

PERANCANGAN DATA WAREHOUSE DAN PENERAPAN METODE ASSOCIATION UNTUK EKSTRAKSI DATA PENJUALAN PADA PT. CNI PALEMBANG

NurArifHidayat (Ndsarif@gmail.com), **Nurtriana** (Nurtriana@gmail.com)
Mardiani, S.Si, M.T.I. (Mardiani@stmik-mdp.net)
Jurusan Teknik Informatika
STMIK MDP

Abstrak: Data *warehouse* merupakan sebuah cara yang digunakan untuk menampung data dalam jumlah besar, untuk membangun sebuah data *warehouse* dibutuhkan sebuah *tools microsoft SQL server 2008, microsoft SQL business intelligence development dan WEKA (Waikato Environment for Knowledge Analysis)*. Dengan menggunakan *tools microsoft SQL server 2008, microsoft SQL business intelligence development dan WEKA (Waikato Environment for Knowledge Analysis)* dapat dibangun sebuah data *warehouse* penjualan dengan mengumpulkan data-data penjualan meliputi *payment, customers id, gender, age, occupation, product id, product name, product cost, dan agency id* yang dimasukkan kedalam sebuah *Ms. Excel* kemudian hasil dari *Ms. Excel* dimasukkan ke dalam *database microsoft SQL server 2008*, setelah data masuk ke dalam *database* kemudian data tersebut di *load* kedalam *microsoft SQL business intelligence development* untuk dianalisis kemudian dimining menggunakan *WEKA (Waikato Environment for Knowledge Analysis)*. Hasil analisis data *warehouse* dan data *mining* yang dibuat dapat memberikan informasi yang lebih jelas untuk pihak PT. CNI dan mempermudah dalam menganalisis barang-barang yang berpasangan dan *association rule* dari pasangan barang tersebut.

Kata kunci: *Data warehouse, data mining, microsoft SQL business intelligence development, microsoft SQL server 2008, WEKA, database, penjualan.*

1 PENDAHULUAN

Perkembangan dunia bisnis yang sangat pesat, mendorong terbentuknya suatu timbunan data-data yang berukuran sangat besar. Informasi yang didapat dari data dalam bentuk tabel-tabel tersebut kurang dapat dimanfaatkan oleh pihak manajemen perusahaan dalam pengambilan keputusan untuk kemajuan perusahaan. Oleh karena itu perlu adanya aktivitas penggalian (ekstraksi) data yang masih tersembunyi untuk selanjutnya diolah menjadi pengetahuan yang bermanfaat dalam pengambilan keputusan.

PT.CNI Palembang memiliki volume produk dan transaksi yang cukup besar, perusahaan ini juga memiliki beberapa cabang

lokasi sehingga menyulitkan perusahaan untuk mengetahui produk mana yang lebih cepat laku dan kurang diminati pelanggan per daerah.

2 LANDASAN TEORI

2.1 Data Warehouse

Menurut Feri Sulianta dan Dominikus dalam buku *Data Mining Meramalkan Bisnis Perusahaan* (2010, hal. 32), *data warehouse* adalah kumpulan data dari berbagai sumber yang ditempatkan menjadi satu dalam tempat penyimpanan berukuran besar lalu diproses menjadi bentuk penyimpanan multi-dimensional dan didesain untuk *querying* dan *reporting*.

Karakteristik *Data Warehouse* :

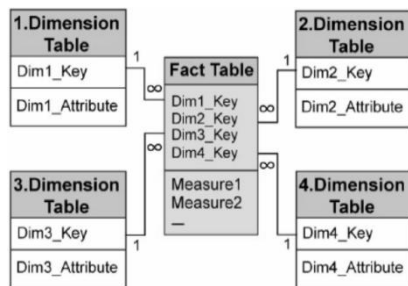
- Subject-oriented* (Berorientasi Subjek)
- Integrated* (Terintegrasi)
- Time Variant* (Rentang Waktu)
- Nonvolatile*

2.2 Dimensional Modelling

Beberapa konsep pemodelan data *Warehouse* pada *dimensionality modeling* yang dikenal pada umumnya, konsep-konsep tersebut adalah *star schema*, *snowflake* dan *fact constellation schema*.

