

Abstrak

Personalitas adalah sekumpulan motivasi, perasaan, dan perilaku unik yang dimiliki manusia. Pendeteksian personalitas di media sosial merupakan topik penelitian yang umum dilakukan dalam ilmu komputer. Model personalitas yang sering digunakan untuk penelitian pendeteksian personalitas adalah model *Big Five Indicator* (BFI) dan *Myers-Briggs Type Indicator* (MBTI). Berbeda dengan BFI yang mengklasifikasikan personalitas berdasarkan sifat-sifat individu, model MBTI mengklasifikasikan personalitas berdasarkan tipe individu. Jadi, MBTI berkinerja lebih baik dalam beberapa skenario daripada model *Big Five*. Banyak penelitian yang menggunakan *machine learning* untuk mendeteksi personalitas di media sosial, seperti *Logistic Regression*, *Naïve Bayes*, dan *Support Vector Machine*. Dengan popularitas *Deep Learning* baru-baru ini, kita dapat menggunakan model bahasa seperti DistilBERT untuk mengklasifikasikan personalitas di media sosial. Dengan kemampuan DistilBERT untuk memproses kalimat besar dan kemampuan paralelisasi berkat arsitektur transformers. Oleh karena itu, penelitian yang diusulkan adalah mendeteksi personalitas MBTI di Reddit menggunakan DistilBERT. Hasil evaluasi menunjukkan bahwa penghapusan *stopwords* pada tahap *preprocessing* data dapat menurunkan performa model, dan dengan penanganan ketidakseimbangan data, DistilBERT memiliki performa yang lebih buruk dibandingkan tanpa penanganan ketidakseimbangan data. Selain itu, sebagai perbandingan, DistilBERT mengungguli pengklasifikasi pembelajaran mesin lainnya seperti *Naïve Bayes*, *SVM*, dan *Logistic Regression* dalam hal *accuracy*, *precision*, *recall*, dan *f1-score*.

Kata Kunci : Deteksi Personalitas, Reddit, DistilBERT.

