A Comparison of Online Investment Application Opinion Based on Sentiment Analysis

Ilucky Salim[1*], Jimmy Tjen[2]
[1] Department of Digital Business, Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak, Indonesia.
[2]Department of Informatics, Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak, Indonesia.
Email: [1] 22430161@widyadharma.ac.id, [2] jimmy.tjen@mathmods.eu

Abstract. As times goes by, the number of online investment applications in Indonesia continues to grow, leaving people confused about which application is best for investing. Reviews of online investment applications on the Google Play Store can provide valuable information for those looking to start investing. Analyzing application reviews requires more than just looking at star rating but also to examine the full content of review comments to understand the intent behind them. A sentiment analysis system can process reviews to extract meaningful information, including sentiment. Therefore, this study aims to compare popular online investment applications in Indonesia to determine the best one using the Naive Bayes classification. A total of 1.550 online investment app user reviews were collected to form the sample dataset. The research stages include data collection, labeling, preprocessing, classification, and evaluation. The data is divided into three categories: negative, neutral, and positive. The results show that Z application has the highest test accuracy, at 94,19%. Based on the Bag of Words analysis, the Z application is more popular among users than other applications due to its ease of use for investing.

Keywords: Naïve Bayes, online investment application, sentiment analysis, bag of words.

Introduction

The rapid growth of financial technology (fintech) has led to the emergence of many online investment applications on the Google Play Store. These apps are designed to simplify investment process and making them accessible to a wider audience. However, the increasing variety of option has created a need for objective assessments to help user select the most suitable app. Each online investment app has its advantages and disadvantages. Common problem encountered when using online investment applications include delays in the purchasing process, extended access times, and difficulties in fund withdrawals. These issues significantly affect user satisfaction, which serves as a critical indicator of the service quality provided by online investment platforms.

Selecting a reliable online investment application necessitates gathering comprehensive based on the experiences of previous users. This process involves analyzing review data available for these applications on the Google Play Store. Sentiment analysis offers a powerful tool to analyze user feedback and reviews, providing valuable insight into user satisfaction and app performance. Sentiment analysis is a critical area of study within natural language processing that focuses on systematically identifying and classifying opinions within text data (Dhamayanthi & others, 2024). The are several ways to do sentiment analysis, one among these is Naïve Bayes.

Naïve Bayes algorithm known for its simplicity and efficiency. It is employed to classify sentiments as positive, neutral, or negative. This method calculates the probabilities for each factor and then selects the outcome with the highest probability (Wisnu, Afif, & Ruldevyani, 2020). By examining user sentiments, this research compares online investment applications with good reputations among users. The findings will provide actionable insight for developers to improve their offerings and for users to make informed decision. Additionally, this study contributes to the growing body of literature on fintech application evaluation through data-driven approaches. The results highlight the importance of user feedback in shaping the development and reputation of online investment platforms. There are several papers that use this classification. (Pristiyono, Ritonga, Ihsan, Anjar, & Rambe, 2021) assesses the opinion of the Indonesia people by analyzing Twitter data with 'Vaccine COVID-19' as keywords using Naïve Bayes Algorithm. The result identified over 3.400 tweets as negative, more than 2.400 tweets as positive, and 301 tweets as neutral during the week-long analysis. (Dieksona, et al., 2023) analyzed the sentiments of Twitter users regarding Traveloka's performance by comparing the Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes Algorithm. The result indicate that the Support Vector Machine (SVM) algorithm demonstrates superior accuracy in analyzing tweets related to Traveloka. (Pratmanto, et al., 2020)classified reviews of the Shopee application on the Google Play Store into positive and negative categories using the Naïve Bayes, achieving an accuracy of 96,67%. (Adam, Rosli, & Soh, 2021) demonstrated that Naïve Bayes can be effectively applied to sentiment analysis tasks for movie reviews, with an accuracy of 89%. (Novendri, Callista, Pratama, & Puspita, 2020) analyzed public opinions in YouTube movie trailer comments using the Naive Bayes algorithm, achieving 81% accuracy, 74,83% precision, and 75,22% recall.

