A Comparison of E-wallet Applications Opinion Based on Sentiment Analysis

Susi Angelina[1*], Jimmy Tjen[2]
[1,2] Department of Informatics, Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak, Indonesia.
*Email: [1]22421586_susiangelina@widyadharma.ac.id, [2]jimmy.tjen@mathmods.eu

Abstract. Nowadays, e-wallets have become a popular payment tool that can be used for various types of transactions. The rapid growth of e-wallets today is closely related to advancements in digital technology, like smartphones. Application X and Y are frequently used by people in Indonesia. With so many reviews about these e-wallets, we can notice that there are many positive and negative reviews. These reviews can be used for sentiment analysis of the applications and classified into positive or negative categories. 1,000 reviews collected from Google Play Store were divided into two sentiment classes: positive sentiment or negative sentiment. This research uses the Naïve Bayes classification method. The algorithm produces an accuracy of 91.87% for the testing model, 86% accuracy for sentiment analysis of Application X and 98% accuracy for sentiment analysis of Application Y. The results indicates that Application X is outperforms than the other one.

Keywords: E-Wallet, Naïve Bayes Algorithm, Sentiment Analysis.

Introduction

Nowadays, digital wallets (E-wallets) have become a popular payment tool, especially among certain communities. E-wallets can be used for various types of transactions, such as online shopping, product payments, food delivery and more (Ningri, Hamidi, & Adrianto, 2023). The rapid growth of e-wallets today is closely related to advancements in digital technology, particularly the development of devices like smartphones, which have made it easier to access and use these services in everyday life (Kathiravan, et al., 2021). E-wallets applications X and Y are widely used by the Indonesian people and have provided various positive or negative impacts.

This research aims to implement the Naïve Bayes method for text classifications. Naïve Bayes is a popular algorithm that used for detecting sentiment (Hayatin, Marthasari, & Nuarini, 2020). Also, Naïve Bayes is recognized as a simple algorithm that delivers high accuracy in various applications and is effective for classifying data based on probabilistic principles (Kristiyanti, Indrayuni, Nurhadi, & Umam, 2020; Kewsuwun & Kajornkasirat, 2022). It works well in this domain due to the ability to calculate the probability of spam text. Naïve Bayes has proven highly effective in various complex real-world applications such as; spam filtering, medical diagnosis, real-time prediction, and weather forecasting (Mansour, Saleh, Badawy, & Ali, 2022). In this research, the Naïve Bayes algorithm is used to analyze and compare two e-wallet applications because of its simplicity, efficiency, and effectiveness in handling large datasets. By applying Naïve Bayes, this research aims to classify e-wallet user reviews, helping to identify which applications perform better and ultimately supporting data-driven decisions on the best applications for diverse user needs.

There are several previous studies related to this sentiment analysis on e-wallets, including (Kristiyanti, Indrayuni, Nurhadi, & Umam, 2020), (Sanjaya, Pudjiantoro, Ningsih, & Renaldi, 2022), (Ruliana, Inayah, & Rais), and (Ningri, Hamidi, & Adrianto, 2023). (Sanjaya, Pudjiantoro, Ningsih, & Renaldi, 2022) stated that the quantity of data that used for this sentiment analysis can significantly impact the accuracy of predictions made using the Naïve Bayes algorithm. In the second trial, (Ruliana, Inayah, & Rais) achieved an accuracy of 97.37% using the Naïve Bayes method for data sharing (80% for data training and 20% for testing). (Ningri, Hamidi, & Adrianto, 2023) used Naïve Bayes to categorized comment data about the e-wallets into two sentiment classes: positive and negative.

This research consists of the following sections: Method, provides a brief explanation of the method used in this research, Result and Discussion discusses the outcomes obtained from applying the method to the collected data, Conclusion, concluded the results of this research based on the method applied to the datasets, expressed as percentage accuracy.

Methods

In this research, the Naïve Bayes method was applied to analyze text data, by utilizing 1,000 review datasets from two selected e-wallet applications on the Google Play Store as the basis for analysis and model evaluation. The flow of the Naïve Bayes Algorithm to be applied is represented in Figure 1.

