

ANALISIS PERBANDINGAN ALGORITMA FP-GROWTH DAN ALGORITMA APRIORI PADA MARKET BASKET ANALYSIS

Anmetatika Angelia Octaviani¹, Arie Ardiyanti Suryani², Imelda Ataina³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Market Basket Analysis (MBA) merupakan salah satu teknik data mining yang bertujuan untuk menemukan produk-produk yang sering dibeli secara bersamaan dari data transaksi. Informasi keterhubungan antar produk dapat dimanfaatkan pihak swalayan untuk meningkatkan keuntungan, misalnya dengan menempatkan produk yang sering dibeli bersamaan ke dalam sebuah area yang berdekatan, merancang tampilan produk di katalog, merancang paket produk. Pada tugas akhir ini diimplementasikan algoritma Frequent Pattern- Growth (FP-Growth) dan algoritma Apriori untuk melakukan market basket analysis. Algoritma FP-Growth dan algoritma Apriori dapat digunakan untuk menentukan himpunan data yang paling sering muncul dalam sekumpulan data.

Analisis yang dilakukan yaitu membandingkan association rule yang dihasilkan oleh algoritma FP-Growth terhadap association rule yang dihasilkan oleh algoritma Apriori, serta waktu pembentukan association rule. Algoritma Apriori merupakan algoritma yang biasa digunakan untuk market basket analysis.

Hasil yang didapat dari pengujian menunjukkan bahwa association rule yang dihasilkan oleh algoritma FP-Growth sama dengan association rule yang dihasilkan oleh algoritma Apriori. Namun, waktu yang diperlukan algoritma FPGrowth jauh lebih cepat daripada algoritma Apriori.

Kata Kunci : Market Basket Analysis, association rule, algoritma FP-Growth,

Abstract

Market Basket Analysis is a data mining technique to find product which purchased together in a transaction from transaction data. This information could be used by seller to increase profit, for example by place product which often be purchased together on near area, designing product catalogue, arranging product package.

This final assessment will implemented Frequent Pattern-Growth (FPGrowth) Algorithm and Apriori Algorithm to do market basket analysis. FPGrowth algorithm and Apriori algorithm could be used to determine data set with most appearance from data collection. Analysis which applied is comparing association rule produced by FP-Growth algorithm with association rule produced by Apriori algorithm, and time required to produce association rule. Apriori algorithm is an algorithm which often used for market basket analysis.

The result from test show that association rule from FP-Growth algorithm same with association rule from Apriori algorithm. Instead the time required by FP-Growth algorithm much less than time required by Apriori algorithm.

Keywords : Market Basket Analysis, association rule, FP-Growth algorithm,

1. Pendahuluan

1.1 Latar belakang

Saat ini, banyak swalayan yang memanfaatkan kemajuan teknologi untuk membantu kelancaran aktivitas perdagangan. Jumlah transaksi penjualan yang dicatat sangat besar, sehingga membutuhkan pengelolaan secara baik. Dari data transaksi yang ada, terdapat banyak informasi yang dapat digali untuk meningkatkan keuntungan swalayan.

Penggalian data merupakan cara yang dapat digunakan untuk mengetahui adanya serangkaian pola informasi dari sejumlah besar data yang ada. Di antara banyak fungsionalitas dari penggalian data, penggalian pola asosiasi merupakan salah satunya. Dalam penggalian dan menganalisis pola asosiasi akan ditemukan atribut-atribut yang menunjukkan kondisi dimana atribut-atribut tersebut sering muncul bersamaan dalam suatu data yang diberikan. Penggalian pola asosiasi ini biasanya sering digunakan dalam menganalisis data transaksi atau keranjang belanja di pasar swalayan, maka dari itu analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Analisis asosiasi atau *association rule mining* bertujuan untuk menemukan pola yang berurutan, asosiasi, dan hubungan sebab akibat antara himpunan data. *Association rule* dapat digunakan untuk mengidentifikasi *item-item* produk yang mungkin dibeli secara bersamaan dengan produk lain. *Association rule* yang dibuat dikatakan *strong* apabila *rule* tersebut memenuhi *minimum support* dan *minimum confidence* yang telah ditentukan oleh user.

