

ABSTRAK

Pembelajaran bahasa sangat dibutuhkan dalam era modern ini, karena dapat menunjang baik di bidang pendidikan maupun di bidang pekerjaan. Salah satu pembelajaran bahasa asing yang banyak dipelajari oleh masyarakat adalah bahasa Inggris. Aplikasi pembelajaran bahasa asing yang saat ini menjadi daya tarik masyarakat yaitu aplikasi Duolingo. Aplikasi ini memiliki banyak manfaat dan sudah dikenal luas, sehingga aplikasi ini memperoleh beragam tanggapan dari penggunanya, baik tanggapan yang positif ataupun negatif. Pada penelitian ini melakukan klasifikasi analisis sentimen terhadap ulasan pengguna Duolingo untuk mengetahui tingkat kepuasan pengguna terhadap aplikasi ini. Maka dari itu, penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) serta seleksi fitur *Chi Square* untuk meningkatkan performa algoritma. Selain itu, dalam penelitian ini akan membandingkan metode *resampling* SMOTE *oversampling* dan *Tomek Links undersampling* dalam menghasilkan akurasi yang optimal. Sebelum data ulasan diproses maka dilakukan tahapan *preprocessing*, pelabelan sentimen menggunakan *library Flair* dan ekstraksi fitur TF-IDF. Dari penelitian yang dilakukan, memperoleh akurasi tertinggi pada metode SVM tanpa *balancing* dengan kernel *gaussian RBF* dan $k\text{-best} = 1000$ sebesar 0.9059 atau 90.59%, SVM dengan SMOTE pada kernel *gaussian RBF* dan $k\text{-best} = 1000$ sebesar 0.9834 atau 98.34%, SVM dengan *Tomek Links* pada kernel *gaussian RBF* dan $k\text{-best} = 1000$ sebesar 0.9828 atau 98.28%. Hasil dari penelitian tersebut menunjukkan bahwa kernel *gaussian RBF* pada SVM memberikan akurasi optimal serta penerapan seleksi fitur *Chi Square* dan metode *resampling* SMOTE, *Tomek Links* dapat memberikan pengaruh dalam memperoleh akurasi yang optimal.

Kata kunci: Pembelajaran bahasa, Duolingo, Analisis Sentimen, *Support Vector Machine*, SMOTE, *Tomek Links*.

ABSTRACT

Language learning is essential in the modern era as it supports both educational and professional fields. One of the foreign languages widely studied by the public is English. An attractive language learning application currently popular among people is Duolingo. This application offers numerous benefits and is widely recognized, thus receiving various responses from its users, both positive and negative. This study conducts sentiment analysis classification on Duolingo user reviews to determine the level of user satisfaction with the application. Therefore, this study employs the Support Vector Machine (SVM) algorithm and Chi-Square feature selection to enhance the algorithm's performance. Additionally, this study compares the resampling methods of SMOTE oversampling and Tomek Links undersampling to achieve optimal accuracy. Before processing the review data, preprocessing steps, sentiment labeling using the Flair library, and TF-IDF feature extraction are performed. The study found that the highest accuracy was achieved using SVM without balancing with a Gaussian RBF kernel and k -best = 1000 at 0.9059 or 90.59%, SVM with SMOTE using a Gaussian RBF kernel and k -best = 1000 at 0.9834 or 98.34%, and SVM with Tomek Links using a Gaussian RBF kernel and k -best = 1000 at 0.9828 or 98.28%. The results of this study indicate that the Gaussian RBF kernel in SVM provides optimal accuracy and that the implementation of Chi-Square feature selection and the resampling methods SMOTE and Tomek Links can significantly influence achieving optimal accuracy.

Keywords: *Language Learning, Duolingo, Sentiment Analysis, Support Vector Machine, SMOTE, Tomek Links.*

