

# 1. Pendahuluan

## 1.1. Latar Belakang

*Online shop* adalah salah satu perwujudan kemudahan teknologi internet, dimana penjual dan pembeli dapat bertransaksi melalui internet. Survei yang dirilis JakPat membuktikan 90% konsumen membeli barang berdasarkan *review* dan hampir 99% konsumen membeli barang dengan membaca setiap *review* produk dari orang lain [1]. Berdasarkan survei tersebut dapat disimpulkan bahwa kebebasan dalam memberikan ulasan pada media membuat pengunjung web kebingungan dan kesulitan dalam mengambil keputusan akan produk tersebut karena mempunyai opini yang bermacam-macam.

Fitur dari suatu produk banyak dikomentari oleh konsumen dalam ulasan opini pada *online shop*. Oleh karena itu pada *sentiment analysis*, proses ekstraksi fitur produk perlu dilakukan untuk menyaring setiap fitur-fitur produk yang disampaikan. Ekstraksi fitur pada ulasan produk sudah banyak dilakukan dan dikembangkan oleh ilmuwan diluar sana. Salah satu metode yang digunakan untuk melakukan ekstraksi fitur adalah *Frequent Itemset Mining* menggunakan Algoritma Apriori.

Berdasarkan penelitian sebelumnya [2] menggunakan metode *Frequent Itemset Mining* pada Algoritma Apriori dan *pruning* untuk ekstraksi fitur dari ulasan produk, hasilnya masih terdapat fitur produk yang belum terekstrak, sehingga mempengaruhi performansi yang diperoleh hanya berkisar 20%-40% saja. Penambahan proses *pruning* yang dilakukan juga tidak memberikan dampak yang cukup signifikan, performansi yang dihasilkan hanya bertambah berkisar 3%-10%. Hal ini terjadi karena beberapa faktor seperti tidak memperhatikan skema *preprocessing*, pengambilan *feature expert judgement* pada sistem tidak terambil semua, atau kesalahan masukan saat proses ekstraksi. Faktor-faktor inilah yang mungkin menyebabkan performansi pada penelitian sebelumnya [2] masih minim. Berdasarkan permasalahan di atas, perlu dilakukan perbaikan dan pengembangan ekstraksi fitur produk pada analisis sebelumnya [2] guna menganalisis apakah

metode baru yang ditambahkan pada *Frequent Itemset Mining* menggunakan Algoritma Apriori ini berpengaruh terhadap hasil ekstraksi fitur produk.

Pada tugas akhir ini akan dilakukan penanganan *Coreference Resolution* dan *IOB-Chunking* di *preprocessing*. Kemudian selain metode *Frequent Itemset Mining* dan *pruning* yang dilakukan pada ekstraksi fitur produk, akan dilakukan ekstraksi fitur yang teridentifikasi memiliki opini dari hasil fitur setelah *pruning*. Setelah itu hasil kandidat fitur tersebut akan diekstraksi kembali untuk melihat taksonomi atau hirarki dari fitur, yang bertujuan agar kelompok-kelompok fitur dari fitur tersebut juga dapat terekstrak. Diharapkan dengan adanya penambahan *preprocessing* dan ekstraksi fitur opini serta taksonomi ini dapat menangani permasalahan pada analisis sebelumnya [2] dan dapat meningkatkan performansi.

## 1.2. Perumusan Masalah

Pada tugas akhir ini memiliki beberapa batasan masalah seperti; berfokus untuk menganalisis proses ekstraksi fitur produk dari tugas akhir sebelumnya [2] yang menggunakan *Frequent Itemset Mining* pada *Association Mining* dengan Algoritma Apriori. Data ulasan produk berupa produk elektronik dalam bahasa inggris dari amazon.com yang sudah pernah diteliti oleh *Minqing Hu and Bing Liu* dalam *paper*-nya "*Mining and Summarizing Customers Reviews*" [3]. Sistem yang dibangun pada tugas akhir ini menggunakan pemrograman java. Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, terdapat beberapa rumusan masalah sebagai berikut:

- a) Bagaimana hasil performansi dari penambahan kombinasi *Coreference Resolution* dan *IOB-Chunking* terhadap ekstraksi fitur menggunakan *Frequent Itemset Mining* pada Algoritma Apriori dan *Pruning*?
- b) Bagaimana hasil ekstraksi fitur menggunakan *Frequent Itemset Mining* pada Algoritma Apriori dan ekstraksi *pruning*?
- c) Bagaimana hasil ekstraksi fitur yang teridentifikasi memiliki opini dari hasil ekstraksi fitur *pruning* yang sudah dilakukan, dan bagaimana hasil ekstraksi fitur taksonomi dengan memanfaatkan *lexicon Wordnet 2.1* dari hasil fitur yang memiliki opini serta *feature expert judgement*?

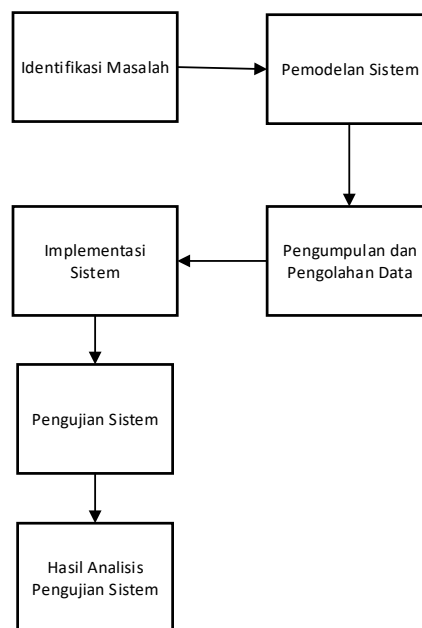
### 1.3. Tujuan

Berdasarkan rumusan masalah yang sudah dideskripsikan di atas dapat disimpulkan bahwa tugas akhir ini memiliki beberapa tujuan, sebagai berikut :

- a) Menganalisis hasil performansi dari penambahan kombinasi *Coreference Resolution* dan *IOB-Chunking* terhadap ekstraksi fitur menggunakan *Frequent Itemset Mining* pada Algoritma Apriori dan *Pruning*.
- b) Menganalisis hasil performansi ekstraksi *Frequent Itemset Mining* menggunakan Algoritma Apriori dan *Pruning* dengan beberapa masukkan *minimum support* serta parameter *compactness pruning* dan *redundancy pruning*.
- c) Menganalisis hasil performansi ekstraksi fitur yang teridentifikasi memiliki opini dari hasil ekstraksi fitur *pruning*, dan menganalisis hasil ekstraksi fitur taksonomi dengan memanfaatkan *lexicon Wordnet 2.1* dari hasil fitur yang memiliki opini serta *feature expert judgement*.

### 1.4. Metodologi Penyelesaian Masalah

Pada bagian ini akan menggambarkan bagan alir beserta penjelasan dari metodologi penyelesaian masalah tugas akhir ini. Bagan alir dapat dilihat pada gambar 1-1 :



Gambar 1-1 Alur Bagan Metodologi Penelitian.

Gambar 1-1 menggambarkan alur proses penelitian pada tugas akhir ini. Dibawah ini terdapat penjabaran dari alur bagan metodologi penelitian tersebut, seperti :

#### **1.4.1. Identifikasi Masalah**

Berdasarkan permasalahan ekstraksi fitur produk pada analisis sebelumnya [2] yang sudah dirumuskan pada sub bab rumusan masalah di atas. Maka pada tugas akhir ini dilakukan perbaikan dan pengembangan dari penelitian sebelumnya [2]. Tujuannya adalah untuk meninjau pengaruh dari metode tersebut. Informasi dikutip dari *paper*, *e-book*, *slide*, buku dan bahan lain yang dapat menunjang pembuatan tugas akhir ini

#### **1.4.2. Pemodelan Sistem**

Pada tugas akhir ini dilakukan pemodelan sistem yang bertujuan untuk memberikan gambaran sistem yang dibangun. *Preprocessing* yang akan diterapkan pada tugas akhir ini adalah *Coreference Resolution*, *Lemmatization*, *Stopword Removal*, *POS-Tagging*, dan *IOB Chunking*. Kemudian pada proses ekstraksi fitur diterapkan metode *Frequent Itemset Mining* menggunakan Algoritma Apriori, *Pruning*, ekstraksi kandidat fitur opini, dan *Taxonomy-Based*.

