

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Interaksi di masa kini dapat dilakukan secara tidak langsung ataupun tidak selalu dilakukan di satu tempat yang sama, salah satu media untuk melakukan interaksi tidak langsung adalah media sosial, dimana media sosial dapat membantu manusia untuk melakukan interaksi, berbagi, dan mencari informasi dengan jarak jauh secara *online* tanpa dibatasi jarak [1].

Perkembangan media sosial sangat pesat baik dari segi kualitas platform maupun kuantitas pengguna, X (Twitter) merupakan salah satu platform media sosial paling banyak digunakan terutama di Indonesia. Terdapat 27,5 juta pengguna X di Indonesia pada Oktober 2023, dengan jumlah sebesar itu menempatkan Indonesia sebagai negara keempat pengguna X paling banyak secara global, dibawah Amerika Serikat, Jepang, dan India [2]. X sendiri merupakan sebuah platform dimana pengguna dapat melakukan pengiriman dan juga membaca pesan yang berbentuk teks hingga 280 karakter dengan sebutan cuitan [3]. Ditinjau dari jumlah yang banyak dan juga interaksi yang berisikan banyak informasi sehingga dapat memudahkan dalam pengumpulan data itulah yang menjadikan X sebagai salah satu media sosial yang cocok untuk dipakai dalam penelitian ini [4].

Kepribadian merupakan gambaran dari sifat yang dimiliki masing-masing individu yang biasanya bersifat unik atau berbeda dari yang lain. Tiap individu harus mengetahui bagaimana kepribadian masing-masing dengan harapan dapat memaksimalkan potensi yang dimiliki tiap individu berdasarkan kepribadiannya [5]. *Big Five* merupakan salah satu teori sifat kepribadian yang diperkenalkan oleh Lewis Goldberg dengan lima dimensi kunci kepribadian, diantaranya adalah *Openness to Experience*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan juga *Neuroticism* [6]. Pemilihan *Big Five* sebagai kelas kepribadian dalam penelitian ini karena *Big Five* memiliki validasi yang kuat dalam segi psikometrik dan sudah banyak digunakan di berbagai bidang seperti analisis sosial media [7].

Klasifikasi kepribadian bisa didapatkan dengan data yang berasal dari media sosial, contohnya adalah X karena pada cuitan yang terdapat pada X dapat diambil data yang berbentuk teks dan dapat dianalisis untuk didapatkan pola yang merujuk

kepada klasifikasi kepribadian. Penelitian ini membandingkan metode *Logistic Regression* dan *Naive Bayes* alasan dipilihnya kedua metode ini karena efektivitasnya dalam menjalankan tugas klasifikasi teks. Dalam penelitian ini juga digunakan data interaksi berupa jumlah *following* dan *followers* dengan tujuan untuk mencari tahu bagaimana pengaruh data interaksi ini bagi kedua metode dalam menjalankan prediksi kepribadian. Dengan menemukan model yang tepat dalam penelitian ini, maka diharapkan dapat membantu mengetahui kepribadian masing-masing orang agar mendapatkan gambaran dalam pengembangan potensi sesuai kepribadian yang dimiliki.

