

## SENTIMENT ANALYSIS OF CUSTOMER REVIEWS ON E-COMMERCE APPLICATIONS: LAZADA, TOKOPEDIA, AND BLIBLI

**Andhika Ihza Adriansyah<sup>1\*</sup>, Muhammad Arifin<sup>1</sup>, Arif Setiawan<sup>1</sup>**

<sup>1</sup>Information Systems, Universitas Muria Kudus

*email: \*202153096@std.umk.ac.id*

**Abstract:** The rapid growth of e-commerce in Indonesia has increased consumer interactions with digital platforms, particularly Lazada, Tokopedia, and Blibli, resulting in a large volume of customer reviews that reflect consumer experiences and perceptions but have not been optimally utilized in business decision-making. The main issue addressed in this study is how to process customer review data to generate meaningful information regarding consumer opinions. This research aims to apply web scraping techniques to collect customer review data and conduct sentiment analysis to identify trends in consumer opinions across the three e-commerce platforms. The dataset consists of 3,000 customer reviews, with 1,000 reviews collected from each platform, covering aspects such as shopping experience, service quality, delivery process, and customer satisfaction. The research methodology includes data collection through web scraping, text preprocessing for data cleaning and normalization, sentiment analysis using machine learning approaches, and visualization of sentiment results. The findings indicate differences in the distribution of positive, negative, and neutral sentiments across platforms, reflecting variations in consumer experiences and service strategies. These results demonstrate that sentiment analysis based on customer reviews can serve as strategic input to improve service quality, business performance, and marketing strategies in Indonesia's e-commerce sector.

**Keywords:** customer reviews; digital services; e-commerce; sentiment analysis; web scarping

**Abstrak:** Pertumbuhan pesat e-commerce di Indonesia meningkatkan interaksi konsumen dengan platform digital, khususnya Lazada, Tokopedia, dan Blibli, yang menghasilkan ulasan pelanggan dalam jumlah besar sebagai cerminan pengalaman dan persepsi konsumen, namun belum dimanfaatkan secara optimal dalam pengambilan keputusan bisnis. Permasalahan utama penelitian ini adalah bagaimana mengolah data ulasan tersebut agar dapat memberikan informasi yang bermakna mengenai opini konsumen. Penelitian ini bertujuan menerapkan web scraping untuk mengumpulkan data ulasan pelanggan serta melakukan analisis sentimen guna mengidentifikasi tren opini konsumen pada ketiga platform e-commerce tersebut. Data yang digunakan berjumlah 3.000 ulasan pelanggan, dengan masing-masing platform diwakili oleh 1.000 ulasan yang mencakup pengalaman berbelanja, kualitas layanan, proses pengiriman, dan tingkat kepuasan pelanggan. Metode penelitian meliputi pengambilan data menggunakan web scraping, pra-pemrosesan teks untuk pembersihan dan normalisasi data, analisis sentimen dengan pendekatan pembelajaran mesin, serta visualisasi hasil sentimen. Hasil penelitian menunjukkan adanya perbedaan distribusi sentimen positif, negatif, dan netral pada setiap platform, yang mencerminkan variasi pengalaman konsumen dan strategi layanan. Temuan ini menunjukkan bahwa analisis sentimen berbasis ulasan pelanggan dapat menjadi masukan strategis untuk meningkatkan kualitas layanan, kinerja bisnis, dan strategi pemasaran e-commerce di Indonesia.

**Kata kunci:** customer reviews; digital services; e-commerce; sentiment analysis; web scarping



## INTRODUCTION

The development of digital technology in Indonesia has driven significant changes in consumer behavior. Currently, mobile-based e-commerce applications have become the primary medium for conducting transactions due to the convenience, speed, and flexibility they offer. [1]Lazada, Tokopedia, and Blibli are three major e-commerce applications in Indonesia that compete to attract and retain users. Application quality plays an important role in influencing user satisfaction and loyalty, making an understanding of user perceptions crucial for application developers [2].

The Google Play Store provides a review feature that allows users to submit evaluations based on their experiences,[3] including both positive and negative opinions covering aspects such as user interface design, system performance, access speed, and application stability [4]. The large volume of reviews poses challenges for manual data processing; however, these reviews also represent a valuable source of information for application evaluation and strategic decision-making by developers [5].

