BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

Comparative Analysis of Machine Learning Algorithms for Indonesian Twitter Sentiment Classification on the Jakarta–Bandung High-Speed Rail Project

Muhammad Noerhadi[1*], Budiman, Sardjono

Information System[1]
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia[1]
https://unibi.ac.id/[1]
Corresponding e-mail: muhammad.n@mhs.unibi.ac.id[1*]

Abstract. The rapid growth of social media in Indonesia has opened up new opportunities to gauge public opinion on major national initiatives. One of the most controversial projects is the Jakarta–Bandung High-Speed Railway (KCJB), which has sparked mixed responses due to its financial, environmental, and socio-political implications. To meet the need for systematic analysis, this study applies sentiment analysis to Indonesian Twitter data to evaluate public perspectives on the KCJB project. This research uses a step-by-step methodology, including data collection via the Twitter API, text preprocessing, manual tagging into positive and negative sentiments, and feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Four machine learning algorithms—Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest—were trained and verified on stratified data splits, with performance evaluated using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The results show that SVM consistently outperforms other models, achieving up to 73% accuracy with balanced precision and recall, as well as the highest AUC value. These findings confirm the robustness of SVM in handling high-dimensional Indonesian text. In addition to its academic contribution to sentiment analysis in languages with limited resources, this research offers practical implications by providing data-driven insights for policymakers and businesses for real-time monitoring, strategic communication, and informed decision-making.

Keywords: sentiment analysis; social media; machine learning; Indonesian language; Support Vector Machine

1. INTRODUCTION

The rapid development of social media has significantly changed the landscape of public communication, allowing individuals to express their opinions and participate in discussions on various social issues [1]. Twitter, in particular, has emerged as one of the most influential platforms thanks to its real-time information dissemination and large user base [2], [3]. In Indonesia, with millions of active Twitter users, this platform provides a rich source of public opinion data that can be used to understand the public's perspective on government policies and infrastructure projects [4], [5]. One of the most discussed topics in recent years has been the Jakarta–Bandung Fast Train (KCJB) project, which has attracted both positive and negative responses from the public due to its financial, environmental, and socio-political implications [6].

Sentiment analysis, also known as opinion mining, offers a systematic approach to extracting and classifying public opinion from unstructured text data [7], [8], [9]. By applying natural language processing (NLP) and machine learning techniques, sentiment analysis enables the identification of positive and negative sentiments in large-scale datasets [10], [11]. Previous studies have explored various algorithms for sentiment classification, including Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest [12], [13], [14]. Although these methods show promising results, the relative performance of these algorithms in the context of Indonesian-language datasets—especially those related to public opinion on national infrastructure projects—has been under-explored.

This study aims to address this research gap by conducting a comparative analysis of four commonly used

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

classification algorithms—Naïve Bayes, SVM, K-NN, and Random Forest—for sentiment analysis of public opinion related to the Jakarta–Bandung High-Speed Railway project. The evaluation was conducted using standard performance metrics such as accuracy, precision, recall, F1-score, and Area Under Curve (AUC). These findings are expected to provide insights into the most effective machine learning approaches for analyzing sentiment in Indonesian social media data, as well as contribute to a broader discussion on public acceptance of large-scale infrastructure initiatives.

The main contributions of this study can be summarized as follows. First, this study provides a comprehensive comparison of four machine learning algorithms commonly used in the context of Indonesian language sentiment analysis, highlighting the relative strengths and limitations of each. Second, this study demonstrates the application of sentiment analysis as a decision-making tool for assessing public response to large-scale infrastructure projects. Third, this study enriches the literature on sentiment analysis in resource-limited languages by providing empirical evidence from an Indonesian case study, thereby expanding the application of machine learning techniques beyond high-resource language contexts.

2. RELATED WORK

Twitter has become an important source of data for understanding public opinion on social issues and significant infrastructure projects, including transportation projects such as the Jakarta–Bandung High-Speed Railway. Previous studies have shown that machine learning and deep learning approaches offer promising results for sentiment classification on Twitter data, but still face challenges related to linguistic context, data diversity, and specific domains.

