

Abstrak

Penggunaan media sosial meningkat dengan cepat. Berbagai macam informasi yang dapat diperoleh dari platform media sosial, salah satunya pada Twitter. Informasi biografi pengguna dan tulisan yang diungkapkan melalui *tweets* merupakan aset penting untuk penelitian yang dapat dimanfaatkan untuk menggambarkan kepribadian *Big 5*, yang terdiri dari *openness*, *conscientiousness*, *extraversion*, *agreeableness*, dan *neuroticism*. Beberapa penelitian sebelumnya telah melakukan prediksi kepribadian *Big 5*. Namun terdapat permasalahan yang ditemukan penulis yakni bagaimana cara meningkatkan performansi dari sistem prediksi kepribadian. Oleh karena itu, penelitian ini bertujuan untuk memprediksi kepribadian *Big 5* pada pengguna Twitter dan meningkatkan performansi dari sistem prediksi kepribadian. Sistem prediksi dibangun menggunakan metode klasifikasi *Support Vector Machine* (SVM). Dalam meningkatkan performansi diterapkan teknik *sampling*, pemilihan fitur dan *hyperparameter tuning*. Penelitian ini juga menerapkan ekstraksi fitur linguistik yaitu TF-IDF dan LIWC. Dengan menggunakan 287 data pengguna twitter yang diperoleh dari *online survey*, dan menerapkan semua teknik optimasi didapat hasil akurasi sebesar 84,22%.

Kata kunci: kepribadian *Big 5*, SVM, LIWC, TF-IDF, optimasi