2.2.1 Star Schema (Skema Bintang)

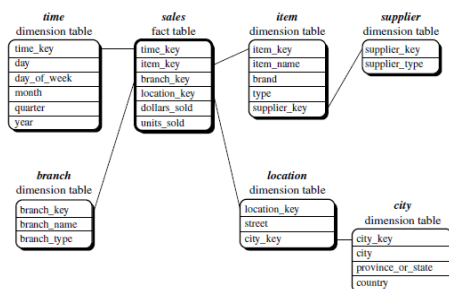
Skema bintang adalah sebuah logical struktur yang mempunyai sebuah table fakta berisi data terbaru di tengah, yang dikelilingi table dimensi yang berisi data referensi.



Gambar 1: Star Schema

2.2.2 Snowflake Schema

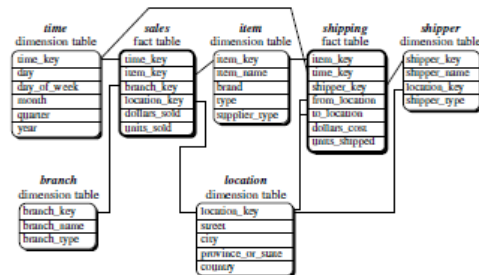
Skema kepingan salju adalah metode normalisasi table dimensi dalam skema bintang ketika anda benar-benar menormalkan semua table dimensi, struktur yang dihasilkan menyerupai kepingan salju dengan table fakta ditengah. Dapat dilihat pada gambar 2.5 (Ponniah 2011, h.259).



Gambar 2: Snowflake Schema

2.2.3 Fact Constellation Schema

Fact constellation schema adalah skema multi dimensional yang berisikan lebih dari satu tabel fakta yang saling berbagi tabel dimensi.



Gambar 3: Constellation Schema

2.3 Data Mining

Menurut Vercellis (2009, h.77) *Data mining* adalah sebuah proses berulang bertujuan untuk menganalisa *database* yang besar dengan tujuan penggalian informasi dan pengetahuan yang berkaitan dengan pengambilan keputusan dan pemecahan masalah.

2.4 ETL (Extract, Transform, Load)

a. Extraction

Menurut Golfarelli dan Rizzi (2009), *Extraction* adalah pengambilan data yang *relevant* atau berkaitan dari sumber data. *Extract* merupakan proses yang pertama kali dilakukan dalam pengisian data *warehouse*.

b. Cleansing

Proses *cleansing* sangatlah penting dalam sistem data *warehouse* untuk meningkatkan kualitas data, karena normalnya data dari sumber data berkualitas rendah.

c. Transformation

Transformation mengubah format data dari sumber data operasional menjadi format data *warehouse* yang lebih spesifik.

2.5 Algoritma Apriori

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Contoh aturan asosiatif dari analisis pembelian disuatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu.

Adapun tahap mencari kombinasi item yang memenuhi syarat minimum *support* dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut :

$$\text{Support}(A) = \frac{\text{JumlahtransaksimengandungA}}{\text{TotalTransaksi}}$$

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2 berikut:

$$\begin{aligned} \text{Support}(A, B) &= P(A \cap B) \\ \text{Support}(A, B) &= \frac{\sum \text{TransaksimengandungAdanB}}{\sum \text{Transaksi}} \end{aligned}$$

Aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$.

