

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Gojek merupakan sebuah perusahaan teknologi asal Indonesia yang memanfaatkan jasa para ojek dengan menggunakan aplikasi berbasis mobile untuk mempermudah masyarakat yang membutuhkan jasa ojek. Untuk meningkatkan pelayanan baik dari segi pelanggan dan non pelanggan maka gojek perlu melakukan peningkatan dari kegunaan aplikasi itu sendiri. Salah satu cara untuk mendapatkan informasi dari pelanggan maka gojek harus mempunyai wadah untuk para konsumennya salah satu sumber data yang bisa digunakan sebagai acuan adalah sosial media.

Rutinitas yang sering dilakukan masyarakat Indonesia adalah membagikan segala aktivitas di sosial media baik itu berupa keluhan, masukan atau bahkan sekedar berkomentar. Dimana dengan meningkatkannya aktivitas ini akan memberikan suatu kontribusi untuk Go-jek itu sendiri, mulai dari nilai positif dan negatif, pandangan masyarakat bahkan meningkatkan mutu dan kualitas.

Dari komentar-komentar yang tersebar di media sosial dapat dijadikan bahan untuk sebagai sarana pengambilan keputusan atau pendapat untuk dilakukan evaluasi atau perbaikan terhadap hal - hal yang menjadi dasar keresahan pengguna pada sosial media.

Untuk 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 ke dalam 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". Didapatkan kesimpulan bahwa penggunaan metode Support Vector Machine (SVM) memberikan hasil paling baik dibandingkan metode lainnya, yaitu dengan tingkat keakuratan hingga 82,2%. [1] Metode SVM sendiri dapat digunakan untuk mengklasifikasikan data opini berdasarkan atribut penilaian yang dimiliki untuk dapat dipisahkan apakah opini termasuk ke dalam kelas bersentimen positif atau negatif. Dengan keunggulan SVM yang dapat di implementasikan seperti generalisasi, stabilitasnya, klasifikasi dan dapat mengolah data linier maupun non linier, ini membuat SVM masuk ke dalam metode klasifikasi yang baik. Dari hasil tersebut maka dalam penelitian ini dibuatlah sebuah analisis sentimen menggunakan algoritma SVM terhadap review pelayanan Go-Jek. Dengan dilakukannya analisis sentimen ini diharapkan akan bisa memberikan manfaat kepada instansi terkait dalam hal ini adalah perusahaan Go-Jek.

## **1.2 Rumusan Masalah**

Perumusan masalah dalam skripsi ini adalah:

1. Menemukan sentimen yang diberikan oleh pengguna Go-Jek dan komentar di Twitter menggunakan Algoritma *Support Vector Machine*.
2. Mencari nilai akurasi dari algoritma *Support Vector Machine* menggunakan dalam klasifikasi sentimen pada review terhadap pelayanan Go-Jek.

### 1.3 Batasan Masalah

Batasan masalah dalam pembuatan Skripsi ini adalah:

1. Data sentimen mengenai Go-Jek diperoleh dari sosial media Twitter dengan hastag @gojekindonesia dan @gojek.
2. Bahasa yang diambil dari twitter adalah Bahasa Indonesia.
3. Analisis sentimen untuk data training dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh selama lima kali dengan jarak waktu tiga minggu.
4. Analisis sentimen untuk data pengujian dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh secara *realtime*.
5. Analisis sentimen dengan metode SVM menggunakan tools yang sudah tersedia.
6. Data sentimen diklasifikasikan kedalam 2 kelas yaitu positif dan negatif.
7. Bahasa pemrograman yang digunakan adalah Python.
8. Python yang digunakan merupakan python versi 3.5.
9. Training data dilakukan secara terpisah dengan test predik suatu data.
10. Jumlah data training 800.

### 1.4 Tujuan Penulisan

Tujuan dari pembuatan Skripsi ini adalah:

1. Menemukan sentimen yang diberikan oleh pengguna Go-Jek dari komentar di twitter menggunakan algoritma Support Vector Machine.

