

# JUITA\_Optimizing+Sentiment+Cl assification+of+E- Commerce\_Rev (1).docx

by Lee.jieun030101@gmail.com 1

---

**Submission date:** 12-Aug-2025 03:23AM (UTC-0500)

**Submission ID:** 2728611846

**File name:** JUITA\_Optimizing\_Sentiment\_Classification\_of\_E-Commerce\_Rev\_1\_.docx (171.62K)

**Word count:** 5361

**Character count:** 32392

# Optimizing Sentiment Classification of E-Commerce Product Reviews: A Comparative Study of Naïve Bayes and SVM with SMO

Riki<sup>1\*</sup>, Sonya Eliese Dameria<sup>2</sup>, Aditiya Hermawan<sup>3</sup>, Junaedi<sup>4</sup>, Yusuf Kurnia<sup>5</sup>  
<sup>1,2,3,4,5</sup>Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia  
\*corr-author: riki@ubd.ac.id

**Abstract** - The rapid growth of e-commerce has led to a surge in user-generated product reviews, making manual sentiment analysis impractical. This study explores automated sentiment classification using two machine learning algorithms: Naïve Bayes and Support Vector Machine (SVM) optimized with Sequential Minimal Optimization (SMO). The dataset comprises 2,000 Shopee product reviews labeled as positive, neutral, or negative. The study focuses on assessing the effectiveness of these algorithms in classifying product reviews, especially in the diverse and high-volume data typical of e-commerce environments. Empirical evaluation shows that Naïve Bayes achieves 68% accuracy, while SVM with SMO attains 79%. Additionally, the study evaluates other important performance metrics, such as precision, recall, and F1-score, showing that SVM with SMO outperforms Naïve Bayes in accurately classifying product reviews. These findings highlight the superior capability of SVM with SMO in handling complex sentiment data, thereby offering a more robust foundation for automated review classification. This research provides insights into selecting suitable classifiers for improving customer experience and strategic decision-making in digital commerce.

**Keywords:** E-commerce; Naïve Bayes Classifier; Review Classification; Sequential Minimal Optimization (SMO); Support Vector Machine (SVM).

## I. INTRODUCTION

Advancements in the digital age continue to influence and transform people's lifestyles. Digital is important to ensure that everyone has digital information and communication skills [1], [2]. The reason is, that most people experience significant changes in their daily activities [3]. One of these changes occurs in people's behavior who prefer online shopping. Online shopping is an activity that is popular and chosen by some people, and its popularity is increasing due to the ongoing Covid-19 pandemic. The process is easy and practical, just use a device connected to the internet.

Based on the research in the article titled 'Challenges of Indonesian Industrialization and Trade Problems in the Industrial Era 4.0 for the Indonesian Economy and Government Policy Breakthrough' published by Eduvest Journal, it was reported that in 2023, the number of internet users in Indonesia reached 215 million people, or about 79.34% of the population [4]. According to the Indonesian Telecommunication Statistics 2022, social media remains the most visited digital platform in Indonesia, with 74.02% of internet users engaging with it during 2021-2022. Additionally, 74.90% of the population accessed the internet to gather information or news, and 69.79% used it for entertainment purposes [5]. While 16.51% of users utilized the internet for online shopping, online chatting applications also played a significant role, being widely used for both personal communication and business transactions. These trends highlight the diverse and growing ways in which Indonesians are integrating the internet into their daily lives.

Online shopping services or websites via the internet are called e-commerce. E-commerce functions as a channel through which consumers transact goods and services electronically via the Internet. E-commerce or Electronic Commerce refers to a marketing system that utilizes electronic media for various functions, including distribution, sales, purchasing, marketing, and service of products carried out in the system. Distribution systems are crucial in establishing a successful business. In the e-commerce sector, business owners must recognize the significance of these systems in their marketing strategies, as they serve as a link between producers and consumers. Optimizing logistics services in the e-commerce industry improves the efficiency of product distribution, leading to higher customer satisfaction and business growth [6]. Further demonstrates how intermediaries play a key role in expanding the accessibility of e-commerce platforms and broadening customer reach [7]. Additionally, the importance of multi-

channel distribution strategies to understand and address the diverse behaviors of consumers [8]. These strategies not only influence how products are distributed but also offer valuable market insights and feedback to producers.

E-commerce platforms offer a product review feature that helps determine if user feedback is positive or negative. This feedback is incredibly valuable to companies and manufacturers in shaping business strategies. User-generated content, such as online reviews, can have a significant impact on purchase intentions in the e-commerce sector [9]. However, analyzing large volumes of reviews or ratings manually can be inefficient and costly, especially when the reviews are in irregular formats that are hard to process [10]. The challenges of conducting large-scale sentiment analysis on e-commerce platforms are significant. Reading through massive amounts of reviews one by one is impractical, making automated sentiment analysis tools essential for efficiently handling large volumes of feedback [11]. Therefore, a smarter approach is needed to identify new reviews and make predictions about product dissatisfaction. Opinion Mining is the process of extracting specific opinions and sentiments related to a product. Sentiment analysis, on the other hand, involves understanding and interpreting the opinions, evaluations, and attitudes expressed in text, enabling automatic data processing to gather useful information. This automated approach not only enhances efficiency but also provides deeper insights into consumer behavior [12].

User comments and reviews on e-commerce platforms are crucial in shaping seller ratings and boosting overall customer satisfaction with online shopping experiences. The ability for users to share their feedback on products has a significant impact on how other potential buyers perceive them, influencing their purchasing decisions and overall satisfaction [13], [14]. Given the large volume of user-generated content, effective text classification becomes vital for managing and analyzing this data. Feature selection, an essential step in the text classification process, helps reduce the complexity of data, removing irrelevant features, and improves the accuracy of classification algorithms [15]. By optimizing feature selection, the performance of sentiment analysis and other text-based predictive models can be enhanced, ensuring that the most relevant aspects of user feedback are effectively highlighted and used to improve e-commerce strategies.

