

BAB I

1 PENDAHULUAN

2 Latar Belakang

Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, sekarang umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap.

Sebagian besar mahasiswa mengambil saran, ulasan atau pendapat dari opini orang lain sebelum mereka menentukan kampus mana yang sesuai dengan keinginan mereka. Media sosial adalah tempat dimana seseorang dapat mengemukakan sentimen, pendapat, ulasan tanpa ada yang membatasi opini mereka. Sehingga dampak dari media sosial berpengaruh baik untuk calon mahasiswa tanpa mereka sadari. Dengan demikian, ulasan dan opini yang tersebar di social media dapat dijadikan bahan sebagai sarana pengambilan keputusan atau pendapat untuk dilakukan evaluasi terhadap hal-hal yang menjadi dasar keresahan pada social media.

Dalam menghasilkan informasi dari data opini yang ada, dapat dilakukan pengolahan data dengan analisis sentimen untuk memperoleh solusi permasalahan yang telah dipaparkan sebelumnya, yaitu dengan memisahkan opini-opini kedalam kelas positif dan negatif. Sebelumnya telah dilakukan penelitian analisis sentimen

terhadap twitter dengan berbagai metode seperti Naïve Bayes Classification, Maximum Entropy, ataupun Support Vector Machine “Twitter Sentiment Classification using Distant”. Pengklasifikasi Naïve Bayes sangat sederhana dan efisien [1]. Selain itu, pengklasifikasi Naïve Bayes adalah teknik machine learning yang populer untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain. Namun, Naïve Bayes memiliki kekurangan yaitu sangat sensitif dalam pemilihan fitur. Terlalu banyak jumlah fitur, tidak hanya meningkatkan waktu penghitungan tapi juga menurunkan akurasi klasifikasi. [2]

Data yang diambil 3 universitas yaitu @UIIYogyakarta (Universitas Islam Yogyakarta), @klik_uad (Universitas Ahmad Dahlan) dan @UMYogya (Universitas Muhammadiyah Yogyakarta), memilih 3 universitas tersebut karena termasuk universitas swasta terkenal di Yogyakarta terdapat banyak fakultas. Universitas Islam Yogyakarta terdapat 8 fakultas, Universitas Ahmad Dahlan 11 fakultas, Universitas Muhammadiyah Yogyakarta terdapat 9 fakultas.

3 Rumusan Masalah

Rumusan masalah dalam skripsi ini adalah:

1. Bagaimana menentukan sentiment yang di berikan oleh mahasiswa dan komentar di Twitter menggunakan Algoritma Naïve Bayes Classification?
2. Bagaimana mencari nilai akurasi dari algoritma naïve Bayes Classification menggunakan dalam klasifikasi sentiment pada review terhadap proses pembelajaran di perguruan tinggi?

4 Batasan Masalah

Batasan masalah dalam pembuatan skripsi ini adalah:

1. Data sentiment mengenai perguruan tinggi diperoleh dari social media Twitter dengan hastag @UIIYogyakarta, @klik_uad dan @UMYogya.
2. Objek data Universitas Islam Indonesia, Universitas Ahmad Dahlan dan Universitas Muhammadiyah Yogyakarta.
3. Bahasa yang diambil dari Twitter adalah Bahasa Indonesia.
4. Analisis sentiment untuk data training dalam penelitian ini menggunakan data tweet dari Twitter yang di unduh selama lima kali dengan jarak waktu tertentu.
5. Data sentiment diklasifikasikan kedalam 3 kelas yaitu positif negative dan netral.
6. Bahasa pemrograman yang digunakan adalah python.
7. Menggunakan aplikasi python yang digunakan merupakan versi 3.7.4.
8. Training yang dilakukan terpisah dengan test predik suatu data.
9. Jumlah data training 500.

5 Tujuan penulisan

Tujuan dari pembuatan skripsi ini adalah:

1. Menentukan sentiment yang di berikan oleh mahasiswa dan orang tua dan komentar di Twitter menggunakan Algoritma Naïve Bayes Classification.
2. Mencari nilai akurasi dari algoritma naïve Bayes Classification menggunakan dalam klasifikasi sentiment pada review terhadap proses pembelajaran di perguruan tinggi.

