Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

Sentiment Analysis of Shopee App User Reviews Based on Naïve Bayes Classifier

Atika Andrian Asmiran*[1)], Azzka Farikatun Nassa [2)], Ayuni Layinah[3)], Warto[4)]
1. Informatika, Dakwah dan Saintek, UIN Prof. K.H. Saifuddin Zuhri, Indonesia
2. Informatika, Dakwah dan Saintek, UIN Prof. K.H. Saifuddin Zuhri, Indonesia
3. Informatika, Dakwah dan Saintek, UIN Prof. K.H. Saifuddin Zuhri, Indonesia
4. Informatika, Dakwah dan Saintek, UIN Prof. K.H. Saifuddin Zuhri, Indonesia

Article Info
Keywords: Sentiment analysis, Shopee, Naïve Bayes, e-commerce

Article history:
Received 12 Agustus 2025
Revised 25 Agustus 2025
Accepted 27 Agustus 2025
Available online 01 November 2025

DOI :
10.48144/suryainformatika.v15i2.2181

* Corresponding author.
Atika Andrian Asmiran
E-mail address:
atikaandrian2@gmail.com

ABSTRACT
Technological developments have made various activities easier, including shopping. Through applications, users can make transactions without having to meet face-to-face with sellers. Although it provides convenience, this service does not always run perfectly. One example of a well-known e-commerce application in Indonesia is Shopee. This study analyzes the Shopee application because it significantly supports the Indonesian economy through digitizing trade. Data from 3,000 Shopee reviews obtained from the Google Play Store website in 2025 were analyzed using Natural Language Processing with the Naïve Bayes method, classifying reviews into positive and negative sentiment categories. From the labeling results, text from each sentiment category was extracted to generate essential and valuable information for decision-making. The sentiment classification using the Naïve Bayes Classifier achieved an accuracy rate of 81.30%. The findings of this study can serve as a foundation for enhancing customer satisfaction and comfort, as well as improving the quality of Shopee's products and services, in line with consumer expectations.

1. INTRODUCTION

Information and communication technology development has rapidly impacted various aspects of life, including the digital commerce or e-commerce sector in Indonesia and Southeast Asia. Shopee has become Indonesia's most dominant e-commerce platform, even leading the Southeast Asian market with a 46.9% market share in 2024 [1]. This dominance reflects the high level of interest among the public in using Shopee compared to competitors like Tokopedia and Lazada. However, a high number of users does not always indicate customer satisfaction. Complaints, such as delivery delays, product quality, and customer service issues, frequently appear in user reviews on the Google Play Store. Therefore, it is crucial to analyze user satisfaction to understand and improve service quality and identify areas for improvement. By collecting and analyzing the sentiment from the reviews that have been written, we can obtain reliable data on the factors most influential to user satisfaction and provide an overview of the social and economic impact of this platform. This situation raises an important question: to what extent does user sentiment toward Shopee reflect their satisfaction, and how can sentiment analysis help understand customer perceptions more deeply.

The Natural Language Processing (NLP) method has been widely used in sentiment analysis to automatically process and understand text [2]. Natural Language Processing (NLP) technology in sentiment analysis enables the efficient and accurate processing of large amounts of data to understand user opinions, emotional responses, and attitudes toward various products and services provided by e-commerce platforms. This study uses a dataset of 2,696 user reviews of the Shopee app, comprising 1,349 positive and 1,347 negative reviews. This method can identify hidden patterns in user reviews, pinpoint aspects of the service that need improvement, and provide a comprehensive overview of users' perceptions of the Shopee app [3]. Based on the principle that the Naïve Bayes Classifier has conditional independence regarding the output it generates. One of the main advantages of using the Naïve Bayes Classifier method is its efficiency in terms of

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

training data requirements, as this algorithm can produce accurate classification parameter estimates even with a limited dataset [4]. In a textual context, using the Python programming language is the primary choice for implementing the entire text mining process and stages, from pre-processing to sentiment classification using the Naïve Bayes Classifier method.