Based on these issues, this study aims to automatically collect user reviews of the Lazada, Tokopedia, and Blibli e-commerce applications from the Google Play Store[6] using web scraping techniques, identify and classify user sentiment toward application quality, analyze user perceptions across each e-commerce platform, and evaluate the performance of the Naive Bayes[7] method in sentiment classification. In addition, this study seeks to assess the performance of the classification model using Accuracy, Precision, Recall, and

F1-Score as objective evaluation metrics.[8] The results of this study are expected to provide strategic recommendations for e-commerce application developers to improve user-based digital service quality [9], while also contributing to academic research on sentiment analysis, text classification algorithm evaluation, and the utilization of user reviews in the context of mobile applications [10].

## METHOD

This study is categorized as applied research using a quantitative approach. An experimental approach is employed because the study involves testing a sentiment classification model to evaluate the algorithm’s performance in processing user review data.[11] The focus of this study is not only on the modeling process but also on the application of sentiment analysis methods to real-world data obtained from e-commerce platforms.

The objects of this study are user reviews of the Lazada, Tokopedia, and Blibli applications collected from the Google Play Store.[12] To maintain the relevance of the analysis, the data are limited to reviews written in Indonesian and collected within a specific time period.

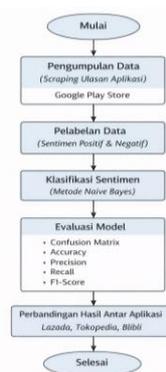


Figure 1 Framework

In general, this study is conducted through several main stages as illustrated in the figure above. These stages include data collection, text preprocessing, sentiment labeling, and sentiment classification using the Naïve Bayes algorithm,[13] model performance evaluation, as well as comparative analysis of the results across the e-commerce applications.

The data in this study were obtained by applying web scraping techniques to the Google Play Store. The review collection process was conducted using the Python programming language with the assistance of the google-play-scraper, BeautifulSoup, and Selenium libraries. The collected data consist of user review texts that reflect users' opinions and experiences with the analyzed e-commerce applications. In this study, user review texts function as the independent variables, while sentiment classes serve as the dependent variables. The sentiment categories are divided into two classes, namely positive sentiment and negative sentiment. Positive sentiment indicates user satisfaction and favorable experiences, whereas negative sentiment represents complaints and less satisfactory experiences [14].

The collected review data still contain various elements that may interfere with the analysis process, such as symbols, numbers, and non-standard words. Therefore, a data preprocessing stage is conducted to improve data quality prior to the classification process. The preprocessing steps include case folding to convert all text to lowercase, removal of punctuation and numbers, stopword removal to eliminate common words that do not convey sentiment meaning, tokenization to divide text into individual words, and stemming to

reduce words to their root forms [15]. This process aims to simplify text representation and make it more suitable for processing by the classification model.

After the preprocessing stage is completed, the review data are labeled as either positive or negative sentiment. This labeling process serves as ground truth data for training and testing the sentiment classification model. The labeled reviews are then analyzed using the Naïve Bayes algorithm to determine whether each review belongs to the positive or negative sentiment category. The Naïve Bayes algorithm is selected due to its effectiveness in handling large-scale text data, relatively fast computation time, and ease of implementation. Additionally, this algorithm has been widely applied in sentiment analysis research, allowing the results of this study to be compared with previous studies.

To evaluate the performance of the classification model, the dataset is divided into 70% training data and 30% testing data. This data split enables the model to learn sentiment patterns from the training data and assess its ability to classify previously unseen data. The evaluation of classification results is carried out using a confusion matrix with four performance metrics: Accuracy, Precision, Recall, and F1-Score. These metrics are used to measure the accuracy, reliability, and consistency of the model in classifying positive and negative sentiments.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2X \frac{\text{Precision} \times \text{Recal}}{\text{Precision} + \text{recall}} \quad (4)$$



### Data Labeling

At this stage, each user review is assigned a positive or negative sentiment label. This labeling aims to determine whether the content of the review reflects satisfaction or, conversely, complaints regarding the e-commerce application being used.

Confusion matrix analysis is used to examine in detail the success and error rates of the classification model in predicting review sentiments. Based on the confusion matrix shown in the figure below, it can be observed that all three applications have a higher number of positive sentiment predictions compared to negative ones. This indicates that user perceptions of the three e-commerce platforms tend to be positive.

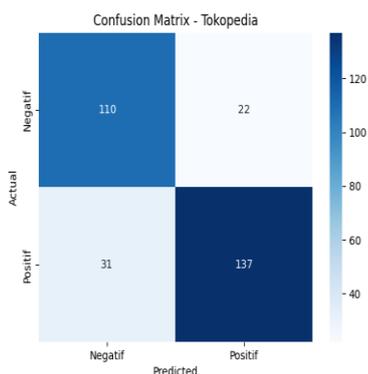


Figure 3 Confusion Matrix Tokopedia

In the Tokopedia application, the model is able to classify most positive and negative reviews accurately, although some misclassifications still occur in both classes. These errors indicate variations in vocabulary and writing styles within the reviews, which cause certain sentiments to be not optimally identified.