Sakirin et al. [15] introduced the ConvBiLSTM model, which combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) with word representations such as Word2Vec and GloVe. This model demonstrates the superiority of deep learning over traditional algorithms in sentiment classification. However, this approach still relies on word representations that may fail to capture the local nuances of the Indonesian language and specific social contexts, such as public opinion on the Jakarta–Bandung High-Speed Railway project.

Feature selection optimization is also an important consideration. Nagamanjula and Pethalakshmi [16] proposed a bi-objective optimization framework with LAN2FIS that improves classification accuracy through feature selection with minimal redundancy and maximum relevance. This approach is relevant for handling the heterogeneity of Twitter data, which includes variations in topics and language styles, but its application in the Indonesian context has rarely been studied.

An aspect-based approach has also been studied to produce a more detailed analysis. Janjua et al. [17] introduced Multilevel Hybrid Aspect-Based Sentiment Classification (MulehyaSC), which combines feature ranking and selection on several Twitter datasets. Although promising, aspect-based analysis requires rich annotation and often faces difficulties in capturing implicit sentiment—a challenge that is magnified when dealing with complex public discourse related to infrastructure projects.

A recent study by Chaudhary et al. [18] summarizes key challenges in deep learning-based sentiment analysis, such as the need for robust preprocessing of unstructured data, variation in emotional expression, and diversity in social context. Barreto et al. [19] emphasize the importance of robust word representation modeling to overcome linguistic noise such as slang, abbreviations, and spelling errors, which are common in Twitter conversations. Similar challenges are reported by Jacob and Vijayakumar [20], who explore clustering-based algorithms and highlight the need for accurate sentiment tagging at scale.

Additionally, several studies analyze the relative performance of various algorithms. Ayyub et al. [21] and AminiMotlagh et al. [22] show that Support Vector Machines (SVM) and deep learning generally outperform classical classification, but their performance is highly dependent on feature diversity and training data quality. Talaat [23] and Lin & Nuha [24] explore hybrid models based on BERT and DistilBERT combined with Bi-LSTM or TCN, resulting in better semantic understanding and higher accuracy. Multilingual research such as Manias et al. [25] highlights the potential of multilingual BERT models for classifying cross-language Twitter data, which is relevant given the use of code-switching on Indonesian Twitter.

In general, the literature shows that although BERT-based deep learning and hybrid approaches offer superior performance, there is still a research gap regarding systematic comparative analysis between algorithms on specific Indonesian datasets. In particular, studies exploring the performance of classical algorithms versus modern deep learning in the context of significant topics such as the Jakarta–Bandung High-Speed Rail project are still limited. Therefore, this study focuses on comparing the performance of various machine learning algorithms on Indonesian-language Twitter data discussing the JBHSR project, to provide empirical insights into the most effective algorithms in this context.

3. METHODS

The research methodology consists of several stages, namely data collection, text preprocessing, feature extraction, model development, and evaluation, as shown in Figure 1. Each stage is designed to ensure that the

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

sentiment analysis process produces valid, reliable, and reproducible results.

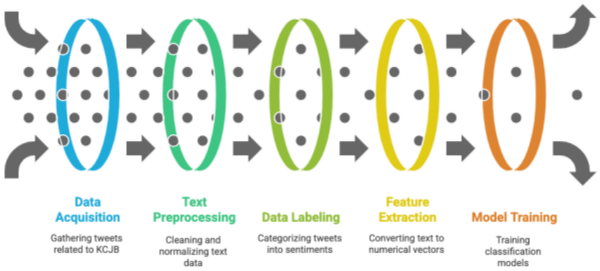

Fig. 1. Research Methodology

3.1. Data Acquisition.

Data Acquisition was carried out via the Twitter API using Python-based tools. Tweets containing specific keywords related to the Jakarta–Bandung Fast Train (KCJB) project were collected during the period from September to November 2023. After the crawling process, a total of 5,047 Indonesian-language tweets were obtained. To ensure data quality, irrelevant entries such as retweets, duplicates, and promotional content were removed. The resulting corpus was then tokenized, so that each tweet could be represented as a sequence of tokens suitable for further processing.