Support Vector Machine (SVM) is a powerful supervised learning algorithm widely used for classification tasks, and it can also address regression problems. SVM is particularly popular when high accuracy is required, and its performance is often improved when combined with optimization techniques like Sequential Minimal Optimization (SMO). SMO is designed to solve the quadratic programming (QP) problem that occurs during SVM training, making the process faster and more scalable, especially when dealing with large datasets. This combination has been effectively used in various studies to enhance the efficiency and accuracy of SVM models.

Recent studies have demonstrated the effectiveness of SMO in overcoming the typically slow and complex nature of SVM training. Studies show that SMO not only alleviates the computational burden but also enhances SVM's overall performance by efficiently managing the optimization process, particularly in large-scale problems [16]. It has also been found that combining SVM with SMO significantly boosts classification accuracy by optimizing the attribute weights used during the classification process. This makes the combination ideal for tasks like sentiment analysis and fault detection [17], [18], [19]. In summary, while SVM is a strong classifier on its own, its integration with SMO greatly enhances performance, making it the preferred choice for high-stakes classification tasks that require both accuracy and computational efficiency.

Recent studies have shown that the Naïve Bayes algorithm is highly efficient for quickly classifying large datasets, particularly in sentiment analysis, and when used alongside Support Vector Machine (SVM), it can effectively analyze sentiment in app reviews, accurately predicting whether the sentiment is positive, neutral, or negative [20]. Another significant study applied Naïve Bayes to predict sentiment in online product reviews, achieving strong classification results. However, it was noted that relying solely on Naïve Bayes without integrating other methods, such as SVM, could limit the model's robustness. By combining Naïve Bayes with complementary algorithms like SVM, the model can benefit from the strengths of both methods, leading to more accurate predictions [21], [22].

Comparative studies on sentiment analysis of e-wallet reviews found that Support Vector Machine (SVM) achieved an accuracy of 91.30%, while Naïve Bayes obtained a slightly higher accuracy of 93.10% when Particle Swarm Optimization (PSO)-based feature selection was applied [23]. Research on classification methods for sentiment analysis of smartphone issues, both with and without feature selection, revealed that SVM, particularly when enhanced with Query Expansion Ranking (QER) feature selection, outperformed Naïve Bayes in terms of accuracy [24]. Further investigation into the impact of feature selection on sentiment analysis demonstrated that SVM with feature selection notably improved accuracy compared to Naïve Bayes, achieving a true positive rate of 86.50% for positive sentiments

and 89.10% for negative sentiment [47]. Collectively, these findings highlight the superior performance of SVM over Naive Bayes, particularly when feature selection techniques are employed in sentiment analysis tasks.

The strong performance of the Support Vector Machine (SVM) and Naive Bayes Classifier algorithms in previous studies inspired the authors to explore optimization techniques, particularly sequential minimal optimization (SMO). Their goal was to compare the effectiveness of these two algorithms in a practical manner, which led to the creation of a web application designed for sentiment analysis.

## II. METHOD

Fig. 1 explains the stages of collecting training data to the stages of calculating methods and algorithms.

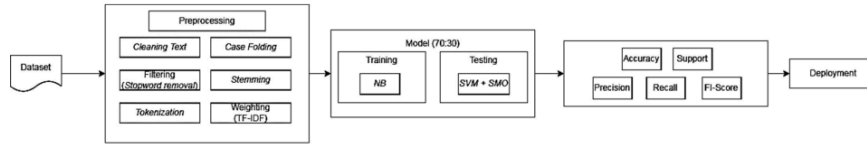


Fig. 1 Research Methods Diagram

### A. Data Retrieval

This training data collection stage is necessary so that the machine learning model can process and learn emotional phrases, allowing the machine to classify the review data based on its analysis and learning, ultimately allowing it to determine the most appropriate classification for Shopee product reviews.

The training data collected is primary. This training data is obtained by extracting Shopee product review text and associated star ratings, which are then labeled positive, negative, or neutral using the Shopee API. The data collection process uses tools that we created, namely retrieving product links from Shopee using Python with the Anaconda Navigator Application.

The dataset used in this study consists of 2,000 product reviews collected from the Shopee platform, which includes textual reviews and corresponding star ratings provided by users. The reviews were automatically retrieved through an API call to the Shopee platform, ensuring that the data collection process was efficient and consistent.

### B. Data Labelling

At this stage, the data is labeled positive, neutral, or negative using the tools used for data collection. Based on the results of the tool processing, 2,000 reviews were collected, each of which was given a 1 to 5-star rating, which was classified as follows: Good (4 to 5 stars), Neutral (3 Stars), and Bad (1 to 2 stars). The results for each category are as follows: 700 positive reviews, 700 negative reviews, and 600 neutral reviews. This distribution reflects a common sentiment distribution in e-commerce reviews. The training data obtained produces the information shown in Table I.

TABLE I  
TRAINING DATA

Parameter	Mark
Rating	Star ratings on Shopee product reviews will be used as labels
Overview	Product Review Text

Once the necessary data has been obtained, the review data available on Shopee e-commerce can be directly processed to obtain qualification and analysis results. This will generate a process that can be used to manage the processed review data.

### C. Text Preprocessing

From the existing review data on Shopee e-commerce, data validation has been carried out at the previous stage. Next, further processing will be carried out to obtain clean data. This stage is carried out to select words that will be used as an index during the data modeling process. The preprocessing process, as illustrated in Fig. 2, involves the use of self-developed tools along with data capture tools.

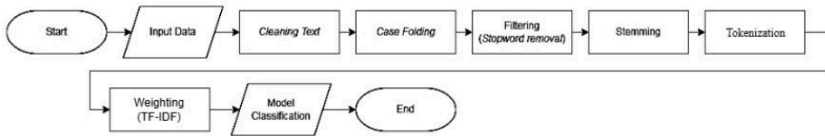


Fig. 2 Preprocessing Process

#### D. Naïve Bayes Classifier Computational Concepts

Calculating the Naïve Bayes Classifier involves using probability theory, specifically Bayes' Theorem, to classify data. This algorithm is considered 'naive' because it assumes all data features are independent. The Naïve Bayes algorithm can perform much better in complex real conditions than expected [26], [27].

