

## ABSTRAK

Halodoc merupakan aplikasi layanan kesehatan digital yang ada di Indonesia, sehingga penting untuk memahami persepsi pengguna melalui analisis sentimen. Penelitian ini bertujuan untuk menganalisis ulasan pengguna di Google Play Store menggunakan metode *Support Vector Machine* (SVM). Dua pendekatan ekstraksi fitur, yaitu *Bag of Words* (BoW) dan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk merepresentasikan data teks ke dalam bentuk numerik. Data ulasan dikategorikan menjadi sentimen positif dan negatif setelah melalui tahapan pra-pemrosesan yang mencakup pembersihan data, normalisasi huruf, tokenisasi, penghapusan *stopword*, dan *stemming*. Model diuji dengan berbagai nilai parameter C pada kernel *linear* untuk menemukan parameter terbaik. Hasil pengujian menunjukkan bahwa metode TF-IDF memberikan performa terbaik dengan akurasi 85%, presisi 86%, dan *F1-score* 86%, sedangkan BoW mencapai akurasi 84%, presisi 82%, dan *F1-score* 85%. Namun, BoW unggul dalam *recall* sebesar 88%, dibandingkan TF-IDF yang mencapai 86%, menunjukkan keunggulannya dalam mendeteksi ulasan positif. Penelitian ini membuktikan bahwa pemilihan metode ekstraksi fitur dan teknik pra-pemrosesan data yang sesuai memiliki dampak terhadap peningkatan performa model klasifikasi sentimen.

**Kata kunci** : Analisis Sentimen, *Support Vector Machine* , Halodoc, BoW, TF-IDF

## ABSTRACT

Halodoc is a digital healthcare service application in Indonesia, making it essential to understand user perceptions through sentiment analysis. This study aims to analyze user reviews on Google Play Store using the Support Vector Machine (SVM) method. Two feature extraction approaches, namely Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), are used to represent text data in numerical form. The review data is categorized into positive and negative sentiments after undergoing preprocessing stages, including data cleaning, letter normalization, tokenization, stopword removal, and stemming. The model is tested with various values of parameter C on the linear kernel to find the optimal parameter. The test results show that the TF-IDF method provides the best performance with an accuracy of 85%, precision of 86%, and F1-score of 86%, while BoW achieves an accuracy of 84%, precision of 82%, and F1-score of 85%. However, BoW excels in recall at 88%, compared to TF-IDF at 86%, demonstrating its superiority in detecting positive reviews. This study proves that selecting the appropriate feature extraction method and data preprocessing techniques significantly impacts the performance of sentiment classification models.

**Keywords** : Sentiment Analysis, Support Vector Machine, Halodoc, BoW, TF-IDF