

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada penelitian ini penulis memilih Tokopedia sebagai objek penelitian. Menurut artikel yang dilansir oleh *kumparan.com*, Tokopedia masih menempati peringkat pertama pada kuartal ketiga tahun ini sebagai *e-commerce* yang paling banyak diakses sejak kuartal ketiga tahun 2018. Data menunjukkan sebanyak 65.953.400 pengunjung telah mengakses situs jual beli ini setiap bulannya. Namun angka tersebut mengalami penurunan jika dibandingkan dengan tahun lalu ([www.kumparan.com](http://www.kumparan.com),2019). Demi menjaga performa dan meningkatkan pelayanan, dibutuhkan masukan atau opini dari pengguna untuk mengetahui kelemahan yang dapat dijadikan bahan evaluasi dan peningkatan pelayanan. Calon pengguna *e-commerce* juga butuh bahan pertimbangan untuk memilih mana aplikasi yang dirasa paling tepat mendukung keberlangsungan usahanya.

Sosial media seringkali menjadi tempat masyarakat dalam menyampaikan opini. Salah satu sosial media yang paling populer untuk berekspresi adalah *Twitter* yang menjadi tempat untuk berinteraksi antara pengguna dengan pihak startup itu sendiri. Melalui *Twitter* pengguna dapat menyampaikan opini atau sentimen hingga 280 karakter dengan jumlah pengguna aktif mencapai 126 juta setiap harinya ([www.tekno.kompas.com](http://www.tekno.kompas.com),2019). *Twitter* menyediakan data yang dapat diakses melalui *twitter API* sehingga dapat mempermudah dalam pengumpulan data.

Data sentimen nantinya akan diolah dengan proses klasifikasi yang merupakan cabang dari *Text Mining*. Tujuan utama dari klasifikasi sentimen adalah

untuk mencari tahu polaritas dari sentimen positif, negatif dan netral [1]. Dalam proses pengolahan data sentimen ada beberapa tahap yang harus dilakukan seperti Pengumpulan data, Preprocessing data, Pembobotan kata kemudian melakukan proses klasifikasi menggunakan algoritma yang telah ditentukan.

Penulis merasa perlu melakukan penelitian ini dengan membuat sebuah sistem untuk menganalisa suatu opini menggunakan algoritma *Support Vector Machine* karena mempunyai kelebihan untuk menemukan hyperplane terbaik yang memisahkan dua buah class. Dalam penelitian yang berjudul "Analisis Sentimen *Cyberbullying* pada Komentar Instagram dengan Metode Klasifikasi *Support Vector Machine*" tingkat akurasi dari hasil klasifikasi algoritma SVM mampu mencapai 90% [2]. Oleh karena itu penulis memilih algoritma *Support Vector Machine* untuk menghasilkan informasi mengenai analisis sentimen masyarakat terhadap *Tokopedia*.

## 1.2 Rumusan Masalah

1. Apakah algoritma *Support Vector Machine* dapat digunakan untuk pengklasifikasian analisis sentimen *twitter* terhadap *Tokopedia* ?
2. Berapa nilai akurasi penggunaan algoritma *Support Vector Machine* dalam pengklasifikasian analisis sentimen *twitter* terhadap *Tokopedia* ?

## 1.3 Batasan Masalah

1. Data yang digunakan berasal dari *tweet* yang menyebutkan kata kunci TOKOPEDIA.
2. *Tweet* hanya berupa teks yang berbahasa Indonesia.

3. Metode yang digunakan untuk klasifikasi adalah *Support Vector Machine* dan tidak membandingkannya dengan metode lain.
4. Proses pemberian label data training dilakukan secara manual.
5. Sistem ini belum terdapat *stopwordremoval* dalam preprocessing data.

#### **1.4 Maksud dan Tujuan Penelitian**

1. Melakukan pembuktian proses pengklasifikasian algoritma *Support Vector Machine* dalam analisis sentimen *twitter* terhadap *Tokopedia*.
2. Mengetahui besarnya akurasi dari penggunaan algoritma *Support Vector Machine* dalam melakukan analisis sentimen *twitter* terhadap *Tokopedia*.

#### **1.5 Manfaat Penelitian**

1. Menjadi bahan pertimbangan dalam menentukan *marketplace* yang paling sesuai dengan pengguna.
2. Mendapatkan informasi tentang kualitas suatu *marketplace*, dimana informasi tersebut dapat dijadikan bahan evaluasi oleh startup untuk meningkatkan pelayanan.