Naïve Bayes is an efficient and straightforward classification technique that predicts the likelihood of an outcome based on a given set of conditions using Bayes' theorem [28]. Application of the Naïve Bayes classification method by considering two probabilities A and B, which can be correlated with  $P(H)$  and  $P(X)$  with conditional probabilities  $P(H|X)$  and  $P(X|H)$ . The relationship between these probabilities is mathematically expressed in Equation (1):

$$p(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)} \quad (1)$$

Information

- $H$  : Hypothesis data is a specific class.
- $X$  : Data with classes that are still unknown.
- $p(H|X)$  : Probability of hypothesis based on conditions.
- $p(H)$  : Hypothesis probability.
- $p(X|B)$  : Probability based on hypothetical conditions.
- $p(X)$  : Probability of X occurring.

To calculate the Naïve Bayes Classifier, a series of systematic steps are followed. The process begins with the collection and preparation of the training data, which involves assembling data samples that include both the relevant features and their corresponding class labels. Once the dataset is ready, the next step is to compute the prior probabilities for each class these represent the overall likelihood of each class appearing within the dataset independent of feature values. Following this, conditional probabilities are calculated for each feature given a particular class. These conditional feature probabilities quantify the likelihood of observing specific feature values assuming the instance belongs to a given class. With both prior and conditional probabilities established, Bayes' Theorem is then applied to calculate posterior probabilities for each class, given the features of the test data. These posterior probabilities represent the probability of each class occurring, conditional on the observed feature values. Finally, the classification phase assigns the test data to the class with the highest posterior probability. In this way, the Naïve Bayes Classifier utilizes probabilistic inference to predict the most likely class for unseen data based on the learned statistical patterns.

#### E. Support Vector Machine Computing Concepts using Sequential Minimal Optimization

Support Vector Machine (SVM) is a method widely explored within the field of machine learning. Developed by Guyon, Boser, and Vapnik, SVM was first introduced in 1992 at the annual Symposium on Computational Learning Theory. The SVM computational process includes several key steps, such as selecting an appropriate kernel, performing optimization, and identifying the hyperplane that effectively separates the data classes.

The Sequential Minimal Optimization (SMO) algorithm operates by addressing a dual-form optimization problem that focuses exclusively on the  $\alpha$  values, distinguishing it from other algorithms. It employs analytical quadratic programming within its inner loop and resolves the challenge of processing two data points during each iteration. This completes the process of finding the best solution [29]. The workflow of the SVM combined with SMO is illustrated

in Fig. 3, which outlines the key steps involved in the process, from kernel selection to optimization and the final test and prediction stages.



Fig. 3 SVM + SMO Workflow

Based on Fig. 3, the implementation of the Support Vector Machine (SVM) algorithm involves a series of methodical steps. The first step is kernel selection, where an appropriate kernel function such as Linear, Polynomial, or Radial Basis Function (RBF) is chosen depending on the nature of the classification problem. Kernels are instrumental in transforming the original feature space into a higher-dimensional space to enable linear separability of non-linearly separable data. Following kernel selection, the SVM training phase commences. During this stage, the algorithm is trained using the selected kernel and labeled training data to construct a model capable of identifying hyperplanes that effectively separate different classes. A hyperplane in this context acts as the decision boundary distinguishing data categories. Subsequently, the optimization phase is carried out, wherein the algorithm identifies the optimal hyperplane by maximizing the margin the shortest distance between the data points and the hyperplane. This stage involves solving a quadratic optimization problem to derive the optimal weight vector that ensures maximal separation between the classes. Finally, in the testing and prediction phase, the trained SVM model is evaluated using previously unseen test data. The model utilizes the learned hyperplane to classify the new instances, thereby assessing its generalization performance and predictive accuracy [30].

### III. RESULT AND DISCUSSION

#### A. Data Understanding

The training data collected is primary. This training data is obtained by taking Shopee product review texts and their stars which will be labeled positive, negative and neutral using the Shopee API. This data collection uses tools that were created ourselves by taking product links from Shopee using Python language with the Anaconda Navigator Application. The following is the flow of the data collection process shown in Fig. 4 below :

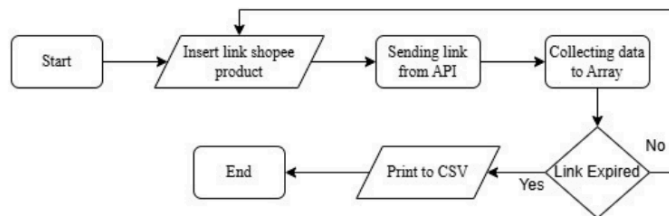


Fig. 4 Data Collecting Workflow

From the results of the processing of the tools used, 2000 reviews were collected, each rating consisting of 1 to 5 stars, which will be classified using the following parameters:

- i. Good (4-5 stars)
- ii. Neutral (3 Star)
- iii. Bad (1-2 Stars)

Based on the results of data collection, as shown in Table II, a total of 700 reviews were labeled as positive, 700 as negative, and 600 as neutral. The examples in Table II illustrate instances of product review texts and their corresponding ratings, where Good corresponds to positive sentiment.

TABLE II  
PRODUCT REVIEW DATA ON SHOPEE

Text	Ratings
Barangnya bagus, pas... terima kasih sis (The item is good, fits... thank you sis 😊😊😊)	Good
Lumayan sih, bekerja dengan baik (Not bad, works well)	Good

### B. Evaluation

In the evaluation phase, the performance of the Naïve Bayes and SVM SMO algorithms, as compared using metrics such as accuracy, precision, recall, and support, derived from confusion matrix. The outcomes of this comparison are presented in Table III and Table IV. Table III provides the confusion matrix results for the Naïve Bayes algorithm, whereas Table IV presents the corresponding results for the SVM SMO algorithm.

