

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini ojek online telah menjadi sarana prasarana yang memudahkan masyarakat dalam beraktifitas. Salah satu startup di bidang jasa transportasi ini adalah Grab. Grab merupakan suatu perusahaan di bidang jasa yang bertujuan untuk mempermudah masyarakat dalam melakukan pemesanan ojek maupun pengiriman barang hanya dengan menggunakan aplikasi berbasis mobile di smartphone. Grab tidak hanya menyediakan jasa transportasi berupa ojek online melainkan Grab juga menyediakan jasa transportasi seperti mobil dan taxi online. Peran konsumen Grab dalam memberikan opini berupa kritik dan saran dapat berguna sebagai feedback untuk Perusahaan Grab agar bisa lebih baik dalam memberikan pelayanannya.

Sebagian besar konsumen mengambil saran, ulasan atau pendapat dari opini orang lain sebelum mereka memutuskan untuk membeli barang atau menggunakan suatu jasa tertentu. Media sosial adalah tempat dimana seseorang dapat mengemukakan sentimen, pendapat, ulasan tanpa ada yang membatasi opini mereka. Dampak media sosial ini sangat berpengaruh baik untuk konsumen maupun produsen tanpa mereka sadari.

. Dengan demikian, dari ulasan-ulasan tersebut dapat memberikan nilai potensial yang dapat diimplementasikan untuk analisis sentimen yang bertujuan guna meningkatkan strategi dalam pemasaran, peningkatan produk/ pelayanan untuk menjadi lebih baik. Salah satu contoh media sosial tempat bertukar opini

adalah Twitter. Dari twitter dapat dilakukan Opinion Mining atau biasa disebut Analisis Sentimen. Untuk melakukan Twitter Mining peneliti akan mengambil data yang digunakan dengan memanfaatkan API Twitter.

Analisis Sentimen merupakan suatu proses pengolahan data berupa teks yang bertujuan untuk mengklasifikasi sentimen positif. Dalam proses pengolahan data sentimen akan melewati beberapa tahapan yaitu Pengumpulan Data, Cleaning Data, Tokenization, Stopword Removal, Stemming, Feature List, Pembobotan kemudian setelah itu di klasifikasi menggunakan algoritma Support Vector Machine. SVM merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning (supervised learning) yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan oleh Vladimir Vapnik. Klasifikasi dilakukan dengan garis pembatas (hyperlane) yang memisahkan antara kelas opini positif dan opini negatif [1]. Dalam penelitian yang berjudul "Sentimen Analysis Using SVM and Naive Bayes Algorithm" yang dilakukan oleh Preety dan Sunny Dahiya [2] dapat disimpulkan bahwa metode ini adalah metode yang paling baik dibandingkan dengan metode lainnya. Dengan menggunakan metode Support Vector Machine, tingkat keakuratan akan mencapai 83,59 %. Oleh karena itu maka dalam penelitian ini akan menggunakan metode Support Vector Machine untuk mengukur nilai akurasi sentimen yang diperoleh dari hasil crawling data twitter Grab.

1.2 Rumusan Masalah

Perumusan masalah dalam skripsi ini adalah:

1. Menemukan sentimen yang diberikan oleh pengguna Grab dari komentar di twitter menggunakan Algoritma Support Vector Machine.
2. Mencari nilai akurasi algoritma Support Vector Machine pada klasifikasi sentimen terhadap pelayanan Grab.

1.3 Batasan Masalah

Batasan masalah dalam pembuatan Skripsi ini adalah:

1. Penelitian hanya dilakukan terhadap data dari twitter dengan hastag yang berhubungan dengan Grab seperti '*Grabbike*', '*GrabIndonesia*', '*GrabID*', '*Grab*'.
2. Bahasa yang diambil dari Twitter adalah Bahasa Indonesia.
3. Analisis sentimen dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh selama tiga kali dalam rentang waktu 2 minggu.
4. Analisa sentimen dengan metode SVM menggunakan tools yang sudah tersedia di library python.
5. Data sentimen diklasifikasikan kedalam 2 kelas yaitu positif, negatif.
6. Bahasa pemrograman yang digunakan adalah Python.
7. Python yang digunakan merupakan python versi 3.5.
8. Training data dilakukan secara terpisah dengan test predik suatu data.
9. Jumlah data training 700 data.

10. Analisis sentimen untuk data pengujian data testing dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh secara *realtime*.

1.4 Tujuan Penulisan

Tujuan dari pembuatan Skripsi ini adalah:

1. Menerapkan metode SVM pada analisis sentimen terhadap data twitter Grab.
2. Menemukan sentimen yang berhubungan dengan Grab dari komentar di twitter menggunakan algoritma Support Vector Machine.
3. Mencari nilai akurasi Algoritma Support Vector Machine dalam mengklasifikasi sentimen terhadap pelayanan Grab.