Nilai *confidence* dari aturan A B di peroleh dari rumus berikut :

$$\begin{aligned} \text{Confidence} &= P(B|A) \\ &= \frac{\sum \text{TransaksimengandungAdanB}}{\sum \text{TransaksimengandungA}} \end{aligned}$$

2.6 Weka (Waikato Environment for Knowledge Analysis)

Menurut (Feri Sulianta, hal.63) *Weka* merupakan aplikasi *datamining* yang berbasis *open source* (GPL) dan berengine *Java*. Pengertian *Weka* menurut situs resminya adalah sebagai berikut : *Weka* adalah koleksi dari algoritma pembelajaran mesin untuk tugas-tugas *data mining*. Algoritma dapat diterapkan secara langsung ke data set atau disebut kode *Java*. *Weka* berisi alat untuk data *pre-processing*,

classification, *regression*, *clustering*, *association rules*, and *visualization*. Hal ini juga cocok untuk mengembangkan skema pembelajaran mesin baru.

2.7 Nine-step Methodology

Nine-step Methodology (Connolly dan Begg, 2005, h.1187). Kesembilan tahap itu yaitu:

1. Pemilihan Proses
2. Pemilihan *Grain*
3. Identifikasi dan penyesuaian
4. Pemilihan Fakta
5. Penyimpanan *pre-calculation* di tabel
6. Memastikan tabel dimensi
7. Pemilihan durasi *database*
8. Melacak perubahan dari dimensi secara perlahan
9. Penentuan prioritas dan model *query*

2.8 CRISP-DM (Cross Industry Standard Process for Data Mining)

Dalam penerapan *Data Mining* digunakan metodologi *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) (Olson & Delen, 2008).

1. *Business understanding*
2. *Data understanding*
3. *Data preparation*
4. *Modeling*
5. *Evaluation*
6. *Deployment*

3 PERANCANGAN DATA WAREHOUSE

3.1 Perancangan Data Warehouse

Perancangan *Data Warehouse* pada PT. CNI Palembang menggunakan metodologi sembilan tahapan (*nine-step methodology*) atau biasa disebut dengan metode Ralph Kimbal, metode ini meliputi :

a. Pemilihan Proses

Proses ini meliputi pendataan tiap-tiap pelanggan yang membeli produk CNI dari setiap agen-agen CNI. Adapun data – data atribut yang digunakan antara lain : Tanggal transaksi penjualan, kelompok transaksi, ID

pelanggan, kelompok kelamin, umur dan pekerjaan dari pelanggan, ID agen, nama agen, kota penjualan agen, ID produk, nama produk, harga produk, kelompok produk, tingkatan kelompok produk, jumlah barang yang dibeli, jumlah pembayaran, pajak pertambahan nilai dan total pembayaran dari transaksi pembelian yang dilakukan.

b. Pemilihan Grain

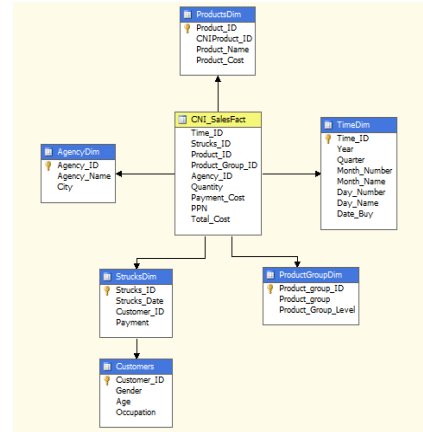
Grain merupakan proses untuk menentukan apa yang digambarkan oleh *record* di dalam tabel fakta. Terdapat empat perhitungan dalam tabel fakta yang akan dibuat di dalam *data warehouse* CNI ini yaitu *Quantity* atau jumlah barang, *payment cost* atau jumlah pembayaran tanpa PPN, adapun PPN atau Pajak Pertambahan Nilai dan *total cost* atau total pembayaran. Berikut adalah *grain* yang ada dalam perancangan *Data Warehouse* pada CNI Palembang.

c. Identifikasi dan Penyesuaian Dimensi

Pada tahap ini dilakukan identifikasi dan penyesuaian dimensi yang terkait dengan fakta yang ditampilkan dalam bentuk matriks. Dimensi yang digunakan dalam tabel fakta penjualan pada CNI Palembang adalah dimensi *time* (waktu), dimensi *strucks* (transaksi) dimensi *agency* (agen penjualan), dimensi *product* dan dimensi *product group*.

d. Pemilihan Fakta

Pada tahap ini dilakukan pemilihan fakta yang akan digunakan pada tabel fakta penjualan CNI. Adapun kolom – kolom pada tabel fakta penjualan CNI meliputi : *Time_ID*, *Strucks_ID*, *Product_ID*, *Product_Group_ID*, *Agency_ID*, *Quantity*, *Payment_Cost*, *PPN*, dan *Total_Cost*.