This study aims to explain in depth and systematically how this sentiment analysis process is carried out, from collecting Shopee user reviews on the Google Play Store platform, the training stage, and testing using the Naïve Bayes Classifier, to create meaningful data. As such, the results of this study are expected to make an essential contribution to the development of e-commerce platforms in Indonesia, particularly in efforts to enhance user experience and customer satisfaction

2. STUDY LITERATUR

According to Muktafin, E.H., analyzing product reviews is essential. This study uses the KNN and TF-IDF algorithms with an NLP approach to classify reviews of "instant hijabs" into two categories (positive and negative). Classification using the NLP approach achieved an accuracy of 76.92%, while without NLP, it only achieved an accuracy of 69.23%.[5]

Furthermore, according to Sihombing, L.O., identifying problems that arise from customer reviews is essential. However, reading and classifying each review takes a long time. Because of this, we analyzed customer sentiment toward the Xiaomi Redmi Note 9 product sold on the Shopee website using the naïve bayes algorithm and found an accuracy rate of 85%.[6]

Meanwhile, Idris, I. S. K. conducted a sentiment analysis on the Shopee application using the Support Vector Machine (SVM) algorithm. The aim was to classify data from comments made by Shopee application users into positive and negative comments. In this study, 3,000 reviews were collected, and the SVM results showed a high accuracy rate of 98%.[7]

Another study by Utami, H. used the Recurrent Neural Network (RNN) method to analyze user sentiment on the Shopee app based on review data. The results showed that approximately 80% of Shopee app users had positive sentiments and 20% had negative sentiments, indicating that the data was unbalanced. In this study, a pre-processing process was conducted using a combination of the synthetic minority oversampling technique (SMOTE) and the Tomek link method to address this condition. The performance achieved was quite good, with an accuracy of 80%, which is better than sentiment analysis without pre-processing to handle imbalanced data.[8]

3. RESEARCH METHODOLOGY

This study uses a quantitative approach with descriptive and associative methods. Based on the sentiment of reviews written by users on the Google Play Store platform, this study aims to analyze how satisfied Shopee app users are [9]. Sentiment analysis uses a Natural Language Processing (NLP)-based approach with the Naïve Bayes Classifier method. Several stages in this study include data collection, labeling, pre-processing, splitting, modeling, and model evaluation. This section aims to provide clear transparency about the stages taken in this study so that readers can understand the results.

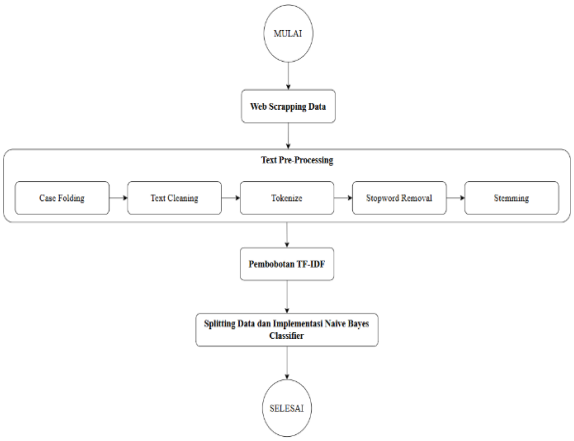

Figure 1. Research Methodology Stages

3.1. *Scrapping Data*

Collecting user reviews from the Google Play Store using web scraping techniques. Web scraping is one of the processes for collecting data from user reviews on the Google Play Store using the Python programming language [10]. Web scraping is performed using the Visual Studio Code code editor and the Python programming language obtained from the Shopee app URL on the Google Play Store, with data collected from 3,000 reviews in 2025. By installing the Google-Play-Scraper library, the existing review data is then extracted for analysis. The results of the scraped data are then exported in CSV format.

Table 1. Example of Scraping Results Document

Dokumen	Ulasan	Score	Label
Dok. 1	Belanja di shopee itu hemat dengan harga, tapi kualitas bgs tapi ga murahan	4	Positif
Dok. 2	Bagus, pengiriman cepat, jaga mutu barang	4	Positif
Dok. 3	Setiap buka shopee pasti muncul halaman tidak tersedia	1	Negatif
Dok. 4	Tolong perbaiiki lagi aplikasinya	1	Negatif
Dok. 5	Mudah aplikasinya, tdk membingungkan, banyak diskon dan cashback	5	Positif
Dok. 6	Selama ada shopee kurang lebih 4 tahun sangat	5	Positif

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

	meringankan ssaya dalam berbelanja		
Dok. 7	Pokoknya happy shopping dech kalo di shopee	5	Positif
Dok. 8	Saya penjual non star yang jarang daper orderan	2	Negatif
...	...	...	...
Dok. x	Terima kasih shopee. Sudah memudahkan berbelanja yang aman	5	Positif

3.2. *EDA (Exploratory Data Analysis)*

Exploratory Data Analysis (EDA) is essential in understanding the characteristics and helping to analyze a dataset. EDA helps identify patterns, relationships, anomalies, understand the structure, characteristics, and distribution of data, and detect potential problems such as duplicate data, missing data, class imbalance, or outliers [11]. The primary purpose of EDA is to gain initial insights into the data before running more complex statistical or machine learning models. EDA standard methods include raw data collection, data processing, data cleaning, and feature visualization [12].

