JOURNAL OF APPLIED DIGITAL AND INFORMATION SYSTEMS
Website: https://ejournal.gws-tech.id/index.php/jadis/

Sentiment Analysis of Shopee Electronic Product Reviews Using TF-IDF, Logistic Regression, and 1D CNN

Analisis Sentimen Ulasan Produk Elektronik Shopee Menggunakan TF-IDF, Regresi Logistik, dan 1D CNN

Mohammad Raihan Akbar, Morin Pita Laura, Wildan Ramdani, Gina Khayatun Nufus

Informatics Study Program, Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Cirebon, 45132, Indonesia
raihan13akbar@gmail.com, morinpitalaura1010@gmail.com, ramdaniwildan879@gmail.com, ginakhn@uinssc.ac.id

Article Information
Submitted: July 9, 2026
Accepted: July 14, 2026
Published: July 15, 2026

Corresponding Author
Gina Khayatun Nufus
ginakhn@uinssc.ac.id

DOI
https://doi.org/10.65897/jadis.v1.i1.98

Copyright © 2026 The Author(s).
Published by GWS Tech Solution
This is an open-access article under the CC BY 4.0 International License.

ABSTRACT

Indonesian e-commerce platforms generate large volumes of product reviews that contain useful evidence about customer satisfaction, product quality, delivery performance, and seller service. This study develops SentiTech-ID, a sentiment classification pipeline for Shopee electronic product reviews. The dataset contains 5,499 reviews collected with Selenium from three electronic products and labeled into positive, neutral, and negative classes using rating-based rules. The study compares two methods: TF-IDF word n-gram representation with balanced Logistic Regression and a 1D Convolutional Neural Network implemented in PyTorch. The preprocessing stage handles informal Indonesian text, removes platform noise, normalizes slang, preserves negation words, and retains English sentiment expressions to reduce information loss in code-switched reviews. The results show that Logistic Regression achieved 93.64% accuracy and 0.5641 macro-F1, while the 1D CNN achieved 94.00% accuracy and 0.4545 macro-F1. Although CNN produced a slightly higher accuracy, Logistic Regression performed better on macro-F1 because it recognized minority sentiment classes more effectively. These findings indicate that accuracy alone is not sufficient for highly imbalanced e-commerce review data. The study contributes a reproducible baseline for Indonesian marketplace sentiment analysis and highlights the need for class balancing, stronger minority-class evaluation, and better treatment of code-switching in future work.

Keywords:

Sentiment Analysis; Shopee Reviews; TF-IDF; Logistic Regression; 1D CNN

ABSTRAK

Platform e-commerce di Indonesia menghasilkan volume ulasan produk yang besar dan memuat informasi penting tentang kepuasan pelanggan, kualitas produk, kinerja pengiriman, dan layanan penjual. Penelitian ini mengembangkan SentiTech-ID, yaitu pipeline klasifikasi sentimen untuk ulasan produk elektronik Shopee. Dataset penelitian terdiri atas 5.499 ulasan yang dikumpulkan menggunakan Selenium dari tiga produk elektronik dan diberi label ke dalam kelas positif, netral, dan negatif berdasarkan aturan rating. Penelitian ini membandingkan dua metode, yaitu representasi TF-IDF word n-gram dengan Logistic Regression berbobot seimbang dan Convolutional Neural Network 1D yang diimplementasikan menggunakan PyTorch. Tahap preprocessing menangani teks informal bahasa Indonesia, menghapus noise dari platform, menormalkan slang, mempertahankan kata negasi, dan tetap menyimpan ekspresi sentimen berbahasa Inggris untuk mengurangi kehilangan informasi pada ulasan yang mengandung code-switching. Hasil penelitian menunjukkan bahwa Logistic Regression memperoleh akurasi sebesar 93,64% dan macro-F1 sebesar 0,5641, sedangkan CNN 1D memperoleh akurasi sebesar 94,00% dan macro-F1 sebesar 0,4545. Meskipun CNN menghasilkan akurasi sedikit lebih tinggi, Logistic Regression menunjukkan kinerja macro-F1 yang lebih baik karena lebih efektif mengenali kelas sentimen minoritas. Temuan ini menunjukkan bahwa akurasi saja tidak cukup untuk mengevaluasi data ulasan e-commerce yang sangat tidak seimbang. Penelitian ini memberikan baseline yang dapat direplikasi untuk analisis sentimen marketplace Indonesia dan menegaskan pentingnya penyeimbangan kelas, evaluasi kelas minoritas yang lebih kuat, serta penanganan code-switching yang lebih baik pada penelitian selanjutnya.

