Jurnal Ilmu Komputer dan Informatika (JIKI) Vol. 5, No. 2, December 2025, Page. 155-164
P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.323

Limitations of Support Vector Machine and Random Forest in Multi-Class Sentiment Analysis: Evidence from Neutral Sentiment Misclassification on Imbalanced Data

Ardy Wicaksono*[1], Suyahman[2], Muhammad Anwar Fauzi[3], Deny Prasetyo[4], Dwi Utari Iswavigra[5], Yulaikha Mar'atullatifah[6], Agatha Pricillia Sekar Tamtomo[7], Muhammad Adi Pratama[8]

[1, 2, 4, 5, 6]Computer Science, Universitas Sugeng Hartono, Indonesia
[3, 7]Digital Business, Universitas Sugeng Hartono, Indonesia
[8]English Language and Culture, Universitas Sugeng Hartono, Indonesia

Email: [1]ardywicaksono166@gmail.com

Received : Dec 09, 2025; Revised : Jan 20, 2026; Accepted : Jan 23, 2026; Published : Mar xx, 2026

Abstract

The rapid growth of mobile applications has generated large volumes of user reviews, making automated sentiment analysis essential for understanding user perceptions. Previous studies have shown that while machine learning models perform well in binary sentiment classification, they often struggle in multi-class settings, particularly in identifying neutral sentiment due to linguistic ambiguity and class imbalance. This study aims to comparatively evaluate the performance of Support Vector Machine (SVM) and Random Forest in multi-class sentiment analysis, with a specific focus on their ability to handle the neutral sentiment category. A supervised learning approach was employed using 2,112 Indonesian-language user reviews collected from the Google Play Store. The data were preprocessed using standard Natural Language Processing techniques and represented using TF-IDF features. Both models were trained and evaluated using accuracy, precision, recall, F1-score, and confusion matrices. The results indicate that SVM achieved an accuracy of 86.52%, outperforming Random Forest, which obtained 83.45%. However, both models completely failed to classify the neutral sentiment class, yielding zero precision and recall for this category. This failure highlights the dominant influence of severe class imbalance and insufficient feature discrimination for neutral sentiment. The findings underscore a critical limitation of traditional machine learning approaches in multi-class sentiment analysis and emphasize the need for improved strategies, such as data resampling, advanced feature representation, or hybrid models, to enhance neutral sentiment detection in real-world applications.

Keywords: *Machine-Learning, Sentiment-Analysis, Support-Vector-Machine, Random-Forest, Text-Classification, Imbalanced-Data.*

1. INTRODUCTION

The rapid proliferation of mobile applications has led to an exponential increase in user-generated content, particularly in the form of online reviews. These reviews have become a critical source of information that influences application reputation, user adoption, and continuous system improvement [1], [2]. From a business and governance perspective, user feedback provides valuable insights into usability, system reliability, and service quality, making it an essential asset for data-driven decision making [3], [4]. However, the massive volume and unstructured nature of textual reviews render manual analysis inefficient and impractical, thereby necessitating automated sentiment analysis techniques [5], [6].

Sentiment analysis, a key task within Natural Language Processing (NLP), focuses on identifying and categorizing subjective opinions expressed in text [5], [6]. Early sentiment analysis research primarily addressed binary classification, distinguishing positive and negative opinions [7]. While binary sentiment classification has demonstrated high predictive performance, it fails to adequately represent real-world opinion distributions, where a substantial proportion of user feedback is neutral,

Jurnal Ilmu Komputer dan Informatika (JIKI) Vol. 5, No. 2, December 2025, Page. 155-164
P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.323

descriptive, or ambiguous in nature [8], [9]. Neutral sentiment often reflects mixed experiences or factual statements and lacks strong emotional polarity, making it considerably more difficult to identify using lexical-based features alone [8], [10].

To address this limitation, multi-class sentiment classification has been increasingly adopted, typically incorporating positive, neutral, and negative sentiment categories [11], [4]. However, multi-class sentiment analysis introduces additional challenges, particularly related to linguistic ambiguity and data distribution imbalance [6], [9]. In real-world application review datasets, positive sentiments often dominate, while neutral sentiments form a minority class, resulting in skewed class distributions [12], [13]. Such imbalance can bias machine learning models toward majority classes and significantly degrade their ability to learn meaningful patterns from minority sentiment categories [14], [15].

