

BAB I

Pendahuluan

1.1 Latar Belakang

Perkembangan internet di Indonesia cukup pesat, hal ini dapat dibuktikan dengan semakin mudahnya akses internet dengan semakin banyaknya provider yang memberikan paket internetan beragam, kemudian perkembangan teknologi smartphone, lalu teknologi 4G yang membuat masyarakat jadi jauh lebih mudah dalam mengakses informasi yang mereka butuhkan dengan cepat melalui internet.

Dengan semakin meningkatnya perkembangan internet maupun penggunaannya. Tentu saja hal ini bisa dimanfaatkan oleh setiap provider untuk mendapatkan pengguna baru agar menggunakan jaringan provider mereka. Tentu saja untuk mendapatkan seorang pengguna atau pelanggan, provider tersebut harus dapat melayani dan memberikan pelayanan maksimal terhadap pengguna provider mereka, entah berupa harga paket internet, kecepatan maupun kekuatan jaringan tersebut. Hal ini dimanfaatkan oleh provider Telkomsel, dimana provider tersebut belakangan mengeluarkan layanan seluler prabayar digital pertama di Indonesia dengan nama By.U. By.U dari Telkomsel ini hadir dengan tagline “Semuanya Semaunya” ini dikeluarkan dengan harapan agar dapat mempertahankan gelar market leader di industry telekomunikasi, dengan tagline tersebut menggambarkan kepada pengguna bahwa pengguna dapat mengontrol secara penuh layanan – layanan yang diinginkan dan dibutuhkan.

Pengguna Internet di Indonesia per Januari 2019 mencapai 150 juta orang dari total sekitar 268 juta penduduk di Indonesia, sedangkan untuk pengguna social media mencapai total 130 juta orang. Ternyata pengguna internet dan social media

di Indonesia mengalami peningkatan dari tahun lalu yang mencapai 13% untuk pengguna internet dan 8,3% untuk pengguna social media [1]. Pengguna social media yang banyak ini bisa dijadikan sebagai referensi terhadap aktivitas pelanggan melalui akun social media mereka baik itu berupa masukan, keluhan atau pun sekedar komentar terhadap penggunaan provider By.U. Apalagi provider tersebut termasuk baru kemudian berasal dari Provider Telkomsel yang dikenal dengan kualitas jaringan serta kecepatannya di Indonesia. Masyarakat biasanya sering memberikan opini maupun pendapat tentang suatu layanan yang digunakan melalui berbagai media, salah satunya media social Twitter. Opini yang diberikan masyarakat terhadap provider By.U ini entah berupa sebuah opini positif atau pun negative terhadap pelayanan yang diterimanya, membuat masyarakat semakin selektif dalam memilih provider internet yang akan dipilih. Banyaknya opini masyarakat di social media ini justru dapat memberikan kontribusi untuk pengembangan By.U ke depannya, sehingga perusahaan mengetahui pelayanan yang telah didapatkan oleh konsumennya dan perusahaan dapat terus memperbaiki standar pelayanannya.

Analisis sentiment adalah studi yang menganalisis opini, sentiment, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik [2]. Untuk memperoleh informasi yang diperlukan dari opini yang sudah diunggah oleh pengguna di social media yaitu perlunya pengolahan data terlebih dahulu kemudian memisahkan opini – opini tersebut ke dalam kelas positif dan negative.

Dalam analisis sentiment ada beberapa pendekatan yang dapat dilakukan salah satunya adalah pembelajaran mesin. Dalam pembelajaran mesin ini terdapat

banyak algoritma yang telah menerapkan implementasinya pada teks atau dokumen, seperti *Support Vector Machine*, *K-Nearest Neighbor*, *Naïve Bayes*, *Decision Tree* dan lain - lain. Di penelitian ini peneliti akan menggunakan 2 algoritma yaitu, *K-Nearest Neighbour* dan *Naïve Bayes Classifier*.