3.3. Labeling data

Labeling is used to assign sentiment categories or classes to each piece of text data, such as "positive," "negative," or "neutral," which serve as variables to be predicted by the model. We labeled each text in the "score" column in this dataset. The labels assigned are Positive and Negative. Out of the 3,000 data points collected, if the 'score' column is 5 or 4, it is classified as Positive (1). If the score is below 3, it is classified as Negative (0). If the value is 3, it is removed. Based on the established criteria, the following results were obtained:

- 1349 reviews classified as Positive sentiment (1)
- 1347 reviews classified as Negative sentiment (0)

It can be seen that the ratio of positive to negative tweets is not significantly different [13].

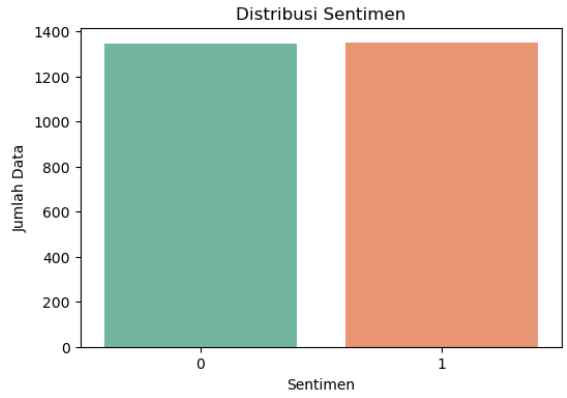

Figure 2. Number of Positive and Negative Sentiments

3.4. Text Pre-Processing

Using the Naïve Bayes Classifier approach, the text pre-processing stage is used in the text processing stage. The text pre-processing process consists of several stages, including:

a. Case Folding
Case folding is a step to convert all capital letters in the text to lowercase letters so that the text is uniform and the data is consistent, thereby avoiding differences due to varying letter writing [14]. At this stage, there is also a character check from the beginning to the end of the data. Characters other than letters and numbers, such as punctuation marks and spaces, are removed because they are considered delimiters (separators). The results of the case folding process can be seen in Table 2.

Table 2. Case Folding Process

belanja di shopee itu hemat dengan harga, tapi kualitas bgs tapi ga murahan

b. Text Cleaning
Text cleaning is the process of cleaning text data prepared for further analysis. Review data will be cleaned, simplified, and standardized at this stage so that the text data becomes more structured. The steps involved in text cleaning include removing characters, numbers, URLs, excess spaces, and emojis from unnecessary or irrelevant data. This is necessary to improve data quality and make it easier for machine learning to read, use, and analyze the data. The results of the text cleaning process can be seen in Table 3.

Table 3. Text Cleaning Process

belanja di shopee itu hemat dengan harga tapi kualitas bgs tapi ga murahan

c. Normalisasi
Normalization involves changing non-standard words to standard words and abbreviations to their original words [15]. For example: "ga", "gue", and others, as well as the frequent use of word fragments, such as: "yg", "brp", 'bgm', "bgs", and others. Words that are not normalized will be recognized as different words. For example, "bgs" and "bagus," which should have the same meaning, will have different meanings because they are written differently. For this reason, words are normalized from non-standard to standard [16].

Table 4. Normalization Process

belanja di shopee itu hemat dengan harga tapi kualitas bagus tapi tidak murahan

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

d. Tokenize
Tokenization is breaking down text or sentences into smaller units called tokens. In this context, tokens can be words, phrases, or characters in user reviews for further analysis using Natural Language Processing (NLP) approaches. The tokenization process involves breaking down and separating text into individual units based on spaces or punctuation marks. By performing tokenization, machine learning algorithms can convert text into a format easier to process. The results of the tokenization process can be seen in Table 5.