TABLE III  
CONFUSION MATRIX NAÏVE BAYES RESULT

	Precision	Recall	F1-Score	Support
Bad	1.00	0.05	0.09	64
Good	68%	1.00	0.81	202
Netral	0.00	0.00	0.00	34
Accuracy	-	-	0.68	300
Macro Avg	0.56	0.35	0.30	300
Weight Avg	0.67	0.68	0.56	300

The evaluation results in Table III indicate that the Naïve Bayes algorithm achieved an overall accuracy of 68%. At the class-specific level, the classifier performs very well in identifying the "Good" class, as evidenced by a precision of 1.00 and recall of 0.81, implying that most positive instances were correctly detected, albeit with some false negatives. However, for the "Bad" class, recall is notably low at 0.05, indicating poor sensitivity in correctly identifying negative reviews, despite a perfect precision of 1.00, which suggests minimal false positives for this class. The "Neutral" class exhibits the weakest performance, with precision, recall, and F1-score approaching zero, highlighting the algorithm's failure to effectively classify neutral reviews. Overall, while Naïve Bayes benefits from simplicity and computational efficiency, its performance is suboptimal, particularly in handling ambiguous or neutral data and in recognizing minority classes, as reflected in the low macro-average recall and F1-score values.

TABLE IV  
CONFUSION MATRIX SVM+SMO RESULT

	Precision	Recall	F1-Score	Support
Bad	0.66	0.72	0.69	64
Good	0.83	0.95	0.89	202
Netral	1.00	0.03	0.06	34
Accuracy	-	-	0.79	300
Macro Avg	0.83	0.56	0.54	300
Weight Avg	0.82	0.79	0.75	300

The analysis of Table IV shows that the Support Vector Machine optimized with Sequential Minimal Optimization (SVM+SMO) significantly outperforms Naïve Bayes, achieving an overall accuracy of 79%. For the "Good" class, the model demonstrates excellent performance with precision of 0.83 and recall of 0.95, indicating a high true positive rate with minimal misclassifications. The "Bad" class also shows improved results, with precision at 0.66 and recall at 0.72, reflecting a more balanced detection capability and fewer false positives and negatives compared to Naïve Bayes. Nevertheless, the "Neutral" class remains challenging, as indicated by a recall of only 0.03, despite perfect precision (1.00), suggesting the model is highly conservative in assigning this class, labeling only very certain instances as neutral. Despite this limitation, the higher macro-average precision and weighted average F1-score confirm that SVM+SMO is more robust and effective in managing imbalanced and complex feature sets, making it better suited for sentiment classification in e-commerce product reviews.

The results of this study demonstrate a clear performance advantage of the Support Vector Machine optimized with Sequential Minimal Optimization (SVM+SMO) over the Naive Bayes classifier in the sentiment analysis of e-commerce product reviews. With an overall accuracy of 79% compared to 68% for Naive Bayes, SVM+SMO exhibits superior capability in handling complex and high-dimensional data, effectively capturing the nuanced differences in user sentiment. The improved recall and F1-scores for the “Good” and “Bad” classes further highlight SVM+SMO’s robustness in identifying strongly polarized sentiments. Nevertheless, both classifiers showed limited effectiveness in accurately classifying the “Neutral” class, reflecting the inherent challenge of distinguishing ambiguous or context-dependent sentiments within short or subtle user reviews. The failure to classify neutral reviews accurately may be due to the short length of reviews, the use of ambiguous words, or shifting context within the sentence, which makes it difficult for the model to distinguish between neutral and slightly positive or negative sentiments.

These findings are consistent with a growing body of literature emphasizing the efficacy of SVM-based models for text classification tasks, particularly when enhanced with feature selection or optimization algorithms, which have been shown to outperform probabilistic methods like Naive Bayes [23], [25]. Similar improvements in classification performance have also been observed in recent studies employing optimization strategies and hybrid deep learning architectures for sentiment analysis, demonstrating the benefits of model enhancement and balanced data handling [31], [32]. Extended experiments utilizing advanced embedding techniques such as Word2Vec and BERT further corroborate contemporary trends that leverage deep contextual embeddings to enhance classification accuracy.

Despite these promising results, this study faces several limitations that suggest directions for future research. The relatively modest dataset size, comprising 2000 reviews, may constrain the model’s ability to generalize across the diverse range of product categories and linguistic variations found in broader e-commerce settings. The small dataset size also poses a risk of overfitting, where the model performs well on the training data but fails to generalize to unseen data. This issue could be addressed by increasing the dataset size or using techniques like cross-validation. The evident class imbalance, especially the underrepresentation of neutral reviews, likely contributed to suboptimal classification performance for this category, underscoring the need for more sophisticated balancing strategies such as synthetic minority oversampling or cost-sensitive learning. Additionally, the exclusive reliance on textual data precludes the incorporation of potentially informative multimodal inputs, including images or user metadata, which could further refine sentiment prediction.

Moreover, the scope of hyperparameter tuning, while methodical, was limited by computational resources, leaving room for further exploration that could uncover more optimal configurations. The error analysis revealed difficulties in detecting subtle linguistic phenomena such as sarcasm or mixed sentiments, highlighting a persistent challenge for automated sentiment classifiers. These challenges are particularly pronounced in neutral reviews, where sentiment may be ambiguous or context-dependent, making it harder for the model to classify with high confidence. Addressing this gap may require models that integrate pragmatic, contextual, and discourse-level features to better interpret complex user expressions.

The practical implications of this study for the e-commerce industry are significant. The findings can be implemented in recommendation systems, where sentiment analysis can help refine product suggestions based on customer sentiment, enhancing user experience. Additionally, automated review moderation can benefit from these results by flagging potentially misleading or irrelevant reviews, streamlining content management on e-commerce platforms.

