

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada beberapa tahun belakangan ini banyak bermunculan provider internet baru yang memikat banyak pengguna internet di Indonesia. Salah satu dari provider tersebut adalah by.u yang merupakan anak dari provider tertua di Indonesia yaitu Telkomsel. Setelah by.u muncul banyak pengguna internet tertarik karena dengan harga yang lebih murah dari Telkomsel namun kecepatan internet yang didapatkan sama, bahkan bisa jadi lebih cepat. Hal ini membuat banyak pengguna internet memberikan komentar mereka melalui sosial media, mulai dari komentar positif hingga negatif. Twitter menjadi salah satu media sosial yang digunakan oleh para pengguna internet untuk memberikan komentar dan review terhadap provider by.u. Dengan adanya fenomena tersebut, penelitian ini akan mengambil komentar dan postingan tweet dalam Twitter tersebut untuk digunakan dalam analisis sentimen terhadap opini kepuasan pelanggan pada provider by.u yang masih baru ini.

Analisis sentimen adalah metode pemrosesan data yang tidak terstruktur di ubah menjadi data terstruktur untuk mendapatkan sentiment [1]. Sentimen biasanya berkaitan dengan emosi, pendapat ataupun komentar yang dinyatakan dengan bentuk data teks. Dalam penelitian ini data text diambil dari komentar dan postingan Twitter dengan kata kunci "by.u". Data ini dapat digunakan sebagai acuan dalam melihat sentiment pengguna internet terhadap kepuasan pada provider baru by.u. Data tersebut akan diolah untuk mengklasifikasikan opini menjadi positif

atau negatif. Untuk melakukan klasifikasi data teks terdapat berbagai macam algoritma yang dapat digunakan.

Naive Bayes Classifier adalah salah satu algoritma yang di gunakan untuk klasifikasi data. Terdapat beberapa penelitian menggunakan metode *Naive Bayes Classifier* untuk dokumen text yang sudah pernah dilakukan di antaranya yaitu analisis sentimen perubahan ke kurikulum 2013 [2], penelitian kepuasan masyarakat terhadap pemilihan presiden 2019 [3], dan analisis sentimen terhadap jasa transportasi online [4]. Dari penelitian tersebut, peneliti memilih menggunakan algoritma *Naive Bayes* untuk melakukan klasifikasi opini positif dan negatif terhadap provider by.u ini. Metode *Naive Bayes Classifier* memiliki tingkat akurasi yang baik untuk klasifikasi data text [5].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka permasalahan yang akan dibahas pada penelitian ini, yaitu :

1. Bagaimana algoritma *Naive Bayes Classifier* dapat mengelompokan tweet yang mengandung sentiment terhadap provider by.u?
2. Berapakah tingkat akurasi dari analisis sentiment menggunakan algoritma *Naive Bayes Classifier*?
3. Apakah jumlah dataset berpengaruh terhadap tingkat akurasi algoritma *Naive Bayes Classifier*?

1.3 Batasan Masalah

Dalam penelitian ini ruang lingkup permasalahan akan dibatasi pada :

1. Menggunakan algoritma Naïve Bayes Classifier.
2. Tweet yang di ambil hanya dalam bahasa Indonesia.
3. Bahasa pemrograman yang digunakan dalam pembuatan sistem adalah Python.
4. Pengambilan data dari Twitter menggunakan Twitter API.
5. Data twitter yang digunakan sebanyak 6000 tweet.
6. Dataset dibagi menjadi 3 bagian untuk training dan testing berdasarkan jumlah tweet yaitu 1000, 2000, dan 3000 dataset.
7. Mempunyai 2 label klasifikasi yaitu sentiment positif dan sentiment negatif.

1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan diatas, tujuan di buatnya penelitian ini yaitu :

1. Mengetahui review dan respon pengguna internet terhadap provider by.u.
2. Mengetahui akurasi dari algoritma *Naïve Bayes Classifier* untuk melakukan klasifikasi *tweet* pada twitter terhadap provider by.u.
3. Mengetahui pengaruh jumlah dataset terhadap akurasi algoritma *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan bermanfaat untuk :

1. Membantu analisis sentiment pada twitter dengan Algoritma *Naive Bayes Classifier*.
2. Mendapatkan kesimpulan mengenai pengaruh jumlah dataset terhadap akurasi algoritma *Naive Bayes Classifier*.
3. Menambah referensi penelitian selanjutnya dan menambah pengetahuan terhadap Algoritma *Naive Bayes Classifier*.

1.6 Metodologi Penelitian

Metode penelitian yang di gunakan adalah :

1.6.1 Metode Pengumpulan Data

Pada penelitian ini metode pengumpulan data dilakukan melalui studi literatur dan pustaka, dengan cara mencari buku, tutorial, jurnal, dan skripsi yang memiliki keterkaitan sebagai rujukan serta memahami langkah-langkah crawling data dari media sosial twitter

1.6.2 Analisis Permasalahan

Untuk mendapatkan pemahaman dari algoritma yang akan digunakan, dilakukan analisa terhadap studi literatur dan pustaka algoritma *Naive Bayes Classifier* sebagai cara menyelesaikan masalah analisis sentiment.

1.6.3 Metode Testing

Pada tahapan ini akan dilakukan testing terhadap 3 dataset dengan jumlah yang berbeda. Testing akan dilakukan dengan tahapan pengumpulan data, pre-processing data, algoritma yang akan digunakan sebagai pengujian data adalah naïve bayes classifier, dan untuk menghitung hasil akurasi menggunakan confusion matrix.

1.7 Sistematika Penulisan

Urutan – urutan dan sistematika penulisan yang akan dilakukan dan ringkasan isi bab pada skripsi ini yaitu :

BAB I: PENDAHULUAN

Bab ini memuat latar belakang masalah, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II: LANDASAN TEORI

Bab ini memuat tentang teori-teori yang akan digunakan sebagai landasan utama dan penjelasan mengenai penelitian yang sudah dilakukan pada analisis sentiment, preprocessing teks, pembobotan kata, dan implementasi algoritma Naïve Bayes Classifier.

BAB III: METODE PENELITIAN

Bab ini memuat alur penelitian analisis sentiment yang akan dirancang untuk algoritma Naïve Bayes Classifier.

BAB IV: HASIL DAN PEMBAHASAN

Bab ini memuat hasil implementasi mengenai rancangan yang sudah direncanakan sebelumnya serta rangkuman hasil penelitian dari program analisis sentiment tersebut.

BAB V: PENUTUP

Bab ini berisi kesimpulan penelitian yang sudah dilakukan dan saran untuk pengembangan penelitian selanjutnya.

