

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Penggunaan internet semakin hari berkembang semakin pesat. Saat ini hampir seluruh perusahaan memanfaatkan internet dalam mempromosikan produk yang mereka jual, bahkan tidak sedikit yang menggunakan internet dalam melakukan transaksi jual beli. Pemanfaatan media elektronik dalam perdagangan biasa disebut *e-commerce* (*electronic commerce*).

Ekspansi yang cepat dari penggunaan *e-commerce*, menjadikan semakin banyak orang yang membeli produk secara *online*. Merupakan hal biasa jika pedagang online meminta pelanggan dalam membagi pengalaman mereka dalam menggunakan produk yang dijual. Banyaknya pelanggan yang memberikan *review* produk menyebabkan kesulitan dalam membaca *review* produk tersebut, terlebih lagi untuk item produk yang populer. Pelanggan akan kesulitan dalam memutuskan item produk mana yang akan dibelinya.

Kesulitan pelanggan dalam membaca *review* produk dalam jumlah banyak, menyebabkan perlunya dilakukan suatu analisis sentimen dalam melihat ekspresi dan opini pelanggan dalam tinjauan produk yang mereka buat, apakah positif atau negatif. *Sentiment analysis* atau *opinion mining* mengacu pada aplikasi dari *Natural Language Processing (NLP)*, *Computational Linguistic*, dan *text analytics* untuk mengidentifikasi dan mengekstrak informasi subjektif dari material sumber [3]. Dalam melihat sebuah review, pelanggan tidak hanya melihat opininya saja, namun juga melihat fitur yang dikomentari. Oleh sebab itu, Tugas Akhir ini mengacu pada fitur serta orientasi opini yang mengikutinya, atau biasa disebut *feature based opinion mining* dimana ekstraksi fitur akan dilakukan pada level kalimat. Fitur tersebut mengacu pada *component*, *feature*, dan *function* dari produk yang dikomentari. Ekstraksi fitur dalam suatu kalimat opini tersebut dapat mempermudah pelanggan dalam menemukan produk yang sesuai dengan keinginannya.

Dalam melakukan analisis sentimen ini, dibentuk suatu kerangka kerja baru yang secara alami merupakan integrasi antara model leksikal dengan *Hidden Markov Model (HMM)*. HMM kemudian akan diintegrasikan dengan model leksikal, *Lexicalized Hidden Markov Model (L-HMM)*. Berbeda dari metode lainnya yang hanya memanfaatkan metode statistik ataupun metode leksikal saja, L-HMM yang merupakan kombinasi keduanya mampu memperkaya model HMM, dimana kata-kata yang bersifat "luar biasa", yaitu kata-kata yang probabilitas distribusinya dalam kategori tertentu berbeda dari yang lain pada suatu analisis sentimen, akan dibuatkan state baru untuk memperkenalkan kata-kata tersebut dalam konteks.

Metode L-HMM yang digunakan, tidak hanya dapat mengenali suatu kalimat opini saja, namun juga dapat digunakan dalam menentukan orientasi opini yang dikenali, apakah opini positif atau opini negatif. Selain itu, metode L-HMM ini juga dapat mengenali pasangan opini beserta fiturnya sekaligus. Metode L-HMM ini kemudian akan dikombinasikan dengan teknik bigram *smoothing*, yaitu teknik yang

digunakan jika menghasilkan probabilitas 0. Teknik ini digunakan untuk membantu perhitungan peluang, sehingga hasil akan semakin baik dengan memaksimalkan kondisi peluang yang ada.

1.2 Perumusan Masalah

Permasalahan yang akan dibahas pada Tugas Akhir ini adalah :

1. Bagaimana mengintegrasikan model leksikal dengan metode statistik menggunakan *Lexicalized Hidden Markov Models* (L-HMM) dalam pengenalan kalimat opini?
2. Bagaimana menggali fitur produk beserta orientasi opini yang mendeskripsikan fitur tersebut dalam sebuah kalimat?
3. Bagaimana pengaruh penggunaan teknik *smoothing* dalam penentuan fitur produk dan orientasi opini?

