

1. Pendahuluan

Kepribadian merupakan suatu pola pikir milik individu yang berlandaskan perilaku, perasaan, dan mental, sehingga membedakan kualitas atau karakteristik masing-masing individu [1]. Kepribadian mampu mencerminkan karakteristik dari pemikiran dan perilaku seseorang [2]. Model Big Five Personality merupakan salah satu model kepribadian yang terkenal efektif dalam mengidentifikasi kepribadian seseorang [3]. Model Big Five ini mampu memprediksi kepribadian seseorang dengan berbagai macam kegunaan seperti analisis calon pelamar kerja, analisis kecocokan dalam hubungan, dan lainnya [4]. Kepribadian seseorang dapat dianalisis melalui hal-hal yang mereka bagikan di media sosial seperti Twitter [5]. Pada penelitian ini akan diimplementasikan model kepribadian Big Five Personality dalam melakukan prediksi kepribadian pengguna Twitter.

Beberapa tahun terakhir Twitter sebagai media sosial kembali aktif dan populer di berbagai kalangan. Pengguna Twitter umumnya cenderung membagikan keluh kesah, suasana hati dan keadaan yang mereka sedang alami dalam bentuk teks. Segala bentuk perilaku seseorang di media sosial akan dengan mudah diperoleh, disimpan, dan dianalisis [6]. Hal tersebut menjadikan Twitter sebagai media sosial yang dapat dimanfaatkan dengan efektif untuk melakukan analisis serta prediksi kepribadian seseorang.

Berdasarkan penelitian [2], penulis membandingkan performansi tiga metode klasifikasi, yaitu K-Nearest Neighbor (KNN), Multinomial Naïve Bayes (MNB), dan Support Vector Machine (SVM) dalam melakukan klasifikasi kepribadian. Berdasarkan perbandingan ketiga metode tersebut, diperoleh kesimpulan bahwa metode klasifikasi yang memiliki performansi terbaik adalah metode MNB dengan nilai akurasi 60%. Penelitian prediksi kepribadian lain yang mengimplementasikan metode Naïve Bayes memperoleh akurasi sebesar 86,66% dengan pengaruh tahapan praproses dan pembobotan kata berdasarkan Term Frequency [7]. Pada penelitian, yang mengimplementasikan metode Support Vector Machine (SVM) dalam memprediksi kepribadian hanya mendapatkan tingkat akurasi terbaik dengan nilai 88,40% [8].

Pendekatan lain yang dapat meningkatkan nilai akurasi dari prediksi kepribadian pada Twitter adalah dengan mengimplementasikan salah satu model Semantic Approach [9]. Bidirectional Encoder Representation from Transformers (BERT) merupakan salah satu bentuk implementasi pendekatan semantik yang mampu mencapai akurasi 92% dalam proses klasifikasi teks [10]. Metode BERT juga telah diterapkan dalam sentiment analisis dengan hasil performansi yang cukup baik dengan nilai akurasi 82% [11]. Penelitian lain menerapkan Linguistic Inquiry Word Count (LIWC) untuk mendeteksi makna psikologis dari berbagai kata [12]. Adanya kedua pendekatan ini mampu memahami penafsiran dari kata-kata yang ada dan menghasilkan tingkat akurasi yang lebih baik.

Berdasarkan beberapa penelitian sebelumnya, motivasi pada Penelitian ini akan dilakukan eksperimen untuk meningkatkan nilai akurasi dalam melakukan prediksi kepribadian pengguna Twitter menggunakan metode Support Vector Machine (SVM) dan BERT sebagai pendekatan semantik. Metode ini dipilih karena memiliki performansi yang lebih baik dan efektif untuk melakukan klasifikasi pada data khususnya text-document-based [13]. Namun, pada penelitian sebelumnya hasil performansi dan nilai akurasi dari penerapan metode ini belum maksimal. Untuk meningkatkan nilai akurasi dan performansi dari metode tersebut akan diimplementasikan LIWC [9].

Rumusan masalah pada penelitian ini adalah mengimplementasikan metode SVM dan BERT dalam melakukan prediksi kepribadian pengguna Twitter di Indonesia. Batasan masalah pada penelitian ini adalah data yang digunakan hanya data dengan Bahasa Indonesia dengan total data 295 pengguna Twitter dengan 511.617 tweets data. Hasil analisis yang ingin dicapai berupa perbandingan nilai performansi dari model yang dibangun.

Tujuan yang ingin dicapai pada penelitian ini adalah mencapai nilai performansi yang optimal pada model prediksi kepribadian pengguna Twitter di Indonesia dengan mengimplementasikan kombinasi metode SVM dan metode BERT sebagai pendekatan semantik.

Penelitian ini terdiri atas empat bagian. Bagian 1 menjelaskan latar belakang, rumusan masalah, dan tujuan dari penelitian ini. Bagian 2 mendiskusikan studi terkait mengenai prediksi kepribadian Big Five. Bagian 3 memaparkan rancangan model yang dibangun. Bagian 4 membahas terkait evaluasi model yang dibangun dengan menjelaskan hasil analisis yang diperoleh. Bagian 5 merupakan bagian penutup yang terdiri dari kesimpulan dari penelitian ini.