Gambar 4: Fakta Skema Snowflake

e. Penyimpanan Pre Calculation Ditabel Fakta

Dalam table fakta terdapat data yang merupakan kalkulasi awal. Hasil dari kalkulasi awal ini disimpan dalam tabel-tabel fakta. Adapun kalkulasi awal yang disimpan pada table CNI_SalesFact.

f. Memastikan Tabel Dimensi

Dalam tahap ini, kembali pada tabel dimensi dan menambahkan gambaran teks terhadap dimensi yang memungkinkan. Gambaran teks harus mudah digunakan dan dimengerti oleh user.

g. Pemilihan Durasi Database

Data yang dimasukan kedalam *Data Warehouse* ialah data transaksi penjualan produk pada PT.CNI Palembang selama tiga tahun.

h. Melacak Perubahan dari Dimensi Secara Perlahan

Dimensi-dimensi dan atribut yang mungkin dapat berubah dapat dilihat pada tabel.

Tabel 1: Kolom Dimensi yang Dapat Berubah

Nama Dimensi	Atribut yang Mungkin Berubah
AgencyDim	Agency_Name City
ProductDim	Product_Cost Product_Name
StrucksDim	Age Occupation

i. Penentuan Prioritas dan Model Query

Dalam tahapan ini yang dilakukan adalah mempertimbangkan pengaruh pada perancangan fisik, seperti keberadaan dari ringkasan (*summaries*) dan penjumlahan (*aggregate*). Dalam perancangan *data warehouse*, kapasitas media penyimpanan merupakan salah satu aspek yang perlu dipertimbangkan.

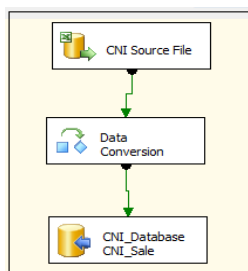
3.2 Extracting Data

Sumber data yang digunakan dalam perancangan *data warehouse* dan sumber data yang digunakan dalam perancangan *data warehouse* pada CNI Palembang ialah data *MS. Excel*.

Tabel 2: Sumber Data Penjualan CNI dalam Bentuk Ms. Excel

Payment	Customer_ID	Gender	Age	Occupation	Product id	Product Name	Product Cost	Agency_ID
Cash	CNIC135	P	42	Pegawai Swasta	HR025	CNI VITASIGI F RASA APEL	42000	CNIA221
Cash	CNIC135	P	42	Pegawai Swasta	N_07	PHYTO LITE SACHET	169000	CNIA221
Cash	CNIC135	P	42	Pegawai Swasta	HR024	CNI VITASIGI F RASA STRAWBERRY	42000	CNIA221
Cash	CNIC135	P	42	Pegawai Swasta	PC091	FIRE SENS NEW LABEL	23000	CNIA221
Cash	CNIC132	L	32	Wiraswasta	FD 385	SEHATIKECAP MANIS	148000	CNIA221
Cash	CNIC132	L	32	Wiraswasta	N_06	NARISH SACHET	175000	CNIA221
Cash	CNIC128	L	27	Wiraswasta	FD 03	PROGENE	610000	CNIA221

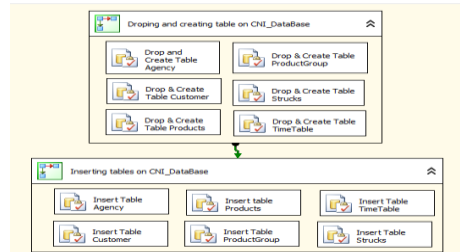
Kemudian data yang telah dikonversi akan dimasukkan kedalam Tabel CNI_Sale. Berikut gambar dari proses *Extracting Data*.



