

ANALISIS DAN IMPLEMENTASI ALGORITMA CT-PRO UNTUK ASOSIASI PEMBELIAN BARANG

Parama Fadli Kurnia¹, Angelina Prima Kurniati², Arie Ardiyanti Suryani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Pertumbuhan data yang sangat cepat, hanya akan mengakibatkan kuburan data atau "data tombs" yang tidak bermanfaat karena data tersebut tidak dipergunakan untuk aplikasi yang berguna. Akumulasi dari pertumbuhan data tersebut, hanya akan menciptakan kondisi "rich of data but poor of information". Untuk mengatasi permasalahan tersebut, data mining merupakan solusinya. Dengan data mining dapat dilakukan pencarian informasi dari kuburan data itu dan menjadikannya knowledge. Knowledge tersebut bisa dipergunakan oleh para pembuat keputusan, contohnya saja di sebuah perusahaan retail. Knowledge yang didapat bisa dipergunakan untuk salah satu tipe analisis data yang paling sering digunakan dalam dunia pemasaran yaitu metode market basket analysis. Hasil dari analisa tersebut adalah berupa pola-pola atau rules tentang kebiasaan konsumen dalam berbelanja. Informasi tersebut dapat menjadi bahan pertimbangan dalam menentukan kebijakan perusahaan. Untuk mendapatkan pola-pola atau rules tersebut, dibutuhkan algoritma untuk mencari frequent pattern.

Pada Tugas Akhir ini bertujuan untuk menemukan pola-pola rules tersebut dengan algoritma CT-Pro menggunakan dua nilai analisis penting yaitu minimum support dan minimum confidence. Setiap rules yang didapatkan berasal dari data transaksi yang dilakukan oleh konsumen dalam berbelanja. Algoritma CT-PRO memiliki tiga tahap yaitu (1) menemukan item-item yang frequent; (2) Membuat struktur data CFP-Tree; (3) Melakukan mining frequent patterns.

Berdasarkan hasil pengujian didapatkan knowledge terdiri dari item yang dibeli dengan item yang dibeli pula beserta nilai support dan confidencenya yang nantinya dapat digunakan oleh perusahaan. Selain itu, didapatkan pula akurasi dari setiap knowledge yang didapatkan.

Kata Kunci : asosiasi, support, confidence, CFP-Tree, CT-Pro

Abstract

Very rapid growth of data, it will only lead to grave data or "data tombs" are not useful because the data is not used for useful applications. The accumulation of data growth, it will only create the conditions "of a data rich but poor of information". To over come these problems, data mining is the solution. With data mining, you can mine information from cemetery data and making knowledge. Knowledge can be used by decision-makers, for example, just in a retail company. Knowledge gained can be used for any type of data analysis most often used in the world of marketing is the method of market basket analysis. The result of the analysis is in the form of patterns or rules about the shopping habits of consumers. Such information maybe material to consider in determining corporate policy. To get the patterns or rules, it takes the algorithm to find frequent patterns.

In this final project to discover patterns such rules with CT-Pro algorithm uses two important analytical values are minimum support and minimum confidence. Any rules can obtained from the data transactions made by consumers in shopping habits. CT-PRO algorithm has three stages: (1) finding the frequent items, (2) Make CFP-Tree data structure, (3) Conducting mining frequent patterns.

Based on the test results obtained knowledge consists of items purchased with your purchased item also its support and confidence value that can later be used by the company. In addition, also obtained accuracy of any knowledge gained.