3.2. Text Preprocessing.

Text preprocessing was applied to clean and normalize the raw text data. This stage included cleaning irrelevant elements such as URLs, hashtags, emoticons, and mentions; applying case folding to standardize all characters to lowercase; tokenizing sentences into individual words; removing stopwords with limited semantic value; and applying stemming to reduce words to their root forms. These steps ensured that the dataset was structured and optimized for feature extraction and classification.

3.3. Data Labeling

The next stage is data labeling, where tweets are manually categorized into two sentiment classes: positive and negative. Based on the labeling process, 2,893 tweets were identified as positive sentiment, while 2,153 tweets were identified as negative sentiment, as shown in Figure 2. This distribution reflects the general tendency of public opinion to be more positive towards the KCJB project. The labeled dataset was then used as ground truth to train and validate the classification model.

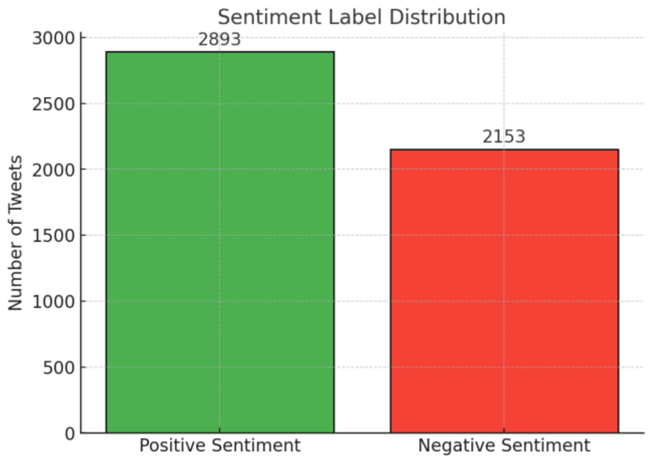

Fig. 2. Sentiment Label Distribution

3.4. Feature Extraction.

For feature extraction, the Term Frequency–Inverse Document Frequency (TF-IDF) method was used. TF-IDF effectively converts text data into numerical vectors by highlighting distinguishing terms while reducing the influence of common words. This representation was then used as the primary input for the classification algorithms.

3.5. Modeling

Four classification models were selected for comparative analysis: Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest. Naïve Bayes was chosen for its probabilistic approach and computational efficiency, SVM for its robustness in handling high-dimensional feature spaces, K-NN for its non-parametric classification based on similarity, and Random Forest for its ability in ensemble learning to reduce overfitting. All models were implemented using Python with support from the Scikit-learn library.

3.6. Model Training and Validation.

The data was divided into training and testing subsets using stratified splitting to maintain class balance.

3.7. Evaluation Metric.

Finally, model performance was assessed using multiple evaluation metrics. Accuracy (Acc) measures the proportion of correctly classified instances and is defined as Eq. (1):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where TP (True Positives) and TN (True Negatives) denote correct predictions, while FP (False Positives) and FN (False Negatives) represent misclassifications.

Precision (P) measures the reliability of positive predictions as Eq. (2):

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

$$P = \frac{TP}{TP + FP} \quad (2)$$

Recall (R), also referred to as sensitivity, measures the ability to identify relevant positive samples as Eq. (3):

$$R = \frac{TP}{TP + FN} \quad (3)$$

The F1-score is the harmonic mean of Precision and Recall as Eq. (4):

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (4)$$

Additionally, the Area Under the Curve (AUC) is derived from the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR). These rates are defined as Eq. (5):

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \quad (5)$$

Thus, the AUC can be expressed as the integral of the ROC curve as Eq. (6):

$$AUC = \int_0^1 TPR(FPR)\ d(FPR) \quad (6)$$

4. RESULTS AND DISCUSSIONS

4.1. Results.

This section shows the results of sentiment classification testing on Twitter data related to the Jakarta–Bandung Fast Train (KCJB) project. The evaluation was conducted using four machine learning algorithms, namely Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest. To ensure fairness and robustness, the dataset was divided into several training-testing ratios, namely 70–30, 80–20, and 90–10. Each model was trained and validated on these different partitions, and its performance was measured using standard metrics including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC).

The comparison results provide insight into how each classification performs under varying proportions of training and testing data. The use of multiple metrics ensures that the analysis not only captures overall accuracy but also reflects the model's ability to balance precision and recall as well as its discriminative power across sentiment classes. Table 1 summarizes the performance results of each algorithm across different data splits.

