

Abstrak

Penelitian ini membahas mengenai penggunaan Support Vector Machine (SVM) dan Naïve Bayes untuk mengklasifikasikan kepribadian seseorang berdasarkan pola postingan yang dilakukan dalam sosial media. Penelitian ini mengintegrasikan fitur tekstual yang didapatkan dari Bag-of-Words (BoW) dan metode Term Frequency-Inverse Document Frequency (TF-IDF), bersamaan dengan perluasan fitur menggunakan Linguistic Inquiry dan Word Count (LIWC), untuk menilai pengaruhnya dalam keakuratan klasifikasi karakteristik kepribadian yang akan dipetakan dari postingan sosial media menggunakan Big Five Inventory (BFI-44). Hasil dari penelitian ini menunjukkan bahwa model SVM yang menggunakan rangkaian fitur TF-IDF + LIWC, menghasilkan performa yang terbaik, dan meraih akurasi 76.60% pada base land model dengan menggunakan kernel linear. Dalam perbandingannya, model Naïve Bayes memberikan performa terbaik dengan feature set yang sama, menghasilkan tingkat keakuratan sebanyak 59.57% dengan smooting parameter dari 1×10^{-2} . Meskipun oversampling meningkatkan recall dan presisi, undersampling ditemukan untuk menghasilkan efek negatif pada performa model. Beberapa penemuan ini menegaskan benefit dari mengombinasikan fitur TF-IDF dan LIWC yang dapat meningkatkan efektivitas model, dengan SVM yang menghasilkan penemuan terbaik secara keseluruhan dalam klasifikasi kepribadian dari data sosial media.