

Sentiment Analysis of Duolingo Application Reviews on Google Play Using the Support Vector Machine Method

Nur Jihan Arifi^{1*}, Novriyenni², Muammar Khadapi³

^{1,2,3}Informatics Engineering, STMIK Kaputama
nurjihanarifibinjai@gmail.com^{1*}, novriyenni.sikumbang@gmail.com², khdafi5@gmail.com³

Abstract

The Duolingo app is one of the most widely used language-learning apps available on Google Play. The large number of reviews posted by users makes sentiment analysis necessary to identify trends in user opinions regarding the app—whether positive, negative, or neutral. This study aims to analyze the sentiment of Duolingo app user reviews on Google Play using the Support Vector Machine (SVM) algorithm. Research data was obtained through web scraping of 1,220 user reviews, followed by preprocessing, sentiment labeling, and word weighting using TF-IDF prior to the classification process using the SVM algorithm. The results show that positive sentiment accounted for 57.1%, negative sentiment for 30.3%, and neutral sentiment for 12.6%. Based on the test results, the Support Vector Machine (SVM) algorithm achieved an accuracy of 79.23%, demonstrating fairly good performance in classifying the sentiment of user reviews for the Duolingo app on Google Play.

Keywords: Sentiment Analysis, Duolingo, Google Play, Support Vector Machine (SVM).

1. Introduction

The rapid development of information and communication technology has significantly transformed various aspects of life, including the education sector. One of the impacts of this transformation is the increasing use of mobile learning applications, which provide flexible and accessible learning experiences. Duolingo is one of the most popular language learning applications, with more than 500 million downloads on Google Play Store. The application offers interactive language courses through gamification and Artificial Intelligence (AI)-based learning features that adapt to users' learning needs [1]. The growing number of Duolingo users has resulted in a large volume of reviews on Google Play Store, reflecting users' experiences, satisfaction, and expectations regarding the application.

User reviews on Google Play Store represent valuable information that can be used to evaluate application quality and identify users' opinions. However, the large number of reviews and their unstructured textual format make manual analysis inefficient. In addition, user reviews often contain informal language, abbreviations, and various writing styles, making them more challenging to process. Therefore, sentiment analysis has become an effective approach for automatically classifying user opinions into positive, negative, and neutral categories. As one of the applications of text mining, sentiment analysis helps developers understand user perceptions and supports decision-making for improving application quality [2].

Support Vector Machine (SVM) is one of the most widely used machine learning algorithms for text classification because of its ability to classify high-dimensional data with high accuracy. Previous studies have demonstrated the effectiveness of SVM in sentiment classification. Achieved an accuracy of 98% in classifying Shopee application reviews using SVM [3]. While obtained an accuracy of 97% in sentiment analysis of the Qur'an Kemenag application after applying the SMOTE technique [4]. These findings indicate that SVM has strong potential for sentiment classification. However, studies focusing on Indonesian-language reviews of the Duolingo application using the SVM algorithm are still limited. Therefore, this study applies the SVM algorithm to analyze the sentiment of Duolingo user reviews on Google Play.

Based on this research gap, this study applies the Support Vector Machine method to analyze Indonesian-language user reviews of the Duolingo application collected from Google Play Store. Before classification, the review data are preprocessed through cleansing, case folding, stopword removal, tokenization, slang word normalization, and stemming. Subsequently, TF-IDF is employed to transform textual data into numerical feature vectors before classification using the SVM algorithm. The objective of this study is to classify user reviews into positive, negative, and neutral sentiments, as well as to evaluate the performance of the Support Vector Machine model using accuracy,

precision, recall, and F1-score. The findings are expected to provide useful insights into users' perceptions of the Duolingo application while demonstrating the effectiveness of SVM for sentiment analysis of Indonesian-language application reviews.

2. Research Methodology

In this study, the Cross-Industry Standard Process for Data Mining (CRISP-DM) was used. This method consists of six main phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. These phases served as a framework for the sentiment analysis of Duolingo app reviews until a suitable classification model was obtained.

