

# 1.

# PENDAHULUAN

## 1.1 Latar Belakang Masalah

Sekarang ini, teknologi sudah semakin berkembang. Maraknya penggunaan internet menjadi salah satu buktinya. Internet dapat diakses darimanapun dan kapanpun, sehingga memudahkan manusia dalam kehidupan sehari-hari. Internet merupakan sumber informasi tanpa batas dimana penggunanya bisa mendapatkan informasi apapun yang diperlukan. Internet tidak hanya digunakan sebagai sumber informasi bagi pengguna, tapi juga sarana berkomunikasi yang menghubungkan pengguna internet dari seluruh dunia. Bahkan, sekarang ini, banyak perusahaan yang menggunakan internet sebagai sarana untuk berbisnis layaknya belanja *online*. Tidak hanya digunakan untuk berjualan barang, tetapi internet juga bisa digunakan untuk melihat tanggapan pelanggan terhadap sebuah produk dari perusahaan melalui ulasan yang diberikan oleh pelanggan. Dengan banyaknya informasi dan sumber-sumber yang tersedia di internet, ide untuk menyaring informasi secara otomatis akan menjadi sangat berguna [2]. Bagi pelanggan, dengan adanya ulasan akan menambah informasi dalam memilih sebuah produk. Sedangkan bagi perusahaan, ulasan yang ada bisa digunakan untuk mengetahui selera pasar [1].

Secara umum pada halaman *website*, ulasan yang tertulis merupakan teks bebas yang tidak terstruktur. Maka dari itu, ulasan pelanggan merupakan data yang tidak terstruktur. Untuk dapat menangani data seperti ini, dibutuhkan proses yang disebut analisis sentimen. Analisis sentimen bertujuan untuk mendapatkan atribut dan komponen dari sebuah objek yang telah dikomentari pada sebuah dokumen. Kemudian, dokumen tersebut dicari orientasinya, positif, negatif, atau netral [4]. Ada beberapa proses yang dilakukan dalam analisis sentimen yaitu, preprosesing, pencarian informasi, ekstraksi informasi, *machine learning*, dan *natural language processing*. Hal ini membuat perhatian pada analisis sentimen semakin meningkat [1]. Sebagian besar sentimen analisis yang telah ada mengambil ulasan produk sebagai datasetnya, padahal masih ada domain ulasan lain yang dapat digunakan, salah satunya yang digunakan pada tugas akhir ini, yaitu ulasan film. Tidak seperti pada ulasan produk, penanganan pada ulasan film memiliki karakteristik yang berbeda. Ketika orang-orang membuat ulasan, mereka tidak hanya memberi komentar tentang elemen dari film saja, tapi juga semua yang berhubungan dengan film, antara lain, aktor dan aktris, sutradara, penulis [3].

Pada penelitian Tugas Akhir ini, akan lebih difokuskan pada permasalahan analisis sentimen berdasarkan fitur karena menurut [4], evaluasi klasifikasi teks pada level dokumen, tidak memberi tahu apakah pemilik opini suka dan tidak suka. Dokumen positif pada sebuah objek tidak berarti pemilik opini berpendapat positif pada semua aspek atau fitur pada objek. Sebaliknya, dokumen negatif bukan berarti pemilik opini tidak menyukai semua tentang objek tersebut. Untuk dapat mengetahui aspek yang detail, level fitur dibutuhkan [4]. Permasalahan yang sama, yaitu analisis sentimen berdasarkan fitur telah dikerjakan pada [8,9] dengan domain pada barang elektronik menggunakan metode ekstraksi fitur CBA (*Classification Based on Association*) dilengkapi dengan *compactness pruning* serta *p-support pruning* dan metode klasifikasi sentimen menggunakan orientasi *opinion words* berbasis kamus

*WordNet*. Pada tugas akhir ini, domain yang difokuskan mengenai analisis sentimen berdasarkan fitur pada film dengan menggunakan metode ekstraksi fitur, yaitu Algoritma Apriori dilengkapi dengan *p-support pruning* dan metode klasifikasi sentimen, yaitu orientasi *opinion words* berbasis kamus *SentiWordNet*. Proses untuk mendapatkan *summary* ini terdiri dari tiga tahap utama, yaitu: 1) mengidentifikasi fitur-fitur yang terdapat dalam ulasan film dengan melakukan ekstraksi fitur 2) klasifikasi sentimen dari sebuah kalimat ulasan, apakah kalimat tersebut positif atau negatif 3) melakukan *summarization*.