Kata Kunci:

Analisis Sentimen; Ulasan Shopee; TF-IDF; Logistic Regression; 1D CNN

1. INTRODUCTION

Online consumer reviews are a strategic form of user-generated content because they describe product experience, perceived quality, delivery reliability, and seller service in natural language [1]. In Indonesian e-commerce, reviews are produced continuously and are too large to evaluate manually. Sentiment analysis therefore offers a practical information system mechanism for converting unstructured review text into actionable polarity labels [2].

Previous Indonesian studies have shown that marketplace and e-commerce application reviews can be analyzed using text mining and natural language processing [3], [4]. Research on the top five Indonesian e-commerce platforms found that app reviews contain positive, neutral, and negative signals about user satisfaction [2]. More specific Shopee studies have applied TF-IDF with Naive Bayes on game-product reviews, Naive Bayes on Google Play reviews, and TF-IDF with Logistic Regression on Shopee application reviews [3], [5]. These studies confirm that Shopee review data are relevant for sentiment mining, but many works still focus on application reviews rather than product-level reviews.

Model selection remains an important issue. Traditional machine learning models can perform well on short Indonesian text when feature extraction is carefully designed [3]. TF-IDF remains competitive for limited and medium-size datasets because it captures discriminative word usage without requiring large-scale pretraining [6]. Deep learning models such as CNN and GRU can learn local and sequential patterns, but they usually require more balanced and more diverse training data to generalize well [2], [6], [7].

Indonesian marketplace reviews also contain linguistic noise. Prior work on Shopee reviews shows that informal language and slang require explicit normalization before classification [8], [9]. Buyers often write informal Indonesian, abbreviations, repeated characters, emoji, and English expressions such as recommended, worth it, fast response, original, and slow delivery [10]. This code-switching or code-mixing phenomenon has been reported as a meaningful challenge in Indonesian-English sentiment analysis because mixed-language expressions can carry sentiment while also increasing vocabulary sparsity [11]. Indonesian NLP resources such as IndoBERT, IndoNLU, and NusaCrowd have improved language coverage, but lightweight baselines are still needed for practical information systems with limited computing resources [12], [13].

The main problem addressed in this study is the performance gap between a lightweight TF-IDF Logistic Regression baseline and a 1D CNN model on highly imbalanced Shopee electronic product reviews. The contribution of this study is threefold. First, it restructures product-level Shopee review sentiment analysis into a reproducible pipeline named SentiTech-ID. Second, it compares Logistic Regression and 1D CNN using accuracy and macro-F1 to avoid majority-class bias. Third, it discusses the effect of class imbalance and code-switching on model reliability.

2. METHODS

2.1. Research Design

This study used an experimental text classification design. The input was Shopee review text, and the output was one of three sentiment classes: negative, neutral, or positive. The pipeline covered data collection, rating-based labeling, preprocessing, feature representation, model training, validation, testing, and error-oriented discussion. Figure 1 summarizes the research workflow.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