Based on that background, we propose the research topic, "A Comparison of Online Investment Application Opinion Based on Sentiment Analysis". This study focuses on three online investment applications widely used in Indonesia. The research has two main objectives. First, to provide recommendations for individuals considering online investments. Second, to offer valuable insights to application developers to improve their services. By analyzing user sentiment, this study aims to support better decision-making for potential investors and optimize services provided by application developers.

This paper is organized as follows: Methods describes the methodology used and the process of data collection. Result and Discussion explains the data processing methods applied to collected data and presents the results, including a comparison of the three online investment applications, highlighting the application with the most positive comments.

Methods

This research reviews the three online investment applications on the Google Play Store by categorizing comments as positive, negative, or neutral. It then analyzes and determines which one has the best reputation in the eye of users. To complete the authors has outlined clear steps as a guide

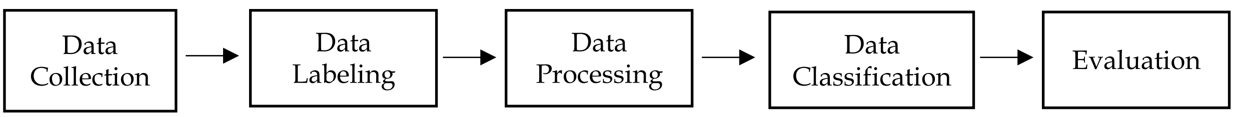

Figure 1: Method Scheme

1. **Data Collection:** The first stage is the process of collecting data on the Google Play Store. The data collection process is carried out through scraping using Google Colab tools, gathering comments from users of three online investment applications. The process involves storing, collecting, and validating data to ensure it can be interpreted as meaningful information.
2. **Data Labeling:** During the labeling process, the data is categorized into three groups: positive, negative, and neutral. The positive category includes emotions such as relaxation, happiness, cheerfulness, and other positive affective states. The negative category encompasses emotions such as sadness, fear, anger, depression, and other states associated with distress or suffering. Meanwhile, the neutral category represents emotionally neutral states that do not exhibit clear positive or negative sentiments, such as factual sentiments or indifferent expressions. This classification is performed manually and relies on subjective judgement.
3. **Data Preprocessing:** The data preprocessing stage is carried out at this stage to process the data that has been collected in an irregular form. There are several stages in this study, there are cleaning, case folding, and tokenize. Cleaning is to removing punctuation marks, certain character, and emote. Case folding is a process in text processing where all letters in a text are converted to lowercase (Erfina & Nurul, 2023). Tokenizing is the process of splitting a text into smaller units, such as sentences and words, which are referred to as tokens, each assigned a unique index for further analysis (Nikmah, et al., 2022).

4. **Data Classification:** This classification employs the Naïve Bayes method with data split for testing and training set at 80% and 20% respectively. The Naïve Bayes method is one of the most effective tools for classification problems, particularly in text mining for sentiment analysis (Pristiyono, Ritonga, Ihsan, Anjar, & Rambe, 2021). Its primary advantage lies in its simplicity, which enables efficient computation while maintaining high performance in text classification tasks.
5. **Evaluation:** The testing data results are evaluated using a confusion matrix, with a focus on accuracy. A confusion matrix displays the predictions alongside the actual conditions of the data generated by the algorithm. Accuracy serves as an evaluation metric to assess the proportion of correct predictions relative to the total dataset, as shown in Equation 1. (AlZoman & Alenazi, 2021)

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

   Recall and precision calculations are also used to measure system performance. As shown in Ewquation 2, precision measures the accuracy of positive predictions by calculating the ratio of correctly predicted sentiments to the total positive predictions. (AlZoman & Alenazi, 2021)

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

   Recall measures the proportion of actual positive cases that are correctly predicted, as described in Equation 3. (AlZoman & Alenazi, 2021)

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

   After obtaining the results of three online investment applications, they will be compared to determine which one is better. In addition, the frequently occurring words in the comments for each investment application will also be identified. This uses the bag of words feature which treats a text (such as a sentence or document) as a collection of individual words, without considering the order or grammar.

