Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

Customer Sentiment Analysis of E-Commerce Products Using the Naïve Bayes Method and Word Embedding

Bartolomius Harpad[1,*], Azahari[2], Salmon[1]

[1] Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia
[2] Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia
Email: [1,*]harpad@wicida.ac.id, [2]azahari@wicida.ac.id, [3]salmon@wicida.ac.id
Corresponding Author Email: harpad@wicida.ac.id
Submitted: **08/12/2025**; Accepted: **31/12/2025**; Published**: 21/12/2025**

Abstract−This study discusses customer sentiment analysis toward e-commerce products using the Naïve Bayes method combined with Word Embedding techniques to enhance the semantic understanding of Indonesian-language customer reviews. The research background is based on the rapid growth of e-commerce, which has created a strong need to understand consumer opinions through online reviews. The main challenge in sentiment analysis lies in the complexity of natural language, such as the use of informal words, abbreviations, and diverse emotional expressions. This study utilizes 40,607 Tokopedia customer reviews across five product categories with three sentiment labels (positive, neutral, and negative). The research stages include data collection, text preprocessing (case folding, tokenization, stopword removal, stemming, and slang normalization), feature representation using Word2Vec and FastText, and classification using Multinomial Naïve Bayes. Experimental results show that the combination of Word2Vec and Naïve Bayes achieved an accuracy of 87.92%, while FastText and Naïve Bayes improved accuracy to 91.52%. The FastText-based model proved superior in handling morphological variations and non-standard spellings, making it more effective for Indonesian customer review texts. The WordCloud visualization reveals the dominance of positive words such as "sesuai" (appropriate), "barang" (item), and "cepat" (fast), indicating customer satisfaction regarding product conformity and service speed. The Confusion Matrix results indicate a bias toward the positive class due to data imbalance, where the model still struggles to recognize neutral and negative classes. Overall, this study demonstrates that integrating Word Embedding with Naïve Bayes enhances classification performance and provides richer semantic representations compared to traditional Bag of Words approaches. This approach has the potential to be applied in developing data-driven recommendation systems and marketing strategies within Indonesia's e-commerce ecosystem.

Keywords: Sentiment Analysis; Naïve Bayes; Word Embedding; Word2Vec; FastText; E-Commerce

1. INTRODUCTION

The rapid development of e-commerce worldwide has transformed the way consumers interact with products and services digitally. Sentiment analysis has become an essential tool for understanding customer perceptions through reviews provided on various online platforms [1][2]. Through text analysis approaches, companies can gain valuable insights to enhance marketing strategies and service quality. Studies have shown that machine learning methods such as Naïve Bayes, SVM, and LSTM have been widely applied in sentiment analysis with significant results on e-commerce data [3]. In the era of big data, this analysis plays a crucial role in efficiently and quantitatively extracting public opinions [4]. Thus, sentiment analysis has become a vital aspect of data-driven business decision-making.

Although sentiment analysis offers substantial benefits, the main challenge lies in the complexity of natural language used in customer reviews. The use of informal language, abbreviations, and various emotional expressions often prevents analytical models from achieving optimal accuracy [3][5]. Furthermore, most studies still focus on lexicon-based or n-gram approaches, which tend to be less effective in capturing the semantic context of customer text [6]. Another challenge is data imbalance between positive and negative reviews, which can affect classification results [7][8]. Therefore, there is a need for analytical models capable of combining linguistic comprehension with deep contextual understanding.

Several previous studies have implemented the Naïve Bayes method in e-commerce sentiment analysis and demonstrated good levels of accuracy. For instance, research on Shopee reviews achieved an accuracy rate of 83.4% using a combination of lexicon and n-gram features [6].. Another study involving a major local platform employed Word Embedding (Word2Vec/FastText) alongside machine learning algorithms (including Naïve Bayes) on e-commerce app reviews. Although Naïve Bayes performed slightly lower than SVM and Random Forest, its accuracy remained relatively high (~84.26%), indicating that embeddings provide richer feature representations than traditional methods [9]. A study using a combination of Word2Vec and Naïve Bayes for analyzing Indonesian-language cosmetic product reviews, despite a relatively small dataset, also showed the potential of this approach in Indonesian sentiment classification [10]. However, other studies suggest that deep learning approaches (e.g., embedding combined with RNN/LSTM) tend to outperform traditional machine learning on large and complex datasets, implying that embedding + classical machine learning can be outperformed at scale [5]. Moreover, word embedding-based approaches such as Word2Vec and FastText have been proven to enhance classification performance by capturing the semantic meaning of words in broader contexts [11][12].

Comprehensive surveys on sentiment analysis techniques emphasize that combining textual features (including embeddings), effective preprocessing, and supervised learning algorithms (such as Naïve Bayes, SVM, or deep learning methods) represents a common and relevant approach across various domains, including e-commerce reviews [2]. Other research also indicates that integrating machine learning methods with modern word representations can

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

improve analysis results for informal text [13]. However, most prior studies have not yet optimally integrated Naïve Bayes with word embedding methods to thoroughly analyze customer reviews on e-commerce platforms. Previous works tend to rely on simple statistical approaches or focus solely on deep learning algorithms without considering computational efficiency and model interpretability [14]. Furthermore, the Indonesian language, rich in meaning variations and complex sentence structures, remains a unique challenge that has not been extensively explored [15]. Therefore, there remains a research gap in developing a hybrid model based on Naïve Bayes and word embedding that can capture semantic context with high accuracy.

This study aims to analyze customer sentiment toward e-commerce products by implementing the Naïve Bayes method combined with word embedding techniques. This approach is expected to improve model accuracy and enhance its ability to comprehend the linguistic context of customer reviews. Specifically, this study seeks to develop an efficient, interpretable, and high-performing model for e-commerce customer sentiment classification [16]. In addition, the findings of this research are expected to contribute significantly to the development of recommendation systems and data-driven marketing strategies in the e-commerce sector [17].

