NUANSA INFORMATIKA
Volume 19 Number 2 July 2025

p-ISSN :1858-3911 , *e-ISSN : 2614-5405*
https://journal.fkom.uniku.ac.id/ilkom

Emoji-Based Sentiment Classification Using Ensemble Learning with Cross-Validation: A Lightweight Approach for Social Media Analysis

Gunthur Bayu Wibisono[1], Nur Alamsyah*[2], Titan Parama Yoga[3], Budiman[4], Acep Hendra[5]
[1,2,3,4,5]Universitas Informatika Dan Bisnis Indonesia, Indonesia
E-mail: [1]gunthur.bw21@student.unibi.ac.id, ***[2]nuralamsyah@unibi.ac.id**, [3]titanparama@unibi.ac.id
[4]budiman@unibi.ac.id, [5]acephendra@unibi.ac.id

Abstract

The increasing use of emojis in online communication reflects emotional expression that is often more immediate and intuitive than text. This study proposes a lightweight sentiment classification approach that utilizes only emoji features extracted from social media posts, without relying on textual content. The importance of this research lies in its relevance to short-form digital content, where textual sentiment cues are minimal or absent. To address the classification problem, we implement and compare multiple machine learning models including Random Forest (RF), Support Vector Machine, and an ensemble Voting Classifier combining both. Emoji tokens were vectorized using character-level count vectorization, and performance was evaluated using 5-fold cross-validation to ensure robustness and generalizability. Results show that the ensemble model achieved the highest average accuracy of 93.6%, outperforming the individual classifiers. These findings confirm that emojis alone can serve as reliable indicators of sentiment and support the deployment of fast, interpretable, and scalable models for social media sentiment analysis.

Keywords— *Emoji Sentiment Analysis, Ensemble Learning, Voting Classifier, Cross-Validation, Social Media*

Submission: May 16, 2025 *Accepted: June 10, 2025* *Published: July 10, 2025*
DOI Article: https://doi.org/10.25134/ilkom.v19i2.396

1. BACKGROUND

The widespread adoption of social media has significantly changed the landscape of digital communication, encouraging users to express opinions, emotions, and reactions not only through text but also through symbolic elements like emojis [1]. Emojis have become essential components of online interactions, often acting as concise yet powerful indicators of sentiment in informal and short-form content such as tweets, captions, and instant messages [2]. This phenomenon creates both opportunities and challenges in sentiment analysis, particularly when textual content is minimal or absent [3].

Traditional sentiment analysis methods generally rely on lexical resources, syntactic features, and word embeddings extracted from textual data to determine polarity or emotion [4] [5]. These techniques have been effective in structured text but tend to perform poorly on informal content like social media posts [6]. The integration of emojis into sentiment models began to gain traction when Wang et al. (2025) [7] analyzed tweets and concluded that emojis are strong sentiment cues, often outperforming text-based features in noisy environments.

Further advancing this domain, Shen et al. (2025) [8] using DeepMoji, a neural network trained on billion tweets containing emojis to predict emotional tone. The model demonstrated that emoji-based supervision could enhance downstream NLP tasks. Fouda et al. (2025) [9] also investigated the semantic behavior of emojis using distributional representations, showing that emojis can cluster based on context and sentiment similarity.

Several researchers have attempted to combine emoji features with text, but few studies have explored the use of emoji-only input for sentiment classification. While deep learning models such as LSTM or transformer-based architectures have yielded high performance, their computational cost and complexity hinder their use in lightweight or real-time systems [10].

More recent studies have experimented with emoji-specific vectorization techniques and feature engineering approaches. For instance, Kusal et al. (2025) [11] proposed Emoji2Vec, a method for generating emoji embeddings based on emoji descriptions, which enabled better generalization in NLP pipelines. Other works have explored the emotional and cultural interpretation of emojis, yet most still rely on combined text-emoji input rather than emojis in isolation [12] [13].

In contrast, lightweight approaches using traditional machine learning models such as Random Forest, SVM, or ensemble strategies remain underexplored in the context of emoji-only sentiment prediction. There is a lack of systematic evaluation that compares these models while ensuring robustness through cross-validation [14].

This study proposes an emoji-based sentiment classification framework that uses only emoji characters as input features, without relying on any textual content. Three classifiers are evaluated: Random Forest, Support Vector Machine (SVM), and an ensemble Voting Classifier combining both [15]. Emoji sequences are encoded using character-level count vectorization, and model performance is assessed using 5-fold cross-validation to ensure robustness.

