

PENGGALIAN KAIDAH MULTILEVEL ASSOCIATION RULE DARI DATA MART SWALAYAN ASGAP

Teguh Pradana ¹⁾

¹⁾ Program Studi/Prodi Teknik Informatika, STMIK Yadika,
email: INTI_PERSADA_SOFTWARE@yahoo.co.id

Abstrak:

Perkembangan teknologi informasi telah mengalami kemajuan yang sangat pesat, termasuk penggunaannya di sektor pemasaran. Namun pada umumnya digunakan sebatas penyimpanan data yang hanya digunakan untuk pencatatan transaksi dan mengolah data transaksi tersebut untuk pembuatan laporan pembelian atau penjualan perusahaan. Bagi penjualan yang hanya menjual beberapa jenis produk atau toko-toko kecil, laporan seperti itu mungkin cukup bagi pemilik untuk melakukan analisis pada pasar dan mengambil keputusan.

Hal ini dapat berpengaruh dalam penjualan suatu jenis produk tertentu, tetapi karena banyaknya data yang harus diolah, informasi tersebut menjadi bias atau bahkan tidak akan ditemukan jika dilakukan secara manual. Oleh karena itu, perusahaan seperti ini perlu menggunakan konsep data mining dengan market basket analysis untuk mengetahui buying habits dari konsumen sehingga dapat membantu pengambilan keputusan. Dengan menggunakan data mart yang ada dapat dijadikan sebagai pendukung untuk diolah dengan menggunakan teknik-teknik yang ada pada data mining.

Multi level association rule bisa memberikan aturan asosiasi pada tingkatan yang berbeda, yang menggunakan data mart yang memiliki data multi dimensi, sehingga dapat memberikan informasi pada tingkatan yang berbeda.

Keywords: *Multilevel Association Rule*

1. Pendahuluan

Swalayan Asgap, sejak didirikan pada tahun 1997 telah melakukan komputerisasi di bidang penjualannya. Analisis telah dilakukan sejauh ini berdasarkan laporan transaksi setiap bulannya. Besarnya volume data yang perlu diolah, menyebabkan swalayan Asgap mengalami kesulitan dalam memantau asosiasi antara penjualan yang satu dengan yang lain. Terkadang hasil pengolahan data dengan cara sederhana tidak memberikan hasil yang mendukung. Kebijakan akhirnya ditempuh berdasarkan informasi singkat berupa laporan yang didukung dengan intuisi manajer. Keberadaan informasi dapat

mempengaruhi segala keputusan dan strategi yang akan diambil oleh manajer.

Melihat bahwa diperlukan adanya berbagai aspek yang menjadi pertimbangan dalam pengambilan keputusan, sehingga dapat dilihat atribut seperti waktu, produk sebagai variabel yang berpengaruh pada informasi yang ingin diperoleh dan dengan konsep *multi-level* untuk pencarian *frequent itemset* dalam pembentukan *association rule*. *Single level* umumnya hanya melihat dari satu konsep *level*, tidak menawarkan informasi *item* dari berbagai abstraksi, sedangkan *multi-level* dapat memberikan informasi dari tingkatan yang berbeda, misalnya berdasarkan katagori barang dan nama barang. *Association Rules* yang dihasilkan disertai dengan tingkat dukungan

data berdasarkan *history* data perusahaan (*support* dan *confidence*). Misalkan ketika konsumen membeli produk A, maka swalayan dapat menawarkan pula produk B. Keputusan ini diambil berdasarkan *history* data transaksi swalayan, terdapat dukungan data dengan *confidence* 80%, bahwa kebutuhan akan produk B akan meningkat seiring dengan permintaan produk A, maka diharapkan dapat membantu pengambilan keputusan untuk *market basket analysis* perusahaan.

