Application of the Support Vector Machine Algorithm for Sentiment Classification of Movin by Mandiri Taspen Application Reviews

Aradi Sebayang, Eko Hariyanto

Abstract

The mobile banking application Movin by Mandiri Taspen is a vital tool that supports digital financial and banking services for customers, particularly retirees and senior citizens. Evaluating user perception and satisfaction through Google Play Store reviews is essential for identifying technical bottlenecks and improving digital service quality. This study aims to analyze and classify user review sentiments for the Movin application using the Support Vector Machine (SVM) algorithm across three sentiment classes (Positive, Neutral, and Negative). A curated dataset of 12,303 review records was processed through text preprocessing stages (including cleaning, case folding, tokenizing, stopword removal, and stemming) and Term Frequency-Inverse Document Frequency (TF-IDF) feature weighting. The dataset was partitioned using an 80:20 ratio into a training set (9,842 samples) and a test set (2,461 samples). Hyperparameter optimization of the SVM algorithm yielded an optimal model configuration using a linear kernel with a regularization penalty parameter C = 10.0. Model evaluation showed strong performance, achieving an Accuracy of 95.98%, Macro Precision of 95.25%, Macro Recall of 94.42%, and a Macro F1-Score of 94.82%. Class-specific evaluations recorded F1-scores of 0.97 for the Neutral class, 0.96 for the Negative class, and 0.91 for the Positive class. The predominance of neutral and negative sentiments underscores the necessity for targeted enhancements in interface responsiveness and system stability. This study offers data-driven strategic insights for application developers to implement targeted technical refinements grounded in public opinion analysis.

Keywords: Sentiment Classification, Movin by Mandiri Taspen, Support Vector Machine, Google Play Store, TF-IDF.

Aradi Sebayang[1]
[1,2] Master of Information Technology, University Pembangunan Panca Budi, Medan, Indonesia
e-mail: aradisebayang23@gmail.com[1]

Eko Hariyanto[2]
e-mail: eko.hariyanto@dosen.pancabudi.ac.id[2]
2nd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

Advances in information technology and mobile devices have transformed mobile banking applications into primary conduits for providing digital financial services and banking transactions. The Movin by Mandiri Taspen application emerged as a dedicated digital solution to support banking services, primarily serving retired and senior customers. Application distribution platforms such as the Google Play Store provide an open forum for users to submit textual reviews reflecting firsthand experiences, positive feedback, and technical grievances regarding application functionality. This user feedback helps software developers assess service quality and drive continuous system improvements. However, accumulating extensive, unstructured textual reviews poses substantial challenges for manual analysis, as manual inspection is time-consuming, inefficient, and susceptible to subjective interpretative bias.

To address the complexity of textual data processing, computational approaches leveraging Natural Language Processing (NLP) and text mining through sentiment analysis are essential for classifying the polarity of public opinion in an automated and objective manner. In Indonesian sentiment analysis, text preprocessing stages encompassing cleaning, case folding, tokenizing, stopword removal, and stemming play a critical role in filtering noise and standardizing text structures. Next, review texts are transformed into high-dimensional vector representations using the Term Frequency-Inverse Document Frequency (TF-IDF) weighting method, which computes a word's relative significance across the document corpus.

Among machine learning algorithms, the Support Vector Machine (SVM) has consistently proven to be one of the most effective methods for text classification and sentiment analysis. SVM's mathematical advantage lies in its ability to construct an optimal separating hyperplane that maximizes the margin between classes in high-dimensional feature spaces. Prior literature underscores the robustness of the SVM algorithm on Indonesian mobile application reviews. A study by [1] on reviews of the Islamic banking application BYOND by BSI showed that integrating TF-IDF with SVM yielded 93.16% accuracy. In an analysis of reviews for the healthcare application PrimaKu, [2] found that evaluating kernel variations favored the linear-kernel SVM, which achieved the highest accuracy (97.5%) and outperformed both polynomial and RBF kernels. [3] further corroborated the reliability of linear-kernel SVM with TF-IDF weighting on Google Maps reviews (96% accuracy), [4] on SIREKAP application reviews (99% accuracy), and [5], which optimized a linear-kernel SVM to achieve 94% accuracy. Furthermore, comparative algorithmic studies by [6] and [7] confirmed that SVM exhibits highly stable generalization capabilities in delineating text patterns within public opinion datasets.

