

PENERAPAN METODE APRIORITID UNTUK MARKET BASKET ANALYSIS PADA SUPERMARKET

Kania Audrint¹, Imelda Atastina², Retno Novi Dayawati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Banyak perusahaan bisnis yang mengumpulkan data dalam jumlah besar setiap harinya. Sebagai contohnya, data pembelian oleh customer dalam jumlah besar dikoleksi setiap harinya di setiap counter pada supermarket. Retailers menaruh perhatian yang sangat besar dalam upaya penganalisaan data-data ini untuk menemukan kebiasaan pembelian customer. Informasi berharga ini teramat penting untuk mendukung strategi promosi, manajemen inventaris, dan manajemen retensi customer.

Terdapat sebuah metodologi untuk menemukan relasi antar dataset yang sangat besar yang dikenal sebagai analisis asosiasi untuk market basket data. Analisis asosiasi akan menghasilkan aturan asosiasi yang mengindikasikan kebiasaan pembelian customer

Meskipun ada beberapa algoritma untuk menemukan aturan asosiasi, namun algoritma-algoritma ini tidak efisien dalam waktu proses. Tujuan dari tugas akhir ini adalah menampilkan hasil analisa dari pengimplementasian algoritma Apriori- TID untuk menemukan aturan asosiasi pada supermarket. Terdapat perbandingan dengan algoritma Apriori sebagai pionir dalam penemuan aturan asosiasi dan apriori-TID sebagai algoritma pengembangan apriori pada performansi waktu proses. Penelitian menunjukkan bahwa penerapan algoritma Apriori-TID menambah efisiensi terhadap waktu proses dibandingkan dengan algoritma Apriori sehingga algoritma apriori-TID sesuai untuk mengekstraksi aturan asosiasi pada dataset besar seperti data transaksi supermarket.

Kata Kunci : analisis asosiasi, market basket data, aturan asosiasi, algoritma apriori, algoritma apriori-TID

Abstract

Many business enterprises accumulate large quantities of data from day-to-day operation. For example, huge amount of customer purchased data are collected daily at check count counter in supermarket. Retailers are interested in analyzing these data to find customer purchased behavior. This valuable information is very important for supporting promotion strategy, inventory management, and customer relationship management

There is methodology for discovering relationship among large data set which is known as association analysis for market basket data. Association analysis will yield association rule which indicate customer purchased behavior. Although some algorithms can find association rule, they can be inefficient in computational time. The aim of this paper is to present the analysis of implementing apriori-TID algorithm for mining association rule in supermarket. There will be comparison of apriori algorithm as pioneer of mining association rule and apriori-TID algorithm as developed apriori algorithm on performance of computational time. The observation shows that apriori-TID algorithm implementation will enhance the efficiency of computational time compare to apriori algorithm thus this Apriori-TID algorithm is suitable for extracting association rule in large dataset as supermarket transaction data.

Keywords : association analysis, market basket data, association rule, apriori algorithm, apriori TID algorithm

1. Pendahuluan

1.1 Latar belakang

Perkembangan era informasi dan teknologi yang semakin pesat membawa konsekuensi terhadap penyimpanan data yang besar. Contohnya saja, sejak ditemukannya teknologi barcode. Teknologi ini memungkinkan pencatatan data transaksi barang dalam jumlah besar yang cepat dan akurat. Semua catatan transaksi ini biasa disebut dengan *basket data*.

Salah satu penghasil data bervolume besar setiap tahunnya yaitu supermarket. Supermarket dengan beribu pelanggan yang setiap harinya melakukan transaksi pelanggan mencatat ratusan basket data lebih. Data ini tentu hanya akan memakan memori penyimpanan apabila tidak diolah untuk kepentingan pemasaran maupun pengambilan keputusan.

Ada beberapa kasus yang dapat diselesaikan dengan pengolahan data diantaranya *market basket analysis* pada pelanggan supermarket. *Market basket analysis* ini merupakan salah satu analisis yang bermanfaat di dunia *marketing* karena dari hasil analisis ini didapat aturan asosiatif dari analisis pembelian. Dengan pengetahuan tersebut sebuah supermarket dapat mengatur strategi marketingnya.

Market basket analysis ini merupakan aturan yang menyatakan asosiasi antara beberapa *item* [4]. Algoritma apriori adalah pencetus pertama penyelesaian aturan asosiasi yang ditemukan oleh Agrawal R pada tahun 1993. Masalah utama algoritma ini ialah dibutuhkan waktu yang sangat lama untuk mengenerate *frequent itemset* [7].

Waktu yang panjang ini disebabkan oleh dua proses yaitu proses pembacaan database yang besar berulang kali serta proses menciptakan *frequent itemset* dengan operasi JOIN pada database [7]. Beberapa algoritma perbaikan apriori telah diciptakan untuk mengurangi waktu proses dalam menemukan aturan asosiasi ini. Diantaranya yaitu algoritma Apriori TID [9].