Numerous studies have demonstrated that class imbalance can lead to misleading performance evaluations when relying solely on overall accuracy [9], [15], [16]. High accuracy values may conceal poor predictive performance on minority classes, particularly neutral sentiment, which is often misclassified or entirely ignored [17], [9]. Consequently, class-level evaluation metrics such as precision, recall, F1-score, and confusion matrix analysis are essential to reveal the true behavior of sentiment classification models in multi-class scenarios [15], [16].

Despite the growing popularity of deep learning and transformer-based architectures, traditional machine learning algorithms such as SVM and Random Forest remain widely used in sentiment analysis research [18], [13], [19]. These models are favored for their interpretability, computational efficiency, and robustness in high-dimensional sparse feature spaces such as TF-IDF representations [20], [13], [21]. In contrast, existing comparative studies predominantly focus on overall accuracy or binary sentiment settings, offering limited insight into how these models handle neutral sentiment under imbalanced data conditions [6], [9], [19].

Recent literature has begun to highlight the persistent difficulty of neutral sentiment classification, emphasizing that neutral opinions are frequently misclassified as either positive or negative due to overlapping lexical features and insufficient representation during model training [9], [10]. Nevertheless, many studies do not explicitly address this limitation in their analysis or discussion, resulting in an incomplete understanding of model reliability in practical sentiment analysis applications [22], [15].

Based on these observations, this study aims to conduct a comparative evaluation of SVM and Random Forest for multi-class sentiment analysis of mobile application reviews, with a specific focus on their ability to handle neutral sentiment in an imbalanced dataset. Unlike prior studies that emphasize accuracy as the primary performance metric, this research adopts a class-level evaluation approach to uncover hidden model weaknesses. The contributions of this study are: (1) providing empirical evidence of neutral sentiment misclassification in traditional machine learning models; (2) demonstrating the limitations of accuracy-based evaluation in imbalanced multi-class sentiment analysis; and (3) offering methodological insights to guide future research toward more robust and reliable sentiment classification frameworks.

2. METHOD

This study employed a structured methodological framework to analyze user sentiment toward the OSS mobile application using supervised machine learning techniques. The methodological process consists of four main stages: data collection, text preprocessing, model training, and model evaluation. Each stage was designed to ensure the integrity, reproducibility, and reliability of the sentiment classification pipeline.

The flowchart below illustrates the complete workflow to offer a visual overview of the sequential procedures involved in transforming raw review data into model-ready features and evaluating classification performance in Figure 1. It presents the overall research workflow that structures the methodological procedures of this study. The following subsections describe each methodological stage in detail.

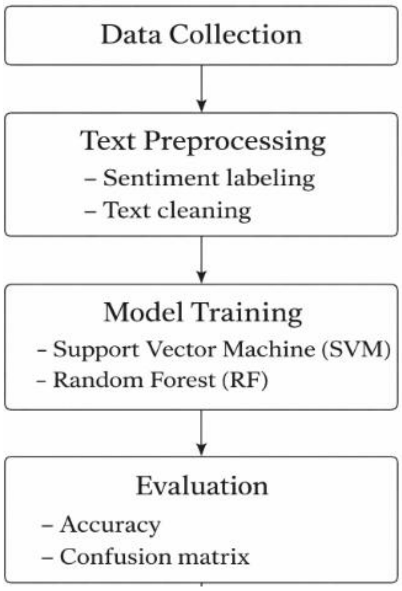

Figure 1. Research Workflow

2.1. Data Collection

The initial stage of this methodological framework involved data collection, which is critical for establishing a representative and robust dataset for sentiment analysis [6]. In the context of mobile application analysis, user feedback is primarily channeled through public review platforms [4]. The foundational dataset for this study was constructed from user-generated reviews extracted from the Google Play Store page of the target Open Sources Software mobile application. The application is publicly accessible via the following URL:

"https://play.google.com/store/apps/details?id=id.go.oss&hl=id"

To facilitate the automated and systematic retrieval of the reviews, the *google-play-scraper Python library* was utilized. This tool was instrumental in extracting publicly available review data in bulk. A total of 2,112 of the most recent user reviews were collected. The acquisition was strictly constrained to the Indonesian language and region (using the parameters $lang = "id"$ and $country = "id"$) to ensure linguistic homogeneity, which is crucial for subsequent language-dependent Natural Language Processing (NLP) and sentiment classification tasks.

