

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pada era digital saat ini, masyarakat dapat memanfaatkan kemudahan dan keefektifan dalam berinteraksi antara satu sama lain. Kegiatan online yang populer di Indonesia adalah penggunaan media sosial. Media sosial merupakan aktivitas sosial yang menggunakan jaringan online berupa bahasa, gambar dan video (Purawinangun, 2020) Aktivitas sosial tersebut dianggap sangat memudahkan seseorang dalam berdiskusi, berbisnis, dan berkomentar secara bebas. (Isnain, Sihabuddin dan Suyanto, 2020) Perkembangan media sosial yang kini digunakan untuk melengkapi dan memudahkan banyak pekerjaan, khususnya di dalam hal berbagi informasi dan menjalani komunikasi dengan banyak pihak. Komunikasi dan berbagi informasi tersebut didukung dengan munculnya media sosial seperti twitter, facebook, tumblr, blog, instagram dan lainnya yang kerap digunakan penggunanya untuk memperoleh informasi serta saling berbagi tanpa dibatasi ruang dan waktu.

Dengan adanya jaringan internet dan perkembangan media sosial tersebut, kini siapapun dapat terhubung dengan siapapun di penjuru dunia manapun. Pengguna internet di indonesia dari total populasi 265 milyar penduduk indonesia, setengah nya atau 50% sebesar 132 milyar penduduknya sudah menggunakan internet dalam aktivitas keseharian mereka. Indonesia tercatat memiliki 130 juta pengguna media sosial aktif pada 2018 mengacu pada laporan dari laman wearesocial.com, Rabu (31/01/2018). Adapun, setidaknya 49% dari total populasi penduduk memiliki dan aktif menggunakan media sosial. Dengan 120 juta pengguna di antaranya menggunakan ponsel untuk terhubung dengan akun media sosialnya. Pengguna twitter di indonesia

mencapai 29 juta (data tahun 2017), dari total 10.6 milyar pengguna global mengacu pada laporan dari laman kompasiana.com.

Twitter adalah media sosial yang memungkinkan orang untuk membangun komunikasi antar pengguna dan dapatkan informasi terbaru atau berita. Twitter sering dijadikan objek penelitian dikarenakan sifatnya yang cepat karena dibatasinya jumlah karakter menjadi maksimal 280 karakter, sehingga lebih langsung menyampaikan maksud pesan. Twitter juga bermanfaat untuk perusahaan mencari informasi kepada calon rekrutmen karyawan. Tweet yang di posting seseorang bisa menggambarkan perasaan dan karakter dari orang tersebut. Karakter adalah kepribadian dari seseorang yang biasanya terbentuk dari internalisasi diri yang di yakinkan menjadi dasar dari cara pandang seseorang, cara berpikir, berperilaku, dan cara bertindak. Dalam ilmu psikologi, dimaknai sebagai hasil proses psikologis perkembangan seseorang yang dapat menjelaskan perbedaan antara satu individu dengan individu lainnya, serta menggambarkan perilaku alami manusia.

Untuk melakukan tes kepribadian dapat dilakukan dengan menggunakan metode DISC, MBTI, Strength Finder dan Big Five. Namun, dalam penelitian ini, menggunakan pendekatan metode DISC. DISC ditemukan oleh seorang ahli psikolog asal Amerika yang bernama William Moulton Marston pada tahun 1928 dalam bukunya yang berjudul *Emotions of Normal People*. DISC adalah model perilaku yang membantu setiap manusia mengapa seseorang melakukan apa yang dia lakukan. DISC membagi karakter seseorang terdiri dari 4 tipe kepribadian yaitu, Dominance (D), Influence (I), Steadiness (S), dan Compliance (C).

Namun untuk melakukan hal ini, diperlukan sebuah metode lain yaitu data mining. Data mining menyangkut database, kecerdasan buatan, statistik. Untuk mengklasifikasikan tweet dari seseorang diperlukan algoritma yang terdapat dalam data

mining. Pada penelitian ini membandingkan kinerja antara algoritma naïve bayes classifier, k-nearest neighbor, dan support vector machine. Algoritma Naïve Bayes Classifier dipilih karena menurut penelitian sebelumnya mengenai aplikasi assessment kebijakan pemerintah terkait operasi ojek online di masa pandemic (covid-19) menghasilkan tingkat akurasi sebesar 92%, sedangkan algoritma K-Nearest Neighbor dipilih karena menurut penelitian sebelumnya untuk menganalisis sentiment ulasan agen perjalanan menghasilkan tingkat akurasi sebesar 87%, dan algoritma Support Vector Machine dipilih karena menurut penelitian sebelumnya mengenai sentiment pengguna twitter terhadap marketplace di Indonesia algoritma Support Vector Machine kernel RBF lebih baik dibandingkan dengan kernel Linier.

