

PENGGALIAN KAIDAH MULTILEVEL ASSOCIATION RULE DARI DATA MART SWALAYAN ASGAP KRIAN

Agung Santoso¹, Teguh Pradana¹, dan Olyontang²

¹Teknik Informatika Universitas Maarif Hasyim Latif Sidoarjo

²Teknik Komputer Universitas Maarif Hasyim Latif Sidoarjo

agung@dosen.umaha.ac.id, teguh@dosen.umaha.ac.id, olyontang@dosen.umaha.ac.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.

Kata kunci: *Data Mining, Market Basket Analysis, Multi Level Association Rule dan Data Mart.*

ABSTRACT

Development of information technology has progressed very rapidly, including its use in the marketing sector. But in general use are only limited data storage is used for recording transactions and processing the transaction data to report the purchase or sale of the company. For sale is simply selling a product or some kind of small shops, such a report may be enough for the owners to do the analysis on the market and make decisions.

This can affect the sold of a certain type of product, but because of data to be processed, the information is biased or not even be found if done manually. Therefore, these companies need to use the concept of data mining with market basket analysis to find out the buying habits of consumers that can help decision making. By using an existing data mart be used as a support to be processed using existing techniques in data mining.

Multi-level association rule can provide association rules at different levels, using data mart that has a multi-dimensional data, so as to provide information on different level.

Keywords: Data Mining, Market Basket Analysis, Multi Level Association Rule and Data Mart.

I. PENDAHULUAN

Swalayan ASGAP, sejak berdiri pada tahun 1997 telah melakukan komputersisasi dalam transaksi penjualannya. Analisis telah dilakukan berdasarkan laporan transaksi bulanan, besarnya volume data yang perlu diolah, menyebabkan swalayan ASGAP mengalami kesulitan dalam meningkatkan penjualan. Terkadang hasil pengolahan data dengan cara sederhana tidak memberikan hasil apapun. Kebijakan akhirnya ditempuh berdasarkan informasi singkat berupa laporan yang didukung dengan intuisi manajer. Keberadaan informasi diharapkan 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 kategori 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.

II. TINJAUAN PUSTAKA

A. Market Basket Analysis

Market Basket Analysis adalah suatu cara yang digunakan untuk menganalisis data penjualan dari suatu perusahaan. Proses ini menganalisis *buying habits* 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 bisa saja terdapat suatu pola pembelian *item* yang tidak terpikirkan sebelumnya, misalnya pembelian minyak goreng dengan deterjen. 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 *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

1. *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).
2. *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-filter itemset* 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. 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.

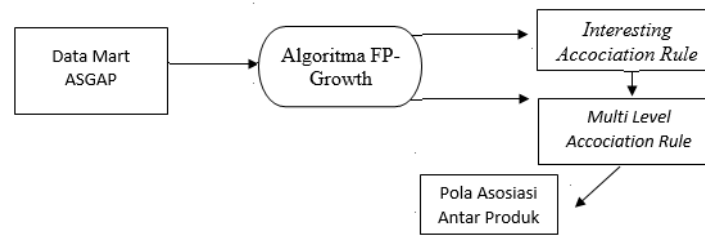
Karakteristik Data mart:

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

Terkadang sulit untuk membedakan antara data warehouse dan data mart karena keduanya hampir sama. Namun, jika dikaji lebih jauh ada beberapa perbedaan yang dimiliki keduanya. Data warehouse merupakan gabungan dari beberapa data mart dan levelnya berada pada perusahaan atau organisasi. Sedangkan data mart merupakan bagian dari data warehouse 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.

III. DESAIN SOFTWARE

Dataset yang digunakan dalam penelitian 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 juga bersumber dari Clipper terdiri dari master penjualan, master transaksi dan master lokasi.



Gambar 1 : Proses Mining Data Mart ASGAP

Desain Software diawali dengan penyiapan data mart kemudian dilanjutkan penentuan himpunan data yang paling sering muncul dalam sekumpulan data dengan menggunakan algoritma FP-Growth selanjutnya himpunan data dikodekan sehingga menghasilkan *encoded transaction table* dan dilakukan perhitungan untuk menentukan *interesting accociation rules* setelah melewati semua proses itu, maka akan menghasilkan pola asosiasi antar produk.

Preprocessing

Tahapan untuk *preprocessing* (persiapan data) tujuannya untuk pemilihan dan pemindahan data, kemudian menggabungkan sumber-sumber data yang ada supaya dapat mempermudah proses menampilkan data dalam jumlah yang lebih mudah dibaca. Proses untuk pembersihan data yang tidak berguna misalnya ada data yang tidak ada tanggalnya atau data kosong tetapi ada nomor notanya. Dalam proses diperlukan kehati-hatian yang sangat tinggi apabila melakukan kesalahan pada data set maka akan mempengaruhi *output klasifikasi*.

