

# 1 Pendahuluan

## 1.1 Latar Belakang

Perdagangan merupakan pekerjaan yang telah lama dilakukan oleh sebagian besar penduduk dunia. Metode dan teknik perdagangan mulai ditekuni sejak awal abad 18 dengan adanya pasar komoditas di Wall Street [1]. Perkembangan teknologi membantu perdagangan menjadi lebih mudah dan cepat dalam transaksinya. Dewasa ini perdagangan sudah umum dilakukan secara daring, perantara yang biasa digunakan adalah media sosial dan Electronic Commerce (EC). Persaingan dagang yang semakin marak membuat produsen harus terus meningkatkan kualitas produknya, sedangkan konsumen harus lebih teliti dalam memilih tempat belanja. Adanya ulasan produk yang berisi tanggapan konsumen yang telah menggunakan dapat membantu calon konsumen memilih tempat berbelanja dan produk yang diinginkan, juga membantu produsen untuk meningkatkan kualitas produk sehingga mampu bersaing. Penelitian *Dimensional Search* menyatakan 90% calon konsumen terpengaruh dengan ulasan produk dengan tanggapan positif, dan 86% calon konsumen terpengaruh dengan ulasan produk dengan tanggapan negatif [2]. Sayangnya banyaknya tanggapan yang masuk menyebabkan penumpukan yang menimbulkan ketidakefisienan sehingga membingungkan pembacanya. *BrightLocal* melakukan penelitian bahwa 85% calon konsumen membaca 10 ulasan dan 7% saja membaca lebih dari 20 ulasan produk [3].

Ringkasan ulasan produk dapat membantu pembaca opini menemukan kesimpulan dari seluruh opini sehingga membantu untuk mengambil keputusan. Proses peringkasan ulasan produk memiliki tiga tahapan yang harus dilakukan. Tahap pertama yaitu ekstraksi fitur produk, ekstraksi yang digunakan dalam penelitian ini adalah *Pattern Knowledge* [4], yaitu melakukan pengekstrasian dengan mengambil kata dari pola tertentu, kemudian hasil tersebut dibandingkan dengan fitur yang ada di *dataset* sehingga didapatkan fitur teridentifikasi.

Tahap kedua yaitu klasifikasi opini dilakukan untuk menentukan orientasi positif dan/atau negatif pada fitur dalam ulasan produk. Klasifikasi dalam penelitian ini menggunakan kamus kata pada SentiWordNet yaitu kamus yang berisikan kata bernilai positif dan negatif cukup lengkap untuk analisis sentimen [5]. Pengklasifikasian pada penelitian ini menggunakan kata sifat yang nantinya dipasangkan dengan fitur yang telah diekstraksi pada tahapan sebelumnya.

Proses terakhir yang dilakukan adalah peringkasan ulasan produk. Peringkasan dilakukan agar penumpukan ulasan produk menjadi lebih mudah dipahami. Berdasarkan penelitian yang ada sebelumnya peringkasan ulasan hanya menggunakan peringkasan ekstrakstif yang sebenarnya masih bisa di spesifikkan lagi dengan pengelompokan kalimat dengan nilai kesamaan tinggi. Penelitian ini menggunakan tf-idf untuk memberikan nilai kepada kalimat terklasifikasi, dilanjutkan dengan pengelompokan kalimat menggunakan *centroid based clustering* [6] sehingga didapatkan hasil peringkasan yang lebih spesifik dan terkelompok.

## 1.2 Perumusan Masalah

Peringkasan dengan pendekatan ekstraktif yang telah dilakukan pada penelitian-penelitian sebelumnya melakukan peringkasan dokumen dengan memilih beberapa kalimat sebagai representasinya. Kalimat representasi tersebut masih dapat dikelompokkan berdasarkan kesamaan yang dimiliki dalam dokumen. Oleh karena itu, peringkasan lebih lanjut dibutuhkan dengan menghitung nilai kalimat menggunakan penilaian tf-idf yang dikelompokkan dengan *k-means clustering* dimana semakin rendah jarak antara *centroid cluster* dengan kalimat maka kalimat tersebut memiliki nilai kesamaan.

## 1.3 Tujuan

Berdasarkan uraian rumusan masalah di atas disimpulkan bahwa tujuan dalam penelitian tugas akhir ini memiliki tahapan yang dilakukan untuk peringkasan ulasan produk, yaitu:

1. Mengekstraksi dan menganalisis daftar fitur dari data ulasan produk menggunakan *Pattern Knowledge*.
2. Mengklasifikasi dan menganalisis hasil orientasi kalimat positif dan/atau negatif untuk dikelompokkan menggunakan SentiWordNet.
3. Menganalisis dan meringkas ulasan dengan penilaian menggunakan tf-idf dan *k-means clustering*.

## 1.4 Batasan Masalah

Tugas akhir ini memiliki batasan masalah sebagai berikut:

1. *Dataset* yang digunakan berupa ulasan produk berbahasa Inggris dengan format .txt.
2. *Dataset* yang digunakan berasal dari sembilan produk yang digunakan pada paper Mingqing Hu dan Bing Liu [7].
3. Ekstraksi fitur yang dilakukan terhadap fitur eksplisit, tidak termasuk fitur implisit pada ulasan produk.
4. Penentuan orientasi opini dilakukan pada level fitur.
5. Bahasa pemrograman yang digunakan adalah bahasa Java.

## 1.5 Metode Penyelesaian Masalah

Tahapan metode penyelesaian masalah yang dilakukan pada penelitian ini, antara lain:

1. Studi literatur

Tahap untuk mengumpulkan informasi, mempelajari, dan mendalami konsep mengenai penambangan data teks khususnya analisis sentimen menggunakan *clustering* kalimat yang diterapkan dalam proposal. Studi literatur dilakukan pada kajian seperti jurnal maupun buku terkait

2. Pengumpulan dan pengolahan data

Bahan penelitian tugas akhir yang disiapkan merupakan kumpulan dokumen berisi opini mengenai ulasan produk dalam bentuk file .txt.

### 3. Pembangunan model

Tahap ini dilakukan untuk mengetahui lebih dalam mengenai pembelajaran dari studi literatur baik studi kasus yang diangkat berikut penyelesaian masalah menggunakan metode terkait dengan.

### 4. Implementasi model

Merancang sistem berdasarkan hasil pembangunan model untuk mengolah data yang telah dipersiapkan. Pengimplementasian rancangan yang digambarkan menggunakan bahasa pemrograman java.

### 5. Analisis dan Pengujian

Berdasarkan implementasi yang telah dilakukan pada tahap sebelumnya dilakukan analisis dan pengujian hasil review berupa ringkasan produk. Hal ini bertujuan untuk meningkatkan kinerja sistem.

### 6. Pembuatan laporan tugas akhir

Hasil penelitian yang dilakukan dijadikan ke dalam buku tugas akhir sebagai dokumentasi lengkap.