This research contributes to the development of fast, interpretable, and resource-efficient sentiment classification techniques suited for modern digital communication. It provides empirical evidence that emojis alone can reliably indicate sentiment, offering a practical solution for lightweight sentiment analysis in social media applications.

2. RESEARCH METHODOLOGY

The proposed method is designed to operate efficiently by leveraging only emoji characters extracted from user-generated social media content, without relying on textual information. The process includes data crawling, emoji preprocessing, feature transformation, model training using ensemble classifiers, and performance evaluation. Figure 1 illustrates the overall architecture of the proposed methodology. The process begins with data crawling from social media sources, followed by emoji extraction using rule-based filtering. The extracted emoji sequences are then transformed into numerical features using a character-level CountVectorizer. These features serve as input for two base classifiers: RF and SVM. Their predictions are combined using a Voting Classifier as the ensemble strategy. The model is evaluated using 5-fold cross-validation, and the final trained model is saved for deployment purposes.

2.1. Data Collection

The dataset used in this research was obtained from the Kaggle platform, which provides a collection of labeled tweets containing various emojis. A total of 3,085 tweets were collected, each representing an emotional expression labeled into one of four predefined sentiment classes (0: sad, 1: happy, 2: angry, 3: love).

The primary objective of this study is to classify these emotions based solely on the emojis embedded in the tweets, without relying on any textual data. This makes the dataset highly relevant for evaluating the feasibility of emoji-based sentiment classification in informal digital communication contexts. Table 1 describes the features available in the dataset.

Table 1. Dataset Features Description

Feature Name	Description	Data Type
Text	The original tweet text containing emojis	String
Sentiment	Labeled emotion category (0: sad, 1: happy, 2: angry, 3: love)	Integer

2.2 Emoji Preprocessing

After the raw tweet data is collected, the next step involves preprocessing to isolate only the emojis used within each message. Since the objective of this study is to classify sentiment based exclusively on emojis—without any reliance on textual content—this preprocessing step is crucial in ensuring that only relevant visual-symbolic features are retained [16].

The emoji extraction process is performed using the emoji library in Python, which can identify and filter emoji characters from mixed-content text. Each tweet is scanned character by character, and only those identified as emoji tokens are extracted. This method eliminates all non-emoji characters including words, numbers, hashtags, and punctuation.

The result of this process is a new feature column that contains sequences of emojis representing the emotional context of the original tweet. Entries with no emojis are discarded to ensure that the learning model is trained solely on emoji-based inputs.

2.3 Feature Transformation

Once the emojis are extracted from each tweet, the next step involves transforming these sequences into numerical feature representations suitable for machine learning models [17]. This transformation is performed using CountVectorizer, a method that converts a collection of character-based emoji tokens into a matrix of token counts.

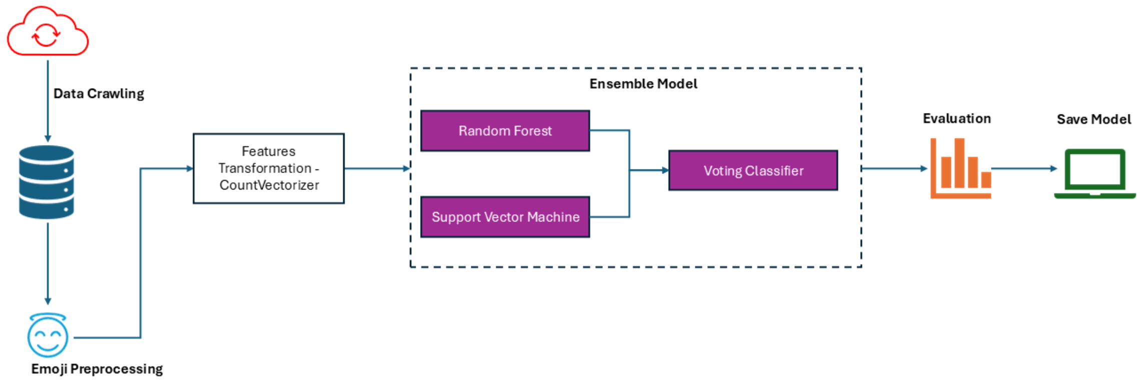

Figure 1. Proposed Method

In this study, character-level vectorization is applied, treating each unique emoji as a feature. Let the set of all unique emojis extracted from the dataset be denoted by:

$$E = \{e_1, e_2, e_3, \dots, e_k\} \quad (1)$$

where k is the total number of distinct emojis across the dataset. Each tweet T_i containing a sequence of emojis is represented as:

$$T_i = \{e_{i1}, e_{i2}, \dots, e_{im}\} \quad (2)$$

The CountVectorizer function transforms each tweet T_i into a feature vector $x_i \in R^k$, where each component x_{ij} represents the count of emoji e_j in tweet T_i:

$$x_{ij} = \text{count}(e_j \in T_i), \quad \forall j = 1,2, \dots, k \quad (3)$$