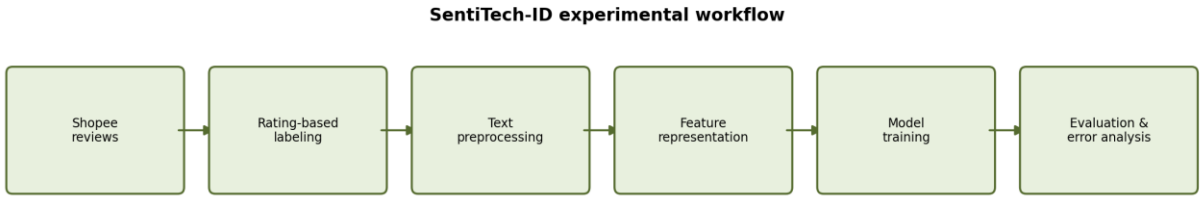

Figure 1. SentiTech-ID Experimental Workflow

2.2. *Dataset and Labeling Strategy*

The dataset consisted of 5,499 Shopee reviews collected using Selenium on May 20, 2026 from three electronic products: Cosmos rice cooker, Hongzhuo electric kettle, and Philips iron. Rating-based labeling was used because star ratings provide an explicit user satisfaction signal in marketplace platforms [14]. Reviews with 1-2 stars were labeled negative, 3-star reviews were labeled neutral, and 4-5-star reviews were labeled positive [14]. This approach follows the practical pattern used in previous Indonesian marketplace sentiment studies, while acknowledging that rating labels may contain noise when review text and star rating disagree [15].

Table 1. Dataset Fields and Labeling Rules

Element	Description
Review text	Customer review sentence or phrase from Shopee product review pages.
Product source	Cosmos rice cooker, Hongzhuo electric kettle, and Philips iron.
Negative label	Reviews with 1-2 star ratings.
Neutral label	Reviews with 3-star ratings.
Positive label	Reviews with 4-5 star ratings.
Split strategy	Stratified 80:20 split: 4,399 training reviews and 1,100 test reviews.

Table 2. Review Distribution by Product and Sentiment Class

Product	Negative	Neutral	Positive	Total
Cosmos Rice Cooker	64	32	1,509	1,605
Hongzhuo Electric Kettle	38	25	1,912	1,975
Philips Iron	35	46	1,838	1,919
Total	137	103	5,259	5,499
Share	2.49%	1.87%	95.64%	100%

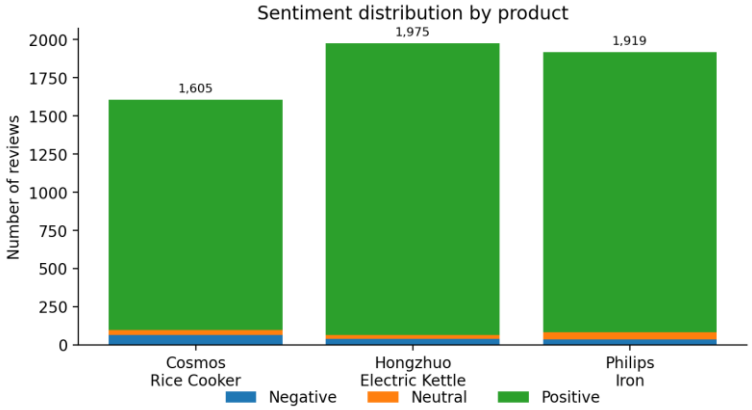

Figure 2. Sentiment Distribution by Product

This is an open-access article under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Table 2 and Figure 2 show a severe class imbalance. Positive reviews represent 95.64% of the dataset, while negative and neutral reviews represent only 2.49% and 1.87%. This distribution reflects a common problem in marketplace review mining: customers who rate products positively dominate the corpus, while minority classes carry important but sparse information about dissatisfaction and ambiguity. For this reason, macro-F1 is treated as the primary comparative metric, while accuracy is interpreted with caution.

2.3. Text Preprocessing

The preprocessing pipeline was designed for informal Indonesian marketplace reviews. It applied case folding, platform-noise removal, URL and number removal, punctuation cleaning, tokenization, slang normalization, and Indonesian stopword removal. Negation words such as *tidak*, *kurang*, and *belum* were retained because they reverse sentiment orientation. English sentiment terms were also retained because many Shopee reviews use code-switched expressions that directly encode polarity.

Table 3. Preprocessing Operations and Rationale

Step	Operation	Rationale
1	Case folding	Reduces duplicated tokens caused by capitalization differences.
2	Noise removal	Deletes platform labels, URLs, newline escapes, numbers, emoji, and punctuation that do not support classification.
3	Tokenization	Splits review text into tokens for TF-IDF and neural input preparation.
4	Slang normalization	Converts informal forms such as gk, bgt, and rekomen into tidak, banget, and rekomendasi.
5	Stopword removal	Removes common Indonesian function words while retaining negation terms.
6	English term preservation	Keeps sentiment-bearing code-switched terms such as recommended, worth it, and slow delivery.