#### **1.6 Metode Penelitian**

Beberapa tahapan yang dilakukan dalam penulisan skripsi ini adalah sebagai berikut :

##### **1.6.1 Metode Pengumpulan Data**

###### **1.6.1.1 Studi literatur**

Studi literatur dilakukan dengan cara mengumpulkan bahan referensi berupa jurnal, paper, makalah dan situs – situs dari internet yang berkaitan dengan algoritma *Support Vector Machine* serta metode *TF-IDF*.

#### 1.6.1.2 Pengumpulan Data Twitter

Mengumpulkan data yang diambil secara langsung dari akun *Tokopedia* dengan memanfaatkan salah satu package dari python.

#### 1.6.2 Metode Analisis

Mengimplementasikan beberapa tahapan analisis seperti dibawah ini :

##### 1. Collecting data

Mengumpulkan data yang dihasilkan dari proses *crawling* akun twitter Tokopedia.

##### 2. Preprocessing data

Ada beberapa tahap dalam melakukan *preprocessing data* seperti berikut :

##### a. Cleansing data

Proses menghilangkan bagian dari tweet yang tidak dibutuhkan dalam proses analisis sentimen.

##### b. Case Folding

Digunakan untuk membuat data tweet menjadi lowercase (huruf kecil).

##### c. Tokenization

Tahap untuk merubah kalimat menjadi sekumpulan kata atau token dan menghilangkan kata yang tidak mempengaruhi hasil sentimen.

d. Stemming

Merupakan tahap untuk mengembalikan kata yang berimbuhan menjadi kata dasar.

3. Pembobotan kata

Perhitungan bobot kata yang akan mempengaruhi hasil sentimen menggunakan fitur TF-IDF.

4. Klasifikasi Cosine Similarity

Proses ini bertujuan untuk menghitung tingkat kemiripan antar dokumen.

5. Klasifikasi Algoritma Support Vector Machine

Tahap mengimplementasikan algoritma SVM dengan melakukan klasifikasi pada data training secara manual dan data uji yang diunduh secara realtime.

6. Tahap Evaluasi

Merupakan tahap untuk menguji tingkat akurasi menggunakan *k-Fold Cross Validation* yang berjumlah  $k=10$ .

### 1.6.3 Metode Perancangan

Pada tahap ini dilakukan perancangan sistem terhadap permasalahan yang ada sesuai dengan batasan masalah.



#### 1.6.4 Implementasi

Melakukan proses implementasi pengkodean program dalam aplikasi komputer menggunakan bahasa pemrograman yang telah dipilih yang sesuai dengan analisis dan perancangan yang sudah dilakukan.

#### 1.6.5 Pengujian sistem

Pada tahap ini melakukan proses pengujian terhadap sistem menggunakan *Algoritma Support Vector Machine* dan hasilnya akan dievaluasi menggunakan *K-fold*

#### 1.7 Sistematika Penulisan

Penulisan skripsi ini terdiri dari 5 bab dengan penjelasan masing – masing bab sebagai berikut :

##### **Bab 1 : Pendahuluan**

Pada bab ini menjelaskan latar belakang masalah yang menjadi alasan penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan metode apa saja yang dilakukan untuk penelitian serta sistematika penulisan skripsi.

##### **Bab 2 : Landasan Teori**

Bab ini berisikan teori – teori yang mendukung penyelesaian skripsi untuk memahami permasalahan yang dibahas pada penelitian ini yaitu Analisis Sentimen Twitter Menggunakan Algoritma Support Vector Machine.

**Bab 3 : Metode Penelitian**

Bab ini membahas tentang penggunaan metode yang dipakai, gambaran umum sistem yang akan dibuat, alat dan bahan yang dibutuhkan dan perancangan antar muka sistem yang akan dibangun.

**Bab 4 : Hasil & Pembahasan**

Pada bab ini menjelaskan tentang bagaimana cara sistem bekerja dan memaparkan hasil dari tahap – tahapan yang dilakukan.

**Bab 5 : Penutup**

Bab ini menerangkan kesimpulan yang dapat diambil dari penelitian skripsi yang dilakukan dan menambahkan pendapat berupa saran yang dapat mendukung penelitian selanjutnya.

