Sentiment Analysis of Shopee App Reviews on Google Play Store Using Machine Learning Models

Fanuel Oley[1,*] and Jutono Gondohanindijo[1]

[1]*Universitas AKI, Semarang, Indonesia*
[*]*Corresponding author. Email: jutono.gondohanindijo@unaki.ac.id*

ABSTRACT
This study aims to analyze user reviews of the Shopee application on the Google Play Store using machine learning models to classify sentiments into positive, neutral, and negative categories. Data collection was carried out by web scraping, resulting in 4,000 review samples. The preprocessing steps included cleaning, tokenization, stopword removal, stemming, and TF-IDF vectorization. Three classification algorithms were tested, namely Naive Bayes, Support Vector Machine (SVM), and Neural Network. Evaluation results showed that the Naive Bayes model achieved the highest accuracy at 81.2%, followed by SVM at 78.6%, and Neural Network at 75%. Furthermore, a sentiment analysis dashboard was developed using Gradio and deployed through Hugging Face to facilitate marketing decision-making based on user perceptions. This research proves the effectiveness of machine learning in sentiment analysis and offers actionable insights to enhance customer satisfaction.

Keywords: *Sentiment Analysis, Shopee, Machine Learning, Naive Bayes, SVM, Neural Network*

1. INTRODUCTION

1.1 Background

The rapid growth of e-commerce platforms in Indonesia has heightened competition among online marketplace providers. Shopee, as one of the leading platforms, must maintain customer satisfaction to sustain its competitive advantage. User reviews on the Google Play Store provide valuable insights into customer perceptions. Negative reviews, if not managed properly, can damage Shopee's reputation and affect user trust (Ahmadi, 2021).

While it is difficult to estimate the direct damage figures, negative reviews can result in decreased downloads, decreased transactions, and ultimately affect a company's revenue. Decreased user trust can lead to decreased loyalty and increased churn rates, which have a significant impact on annual revenue.

Possible limitations include time and resource constraints to manually analyze review data, technical capabilities in understanding and effectively implementing data analysis, and challenges in integrating findings from the analysis into product development strategies in real time (Hudaya, 2019).

The solution offered is the development of a Shopee app review sentiment analysis dashboard that uses the Naive Bayes algorithm. This dashboard will allow business users to quickly and accurately identify sentiment trends, group reviews based on positive, negative, or neutral sentiment, and display this data in an easy-to-understand visual format (Liu, 2012).

1.2 Problem Statement

Although Shopee collects a large number of reviews, automatic sentiment classification is still underdeveloped. Determining which machine learning model yields the best sentiment classification performance remains a challenge.

1.3 Research Objectives

This research aims to evaluate the effectiveness of Naive Bayes, SVM, and Neural Network models in classifying user review sentiments. It also seeks to develop a dashboard that can provide marketing teams with insights from user feedback.

1.4 Scope and Limitation

This study only considers Shopee application reviews collected from the Google Play Store, written in Indonesian, during a specific period.

2. RESEARCH METHODS

2.1 Data Collection

This dataset comes from Shopee app user reviews taken from the Google Play Store. These reviews reflect users' opinions about their experience using the app, including criticism, suggestions, and appreciation. Review data was obtained by scraping the Shopee app page on Google Play Store using Python-based scraping tools. A total of 4,000 review entries were collected.

2.2 Text Preprocessing

Text Preprocessing is a series of steps to prepare raw text data into a clean, structured format that is ready to be used in analysis or modeling. This process is very important in text analysis, such as natural language processing (NLP), because text data usually has many variations, including noise, non-uniform formats, and irrelevant elements. The text preprocessing process involves several important steps to prepare the review data, namely (Hapsari, 2018) :

1. Case Folding
   - Changed all review text to lowercase to maintain consistency.
   - Example : " Very Good Item " → " very good stuff ".
2. Stopword Removal
   - Remove common words that do not contribute significantly to sentiment analysis, like "which", "and", "with".
   - Stopwords are removed using libraries such as nltk
3. Tokenizing
   - Breaking text into small units in the form of words or tokens.
   - Example : " very good stuff " → ["goods", "very", "good"].
4. Stemming
   - Change words to their base form to avoid redundancy.
   - Example : " delivery " → "send", "the best" → "very".

2.3 Modelling

Three machine learning models were developed :

- **Naive Bayes :** a probabilistic classifier based on Bayes' theorem.
- **SVM :** a supervised learning model that finds the optimal decision boundary.
- **Neural Network :** a basic feed-forward neural architecture.