2.4. Feature Representation and Model Configuration

Two model families were evaluated. The first model used TF-IDF word n-gram features and balanced Logistic Regression. TF-IDF assigns higher weights to terms that occur frequently in a review but are less frequent across the corpus. The second model used a 1D CNN that learns local n-gram-like patterns from token sequences through convolutional filters. A 5-fold cross-validation experiment compared word n-gram and character n-gram TF-IDF representations before the final test evaluation.

$$tf - idf(t,d) = tf(t,d) \times log\big(N/(df(t)+1)\big) \quad (1)$$

Table 4. Model Configuration

Component	Logistic Regression	1D CNN
Input representation	TF-IDF word n-gram (1,2)	Token sequence embedding
Main architecture	Linear multinomial classifier	Embedding(128) -> Conv1D(128,k=3) -> Conv1D(64,k=5) -> GlobalMaxPool -> Dropout -> Dense
Class imbalance strategy	class_weight=balanced	Weighted CrossEntropyLoss
Optimizer	scikit-learn solver configuration	Adam optimizer
Training setting	max_iter=1000	10 epochs, batch size 32
Implementation	scikit-learn [20]	PyTorch [21]

The Logistic Regression classifier used class_weight=balanced to increase the loss contribution of minority classes, a common approach in Indonesian sentiment analysis to mitigate the impact of imbalanced datasets [16], [17]. This choice follows the class-imbalance concern prevalent in Indonesian marketplace reviews, where standard accuracy metrics can be misleading and macro-level

This is an open-access article under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

metrics often reveal weaknesses that are otherwise hidden [18], [19]. Oversampling is another valid alternative for future experiments, but it was not applied in the present baseline comparison.

2.5. Evaluation Metrics

The evaluation used accuracy, precision, recall, F1-score, and macro-F1. Accuracy measures the proportion of correct predictions, but it can be misleading when one class dominates the test set. Macro-F1 calculates the unweighted mean of F1-scores across classes, so it gives equal importance to negative, neutral, and positive classes. This metric is more suitable for the current dataset because positive reviews dominate the corpus.

$$Macro - F1 = (1/K)\Sigma F1_k \quad (2)$$

3. RESULTS

3.1. TF-IDF Representation Selection

The 5-fold cross-validation comparison showed that word n-gram TF-IDF was more effective than character n-gram TF-IDF. Word n-gram (1,2) reached an average macro-F1 of 0.5276 with a standard deviation of 0.0420, while character n-gram (3,5) reached 0.4867 with a standard deviation of 0.0381. Word n-gram TF-IDF was therefore selected for the final Logistic Regression experiment.

Table 5. Cross-Validation Result for TF-IDF Representation

Feature representation	Cross-validation setting	Macro-F1 mean	Standard deviation	Decision
Word n-gram (1,2)	5-fold	0.5276	0.0420	Selected
Character n-gram (3,5)	5-fold	0.4867	0.0381	Not selected

3.2. Classification Results on Test Set

The final evaluation used 1,100 stratified test reviews. Table 6 presents class-level classification results. Logistic Regression produced stronger F1-scores for the negative and neutral classes than the 1D CNN. The gap was most visible in the neutral class, where Logistic Regression reached 0.21 F1-score, while the 1D CNN reached only 0.07. Both models performed well on the positive class because this class dominated the training set.

Table 6. Class-Level Classification Results on the Test Set

Model / Class	Precision	Recall	F1-score	Support
LR - Negative	0.44	0.63	0.52	27
LR - Neutral	0.16	0.29	0.21	21
LR - Positive	0.98	0.96	0.97	1,052
CNN - Negative	0.27	0.37	0.31	27
CNN - Neutral	0.05	0.10	0.07	21
CNN - Positive	0.97	0.96	0.96	1,052

Table 7. Overall Model Performance

Model	Accuracy	Macro-F1	Interpretation
Logistic Regression with TF-IDF	0.9364	0.5641	Lower accuracy but better minority-class balance.
1D CNN with PyTorch	0.9400	0.4545	Higher accuracy but weaker macro-level balance.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

