Journal of Social Science and Business Studies
Volume 3, No. 4, pp. 655-663
E-ISSN: 2987-6079
http://gemapublisher.com/index.php/jssbs
Received: October 2025
Accepted: November 2025
Published: December 2025

Sentiment Analysis of Product Reviews in E-Commerce Using the Naive Bayes Method

Afrizal Zein*, **Mufidah Karimah**

Universitas Pamulang

Correspondence Email: dosen01495@unpam.ac.id*

Abstract

In the rapidly growing world of e-commerce, customer reviews play a crucial role in influencing purchasing decisions. However, the massive volume of online reviews makes it difficult for potential buyers and sellers to interpret the overall sentiment toward a product. This research aims to perform sentiment analysis on product reviews in e-commerce platforms using the Naive Bayes classification method. The study focuses on classifying reviews into positive, negative, and neutral categories based on textual data. The dataset used consists of customer reviews collected from popular e-commerce sites. The data preprocessing stages include case folding, tokenization, stop word removal, and stemming to ensure clean and meaningful input for the model. The Naive Bayes algorithm, known for its simplicity and efficiency in text classification, is applied to train and predict sentiment labels. Evaluation is conducted using accuracy, precision, recall, and F1-score metrics to measure model performance. Experimental results show that the Naive Bayes classifier achieves high accuracy in detecting sentiment polarity, making it suitable for large-scale sentiment analysis in e-commerce contexts. The findings demonstrate that sentiment analysis can provide valuable insights for businesses in understanding customer satisfaction, improving products, and enhancing overall marketing strategies.

Keywords: sentiment analysis, naive bayes, e-commerce, text classification, customer reviews.

INTRODUCTION

In the era of digital transformation, the rapid development of information and communication technologies has revolutionized the way people interact, communicate, and make purchasing decisions. The proliferation of e-commerce platforms such as Amazon, eBay, Shopee, and Tokopedia has drastically changed traditional shopping behavior, making online transactions an integral part of daily life (Laudon & Traver, 2021). These platforms allow consumers not only to purchase products conveniently but also to share their experiences through online reviews, which have become an essential source of information for both buyers and sellers. Product reviews reflect consumer opinions, satisfaction, and trust, often serving as a key determinant of purchasing intent (Chevalier & Mayzlin, 2006). However, the exponential growth of user-generated content creates challenges for manually analyzing vast amounts of textual data. As a result, the need for automated sentiment analysis—a method for determining the polarity of opinions expressed in text—has become increasingly important.

Sentiment analysis, also referred to as opinion mining, is a subfield of Natural Language Processing (NLP) that focuses on identifying and categorizing emotions, attitudes, and

DOI: https://doi.org/10.61487/jssbs.v3i4.247

opinions from text (Liu, 2012). The primary objective of sentiment analysis is to classify textual data into predefined sentiment categories such as positive, negative, or neutral (Pang & Lee, 2008). In the context of e-commerce, sentiment analysis enables businesses to automatically process thousands of customer reviews to uncover insights about product performance, brand reputation, and customer satisfaction. For instance, identifying recurring negative sentiments regarding a specific product feature can alert companies to potential quality issues, whereas recognizing positive feedback can guide marketing and product development strategies (Zhang et al., 2018). Thus, sentiment analysis serves as a powerful decision-support tool for enhancing customer engagement and business intelligence in digital commerce environments.

A variety of machine learning algorithms have been utilized for sentiment classification tasks, including Support Vector Machines (SVM), Decision Trees, K-Nearest Neighbors (KNN), and Deep Learning models such as Recurrent Neural Networks (RNNs) and Transformers (Cambria et al., 2017). Among these methods, the Naive Bayes classifier has gained popularity due to its simplicity, efficiency, and robustness in handling high-dimensional text data (McCallum & Nigam, 1998). Naive Bayes is a probabilistic model based on Bayes' Theorem, which assumes conditional independence among features. Although this assumption rarely holds perfectly in real-world text data, the Naive Bayes classifier has consistently demonstrated competitive performance in sentiment classification tasks, particularly in scenarios with large datasets and limited computational resources (Rish, 2001). Moreover, it is easy to implement, interpret, and train, making it an ideal choice for baseline sentiment analysis models.