Although extensive sentiment analysis research has been conducted on application reviews, most prior studies have focused on e-commerce platforms, social media, or general-purpose applications, often neglecting hyperparameter optimization of the regularization penalty (C) under class imbalance. Empirical investigations that specifically evaluate specialized mobile banking applications such as Movin by Mandiri Taspen, whose user base is predominantly senior citizens with distinct technical pain points, remain exceedingly scarce. Consequently, this study aims to classify user review sentiments for the Movin by Mandiri Taspen application on the Google Play Store into three sentiment classes (Positive, Neutral, and Negative) utilizing a TF-IDF-based Support Vector Machine (SVM) algorithm paired with hyperparameter optimization (C = 10.0, linear kernel). By analyzing a clean dataset of 12,303 review records partitioned via an 80:20 train-test split (9,842 training samples and 2,461 testing samples) and evaluating performance primarily through the Macro F1-Score (94.82%), this study is intended to provide strategic, data-driven recommendations for application

developers to sustainably elevate customer satisfaction and digital banking service delivery.

Literature Review

The deployment of Natural Language Processing (NLP) and machine learning to analyze user review sentiments across mobile applications has advanced considerably within various digital platform domains in Indonesia. In public services and digital banking, user reviews on the Google Play Store serve as primary empirical data for evaluating satisfaction levels and identifying technical software bottlenecks. A study by [1] investigated 35,000 reviews of the Sharia banking application BYOND by BSI using Term Frequency-Inverse Document Frequency (TF-IDF) alongside Support Vector Machine (SVM) and Random Forest classifiers. Their findings confirmed that SVM achieved the highest accuracy (93.16%), with a more stable learning curve in capturing users' expressions of technical grievances. [4] also documented SVM's superiority in the public sector across 15,000 SIREKAP reviews, where three-class sentiment modeling using a linear-kernel SVM achieved 99% accuracy. Additionally, [7] evaluated the efficacy of SVM and Naïve Bayes on user reviews of the SMILE Indonesia healthcare logistics service application, while [2] evaluated reviews of the PrimaKu application, demonstrating that linear-kernel SVM classification achieved 97.5% accuracy, surpassing polynomial (96.5%) and RBF kernels (97%).

Methodologically, the Support Vector Machine (SVM) framework rests on a robust theoretical foundation for resolving high-dimensional text classification problems. SVM functions by determining the optimal separating hyperplane that maximizes the margin between classes within the vector feature space constructed through TF-IDF term weighting [8], [9]. In an analysis of the financial application Ajaib, [6] benchmarked Naïve Bayes, SVM, and Random Forest, with SVM achieving the highest accuracy (91%). The consistency of pairing TF-IDF with SVM to attain classification performance exceeding 90% was likewise affirmed by [10] on Shopee digital marketplace reviews, achieving 98% accuracy and F1-score, as well as by [11], who established that SVM (94.5% accuracy) systematically outperformed Random Forest (92.7%) and Decision Tree (88.2%) classifiers on an Indonesian textual evaluation corpus.

Regarding model optimization and textual data structures, several researchers have focused on hyperparameter tuning and mitigating class imbalance. [5] implemented GridSearch techniques to optimize the hyperparameters of a linear-kernel SVM on skincare product reviews, improving accuracy to 94% with precision, recall, and F1-score values exceeding 0.90. Concurrently, [3] applied LinearSVC using an 80:20 data split on Google Maps reviews of mobile phone repair services, achieving 96% accuracy and an F1-score. Under severe class imbalance, [12] applied the Synthetic Minority Oversampling Technique (SMOTE) before testing an SVM model on social media platform X tweets, achieving 95.7% accuracy. Nevertheless, deploying synthetic oversampling methods such as SMOTE in high-dimensional, sparse text feature matrices risks generating syntactic redundancy or artificial noise.

