**ANALISA DATA DENGAN TEKNIK ASSOCIATION RULE DALAM DATA MINING**

Ermatita1, Saparudin2

Fakultas Ilmu Komputer Universitas Sriwijaya

Palembang, Indonesia

e-mail: ermatitaz@yahoo.com

ABSTRACT

Retrieving data from large databases is very important for company or organization. The development of information technology such as data mining has helped in the process of information retrieval from a large database. The process of information retrieval with data mining has been offering a variety of techniques. One technique that can be used in the excavation and retrieval of information from these databases is the Association's rules. Association rule techniques help dig up information that can be used in developing the company or for the purposes of decision making in running the company activity. In this paper simulation of an association rule is made in making information on a business activity, so it can provide what the decision will be made for business development.

Keywords: Association Rule, Data mining, information retrieval, Decision Making

INTISARI

Data mining membantu analisis untuk memperkirakan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting dalam menunjang aktivitas dan pengembangan perusahaan. Analisis dengan data mining dilakukan dengan otomatisasi sehingga dapat mengurangi penggunaan waktu dan biaya yang tinggi. Data Mining mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi untuk memprediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka.Perkembangan *data mining* (DM) yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi.

Berbagai teknik dalam proses data mining telah berkembang, salah satu teknik yang banyak digunakan adalah teknik association rule.Teknik Association Rule dapat membantu menggali informasi pola pembelian pelanggan, sehingga dapat di buat suatu keputusan untuk mengembangkan bisnis.

1. **PENDAHULUAN**

Perkembangan teknologi telah membawa dampak yang sangat besar di berbagai sisi kehidupan manusia. Pengolahan data yang merupakan aset bagi perusahaan ataupun organisasi sudah menjadi kegiatan yang sangat penting untuk menunjang aktivitas dan kemajuan bagi suatu perusahaan atau organisasi. Data Mining merupakan salah satu perkembangan teknologi yang sangat berguna untuk membantu perusahaan atau organisasi dalam mengolah data dan menggali informasi yang sangat di butuhkan untuk pengembangan perusahaan atau organisasi.

Data mining membantu analisis untuk memperkirakan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting dalam menunjang aktivitas dan pengembangan perusahaan. Analisis dengan data mining dilakukan dengan otomatisasi sehingga dapat mengurangi penggunaan waktu dan biaya yang tinggi. Data Mining mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi untuk memprediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka.Perkembangan *data mining* (DM) yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi.

Berbagai teknik dalam proses data mining telah berkembang, salah satu teknik yang banyak digunakan adalah teknik association rule.

1. DATA MINING DAN TEKNIK ASSOCIATION RULE
2. **Data Mining**

Data mining adalah Ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar. Beberapa definisi data mining seperti berikut ini:

* Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.
* Data mining adalah analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya

Dari definisi-definisi itu, dapat dilihat ada beberapa faktor yang mendefinisikan data mining :

1. data mining adalah proses otomatis terhadap data yang dikumpulkan di masa lalu
2. objek dari data mining adalah data yang berjumlah besar atau kompleks
3. tujuan dari data mining adalah menemukan hubungan-hubungan atau pola-pola yang mungkin memberikan indikasi yang bermanfaat . (Pramudiono,2006)

Jadi data mining merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis dalam gudang data untuk menemukan keteraturan, pola atau hubungan dalam sekumpulan data yang berukuran besar.

Data mining merupakan bagian integral dari Knowledge discovery dalam databases. Proses keseluruhan dalam data mining dan Knowledge discovery adalah mengubah raw data menjadi informasi yang berguna. Berikut ini adalah proses data preprocessing ke postprocessing pada data mining

Postprocessing

Data

Mining

Data

Preprocessing

Input data Information

Filtering Patterns visualization

Pattern Interpretation

Feature selection Dimensionality reduction Normalization

Data Subsetting

Gambar.1. Proses pada Knowledge discovery in Databases (KDD).(Tan, 2006)

Proses preprocessing adalah mentransformasi data input ke dalam format yang tepat untuk analisis data. Tahapan dalam *preprocessing* data menggabungkan data dari *multiple sources*, *cleaning* data untuk menghilangkan noise dan observasi ganda, pemilihan record dan *features*. Tahapan *Post processing* memastikan bahwa hanya hasil yang valid dan bermanfaat dalam sistem pengambilan keputusan. Keluaran dari data mining dapat dipakai untuk memperbaiki dalam pengambilan keputusan di masa depan.