Keywords : association, support, confidence, CFP-Tree, CT-Pro

1. Pendahuluan

1.1. Latar belakang masalah

Teknologi Informasi sekarang ini telah digunakan hampir di semua aspek kehidupan, contohnya dalam sebuah perusahaan ritel. Dengan sistem yang telah terkomputerisasi, sebuah perusahaan ritel dapat mengumpulkan data transaksi dengan cepat serta menghasilkan data yang sangat besar. Tetapi pertumbuhan data yang pesat itu, telah menciptakan kondisi “*rich of data, but poor of information*” (Iko Pramudiono, 2003:1) karena data yang terkumpul itu tidak digunakan untuk aplikasi yang berguna. Tidak jarang kumpulan data itu dibiarkan begitu saja seakan-akan menjadi “kuburan data” (*data tombs*). Padahal sebenarnya dapat dilakukan pencarian informasi-informasi dari data yang terkubur itu dan menjadikannya informasi penting bagi perusahaan untuk mendukung keputusan ataupun membantu dalam menentukan strategi pemasaran.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Salah satu teknik yang terkenal dalam *data mining* adalah *association rule mining*. *Association rule mining* adalah teknik *mining* untuk menemukan aturan asosiatif antara suatu kombinasi item (Iko Pramudiono, 2003). Dalam *association rule mining* terdapat beberapa algoritma diantaranya Apriori dan *Frequent Pattern-Growth* (FP-Growth). Dibandingkan dengan Apriori, FP-Growth merupakan algoritma yang sangat efisien dalam pencarian *frequent itemset*. Algoritma ini menyimpan informasi mengenai *frequent itemset* dalam bentuk struktur *prefix-tree* atau sering disebut *FP-tree*. Algoritma ini tidak melakukan *candidate generation* dalam proses pencarian *frequent itemset*, sehingga dapat mengurangi *scan* database secara berulang dalam proses *mining* dan dapat berlangsung lebih cepat. Walaupun kinerja algoritma FP-Growth cukup baik, namun jika terdapat database yang berukuran besar kurang efektif jika membangun *FP-tree* pada *main memory*. Maka dari itu, dipakai algoritma *CT-Pro* dengan struktur data *CFP-Tree* untuk memodifikasi struktur *FP-tree* ke dalam bentuk *Compressed FP-Tree* dimana informasi dari sebuah *FP-Tree* diringkas dengan struktur yang lebih kecil atau ringan, sehingga lebih cepat dalam pembentukan *tree* maupun melakukan *frequent itemset mining*.

Teknik *association rule mining* dapat diimplementasikan pada *market basket analysis*. *Market basket analysis* merupakan sebuah analisis terhadap kebiasaan konsumen dalam berbelanja, dimana hasil analisisnya adalah menemukan pola keterkaitan antar produk yang dibeli oleh konsumen dalam bentuk *rules*. Pola-pola atau *rules* yang diperoleh dapat dijadikan acuan untuk dijadikan pedoman penentuan tata letak produk yang memudahkan konsumen untuk mendapatkannya secara bersamaan, mengenali perilaku konsumen, membaca kecenderungan selera pasar di masa depan, serta untuk mendapatkan pasar baru.

1.2. Perumusan masalah

Berdasarkan latar belakang diatas, masalah yang akan ditangani pada Tugas Akhir ini adalah :

1. Bagaimana cara menerapkan Algoritma CT-Pro dengan struktur data *CFP-Tree* pada *association rule mining*?
2. Apa saja faktor-faktor yang mempengaruhi algoritma *CT-Pro*?
3. Bagaimana model Aplikasi berbasis komputer dengan menggunakan algoritma *CT-Pro* dengan struktur data *CFP-Tree*?

1.3. Batasan masalah

Adapun batasan masalah dalam Tugas Akhir ini adalah :

1. *Knowledge* yang dihasilkan berupa item yang dibeli dengan item yang dibeli pula beserta nilai *support* dan *confidencenya*.
2. Tidak dilakukan prediksi terhadap item yang akan dibeli pada masa depan.
3. Data yang digunakan berasal dari data transaksi penjualan butik busana muslim Siti Hajar Tren Islami yang terdiri dari 5 outlet.

1.4. Tujuan

Berdasarkan rumusan masalah diatas, tujuan dari Tugas Akhir ini adalah :

1. Menentukan *knowledge* berupa item yang dibeli dengan item yang dibeli pula beserta nilai *support* dan *confidencenya* menggunakan Algoritma CT-Pro dengan struktur data *CFP-Tree*
2. Menganalisa pengaruh nilai *support* dan *confidence* pada algoritma *CT-Pro*.
3. Membangun perangkat lunak untuk memanfaatkan data yang besar dengan menerapkan algoritma *CT-Pro* dengan struktur data *CFP-Tree* .

