

BAB I

PENDAHULUAN

1.1 Latar Belakang

Twitter merupakan salah satu media sosial yang umum digunakan oleh masyarakat terutama di Indonesia. Tercatat bahwa pengguna *Twitter* di Indonesia sendiri mencapai 6.430.000 pengguna aktif pada tahun 2019 [1]. Angka ini menunjukkan bahwa banyak masyarakat Indonesia yang menggunakan media sosial *Twitter* untuk berkomunikasi ataupun menyampaikan pendapatnya. Dalam berkomunikasi maupun menyampaikan pendapat, kata-kata yang digunakan pada umumnya adalah kata-kata yang sering digunakan dalam kesehariannya, sehingga dapat dilihat secara garis besar sikap dan karakteristik individu tersebut dari *tweet* yang dibuatnya.

Dari beberapa kata yang ditemukan pada tiap *tweet* individu tersebut dapat dianalisis menggunakan metode DISC (*Dominance, Influence, Steadiness, Compliance*), RIASEC (*Realistic, Investigative, Artistic, Social, Enterprising, Conventional*), dan MBTI (*Myers Briggs Type Indicator*) untuk mengetahui karakter dari individu tersebut [2] [3] [4]. Pada penelitian ini akan dipilih metode DISC karena dapat membantu setiap manusia memahami “mengapa seseorang melakukan apa yang dilakukannya” [2]. Dengan mengetahui karakter individu tersebut, hal ini bisa digunakan sebagai tolak ukur seperti saat melakukan seleksi karyawan. Sehingga pihak

HRD (*Human Resource Department*) tidak perlu mengeluarkan tenaga lebih untuk mengetahui karakter individu tersebut di dunia maya.

Namun untuk melakukan hal ini, diperlukan sebuah metode lain yaitu Data Mining. Untuk dapat mengklasifikasikan tiap *tweet* dari individu tersebut diperlukan algoritma yang terdapat dalam *Data Mining*. Beberapa algoritma yang dapat digunakan untuk mengklasifikasikan *tweet* tersebut yaitu *C4.5*, *ID3*, *CART*, *K - Nearest Neighbor* dan *Naïve Bayes Classifier* [5][6][7]. Pada penelitian ini akan membandingkan kinerja antara algoritma *K - Nearest Neighbor* dengan *Naïve Bayes Classifier*. Algoritma *K - Nearest Neighbor* dipilih karena menurut Mei Lestari, algoritma *K - Nearest Neighbor* memiliki kelebihan berupa tangguh terhadap data training yang memiliki banyak *noise* [7], sedangkan algoritma *Naïve Bayes Classifier* dipilih karena menurut Ridwan, Suyono, dan Sarosa, algoritma *Naïve Bayes Classifier* merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan teorema Bayes [6]. Kedua algoritma ini merupakan bagian dari cabang *Data Mining* yaitu klasifikasi [5][6]. Dimana klasifikasi merupakan cabang *Data Mining* yang digunakan untuk mengetahui termasuk klasifikasi apakah inputan dengan membandingkan parameter-parameter yang ada.

Untuk mengklasifikasikan *tweet* menggunakan kedua algoritma tadi, harus ditentukan terlebih dahulu keyword-keyword mana yang akan menjadi penentu klasifikasi. Hal ini tidak lepas dari proses awal data mining pada *tweet* yang harus melalui proses *preprocessing* terlebih dahulu [6]. Setelah dilakukan *preprocessing* akan dilakukan pembobotan pada *tweet* menggunakan TF-IDF. Baru setelah itu akan

digunakan kedua algoritma *K – Nearest Neighbor* dan *Naïve Bayes Classifier* untuk mengklasifikasikan akun *Twitter* kedalam Metode DISC.

Diharapkan dengan membandingkan kinerja Algoritma *K – Nearest Neighbor* dengan *Naïve Bayes Classifier* dengan tahap *preprocessing* dan pembobotan TF-IDF, dapat diketahui algoritma mana yang lebih cocok untuk mengklasifikasi karakter individu dengan metode DISC. Tentunya diharapkan juga penelitian ini dapat berguna sebagai penunjang seperti pada pihak HRD perusahaan untuk mengetahui karakter individu dari *tweet* pada linimasa akun *Twitter*nya, atau untuk pihak konseling pada sekolah untuk mengetahui karakter siswa sehingga dapat membantu siswa melakukan penggalian potensi yang dimiliki berdasarkan karakter yang dimiliki.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan diatas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimanakah algoritma *K – Nearest Neighbor* dan *Naïve Bayes Classifier* dapat mengklasifikasikan akun *Twitter* kedalam metode DISC?
2. Bagaimanakah tingkat akurasi algoritma *K – Nearest Neighbor* dibandingkan dengan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan akun *Twitter* kedalam metode DISC?

