved 0.87, and Random Forest achieved 0.87. The improvement from Naive Bayes to the other models is substantial, consistent with the advantage of margin based and ensemble approaches on high dimensional sparse text representations. The difference between Support Vector Machine and Random Forest is comparatively small, indicating that both provide strong baselines for sentiment classification in this setting.

Despite SMOTE balancing, the positive class remains the most challenging category across models. This outcome can be explained by the smaller diversity of positive expressions, overlapping vocabulary between neutral and

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

supportive statements, and potential labeling noise from keyword based annotation. Practically, these findings suggest that public discourse on teacher bullying is dominated by negative and neutral content, while supportive sentiment is less frequent. From a policy perspective, the sentiment landscape can inform targeted communication and intervention strategies, especially to strengthen teacher protection and promote constructive digital engagement.

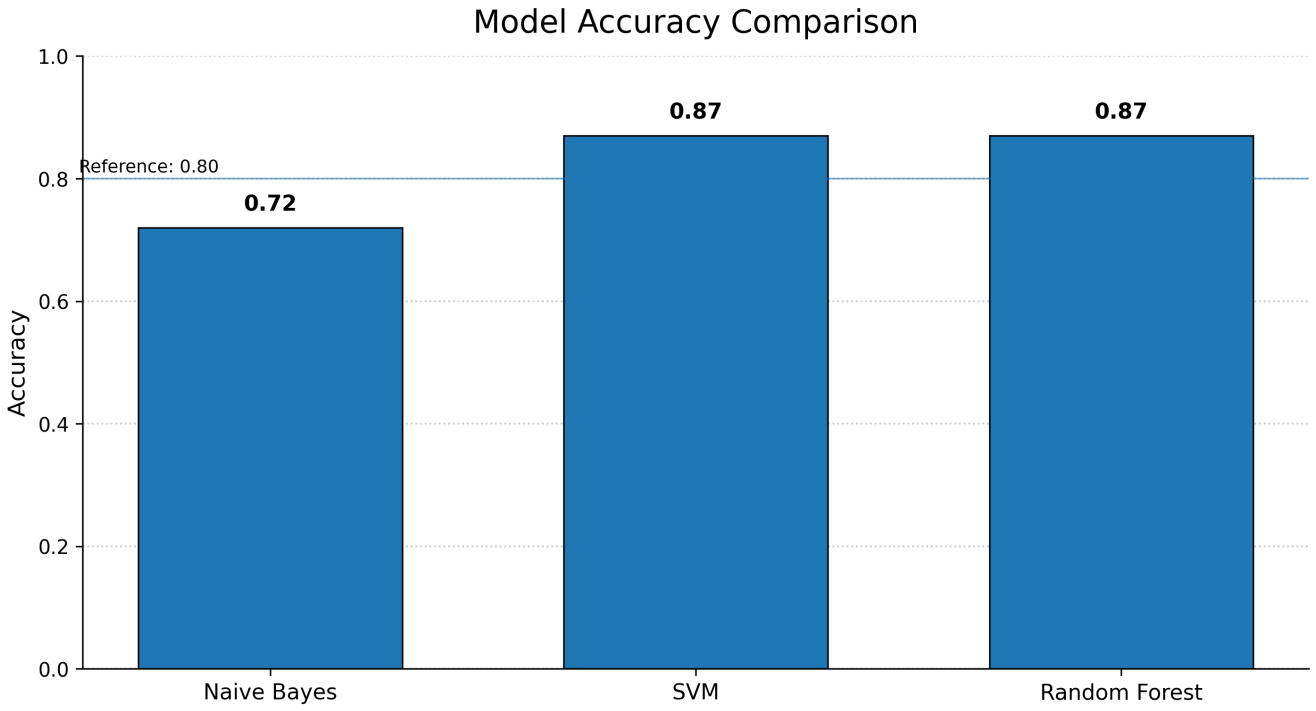


Figure 7. Accuracy comparison of Naive Bayes, Support Vector Machine, and Random Forest.

Conclusion

This study reports comparative classification accuracies of 71.83% for Naive Bayes, 86.89% for Support Vector Machine (SVM), and 87.48% for Random Forest. The dataset comprised 3,355 raw tweets collected from X, and 3,351 tweets remained after preprocessing. Sentiment analysis was conducted using three polarity labels, namely positive, negative, and neutral. Labels were assigned automatically using a rule based classification approach grounded in a predefined keyword lexicon. Based on the experimental evaluation, Naive Bayes achieved the lowest accuracy, whereas Random Forest delivered the strongest performance for sentiment detection related to teacher bullying in Indonesia. These results indicate that Random Forest is more effective in capturing sentiment patterns and class distribution in the dataset than the other two methods.

For future work, it is recommended to expand the volume of tweet data to improve model robustness and potentially increase predictive accuracy. In addition, subsequent studies should consider deep learning approaches for sentiment analysis on teacher bullying discourse in Indonesia, as such methods are more capable of modeling nuanced linguistic context and learning complex sentiment representations beyond keyword driven signals.

References

Adi, S., Mola, S. A. S., Baun, D. L. B., & Nunes, I. O. (2024). Analisis sentimen aplikasi Halo BCA di Google Play Store menggunakan Naive Bayes dan Random Forest. HOAQ, 15(2), 69–79. https://doi.org/10.52972/hoaq.vol15no2.p69-79

Andryawan, A., Laurencia, C., & Putri, M. P. T. (2023). Peran guru dalam pencegahan perundungan di lingkungan pendidikan. INNOVATIVE: Journal of Social Science Research, 3(6), 2837–2850.

Fachriza, M., & Munawar, M. (2023). Analisis sentimen kalimat depresi di Twitter. Komputek, 7(2), 49–58. https://doi.org/10.24269/jkt.v7i2.2218

Raisa, N., & Riza, N. (2023). Sentimen analisis terhadap opini masyarakat pada media sosial. JATI (Jurnal Mahasiswa Teknik Informatika), 7(2), 1312–1320. https://doi.org/10.36040/jati.v7i2.6765
Susanto, A., & Dzulkarnain, I. A. (2023). Analisis sentimen pada topik ekonomi dan industri. Zenodo. https://doi.org/10.5281/zenodo.8398895

Zamsuri, A., & Wileks. (2024). Sentiment analysis of Saudi e commerce discourse using classical machine learning. International Journal of Data and Network Science, 8(3), 1607–1612. https://doi.org/10.5267/j.ijdns.2024.3.006