

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pariwisata merupakan salah satu sektor yang sedang digalakkan dan dikembangkan oleh pemerintah Indonesia setiap daerah di Indonesia memiliki potensi wisata yang unik dan beragam jenisnya. Pemerintahan setiap daerah perlu memetakan potensi di daerahnya masing-masing sebagai saran dan data dalam pengambilan kebijakan pengembangan wisatanya. Salah satu sumber data yang bisa dijadikan acuan adalah media sosial. Para wisatawan baik domestik maupun mancanegara kerap mengunggah aktivitas pesan dan kesan tentang suatu objek wisata di media sosial seperti Twitter, Facebook, Instagram dan lain-lain.

"Cuitan-cuitan" pengguna di media sosial sering merupakan ekspresi dan ungkapan emosi penggunaannya atas suatu kejadian atau peristiwa "Cuitan - cuitan" di Twitter tentang wisata di Yogyakarta merupakan suatu ungkapan wisatawan setelah berkunjung ke wisata di Yogyakarta. "Cuitan" ini bisa diambil pengetahuannya menggunakan *Twitter mining*. Dengan melakukan *Twitter mining* dapat diambil suatu Sentiment Analysis guna dilakukan pengklasifikasi informasi itu termasuk negatif, positif, dan netral. Keseluruhan data didapat dari text mining yang terdapat dalam media sosial Twitter dengan memanfaatkan API di bahasa pemrograman python yang telah banyak mempermudah untuk kegiatan tweet mining. *Sentiment analysis* dapat dibagi menjadi dua data yaitu : *Course – grained*

sentiment analysis dan Fined – grained sentiment analysis. Tapi yang sedang terkenal sekarang adalah Findle-grained sentiment analysis yaitu melakukan pengelompokan perdokumen yang akan dapat digolongkan menjadi positif dan negatif. Pendekatan klasifikasi yang dilakukan menggunakan pendekatan SVM (*Support Vector Machine*) Prinsip dasar SVM adalah linear klasifikasi, dan selanjutnya dikembangkan agar dapat bekerja pada problem non-linear. Dengan memasukkan konsep kernel trick pada ruang kerja berdimensi tinggi.

Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Pembelajaran dengan cara ini disebut dengan pembelajaran terarah (*supervised learning*). Dengan pembelajaran terarah ini akan diperoleh fungsi yang menggambarkan bentuk ketergantungan input dan outputnya. Selanjutnya, diharapkan fungsi yang diperoleh mempunyai kemampuan generalisasi yang baik, dalam arti bahwa fungsi tersebut dapat digunakan untuk data input di luar data pembelajaran. SVM ini akan didukung dengan algoritma *Particle Swarm Optimization* yang digunakan untuk mengeser partikel mendekati partikel yang dianggap memiliki Gbest atau nilai terbaik. Dan dapat dijadikan untuk Feature Selection. Feature Selection merupakan teknik pemilihan atribut yang akan dipakai dalam klasifikasi yang mengakibatkan meningkatnya akurasi dari SVM.

## 1.2 Rumusan Masalah

Perumusan masalah dalam pembuatan Skripsi ini adalah:

Bagaimana melakukan optimalisasi atau perbaikan terhadap partikel yang dijadikan parameter algoritma SVM menggunakan PSO sebagai *Feature Selection* ?

### 1.3 Batasan Masalah

Batasan masalah dalam pembuatan Skripsi ini adalah:

1. Analisa sentimen dengan metode SVM menggunakan *tools* yang sudah tersedia dan mengklasifikasikan ke dalam tiga kelas sentimen, positif, negatif, dan netral. Pekerjaan yang dilakukan mencakup mengunduh *tweet*, *tweet preprocessing*, klasifikasi *tweet*, pengujian data, serta pemanfaatan *tools* atau *library* yang tersedia untuk analisis sentimen maupun *tweet preprocessing*.
2. Fokus dari penelitian ini yaitu menerapkan metode *Wrapper* pada *Feature Selection* sebelum masuk ke tahap klasifikasi yang dilakukan oleh SVM.
3. Analisis sentimen dalam penelitian ini menggunakan data *tweet* dari Twitter yang di unduh selama satu hari untuk pencarian *tweet* yang terkait dengan pariwisata Indonesia.
4. Bahasa pemrograman yang digunakan adalah python.
5. Python yang digunakan merupakan python versi 3.6.
6. Training data dilakukan secara terpisah dengan test atau predik suatu data.

### 1.4 Tujuan Penulisan

Tujuan dalam pembuatan Skripsi ini adalah:

1. Mengukur tingkat akurasi ketika dikombinasikan dengan algoritma (*Particle Swarm Optimization* untuk *Feature Selection*).
2. Memenuhi persyaratan kelulusan Strata 1 Teknik Informatika.

3. Meningkatkan kemampuan sebagai analis dan programmer.
4. Menambah pengetahuan tentang Python.

### 1.5 Manfaat Penulisan

Manfaat yang diharapkan dari penyusunan Skripsi ini antara lain:

- a. Bagi Pembaca
  1. Memperoleh informasi mengenai bagaimana tentang mengolah teks untuk mendapatkan suatu *Sentiment Analysis* dengan menggunakan *Support Vector Machine* (SVM)
  2. Bagaimana implementasi text mining dalam pemrograman Python
- b. Bagi Penulis
  1. Meningkatkan kemampuan dalam data mining dan analisis.
  2. Menerapkan ilmu pengetahuan yang telah didapat selama kuliah.

### 1.6 Metode Pengumpulan Data

#### 1.6.1 Studi pustaka

Mencari literatur atau jurnal yang terkait dengan *Text Mining*, *Support Vector Machine*, *Sentiment Analysis* dan *Particle Swarm Optimization*.

#### 1.6.2 Interview

Melakukan interview terhadap seorang yang memang sudah ahli atau sudah pernah melakukan sebuah *Sentiment Analysis*.

### 1.6.3 Survey

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

### 1.6.4 Ekperimen (Percobaan)

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

## 1.7 Sistematika Penulisan

Laporan penelitian ini akan disusun secara sistematis kedalam 5 bab, masing-masing bab akan diurutkan sebagai berikut :

### BAB I PENDAHULUAN

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

### BAB II LANDASAN TEORI

Pada bab ini akan diuraikan tentang teori yang digunakan sebagai dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi Pengertian Data Mining, *Support Vector Machine (SVM)*, Pengertian *Sentiment Analysis*, *Pre-Processing*, *Feature Selection*, *Term-Frequency-Inverse Document Frequency*, *Cosine Similarity*, Pengertian *Particle Swarm Optimization (PSO)*, *K-Fold Cross Validation*, *UML* dan *Confusion Matrix*.

### BAB III ANALISIS DAN PERANCANGAN

Pada bab ini akan diuraikan tentang perancangan dan model cara kerja yang akan dilakukan di dalam studi kasus sentiment analysis tweet pariwisata indonesia.

### BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini membahas *Sentiment Analysis* Twitter mining dari *Pre-Processing* data, *Tokenization*, *Steaming*, *TF-IDF*, *SVM*, *PSO* 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.

