

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data ulasan suatu produk sangat penting untuk konsumen yang ingin membeli sebuah produk dan untuk produsen yang ingin melihat tanggapan konsumen. Konsumen dan produsen tidak lagi terbatas meminta teman atau keluarga berpendapat dalam membeli ataupun meminta tanggapan suatu produk, karena sudah banyaknya ulasan yang dikemukakan di forum publik web dan jejaring sosial. Banyak konsumen dan produsen yang ingin tahu apakah ulasan produk tersebut bersifat negatif atau positif. Maka diperlukan analisis sentimen untuk mendeteksi secara otomatis label positif dan negatif dari ulasan produk dan menjadikan ringkasan. Ringkasan tersebut dapat memudahkan konsumen menyaring informasi dan dapat membantu mengambil keputusan dalam membeli sebuah produk maupun memudahkan produsen untuk melihat tanggapan konsumen dan menarik kesimpulan dari produk.

Pada penelitian analisis sentimen yang telah dilakukan pada tugas akhir sebelumnya [8], khususnya pada bagian klasifikasi masih dapat untuk dikembangkan. dengan menambahkan beberapa fitur ciri seperti semantik dan negasi yang dapat menambah informasi *classifier*. Pada tugas akhir sebelumnya [8] masih belum terdapat informasi fitur ciri semantik, sehingga sulit mendapatkan informasi fitur produk yang memiliki makna semantik, sehingga masih terdapat fitur produk yang khususnya memiliki makna semantik masih sulit dikenali yang menyebabkan *classifier* tidak dapat mengklasifikasi fitur produk semantik. Selanjutnya, masih belum terdapat informasi fitur ciri bernegasi yang dapat membalikan orientasi fitur produk yaitu dari positif ke negatif atau dari negatif ke positif sehingga *classifier* masih sulit untuk mengkategorisasikan fitur produk yang bernegasi dan dapat menyebabkan kesalahan kategorisasi.

Pada tugas akhir ini menawarkan solusi untuk menangani masalah yang dipaparkan di atas yaitu dengan menggunakan pendekatan *Semantic role labeling* dan *Negation handling*. *Semantic role labeling* bertujuan untuk mengidentifikasi

hubungan semantik pada fitur produk yang memiliki makna implisit yang digunakan sebagai salah satu fitur ciri klasifikasi. *Negation handling* bertujuan untuk mengidentifikasi kata yang bernegasi yang digunakan sebagai salah satu fitur ciri klasifikasi. Penambahan fitur ciri *Semantic role labeling* dan *Negation handling* pada algoritma *Random forest* bertujuan untuk menambahkan informasi pemodelan pada klasifikasi agar algoritma *random forest* dapat lebih mudah mengklasifikasikan fitur prduk yang memiliki makna semantik dan negasi. Algoritma *random forest* memiliki akurasi baik pada analisis sentimen jika menggunakan karakteristik data unigram [21] seperti yang digunakan pada tugas akhir ini sebagai fitur produk. Maka diharapkan hasil klasifikasi ini dapat lebih akurat agar nantinya dapat dijadikan ringkasan yang lebih baik untuk meningkatkan ketepatan dalam penarikan kesimpulan dari ulasan produk.

1.2 Rumusan Masalah

Perumusan masalah berfokus kepada beberapa hal yaitu sebagai berikut:

1. Bagaimana mengekstrak informasi dari *semantic role labeling* dan *negation handling* agar dapat dijadikan fitur ciri pada klasifikasi?
2. Bagaimana akurasi klasifikasi setelah diterapkan *semantic role labeling* dan *negation handling*?

Adapun batasan masalah pada tugas akhir ini yaitu sebagai berikut:

1. Bahasa pemrograman yang digunakan adalah bahasa pemrograman java.
2. Dataset yang digunakan adalah ulasan produk berbahasa inggris dari data ulasan amazon.com yang digunakan pada penelitian Minqing Hu dan Bing Liu.
3. Hanya berfokus pada pengembangan klasifikasi fitur produk dari Tugas Akhir sebelumnya [8] yang berjudul “Analisis Sentimen dan Peringkasan Opini pada Ulasan Produk Menggunakan Algoritma Random Forest”.
4. Pada penerapan metode *semantic role labeling* digunakan *library Clear NLP* dan pada penerapan Algoritma *Random forest* digunakan *library Weka*.