Pada algoritma Apriori TID, pembacaan database hanya dilakukan pertama kali diawal, setelahnya akan dibaca database yang berisi *candidate itemset*. Pembacaan ini akan mengurangi waktu operasi I/O karena database yang berisi *candidate itemset* jauh lebih kecil dari database sebelumnya. Hal inilah yang menambah efisiensi dari algoritma Apriori TID [7].

Penggunaan algoritma Apriori TID ini memungkinkan penemuan aturan asosiatif dengan waktu yang relatif lebih cepat. Oleh sebab itu, metode Apriori TID paling sesuai untuk digunakan dalam memecahkan permasalahan aturan asosiasi pada basket data di Supermarket.

1.2 Perumusan masalah

Masalah yang menjadi acuan dalam pembuatan TA ini adalah :

1. Bagaimana mencari aturan asosiasi pada data supermarket
2. Bagaimana menemukan aturan asosiasi dengan metode Apriori TID
3. Bagaimana membangun sistem berdasar metode Apriori TID untuk melakukan market basket analysis pada data supermarket
4. Bagaimana performansi penerapan metode Apriori TID terhadap pencarian aturan asosiasi pada data supermarket

1.3 Batasan Masalah

Batasan masalah dalam Tugas Akhir ini meliputi :

1. Dataset yang digunakan sebagai input berupa data set transaksi pelanggan supermarket YOMART Margacinta Bandung tahun 2009
2. Asumsi data bersih.

1.4 Tujuan

Tujuan dari tugas akhir ini yaitu:

1. Melakukan proses market basket analysis dengan menggunakan metode Apriori TID.
2. Mengimplementasikan metode Apriori TID dalam bentuk perangkat lunak
3. Mencari tingkat keakuratan output sistem.
4. Mencari pengaruh minimum support dan minimum confidence terhadap jumlah frequent itemset dan aturan asosiasi.
5. Mencari nilai optimal minimum support dan minimum confidence untuk mendapat rata-rata nilai lift tertinggi dari aturan asosiasi yang dihasilkan oleh sistem.
6. Mengukur seberapa baik performansi dari penerapan metode Apriori TID dalam menemukan aturan asosiasi dengan parameter pengujian minimum support,skalabilitas,jumlah k-itemset,dan waktu proses.

1.5 Metodologi penyelesaian masalah

Pengerjaan Tugas Akhir ini menggunakan metodologi :

1. Studi Literatur :
Pencarian referensi-referensi dan sumber-sumber lain yang layak yang berhubungan dengan *metode Apriori TID, market basket analysis, data mining,dan association rule*.
2. Mengumpulkan dan memahami data transaksi sebuah supermarket
3. Mempelajari konsep dari metode Apriori TID beserta pemecahannya.
4. Melakukan analisis metode Apriori TID dalam upaya pembuatanperangkat lunaknya.
5. Melakukan implementasi perancangan perangkat lunak.

6. Melakukan pengujian akurasi dari hasil perangkat lunak yang telah dibangun
7. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.

1.6 Sistematika Penulisan

Tugas akhir ini akan disusun dengan sistematika penulisan sebagai berikut:

1. PENDAHULUAN

Bab ini memaparkan latar belakang dilakukannya penelitian, perumusan masalah yang akan dibahas, pembatasan masalah, tujuan yang ingin dicapai melalui penelitian ini, metode penyelesaian masalah dan sistematika pembahasan.

2. DASAR TEORI

Bab ini memuat berbagai dasar teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu mengenai konsep dari metode Apriori TID, market basket analysis, data mining

3. ANALISIS DAN PERANCANGAN SISTEM

Bab ini menguraikan tentang tahapan yang dilakukan untuk membangun perangkat lunak sebagai pembantu dalam mendapatkan data untuk proses analisa, alur kerja (*work flow*) dari perangkat lunak yang dibuat, bagaimana keterhubungan antar objek dan kelas-kelas yang terbentuk.

4. IMPLEMENTASI DAN HASIL PENGUJIAN

Bab ini menyajikan tentang implementasi hasil analisis dan perancangan sistem ke dalam bentuk pemrograman aplikasi. Melakukan pengujian terhadap aplikasi menggunakan kasus yang sederhana.