The scraping process focused on extracting two primary attributes from each review instance, which constitute the core components for the sentiment analysis pipeline:

a) Textual Review Content (Content): The raw text provided by the user.
b) Associated Star Rating (Score): The numerical rating (1–5 stars), which serves as the ground truth label for training the supervised machine learning model.

Data integrity was enforced during the acquisition phase through immediate filtering. Reviews lacking textual content, duplicate entries, or those identified as non-Indonesian language content were systematically excluded. Furthermore, only publicly visible review data were collected; no personally identifiable information (such as usernames or email addresses) was accessed or stored, aligning with established ethical research guidelines.

All retrieved data were organized and stored in a structured tabular format to facilitate seamless integration into the subsequent analytical procedures. The data collection methodology fully complied with Google Play's terms of service regarding the use of publicly posted user reviews. No attempts were made to infer user identities. For the sake of research transparency and reproducibility, detailed scraping scripts and timestamp metadata are archived and can be provided upon request.

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.323

Table 1. Sample of Filtered Google Play User Review Data

No	content	score	sentiment
1	sekarang jadi mudah bikin nib sendiri	5	positif
2	Oke	5	positif
3	Terimakasih OSS yg bisa memberikan NIB secara ...	5	positif
4	harus pakai aplikasi ini biar mudah	5	positif
5	mantab, terimakasih 🙏	5	positif
6	aplikasi ga bisa di pake buat nib malah jd ribet	1	negatif
7	ko ga bisa akses kelola nib ya . . selalu kelu...	1	negatif
8	uhh bikin susah ini mah	1	negatif
9	super ribet banget aplikasi nyusahin rakyat ya...	1	negatif
10	Gak berfungsi	1	negatif

2.2. Text Preprocessing

The text preprocessing stage is essential in Natural Language Processing (NLP) methodologies, serving to transform raw, noisy, and unstructured textual data into a refined, consistent, and numerically representable format that is optimal for machine learning algorithms [20]. This process is paramount for enhancing feature quality and maximizing the classification model's performance [23].

First, each review was assigned a categorical sentiment label based on its associated numerical star rating, establishing the ground truth for classification [12]. This rule-based approach, which follows common practices in prior sentiment analysis studies, maps the ratings into a ternary classification schema: reviews rated 1–2 stars were labeled as Negative, ratings of 3 stars as Neutral, and ratings of 4–5 stars as Positive.

Following sentiment labeling, the textual content underwent rigorous cleaning and normalization to reduce noise and ensure linguistic consistency. This included converting all text to lowercase, removing non-alphanumeric characters using regular expressions, and normalizing excessive whitespace. These operations produced a standardized text corpus suitable for feature extraction.

Feature representation was conducted using the Term Frequency-Inverse Document Frequency (TF-IDF) method, implemented via the $TfidfVectorizer$ in the $scikit-learn$ library. TF-IDF assigns a weight to each term based on its frequency within a document and its inverse rarity across the entire dataset [13]. The TF-IDF score for a term t in document d within corpus D is calculated as:

$$RF-IDF(t,d,D) = TF(t,d)xIDF(t,D) \tag{1}$$

Equation (1) Shows the TF-DF Score

A maximum vocabulary size of 5000 features was adopted to balance model expressiveness with computational efficiency. The transformation generated a sparse matrix in which each review was represented as a weighted vector capturing its most informative terms.

Finally, the resultant feature set was partitioned into training and testing subsets using an 80:20 train-test split. Crucially, stratification by sentiment class was applied to preserve the original class distribution across both sets. This mitigates the risk of model bias toward overrepresented categories and ensures that the testing set provides an unbiased basis for performance evaluation. The resulting training set was used to fit the classification models.

2.3. Model Training and Classifier Implementation

The penultimate stage of the methodology involved training the selected supervised machine learning models using the feature-engineered dataset [22]. This study employed two distinct algorithms a linear classifier and an ensemble-based method to benchmark classification performance on the Indonesian sentiment corpus [19]. Both models were exclusively trained on the stratified 80% training subset of the dataset, ensuring a proportional and unbiased representation of the Positive, Neutral, and Negative sentiment classes.