Naïve Bayes Classifier merupakan metode klasifikasi sederhana dan paling umum digunakan. *Naive Bayes Classifier* menghitung probabilitas kelas berdasarkan distribusi kata – kata yang ada dalam dokumen. *Naïve Bayes* memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi. Sedangkan *K-Nearest Neighbour* merupakan salah satu algoritma supervised learning berarti membutuhkan data training untuk mengklasifikasikan objek yang jaraknya terdekat, prinsip kerjanya yaitu mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (neighbour) dalam data training [3].

Oleh karena itu, dilakukanlah penelitian perbandingan antara dua algoritma klasifikasi *K-Nearest Neighbour* dan *Naïve Bayes Classifier* untuk mengetahui akurasi dan waktu proses dalam analisis sentiment terhadap penggunaan provider By.U. Perbandingan ini bertujuan untuk mengetahui algoritma mana yang memiliki akurasi terbaik dalam memproses data klasifikasi terhadap data training yang dijadikan objek.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disampaikan, maka rumusan masalah penelitian ini adalah : Diantara algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier*, manakah algoritma yang memiliki akurasi terbaik?

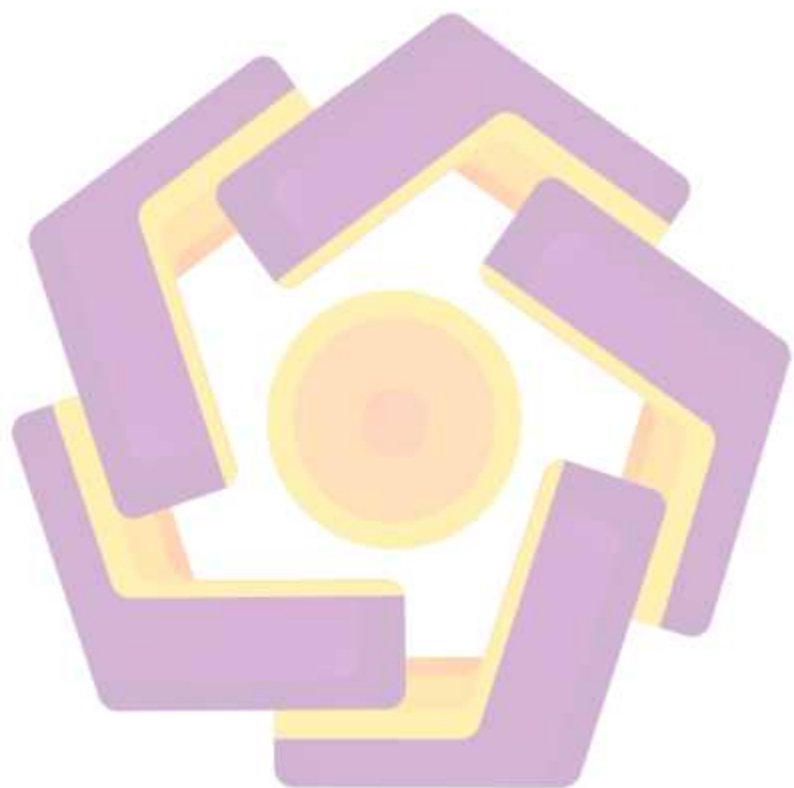
1.3 Batasan Masalah

Batasan masalah pada skripsi ini yaitu :

1. Menggunakan sentiment yang terkait dengan provider By.U yang diberikan oleh pengguna di twitter menggunakan algoritma *k-nearest neighbour* dan *naïve bayes classifier*.
2. Penelitian ini menggunakan 2 model Algoritma klasifikasi, yaitu *k-nearest neighbour* dan *naïve bayes classifier*.
3. Penelitian ini hanya mengukur akurasi dari kedua model dari masing – masing algoritma.
4. *Tweet* yang digunakan adalah tweet berbahasa Indonesia.
5. Bahasa pemrograman yang dipakai adalah Python.