Table 5. Tokenization Process

['belanja', 'di' 'shopee', 'itu', 'hemat', 'dengan', 'harga', 'tapi, 'kualitas','bagus', 'tapi', 'tidak', 'murahan']

e. Stopword Removal
Stopword removal at this stage removes common words with irrelevant meanings in the sentiment analysis of Shopee user reviews. These familiar words usually consist of conjunctions, pronouns, or words that do not contribute significantly to a sentence or text. Examples include words like "in," "that," "with," and "but." This process enhances the accuracy of the Naïve Bayes Classifier approach in machine learning. The results of the stopword removal process can be seen in Table 6.

Tabel 6. Penghapusan Kata Henti

['belanja','shopee','hemat','harga', 'kualitas','bagus','murahan']

f. Stemming
Stemming in this stage is converting affixed words into root words or stems. This stage is carried out to reduce the variation of words with the same meaning and simplify them, thereby improving the efficiency of data analysis in user reviews and making them easier to process in sentiment analysis based on Natural Language Processing (NLP). The results of the stemming process can be seen in Table 6 [17].

Table 7. Proses Stemming

['belanja','shopee','hemat','harga', 'kualitas','bagus','murah]

3.5. Spliting Data

The stages of data splitting in the classification process, using the Hold-out method, involve dividing the dataset into two parts: training and testing data. In the processed data, 80% was obtained for training and 20% for testing. In this dataset division process, the Python programming language was used as follows.

Table 8. Data Splitting Process

X_train, X_test, y_train, y_test = train_test_split (df_baru ['stemmed_text'], df_baru ['Label'], test_size = 0.20, random_state = 42, stratify=y

3.6. TF-IDF Weighting

TF-IDF aims to convert user review data text into numerical form, so that the processed data can be used as input for machine learning, especially in the Naïve Bayes Classifier approach. In the TF-IDF process, each word in the review is weighted based on two main factors: how often the word appears in the review (Term Frequency) and how rarely the word appears in the entire review data (Inverse Document Frequency).

a. Term Frequency Calculation (TF)
In this calculation, term frequency measures how often a word appears in a review, normalized by dividing the number of occurrences by the total number of words in the review.

$$TF(t, d) = \frac{ft}{N}$$

Description:
- Ft = number of occurrences of word (t) in document (d)
- N = Total Number of words in the document (d)

With the example of TF value calculation in document review-1 as follows: ['shopping', 'shopee', 'savings', 'price', 'quality', 'good', 'cheap']. If a calculation is performed from this data, the result will be shown in Table 9.

Table 9. TF calculation results

Kata	Ft (Jumlah kemunculan)	TF=$\frac{ft}{N}$
belanja	1	$\frac{1}{7} = 0.143$
shopee	1	$\frac{1}{7} = 0.143$
hemat	1	$\frac{1}{7} = 0.143$
harga	1	$\frac{1}{7} = 0.143$
kualitas	1	$\frac{1}{7} = 0.143$
bagus	1	$\frac{1}{7} = 0.143$
murah	1	$\frac{1}{7} = 0.143$

All words in document-1 review have the same TF result because they have the same number, which is one each in 7 words.

b. Inverse Document Frequency Calculation (IDF)
In this process, IDF is used to calculate reviews to determine how rarely a word appears in the existing data [18].

$$IDF\ (t) = \log \frac{D}{dt}$$

Description:
- t = certain words in a document
- D = total number of documents in a document collection

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

- Dt = number of documents containing the word (t)

An example of the IDF equation calculation application in document review-1 is ['shopping', 'shopee', 'economical', 'price', 'quality', 'good', 'cheap']. The number of words appearing in a single dataset has been obtained through the coding process in Table 10.

Table 10. Frequency of Words Appearing

Kata	Frekuensi
belanja	347
shopee	746
hemat	39
harga	138
kualitas	23
bagus	202
murah	80

Table 10 shows the frequency or number of words in document-1. Thus, the IDF results for the reviews in document-1 can be seen in Table 11.

