

1. Pendahuluan

Perkembangan teknologi pada saat ini sudah semakin pesat dengan adanya media sosial seperti Facebook, Twitter, Instagram dan sejenisnya. Pengguna sosial media pada January 2005 sudah mencapai 115 juta member[1]. Salah satu media sosial twitter. Twitter adalah sosial media yang berbasis *microblogging* yang peluncurannya pada tanggal 13 Juli 2006[2]. Pengguna twitter saat pada tahun 2009 sudah mencapai lebih dari 41 juta orang dan terus meningkat hingga saat ini[3]. *Tweet* yang yang di-*posting* seseorang itu bisa menggambarkan perasaan dan kepribadian dari orang tersebut, oleh karena itu dalam penulisan penelitian ini penulis ingin mengambil informasi *tweet* dari twitter sebagai bahan acuan kepribadian seseorang.

Untuk menilai kepribadian seseorang biasanya bisa menggunakan beberapa metode seperti kuisioner, wawancara dan juga menggunakan jasa psikolog tetapi membutuhkan banyak waktu dan dana yang lebih. Saat ini sudah banyak metode yang dapat digunakan untuk menilai kepribadian. Salah satu metode yang bisa digunakan ialah kepribadian DISC.

Kepribadian DISC dikemukakan oleh seorang ahli psikolog asal Amerika yang bernama William Moulton Marston pada tahun 1928 dalam bukunya yang berjudul *Emotions of Normal People*. Ia ber teori bahwa ekspresi perilaku emosi bisa dikategorikan menjadi 4 tipe perilaku individu ketika berinteraksi dengan lingkungannya yaitu *Dominance (D)*, *Influence (I)*, *Steadiness (S)*, dan *Compliance (C)*[4]. Kepribadian DISC merupakan salah satu alat penilai kepribadian seseorang tetapi belum banyak digunakan untuk penelitian oleh karena itu DISC merupakan alternatif yang baru dalam melakukan penelitian[4]. Dalam penelitian sebelumnya[4] metode kepribadian DISC mengklasifikasikan kepribadian dengan data yang pengambilannya menggunakan test kepada setiap mahasiswa. Oleh karena itu penulis ingin melakukan penelitian penilaian kepribadian dengan menggunakan data *tweet* di twitter.

Algoritma yang terdapat pada data *mining* dapat digunakan sebagai salah satu metode dalam penelitian ini. Data *mining* bisa di sebut juga pencarian pengetahuan dari data yang dimana data *mining* akan mengesktrak secara otomatis pola atau pengetahuan yang menarik, tersembunyi, tidak diketahui sebelumnya, dari data dalam jumlah yang sangat besar[5]. Data mining merupakan ilmu yang relatif baru, dan sampai sekarang orang masih memperdebatkan untuk menempatkan data mining di bidang ilmu mana karena data mining menyangkut database, kecerdasan buatan, statistik[6]. Salah satu algoritma pengklasifikasian data yang terdapat dalam data mining adalah *K- Nearest Neighbors Algorithm (KNN)*[7].

KNN adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan K buah data latih yang jaraknya paling dekat dengan objek tersebut[8].

Dalam penelitian ini, penulis menggunakan pembobotan *Term Frequency Inverse Document Frequency (TF-IDF)* dan *TF-Chi Square*. TF-IDF yang merupakan metode yang digunakan dalam melakukan pembobotan terhadap kemunculan kata dalam suatu dokumen[7]. Uji *TF-Chi Square* berguna untuk menguji hubungan dua buah variabel nominal dan mengukur kuatnya hubungan antara variabel yang satu dengan variabel nominal lainnya. Oleh Karena itu, pada penelitian ini akan melihat seberapa akurat penggunaan algoritma KNN dalam pengklasifikasian kepribadian pengguna twitter serta seberapa besar perbedaan antara dua metode pembobotan yang digunakan.