

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan ilmu pengetahuan dan teknologi ditandai oleh kemajuan informasi dan komunikasi. Merambahnya teknologi informasi dalam kehidupan sehari-hari menimbulkan luapan informasi dalam bentuk digital maupun elektronik. Sudarsono (dalam Koswara, 1998: 87) menyatakan bahwa informasi dalam bentuk digital dapat didistribusikan melalui media internet. Internet menawarkan *alternative* baru dalam memperoleh informasi baru sekaligus menyebarkan informasi itu sendiri [1].

Internet merupakan jaringan *computer* berskala internasional yang dapat membuat masing-masing *computer* saling berkomunikasi dari berbagai tempat, negara, benua sehingga dapat berkomunikasi dengan *protocol* tertentu, saling mengirimkan dan mendapat informasi (hartono, 2000; 341). Internet secara tidak langsung merubah paradigma dalam mendapatkan informasi dan komunikasi, yang tidak lagi dibatasi oleh dimensi ruang dan waktu. Melalui internet semua orang bisa mendapatkan informasi dan berkomunikasi sesuai pribadi masing-masing [1]. Perkembangan internet dewasa ini berkembang sangat pesat dengan ditandai meningkatnya pengguna internet dunia. Kemajuan dari *Web 2.0* mendorong penciptaan konten *web* pribadi yang melibatkan sentimen seperti *blog*, *tweets*, *web* forum dan jenis-jenis *media sosial*. Dengan adanya keragaman aplikasi yang tumbuh di internet sebagian orang atau suatu badan dapat mengambil keuntungan dengan memanfaatkan pendekatan *opinion mining*, misalnya kampanye pemasaran dapat menerima dan mengevaluasi umpan balik dari basis pengguna yang besar, politisi dapat memperkirakan popularitas mereka, produsen dapat mengidentifikasi kelemahan dari produk mereka, pemerintah dapat meningkatkan kualitas layanan informasi yang ditawarkan kepada masyarakat sehingga pemerintah lebih bertanggung jawab kepada masyarakatnya dan dapat

memajukan transparansi di sektor publik. Salah satu tujuan dari *opinion mining* dan *sentiment analysis* adalah untuk membantu orang menemukan informasi yang berharga dari sejumlah data yang tidak terstruktur [2].

Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, sentimen, dll) bersifat positif, negatif atau netral. Analisis sentiment atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Banyak situs yang menyediakan *review* tentang suatu produk yang dapat mencerminkan pendapat pengguna. Salah satu contohnya adalah situs Internet Movie Database (IMDb) [3].

Internet Movie Database (IMDb) adalah situs *web* yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis, penata rias, *soundtrack*, nominasi yang pernah didapat, dan *review* dari para pengunjung. Ini mempengaruhi bagaimana cara kita mendapatkan informasi, dimana informasi dan hiburan dengan mudahnya kita akses lewat media digital. Selain itu, kita juga dapat berinteraksi satu sama lain antar sesama pengunjung dan memberikan tanggapan mengenai suatu film. Pengguna IMDb di Indonesia mencapai 1.3% dari seluruh dunia. Yaitu sekitar 1.3% dari 57 juta *visitors*. Ini membuat Indonesia berada diperingkat 12 menurut statistik dari *hypestat*. Kebanyakan pengguna IMDb merupakan bukan pengguna *unique*. Dalam arti kata pengguna tidak menjadi member hanya mengunjungi situs dan mencari beberapa informasi tentang film. Perbandingannya sekitar 1 dari 5 mahasiswa di Bandung menjadikan IMDb acuan sebelum menonton film baik bioskop ataupun *streaming*. Dari satu forum menunjukkan tujuan para *users* menggunakan IMDb.com untuk melihat *rating*, dan melihat profil *actor* atau

aktris kesukaan mereka sebelum menonton film. Ini memperlihatkan walaupun bukan sebagai pengguna *unique*, calon penonton tetap menggunakan IMDb sebagai sumber referensi untuk film yang akan ditonton maupun yang sudah ditonton. Untuk yang telah menonton film tersebut, *visitors* menggunakan IMDb untuk mencari *cast* yang ada di film tersebut. Berdasarkan pengamatan penulis, maka terlihat besarnya peran sebuah media (IMDb) untuk mengangkat serta suatu film agar dapat mempengaruhi keputusan calon penonton untuk menentukan pilihan menonton serta menilai sebuah kualitas film. IMDb memiliki 1 fitur berupa *rating* yang hanya bisa digunakan oleh *unique users*. Maksudnya adalah hanya yang melakukan registrasi saja yang bisa menilai atau memberikan *score* di satu film [1].

