

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi telah mengantarkan manusia untuk menciptakan bentuk baru dalam berinteraksi dan bersosialisasi, salah satunya adalah inovasi teknologi komunikasi berupa media sosial. Sebagai sarana komunikasi yang efektif, media sosial sangat berperan dalam aktifitas keseharian bersosial di masyarakat [1]. Kegiatan masyarakat pada saat ini tidak terlepas dari media sosial. Media sosial yang sedang banyak diminati oleh khalayak saat ini adalah Instagram. Survey baru-baru ini menunjukkan bahwa Instagram adalah platform media sosial terpopuler kedua, dengan 59% pengguna online usia 18-29 tahun menggunakan Instagram. Kehadiran media sosial Instagram di kalangan remaja menjadi sebuah fenomena yang menarik. Instagram merupakan aplikasi sharing foto yang meningkat popularitasnya sejak 2010, dengan lebih dari 500 juta pengguna aktif (A. Jackson, 2017). Di Indonesia tercatat lebih dari 12.000 pengguna aktif yang tergabung dalam komunitas pengguna instagram yang bernama instameet [2].

Didalam Instagram ini berisikan kumpulan foto dan caption berupa text. Dimana caption ini digunakan untuk mengungkapkan tentang diri mereka sendiri dan hal yang sedang dirasakannya atau menggambarkan tentang diri mereka sendiri maka dari data yang ada dapat dimanfaatkan untuk bidang psikologi yaitu menganalisis kepribadian seseorang secara digital. Data caption-caption yang berupa text tersebut digunakan sebagai sumber data penelitian untuk menganalisis profil kepribadian seseorang seperti dalam penelitian [3]. Jumlah pengguna dan aktivitas yang melimpah per hari memberikan peneliti data yang besar untuk diuji. Melakukan prediksi manual bisa melelahkan karena

alasan yang disebutkan. Dengan teknologi yang tersedia saat ini, prediksi otomatis menggunakan komputer dimungkinkan [4].

Penelitian terkait analisis kepribadian dengan pendekatan DISC sebelumnya pernah dilakukan oleh peneliti sebelumnya menggunakan metode Multinomial Naïve Bayes Classifier dan mendapatkan hasil akurasi 48%. Didasari oleh penelitian tersebutlah maka dilakukan penelitian ini dengan menggunakan metode dan teknik pengolahan data yang berbeda dan diharapkan mampu meningkatkan nilai akurasi dari penelitian sebelumnya [5].

Prediksi kepribadian dilakukan dengan pendekatan DISC yaitu dengan membedakan tipe kepribadian menjadi empat tipe: Dominance (Dominasi), Influence (Pengaruh), Steadiness (Stabilitas), dan Compliance (Kepatuhan). Prediksi kepribadian dengan pendekatan DISC merupakan alternatif yang baru dalam melakukan penelitian kepribadian seseorang [6]. Data mining merupakan proses untuk memanipulasi data dengan mengekstraksi informasi yang sebelumnya tidak diketahui dari dataset yang berukuran besar [7].

Text Mining merupakan penemuan informasi baru yang sebelumnya tidak diketahui oleh komputer, dengan mengekstrak informasi secara otomatis dari berbagai sumber tertulis. Elemen kuncinya adalah menghubungkan informasi yang diekstraksi bersama-sama untuk membentuk fakta baru atau hipotesis baru untuk dieksplorasi lebih lanjut dengan cara eksperimen yang lebih konvensional [8].

Prediksi kepribadian dilakukan dengan mengklasifikasi caption instagram bahasa Indonesia. Metode klasifikasi yang digunakan berfokus pada metode Random Forest Classifier. Metode random forest merupakan pengembangan dari metode CART (Classification and Regression Tree). CART merupakan metode eksplorasi data yang didasarkan pada teknik pohon keputusan. Pohon klasifikasi dihasilkan saat peubah respons berupa data kategorik, sedangkan pohon regresi dihasilkan saat peubah respons berupa data numerik [9].

Model Random Forest adalah model kombinasi dari tree yang menggunakan vektor acak yang diambil secara terpisah dari vektor input, dan setiap tree memberikan klas popularnya untuk mengklasifikasikan vektor masukan [10].