Figure 1. Method Scheme

1. Data Collection
   In the first stage of the method, datasets were collected from the Google Play Store using the google-play-scraper library, which was utilized to extract reviews and relevant information from two different applications in a structured format. The data collected consisted of 1,000 reviews obtained through this data scraping process for further analysis.
2. Data Labelling
   During the data labelling stage, the collected datasets are initially labeled as either "Positive" or "Negative" before proceeding to the next step in the process. Sentiment labels, "Positive" and "Negative", are assigned based on text analysis rather than the given rating. Positive sentiment is identified by the presence of words indicating satisfaction, appreciation, or recommendation. Meanwhile, negative sentiment is characterized by words expressing criticism, complaints or dissatisfaction.
3. Data Preprocessing
   In this data preprocessing stage, the collected data are first subjected to text normalization, where the text is standardized to ensure uniformity and punctuation removal to eliminate unnecessary characters. The process continues with tokenization of the datasets, where each review is broken down into smaller units to facilitate easier analysis.
4. Naïve Bayes Classification
   The Naive Bayes algorithm is a commonly used algorithm for data classification, known for its ability to calculate probabilities based on processed data (Lutfi, Saputra, & Fa'rifah, 2023). In the context of this research, Naïve Bayes will be applied to the collected datasets to evaluate the accuracy of the model in making predictions. (Sanjaya, Pudjiantoro, Ningsih, & Renaldi, 2022)
5. Evaluation
   After implementing the Naïve Bayes algorithm, an evaluation was conducted to assess the performance of the resulting model. The evaluation involves metrics such as accuracy, recall, and precision to ensure that the model effectively classifies sentiment.

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

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

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

   The results of this evaluation will be used to assess the effectiveness of the method and identify areas for improvement. These findings will be presented in the Result and Discussion section, following the application of the Naïve Bayes algorithm to determine which of the two selected e-wallet applications perform better.

Result and Discussion

This section presents the analysis results along with a concise discussion:
The collected data consists of 1,000 datasets, with 500 reviews from application X and 500 from Application Y. The data was then labeled as positive or negative.

Table 1: Labelling Result

Reviews	Sentiment
Aplikasi tidak sesuai sama janjinya ke driver??menyedihkan...notifikasinya aja mengiurkan...pinjaman kendaraan juga gak bisa di akses untuk melanjutkan...beda sama sebelah pinjaman muda di acc yg penting driver rajin	Negative
Woi sumpah iklan aplikasi ini ganggu banget gw lagi main game ada aja iklan aplikasi ini tolong diperhatikan	Negative
bagus juga sering ada promo dan mudah dipahami pemakaian aplikasi nya. Semoga jadi aplikasi terbaik dan memudahkan semua pengguna ya!	Positive
Fasilitas dan fitur yg mudah. Aman. Murah dalam berbagai transaksi. Recommended	Positive
Bagus ...banyak fitur yg praktis dan mudah digunakan..next agar bisa lebih bagus . terutama harga bisa lebih murah	Positive

After the data labelling process, the next step is the normalization of data reviews. Each review is first converted to lowercase to ensure uniformity and eliminate any inconsistencies related to case sensitivity. Punctuation is then removed to focus solely on the words themselves. Then, the reviews are tokenized, breaking each sentence into individual words (Masturoh, Pratiwi, Saelan, Radiyah, & others, 2023). This process facilitates a more precise analysis of the content, enabling the model to accurately interpret the meaning and sentiment of each review.