TABLE 1. Summarizes the performance outcomes for each algorithm across different data splits

Algorithm	Train–Test Split	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Naïve Bayes	70:30	69	69	83	76	66
	80:20	71	70	85	77	69
	90:10	72	72	83	78	70
SVM	70:30	69	68	88	77	66
	80:20	69	68	88	77	66
	90:10	71	70	88	78	68
K-NN	70:30	71	71	83	77	69
	80:20	73	72	85	78	70
	90:10	73	74	83	78	71
Random Forest	70:30	70	69	86	76	66
	80:20	70	69	87	77	67
	90:10	71	70	87	78	68

Naïve Bayes shows relatively consistent performance across all data split scenarios. Its accuracy gradually increases from 69% at a 60:40 ratio to 71% at 70:30, and reaches 72% at 80:20. This stability reflects that the model can perform well even when the size of the training data varies. The main advantage of Naïve Bayes is its high recall value, consistently in the range of 83 to 85%, which demonstrates its ability to capture most opinions, both positive and negative. The F1-score value, which increased from 76% to 78%, shows a balance between precision and recall. However, the relatively low AUC value, around 66% to 70%, indicates limitations in its discriminatory ability, making this model less than optimal for distinguishing more subtle or ambiguous sentiments. This weakness is generally related to the assumption of feature independence, which is often not met in Indonesian text due to the semantic relationships between words.

The performance of K-Nearest Neighbors tends to be below that of other models, with accuracy stagnating at around 69% at ratios of 60:40 and 70:30, and only increasing slightly to 71% at 80:20. Although recall is very high, reaching 88%, the precision value is lower, only around 68% to 70%, so the model often produces false positive predictions. This condition causes an imbalance between prediction accuracy and completeness, reflected in the F1 score, which falls within the range of 77% to 78%. The limited AUC value, around 66% to 68%, further confirms that K-Nearest Neighbors is less reliable in high-dimensional text classification. Its dependence on distance calculations in a complex vector space makes this model difficult to compete with large datasets that have high linguistic variation. Although simple and easy to understand, K-Nearest Neighbors is not suitable as the main model for sentiment analysis in this case.

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

Random Forest produces more competitive performance than Naïve Bayes and K-Nearest Neighbors, with stable accuracy in the range of 70 to 71% across all data splits. Recall values are quite high, between 83% and 87%, while precision ranges from 69% to 71%, which collectively result in consistent F1 scores at the 76% to 78% level. This shows that Random Forest can maintain a balance of performance, especially on datasets with a relatively balanced class distribution. The AUC value, which ranges from 66% to 69%, shows better discrimination capabilities than K-Nearest Neighbors, although it is still not as strong as the best model. The main advantage of Random Forest is its prediction stability thanks to its ensemble mechanism, but its weakness lies in its higher computational cost. In the context of real-time big data processing, this limitation is an important consideration for practical implementation.

SVM emerged as the best model in this study. Its accuracy reached 71% at a 60:40 ratio, increased to 73% at 70:30, and remained at 73% at 80:20. This model also showed a balance between precision, ranging from 71% to 74%, and recall, ranging from 83% to 85%, resulting in a consistent F1 score of 77% to 78%. The AUC value, which reached 71% at a ratio of 80:20, shows SVM's ability to distinguish positive and negative sentiments more effectively than other algorithms. This advantage is closely related to the maximum margin principle used, making the model more resistant to overfitting even with high-dimensional data. These results reinforce previous research findings that place SVM as one of the most reliable methods in text classification.

Based on these comparative results, SVM can be designated as the best model for sentiment analysis related to the KCJB project. Its advantages include stable performance across various data division ratios, resistance to high feature dimensions, and a better balance of evaluation metrics compared to other algorithms. The selection of SVM as the main model is in line with the research objective of finding an effective method for Indonesian sentiment analysis. In terms of contribution, this research confirms the position of SVM as a solid foundation for further development, for example, through integration with deep learning-based feature representations such as word embeddings or transformer models.