Table 11. IDF Results

Kata	dt	IDF(t) = $\log(\frac{1325}{dt})$
belanja	347	$\log(\frac{1325}{347})$=0.58
shopee	746	$\log(\frac{1325}{746})$=0.25
hemat	39	$\log(\frac{1325}{39})$=1.53
harga	138	$\log(\frac{1325}{138})$=0.98
kualitas	23	$\log(\frac{1325}{23})$=1.76
bagus	202	$\log(\frac{1325}{202})$=0.82
murah	80	$\log(\frac{1325}{80})$=1.22

Table 11 shows that the calculation of the word "Shopee" has the lowest IDF with a result of 0.25 and a word frequency of 746 in the data, so that word is considered a common word and does not provide vital information in this sentiment analysis. Meanwhile, the word "quality" has the highest IDF value of 1.76, with 23 occurrences across the entire dataset. This value indicates that the word "quality" significantly influences the sentiment analysis. On the other hand, the words "price," "cheap," and "good" have moderate IDF values ranging from 0.82 to 1.22. These words appear frequently but still hold relevant value in the sentiment analysis.

c. TF-IDF calculation

Based on the results of the Term Frequency (TF) and Inverse Document Frequency (IDF) calculations in Tables 7 and 9, TF-IDF calculations can be performed to determine the weight of each word in the Shopee app user review data on the Google Play Store. By calculating the TF and IDF values and then multiplying them, the TF-IDF calculation can assign a higher weight to words that appear infrequently but have a high value for distinguishing sentiment analysis. The following is the TF-IDF calculation shown in Table 12.

Table 12. TF-IDF results

Kata	TF	IDF	TF-IDF
Belanja	0.143	0.58	0.083
Shopee	0.143	0.25	0.036
Hemat	0.143	1.53	0.219
Harga	0.143	0.98	0.140
Kualitas	0.143	1.76	0.252
bagus	0.143	0.82	0.117
Murah	0.143	1.22	0.174

Based on the TF-IDF calculation results in Table 12, it is known that the word "quality" has the highest value of 0.252. This indicates that the word is rarely used and appears, but it has a high weight in distinguishing the sentiment of Shopee app user reviews. Thus, it can be interpreted that the quality of products and services provided on the Shopee app is crucial to users. Words with the highest TF-IDF weights are more critical in distinguishing between positive and negative sentiments in user reviews. This enables the Naïve Bayes classifier to focus on words that significantly impact the interpretation of user reviews on the Shopee app.

3.7. Naïve Bayes Model Training

The Naïve Bayes classification algorithm is a classification technique that utilizes probability and statistical methods. As a supervised learning method, Naïve Bayes Multinomial requires data labeling before training. The equation can be seen as follows:

$$P(x|y) = \frac{P(x|y)\,.P(y)}{P(x)}$$

- $P(x|y)$: Calculate the probability of the appearance of a set of feature values x when the target label value y is known. Then multiply that value by $P(y)$.
- $P(y)$: probability of target label value y. Next, that value is divided by $P(x)$.
- $P(x)$: the probability value of a set of feature values x [13].

3.8. Prediction & Evaluasi

This test was conducted to determine the results of the model created and assess the model's ability to classify sentiment based on input text automatically. The prediction results of the model were then evaluated using several performance metrics, namely the confusion matrix, accuracy, precision, recall, and F1-score [19].

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 15, No. 2. November 2025, Pp. 91-98

```
Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.86      0.82       270
           1       0.84      0.77      0.80       270

    accuracy                           0.81       540
   macro avg       0.82      0.81      0.81       540
weighted avg       0.82      0.81      0.81       540

Accuracy: 81.30%
```

Figure 3. Evaluation Metrics

To evaluate the model that has been created, the confusion matrix and cross-validation score methods are used. The confusion matrix is used to show the number of correct and incorrect classification results from the model. Meanwhile, the cross-validation score is used to measure how well the model performs in predicting data that has not been seen before [20].

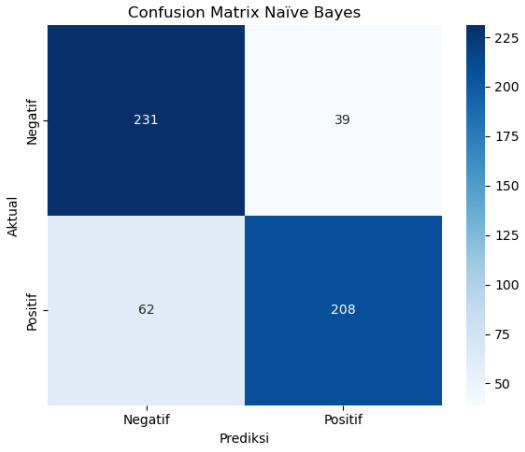

Figure 4. Confusion Matrix

In addition to using accuracy, precision, recall, and F1-score as evaluation metrics, this study also utilizes AUC–ROC (Area Under the Curve-Receiver Operating Characteristic) as an additional indicator to assess the performance of binary classification models.