Metode yang digunakan dalam penelitian ini adalah Algoritma Apriori, *P-support Pruning* dan Orientasi *Opinion Words*. Algoritma Apriori, salah satu metode yang terdapat pada *Association Mining*, digunakan untuk ekstraksi fitur. *P-support Pruning* adalah salah satu metode untuk melakukan *pruning* yang digunakan untuk melengkapi algoritma apriori. Hasil algoritma apriori memungkinkan mengandung kata yang bukan fitur dan tidak penting, sehingga dilakukan *p-support pruning* untuk menghapus kata-kata yang bukan fitur dan tidak penting tersebut. *Orientasi Opinion Words* merupakan salah satu metode yang menggunakan pendekatan berbasis kamus yang digunakan untuk melakukan klasifikasi sentimen.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah dikemukakan, maka perumusan masalahnya sebagai berikut.

1. Bagaimana mengekstraksi fitur yang terdapat dalam kalimat opini dengan menggunakan algoritma apriori dilengkapi dengan *p-support pruning*?
2. Bagaimana melakukan klasifikasi sentimen untuk mencari orientasi apakah kalimat opini tersebut positif atau negatif menggunakan orientasi *opinion words*?
3. Bagaimana melakukan *summarization* berdasarkan fitur, orientasi *opinion word*, dan sentimen?
4. Bagaimana tingkat performansi dari ekstraksi fitur pada algoritma apriori dan pada saat algoritma apriori dilengkapi dengan *p-support pruning* menggunakan parameter *recall* dan *precision*?
5. Bagaimana tingkat performansi dari klasifikasi sentimen dengan parameter akurasi, *recall*, *precision*, dan *F-score*?

## 1.3 Tujuan Penelitian

Tujuan dari penelitian tugas akhir ini adalah

1. Menganalisa dan mengimplementasikan algoritma apriori dan *p-support Pruning* pada ekstraksi fitur.
2. Menganalisa dan mengimplementasikan orientasi *opinion words* pada klasifikasi sentimen.
3. Menghasilkan *summarization* berdasarkan fitur, orientasi *opinion words*, dan orientasi sentimen.
4. Menguji dan menganalisa performansi ekstraksi fitur pada algoritma apriori dan pada saat algoritma apriori dilengkapi dengan *p-support pruning* dengan parameter *precision* dan *recall*.
5. Menguji dan menganalisa performansi klasifikasi sentimen pada sistem dengan parameter akurasi, *precision*, *recall*, dan *F-score*.

## 1.4 Batasan Masalah

Batasan masalah dalam penelitian tugas akhir ini, antara lain :

1. Dataset yang digunakan merupakan dataset berbahasa Inggris.
2. Dataset menggunakan kalimat sempurna yang penulisannya tidak disingkat.

## 1.5 Metodologi Penelitian

Metodologi yang digunakan untuk menyelesaikan tugas akhir ini adalah :

1. Identifikasi Masalah  
Pada tahap ini dilakukan identifikasi masalah yang akan dibahas pada tugas akhir ini, yaitu mengenai analisis sentiment berdasarkan fitur dari ulasan film.
2. Studi Literatur  
Pada tahap ini, dilakukan pencarian materi-materi yang berhubungan dengan tugas akhir ini. Referensi yang digunakan berupa jurnal dan buku berhubungan dengan analisis sentiment berdasarkan fitur, *summarization*, orientasi *opinion words* dan Algoritma Apriori.
3. Pengumpulan Data  
Pengumpulan dataset yang akan digunakan pada tugas akhir ini, diperoleh dari <http://www.imdb.com>.
4. Preprocessing Data  
Melakukan preprosesing data pada dataset. Proses-proses yang dilakukan yaitu, tokenisasi, mengubah sebuah kalimat menjadi per kata. *Part-Of-Speech Tagging*, untuk mengidentifikasi kata yang berupa *noun*, *verb*, *adjective*, *adverb*.
5. Perancangan Sistem  
Perancangan sistem dibuat sebagai gambaran umum tentang sistem yang akan dibuat dalam menyelesaikan masalah yang dibahas pada tugas akhir ini. Baik dari bahasa pemrograman, *software* atau *tools* pendukung, serta cara kerja sistem tersebut.
6. Pengujian dan Analisis Hasil  
Pengujian dan analisis akan dilakukan pada hasil yang didapat dari proses ekstraksi fitur dan klasifikasi sentimen.
7. Perumusan Kesimpulan dan Penyusunan Laporan Tugas Akhir  
Perumusan kesimpulan dibuat berdasarkan pengujian dan analisis hasil yang didapat. Setelah diambil kesimpulan, selanjutnya akan dilakukan penyusunan laporan tugas akhir.

