

ABSTRAK

Kepribadian mencerminkan pola pikir, emosi, dan perilaku individu, dan salah satu model yang digunakan untuk menganalisisnya adalah *Big Five Personality*, yang terdiri dari lima dimensi utama yaitu *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism*. Penelitian ini bertujuan untuk mengklasifikasikan kepribadian *Big Five* menggunakan data dari media sosial X dengan metode RoBERTa. Kendala utama dalam penelitian ini adalah keterbatasan *dataset* yang hanya mencakup 381 pengguna serta ketidakseimbangan distribusi data antar label, yang memengaruhi akurasi prediksi.

Untuk mengatasi masalah ini, digunakan teknik *Random Oversampling* dan *Easy Data Augmentation* untuk menyeimbangkan dan memperkaya data. Model diuji dengan dua jenis *preprocessing*, yaitu *Half Preprocessing* dan *Full Preprocessing*, serta *tuning hyperparameter* menggunakan *Random Search*. Hasil eksperimen menunjukkan bahwa kombinasi *Data Augmentation* dan *Random Oversampling* memberikan performa terbaik, dengan akurasi 73,73% dan *F1-score* 73,37% pada skenario *Full Preprocessing*. Penelitian ini memberikan kontribusi dalam pengembangan metode klasifikasi berbasis teks untuk analisis kepribadian pada *dataset* kecil.

Kata Kunci: kepribadian *Big Five*, RoBERTa, *Random Oversampling*, *Data Augmentation*, media sosial, klasifikasi teks.