The final output is a document-term matrix $X \in R^{n \times k}$, where n is the number of tweets (documents), k is the number of unique emojis (features), X_{ij} is the frequency of emoji e_j in tweet T_i. This vectorized representation captures the emoji usage patterns across tweets and serves as input to the classification models.

2.4 Ensemble Model

To improve classification performance and generalizability, this study employs an ensemble learning approach using a Voting Classifier, which combines the predictions of multiple base classifiers. Specifically, we integrate two widely used algorithms: Random Forest (RF) and Support Vector Machine (SVM). The goal of this ensemble is to leverage the strengths of each model—Random Forest's robustness in handling high-dimensional data and SVM's ability to find optimal decision boundaries.

```
Algorithm 1: Emoji-Based Sentiment Classification Using Ensemble

Input  : X (emoji feature matrix), y (sentiment labels)
Output : y_pred (final sentiment prediction)

Initialize base models:
    model_1 ← RandomForestClassifier()
    model_2 ← SupportVectorMachine(probability=True)

Train both models on training data:
    model_1.fit(X_train, y_train)
    model_2.fit(X_train, y_train)

Predict probabilities on test data:
    prob_1 ← model_1.predict_proba(X_test)
    prob_2 ← model_2.predict_proba(X_test)

Compute soft voting average:
    prob_avg ← (prob_1 + prob_2) / 2

Assign final class based on max probability:
    y_pred ← argmax(prob_avg, axis=1)

Return y_pred
```

Algorithm 1 illustrates the steps involved in constructing the ensemble model for emoji-based sentiment classification using soft voting. The algorithm begins by initializing two base classifiers—Random Forest and Support Vector Machine. Both models are independently trained using the same training dataset composed of emoji features and corresponding sentiment labels.

Once trained, each model is used to predict the probability distribution over the sentiment classes for the test data. These probability outputs are then averaged in a soft voting scheme, which computes the mean predicted probability for each class across both models. The final predicted class label for each instance is determined by selecting the class with the highest averaged probability.

This approach enables the ensemble model to leverage the strengths of both classifiers while mitigating their individual weaknesses. By averaging the prediction probabilities, the Voting Classifier can produce more balanced and robust predictions, especially when working with sparse or symbolic input data such as emojis.

The Voting Classifier operates by aggregating the predictions from both classifiers

p-ISSN :1858-3911 , e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

and determining the final class label based on the majority (hard voting) or average predicted probability (soft voting). In this study, soft voting is used, where the class probabilities predicted by each base model are averaged, and the class with the highest average probability is selected.

Let C be the number of sentiment classes (in this case, $C = 4$), $M = \{m_1, m_2\}$ represent the set of base classifiers (RF and SVM), $P^{(c)}_{m_i}(x)$ denote the probability predicted by model m_i for class c, given input feature vector x.

The ensemble model's prediction for class c is computed as:

$$P^{(c)}_{\text{ensemble}}(x) = \frac{1}{|M|}\sum_{i=1}^{|M|} P^{(c)}_{m_i}(x) \quad (4)$$

The final predicted class $\hat{y}$ is given by:

$$\hat{y} = \arg\max_{c\in\{1,2,\dots,C\}} P^{(c)}_{\text{ensemble}}(x) \quad (5)$$

This ensemble strategy enhances classification robustness by reducing variance and balancing out the biases of individual classifiers.

2.5 Evaluation

To ensure that the classification models generalize well to unseen data and are not overfitting to specific subsets, evaluation is performed using k-Fold Cross-Validation [18]. This technique provides a robust estimate of the model's performance by splitting the dataset into multiple train-test partitions.

In this study, we employ 5-Fold Cross-Validation (k = 5). The dataset is divided into five equal-sized subsets (folds). During each iteration:

- Four folds are used to train the model.
- The remaining one fold is used to test the model.
- This process is repeated five times, such that each fold is used exactly once as the validation set.