Table 2: Data Preprocessing

	Before	After
Text Cleaning	Bagus ...banyak fitur yg praktis dan mudah digunakan..next agar bisa lebih bagus . terutama harga bisa lebih murah	bagus banyak fitur yg praktis dan mudah digunakan next agar bisa lebih bagus terutama harga bisa lebih murah
Tokenization	fasilitas dan fitur yg mudah aman murah dalam berbagai transaksi recommended	‘fasilitas’, ‘dan', 'fitur', 'yg', 'mudah', ‘aman', 'murah', 'dalam', 'berbagai’, 'transaksi’, 'recommended’

After tokenization, the data preprocessing step continues with the creation of a word dictionary from the list of generated words. This dictionary is a collection of unique words extracted from all reviews that taken, a bag-of-words representation is created to represent the frequency of each word in each review. This approach allows for identifying and evaluating the most dominant words based on their occurrences.

Table 3: Frequently Words used

Words	Application X			Application Y		
	Positive Probability	Negative Probability	Amount	Positive Probability	Negative Probability	Amount
Aplikasi	18.20%	81.80%	254	2.91%	97.09%	332
Saldo	5.40%	94.60%	146	4.58%	95.42%	277
Transaksi	12.75%	87.25%	133	6.50%	93.50%	165
Transfer	18.39%	81.61%	77	6.73%	93.27%	146

Bagus	49.89%	50.11%	39	15.77%	84.23%	32

Table 3 presents the words frequently used in user reviews for two e-wallet applications, Applications X and Y, analyzed through a sentiment analysis process. The table displays the probabilities of positive and negative sentiments associated with each word, as well as the total of occurrences in the reviews. For Application X, words such as 'aplikasi' and 'saldo' also appear more frequently, with 254 and 146 occurrences, respectively, and show predominantly negative sentiment probabilities. In Application Y, the same word also occurs frequently with 'aplikasi' appearing 332 times and 'saldo' appearing 277 times, similarly demonstrating high negative sentiment probabilities of 97.09% and 95.42%. This analysis provides valuable insights into the area of user satisfaction and dissatisfaction, helping to understand the positive and negative perceptions for each application.

For model testing, 80% of the data from each application was used. The total dataset for this model consists of 800 entries, divided into training data and testing data with a ratio of 80:20. The evaluation of this model testing, generated using 80% of the dataset consisting of 800 entries, utilizes 640 data that selected for model training and presented in the form of a confusion matrix table.

Table 4: Confusion Matrix for Model Testing (Training Data)

Classification Category	Predicted Positive	Predicted Negative	Recall
True Positive	50	14	78%
True Negative	0	576	100%
Precision	100%	98%	

And the evaluation of this model testing, generated using 20% of the dataset consisting of 800 entries, utilizes 160 data that selected for model training and presented in the form of a confusion matrix table.

Table 5: Confusion Matrix for Model Testing (Testing Data)

Classification Category	Predicted Positive	Predicted Negative	Recall
True Positive	6	10	38%
True Negative	3	141	98%
Precision	67%	93%	

This model achieved an accuracy of 97.81% on the training data and 91.87% on the testing data, by applying Naïve Bayes classification method to 640 data training and 160 data testing, as represented by Table 4 and Table 5. This testing model will then be applied to the testing data for application X and application Y. The results of the model evaluation on the testing data for application X and Y are presented in the form of these confusion matrix tables.

Table 6: Confusion Matrix Naïve Bayes for Application X

Classification Category	Predicted Positive	Predicted Negative	Recall
True Positive	8	0	100%
True Negative	14	78	85%
Precision	36%	100%	

Based on Table 6, the sentiment analysis of Application X achieves 86% accuracy. With a total of 100 reviews, 78 true negatives and 22 true positives were obtained. For recall, the percentage obtained 100% for negative and 36.36% for positive. Meanwhile, for precision, the percentage obtained was 84.78% for negative and 100% for positive.

Table 7: Confusion Matrix Naïve Bayes for Application Y

Classification Category	Predicted Positive	Predicted Negative	Recall
True Positive	1	0	100%
True Negative	5	94	95%
Precision	17%	100%	

Based on Table 4, the sentiment analysis of Application Y achieves 95% accuracy. With a total of 100 reviews, 94 true negatives and 6 true positives were obtained. For recall, the percentage obtained 95% for negative and 100% for positive. Meanwhile, for precision, the percentage obtained was 100% for negative and 17% for positive.