From a practical standpoint, these findings have significant implications for both public policy and the business world. In large-scale national infrastructure projects such as KCJB, SVM-based sentiment analysis can be used to monitor public perception in real time, identify issues that cause social resistance, and assist the government in designing more adaptive public communication strategies. In a business context, a similar approach can be applied to measure consumer response to products or services, evaluate the effectiveness of marketing campaigns, and improve service quality based on customer feedback. Therefore, this research not only contributes to the academic literature on Indonesian text classification but also provides practical contributions to strategic decision-making in the real world.

4.2. Discussions.

The results show that all classical machine learning algorithms—Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, and Random Forest—achieve relatively stable F1 scores in the range of 76% to 78% in all data partitioning scenarios. Naïve Bayes excels in consistently high recall values despite its low Area Under Curve, while Random Forest offers stability thanks to its ensemble mechanism. K-Nearest Neighbors produced the highest accuracy and AUC when the training data set was expanded, but its precision was lower. Support Vector Machine proved to be the best model with 73% accuracy, 88% recall, and the highest AUC, confirming its superiority in high-dimensional text classification as reported in previous studies [21], [22]. These findings are in line with the literature that places margin-based classification as a strong benchmark for Twitter sentiment classification [21], [22]. However, other studies show that deep learning-based models such as ConvBiLSTM and transformer architectures such as BERT are capable of outperforming classical algorithms when supported by contextual word representations [15], [23], [24], [26]. For Indonesian language corpora, BERT-based hybrid strategies combined with Bi-LSTM or TCN have been shown to improve semantic understanding and classification accuracy [24], while multilingual models such as mBERT are better able to handle code-switching phenomena [25].

The main limitation of classical models lies in their inability to capture linguistic nuances, slang, and non-linguistic symbols such as emojis that are often found on Twitter [19], [27]. This points to the need for further research that integrates contextual word representations, aspect-based analysis, and time-based evaluation to handle dynamic vocabulary shifts. Therefore, this study can serve as an important initial benchmark for KCJB project sentiment analysis and encourage the development of transformer-based models that are more adaptive to the characteristics of Indonesian social media.

5. CONCLUSIONS

This study aims to address the research gap in Indonesian sentiment analysis by conducting a comparative evaluation of four commonly used classification algorithms—Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest—using Twitter data related to public opinion on the Jakarta–Bandung Fast Train (KCJB) project. The methodological framework includes data collection, preprocessing, feature extraction using TF-IDF, and comprehensive model evaluation on various training-testing ratios using accuracy, precision, recall, F1-score, and AUC as performance metrics.

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

The results highlight the strengths and limitations of each algorithm. Naïve Bayes shows strong recall but limited discriminatory power due to its independence assumption, while K-NN produces consistently high recall but suffers from low precision, reducing its robustness for large-scale text classification. Random Forest provides stable and balanced performance across all metrics, but at a higher computational cost. Among the models tested, SVM consistently outperformed other models, achieving the highest accuracy up to 73%, balanced precision and recall, and superior AUC values, confirming its robustness in handling high-dimensional text data.

The contributions of this study can be summarized in three aspects. First, this study provides an empirical comparison of four classification algorithms in the context of Indonesian social media data, enriching knowledge about sentiment analysis in languages with limited resources. Second, this study demonstrates the practical application of sentiment analysis as a decision-making tool for monitoring public opinion on large-scale infrastructure initiatives, providing insights for policymakers in understanding public acceptance and resistance. Third, this study establishes SVM as a baseline model for future studies while suggesting opportunities to integrate advanced deep learning techniques such as LSTM or transformer-based models to further improve performance.

Practically, the findings of this study provide applicable insights for the public and private sectors. For policymakers, sentiment analysis can serve as a real-time monitoring system to capture public attitudes toward national development projects, facilitating more adaptive communication strategies and informed decision-making. In the business realm, similar approaches can be applied to track consumer sentiment, evaluate marketing campaigns, and refine customer engagement strategies. Thus, beyond its academic contributions, this research highlights the role of sentiment analysis as a strategic tool that bridges data-driven insights with real-world decision-making.

REFERENCES

[1] S. Ezung and K. O. Koza, 'The Influence of Social Media on Public Discourse: Examining How Social Media Shape Popular Opinion', Dec. 2024, doi: 10.5281/ZENODO.14252454.