2.4 Evaluation Metrics

Model performance was evaluated using :

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

3. RESULTS AND DISCUSSIONS

3.1 Data Description

The data was taken directly from the Google Play Store using web scraping techniques. The files used are in CSV format, making it easier for further analysis. This dataset provides a comprehensive picture of the user experience of the Shopee application, making it a highly relevant data source for sentiment analysis research. The dataset included review texts, ratings (1-5 stars), and sentiment labels generated based on the rating scores.

The dataset is divided into two parts :

- Training Data (Training Data) : 80% from the dataset, used to train the model.
- Testing Data : 20% from the dataset, used to measure model performance.

The splitting is done using the train_test_split() function from the sklearn library. This process is important to ensure that the model can generalize well on new data.

TF-IDF (Term Frequency - Inverse Document Frequency) is used to calculate the weight of each word in the review text. TF-IDF gives higher values to words that occur frequently in a single review but rarely occur across the dataset, thus helping to identify significant words.

3.3 Model Training and Evaluation

The models are trained using various machine learning algorithms, including Naive Bayes, SVM (Support Vector Machine), Neural Network.

Model performance visualizations, such as accuracy comparison graphs or confusion matrices, are created to provide further insight into the effectiveness of the model. The following are the evaluation results of each model..

Each model's performance was tested on the same dataset :

- Naive Bayes achieved 81.2% accuracy.
- SVM achieved 78.6% accuracy.
- Neural Network achieved 75% accuracy.

Table 1. The Comparison of Machine Learning in Accuracy

Model	Accuray
Naive Bayes	81.2%
Support Vector Machine (SVM)	78.6%
Neural Network	75%

3.4 Visualization Data

The Gradio dashboard displayed the real-time sentiment prediction results. Users could input text and receive instant sentiment classification along with word frequency visualizations.