The Support Vector Machine (SVM) algorithm, specifically the Linear Support Vector Classifier ($Linear\ SVC$), was utilized. SVM operates by finding the optimal hyperplane that maximally separates the data points (TF-IDF feature vectors) belonging to different sentiment classes in the high-dimensional

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.323

feature space. The ($Linear\ SVC$) was specifically chosen due to its inherent efficiency and proven robust performance on high-dimensional and sparse text data, characteristic of TF-IDF representations [24]. The model was trained using the default hyperparameter settings, establishing a competitive and reproducible baseline performance without relying on resource-intensive optimization techniques.

The Random Forest (RF) algorithm was selected as the representative ensemble-based classifier. RF operates by constructing a multitude of decision trees during the training phase and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This method excels at capturing complex, non-linear relationships within the feature space [21]. The RF classifier was configured with 300 decision trees ($n_{estimators} = 300$). This specific configuration was chosen to achieve an optimal balance between enhancing model stability (reducing variance) and managing computational cost. The $random\_state$ parameter was fixed to ensure the computational experiment is fully reproducible.

Both models learned the underlying patterns and weighted relationships between the TF-IDF features and the assigned sentiment polarity. Following the training completion, the resulting classifiers were subjected to rigorous testing on the reserved 20% test set to assess their generalization capabilities, an evaluation process detailed in the subsequent section.

2.4. Model Evaluation and Model Metrics

The final stage of the methodological workflow involved a rigorous Model Evaluation to objectively assess the predictive performance and generalization capability of the trained classifiers (SVM and RF) [25]. Evaluation is a crucial step that validates the utility of the constructed models using a reserved, unseen dataset. For classification tasks, performance must be quantified using metrics that provide both a high-level summary and granular insight into class-specific behavior.

Accuracy was selected as the primary metric to summarize overall model performance. Accuracy reflects the ratio of correctly predicted instances to the total number of samples in the test set [15]. Due to its direct interpretability, it is highly suitable for providing a clear, high-level reflection of the model's success in this multiclass sentiment classification context. Accuracy is formally defined as:

$$Accuracy = \frac{True\ Positivies + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (2)$$

Equation (2) Shows the Accuracy

In addition to overall accuracy, Confusion Matrices were generated for both algorithms. Confusion Matrices provide a crucial, granular understanding of classification behavior across the three sentiment categories (Positive, Neutral, and Negative). By mapping the actual classes against the predicted classes, the matrix allows for the quantification of True Positives (correctly classified), False Positives (Type I error), and False Negatives (Type II error).

This level of detail was essential, particularly considering the potential for class imbalance often inherent in user-generated review datasets (where positive sentiments tend to dominate). The Confusion Matrix facilitated the identification of systematic misclassification patterns, such as the tendency to confuse Neutral reviews with either Positive or Negative sentiments.

To further dissect the model's performance on individual classes, metrics derived from the Confusion Matrix namely Precision and Recall were analyzed. These metrics provide insight into the model's predictive reliability for each category:

a) Precision: The proportion of positive identifications that were actually correct.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3)$$

Equation (3) Shows the Precision

b) Recall: The proportion of actual positives that were identified correctly.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (4)$$

Equation (4) Shows the Recall

Together, the overall accuracy, the visual interpretation of the Confusion Matrices, and the derived class-level metrics provided a comprehensive assessment of the models' predictive performance and reliability in the ternary sentiment classification task.

3. RESULT

3.1. Exploratory Data Analysis: Sentiment Word Clouds

Initial analysis of user sentiment was performed using Word Clouds, which visualize the frequency of words in the text data. This step provides an intuitive understanding of the prevailing themes in positive and negative reviews

3.1.1. Positive Sentiment

The Word Cloud for positive reviews (*Figure 2*) is dominated by words expressing satisfaction and utility. Key terms such as "mantap" (excellent), "bisa" (can), "baik" (good), "bagus" (great), and "membantu" (helpful) are prominent. This indicates that users primarily appreciate the system's core functionality, ease of use, and overall service quality.

Figure 2. Word Cloud for Positive Sentiment.

3.1.2. Negative Sentiment

Conversely, the negative sentiment Word Cloud () highlights issues related to complexity and technical difficulties. The most frequent words include "aplikasi" (application), "ribet" (complicated), "susah" (difficult), "email", "verifikasi" (verification), and "login". This strong pattern suggests that negative feedback is centered on specific pain points in the user experience, particularly concerning access, registration, and complicated procedural steps.