2. Tinjauan Pustaka

a. Market Basket Analysis

Market Basket Analysis adalah suatu cara yang digunakan untuk menganalisis data penjualan dari suatu perusahaan. Proses ini menganalisis *buyinghabits* konsumen dengan menemukan asosiasi antar *item-item* yang berbeda yang diletakkan konsumen dalam *shopping basket* (Gregorius S Budhi, 2007). Hasil yang telah didapatkan ini nantinya dapat dimanfaatkan oleh perusahaan retail seperti toko atau swalayan untuk mengembangkan strategi pemasaran dengan melihat *item-item* mana saja yang sering dibeli secara bersamaan oleh konsumen.

Untuk beberapa kasus, pola dari *item-item* yang dibeli secara bersamaan oleh konsumen mudah ditebak, misalnya susu dibeli secara bersamaan dengan roti, Namun, mungkin saja terdapat suatu pola pembelian *item* yang tidak terpikirkan sebelumnya, misalnya pembelian minyak goreng dengan deterjen. Mungkin saja pola ini tidak pernah terpikirkan sebelumnya karena minyak goreng dan deterjen tidak ada hubungan sama sekali, baik sebagai barang pelengkap maupun barang pengganti. Hal ini mungkin tidak terpikirkan sebelumnya sehingga tidak dapat diantisipasi jika terjadi sesuatu, seperti kekurangan stok deterjen misalnya. Inilah salah satu manfaat yang dapat diperoleh dari melakukan *market*

basket analysis. Dengan melakukan proses ini secara otomatis seorang manajer tidak perlu mengalami kesulitan untuk menemukan pola *item* apa saja yang mungkin dibeli secara bersamaan.

b. Association rule

Association Rule Mining meliputi dua tahap (Ulmer, David, 2002) :

1. Mencari kombinasi yang paling sering terjadi dari suatu *itemset*.
2. Mendefinisikan *Condition* dan *Result* (untuk *conditional association rule*).

Dalam menentukan suatu *association rule*, terdapat suatu *interestingness measure* (ukuran kepercayaan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Umumnya ada dua ukuran, yaitu

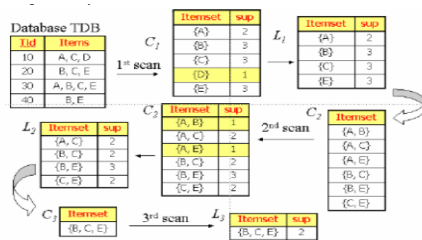
- a. *Support*, yaitu suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *itemset* layak untuk dicari *confidence*-nya (misalnya, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa *item* A dan B dibeli bersamaan).
- b. *Confidence*, yaitu suatu ukuran yang menunjukkan hubungan antar dua item secara *conditional* (misal, seberapa sering *item* B dibeli jika orang membeli *item* A). Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh user. Batasan tersebut umumnya terdiri dari *minimum support* dan *minimum confidence*.

c. Algoritma Apriori

Apriori adalah suatu algoritma yang sudah sangat dikenal dalam melakukan *frequent itemset* dengan *association rule*. Sesuai dengan namanya, algoritma ini menggunakan *knowledge* mengenai *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya.

Algoritma inilah yang biasanya dipakai dalam proses *data mining* untuk *market basket analysis*. Algoritma apriori menggunakan pendekatan *iterative (level-wise search)*, dimana *k-itemset* dipakai untuk menyelidiki $(k+1)$ -*itemset*. Langkah-langkah dari algoritma ini adalah sebagai berikut (Ulmer, David, 2002) :

1. Set $k=1$ (menunjuk pada *itemset* ke-1)
2. Hitung semua *k-itemset (itemset yang mempunyai k item)*
3. Hitung support dari semua calon *itemset-filteritemset* tersebut berdasarkan perhitungan *minimum support*.
4. Gabungkan semua *k-sized itemset* untuk menghasilkan calon *itemset* $k+1$.
5. Set $k=k+1$.
6. Ulangi langkah 3-5 sampai tidak ada *itemset* yang lebih besar yang dapat dibentuk.
7. Buat *final set* dari *itemset* dengan menciptakan suatu *union* dari semua *k-itemset*.



