

BAB I

PENDAHULUAN

1.1 Latar Belakang

Twitter sebagai salah satu media sosial terbesar dengan jumlah pengguna pada 326 juta pengguna aktif pada Januari 2019 khususnya di Indonesia [1], layanan yang diberikan oleh media sosial ini adalah layanan *microblogging* yaitu sebuah varian *blogging* yang kontennya sangat pendek, ada batasan 140 karakter seperti SMS untuk setiap cuitan. Media sosial tradisional mengadopsi media sosial sebagai cara untuk menjangkau khalayak yang lebih luas, dan sebagian besar selebriti memiliki akun twitter untuk tetap terhubung dengan penggemar mereka. Pengguna mendiskusikan apa yang terjadi secara *realtime*, termasuk perayaan, acara tv, acara olahraga, pemilihan politik dan lain sebagainya [2].

Data akun personal pada twitter dapat digunakan untuk banyak tujuan diantaranya adalah dapat digunakan dalam bidang industri atau lebih spesifik ke sumber daya manusia yang ada di perusahaan dapat membantu memberikan rekomendasi dalam melakukan perekrutan dan penyaringan karyawan agar mendapatkan karyawan yang lebih berkualitas dan mumpuni serta profesional [3], dalam bidang pendidikan atau sekolah menengah sampai dengan perguruan tinggi dapat digunakan untuk mencari karakter anak untuk membantunya menutupi kekurangan dari karakter yang dimiliki agar dapat meraih impian yang diinginkan [8] dan dalam bidang marketing atau bisnis dapat digunakan untuk menargetkan iklan

kepada target yang lebih relevan dan spesifik dari kecenderungan seseorang terhadap suatu hal [9].

Text mining proses atau metode penambangan teks yang berguna untuk mengambil makna penting didalamnya [2], salah satu contoh kasus dalam bidang sumber daya manusia yang ada di perusahaan dapat dilakukan dengan *text mining* adalah dengan melakukan *text mining* ke dalam setiap cuitan yang dibuat oleh calon karyawan pada twitter mereka dapat dilakukan untuk mengetahui karakter dari orang tersebut guna membantu dalam mengindikasi karakter orang tersebut dalam penggunaan bermedia sosial, salah satu proses yang harus dilakukan dalam *text mining* adalah melakukan klasifikasi, terdapat beberapa metode untuk dapat melakukan klasifikasi seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN) dan lainnya, akan tetapi penulis hanya akan menggunakan algoritma *Support Vector Machine* (SVM) sebagai metode untuk melakukan klasifikasi [4].

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* yang memprediksi kelas berdasarkan model atau pola dari proses training. Klasifikasi dilakukan dengan mencari garis pembatas yang memisahkan antara satu kelas dengan kelas yang lain, yang dalam kasus ini garis - garis tersebut berguna untuk memisahkan setiap karakter yang ada[5][6]. SVM dipilih pada penelitian ini karena pada penelitian sebelumnya oleh Lilyani pada tahun 2017 berjudul "Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma *Support Vector Machine* Dan *K-nearest Neighbor* Berbasis

Particle Swarm Optimization” menunjukkan bahwa SVM memiliki tingkat akurasi tinggi yaitu 80,83% dalam penggunaannya untuk melakukan klasifikasi dalam analisis sentimen [7].

Penulis berharap dengan adanya penelitian ini dapat membuahkan sebuah aplikasi yang dapat mengambil cuitan tiap pengguna twitter yang dapat melakukan klasifikasi karakter terhadap setiap akun personal dan juga dapat dikembangkan ke penelitian selanjutnya baik untuk perbaikan dalam penelitian ini maupun penelitian terapan yang dapat diterapkan ke berbagai bidang seperti industri, pendidikan, bisnis dan bidang lainnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, maka dapat dirumuskan beberapa masalah yang ingin diselesaikan dalam penelitian ini adalah:

1. Bagaimana hasil aplikasi dari *text mining* dan klasifikasi karakter dari cuitan di twitter dengan pendekatan DISC ?
2. Bagaimana mengukur akurasi algoritma SVM yang digunakan untuk melakukan klasifikasi karakter pada cuitan twitter ?

1.3 Batasan Masalah

Beberapa batasan dalam penelitian ini adalah sebagai berikut ini :

1. Aplikasi yang digunakan hanya berbasis web.
2. Algoritma yang digunakan untuk melakukan klasifikasi adalah SVM.

3. Objek yang digunakan dalam penelitian ini adalah cuitan twitter akun personal pada lini masanya sebanyak 109 akun.
4. Klasifikasi karakter dengan menggunakan pendekatan DISC.
5. Pada klasifikasi yang dilakukan satu akun hanya diklasifikasi dengan satu karakter.
6. Proses pengambilan cuitan twitter menggunakan interface web twitter sebelum 15 Juli 2019.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan penelitian ini adalah untuk :

1. Untuk membuat aplikasi yang melakukan klasifikasi karakter pada cuitan akun personal twitter dengan pendekatan DISC
2. Untuk mencari teknik mengukur akurasi algoritma SVM yang digunakan untuk melakukan klasifikasi karakter pada cuitan akun personal twitter.

1.5 Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan dari penelitian ini :

1. Bagi instansi pendidikan terkait, dapat digunakan sebagai media konseling siswa untuk mengetahui karakter siswa agar dapat mengatasi kelemahan dan meraih impiannya.
2. Bagi perusahaan, dapat dikembangkan menjadi aplikasi sebagai alat bantu dalam proses perekrutan karyawan baru.

1.6 Metodologi Penelitian

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

1. Pengumpulan Data

Pada tahapan pengumpulan data ini dilakukan pengambilan data latih dari Twitter menggunakan tool scrapy yang berupa seluruh cuitan dari beberapa akun pada linimasa mereka. Setelah data diperoleh maka selanjutnya akan dilakukan pelabelan data secara manual dan kemudian dilakukan pengklasifikasian menggunakan algoritma yang dipilih yaitu SVM.

2. Analisis

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

3. Perancangan

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

4. Implementasi

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

5. Pembahasan

Pada tahap ini dilakukan perhitungan akurasi dari implementasi yang sudah dilakukan. Metode pengujian terhadap algoritma yang digunakan adalah menggunakan metode *Confusion Matrix*. Dalam penelitian ini, tahap pengujian

menggunakan *Confusion Matrix* sebagai parameter untuk melakukan pengujian. Pengujian yang dilakukan adalah mengukur akurasi klasifikasi yang dihasilkan oleh algoritma SVM.

1.7 Sistematika Penulisan

Sistematika dari penulisan skripsi adalah sebagai berikut :

BAB 1. PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.

BAB 2. LANDASAN TEORI

Bab ini akan menjelaskan tentang penelitian sebelumnya dan perbedaan dari penelitian sebelumnya dengan penelitian ini dan berisi teori-teori yang digunakan dalam merancang analisis, data, penulisan serta perangkat yang dibangun sebagai dasar penulisan skripsi.

BAB 3. ANALISIS DAN PERANCANGAN

Dalam bab ini penulis mengemukakan deskripsi umum sistem yang mencakup dalam analisa, perancangan, dan implementasi.

BAB 4. IMPLEMENTASI DAN PEMBAHASAN

Menjelaskan bentuk implementasi aplikasi beserta pembahasan sistem yang telah diselesaikan.

BAB 5. PENUTUP

Pada bagian ini dijelaskan mengenai kesimpulan dan saran guna memperbaiki kelemahan yang terdapat pada aplikasi tersebut