1.3 Batasan masalah

Batasan masalah yang diberikan dalam pengerjaan Tugas Akhir ini adalah :

1. Dataset yang akan digunakan berasal dari *online review* [Amazon.com](https://www.amazon.com)
2. Teknik *smoothing* yang digunakan adalah teknik *smoothing* Jelinec-Mercer.

1.4 Tujuan

Tujuan Tugas Akhir ini adalah :

1. Mengintegrasikan model leksikal dengan metode statistik dalam *Lexicalized Hidden Markov Model* (L-HMM) dalam pengenalan kalimat opini.
2. Menggali fitur produk beserta orientasi opini yang mendeskripsikan fitur tersebut dalam sebuah kalimat.
3. Mengetahui pengaruh penggunaan teknik *smoothing* dalam penentuan fitur produk dan orientasi opini.

1.5 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan dilakukan pada penelitian Tugas Akhir ini terdiri dari 6 tahap. Tahapan-tahapan tersebut adalah sebagai berikut :

1. Studi literatur
Melakukan proses pencarian dan pengumpulan referensi yang berkaitan dengan analisis sentiment, *Lexicalized Hidden Markov Model* (L-HMM) dan teknik bootstrapping melalui buku, jurnal, dan referensi lainnya untuk kemudian dipelajari lebih dalam.
2. Pengumpulan data
Pada tahap ini akan dilakukan pengumpulan data yang berasal dari *online review*, yaitu [Amazon.com](https://www.amazon.com)
3. Analisis dan perancangan sistem

Melakukan analisa kebutuhan dan perancangan sistem yang akan dibangun sesuai dengan identifikasi kebutuhan. Perancangan sistem yang akan dibangun adalah sebagai berikut:

- a. *Pre-processing* : Melakukan penyisiran data, kemudian data dibersihkan sehingga menghasilkan suatu kumpulan tulisan yang akan ditinjau
 - b. *Evaluation Data* : Melakukan *tagging* berdasarkan *Part-of-Speech* (POS) oleh system ahli
 - c. *Training Corpus* : Melakukan *tagging* berdasarkan *Part-of-Speech* (POS) dan sebagian lagi dilakukan *tagging manual* untuk kemudian dilakukan *bootstrapping* data
 - d. *Bootsraping data* : Melakukan pendekatan *bootstrapping* untuk mengekstrak data agar dapat melakukan pembelajaran mandiri
 - e. *Classifier* : Melakukan klasifikasi data berdasarkan metode *Lexicalized Hidden Markov Model* (L-HMM) kemudian melakukan *decoding* untuk mencari pola tag terbaik
 - f. *Post-processing* : Melakukan penghapusan kalimat yang bukan merupakan ekspresi dari pelanggan dan untuk entitas produk yang berbeda serta menentukan orientasi suatu kalimat opini, apakah merupakan kalimat positif atau kalimat negatif.
4. Implementasi dan pembangunan sistem
Melakukan implementasi dari hasil perancangan yang telah dibuat sebelumnya. Implementasi ini dilakukan dengan membangun sebuah sistem berupa program yang mampu memberikan label terhadap dataset yang diberikan.
 5. Pengujian sistem dan analisa hasil
Melakukan pengujian sistem yang telah dibuat, apakah sesuai dengan perancangan sistem yang diinginkan. Kemudian melakukan analisa terhadap hasil empirik yang diberikan oleh sistem.
 6. Pengambilan kesimpulan dan pembuatan laporan
Tahap ini merupakan tahap terakhir, dimana berdasarkan analisis dari hasil yang diberikan saat pengujian, ditarik kesimpulan dan juga dilakukan proses dokumentasi sebagai laporan untuk Tugas Akhir.