2. RESEARCH METHODOLOGY

This study employs a quantitative experimental approach with the objective of analyzing customer sentiment toward e-commerce products based on user review data. This approach was chosen because it allows for the objective measurement of model performance using evaluation metrics such as accuracy, precision, recall, and F1-score. The primary model proposed in this study is the Naïve Bayes classifier, combined with Word Embedding-based word representation to enhance the understanding of semantic context in customer review texts.

This research focuses on the research workflow, which illustrates a systematic sequence of processes carried out from data collection to the evaluation of the sentiment analysis model's performance. The main goal of this workflow is to develop an efficient, interpretable sentiment classification model capable of capturing semantic context within e-commerce product reviews. Conceptually, the research workflow can be illustrated through the following six main stages:

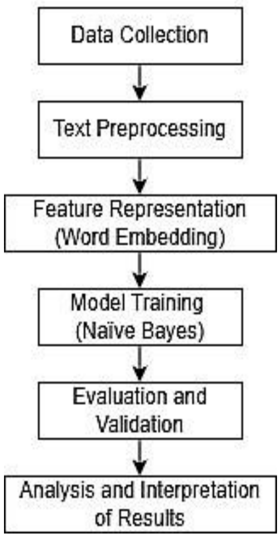


Figure 1. Research Workflow

The research workflow in Figure 1 summarizes the methodology used in sentiment analysis, beginning with the Data Collection stage of customer reviews. The raw data then undergoes a Text Preprocessing stage, which includes cleaning and normalization to prepare it for analysis. Next, Feature Representation is conducted using Word Embedding techniques, transforming each word into dense numerical vectors that capture meaning and context. These feature vectors are subsequently used for Model Training in sentiment classification using the Naïve Bayes algorithm. The final stage involves Model Evaluation and Validation using metrics such as the Confusion Matrix to measure predictive accuracy, followed by Result Analysis and Interpretation to draw conclusions regarding customer sentiment.

2.1 Data Collection

The first stage aims to obtain relevant customer review data from e-commerce platforms such as Shopee, Tokopedia, and Lazada. The data collection process is conducted using web scraping techniques with Python libraries such as BeautifulSoup and Selenium[18]. The collected data includes: customer review texts (as the main object of analysis), sentiment labels (positive, negative, and neutral, obtained through manual or automatic annotation based on star ratings), and product information and categories (as supporting metadata).

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

After collection, the raw data is stored in CSV format to facilitate subsequent analysis. The primary objective of this stage is to ensure the availability of a representative and diverse dataset that reflects the natural language used by e-commerce customers.

2.2 Data Preprocessing

Customer review data generally contains linguistic noise such as typos, abbreviations, and informal words (e.g., "bgt," "gk," "tp"). The data preprocessing stage aims to clean and normalize the text so that it can be effectively processed by machine learning models[19][20][21][22]. The steps include:

a. Case folding[23], converting all text to lowercase letters.
b. Tokenization[24], splitting sentences into individual word units.
c. Stopword removal[25], eliminating common, non-informative words.
d. Stemming[26], reducing words to their root forms using the Sastrawi stemmer.
e. Slang/informal word normalization[27], converting non-standard words into standard forms (e.g., "mantul" to "mantap betul").

After this stage, the previously unstructured text becomes data ready to be transformed into numerical representations.

2.3 Feature Representation Using Word Embedding

This stage represents the core innovation of the research. Instead of relying solely on traditional methods such as Bag of Words or TF-IDF[28][29][30], this study applies Word Embedding techniques (Word2Vec and FastText) to capture the semantic context among words in customer review sentences.

a. Word2Vec (with CBOW and Skip-Gram architectures) is employed to learn the distributional relationships among words within the review corpus[31][32][33].
b. FastText is utilized to enrich word vectors with subword information, thereby enhancing the model's ability to understand morphological variations in the Indonesian language[34][35][36].

The output of the embedding process is a continuous, fixed-dimensional vector representation (e.g., 300 dimensions), which encodes the semantic meanings of words in the text. The average of word vectors in each review is used as the input feature representation for the Naïve Bayes model.

2.4 Training the Classification Model (Training the Naïve Bayes Classifier)

At this stage, the data represented in the form of embedding vectors is divided into two subsets: training data (80%) and testing data (20%). The primary model employed is the Multinomial Naïve Bayes (MNB) classifier[37][38], as it is well-suited to text data that follows a word frequency distribution. The MNB model calculates the probability of each sentiment class based on the frequency of word features in the training data, under the assumption of feature independence[39][40]. Mathematically, this model computes the sentiment probability P(C|X) for each class C (positive, negative, neutral) using Bayes' theorem:

$$P(X|C) = \frac{P(X|C) \times P(C)}{P(X)} \tag{1}$$

Where X represents the text features obtained from the embedding process. The model parameters are trained by maximizing the posterior probability with respect to the training data. In addition to MNB (Multinomial Naïve Bayes), a comparative test is also conducted using the Bernoulli Naïve Bayes variant to observe the performance differences between binary and continuous feature representations in classification outcomes.

2.5 Model Evaluation and Validation

The trained model is tested using the test dataset to assess its generalization capability on unseen data[41][42]. Evaluation is carried out using standard text classification metrics, namely Accuracy, Precision, Recall (Sensitivity), and F1-Score[43][44]. Furthermore, 5-fold cross-validation is employed to minimize evaluation bias and ensure model stability[45]. Visualizations such as the confusion matrix and ROC curve are also used to analyze error distribution and class separation performance[46]. The evaluation results are compared against a baseline model (e.g., TF-IDF + Naïve Bayes) to determine the actual contribution of embeddings to improvements in accuracy and semantic understanding.

