

BAB I PENDAHULUAN

1.1.Latar Belakang

Twitter adalah sebuah situs jejaring sosial populer dan alat *micro-blogging* yang diluncurkan pada tahun 2006 [1]. Pengguna Twitter dapat mengirim dan membaca *post* teks maksimal 280 karakter. Cara kerja *tweet* (*post* pada Twitter), ketika pengguna Twitter mengirim *tweet*, *tweet* tersebut langsung dapat dilihat oleh semua *followers* (pengikut akun pembuat *tweet*) [2]. Dengan keterbatasan karakter yang dapat di *tweet*, sebuah *tweet* dapat mengandung informasi yang efektif untuk dibaca.

Informasi yang terkandung dalam sebuah *tweet* dapat mencerminkan kepribadian penggunanya. Kepribadian adalah konstruksi psikologis yang memperhitungkan perbedaan setiap individu pada manusia [3]. Prediksi kepribadian adalah salah satu tugas yang sulit dalam komputasi stilometri, ini melibatkan pendeteksian ciri kepribadian berdasarkan gaya penulisan *tweet*. Ada beberapa tipologi ciri kepribadian, salah satu yang terkenal adalah kepribadian *Big Five* [4].

Kepribadian *Big Five* merupakan salah satu metode yang dikenal dalam dunia psikologi untuk menginterpretasi kepribadian seseorang, terutama untuk menemukan hubungan kepribadian dengan lingkungan pekerjaan [5]. Kepribadian *Big Five* terdiri dari *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism*. Kepribadian *Big Five* telah terbukti berhubungan secara signifikan dengan perilaku pengguna di media sosial [6]. Maka, sebuah prediksi kepribadian *Big Five* pada pengguna Twitter melalui *tweet* yang dikirimnya dapat dilakukan dan memiliki manfaat dalam jangka Panjang.

Dalam penelitian sebelumnya [7], untuk prediksi kepribadian *Big Five* pada Twitter dengan menggunakan tiga algoritma klasifikasi, yaitu *Naïve Bayes* (NB), *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM). Hasil dari masing-masing algoritma diatas sebesar 63% menggunakan *Naïve Bayes*, 60% menggunakan KNN, dan 61% menggunakan SVM. Hasil dari penelitian ini dapat dijadikan acuan untuk bagian sumber daya manusia ketika melakukan wawancara kerja. Namun, penelitian sebelumnya hanya menggunakan satu pembobotan saja yaitu TF-IDF.

Dalam tugas akhir ini, penulis menggunakan metode *AdaBoost* dengan dua pembobotan, yaitu TF-IDF dan WIDF. Penelitian ini akan mencoba memperbaiki akurasi dari penelitian sebelumnya. Dalam implementasinya, akan membandingkan penggunaan metode *AdaBoost* dengan pembobotan TF-IDF dan metode *AdaBoost* dengan pembobotan WIDF. Hal

lain yang membedakan penelitian [7] dengan penelitian ini adalah metode penelitian yang menggunakan metode *AdaBoost*.

1.2. Rumusan Masalah

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

1. Bagaimana proses implementasi metode *AdaBoost* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF) dalam prediksi kepribadian seseorang dari kiriman *post* Twitternya?
2. Bagaimana menguji dan menganalisis performansi dari sistem prediksi kepribadian seseorang dari kiriman *post* Twitternya dengan metode *AdaBoost* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF)?

1.3. Batasan Masalah

Batasan masalah untuk penelitian tugas akhir ini adalah :

1. Akun Twitter bahasa Indonesia.
2. Data *tweet* diambil secara mingguan selama 4 bulan.
3. Jumlah data *tweet* diambil maksimal 100 *tweet* per akun.

1.4. Tujuan Penelitian

Berdasarkan masalah yang dipaparkan, tujuan dari penelitian tugas akhir ini adalah:

1. Mengimplementasikan metode *AdaBoost* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF) dalam prediksi kepribadian seseorang dari kiriman *post* Twitternya.
2. Menguji dan menganalisis performansi dari sistem prediksi kepribadian seseorang dari kiriman *post* Twitternya dengan metode *AdaBoost* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF).

1.5. Rencana Kegiatan

Untuk memenuhi tujuan penelitian, maka disusun rencana kegiatan pada tugas akhir ini sebagai berikut :

1. Studi Literatur

Pada tahap ini, akan dilakukan pencarian dan pengumpulan referensi tentang penelitian yang berkaitan dengan *Big Five Personality*, *data mining*, *machine learning*, klasifikasi menggunakan *AdaBoost*, dan pembobotan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF). Materi-materi yang dikumpulkan kemudian dipahami untuk mendukung penyelesaian penelitian tugas akhir ini.

2. Pengumpulan Data

Pada langkah ini, dilakukan pengambilan data (*crawling*) dari akun pengguna Twitter menggunakan API Twitter yang sudah ada.

3. Perancangan Sistem

Pada tahap ini, dilakukan perancangan sistem yang dirancang untuk prediksi kepribadian pengguna Twitter berdasarkan *Big Five Personality* menggunakan algoritma *AdaBoost* dengan pembobotan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Weighted Inverse Document Frequency* (WIDF).

4. Implementasi Sistem

Pada tahap ini, algoritma *AdaBoost* diimplementasikan ke dalam sistem untuk klasifikasi pengguna Twitter sesuai kepribadian berdasarkan teori *Big Five Personality*.

4. Analisis dan Hasil Implementasi

Pada tahap ini, hasil implementasi sistem dianalisis untuk mengetahui kebenaran klasifikasi data dan tingkat akurasi.

5. Penyusunan Laporan

Pada tahap ini, dilakukan dokumentasi penyelesaian tugas akhir ke dalam laporan tertulis yang bertujuan untuk menginformasikan hasil analisa.

1.6.Jadwal kegiatan

No	Kegiatan	Bulan					
		1	2	3	4	5	6
1	Studi Literatur						
2	Analisis dan pengumpulan Data						

3	Perancangan Sistem						
4	Implementasi Sistem						
5	Analisis dan hasil Implementasi						
6	Penyusunan Laporan						