#### **1.4.3. Pengumpulan dan Pengolahan Data**

Tahap pengumpulan *resource* adalah tahap yang harus dilakukan agar tercapainya suatu penelitian ilmiah tersebut. Tugas akhir ini memanfaatkan 6 dataset yang berasal dari *paper* Minqing Hu dan Bing Liu [3]. Selain dataset, tugas akhir ini menggunakan daftar *Stopword* pada proses *Stopword Removal* dan kamus *Wordnet 2.1* pada proses *Taxonomy-Based*. Pada tabel 1-1 terdapat informasi mengenai dataset yang digunakan.

Tabel 1-1 Informasi Dataset.

No	Dataset	Total Sentence
1	Apex AD2600	739
2	Canon G3	597
3	Canon PowerShot SD500	229
4	Canon S100	298
5	Diaper Champ	375
6	Hitachi router	312

Selain dataset di atas, tugas akhir ini memanfaatkan beberapa *library* pada sistem yang dibangun dengan bahasa pemrograman java sebagai pendukung dalam mengimplementasikan sistem. Pada tabel 1-2 terdapat informasi mengenai *library* yang digunakan.

Tabel 1-2 Informasi Library.

No	Library	Proses
1	Library Stanford NLP	Coreference Resolution, Lemmatization, POS-Tagging
2	Library Illinois	IOB Chunking
3	Library SPMF	Ekstraksi fitur apriori

#### 1.4.4. Implementasi Sistem

Pada tahap ini, menerapkan metode pada sistem yang akan dibangun menggunakan bahasa pemrograman java berdasarkan permasalahan pada penelitian sebelumnya yang sudah dijelaskan di sub bab rumusan masalah. Untuk penjelasan proses implementasi sistem pada tugas akhir ini dapat dilihat pada bab 3 selanjutnya.

#### 1.4.5. Pengujian Sistem

Pada tahap ini akan dilakukan pengujian sistem terhadap 6 dataset yang berasal dari *paper* Minqing Hu dan Bing Liu [3]. Pengukuran pada pengujian ini ditinjau dari hasil *Recall*, *Precision*, dan *F1-Score* yang dihasilkan untuk setiap metode yang diterapkan.

### 1.5. Sistematika Penulisan

Untuk lebih jelas memahami laporan tugas akhir ini, berikut terdapat gambaran singkat mengenai informasi setiap bab dalam laporan tugas akhir ini.

## BAB I Pendahuluan

Pada bagian ini berisi tentang latar belakang, perumusan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan.

## BAB II Kajian Pustaka

Pada bagian ini berisikan tentang teori-teori yang digunakan pada analisis ini, berupa pengertian, definisi yang diambil dari kutipan buku, jurnal, *e-book*, serta beberapa literatur *review* yang berhubungan dengan penelitian.

## BAB III Perancangan Sistem

Pada bagian ini menjelaskan tentang proses sistem yang diusulkan, dan detail tentang proses serta hasil kerja dari sistem yang di implementasikan. Kemudian dijabarkan secara satu persatu dengan menerapkan konsep sesudah adanya sistem yang diusulkan.

## BAB IV Pengujian dan Analisis

Pada bagian ini berisikan hasil pengujian dan analisis dari pengujian yang sudah dilakukan. Dijelaskan secara detail satu persatu berdasarkan proses yang telah diuji.

## BAB V Kesimpulan dan Saran

Pada bagian ini berisi kesimpulan dan saran yang berkaitan dengan analisis berdasarkan yang telah diuraikan pada bab-bab sebelumnya.