While prior literature has widely substantiated the resilience of the SVM algorithm for application review classification, earlier works predominantly relied on default regularization parameters or focused on generic applications, often yielding evaluations biased toward the majority class. This study bridges that gap by rigorously tuning the regularization penalty ($C = 10.0$) with a linear kernel, without altering the data distribution through synthetic oversampling. The empirical evaluation is conducted on a clean dataset of 12,303 reviews of the Movin by Mandiri Taspen mobile banking application, characterized by a unique demographic profile dominated by retirees and senior citizens. Using an 80:20 stratified split (9,842 training samples and 2,461 testing samples) and evaluating model performance with Macro F1-Score (94.82%) and overall Accuracy (95.98%), this study contributes a methodological framework for objective, balanced, and unbiased three-class

sentiment modeling.

Research Methods

2.1 Data Collection and Dataset Distribution

This study adopts a supervised machine learning experimental approach to classify user sentiment into three categories: Negative, Positive, and Neutral. The quantitative experimental dataset comprises user reviews of the Movin by Mandiri Taspen application extracted from the Google Play Store. Following data scrubbing and preliminary filtering, a refined dataset of 12,303 review records was compiled. The dataset is structured into three sentiment classes: Neutral (5,971 reviews; 48.53%), Negative (4,714 reviews; 38.32%), and Positive (1,618 reviews; 13.15%). This class imbalance is central to designing the preprocessing workflow and performance evaluation metrics. The overarching research workflow is structured systematically into six primary phases:

1. Data Collection and Curation: Compiling and preparing the verified review dataset annotated with sentiment ground-truth labels.
2. Text Preprocessing: Cleansing, filtering, and standardizing the raw textual corpus.
3. Feature Extraction: Converting unstructured text data into numerical representations using Term Frequency-Inverse Document Frequency (TF-IDF) weighting.
4. Data Splitting: Partitioning the feature space into training and testing subsets using an 80:20 stratified ratio.
5. Hyperparameter Optimization and Model Training: Determining optimal parameters through systematic hyperparameter tuning and training the Support Vector Machine classifier.
6. Model Evaluation: Quantifying model generalization capability using a multi-class confusion matrix, overall accuracy, and macro-averaged metrics.

2.2 Text Preprocessing

Unstructured review text is processed sequentially across five preprocessing stages to eliminate informational noise and standardize textual syntax:

1. Cleansing: Stripping punctuation, digits, special symbols, emojis/emoticons, and non-informative URLs.
2. Case Folding: Converting all uppercase characters into uniform lowercase text.
3. Tokenizing: Segmenting full review strings into individual constituent words (tokens).
4. Stopword Removal: Eliminating common functional words that carry negligible sentiment orientation.
5. Stemming & Normalization: Reducing affixed words to their morphological root forms according to standard Indonesian linguistic rules, while normalizing colloquial terms, slang, and non-standard abbreviations into formal lexicon.

2.3 Feature Extraction (TF-IDF Weighting)

Preprocessed review texts are converted into high-dimensional numerical feature spaces via Term Frequency-Inverse Document Frequency (TF-IDF) weighting. This weighting metric measures a term's relative importance based on its local frequency within an individual review, scaled by its inverse distribution across the entire document corpus. Words that occur frequently in an isolated review but infrequently across the broader collection receive higher weights, creating highly discriminative features for the classification algorithm.

2.4 Data Splitting

The numerical TF-IDF feature matrix was partitioned using a Stratified Train-Test Split with an 80% training and 20% testing ratio. This stratified split yielded 9,842 training samples for model optimization and 2,461 testing samples to evaluate generalization. The resulting test set preserves class proportions, consisting of 1,194 Neutral reviews, 943 Negative reviews, and 324 Positive reviews.

2.5 Classification and SVM Hyperparameter Optimization

Sentiment classification was executed using the Support Vector Machine (SVM) algorithm

configured with a One-vs-Rest (OvR) decision strategy to handle the three-class formulation. SVM constructs an optimal separating hyperplane that maximizes the margin separating discrete classes within the high-dimensional feature space. To address class imbalance constraints, hyperparameter tuning was conducted using Grid Search. The superior model configuration was obtained utilizing a Linear Kernel with a regularization penalty parameter C = 10.0. The linear kernel is computationally efficient and well suited to high-dimensional, linearly separable text representations, while C = 10.0 applies a strong penalty for training classification errors without inducing overfitting.