Overall, this study substantiates the superiority of SVM+SMO, particularly when combined with advanced embeddings, in classifying e-commerce sentiment data. Yet, it also underscores the ongoing challenges in handling ambiguous sentiments and the necessity for larger, more diverse datasets and multimodal approaches to enhance model robustness and applicability. This study’s novelty lies in the comparative analysis of Naive Bayes and SVM+SMO for e-commerce sentiment analysis, providing clear insights into the performance of these models in a real-world setting and offering practical recommendations for improving automated review classification.

#### IV. CONCLUSION

The comparative analysis presented in this study highlights significant differences in the performance of the Naive Bayes Classifier and Support Vector Machine (SVM) methods when applied to sentiment analysis of product reviews. Nevertheless, the SVM, particularly when enhanced with Sequential Minimal Optimization (SMO), achieved superior accuracy, achieving 79% compared to 68% for the Naive Bayes Classifier. This performance advantage is mainly due

to SVM's ability to handle non-linear relationships and optimize classification boundaries, making it more suitable for the complex and diverse datasets commonly encountered in e-commerce platforms. The findings indicate that Naive Bayes is advantageous for its simplicity and computational efficiency, but it may struggle with more intricate data structures or contexts where high classification precision is required. In contrast, SVM with SMO provides a more robust and scalable solution for sentiment classification tasks, particularly in extracting actionable insights from consumer reviews. This study provides a thorough comparison between the two algorithms and affirms the suitability of SVM with SMO in dealing with high-dimensional and imbalanced textual data. This research underscores the importance of selecting the right machine learning algorithms for improving product classification tasks, offering valuable insights for both researchers and practitioners focused on enhancing the accuracy and reliability of sentiment analysis models in e-commerce settings. Future research should consider expanding the dataset to encompass more diverse product types, applying advanced feature engineering techniques, incorporating transformer-based models such as BERT to capture contextual dependencies, and exploring hybrid or ensemble approaches to better address class imbalance and improve the detection of neutral or ambiguous sentiments.

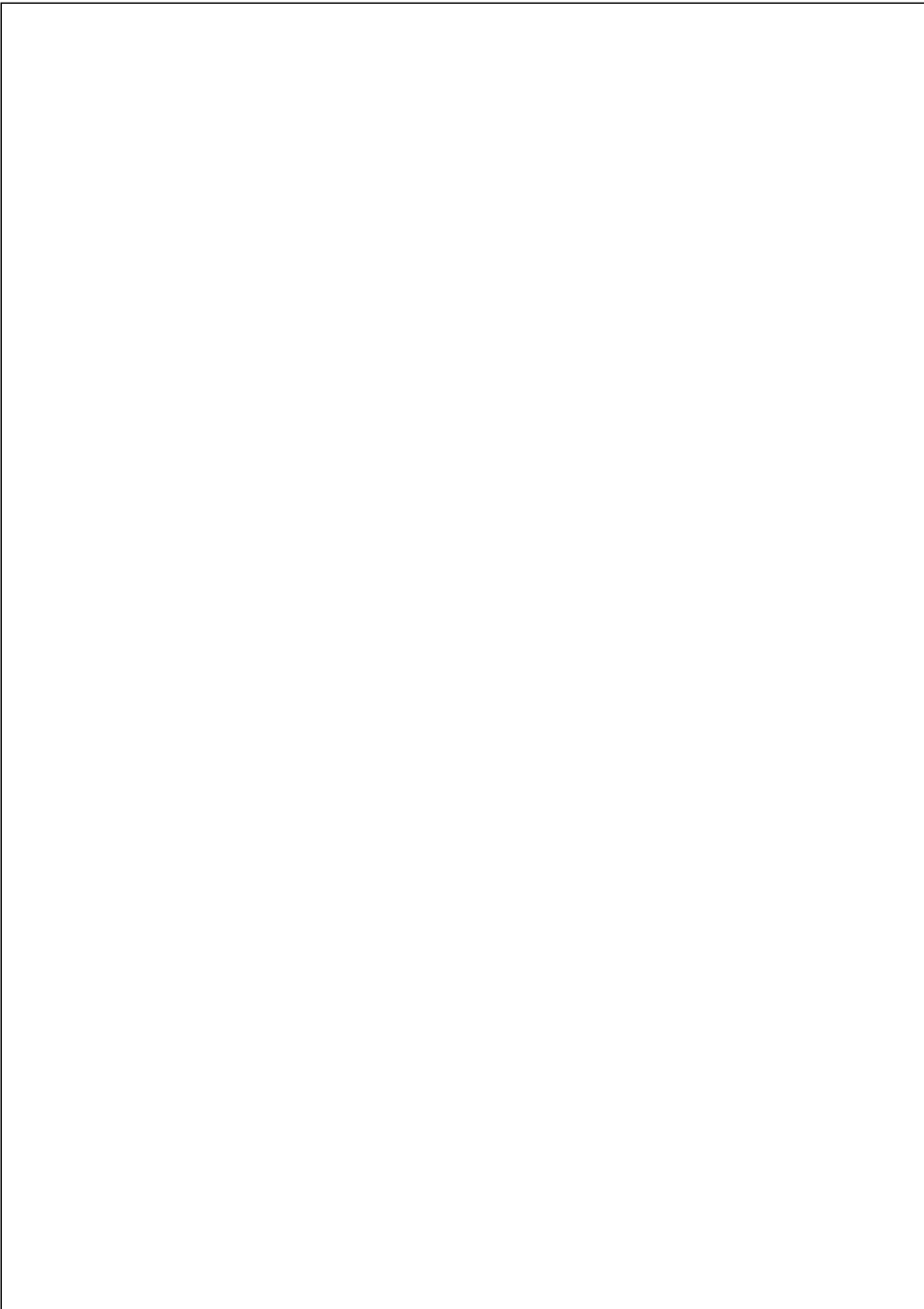
#### ACKNOWLEDGEMENT

This research was supported by funding from Buddhi Dharma University, and we would like to express our gratitude for their financial assistance and continued support throughout this study.

