International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

Sentiment analysis of Indonesian e-commerce product reviews using machine learning classifiers

Wahyu Widyananda *, Maskur and Ahmad Fauzi

State Polytechnic of Malang, Malang, Indonesia.

International Journal of Science and Research Archive, 2025, 16(02), 445-450

Publication history: Received on 29 June 2025; revised on 06 August 2025; accepted on 09 August 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.2.2345

Abstract

The rapid growth of e-commerce in Indonesia has generated a massive volume of user-generated content in the form of product reviews. This textual data is a rich resource for businesses to understand customer satisfaction and product performance. This study aims to perform sentiment analysis on Indonesian e-commerce product reviews to automatically classify them into positive and negative sentiments. A publicly available dataset of Indonesian marketplace product reviews was utilized. The methodology involved text preprocessing steps, including case folding, tokenization, stopword removal, and stemming, followed by feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). Two common machine learning classifiers, Naïve Bayes and Support Vector Machine (SVM), were trained and evaluated on the dataset. The performance of the models was measured using accuracy, precision, recall, and F1-score. The results indicate that the Support Vector Machine (SVM) classifier achieved a higher accuracy of 89.5% compared to the Naïve Bayes classifier, which achieved an accuracy of 84.2%. These findings demonstrate the effectiveness of machine learning techniques in analyzing consumer sentiment in the Indonesian language, providing a valuable tool for market intelligence and business decision-making.

Keywords: Sentiment Analysis; Data Mining; E-commerce; Machine Learning; Naïve Bayes; Support Vector Machine; Text Mining; Indonesian Language

1. Introduction

The digital transformation across the globe has been accelerated by the expansion of the internet access worldwide, causing people to change how they shop and interact online. Indonesia, as one of the largest and fastest-growing digital economies in Southeast Asia, has witnessed an exponential rise in e-commerce platforms and users [1]. The growth of the digital economy is a key government focus, with studies highlighting the critical role of user trust and technology adoption in this expansion [2]. Platforms like Tokopedia, Shopee, and Bukalapak have become key component to the retail landscape, facilitating millions of transactions daily. A direct consequence of this growth is the creation of vast amounts of user-generated data, particularly in the form of product reviews. These reviews contain valuable opinions, sentiments, and feedback from customers, representing a goldmine of information for businesses [3].

Manually reviewing large volumes of customer feedback is inefficient, time-consuming, and prone to human bias. This challenge highlights the need for automated methods to process and extract insights from this data. Data mining, specifically through the subfield of text mining and Natural Language Processing (NLP), offers a powerful solution [4]. Sentiment analysis, or opinion mining, is a technique used to systematically identify, extract, and quantify affective states and subjective information from text data [5]. By applying sentiment analysis, businesses can automatically assess public attitudes and opinions, monitor brand and product perception, and gain actionable insights to improve products and services [6].

* Corresponding author: Wahyu Widyananda

The strategic value of these automated insights is immense. By systematically analyzing customer sentiment, e-commerce companies can move beyond reactive problem-solving and towards proactive strategy development. For instance, consistently negative sentiment towards a product can trigger a review of its quality or description, while positive sentiment can be leveraged in marketing campaigns. This data-driven approach allows businesses to rapidly adapt to market demands, enhance customer loyalty, and ultimately gain a significant competitive advantage in a crowded marketplace.

This research focuses on applying and comparing two widely used machine learning algorithms for the task of sentiment analysis on Indonesian e-commerce reviews. We investigate the performance of a probabilistic classifier, Naïve Bayes, and a linear classifier, Support Vector Machine (SVM), in categorizing product reviews as either 'positive' or 'negative'. The primary objective is to determine which model provides better performance for this specific context, thereby providing a practical framework for sentiment classification of Indonesian text.

2. Literature Review

Sentiment analysis has been a subject of extensive research for over two decades. Early work primarily focused on the English language, but there is a growing body of research on other languages, including Indonesian [7], [8].

Machine learning approaches for sentiment analysis can be broadly categorized into supervised, unsupervised, and semi-supervised methods. Supervised learning is the most common approach, where a labelled dataset (text documents pre-tagged with their sentiment) is used to train a classifier [9]. Several algorithms have been successfully applied. The Naïve Bayes classifier, based on Bayes' theorem with an assumption of conditional independence between features, is known for its simplicity and efficiency, especially in text classification tasks [10]. Despite its "naïve" assumption, it often performs surprisingly well and serves as a strong baseline model [11].

