

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kepribadian merupakan bagian dari jiwa manusia yang membangun keberadaannya secara utuh dan tidak terpecah – pecah. Kepribadian dipakai untuk mengidentifikasi suatu individu dan membedakannya dengan individu lain. Kepribadian bersifat awet dan tidak mudah berubah. Jika kepribadian mengalami perubahan, hal itu dilakukan secara bertahap atau akibat atas kejadian luar biasa yang menimpa individu [1]. Kepribadian dapat dilihat dari bagaimana individu bertingkah laku di kesehariannya. Satu atau dua orang mungkin memiliki kesamaan dalam hal tertentu, namun tetap saja mereka memiliki kepribadian yang unik.

Karakteristik kepribadian yang menggambarkan kebiasaan seseorang dapat dikelompokkan menjadi empat dan semuanya direpresentasikan oleh The Four Temperaments [2]. Karakteristik ini meliputi korelis (pemarah, bergairah, dan mudah emosi), sanguinis (ceria, percaya diri dan optimis), plegmatis (pendiam, tidak emosional dan tidak semangat), dan melankolis (sedih, termenung, depresi). Kita dapat mengetahui karakteristik kepribadian diri kita sendiri maupun orang lain melalui tes psikologi. Tes tersebut adalah menjawab pertanyaan atau memilih pernyataan yang paling sesuai dengan diri kita. Banyak manfaat yang bisa diperoleh dengan mengetahui karakteristik kepribadian seseorang. Salah satunya ialah penelitian yang dilakukan oleh Yessivha Imanuela Claudy, dkk [3]. Mereka membuat penelitian tentang klasifikasi kepribadian seseorang untuk mengetahui karakter calon karyawan. Karakteristik kepribadian calon karyawan menunjukkan

efektivitas kinerja dan kerjasama dalam kelompok sehingga dapat diketahui mana calon karyawan yang berkualitas.

Situs jejaring sosial menyediakan fitur dan layanan yang memungkinkan penggunanya dapat berkomunikasi, memberikan opini, dan bertukar informasi. Salah satu contoh jejaring sosial yang banyak digunakan adalah Twitter. Pengguna dapat menulis cuitan, mengunggah gambar, maupun membagikan video mereka. Cuitan dalam bentuk teks adalah yang paling digemari karena relatif cepat dan mudah dilakukan. Kebiasaan seseorang dalam beropini di Twitter tentu berbeda-beda. Kebiasaan tersebut dapat merepresentasikan karakteristik kepribadian mereka. Sering kali kita berkomunikasi dengan orang lain melalui Twitter tanpa tahu kepribadian mereka. Oleh karena itu terkadang kita dapat salah dalam menyampaikan opini atau memilih kata yang kurang tepat. Hal ini mengakibatkan perbedaan persepsi dan dapat memicu konflik, padahal kita tidak bermaksud demikian.

Untuk itu dalam penelitian ini menggunakan media jejaring sosial Twitter guna mencari tahu karakteristik kepribadian seseorang. Disamping kita dapat menghindari kesalahpahaman dalam berkomunikasi, penelitian ini juga dapat menghasilkan peluang yaitu dapat mempermudah perekrutan calon karyawan atau anggota organisasi. Hal ini tentu sangat bermanfaat bagi perusahaan yang membutuhkan tenaga kerja baru. Perusahaan tidak harus melakukan prosedur wawancara yang panjang untuk mencari calon karyawan yang tepat.

Untuk mengetahui kepribadian seseorang dari Twitter miliknya dapat dilakukan ekstraksi dengan *text mining*. Peneliti akan memanfaatkan API Twitter untuk mendapatkan dataset berupa teks. Setelah itu, dataset diolah untuk dicari

makna yang terkandung didalamnya agar dapat dilakukan klasifikasi kepribadian. Proses pengolahan dataset terdiri dari beberapa tahapan yakni pengumpulan data, *preprocessing*, pembobotan, klasifikasi, implementasi sistem dan testing.

Naïve Bayes Classifier merupakan salah satu algoritma yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. NBC memiliki tingkat akurasi dan kecepatan yang tinggi saat diaplikasikan dengan data yang besar [4]. Dalam penelitian yang berjudul "Klasifikasi Kepribadian Big Five Pengguna Twitter dengan Metode Naïve Bayes" yang dilakukan oleh Yusra, dkk [5] menghasilkan akurasi sebesar 86,66 % dengan perbandingan data latih dan uji 70:30. Oleh karena itu, dalam penelitian ini akan menggunakan metode algoritma NBC untuk menentukan kepribadian seseorang.

Dari penelitian ini akan diketahui bahwa algoritma Naïve Bayes Classifier dapat digunakan untuk mengetahui kepribadian seseorang atau tidak. Penggunaan algoritma NBC yang memiliki akurasi tinggi juga diharapkan mampu memberikan ketepatan dalam menentukan kepribadian seseorang berdasarkan Twitter miliknya.

1.2 Rumusan Masalah

Perumusan masalah yang akan diselesaikan pada penelitian ini adalah.

1. Apakah algoritma Naïve Bayes Classifier dapat menentukan karakteristik kepribadian dari cuitan Twitter seseorang.
2. Berapa nilai akurasi algoritma Naïve Bayes Classifier pada klasifikasi kepribadian berdasarkan cuitan Twitter.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah.