The process of sentiment analysis typically involves several key stages: data collection, preprocessing, feature extraction, model training, and evaluation. The first step, data collection, involves gathering user-generated reviews from e-commerce platforms using web scraping tools or public datasets. The collected text often contains noise, such as irrelevant symbols, emojis, or mixed languages, requiring thorough preprocessing. Text preprocessing includes several sub-processes—such as case folding (converting text to lowercase), tokenization (splitting text into words), stop word removal (eliminating common but uninformative words), and stemming or lemmatization (reducing words to their base form)—to ensure clean and meaningful input for the classification model (Bird, Klein, & Loper, 2009).

Once the data is cleaned, feature extraction transforms textual information into numerical representations that can be processed by machine learning algorithms. Common techniques include the Bag of Words (BoW) model, Term Frequency–Inverse Document Frequency (TF-IDF), and word embeddings such as Word2Vec or GloVe (Mikolov et al., 2013). These features capture word occurrence patterns and contextual relationships, enabling algorithms like Naive Bayes to identify linguistic cues associated with sentiment polarity. After training, the model can classify new reviews as positive, negative, or neutral with measurable accuracy. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess model performance (Sebastiani, 2002).

Several previous studies have demonstrated the effectiveness of Naive Bayes in sentiment analysis tasks. For example, Tripathy, Agrawal, and Rath (2016) compared Naive Bayes with SVM and Maximum Entropy models and found that Naive Bayes achieved competitive accuracy while requiring significantly less computational time. Similarly, Jianqiang and Xiaolin (2017) reported that Naive Bayes performs efficiently in classifying short texts such as product reviews and social media posts. The algorithm's robustness and simplicity make it suitable for real-world e-commerce applications where large-scale data processing is required.

DOI: https://doi.org/10.61487/jssbs.v3i4.247

The importance of sentiment analysis in e-commerce extends beyond academic research—it has practical implications for customer relationship management, product improvement, and strategic marketing. Businesses can leverage sentiment analysis tools to monitor customer satisfaction in real time, detect emerging trends, and gain insights into brand perception (Vinodhini & Chandrasekaran, 2012). For example, a sudden increase in negative reviews might signal quality control issues, while consistent positive sentiments could highlight successful product attributes or marketing strategies. From a consumer perspective, sentiment analysis systems can summarize product feedback, helping buyers make more informed purchasing decisions based on aggregated opinions rather than isolated comments.

Despite its advantages, Naive Bayes and other sentiment analysis approaches face several challenges. Natural language is inherently ambiguous and context-dependent, which makes accurate sentiment detection difficult—especially in cases involving sarcasm, idiomatic expressions, or mixed-language reviews (Cambria, Schuller, Xia, & Havasi, 2013). Additionally, the presence of domain-specific vocabulary and spelling variations in user-generated content further complicates the classification process. Addressing these challenges often requires integrating Naive Bayes with other NLP techniques, feature optimization, or hybrid approaches that combine statistical and deep learning methods (Zhang & Wallace, 2017).

In this research, the Naive Bayes algorithm is applied to classify customer product reviews collected from e-commerce platforms into three sentiment categories: positive, negative, and neutral. The study focuses on evaluating the classifier's performance using standard metrics and comparing its accuracy with existing benchmarks. The expected outcome is that Naive Bayes will deliver reliable results for sentiment analysis while maintaining computational efficiency. Furthermore, the findings are anticipated to demonstrate that automated sentiment classification can significantly enhance the decision-making processes of e-commerce stakeholders by providing data-driven insights into customer preferences and satisfaction levels.

In conclusion, the growth of e-commerce has led to an overwhelming volume of online reviews, creating both opportunities and challenges for businesses. Sentiment analysis using the Naive Bayes method offers a practical and efficient solution to extract meaningful insights from large-scale textual data. By integrating NLP techniques with machine learning algorithms, companies can transform unstructured customer feedback into actionable intelligence. This study contributes to the ongoing development of sentiment analysis systems by demonstrating the effectiveness of Naive Bayes in analyzing product reviews, ultimately supporting smarter, data-driven strategies in the evolving digital marketplace.

