

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi yang semakin pesat merupakan bukti bahwa dewasa ini masyarakat semakin peka dan semakin membutuhkan akan adanya teknologi di segala aspek kehidupan mereka, mulai dari hal-hal kecil seperti media untuk berbagi informasi seperti *smartphone*, komputer, *internet* hingga kebutuhan transportasi seperti mobil, pesawat dan kapal. Semua hal itu tentunya tidak lepas dari kenyataan bahwa manusia menciptakan berbagai macam alat/benda nyata maupun benda *virtual* seperti *software* dengan tujuan untuk mempermudah pekerjaannya.

Tidak hanya untuk mempermudah pekerjaan saja, bahkan untuk melakukan sosialisasi dan bergaul pun kini dapat dilakukan dengan lebih mudah karena adanya berbagai media sosial. Hal ini tentu sangat dapat membantu untuk dapat saling berkomunikasi dengan keluarga dan kerabat yang terpisah karena jarak, untuk bertatap muka pun tak harus pergi jauh-jauh karena sudah ada *video call*, untuk saling mengetahui kabarpun kadang tidak perlu lagi harus menelfon karena sudah dapat dilihat dari update postingan dari keluarga atau kerabat tersebut. Indonesia sendiri merupakan salah satu negara pengguna media sosial aktif terbesar di dunia seperti Facebook, Twitter, Instagram dan lain-lain.

Ketertarikan masyarakat terhadap media sosial memang sangatlah besar, akan tetapi dengan jumlah yang sangat besar itu tidak semua menimbulkan efek positif, contoh kecilnya saja ketika bersosialisasi secara langsung atau berkumpul bersama teman-teman, masih banyak yang sibuk dengan *gadget*-nya masing-masing, belum lagi banyak kasus *bullying* dan perkelahian antar penggunanya di sosial media. Hal ini tidak terlepas dari sifat dan kepribadian dari masing-masing penggunanya juga, sama halnya dengan bersosial secara langsung, pasti ada juga pribadi yang tidak cocok dengan pribadi lainnya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pendekatan yang digunakan dalam psikologi untuk melihat kepribadian manusia melalui *trait* yang tersusun dalam lima buah domain kepribadian yang telah di bentuk dengan menggunakan analisis faktor. Adapun lima buah domain atau model pembagian teori *big five* ini lebih dikernal dengan OCEAN yaitu singkatan dari *Openess to experience, Conscientiousness, Extraversion, Agreeableness* dan *Neuroticism*.

Maka dari itu peneliti mengangkat topik tentang klasifikasi kepribadian di sosial media menggunakan algoritma *K-Nearest Neighbor* dan teori kepribadian *big five* dengan menggunakan data berupa kumpulan *tweet* di Twitter sebagai datanya. Peneliti menggunakan media sosial Twitter sebagai bahan penelitian karena Twitter merupakan salah satu media sosial paling populer di dunia termasuk di Indonesia, *K-Nearest Neighbor*. *K-Nearest Neighbor* menurut (Santoso, 2007) merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*). *K-Nearest Neighbor* memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak *noise* dan efektif apabila *training* datanya besar.

Penelitian dengan kasus yang sama sebelumnya sudah pernah dilakukan oleh Rinaldi Syafrianto, yang menggunakan data yang sama dengan penelitian ini akan tetapi menggunakan metode yang berbeda dalam skripsi nya yang berjudul *Klasifikasi Kepribadian Akun Twitter Berdasarkan tweet menggunakan Metode Naïve Bayes* dan menghasilkan tingkat akurasi tertinggi 86,66% dengan perbandingan data latih dan data uji sebesar 70:30. Penelitian inilah yang akan menjadi rujukan utama peneliti dalam melakukan penelitian ini yang nantinya akan dibandingkan hasil akhirnya.