5. KESIMPULAN DAN SARAN

Berisi kesimpulan dari hasil penelitian tugas akhir ini serta saran-saran untuk pengembangan lebih lanjut

Telkom
University

5. Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari Tugas Akhir ini yaitu:

- a) Algoritma AprioriTID dapat diimplementasikan pada perangkat lunak dan mampu menghasilkan output aturan asosiasi dengan akurasi 100%
- b) Nilai minimum support dan minimum confidence memiliki korelasi yang kuat terhadap jumlah frequent itemset dan aturan asosiasi yang dibangkitkan oleh sistem. Semakin besar nilai minimum support semakin sedikit jumlah frequent itemset dan semakin besar nilai minimum confidence semakin sedikit jumlah aturan asosiasi yang terbentuk
- c) Nilai lift dari aturan asosiasi dapat digunakan untuk memperkirakan nilai minimum support dan minimum confidence yang optimal dalam menghasilkan aturan asosiasi dengan nilai interest yang baik.
- d) Performansi sistem algoritma apriori dan aprioriTID memiliki korelasi yang kuat dengan nilai minimum support. Terdapat hubungan dimana apabila nilai minimum support kecil dan mencapai nilai optimal, performansi aprioriTID mengungguli algoritma apriori.
- e) Performansi sistem berdasarkan jumlah record terhadap waktu proses oleh algoritma aprioriTID unggul dari algoritma apriori selama nilai minimum support yang digunakan mencapai nilai optimal, dimana optimal berarti jumlah kandidat frequent itemset yang dibentuk besar dan mencukupi kapasitas memori.
- f) Performansi algoritma aprioriTID juga dipengaruhi oleh faktor k-itemset. Semakin besar jumlah k-itemset pada database transaksi semakin lama waktu proses yang dibutuhkan sistem.
- g) Karakteristik data dengan variansi kecil dan jumlah k-itemset kecil akan menghasilkan aturan asosiasi dengan nilai lift yang besar

5.2 Saran

Saran yang dapat diberikan pada tugas akhir ini sebagai bentuk pengembangan sistem yaitu:

- a) Data transaksi sebagai data uji percobaan pada tugas akhir ini masih bersifat statis (tidak berubah). Sebaiknya data transaksi yang dipakai menggunakan data transaksi realtime yang bersifat dinamis (berubah) agar hasil aturan asosiasinya lebih akurat
- b) Kriteria pengukuran nilai *nilai interestingness* dari sebuah aturan asosiasi perlu digali lebih dalam agar user benar-benar menerima rule dengan *nilai interestingness* tinggi

Daftar Pustaka

- [1] A.Rakesh, and S.Ramakrishnan."Fast Algorithm for Mining Association Rules".Proceedings of the 20th international conference on very large databases, 478-499, 1994
- [2] B.Ferenc, "A fast Apriori implementation",Informatics Laboratory Computer and Automation Research Institute, Hungarian Academy of Sciences,2003
- [3] Jiawei Han, Datamining:Concepts and Techniques , 2010, <http://www.cs.uiuc.edu/homes/hanj/bk2/toc.pdf> didownload pada tanggal 09 Januari 2011.
- [4] Kusriani, Luthfi E, 2007. Algoritma Data Mining. Edisi Pertama. STMIK AMIKOM, Yogyakarta
- [5] Motoda Hiroshi and Ohara Kouzou,The top ten algorithm in datamining, <http://www.inst-informatica.pt/servicos/informacao-e-documentacao/dossiers-tematicos/dossier-tematico-no-8-business-intelligence-abril-2010/livros-recomendados/the-top-ten-algorithms-in-data-mining-indice> didownload pada tanggal 1 januari 2011
- [6] NguyenHung Son, Transaction data analysis and associationrules , 2006, http://www.mimuw.edu.pl/~son/datamining/DM/2-asso_rules.pdf didownload pada tanggal 08 Januari 2011
- [7] P.Li, J. Chen, F.Bian."A Developed Algorithm of Apriori Based On Association Analysis".Wuhan University
- [8] Qiankun Zhao, Association Rule Mining: A Survey, 2010, <http://sci2s.ugr.es/keel/pdf/specific/report/zhao03ars.pdf> didownload pada tanggal 01 Januari 2011
- [9] Santosa,Budi.2007. "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis Teori dan Aplikasi".Graha Ilmu, Yogyakarta.
- [10] Srikant Ramakrishnan, Mining generalized Association Rule, 2010, <http://www.sop.inria.fr/axis/teaching/MasterPLMT/articles/Srikant.pdf> didownload pada tanggal 09 Januari 2011.
- [11] http://en.wikipedia.org/wiki/A_priori_and_a_posteriori. Diakses pada tanggal 26 Oktober 2009 pukul 15.35
- [12] <http://www.iep.utm.edu/apriori/>. Diakses pada tanggal 26 Oktober 2009 pukul 15.35
- [13] <http://www.computer.org/portal/web/csdl/doi/10.1109/ESIAT.2009.64>. Diakses pada tanggal 26 Oktober 2009 pukul 15.35
- [14] <http://www.citeulike.org/user/seninp/article/5972687>. Diakses pada tanggal 26 Oktober 2009 pukul 15.35
- [15] http://www.paper.edu.cn/en/paper.php?serial_number=200903-333. Diakses pada tanggal 26 Oktober 2009 pukul 15.35