Result and Discussion

1. **Data Collection:** The data collection performed via web scraping, was executed using Google Colab. The result of the scrapping stage for three online investment applications from the Play Store yielded a total 1.550 comments based on the latest data.
2. **Data Labeling:** The scrapped data that has been saved is then labeled to distinguish positive, negative, and neutral responses. The authors chose to perform manual labeling because some users may give a 5-star rating but leave a negative comment. If a machine were to be used for labeling, it would classify 5-star ratings as positive comments. The following is an example of the labeling process shown in Table 1.

Table 1: Labeling Result

Content	Label
Susah banget verifikasi penarikan dan penambahan akun bank	Positif
Aplikasinya udah bagus, tapi mohon tambahan min di fitur sip, buat perminggu. Kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya.	Netral
sekarang bisa pakai margin …mantap	Positif
Proses jual sampai bisa di tarik dana.lama.seminggu	Negatif

3. **Data Processing:** In this study, preprocessing consists of three stages. There are:
   a. **Cleaning**
   The removal of punctuations, numbers, Unicode, and characters in a data is done in this stage. The following is an example of the cleaning process shown in Table 2.

Table 2: Cleaning Result

Data before cleaning	Data after cleaning
Susah banget verifikasi penarikan dan penambahan akun bank	Susah banget verifikasi penarikan dan penambahan akun bank
Aplikasinya udah bagus, tapi mohon tambahan min di fitur sip, buat perminggu. Kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya.	Aplikasinya udah bagus tapi mohon tambahan min di fitur sip buat perminggu Kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya
sekarang bisa pakai margin …mantap	sekarang bisa pakai margin mantap

Proses jual sampai bisa di tarik dana.lama.seminggu	Proses jual sampai bisa di tarik dana lama seminggu

b. **Case folding**
The data will transform to lowercase in this stage. The following is an example of the case folding process shown in Table 3.

Table 3: Case Folding Result

Data before case folding	**Data after case folding**
Susah banget verifikasi penarikan dan penambahan akun bank	susah banget verifikasi penarikan dan penambahan akun bank
Aplikasinya udah bagus tapi mohon tambahan min di fitur sip buat perminggu Kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya	aplikasinya udah bagus tapi mohon tambahan min di fitur sip buat perminggu kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya
sekarang bisa pakai margin mantap	sekarang bisa pakai margin mantap
Proses jual sampai bisa di tarik dana lama seminggu	proses jual sampai bisa di tarik dana lama seminggu

c. **Tokenizing**
This stage aims to separate word and symbols from a given text**.** The following is an example of the tokenizing process shown in Table 4.

Table 4: Tokenizing Result

Data before tokenizing	**Data after tokenizing**
susah banget verifikasi penarikan dan penambahan akun bank	'susah', 'banget', 'verifikasi', 'penarikan', 'dan', 'penambahan', 'akun', 'bank'
aplikasinya udah bagus tapi mohon tambahan min di fitur sip buat perminggu kayak ada fitur 2 minggu dan 3 minggu terimakasih sebelumnya	'aplikasinya', 'udah', 'bagus', 'tapi', 'mohon', 'tambahan', 'min', 'di', 'fitur', 'sip', 'buat', 'perminggu', 'kayak', 'ada', 'fitur', '2', 'minggu', 'dan', '3', 'minggu', 'terimakasih', 'sebelumnya'
sekarang bisa pakai margin mantap	'sekarang', 'bisa', 'pakai', 'margin', 'mantap'
proses jual sampai bisa di tarik dana lama seminggu	'proses', 'jual', 'sampai', 'bisa', 'di', 'tarik', 'dana', 'lama', 'seminggu'