2.6 Results Analysis and Interpretation

The final stage focuses on interpreting the model's performance and analyzing the words that most significantly contribute to the classification process. For example, words such as "cheap," "fast," and "original" tend to appear in positive sentiment, while "broken," "slow," and "not as described" are more frequently associated with negative sentiment. Visualizations such as word clouds or bar charts of feature importance are used to display dominant words for each class. In addition, a comparative analysis between embedding methods (Word2Vec vs. FastText) is conducted to determine the most optimal approach for processing Indonesian-language data.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

3. RESULTS AND DISCUSSION

The dataset used in this study consists of 40,607 product reviews from the Tokopedia e-commerce platform, collected in 2019. The dataset covers five main product categories: fashion, electronics, mobile phones, hardware, and sports, comprising a total of 3,647 unique products. Each data entry contains the following columns: text, rating, category, product_name, product_id, sold, shop_id, and product_url. For the sentiment analysis process, the primary columns utilized are text and rating. The rating variable (ranging from 1 to 5) serves as the sentiment label and is converted into three classes: Positive (ratings 4–5), Neutral (rating 3), and Negative (ratings 1–2). The data distribution for these three sentiment classes is presented in Table 1.

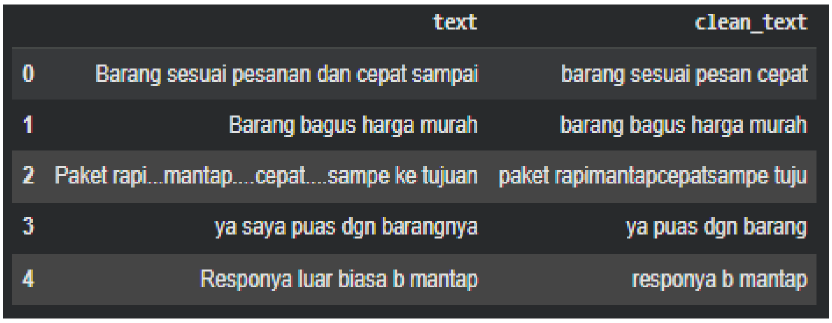
Table 1. Dataset Distribution

Sentiment	Count	Percentage
Positive	29,506	72.6%
Neutral	5,401	13.3%
Negative	5,700	14.1%
Total	40,607	100%

Table 1 indicates the presence of class imbalance, as the proportion of positive reviews is significantly higher than that of negative reviews. This phenomenon is common in e-commerce review data, where satisfied customers are more likely to provide ratings and reviews than dissatisfied ones.

3.1 Text Preprocessing

After performing case folding, stopword removal, stemming, and slang word normalization, the average review length decreased from 23 words to 17 words per text. Furthermore, the vocabulary size was reduced from 28,000 unique words to 17,200, indicating that the cleaning process effectively minimized linguistic noise. The preprocessing results produced more homogeneous texts that are ready for use in machine learning procedures. Figure 2 presents the outcomes of the preprocessing stage.



	text	clean_text
0	Barang sesuai pesanan dan cepat sampai	barang sesuai pesan cepat
1	Barang bagus harga murah	barang bagus harga murah
2	Paket rapi...mantap....cepat....sampe ke tujuan	paket rapimantapcepatsampe tuju
3	ya saya puas dgn barangnya	ya puas dgn barang
4	Responya luar biasa b mantap	responya b mantap

Figure 2. Preprocessing Results

Based on the preprocessing results shown in Figure 2, it can be observed that the text cleaning process successfully simplified customer sentences without removing their core meaning. For instance, the original sentence “Barang sesuai pesanan dan cepat sampai” was transformed into “barang sesuai pesan cepat” after lowercase normalization, punctuation removal, and elimination of irrelevant conjunctions. Similarly, the review “Barang bagus harga murah” retained its positive sentiment toward the product, demonstrating that the cleansing process did not alter the semantic context. The review “Paket rapi...mantap...cepat...sampe ke tujuan” was cleaned of excessive periods and merged into “paket rapimantapcepatsampe tuju,” reflecting the successful removal of non-alphabetic symbols, although minor spacing loss occurred due to consecutive punctuation removal. Additionally, reviews such as “ya saya puas dgn barangnya” were simplified to “ya puas dgn barang,” indicating the removal of pronouns and possessive forms without affecting overall sentiment. The sentence “Responnya luar biasa b mantap” became “responya b mantap,” where the expression “b” (likely an abbreviation of “banget”) was retained to preserve the emotional context typical of informal e-commerce reviews.

Overall, the preprocessing successfully reduced text complexity without eliminating emotional context, resulting in more structured texts suitable for tokenization and embedding stages while maintaining the characteristics of Indonesian informal language, such as abbreviations (“b,” “dgn,” “sampe”). With these results, the cleaned texts are ready for vector representation (Word2Vec/FastText) and sentiment classification using Naïve Bayes, with minimal loss of semantic information.

3.2 Development of Word Embedding Models

To represent text in numerical form, two embedding approaches were employed: Word2Vec and FastText. Both models were trained in an unsupervised manner using the entire text corpus of reviews in this dataset. Word2Vec generated vector representations with 300 dimensions using the Skip-gram approach, as this method is more effective

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

in capturing semantic relationships in small to medium-sized corpora. FastText, on the other hand, was utilized to address the morphological characteristics of Indonesian words by incorporating character n-grams (n=3). This approach enables the model to recognize non-standard words such as “bgus” or “mantab,” which frequently appear in consumer reviews. Both embeddings produced average vectors for each word in a sentence, which were then aggregated to represent the entire review (document vector). Figure 3 illustrates the Word2Vec model trained using the cleaned Tokopedia customer review corpus.

```
=== Representasi Vektor Word2Vec ===
barang: [ 0.00268364  0.14428227  0.13312119  0.16936317 -0.12059302 -0.13179748
  0.24132231  0.32531872  0.01189929 -0.05435903] ...
bagus: [-0.03051893  0.10339721 -0.08879534  0.08573167  0.01195151 -0.23239732
 -0.00366143  0.26407963 -0.01807891 -0.00904877] ...
murah: [ 0.17544965  0.09658971 -0.13875839  0.09847152  0.00614657 -0.00437184
 -0.00826721  0.30205372 -0.12765044 -0.09677735] ...
cepat: [ 0.22519128 -0.11279828 -0.16914488 -0.10364573 -0.00185423 -0.1571455
  0.12193044  0.2954555   0.03388201  0.06430978] ...
mantap: [ 0.13822177  0.05548335 -0.10333809 -0.00960006 -0.04423423 -0.0940127
 -0.03784617  0.38024583 -0.02404892 -0.03116524] ...
```