Dari gambar di atas proses dalam data mining adalah memproses data dengan teknik dalam data mining yang menghasilkan keluaran berupa informasi yang dapat digunakan dalam pengambilan keputusan. Salah satu teknik dalam proses data mining adalah *Association Rule*.

1. **Association Rule**

Association Rule adalah teknik data mining untuk menemukan aturan assosiatif dengan kombinasi item. Asosiasi adalah aturan yang mendukung pengolahan data, mengidentifikasi i korelasi (generalisasi asosiasi), dan mempertimbangkan urutan pengambilan materi sebagai suatu basis pembangkit aturan.

Menurut Agrawal dan Srikant dalam (Ridowati,2003), masalah dalam pencarian seluruh *association rule* dapat dibagi menjadi dua, yaitu bagimana menemukan seluruh item-item (itemset) yang memiliki support di atas minimum support yang telah ditentukan. Support untuk sebuah itemset adalah jumlah transaksi yang terdapat dalam itemset itu. ltemset dengan minimum support disebut large itemset (I-itemset). Masalah yang kedua adalah bagaimana meggunakan large itemset untuk menghasilkan aturan-aturan yang diperlukan. Aturan yang hendak dihasilkan adalah dalam bentuk a→ (I-a), yaitu perbandingan antara support (I) dengan support (a) haruslah lebih besar dari minconf (minimum confidence). Misalkan [I]={I1,I2,..Im) adalah hirnpunan literal yang disebut item. Himpunan item-item disebut dengan itemset. D adalah himpunan transaksi, di mana setiap transaksi T itemset sehingga ***TE.*** I. Sebuah association rule adalah sebuah implikasi berbentuk X → Y, di mana XСI, Y С I,danX X∩ Y=∞.

Ada dua pengukuran yang penting untuk association rule, yaitu:

1. Support (dukungan)

Aturan X → Y memiliki supports dalam transaksi D jika s% dari transaksi berisi Xu Y. Atau

dengan kata lain, support untuk X→ Y adalah support untuk sekumpulan item Xu Y (item

x→Y).

2. Confidence (kepercayaan)

Confidence untuk aturan X → Y adalah prosentase dari transaksi dalarn D yang berisi item

X juga berisi item Y.(Gunawan,R,2003)

Misalkan analisis terhadap transaksi di supermarket “suatu pelanggan yang membeli item X dan kemudian membeli item Y misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Di sini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang database transakasi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa

Banyak teori dan pendekatan yang dikembangkan untuk memperoleh hasil penemuan kaidah asosiasi dan pola. Assosiation rule merupakan studi tentang hubungan antara ‘apa bersama apa’. Dengan teknik ini database transaksi pelanggan dapat di lihat untuk menentukan kebiasaan suatu produk dibeli bersama produk apa Ketersediaan database mengenai catatan transaksi bisnis mendorong pengembangan teknik-teknik yang secara otomatis menemukan assosiasi produk atau item-item yang tersimpan dalam database tersebut disebut dengan Association rule.

Dalam tahapan pada proses Association rule terdapat berbagai istilah sebagai berikut:

1. I adalah himpunan yang tengah dibicarakan.

Contoh:

{Gula, Kopi, Teh, Susu, … , Ginseng}

1. D adalah Himpunan seluruh transaksi yang tengah dibicarakan

Contoh:

{Transaksi 1, transaksi 2, …, transaksi 14}

1. Proper Subset adalah Himpunan Bagian murni

Contoh:

Misalkan himpunan A={a,b,c,}

1. Himpunan Bagian dari A adalah
2. Himpunan Kosong = {}
3. Himpunan 1 Unsur = {a},{b},{c}
4. Himpunan 2 Unsur = {a,b},{a,c},{b,c}
5. Himpunan 3 Unsur = {a,b,c,}
6. Proper subset nya adalah Himpunan 1unsur dan Himpunan 2 Unsur
7. Item set adalah Himpunan item atau item-item di I

