

1. Pendahuluan

1.1. Latar Belakang

Pemanfaatan teknologi internet pada saat ini menyebabkan banyaknya penggunaan *website* untuk kegiatan transaksi jual beli suatu produk. Hal ini membuat para pembeli dapat dengan bebas memberikan opininya tentang produk yang mereka beli, karena dalam *website* jual beli atau yang lebih dikenal dengan *e-commerce* selalu disediakan ruang untuk pembeli mengeluarkan saran, pendapat atau opininya terhadap barang yang sudah dibeli ataupun barang yang akan dibeli. Data – data opini tersebut tidak hanya berguna untuk pembeli melainkan juga berguna untuk manufaktur penjualan produk. Oleh karena itu butuh dilakukan *review* terhadap suatu produk tertentu untuk mengidentifikasi kelebihan dan kekurangan produk dari sisi pembeli.

Terlalu banyaknya opini yang muncul tentang suatu produk akan menyulitkan pembeli yang potensial untuk membacanya dan mendapatkan informasi dalam hal pengambilan keputusan pembelian sebuah produk. Oleh karena itu akan dilakukan *mining* dan *summarization* pada *customer review* dari sebuah produk untuk mengatasi permasalahan tersebut. *Summarization* yang dilakukan berbeda dengan *text summarization* biasa karena pada *opinion summarization* untuk *review* sebuah produk hanya dilakukan *mining* fitur dari produk tersebut dengan *review* yang berupa opini dari *customer* dan menentukan apakah opini tersebut positif atau negatif.

Dalam proses *summarization* akan dilakukan *mining* terhadap opini – opini yang sudah terkumpul terlebih dahulu dengan pendekatan konsep *association rule*. Konsep ini bertujuan untuk membuat *rule – rule* yang mendefinisikan keterkaitan antara fitur dan opini dari sebuah produk. Penggunaan konsep *association rule* didasari dengan prinsip apriori dimana jika set fitur-opini sering muncul dalam sebuah opini maka semua subset akan mengikutinya. Oleh karena itu konsep ini dapat digunakan untuk menentukan keterkaitan antara opini umum dari suatu produk dan menentukan fitur produk yang sering muncul beserta opini dari fitur tersebut.

Penggunaan data pada sistem yang akan dibuat merupakan data yang bersifat *lexical* atau data yang berhubungan dengan bahasa, sehingga algoritma PMI – IR (*Pointwise Mutual Information – Information Retrieval*) dapat digunakan untuk proses *opinion summarization* karena merupakan algoritma yang menggunakan pendekatan teori informasi untuk menemukan keterkaitan antara kata – kata.. Algoritma PMI – IR menghitung seberapa banyak sebuah kata disebutkan dalam sebuah opini agar dapat ditarik sebuah informasi yang akan kita dapatkan tentang suatu produk. Algoritma PMI – IR merupakan pengembangan dari algoritma PMI dengan menambahkan konsep *Information Retrieval* untuk meningkatkan ketepatan perhitungan kemunculan pasangan kata dalam sebuah opini dan mencatatnya sebagai nilai *hits* yang dihasilkan. Oleh karena itu diharapkan algoritma ini dapat meningkatkan nilai akurasi dari algoritma sebelumnya sehingga dapat menghasilkan *summary* atau kesimpulan yang lebih tepat dalam hal

penentuan kelebihan dan kekurangan suatu produk dari sebuah *customer review*.

1.2. Perumusan masalah

Berdasarkan latar belakang tersebut, permasalahan yang akan dibahas pada tugas akhir ini antara lain :

1. Bagaimana cara meringkas opini berdasarkan fitur dan orientasinya ?
2. Bagaimana cara mengekstrak fitur produk dan orientasi opini dari fitur tersebut dalam sebuah review produk ?
3. Bagaimana cara menentukan pasangan kata fitur dengan orientasi opininya yang muncul bersamaan ?
4. Bagaimana cara menghitung nilai ketergantungan pasangan kata yang merupakan fitur produk dan orientasinya dengan algoritma *Pointwise Mutual Information – Information Retrieval* (PMI – IR) ?

