# Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier

**Adelia Nopiazinda Br Panjaitan[1*], Ismail Husein[2]**

[1*,2]Mathematics, Universitas Islam Negeri Sumatera Utara Medan, Tuntungan, Medan, Indonesia

| Article Info | ABSTRACT |
|---|---|
| **Keywords:**<br>Naïve Bayes Classifier, Shope Application, User Reviews, Sentiment Analysis. | Shopee is currently the number one online shopping application in Indonesia. The purpose of the study is to find out the classification of shopee application user reviews and the results of prediction accuracy by classifying naïve bayes classifiers. Positive reviews, negative reviews, total of all reviews, and review status are the variables of the study. Simple random sampling technique to determine the sample, the results were selected by Indonesian user reviews. Scraping the Google Play Store website page generates attribute variable data, which is then determined using rating, date, and review criteria. To create a naïve bayes classifier, the before and after probabilities are calculated using rapid miner software. The results showed that naïve bayes classifier obtained an accuracy score with a 50% training data trial and a 50% data test resulting in an accuracy of 90.53%, a precision of 90.53%, and a recall of 100% with an average accuracy of 90.53%. It can be concluded that naïve bayes classifier can be used in predicting the review rate of Shopee application users effectively and efficiently |
| | **Corresponding Author:**<br>Adelia Nopiazinda Br Panjaitan<br>UIN Sumatera Utara Medan<br>Jl. Lapangan Golf, Medan<br>adelia0703202035@uinsu.ac.id |

## INTRODUCTION

Shopee is one of the most popular applications among the public today because it provides an online service system. Through the live chat feature, sellers and buyers can be involved effectively (A. Oktian Permana & Sudin Saepudin, 2023). So there is no need to make personal contact when buying and selling promoted goods (Ernawati, 2016). Therefore, Shopee must maintain its popularity by providing good service to increase user trust and satisfaction (Pangestu et al., 2024).

The Shopee application is known as an e-commerce platform that seeks to reach all groups by recruiting the best brands (Juwira, 2024). As a result, every year Shopee continues to experience extraordinary progress because it is able to provide comprehensive information and pages for sellers and buyers. Thus, users' feelings, behaviors, and trust in this application grow when making buying and selling transactions (Steven & Ramli, 2023). The following top ten e-commerce applications in Indonesia in the first quarter of 2023 are presented below based on visitor volume.

Shopee has the most visits, namely 71.53 million, beating several other well-known online applications. The number of visits to similar websites in October 2023 was more than

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**317** | P a g e

243 million times, proving that users are more dominant on Shopee to make buying and selling transactions, meet their needs, and desires. The reason is, this application offers interesting features, great products, and affordable prices (Huda et al., 2023). Therefore, in order for Shopee and its customers to remain stable, a breakthrough is needed in looking at opinions, suggestions, and input, one of which is with user reviews (Hafid, 2023).

User reviews as an effective and efficient way to search for and uncover important information. In addition, with existing reviews, each user can easily understand the positive and negative judgments of other users (Saifurridho et al., 2024). Thus, shopee and users can improve their performance in producing and selling goods and services. Therefore, to understand user reviews of a product and service, sentiment analysis is needed (Amalia et al., 2023). Sentiment analysis is one of the useful tools for classifying an object into two positive and negative categories. In addition, this analysis can also show the level of a person's behavior in expressing ideas on certain topics using natural language. So, to see the response of Shopee application users to the incident, it is necessary to conduct a sentiment analysis (Permata & Haryanto, 2022).

But unfortunately, many potential customers still cannot understand other users' reviews because the language used is not standard (Hasto, 2024). As a result, it is difficult for other users to get information about the quality, price, material, color, shape, size, and taste of the goods offered. In fact, positive and negative user reviews are very important for every online application and other users when marketing goods, services, and even buying and selling (Yang & Sihotang, 2023).

Various online sentiment analysis of shopee application user reviews has actually been carried out.  Some of these studies have been carried out with several approaches, such as Machine Learning, Natural Language Processing, and  Long Short Term Memory. However, it seems that most of the analysis carried out is more focused on students. In addition, the increase in the number of impulse purchases is based on price discounts (Andriana et al., 2021). In line with this information, research that examines positive, negative, and neutral user reviews on the increase in the number of purchases on the Shoppe application using the Naïve Bayes Classifier has never been conducted. This kind of research needs to be carried out in order to provide knowledge and understanding to consumers and Shopee users to get the desired satisfaction when making buying and selling transactions (Utami, 2022). The results can also be used as evaluation material in improving the image of Shopee. Therefore, the purpose of this study is to find out the classification of Shopee application user reviews and the results of prediction accuracy by conducting a naïve bayes classifier (Idris et al., 2023).

