

BAB 1

PENDAHULUAN

Saat ini, kebutuhan manusia akan informasi sangat besar. Informasi yang penting harus disimpan karena bisa saja digunakan untuk masa yang akan datang. Sebagai contoh informasi penjualan produk sebuah perusahaan. Tiap tahun bahkan tiap bulan, informasi mengenai jumlah penjualan produk maupun wilayahnya harus disimpan dan jumlahnya sangat besar. Untuk menyimpan data yang jumlahnya sangat besar diperlukan suatu sistem yang baik. Sistem ini dinamakan *database*. *Database* tidak hanya mampu menyimpan data dengan jumlah besar namun juga mampu membantu kita untuk memperoleh data dengan mudah jika sewaktu-waktu dibutuhkan.

Namun demikian, *database* tidak bisa digunakan untuk menjawab pertanyaan seperti ”apakah ada pengaruh penjualan pasta gigi terhadap sikat gigi?”. Oleh sebab itu, untuk menjawabnya dibutuhkan *data mining*. *Data mining* merupakan suatu proses iterasi menganalisa data untuk mencari informasi (Kantardzic, 2003). Seperti contoh diatas, *data mining* dapat digunakan untuk mengetahui hubungan antara penjualan pasta gigi dengan sikat gigi.

Teknik-teknik dalam *data mining* sangat banyak dan memiliki keunggulan masing-masing. Penggunaan teknik tersebut disesuaikan dengan kebutuhan informasi yang ingin didapat. Pada perkembangan selanjutnya, setiap teknik bermunculan algoritma-algoritma sesuai dengan ciri teknik tersebut. Algoritma-algoritma ini berkembang seiring kebutuhan akan proses dalam *data mining* yang efektif dan efisien.

Pada bab ini akan dijelaskan mengenai latar belakang, tujuan penelitian, ruang lingkup tugas akhir, metode penelitian dan sistematika penulisan laporan.

1.1 Latar Belakang

Dalam setiap transaksi, baik pembelian maupun penjualan suatu barang atau jasa, selalu ada data yang disimpan. Data tersebut berisi informasi mengenai

setiap transaksi yang terjadi setiap waktu. Data tersebut hanya berisi informasi tertentu, sebagai contoh dalam penjualan barang berupa nama barang, jumlah yang terjual dan harga barang.

Suatu perusahaan dalam waktu-waktu tertentu selalu memiliki rencana untuk meningkatkan penjualan barang atau jasa miliknya. Setiap rencana selalu memerlukan pertimbangan atau analisa mengenai penjualannya. Analisa tersebut salah satunya didapat dari data-data penjualan yang telah terjadi. Sebagai contoh, suatu perusahaan ingin menjual barang hanya pada daerah yang penjualannya diatas 100.000 barang tiap tahun untuk mengurangi biaya pengiriman, maka perusahaan tersebut dapat mengetahuinya dari *database* penjualan.

Seperti yang telah dijelaskan sebelumnya, ada beberapa pertanyaan yang tidak dapat diperoleh hanya dengan melihat *database* saja. Sebagai contoh, bagaimana jika perusahaan ingin mengetahui hubungan dari suatu barang dengan barang yang lain, seperti apakah dalam pembelian sikat gigi selalu terjadi pembelian pasta gigi, atau apakah dalam setiap pembelian gula selalu ada pembelian kopi. Untuk mengetahui jawaban dari pertanyaan tersebut, maka digunakanlah *data mining*.

Data mining biasanya terdapat pada *business intelligent tools*, dimana *data mining* merupakan bagian yang penting untuk mengolah data. Walaupun saat ini banyak sekali aplikasi *business intelligent tools*, namun sedikit aplikasi yang hanya berupa *tools data mining*.

Salah satu *tools data mining* yang banyak digunakan dan *open source* yaitu WEKA (*Waikato Environment for Knowledge Analysis*). WEKA dibuat oleh universitas Waikato, New Zealand, yang mengimplementasikan teknik-teknik *data mining*, antara lain *Classification*, *clustering* dan *Association*. Algoritma-algoritma yang diimplementasikan dalam WEKA antara lain NaiveBayes, ZeroR (*Classification*), EM, *Cobweb* (*clustering*), Apriori, Tertius (*Association*). Namun, algoritma-algoritma terbaru dan populer tidak diimplementasikan dalam

WEKA seperti *Iterate*, *FP Growth*, *CMAR*. Algoritma-algoritma tersebut diimplementasikan dengan menggunakan bahasa Java.

Fakultas Ilmu Komputer Universitas Indonesia (Fasilkom UI) melalui RUUI 2007, mengembangkan *data mining Algorithms Collection* sebagai *workbench data mining* dengan mengimplementasi algoritma-algoritma menggunakan bahasa pemrograman C/C++ yang diharapkan lebih cepat dan efisien dibandingkan Java.

Tugas Akhir ini merupakan salah satu bagian dalam *data mining Algorithms Collection*, dimana penulis mengimplementasikan 2 buah algoritma yang menggunakan teknik *clustering*. Algoritma tersebut yaitu algoritma *Iterate* dan *Cobweb*. Kedua algoritma tersebut diimplementasikan dengan menggunakan bahasa pemrograman Visual C++.

