

ABSTRACT

The development of digital healthcare services has driven an increase in users of telemedicine applications, one of which is Halodoc. User reviews available on digital platforms serve as an important source of information to understand user perceptions and satisfaction levels regarding the services provided. However, manual review analysis has limitations, primarily due to the large volume of data, the use of non-standard language, and the imbalanced distribution of sentiment classes. Therefore, a computational approach is required to perform sentiment analysis automatically. This study discusses the sentiment classification of Halodoc application user reviews using the Support Vector Machine (SVM) method and TF-IDF feature extraction. Review data was obtained from the Google Play Store and processed through several preprocessing stages, namely cleansing, case folding, tokenization, normalization, and stopword removal. Data labeling was conducted using two approaches: manual labeling and lexicon-based labeling. Furthermore, this study applied SMOTE and undersampling methods to address the data imbalance between sentiment classes. Testing was carried out under several scenarios using a linear SVM kernel and variations in the C parameter value, both with and without SMOTE. Model evaluation was performed using accuracy, precision, recall, and F1-score metrics. The test results indicate that the lexicon labeling scenario with the application of SMOTE yielded the best performance, achieving an accuracy of 0.979, precision of 0.993, recall of 0.983, and an F1-score of 0.988 at a C value of 1. Based on these results, it can be concluded that the SVM method combined with TF-IDF is capable of effectively classifying the sentiment of Halodoc user reviews, and the use of SMOTE can improve model performance on imbalanced data.

Keywords: Halodoc, Support Vector Machine, TF-IDF, SMOTE, Sentiment analysis.

ABSTRAK

Perkembangan layanan kesehatan digital mendorong meningkatnya pengguna aplikasi telemedicine, salah satunya adalah halodoc. Ulasan pengguna yang tersedia pada platform digital menjadi sumber informasi penting untuk mengetahui persepsi dan tingkat kepuasan pengguna terhadap layanan yang diberikan. Namun, analisis ulasan secara manual memiliki keterbatasan, terutama karena jumlah data yang besar, penggunaan Bahasa tidak baku, serta ketidakseimbangan distribusi kelas sentiment. Oleh sebab itu, diperlukan pendekatan komputasional untuk melakukan analisis sentimen secara otomatis. Penelitian ini membahas klasifikasi sentiment ulasan pengguna aplikasi Halodoc dengan menggunakan metode Support Vector Machine (SVM) dan ekstraksi fitur TF-IDF. Data ulasan diperoleh dari Google Play Store dan diproses melalui beberapa tahapan preprocessing, yaitu Cleansing, case folding, tokenization, Normalization, stopword removal. Pelabelan data dilakukan menggunakan dua pendekatan, yaitu pelabelan manual dan pelabelan berbasis leksikon. Selain itu, penelitian ini juga menerapkan metode SMOTE dan undersampling untuk menangani ketidakseimbangan jumlah data antar kelas sentiment. Pengujian dilakukan dengan beberapa skenario menggunakan kernel linear SVM dan variasi nilai parameter C, menggunakan SMOTE maupun tidak. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa skenario pelabelan leksikon dengan penerapan SMOTE memberikan hasil terbaik, dengan akurasi sebesar 0,979, presisi 0,993, recall 0,983, dan F1-score 0,988 pada Nilai C = 1. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode SVM dengan TF-IDF mampu mengklasifikasikan sentiment ulasan pengguna Halodoc dengan baik, serta penggunaan SMOTE dapat meningkatkan performa model pada data yang tidak seimbang.

Kata Kunci: Halodoc, *Support Vector Machine*, TF-IDF, SMOTE , Analisis sentimen.