Penelitian terkait metode *K-Nearest Neighbor* ini juga sudah banyak dilakukan seperti yang dilakukan oleh Windu Gata dan Purnomo di dalam Jurnal *Format Volume 6 Nomor 1 Tahun 2017 :: ISSN : 2089 -5615* yg berjudul *Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS yang berhasil mendapatkan tingkat akurasi sebesar 96,15%*. Ada juga penelitian yang dilakukan oleh Mustakim dan Giantika Oktavini F dalam Jurnal *Sains, Teknologi dan Industri, Vol. 13, No.2, Juni 2016, pp.195 – 202* yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berjudul Algoritma *K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa yang menghasilkan tingkat akurasi sebesar 82%. Berikutnya ada penelitian yang dilakukan oleh Zuliar Efendi dan Mustakim dalam jurnal Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9 halaman 2579-5406 pada tahun 2017 yang berjudul *Text Mining Classification* Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi yang juga menggunakan algoritma *K-Nearest Neighbor* berhasil mendapatkan tingkat akurasi sebesar 60% dengan nilai K sebanyak 15.

Dari berbagai permasalahan diatas maka peneliti akan melakukan pengujian dengan algoritma KNN untuk melakukan klasifikasi kepribadian dengan menggunakan teori kepribadian *big five personality* terhadap data yang merupakan kumpulan *tweet* dari orang-orang yang telah mengisi angket tes kepribadian yang telah disebar oleh peneliti sebelumnya dan diharapkan dapat mengetahui bagaimana hasil akurasi, apakah lebih tinggi atau lebih rendah dari penelitian yang dilakukan oleh peneliti sebelumnya.

1.2. Rumusan Masalah

Berdasarkan dari permasalahan yang telah dibahas pada latar belakang, maka dapat disimpulkan bahwa rumusan masalah pada penelitian ini adalah bagaimana menerapkan algoritma *K-Nearest Neighbor* untuk mengklasifikasikan kepribadian pemilik akun Twitter dengan teori kepribadian *Big Five Personality* berdasarkan *tweet*-nya.

1.3. Tujuan Penelitian

Adapun beberapa tujuan yang ingin dicapai dari penelitian ini adalah:

1. Menenerapkan *K-Nearest Neighbor* untuk mengklasifikasikan kepribadian pemilik akun Twitter berdasarkan teori kepribadian *Big Five Personality*.
2. Mengetahui tingkat akurasi penerapan *K-Nearest Neighbor* dan membandingkan hasil penelitian ini dengan penelitian sebelumnya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.4. Batasan Masalah

Batasan masalah pada penelitian ini bertujuan agar pembahasan tidak meluas dan dapat terfokus pada bidang kajiannya maka batasan masalahnya sebagai berikut:

1. Data yang digunakan adalah data sekunder yang merupakan kumpulan *tweet* dari orang-orang yang telah mengisi angket tes psikologi dengan menggunakan teori kepribadian *Big Five Personality*.
2. Jumlah data yang digunakan adalah sebanyak 1500 *tweet* dari total 15 orang.
3. Variabel inputan nya adalah field teks yang merupakan kutipan *tweet* dan kelas yang merupakan hasil klasifikasi kepribadian dengan menggunakan angket.
4. Kelas terdiri atas 5 yaitu Openess, Conscientiouness, Extraversion, Agreeableness dan Neuroticism yang akan disingkat menjadi O, C, E, A, N.

1.5. Sistematika Penulisan

Berikut ini merupakan rencana susunan sistematika penulisan laporan Tugas Akhir yang akan dibuat:

Bab I Pendahuluan

Bagian ini berisi tentang deskripsi umum penelitian yang meliputi latar belakang masalah, perumusan masalah, tujuan penelitian, batasan penelitian, serta sistematika penulisan.

Bab II Landasan Teori

Bagian ini menjelaskan tentang teori-teori yang digunakan dalam penelitian meliputi teori umum *Text Mining*, *K-Nearest Neighbor* serta Teori Kepribadian *Big Five Personality*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bab III Metodologi Penelitian

Bagian ini menjelaskan tentang metodologi yang digunakan dalam penelitian, mulai dari identifikasi masalah, teknik pengumpulan data, analisa algoritma dan alat bantu dalam penelitian.

Bab IV Analisa

Bab ini menjelaskan tentang analisa kebutuhan data serta analisa kebutuhan data *mining* selama proses penelitian.

Bab V Pengujian

Bab ini menjelaskan proses pengujian dengan *tools* yang telah ditentukan serta tahap evaluasi terhadap hasil pengujian.

Bab VI Penutup

Bab ini berisi tentang kesimpulan dari keseluruhan penelitian dan saran-saran yang perlu disampaikan kepada pembaca.