4. **Data Classification:** The data generated during the preprocessing stage is subsequently split into training and testing sets in an 80%:20% ratio. Eighty percent is used as training data while twenty percent is used as testing data. So, from a total of 1.550 comments, 1.240 were used as training data, while 310 were used as test data. The data is used to train a model using the review data and predict outcomes to achieve good performance. The machine will be trained based on the training data and will be tested with the testing data. This classification model utilizes the Naïve Bayes Classifier.
5. **Evaluation and Performance Analysis of Classification:** The evaluation needs a confusion matrix to evaluate how well the Naïve Bayes Classifier perform in classifying data.

Table 5: Confusion Matrix of X Application

Confusion Matrix (80%:20%)		Actual				
		Positive (1)	Negative (2)	Neutral (3)	Total	Recall
Prediction	Positive (1)	243	3	8	254	97,59%
	Negative (2)	6	38	5	49	92,68%
	Neutral (3)	5	1	1	7	0
	Total	254	42	14	310	
	Precision	98,78%	86,36%	0		

Based on table 5, there are 310 comments divided into three categories: 254 comments are positive responses, 42 are negative responses, and 14 are neutral responses. Out of the 254 comments predicted to positive responses, 243 comments actually positive responses, 6 comment are predicted to contain negative responses, and 5 comments are predicted to contain neutral responses. Then, out of the 42 comments predicted to negative responses, 38 actually contain negative responses, 3 comments are predicted to contain positive responses, and 1 comment is predicted to contain neutral responses. Out of the 14 comments predicted to neutral responses, 1 comment actually neutral response, 8 comment are predicted to contain positive responses, and 5 comments are predicted to contain negative responses. The test accuracy obtained from the Naïve Bayes Algorithm for the X application is 90,97%, while the training accuracy is 94,84%. Positive sentiment produces a recall of 97,59% and a precision of 98,78%, while negative sentiment produces a recall of 92,68% and a precision of 86,36%.

Table 6: Confusion Matrix of Y Application

Confusion Matrix (80%:20%)		Actual				
		Positive (1)	Negative (2)	Neutral (3)	Total	Recall
Predicted	Positive (1)	212	5	15	232	98,60%
	Negative (2)	3	65	9	77	92,86%
	Neutral (3)	0	0	1	1	0
	Total	215	70	25	310	
	Precision	97,70%	95,59%	0		

Based on table 6, there are 310 comments divided into 3 categories. 215 comments are positive responses, 70 comments are negative responses, and 25 comments are neutral responses. Out of the 215 comments predicted to positive responses, 212 comments actually positive responses, 3 comment are predicted to contain negative responses, and 0 comment is predicted to contain neutral response. Then, out of the 70 comments predicted to negative responses, 65 comments actually contain negative responses, 5 comments are predicted to contain positive responses, and 0 comment is predicted to contain neutral response. Out of the 25 comments predicted to neutral responses, 1 comment actually neutral response, 15 comments are predicted to contain positive responses, and 9 comments are predicted to contain negative responses. The test accuracy obtained from the Naïve Bayes Algorithm for the Y application is 89,68%, while the training accuracy is 94,84%. Positive sentiment produces a recall of 98,60% and a precision of 97,70%, while negative sentiment produces a recall of 92,86% and a precision of 95,59%.

Table 7: Confusion Matrix of Z Application

Confusion Matrix (80%:20%)		Actual				
		Positive (1)	Negative (2)	Neutral (3)	Total	Recall
Predicted	Positive (1)	273	4	5	282	97,50%
	Negative (2)	7	19	1	27	82,61%
	Neutral (3)	1	0	0	1	0
	Total	281	23	6	310	
	Precision	98,56%	73,08%	0		