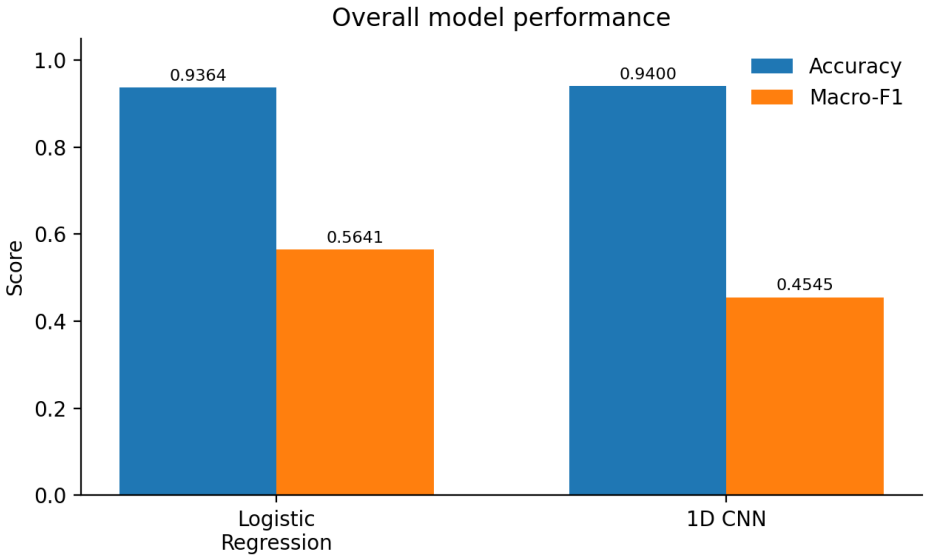

Figure 3. Overall Model Performance Comparison

3.3. CNN Training Behavior

The CNN learning curve indicates that training accuracy increased across epochs, while validation loss became unstable after the early epochs. This pattern suggests that the CNN learned the dominant positive-review pattern more quickly than the minority-class patterns. The limited number of negative and neutral samples reduced the model's ability to form stable representations for those classes.

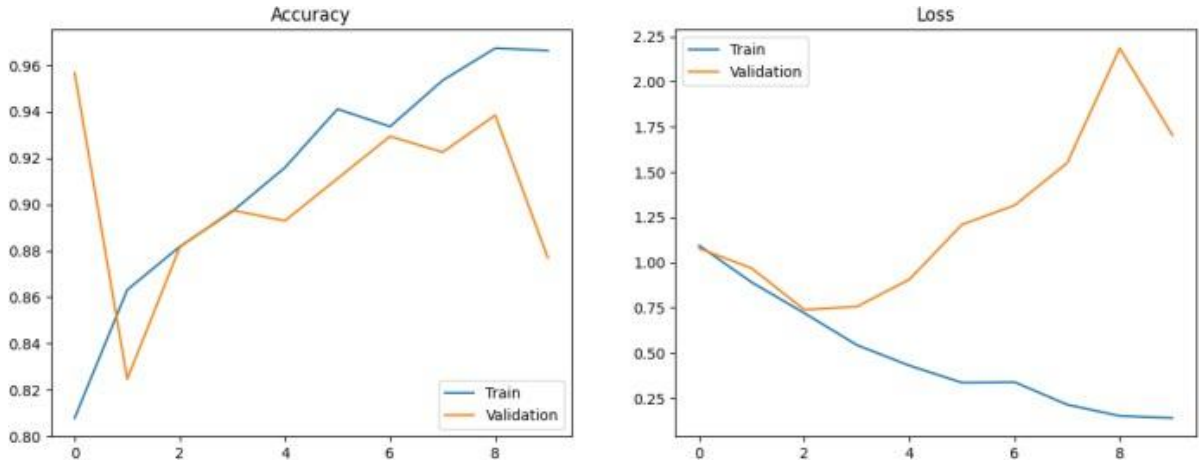

Figure 4. CNN Training Accuracy and Loss Curves

4. DISCUSSION

4.1. Why Logistic Regression Outperformed CNN on Macro-F1

The results show that the 1D CNN achieved a slightly higher accuracy than Logistic Regression, but its macro-F1 was lower. This means that CNN classified the majority positive class well but performed poorly when the test samples belonged to the negative and neutral classes. In a dataset where more than 95% of reviews are positive, a model can obtain high accuracy by learning a majority-class tendency. Macro-F1 exposes this weakness because it gives equal weight to all classes.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

