

BAB 1

1. Pendahuluan

1.1 Latar Belakang

Kepribadian mencakup suasana hati, sikap, dan pendapat. Dimana cakupan dari semua itu paling jelas diekspresikan dalam interaksi dengan orang lain. Kepribadian mencakup sifat-sifat perilaku bawaan dan didapat untuk membedakan satu orang dari orang lain dan dapat diamati dalam kaitannya dengan lingkungan dan kelompok social [1]. Teori sifat Kepribadian Model Lima Besar atau *Big Five Personality Traits Model* tersebut terdiri dari 5 dimensi kunci yaitu *Openness, Conscientiousness, Extraversion, Agreeableness dan Neuroticism* [2].

Penelitian ini memanfaatkan Twitter sebagai media untuk memprediksi kepribadian seseorang. Dimana Twitter saat ini memiliki lebih dari 126 juta pengguna aktif [3]. Banyaknya *tweet* pada Twitter setiap harinya memungkinkan pengguna tidak sadar memberikan informasi terkait kepribadiannya melalui *tweet* tersebut.

Logistic Regression adalah Teknik untuk mendapatkan nama dari jenis kurva yang digunakan untuk mendekati data. *Logistic Regression* juga dapat memprediksi probabilitas terkait. Dalam melakukan klasifikasi pada penelitian ini penulis menggunakan teori OCEAN seperti yang sudah dijelaskan diatas. Dan klasifikasi yang digunakan untuk mendukung metode *Logistic Regression* adalah IndoBERT.

Penelitian ini menggunakan referensi pada penelitian sebelumnya [4] dimana dalam penelitiannya menggunakan SVM, LIWC, SMOTE, dan IndoBERT untuk sistem prediksi kepribadian. Penggunaan IndoBERT dalam sistem prediksi kepribadian mendapatkan peningkatan dalam prosesnya. Hal ini dikarenakan IndoBERT dibuat untuk digunakan pada dataset Indonesia, dimana penelitian ini menggunakan dataset dengan Bahasa Indonesia. Sama dengan halnya dalam penggunaan LIWC dalam melakukan prediksi kepribadian dikarenakan LIWC cocok dalam penggunaan dengan data yang meliputi kata-kata.

1.2 Topik dan Batasannya

Berdasarkan latar belakang, penelitian ini membangun sistem prediksi kepribadian *Big Five Personality* dengan menggunakan Twitter sebagai media untuk mendapatkan *tweets* dari pengguna, *tweet* digunakan sebagai fitur untuk menilai kepribadian pengguna. Sistem yang digunakan pada penelitian ini menggunakan metode *Logistic Regression* dan IndoBERT dengan fitur LIWC, SMOTE, dan ANOVA F-Classif. Batasan pada penelitian ini menggunakan dataset yang bersumber dari twitter dimana dataset yang digunakan menggunakan Bahasa Indonesia.

