

BAB 1

PENDAHULUAN

1.1. Latar belakang

Industry 4.0 merupakan industri yang berkaitan erat dengan perkembangan teknologi, salah satu teknologi yang berkembang dengan pesat adalah teknologi informasi, dengan media sosial sebagai salah satu contohnya. Media sosial merupakan tempat untuk berbagi dan menyebarkan informasi maupun opini. Beberapa media yang populer digunakan ialah Facebook, Instagram serta Twitter. Pada survei yang dilakukan di tahun 2013 Indonesia menempati posisi ke 5 pengguna Twitter terbesar di dunia[1].

Hasil survei terbaru pada tahun 2020 Indonesia Memiliki 175,4 juta pengguna internet dengan total pengguna media sosial 160 juta pengguna aktif meningkat 12 juta pengguna dari rentang April 2019 sampai dengan Januari 2020. Dari hasil ini dapat disimpulkan bahwa minat dan penggunaan dalam mengakses media sosial sangatlah besar di Indonesia[2].

Twitter adalah salah satu media penyebaran informasi, mulai dari berita, kritik, opini dengan data yang berlimpah. Sehingga memudahkan untuk melakukan Analisa terhadap suatu produk atau jasa. Penulis tertarik untuk mencari informasi dan opini yang disampaikan oleh pengguna di media sosial terkait tanggapan provider By.u. By.u merupakan produk baru dari telkomsel yang diluncurkan pada tahun 2019 dengan taglinenya “semuanya semauanya”, dari tagline tersebut

menjelaskan bahwa pengguna dapat mengontrol secara penuh layanan-layanan yang diinginkan.

Opini yang diberikan oleh masyarakat mengenai provider By.U, baik opini positif, netral, maupun negatif terhadap pelayanan yang diterima oleh masyarakat, dari banyaknya opini masyarakat dimedia sosial dapat menjadi kontribusi untuk perkembangan By.U sehingga perusahaan dapat mengetahui pelayanan yang didapatkan oleh konsumennya serta perusahaan dapat terus melakukan perbaikan pada setandainya pelayanannya. Untuk menganalisis dan mengklasifikasi opini yang diberikan oleh masyarakat, apakah opini tersebut merupakan opini positif, netral, atau negatif, hal ini dapat dilakukan dengan Text Mining.

Metode multinomial naïve bayes yang merupakan salah satu algoritma yang populer dalam pengklasifikasian text. Multinomial Naïve bayes merupakan algoritma yang efektif, multinomial naïve bayes menggunakan data ada masa lampau untuk melakukan prediksi disama mendatang, pada praktiknya pengklasifikasian text menggunakan multinomial naïve bayes memiliki hasil pengujian akurasi yang cukup baik[3], dengan data latih pada masing masing kategori sangat berpengaruh pada hasil akurasi yang didapatkan, semakin banyak data latih pada masing-masing kategori maka akan semakin baik pula hasil yang didapat[4].

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka permasalahan yang dapat dirumuskan adalah :

1. Apakah metode Multinomial Naïve Bayes dapat digunakan dalam pengklasifikasian Analisis Sentimen Provider By.U?
2. Seberapa baik performa metode Multinomial Naïve Bayes dalam pengklasifikasian Analisis Sentimen Provider By.U?

1.3. Batasan Masalah

Berdasarkan rumusan masalah yang dipaparkan di atas, beberapa hal batasan masalah antara lain :

1. Tweet yang di gunakan hanya tweet dengan bahasa Indonesia.
2. Tweet yang di gunakan adalah tweet dengan kata kunci "By.U".
3. Menggunakan bahasa pemrograman Python.

1.4. Maksud dan Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini antara lain :

1. Menguji akurasi yang dihasilkan dari metode Multinomial Naïve Bayes terhadap data tweet provider By.U.
2. Menganalisis dan mengklasifikasi sentiment pada sebuah data tweet menggunakan Multinomial Naïve Bayes untuk mendapatkan kategori sentiment yang sesuai.