## METHODS

This research was conducted in the even semester of 2023 at the State Islamic University of North Sumatra Medan. The variables of the research attributes were positive reviews, negative reviews, and total user reviews. This study involves all reviews of English and Indonesian shoppe application users who are in the google play store website database as the research population. Therefore, the simple random sampling technique is used in the sampling process.

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier—
Adelia Nopiazinda Br Panjaitan et.al
**318** | P a g e

As a result, reviews of Indonesian-language shopee application users were selected as a sample of this study with the number of reviews of Indonesian-language shopee application users from January to December 2023 as much as 1.032.126.

The attribute variables of this study are positive reviews, negative reviews, and the total of all reviews of Shopee application users. Each data from the attribute variable is obtained from the scraping results on the Google Play Store website page. Then, the determination of attribute variables in classifying user reviews is carried out by rating, date, and review. After the user review data was obtained, the quality analysis of the naïve bayes classifier classification data was carried out by distinguishing all data into a dataset through a 2 x 2 confusion matrix, as follows:

**Table 1**. Confusion Matrix

| Class | Positive | Negative |
|---|---|---|
| Positive | True Positive (TP) | True Negative (TN) |
| Negative | False Positive (FP) | False Negative (FN) |

Next, calculate the classification performance to generate accuracy, precision, and recall values using the formula:

$$\text{Accuracy} = \frac{TP + FN}{TP+FP+TN+FN} \tag{1}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive+False\ Positive} \tag{2}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive+False\ Negative} \tag{3}$$

In the development of classification data, it is carried out through the calculation of posterior and prior probabilities to obtain each class data using the equation formula:

$$P (H = x \mid C = j) = \frac{1}{\sqrt{2\ phi\ x\ s^2}}\ xe^{-\frac{(x-mean)^2}{2\ x\ s^2}} \tag{4}$$

$$P (A) = \frac{x_A}{n} \tag{5}$$

The result of the naïve Bayes classification with the probability value closest to the lift of 1 is the result of the classification. After that, the data is separated into training and testing data. This research involves descriptive analysis in producing a general image of Shopee application user reviews obtained through the Google Play Store website. Then, sentiment analysis is positioned to label data into sentiment classes using lexicon dictionaries. Meanwhile, the naïve bayes classifier exists to classify positive, negative, and total user reviews using the following equation formula.

$$\text{Mean} = \frac{x_1+x_2+\cdots x_n}{n} \tag{6}$$

$$\text{Standar Deviasi} = \sqrt{\frac{\sum(X- mean)^2}{n-1}} \tag{7}$$

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**319** | P a g e

## RESULTS AND DISCUSSION

In this study, the descriptive analysis of its existence to find out the reviews of Shopee application users by using a table to describe the rating data obtained from January to December 2023. The data was obtained directly from the Google Play Store website, then data cleaning was carried out to eliminate data noise that was inconsistent, irrelevant, and unused. It can be presented as follows.

**Table 2.** Shopee Application User Review

| Moon | Rating | | Total | Category |
|------|--------|--------|-------|----------|
| | Positive | Negative | | Class |
| January | 65.409 | 5.635 | 71.044 | Go down |
| February | 79.707 | 4.067 | 83.774 | Climb |
| March | 81.145 | 5.672 | 86.817 | Climb |
| April | 67.232 | 9.793 | 77.025 | Go down |
| May | 73.402 | 8.378 | 81.780 | Climb |
| June | 72.602 | 9.984 | 82.586 | Climb |
| July | 59.340 | 8.129 | 67.469 | Go down |
| August | 98.403 | 9.140 | 107.543 | Climb |
| September | 76.284 | 8.114 | 84.398 | Go down |
| October | 81.913 | 9.896 | 91.809 | Climb |
| November | 95.232 | 10.948 | 106.180 | Climb |
| December | 83.753 | 7.948 | 91.701 | Go down |
| Total | 934.422 | 97.704 | 1.032.126 | |

Based on table 2 above, it is informed that the data obtained is consistent because the three data groups are taken entirely without cleaning, so that the number of attributes and records in the data group is fixed and clean to use.

**Table 3.** Data Training and Testing

| Data | Total |
|------|-------|
| Training | 9 |
| Testing | 3 |
| Total | 12 |

Based on table 3 above, it shows that the total data used is 12 with a composition of 80% for training data amounting to 9 and testing data amounting to 3. Then, the 3 existing testing data will be tested to get classification results.

**Table 4.** Data Testing

| Month | Rating | | |
|-------|--------|--------|-------|
| | Positive | Negative | Total |
| October | 81.913 | 9.896 | 91.809 |
| November | 95.232 | 10.948 | 106.180 |
| December | 83.753 | 7.948 | 91.701 |

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**320** | P a g e

Based on table 4 above, it shows that the classification of shoppe application user review data uses a data set of 3. Furthermore, the mean value and standard deviation will be searched as parameters for positive reviews, negative reviews, and total reviews.