Algoritma *Cobweb*, yang merupakan salah satu algoritma *clustering* yang banyak dikenal, dikemukakan oleh Douglas H. Fisher pada tahun 1987. Algoritma ini menggunakan *incremental hierarchy conceptual clustering* yaitu terus-menerus (*incremental*) mengobservasi obyek-obyek ke dalam klasifikasi *tree*. Dalam klasifikasi *tree* tersebut, setiap *node* merupakan sebuah kelas (*concept*) yang memiliki probabilitas kelas dan *attribute-attribute* dari obyek-obyek yang disimpannya (Fisher, 1987).

Algoritma *Iterate* dikemukakan oleh Gautam Biswas, Jerry B. Weinberg, dan C. Li pada tahun 1995. Algoritma ini juga menggunakan klasifikasi *tree*, namun tidak seperti *Cobweb* yang menggunakan proses *merge*, *split* dan *new node* untuk memperbaiki klasifikasi *tree*, *Iterate* mengumpulkan *node-node* yang nilainya rendah untuk memperbaiki *tree* dan menggunakan fungsi ADO untuk mengurutkan obyek sebelum klasifikasi *tree* dibentuk.

1.2 Tujuan penelitian

Tujuan khusus pelaksanaan dari Tugas Akhir ini adalah untuk melakukan analisa dan mengimplementasikan algoritma *Cobweb* dan *Iterate* sebagai bagian dari *data mining Algorithms Collection*, serta melakukan uji coba untuk membandingkan kualitas dan waktu eksekusi implementasi algoritma *Cobweb*

hasil implementasi dengan algoritma *Cobweb* pada WEKA dan implementasi algoritma *Iterate*. Selain itu, tujuan umum dari Tugas Akhir ini adalah untuk memenuhi syarat memperoleh gelar sarjana S1 Ilmu Komputer serta agar penulis dapat menerapkan ilmu yang didapat dalam perkuliahan.

1.3 Ruang Lingkup Tugas Akhir

Tugas akhir ini meliputi analisa dan implementasi algoritma *Cobweb* dan *Iterate*. Pada tahap implementasi program, dilakukan implementasi dengan menggunakan MFC (*Microsoft Foundation Class*) dan bahasa pemrograman Visual C++ pada Framework Visual Studio .NET 2005. Kedua algoritma tersebut diimplementasi dengan menggunakan sebuah komputer Pentium 4 2,00 GHz, memory 256 MB DDR.

Tahap analisa algoritma dilakukan untuk memahami algoritma *Cobweb* dan *Iterate*, melakukan perencanaan kebutuhan *input* dan *output* program, dan menganalisa kebutuhan struktur data.

1.4 Metode Penelitian

Metode yang digunakan dalam Laporan Tugas Akhir ini adalah studi literatur, analisa algoritma, penerapan algoritma dan percobaan. Literatur yang menjadi sumber penulisan dan implementasi algoritma *Cobweb* dan *Iterate* ini antara lain buku, situs *web*, dan berbagai makalah tentang *data mining*, *clustering*, *Cobweb* dan *Iterate*. Analisa penerapan dilakukan yaitu dengan membaca hasil analisa dari para ahli dan membandingkannya dengan hasil analisa yang didapat dari penelitian ini. Penerapan algoritma yang dilakukan penulis hanya pada algoritma *clustering* yaitu *Cobweb* dan *Iterate*. Pengujian yang dilakukan dengan menggunakan *dataset* kecil dahulu agar mudah dilakukan analisa sehingga dapat dilakukan perbandingan kualitas partisi. Pengujian selanjutnya menggunakan *dataset* dalam jumlah besar untuk membandingkan waktu eksekusi program.

1.5 Sistematika Penulisan Laporan

Sistematika penulisan laporan Tugas Akhir ini dibagi menjadi beberapa bagian, antara lain:

Bab 1: Pendahuluan

Pada bab ini dijelaskan mengenai latar belakang penulisan, tujuan penelitian, ruang lingkup Tugas Akhir, metode penelitian, dan sistematika penulisan laporan.

Bab 2: Landasan Teori

Pada bab ini dijelaskan mengenai landasan teori Tugas Akhir yaitu penjelasan mengenai teknik-teknik dalam *data mining*, penjelasan teknik *clustering* dan contoh algoritma *clustering*, penjelasan algoritma *Cobweb* yang terdiri dari fungsi evaluasi heuristik, representasi *state*, operator, strategi kontrol *Cobweb*, dan penjelasan langkah-langkah dalam algoritma *Iterate*.

Bab 3: Analisa dan Perencanaan

Pada bab ini dijelaskan mengenai analisa yang dibutuhkan dalam implementasi algoritma. Bab ini juga menjelaskan mengenai perencanaan langkah-langkah yang akan dilakukan dalam implementasi algoritma.

Bab 4: Implementasi Algoritma

Pada bab ini dijelaskan mengenai algoritma *Cobweb* dan *Iterate*, implementasi algoritma *Cobweb* dan *Iterate*.

Bab 5: Hasil Uji Coba Implementasi

Pada bab ini dijelaskan mengenai hasil uji coba yang menggunakan *dataset* kecil dan *dataset* besar serta analisa hasil uji coba.

Bab 6: Kesimpulan dan Saran

Bab ini berisikan kesimpulan dari Tugas Akhir serta saran dari penulis untuk penelitian *data mining* lebih lanjut dalam *clustering*, khususnya algoritma *Cobweb* dan *Iterate*.