#### REFERENCES

- [1] F. Tamimi and S. Munawaroh, "Teknologi Sebagai Kegiatan Manusia Dalam Era Modern Kehidupan Masyarakat," vol. 2, no. 3, pp. 66–74, 2024, doi: <https://doi.org/10.61132/saturnus.v2i3.157>.
- [2] A. Sacavém, A. de Bem Machado, J. R. dos Santos, A. Palma-Moreira, H. Belchior-Rocha, and M. Au-Yong-Oliveira, "Leading in the Digital Age: The Role of Leadership in Organizational Digital Transformation," *Adm. Sci.*, vol. 15, no. 2, pp. 1–21, 2025, doi: [10.3390/admsci15020043](https://doi.org/10.3390/admsci15020043).
- [3] N. Shaw, B. Eschenbrenner, and D. Baier, "Online shopping continuance after COVID-19: A comparison of Canada, Germany and the United States," *J. Retail. Consum. Serv.*, vol. 69, p. 103100, 2022, doi: <https://doi.org/10.1016/j.jretconser.2022.103100>.
- [4] Y. N. Kusumawati, S. Indarto, B. Education, and Y. State, "Challenges Of Indonesian Industrialization And Trade Problems In The Industrial Era 4.0 For The Indonesian Economy And Government Policy," vol. 4, no. 07, pp. 6037–6046, 2024, doi: <https://doi.org/10.59188/eduvest.v4i7.1575>.
- [5] T. Sutarsih and K. Maharani, *Telecommunication Statistics in Indonesia*. BPS-Statistics Indonesia, 2022. [Online]. Available: <https://www.bps.go.id/id/publication/2023/08/31/131385d0253c6aae7c7a59fa/statistik-telekomunikasi-indonesia-2022.html>
- [6] C. B. Sivaparhipan, H. Alabdeli, P. Harshitha, Y. Nagendar, and G. G. Kulkarni, "An Optimization Model for Logistics Services in E-Commerce Market," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 2024, pp. 1–5. doi: [10.1109/ICDSIS61070.2024.10594347](https://doi.org/10.1109/ICDSIS61070.2024.10594347).
- [7] C. V. Kalyan, V. V. S. N. Akhila Sree Rajeswari, Y. S. D. Phaneendra, and S. R. Kishan, "Lokart: Empowering Local Artisans through Mobile E-commerce," in *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, 2024, pp. 1–6. doi: [10.1109/ICETCS61022.2024.10543943](https://doi.org/10.1109/ICETCS61022.2024.10543943).
- [8] Bijith Marakarkandy, "Enhancing Multi-Channel Consumer Behavior Analysis: A Data-Driven Approach using the Optimized Apriori Algorithm," *J. Electr. Syst.*, vol. 20, no. 2s, pp. 700–708, 2024, doi: [10.52783/jes.1536](https://doi.org/10.52783/jes.1536).
- [9] A. Prasetyo, N. A. Witasryah, and Indrawati, "The effect of e-WOM on purchase intention in e-commerce in Indonesia through the expansion of the information adoption model," *Int. J. Data Netw. Sci.*, vol. 8, no. 3, pp. 1959–1968, 2024, doi: [10.52677/ijdns.2024.1.017](https://doi.org/10.52677/ijdns.2024.1.017).
- [10] H. Huang, A. Asemi, and M. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," *IEEE Access*, vol. 11, pp. 90367–90382, 2023, doi: [10.1109/ACCESS.2023.3307308](https://doi.org/10.1109/ACCESS.2023.3307308).
- [11] X. Ma, Y. Li, and M. Asif, "E-Commerce Review Sentiment Analysis and Purchase Intention Prediction Based on Deep Learning Technology," *J. Organ. End User Comput.*, vol. 36, no. 1, pp. 1–29, 2024, doi: [10.4018/JOEUC.335122](https://doi.org/10.4018/JOEUC.335122).
- [12] S. Gupta, P. S. Kushwaha, U. Badhera, P. Chatterjee, and E. D. R. S. Gonzalez, "Identification of benefits, challenges, and pathways in E-commerce industries: An integrated two-phase decision-making model," *Sustain. Oper. Comput.*, vol. 4, no. 4, pp. 200–218, 2023, doi: [10.1016/j.susoc.2023.08.005](https://doi.org/10.1016/j.susoc.2023.08.005).