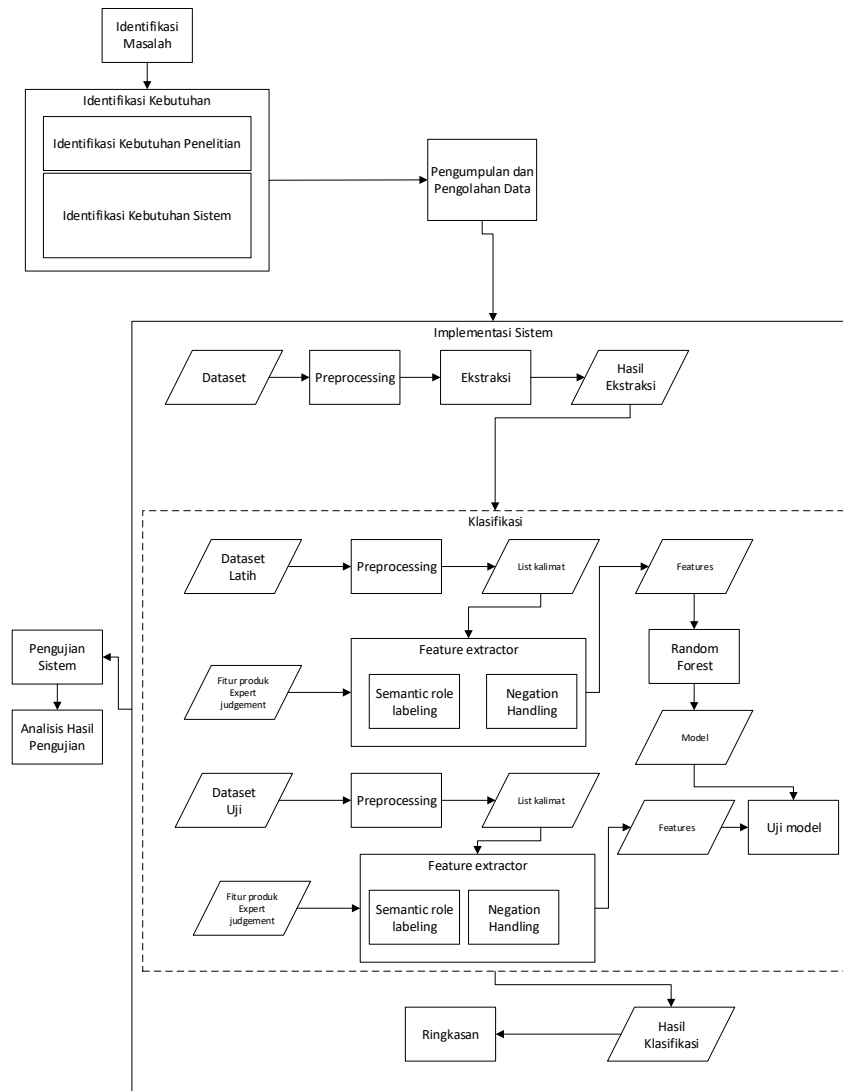
1.3 Tujuan

Terdapat beberapa tujuan dari tugas akhir ini yaitu sebagai berikut:

1. Menganalisis akurasi klasifikasi fitur produk setelah menggunakan semua fitur ciri.
2. Menganalisis akurasi klasifikasi setelah menambahkan fitur ciri dari *Negation handling* dan *semantic role labeling*.
3. Menganalisis pengaruh dataset yang memiliki karakteristik yang berbeda-beda terhadap *negation handling* dan *semantic role labeling*.

1.4 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah dalam penelitian ini adalah:



Gambar 1-1 Metodologi penelitian tugas akhir

1.4.1 Identifikasi Masalah

Masalah pada Tugas Akhir sebelumnya[8] seperti yang sudah dipaparkan pada latar belakang yaitu masih belum terdapat informasi hubungan semantik dari fitur produk belum terdapat informasi dari fitur produk yang bernegasi yang bisa menyebabkan kesalahan dalam menentukan orientasi dari fitur produk.

1.4.2 Identifikasi Kebutuhan

Terdapat dua kebutuhan pada tugas akhir ini yaitu:

1.4.2.1 Identifikasi kebutuhan penelitian

Identifikasi kebutuhan dari penelitian ini yaitu melakukan studi literatur dengan mempelajari tugas akhir sebelumnya[8], dari paper dan buku terkait yang dapat menangani masalah yang ditemukan pada tugas akhir yang sebelumnya[8]. Studi literatur juga dilakukan dengan dosen pembimbing dan dosen di kelas yang berhubungan dengan tugas akhir ini.

