1. PENDAHULUAN

Kepribadian merupakan kombinasi perilaku dan emosi yang tergantung pada lingkungan dan penyebab biologis[1]. Kepribadian adalah kecenderungan psikologis seseorang untuk melakukan tingkah laku sosial tertentu, baik berupa perasaan, berpikir, bersikap, dan berkehendak maupun perbuatan. Menurut M.a.W Bower, kepribadian adalah keseluruhan perilaku dari seorang individu dengan system kecenderungan tertentu yang berinteraksi dengan serangkaian instruksi. Setiap orang pasti mempunyai kepribadian yang berbeda – beda tergantung bagaimana individu tersebut membawa dirinya. Ada beberapa model kepribadian yang digunakan dalam memprediksi kepribadian, seperti Big Five Personality, MBTI (Myers Briggs Type Indicato) atau DISC (Dominance Influence Steadiness Conscientiousness)[2].

Big Five merupakan model kepribadian yang paling populer dan banyak digunakan, oleh karena itu proposal kali ini menggunakan model BigFive Personality. BigFive Personality meliputi 5 sifat utama yaitu *Extraversion*, *Agreeableness*, *Openness*, *Conscientiousness*, dan *Neuroticism*. *Extraversion* memiliki sifat yang tegas, mudah bergaul dengan lingkungan, aktif, cenderung mengejar kegembiraan dan juga tantangan. *Agreeableness* memiliki sifat yang lembut, kooperatif, rendah hati dan baik hati, individu yang memiliki sifat ini juga cenderung menerima dan mematuhi budaya dan norma. *Openness* memiliki sifat yang sensitif, berpikiran terbuka, cenderung memiliki intelektual, dan juga imajinatif yang tinggi. *Conscientiousness* cenderung memiliki sifat yang teorganisir, bertanggung jawab, dan juga teliti dalam segala hal namun juga berhati-hati. *Neuroticism* memiliki sifat yang cenderung cemas akan sesuatu, mudah marah dan merasa depresi. Orang yg memiliki sifat ini juga mudah merasa tidak tenang [3].

Manusia melakukan interaksi melalui media sosial di era teknologi seperti sekarang ini. Yang dimaksud media sosial adalah media online yang penggunaannya dapat dengan mudah dalam menciptakan ide, menyampaikan pendapat secara bebas, dan membuat konten. Kegiatan dalam menggunakan media sosial sebagai jaringan komunikasi *online* yang meliputi teks, gambar dan video sangat membantu dalam kegiatan sehari-hari seperti berdiskusi hingga berbisnis.

Twitter merupakan media sosial yang digunakan oleh semua kalangan[4]. Menurut data *We Are Social* Januari 2020, Twitter merupakan jejaring sosial media terpopuler yang digunakan semua kalangan masyarakat Indonesia dengan sekitar 10,65 juta pengguna, dan juga termasuk peringkat nomor satu diantara semua platform media sosial [5]. Di Twitter orang sering mengungkapkan apa yang mereka pikirkan dengan satu *tweet* dengan tidak lebih dari 140 karakter secara realtime. Dari cuitan Twitter ini dijadikan dataset kemudian dianalisis. *Machine learning* atau pembelajaran mesin adalah metode yang banyak digunakan dalam analisa permasalahan seperti deteksi depresi dan deteksi kepribadian[6].

Deteksi kepribadian di era teknologi ini sangat dibutuhkan karena banyak informasi tentang penggunaannya, seperti perusahaan yang menggunakan deteksi kepribadian untuk mendeteksi kepribadian calon karyawannya. Jika deteksi kepribadian tidak dilakukan, perusahaan dapat mengalami kesalahan dalam merekrut sumber daya manusianya [7].

Salah satu metode yang banyak digunakan yaitu menggunakan metode *Support Vector Machine (SVM)*. Berdasarkan penelitian sebelumnya, yang sudah dilakukan oleh Shantika Valerin Therik dan Erwin Budi Setiawan yang berjudul “Deteksi Kepribadian *Big Five* Pengguna Twitter Dengan Metode C4.5”, dimana akurasi tertinggi didapatkan sebesar 76,92% yang diperoleh dari perbandingan data *Baseline+LIWC+TF-IDF* dan menerapkan metode SMOTE [7]. Penelitian yang dilakukan oleh Maulana Amsor yang menggunakan metode *Support Vector Machine (SVM)* dan metode *Principal Component Analysis (PCA)* didapatkan hasil akurasi sebesar 70.9% [8]. Penelitian yang dilakukan oleh Y B N D Artissa, I Asror, dan S A Faraby dengan judul “*Personality Classification based on Facebook Status Text Using Multinomial Naive Bayes Method*” memperoleh akurasi hanya sebesar 59,9% [9]. Kemudian penelitian yang dilakukan oleh Bayu Yudha Pratama dan Rianarto Sarno dengan judul “*Personality Classification Based on Twitter Text Using Naive Bayes, KNN, and SVM*” diperoleh akurasi sebesar 65%[10]. Penelitian lain yang dilakukan oleh Tri Ayu Syifa’ur Rohmah yang berjudul “Deteksi Kepribadian di Media Sosial Twitter Menggunakan Metode IndoBERT” mendapatkan akurasi tertinggi sebesar 0,46. Penelitian ini menggunakan 258 dataset crawling twitter yang skenario ujinya terletak pada dataset selisih *preprocessing*[11]. Berdasarkan penelitian sebelumnya, penelitian ini berfokus untuk mengidentifikasi informasi pribadi tentang orang yang menggunakan situs media sosial seperti Twitter dengan menggunakan *Support Vector Machine (SVM)* karena memiliki waktu komputasi yang optimal [12], serta fungsi pemisal yang optimal dan mempunyai generalisasi data yang tinggi [13]. Akhir – akhir ini metode SVM berhasil diterapkan pada situasi *real-word problems* (problema dunia nyata). Selain itu, SVM menawarkan solusi yang lebih baik daripada metode konvensional lainnya [14]. SVM mempunyai keunggulan yaitu SVM bisa diperlebar atau diperluas untuk menyelesaikan permasalahan non-linier, yakni SVM linier bisa diubah menjadi SVM non-linier dengan memakai kernel [15]. Kelebihan SVM yang lain yaitu SVM mempunyai landasan teori yang dapat dianalisis secara jelas dan tidak bersifat *black box*. Strategi generalisasi selanjutnya adalah kemampuan metode saat ini untuk

mengklasifikasikan pola tertentu yang tidak menyertakan data yang digunakan dalam metode tersebut. Dua penyebab utama dalam *generalized error* adalah *error* yang melibatkan *set training* dan *error* yang disebabkan oleh VC (*Vapnik-Chervokinensis*) [14]. Penelitian ini akan menggunakan metode SVM dengan menggunakan fitur *Count Vectorizer*. Tujuan dari penelitian ini adalah mendapatkan kinerja paling maksimal pada pendekatan model kernel *Support Vector Machine (SVM)* dalam melakukan pendeteksian kepribadian pada pengguna Twitter dan mengetahui tingkat akurasi dari setiap pendekatan kernel yang digunakan SVM.