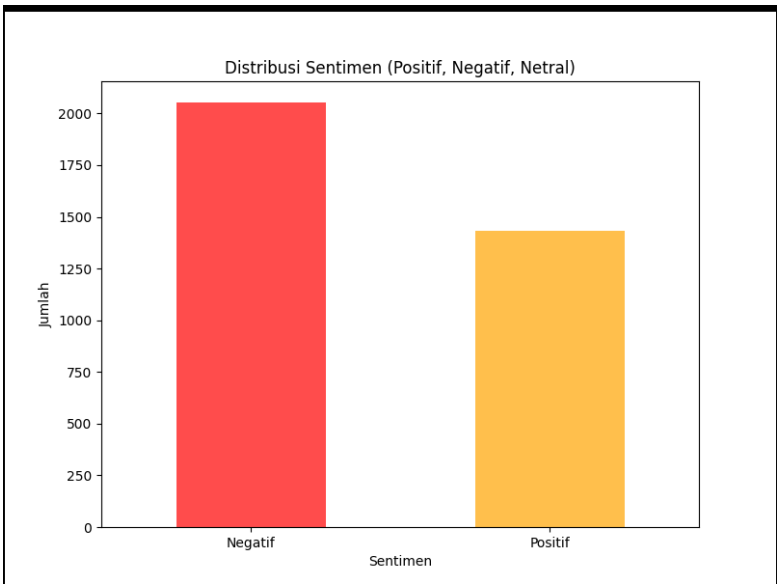

Figure 1 Sentiment Distribution

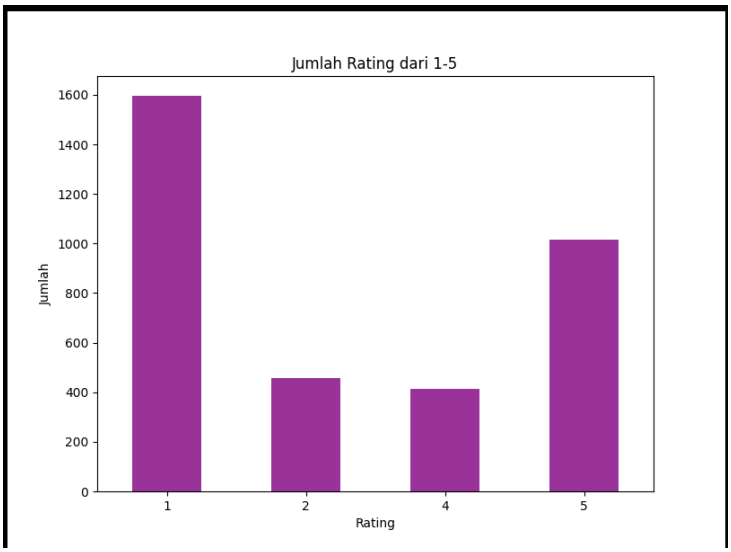

Figure 2 Number of Ratings from Label

Figure 3 WordCloud

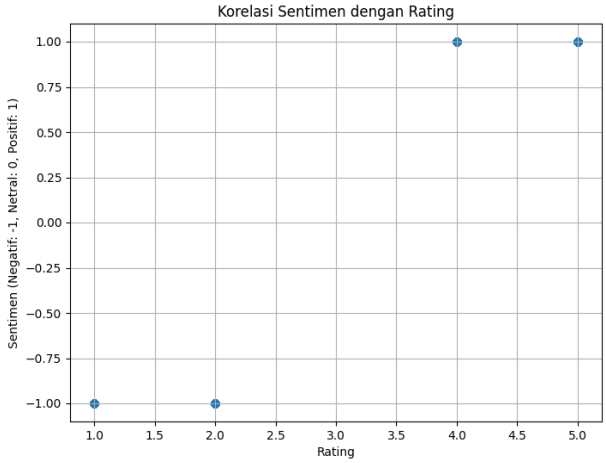

Figure 4 Correlation of Sentiment and Rating

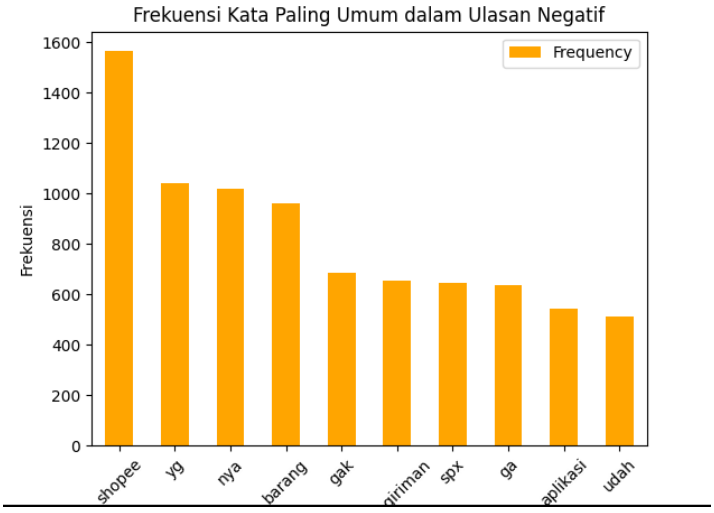

Figure 5 Reviews That Influence Negative Sentiment

4. CONCLUSION

This study successfully built a sentiment analysis system for Shopee app reviews using machine learning techniques. Naive Bayes outperformed SVM and Neural Network models based on evaluation metrics. Visualization through a Gradio dashboard proved useful in providing Shopee's marketing team with actionable insights. Future research should explore the use of deep learning methods and broader datasets.

REFERENCES

Ahmadi, M., Gustian, & Sembiring, Y. (2021). Implementation of Naive Bayes in Sentiment Analysis on Twitter. Jurnal Teknologi Informasi, 12(2), 11–16.

Falahah, and Dyar Dwiki Adriadi Nur. (2015). Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes (Studi Kasus Sentiment Analysis Dari Media Twitter). Seminar Nasional Sistem Informasi Indonesia (November): 335–40.

Hapsari, Yustia, Muhammad Fikri Hidayattullah, and Mohammad Khambali. (2018). Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes. Jurnal Informatika:Jurnal Pengembangan IT (JPIT) 03(02): 233–36.

Hayuningtyas, Ratih Yulia, Retno Sari, and Teknik Informatika. (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap. TECHNO Nusa Mandiri 16(1): 37–42.

Hudaya, A., Fakhrurroja, R., & Alamsyah, R. (2019). Sentiment Analysis using Text Mining Techniques. Jurnal Teknologi Informasi dan Ilmu Komputer, 6(1), 23–30.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

Siringoringo, Rimbun, and Jamaludin Jamaludin. (2019). Text Mining Dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online. Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP) 2(1): 41–48.

Wilana, Wivadario et al. 2018. Analisis Sentimen Terhadap Opini Masyarakat Indonesia Mengenai Bukalapak. Seminar Nasional Teknologi Informasi dan Multimedia 2018: 1–6.

Yane Marita Febrianti. (2018). Analisis Sentimen Pada Ulasan 'Lazada' Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-Nn). Repository.Ub.Ac.Id 2(Analisis Sentimen Pada Ulasan Lazada Berbahasa Indonesia Menggunakan K-Nearest Neighbor(K-NN)): 3689–98.