[2] H. U. Khan, S. Nasir, K. Nasim, D. Shabbir, and A. Mahmood, 'Twitter trends: A ranking algorithm analysis on real time data', *Expert Systems with Applications*, vol. 164, p. 113990, Feb. 2021, doi: 10.1016/j.eswa.2020.113990.

[3] Y. Özkent, 'Social media usage to share information in communication journals: An analysis of social media activity and article citations', *PLOS ONE*, vol. 17, no. 2, p. e0263725, Feb. 2022, doi: 10.1371/journal.pone.0263725.

[4] K. Khoirunurrofik, C. Endrina Dewi, and A. Marwah Zulkarnain, 'Exploring the public sentiment of local community on major infrastructure development: Evidence from media news and Twitter data', *Journal of Human Behavior in the Social Environment*, vol. 34, no. 3, pp. 423–443, Apr. 2024, doi: 10.1080/10911359.2023.2249962.

[5] J. Widodo and K. Kusnan, 'Mapping and Analyzing E-Government Sentiments in Local Governments in Indonesia', *Society*, vol. 11, no. 2, pp. 434–457, Dec. 2023, doi: 10.33019/society.v11i2.558.

[6] T. Wijaya, '"Risk is not Measured, but Contested and Compromised": A Case Study of Jakarta–Bandung High-Speed Railway', *Journal of Contemporary Asia*, vol. 55, no. 3, pp. 405–429, May 2025, doi: 10.1080/00472336.2024.2378856.

[7] A. Ullah, S. N. Khan, and N. M. Nawi, 'Review on sentiment analysis for text classification techniques from 2010 to 2021', *Multimed Tools Appl*, vol. 82, no. 6, pp. 8137–8193, Mar. 2023, doi: 10.1007/s11042-022-14112-3.

[8] B. Budiman, Z. S. Anggraeni, C. Habibi, and N. Alamsyah, 'Analisis Sentimen Publik pada Media Sosial Twitter Terhadap Tiket.com Menggunakan Algoritma Klasifikasi', *Jurnal Informatika*, vol. 11, no. 1, Art. no. 1, Apr. 2024, doi: 10.31294/inf.v11i1.17988.

[9] D. P. Sari, Budiman, and N. Alamsyah, 'Analisis Perbandingan Sentimen Pengguna Twitter Terhadap Layanan Salah Satu Provider Internet Di Indonesia Menggunakan Metode Klasifikasi', *TEMATIK*, vol. 10, no. 2, Art. no. 2, Dec. 2023, doi: 10.38204/tematik.v10i2.1578.

[10] N. Alamsyah, T. P. Yoga, Budiman, I. Akbar, A. Hendra, and A. J. Prima, 'A Bidirectional GRU Approach with Hyperparameter Optimization for Sentiment Classification in Game Reviews : Pendekatan GRU Dua Arah dengan Optimasi Hiperparameter untuk Klasifikasi Sentimen dalam Ulasan Game', *NUANSA INFORMATIKA*, vol. 19, no. 2, pp. 88–95, Jul. 2025, doi: 10.25134/ilkom.v19i2.399.

[11] N. Alamsyah, G. B. Wibisono, T. P. Yoga, Budiman, and A. Hendra, 'Emoji-Based Sentiment Classification Using Ensemble Learning with Cross-Validation: A Lightweight Approach for Social Media Analysis: Klasifikasi Sentimen Berbasis Emoji Menggunakan Ensemble Learning dengan Validasi Silang: Pendekatan Ringan untuk Analisis Media Sosial', *NUANSA INFORMATIKA*, vol. 19, no. 2, pp. 81–87, Jul. 2025, doi: 10.25134/ilkom.v19i2.396.

[12] K. H. Oo, 'Comparing Accuracy Between SVM, Random Forest, K-NN Text Classifier Algorithms for Detecting Syntactic Ambiguity in Software Requirements', in *International Conference on Information Systems and Intelligent Applications*, M. Al-Emran, M. A. Al-Sharafi, and K. Shaalan, Eds., Cham: Springer International Publishing, 2023, pp. 43–58. doi: 10.1007/978-3-031-16865-9_4.