The overall performance is computed as the average of accuracy scores across all folds. Let A_i denote the accuracy of the model on the $i-\text{th}$ fold. Then, the mean accuracy is calculated as:

$$\text{Mean Accuracy} = \frac{1}{k}\sum_{i=1}^{k} A_i \quad (6)$$

and the standard deviation is computed to assess the model's stability:

$$\text{Standard Deviation} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(A_i - \bar{A})^2} \quad (7)$$

where $\bar{A}$ is the mean accuracy across all folds.

This cross-validation approach is applied consistently across all models tested in this study: Random Forest, SVM, and the ensemble Voting Classifier. The results from this evaluation provide a reliable estimate of model performance and allow for a fair comparison between classifiers.

2.6 Save Model

After the model has been trained and evaluated, the final step involves saving the best-performing classifier for future use. In this study, the Voting Classifier—which demonstrated the highest average accuracy and stability during cross-validation—is selected as the final model.

Model persistence is essential for deployment in practical applications, enabling the reuse of trained models without the need to retrain them. The model is serialized using an appropriate object serialization library that supports machine learning models with large numerical structures.

Alongside the classifier, the vectorization object used to convert emoji characters into numerical features is also saved. This ensures consistency between the training phase and any future prediction tasks, as the same feature transformation process must be applied to new input data.

3. RESEARCH RESULT

This study evaluates the performance of three classification models—Random Forest (RF), Support Vector Machine (SVM), and an ensemble Voting Classifier (VC)—using emoji-only input features. Each model was assessed through 5-Fold Cross-Validation to ensure robust evaluation of generalizability and stability.

The Random Forest classifier yielded strong results, achieving an average accuracy of 93.33% across the five folds. However, it showed a relatively high standard deviation of 3.59%, indicating moderate variation in performance across different folds. This suggests that while RF performs well overall, it may be sensitive to data distribution in some folds, especially when class balance is not uniform.

Support Vector Machine, using a linear kernel, achieved slightly higher average accuracy at 93.36%, and notably exhibited greater consistency across folds, as indicated by its lower standard deviation of 2.85%. The model's ability to generalize well from sparse emoji-based features demonstrates the effectiveness of margin-based classification even in non-textual data contexts.

The best performance was obtained from the ensemble model using the Voting Classifier. By combining the predictions from both Random Forest and SVM through soft voting, the ensemble achieved an average accuracy of 93.63% with a standard deviation of 3.38%. This indicates that the ensemble approach not only improves predictive performance but also balances out the weaknesses of the individual models, offering more stable results across all validation folds.

A summary of these results is presented in Table 2. The table compares mean accuracy, standard deviation, best and lowest accuracy across all validation folds for each model.

Table 2. Comparison of Classification Performance

Model	Mean Accuracy	Standard Deviation	Best Fold Accuracy	Lowest Fold Accuracy
Random Forest	0.9333	0.0359	0.9683	0.8696
SVM (Linear)	0.9336	0.0285	0.9633	0.8830
Ensemble Model	**0.9363**	0.0338	**0.9720**	0.8813

Figure 2 presents the cross-validation accuracy obtained from the ensemble model using Voting Classifier, which combines predictions from Random Forest and Support Vector Machine. The figure clearly illustrates the model's performance across five different data folds.

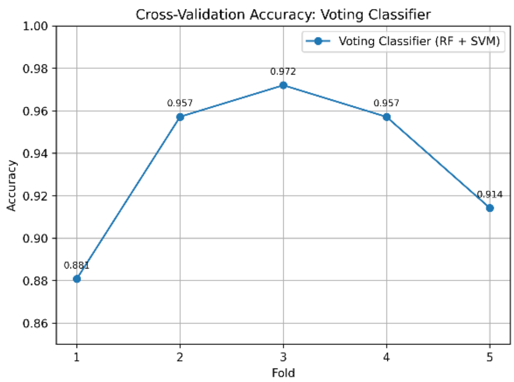

Figure 2 Cross-Validation Accuracy

From the chart, it can be observed that the accuracy ranges between 88.1% and 97.2%, with the highest performance achieved in fold 3 and the lowest in fold 1. The results indicate a generally high and consistent level of accuracy across the folds, reinforcing the model's robustness in classifying sentiment based solely on emoji input.

The slight variation in accuracy between folds reflects natural differences in data distribution and class representation within each subset, which is typical in cross-validation. Despite this, the ensemble model maintains stable performance, as evidenced by three folds yielding accuracies above 95%.