Penelitian ini juga mengacu dari beberapa penelitian sebelumnya, Penelitian yang dilakukan oleh Mawadatul dalam klasifikasi kepribadian menggunakan beberapa algoritma *machine learning* menunjukkan bahwa metode *Logistic Regression* memiliki nilai akurasi sebesar 87,21% yang merupakan nilai terbesar dibanding algoritma lain seperti *Adam's Optimizer* dengan hasil akurasi 80,73% [8]. Penelitian yang dilakukan oleh Aristin dan Ema dalam penelitian tentang perbandingan beberapa algoritma untuk klasifikasi karakter individu menunjukkan hasil jika metode *Naive Bayes* memiliki akurasi paling tinggi dengan nilai sebesar 31,5% dibanding metode *k-Nearest Neighbor* dengan nilai akurasi sebesar 23,8% dan metode *Support Vector Machine* dengan nilai akurasi sebesar 28,4% [9]. Pada penelitian yang dilakukan oleh Andi, Kusriani, dan Asro tentang prediksi kepribadian berdasarkan status sosial media Facebook juga menunjukkan bahwa metode *Naive Bayes* memiliki nilai akurasi sampai 100% dibanding dengan metode *K-Nearest Neighbor* yang hanya memiliki nilai akurasi 58,96% [10]. Penelitian yang dilakukan oleh Mutiara, Andi, dan Aliyah tentang penerapan SMOTE untuk mengatasi *imbalance class* dalam klasifikasi kepribadian MBTI menggunakan *Naive Bayes Classifier* menunjukkan jika metode *Logistic Regression* memiliki akurasi paling tinggi dengan nilai 80%, namun performanya rendah untuk kelas *Artisan* dan *Guardian*, yang mana metode *Naive Bayes* memiliki performa yang lebih baik untuk kelas *Artisan* dan *Guardian* [11]. Pada penelitian yang dilakukan oleh Derwin, Marcella, dan Stefanny tentang *Machine Learning* untuk prediksi kepribadian berdasarkan Facebook *post* juga menunjukkan bahwa *Naive Bayes* menjadi algoritma terbaik dengan nilai *f1-score* sebesar 82.31%, lebih besar

daripada algoritma lain seperti *SVM*, *Naive Bayes*, *Decision Tree*, *KNN*, dan *Logistic Regression* [12]. Berdasarkan beberapa penelitian sebelumnya, mengapa metode *Logistic Regression* dan *Naive Bayes* dipilih dalam penelitian ini, karena kedua metode tersebut terbukti memiliki akurasi yang tinggi untuk kasus prediksi kepribadian berdasarkan data dari media sosial terutama dalam bentuk teks.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian ini, didapatkan rumusan masalah seperti berikut:

1. Bagaimana perbedaan kinerja metode *Logistic Regression* dan *Naive Bayes* dalam memprediksi kepribadian berdasarkan interaksi di media sosial X?
2. Faktor apa saja yang mempengaruhi perbedaan kinerja antara kedua metode?

1.3. Tujuan dan Manfaat

Penelitian ini memiliki tujuan untuk memahami bagaimana perbedaan *Logistic Regression* dan *Naive Bayes* dalam melakukan prediksi kepribadian berdasarkan data dari X serta faktor apa saja yang dapat mempengaruhi performa dari kedua model terutama untuk peran dari data interaksi. Harapannya, penelitian ini dapat berkontribusi dengan memberikan gambaran untuk pengembangan yang lebih lanjut untuk model prediksi kepribadian dengan data dari X khususnya yang menggunakan data interaksi.

1.4. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan sebagian besar berbahasa Indonesia, karena proses *crawling* yang dilakukan hanya berasal dari pengguna asal Indonesia.
2. Dataset hanya berasal dari satu platform yaitu media sosial X.

3. Metode yang dipakai hanya menggunakan metode *Logistic Regression* dan *Naive Bayes*.

1.5. Metode Penelitian

Metode yang dilakukan dalam penelitian ini melalui pendekatan sebagai berikut:

1. Studi Literatur
Merupakan kegiatan pengumpulan informasi tentang studi terkait dari berbagai sumber seperti jurnal baik internasional maupun nasional dan juga buku yang kemudian bisa digunakan sebagai referensi untuk digunakan dalam penelitian.
2. Pengumpulan Data
Data diambil berdasarkan kuisioner BFI-44 dan juga kesediaan dari responden untuk data pada akun X responden dapat digunakan dalam penelitian ini.
3. Perancangan Sistem
Membuat rancangan tentang bagaimana tahapan sistem dikerjakan untuk penelitian.
4. Implementasi Metode
Proses pengerjaan penelitian berdasarkan perancangan sistem untuk mencapai tujuan penelitian..
5. Analisis Hasil
Mengidentifikasi apakah hasil dari implementasi metode dapat mencapai tujuan yang diharapkan pada saat perancangan sistem.