2.6 Model Performance Evaluation

Model assessment was performed on the 2,461 testing samples utilizing a Multi-Class Confusion Matrix. Evaluated metrics include Accuracy, Precision, Recall, and F1-Score. Given the skewed sentiment class distribution, aggregate model performance focuses primarily on macro-averaged indicators (Macro Precision, Macro Recall, and Macro F1-Score). Macro-averaged metrics compute performance for each class independently and weight them equally, ensuring predictive capabilities are evaluated objectively without disproportionate skew from majority classes.

Results and Discussion

3.1 User Review Dataset Distribution

Acquiring Movin by Mandiri Taspen user reviews from the Google Play Store, followed by text preprocessing, yielded a final clean dataset of 12,303 records. Based on annotations across three sentiment categories, the empirical frequency distribution is presented in Table 1.

Table 1. Sentiment Class Distribution of the Movin by Mandiri Taspen Dataset

Sentiment Class	Number of Reviews (Rows)	Percentage (%)
Neutral	5,971	48.53%
Negative	4,714	38.32%
Positive	1,618	13.15%
Total Dataset	**12,303**	**100.00%**

Table 1 highlights a clear class imbalance, with Neutral and Negative reviews collectively constituting 86.85% of the corpus. Conversely, the Positive class represents a distinct minority at 13.15% (1,618 reviews). Applying the Stratified Train-Test Split (80:20) generated 9,842 training instances and 2,461 testing instances. The test set comprises 1,194 Neutral reviews, 943 Negative reviews, and 324 Positive reviews.

3.2 SVM Hyperparameter Optimization Results

Hyperparameter tuning across the 9,842 training samples using TF-IDF features was performed to determine the most precise decision boundaries. Evaluating diverse kernel functions alongside variations in the regularization parameter (C) confirmed that the optimal configuration was attained using a Linear Kernel with C = 10.0.

The high-dimensional feature space generated by TF-IDF vectorization is linearly separable; therefore, a linear kernel ensures computational efficiency while maintaining classification stability. Setting C = 10.0 imposes a tight penalty on training misclassification margins, creating firm decision boundaries without causing model overfitting.

3.3 Confusion Matrix Analysis

The predictive performance of the linear-kernel SVM model (C = 10.0) was evaluated on the 2,461 test samples. The visualization of the multi-class Confusion Matrix across the three sentiment classes is depicted in Figure 1.

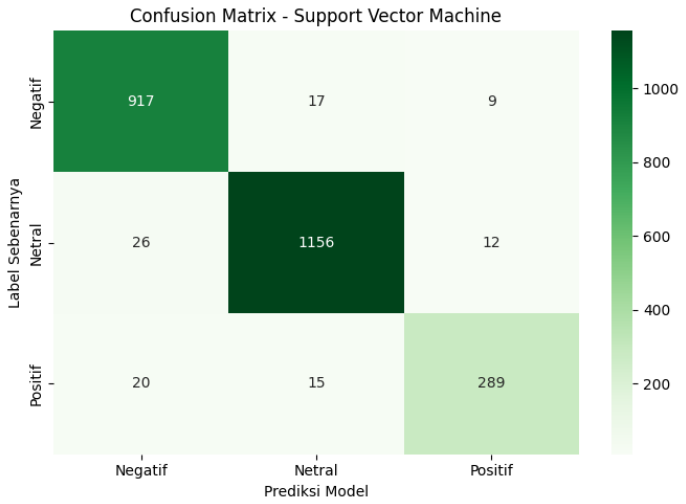


Figure 1: Confusion Matrix – Support Vector Machine (C=10.0, Linear Kernel)

Based on Figure 1, the itemized breakdown of model predictions relative to actual ground-truth labels across the 2,461 test records is detailed as follows:

1. Negative Class (943 reviews total): 917 reviews were correctly classified as Negative (True Negatives). 17 reviews were erroneously classified as Neutral, and 9 reviews were misclassified as Positive.
2. Neutral Class (1,194 reviews total): 1,156 reviews were correctly classified as Neutral (True Neutrals). 26 reviews were misclassified as Negative, and 12 reviews were misclassified as Positive.
3. Positive Class (324 reviews total): 289 reviews were correctly classified as Positive (True Positives). 20 reviews were misclassified as Negative, and 15 reviews were misclassified as Neutral.