Overall, the visual result supports the numerical findings presented earlier, confirming that the Voting Classifier delivers strong and reliable sentiment classification performance even when limited to non-textual, emoji-based features.

4. DISCUSSION

The results obtained in this study demonstrate that sentiment classification based solely on emoji usage is not only feasible but can also achieve a high level of accuracy. All three tested models—Random Forest, SVM, and Voting Classifier—produced consistent and strong performance, with average accuracies exceeding 93%. This confirms that emojis, despite their simplicity and non-verbal nature, carry significant emotional cues that can be effectively leveraged for classification tasks.

Among the three models, the ensemble approach using a Voting Classifier yielded the best results. Its ability to combine the strengths of both Random Forest and SVM enabled it to achieve the highest average accuracy and strong stability across all validation folds. This outcome highlights the effectiveness of ensemble methods in improving predictive performance, especially in scenarios where the feature space is minimal and symbolic, such as emoji sequences.

The importance of these findings lies in the practical implication that sentiment analysis systems do not always require complex language processing pipelines. By focusing on a lightweight input source like emojis, it is possible to design models that are efficient in terms of computation while still maintaining high accuracy. This is particularly useful in environments with resource constraints or where input data is limited to short, informal messages such as those found on social media platforms.

p-ISSN :1858-3911 , e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

Furthermore, the robustness of the Voting Classifier across various data folds implies that this approach can be generalized to other emoji-rich datasets or even extended to multilingual or cross-cultural contexts where textual analysis may encounter limitations. The results support the idea that visual-symbolic communication, which is increasingly prevalent in digital interactions, can be analyzed with the same rigor as traditional text-based sentiment analysis.

5. CONCLUSION

This study has successfully demonstrated that sentiment classification based solely on emojis can achieve high accuracy using conventional machine learning techniques. By transforming emoji characters into numerical features via character-level Count Vectorizer and applying classification models such as Random Forest, Support Vector Machine, and Voting Classifier, the system was able to perform reliably in identifying emotional categories within tweets.

The ensemble model using Voting Classifier outperformed individual models, achieving the highest average accuracy of 93.63% and showing stable performance across all validation folds. This highlights the strength of ensemble learning in combining complementary classifiers to enhance predictive capability, especially in domains with symbolic or non-verbal inputs.

One of the key advantages of the proposed method lies in its simplicity and efficiency. The system operates without the need for complex text preprocessing or deep learning architectures, making it suitable for real-time and resource-constrained applications. In addition, the focus on emojis as primary features addresses the growing need for models that can handle informal and visual modes of digital communication.

However, the approach also has limitations. The reliance solely on emoji input may not fully capture nuanced sentiments expressed in more complex messages, especially those involving sarcasm or mixed emotions. Furthermore, the model's performance may vary depending on the distribution and cultural interpretation of emojis in different datasets.

Future development may include integrating emoji features with minimal text input to enhance context understanding, experimenting with emoji embeddings or deep learning methods such as attention mechanisms, and testing the model across diverse social media platforms and languages. Such enhancements could further improve accuracy and expand the applicability of emoji-based sentiment analysis in practical scenarios.

6. SUGGESTION

Based on the limitations identified in this study, several directions for future research are recommended. First, although the model performed well using only emojis as input, future studies may explore the integration of minimal text-based features to capture more nuanced sentiment, especially in cases where emojis are used ambiguously or contextually.

Second, the use of character-level Count Vectorizer could be enhanced by employing more sophisticated representation techniques, such as emoji embeddings or pre-trained vector spaces specifically designed for emoji semantics. This would allow models to learn more about the relationships between emojis beyond simple frequency counts.

Third, future research should consider evaluating the model on larger and more diverse datasets, particularly those containing multilingual content or region-specific emoji usage. This will help determine the generalizability of the proposed approach across different cultural and linguistic contexts.

Lastly, incorporating deep learning-based ensemble models, such as stacking or blending, could be investigated to assess whether more complex combinations of base classifiers lead to further improvements in accuracy and robustness.

These suggestions aim to address the current limitations of the study and provide a roadmap for enhancing the capability and applicability of emoji-based sentiment classification models in subsequent research.

REFERENCES

[1] Q. A. Xu, C. Jayne, and V. Chang, "An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews," *Technol. Forecast. Soc. Change*, vol. 202, p. 123326, 2024.

