

1. Pendahuluan

1.1. Latar Belakang

Media sosial merupakan media daring yang dapat memfasilitasi pengguna untuk bersosialisasi, berbagi informasi, serta mengekspresikan diri mereka dalam sebuah jejaring sosial. Salah satu media sosial dengan pengguna terbanyak adalah Twitter, dengan memiliki lebih dari 330 milyar pengguna aktif perbulan [3]. Twitter memberi kemudahan bagi pengguna untuk berbagi pikiran melalui postingan berupa teks atau biasa disebut *tweet*. Aktivitas tersebut tentu bersifat personal dan sangat terkait dengan psikososial seseorang sehingga dari kumpulan teks berbentuk *tweet* tersebut dapat diambil informasi seperti jenis emosi, sentimen, dan kepribadian seseorang.

Kepribadian seseorang penting untuk diketahui demi memahami potensi kebutuhan pengguna berdasarkan peristiwa yang berbeda-beda. Mendeteksi kepribadian pengguna juga berguna untuk meningkatkan akurasi sistem rekomendasi, pemasaran, dan membantu *human resource* untuk menentukan pekerjaan yang sesuai [1]. Kepribadian akan diprediksi dan diklasifikasikan menggunakan teori model *Big Five*, dimana dalam dunia psikologi dipercaya dapat mendeskripsikan struktur dasar dari kepribadian manusia. Model *Big Five* berpendapat bahwa kepribadian seseorang dapat diukur dengan menganalisis peringkat mereka kedalam lima kategori yaitu; *extraversion*, *agreeableness*, *conscientiousness*, *neuroticism*, dan *openness to experience* [11]. Model *Big Five* dipercaya merupakan model yang paling cocok digunakan untuk memprediksi kepribadian dari perilaku seseorang di media sosial [4]

Banyak penelitian sebelumnya yang berhasil menemukan kepribadian pengguna media sosial berdasarkan data bahasa Inggris seperti yang dilakukan Tadesse et al. [1] dengan menggunakan fitur sosial dan Majumder et al. [19] yang menggunakan fitur deteksi emosi dan analisis sentimen. Namun untuk bahasa Indonesia masih belum dipahami dan diteliti dengan baik. Maka dari itu penelitian ini mencoba menggunakan data *tweet* bahasa Indonesia dengan penggunaan fitur sosial, analisis sentimen, dan deteksi emosi. Pembobotan kata dilakukan menggunakan TF-IDF seperti yang dilakukan Mohsen et al. [15]. Metode yang digunakan untuk mengklasifikasi tipe kepribadian untuk penelitian ini adalah *Naive Bayes*. *Naive Bayes* memiliki kelebihan implementasi yang cenderung tidak rumit dan memiliki waktu eksekusi yang cepat untuk data yang banyak serta efisiensi ruang [4].

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah disampaikan sebelumnya, maka rumusan masalah yang dapat diambil adalah:

1. Penggunaan fitur manakah yang akurasinya paling optimal untuk mendeteksi kepribadian *Big Five*?
2. Apa karakteristik masing-masing kepribadian *Big Five* menurut fitur emosi, sentimen, dan sosialnya?

1.3. Batasan Masalah

Adapun ruang lingkup yang menjadi batasan penelitian tugas akhir ini adalah:

1. Responden yang dipilih menjadi dataset adalah yang akunnya tidak dikunci atau *private*.
2. Akun yang dipilih harus memiliki minimal 200 jumlah *tweet*, dimana *retweet* tidak termasuk.
3. Data *tweet* diambil sebanyak 200 *tweet* dari yang paling baru.
4. Kamus emosi dan sentimen menggunakan kata berbahasa Indonesia yang sudah dilakukan *stemming*.

1.4. Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui penggunaan fitur yang akurasinya paling optimal untuk mendeteksi kepribadian *Big Five* diantara fitur sosial, deteksi emosi, dan analisis sentimen. Kemudian menganalisis karakteristik masing-masing kepribadian *Big Five* menurut fitur emosi, sentimen, dan sosialnya.

1.5. Organisasi Tulisan

Organisasi tulisan untuk jurnal ini terdiri atas lima bagian, bagian 1 merupakan penjelasan latar belakang masalah, perumusan masalah, batasan masalah, dan tujuan dari penelitian. Pada bagian 2 dijelaskan studi yang terkait dengan penelitian ini. Bagian 3 menjelaskan tentang sistem yang dibangun, bagaimana metode yang akan digunakan dan alurnya. Kemudian bagian 4 berisikan evaluasi terkait sistem dan metode yang dijelaskan pada bagian sebelumnya. Dan yang terakhir pada bagian 5 memberikan kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian berikutnya.