Semua data berasal dari data mentah dan hasilnya akan menjadi data yang nantinya siap untuk diolah menjadi *data mart*, spesifikasi data sebagai berikut:

- Jumlah record tabel master barang sebanyak 3,442.
- Jumlah record tabel transaksi penjualan sebanyak 19.929.
- Jumlah record master lokasi sebanyak 2 .

Design Data Mart

Data mart adalah dataset yang mempunyai model dimensional dan merupakan rancangan logika yang bertujuan untuk menampilkan data dalam bentuk standart dan intuitif yang memperbolehkan akses dengan performa yang tinggi. Model dimensional menggunakan konsep model hubungan antar entity (ER) dengan beberapa batasan yang penting. Setiap model dimensi terdiri dari sebuah tabel dengan sebuah komposit *primary key*, disebut dengan tabel fakta dan satu set tabel yang lebih kecil disebut tabel dimensi.

Selanjutnya dari ketiga tabel tersebut direlasikan kemudian dibentuk sebuah data mart yang mempunyai dimensi produk yang berisi nama produk dalam transaksi serta total penjualan mulai Januari sampai Desember. Dari data mart tersebut bisa di *drill down* berdasarkan produk, juga di *drill down* berdasarkan kategori, selanjutnya di *drill down* berdasarkan bulan penjualan, di *drill down* lagi berdasarkan tanggal transaksi, di *drill down* berdasarkan lokasi.

Frequent Itemset

Data mart dijadikan acuan untuk mencari *frequent itemset* (item yang sering muncul). Untuk menghitung itemset yang sering muncul tidak berdasarkan banyaknya

jumlah item yang dibeli dalam satu nota dalam satu item. Sehingga didapat item-item yang sering muncul dalam penelitian ini dibatasi sampai 100 kali kemunculan saja. Jadi perhitungan keseluruhan transaksi dengan minimum support =100 yang berarti minimal muncul 100 kali dalam suatu transaksi.

Interestingness Measure

Dari hasil kombinasi antar item kemudian dihitung kembali dengan *Interestingness Measure* (ukuran kepercayaan), perhitungan ada dua macam.

Support: ukuran yang menunjukkan seberapa besar tingkat dominasi suatu itemset dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu itemset layak untuk dicari *confidence* nya, dengan rumus perhitungan sebagai berikut:

$$\text{Nilai Support } (A \Rightarrow B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\% \quad (1)$$

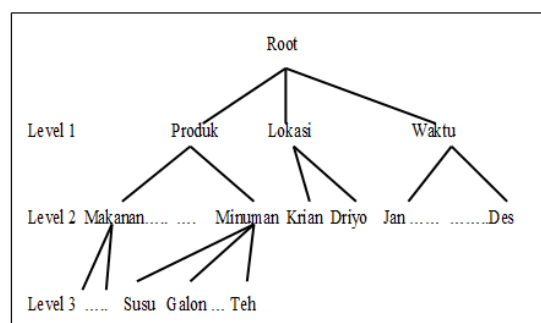
Confidence: yaitu suatu ukuran yang menunjukkan hubungan antar dua item secara *conditional* artinya seberapa sering item B dibeli jika orang membeli item A.

$$\begin{aligned} \text{Nilai Confidence } (A \Rightarrow B) &= P(B|A) \\ &= \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}} \times 100\% \quad \dots\dots\dots (2) \end{aligned}$$

Kedua ukuran ini nantinya berguna dalam menentukan *interesting accociation rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh user. Batasan tersebut umumnya terdiri dari *minimum support* dan *minimum confidence*.

Multilevel Association Rule

Untuk mendapatkan *Multilevel Accociation Rule* perlu terlebih dahulu dibentuk suatu *Concept Hierarchy Tree* dari data-data yang ada kemudian menyusunnya ke dalam sebuah *generalized description table*. Selanjutnya datamart yang akan dimining dirubah atau ditransformasikan menjadi *encoded transaction table* dimana item-item yang muncul pada *concept hierarchy* dari sebuah transaksi dikodekan sesuai dengan nilai *GID* nya pada *geralized description table*. Dalam menggali *Multilevel Accociation Rule* kuncinya adalah dengan membangun sebuah *concept hierarchy tree* dari sebuah predikat.



Gambar 2. Concept Hierarchy Tree

Dari pohon skema pada gambar *Concept Hierarchy Tree* selanjutnya diubah dalam bentuk tabel yaitu sebuah *generalized description table*, seperti terlihat pada Tabel 1 (Scott Fortin, Ling Liu, Randy Goebel 1996).