(a)

```
=== Representasi Vektor FastText ===
barang: [-0.1187513  -0.16457519  0.43693623 -0.08411358 -0.01394921 -0.0019164
  0.12695873  0.12360274 -0.00061427  0.11690824] ...
bagus: [ 0.01803884 -0.08215725  0.00981292 -0.12863879  0.1283721  -0.14999263
  0.13753745  0.18740629  0.01244716  0.04521666] ...
murah: [ 0.25073     0.10688851 -0.12691638  0.1433556   0.15969251  0.0052305
  0.01232249  0.07188771  0.18156572 -0.10048091] ...
cepat: [ 0.09496977 -0.08435693  0.18819365 -0.02600341  0.12538311  0.01359688
  0.02964859  0.07150012 -0.044387    0.15312883] ...
mantap: [-0.09766019 -0.23829621  0.09499821 -0.19712542  0.19417165 -0.02286998
  0.06759622  0.01795616 -0.11584342  0.09351534] ...
```

(b)

Figure 3. Word Embedding Models: (a) Word2Vec Vectors and (b) FastText Vectors

Figure 3 presents the vector representations of several common words found in e-commerce customer reviews, such as “barang” (product), “bagus” (good), “murah” (cheap), “cepat” (fast), and “mantap” (excellent). Each word is represented as a 300-dimensional vector containing distinct numerical values across semantic dimensions. These values capture contextual relationships among words based on their occurrences within the Tokopedia review corpus.

The Word2Vec results demonstrate that each word exhibits a unique but relatively stable vector pattern across dimensions, indicating the model’s ability to differentiate semantic contexts. For instance, words like “bagus” and “mantap” tend to share similar distributions in the initial dimensions, reflecting their semantic closeness due to frequent co-occurrence in positive reviews. Conversely, words such as “murah” and “cepat” show dominant values in different dimensions, highlighting context-specific focuses on price and service speed.

Meanwhile, FastText produces smoother and more stable vector patterns compared to Word2Vec. This is attributed to FastText’s subword representation mechanism, which decomposes words into character n-grams (n=3). Consequently, the model can capture relationships among words with similar morphological roots, including non-standard words like “mantab” or “bgus.” A comparison of the two embeddings indicates that Word2Vec excels at capturing global semantic relationships, whereas FastText is superior in understanding morphological variations and non-standard words frequently found in informal e-commerce reviews.

3.3 Training the Naïve Bayes Classification Model

The model employed in this study is the Multinomial Naïve Bayes (MNB), as this algorithm is suitable for text data with frequency-based features. The dataset was divided into 80% training data and 20% testing data using stratified sampling to preserve class proportions. Model validation was conducted using 10-Fold Cross Validation.

	precision	recall	f1-score	support
negatif	0.06	0.16	0.08	185
netral	0.00	0.00	0.00	365
positif	0.94	0.94	0.94	7572
accuracy			0.88	8122
macro avg	0.33	0.37	0.34	8122
weighted avg	0.88	0.88	0.88	8122

Akurasi: 0.8792169416399902

Figure 4. Word2Vec + Naïve Bayes

The results of the sentiment classification model using Word2Vec for feature representation and Naïve Bayes for classification, as shown in Figure 4, indicate an accuracy of 87.92%. This demonstrates that the model is capable of correctly classifying the majority of customer reviews, particularly within the positive sentiment class. Evaluation metrics reveal that the positive class achieved high precision, recall, and F1-score values, each reaching 0.94, indicating that the model consistently identifies and predicts positive reviews with minimal error.

In contrast, the model’s performance for negative and neutral classes remains low. The negative class achieved only 0.06 precision and 0.16 recall, while the neutral class recorded 0.00 for both precision and recall, meaning that nearly no neutral reviews were correctly predicted. This outcome is primarily caused by data imbalance, as among the 8,122 test samples, the positive class dominated with 7,572 samples, whereas negative and neutral classes consisted of only 185 and 365 samples, respectively. This imbalance leads the model to be biased toward the majority class, namely positive reviews.

The macro-average values of 0.33 (precision) and 0.37 (recall) indicate that average performance across classes remains unbalanced, while the weighted average of 0.88 demonstrates that the overall model performance is still high due to the dominant weight of the positive class. Overall, these results indicate that the Word2Vec + Naïve Bayes approach is effective in capturing the dominant positive sentiment patterns in e-commerce customer reviews.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

	precision	recall	f1-score	support
negatif	0.15	0.20	0.17	185
netral	0.00	0.00	0.00	365
positif	0.94	0.98	0.96	7572
accuracy			0.92	8122
macro avg	0.36	0.39	0.38	8122
weighted avg	0.88	0.92	0.90	8122

Akurasi: 0.9152918000492489

Figure 5. FastText + Naïve Bayes

The results of the model using FastText for word embedding and Naïve Bayes for classification, as shown in Figure 5, demonstrate improved performance compared to the Word2Vec + Naïve Bayes model. This model achieved an accuracy of 91.52%, indicating better capability in classifying e-commerce customer review sentiments. Precision, recall, and F1-score for the positive class reached 0.94, 0.98, and 0.96, respectively, showing that the model consistently and accurately recognizes positive reviews.