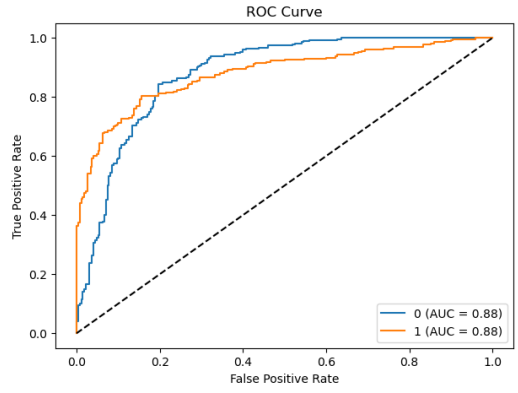

Figure 5. ROC Curve

The ROC curve above shows the classification performance of the Naïve Bayes model in distinguishing between two sentiment labels (positive and negative). With an AUC value of 0.88 for both classes, both curves are well above the diagonal line, indicating performance that is far better than random guessing

4. RESULTS AND DISCUSSION

In this study, sentiment analysis was conducted on user reviews of the Shopee application using the Naïve Bayes Classifier method based on Natural Language Processing (NLP). The purpose of this study was to automatically classify user reviews into two sentiment categories, namely positive and negative. The selection of Naïve Bayes was based on its advantages as a simple algorithm, fast in the training process, and efficient in the use of computational resources. Additionally, this method has proven effective in processing text data with TF-IDF representation and is often used as a baseline in sentiment analysis research. This process is expected to provide a clear picture of consumer satisfaction with the Shopee app. The results of this sentiment analysis are not only beneficial for users in making purchasing decisions but are also important for sellers and app developers to evaluate and improve product and service quality based on user feedback.

The data used in this study was collected through web scraping from the Google Play Store platform, with 3,000 reviews of 1,349 positive and 1,347 negative reviews. The review data was then processed through labeling and pre-processing, such as case folding, text cleaning, normalization, tokenization, stopword removal, and stemming, before being converted into numerical form using the TF-IDF method. This process aimed to prepare the data for use in machine learning algorithms. The model was then built and implemented using the Python programming language, with data split using the Hold-out technique, allocating 80% for training data and 20% for testing data.

The results of the model performance evaluation show that the Naïve Bayes method can produce an accuracy of 81.30%, precision of 84.21%, recall of 77.04%, and an F1-score of 80.46%. This indicates that the model has a pretty good ability to classify user reviews correctly.

However, several challenges were also identified during this analysis, namely that improving accuracy was difficult despite several optimizations. This is thought to be due to the model's limitations in understanding complex linguistic contexts and the ambiguity of the review data [21].

This study demonstrates that applying the Naïve Bayes Classifier method in analyzing user sentiment in Shopee app reviews can provide accurate and helpful information. These results are expected to be a foundation for developing e-commerce services that are more responsive to user needs and expectations.

5. CONCLUSION AND SUGGESTION

Based on the results of the analysis attached, it can be concluded that the application of the Naïve Bayes Classifier method in analyzing user sentiment reviews on the Shopee app is capable of classifying reviews into two sentiment categories, namely positive and negative, out of 3,000 Shopee user reviews, comprising 1,349 positive reviews and 1,347 negative reviews, achieving an accuracy rate of 81.30%. This analysis process was supported by Natural Language Processing (NLP) techniques through text pre-processing, TF-IDF weighting, and data separation using the Hold-out method. These findings are expected to serve as a basis for evaluation by developers and sellers in improving services and product quality on the Shopee platform.

6. ACKNOWLEDGEMENTS

The author would like to express his deepest gratitude to all those who have supported and assisted in preparing this scientific paper. Thank you to the supervising lecturer and teammates who have worked well together. In the context of e-commerce, user satisfaction is one of the most crucial aspects to ensure service quality. Therefore, reviews on platforms such as the Google Play Store are considered important indicators, as they directly reflect user experiences and opinions. By analyzing these reviews through sentiment analysis, this work is expected to provide benefits and contributions to the development of science, particularly in the areas of sentiment analysis and e-commerce technology.