6 Manfaat Penulisan

Manfaat yang diharapkan dari penulisan Skripsi ini antara lain:

a. Bagi universitas

Mendapatkan sentiment publik dari pengolahan data twitter yang menjadi tolak ukur terhadap review pelayanan pembelajaran di perguruan tinggi.

b. Bagi ilmu pengetahuan

1. Memperoleh informasi mengenai bagaimana tentang mengolah teks untuk mendapatkan suatu sentiment analysis dengan menggunakan Algoritma naïve bayes.
2. Mengetahui bagaimana implementasi text mining dalam pemrograman python.

7 Metode Penelitian

Mengacu pada penelitian sebelumnya pada tahapan sentiment analysis, berikut beberapa penjelasan tentang proses sentiment analysis:

8 Metode Pengumpulan data

9 Study Pustaka

Mencari literatur atau jurnal yang bersangkutan dengan text mining, sentiment analysis dan naïve bayes.

10 Survey

Mengamati kebutuhan koresponden dengan melakukan penyebaran suatu

pertanyaan yang nantinya menjadi dasar dalam pemberian suatu label pada data twitter yang akan dilatih.

11 Eksperimen

Menguji kode program agar mudah untuk dikembangkan dan mencari metode penulisan program yang tepat. Membuat user interface yang mudah dipahami dan memberikan banyak informasi.

12 Metode analisis

Dalam metode analisis ini terdapat beberapa tahapan seperti dibawah ini:

1. Data Colletion

Data colletion tahap awal ini adalah data yang didapatkan dari social media seperti twitter diambil @UIIYogyakarta, @klik_uad dan @UMYogya yang diundul 5 kali dengan rentang waktu 3 minggu dengan klasifikasi positif dan negatif.

2. Data Preprocessing

Tahap preprocessing adalah tahap awal dalam pengolahan data teks menjadi Analisa sentiment. Terdapat beberapa metode yang digunakan pada tahap preprocessing diantaranya adalah:

- a. Tokenization merupakan proses untuk memisahkan data teks menjadi beberapa token.

- b. Lower casing kebanyakan data ditwitter menggunakan huruf kapital dan huruf kecil. Proses ini menjadikan tulisannya menjadi huruf kecil semua.
- c. Cleaning merupakan proses pembersihan data teks yang tidak diperlukan yaitu dengan menghilangkan data yang tidak konsisten dan simbol.
- d. Stemming merupakan tahapan merubah bentuk kata menjadi bentuk dasarnya dengan menghilangkan kata imbuhan awalan dan akhiran. Contohnya “menyenangkan” menjadi kata dasar yaitu “senang”.

3. Klasifikasi naïve bayes

Tahapan selanjutnya yaitu klasifikasi menggunakan naïve bayes, pada tahapan ini akan melalui 2 tahap yaitu data training dan data testing.

Data training yang digunakan dalam pengklasifikasian teks ini terdiri dari 500 data komentar dengan klasifikasi positif dan negatif. Data tersebut masih berupa sekumpulan teks yang terpisah dalam bentuk komentar yang diunduh secara realtime.

4. Menghitung akurasi

Pada tahapan ini akan dilakukan perhitungan akurasi dari implementasi yang telah dilakukan. Metode yang digunakan yaitu confusion matrix.

13 *Metode Implementasi*

1. Implementasi Sistem di shell python menggunakan bahasa python.
2. Manual Program.

3. Hasil Implementasi.

14 Metode testing

Metode yang digunakan untuk pengujian aplikasi menggunakan blackbox testing.

15 Sistematika Penulisan

Sistematika penulisan berdasarkan pokok permasalahan agar lebih terarah dapat diurutkan sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang latar belakang, perumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan teori yang berupa pengertian dan definisi yang diambil dari kutipan buku yang berkaitan dengan penyusunan laporan skripsi serta beberapa literature review yang berhubungan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan data yang telah dikumpulkan dan langkah-langkah investigasi data yang telah didapatkan.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini menjelaskan proses dan hasil sentiment analysis twitter mining dari preprocessing data, tokenization, stemming, TF-IDF, naïve bayes, dan confusion Matrix.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa dan optimalisasi sistem berdasarkan yang telah diuraikan pada bab-bab sebelumnya.

