

1. Pendahuluan

Kepribadian adalah suatu ciri yang paling dominan keluar dari diri seseorang, ciri tersebut dapat berupa tingkah laku atau sifat dari orang tersebut. Dari kepribadian seseorang kita dapat mengetahui bagaimana pola pikir seseorang untuk mengambil sebuah keputusan dan bereaksi terhadap sesuatu. Pengetahuan tentang kepribadian seseorang dianggap penting karena mempengaruhi tingkah laku dalam mengambil keputusan yang berdampak baik atau buruk [1]. Penggunaan jasa psikolog atau mengikuti serangkaian tes psikologi untuk menentukan kepribadian seseorang akan memakan waktu serta dana yang cukup besar.. Saat ini sudah banyak penelitian menggunakan sarana sosial media untuk memprediksi kepribadian seseorang salah satunya menggunakan media sosial Twitter [1].

Twitter telah menghasilkan 110 juta *tweets* setiap hari dan memiliki lebih dari 200 juta pengguna [2]. Dalam menganalisis kepribadian seseorang melalui Twitter dibutuhkan metodologi yang tepat untuk mendapatkan hasil yang akurat. untuk menilai kepribadian pengguna *Twitter* diperlukan metode untuk megklasifikasi data teks pada *Twitter*. Algoritma untuk mengklasifikasikan data teks yang dapat digunakan pada penelitian ini terdapat pada *machine learning*. *Machine Learning* adalah perancangan dan pengembangan algoritma yang memungkinkan komputer untuk mengenali perilaku yang berdasarkan kepada data yang ada dengan menangkap ciri-ciri yang diperlukan [3]. Salah satu contoh algortima klasifikasi yang terdapat pada *machine learning* adalah *Support Vector Machine*.

Pada tugas akhir ini penulis melakukan analisis yaitu pendekatan terhadap perilaku sosial pengguna Twitter dan penggunaan linguistik atau kata yang digunakan pengguna saat menuliskan *tweet*. Metode klasifikasi yang digunakan adalah *Support Vector Machine* karena sangat efektif dalam berbagai bidang seperti data dan penambangan teks, ekstraksi informasi, pembelajaran mesin, dan pengenalan pola [4]. Hal lain yang membedakan penelitian ini adalah pembobotan setiap kata menggunakan metode pembobotan *TF-IDF* dan *WIDF* dengan penambahan fitur baru untuk pendekatan berdasarkan perilaku sosial pengguna seperti jumlah karakter pada *tweet*, rata-rata karakter pada setiap *tweet*, rata-rata kata pada *tweet*, media URL yang membaca seberapa banyak pengguna mengunggah foto atau video, tanda baca yaitu menghitung jumlah tanda Tanya (?) dan tanda seru (!) pada setiap *tweet*, menghitung huruf besar, dan emoji pada *tweet* yang dapat dideteksi sebanyak 2.552 karakter yang berbeda-beda.