

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan penggunaan media sosial di Indonesia sangat berkembang pesat. Bila dilihat dari jumlah pengguna internet, maka bisa dibilang seluruh pengguna media sosial di Indonesia. *We Are Social* mengatakan 132,7 juta pengguna internet, 130 juta diantaranya pengguna aktif di media sosial. Media sosial adalah sebuah media *online*, dengan para penggunanya dengan mudah berpartisipasi, dan berbagi. Blog dan jejaring social merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Seorang pasti memiliki macam-macam motivasi dalam menggunakan medsos. Untuk mencari kabar perkembangan sesuatu, untuk berbagi informasi maupun untuk mengikuti salah satu yang menjadi *trend* saat ini yaitu menggunakan medsos sebagai untuk apresiasi diri. *Facebook* salah satu medsos yang sangat populer dan banyak digemari orang di semua dunia. Dari anak orangtua sudah menggunakan *facebook*. Kebanyakan dari mereka memakai *facebook* untuk menjalin hubungan pertemanan dan berkomunikasi dengan banyak orang, yaitu sahabat bahkan teman jauh. Pengertian *facebook* secara lengkap ialah sebuah situs internet sosial yang memungkinkan pengguna dapat saling berinteraksi dengan pengguna lain di seluruh dunia maya. Mark Zuckerberg mendirikan *facebook* saat ia belajar di fakultas psikologi di *Harvard University*. Pada Februari tahun 2004,

Mark menegakkan *Facebook*, dimana nama itu berasal dari lembaran *profil* yang dibagikannya kepada mahasiswa baru di *Harvard*. Pada awalnya, *facebook* dikenalkan hanya untuk kalangan mahasiswa. Hal itu digunakan sebagai media perkenalan bagi para mahasiswa *Harvard*. Seseorang yang menjadikan *Facebook* sebagai media sosial favorit mereka adalah orang yang suka menjalin persahabatan dan juga berbagi. Karakter lain yang dapat terlihat dari pengguna *Facebook* adalah orang yang *ekstrovert*. Hal ini dilihat dari orang yang suka menggunakan *Facebook* akan menunjukkan segala kegiatannya di *wall* mereka. Masyarakat dalam mengakses media sosial tentu bukan lagi menyangkut persoalan pribadi saja, melainkan sudah berkembang pada *aksesibilitas* ranah publik yang tinggi. Batas privasi seakan menjadi sampingan manakala menu perhatian ruang publik semakin meningkat. Derasnya perhatian ruang publik yang dikonsumsi menandakan demokrasi semakin menunjukkan tarungnya. Eksistensi dan perhatian masyarakat terhadap apa yang menjadi kehidupan ruang publik sangat diperhatikan. Hal itu ditandai dengan munculnya kepribadian yang berbeda-beda dimana perkataan yang mereka tulis diberanda media social ada yang menghujat, memberitakan berita yang tidak nyata (*hoax*) dan lain sebagainya.

Banyak ada perkembangan dari media sosial ini terhadap pendidikan anak dibawah umur sekolah dasar, ada dampak *positif* ataupun dampak *negatif*, terlebih lagi pada dampak yang dapat membuat perubahan kepada sosial anak. Seluruh masyarakat kita hampir semuanya kelengkapan bagi penghuni rumahnya selalu berbagi bentuk hasil karya teknologi. Hasil informasi dan teknologi komunikasi

seperti media sosial menjadikan seorang anak “orang asing” yang mengakibatkan dunia menjadi begitu luas hadir di pertengahan keluarga, mengajari *user* apa saja setiap waktu, mengubah pola kehidupan, berkomunikasi dengan orang lain dapat merubah perilaku kebutuhan teknologi. Zaman sekarang teknologi dibuat begitu simple untuk *user*, menikmati fitur-fitur aplikasi yang ada dengan medsos. Bahkan anak-anak sekolah dasar bisa cepat dalam belajari pengguna perangkat teknologi yang banyak-banyak dipakai oleh orang besar seperti handphone genggam maupun PC yang bersambung dengan internet sehingga diberikan kenyamanan internet yang luar biasa ke berbagai macam-macam situs. Yang banyak disediakan secara *free*. Banyak dari jenis *merk* perangkat handphone saat ini menyediakan fitur sosmed yang *free* atau tidak bayar dan simple di unduh (Fitri, 2017)

Algoritma *Nearest Neighbor (k-nearest neighbor)* adalah algoritma untuk melakukan *classifikasi* terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Kasus khusus di mana *classifikasi* diprediksikan berdasarkan data pembelajaran yang paling dekat disebut algoritma *k-nearest neighbor*. Dalam metod KNN penggunaan atribut kata yang hadir menjadi dasar klasifikasi untuk penggolongan atau klasifikasi teks. Dari penjelasan tersebut peneliti ingin melakukan pengklasifikasian teks dengan menggunakan metode KNN untuk mendapatkan hasil. Untuk selanjutnya di golongan mana yang akan masuk dalam kepribadian *Big Five Personality* yaitu *Openness, Conscientiousness, Extraversion, Agreeableness dan Neuroticism*.

Berdasarkan latar belakang tersebut, maka penulis mengambil judul **“Penerapan *Text Mining* dalam menganalisis kepribadian pengguna media sosial dengan *K-Nearest Neighbor* (Studi Kasus :*Facebook*)**

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut penulis akan mengkasifikasikan teks yang didapatkan dan menggolongkan teks tersebut ke dalam *Big Five Personality* untuk mengetahui kepribadian pengguna media sosial menggunakan metode *K-Nearest Neighbor*?