[13] 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, doi: 10.1109/ACCESS.2022.3149482.

[14] E. D. Madyatmadja, C. P. M. Sianipar, C. Wijaya, and D. J. M. Sembiring, 'Classifying Crowdsourced Citizen Complaints through Data Mining: Accuracy Testing of k-Nearest Neighbors, Random Forest, Support Vector Machine, and AdaBoost', *Informatics*, vol. 10, no. 4, p. 84, Dec. 2023, doi: 10.3390/informatics10040084.

[15] S. Tam, R. B. Said, and O. O. Tanriover, 'A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification', *IEEE Access*, vol. 9, pp. 41283–41293, 2021, doi: 10.1109/ACCESS.2021.3064830.

[16] R. Nagamanjula and A. Pethalakshmi, 'A novel framework based on bi-objective optimization and LAN2FIS for Twitter sentiment analysis', *Soc. Netw. Anal. Min.*, vol. 10, no. 1, p. 34, Dec. 2020, doi: 10.1007/s13278-020-00648-5.

[17] S. H. Janjua, G. F. Siddiqui, M. A. Sindhu, and U. Rashid, 'Multilevel aspect based sentiment classification of Twitter data: using hybrid approach in deep learning', *PeerJ Computer Science*, vol. 7, p. e433, Apr. 2021, doi: 10.7717/peerj-cs.433.

[18] L. Chaudhary, N. Girdhar, D. Sharma, J. Andreu-Perez, A. Doucet, and M. Renz, 'A Review of Deep Learning Models for Twitter Sentiment Analysis: Challenges and Opportunities', *IEEE Trans. Comput. Soc. Syst.*, vol. 11, no. 3, pp. 3550–3579, Jun. 2024, doi: 10.1109/TCSS.2023.3322002.

[19] S. Barreto, R. Moura, J. Carvalho, A. Paes, and A. Plastino, 'Sentiment analysis in tweets: an assessment study from classical to modern word representation models', *Data Min Knowl Disc*, vol. 37, no. 1, pp. 318–380, Jan. 2023, doi: 10.1007/s10618-022-00853-0.

[20] S. S. Jacob and R. Vijayakumar, 'Sentimental analysis over twitter data using clustering based machine learning algorithm', *J*

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.3

Ambient Intell Human Comput, Jan. 2021, doi: 10.1007/s12652-020-02771-9.

[21] K. Ayyub, S. Iqbal, E. U. Munir, M. W. Nisar, and M. Abbasi, 'Exploring Diverse Features for Sentiment Quantification Using Machine Learning Algorithms', *IEEE Access*, vol. 8, pp. 142819–142831, 2020, doi: 10.1109/ACCESS.2020.3011202.

[22] M. AminiMotlagh, H. Shahhoseini, and N. Fatehi, 'A reliable sentiment analysis for classification of tweets in social networks', *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 7, Dec. 2022, doi: 10.1007/s13278-022-00998-2.

[23] A. S. Talaat, 'Sentiment analysis classification system using hybrid BERT models', *J Big Data*, vol. 10, no. 1, p. 110, Jun. 2023, doi: 10.1186/s40537-023-00781-w.

[24] C.-H. Lin and U. Nuha, 'Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy', *J Big Data*, vol. 10, no. 1, p. 88, May 2023, doi: 10.1186/s40537-023-00782-9.

[25] G. Manias, A. Mavrogiorgou, A. Kiourtis, C. Symvoulidis, and D. Kyriazis, 'Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data', *Neural Comput & Applic*, vol. 35, no. 29, pp. 21415–21431, Oct. 2023, doi: 10.1007/s00521-023-08629-3.

[26] A. Bello, S.-C. Ng, and M.-F. Leung, 'A BERT Framework to Sentiment Analysis of Tweets', *Sensors*, vol. 23, no. 1, p. 506, Jan. 2023, doi: 10.3390/s23010506.

[27] B. Wang, D. Shan, A. Fan, L. Liu, and J. Gao, 'A Sentiment Classification Method of Web Social Media Based on Multidimensional and Multilevel Modeling', *IEEE Trans. Ind. Inf.*, vol. 18, no. 2, pp. 1240–1249, Feb. 2022, doi: 10.1109/TII.2021.3085663.