

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Memasuki era *e-commerce* 4.0 di Indonesia yang ditandai dengan integrasi antara bisnis *online* dan bisnis *offline*. Ditambah menjamurnya *e-commerce* berbentuk situs jual beli *online* di pasar sudah menjadi tanda-tanda dimulainya evolusi era ini. Banyaknya situs jual beli online akan menimbulkan persaingan yang ketat untuk menarik pengguna layanannya. Salah satu marketplace yang sudah beredar di Indonesia adalah Blibli.com.

Blibli.com merupakan salah satu situs jual beli online yang memanfaatkan website dan aplikasi berbasis mobile untuk mempermudah pengguna yang membutuhkan barang ataupun kebutuhan lainnya. Untuk memberikan pelayanan yang baik Blibli.com menyediakan layanan pelanggan yang dapat diakses selama 24 jam.

Salah satu cara untuk memberikan pelayanan kepada pelanggan blibli.com mempunyai wadah untuk para pengguna salah satunya menggunakan akun sosial media. Selain dapat memberikan pelayanan kepada pelanggan dengan akun sosial media, Blibli.com dapat menggunakan data yang tersedia untuk diolah menjadi informasi untuk meningkatkan pelayanan.

Rutinitas yang bisa dilakukan oleh masyarakat Indonesia biasanya membagikan segala aktivitas di sosial media baik berupa masukan, komentar atau sekedar berkomentar. Dengan adanya peningkatan aktivitas ini dapat menghasilkan suatu pola data yang dapat digunakan untuk memberikan suatu kontribusi untuk Blibli.com itu sendiri, mulai dari nilai positif dan negatif

bahkan mengukur tingkat kualitas pelayanan.

Untuk mendapatkan informasi dari pola tersebut dibutuhkan teknik pengolahan data analisis sentimen untuk memperoleh solusi permasalahan yang telah dipaparkan sebelumnya, dengan cara memisahkan komentar-komentar yang ada ke dalam kelas positif dan negatif. Terdapat beberapa metode yang biasa digunakan untuk melakukan analisis sentiment terhadap twitter salah satunya *Support Vector Machine(SVM)*.

Metode SVM adalah metode klasifikasi yang populer dan sering digunakan serta memiliki akurasi yang baik dalam klasifikasi analisis teks[1]. Dimana metode *Support Vector Machine* dapat diterapkan untuk melakukan klasifikasi data *multiclass*[2].Sebelumnya telah dilakukan penelitian sentimen analisis terhadap twitter dengan berbagai metode seperti SVM, *Naïve Bayes (NB)*, *Decision Tree*, *K-nearest neighbor (kNN)*, dan *Rocchio classifier* "Comparison of Text Classifiers on News Articles". Didapatkan kesimpulan bahwa metode *Support Vector Machine* menghasilkan tingkat akurasi sebesar 86,7% dengan waktu *running time* yang lebih cepat dibandingkan metode lainnya[3]. Kemudian penelitian oleh Elly Indrayuni mengenai perbandingan metode *Naïve Bayes (NB)* dengan SVM menghasilkan nilai akurasi sebesar 90% untuk metode SVM[4]. Dengan keunggulan SVM dapat mengolah data linear dan non-linear, maka membuat metode ini sangat baik untuk klasifikasi. Dari hasil tersebut maka dalam penelitian ini dibuatlah analisis sentimen menggunakan algoritma SVM terhadap review pelayanan Blibli.com. Diharapkan dengan adanya penelitian ini dapat memberikan manfaat kepada perusahaan terkait dalam hal ini adalah Blibli.

## 1.2 Rumusan Masalah

Perumusan masalah dalam penelitian ini adalah:

1. Menemukan sentimen yang diberikan pengguna Blibli.com serta komentar di Twitter menggunakan algoritma SVM.
2. Mencari nilai akurasi dari algoritma SVM menggunakan klasifikasi sentimen pada *review* terhadap pelayanan Blibli.com.

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data sentimen mengenai Blibli.com dan diperoleh dari sosial media Twitter dengan hastag @BlibliCare
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 dalam rentang bulan agustus 2020 – januari 2021.
4. Analisis sentimen untuk data pengujian dalam penelitian ini menggunakan data *tweet* dari Twitter yang diunduh secara langsung.
5. Analisis sentimen menggunakan metode SVM dengan tools yang sudah tersedia.
6. Data sentimen diklasifikasikan kedalam 2 kelas yaitu positif dan negatif.
7. Training data dilakukan secara terpisah dengan test predik suatu data.
8. Data training yang digunakan berjumlah 800.
9. Bahasa pemrograman yang digunakan adalah Python.