Contoh:

Ada suatu himpunan A={a,b,c,}

1. Item set nya adalah

{a};{b}:{c};{a,b};{a,c};{b,c}

1. K- item set adalah Item set yang terdiri dari K buah item yang ada pada
2. K merupakan jumlah unsur yang terdapat pada suatu Himpunan

Contoh:

3-item set adalah yang bersifat 3 unsur

1. Frekuensi item set adalah Jumlah transaksi di I yang mengandung jumlah item set tertentu.

Contoh:

1. frekuensi Item set yang sekaligus membeli Teh dan gula adalah 3, berarti ada tiga transaksi yang membeli teh sekaligus juga membeli gula
2. Frekuen Item Set adalah item set yang muncul sekurang-kurangnya “Ф” kali di D. Ф merupakan batas minimum dalam suatu transaksi

**Data mining dengan teknik Association Rule membantu :**

* memanfaatkan data transaksi penjualan untuk dianalisis sehingga dapat ditemukan pola berupa item-item yang cenderung muncul bersama dalam sebuah transaksi.
* untuk menganalisis kebiasaan pelanggan dalam menyimpan item-item yang akan dibeli ke dalam keranjang belanjaannya
* memanfaatkan data transaksi penjualan untuk dianalisis sehingga dapat ditemukan pola berupa item-item yang cenderung muncul bersama dalam sebuah transaksi
1. **PEMBAHASAN**

Data transaksi dari suatu perusahaan yang bergerak di bidang penjualan seperti supermarket saat ini sudah menjadi aset yang sangat berguna bagi perusahaan. Ketersediaan database mengenai catatan transaksi pembelian para pelanggan telah mendorong pengembangan dan pemakaian teknik-teknik yang secara otomatis menemukan asosiasi produk atau item-item yang tersimpan dalam database . Teknik Association rule telah banyak digunakan untuk kepentingan ini.

Database yang mengandung record dalam jumlah sangat besar dalam suatu supermarket dapat di manfaatkan untuk pengembangan bisnis. Setiap record dalam database mempunyai daftar semua item yang dibeli oleh seorang pelanggan secara bersamaan dalam suatu transaksi pembelian. Untuk mengetahui apakah suatu kelompok item selalu dibeli secara bersama-sama. Aturan asosiasi memberikan informasi hubungan ‘if-then’ atau ‘jika-maka’ antara satu item dengan item yang lainnya.

Frequen item set dengan satu item harus dilakukan untuk menghitung transaksi yang mengandung item-item tertentu. Setiap item di hitung , berapa banyak transaksi yang mengandung item tersebut. Jumlah transaksi yang didapat adalah support untuk set satu item tersebut. Selanjutnya kita bisa menghilangkan set satu item yang nilai supportnya sesuai dengan nilai yang telah kita tetapkan untuk mendapatkan daftar frequen item set dengan satu item. Sebagai ilustrasi berikut ini akan dicontohkan penggunaan tahapan dalam association rule dalam menemukan rule dalam data transaksi penjualan. Data di simulasikan pada transaksi penjualan di suatu supermarket:

Tabel. 1 Transaksi Penjualan di supermarket

|  |  |
| --- | --- |
| **Transaksi** | **Keterangan** |
| 1 | Bawang Merah,Bawang Putih, Sawi Hijau |
| 2 | Barang Putih, Tomat |
| 3 | Bawang Putih, Buncis |
| 4 | B awangMerah, Bawang Putih, Tomat |
| 5 | Bawang Merah, Buncis |
| 6 | Bawang Putih, Buncis |
| 7 | Bawang Putih, Tomat |
| 8 | Bawang Merah, Bawang Putih, Buncis, Sawi Hijau |
| 9 | Bawang Merah, Bawang Putih, Buncis |
| 10 | Cabe |
|  |  |
|  |  |

* 1. Tahapan awal yang dilakukan dalam Association Rule adalah : mendaftar masing-masing item seperti berikut ini:

Tabel 2. Masing-masing item

|  |
| --- |
| Item yang dibeli |
| Bawang Merah |
| Bawang Putih |
| Sawi Hijau |
| Tomat |
| Buncis |
| Cabe |

* 1. Menghitung jumlah banyaknya pembelian untuk setiap item
* Tabel. 3 Transaksi tiap Item

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Transaksi | Bawang Merah | Bawang Putih | Sawi Hijau | Tomat | Buncis | Cabe |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 | 0 | 1 | 0 |
| 4 | 1 | 1 | 0 | 1 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 1 | 0 | 0 | 1 | 0 |
| 7 | 0 | 1 | 0 | 1 | 0 | 0 |
| 8 | 1 | 1 | 1 | 0 | 1 | 0 |
| 9 | 1 | 1 | 0 | 0 | 1 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 1 |
| ∑ | 5 | 8 | 2 | 3 | 5 | 1 |

Item set Frekuensi adalah Jumlah transaksi di I yang mengandung jumlah item set tertentu. Intinya jumlah transaksi yang membeli suatu item set.

Untuk k = 2, diperlukan tabel untuk tiap pasang item sehingga didapat tabel pasangan untuk masing-masing item, contoh salah satu pasangan item untuk 2 item dari tabel transaksi di atas::

- frekuensi Item set yang sekaligus membeli bawang merah dan bawang putih dapat dilihat dari tabel di bawah ini :

Tabel.4 Transaksi Bawang merah dan bawang putih

|  |  |  |  |
| --- | --- | --- | --- |
| Transaksi | Bawang merah | Bawang putih | F |
| 1 | 1 | 1 | P |
| 2 | 0 | 1 | S |
| 3 | 0 | 1 | S |
| 4 | 1 | 1 | P |
| 5 | 1 | 0 | S |
| 6 | 0 | 1 | S |
| 7 | 0 | 1 | S |
| 8 | 1 | 1 | P |
| 9 | 1 | 1 | P |
| 10 | 0 | 0 | S |
|  |  | ∑P | 4 |

 Dari tabel di atas frequensi item set untuk dua item bawang merah dan bawang putih adalah 4. Selanjutnya ditentukan Frequent item set. Frekuen Item Set adalah frequensi item set yang muncul biasanya disimbolkankan dengan Ф. Ф merupakan batas minimum dalam suatu transaksi berikut ini contoh perhitungan frequent item set (Ф):

Frequent item set kita tentukan Ф = 3. Jika Ф=3 untuk {bawang merah, Bawang putih} maka frequent item setnhya akan di hitung sebagai berikut: Dari tabel di atas jumlah transaksi yang membeli bawang merah sekaligus membeli bawang putih adalah 4

Karena 4 >= 3 maka {Bawang merah, Bawang putih} merupakan Frekuen Item set.

Rule ditentukan dengan menggunakan rule ***If x Then y,*** dimana x adalah Antecedent dan y adalah Consequent. Dari rule di atas dibutuhkan 2 item , yaitu satu item sebagai antecedent dan yang lainnya sebagai Consequent.

Untuk {bawang merah, Bawang putih}, kita misalkan (antecedent) = bawang merah, dan consecuent = Bawang putih

Maka => *If Buy Bawang merah then BuyBawang putih*

Langkah selanjutnya adakah menentukan *Support* dan *Confidence.* *Support* dan *confidence* dengan cara-cara sebagai berikut:

Support di hitung dengan rumus seperti berikut**:**

Untuk Σ Item yang dibeli sekaligus contohnya pada tabel 2 unsur untuk {Bawang merah ,Bawang putih}, ada 4 dari 10 transaksi sehingga Support nya adalah :

 4\_ X 100% = 40%

 10

Didapat Support = 40 %

 Sedangkan Confidence di hitung dengan rumus sebagai berikut:

Untuk Σ Item yang dibeli sekaligus contohnya pada tabel di atas unsur untuk

{Bawang merah, Bawang putih}, ada 4 sedangkan yang membeli bawang merah adalah 5 sehingga untuk Confident dihitung sebagai berikut:

Setelah support dan confident didapat lalu dikalikan sehingga didapat prosentase *support* dan *confident*. Hal ini dilakukan untuk semua kemungkinan item set. Untuk mendapatkan rule yang didapat adalah yang menghasilkan nilai *support confident* paling tinggi prosentasenya.