Balanced Logistic Regression performed better because TF-IDF created sparse but discriminative features for short reviews, while class weighting increased the influence of negative and neutral samples during optimization [16], [17]. This finding is consistent with Indonesian sentiment studies showing that traditional models can remain competitive when datasets are short, noisy, and limited in size [16]. The CNN architecture is not inherently weak, but the training condition was unfavorable because the minority classes were too small to support robust convolutional pattern learning, a known challenge where positive-class dominance can lead to poor minority-class representation [18], [19].

4.2. Code-Switching and Informal Marketplace Language

Code-switching affects both representations. In TF-IDF, English expressions create additional vocabulary items. This can be useful when the terms carry strong polarity, but it also increases sparsity when users spell words inconsistently. In CNN, code-switched tokens can become rare or out-of-vocabulary items that weaken the embedding representation. For this reason, the preprocessing pipeline retained English sentiment words instead of deleting them. Removing terms such as recommended, worth it, fast delivery, and worst would remove valuable sentiment cues.

Table 8. Examples of Code-Switched Expressions and Preprocessing Decisions

Expression type	Example	Decision	Reason
Positive English phrase	recommended, worth it	Retained	The phrase directly signals satisfaction.
Service-related phrase	fast delivery, slow respon	Retained and normalized when possible	The phrase describes delivery or service quality.
Informal Indonesian slang	gk, bgt, rekomen	Normalized	Normalization reduces sparse variants.
Negation	tidak bagus, kurang cepat	Retained	Negation changes polarity direction.
Platform noise	barang sudah diterima	Removed when detected as template text	Template text can dominate features without reflecting sentiment.

4.3. Implications for Information Systems

From an information systems perspective, the model is useful as an automated review-monitoring component for sellers or marketplace analysts. Logistic Regression is more suitable as a practical baseline because it is interpretable, computationally light, and more stable under the current imbalance condition. The model can support dashboards that summarize negative complaints, neutral uncertainty, and positive satisfaction trends. However, deployment should include confidence scoring and manual review for minority-class predictions because the neutral class remains difficult to classify.

4.4. Limitations

This study has four limitations. First, the dataset came from only three electronic products, so the findings cannot represent all Shopee product categories. Second, rating-based labels may contain noise because some users assign high ratings while writing negative comments, or assign low ratings for delivery problems rather than product quality. Third, the class distribution was extremely imbalanced, which limited minority-class learning. Fourth, the code-switching analysis was discussed from observed preprocessing and error tendencies, but it was not yet supported by a separate manually validated bilingual subset. These limitations should guide future development rather than weaken the value of the baseline experiment.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

5. CONCLUSION

This study developed SentiTech-ID for three-class sentiment analysis of Shopee electronic product reviews. The dataset contained 5,499 reviews from three electronic products and showed extreme class imbalance, with positive reviews representing 95.64% of all samples. Logistic Regression with TF-IDF word n-gram achieved 93.64% accuracy and 0.5641 macro-F1, while the 1D CNN achieved 94.00% accuracy and 0.4545 macro-F1. The main finding is clear: the slightly higher CNN accuracy does not indicate better overall classification quality because its minority-class F1-scores were lower. Logistic Regression is the stronger baseline for this dataset because it handles sparse short-text features and class weighting more effectively.

Future research should expand the dataset to more product categories, manually validate a subset of rating-based labels, apply oversampling or data augmentation for minority classes, compare Logistic Regression with SVM and transformer-based models such as IndoBERT, and evaluate code-switched reviews using a dedicated bilingual test set. These steps would improve robustness and support deployment as a review intelligence component in Indonesian e-commerce information systems.

REFERENCES

[1] V. Gooljar, T. Issa, S. Hardin-Ramanan, and B. Abu-Salih, "Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review," *Journal Of Big Data*, vol. 11, no. 1, Aug. 2024, doi: 10.1186/s40537-024-00947-0.