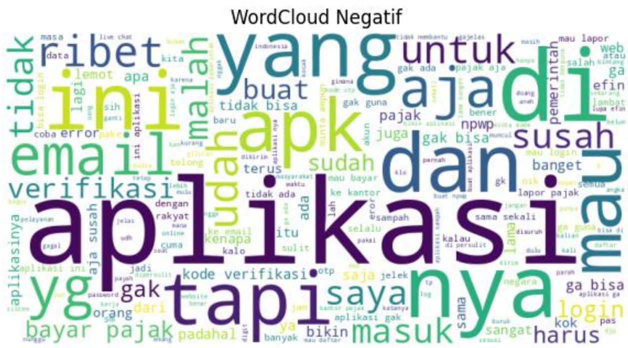

Figure 3. Word Cloud for Negative Sentiment.

3.2. Model Performance Evaluation

Two machine learning models, Support Vector Machine (SVM) and Random Forest, were evaluated for their performance in classifying the three sentiment classes (Negative, Neutral, Positive).

3.2.1. Support Vector Machine (SVM) Perfomance

The SVM model achieved an overall accuracy of 86.52%. Detailed performance metrics for each class are presented in Table 2.

Table 2. Classification Report for Random Forest.

Class	Precision	Recall	F1-Score	Support
Negative	0.86	0.89	0.87	197
Neutral	0.00	0.00	0.00	19
Positive	0.88	0.92	0.90	207
Accuracy			0.87	423
Macro Avg	0.58	0.60	0.59	423
Weighted Avg	0.83	0.87	0.85	423

The high F1-Scores for the Negative (0.87) and Positive (0.90) classes demonstrate the model's high capability in identifying clear sentiment orientations. However, the Confusion Matrix (Figure 4*)* clearly illustrates the model's major deficiency: the Neutral class. The model failed to correctly predict any neutral samples (True Positive = 0), resulting in zero values for Precision, Recall, and F1-Score for this class. The 19 neutral samples were incorrectly distributed across the Negative (14 samples) and Positive (5 samples) categories.

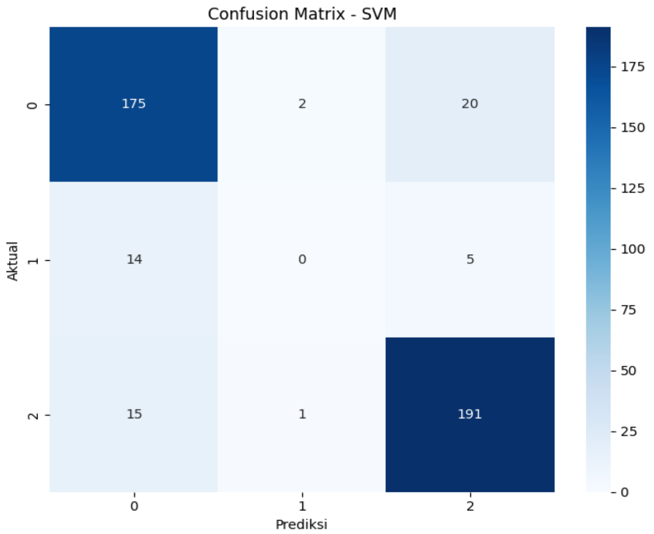

Figure 4. Confusion Matrix for SVM Model.

3.2.2. Random Forest Performance

The Random Forest model yielded an overall accuracy of 83.45%. The classification report is summarized in Table 3.

Table 3. Classification Report for Random Forest.

Class	Precision	Recall	F1-Score	Support
Negative	0.81	0.88	0.84	197
Neutral	0.00	0.00	0.00	19
Positive	0.86	0.86	0.86	207
Accuracy			0.83	423
Macro Avg	0.56	0.58	0.57	423
Weighted Avg	0.80	0.83	0.82	423

While exhibiting strong performance on the majority classes, the Random Forest model shared the same fundamental weakness as the SVM. As shown in the Confusion Matrix (Figure 5), the model was completely unable to correctly classify the Neutral class, with a True Positive count of 0, confirming the 0.00 F1-Score.