Cumulatively, across the 2,461 test samples, the optimized SVM model correctly predicted 2,362 reviews (the sum of the main diagonal: 917 + 1,156 + 289) and misclassified only 99 instances.

3.4 Classification Model Performance Evaluation

Testing the Support Vector Machine (SVM) model (C = 10.0, linear kernel) on the 2,461 test samples showed consistently strong and balanced performance across all aggregate metrics, as summarized in Table 2.

Table 2. Summary of Aggregate Performance Evaluation for the SVM Model

Global Evaluation Metric	Performance Value (%)
Accuracy	95.98%
Precision (Macro Avg)	95.25%
Recall (Macro Avg)	94.42%
Macro F1-Score (Primary Metric)	**94.82%**

The SVM model achieved an overall Accuracy of 95.98% and a Macro F1-Score of 94.82%. The close alignment between the Macro F1-Score and Macro Precision (95.25%) and Macro Recall (94.42%) confirms that the model remains robust to class imbalance. The classifier demonstrates an aptitude for discerning each opinion category independently and equitably, without bias toward the majority Neutral or Negative classes.

3.5 Class-Specific Performance Analysis

To evaluate the model's granular efficacy within each sentiment orientation, Table 3 presents the detailed per-class Classification Report.

Table 3. SVM Classification Report Across Sentiment Classes

Sentiment Class	Precision	Recall	F1-Score	Support (Test Samples)
Negative	0.95	0.97	0.96	943
Neutral	0.97	0.97	0.97	1,194
Positive	0.93	0.89	0.91	324
Accuracy	—	—	**0.96 (95.98%)**	**2,461**
Macro Average	**0.95**	**0.94**	**0.95 (94.82%)**	**2,461**
Weighted Average	**0.96**	**0.96**	**0.96**	**2,461**

In reference to Table 3, the analysis of per-class predictive characteristics is structured as follows:

1. Neutral Class (F1-Score: 0.97): Recorded the highest predictive accuracy, with both precision and recall achieving 0.97 across 1,194 test samples. This strong performance is attributed to the syntactic structure of neutral reviews, which largely consist of direct inquiries or factual statements.
2. Negative Class (F1-Score: 0.96): Attained a precision of 0.95 and a recall of 0.97 across 943 test samples. The 0.97 recall underscores the model's acute sensitivity in capturing critical user complaints. Distinct TF-IDF weights applied to salient negative terms such as "gagal" (failed), "error", "kendala" (issue/obstacle), "lambat" (slow), and "sulit" (difficult) enabled the SVM to define robust hyperplanes demarcating negative instances.
3. Positive Class (F1-Score: 0.91): Despite serving as the minority class with only 324 test instances, the classifier achieved strong results, recording a precision of 0.93 and a recall of 0.89. An F1-score of 0.91 confirms that TF-IDF weighting, coupled with C = 10.0 regularization, adequately preserved affirmative linguistic features and prevented them from being overshadowed by majority-class distributions.

3.6 Key Findings and Practical Implications

The observed prevalence of negative (38.32%) and neutral (48.53%) sentiments reflects user experiences with the Movin by Mandiri Taspen application. An analysis of heavily weighted terms in the negative class identifies systemic challenges centered on application instability after version updates, login and authentication failures, and sluggish user interface response times.

Because Movin's target user base is primarily pensioners and elderly individuals, transactional certainty and frictionless navigation are paramount operational criteria. These empirical insights guide development teams to prioritize back-end server infrastructure improvements, reinforce transaction gateway stability, and systematically streamline user interface and user experience (UI/UX) workflows.