Gambar 5: Proses Ekstrak Data Ms. Excel

3.3 Transformation

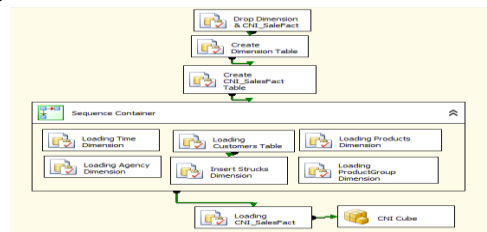
Setelah data-data berhasil dimasukkan ke dalam tabel CNI_sale pada *database* CNI, isi dari tabel CNI_sale akan diisi ke beberapa tabel baru yang dibuat di dalam *database* CNI yaitu : tabel Agency, Customer, Products, Strucks, ProductGroup dan TimeTable.



Gambar 6: Paket Transformasi pada SSIS

3.4 Loading Tabel Dimensi dan Tabel Fakta

Proses ini merupakan tahapan pembuatan tabel-tabel dimensi pada *database* DW_CNI yang bertujuan untuk mengisi tiap tabel-tabel dimensi, setiap isi tabel dimensi berasal dari tabel *database* CNI yang sudah terpisah. Berikut gambar proses *loading* tabel dimensi dan tabel fakta :



Gambar 7: Proses Loading Dimensi dan Fakta

4 ANALISIS DATA WAREHOUSE DAN HASIL DATA MINING

4.1 Presentasi Data Warehouse

Adapun hasil informasi yang didapatkan pada *data warehouse* PT. CNI Palembang yang berisikan data-data penjualan produk selama 3 tahun terakhir dari tahun 2011-2013 untuk setiap periode waktu, antara lain :

confidence 0,75, “Jika membeli Super MPC Sachet new maka membeli Narish Sachet”.

confidence 0,74, “Jika membeli Bio Plus 1.2 gr isi 10 stick maka membeli Progene”.

confidence 0,73, “Jika membeli Sigi-F Hand & Body Lotion (new bottle) maka membeli CNI Vitasigi F Rasa Strawberry”.

confidence 0,72, “Jika membeli CNI Sarung Sejahtera maka membeli Mie Ginseng Soto Ayam – Ekonomis”, “ Jika membeli Mie Ginseng Soto Ayam – Ekonomis maka membeli CNI Sarung Sejahtera”, dan “Jika membeli CNI Vitasigi F Rasa Strawberry maka membeli Sigi-F Hand & Body Lotion (New Bottle)”.

confidence 0,68, “Jika membeli Progene (Istrip) maka membeli Bio Plus 1.2 Gr Isi 10 Stick”.

Untuk melihat gambaran hasil prediksi yang telah dilakukan dengan menggunakan WEKA dapat dilihat pada gambar dibawah ini.

```
Best rules found:
1. SEHATI RECAP MANIS=c 1294 ==> NARISH SACHET=c 1037 conf:(0.8)
2. SUPER MPC SACHET new=c 1560 ==> NARISH SACHET=c 1169 conf:(0.75)
3. BIO PLUS 1.2 GR ISI 10 STICK=c 1565 ==> PROGENE (ISTRIP)=c 1165 conf:(0.74)
4. SIGI-F HAND & BODY LOTION (NEW BOTTLE)=c 1170 ==> CNI VITASIGI F RASA STRAWBERRY=c 861 conf:(0.73)
5. CNI SARUNG SEJAHTERA=c 1439 ==> MIE GINSENG SOTO AYAM - EKONOMIS=c 1042 conf:(0.72)
6. MIE GINSENG SOTO AYAM - EKONOMIS=c 1441 ==> CNI SARUNG SEJAHTERA=c 1042 conf:(0.72)
7. CNI VITASIGI F RASA STRAWBERRY=c 1183 ==> SIGI-F HAND & BODY LOTION (NEW BOTTLE)=c 861 conf:(0.72)
8. PROGENE (ISTRIP)=c 1717 ==> BIO PLUS 1.2 GR ISI 10 STICK=c 1165 conf:(0.68)
9. NARISH SACHET=c 2635 ==> SUPER MPC SACHET new=c 1169 conf:(0.44)
10. NARISH SACHET=c 2635 ==> SEHATI RECAP MANIS=c 1037 conf:(0.39)
```