2. Mencari nilai akurasi Algoritma Support Vector Machine dalam mengklasifikasi sentimen terhadap pelayanan Go-Jek.

### 1.5 Manfaat Penulisan

Manfaat yang diharapkan dari penulisan Skripsi ini antara lain:

- a. Bagi Perusahaan Go-Jek

Mendapatkan sentimen publik dari pengolahan data twitter Go-jek yang menjadi tolak ukur terhadap review pelayanan Go-jek.

- b. Bagi Ilmu Pengetahuan

1. Memperoleh informasi mengenai bagaimana tentang mengolah teks untuk mendapatkan suatu *Sentiment Analysis* dengan menggunakan *Support Vector Machine (SVM)*.

2. Mengetahui bagaimana implementasi *text mining* dalam pemrograman Python.

### 1.6 Metode Penelitian

Mengacu pada penelitian sebelumnya pada tahapan sentimen analisis, berikut beberapa penjelasan tentang proses sentimen analisis:

#### 1.6.1. Metode Pengumpulan Data

##### 1.6.1.1. Studi Pustaka

Mencari literatur atau jurnal yang bersangkutan dengan Text Mining, Sentimen Analisis dan Support Vector Machine.

##### 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. Data Collection

*Data collection* tahap awal ini adalah data yang didapatkan dari sosial media seperti twitter lalu dari twitter diambil @Gojekindonesia dan @gojek yang diunduh selama 5 kali dengan rentang waktu tiga minggu dengan klasifikasi komentar positif dan negatif.

#### 2. Data Preprocessing

Tahap *preprocessing* adalah tahap awal dalam pengolahan data teks menjadi analisa sentimen. 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 gabungan 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 “menyebalkan” dirubah mejadi kata dasarnya yaitu “sebal”.

### 3. Feature Weighting

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

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

### 4. Klasifikasi Support Vector Machine

Tahapan selanjutnya yaitu klasifikasi menggunakan *Support Vector Machine*, pada tahapan ini akan melalui dua tahap lagi yaitu data *training* dan data *testing*. Data training berjumlah 800 data komentar dengan klasifikasi positif dan negatif yang diberikan label secara manual dan data pengujian yang diunduh secara *realtime*.

### 5. Menghitung akurasi

Pada tahapan ini akan dilakukan perhitungan akurasi dari implementasi yang telah dilakukan. Metode yang digunakan yaitu *K-Fold Cross Validation*. Dalam penelitian ini, tahap pengujian menggunakan 10-Fold sebagai parameter untuk melakukan Cross Validation. Pengujian yang dilakukan menggunakan *tools* pada Python untuk menguji seberapa akurat algoritma *Support Vector Machine* pada dataset penelitian ini.

#### 1.6.3. Metode Perancangan

1. Perancangan Aplikasi menggunakan flowchart.
2. Perancangan Antarmuka Pengguna.

#### 1.6.4. Metode Implementasi

1. Implementasi Sistem
2. Manual Program
3. Hasil Implementasi

#### 1.6.5. Metode Testing

Metode yang digunakan untuk mengetahui rata-rata keberhasilan suatu sistem menggunakan *K-Fold*.

#### 1.7 Sistematika Penulisan

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

##### 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*, *Support Vector Machine(SVM)*, *Sentiment Analysis*, *Pre-Processing*, *Term-Frequency-Inverse Document Frequency*, *Consine Similarity*, *K-Fold Cross Validation*, *UML* dan *Confusion Matrix* yang digunakan sebagai landasan dalam pembuatan skripsi.

##### BAB III ANALISIS DAN PERANCANGAN

Pada bab ini akan menjelaskan tentang pembuatan dan perancangan model cara kerja yang akan dilakukan didalam studi kasus sentiment analysis tweet Go-Jek.

## BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini memaparkan proses dan hasil *Sentiment Analysis* Twitter mining dari *Pre-Processing* data, *Tokenization*, *Steaming*, *TF-IDF*, *SVM*, dan *Confusion Matrix*.

## BAB V PENUTUP

Bab ini berisi kesimpulan yang didapat dari hasil analisis, perancangan dan implementasi website serta saran untuk pengembangan lebih baik.