Menggunakan ketiga algoritma tersebut untuk mengklasifikasikan data tweet harus membandingkan penelitian sebelumnya algoritma mana yang menghasilkan tingkat akurasi lebih baik. Proses awal data mining pada data tweet yang sudah dicrawling harus melalui proses preprocessing terlebih dahulu. Setelah melakukan preprocessing dilakukan pembobotan data tweet menggunakan TF-IDF. Kemudian akan mengklasifikasikan akun twitter kedalam metode DISC digunakan ketiga algoritma yaitu algoritma Naïve Bayes Classifier, K-Nearest Neighbor dan Support Vector Machine. Diharapkan dengan membandingkan kinerja ketiga algoritma tersebut dengan tahapan preprocessing, pembobotan TF-IDF dan mengklasifikasikan akun twitter kedalam metode DISC dapat diketahui algoritma mana yang lebih cocok.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mendapatkan data tweet dari pengguna akun twitter?

2. Bagaimana algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine dapat mengklasifikasikan akun twitter seseorang kedalam pendekatan metode DISC?
3. Bagaimana tingkat akurasi perbandingan algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine dalam mengklasifikasikan akun twitter seseorang kedalam pendekatan metode DISC?

### 1.3 Batasan Masalah

Berdasarkan batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan akun twitter yang bersifat personal yang sudah dilakukan pelabelan oleh psikolog secara manual berjumlah 130 akun twitter dengan masing-masing akun berjumlah 200 tweet.
2. Penelitian ini memiliki dataset berupa kata-kata tweet seseorang yang sudah dilakukan pelabelan oleh psikolog dalam pendekatan metode DISC.
3. Hasil dari klasifikasi dengan algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine hanya ke dalam pendekatan metode DISC pada setiap akun twitter.
4. Penelitian ini perhitungan evaluasi yang dilakukan dengan confusion matrix.
5. Penelitian ini akan membandingkan tingkat akurasi, presisi, dan recall pada algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine dalam klasifikasi karakter seseorang.
6. Penelitian ini untuk mengklasifikasikan akun twitter kedalam karakter DISC menggunakan bahasa pemrograman python.

### 1.4 Tujuan Penelitian



Tujuan utama dari penelitian ini adalah untuk mengetahui manfaat media sosial berupa twitter dalam mengklasifikasikan karakter seseorang. Adapun tujuan khusus dari penelitian ini adalah sebagai berikut :

1. Penelitian ini untuk membandingkan algoritma naïve bayes classifier, k-nearest neighbor, dan support vector machine dalam mengklasifikasi karakter seseorang dengan pendekatan metode DISC.
2. Penelitian ini untuk memprediksi apakah tweet dapat di klasifikasikan karakter seseorang dengan algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine.

### **1.5 Manfaat Penelitian**

Melalui penelitian ini diharapkan dapat diperoleh manfaat sebagai berikut :

1. Peneliti dapat mengetahui kinerja algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Support Vector Machine dalam mengklasifikasi akun twitter kedalam metode DISC.
2. Bagi psikolog membantu pekerjaan mengambil keputusan karakter seseorang yang didapatkan dari tweet seseorang pada akun twitternya.

### **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 tahap ini dilakukan pengumpulan data dari data twitter menggunakan crawling dengan library yang disediakan oleh python yaitu tweepy yang berupa seluruh tweet dari beberapa akun. Setelah data diperoleh maka dilakukan pelabelan data secara manual dan kemudian dilakukan

pengklasifikasian menggunakan algoritma yang dipilih yaitu naïve bayes classifier, k-nearest neighbor, dan support vector machine.

#### **1.6.2 Metode Analisis**

Pada tahap ini dilakukan analisis terhadap data yang sudah didapatkan dari crawling seperti preprocessing, analisis akun, dan pembobotan.

#### **1.6.3 Metode Perancangan**

Pada tahap ini dilakukan perancangan sistem seperti alur dari preprocessing sampai pengklasifikasian menggunakan algoritma.

#### **1.6.4 Implementasi**

Pada tahap ini dilakukan pembuatan code program menggunakan python untuk 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 naïve bayes classifier, k-nearest neighbor, dan support vector machine.

### **1.7 Sistematika Penulis**

Sistematika penulis bertujuan untuk mempermudah dalam penyusunan skripsi, maka diperlukan sistematika penulisan yang baik. Sistematika penulis terdiri atas beberapa bagian sebagai berikut :

#### **BAB I      Pendahuluan**

Bab ini mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan, manfaat, metodologi yang digunakan dan sistematika penulisan.

**BAB II Landasan Teori**

Bab ini membahas beberapa teori yang digunakan dalam penelitian, perenvangan dan pembuatan sistem.

**BAB III Analisis dan Perancangan**

Bab ini mengemukakan metode penelitian yang dilakukan dalam analisa, perancangan dan implementasi.

**BAB IV Implementasi dan Pembahasan**

Bab ini memaparkan hasil tahapan penelitian, mulai dari analisis, desain, hasil testing dan implementasi.

**BAB V Penutup**

Bab ini berisi kesimpulan dan saran dari hasil penelitian yang dilakukan.