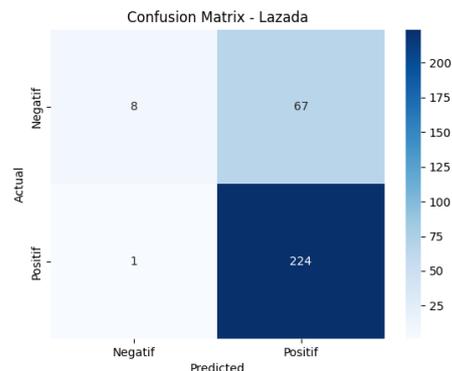


Figure 4 Confusion Matrix Lazada

In the Lazada application, the confusion matrix shows a dominance of positive sentiment predictions. The model performs very well in recognizing positive reviews, but a considerable number of negative reviews are still misclassified as positive. This condition indicates a tendency of the model to be biased toward the positive class in the Lazada dataset.

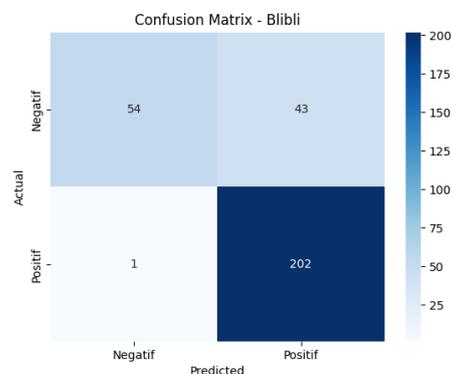


Figure 5 Confusion Matrix Blibli

Meanwhile, in the Blibli application, the confusion matrix shows a relatively more balanced performance. The model is able to classify positive reviews very well, and the misclassification rate for the positive class is relatively low, although there are still some negative reviews that are incorrectly predicted as positive.

### Implementasi Naïve Bayes

In this study, the sentiment classification process is carried out using the Naïve Bayes algorithm. The labeled review data are then transformed into numerical form using the TF-IDF method, so that each word is assigned a weight according to its frequency of occurrence in the reviews.

The data are divided into two parts: 70% as training data and 30% as testing data. The training data are used to train the model to recognize sentiment patterns, while the testing data are used to evaluate the model's ability to classify reviews that have not been previously seen.

	precision	recall	f1-score	support
Negatif	0.98	0.56	0.71	97
Positif	0.82	1.00	0.90	203
accuracy			0.85	300
macro avg	0.90	0.78	0.81	300
weighted avg	0.88	0.85	0.84	300

Figure 6 values for the blibli application

	precision	recall	f1-score	support
Negatif	0.89	0.11	0.19	75
Positif	0.77	1.00	0.87	225
accuracy			0.77	300
macro avg	0.83	0.55	0.53	300
weighted avg	0.80	0.77	0.70	300

Figure 7 values for the lazada application

	precision	recall	f1-score	support
Negatif	0.77	0.83	0.80	132
Positif	0.86	0.81	0.83	168
accuracy			0.82	300
macro avg	0.82	0.82	0.82	300
weighted avg	0.82	0.82	0.82	300

Figure 8 values for the Tokopedia application

### Evaluation of the Naïve Bayes Model Performance

The model performance evaluation is conducted using four main metrics: Accuracy, Precision, Recall, and F1-Score. The metric values are calculated based on the model's predictions on the

testing data, which constitute 30% of the total dataset, while the remaining 70% is used as training data. This data split is intended to ensure that the model has sufficient data to learn sentiment patterns while also providing a representative evaluation of its generalization capability.

Table 2 Performance Comparison Across Platforms

Dataset	Accur acy	Precisi on	Reca ll	F1- score
Blibli	85.33 %	82.45 %	99.51 %	90.18 %
Lazada	77.33 %	76.98 %	99.56 %	86.82 %
Tokope dia	82.33 %	86.16 %	81.55 %	83.79 %

The evaluation results show that the Blibli application has the best model performance, with an Accuracy of 85.33% and an F1-Score of 90.18%. This indicates that the model is able to classify Blibli review sentiments consistently and in a balanced manner.

