

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kepribadian merupakan keseluruhan bentuk perilaku seseorang individu yang dipunyai sebagai latar belakang pemiliknya terhadap lingkungannya. Setiap individu mempunyai ciri khas kepribadian masing-masing yang dibentuk dari kecil. Kepribadian dapat mempengaruhi pilihan hidup seseorang dalam berbagai hal dan mempengaruhi interaksi dengan orang lain dan lingkungan. Ada banyak metode klasifikasi kepribadian salah satunya yang banyak digunakan yaitu metode *Big five personality* yang terdiri dari *openness*, *conscientiousness*, *extraversion*, *agreeableness*, dan *neuroticism*[1]. Kepribadian dapat digunakan sebagai penilaian dalam berbagai hal seperti perekrutan, karir, dan kesehatan dengan melakukan tes kepribadian untuk mengetahui kepribadiannya[2]. Banyak metode untuk melakukan tes kepribadian dengan cara wawancara, observasi, atau survei kuesioner, namun metode ini masih kurang praktis dan mahal. Dalam sebuah studi baru-baru ini menunjukkan bahwa tes kepribadian dapat diperoleh dari media sosial dengan menganalisa teks yang mereka tulis[3][4][5], metode tes kepribadian ini lebih praktis dan murah.

Media sosial merupakan tempat seseorang untuk berinteraksi dengan individu dalam dunia maya, akun media sosial ini bersifat pribadi sehingga mencerminkan kepribadian seseorang tersebut. Salah satu media sosial yang banyak digunakan yaitu Twitter dengan jumlah pengguna 328 juta pada tahun 2017 di seluruh dunia. Penelitian tugas akhir ini menggunakan Twitter dikarenakan banyak digunakan dan pengguna twitter cenderung mencurahkan pendapatnya lewat kata-kata yang dibatasi 280 karakter sehingga lebih langsung menyampaikan maksud pesan. Kata-kata bisa menggambarkan kepribadian seseorang dengan menganalisa untuk mendapatkan informasi dari pengguna Twitter. Metode yang digunakan untuk menganalisa kata menggunakan *Linguistic Inquiry and Word Count (LIWC)* yaitu program analisis teks di bidang ilmu psikologi.

Pada penelitian sebelumnya [4][5] yang berkaitan dengan prediksi kepribadian melalui media sosial twitter memanfaatkan perpaduan teknik dari ilmu *data mining* dan *machine learning*. *Data mining* untuk mencari informasi yang berguna dari banyak data yang terdapat dalam twitter, sedangkan *machine learning* untuk mempelajari pola untuk pembelajaran mendapatkan klasifikasi kepribadian. Algoritma yang digunakan untuk prediksi kepribadian melalui Twitter pada penelitian [4] menggunakan *Naïve Bayes Classifier (NBC)*, *K-Nearest Neighbor (KNN)*, dan *Support Vector Machine (SVM)* dengan pembobotan TF-IDF menghasilkan nilai akurasi masing-masing yaitu NBC 68%, KNN 60%, dan SVM 61%.

Dalam penelitian tugas akhir ini mencoba melakukan prediksi kepribadian *big 5* melalui Twitter dengan menggunakan algoritma *Naïve Bayes Classifier* dikarenakan implementasi yang sederhana untuk dibangun dengan waktu komputasi yang relatif cepat dan tingkat akurasi yang tinggi dibandingkan dengan

algoritma yang lain[4]. Hal lain yang membedakan dengan penelitian [4] yaitu pengguna twitter berbahasa Indonesia dan menggunakan pembobotan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF) untuk membandingkan nilai akurasi kedua pembobotan tersebut.

1.2 Rumusan Masalah

Dari uraian latar belakang terdapat beberapa masalah, maka rumusan masalah pada tugas akhir ini adalah :

1. Bagaimana proses implementasi metode *Naive Bayes Clasifier* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF) dalam prediksi kepribadian seseorang dari postingan Twitternya ?
2. Bagaimana menguji dan menganalisis perfomansi dari sistem prediksi kepribadian seseorang dari postingan Twitternya dengan metode *Naive Bayes Clasifier* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF)?

1.3 Batasan Masalah

Batasan masalah untuk penelitian tugas akhir ini adalah :

1. Akun Twitter bahasa Indonesia.
2. Data tweet diambil dalam jangka waktu tertentu.
3. Jumlah data tweet diambil maksimal 100 tweet per akun.

1.4 Tujuan Penelitian

Tujuan dari penelitian tugas akhir ini adalah:

■ Mengimplementasikan metode *Naive Bayes Clasifier* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF) dalam prediksi kepribadian seseorang dari postingan Twitternya.

■ Menguji dan menganalisis perfomansi dari sistem prediksi kepribadian seseorang dari postingan Twitternya dengan metode *Naive Bayes Clasifie* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF).

1.5 Metodologi Penyelesaian Masalah

Adapun metodologi penelitian pada tugas akhir ini adalah :

1. Studi Literatur
Studi literatur yang dilakukan adalah pencarian dan pengumpulan referensi tentang penelitian yang berkaitan Twitter, *Big Five Personality*, *data mining*, klasifikasi menggunakan *Naive Bayes Clasifier*, dan pembobotan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF). Materi-materi yang didapat kemudian dipahami untuk mendukung penyelesaian penelitian tugas akhir ini.

2. Pengumpulan Data
Tahapan dalam penelitian ini dilakukan dua tahapan yaitu tahap pertama digunakan untuk pelabelan dan tahap kedua digunakan sebagai data pembelajaran.
3. Perancangan Sistem
Perancangan sistem merupakan proses membuat rancangan system yang dirancang untuk prediksi kepribadian berdasarkan *Big 5 Personality* pengguna Twitter menggunakan algoritma *Naïve Bayes Classifier* dengan pembobotan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF).
4. Implementasi
Tahap implementasi algoritma *Naïve Bayes Classifier* ke dalam sistem untuk klasifikasi pengguna Twitter sesuai kepribadian berdasarkan teori *Big 5 personality*.
5. Pengujian dan Analisis Hasil
Menganalisis hasil implementasi sistem untuk mengetahui kebenaran klasifikasi data dan tingkat akurasi.
6. Penyusunan Laporan Tugas Akhir
Mendokumentasikan penyelesaian Tugas Akhir ke dalam laporan tertulis yang bertujuan untuk menginformasikan hasil analisa.

1.6 Jadwal kegiatan

No	Kegiatan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur	■	■	■	■	■	■
2	Pengumpulan Data		■	■	■		
3	Perancangan Sistem			■	■		
4	Implementasi Sistem			■	■		
5	Analisis Hasil Pengujian				■	■	■
6	Penyusunan Laporan			■	■	■	■