REFERENCES

[1] Affifa, “Melihat Pangsa Pasar E-commerce Asia Tenggara, dengan Shopee Memimpin,” CUBE. [Online]. Available: https://cube.asia/read/ecommerce-market-share-shopee-leads-the-pack/

[2] O. N. Julianti, N. Suarna, and W. Prihartono, “Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online Di Media Sosial Twitter,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2936–2941, 2024, doi: 10.36040/jati.v8i3.9613.

[3] V. No, M. D. Ikhsan, B. Huda, A. Hananto, F. Nurapriani, and V. No, “Infotek : Jurnal Informatika dan Teknologi Analisis Sentimen Ulasan Pengguna Alikasi Traveloka Pada Google Play Store Menggunakan Algoritma Naive Bayes Infotek : Jurnal Informatika dan Teknologi Dengan pesatnya kemajuan internet dan teknologi , aplikasi p,” vol. 8, no. 2, pp. 367–377, 2025.

[4] M. A.-Z. Faradeya and E. R. Subhiyakto, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naïve Bayes," *J. Algoritm.*, vol. 22, no. 1, pp. 115–127, 2025, doi: 10.33364/algoritma/v.22-1.2178.

[5] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.

[6] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.

[7] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[8] 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.

[9] N. Huda, F. Habrizons, A. Satriawan, M. Iranda, and T. Pramuda, "Analisis Usability Testing Menggunakan Metode SUS (System Usability Scale) Terhadap Kepuasan Pengguna Aplikasi Shopee," *Simkom*, vol. 8, no. 2, pp. 208–220, 2023, doi: 10.51717/simkom.v8i2.158.

[10] N. Nurzaman, N. Suarna, and W. Prihartono, “Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.

[11] M. Rizqi, A. Rustiawan, and P. T. Prasetyaningrum, “Analisis Sentimen Terhadap Klinik Natasha Skincare di Yogyakarta Dengan Metode Google Review,” *J. Inf. Technol. Ampera*, vol. 5, no. 1, pp. 2774–2121, 2024, doi: 10.51519/journalita.v5i1.556.

[12] R. Merdiansah, S. Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[13] D. F. Rahman, “Analisis ChatGPT tweet menggunakan EDA dan sentiment analysis: Data pengguna Twitter di Indonesia,” *no. Sept.*, no. September, 2023, [Online]. Available: https://www.researchgate.net/profile/Dzul-Rahman-4/publication/376356578_Analisis_Chatgpt_Tweet_Menggunakan_Eda_Dan_Sentiment_An

alysis_Data_Pengguna_Twitter_Di_Indonesia/links/6573ec0ccbd2c535ea0c1c4c/Analisis-Chatgpt-Tweet-Menggunakan-Eda-Dan-Sentiment-A

[14] T. Atha Anastasya, A. Diani Putri Saka, M. Juventus Dappa Deke, and A. Mustika Rizki, "Optimasi Algoritma Svm Dengan Pso Untuk Analisis Sentimen Pada Media Sosial X Terhadap Kinerja Timnas Di Era Shin Tae-Yong," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 384–391, 2024, doi: 10.36040/jati.v9i1.12298.

[15] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.

[16] S. Wardani, "Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm," *J. Din. Inform.*, vol. 5, no. November, pp. 1–13, 2015.

[17] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.

[18] G. Noer, "Implementasi Algoritma Naïve Bayes dan TF-IDF Dalam Analisis Sentimen Data Ulasan (Studi Kasus: Ulasan Review Aplikasi E-commerce Shopee di Situs Google …," *Repository.Uinjkt.Ac.Id*, 2023, [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/68747%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/68747/1/GERALD HALIM AL RASYID NOER-FST.pdf

[19] T. M. Iryana, I. Indriati, and P. P. Adikara, "Analisis Sentimen Masyarakat Terhadap Mass Rapid Transit Jakarta Menggunakan Metode Naïve Bayes Dengan Normalisasi Kata," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2753–2760, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9421

[20] 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.

[21] H. Horacek, "Natural language processing," *Comput. Phys. Commun.*, vol. 61, no. 1–2, pp. 76–92, 1990, doi: 10.1016/0010-4655(90)90107-C.