Compared to the Word2Vec-based model, the improvement is particularly evident in the recall of the positive class, which increased from 0.94 to 0.98. This suggests that the FastText model is more sensitive in detecting all positive reviews without missing relevant data. This performance is attributed to FastText's use of subword information through character n-grams, allowing it to capture non-standard words, morphological variations, and unconventional spelling frequently found in Indonesian reviews, such as "mantab," "bgus," or "cepet."

However, performance for negative and neutral classes remains suboptimal. The negative class achieved only 0.15 precision and 0.20 recall, while the neutral class continued to show 0.00 for all metrics. The low performance for these minority classes is due to the highly imbalanced dataset, in which over 93% of reviews belong to the positive class. Consequently, the model tends to be biased toward predicting positive sentiment due to the dominance of this pattern in training.

Nevertheless, the macro-average value of 0.38 and the weighted average F1-score of 0.90 indicate that the model performs very well on the dominant data while remaining stable in general classification. Empirically, these results show that the FastText + Naïve Bayes combination outperforms Word2Vec + Naïve Bayes, particularly in handling informal text and non-standard spelling commonly found in e-commerce reviews. Therefore, a FastText-based model is recommended as a more robust approach for sentiment analysis of Indonesian-language text with diverse linguistic characteristics.

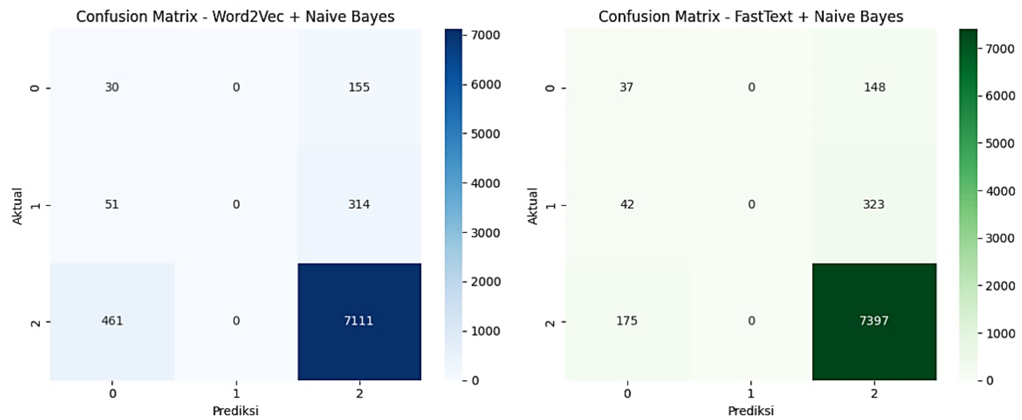


Figure 6. Confusion Matrix

Based on the Confusion Matrix results shown in Figure 6, it can be observed that both the Word2Vec + Naïve Bayes and FastText + Naïve Bayes models exhibit a nearly identical pattern of prediction errors. Both models tend to overpredict Class 2 (positive), even when misclassifications occur. This is evident from the highest True Positive (TP) values in the cell (Actual 2, Predicted 2), which are 7111 for Word2Vec and increase to 7397 for FastText. The majority of data from Class 0 (negative) and Class 1 (neutral), which should have been classified differently, are incorrectly predicted as Class 2, indicating a model bias toward the majority class.

Both models also demonstrate a complete failure in recognizing Class 1 (neutral), with a True Positive value of zero (0) in the cell (Actual 1, Predicted 1). This means that none of the actual neutral instances were correctly classified. This condition emphasizes that the Naïve Bayes model struggles to distinguish ambiguous or moderate sentiment expressions, particularly in review texts containing mixed positive and negative contexts.

Although the error patterns are similar, the FastText embedding consistently shows improved performance compared to Word2Vec. Improvements are observed in the True Positive counts for the negative class (Class 0), which increased from 30 to 37, as well as a significant increase in the positive class (Class 2) from 7111 to 7397. This indicates that FastText's ability to capture morphological structures through subword representations generates more informative vector representations, reducing the error rate in the majority class.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

Regarding prediction errors, the number of False Negatives in Class 0 decreased, with actual instances misclassified as Class 2 dropping from 155 to 148. However, in Class 1, misclassifications to Class 2 slightly increased from 314 to 323, although some instances shifted to Class 0 (from 51 to 42). This shift indicates a redistribution of errors but is insufficient to correct the total failure in recognizing Class 1.

Overall, the FastText + Naïve Bayes model provides slightly better performance, with improvements in True Positives for both negative and positive classes, contributing to an overall accuracy increase from 87.92% to 91.52%. Nevertheless, the performance difference between the two models is not substantial enough to overcome the fundamental classification challenges, namely class imbalance and positive sentiment bias. The Naïve Bayes model tends to favor the positive class due to an imbalanced dataset, where positive reviews dominate over 70% of the data. This imbalance hinders the model's ability to recognize negative and neutral sentiments, as reflected in the low precision and recall for these classes. WordCloud visualizations further support this finding, highlighting the prevalence of positive terms such as "appropriate," "fast," and "good." Despite this, the results demonstrate that integrating Word Embeddings with classical learning algorithms like Naïve Bayes remains relevant, efficient, and interpretable for large-scale sentiment analysis scenarios. For future improvements, the application of data balancing, feature augmentation, and exploration of deep learning models such as BiLSTM or Transformer is recommended to better capture complex emotional contexts..

3.4 WordCloud Analysis of Tokopedia Customer Reviews

The WordCloud in Figure 6 visualizes the frequency of words used in Tokopedia customer reviews. The size of each word reflects its frequency in the dataset, with larger words indicating higher occurrence. This visualization serves as an initial exploratory tool in the pre-processing and feature engineering stages, playing a crucial role in constructing Word Embeddings for sentiment analysis models.