1.5. Metodologi penyelesaian masalah

Metodologi yang digunakan dalam memecahkan permasalahan-permasalahan dalam Tugas Akhir ini terdiri dari 6 tahap, yaitu:

1. Tahap Studi Literatur
Pada tahap ini, akan dilakukan membaca literatur yang ada dan mencari literatur tambahan yang dibutuhkan dalam pendalaman materi terhadap konsep dan teori *association rule mining* dengan algoritma *CT-Pro* dengan struktur data *CFP-Tree* .
2. Tahap Perancangan Sistem
Pada tahap ini, akan dilakukan perancangan sistem.
3. Tahap Implementasi
Pada tahap ini, akan dilakukan implementasi berdasarkan rancangan yang telah dibuat sebelumnya.
4. Tahap Pengujian
Pada tahap ini, akan dilakukan pengujian terhadap sistem.
5. Tahap Analisis Hasil Pengujian
Pada tahap ini dilakukan analisis berdasarkan data empirik hasil dari tahap pengujian sebelumnya.
6. Tahap Pembuatan Laporan
Pada tahap ini, akan dilakukan penyusunan laporan akhir dan pengumpulan dokumentasi berdasarkan analisis hasil penelitian Tugas Akhir ini.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan pengujian serta analisis yang telah dibahas dan dilaksanakan pada bab sebelumnya (Bab 4), maka dapat disimpulkan beberapa hal sebagai berikut:

1. Dapat menghasilkan program sebagai implementasi algoritma *CT-Pro*. Dari program tersebut didapatkan *knowledge* bagi perusahaan berupa informasi item yang dibeli dengan item yang dibeli pula beserta nilai *support* dan *confidencenya*.
2. Nilai *support* mempengaruhi waktu eksekusi program, jumlah *knowledge* dan akurasi dari *knowledge* yang didapatkan. Semakin besar nilai *support* maka akan semakin sedikit waktu eksekusi program dan jumlah *knowledge* yang dihasilkan.
3. Nilai *confidence* tidak mempengaruhi waktu eksekusi program, jumlah *knowledge* dan akurasi dari *knowledge* yang didapatkan. Hal ini terjadi karena digunakan nilai minimum *support* yang tetap yakni 50 % sehingga item yang akan menjadi kandidat *knowledge* akan tereduksi sebesar 50 % di setiap pengujian pada nilai minimum *confidence* yang berbeda-beda.

5.2. Saran

1. Proses pengambilan data dapat dikembangkan secara *online*, sehingga proses pengambilan data dapat dilakukan secara otomatis.
2. Implementasi pembuatan program menggunakan bahasa pemrograman C/C++ agar waktu eksekusi program dalam pembentukan *knowledge* lebih cepat.

6. Daftar Pustaka

- [1] Chandrawati, Evi. *Implementasi Algoritma SQL-Based Frequent Pattern Mining Dengan Algoritma Frequent Pattern-Growth Pada Metode Market Basket Analysis*. Universitas Pendidikan Indonesia.
- [2] J. Han, J. Pei, Y. Yin, and R. Mao. 2004. *Mining Frequent Patterns without Candidate Generation: A Frequentpattern Tree Approach*. *Data Mining and KnowledgeDiscovery: An International Journal*, KluwerAcademic Publishers.
- [3] Jiawei han & Micheline Kamber Harcourt. 2006. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers.
- [4] M. El-Hajj and O. R. Zaiane. 2003. *COFI-tree Mining: A New Approach to Pattern Growth with Reduced Candidacy Generation*. *Proceedings of the IEEE ICDM Workshop on Frequent itemset Mining Implementations*, Melbourne, USA.
- [5] Nugroho, Adi. 2009. *Rekayasa perangkat lunak menggunakan UML dan Java*. Yogyakarta : Andi.
- [6] R. Agarwal, C. Aggarwal, and V. V. V. Prasad. 2000. *A Tree Projection Algorithm for Generation of Frequent itemsets*. *Journal of Parallel and Distributed Computing (Special Issue on High Performance Data Mining)*.
- [7] Ruldeviyani, Yova, Muhammad Fahrian. 2008. *Implementasi Algoritma- Algoritma Associations Rules Sebagai Bagian dari Pengembangan Data Mining Algorithms Collections*. Universitas Indonesia.
- [8] Sucahyo, Yudho Giri, Raj P Gopalan. 2005. *CT-PRO: A Bottom-Up Non Recursive Frequent itemset Mining Algorithm Using Compressed FP-Tree Data Structure*. Curtin University of Technology.
- [9] Sucahyo, Yudho Giri, Raj P Gopalan. 2004. *High Performance Frequent Patterns Extraction using Compressed FP-Tree*. Curtin University of Technology.
- [10] Wicaksono Alfian & Purwarianti Ayu. 2010. *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*. Institut teknologi Bandung.