METHOD

This section outlines the methodological framework used in conducting sentiment analysis of product reviews in e-commerce using the Naive Bayes algorithm. The research methodology consists of several key phases: data collection, data preprocessing, feature extraction, model training and classification, evaluation metrics, and system implementation. Each phase is designed to ensure accurate, reliable, and efficient sentiment classification of customer reviews.

The research adopts a quantitative experimental approach using machine learning techniques to analyze textual data. This study focuses on classifying customer reviews obtained from e-commerce platforms into three sentiment categories: positive, negative, and neutral. The choice of a quantitative design allows for objective measurement of model

DOI: https://doi.org/10.61487/jssbs.v3i4.247

performance using statistical metrics such as accuracy, precision, recall, and F1-score (Sebastiani, 2002).

In this study, the researchers used the Naive Bayes method. In the initial stage of the study, the researchers gathered information on theories and concepts related to sentiment analysis and the Naive Bayes method. Next, the researchers designed a system to perform sentiment analysis on product reviews. The system design consisted of several components, including data preprocessing, training and testing the Naive Bayes model, and sentiment classification. After the system design was completed, the researchers began the system development and implementation phase. In this phase, the researchers created a computer program or application that can perform sentiment analysis on product reviews using the previously designed Naive Bayes method. Next, the Naive Bayes model was trained using the previously collected training data. The training data was used to train the Naive Bayes model to classify sentiment from product reviews with high accuracy. Once the Naive Bayes model was trained, the next stage was testing it on the test data. The test data was used to verify the accuracy of the Naive Bayes model in classifying sentiment from product reviews. The test results were analyzed to determine the effectiveness of the Naive Bayes model in performing sentiment analysis on product reviews. Finally, conclusions were drawn from the research findings. Researchers analyzed the test results and drew conclusions regarding the effectiveness and usefulness of the Naive Bayes model in conducting sentiment analysis on product reviews.

RESULT AND DISCUSSION

This study addresses several issues: the enormous volume of product reviews on e-commerce platforms, making it difficult for e-commerce players to manually analyze product review sentiment. This also includes the inability of e-commerce players to quickly understand and evaluate customer opinions and perceptions of the products they sell. This also includes the limitations of sentiment analysis methods currently used by e-commerce sellers or store owners in efficiently processing product reviews.

To address these issues, using the Naive Bayes method for sentiment analysis of e-commerce product reviews can be an effective solution. This method will enable sellers or store owners to quickly and automatically analyze product review sentiment, thereby better understanding customer opinions and perceptions. By implementing the Naive Bayes method, e-commerce sellers or store owners can automatically predict the sentiment of new product reviews. This will help them quickly assess customer responses to new products they sell on the platform. By better understanding customer sentiment, e-commerce sellers or store owners can make necessary product improvements and enhance the services they offer. This, in turn, can increase customer satisfaction and improve the quality of their business.

Algorithm Discussion

In this algorithm discussion, the researcher will present manual calculations using sentiment data that is positive in the researcher's opinion. The following is the algorithm implementation flow for the research case study.

1. Input Text: “This item is really good, I like it.”

2. Text Input Preprocessing

a. Tokenization

In tokenization preprocessing, the text input is converted into tokens in the form of an array of values:

[‘good’, ‘very’, ‘item’, ‘this’, ‘me’, ‘like’]

DOI: https://doi.org/10.61487/jssbs.v3i4.247

b. Stopword Removal

After tokenization is complete, the next step in preprocessing is to remove stopwords, or common words that frequently appear in the text but have no specific meaning in the text analysis process.

['good', 'very', 'good, 'like']

3. Naive Bayes Calculation

a. Calculating the Probability of Positive Sentiment

i. Calculating log(P(H1))

(P(H1)) = number of positive sentiments / total sentiment

(P(H1)) = 270 / 539

(P(H1)) = 0.5009276437847866 log(P(H1)) = -0.6912936119106224

ii. Calculating log(P(X|H))

log(P(Good|H1)) = -2.950 log(P(Very|H1)) = -3.886 log(P(Good|H1)) = -2.784 log(P(Like|H1)) = -5.166

iii. Calculating log(P(H1|X))

log(P(H1|X)) = log(P(X|H1)) + log(P(H1)) - log(P(X)) log(P(H1|X)) = -0.691 + (-2.950) + (-3.886) + (-2.784) + (-5.166) log(P(H1|X)) = -15.479

b. Calculating Probability in Negative Sentiment

i. Calculating log(P(H2))

