

# BAB I PENDAHULUAN

## 1.1 Latar Belakang Masalah

Perkembangan informasi yang pesat tidak terlepas dari pengguna web yang menyediakan informasi yang beragam. Sebagian besar konten yang terdapat pada web bersifat subjektif atau merefleksikan opini dari orang yang menulis. Dengan pesatnya perkembangan web, semakin banyak orang menuliskan opini mereka tentang produk atau jasa. Hal ini dimanfaatkan sebagai acuan para *customer* untuk mengetahui pendapat orang lain tentang produk atau jasa yang akan digunakan. Sebaliknya bagi produsen dapat digunakan sebagai media pelacak tentang kepuasan pelanggan (Bo Pang, 2008). Berdasarkan (Bo Pang, 2002), 81% pengguna internet melakukan pencarian *review* secara *online* minimal sekali dan 73% sampai 87% pembaca *review* secara *online* mengakui pembacaan *review* mempengaruhi mereka dalam membeli produk atau menggunakan jasa.

Salah satu hal yang penting dalam sebuah *review* adalah opini yang terkandung didalamnya. Opini yang terkandung dalam *review* bisa berupa opini positif atau opini negatif (Hang Cui, 2008), maka dari itu dibutuhkan sebuah pengklasifikasian *review* yang dapat membedakan opini negatif dan opini positif. Salah satu pengklasifikasian yang dapat digunakan adalah *Text Mining*. *Text Mining* dapat didefinisikan sebagai pengambilan informasi yang bersumberkan dari beberapa dokumen. Dalam *Text Mining* terdapat beberapa tujuan penggunaan yang khas diantaranya adalah *text categorization*, *text clustering*, dan *sentiment analysis* (Ronen Feldman, 2007). Dalam kebutuhan untuk pengklasifikasian opini yang telah dijelaskan, fungsi *text mining* yang dapat digunakan adalah *sentiment analysis*. *Sentiment analysis* adalah teknik *text mining* yang dapat mengklasifikasikan sebuah opini ke dalam opini positif atau opini negatif.

*Sentiment analysis* atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi *sentiment* 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 (Bo Pang, 2002).

Penelitian di bidang *sentiment analysis* atau *opinion mining* mulai marak pada tahun 2002. Pang et.al. pada tahun 2002 mengklasifikasikan *review* dari film pada level dokumen yang memiliki pendapat positif atau negatif dengan menggunakan teknik *supervised learning*. Sekumpulan dari *review* film yang sebelumnya telah ditentukan menjadi baik positif ataupun negatif digunakan sebagai data latihan untuk beberapa algoritma *machine learning* yang sudah ada. Akurasi yang didapatkan berkisar antara 72% sampai 83% (Hang Cui, 2008).

Dalam penelitian ini metode yang digunakan adalah K-NN (*K-Nearest Neighbor*). *K-Nearest Neighbor* merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data latih yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut (Ronen Feldman, 2007).

Penelitian ini bertujuan untuk melakukan pengklasifikasian terhadap *sentiment* positif dan negatif dari *review* film berbahasa Inggris serta mencari akurasi dari metode yang digunakan yaitu *K-Nearest Neighbor* yang akan dibandingkan dengan Metode *Naive Bayes*. Sumber *review* film berbahasa Inggris dipilih karena ketersediaan data *corpus* dalam jumlah besar dan tersedianya algoritma *stemming* dalam bahasa Inggris yang digunakan dalam proses *text preprocessing*, sedangkan *K-Nearest Neighbor* dipilih karena pada *sentiment analysis* belum banyak digunakan.

Berdasarkan uraian di atas, maka disusun suatu laporan tugas akhir yang berjudul “***Sentiment analysis pada Review Film dengan Menggunakan Metode K-Nearest Neighbor***”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, rumusan masalah penelitian ini adalah sebagai berikut :

1. Bagaimana pengklasifikasian *sentiment* negatif dan *sentiment* positif terhadap data uji dengan menggunakan metode *K-Nearest Neighbor* ?
2. Bagaimana tingkat akurasi metode *K-Nearest Neighbor* terhadap data uji pada pembentukan *sentiment analysis* dibandingkan dengan metode *Naive Bayes* ?

## 1.3 Tujuan

Berdasarkan latar belakang yang telah dikemukakan, tujuan penelitian ini adalah sebagai berikut :

1. Melakukan pengklasifikasian terhadap data uji dengan menggunakan metode *K-Nearest Neighbor*.
2. Menguji akurasi metode yang digunakan yaitu *K-Nearest Neighbor* dalam *Sentiment analysis* dan membandingkannya dengan metode *Naive Bayes*.

## 1.4 Batasan Masalah

Dalam penelitian *Sentiment analysis* ini dibatasi dengan hal-hal sebagai berikut :

1. Sumber *review* film dalam bahasa Inggris yang diambil dari *Large Movie Review Dataset v1.0* dari *cornell movie review* data dengan jumlah data 700 dokumen.
2. Metode *Text Mining* yang digunakan adalah *K-Nearest Neighbor* yang akan dibandingkan dengan metode *Naive Bayes*.
3. *Sentiment analysis* yang dilakukan adalah *Coarse-grained sentiment analysis*.
4. Proses stemming menggunakan *Snowball Stemming Algorithms*. Proses stemmer tidak dijelaskan keseluruhan dikarenakan menggunakan *library* yang disediakan oleh Snowball Tartarus Organization.

5. Nilai  $K$  pada metode *K-Nearest Neighbor* yang diujikan adalah 1 , 3 , 5 , 7, 9, 11, 13, 15.
6. Nilai *Feature Selection* yang digunakan dalam pengujian adalah 100 , 200, 400 dan 800.

## 1.5 Sistematika Penulisan

Adapun sistematika penulisan laporan tugas akhir ini terdiri dari enam bab, dengan rincian sebagai berikut :

- BAB I**            **Pendahuluan**, Membahas mengenai Latar Belakang Masalah Tugas Akhir, Rumusan Masalah Tugas Akhir, Tujuan Tugas Akhir, Batasan Permasalahan Tugas Akhir dan Sistematika Penulisan Tugas Akhir.
- BAB II**            **Landasan Teori**, Pada Bagian ini berisi uraian tentang teori-teori yang mendukung dalam mengerjakan Tugas Akhir.
- BAB III**           **Metodologi**, Pada bagian ini akan dibahas metodologi penelitian yang dilakukan.
- BAB IV**           **Analisis dan Perancangan**, Pada bagian ini akan dipaparkan kebutuhan-kebutuhan ketika melakukan penelitian. Untuk itu pembahasan ini akan terbagi atas analisis dan perancangan.
- BAB V**            **Implementasi dan Hasil**, Pada bagian ini akan dibahas tentang tahapan implementasi. Pada bagian implementasi akan melakukan pengujian pada metode-metode yang telah dibuat pada perancangan sistem dan melakukan penghitungan akurasi metode. Bab ini juga membahas tentang hasil dari implementasi dan analisis.
- BAB VI**           **Kesimpulan dan Saran**, Pada bagian ini dibahas tentang kesimpulan dan saran-saran terhadap penelitian yang telah dilakukan.