Aplikasi Tokopedia memperoleh nilai Accuracy sebesar 82,33% dan F1-Score sebesar 83,79%, yang menunjukkan performa model cukup stabil meskipun masih terdapat kesalahan klasifikasi akibat variasi bahasa pada ulasan pengguna.

Meanwhile, the Lazada application achieves an Accuracy of 77.33% and an F1-Score of 86.82%. The high Recall value indicates that the model is very effective at detecting positive sentiment, while the lower Precision suggests a tendency for the model to misclassify negative reviews as positive.

Thus, this study identifies user sentiment trends toward e-commerce applications and evaluates the performance of the Naïve Bayes

algorithm in classifying user reviews. The results indicate that sentiment analysis can provide an overall picture of user perceptions for each application. Furthermore, the model performance evaluation shows that the Naïve Bayes algorithm is fairly effective for classifying Indonesian-language text sentiment, although it still has limitations in handling language variations and expressions of user complaints. Therefore, this study not only achieves the technical objectives of classification but also provides a deeper understanding of user experiences with e-commerce applicati

## CONCLUSION

Based on the results of this study, it can be concluded that the application of web scraping and sentiment analysis using the Naïve Bayes algorithm is able to effectively classify user reviews of Lazada, Tokopedia, and Blibli e-commerce applications on the Google Play Store. The sentiment analysis results show that positive sentiment dominates across all three applications, with Lazada obtaining 751 positive reviews and 249 negative reviews, Blibli obtaining 676 positive reviews and 324 negative reviews, and Tokopedia obtaining 559 positive reviews and 441 negative reviews out of 1,000 reviews for each application. This comparison indicates that user satisfaction with Lazada and Blibli is relatively higher, while Tokopedia shows a more balanced proportion between positive and negative sentiments. Furthermore, the evaluation of model performance using Accuracy, Precision, Recall, and F1-Score demonstrates that the Naïve Bayes algorithm performs well in classifying

Indonesian-language text, with Blibli achieving the best overall model performance. These findings confirm that sentiment analysis based on user reviews can provide valuable insights for improving service quality and enhancing the overall performance of e-commerce applications.

## BIBLIOGRAPHY

- [1] A. M. Pravina, K. N. Sani, H. D. Harfianto, A. Pratama, A. Fahrina, and Y. Ruldeviyani, "Sentiment Analysis Of Delivery Service Opinions On Twitter Documents Using K-Nearest Neighbor," vol. 9, no. 2, pp. 996–1012, 2022.
- [2] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [3] E. Rahman, L. Anas, U. Pertiwi, U. Pertiwi, and U. M. Makassar, "SENTIMENT ANALYSIS USING NAIVE BAYES ALGORITHM CASE STUDY," vol. XI, no. 3, pp. 457–462, 2025.
- [4] A. Erfina and M. F. Al-shufi, "ANALISIS SENTIMEN APLIKASI JASA KURIR DI PLAY STORE," vol. 5, no. 2, pp. 103–110, 2022.
- [5] R. Hardian, L. D. Oktaviana, A. Hamdi, U. A. Purwokerto, and U. A. Purwokerto, "Sentiment analysis of pegipegi.com on google playstore with naive bayes algorithm," vol. X, no. 3, 2024.
- [6] E. Jurnal, "Analisis Sentimen Ulasan E-Commerce Menggunakan Metode SVM," vol.

- 18, no. 2, pp. 164–175, 2025.
- [7] A. Erfina and R. A. Lestari, “Analisis Sentimen terhadap Kendaraan Listrik menggunakan Algoritma Naïve Bayes Sentiment Analysis of Electric Vehicles using the Naïve Bayes Algorithm,” vol. 12, pp. 178–185, 2023.
- [8] M. I. Putri and I. Kharisudin, “Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 759–766, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [9] I. H. Kusuma and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [10] B. Z. Ramadhan, R. I. Adam, and I. Maulana, “Analisis Sentimen Ulasan pada Aplikasi E-Commerce dengan Menggunakan Algoritma Naïve Bayes,” *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 220–225, 2022, doi: 10.30871/jaic.v6i2.4725.
- [11] S. N. Fadhilah and F. S. Utomo, “Algoritma Naïve Bayes untuk Analisis Sentiment ReviewBibli.com di Google Play Store,” *Sist. J. Sist. Inf.*, vol. 13, no. 2, pp. 834–4, 2024, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [12] B. Ramadhani and R. R. Suryono, “Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.
- [13] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes,” *J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.
- [14] H. Utami, “Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, 2022, doi: 10.13057/ijas.v5i1.56825.
- [15] J. Supriyanto, D. Alita, and A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.