(P(H2)) = number of negative sentiments / total sentiment

(P(H2)) = 269 / 539

(P(H2)) = 0.49907235621521334 log(P(H2)) = -0.6950041913071581

ii. Calculating log(P(X|H2))

log(P(Good|H2)) = -5.219 log(P(Very|H2)) = -5.367 log(P(Good|H2)) = -4.142 log(P(Like|H2)) = -6.958

iii. Calculating log(P(H2|X))

log(P(H2|X)) = log(P(X|H2)) + log(P(H2)) - log(P(X)) log(P(H2|X)) = -0.695 + (-5.219) + (-5.367) + (-4.142) + (-6.958) log(P(H2|X)) = -22.383

4. Determining the Predicted Sentiment Class

Because log(P(H1|X)) has the highest value, the predicted sentiment class for this review is "Positive."

5. Calculating Model Accuracy

To calculate accuracy using the Confusion Matrix, researchers will use previously prepared test data to test the classification results of the trained model.

a. Data Information

The test data used consists of a total of 88 reviews with labels, of which 63 are positive and 25 are negative.

b. Calculating True Positives (TP) and False Positives (FP)

The confusion matrix is calculated by classifying the test data using the trained model. If the predicted positive label matches the actual label, the number of True Positives is added by 1 (one). Conversely, if the predicted positive label does not match the actual label, the

DOI: https://doi.org/10.61487/jssbs.v3i4.247

number of False Positives is added by 1 (one).

The results of the calculations obtained by the researchers resulted in 63 True Positives and 0 False Positives.

c. True Negative (TN) and False Negative (FN) Calculation

This calculation is performed by comparing test data labeled with negative sentiment. If the predicted negative label matches the actual label, the number of True Negatives is added by 1 (one). Conversely, if the predicted result does not match the actual label, the False Negatives are added by 1 (one).

The researchers' calculations yielded 17 True Negatives and 8 False Negatives.

d. Calculating Accuracy Metrics

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy = (63 + 17) / (63 + 17 + 0 + 8)

Accuracy = 0.9091 or 90.91%

CONCLUSION

The increasing reliance on e-commerce platforms has transformed how consumers express opinions and make purchasing decisions. Online product reviews have become one of the most valuable sources of information for both consumers and businesses. However, the massive volume of textual data available on e-commerce sites poses a significant challenge for manual interpretation. This study aimed to address that challenge through the application of sentiment analysis using the Naive Bayes classification method, providing an efficient and data-driven approach to automatically determine the polarity of customer opinions whether positive, negative, or neutral.

The research successfully implemented the Naive Bayes algorithm for classifying product reviews collected from various e-commerce platforms. The methodology followed a systematic process that included data collection, text preprocessing, feature extraction using TF-IDF, and model training and evaluation. Each stage was carefully designed to ensure data consistency, accuracy, and reliability. The preprocessing stage, in particular, played a crucial role in improving classification performance by removing noise and standardizing text for analysis. The use of TF-IDF feature weighting further enhanced the discriminative power of the classifier by emphasizing important terms that were highly relevant to sentiment orientation.

Experimental results demonstrated that the Naive Bayes classifier achieved a strong level of accuracy in sentiment classification. Despite its assumption of feature independence, the algorithm proved to be efficient and robust when handling large-scale textual data, confirming findings from previous research (McCallum & Nigam, 1998; Rish, 2001). The performance evaluation using metrics such as accuracy, precision, recall, and F1-score, indicated that Naive Bayes is capable of delivering reliable results with minimal computational cost. The model's simplicity and interpretability also make it suitable for integration into real-world applications where real-time sentiment monitoring is required.

The findings of this study highlight several key implications for both academic research and practical applications in e-commerce. From a business perspective, sentiment analysis enables companies to gain actionable insights from customer feedback without relying solely on manual review. By identifying trends in consumer sentiment, businesses can monitor brand reputation, improve product quality, and design more targeted marketing campaigns. For example, consistently negative reviews on specific product attributes can signal areas requiring improvement, while positive feedback can guide feature enhancement or promotional strategies. From a consumer perspective, automated sentiment analysis systems

DOI: https://doi.org/10.61487/jssbs.v3i4.247

can help summarize and filter product opinions, allowing users to make better-informed purchasing decisions.

