

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial Twitter merupakan microblog yang memungkinkan penggunanya untuk menuliskan opini-opini tentang berbagai topik dan membahas isu-isu yang sedang terjadi. Pengguna Twitter, berdasarkan data PT Bakrie Telecom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global.[1] Twitter memberikan layanan mengirim dan membaca tweets yang berupa teks maksimal sebanyak 280 karakter pada penggunanya. Adanya media sosial Twitter membuat banyak orang menyampaikan pendapat/opininya melalui media sosial tersebut daripada menyampaikannya secara langsung. Sebanyak 19% pengguna media sosial Twitter memberikan berbagai opini mengenai merk dan produk (Jansen et al. 2009), serta ekspresi perasaan mereka terhadap operator seluler dengan akurasi memprediksi sentimen mencapai 80% (Wijaya et al. 2013).

Perusahaan operator seluler merupakan penyedia layanan telekomunikasi untuk pengguna telepon seluler. Pertumbuhan pengguna telekomunikasi setiap tahun terus bertambah banyak dengan cepat. Hal tersebut seiring dengan kebutuhan publik akan komunikasi yang terus meningkat melalui sms, telepon maupun data. Dengan demikian terjadi persaingan para provider telekomunikasi Indonesia untuk menarik atau mempertahankan pelanggannya. Berbagai opini yang dikemukakan oleh

pelanggan provider telekomunikasi khususnya Telkomsel dapat diketahui melalui media sosial Twitter. Namun, dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu pihak Telkomsel untuk mendukung suatu keputusan atau pilihan.

Text mining adalah sebuah proses pengambilan informasi dari data tekstual yang memiliki kualitas tinggi serta dapat mengetahui permasalahan dalam teks dari sebuah topik tertentu. Text mining dalam analisa sentimen mampu mengidentifikasi emosional terhadap suatu pernyataan (Zhang, et al, 2015). Analisis sentimen merupakan teknik yang digunakan untuk melakukan evaluasi dan mengidentifikasi emosi dan opini baik positif maupun negatif (Wilson et al. 2009). Dengan menggunakan analisis sentimen pengukuran kualitas layanan provider Telkomsel akan lebih cepat dan mudah.

1.2 Rumusan Masalah

Perumusan masalah dalam skripsi ini adalah:

1. Menentukan sentimen yang diberikan oleh pengguna provider Telkomsel melalui komentar di Twitter menggunakan Algoritma NBC (Naïve Bayes Classifier).
2. Mencari nilai akurasi dari Algoritma Naïve Bayes Classifier menggunakan klasifikasi sentimen pada komentar twitter terhadap pelayanan provider Telkomsel.

1.3 Batasan Masalah

Batasan masalah dalam pembuatan skripsi ini adalah:

1. Analisis sentimen dalam penelitian ini menggunakan data tweet yang diunduh menggunakan Twitter Scraper.
2. Selain mengunduh tweet kegiatan yang dilakukan yaitu tweet preprocessing, klasifikasi tweet, dan pengujian akurasi algoritma Naïve Bayes Classifier.
3. Fokus dari penelitian ini yaitu menerapkan algoritma Naïve Bayes Classifier untuk proses klasifikasi tweet.
4. Data sentimen diklasifikasikan kedalam 3 kelas yaitu: positif, negatif, dan netral.

1.4 Tujuan Penelitian

Tujuan dari pembuatan skripsi ini adalah:

1. Menemukan sentimen yang diberikan oleh pengguna provider Telkomsel dari komentar di Twitter menggunakan Algoritma Naïve Bayes Classifier.
2. Mencari nilai akurasi Algoritma Naïve Bayes Classifier dalam mengklasifikasi sentimen terhadap pelayanan provider Telkomsel.