Based on table 7, there are 310 comments divided into 3 categories. 281 comments are positive responses and 23 are negative responses, and 6 comments are neutral responses. Out of the 281 comments predicted to positive responses, 273 comments actually positive responses, 7 comment are predicted to contain negative responses, and 1 comment is predicted to contain neutral response. Then, out of the 23 comments predicted to negative responses, 19 actually contain negative responses, 4 comments are predicted to contain positive responses, and 0 comment is predicted to contain neutral response. Out of the 6 comments predicted to neutral responses, 0 comment actually neutral response, 5 comment are predicted to contain positive responses, and 1 comment are predicted to contain negative responses. The test accuracy value obtained from the Naïve Bayes Algorithm for Z application is 94,19%, while the training accuracy is 94,76%. Positive sentiment produces a recall of 97,59% and a precision of 98,78%, while negative sentiment produces a recall of 92,68% and a precision of 86,36%.

Among the three applications, the Z application achieves the highest accuracy for both testing (94,19%) and training (94,76%). This demonstrates a strong ability to correctly classify sentiment in both phases. The small difference between training and testing accuracy indicates that the model does not suffer from overfitting, ensuring consistent performance across a wide range of datasets. The second highest accuracy is application X with the training accuracy of 94,84% and testing accuracy of 90,97%. It still high and good but less robust than application Z, which has a much smaller difference between its training and testing accuracies. Last is Y application with testing accuracy of 89,68% and training accuracy is 94,84%. It performs well during training but for the testing accuracy suggests that the model might be overfitting. Overfitting happens when the model shows high performance on the training data but has difficulty generalizing to new, unseen data, as seen in the lower testing accuracy. So, it needs some adjustments to improve its generalization ability.

Bag of Word: The BoW model simplifies text analysis by converting words into numerical values based on their frequency, without considering the sequence in which they appear. (Singh, Prabhu Shankar, & Chinnaiyan, 2023). The following table lists the words that frequently appear in applications X, Y, and Z:

Table 8: Frequently Words in the X, Y, and Z application

Words	X Application		Y Application		Z Application	
	Probability (Positive)	Amount	Probability (Positive)	Amount	Probability (Positive)	Amount
mudah	99,41%	254	99,28%	253	99,58%	344
bagus	96,97%	232	96,16%	228	97,01%	226
investasi	96,52%	196	95,84%	210	97,30%	283
saham	88,77%	210	87,98%	197	91,58%	273
pemula	99,33%	156	99,04%	165	99,47%	225
fitur	95,45%	100	93,18%	86	93,96%	111
mantap	98,34%	88	98,19%	81	97,86%	78
membantu	98,12%	63	98,97%	63	99,12%	80
cocok	98,86%	54	98,82%	55	99,12%	80
terbaik	98,89%	56	98,73%	51	98,89%	63

Based on the table 8, there are 10 words that frequently appear across these three investment applications. The words like "mudah", "bagus", "pemula" and "investasi" are common across the X, Y, and Z applications. It suggests that users frequently associate these words with positive experiences. It means that the three applications are shown to be easy to use for investing and suitable for beginners. These words have consistently high positive sentiment probabilities. For example, "mudah" has a positive sentiment probability of 99,41% in X, 99,28% in Y, and 99,58% in Z. It reflecting user's consistent appreciation of ease of use in all three applications. While, the word "saham" has a noticeably lower positive sentiment in Y (87,98%) compared to X (88,77%), and Z (91,58%), potentially indicating a weaker association with satisfaction in Y. It means that apps X and Z offer better stock related functionality than Y, such as clearer investment guides, better tools, and more engaging content, all of which users find more useful and user-friendly. To address this, app Y can focus enhancing stock-related functionality, increasing user engagement, and gathering targeted feedback to address user concerns. The "amount" column shows that certain words, such as "mudah" (344 comments) and "investasi" (283 comments) are mentioned more frequently in application Z. This suggests that these topics are highly relevant and significant to user of app Z.

Overall, application Z appears to perform slightly better in terms of positive sentiment and word frequency, indicating higher user satisfaction. Meanwhile, apps X and Y could improve their sentiment scores in certain areas, such as stocks, to better align with user expectations. Addressing these differences could enhance user satisfaction and ensure consistent positive feedback across all platforms.

