

# 1. Pendahuluan

## 1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan internet telah menyebabkan perubahan dan cara pandang hidup manusia. Perkembangan yang sedemikian pesatnya telah membawa dunia memasuki era baru yang lebih cepat. Perkembangan internet sudah menjadi kebutuhan sehari-hari yang sangat vital di beberapa Negara, termasuk Indonesia.

Perkembangan teknologi informasi dan internet banyak dimanfaatkan oleh perusahaan dan perorangan untuk menjalankan beberapa bisnis secara *online*. Salah satu diantaranya adalah dengan membuat sebuah halaman *website* belanja *online*. Selain memudahkan dalam berbelanja, pada *website* belanja *online* juga sering ditemukan fitur *review product* atau tanggapan dari barang yang dijual pada *website* belanja *online*.

Tanggapan mengenai sebuah barang pada satu *website online* sering dijadikan sebagian acuan terhadap sebuah kualitas barang. Bahkan menurut sumber dari *Invesp More Conversions*, 90% dari pengguna *website online* membaca tanggapan dari sebuah produk, dan 80% dari mereka mempercayai hasil dari tanggapan terhadap suatu produk [1]. Namun demikian, banyaknya tanggapan terhadap sebuah barang dalam satu *website online*, menjadikan kesulitan tersendiri untuk menyimpulkan hasil dari tanggapan barang tersebut.

Maka dari itu untuk mempermudah dalam menyimpulkan hasil tanggapan dari sebuah barang pada satu *website online*, perlu dibuat sebuah *system* untuk menganalisa hasil tanggapan dari konsumen yang telah dikirim. Sistem akan menganalisa aspek yang terdapat pada kalimat dan menentukan apakah termasuk ke dalam karakter *positive* atau *negative*. Dengan demikian hasil yang didapat bisa dijadikan kesimpulan terhadap tanggapan sebuah barang dalam suatu *website online*.

Sistem peringkasan ini dibangun menggunakan metode CRF (*Conditional Random Fields*) untuk ekstraksi aspeknya dan K-NN untuk klasifikasinya. Metode CRF dan KNN dipilih karena metode ini dapat digunakan untuk mengolah data berupa *text* dan menghasilkan hasil yang baik.[4,10] Parameter yang dibutuhkan pada sistem ini adalah persentase data training, penggunaan *lemmatization* pada *preprocessing*, nilai standar deviasi pada ekstraksi aspek, nilai *learning rate* pada ekstraksi aspek, *threshold* pada ekstraksi opini, dan nilai *k* pada klasifikasi. Penelitian ini dilakukan untuk mencari nilai efektif parameter inputan.

## 1.2 Perumusan Masalah

Pada tugas akhir ini, beberapa rumusan masalah yang ingin diselesaikan adalah :

1. Apakah data training berpengaruh terhadap ekstraksi dan klasifikasi pada penelitian ini?
2. Bagaimana pengaruh parameter standar deviasi dan *learning rate* terhadap metode *conditional random field* pada ekstraksi fitur?
3. Bagaimana pengaruh *threshold* pada ekstraksi opini terhadap proses klasifikasi?
4. Bagaimana pengaruh parameter nilai k terhadap metode K-NN pada klasifikasi?

## 1.3 Tujuan

Tujuan yang ingin dicapai dari pengerjaan Tugas Akhir ini adalah:

1. Menganalisis pengaruh data training terhadap ekstraksi dan klasifikasi pada penelitian ini.
2. Menganalisis pengaruh parameter standar deviasi dan *learning rate* terhadap metode *conditional random field* yang dilakukan pada ekstraksi fitur.
3. Menganalisis pengaruh *threshold* pada ekstraksi opini terhadap proses klasifikasi.
4. Menganalisis pengaruh parameter nilai k terhadap metode K-NN yang dilakukan pada klasifikasi.

## 1.4 Batasan Masalah

Batasan masalah dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Data berupa *text* (.txt)
2. Data dalam bahasa inggris
3. Data yang digunakan adalah data yang sudah diberi aspek dan *polarity score*
4. Sistem berjalan secara *offline*
5. Sistem berfokus pada fitur eksplisit
6. Sistem dibangun dalam bahasa java menggunakan *Netbeans IDE*
7. Hasil dari peringkasan berupa *polarity* (+) atau (-)

## 1.5 Metodologi Penelitian

Susunan metodologi penelitian yang dilakukan adalah sebagai berikut :

1. Studi Literatur

Mencari dan mengumpulkan materi yang terkait dengan penelitian tugas akhir. Materi yang telah terkumpul dipelajari untuk dijadikan referensi penelitian tugas akhir.

## 2. Pemilihan Data

Menyiapkan data yang akan digunakan pada penelitian tugas akhir. Dalam penelitian ini data yang digunakan berbentuk .txt yang berisi *review* produk. Data yang ada sudah memiliki aspek dan *polarity score*-nya.

## 3. Perancangan Sistem

Membangun sistem sesuai dengan spesifikasi yang akan digunakan. Sistem ini mencakup proses *preprocessing* (Tokenisasi, *Stopword Removal*, *Lemmatization*, *POS Tagging*), ekstraksi aspek (*Conditional Random Field*), ekstraksi opini (*Nearest Opinion Word*) dan klasifikasi (*K-NN*).

## 4. Implementasi dan Analisis

Sistem yang sudah dibangun akan diuji dengan dataset dan data *input* yang berbeda-beda. Setelah itu, hasil *output* sistem dengan data *input* dan dataset yang berbeda-beda dianalisis.

## 5. Kesimpulan

Penarikan kesimpulan akhir berdasarkan studi literatur, hasil pengujian sistem, dan analisis.