d. Algoritma FP-Growth

Algoritma FP-Growth merupakan pengembangan dari algoritma apriori, sehingga kekurangan dari algoritma apriori diperbaiki oleh algoritma FP-Growth. *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sekumpulan data^[8]. Struktur data yang digunakan untuk mencari

frequent itemset dengan algoritma FP-Growth adalah perluasan dari sebuah pohon prefix, yang biasa disebut FP-Tree.

e. Data Mart

Data mart adalah suatu bagian pada data warehouse yang mendukung pembuatan laporan dan analisa data pada suatu unit, bagian atau operasi pada suatu perusahaan. Dalam beberapa implementasi data warehouse, data mart adalah miniature data warehouse. Data mart sering digunakan untuk memberikan informasi kepada segmen fungsional organisasi. Contoh umum data mart adalah untuk departemen penjualan, departemen persediaan dan pengiriman, departemen keuangan, manajemen tingkat atas, dan seterusnya.

Karakteristik Data mart :

- Data mart memfokuskan hanya pada kebutuhan-kebutuhan pemakai yang terkait dalam sebuah departemen atau fungsi bisnis.
- Data mart biasanya tidak mengandung data operasional yang rinci seperti pada *data warehouse*.
- Data mart hanya mengandung sedikit informasi dibandingkan dengan *data warehouse*. Data mart lebih mudah dipahami.

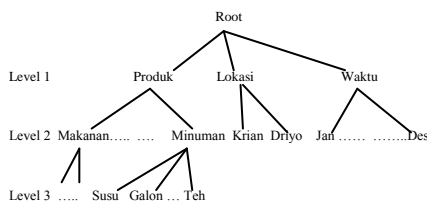
Kadang kala sulit untuk membedakan antara datawarehouse dan data mart karena keduanya hampir sama. Namun, jika dikaji lebih jauh ada beberapa perbedaan yang dimiliki keduanya. Datawarehouse merupakan gabungan dari beberapa data mart dan levelnya berada pada perusahaan atau organisasi. Sedangkan data mart merupakan bagian dari datawarehouse dan berada level departemen pada perusahaan atau organisasi tersebut. Data mart menangani sebuah business proses, misalkan penjualan, maka hanya proses penjualan saja yang ditangani pada data mart.

Tiga fungsi utama yang perlu dilakukan untuk membuat data siap digunakan pada

datawarehouse adalah *extraction, transformation* dan *loading* (ETL). Ketiga fungsi ini terdapat pada *staging area*. Pada *data staging* ini, disediakan tempat dan area dengan beberapa fungsi seperti data *cleansing, change, convert*, dan menyiapkan data untuk disimpan serta digunakan oleh datawarehouse.

f. Multilevel Association Rule

Banyak aplikasi *data mining* asosiasi yang membutuhkan pemrosesan pada *multi-level* abstraksi. Dibandingkan dengan *single-level, multi-level* dapat memberikan informasi yang lebih spesifik dan lebih fokus karena dapat memberikan informasi dari tingkatan abstraksi yang berbeda [11]. Pada penelitian sebelumnya yang dilakukan oleh Scott Fortin, Ling Liu, Randy Goebel (June 1996)[13]. Untuk mendapatkan multilevel association rule, perlu terlebih dahulu dibentuk suatu *concept hierarchy tree* dari data - data yang ada, seperti terlihat pada Gambar 1 dan menyusunnya ke dalam sebuah *generalized description table*, seperti terlihat pada Tabel 1. Selanjutnya data transaksi yang akan di-mining dirubah / di-transformasi menjadi *encoded transaction table*, dimana item - item yang ada pada sebuah transaksi dikodekan sesuai dengan nilai GID-nya pada *generalized description table* [11]. Sebagai contoh, lihat Tabel 1.