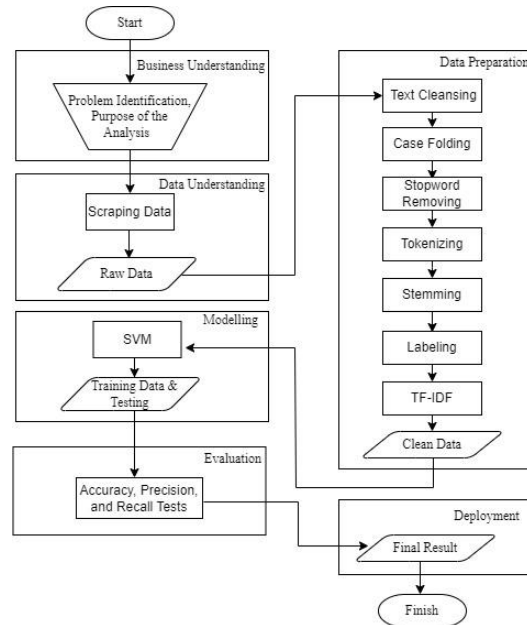


Fig. 1: Methodology

2.1. Data Collection

The data collection phase was conducted to obtain user reviews of the Duolingo app for use in the sentiment analysis process. The data was obtained from the Google Play Store using the google-play-scraper library, which was run on Google Colab. The data collection process focused on Indonesian-language reviews published between March 10, 2025, and March 26, 2026, resulting in a total of 1,220 reviews. All successfully collected data was then saved in CSV format to facilitate processing and analysis in subsequent stages.

2.2. Text Preprocessing

After the data collection stage, text preprocessing was performed to improve the quality of the data before sentiment classification. This process aims to reduce noise and normalize the text, making the data ready for further analysis. In this study, the text preprocessing stage consisted of cleaning, case folding, stopwords removal, tokenization, slang word normalization, and stemming [5]. Cleaning was performed to remove unnecessary characters, symbols, numbers, and punctuation marks. Case folding converted all text into lowercase letters. Stopword removal eliminated common words that did not contribute significantly to sentiment analysis. Tokenization split each sentence into individual words or tokens. Slang word normalization transformed non-standard words into their standard forms using a normalization dictionary. Finally, stemming reduced words to their root forms to produce a more consistent text representation.

2.3. Sentiment Labeling

After the text preprocessing stage, sentiment labeling was performed using a lexicon-based approach. Each review was assigned a sentiment score by calculating the total weight of positive and negative words contained in the text. Reviews with a positive total score were labeled as positive, reviews with a negative total score were labeled as negative, and reviews with a score of zero were labeled as neutral. The sentiment labels obtained from this process were then used as the target classes for the Support Vector Machine classification model.

2.4. Split Dataset

After the sentiment labeling stage, the dataset was divided into training and testing data using a 70:30 ratio. The training data were used to develop the Support Vector Machine (SVM) classification model, while the testing data were used to evaluate its performance. This data split was performed to ensure that the model could be trained effectively and assessed objectively using evaluation metrics, including accuracy, precision, recall, and F1-score.

2.5. Term Frequency-Inverse Document Frequency Weighting (TF-IDF)

After the data splitting stage, the review texts were transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting method. TF-IDF was applied to measure the importance of each word based on its frequency within a review and across the entire dataset. The resulting feature vectors were then used as input for the Support Vector Machine (SVM) classification model [6].

2.6. Support Vector Machine (SVM) Training

Following the TF-IDF feature extraction stage, the sentiment classification model was trained using the Support Vector Machine (SVM) algorithm with a linear kernel. The training data represented as TF-IDF feature vectors were used to build the classification model, while the testing data were used to evaluate its performance. The linear kernel was selected because it is effective for text classification tasks and provides efficient computation by separating data using a linear decision boundary. The trained model was then used to classify Duolingo user reviews into positive, negative, and neutral sentiment categories [7].