The Support Vector Machine (SVM) is another powerful supervised learning algorithm. SVM works by finding an optimal hyperplane that separates data points of different classes in a high-dimensional space [12]. It is highly effective in high-dimensional spaces, which is characteristic of text data where each word can be a feature. Studies have frequently shown SVM to be one of the top-performing classifiers for text sentiment analysis, often outperforming other traditional machine learning models [13], [14].

Research in the Indonesian context has explored various techniques. A study by [15] on Indonesian hotel reviews using Long Short-Term Memory (LSTM) networks showed the promise of deep learning models. Another study by [16] compared Decision Tree, Random Forest, and SVM for Indonesian e-commerce reviews, finding that SVM yielded high accuracy. More recent work has also focused on creating robust datasets and models for the Indonesian language, such as IndoBERT, which leverages the transformer architecture for better contextual understanding [17]. This study builds upon existing work by providing a direct comparison of Naïve Bayes and SVM on a publicly accessible Indonesian marketplace dataset, contributing to the body of knowledge on sentiment analysis for this specific linguistic and commercial environment [18].

3. Methodology

This research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which provides a structured approach for data mining projects [19]. The process includes data understanding, data preparation, modelling, and evaluation.

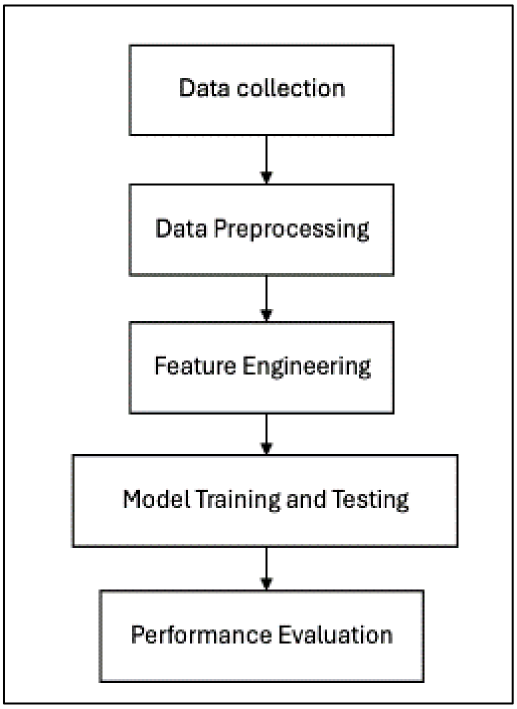

Figure 1 Research Methodology Design

3.1. Dataset

The dataset used in this study is the "Indonesian Marketplace Product Reviews" dataset, publicly available on the Kaggle platform. The dataset contains several hundred reviews, each with a text review in the Indonesian language and a corresponding sentiment label ('positive' or 'negative') [20].

3.2. Data Preprocessing

Raw text data is unstructured and contains noise. To prepare the data for the machine learning models, a series of standard NLP preprocessing steps were performed [21]:

- **Case Folding:** All text was converted to lowercase to ensure uniformity (e.g., "Bagus" and "bagus" are treated as the same word).
- **Tokenization:** The text was broken down into individual words or tokens.
- **Stopword Removal:** Common Indonesian words that do not carry significant meaning (e.g., "dan", "di", "yang") were removed using a standard Indonesian stopword list. The importance of a language-specific stopword list is crucial for effective feature selection [22].
- **Stemming:** Words were reduced to their root form (e.g., "membeli" becomes "beli"). The Sastrawi library, a popular rule-based stemmer for the Indonesian language, was used for this purpose [23].

3.3. Feature Extraction

After preprocessing, the textual data needs to be converted into a numerical format that machine learning algorithms can understand. This was achieved using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique. TF-IDF evaluates how relevant a word is to a document in a collection of documents. It assigns a weight to each word, with higher weights given to words that are frequent in a document but rare across all documents [4].

3.4. Modeling and Evaluation

The pre-processed and vectorized data was split into a training set (80%) and a testing set (20%). Two classification models were built using the training data:

- **Naïve Bayes:** Specifically, the Multinomial Naïve Bayes classifier was used, which is well-suited for text classification problems with discrete features like word counts [10].

- **Support Vector Machine (SVM):** A linear kernel was used for the SVM classifier due to its efficiency and effectiveness with high-dimensional text data [13].