[2] H. Murfi, Syamsyuriani, T. Gowandi, G. Ardaneswari, and S. Nurrohmah, "BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis," *Applied Soft Computing*, vol. 151, pp. 111112–111112, Dec. 2023, doi: 10.1016/j.asoc.2023.111112.

[3] W. Widyananda, M. Maskur, and M. F. A. Fauzi, "Sentiment analysis of Indonesian e-commerce product reviews using machine learning classifiers," *International Journal of Science and Research Archive*, vol. 16, no. 2, pp. 445–450, Aug. 2025, doi: 10.30574/ijsra.2025.16.2.2345.

[4] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine," *Journal of Information Systems Engineering and Business Intelligence*, vol. 4, no. 1, pp. 57–57, Apr. 2018, doi: 10.20473/jisebi.4.1.57-64.

[5] N. Nabiilah, S. Rohimah, and S. K. P. Loka, "Analyzing Customer Sentiment Towards Marketplace Reviews Using Classification Algorithms," *Indonesian Journal of Applied Technology and Innovation Science*, vol. 2, no. 1, pp. 23–30, Feb. 2025, doi: 10.57152/ijatis.v2i1.1774.

[6] C. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *Journal Of Big Data*, vol. 10, no. 1, pp. 88–88, May 2023, doi: 10.1186/s40537-023-00782-9.

[7] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, "Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/2826773.

[8] M. A. Ayu and A. H. Muhendra, "Preprocessing of Slang Words for Sentiment Analysis on Public Perceptions in Twitter," in *Contact-less Assessment of In-vivo Body Signals Using Microwave Doppler Radar (InTech)*, Intech (Slovakia), 2023. doi: 10.5772/intechopen.113725.

[9] J. Li, Advances in Sentiment Analysis - Techniques, Applications, and Challenges. IntechOpen, 2023. doi: 10.5772/intechopen.111293.

[10] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 1, pp. 53–53, Jan. 2019, doi: 10.22146/ijccs.41136.

[11] L. W. Astuti, Y. Sari, and S. Suprapto, "Code-Mixed Sentiment Analysis using Transformer for Twitter Social Media Data," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, Jan. 2023, doi: 10.14569/ijacsa.2023.0141053.

[12] S. Cahyawijaya *et al.*, "NusaCrowd: Open Source Initiative for Indonesian NLP Resources," Dec. 19, 2022, *Cornell University*. doi: 10.48550/arxiv.2212.09648.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

[13] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Sep. 11, 2020, *Cornell University*. doi: 10.48550/arxiv.2009.05387.

[14] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.

[15] A. Setiyawan, A. W. Wijayanto, and Y. He, "Extracting Consumer Opinion on Indonesian E-Commerce: A Rating Evaluation and Lexicon-Based Sentiment Analysis," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2021, no. 1, pp. 2–11, Jan. 2022, doi: 10.34123/icdsos.v2021i1.22.

[16] Herlinah, S. Aziz, H. M. I. Sultan, Yuyun, M. A. Nur, and Nurfaedah, "Comparison of Machine Learning Models for Indonesian Housewifes' Digital Discourse," pp. 422–427, Dec. 2025, doi: 10.1109/bts-i2c67944.2025.11399412.

[17] N. Chamidah, D. Widiyanto, H. B. Seta, and A. A. Aziz, "The Impact of Oversampling and Undersampling on Aspect-Based Sentiment Analysis of Indramayu Tourism Using Logistic Regression," *Revue d intelligence artificielle*, vol. 38, no. 3, pp. 795–804, Jun. 2024, doi: 10.18280/ria.380306.

[18] A. Romadhony, S. A. Faraby, R. Rismala, U. N. Wisesty, and A. Arifianto, "Sentiment Analysis on a Large Indonesian Product Review Dataset," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 167–178, Feb. 2024, doi: 10.20473/jisebi.10.1.167-178.

[19] Y. A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," *Journal of Information System Research (JOSH)*, vol. 6, no. 2, pp. 976–986, Jan. 2025, doi: 10.47065/josh.v6i2.6505.

This is an open-access article under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).