

1. Pendahuluan

1.1 Latar Belakang

Teknologi Internet yang berkembang dengan pesat membuat penggunaanya bertambah seiring waktu. Melalui Internet, masyarakat bisa mendapatkan banyak informasi ataupun melakukan hal komersil seperti jual beli. Jual beli *online* seperti ini dimanfaatkan oleh para produsen untuk mempromosikan produk yang mereka jual. Dalam situs promosi tersebut biasanya disediakan kolom untuk para pembeli atau pembaca untuk memberikan ulasan tentang produk tersebut

Negara Indonesia yang menganut kebebasan berpendapat membuat masyarakat tidak segan untuk menulis ulasan. Ditambah lagi dengan semakin banyaknya produk yang dipromosikan oleh produsen lewat internet membuat masyarakat dapat dengan nyaman untuk menulis ulasan sesuai dengan pendapat mereka. Akibatnya, jumlah ulasan dari produk mengalami peningkatan yang sangat besar. Produk-produk yang dibahas pada sebuah situs terkenal seperti www.teknoup.com bahkan mempunyai ratusan ulasan pembaca didalamnya. Hal ini dapat mengakibatkan calon pembeli dapat mengalami kesulitan dalam membaca ulasan dari produk tersebut. Padahal, ulasan tersebut dapat mempengaruhi mereka dalam mengambil keputusan untuk membeli suatu produk[2]. Masalah inilah yang mendasari sehingga diperlukan sebuah kemudahan bagi pelanggan untuk melihat ulasan dengan mudah dan tepat guna[13].

Dalam data mining, dikenal istilah *opinion summarization* untuk memberikan kemudahan pada pembaca dalam membeli sebuah produk. Pada *opinion summarization* ada dua tahap utama yang dilakukan yaitu mengidentifikasi fitur-fitur dari produk diulas oleh pembaca sekaligus mengidentifikasi opini yang diekspresikan pada fitur tersebut kemudian untuk setiap fitur, diidentifikasi berapa banyak review pelanggan yang memiliki opini positif maupun opini negatif[13]. Namun, pada Tugas Akhir ini hanya akan menyelesaikan tahap pertama, yaitu mengekstraksi fitur dan opini dari rangkaian ulasan pembaca.

Pada penelitian yang dilakukan sebelumnya, ekstraksi fitur dan opini dilakukan dalam dua proses yang berbeda. Proses pertama adalah mengekstrak fitur pada produk kemudian mengekstrak opini dengan mencari kata sifat atau kata keterangan disekitar fitur produk. Hal ini dapat menyebabkan kesalahan yang terjadi pada ekstraksi fitur akan diwariskan pada proses selanjutnya yaitu ekstraksi opini. Oleh karena itu, ekstraksi fitur dan opini perlu dilakukan dalam satu proses yang sama[14].

Dalam Tugas Akhir ini, ekstraksi fitur dan opini dilakukan dengan menggunakan metode *Conditional Random Fields (CRF)*. CRF adalah model probabilistik yang memetakan distribusi probabilitas kondisional dari kumpulan keluaran berupa label yang mungkin untuk suatu sekuens data. Setiap kalimat dalam ulasan akan dianggap sebagai sekuens data. Kemudian dari sekuens data tersebut akan dilakukan pelabelan fitur dan opini. CRF terbukti sangat efektif dan memberikan hasil yang baik dalam pelabelan sekuens data[11]

1.2 Rumusan Masalah

Masalah yang akan ditangani dalam Tugas Akhir ini adalah :

1. Bagaimana mengekstrak fitur dan opini pada dokumen berbahasa Indonesia dengan menggunakan metode *Conditional Random Fields*.
2. Bagaimana performansi metode *Conditional Random Fields* terhadap karakteristik data input ?
3. Bagaimana performansi metode *Conditional Random Fields* terhadap parameter *Gaussian prior*, *learning rate*, dan *feature function* yang digunakan ?

1.3 Batasan Masalah

Adapun batasan masalah dalam Tugas Akhir ini adalah sebagai berikut :

1. Ekstraksi fitur dan opini dilakukan pada data ulasan berbahasa Indonesia mengenai produk ponsel.
2. Dataset yang digunakan hanya terbatas pada produk ponsel.
3. Hanya terdapat 3 jenis label yang akan digunakan dalam tugas akhir ini yaitu A untuk label kata yang berupa fitur, O untuk label kata yang berupa opini, dan BG untuk label kata selain fitur dan opini.
4. Model CRF yang digunakan adalah *Linear-chained CRF Model*.
5. Sistem yang akan dibangun tidak terhubung ke internet.

1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Mengekstrak fitur dan opini yang ada pada ulasan.
2. Menganalisis pengaruh karakteristik data terhadap performansi yang dihasilkan
3. Menganalisis pengaruh parameter *Gaussian prior*, *learning rate*, dan *feature function* yang digunakan terhadap performansi yang dihasilkan.

1.5 Metodologi Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian Tugas Akhir ini adalah sebagai berikut :

1. Studi Literatur
Tahap ini merupakan tahap persiapan yang meliputi pencarian dan pengumpulan informasi yang berkaitan dengan tugas akhir. Informasi didapat dari jurnal-jurnal ilmiah, buku-buku, atau hal lain yang berkaitan dengan *Conditional Random Fields* dan ekstraksi informasi
2. Pengumpulan Data
Pada tahap ini, akan dilakukan pengumpulan data berupa dokumen berbahasa Indonesia. Jenis dokumen yang digunakan adalah ulasan dari pembaca terhadap produk ponsel pada situs www.teknoup.com Pengambilan data dilakukan dengan cara menyalin ulasan kedalam sebuah dokumen.
3. Perancangan Sistem
Melakukan analisis kebutuhan yang diperlukan dalam sistem serta merancang sistem ekstraksi fitur dan opini pada dokumen berbahasa Indonesia dengan menggunakan metode *Conditional Random Fields*. Penentuan alat bantu pemrograman yang digunakan juga dilakukan pada tahap ini.

4. Implementasi

Melakukan implementasi berdasarkan rancangan yang telah dibuat sebelumnya. Secara umum tahap implementasi dalam ekstraksi informasi dengan *Conditional Random Fields* antara lain :

- Ekstraksi *feature function*. *Feature function* dibentuk dengan mengasosiasikan *feature* dengan labelnya yang bersesuaian pada data *training*.
- *Training*. Data *training* yang telah dilabeli sebelumnya dan parameter *gaussian prior* dan *learning rate* menjadi inputan pada tahap ini. Dalam *training* diformulasikan *likelihood* data dan memaksimalkannya untuk mendapatkan nilai optimal dari parameter pendukung.
- *Decoding*. Parameter pendukung *feature function* yang dihasilkan pada fase *training*, kemudian menjadi inputan bersama data *testing* dalam melakukan decoding untuk mendapatkan label yang terbaik untuk setiap data yang diuji.

5. Pengujian

Pada tahap ini, dilakukan pengujian pada data yang sudah dikumpulkan terhadap sistem yang telah dibuat dalam mengekstraksi informasi dengan metode *Conditional Random Fields*. Skenario untuk menguji akurasi disiapkan untuk melihat pengaruh setiap parameter yang dapat mempengaruhi performansi dari sistem yang telah dibuat.

6. Analisa Hasil

Analisis akurasi terhadap hasil pengujian yang didapat berdasarkan skenario uji yang telah dibuat sebelumnya

7. Pengambilan Kesimpulan dan Pembuatan Laporan

Membuat kesimpulan dari analisa yang telah dilakukan dan membuat laporan tugas akhir yang mendokumentasikan tahap-tahap penelitian.