Tahapan dalam teknik *Association Rule* tersebut di atas akan menemukan pola. Selanjutnya pola yang ditemukan dapat dimanfaatkan untuk merancang strategi penjualan atau pemasaran yang efektif, yaitu dengan menempatkan item-item yang sering dibeli bersamaan ke dalam sebuah area yang berdekatan, merancang tampilan item-item di katalog, merancang kupon diskon (untuk diberikan kepada pelanggan yang membeli item tertentu), merancang penjualan item-item dalam bentuk paket, dan sebagainya.

1. KKESIMPULAN

Pengambilan informasi dari databse dengan record yang sangat besar saat ini bukanlah hal yang mustahil lagi. Teknik dalam data mining telah banyak membantu dalam proses pengambilan dan penggalian informasi yang dibutuhkan oleh organisasi atau perusahaan. Penggalian informasi dari data yang sangat besar pada suatu perusahaan sangat penting. Hal ini dapat membantu dalam pengembangan bisnis dan menunjang setiap aktivitas dalam perusahaan.

Teknik Association rule sebagai salah satu teknik dalam penggalian informasi dalam database, telah banyak digunakan untuk melakukan analisis terhadap data untuk memperkirakan pola dan tren bisnis yang dapat membantu dalam pengembangan bisnis.

Abdullahtoni, **Metode Association Analysis,** <http://abdullahtoni.wordpress.com/category/association-rule/> di akses 15 September 2010

Han, J and Kamber, M, 2001, *Data mining : Concepts and Techniques.* Morgan Kaufmann

Gunawan,R,2003, Pencarian Pola dari Data Web Server Log dengan Menggunakan Teknik Association Rule,  *SIGMA, Vol. 6, No.1,* Januari *2003: 43-50 ISSN: 141 0-5888*

Pramudiono, I, 2006, Indo data mining

 Santosa, B, 2007, Data Mining: Teknik Pemanfaatan Data untuk Keperluan bisnis Teori dan aplikasi, Graha Ilmu, Yogyakarta.

Tan P. N., Steinbach, M., Kumar, V., 2006, *Introduction to Data mining*, Addison Wesley.

<http://www.resample.com/xlminer/help/Assocrules/associationrules_intro.htm.......introduction>

[http://infolab.stanford.edu/~ullman/mining/assocrules.pdf ....assosiason](http://infolab.stanford.edu/~ullman/mining/assocrules.pdf%20....assosiason)

<http://www.eecs.qmul.ac.uk/~christof/html/courses/ml4dm/week10-association-4pp.pdf>

<http://id.wikipedia.org/wiki/Penggalian_data>

http://iatt.kemenperin.go.id/tik/fullpaper/fullpaper113\_widodo.pdf

Nama Lengkap ( dengan gelar ) : Dra. Ermatita, M.KOM

Tempat / Tanggal Lahir : Lahat, 13 September 1967

Alamat lengkap kantor : Jalan Palembang Prabumulih KM.32 Inderalaya

E-mail : ermatitaz@yahoo.com

Asal S1, Bidang Ilmu : Matematika

Asal S2, Bidang Ilmu : Ilmu Komputer

Asal S3, Bidang Ilmu : -

Spesialisasi dan minat keilmuan : RPL dan Sistem Informasi

Nama Lengkap ( dengan gelar ) : Drs. Saparudin, M.T

Tempat / Tanggal Lahir : Bangka, 12 April 1969

Alamat lengkap kantor : Jalan Palembang Prabumulih Km.32 Inderalaya

E-mail : saparudin@ilkom.unsri.ac.id

Asal S1, Bidang Ilmu : matematika

Asal S2, Bidang Ilmu : Informatika

Asal S3, Bidang Ilmu : -

Spesialisasi dan minat keilmuan : RPL