#### 1.4 Maksud dan Tujuan Penelitian

Tujuan dalam penelitian ini adalah:

1. Menemukan sentimen yang diberikan oleh pengguna Blibli.com dari tweet di twitter menggunakan algoritma SVM.
2. Mencari nilai akurasi Algoritma SVM dalam mengklasifikasikan sentimen pelayanan Blibli.com.

#### 1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penulisan penelitian ini antara lain:

- a. Bagi Perusahaan Blibli.com
  - Mendapatkan sentimen publik dari pengolahan data twitter Blibli.com yang menjadi tolak ukur terhadap pelayanan Blibli.com.
- b. Bagi Ilmu Pengetahuan
  - Memperoleh informasi mengenai bagaimana mengolah teks untuk mendapatkan suatu *Sentiment Analysis* dengan menggunakan algoritma *Support Vector Machine (SVM)*.

#### 1.6 Metode Penelitian

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

##### 1.6.1 Metode Pengumpulan Data

###### 1.6.1.1 Studi Pustaka

Mencari jurnal atau literatur yang bersangkutan dengan *text mining*, *sentiment analysis*, 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 memuat informasi yang banyak.

### 1.6.2 Metode Analisis

Dalam metode analisis ini terdapat beberapa tahapan seperti dibawah ini:

#### 1. Data Collection

*Data collection* pada tahap awal ini adalah data yang didapatkan dari sosial media twitter yang diambil melalui hastag @BibliCare yang diunduh selama rentang waktu enam bulan dengan klasifikasi positif dan negatif.

#### 2. Data Preprocessing

Tahapan *preprocessing* adalah tahap awal dalam pengolahan data teks menjadi analisis sentimen. Terdapat beberapa metode yang digunakan pada tahapan preprocessing di antaranya adalah:

- a. *Tokenization* merupakan proses untuk memecah data teks menjadi beberapa token.
- b. *Lower Casing* merupakan proses untuk menjadikan tulisan menjadi huruf kecil semua.
- c. *Cleaning* merupakan proses pembersihan data teks yang tidak diperlukan dengan menghilangkan data tidak konsisten dan

simbol.

- d. *Stemming* merupakan tahapan merubah bentuk kata menjadi bentuk dasarnya dengan menghilangkan kata imbuhan awalan dan akhiran. Contohnya “keadaan” diubah menjadi kata dasarnya yaitu “ada”.

### 3. Feature Weighting

- a. *Term Frequency (TF)* merupakan frekuensi kemunculan kata pada kalimat.
- b. *Document Frequency (IDF)* merupakan kemunculan berapa banyak kata dalam satu kalimat.

### 4. Klasifikasi Support Vector Machine

Pada tahapan ini akan melalui dua tahap lagi yaitu data *training* dan data *testing*. Data *training* menggunakan 800 data komentar dengan sentimen positif dan negatif yang diberikan label secara manual. dan data pengujian menggunakan data yang diunduh secara *realtime*. Pada data *training* data akan dibagi menjadi 400 data sentimen positif dan 400 data sentimen negatif. Pembagian data dimaksudkan agar hasil akurasi klasifikasi lebih optimal [5].

### 5. Menghitung Akurasi

Pada tahapan ini dilakukan perhitungan akurasi dari implementasi yang sudah dilakukan. Menggunakan metode *K-Fold Cross Validation*. Dalam penelitian ini, tahap pengujian menggunakan 10-Fold sebagai parameter untuk melakukan *Cross Validation*. Pengujian ini 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 keberhasilan suatu system 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**

Pada bab ini akan membahas mengenai beberapa teori dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi pengertian *Data mining*, *Support Vector Machine (SVM)*, *Sentiment Analysis*, *Preprocessing*, *Term-Frequency-Inverse Document Frequency*, *Cosine Similarity*, *K-Fold Cross*

*Validation*, *UML*, dan *Confusion Matrix* yang digunakan untuk landasan teori dalam pembuatan skripsi.

### BAB III ANALISIS DAN PERANCANGAN

Bab ini akan menjelaskan tentang pembuatan dan perancangan model cara kerja yang dilakukan dalam studi kasus sentimen analisis tweet Blibli.com.

### BAB IV IMPLEMENTASI DAN PEMBAHASAN

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

### BAB V PENUTUP

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