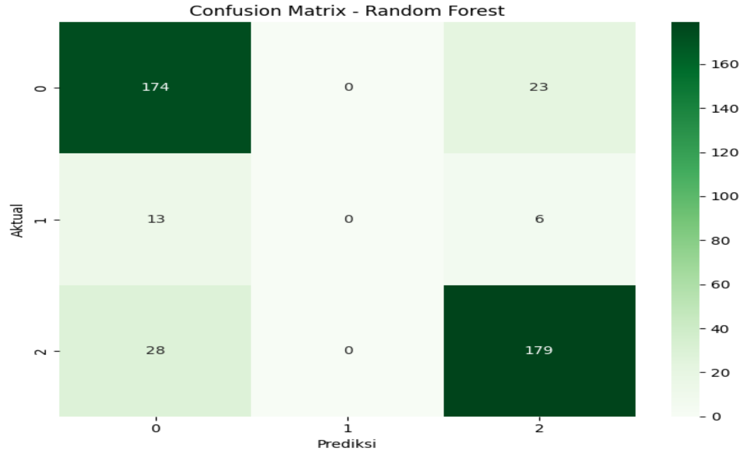

Figure 5. Confusion Matrix for Random Forest Model.

4. DISCUSSIONS

In accordance with the finished analysis result, the implication is obviously validated. Comparative Analysis: The SVM model (Accuracy: 86.52%) outperformed the Random Forest model (Accuracy: 83.45%) by approximately 3%, establishing SVM as the superior classifier for this specific dataset.

In regards to the discussion on Class Imbalance, the most critical finding is the pervasive failure of both advanced models to classify the Neutral class. This is directly attributable to severe class imbalance, as the Neutral class accounts for only 4.5% of the total data (19 samples). The high overall accuracy figures are thus misleading, as they primarily reflect the excellent classification of the majority classes (Negative and Positive). In a real-world scenario, this model limitation would render the classification of truly neutral feedback unreliable.

5. CONCLUSION

To sum up, as this study investigated the performance of SVM and Random Forest algorithms for multi-class sentiment analysis, the results indicate that the SVM model is the better classifier in terms of the effectiveness area, achieving an overall accuracy of 86.52%. Both models demonstrated high proficiency in distinguishing between Negative and Positive sentiments, as evidenced by their high F1-scores in these categories.

However, the key conclusion is the complete failure of both models to predict the Neutral sentiment class (F1-score 0.00). This critical limitation stems from the severe class imbalance in the dataset, where the Neutral class was overwhelmingly underrepresented.

For future research, it is strongly recommended that methods to mitigate class imbalance, such as oversampling or undersampling, are implemented to ensure that the classifier could learn meaningful patterns from the minority class. Alternatively, transforming the problem into a binary classification task (Negative vs. Positive) may yield more robust and reliable results given the current data distribution.

REFERENCES

[1] A. M. Zahran and B. Rolando, "Understanding The Influence Of Online Reviews On," *Int. J. Econ. Bus. Stud.*, vol. 2, no. 1, pp. 1–23, 2025, doi: https://doi.org/10.1234/ijebs.v2i1.109.

[2] P. Weichbroth, "Usability of mobile applications: A systematic literature study," *IEEE Access*, vol. 8, pp. 55563–55577, 2020, doi: 10.1109/ACCESS.2020.2981892.

[3] M. Masumbika and N. Patrick, "Discover Sustainability Enhancing environmental decision - making : a systematic review of data analytics applications in monitoring and management," *Discov. Sustain.*, 2024, doi: 10.1007/s43621-024-00510-0.

[4] F. Faridh, "Respons Publik Terhadap Layanan Pengaduan SP4N LAPOR!: Analisis Sentimen

Multi-Platform.," *J. Audiens*, vol. 6, no. 2, pp. 295–311, 2025, doi: https://doi.org/10.18196/jas.v6i2.616.

[5] T. Kaur and G. Kaur, "Sentiment Analysis for Low Resource Language Using Machine Learning and Deep Learning," *AIJR Proceeding*, vol. 7, no. 6, pp. 179–185, 2025, doi: 10.21467/proceedings.7.6.

[6] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, "Generalizing sentiment analysis : a review of progress , challenges , and emerging directions," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.1007/s13278-025-01461-8.

[7] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[8] S. Simonsen and S. Simonsen, "Social Actor Representation in Media Discourse : How Neutral Linguistic Cues Get Endowed with Meaning that Signifies Ethnicity," *Lang. Discourse Soc.*, vol. 11, no. 1, 2023.