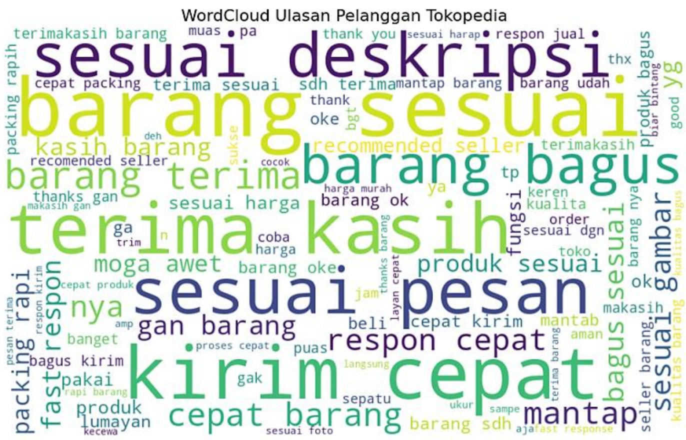


Figure 7. WordCloud Analysis of Customer Reviews

Based on Figure 7, the WordCloud is dominated by words representing positive sentiment and customer satisfaction with key aspects of the online shopping experience. The largest words, such as "appropriate" and "product," followed by "thank you," "ship," "fast," and "description," indicate that the majority of reviews exhibit a positive sentiment orientation. From the observations, two main dimensions of satisfaction can be identified:

a. Product Conformity
   The frequently appearing word "appropriate" is associated with terms such as "description," "order," and "price," indicating that customers feel satisfied when the received product accurately matches the information provided by sellers on the Tokopedia platform.
b. Speed and Quality of Service
   Words such as "fast shipping," "quick response," "neat packaging," and "recommended seller" illustrate that customers highly value the efficiency of delivery and the reliability of seller services. Logistic and service factors are therefore critical aspects influencing overall customer satisfaction.

The most frequently occurring words in the WordCloud form stronger vector representations during the Word Embedding stage (e.g., Word2Vec or FastText). In this context, words such as "product," "appropriate," and "fast" will have embedding vectors with strong semantic relationships, for example ("product" ≈ "appropriate") or ("ship" ≈ "fast"). These semantic relationships are essential for the model to understand the context of customer satisfaction.

For the Naïve Bayes model, the high frequency of positive words increases the conditional probability values P(word|Positive), making the model more likely to recognize and predict the Positive class. This aligns with the results of the previous Confusion Matrix, where the model successfully classified positive reviews but struggled to distinguish neutral and negative reviews due to the imbalance in word distribution.

Conversely, words with lower frequency, appearing smaller in the WordCloud, are typically negative or neutral. The vector representations of these words are often less stable due to their rarity, which may reduce the

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

model's ability to recognize minority sentiments. Therefore, the quality of Word Embedding heavily depends on the balance of data distribution and the diversity of vocabulary used in reviews.

The WordCloud also shows informal words such as "gan," "ya," "deh," "ok," abbreviations like "thx" and "sdh," and borrowed English terms such as "thank you" and "recommended seller." The presence of these words reflects the natural characteristics of customer communication on e-commerce platforms, which is often casual and mixed-language. Consequently, text normalization processes (e.g., stemming, case-folding, and slang dictionary mapping) are crucial before training Word Embeddings to ensure accurate and consistent vector representations.

Overall, this WordCloud provides preliminary evidence that the Tokopedia review dataset is dominated by positive sentiment, focusing on product satisfaction and service speed. The predominance of positive words explains why the Naïve Bayes model performs well on the positive class but weakly recognizes neutral and negative classes. Thus, the overall success of the sentiment analysis system is strongly influenced by the effectiveness of Word Embedding in distinguishing dominant positive sentiment features from infrequent negative ones.

4. CONCLUSION

This study successfully demonstrates that the combination of the Naïve Bayes method with Word Embedding techniques (Word2Vec and FastText) is effective in analyzing customer sentiment toward Indonesian-language e-commerce products. Experimental results indicate that the FastText + Naïve Bayes model achieved the best performance, with an accuracy of 91.52%, outperforming the Word2Vec + Naïve Bayes model, which reached 87.92%. This improvement is attributed to FastText's ability to capture morphological variations and non-standard words commonly used in e-commerce customer reviews. Nonetheless, the model still faces challenges in recognizing neutral and negative sentiment classes due to data imbalance, resulting in a bias toward positive sentiment. Overall, Word Embedding–based approaches have been shown to enrich the semantic representation of text and enhance classification effectiveness compared to traditional methods such as TF-IDF. Therefore, this method can serve as an efficient, interpretable, and computationally lightweight solution for large-scale sentiment analysis. Future research is recommended to implement data balancing techniques and to combine this approach with deep learning algorithms, such as SVM or LSTM, to improve the model's capacity for understanding more complex and diverse sentiment contexts within the e-commerce domain.

