

1. Pendahuluan

Platform media sosial menjadi sarana untuk mengekspresikan diri dan berkomunikasi sosial secara online. Oleh karena itu, platform media sosial menjadi salah satu tambang terbesar untuk memperoleh informasi pribadi seseorang. Kemudahan mengakses media sosial memungkinkan pengguna untuk membuat identitas online, membagikan konten (teks, tautan, atau gambar) dan berinteraksi dengan orang lain. Twitter merupakan salah satu platform media sosial yang saat ini populer digunakan. Dengan *tweet* sebagai media untuk menulis dan membagikan tulisan kepada orang lain, maka perilaku dan tulisan yang ditinggalkan pengguna dapat diperoleh dengan mudah dan dianalisis kepribadian pengguna Twitter tersebut [1][2][3].

Kepribadian seseorang dapat mempengaruhi prospek kehidupan, karier, dan romansa. Bagi perusahaan mengetahui kepribadian seseorang dapat digunakan untuk rekrutmen karyawan, konseling karier, dan kesehatan [4]. Kepribadian pengguna dapat diprediksi dengan beberapa model kepribadian, seperti *Mayers-Briggs Type Indicator* (MBTI), *Big Five Personality*, atau *Dominance Influence Steadiness Conscientiousness* (DISC). Namun, *Big Five Personality* merupakan yang paling populer dalam dunia psikologi dan cukup tepat dalam mencirikan kepribadian seseorang [5][6]. Kepribadian *Big Five* tersebut terdiri dari *openness* (O), *conscientiousness* (C), *extraversion* (E), *agreeableness* (A), dan *neuroticism* (N).

Pada tugas akhir ini penulis melakukan penelitian menggunakan metode *Support Vector Machine* (SVM) sebagai model prediksi kepribadian pengguna. Metode ini terbukti memberikan kinerja yang baik pada penelitian sebelumnya [7][8]. Penulis menggunakan tiga kernel yang berbeda pada SVM, yaitu *Linear*, *Radian Base Function* (RBF) dan *Polynomial* untuk kernel mana yang menghasilkan akurasi terbaik. Pada penelitian ini juga dilakukan analisis terhadap perilaku sosial dan penggunaan dua fitur linguistik yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Linguistic Inquiry and Word Count* (LIWC). Penelitian sebelumnya telah dilakukan namun hasil akurasi yang diperoleh masih cukup rendah dan data yang digunakan tidak seimbang [9]. Maka pada penelitian ini dilakukan optimisasi menggunakan teknik *sampling*, pemilihan fitur, dan *hyperparameter tuning* untuk meningkatkan performansi pada sistem.

Penyusunan struktur pada tugas akhir ini sebagai berikut. Bagian 2 akan membahas teori/studi yang berkaitan erat dengan penelitian. Bagian 3 menjelaskan metode atau rancangan sistem penelitian prediksi kepribadian. Bagian 4 menyajikan hasil dan analisis penelitian yang dilakukan. Selanjutnya pada bagian 5 disampaikan kesimpulan dan saran yang diperoleh dari penelitian yang dilakukan.