

## 1. Pendahuluan

### 1.1 Latar Belakang

Kepribadian merupakan salah satu faktor yang mempengaruhi perilaku dan tindakan manusia. Menurut para ahli, kepribadian adalah atribut yang menampilkan apa, mengapa, dan bagaimana seseorang berperilaku di lingkungannya [1]. Banyak tipe kepribadian yang telah ditemukan, seperti Big Five, MBTI, dan DISC. Di antara tipe-tipe tersebut, Big Five merupakan model yang paling sering digunakan oleh para peneliti. Model ini memiliki pemahaman yang benar-benar integratif mengenai struktur kepribadian yang mencakup sifat-sifat kepribadian yang esensial dan psikiatri klinis [2]. Big Five, seperti yang dijelaskan oleh Costa & McCrae dalam Utami [3], merupakan alat tes yang dapat meningkatkan teori kepribadian, yang menggambarkan model untuk menemukan karakteristik kunci dalam menggambarkan kepribadian. Model Big Five juga sering digunakan dalam penelitian psikologi untuk mengevaluasi kontribusi individu dalam sebuah organisasi. Salah satu penelitian tersebut dilakukan oleh Febriyanti, dkk. [4], yang menggunakan Big Five Model untuk mengukur kesejahteraan sosial tenaga kerja mereka. Lebih lanjut, Celli dan Leprib [5] membuktikan bahwa Big Five menawarkan data dan aturan yang lebih mendalam daripada MBTI. Big Five terdiri dari Neuroticism (stabilitas emosi), Extroversion (kemampuan bersosialisasi), Openness (intelektualisme), Agreeableness (tingkat kepekaan), dan Conscientiousness (disiplin diri) [6].

Mendeteksi kepribadian seseorang adalah proses rumit yang sering kali melibatkan pengisian kuesioner panjang yang mengevaluasi banyak sifat kepribadian. Namun, hasilnya mungkin tidak akurat atau tidak sesuai jika responden perlu memahami pertanyaan sepenuhnya atau tidak mau memberikan jawaban yang jujur [7]. Oleh karena itu, diperlukan sebuah sistem yang dapat mendeteksi kepribadian seseorang tanpa perlu kuesioner yang panjang dan dengan hasil yang akurat. Di era teknologi ini, mendeteksi kepribadian seseorang dapat dilihat melalui interaksi online mereka. Karena pengguna media sosial memiliki banyak informasi yang tersedia mengenai dirinya, maka diperlukan sebuah sistem yang dapat mendeteksi individu melalui media sosial yang digunakannya [8]. Menurut Nuo Han dkk., media sosial dapat mengungkapkan kepribadian seseorang. Oleh karena itu, para penulis menyatakan bahwa ekspresi media sosial dapat mengungkapkan ciri-ciri kepribadian karena pengguna media sosial sering mengekspresikan perasaan mereka [9].

Salah satu platform media sosial yang banyak digunakan saat ini adalah Twitter. Ruby [10] menuliskan bahwa Twitter merupakan platform media sosial ke-15 yang paling sering digunakan secara global, dengan jumlah pengguna aktif Twitter sebanyak 450 juta orang pada tahun 2022. Hal ini menunjukkan bahwa Twitter merupakan platform media sosial yang paling banyak digunakan di Indonesia. Menurut Kemp [11] jumlah pengguna aktif Twitter di Indonesia pada tahun 2022 telah mencapai 18,45 juta. Jumlah pengguna aktif yang besar mengindikasikan bahwa jumlah tweet yang dihasilkan juga sama. Dengan menggunakan tweet, manusia dapat berkomunikasi satu sama lain untuk mengekspresikan perasaan mereka.

Banyak peneliti yang telah meneliti tentang deteksi kepribadian pada pengguna Twitter. Namun, hasil dari penelitian-penelitian tersebut hanya dapat dianggap kurang lebih akurat. Sebagai contoh, penelitian Angsaweni menunjukkan bahwa tingkat akurasi untuk mengidentifikasi ciri-ciri kepribadian Big Five dengan menggunakan metode AdaBoost hanya sebesar 53,57% [12]. Dalam penelitian Lydia, SVM digunakan untuk mengidentifikasi kepribadian tipe DISC, tetapi akurasi model hanya 53% [13]. Pratama [14] melakukan penelitian serupa dengan menggunakan metode yang berbeda, yaitu Random Forest dan mencapai tingkat akurasi sebesar 69,23%. Melihat tingkat akurasi yang dicapai pada penelitian-penelitian tersebut, maka diperlukan model klasifikasi yang dapat memberikan tingkat akurasi yang lebih tinggi.

Oleh karena itu, penelitian ini akan menggunakan algoritma XGBoost. XGBoost (Extreme Gradient Boosting) merupakan salah satu model turunan dari Gradient Boosting Model yang memiliki perhitungan yang cepat dan akurat, sehingga menjadi model yang memiliki performa terbaik di antara model-model Decision Tree yang ada. Algoritma ini sebelumnya telah digunakan oleh Nasution dan timnya [15] untuk menghitung tingkat akurasi klasifikasi diabetes, dan ditemukan bahwa XGBoost memiliki tingkat akurasi tertinggi (90,10%) dibandingkan dengan model pembanding, Naive Bayes (79,68%). Kurniawanda dan Tobing [16] juga melakukan penelitian dengan menggunakan XGBoost untuk menganalisa sentimen komentar di Instagram, dengan hasil tingkat akurasi sebesar 75,20%. Qi [17] membuktikan dalam penelitiannya mengenai klasifikasi kejahatan pencurian berbasis teks dengan menggunakan algoritma XGBoost dan menghasilkan nilai precision sebesar 0,96, recall 0,96 dan f1-score 0,96. Oleh karena itu, penerapan Algoritma XGBoost pada penelitian ini adalah untuk melakukan identifikasi kepribadian pada pengguna Twitter dengan hasil performa yang baik. Selain itu, Algoritma XGBoost mampu memprediksi kepribadian pengguna Twitter berdasarkan kepribadian pengguna yang sebenarnya.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka berikut beberapa rumusan masalah yang akan diangkat pada penelitian ini adalah:

1. Bagaimana metode Algoritma XGBoost untuk mendeteksi kepribadian Big Five pada pengguna Twitter ?
2. Apa saja faktor yang mempengaruhi performansi tingkat akurasi metode Algoritma XGBoost dalam mendeteksi kepribadian Big Five pada pengguna Twitter?

## 1.3 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini adalah sebagai berikut:

1. Data pengguna Twitter yang akan digunakan hanya yang berasal dari Indonesia.
2. Bahasa tweet yang akan digunakan hanya yang berbahasa Indonesia.
3. Jenis kepribadian yang akan digunakan yaitu model Big Five Personality.

## 1.4 Tujuan

Tujuan penelitian yang akan dicapai berdasarkan rumusan masalah di atas adalah sebagai berikut:

1. Menerapkan metode Algoritma XGBoost untuk mendeteksi kepribadian Big Five pada pengguna Twitter.
2. Menganalisis faktor yang mempengaruhi nilai performansi akurasi metode Algoritma XGBoost dalam mendeteksi kepribadian Big Five.