[2] N. Alamsyah, A. P. Kurniati, and others, "Airfare Fluctuation Analysis with Event and Sentiment Features by Stacking Ensemble Model," in *2024 Ninth International Conference on Informatics and Computing (ICIC)*, IEEE, 2024, pp. 1–6.

[3] W. Erpurini, A. G. Putrada, N. Alamsyah, S. F. Pane, and M. N. Fauzan,

"Confirmatory Factor Analysis for The Impact of Students' Social Medial on University Digital Marketing," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, IEEE, 2023, pp. 615–620.

[4] J. Yu and C. Qi, "Machine Learning-Based Sentiment Analysis in English Literature: Using Deep Learning Models to Analyze Emotional and Thematic Content in Texts," *IEEE Access*, 2025.

[5] C. M. Liapis, A. Karanikola, and S. Kotsiantis, "Enhancing sentiment analysis with distributional emotion embeddings," *Neurocomputing*, vol. 634, p. 129822, 2025.

[6] R. Ahamad and K. N. Mishra, "Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach," *J. Big Data*, vol. 12, no. 1, p. 11, 2025.

[7] S. Wang, Q. Liu, Y. Hu, and H. Liu, "Public Opinion Evolution Based on the Two-Dimensional Theory of Emotion and Top2Vec-RoBERTa," *Symmetry*, vol. 17, no. 2, p. 190, 2025.

[8] Y. Shen *et al.*, "The DeepMoji algorithm for fast and accurate classification of massive books based on emotion encoding with redefined weights in multihead attention," in *Third International Conference on Algorithms, Network, and Communication Technology (ICANCT 2024)*, SPIE, 2025, pp. 189–195.

[9] W. Fouda, A. Hegazy, N. M. Alnaqbi, E. Ozbilge, and E. Özbilge, "Enhancing educational environments with Social Media Feedback Evaluation Employing Hybrid Neutrosophic Decision Optimization (HNDO) and Neutrosophic Sentiment Fusion (NSF).," *Int. J. Neutrosophic Sci. IJNS*, vol. 26, no. 1, 2025.

[10] N. Alamsyah, T. P. Yoga, B. Budiman, and others, "IMPROVING TRAFFIC DENSITY PREDICTION USING LSTM WITH PARAMETRIC ReLU (PReLU) ACTIVATION," *JITK J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 9, no. 2, pp. 154–160, 2024.

[11] S. Kusal, S. Patil, and K. Kotecha, "Multimodal text-emoji fusion using deep neural networks for text-based emotion detection in online communication," *J. Big Data*, vol. 12, no. 1, pp. 1–25, 2025.

[12] P. M. Hancock, C. Hilverman, S. W. Cook, and K. M. Halvorson, "Emoji as gesture in digital communication: Emoji improve comprehension of indirect speech," *Psychon. Bull. Rev.*, vol. 31, no. 3, pp. 1335–1347, 2024.

[13] A. G. Putrada, N. Alamsyah, and M. N. Fauzan, "BERT for sentiment analysis on rotten tomatoes reviews," in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, 2023, pp. 111–116.

[14] N. Alamsyah, A. P. Kurniati, and others, "Event Detection Optimization Through Stacking Ensemble and BERT Fine-tuning For Dynamic Pricing of Airline Tickets," *IEEE Access*, 2024.

[15] N. Alamsyah and others, "Analisis Perbandingan Sentimen Pengguna Twitter Terhadap Layanan Salah Satu Provider Internet Di Indonesia Menggunakan Metode Klasifikasi," *TEMATIK*, vol. 10, no. 2, pp. 246–251, 2023.

[16] N. Alamsyah, B. Budiman, R. Nursyanti, E. Setiana, and V. R. Danestiara, "Approximate Bayesian Inference for Bayesian Confidence Quantification in DNA Sequence Classification Using Monte Carlo Dropout Approach," *Innov. Innov. Res. Inform.*, vol. 7, no. 1, 2025.

[17] A. Khan, D. Majumdar, and B. Mondal, "Sentiment analysis of emoji fused reviews using machine learning and Bert," *Sci. Rep.*, vol. 15, no. 1, p. 7538, 2025.

[18] N. Alamsyah, V. R. Danestiara, B. Budiman, R. Nursyanti, E. Setiana, and A. Hendra, "OPTIMIZED FACEBOOK PROPHET FOR MPOX FORECASTING: ENHANCING PREDICTIVE ACCURACY WITH HYPERPARAMETER TUNING," *J. Techno Nusa Mandiri*, vol. 22, no. 1, pp. 90–98, 2025.