The performance of each model was evaluated on the unseen testing set using a confusion matrix. From the confusion matrix, the following metrics were calculated [24]:

- **Accuracy:** The proportion of correctly classified reviews.
- **Precision:** The proportion of positive predictions that were actually correct.
- **Recall (Sensitivity):** The proportion of actual positives that were identified correctly.
- **F1-Score:** The harmonic mean of precision and recall, providing a single score that balances both metrics.

4. Results and Discussion

The two machine learning models were trained and tested, and their performance metrics were recorded. The results are visualized in Figure 2.

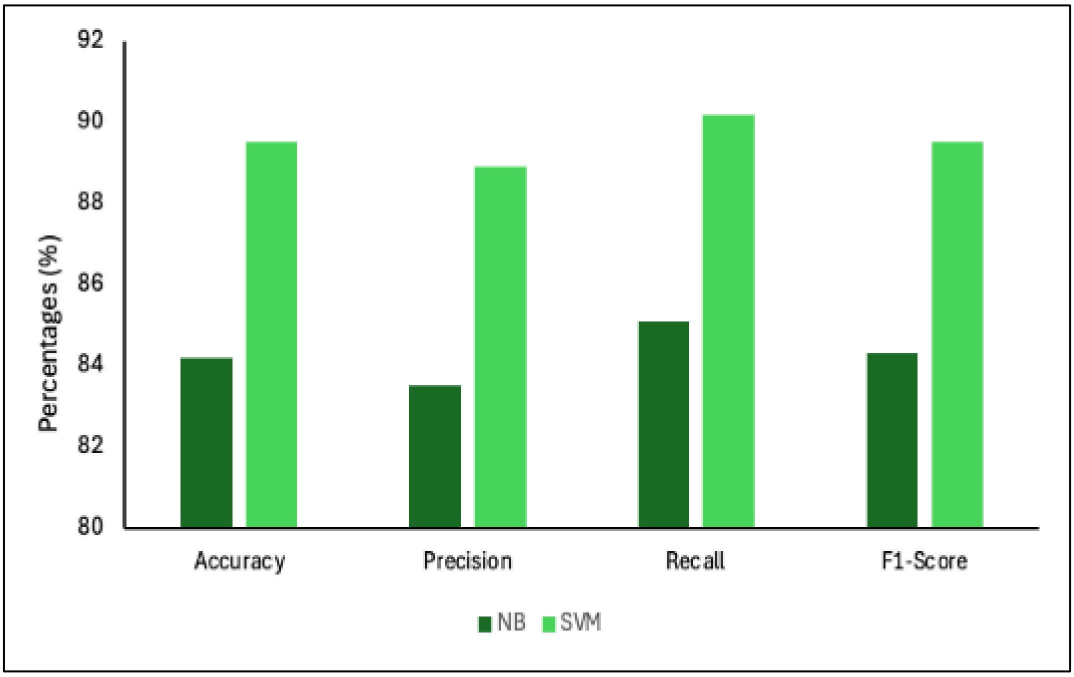

Figure 2 Comparison of Model Performance Metrics

The results clearly show that the Support Vector Machine (SVM) classifier outperformed the Naïve Bayes classifier across all four evaluation metrics. The SVM model achieved an accuracy of 89.5%, meaning it correctly classified the sentiment of nearly 9 out of 10 percent of reviews in the test set. This is 5.3 percentage points higher than the Naïve Bayes model's accuracy.

Table 1 Performance Metrics of Classification Models

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	84.2%	83.5%	85.1%	84.3%
Support Vector Machine (SVM)	89.5%	88.9%	90.2%	89.5%

The higher precision and recall of the SVM model indicate that it is not only more accurate overall but also better at correctly identifying positive reviews (precision) and capturing the majority of all positive reviews in the dataset (recall). The F1-Score, which balances precision and recall, further confirms the superior and more robust performance of the SVM model for this task.

The better performance of SVM can be attributed to its ability to handle high-dimensional data effectively. In text classification, the vocabulary size can be very large, resulting in a high-dimensional feature space. SVM's mechanism of finding the optimal separating hyperplane is particularly powerful in this scenario, whereas the Naïve Bayes model's strong independence assumption may not fully hold true for natural language, limiting its performance [25].

5. Conclusion

This study successfully applied and compared two machine learning classifiers, Naïve Bayes and Support Vector Machine, for sentiment analysis of Indonesian e-commerce product reviews. The results demonstrated that both models are capable of classifying sentiment with a reasonable degree of accuracy. However, the Support Vector Machine (SVM) model showed significantly better performance, achieving an accuracy of 89.5%.