Gambar 1. Contoh *Concept Hierarchy Tree*

Tabel 1. Contoh *Generalized Description Table*

TID	Items
T ₁	{111, 121, 211, 221}
T ₂	{111, 211, 222, 323}
T ₃	{112, 122, 221, 411}
T ₄	{111, 121}
T ₅	{111, 122, 211, 221, 413}
T ₆	{211, 323, 524}
T ₇	{323, 411, 524, 713}

3. Implementasi Sistem

Dataset yang akan digunakan dalam studi kasus ini adalah dataset retail yang diambil dari swalayan “Asgap”. Swalayan Asgap yang berlokasi di Krian berdiri pada tahun 1997 dan yang berlokasi di Driyorejo berdiri tahun 2004. Aplikasi penjualan menggunakan bahasa pemrograman Clipper (*under DOS*). Database yang digunakan terdiri dari master penjualan item barang. Setiap periode baru aplikasi akan *create* tabel penjualan berdasarkan periode, sehingga tabel akan semakin banyak berdasarkan berjalannya waktu periode.

- Konsumen yang membeli barang dengan katagori “ROKOK” dan “PASTA” cenderung membeli barang dengan katagori “KOPI”
[support : 0.27% confidence : 3.70%]
- Konsumen yang membeli barang di wilayah “DRIYOREJO” cenderung membeli barang dengan katagori “MIE INSTANT” dan “KOPI”
[support : 3.42% confidence : 2.94%]
- Konsumen yang membeli barang di bulan “JANUARI” cenderung membeli barang dengan katagori “MINUMAN” dan “SABUN CUCI”
[support : 3.97% confidence : 0.38%]
- Konsumen yang membeli barang dengan merk “GG SURYA” cenderung membeli barang dengan merk “ROMA”
[support : 9.98% confidence : 0.45%]
Dan seterusnya

4. Penutup

Data mart dan data mining dapat membantu manajemen dalam pengambilan tindakan-tindakan bisnis dengan membekali pengetahuan berupa pola yang berasal dari data-data masa lalu. Dalam penelitian ini diperoleh pengetahuan berupa pola asosiasi

antara satu produk dengan produk yang lain. Aturan yang dihasilkan harus dievaluasi terlebih dahulu sebelum diaplikasikan. Hasil dari evaluasi memberikan hasil yang memuaskan, di mana aturan asosiasi memberikan manfaat yang lebih besar daripada tidak menggunakan aturan sama sekali.

5. Kesimpulan dan Saran

1. Data mart bisa digunakan untuk menentukan *frequent itemset*, dan data mart bisa mendukung data mining khususnya multi level association rule, dengan memberikan dimensi data yang berbeda yang dibutuhkan oleh data mining.
2. Multi Level Association Rule bisa menghasilkan kombinasi item dari tingkatan level yang berbeda, seperti produk, waktu, serta wilayah.
3. Dari hasil kombinasi item, yang mempunyai nilai *confidence* yang besar, bisa dijadikan sebagai pedoman dalam meningkatkan penjualan, misalnya manajemen dapat menginstruksikan kepada tenaga penjual untuk selalu mengingatkan konsumen yang membeli item untuk membeli item yang lain yang mempunyai nilai *confidence* yang besar terhadap item yang lain.

Saran yang bisa diberikan berkaitan dengan penelitian ini untuk pengembangan selanjutnya adalah :

- Adanya pengembangan penelitian dengan menggunakan dimensi yang lebih banyak dari penelitian ini (produk, waktu dan wilayah).

Daftar Pustaka

1. Yinbo WAN, Yong LIANG, Liya DING. "Mining Association Rules From Primitive Frequent Itemset". Vol. 3 No. 1. June 30, 2009.
2. Virendra Kumar Shrivastava, Parveen Kumar, K. R. Pardasani. "Discovery of

Multi-level Association Rules from Primitive Level Frequent Patterns Tree". Vol. 3 No. 1, July 2010 (ISSN 0974-3375).

3. Berlin Chen (2006). "Association Rules".
4. Pratima Gautam, Dr. K. R. Pardasani. "Algorithm for Efficient Multilevel Association Rule Mining". Vol. 02, No. 05, 2010, 1700-1704.
5. Scott Fortin, Ling Liu. "An Object-Oriented Approach to Multi-Level Association Rule Mining".