1.4.2.2 Identifikasi kebutuhan sistem

Identifikasi kebutuhan sistem adalah kebutuhan akan penggunaan metode dalam menyelesaikan masalah. Metode yang digunakan pada penelitian ini yaitu *Semantic role labeling* dan *Negation handling*. *Semantic role labeling* bertujuan untuk mengidentifikasi hubungan semantik dari fitur produk. *Negation handling* bertujuan untuk mengidentifikasi kata yang bernegasi yang dapat merubah polaritas dari ulasan. *Semantic role labeling* dan *Negation handling* akan dijadikan sebagai fitur ciri klasifikasi yang bertujuan untuk menambah informasi dari *classifier*.

1.4.3 Pengumpulan dan Pengolahan Data

Data yang digunakan dalam tugas akhir ini adalah melanjutkan dataset yang digunakan pada tugas akhir sebelumnya[8] yang diambil dari ulasan produk yang terdapat pada situs amazon.com yang digunakan pada penelitian Minqing Hu dan Bing Liu. Format data yang digunakan merupakan *file txt* yang berupa kalimat *review* beserta label fitur dan opini untuk setiap kalimatnya. Ulasan produk yang akan digunakan dalam tugas akhir ini terdapat 5 jenis , yaitu Apex DVD Player, Canon G3, Nikon Coolpix 4300, Nokia 6610, Zen Mp3 Player.

1.4.4 Implementasi Sistem

Implementasi sistem pada Tugas Akhir ini yaitu melakukan perancangan sistem yang akan di paparkan pada bab 3. Implementasi meliputi dalam pengembangan klasifikasi dari tugas akhirnya sebelumnya[8] dengan menggunakan beberapa metode yang diusulkan. Pada tugas akhir ini di fokuskan pada proses klasifikasi dan akan mengembangkan dari proses klasifikasi sebelumnya[8]. Pada tahap awal yaitu sistem membaca dataset dan melakukan pengekstrakan fitur produk dengan *expert judgment* dan dilakukan pengekstrakan label dari fitur produk. Dataset yang sudah di *tokenize* dijadikan sebagai inputan *semantic role labeling*. *Semantic role labeling* untuk mengidentifikasi hubungan semantik dari setiap kata. *Dependency parser* yang digunakan dari *semantic role labeling* akan digunakan juga pada *Negation handling* untuk mengidentifikasi kata yang bernegasi. Setelah melakukan beberapa metode tersebut akan diekstrak informasi dari setiap metode yang dapat dijadikan sebagai fitur ciri klasifikasi yang akan di proses oleh algoritma *random forest*.

1.4.5 Pengujian Sistem

Pengujian yang dilakukan yaitu dengan menguji akurasi setiap atribut pada klasifikasi dan pengujian terhadap akurasi dari klasifikasi setelah menggunakan beberapa metode meliputi *semantic role labeling* dan *negation handling*. Pengujian juga dilakukan pada semua dataset untuk mengetahui akurasi masing-masing dari klasifikasi.

1.4.6 Analisis Pengujian Sistem

Analisis yang dilakukan meliputi dari hasil pengujian yang dilakukan yaitu menganalisis pengaruh dari setiap atribut yang digunakan dan menganalisis hasil setelah diterapkan *semantic role labeling* dan *Negation Handling* sebagai fitur ciri klasifikasi. Menganalisis pengaruh dari dataset yang memiliki karakteristik yang berbeda-beda.

1.5 Sistematika Penulisan

Tugas akhir yang berjudul “Pengembangan Klasifikasi Opini Pada Ulasan Produk Menggunakan Algoritma *Random Forest*” disusun dengan sistematika penulisan sebagai berikut:

1. Pendahuluan

Pada Bab 1, menjelaskan latar belakang penulisan dan masalah yang diangkat, perumusan masalah, tujuan, batasan masalah, metologi penyelesaian masalah serta sistematika penulisan dari tugas akhir ini.

2. Tinjauan Pustaka

Pada Bab 2, menjelaskan landasan teori yang mendukung dan terkait dalam tugas akhir ini.

3. Perancangan Sistem

Pada Bab 3, berisikan gambaran umum sistem yang akan dibangun, langkah-langkah pengumpulan data, tahapan setiap proses dalam *semantic role labeling*, *negation handling* dan klasifikasi *random forest*.

4. Pengujian dan Analisis

Pada Bab 4, memuat hasil analisis terhadap data yang sudah dilakukan klasifikasi dengan melakukan perhitungan berupa akurasi.

5. Kesimpulan dan Saran

Pada Bab 5, menjelaskan kesimpulan hasil dari penelitian yang sudah dilakukan dan saran untuk pengembang penelitian ini.