2.7. Model Evaluation

The model evaluation stage was conducted to assess the performance of the Support Vector Machine (SVM) model. The trained model was tested using the testing dataset to determine its ability to classify the sentiment of Duolingo user reviews. The model was evaluated using a confusion matrix with four performance metrics: accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model's predictions. Precision evaluates how accurately the model classifies each sentiment class. Recall measures the model's ability to identify all instances belonging to the correct sentiment class. Meanwhile, F1-score combines precision and recall into a single metric to provide a balanced evaluation of the model's performance.

3. Result and Discussion

This section presents the evaluation results of the Support Vector Machine (SVM) model developed for classifying the sentiment of Duolingo user reviews. The model performance was evaluated using a confusion matrix based on accuracy, precision, recall, and F1-score metrics. In addition, this section presents the sentiment distribution obtained from the classification results to illustrate the overall tendency of user opinions toward the Duolingo application.

3.1. Accuracy

Based on the experimental results, the Support Vector Machine (SVM) model achieved an accuracy of 79.23% in classifying the sentiment of Duolingo user reviews from Google Play. According to the classification report, the positive class obtained a precision of 0.89, a recall of 0.82, and an F1-score of 0.86. The negative class achieved 0.80 for precision, recall, and F1-score. Meanwhile, the neutral class obtained a precision of 0.30, a recall of 0.50, and an F1-score of 0.38. These results indicate that the proposed model performs well in classifying positive and negative sentiments, while its performance in identifying neutral sentiments remains relatively low.

HASIL EVALUASI MODEL SVM

Akurasi : 79.23%

Classification Report

	precision	recall	f1-score	support
negatif	0.80	0.80	0.80	111
netral	0.30	0.50	0.38	28
positif	0.89	0.82	0.86	227
accuracy			0.79	366
macro avg	0.67	0.71	0.68	366
weighted avg	0.82	0.79	0.80	366

Fig. 2: Classification Report of the SVM Model

3.2. Confusion Matrix

The confusion matrix results indicate that the Support Vector Machine (SVM) model correctly classified 89 negative reviews, 14 neutral reviews, and 187 positive reviews. These values are located on the main diagonal of the confusion matrix, representing the number of correctly classified instances. Meanwhile, the values outside the main diagonal indicate that several reviews were misclassified into different sentiment categories. Nevertheless, the number of correct predictions was higher than the number of misclassifications, indicating that the model was able to distinguish positive, negative, and neutral sentiments effectively. These findings are consistent with the obtained accuracy score, demonstrating that the Support Vector Machine (SVM) algorithm achieved satisfactory performance in classifying the sentiment of Duolingo user reviews on Google Play.