Conclusion

This study demonstrates that the Support Vector Machine (SVM) algorithm, using TF-IDF feature weighting and hyperparameter optimization (C = 10.0, linear kernel), delivers highly accurate and balanced sentiment classification on user reviews of the Movin by Mandiri Taspen application from the Google Play Store. Across the clean dataset of 12,303 reviews, Neutral (48.53%) and Negative (38.32%) classes dominated the sentiment distribution, while Positive reviews accounted for 13.15%. Model evaluation on 2,461 test samples showed strong classification performance, with an overall Accuracy of 95.98% and a Macro F1-Score of 94.82% (Macro Precision of 95.25% and Macro Recall of 94.42%). Consistently high class-level metrics with F1-scores of 0.97 for Neutral, 0.96 for Negative, and 0.91 for Positive affirm that the tuned model is resilient against class imbalance and operates without discernible bias toward majority classes.

References

[1] F. N. Firzatullah and Nuroji, “Analisis Sentimen Pengguna Aplikasi Byond BSI pada Google Play Store Menggunakan Algoritma SVM dan Random Forest,” STORAGE J. Ilm. Tek. dan Ilmu Komput., vol. 4, no. 3, pp. 265–276, Aug. 2025, doi: 10.55123/storage.v4i3.6027.

[2] T. Tinaliah and T. Elizabeth, “Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine,” JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.

[3] R. T. Bagastama, “Implementasi Support Vector Machine untuk Analisis Sentimen Ulasan Google Maps pada Layanan Perbaikan Ponsel,” Pros. Semin. Nas. Teknol. Dan Sains, vol. 5, pp. 218–225, 2026.

[4] M. Z. Abidin, R. Makhfuddin, and A. Syaifuddin, “Analisis Sentimen Ulasan Aplikasi Sirekap Menggunakan Algoritma Support Vector Machine,” J. Inform. dan Tek. Elektro Terap., vol. 14, no. 3, pp. 679–691, Aug. 2026, doi: 10.23960/jitet.v14i3.9549.

[5] H. Harnelia, “Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (SVM)," J. Inform. dan Tek. Elektro Terap., vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4095.

[6] N. D. Kurniawan, P. R. Ferdian, and N. Hidayati, “Analisis Sentimen Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Pada Ulasan Aplikasi Ajaib,” J. Inform. Terpadu, vol. 11, no. 2, pp. 134–143, 2025, doi: 10.54914/jit.v11i2.1884.

[7] L. Rhomaningtias, A. Khairunisa, S. Shella May Wara, and K. Maulida Hindrayani, “Analisis Sentimen Ulasan Aplikasi Smile Indonesia Menggunakan Metode Naive Bayes dan Support Vector Machine (Svm),” HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf., vol. 16, no. 1, pp. 79–91, 2025, doi: 10.52972/hoaq.vol16no1.p79-91.

[8] J. I. W. Siregar, A. P. U. Siahaan, M. Iqbal, D. Nasution, and R. F. Wijaya, “Sentiment Analysis Classification of E-commerce User Reviews Using Natural Language Processing (NLP) and Support Vector Machine (SVM) Methods,” vol. 6, no. 2, pp. 195–205, 2025, doi: 10.47065/bit.v5i2.2018.

[9] J. Ipmawati, S. Saifulloh, and K. Kusnawi, “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine: Sentiment Analysis of Tourist Attractions Based on Reviews on Google Maps Using the Support Vector Machine Algorithm,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 1, pp. 247–256, 2024.

[10] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” Jambura J. Electr. Electron. Eng., vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[11] K. H. Hanif, N. R. Muntiari, D. Harto, and D. S. Wiranata, “Perbandingan Analisis Sentimen Komentar Mahasiswa Prodi Teknik Komputer Menggunakan Algoritma Decision Tree, Support Vector Machine (SVM), dan Random Forest,” Insect (Informatics Secur.). J. Tek. Inform., vol. 12, no. 01, pp. 24–33, 2026, doi: 10.33506/insect.v12i01.5144.

[12] D. C. Novitasari and R. R. Putri, "Analisis Sentimen Cuitan X terhadap Pendaki Asing Gunung Rinjani Menggunakan Algoritma SVM dan Random Forest,” INTEGER J. Inf. Technol., vol. 10, no. 2, pp. 239–252, 2025, doi: 10.31284/j.integer.0.v10i2.8119.