[9] M. Lango, "Tackling the Problem of Class Imbalance in Multi-class Sentiment Classification : An Experimental Study," *Found. Comput. Decis. Sci*, vol. 44, no. 2, pp. 151–178, 2019, doi: 10.2478/fcds-2019-0009.

[10] 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.

[11] J. C. Melaka, "Exploring Enhancement In Sentiment Analysis Using Descriptive-Semantic Techniques : A Systematic Literature Review," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 13, pp. 4819–4837, 2025.

[12] M. Ramdan, A. Surya, and U. Hayati, "Analisis Sentimen Ulasan Pengguna Ovo Menggunakan," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2780–2786, 2024.

[13] F. Fitriana and H. Setiawan, "Performance Analysis of SVM In Emotion Classification : A Comparative Study Of TF-IDF and Countvectorizer," *J. Embed. Syst. Secur. Intell. Syst.*, vol. 6, no. 2, pp. 133–145, 2025.

[14] S. Gholampour, "Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets : Clinical Validity of SMOTE Is Questionable," *MAKE Mach. Learn. Knowl. Extr.*, vol. 6, no. 2, pp. 827–841, 2024.

[15] M. Owusu-Adjei, J. Ben Hayfron-Acquah, T. Frimpong, and G. Abdul-Salaam, "Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems," *PLOS Digit. Heal.*, vol. 2, no. 10, pp. 1–19, 2023, doi: 10.1371/journal.pdig.0000290.

[16] T. Cai and X. Zhang, "Imbalanced Text Sentiment Classification Based on MultiChannel BLTCN-BLSTM SelfAttention," *Sensors*, vol. 23, no. 4, pp. 2–15, 2023.

[17] P. Putra, M. K. Anam, A. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques," *J. Appl. Data Sci.*, vol. 6, no. 2, pp. 921–935, 2025, doi: https://doi.org/10.47738/jads.v6i2.618.

[18] A. Wilson and M. R. Anwar, "The Future of Adaptive Machine Learning Algorithms in High-Dimensional Data Processing," *Int. Trans. Artif. Intell.*, vol. 3, no. 1, pp. 97–107, 2024.

[19] M. K. Anam, T. P. Lestari, H. Yenni, T. Nasution, and M. B. Firdaus, "Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble.pdf," *ECTITransactionsonComputerandInformationTechnology*, vol. 19, no. 2, pp. 159–167, 2025, doi: https://doi.org/10.37936/ecti-cit.2025192.257815.

[20] A. Tabassum and R. R. Patil, "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing," *Int. Res. J. Eng. Technol.*, vol. 07, no. 06, pp. 4864–4867, 2020, [Online]. Available: www.irjet.net

[21] Sunardi and Suyahman, "Analisis Komparasi Prediksi Serangan DDoS Menggunakan Machine Learning," *Proceeding Informatics Collab. Dessimenation Meet.*, vol. 1, no. 1, pp. 84–91, 2025, [Online]. Available: https://www.kaggle.com/datasets/oktayrdeki/ddos-traffic-dataset/

[22] A. Javeed, A. L. Dallora, J. S. Berglund, A. Ali, L. Ali, and P. Anderberg, "Machine Learning

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.323

for Dementia Prediction: A Systematic Review and Future Research Directions," *J. Med. Syst.*, vol. 47, no. 1, 2023, doi: 10.1007/s10916-023-01906-7.

[23] W. Hussain *et al.*, "Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning," *Sci. Rep.*, vol. 15, no. 1, pp. 1–24, 2025, doi: 10.1038/s41598-024-76178-3.

[24] S. Muawanah, U. Muzayanah, M. G. R. Pandin, M. D. S. Alam, and J. P. N. Trisnaningtyas, "Stress and Coping Strategies of Madrasah's Teachers on Applying Distance Learning During COVID-19 Pandemic in Indonesia," *Qubahan Acad. J.*, vol. 3, no. 4, pp. 206–218, 2023, doi: 10.48161/Issn.2709-8206.

[25] G. Papoutsoglou *et al.*, "Machine learning approaches in microbiome research: challenges and best practices," *Front. Microbiol.*, vol. 14, no. 09, pp. 1–21, 2023, doi: 10.3389/fmicb.2023.1261889.