REFERENCES

[1] A. R. Susanti and E. N. Ilahi, "Sentiment Analysis of User Reviews of E-commerce Applications: Case Study on the Shoppe Platform," *J. Soc. Sci.*, vol. 5, no. 4, pp. 983–988, 2024, doi: https://doi.org/10.46799/jss.v5i4.885.
[2] F. Aftab *et al.*, "A Comprehensive Survey on Sentiment Analysis Techniques," *Int. J. Technol.*, vol. 14, no. 6, pp. 291–319, 2023, doi: https://doi.org/10.14716/ijtech.v14i6.6632.
[3] A. H. I. Ahmad Azrir, P. Naveen, and S. C. Haw, "Sentiment Analysis using Machine Learning Models on Shopee Reviews," *J. Syst. Manag. Sci.*, vol. 14, no. 2, pp. 214–228, 2024, doi: https://doi.org/10.33168/jsms.2024.0213.
[4] F. A. Ramadhan, R. R. P. Ruslan, and A. Zahra, "Sentiment Analysis Of E-Commerce Product Reviews For Content Interaction Using Machine Learning," *Cakrawala Repos. IMWI*, vol. 6, no. 1, pp. 207–220, 2023, doi: https://doi.org/10.52851/cakrawala.v6i1.219.
[5] A. Daza, N. D. G. Rueda, M. S. A. Sánchez, W. F. R. Espíritu, and M. E. C. Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, p. 100267, 2024, doi: https://doi.org/10.1016/j.jjimei.2024.100267.
[6] N. M. Al Ghazali and Y. Sibaroni, "Sentiment Classification in E-Commerce Using Naïve Bayes and Combined Lexicon - N-Gram Features," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 2, pp. 1257–1271, 2025, doi: https://doi.org/10.29100/jipi.v10i2.6157.
[7] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci. (Ny).*, vol. 513, pp. 429–441, 2020, doi: https://doi.org/10.1016/j.ins.2019.11.004.
[8] 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: https://doi.org/10.1109/ACCESS.2022.3149482.
[9] M. Idris, A. Rifai, and K. D. Tania, "Sentiment Analysis of Tokopedia App Reviews using Machine Learning and Word Embeddings," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 1, pp. 210–219, 2025, doi: https://doi.org/10.33395/sinkron.v9i1.14278.
[10] C. C. P. Hapsari, W. Astuti, and M. D. Purbolaksono, "Naive Bayes Classifier and Word2Vec for Sentiment Analysis on Bahasa Indonesia Cosmetic Product Reviews," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*, 2021, pp. 22–27. doi: https://doi.org/10.1109/ICoDSA53588.2021.9617544.
[11] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artif. Intell. Rev.*, vol. 56, pp. 10345–10425, 2023, doi: https://doi.org/10.1007/s10462-023-10419-1.
[12] H. Abdelmotaleb, C. Mcneile, and M. Wojtyś, "A comparative study of word embedding techniques for classification of star ratings," *Expert Syst. Appl.*, vol. 297, no. A, p. 129037, 2025, doi: https://doi.org/10.1016/j.eswa.2025.129037.
[13] N. Tabassum *et al.*, "Semantic Analysis of Urdu English Tweets Empowered by Machine Learning," *Intell. Autom. Soft Comput.*, vol. 30, no. 1, pp. 175–186, 2021, doi: http://dx.doi.org/10.32604/iasc.2021.018998.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