Secara umum ada 2 tahap dalam melakukan *Association Rule Mining* yaitu :

1. Menemukan semua *frequent itemset*.
2. Meng-generate association rule dari *frequent itemset*.

Istilah *Market Basket* digunakan untuk menggambarkan kelompok *item* (terdiri dari dua atau lebih *item*) yang cenderung dibeli oleh seorang customer sewaktu berbelanja di swalayan dalam satu transaksi pembelian. Analisis *market basket* merupakan analisis dari kebiasaan customer dengan mencari asosiasi dan korelasi antar *item-item* berbeda yang diletakkan customer dalam keranjang belanjanya. Informasi mengenai keterhubungan antar *item* dibutuhkan untuk membuat suatu strategi pemasaran yang baik. Misalnya, kecenderungan customer untuk membeli *item* Y jika dia membeli *item* X, atau sebaliknya. Jika hal itu dapat diketahui oleh pihak *swalayan*, maka bisa meningkatkan penjualan *item* X dan *item* Y dengan cara menempatkan keduanya ditempat yang saling berdekatan. Dengan cara ini, diharapkan customer akan tertarik untuk membeli *item* Y saat ia membeli *item* X, atau sebaliknya. Selain itu dapat juga digunakan untuk mengatur *item-item* apa saja yang perlu dimasukkan dalam katalog produk untuk keperluan promosi produk.

Proses *market basket analysis* dapat diterapkan dengan beberapa algoritma, contohnya algoritma *Apriori* dan (*Frequent Pattern*) *FP-Growth*. *FP-Growth* merupakan salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *FP-Growth* menggunakan pendekatan yang berbeda dari

algoritma *Apriori* sehingga algoritma *FP-Growth* berpeluang menghasilkan rule yang berbeda. Dengan Algoritma *FP-Growth* kita dapat menemukan *Frequent Itemset* tanpa perlu melakukan *Candidate Generation*. Dengan cara ini *scan database* hanya dilakukan 2 kali, tidak perlu dilakukan berulang-ulang [5]. Data akan direpresentasikan dalam bentuk *FP-Tree*, setelah *FP-Tree* terbentuk digunakan pendekatan *Divide and Conquer* untuk memperoleh *Frequent Itemset*. *FP-Tree* merupakan struktur data yang baik sekali untuk *Frequent Pattern Mining*. Struktur ini memberikan informasi yang lengkap untuk membentuk *frequent pattern*. [11]

Algoritma *FP-Growth* akan efisien dalam menemukan *frequent itemset* [5], maka pada Tugas Akhir yang berjudul *Penerapan Algoritma FP-Growth Pada Market Basket Analysis* akan dianalisis efisiensi dari algoritma *FP-Growth* dalam menemukan *frequent itemset* dan pembuatan *association rule*.

1.2 Perumusan masalah

Perumusan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana membangun sebuah proses *market basket analysis* pada data transaksi dengan menggunakan algoritma *FP-Growth* dan algoritma *Apriori*?
2. Bagaimana performansi dari penerapan algoritma *FP-Growth* dan algoritma *Apriori* pada proses *market basket analysis* ?

Sedangkan batasan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Proses yang dilakukan hanya sebatas penerapan algoritma *FP-Growth* dan algoritma *Apriori* untuk pembentukan *association rule*.
2. Aplikasi yang akan dibuat juga menerapkan algoritma *Apriori*, hal ini agar dapat dilakukan perbandingan hasil secara adil.
3. Ada 3 dataset yang akan digunakan untuk pengujian.
4. Perbandingan hasil dari algoritma *FP-growth* dan algoritma *Apriori* yaitu dengan melihat nilai rata-rata *confidence* dari *association rule*

Hipotesa awal:

Pencarian *frequent itemset* dengan algoritma *FP-Growth* akan lebih efisien dibandingkan algoritma *Apriori*.

1.3 Tujuan

Tujuan yang ingin dicapai dalam tugas akhir ini adalah:

1. Menerapkan algoritma *FP-Growth* dan algoritma *Apriori* pada proses *market basket analysis* pada data transaksi penjualan, sehingga mendapatkan *association rule* yang baik agar dapat dijadikan informasi bagi swalayan untuk dapat mengatur strategi pemasaran produk.
2. Menganalisa performansi dari implementasi *FP-Growth* dan algoritma *Apriori* berdasarkan waktu dan rata-rata nilai *confidence*.

1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Studi literatur
Mencari sumber-sumber yang dapat menunjang topik Tugas Akhir ini. Yaitu mencari referensi mengenai algoritma *Apriori* dan algoritma *FP-Growth* serta *Association rule*.
2. Pengumpulan data-data penunjang Tugas Akhir
Mengumpulkan beberapa dataset yang akan digunakan untuk pengujian dan analisis Tugas Akhir.
3. Analisa dan perancangan sistem
Menganalisis masalah dan perancangan perangkat lunak untuk proses *market basket* yang mengimplementasikan algoritma *Apriori* dan algoritma *FP-Growth*.
4. Implementasi
Merealisasikan sistem dari rancangan yang telah dibuat. Sistem direalisasikan menggunakan program aplikasi NetBeans IDE 6.0
5. Pengujian sistem
Menguji perangkat lunak yang telah dibuat dengan inputan berupa data transaksi penjualan, nilai *minimum support*, dan *minimum confidence*. Kemudian membandingkan *association rule* dari algoritma *FP-Growth* dengan algoritma *Apriori* kemudian menganalisisnya.
6. Penyusunan laporan tugas akhir
Mengambil kesimpulan dari hasil analisa yang telah dilakukan, kemudian menyusun laporan tugas akhir.