1.4 Maksud dan Tujuan Penelttian

1. Penelitian ini bertujuan untuk membandingkan algoritma *k-nearest neighbour* dan juga *naïve bayes classifier* untuk melihat hasil akurasi antara kedua algoritma yang telah dibuat.
2. Untuk mengetahui sentiment pelanggan yang diperoleh dari hasil penelitian yang dilakukan terhadap objek.



1.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut :

1. Penelitian ini dapat memberikan informasi tentang pengolahan teks analisis sentiment twitter dengan menggunakan Algoritma *k-nearest neighbour* dan *Naïve Bayes Classifier*.
2. Penelitian ini bisa berguna untuk mengetahui algoritma mana yang jauh lebih baik dalam analisis sentimen dengan objek tersebut.
3. Hasil Penelitian dapat dijadikan referensi dan evaluasi kinerja kedua Algoritma yaitu *k-nearest neighbour* dan *Naive Bayes classifier* dalam mengklasifikasi sentiment terhadap penggunaan aplikasi Gojek.
4. Bagi pengguna internet penelitian ini bisa dijadikan sebagai pertimbangan terhadap citra dari objek.

1.6 Metode Penelitian

Mengacu kepada penelitian yang dilakukan oleh peneliti sebelumnya berikut beberapa metode penelitiannya :

1.6.1 Metode Pengumpulan Data

a. Studi Litelatur

Studi litelatur dilakukan dengan mempelajari dan mengevaluasi peneltian – penelitian sejenis yang telah dilakukan sebelumnya. Penelitian tersebut juga dijadikan sebagai referensi terhadap penelitian ini, referensi didapatkan bukan hanya berdasarkan penelitian sebelumnya tetapi juga dapat berasal dari buku, jurnal, thesis, dan lainnya yang sesuai dengan penelitian ini.

b. Pengumpulan data Twitter

Mengumpulkan data tweet dengan keyword “by.u”, “byu_id” dan “byu.id” dengan memanfaatkan salah satu package dari python.

1.6.2 Metode Analisis

Dalam metode ini terdapat beberapa tahapan seperti dibawah ini:

1. Pengambilan Data

Tahap awal pengambilan data dilakukan dari social media twitter dengan menggunakan API twitter.

2. Data Pre-processing

Tahap selanjutnya yaitu pre-processing yaitu tahap ini merupakan awal untuk mempersiapkan data yang akan diolah selanjutnya.

3. Fitur Pembobotan

a. *Term Frequency (TF)* merupakan frekuensi kemunculan kata pada kalimat.

b. *Document Frequency (IDF)* Kemunculan berapa banyak kata dalam satu kalimat

4. Klasifikasi k-nearest neighbour dan Naïve Bayes Classifier

Tahap berikutnya yaitu klasifikasi sentimen menggunakan *k-nearest neighbour* dan *Naïve Bayes Classifier*.

5. Pengujian

Pengujian dilakukan dengan menggunakan model yang telah dibuat. Proses ini dilakukan untuk mengetahui nilai akurasi dan kecepatan proses dari kedua model yang sudah dibuat. Kemudian dari hasil yang sudah diperoleh akan dibandingkan untuk mengetahui mana yang

memiliki nilai akurasi maupun kecepatan yang lebih baik diantara keduanya.

1.7 Sistematika Penulisan

Untuk memudahkan dalam memahami isi dari karya ilmiah ini, maka penulis membuat sistematika penyusunan sebagai berikut :

BAB I PENDAHULUAN

Dalam bab ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, Batasan masalah, manfaat dan tujuan penelitian, metode penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan menguraikan lebih lanjut mengenai landasan teori serta penelitian sebelumnya yang digunakan sebagai referensi yang berkaitan dengan penelitian ini.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini akan membuat Analisa serta rancangan model yang akan dibuat.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini akan berisi hasil dan pembahasan dari penelitian yang telah dilakukan.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran untuk penelitian selanjutnya untuk pengembangan penelitian yang lebih baik.