1.3. Batasan masalah

Batasan masalah yang diberikan dalam pengerjaan tugas akhir ini adalah :

1. Penulis tidak melakukan penelusuran web (*web crawling*) untuk mendapatkan data melainkan dengan menggunakan data set online review Amazon.com yang telah tersedia.
2. Data set diasumsikan sudah merupakan sebuah opini sehingga tidak diperlukan tahapan untuk penentuan opini atau bukan dari sebuah *statement* atau pernyataan *customer*.
3. Review produk yang digunakan merupakan review yang secara eksplisit menyebutkan fitur dan orientasinya.
4. Proses pengekstrakan fitur dan opini dari sebuah review produk menggunakan algoritma ekstraksi.
5. Algoritma ekstraksi tidak melakukan proses ekstraksi terhadap kata fitur yang berupa fitur frase.
6. Peringkasan yang dilakukan hanya merupakan pengelompokan kata berdasarkan fitur dan orientasinya.
7. Kata opini yang diekstrak oleh algoritma ekstraksi hanya kata opini yang memiliki tag *adjective*

1.4. Tujuan

Berdasarkan rumusan masalah diatas, tujuan dari Tugas Akhir ini adalah :

1. Menganalisis dan meringkas opini berdasarkan fitur dan orientasinya.

2. Menganalisis dan mengekstrak fitur produk dan orientasi opini dari fitur tersebut dengan algoritma ekstraksi (*extraction algorithm*).
3. Menganalisis dan menentukan pasangan kata fitur dengan orientasi opininya sebagai suatu *itemset* dengan *Association Rule Mining*.
4. Menganalisis dan menghitung nilai ketergantungan pasangan kata yang merupakan fitur produk dan orientasinya dengan algoritma *Pointwise Mutual Information – Information Retrieval* (PMI – IR).

1.5. Metodologi penyelesaian masalah

Metodologi penyelesaian yang digunakan pada penelitian tugas akhir ini terdiri dari 6 tahap yaitu :

1. Studi Literatur
 - Pencarian materi yang dibutuhkan dan berhubungan dengan permasalahan, seperti materi tentang *opinion mining*, *opinion summarization*, algoritma PMI - IR dan materi pendukung lainnya.
 - Pemahaman materi yang telah diperoleh untuk membantu proses penyelesaian masalah.
2. Pengumpulan data

Pencarian dan pengumpulan data set berupa review produk yang berasal dari online review Amazon.com.
3. Tahap perancangan sistem

Perancangan sistem yang akan dibangun adalah sebagai berikut :

 - a. Data cleaning : Melakukan analisis terhadap dataset asli, kemudian dilakukan pembersihan data sehingga menghasilkan data yang siap untuk diproses.
 - b. Pre-processing : Melakukan tagging berdasarkan *Part-of-Speech* (POS) oleh sistem ahli dan mengekstraksi kata yang berupa fitur produk dan opini yang berupa orientasi dari fitur tersebut.
 - c. Pembuatan Rule : Melakukan pencarian frequent item set dari kata – kata yang telah diekstrak dengan metode *Association Mining Rule*.
 - d. Perhitungan PMI–IR : melakukan perhitungan nilai ketergantungan antara fitur dan orientasinya pada *rule* yang telah terekstrak, kemudian dilakukan pemilihan *rule* dengan nilai PMI-IR yang terbaik jika terdapat *rule* dengan fitur yang sama.
 - e. Pengorientasian Opini : Mengidentifikasi orientasi kata opini apakah merupakan kata positif atau kata negatif.
 - f. Summarization : pengelompokan pasangan kata berdasarkan orientasinya untuk menentukan kelebihan dan kekurangan suatu produk.

5. Tahap Implementasi
Pada tahap ini, akan dilakukan implementasi berdasarkan rancangan yang telah dibuat sebelumnya.
6. Tahap Pengujian dan analisa hasil
Melakukan Pengujian sistem yang dibangun berdasarkan perancangan sebelumnya dengan menggunakan data uji dan menganalisis hasil pengujian berdasarkan performansinya.
7. Tahap penyusunan laporan
Merupakan tahapan terakhir dalam penelitian untuk mengambil kesimpulan berdasarkan hasil pengujian dan dilakukan penyusunan laporan tugas akhir.