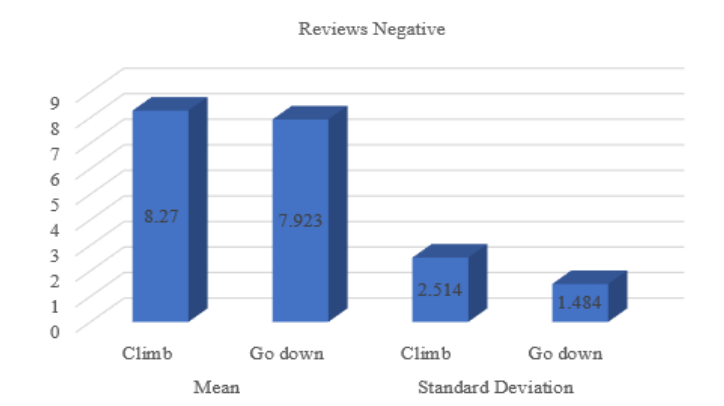
**Figure 1.** Graphs Positive Reviews
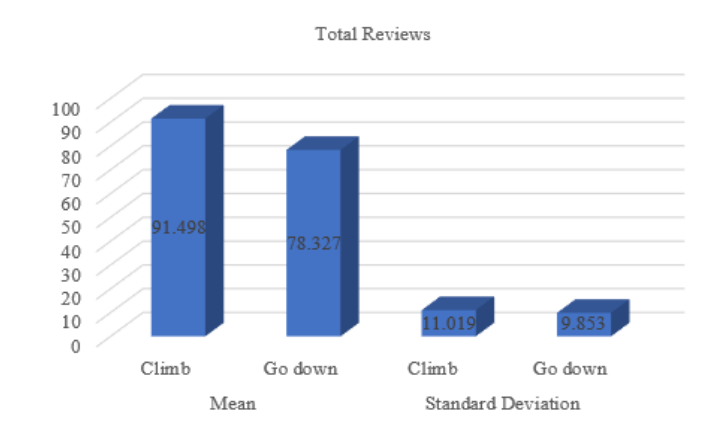


**Figure 2.** Graphs Negative Reviews



**Figure 3.** Graphs Total Reviews

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**321** | P a g e

Table 5. Naïve Bayes Classification Results

| Moon | Rating | | Total | Category Classification |
|---|---|---|---|---|
| | Positive Reviews | Negative Reviews | Reviews | |
| October | 81.913 | 9.896 | 91.809 | Climb |
| November | 95.232 | 10.948 | 106.180 | Climb |
| December | 83.753 | 7.948 | 91.701 | Climb |

Table 6. Prediction Accuracy Value

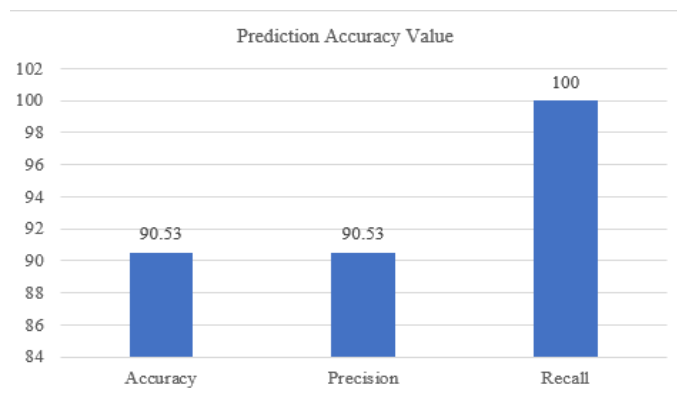| Actual Data | Prediction Data | |
|---|---|---|
| | Positive | Negative |
| + | 467.211(TP) | 0 (FN) |
| - | 48.852 (FP) | 2 (TN) |
| Accuracy | 90,53% | |



Figure 4. Graphs Prediction Accuracy Value

Based on table 5,6 and figure 5 above, it shows that the tests carried out obtained an accuracy value with a 50% training data trial and a 50% data testing test. Thus, it produces an accuracy of 90.53%, precision of 90.53%, and recall of 100% with an average accuracy of 90.53%.