This research confirms that data mining techniques are highly effective for extracting valuable insights from customer reviews in the Indonesian language. E-commerce businesses can leverage these models to automate the process of sentiment analysis, enabling them to monitor customer feedback in real-time, identify areas for product improvement, and enhance overall customer satisfaction. Future work could explore more advanced deep learning models like LSTM or Transformers (e.g., IndoBERT) and expand the analysis to include multi-class sentiment (e.g., positive, negative, neutral) or aspect-based sentiment analysis.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] K. D. Pratiwi, "E-Commerce and Economic Growth in Indonesia: Analysis of Panel Data Regression," Gema Publica, vol. 7, no. 1, pp. 171–186, 2022, doi: 10.14710/gp.7.1.2022.171-186.

[2] D. P. Syaharani and N. N. K. Yasa, "The Role of Trust as Mediation between Perceived Usefulness and Perceived Ease of Use on Repurchase Intention," European Journal of Development Studies, vol. 2, no. 3, pp. 36–40, 2022, doi: 10.24018/ejdevelop.2022.2.3.91.

[3] M. Hu and B. Liu, "Mining and summarizing customer reviews," KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177, 2004, doi: 10.1145/1014052.1014073.

[4] A. Rajaraman and J. D. Ullman, "Mining of massive datasets," Mining of Massive Datasets, vol. 9781107015, pp. 1–315, 2011, doi: 10.1017/CBO9781139058452.

[5] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–184, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[6] R. Feldman, "Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science," Commun ACM, vol. 56, no. 4, pp. 82–89, 2013, doi: 10.1145/2436256.2436274.

[7] A. Ilmania, Abdurrahman, S. Cahyawijaya, and A. Purwarianti, "Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis," Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018, pp. 62–67, 2019, doi: 10.1109/IALP.2018.8629181.

[8] A. R. Prananda and I. Thalib, "Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion," Journal of Information Systems Engineering and Business Intelligence, vol. 6, no. 1, p. 1, 2020, doi: 10.20473/jisebi.6.1.1-8.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Association for Computational Linguistics, 2002, pp. 79–86. doi: 10.3115/1118693.1118704.

[10] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41–48, 1998, doi: 10.1.1.46.1529.

[11] Irina Rish, "An empirical study of the naive Bayes classifier," IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, pp. 41–46, 2001.

[12] C. Cortes and V. Vapnik, "Support-vector networks," Mach Learn, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/bf00994018.

[13] T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," pp. 137–142, 1998, doi: 10.1007/bfb0026683.

[14] Y. Xu and H. Wang, "A new feature selection method based on support vector machines for text categorization," International Journal of Data Analysis Techniques and Strategies, vol. 3, no. 1, pp. 1–20, 2011, doi: 10.1504/IJDATS.2011.038803.

[15] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews," Procedia Comput Sci, vol. 179, pp. 728–735, 2021, doi: 10.1016/j.procs.2021.01.061.

[16] C. C. P. Hapsari, W. Astuti, and M. D. Purbolaksono, "Naive Bayes Classifier and Word2Vec for Sentiment Analysis on Bahasa Indonesia Cosmetic Product Reviews," 2021 International Conference on Data Science and Its Applications, ICoDSA 2021, pp. 22–27, 2021, doi: 10.1109/ICoDSA53588.2021.9617544.

[17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.

[18] R. Sutoyo, S. Achmad, A. Chowanda, E. W. Andangsari, and S. M. Isa, "PRDECT-ID: Indonesian product reviews dataset for emotions classification tasks," Data Brief, vol. 44, 2022, doi: 10.1016/j.dib.2022.108554.

[19] C. Shearer et al., "The CRISP-DM model: The New Blueprint for Data Mining," Journal of Data Warehousing, 2000.

[20] Taqiyya Ghazi, Indonesian Marketplace Product Reviews. 2022. [Online]. Available: https://www.kaggle.com/datasets/taqiyyaghazi/indonesian-marketplace-product-reviews

[21] D. Jurafsky and J. H. Martin, Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition, no. 3rd. 2025. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/

[22] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7649 LNCS, no. PART 1, pp. 508–524, 2012, doi: 10.1007/978-3-642-35176-1_32.

[23] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian," ACM Transactions on Asian Language Information Processing, vol. 6, no. 4, pp. 1–33, 2007, doi: 10.1145/1316457.1316459.

[24] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, [Online]. Available: http://arxiv.org/abs/2010.16061

[25] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," Soft comput, vol. 25, no. 3, pp. 2277–2293, 2021, doi: 10.1007/s00500-020-05297-6.