Conclusions

From the tests that have been done using the naïve bayes algorithm method with a total of 1.550 comments, the test accuracy value of the X application is 90,97%, the Y investment application is 89,68%, and the Z Application is 94,19%. Of the three applications, the Z application has the highest accuracy. Across all three applications, users generally associate positive sentiments with words related to ease of use, quality, and relevance to investment, as shown by high positive probabilities. Application Z generally records the highest positive sentiment probabilities for most words, indicating it might provide the best user experience among the three. The second-best application is X, while the lowest performing one is Y, based on the accuracy achieved using the Naïve Bayes method and the bag of words approach.

References

Adam, N. L., Rosli, N. H., & Soh, S. C. (2021). Sentiment analysis on movie review using Naive Bayes. *2021 2nd international conference on artificial intelligence and data sciences (AiDAS)* (pp. 1--6). IEEE. doi:10.1109/AiDAS53897.2021.9574419

AlZoman, R. M. (2021). A comparative study of traffic classification techniques for smart city networks. *Sensors, 21*(14), 4677. doi:https://doi.org/10.3390/s21144677

Dhamayanthi, N., & others. (2024). The Role Of Naive Bayes, SVM, And Decision Trees In Sentiment Analysis. *Educational Administration: Theory and Practice, 30*(4), 6377--6381. doi:https://doi.org/10.53555/kuey.v30i4.2392

Dieksona, Z. A., Prakosoa, M. R., Qalby, M. S., Putraa, M., Achmada, S., & Sutoyoa, R. (2023). Sentiment analysis for customer review: Case study of Traveloka. *Procedia Computer Science, 216*, 682--690. Retrieved from https://www.researchgate.net/profile/Rhio-Sutoyo/publication/366990352_Sentiment_analysis_for_customer_review_Case_study_of_Traveloka/links/65a2bceebc30165e6e37871f/Sentiment-analysis-for-customer-review-Case-study-of-Traveloka.pdf

Erfina, A., & Nurul, M. R. (2023). Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. *Data and Metadata, 2*, 45--45. doi:https://doi.org/10.56294/dm202345

Kundana, D., & others. (2023). Data Driven Analysis of Borobudur Ticket Sentiment Using Naive Bayes. *Aptisi Transactions on Technopreneurship (ATT), 5*(2sp), 221-233. doi:https://doi.org/10.34306/att.v5i2sp.353

Nikmah, T. L., Ammar, M. Z., Allatif, Y. R., Husna, R. M., Kurniasari, P. A., & Bahri, A. S. (2022). Comparison of LSTM, SVM, and naive bayes for classifying sexual harassment tweets. *Journal of Soft Computing Exploration, 3*(2), 131--137. doi:https://doi.org/10.52465/joscex.v3i2.85

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment analysis of YouTube movie trailer comments using naive bayes. *Bulletin of Computer Science and Electrical Engineering, 1*(1), 26--32. doi:https://doi.org/10.25008/bcsee.v1i1.5

Pratmanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, *1641*, p. 012043. doi:10.1088/1742-6596/1641/1/012043

Pristiyono, Ritonga, M., Ihsan, M. A., Anjar, A., & Rambe, F. H. (2021). Sentiment analysis of COVID-19 vaccine in Indonesia using Naive Bayes Algorithm. *IOP Conference Series: Materials Science and Engineering. 1088*, p. 012045. IOP Publishing. doi:10.1088/1757-899X/1088/1/012045

Singh, U. K., Prabhu Shankar, B., & Chinnaiyan, R. a. (2023). Machine Learning-Based Text Categorization with Bag of Words. *The International Conference on Recent Innovations in Computing* (pp. 577--587). Springer.

Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naive Bayes. *Journal of Physics: Conference Series. 1444*, p. 010234. IOP Publishing. doi:10.1088/1742-6596/1444/1/012034