1. Penelitian dilakukan terhadap *tweet* yang berbahasa Indonesia.
2. Klasifikasi dengan metode NBC menggunakan *tools* yang sudah tersedia di *library* Python.
3. Pengunduhan dataset dilakukan dari tanggal 4 September 2020 sampai 5 September 2020.
4. Dari dataset yang ada diklasifikasi menjadi empat kelas yaitu sanguinis, korelis, plagmatis, dan melankolis.
5. Dataset yang digunakan dibagi menjadi data uji dan data latih dengan perbandingan 25% dan 75%.
6. Dataset yang digunakan terdiri dari dua jenis, yaitu dataset yang *balance* dan dataset yang tidak *balance*.
7. Pengunduhan *tweet* pada proses prediksi kepribadian dibatasi sebanyak 100 *tweet* terbaru setiap akun pengguna.
8. Antarmuka yang dibuat menggunakan PHP dan HTML.

1.4 Maksud dan Tujuan Penelitian

Tujuan dari penelitian ini adalah.

1. Menemukan karakteristik kepribadian berdasarkan cuitan Twitter.
2. Menerapkan algoritma Naïve Bayes Classifier pada analisis karakteristik kepribadian terhadap data yang bersumber dari Twitter.
3. Pengujian nilai akurasi yang dihasilkan oleh algoritma Naïve Bayes Classifier dalam mengklasifikasikan kepribadian seseorang.

1.5 Manfaat Penelitian

a. Bagi umum

Mengetahui bagaimana karakteristik kepribadian pengguna Twitter berdasarkan The Four Temperaments.

b. Bagi ilmu pengetahuan

1. Mengetahui bagaimana implementasi *text mining* terhadap data yang berasal dari Twitter menggunakan bahasa pemrograman Python.
2. Mengetahui cara mengolah dan mengklasifikasi data menggunakan algoritma Naïve Bayes Classifier.
3. Mengetahui nilai akurasi algoritma Naïve Bayes Classifier dalam mengklasifikasi data.

1.6 Metode Penelitian

Dalam penelitian ini akan menjelaskan tahapan – tahapan klasifikasi kepribadian pada data Twitter sebagai berikut.

1.6.1 Metode Pengumpulan Data

1.6.1.1 Studi Pustaka

Mencari literatur atau jurnal yang memiliki kaitan dengan klasifikasi data, *text mining*, algoritma Naïve Bayes Classifier, dan pengukuran akurasi menggunakan *K-Fold Cross Validation*.

1.6.1.2 Survey

Mendapatkan dataset berupa *username* Twitter dan kepribadian yang sudah ditentukan berdasarkan kuesioner. Dataset tersebut digunakan sebagai acuan untuk *text mining* dan validasi hasil testing program.

1.6.1.3 Text Mining

Mengumpulkan cuitan dari akun Twitter sesuai dengan *username* yang didapatkan dari survey. Cuitan tersebut diperoleh dengan cara *crawling* menggunakan Twitter API.

1.6.2 Metode Analisis

Dalam metode analisis terdapat beberapa tahapan sebagai berikut.

1. Data Preprocessing

Tahap *preprocessing* data dari Twitter meliputi *cleaning*, *tokenization*, *slang replacement*, *stemming*, *stopword removal*, *POS Tagging*, *POS Filtering*, dan *common word removal*. Tahap ini bertujuan untuk mendapatkan data berupa kata – kata yang bersih dan agar dapat lebih mudah diolah.

2. Pembobotan Kata

Pada tahap ini kata – kata yang telah bersih dihitung frekuensi kemunculannya. Pembobotan kata tersebut menggunakan *Term Frequency*.

3. Klasifikasi Naïve Bayes Classifier

Pada tahap ini data diklasifikasi menggunakan algoritma Naïve Bayes Classification sehingga data dapat terbagi ke dalam kelas – kelas yang sudah ditentukan. Tahap ini dibagi menjadi dua yaitu tahap *data training* dengan pemberian label secara manual dan tahap *data testing* berdasarkan data yang diunduh dari Twitter secara *realtime*.

1.6.3 Metode Perancangan

Merancang alur dari sistem yang akan dibuat menggunakan flowchart, merancang database untuk menyimpan data, dan membuat tampilan antarmuka pengguna.

1.6.4 Metode Implementasi

Implementasi algoritma menggunakan bahasa pemrograman Python yang dibangun menggunakan *framework* Flask dan diintegrasikan ke dalam sistem dengan bahasa pemrograman PHP sebagai antarmukanya.

1.6.5 Metode Pengujian

Pengujian sistem yang telah dibuat menggunakan metode *K-Fold Cross Validation* dan menggunakan data uji untuk mendapatkan nilai akurasi ketepatan sistem dalam menentukan kepribadian seseorang.

1.7 Sistematika Penulisan

Sistematika penulisan pada skripsi ini terdiri dari beberapa bagian yang diuraikan dalam bentuk bab – bab sebagai berikut.

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Pada bab ini menjelaskan mengenai teori yang menjadi dasar dalam penelitian.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini menjelaskan mengenai metode yang dipakai dalam penelitian serta alur perancangan sistem yang akan dibangun dalam studi kasus klasifikasi kepribadian berdasarkan cuitan Twitter.

BAB IV IMPLEMENTASI

Pada bab ini menjelaskan mengenai implementasi algoritma *Naïve Bayes Classification* dalam mengklasifikasikan data yang bersumber dari Twitter berdasarkan *The Four Temperaments*. Bab ini juga membahas mengenai proses pengujian dari sistem yang dibuat.

BAB V PENUTUP

Pada bab ini menjelaskan kesimpulan yang diperoleh dari hasil analisis penelitian dan saran untuk pengembangan yang lebih baik lagi.