

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi NewSakpole berdasarkan ulasan di Google Play Store setelah pembaruan aplikasi pada periode 10 Januari 2025 hingga 6 Juni 2025. Penelitian ini menggunakan pendekatan berbasis lexicon dan representasi teks Term Frequency-Inverse Document Frequency (TF-IDF) yang dipadukan dengan algoritma Support Vector Machine (SVM) berkernel linier. Sentimen diklasifikasikan ke dalam dua kategori, yaitu positif dan negatif. Tahapan penelitian meliputi preprocessing teks, labeling sentimen, transformasi teks menggunakan TF-IDF, klasifikasi menggunakan SVM, serta penanganan ketidakseimbangan kelas dengan metode Synthetic Minority Oversampling Technique (SMOTE) dan penambahan data eksternal berlabel positif *external augmentation*. Hasil menunjukkan bahwa mayoritas ulasan pengguna mengandung sentimen negatif, terutama terkait masalah teknis seperti error aplikasi, fitur yang tidak berfungsi, dan kendala pembayaran pajak online. Model baseline SVM menghasilkan akurasi sebesar 92,01% dan meningkat menjadi 92,78% setelah penerapan SMOTE. Penambahan data eksternal positif meningkatkan akurasi hingga 94,43%, dan kombinasi SMOTE dengan data eksternal menghasilkan akurasi terbaik sebesar 94,97%, semuanya termasuk dalam kategori *excellent classification*. Teknik SMOTE dan penambahan data eksternal terbukti efektif dalam menyeimbangkan distribusi kelas, meskipun SMOTE tidak lagi signifikan pada dataset yang telah seimbang. Secara keseluruhan, integrasi pendekatan leksikon, TF-IDF, SVM, serta strategi penyeimbangan kelas berhasil meningkatkan performa klasifikasi sentimen terhadap ulasan aplikasi NewSakpole.

Kata kunci: analisis sentimen, SVM, TF-IDF, lexicon, NewSakpole.

ABSTRACT

This study aims to analyze user sentiment toward the NewSakpole application based on user reviews from the Google Play Store following the app update that took place between January 10, 2025 and June 6, 2025. The research adopts a lexicon-based approach and Term Frequency-Inverse Document Frequency (TF-IDF) text representation, combined with the Support Vector Machine (SVM) algorithm using a linear kernel. Sentiment is classified into two categories: positive and negative. The research stages include text preprocessing, sentiment labeling, text transformation using TF-IDF, classification using SVM, and handling class imbalance through the Synthetic Minority Oversampling Technique (SMOTE) and external augmentation with positively labeled data. The results show that most user reviews express negative sentiment, particularly concerning technical issues such as application errors, non-functional features, and problems with online tax payments. The baseline SVM model achieved an accuracy of 92.01%, which increased to 92.78% after applying SMOTE. The addition of positive external data improved the accuracy to 94.43%, and the combination of SMOTE with external data yielded the highest accuracy of 94.97%, all falling under the category of excellent classification. Both SMOTE and external data augmentation proved effective in balancing the class distribution, although SMOTE had minimal impact once the dataset was already balanced. Overall, the integration of lexicon-based scoring, TF-IDF, SVM, and class balancing strategies significantly enhanced the performance of sentiment classification on NewSakpole application reviews.

Keywords: sentiment analysis, SVM, TF-IDF, lexicon, NewSakpole.