

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penggalian kaidah asosiasi (*mining association rules*) merupakan salah satu proses *data mining* yang mampu menemukan pola dan aturan (*rule*) dari sekumpulan data yang besar. Pada manajemen bisnis proses ini digunakan untuk menemukan pola-pola dan perilaku pelanggan dalam melakukan suatu transaksi dimana pola-pola ini merupakan kumpulan item (*itemset*) yang sering dibeli oleh pelanggan. Dengan mengetahui pola tersebut, proses manajemen bisnis menjadi lebih mudah.

Proses pencarian item-item yang sering dibeli atau sering muncul dalam transaksi (*frequent itemset*) pada data yang besar membutuhkan waktu yang sangat lama, dan hal ini merupakan syarat untuk membentuk kaidah asosiasi. Dari kondisi tersebut, diperlukan suatu algoritma yang bisa meminimalisasi pembacaan basis data, sehingga bisa mengoptimasi waktu yang dibutuhkan.

Algoritma yang paling populer saat ini adalah algoritma Apriori yang menggunakan prinsip Apriori, yaitu jika suatu *itemset* merupakan *frequent* (yang sering muncul), maka semua *subset*-nya akan berupa *frequent*. Algoritma ini merupakan pionir dalam menggunakan *support-base pruning* untuk mengontrol pertumbuhan eksponensial dari kandidat *itemset* secara sistematis. Proses pada algoritma ini membangkitkan *frequent itemset* per level, dimulai dari level 1-*itemset* sampai ke *longest itemset*, kandidat level yang baru dibentuk dari *frequent itemset* yang ditemukan di level sebelumnya lalu menentukan nilai *support*-nya.

Kelemahan yang paling kritis pada Apriori adalah bahwa ketika kardinalitas pada *longest frequent itemset* berupa k , Apriori membutuhkan sebanyak k pembacaan basis data. Hal lainnya adalah algoritma Apriori memiliki sifat *computation-intensive* dalam membangkitkan kandidat *itemset* dan

penghitungan nilai *support*, khususnya untuk aplikasi yang memiliki nilai *support* yang sangat rendah dan atau item yang sangat banyak.

Ada banyak macam algoritma yang sudah diajukan untuk memperbaiki efisiensi, termasuk DHP, Partition, DIC, Eclat, Top-down, FP-growth, dan yang lainnya. Walaupun semua macam algoritma ini mengadopsi teknik yang berbeda dan sudut pandang yang berbeda, tapi dapat dikelompokkan menjadi tiga aspek yang berbeda, yaitu :

1. Strategi arah pencarian: *top-down* dan *bottom-up*
2. Strategi pencarian: *breadth first search (BFS)* dan *Depth first search (DFS)*
3. Strategi penghitungan : *vertical intersection* dan *horizontal counting*

Algoritma *Cut Both Ways (CBW)* menggunakan gabungan dari beberapa teknik ini. Kebanyakan algoritma yang mirip dengan algoritma Apriori, lebih menitikberatkan pada pembatasan dengan nilai *support* minimal untuk membuang sejumlah besar *infrequent itemset*, hal inilah yang menjadi motivasi dibentuknya algoritma CBW. Oleh karena itu algoritma ini menggunakan *cutting level* (α) untuk membagi ruang pencarian menjadi dua bagian. Nilai *cutting level* merupakan nilai rata-rata dari kardinalitas *frequent itemset*, yang diharapkan pada level ini akan ditemukan banyak *frequent itemset*.

Setelah menentukan nilai *cutting level*, algoritma ini melakukan pencarian ke bagian bawah untuk menemukan *frequent itemset* yang nilai kardinalitasnya kurang dari dan sama dengan nilai *cutting level* dan menentukan nilai *support*-nya. Selanjutnya melakukan pencarian ke bagian atas untuk menentukan semua *frequent itemset* dengan kardinalitas lebih dari α dengan mengikuti paradigma Apriori (membangkitkan *frequent itemset* per level) dan menggunakan *vertical intersection* untuk mengurangi penghitungan nilai *support* yang tidak perlu.