1.5 Manfaat Penulisan

a. Bagi Perusahaan Grab

Mendapatkan presentase sentimen publik dari data twitter Grab terhadap pelayanan dari Grab guna meningkatkan pelayanan dan mutu dari Grab.

b. Bagi Ilmu Pengetahuan

1. Mengetahui bagaimana implementasi opinion mining dalam pemograman python
2. Mengetahui cara memperoleh, mengolah data untuk sentimen analisis

1.6 Metode Penelitian

Dalam penelitian ini akan menjelaskan tahapan – tahapan sentimen analisis sebagai berikut.

1.6.1. Metode Pengumpulan Data

1.6.1.1. Studi Pustaka

Mencari literatur atau jurnal yang bersangkutan dengan Analisis Sentimen/Opinion Mining, Algoritma Support Vector Machine, K-Fold Cross Validaton , .

1.6.1.2. Survei

Mengamati kebutuhan koresponden dengan melakukan penyebaran suatu pernyataan yang nantinya menjadi dasar dalam pemberian suatu label pada data Twitter yang akan dilatih.

1.6.1.3. Eksperimen

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

1.6.2. Metode Analisis

Dalam metode analisis ini terdapat beberapa tahapan seperti dibawah ini:

1. Collecting Data

Data yang dikumpulkan diperoleh dari crawling data twitter menggunakan Twitter API selama tiga kali dalam rentang waktu 2 minggu.

2. Data Preprocessing

Tahap *preprocessing* adalah tahap awal pengolahan data dalam analisis sentimen. Berikut tahap - tahapannya:

- a. Cleaning Data merupakan tahapan penghapusan kata – kata yang tidak dipakai seperti hastag dan username.

- b. Case folding atau Lower Case digunakan untuk membuat keseluruhan data tweet menjadi lowercase (huruf kecil).
 - c. Tokenization merupakan tahap untuk memisahkan kalimat menjadi kata dan menghilangkan data yg tidak bermakna.
 - d. Stemming merupakan tahapan untuk menghilangkan imbuhan awal atau akhir pada kata tersebut seperti memasak menjadi masak.
3. Feature Weighting (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 dengan feature Cosine Similarity
- Feature Cosine Similarity adalah fitur untuk mencari tingkat kesamaan data untuk meningkatkan akurasi data.
5. Klasifikasi Support Vector Machine
- Pada tahap ini data tersebut diklasifikasi menggunakan Algoritma SVM dengan menentukan suatu kelas berdasarkan pola yang terbentuk dari hasil proses training. Disini data diklasifikasi menjadi 2 tahap yaitu dengan tahap training berjumlah 700 data yg kemudian akan diberikan label secara manual dan untuk data pengujian yang diunduh secara realtime.
6. Menghitung akurasi dari pengujian data

Untuk menguji seberapa akurat penelitian yang dilakukan maka akan dilakukan pengujian data dengan 10-Fold sebagai parameternya untuk melakukan Cross Validation yang diuji menggunakan tools pada python.

1.6.3. Metode Perancangan

1. Perancangan Aplikasi menggunakan flowchart
2. Perancangan Antarmuka Pengguna

1.6.4. Metode Implementasi

1. Implementasi Sistem di shell python dengan menggunakan bahasa Python.
2. Manual Program
3. Hasil Implementasi

1.6.5. Metode Testing

K-fold merupakan metode yang digunakan untuk mengetahui rata-rata keberhasilan suatu sistem.

1.7 Sistematika Penulisan

Laporan penelitian ini akan disusun secara sistematis yang akan diurutkan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini akan menjelaskan mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan, manfaat dan sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Pada bab ini menjelaskan tentang mengenai teori yang digunakan sebagai dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi *Pengertian Data mining, Support Vector Machine(SVM), Sentiment Analysis, Pre-Processing, Term-Frequency-Inverse Document Frequency, Cosine Similarity, K-Fold Cross Validation* dan *Confusion Matrix*.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini akan menjelaskan mengenai perancangan dan model cara kerja yang akan dilakukan di dalam studi kasus sentimen analysis tweet Grab.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini akan menjelaskan *Analisis Sentimen* dari *Pre-Processing* data, *Tokenization*, *Stemming*, *TF-IDF*, *SVM*, dan *Confusion Matrix*.

BAB V PENUTUP

Bab ini akan menjelaskan kesimpulan yang didapat dari hasil analisis, perancangan dan website serta saran untuk pengembangan lebih baik

