

1. Pendahuluan

1.1. Latar Belakang

Personalitas adalah seperangkat motivasi, perasaan, dan perilaku unik yang dimiliki manusia. Personalitas sangat memengaruhi kehidupan, pilihan hidup, sikap, kesehatan, dan keinginan seseorang [1]. Setiap orang memiliki personalitas yang berbeda. Telah banyak penelitian mengenai personalitas pada seorang individu atau sekelompok individu. Pendeteksian personalitas pada media sosial merupakan salah satu topik penelitian yang banyak dilakukan di bidang Ilmu Komputer. Sebagai contoh Shantika, et al. melakukan penelitian mengenai pendeteksian personalitas *Big Five* pengguna Twitter, dimana dari 549.151 *tweet* dan 287 akun Twitter, mayoritas pengguna memiliki personalitas *Openness* sebanyak 91 pengguna sedangkan kelompok personalitas yang paling sedikit penggunaannya adalah personalitas *Neuroticism* dengan 43 pengguna Twitter [2].

Selain sifat personalitas *Big Five*, personalitas juga dapat diidentifikasi dengan menggunakan model *Myers-Briggs Type Indicator* (MBTI). *Extraversion-Introversion* (E/I), *Sensing-Intuition* (S/N), *Thinking Feeling* (T/F), dan *Judging-Perceiving* (J/P) merupakan empat kelas biner yang mendefinisikan MBTI yang kemudian dikombinasikan menjadi 16 tipe personalitas, sedangkan BFI mendefinisikan personalitas seseorang berdasarkan berdasarkan 5 skalar yaitu *Extraversion*, *Neuroticism*, *Agreeableness*, *Conscientiousness*, dan *Openness* [3]. Berbeda dengan model *Big Five* yang mengklasifikasikan personalitas berdasarkan sifat-sifat individu, MBTI mengklasifikasikan personalitas berdasarkan tipe individu itu sendiri [4]. Model personalitas MBTI berkinerja lebih baik dalam beberapa skenario daripada model personalitas *Big Five* [3].

Sebagian besar penelitian tentang deteksi personalitas dilakukan di media sosial Twitter karena memiliki API (*Application Program Interface*) yang bersifat publik sehingga memudahkan para peneliti untuk mendapatkan data yang berkualitas. [5]. Namun, hanya beberapa penelitian tentang deteksi personalitas yang dilakukan di forum *online* Reddit. Menurut Proferes, et al., Reddit telah mulai mendekati Twitter untuk kemudahan pengumpulan data dan memberikan manfaat tambahan untuk penelitian [6]. Sebagai contoh, struktur *subreddit* memudahkan peneliti untuk mencari data berdasarkan topik yang relevan dengan penelitian mereka. Topik yang didapatkan di Reddit juga lebih berkualitas karena adanya moderator dan aturan yang harus ditaati dalam sebuah *subreddit* [7]. Oleh karena itu, dalam penelitian ini, dilakukan deteksi personalitas MBTI di Reddit menggunakan DistilBERT.

1.2. Topik dan Batasannya

Topik dan batasan masalah pada penelitian ini adalah mengetahui performa model bahasa DistilBERT dalam mendeteksi personalitas di komunitas Reddit dengan beberapa tipe pengujian. Dataset yang digunakan adalah komentar yang di unggah di komunitas Reddit.

1.3. Tujuan

Tujuan dari penelitian ini adalah melakukan deteksi personalitas di komunitas Reddit dengan model bahasa DistilBERT, mengetahui performa hasil prediksi personalitas DistilBERT dengan tahap *preprocessing* data yang berbeda-beda, dan mengetahui perbandingan performa metode DistilBERT dengan metode pembelajaran mesin lainnya