1.5. Metode Penelitian

Adapun metode penelitian yang digunakan dalam penelitian ini adalah sebagai berikut:

1.5.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini sebagai berikut :

a. Crawling data

Pada proses ini pengambilan data set dilakukan menggunakan metode crawling yaitu pengambilan sebuah data pada sosial media twitter dengan kata kunci "By.U" sebagai data set. Dengan memanfaatkan library tweepy yang ada pada python.

b. Studi Litelatur

Pengumpulan data dengan cara mencari referensi teori yang relevan diantaranya jurnal, paper, buku dan hasil penelitian yang berkaitan dengan sentimen analisis dan Naïve Bayes.

1.5.2. Metode Analisis

Pada tahapan ini terdapat beberapa tahapan antara lain yaitu:

a. Preprocessing

Preprocessing adalah tahap yang dilakukan sebelum proses pengklasifikasian sentimen. Pada proses dataset diolah agar siap dan dapat dibaca oleh machine learning. Terdapat beberapa tahapan dalam preprocessing, di antaranya adalah :

1. Filtering

Pada proses ini data yang telah didapatkan dari proses crawling akan dilakukan pemfilteran, dengan tujuan untuk mengambil kata-kata penting

dari hasil crawling, filtering dapat menggunakan dua algoritma yaitu *word list* (menyimpan kata penting) atau *stop list* (membuang kata yang kurang penting).

2. Tokenizing

Tokenisasi adalah sebuah proses yang dilakukan untuk memotong atau memecah kalimat menjadi beberapa bagian atau kata. Dengan hasil pemotongan disebut token.

3. Normalisasi

Normalisasi adalah tahapan untuk merubah kalimat-kalimat singkatan atau kata-kata yang tidak sesuai dengan KBBI diubah sesuai dengan KBBI.

4. Cleansing

Cleansing adalah tahap penghapusan karakter atau tanda baca yang tidak diperlukan dihapus dari teks. Contoh karakter yang dibuang adalah tanda seru, tanda tanya, koma dan titik.

5. Transform Case

Transform Case merupakan proses yang dilakukan untuk merubah seluruh karakter dengan huruf kapital menjadi huruf standar.

6. Stemming

Stemming adalah tahapan untuk merubah kata-kata yang berimbuhan menjadi kata dasar dengan cara menghilangkan prefix dan suffix dari kata hasil filtering, dengan tujuan untuk merubah bentuk kata menjadi kata dasar dari kata tersebut.

b. Pembobotan kata

Term Frequency–Inverse Document Frequency (TF-IDF) adalah adalah Teknik pembobot kata berdasarkan frekuensi atau banyak kemunculan suatu kata dalam sebuah dokumen. Pada penelitian ini metode pembobotan yang digunakan adalah TF-IDF.

c. Klasifikasi

Metode yang digunakan dalam melakukan klasifikasi adalah Multinomial naïve bayes. Multinomial naïve bayes merupakan pengembangan dari model naïve bayes dengan menggunakan metode probabilitas dan statistik yang dikemukakan untuk dapat memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

1.5.3. Metode Evaluasi

Dari hasil analisis dilakukan evaluasi dengan menghitung precision, recall, fl, dan accuracy untuk memberi tahu seberapa baik model yang kita buat menggunakan confusion matrix.

1.6. Sistematika Penulisan

Berikut ini adalah sistematika penulisan yang dilakukan penulis:

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan penelitian yang telah dilakukan sebelumnya serta landasan teori yang berhubungan dengan topik yang akan dibahas dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini membahas metode yang dilakukan dalam penelitian serta langkah-langkah yang harus dilakukan untuk mencapai tujuan dalam penelitian tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini berisi uraian tentang hasil penelitian dan pembahasan terhadap hasil penelitian yang dilakukan.

BAB V PENUTUP

Bab ini menjelaskan kesimpulan dari seluruh tahapan penelitian serta saran terhadap penelitian yang dilakukan.