Dengan kata lain, model ini menggunakan fitur yang dipilih secara acak atau kombinasi fitur disetiap simpul untuk membangkitkan tree [11]. Penelitian ini menggunakan metode Random Forest Classifier karena memiliki beberapa kelebihan diantaranya untuk banyak dataset, model pembelajaran ini menghasilkan classifier yang sangat akurat, untuk dataset yang berukuran besar, model ini berjalan dengan efisien, dan model ini efektif ketika berhadapan dengan data yang memiliki missing value dengan tetap mempertahankan akurasi bahkan ketika sebagian besar data hilang [12].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah:

1. Apakah algoritma *Random Forest Classifier* dapat diimplementasikan untuk mengklasifikasi kepribadian DISC seseorang berdasarkan caption instagram bahasa Indonesia?
2. Berapa akurasi dari algoritma *Random Forest Classifier* dalam mengklasifikasi kepribadian DISC berdasarkan caption instagram bahasa Indonesia?
3. Berapa perbandingan tingkat akurasi metode *Random Forest* dibandingkan dengan metode *MNB*, *Decision Tree*, dan *KNN*?

1.3. Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Penelitian ditekankan pada pengujian metode Random Forest Classifier.
2. Dataset yang digunakan hanya caption Instagram Bahasa Indonesia.
3. Analisis yang digunakan pada kepribadian DISC saja.
4. Sistem hanya dapat memproses akun instagram yang bersifat publik.

1.4. Maksud Dan Tujuan Penelitian

Adapun maksud dan tujuan yang ingin dicapai dari penelitian ini adalah :

1. Untuk mengetahui apakah metode Random Forest Classifier dapat digunakan dalam menganalisis kepribadian DISC dari seseorang.
2. Mengetahui tingkat akurasi sistem dalam menganalisis kepribadian DISC dengan caption Bahasa Indonesia di instagram seseorang menggunakan metode Random Forest Classifier.
3. Mengetahui performa metode *Random Forest* dibandingkan dengan metode *MNB*, *Decision Tree*, dan *KNN*.

1.5. Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan pada pembuatan penelitian ini adalah:

1. Penelitian ini diharapkan dapat menambah dan mengembangkan wawasan, informasi dan pengetahuan kepada pihak yang berkepentingan.
2. Untuk peneliti diharapkan dapat menerapkan ilmu yang diperoleh peneliti dan berguna bagi kemajuan teknologi dan ilmu pengetahuan.

1.6. Metode Penelitian

1.6.1. Metode Studi Literatur

Metode studi literatur dalam penelitian ini dilakukan dengan mempelajari dan mengolah referensi bahan penelitian yang diperoleh dari pengumpulan data pustaka. Dalam hal ini dengan mengkaji dan mempelajari banyak referensi dari artikel ilmiah, jurnal-jurnal, dan sumber lainnya yang berkaitan dengan kebutuhan penelitian yang dilakukan secara offline maupun online sebagai penguat dari sumber penelitian.

1.6.2. Metode Pengumpulan Data

Metode pengumpulan data merupakan tahap penelitian untuk mengumpulkan data sebagai bahan pengujian. Data yang telah dikumpulkan digunakan sebagai dataset yang telah diberi label DISC dan validasi dari psikolog.

1.6.3. Metode Analisis

Metode analisis masalah dilakukan berdasarkan informasi yang telah diperoleh dari tahap sebelumnya untuk kemudian dilakukan pemilihan metode sebagai pemecahan masalah sesuai dengan analisis masalah dalam penelitian.

1.6.4. Implementasi

Implementasi pada penelitian ini dilakukan berdasarkan analisis desain yang dilakukan dalam bentuk program.

1.6.5. Testing

Metode pengujian atau testing dalam penelitian ini dilakukan sebagai tahap untuk menguji performa dari metode yang telah dibangun, teknik yang digunakan yaitu *Confusion Matrix*.

1.7. Sistematika Penulisan

Laporan skripsi dikerjakan sesuai dengan petunjuk pengerjaan yang berlaku di Universitas Amikom Yogyakarta. Secara keseluruhan terdiri dari lima (5) bab, adapun sistematika penulisan skripsi sebagai berikut :

BAB I PENDAHULUAN

Bab I berisikan tentang penjelasan latar belakang dari penelitian, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab II berisikan pembahasan penelitian sebelumnya yang pernah dilakukan oleh peneliti lain sebagai referensi dari penelitian ini, dan juga berisikan teori-teori yang digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab III ini berisikan tentang tahapan yang dilakukan dalam pengimplementasian metode dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan hasil dan pembahasan dari implementasi algoritma *Random Forest Classifier*, mulai dari alur kerja metode, dan source code yang digunakan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari pembahasan penelitian yang diperoleh dari rumusan masalah, serta saran untuk pengembangan sistem berikutnya agar lebih baik lagi.