- [13] M. A. Salam and M. Ali, "Optimizing Extreme Learning Machine using GWO Algorithm for Sentiment Analysis," *Int. J. Comput. Appl.*, vol. 176, no. 38, pp. 22–28, 2020, doi: 10.5120/ijca2020920450.
- [14] P. H. Prastyo, R. Hidayat, and I. Ardiyanto, "Enhancing sentiment classification performance using hybrid Query Expansion Ranking and Binary Particle Swarm Optimization with Adaptive Inertia Weights," *ICT Express*, vol. 8, no. 2, pp. 189–197, 2022, doi: <https://doi.org/10.1016/j.ict.2021.04.009>.
- [15] P. Kumar and Monika, "Improved Tweet Sentiment Analysis by Features Weight Optimize by GWO and Classify by XG-Boost," *Advances in Intelligent Systems and Computing*, pp. 607–614, 2023, doi: 10.1007/978-981-19-5443-6\_46.
- [16] V. R. Modhugu and S. Ponnusamy, "Comparative Analysis of Machine Learning Algorithms for Liver Disease Prediction: SVM, Logistic Regression, and Decision Tree," *Asian J. Res. Comput. Sci.*, vol. 17, pp. 188–201, May 2024, doi: 10.9734/ajrcos/2024/v17i6467.
- [17] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Inf.*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.
- [18] H. Wang, G. Li, and Z. Wang, "Fast SVM classifier for large-scale classification problems," *Inf. Sci. (Ny.)*, vol. 642, p. 119136, 2023, doi: <https://doi.org/10.1016/j.ins.2023.119136>.
- [19] Y. Wan, Z. Wang, and T.-Y. Lee, "Incorporating support vector machine with sequential minimal optimization to identify anticancer peptides," *BMC Bioinformatics*, vol. 22, no. 1, p. 286, 2021, doi: 10.1186/s12859-021-03965-4.
- [20] M. Madjid, D. Eka Ratnawati, and B. Rahayudi, "Sentiment Analysis on App Reviews Using Support Vector Machine and Naïve Bayes Classification," *Sinkron*, vol. 8, no. 1, pp. 556–562, Feb. 2023, doi: 10.33395/sinkron.v8i1.12161.
- [21] A. Firdaus, A. Yudhana, and I. Riadi, "Prediction of Indonesian Presidential Election Results using Sentiment Analysis with Naïve Bayes Method," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, pp. 41–50, Jan. 2024, doi: 10.30865/mib.v8i1.7007.
- [22] Fadlila Nurwanda, Winita Sulandari, Yuliana Susanti, and Zakya Reyhana, "Comparative Analysis Of Performance Levels Of Svm And Naïve Bayes Algorithm For Lifestyle Classification On Twitter Social Media," *Int. Conf. Digit. Adv. Tour. Manag. Technol.*, vol. 1, no. 1, pp. 215–230, 2023, doi: 10.56910/ictmt.v1i1.65.
- [23] D. A. Putri, D. A. Kristiyanti, E. Indrayuni, A. Nurhadi, and D. R. Hadinata, "Comparison of Naïve Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, p. 012085, 2020, doi: 10.1088/1742-6596/1641/1/012085.
- [24] G. Novilia, R. Adriman, and T. F. Abidin, "Comparison of Classification Methods Using Feature Selection for Smartphone Sentiment Analysis," in *2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, 2022, pp. 44–48, doi: 10.1109/ICEECIT55908.2022.10030703.
- [25] P. Pongthanoon and W. Songpan, "Feature Selection and Reduction Based on SMOTE and Information Gain for Sentiment Mining," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, vol. 8, pp. 109–114, May 2020, doi: 10.1109/ICCCS49078.2020.9118467.
- [26] H. Rahadian, A. Yudhana, and others, "Naïve Bayes Method Monitoring Macro Nutrition and Soil Moisture Using Naïve Bayes Method based on Internet of Things," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 2, pp. 251–259, 2023, doi: 10.12928/biste.v5i2.6305.
- [27] F. Ramadhani, Al-Khowarizmi, and I. P. Sari, "Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain Ratio and Adaboost," *2021 International Conference on Computer Science and Engineering (IC2SE)*, pp. 1–5, Nov. 2021, doi: <https://doi.org/10.1109/ic2se52832.2021.9792027>.
- [28] H. Taherdoost, "Machine Learning Algorithms: Features and Applications," 2022, pp. 938–960, doi: 10.4018/978-1-7998-9220-5.ch054.
- [29] H. Noviyanto and B. Mukti, "Period Study Accuracy Prediction using Sequential Minimal Optimization Algorithm," *Sinkron*, vol. 5, no. 1, pp. 164–169, Oct. 2020, doi: 10.33395/sinkron.v5i1.10621.
- [30] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020, doi: 10.1089/big.2020.0062.
- [31] Y. N. Kunang and W. P. Mentari, "Analysis of the Impact of Vectorization Methods on Machine Learning-Based Sentiment Analysis of Tweets Regarding Readiness for Offline Learning," *JUITA : Jurnal Informatika*, vol. 11, no. 2, p. 271, Nov. 2023, doi: <https://doi.org/10.30595/juita.v11i2.17568>.
- [32] S. Riyadi, M. D. Mubarak, C. Damarjati, and A. J. Ishak, "Improving Sentiment Analysis Accuracy Using CRNN on Imbalanced Data: a Case Study of Indonesian National Football Coach," *JUITA: Jurnal Informatika*, vol. 12, no. 2, p. 159, Nov. 2024, doi: <https://doi.org/10.30595/juita.v12i2.21847>.



# JUITA\_Optimizing+Sentiment+Classification+of+E-Commerce\_Rev (1).docx

## ORIGINALITY REPORT

14%

SIMILARITY INDEX

11%

INTERNET SOURCES

10%

PUBLICATIONS

2%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://www.mdpi.com">www.mdpi.com</a> Internet Source	1%
2	<a href="http://repository.mercubuana.ac.id">repository.mercubuana.ac.id</a> Internet Source	1%
3	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	1%
4	<a href="http://repository.uin-suska.ac.id">repository.uin-suska.ac.id</a> Internet Source	1%
5	<a href="http://eduvest.greenvest.co.id">eduvest.greenvest.co.id</a> Internet Source	1%
6	Zahra Hashemi, Maysam Gholampour, Ming Chang Wu, Ting Ya Liu, Chuan Yi Liang, Chi-Chuan Wang. "A physics-informed neural networks modeling with coupled fluid flow and heat transfer – Revisit of natural convection in cavity", International Communications in Heat and Mass Transfer, 2024 Publication	<1%
7	<a href="http://journalofbigdata.springeropen.com">journalofbigdata.springeropen.com</a> Internet Source	<1%
8	<a href="http://www.isteonline.in">www.isteonline.in</a> Internet Source	<1%
9	<a href="http://www.jatit.org">www.jatit.org</a> Internet Source	<1%

10 Sara Raquel Araujo Leal, Franciely Alves De Souza, Lucas Candeia Teixeira, Bruna Roberta Araujo Leal et al. "Towards Screening of Children Students with Autism Spectrum Disorder Based on Executive Functions with Serious Game and Machine Learning Approaches", Expert Systems with Applications, 2025  
Publication

<1 %

11 [btu.edu.ge](http://btu.edu.ge)  
Internet Source

<1 %

12 [jurnalnasional.ump.ac.id](http://jurnalnasional.ump.ac.id)  
Internet Source

<1 %

13 [migrationletters.com](http://migrationletters.com)  
Internet Source

<1 %

14 "Exploring Trends, Innovations, and Digitalization of Entrepreneurship", Springer Science and Business Media LLC, 2025  
Publication