## CONCLUSION

This study shows that naïve bayes classifiers are able to produce strong predictions, this is evidenced by the level of reviews of Shopee application users for October 2023 with 81,913 positive reviews, 9,896 negative reviews and a total number of 91,809 reviews. Then, in November 2023 with 95,232 positive reviews, 10,948 negative reviews, and a total number of 106,180 reviews. Meanwhile, in December 2023 with 83,753 positive reviews, 7,948 negative reviews, and a total of 91,701 reviews. The results obtained are similar to which state that this classification system is very possible to be further developed according to the increasing needs so that it can improve the accuracy of the system. Then, the results obtained from stated that the performance of the classification system was able to produce better accuracy. Tests conducted with naïve bayes classifier also showed that in October 2023 the

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**322** | P a g e

final value of probability increased by 0.8070762875 and the final value of probability decreased by 0.1929237125. Then, in November 2023, the final value of probability increased by 0.9930399667 and the final value of probability decreased by 0.00696002996. Meanwhile, in December 2023, the final value of probability increased by 0.6333040371 and the final value of probability decreased by 0.36669596239. Thus, the accuracy score for October, November, and December 2023 was 90.53%. This result is similar to which states that the classification process with the right method will produce the highest level of accuracy, which is 91%. Meanwhile, the results conducted by stated that the classification test obtained an accuracy level of 86.81% in the 7-fold cross validation test for the Sigmoid kernel type. The results achieved are quite satisfactory according to the researcher's wishes, but not without shortcomings. For example, selecting only a few attribute variables and using a specific method can narrow down the findings to a more effective and efficient level of accuracy. Therefore, in the future, we need to increase the number of attribute variables and compare the methods used to determine which method is more accurate.

## REFERENCE

A. Oktian Permana, & Sudin Saepudin. (2023). Perbandingan Algoritma K-Nearst Neighbor dan Naïve Bayes Pada Aplikasi Shopee. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(1), 25–32.

Amalia, A., Fathiyya Nuha, S., Zikrinawati, K., & Fahmy, Z. (2023). Pengaruh Brand Ambassador dan Persepsi Kualitas Produk terhadap Minat Pembelian Mahasiswa Pengguna Aplikasi Shopee. *Jurnal Mahasiswa Kreatif*, *1*(2), 41–52.

Andriana, A. N., Aisha, Z., Kusuma, A. N., & ... (2021). Pelatihan Penggunaan Aplikasi Shopee untuk Meningkatkan Minat Berbisnis Secara Online. *JMM (Jurnal ...*, *5*(6), 4–12.

Ernawati, S. (2016). *Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naive Bayes*. *66*, 37–39.

Hafid, H. (2023). Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia. *Journal of Mathematics*, *6*(2), 161–168.

Hasto, N. C. (2024). Pengaruh Diskon, Brand Image, dan Kemudahan Akses terhadap Keputusan Pembelian Mahasiswa Pengguna Aplikasi Shopee Melalui Minat Beli Sebagai Variabel Intervening. *Reslaj: Religion Education Social Laa Roiba Journal*, *6*(4), 1423–1441.

Huda, N., Habrizons, F., Satriawan, A., Iranda, M., & Pramuda, T. (2023). Analisis Usability Testing Menggunakan Metode SUS (System Usability Scale) Terhadap Kepuasan Pengguna Aplikasi Shopee. *Simkom*, *8*(2), 208–220.

Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, *5*(1), 32–35.

Juwira, R. S. T. (2024). Analisis Tingkat Kepuasaan Pengguna Aplikasi Shopee Dikalangan GEN Z Mengunakan Regresi Sederhana. *J-ENSISTEC (Journal of Engineering and*

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**323** | P a g e

*Sustainable Technology)*, *10*(2), 10101–10106.

Pangestu, A., Tajul Arifin, Y., & Ade Safitri, R. (2024). Analisis Sentimen Review Publik Pengguna Game Online Pada Platform Steam Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3106–3113.

Permata, S., & Haryanto, H. (2022). Perlindungan Hukum Terhadap Pengguna Aplikasi Shopee Pay Later. *Krisna Law : Jurnal Mahasiswa Fakultas Hukum Universitas Krisnadwipayana*, *4*(1), 33–47.

Saifurridho, M., Martanto, M., & Hayati, U. (2024). Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi Shopee. *Jurnal Informatika Terpadu*, *10*(1), 21–26.

Steven, J., & Ramli, A. H. (2023). E-Service Quality, E-Wallet Dan Kepercayaan Terhadap Minat Beli Pada Pengguna Pembayaran Non Tunai Aplikasi Shopee. *Jurnal Ilmiah Manajemen Kesatuan*, *11*(2), 267–278.

Utami, H. (2022). Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network. *Indonesian Journal of Applied Statistics*, *5*(1), 31.

Yang, M. Z., & Sihotang, J. I. (2023). Analisis Kepuasan Pengguna Terhadap User Interface Aplikasi E-Commerce Shopee Menggunakan Metode EUCS di Jakarta Barat. *Informatics and Digital Expert (INDEX)*, *4*(2), 53–60.

Sentiment Analysis Of Shopee Application User Reviews Using Naïve Bayes Classifier–
Adelia Nopiazinda Br Panjaitan et.al
**324** | P a g e