Between these two applications, it was found that Application Y had higher accuracy than Application X. Application Y achieved an accuracy rate of 95% compared to 86% for Application X. However, Application X received more positive feedback than Application Y. Out of 100 reviews for each application tested on the training model, Application X received 22 positive feedback responses, while Application Y received only 6 positive feedback responses. This indicates that Application X is notably better than Application Y in terms of positive feedback.

Conclusion

Based on the research results on sentiment analysis of e-wallet applications X and Y, it was found that the reviews tend to be negative. Using the Naïve Bayes classification method, the model achieved an accuracy of 91.87% on user review data labeled as positive and negative. The accuracy of the training model in sentiment analysis can be considered good, as it achieved 91.87% accuracy. When the model is applied to the test data for each application, 86% accuracy is achieved for Application X and 98% accuracy for Application Y. However, when compared these two applications with higher levels of positive feedback, Application X stands out with 22 instances of positive feedback, whereas Application Y only has 6. Therefore, Application X is better than Application Y in terms of positive feedback.

In future research, it is hoped to increase the amount of data used to improve the accuracy of positive and negative sentiment classification. Additionally, other classification algorithms, such as Support Vector Machine (SVM) or K-Nearest Neighbor (KNN), are expected to be used. Furthermore, online applications can include more than two e-wallet applications, and it is hoped that this research can be used as a reference for further research.

References

Hayatin, N., Marthasari, G. I., & Nuarini, L. (2020). Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization. *Jurnal Online Informatika, 5*(1), 81-88.

Kathiravan, D. C., Rajasekar, A., Velmurgan, S., Mahalakshmi, P., Chandramouli, E. a., Padmaja, B., & Dhanalakshmi, K. (2021). Sentiment analysis and text mining of online customer reviews for digital wallet apps of Fintech industry. *Int. J. of Aquatic Science, 12*(3), 2139-2150.

Kewsuwun, N., & Kajornkasirat, S. (2022). A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. *International Journal of Electrical & Computer Engineering (2088-8708), 12*(3).

Kristiyanti, D. A., Indrayuni, E., Nurhadi, A., & Umam, A. H. (2020). E-Wallet Sentiment Analysis Using Naive Bayes and Support Vector Machine Algorithm. *Journal of Physics: Conference Series, 1641*(1), 012079.

Lutfi, Y. T., Saputra, M., & Fa'rifah, R. Y. (2023). Aspect-Based Sentiment Analysis in Identifying Factors Causing Technostress in Fintech Users Using Na{\"\i}ve Bayes Algorithm. *International Conference on Enterprise and Industrial Systems (ICOEINS 2023)*, 107-117.

Mansour, N. A., Saleh, A. I., Badawy, M., & Ali, H. A. (2022). Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. *Journal of ambient intelligence and humanized computing*, 1-33.

Masturoh, S., Pratiwi, R. L., Saelan, M. R., Radiyah, U., & others. (2023). Application of the K-Nearest Neighbor (Knn) Algorithm in Sentiment Analysis of the Ovo E-Wallet Application. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 8*(2), 84-89.

Ningri, L. J., Hamidi, M., & Adrianto, F. (2023). Sentiment Analysis Against Digital Payment "GoPay","OVO","DANA", and "ShopeePay" Using Naïve Bayes Classifier Algorithm. *Indonesian Journal of Economics and Management, 3*(2), 322-336.

Ruliana, R., Inayah, A., & Rais, Z. (n.d.). Analysis of naïve bayes classification algorithm of Go-pay user satisfaction. *AIP Conference Proceedings, 2704*(1).

Sanjaya, F. A., Pudjiantoro, T. H., Ningsih, A. K., & Renaldi, F. (2022). Sentiment Analysis Of E-Wallets on Twitter social media With Naïve Bayes and Lexicon-Based Methods. *Proceedings of the International Conference on Industrial Engineering and Operations Management, Michigan, USA: IEOM Society International*, 1002-1010.