<1 %

15 [1library.net](http://1library.net)  
Internet Source

<1 %

16 [jurnalilmukomunikasi.uho.ac.id](http://jurnalilmukomunikasi.uho.ac.id)  
Internet Source

<1 %

17 Submitted to CSU, San Jose State University  
Student Paper

<1 %

18 Sevtian Ferdian, Aditiya Hermawan, Edy Edy. "Designing a Chatbot Based on Full-Text Search and 3D Modelling as a Promotional Media", JUITA : Jurnal Informatika, 2023  
Publication

<1 %

19 [itijournal.org](http://itijournal.org)  
Internet Source

<1 %

20 [link.springer.com](http://link.springer.com)  
Internet Source

<1 %

21	nepjol.info Internet Source	<1 %
22	www.frontiersin.org Internet Source	<1 %
23	www.ncbi.nlm.nih.gov Internet Source	<1 %
24	Ira Puspita Sari, Budi Warsito, Oky Dwi Nurhayati. "Automated Classification of Parasitic Worm Eggs Based on Transfer Learning and Fine-Tuned CNN Models", International Journal of Advanced Computer Science and Applications, 2025 Publication	<1 %
25	Naveen Kumar Pareek, Deepika Soni, Sheshang Degadwala. "Prediction of CKD Using Expert System Fuzzy Logic & AI", 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2023 Publication	<1 %
26	Xu, Hongyi. "Constructing and Evaluating a Predictive Model for Cardiovascular Disease Based on BRFSS Data", Yale University, 2025 Publication	<1 %
27	www.science.gov Internet Source	<1 %
28	Louw, Claudette. "To Investigate the Role of the 3R's (Review, Rating and Recommendation Systems) in Shaping Overall Online Customer Experience", University of the Witwatersrand, Johannesburg (South Africa) Publication	<1 %
29	Submitted to The University of the West of Scotland Student Paper	<1 %

---

30 Winda Ayu Fazraningtyas, Bahbib  
Rahmatullah, Husni Naparin, Mohammad  
Basit, Nor Asiah Razak. "A predictive model  
for postpartum depression: ensemble  
learning strategies in machine learning",  
Indonesian Journal of Electrical Engineering  
and Computer Science, 2025  
Publication

<1 %

---

31 dokumen.pub  
Internet Source

<1 %

---

32 static.frontiersin.org  
Internet Source

<1 %

---

33 thesai.org  
Internet Source

<1 %

---

34 www.fastercapital.com  
Internet Source

<1 %

---

35 "Proceedings of International Conference on  
Communication and Computational  
Technologies", Springer Science and Business  
Media LLC, 2023  
Publication

<1 %

---

36 Jun Zheng, Wang Zheng, Ankai Chen, Jinkang  
Yao, Yicheng Ren, Chen Zhou, Jian Wu, Wei  
Ling, Bing Bai, Wei Wang, Zhongwei Zhang.  
"Sustainability of unconventional machining  
industry considering impact factors and  
reduction methods of energy consumption: A  
review and analysis", Science of The Total  
Environment, 2020  
Publication

<1 %

---

37 Md. Rezaul Islam, Sultana Kowser, Aniruddha  
Islam Chowdhury, Sharmin Shama, Md.  
Masudul Hasan Lamyaa. "Early Prediction of  
Liver Disease Using Ensemble Machine  
Learning Models and Explainable AI

<1 %

## Techniques", Biomedical Materials & Devices, 2025

Publication

- 
- |    |   |      |
|----|---|------|
| 38 | fastercapital.com<br>Internet Source  | <1 % |
| 39 | ijair.id<br>Internet Source   | <1 % |
| 40 | www.growingscience.com<br>Internet Source   | <1 % |
| 41 | "Data Engineering and Applications", Springer Science and Business Media LLC, 2024<br>Publication   | <1 % |
| 42 | Dinda Julia Arfah, Masrizal Masrizal, Irmayanti Irmayanti. "Machine Learning to Predict Student Satisfaction Level Using KNN Method and Naive Bayes Method", sinkron, 2024<br>Publication   | <1 % |
| 43 | Muhammad Yassir, Wanda Cahyani. "Prediction Analysis of Package C Student Graduation at the Bollo DMansel Community Learning Activity Center (PKBM) with the Naïve Bayes Algorithm Method", Internet of Things and Artificial Intelligence Journal, 2024<br>Publication | <1 % |
| 44 | Mujiono Sadikin, Fahri Alfiandi. "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk", International Journal of Electrical and Computer Engineering (IJECE), 2018<br>Publication                                       | <1 % |
| 45 | Sidharta Sidharta, Yaya Heryadi, Lukas Lukas, Wayan Suparta, Yulyani Arifin. "Sentiment Analysis of Hashtag Activism on Social Media Twitter", 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023               | <1 % |

46 Esti Suryani, Wiharto -, Adi Prasetya Putra, Wisnu Widiarto. "Detection of Acute Myeloid Leukemia based on White Blood Cell Morphological Imaging using Naïve Bayesian Algorithm", International Journal of Advanced Computer Science and Applications, 2021  
Publication

---

47 Sirmayanti, Pulung Hendro Prastyo, Mahyati. "Enhancing Diabetes Prediction Performance using Feature Selection Based on Grey Wolf Optimizer with Autophagy Mechanism", Computer Methods and Programs in Biomedicine Update, 2025  
Publication

---

48 Slamet Riyadi, Muhammad Dzaki Mubarok, Cahya Damarjati, Asnor Juraiza Ishak. "Improving Sentiment Analysis Accuracy Using CRNN on Imbalanced Data: a Case Study of Indonesian National Football Coach", JUITA: Jurnal Informatika, 2024  
Publication

---

Exclude quotes  Off

Exclude matches  Off

Exclude bibliography  On