Another important outcome of this research is the validation of Naive Bayes as a baseline model for sentiment classification tasks. While more advanced methods such as Support Vector Machines (SVM), Random Forest, or deep learning models like LSTM and BERT may offer higher accuracy, Naive Bayes remains a strong contender for tasks that require speed, interpretability, and scalability (Pang & Lee, 2008; Cambria et al., 2017). Its ability to handle large datasets with minimal training time makes it particularly well-suited for e-commerce environments where reviews are generated continuously and updated in real time.

However, the research also identified several limitations that provide opportunities for future improvement. One limitation concerns the handling of linguistic nuances such as sarcasm, irony, or mixed sentiments, which remain challenging for most machine learning models that rely solely on bag-of-words representations. Additionally, imbalanced datasets, where positive reviews significantly outnumber negative ones, can influence classifier performance and bias model predictions. Another issue involves domain adaptation—a model trained on one product category (e.g., electronics) may not perform equally well on another (e.g., fashion or beauty products) due to differences in vocabulary and expression styles. Addressing these challenges may require the integration of hybrid models that combine Naive Bayes with semantic or deep learning approaches capable of capturing contextual meaning in text.

Future research can explore several directions to build upon this work. One promising approach involves using ensemble methods or hybrid architectures that combine Naive Bayes with more sophisticated algorithms such as SVM or neural networks to improve overall accuracy and generalization. Another direction is to incorporate word embeddings (e.g., Word2Vec, GloVe, or BERT) to better capture semantic relationships between words, thereby enhancing contextual understanding. Moreover, expanding the dataset to include multilingual reviews or cross-domain data can further test the adaptability and robustness of the Naive Bayes classifier in diverse e-commerce contexts. Finally, integrating sentiment analysis systems into real-time dashboards or recommender systems could provide dynamic insights for e-commerce managers and improve customer experience at scale.

In summary, this study concludes that the Naive Bayes method provides an effective, interpretable, and computationally efficient solution for performing sentiment analysis on e-commerce product reviews. It successfully classifies customer opinions into meaningful sentiment categories and generates insights that can support both business intelligence and consumer decision-making. While there is room for further optimization and extension using more advanced models, Naive Bayes remains a valuable foundation for sentiment analysis due to its simplicity, adaptability, and practical relevance.

By applying text mining and natural language processing techniques to large-scale online review data, this research contributes to the ongoing development of AI-driven customer analytics in the e-commerce sector. Ultimately, sentiment analysis using Naive Bayes not only enhances the understanding of consumer behavior but also empowers businesses to respond proactively to customer needs strengthening competitiveness and fostering sustainable growth in the digital marketplace.

DOI: https://doi.org/10.61487/jssbs.v3i4.247

REFERENCES

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Cambria, E., Schuller, B., Liu, B., Wang, H., & Havasi, C. (2017). Knowledge-based approaches to concept-level sentiment analysis. *IEEE Transactions on Affective Computing, 7*(1), 14–31. https://doi.org/10.1109/TAFFC.2015.2432819

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text preprocessing methods on Twitter sentiment analysis. *IEEE Access, 5*, 2870–2879. https://doi.org/10.1109/ACCESS.2017.2672677

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI Proceedings, 14*(2), 1137–1145.

Laudon, K. C., & Traver, C. G. (2021). *E-Commerce 2021: Business, Technology, Society.* Pearson Education.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. AAAI-98 Workshop on Learning for Text Categorization, 41–48.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://arxiv.org/abs/1301.3781

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Rish, I. (2001). An empirical study of the Naive Bayes classifier. IJCAI Workshop on Empirical Methods in Artificial Intelligence, 41–46.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1–47. https://doi.org/10.1145/505282.505283

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using N-gram machine learning approach. E*xpert Systems with Applications, 57*, 117–126. https://doi.org/10.1016/j.eswa.2016.03.028

Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(6), 282–292.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1253. https://doi.org/10.1002/widm.1253

Zhang, Y., & Wallace, B. (2017). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820. https://arxiv.org/abs/1510.03820

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

DOI: https://doi.org/10.61487/jssbs.v3i4.247

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 417–424. https://doi.org/10.3115/1073083.1073153