

# 1 Pendahuluan

## 1.1 Latar belakang

Seiring dengan bertambah pesatnya sejumlah informasi yang tersimpan dalam *text database* maka semakin besar usaha manusia untuk menambang kembali informasi. Teks data mining atau lebih dikenal dengan *teks mining* menjadi ‘perkakas’ penting untuk pertambangan pengetahuan di dalam teks. Salah satu tipe *teks mining* yang cukup populer adalah *keyword based association analysis*. Motivasi dari *keyword based association analysis* ini adalah mengoleksi kumpulan *keywords* atau *term* yang sering terjadi bersamaan dan kemudian menemukan hubungan korelasi atau asosiasi diantara kumpulan *keyword* tersebut. Sesuai dengan motivasinya, penerapan *keyword based association analysis* akan sangat bermanfaat dalam otomatisasi penandaan dokumen terkait keberadaan suatu frase atau istilah yang ada pada dokumen dengan cepat dan tepat. Untuk mengimplementasikan teknik ini terdapat dua sub proses yaitu *preprocessing* dokumen dan *generate rule* asosiasi.

Pada proses *generate rule* asosiasi terdapat dua parameter batasan yang dispesifikasikan *user*, yaitu *minsup* dan *minconf*. *Minsup* digunakan untuk mengontrol term-term yang terlibat dalam pembentukan rule, sedangkan *minconf* mengontrol kekuatan prediksi pada rule yang terbentuk. *Generate rule* asosiasi yang sangat populer saat ini pada umumnya menggunakan *single minsup*. Sehingga semua *term* yang terlibat dianggap memiliki sifat dan frekuensi yang sama. Padahal kasus tersebut sangat jarang terjadi di dunia nyata. Kebanyakan kasus di dunia nyata *term-term* yang terlibat memiliki sifat dan frekuensi yang bervariasi. Pendekatan yang agar variasi frekuensi *term* yang ada pada sekumpulan dokumen dapat ter-cover, penulis mencoba mengimplementasikan *keyword based association analysis* yang menerapkan *multiple minimum support*. Dengan menerapkan *multiple minimum support* setiap *rule* yang dihasilkan memiliki batas *minsup* yang berbeda-beda sesuai dengan *item* yang terlibat dalam *rule* tersebut.

## 1.2 Perumusan masalah

Dalam pengerjaan Tugas Akhir ini, terdapat beberapa rumusan masalah yang ingin diselesaikan:

- a. Bagaimana mencari aturan asosiasi pada sebuah dokumen
- b. Bagaimana menemukan aturan asosiasi dengan algoritma MS Apriori
- c. Bagaimana membangun sistem berdasarkan Algoritma MS Apriori untuk melakukan *keyword based association analysis* pada dokumen

- d. Bagaimana performansi penerapan algoritma MS Apriori terhadap pencarian aturan asosiasi pada dokumen

Adapun batasan masalah pada tugas akhir ini yaitu:

Data transaksi yang digunakan merupakan kumpulan dokumen bahasa Indonesia dengan asumsi data bersih.

### 1.3 Tujuan

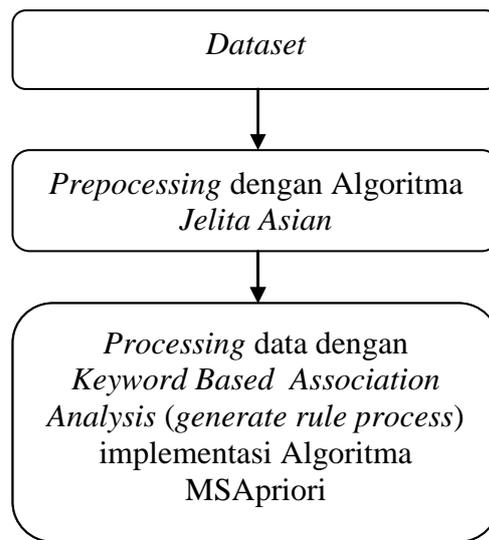
Tujuan dari Tugas Akhir ini adalah :

1. Mengukur efisiensi penerapan *multiple minimum support* dengan algoritma MS Apriori dalam menemukan aturan asosiasi pada proses *generate rule keyword based association analysis*
2. Mengukur efektifitas penerapan *multiple minimum support* dengan algoritma MS Apriori dalam menemukan *frequent itemset* dan aturan asosiasi pada proses *generate rule keyword based association analysis*

### 1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam tugas akhir ini adalah :

- a. Studi literatur  
Mencari dan mengumpulkan informasi serta memahami dan mempelajari tentang metode *Keyword based association analysis*, Algoritma MS Apriori, serta konsep *Text Mining* melalui literatur berupa makalah, buku atau jurnal.
- b. Pencarian dan pengumpulan data.  
Data yang akan digunakan berupa artikel berbahasa Indonesia yang diperoleh dari web.
- c. Analisis dan Perancangan Perangkat Lunak  
Pada tahap ini akan dilakukan analisis terhadap perangkat lunak yang diperlukan dalam mengimplementasikan metode *Keyword based association analysis* . Selain itu juga dilakukan pemilihan model untuk perangkat lunak dan bahasa pemrograman yang digunakan pada tahap perancangan.



**Gambar 1.1 Bagan Alur Proses Pembuatan Sistem**

- d. Implementasi dan *Prototyping*  
Setelah perancangan, dilakukan implementasi dan *Prototyping* untuk membangun aplikasi generate rule
- e. Pengujian  
Pada tahap ini dilakukan pengujian terhadap hasil implementasi dan *prototyping* untuk meminimalisir adanya *bug/error*. Jika *bug/error* ditemukan, maka dilakukan implementasi dan *prototyping* ulang.
- f. Analisis Hasil Implementasi  
Pada tugas akhir ini dilakukan analisa aturan asosiasi untuk mendapatkan *rule* yang paling optimal untuk ditanamkan pada sistem.
- g. Pembuatan laporan Tugas Akhir  
Pada tahap akhir, dilakukan pembuatan dokumentasi yang berupa laporan Tugas Akhir.