Tabel 1. Generalized Description Table

Tanggal	Kd_brg	Jenis	Merk	Lokasi
01-01-2010	AQUA G	MINUMAN	AQUA	KRIAN
01-01-2010	HANDY A	ATK	HANDY	DRIYOREJO
01-01-2010	LPG	GAS	LPG	DRIYOREJO
01-01-2010	MOLTO FI	PEWANGI	MOLTO	KRIAN
01-01-2010	CDR	OBAT	CDR	KRIAN
01-01-2010	DJISAMSU	ROKOK	DJISAMSU	DRIYOREJO
01-01-2010	ANTI	OBAT	ANTIMO	KRIAN
01-01-2010	ANTIM	OBAT	ANTIMO	KRIAN
01-01-2010	BLU	GAS	BLUE	KRIAN
01-01-2010	BUKU FOLIO	BUKU	FOLIO MIRAGE	KRIAN

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 kode-nya pada *generalized description*. Langkah berikutnya adalah melakukan *join* tabel transaksi yang biasanya bersifat multi dimensi dengan tabel detail transaksi yang biasanya bersifat *multi level* namun *single* dimensi, seperti yang terlihat pada tabel di bawah ini:

Tabel 2. Tabel Transaksi Penjualan

Nota	Tanggal	Lokasi
1	01-01-2010	DRIYOREJO
2	01-01-2010	KRIAN
3	01-01-2010	DRIYOREJO
4	01-01-2010	DRIYOREJO
5	01-01-2010	DRIYOREJO
6	01-01-2010	DRIYOREJO
7	01-01-2010	DRIYOREJO
8	01-01-2010	DRIYOREJO
9	01-01-2010	DRIYOREJO
10	01-01-2010	DRIYOREJO

Proses akhir dari *encoded transaction table* adalah perhitungan untuk menentukan *interesting accociation rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) terdiri dari *minimum support* dan *minimum confidence*. Dengan cara menghitung itemset yang sering muncul tidak berdasarkan banyaknya jumlah item yang dibeli dalam satu nota dalam satu item dengan rule sebagai berikut :

- Jika hanya dua kombinasi item, maka hitung kategori pertama, misalnya item yang muncul "minuman" kemudian hitung pada kolom pertama yang mengandung kategori "minuman" kemudian hitung jumlah transaksi keseluruhan kemudian jumlah kolom pertama dibagi dengan jumlah total transaksi dikalikan 100%.
- Jika muncul tiga kombinasi item, maka gabungkan kategori pertama dan katagori kedua, misalnya item yang muncul "minuman" dan "minyak" kemudian hitung pada kolom pertama dan kolom kedua yang yang mengandung kategori "minuman" dan "minyak" lalu hitung jumlah traksaksi keseluruhan kemudian

jumlah kolom pertama dan kolom kedua dibagi dengan jumlah total transaksi dikalikan 100%.

Rule tersebut juga akan digunakan untuk menghitung seluruh Proses (pengelompokan) yang terdiri diantaranya:

Proses 1 : Hasil perhitungan *multi level association rule* dengan pengelompokan berdasarkan Jenis (katagori) dengan output besaran nilai support dan confidence.

Proses 2 : Hasil perhitungan *multi level association rule* dengan pengelompokan berdasarkan Merk Barang dengan output besaran nilai support dan confidence.

Proses 3 : Hasil perhitungan *multi level association rule* dengan pengelompokan berdasarkan Nama Barang dengan output besaran nilai support dan confidence.

Dari pemrosesan mulai dari proses 1 sampai proses 3 diperoleh pengetahuan berupa pola asosiasi antara satu produk dengan produk yang lain dengan hasil sebagai berikut:

- Konsumen yang membeli barang dengan katagori "MINUMAN" cenderung membeli barang dengan katagori "MINYAK"
[support : 3.97% confidence : 0.51%]
- Konsumen yang membeli barang dengan katagori "MINUMAN" cenderung membeli barang dengan katagori "ROKOK"
[support : 3.97% confidence : 22.10%]
- 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

IV. 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. Hasil dari evaluasi memberikan hasil yang memuaskan, di mana aturan asosiasi memberikan manfaat yang lebih besar daripada tidak menggunakan aturan sama sekali.

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.

V. DAFTAR PUSTAKA

- [1] Gregorius S. Budhi, Resmana Lim , *Penggunaan Metode Fuzzy C-covering untuk Analisa Market Basket Padasupermarket*, <http://puslit.petra.ac.id/journals/informatics/>
- [2] David Ulmer, *Mining an Online Auctions Data Warehouse*, Proceedings of MASPLAS'02 The Mid-Atlantic Student Workshop on Programming Languages and Systems Pace University, April 19, 2002.
- [3] Pratima Gautam, Dr. K. R. Pardasani. *Algorithm for Efficient Multilevel Association Rule Mining*. Vol. 02, No. 05, 2010, 1700-1704.
- [4] Scott Fortin, Ling Liu, Randy Goebel. *An Object-Oriented Approach to Multi-Level Association Rule Mining*.
- [5] Yinbo WAN, Yong LIANG, Liya DING. *Mining Association Rules From Primitive Frequent Itemset*. Vol. 3 No. 1. June 30, 2009.
- [6] 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).