5. Penutup

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari hasil pengujian dan analisis yang telah dilakukan sebelumnya, yaitu :

1. Algoritma *FP-Growth* dapat diterapkan untuk proses Market Basket Analysis
2. Algoritma *FP-Growth* dapat membentuk association rule yang sama baik dengan association rule yang dibentuk oleh algoritma yang biasa digunakan untuk market basket analysis, yakni algoritma *Apriori*.
3. Dalam melakukan proses market basket analysis, algoritma *FP-Growth* dapat bekerja lebih cepat diukur dari segi waktu dibandingkan algoritma *Apriori*.
4. Nilai optimal *FP-Growth* jika dibandingkan dengan algoritma *Apriori* tercapai pada dataset dengan jumlah records lebih dari 700 records. Atau tercapai pada dataset dengan jumlah item lebih dari 1807 item.

5.2 Saran

Saran-saran yang dapat diberikan jika dilakukan pengembangan terhadap TA ini adalah sebagai berikut :

1. Perlunya penelitian lebih lanjut tentang penentuan minimum support yang tepat agar kualitas association rule yang dihasilkan semakin baik.
2. Pentingnya pemilihan data yang baik, yaitu yang lebih tersebar merata, agar association rule yang dihasilkan semakin baik.

Daftar Pustaka

- [1] Berry, Michael J.A dan Linoff, Gordon S., "*Data Mining Techniques for Marketing, Sales, Customer Relationship Management*", Second Edition, Wiley Publishing, Inc., 2004
- [2] Christian, Borhelt, "*An Implementation of the FP-Growth Algorithm*", <http://osdm.ua.ac.be/papers/p1-borgelt.pdf> didownload pada tanggal 10 Desember 2008
- [3] Fung, Benjamin C.M, Ke Wang dan Martin Ester., "*Hierarchical Document Clustering*", Canada: Simon Fraser University, 2002
www.cs.sfu.ca/~ester/papers/Encyclopedia.pdf didownload pada tanggal 12 Desember 2008
- [4] Han, Jiawei and Kamber, Micheline, "*Data Mining Concepts and Techniques*", Morgan Kaufmann Publishers, California, 2006
- [5] Han, Jiawei, Pei, Jian and Yin, Yiwen, "*Mining Frequent Patterns without Candidate Generation*",
<http://www.cs.ualberta.ca/~zaiane/courses/cmpu695-00/papers/fptree.pdf> didownload pada tanggal 13 Desember 2008
- [6] Kantardzic, Mehmed, "*Data Mining : Concepts, Models, Methods, and Algorithms*", John Wiley & Sons, Inc., 2003
- [7] Kumar, A.V.Senthil and Wahidabanu, R.S.D, "*Discovery of Frequent Itemset : Frequent Item Tree-Based Approach*",
<http://proceedings.itb.ac.id/download.php?file=C07004.pdf&id=239&cup=18> didownload pada tanggal 13 Desember 2008
- [8] Larose, Daniel T, "*Discovering Knowledge in Data: An Introduction to Data Mining*", John Wiley & Sons, Inc., 2005
- [9] Ranka, Sanjay, "*Association Analysis*",
<http://www.cise.ufl.edu/class/cis6930fa03dm/notes/dm4part1.pdf> didownload pada tanggal 23 Oktober 2008
- [10] Samuel, David, "*Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset*",
<http://www.informatika.org/~rinaldi/Matdis/2007-2008/Makalah/MakalahIF2153-0708-006.pdf> didownload pada tanggal 13 Desember 2008
- [11] Soelaiman, Rully and Arini, Ni Made, "*Analisis Kinerja Algoritma FOLD-Growth dan FP-Growth pada Penggalan Pola Asosiasi*",
<http://www.si.its.ac.id/Penelitian/JURNAL/Arin.pdf> didownload pada tanggal 13 Desember 2008
- [12] Srikant, Ramakrishnan and Agrawal, Rakesh, "*Fast Algorithm for Mining Association Rules*", <http://140.115.82.191/old/Agr0801/vldb94.ps> didownload pada tanggal 13 Desember 2008
- [13] Srikant, Ramakrishnan and Agrawal, Rakesh, "*Mining Generalized Association Rules*", http://140.115.82.191/old/Agr0801/vldb95_tax.ps didownload pada tanggal 13 Desember 2008