Adapun contoh tampilan *review* film dari *user* di IMDb ditunjukkan pada gambar 1.1



**Gambar 1. 1 Tampilan *User Review* di IMDb**

Dengan jumlah film dan *review* yang banyak, untuk memudahkan *user* dalam melihat *review* tersebut dibutuhkan Analisis Sentimen secara otomatis. Pada penelitian ini Analisis Sentimen digunakan untuk mengklasifikasi *review* film yang ada di IMDb kedalam sentimen positif, dan sentimen negatif. Sehingga kebutuhan *user* untuk menentukan akan menonton suatu film dengan melihat *review* penonton menjadi lebih cepat dan efisien. Metode *Machine Learning* yang digunakan peneliti untuk klasifikasi teks adalah *Support Vector Machine* (SVM). SVM merupakan suatu teknik untuk menemukan *hyperplane* yang bisa memisahkan dua set data dari dua kelas yang berbeda. SVM memiliki kelebihan diantaranya adalah dalam

menentukan jarak menggunakan *Support Vector* sehingga proses komputasi menjadi cepat [4]. Penelitian tentang metode SVM telah dilakukan sebelumnya yaitu klasifikasi komentar yang mengandung *bullying* pada Instagram yang kemudian mendapatkan hasil akurasi sebesar 90% [5]. Selain itu pada penelitian lain SVM digunakan untuk klasifikasi pada data akreditasi sekolah yang kemudian mendapatkan hasil akurasi sebesar 93,902% [4].

Dengan hasil akurasi yang cukup tinggi dan kemampun *Support Vector Machine* (SVM) yang mampu memisahkan dua set data dari dua kelas yang berbeda, maka pada penelitian ini penulis menerapkan metode *Support Vector Machine* untuk melakukan Klasifikasi Data *Review* IMDb Berdasarkan Analisis Sentimen dengan dataset yang digunakan yaitu berasal dari github.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah sebagai berikut.

- a. Bagaimana mengimplementasikan metode *Support Vector Machine* terhadap Analisis Sentimen untuk klasifikasi *review* di IMDb?
- b. Bagaimana hasil implementasi metode *Support Vector Machine* terhadap Analisis Sentimen untuk klasifikasi *review* di IMDb?

### 1.3 Batasan Masalah

Agar pembahasan tidak menyimpang, beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

- a. Metode yang digunakan untuk klasifikasi adalah *Support Vector Machine*.
- b. Dataset yang digunakan yaitu data *Review* di IMDb yang didapat dari github.
- c. Sentimen yang dianalisis meliputi sentimen positif dan sentimen negatif.
- d. Dataset yang digunakan berbahasa inggris.
- e. Hasil klasifikasi hanya menampilkan jumlah sentimen dan nilai akurasi sistem.

- f. Sistem digunakan sebagai pembuktian bahwa algoritma yang diimplementasikan berjalan.

#### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian klasifikasi data *review* IMDb adalah untuk mengetahui tingkat akurasi *Support Vector Machine* terhadap analisis sentimen untuk klasifikasi *review* di IMDb.

#### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian yang dilakukan adalah sebagai berikut.

a. Bagi Peneliti

Menambah wawasan dan ilmu pengetahuan seputar Analisis Sentimen dengan metode *Support Vector Machine*. Penelitian ini juga sebagai bentuk implementasi pembelajaran yang telah didapat selama perkuliahan.

b. Bagi Ilmu Pengetahuan

Mengimplementasi algoritma *Support Vector Machine* terhadap Analisis Sentimen, serta memberikan kontribusi terhadap ilmu pengetahuan tentang pengujian algoritma pada penelitian ini.

c. Bagi Peneliti Lanjutan

Hasil dari penelitian ini dapat digunakan sebagai acuan referensi untuk melakukan penelitian yang baru atau penelitian selanjutnya dengan menggunakan algoritma klasifikasi lainnya.

## 1.6 Metode Penelitian

Metode penelitian dapat disebut juga tahapan - tahapan atau langkah - langkah yang dilakukan dalam melakukan suatu penelitian. Berikut tahapan - tahapan metode yang dipakai dalam penelitian ini.

### a. Studi Pustaka

Studi Pustaka dilakukan oleh peneliti untuk mengumpulkan informasi dari penelitian - penelitian terkait untuk digunakan sebagai acuan peneliti dalam melakukan penelitian.

### b. Pengumpulan Data

Data input yang digunakan bersumber dari *review* di IMDb. Pengumpulan data diambil dari *github*. Data disimpan di Microsoft Excel dengan format penyimpanan *.csv*.

### c. Analisis dan Perancangan

Pada tahapan metodologi penelitian yaitu melakukan perancangan sistem yang dapat menganalisis sentimen. Tahapan dalam proses perancangan sistem ini antara lain perancangan *pre-processing*, perancangan pembobotan, perancangan algoritma klasifikasi *Support Vector Machine*, dan perancangan antarmuka sistem.

### d. Implementasi

Implementasi dilakukan dengan menggunakan rancangan sistem yang telah dibuat pada tahapan sebelumnya.

### e. Pengujian

Pengujian dilakukan ke sistem yang telah diimplementasikan. Pengujian dilakukan untuk evaluasi model dan mengetahui nilai performa dari algoritma yang digunakan.

### f. Penulisan Laporan

Hasil dari penelitian kemudian dituliskan dalam sebuah laporan. Pada penulisan laporan penelitian akan ditarik sebuah kesimpulan berdasarkan hasil pembahasan dan pengujian, serta memberikan saran untuk penelitian selanjutnya.

## 1.7 Sistematika Penulisan

Sistematika penulisan memberikan gambaran yang jelas serta menjadi pedoman dalam penulisan. Adapun sistematika penulisan pada penelitian ini disusun sebagai berikut.

### a. BAB I PENDAHULUAN

Bab ini berisi penjelasan dasar-dasar dari usulan penelitian, yang terdiri dari latar belakang, rumusan masalah, maksud dan tujuan penelitian, serta sistematika penulisan skripsi.

### b. BAB II LANDASAN TEORI

Bab ini berisi penjelasan mengenai landasan teori, konsep, model, dan metode dari berbagai penelitian yang terkait sebelumnya yang digunakan sebagai dasar untuk melakukan penelitian agar memahami konsep dan teori terhadap permasalahan yang akan diteliti. Bab ini meliputi kajian pustaka dan dasar teori dari penelitian yang akan dilakukan. Dasar teori yang digunakan untuk mendukung penelitian ini adalah Teks Mining, Analisis Sentimen, *Preprocessing*, dan Bahasa Pemrograman yang akan digunakan.

### c. BAB III METODOLOGI PENELITIAN

Bab ini berisi komponen alat dan bahan apa saja yang dibutuhkan dalam penelitian serta uraian mengenai langkah – langkah menyelesaikan penelitian ini.

### d. BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini berisi penjelasan mengenai proses implementasi dan pembahasan hasil dari perancangan yang dibuat oleh peneliti.

### e. BAB V KESIMPULAN DAN SARAN

Bab ini berisi pembahasan mengenai hasil dari keseluruhan penelitian dan juga membahas mengenai saran bagi penelitian yang akan datang tentang kekurangan penelitian ini yang nantinya bisa dikembangkan pada penelitian selanjutnya.

### f. DAFTAR PUSTAKA

Bab ini berisi daftar pustaka yang dipakai sebagai acuan referensi *literature* dalam penelitian ini.