Permasalahan efisiensi dalam pencarian *frequent itemset* adalah proses minimalisasi pembacaan basis data dan penghitungan nilai *support*. Algoritma CBW menggunakan nilai *cutting level* untuk membagi ruang pencarian dan menggunakan *vertical intersection* untuk menentukan nilai *support* pada proses pencarian *bottom-up*.

1.2 Perumusan Masalah

Pencarian *frequent itemset* merupakan syarat dalam penggalian kaidah asosiasi dan memerlukan banyak waktu. Pada beberapa kasus, algoritma Apriori kurang baik digunakan untuk pencarian *frequent itemset* terutama untuk basis data yang besar, item yang sangat banyak dan nilai minimum *support* yang kecil. Algoritma CBW menggabungkan beberapa teknik dan membagi ruang pencarian menjadi dua bagian dengan nilai *cutting level*, sehingga memungkinkan algoritma ini lebih baik dibanding algoritma Apriori walau dengan kondisi yang sebelumnya disebutkan, tetapi mungkin saja algoritma Apriori lebih baik digunakan untuk keadaan yang sebaliknya.

1.3 Tujuan

Tujuan dari tugas akhir ini adalah untuk mengimplementasikan algoritma Apriori dan algoritma CBW untuk proses pencarian *frequent itemset*. Selanjutnya membandingkan kinerjanya untuk menentukan kelebihan dan kekurangannya terhadap kondisi data, sekaligus membuktikan efisiensi yang dihasilkan dari penggabungan beberapa teknik yang digunakan dan penggunaan nilai *cutting level*.

1.4 Batasan Masalah

Batasan permasalahan dalam tugas akhir ini antara lain :

1. Proses yang dilakukan sampai ditemukan *frequent itemset* maksimal, tidak dilanjutkan dengan membentuk kaidah asosiasi.
2. Jenis data yang digunakan adalah data transaksi yang telah melalui proses *data cleaning*.
3. Analisis kompleksitas algoritma hanya dilakukan untuk kompleksitas waktu.

1.5 Metodologi

Metodologi yang digunakan untuk menyelesaikan tugas akhir ini adalah sebagai berikut :

1. Studi Pustaka

Pada tahap ini yang dilakukan adalah mencari, mengumpulkan dan mempelajari segala macam informasi yang berhubungan dengan *data mining*, *association rules*, algoritma Apriori, teknik-teknik pencarian *frequent itemset* dan algoritma CBW.

2. Perancangan Perangkat Lunak

Pada tahap ini dilakukan perancangan perangkat lunak, dimana ada 3 perancangan yang dibuat yaitu perancangan data, perancangan proses, dan perancangan antarmuka.

3. Pembuatan Program

Pada pembuatan program dilakukan implementasi dari tahap sebelumnya. Pembuatan perangkat lunak untuk tugas akhir ini menggunakan bahasa pemrograman C#.Net .Sistem operasi yang digunakan adalah Windows XP. Basis data yang digunakan adalah SQL Server 2000.

4. Uji Coba dan Evaluasi

Pada tahap ini dilakukan uji coba dari program yang telah dibuat dengan menggunakan data yang telah dipersiapkan sebelumnya. Selanjutnya, hasil dari pengujian program akan dievaluasi untuk menentukan kebenaran dan kehandalan.

5. Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan dari mulai tahap studi literatur sampai tahap pengujian dari evaluasi.

1.6 Sistematika Penulisan

Sistematika penulisan dibagi menjadi :

BAB I PENDAHULUAN

Pada bab ini diuraikan latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi serta sistematika penulisan.

BAB II DASAR TEORI

Pada bab ini dibahas teori umum yang meliputi definisi dan uraian singkat tentang *Data mining* dan *Association rule*, uraian

singkat dari algoritma Apriori, algoritma CBW dan teknik-teknik pencarian *frequent itemset* yang digunakan.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi analisis algoritma Apriori dan algoritma CBW dari segi teori dan kompleksitas, analisis dan perancangan perangkat lunak sebagai alat bantu dalam proses analisa.

BAB IV IMPLEMENTASI DAN HASIL PENGUJIAN

Bab ini memuat tentang analisis terhadap performansi algoritma Apriori dan algoritma CBW, membandingkan kinerja kedua algoritma berdasarkan parameter yang ada.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diambil dari pembahasan bab-bab sebelumnya serta saran-saran untuk pengembangan selanjutnya.