[14] A. Chakravarthy, P. Desai, S. Deshmukh, S. Gawande, and I. Saha, "Hybrid Architecture for Sentiment Analysis using Deep Learning," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 1, pp. 735–738, 2018, doi: http://dx.doi.org/10.26483/ijarcs.v9i1.5388.
[15] A. T. Rizkya, R. Rianto, and A. I. Gufroni, "Implementation of the Naive Bayes Classifier for Sentiment Analysis of Shopee E-Commerce Application Review Data on the Google Play Store," *Int. J. Appl. Inf. Syst. Informatics*, vol. 1, no. 1, 2023, doi: https://doi.org/10.37058/jaisi.v1i1.8993.
[16] S. Xiao, H. Wang, Z. Ling, L. Wang, and Z. Tang, "Sentiment Analysis For Product Reviews Based on Deep Learning," in *Journal of Physics: Conference Series*, 2020, vol. 1651, p. 12103. doi: 10.1088/1742-6596/1651/1/012103.
[17] S. N. Fadhilah and F. S. Utomo, "Naïve Bayes Algorithm for Sentiment Analysis of Blibli. com Review on Google Play Store," *Sist. J. Sist. Inf.*, vol. 13, no. 2, pp. 831–840, 2024, doi: https://doi.org/10.32520/stmsi.v13i2.3887.
[18] M. A. Wahed, M. S. Alzboon, M. Alqaraleh, J. Ayman, M. Al-Batah, and A. F. Bader, "Automating web data collection: Challenges, solutions, and python-based strategies for effective web scraping," in *2024 7th International Conference on Internet Applications, Protocols, and Services (NETAPPS)*, 2024, pp. 1–6. doi: https://doi.org/10.1109/NETAPPS63333.2024.10823528.
[19] B. Bala and S. Behal, "A brief survey of data preprocessing in machine learning and deep learning techniques," in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2024, pp. 1755–1762. doi: https://doi.org/10.1109/I-SMAC61858.2024.10714767.
[20] D. A. A. Prakash, "Pre-processing techniques for preparing clean and high-quality data for diabetes prediction," *Int. J. Res. Publ. Rev.*, vol. 5, no. 2, pp. 458–465, 2024, doi: https://doi.org/10.55248/gengpi.5.0224.0412.
[21] N. G. Ramadhan, W. Maharani, and A. A. Gozali, "Chronic diseases prediction using machine learning with data preprocessing handling: A critical review," *IEEE Access*, vol. 12, pp. 80698–80730, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3406748.
[22] P. Koukaras and C. Tjortjis, "Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices," *AI*, vol. 6, no. 10, 2025, doi: 10.3390/ai6100257.
[23] L. Afuan, N. Hidayat, N. Nofiyati, and M. F. As' ad, "Sentiment Analysis of the Kampus Merdeka Program on Twitter Using Support Vector Machine and a Feature Extraction Comparison: TF-IDF vs. FastText," *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1738–1753, 2024, doi: https://doi.org/10.47738/jads.v5i4.436.
[24] E. Dotan, G. Jaschek, T. Pupko, and Y. Belinkov, "Effect of tokenization on transformers for biological sequences," *Bioinformatics*, vol. 40, no. 4, 2024, doi: https://doi.org/10.1093/bioinformatics/btae196.
[25] M. E. Samie, E. Bahmani, and N. Mozafari, "Analytical Comparison of Stop Word Recognition Methods in Persian Texts," *Int. J. Inf. Sci. Manag.*, vol. 23, no. 1, pp. 91–107, 2025, doi: https://doi.org/10.22034/ijism.2025.2017335.1322.
[26] V. P. Carolina, E. Utami, and A. Yaqin, "Exploring Stemming Techniques in Ambon Malay Languages: A Systematic Literature Review," *Jambura J. Informatics*, vol. 6, no. 1, pp. 1–13, 2024, doi: https://doi.org/10.37905/jji.v6i1.24954.
[27] B. A. Mustofa and W. L. Y. Saptomo, "Use of Natural Language Processing in Social Media Text Analysis," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 2, pp. 1235–1238, 2025, doi: https://doi.org/10.59934/jaiea.v4i2.875.
[28] K. Hadi and E. Utami, "Analysis of K-NN with the Integration of Bag of Words, TF-IDF, and N-Grams for Hate Speech Classification on Twitter," *JUITA J. Inform.*, vol. 12, no. 2, pp. 289–298, 2024, doi: https://doi.org/10.30595/juita.v12i2.23829.
[29] E. De Santis, A. Martino, F. Ronci, and A. Rizzi, "From bag-of-words to transformers: A comparative study for text classification in healthcare discussions in social media," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 9, no. 1, pp. 1063–1077, 2024, doi: https://doi.org/10.1109/TETCI.2024.3423444.
[30] M. Parmar and A. Tiwari, "Enhancing Text Classification Performance using Stacking Ensemble Method with TF-IDF Feature Extraction," in *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2024, pp. 166–174. doi: https://doi.org/10.1109/ICMCSI61536.2024.00031.
[31] T. S. Y. Win, "Authorship Identification System Using Word2Vec Word Embedding Model," in *2024 IEEE Conference on Computer Applications (ICCA)*, 2024, pp. 1–9. doi: https://doi.org/10.1109/ICCA62361.2024.10533018.
[32] B. P. Kamath, M. Geetha, U. D. Acharya, D. Singh, and A. Rao, "Comprehensive Analysis of Word Embedding Models and Design of Effective Feature Vector for Classification of Amazon Product Reviews," *IEEE Access*, vol. 13, pp. 25239–25255, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3536631.
[33] H. Mahmud, M. Hasan, F. R. Kabir, and M. Z. Aqib, "A Systematic Literature Review of Similarity Analysis Techniques for Bangla Text," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 10, pp. 3051–3058, 2024, doi: https://doi.org/10.5281/zenodo.14730649.
[34] L. A. Widjayanto and E. B. Setiawan, "Depression Detection using Convolutional Neural Networks and Bidirectional Long Short-Term Memory with BERT variations and FastText Methods," *J. Tek. Inform.*, vol. 6, no. 3, pp. 1555–1568, 2025, doi: https://doi.org/10.52436/1.jutif.2025.6.3.4874.
[35] N. A. Nasution, E. B. Nababan, and H. Mawengkang, "Comparing LSTM Algorithm with Word Embedding: FastText and Word2Vec in Bahasa Batak-English Translation," in *2024 12th International Conference on Information and Communication Technology (ICoICT)*, 2024, pp. 306–313. doi: https://doi.org/10.1109/ICoICT61617.2024.10698481.
[36] A. R. Mandasari, E. B. Nababan, and H. Mawengkang, "Long Short Term Memory (LSTM) Improvement Accuracy Using FastText and Glove for Language Translation Batak-Indonesian," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, pp. 1–5. doi: https://doi.org/10.1109/ICTIIA61827.2024.10761640.
[37] T. O. Atoyebi, R. F. Olanrewaju, N. V Blamah, and E. C. Uwazie, "Comparison of Multinomial Naive Bayes (MNB), Gaussian Naive Bayes (GNB) and Random Forest (RF) Algorithm in Malaria Disease Diagnosis," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, 2024, pp. 1–6. doi: https://doi.org/10.1109/SEB4SDG60871.2024.10630308.
[38] E. Mulyani, M. Sari, and F. Ishlakhuddin, "Multinomial Naïve Bayes Optimization with Information Gain for Library Book Classification," *TeknoIS J. Ilm. Teknol. Inf. dan Sains*, vol. 15, no. 1, pp. 104–110, 2025, doi: https://doi.org/10.36350/jbs.v15i1.300.
[39] M. K. Matarat, "Aspect-Based Sentiment Analysis in Thai Texts: A Comparative Study of Machine Learning and Neural Network Approaches," *Indochina Appl. Sci.*, vol. 14, no. 3, p. 262648, 2025, doi: https://doi.org/10.55674/ias.v14i3.262648.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2025−2034
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8879

[40] M. Hasan, T. Ahmed, M. R. Islam, and M. P. Uddin, "Leveraging textual information for social media news categorization and sentiment analysis," *PLoS One*, vol. 19, no. 7, p. e0307027, 2024, doi: https://doi.org/10.1371/journal.pone.0307027.
[41] J. Wang *et al.*, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 8052–8072, 2022, doi: https://doi.org/10.1109/TKDE.2022.3178128.
[42] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021, doi: https://doi.org/10.1145/3446776.
[43] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustain. Oper. Comput.*, vol. 3, pp. 238–248, 2022, doi: https://doi.org/10.1016/j.susoc.2022.03.001.
[44] J. C. Obi, "A comparative study of several classification metrics and their performances on data," *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 308–314, 2023, doi: https://doi.org/10.30574/wjaets.2023.8.1.0054.
[45] A. Geroldinger, L. Lusa, M. Nold, and G. Heinze, "Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures—a simulation study," *Diagnostic Progn. Res.*, vol. 7, no. 1, p. 9, 2023, doi: https://doi.org/10.1186/s41512-023-00146-0.
[46] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," *African J. Biomed. Res.*, vol. 27, no. 4S, pp. 4023–4031, 2024, doi: https://doi.org/10.53555/AJBR.v27i4S.4345.