1.3 Batasan Masalah

Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Akun *Twitter* yang akan digunakan untuk penelitian ini merupakan akun *Twitter* yang bersifat personal yang sudah divalidasi oleh psikolog secara manual yang berjumlah 120 akun.
2. Penelitian ini memiliki dataset berupa kata-kata yang sudah divalidasi sebelumnya oleh psikolog kedalam beberapa kelas dalam metode pendekatan DISC.
3. Hasil dari klasifikasi dengan Algoritma K – *Nearest Neighbor* dan Naïve Bayes Classifier hanya ke dalam metode DISC ke setiap akun *Twitter*.
4. Penelitian ini hanya membandingkan Algoritma K – *Nearest Neighbor* dan Naïve Bayes Classifier dalam klasifikasi karakter seseorang dengan membandingkan tingkat akurasi, *precision*, dan *recall*.

1.4 Maksud dan Tujuan Penelitian

Tujuan umum yang ingin dicapai dalam penelitian ini adalah untuk mengetahui bagaimana pemanfaatan media sosial *Twitter* dalam mengklasifikasi karakteristik seseorang. Tujuan khusus penelitian ini adalah sebagai berikut :

1. Penelitian ini ditujukan untuk membandingkan Algoritma K – *Nearest Neighbor* dengan Naïve Bayes Classifier dalam mengklasifikasi karakter seseorang dengan metode DISC.
2. Penelitian ini juga ditujukan untuk mengetahui apakah *tweet* dapat untuk mengklasifikasikan karakter seseorang dengan Algoritma K – *Nearest Neighbor* dan Naïve Bayes Classifier dengan metode DISC.

1.5 Manfaat Penelitian

Melalui penelitian ini diharapkan dapat diperoleh manfaat sebagai berikut :

1. Bagi peneliti mengetahui kinerja Algoritma K -*Nearest Neighbor* dan Naïve Bayes Classifier dalam mengklasifikasi karakter seseorang dengan metode DISC pada *tweet* seseorang.
2. Bagi Perusahaan mempermudah dalam perekrutan tenaga kerja baru ditinjau dari karakter seseorang yang didapatkan dari *tweet* di akun *Twitter*nya.
3. Bagi Konseling sekolah membantu penggalian potensi siswa ditinjau dari karakter siswa yang didapatkan dari *tweet* di akun *Twitter*nya.
4. Bagi Psikolog membantu pekerjaan sebagai penunjang keputusan karakter seseorang yang didapatkan dari *tweet* di akun *Twitter*nya.

1.6 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah metode eksperimental dan dilaksanakan dalam tahapan-tahapan berikut:

1.6.1 Metode Pengumpulan Data

Pada tahapan pengumpulan data ini dilakukan pengambilan data latih dari *Twitter* dengan menggunakan cara *Scraping* dengan library yang disediakan oleh python yaitu Scrapy yang berupa seluruh *tweet* dari beberapa akun pada linimasa mereka. Setelah data diperoleh maka selanjutnya akan di lakukan pelabelan data secara manual dan kemudian dilakukan pengklasifikasian menggunakan algoritma yang dipilih yaitu K -*Nearest Neighbor* dan Naïve Bayes Classifier.

1.6.2 Metode Analisis

Pada tahap ini dilakukan beberapa bagian yaitu melakukan analisis terhadap data yang sudah didapatkan dari *scraping* seperti *preprocessing*, analisis akun personal dan juga pembobotan.

1.6.3 Metode Perancangan

Pada tahap ini dilakukan perancangan cara mengimplementasikan komponen-komponen sistem seperti pembuatan data latih serta alur dari *preprocessing* sampai pengklasifikasian menggunakan algoritma.

1.6.4 Implementasi

Pada tahap ini dilakukan pembuatan kode program sampai proses klasifikasi *tweet* menggunakan algoritma.

1.6.5 Pengujian

Pada tahap ini dilakukan perhitungan akurasi dari implementasi yang sudah dilakukan. Metode yang digunakan yaitu *Confusion Matrix*. Pengujian yang dilakukan adalah perbandingan akurasi klasifikasi yang dihasilkan *K-Nearest Neighbor* dengan *Naïve Bayes Classifier*.

1.7 Sistematika Penulisan

Sistematika penulisan dibuat untuk mempermudah dalam penyusunan skripsi ini maka perlu ditentukan sistematika penulisan yang baik. Sistematika penulisannya adalah sebagai berikut :

BAB I Pendahuluan

Bab pendahuluan mendeskripsikan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan, manfaat dan sistematika penulisan.

BAB II Landasan Teori

Berisi tentang teori-teori yang digunakan dalam penelitian, perancangan dan pembuatan sistem.

BAB III Analisis dan Perancangan

Dalam bab ini penulis mengemukakan metode penelitian yang dilakukan dalam analisa, perancangan, dan implementasi.

BAB IV Hasil dan Pembahasan

Memaparkan dari hasil-hasil tahapan penelitian, mulai dari analisis, desain, hasil testing dan implementasinya.

BAB V Penutup

Berisi kesimpulan dan saran dari seluruh penelitian yang telah dilakukan.