1.5 Manfaat Penulisan

Manfaat yang diharapkan dari penulisan Skripsi ini antara lain:

- a. Bagi Perusahaan Telkomsel

Mendapatkan sentimen public dari pengolahan data twitter provider Telkomsel yang dapat dijadikan sebagai tolak ukur terhadap review pelayanan Telkomsel.

b. Bagi Ilmu Pengetahuan

Mengetahui akurasi Algoritma Naïve Bayes Classifier dan memperoleh informasi mengenai tentang bagaimana mengolah teks untuk mendapatkan suatu Sentiment Analysis dengan menggunakan Algoritma Naïve Bayes Classifier.

1.6 Metodologi Penelitian

Metodologi yang digunakan dalam pengerjaan tugas akhir ini adalah metode experimental dan dilaksanakan dalam tahapan-tahapan berikut:

1. Studi Literatur

Pengumpulan bahan referensi, seperti jurnal penelitian, tesis, buku-buku teori dan sumber lain termasuk informasi yang diperoleh dari internet. Mempelajari teori pendukung penelitian mengenai klasifikasi khususnya Naïve Bayes Classifier.

2. Analisis Sistem

Analisa terhadap syarat yang dibutuhkan dalam pembuatan dan pengembangan sistem.

3. Perancangan

Data dalam penelitian ini diambil menggunakan Twitter Scraper yang sudah disediakan, selanjutnya file tersebut disimpan dalam file

dokumen. Perancangan sistem meliputi training dan testing. Training sendiri terdapat beberapa tahap meliputi data preprocessing dan proses training sendiri. Pertama – tama, secara manual, data training akan dikelompokkan menjadi dalam 3 kelompok, positif, negative, dan netral. Lalu dilanjutkan ke tahap preprocessing. Dalam tahap preprocessing sendiri menggunakan library yang sudah ada. Ada beberapa hal yang dilakukan pada tahap preprocessing, yaitu:

- a. Clear invalid UTF 8
- b. Casefolding
- c. Remove URL
- d. Remove @username
- e. Remove #hashtag
- f. Remove Punctuation
- g. Remove Stopword
- h. Part Of Speech Tagging

Setelah tahap preprocessing data tweet yang sudah diberi label kemudian dihitung bobotnya. Bobot tersebut akan disimpan dalam berkas dan selanjutnya akan dilakukan proses pelatihan. Hasil dari pelatihan adalah Naïve Bayes model yang disimpan dalam berkas.

Pada proses testing akan menggunakan metode Naïve Bayes Classifier yang menggunakan library Sklearn. Dokumen yang berisikan tweet

akan dianalisa sentimennya melalui Naïve Bayes Classifier model yang telah dibuat. Setelah melalui tahap preprocessing data tweet akan diidentifikasi bobotnya dengan melakukan pembobotan Term Frequency terhadap tweet tersebut. Langkah terakhir yaitu ditentukan labelnya melalui klasifikasi yang dilakukan oleh model Naïve Bayes Classifier.

4. Implementasi

Pada tahap ini akan dilakukan pembuatan kode program sampai proses klasifikasi sentimen dokumen.

5. Pengujian

Tahap pengujian yaitu tahap perhitungan akurasi dari implementasi yang telah dilakukan sebelumnya. Metode yang digunakan untuk menguji yaitu K-Fold Cross Validation. Pengujian yang dilakukan adalah perbandingan akurasi klasifikasi yang dihasilkan Naïve Bayes Classifier.

1.7 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini memberikan arahan tentang masalah yang akan dibahas mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas tentang beberapa teori dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi pengertian Data Mining, Naïve Bayes Classifier, Sentimen Analysis, K-Fold Cross Validation, Flowchart, dan Confusion Matrix yang digunakan sebagai landasan dalam pembuatan skripsi.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas dan menjelaskan tentang pembuatan dan perancangan model cara kerja yang akan dilakukan dalam studi kasus sentiment analysis tweet Telkomsel.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini memaparkan proses dan hasil Sentiment Analysis Twitter mining dari Pre-Processing data, Tokenization, Steaming, TF-IDF, Naïve Bayes Classifier, KFold Cross Validation, dan Confusion Matrix.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari hasil penelitian yang telah dibuat dan saran yang berguna untuk penelitian selanjutnya.