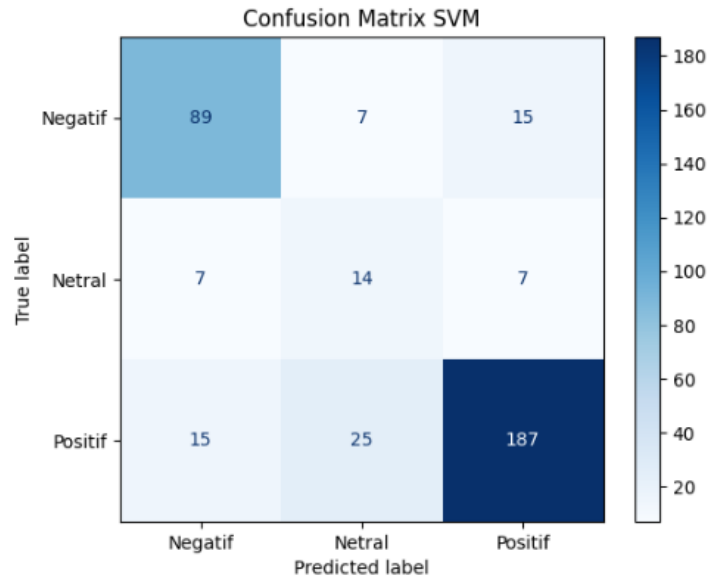


Fig. 3: Confusion Matrix of the SVM Model

3.3. Sentiment Distribution

The sentiment distribution analysis of 1,220 Duolingo user reviews shows that positive sentiment dominates with 57.1%, followed by negative sentiment at 30.3% and neutral sentiment at 12.6%. These results indicate that most users have a positive perception of the Duolingo application, suggesting that its features and learning experience have generally met user expectations. However, the presence of negative and neutral reviews indicates that some users still encountered issues or expressed concerns regarding certain aspects of the application. Therefore, these findings can serve as valuable feedback for developers to further improve the quality and user experience of the Duolingo application.

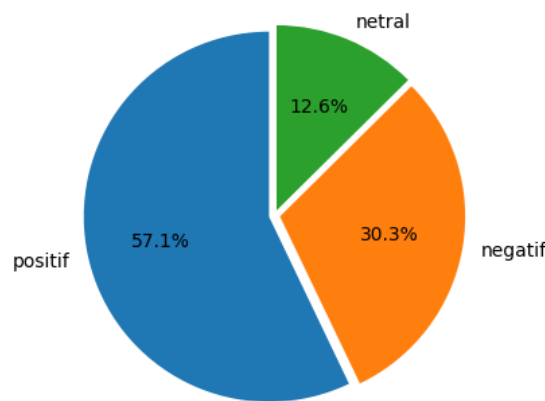


Fig. 4: Sentiment Distribution of Duolingo User Reviews

4. Conclusion

This study analyzed the sentiment of Duolingo user reviews on Google Play using the Support Vector Machine (SVM) algorithm with TF-IDF feature weighting. The results showed that user reviews were classified into three sentiment categories, namely 57.1% positive, 30.3% negative, and 12.6% neutral, indicating that most users expressed positive opinions toward the application. Furthermore, the proposed SVM model achieved an accuracy of 79.23%, demonstrating satisfactory performance in classifying user reviews into positive, negative, and neutral sentiment categories. These findings indicate that the combination of TF-IDF and SVM is effective for sentiment analysis of application reviews and can provide valuable insights for developers to evaluate and improve the quality of the Duolingo application.

References

- [1] A. Saninah, W. Prihartono, and C. L. Rohmat, "Analisis Sentimen Pengguna Terhadap Aplikasi Duolingo Dengan Naïve Bayes Classifier," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 1, pp. 620–629, 2025, doi: 10.23960/jitet.v13i1.5691.
- [2] M. F. Alamsah and R. Ardianto, "Memahami Sentimen Pengguna: Analisis Ulasan Aplikasi STRAVA di Google Play Store," *Jurnal Kolaborasi Riset Sarjana*, vol. 2, no. 2, pp. 36–44, 2025.
- [3] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.

- [4] Y. Yarkhamsetiawan, M. Akbar, and Satrianansyah, "Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 15, no. 2, pp. 181–193, 2025, doi: 10.34010/jamika.v15i2.16843..
- [5] M. I. Ghazali, W. H. Sugiharto, and A. F. Iskandar, "Analisis Sentimen Pinjaman Online di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1340–1348, 2023, doi: 10.30865/klik.v3i6.936.
- [6] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, pp. 2755–2762, 2024, doi: 10.23960/jitet.v12i3.4902.
- [7] M. A. Ramadhani, A. Khumaidi, A. M. Disrinama, M. A. Jami'in, Adianto, A. Z. Arfianto, I. Rachman, M. K. Hasin, Y. Widiarti, R. Y. Adhitya, Z. M. A. Putra, I. Munadhif, and D. P. Riananda, "Implementasi Algoritma Support Vector Machine (SVM) untuk Diagnosis Kesehatan Manusia Berbasis Web Application," *Jurnal Ners*, vol. 9, no. 1, pp. 896–902, 2025, doi: 10.31004/jn.v9i1.31481.