1.3. Batasan Masalah

1. Mengklasifikasikan kepribadian menggunakan metode klasifikasi *K-Nearest Neighbor*(KNN) .
2. Sumber data yang digunakan hanya dari social media yaitu *Facebook*

1.4. Tujuan dan Manfaat Penulisan

1.4.1. Tujuan Penelitian

1. Dapat mengetahui kepribadian pengguna media social berdasarkan teori *Big Five Personality*.
2. Mengklasifikasikan jenis kepribadian melalui social media menggunakan *K-Nearest Neighbor* dan teori *Big Five*
3. Mengetahui hasil dari klasifikasi teks menggunakan *K-Nearest Neighbor*

1.4.2. Manfaat Penelitian

1. Menambah pengetahuan mengenai analisis kepribadian pengguna media social.
2. Mengetahui bahwa teori *Big Five Personality* dapat menyimpulkan kepribadian pengguna media social.
3. Membantu seseorang untuk mengetahui kepribadiannya.

1.5 Metodologi Penelitian

1.5.1 Waktu Penelitian

Waktu penelitian dilakukan selama 9 bulan yaitu mulai dari bulan Mei 2019 sampai dengan february 2020.

1.5.2 Data Penelitian

Dalam penelitian ini menggunakan jenis data sekunder. Data sekunder adalah pengumpulan data melalui cara tidak langsung atau harus melakukan pencarian mendalam dahulu seperti melalui internet, literatur, statistik, dan lain-lain. Penelitian ini mengambil data melalui internet melalui media sosial facebook untuk diambil status pengguna nya sebanyak 550 data status.

1.5.3 Metode Penelitian

Pada penelitian ini metode yang digunakan yaitu metode deskriptif. Metode deskriptif adalah metode yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk generalisasi.

1.5.4 Metode Pengumpulan Data

Dalam penelitian ini bagian – bagian penelitian akan dilakukan secara *komputerisasi*. Mulai dari pengambilan data dengan teknik *Web Scrapping* hingga analisis kepribadian pengguna media social yang menggunakan metode *K-Nearest Neighbor*.

1. *Web Scrapping*

Web Scraping adalah pengambilan suatu dokument semi-terstruktur dari internet, umumnya berupa halaman *Web* yang dibangun dengan bahasa *markup* seperti HTML atau XHTML yang tujuannya untuk mengambil informasi dari halaman tersebut baik secara seluruh atau bagian digunakan untuk kepentingan lain. Umumnya, ada empat tahap dalam penggunaan *Web Scraping* untuk mengambil data untuk mengambil data secara otomatis dari sebuah laman *Web* sebagai berikut :

1. Belajar document HTML dari *website* yang akan diambil informasinya untuk *tag* HTML yang mengapit informasi yang akan diambil.
2. Menelusuri mekanisme navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat.
3. Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi *Web scarper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang akan ditentukan
4. Informasi yang didapat dari langkah 3 disimpan dalam format data tertentu.

Dalam penelitian ini *web scraping* digunakan untuk mengambil data dari sebuah laman *web* kemudian melakukan transformasi dari bentuk yang tidak terstruktur, umumnya dalam format HTML menjadi suatu format data terstruktur yang dapat disimpan ke dalam *database* untuk keperluan repository maupun analisis lebih lanjut.(Priyanto & Ma'arif, 2018)

1.5.5 Teknik Analisis

Algoritma *K-Nearest Neighbor* merupakan suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada KNN. Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Algoritma KNN bekerja berdasarkan jarak terpendek dari query instance ke training data untuk menentukan KNN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *euclidian distance*(Andono, 2012). Berikut ialah rumus *Ecludian Distance* :

$$\left(\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right)^{1/2}$$

Dimana,

X_{ik} = nilai X pada training data

X_{jk} = nilai X pada testing data

m = batas jumlah banyaknya data

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka

akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai k yang terbaik itu tergantung pada jumlah data. Ukuran nilai k yang besar belum tentu menjadi nilai k yang terbaik begitupun juga sebaliknya.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini dimaksudkan agar dapat menjadi pedoman atau garis besar penulisan dan dapat menggambarkan secara jelas isi dari laporan penelitian sehingga terlihat hubungan antara bab awal hingga bab terakhir. Sistem penulisan laporan penelitian ini terdiri atas :

BAB I PENDAHULUAN

Pada bab pertama dibahas tentang latar belakang penelitian, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, tentang waktu, data penelitian yang digunakan, metod penelitian, metode pengumpulan data, teknik analisis serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi landasan teori yaitu teori-teori umum dan khusus yang mendukung penulisan skripsi ini. Hal-hal yang tercakup didalamnya adalah pembahasan Text Mining, *K-Nearest Neighbor*, Kepribadian dan *Facebook* serta berisi penelitian terdahulu dan kerangka berpikir.

BAB III ANALISIS TEXT MINING

Bab ini membahas tentang *procedure* pengambilan data dan analisis text mining yang menerapkan tahap-tahap *preprocessing*, *tokenizing*, *filtering*, *steaming*, dan *tagging*. lalu menganalisis dengan tahapan : Pembersihan data (*data cleaning*), *feature selection*, dan *feature ekstraksi*

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang lanjutan dari KDD yang telah dilakukan pada bab IV yaitu proses *mining*, evaluasi pola (*pattern evaluation*), presentasi pengetahuan (*knowledge presentation*) serta hasil dan pembahasan perhitungan dari teknik *classification* dengan pendekatan menggunakan algoritma *k-nearest neighbor* yang dihasilkan oleh penelitian ini.

BAB V KESIMPULAN DAN SARAN

Bab V berisi kesimpulan yang didapatkan dari hasil penelitian dan saran untuk perbaikan atau pengembangan seterusnya dari hasil penelitian ini.