Gambar 12: Rules hasil Pengujian

5 PENUTUP

5.1 Kesimpulan

- Terjadi penurunan drastis dua tahun sebelumnya pada agen ‘CV.ANEKA CEMERLANG’ Kota ‘Jambi’ dengan 1382 item produk terjual.
- Pada tahun 2011 dapat dilihat berdasarkan dari nama produk Super SC 88 Laundry Compound – Sachet yang jumlahnya ada 38 unit dengan total pembayaran sebesar Rp. 1.669.145,6876.

- Dapat dilihat tingkatan kelompok barang yang ke 3 pada tahun 2011 dan kelompok produk dari tahun 2011 sebanyak 159 item dengan total pembayaran sebesar Rp. 8.347.070,1097.
- Terjadi peningkatan jumlah penjualan produk pada Kota ‘Bengkulu’ sebesar 83.26% dari bulan Januari ke bulan Februari 2013 dan total pembayaran transaksi sebanyak 124,11%.
- Transaksi penjualan produk selama tiga tahun pada PT.CNI Palembang terdiri dari 8443 baris transaksi dan 293 kolom nama produk.
- Pembentukan *Association Rules* menggunakan *Algoritma Apriori* pada aplikasi *WEKA* dijalankan dengan *support* 0,1 dan *confidence* 0,25.
- Dapat dilihat jumlah penjualan produk dari pelanggan PT.CNI Palembang dengan status pekerjaan, umur, jenis kelamin perbulannya.

5.2 Saran

- Diharapkan data yang digunakan dalam pembentukan *datawarehouse* ini bukan hanya pada wilayah Sumatera Selatan dan sekitarnya saja, melainkan seluruh cabang penjualan PT. CNI yang ada di Indonesia.
- PT. CNI Palembang agar menambah proses lain selain penjualan barang, misalnya pembelian barang ataupun proses lainnya.
- Digunakan *software* data *mining* yang lain selain Weka misalnya Ms. Visual Studio, Rapidminer ataupun software lainnya dengan algoritma yang sama ataupun berbeda sehingga pihak PT. CNI dapat membandingkan *association rules* yang dihasilkan.

DAFTAR PUSTAKA

- [1] Connolly, Thomas, C. B 2010, *Database System :A Practical Approach to Design, Implementation, and Management Fifth Edition*, Pearson Education Inc.
- [2] Cuzzocrea, Alfredo 2011, *Data Warehouseing and Knowledge Discovery*. London : Springer-Verlag Berlin Heidelberg.
- [3] Indrajani 2009, *Sistem Basis Data dalam Paket Five In One*, PT. Elex Media Komputindo, Jakarta
- [4] Kimball, R, Margy R, Warren T, Joy M and Bob B 2008. *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing Inc, Canada
- [5] Komputer, Wahana 2010, *Panduan Belajar MySQL Database Server*, Media Kita, Jakarta Selatan
- [6] Kusrini 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta
- [7] Nuraida, Ida 2008, *Manajemen Administrasi Perkantoran*, Kanisius, Yogyakarta
- [8] Ponniah, Paulraj 2011, *Data Warehouseing*, John Wiley & Sons Inc, Canada
- [9] Rainaldi, Vincent 2008, *Building A Data Warehouse With Examples In SQL Server*, Apress
- [10] Sulianta, Feri dan Dominikus Juju 2010, *Data Mining Meramalkan Bisnis Perusahaan*, Elex Media Komputindo, Jakarta
- [11] Tantra, Rudi 2012, *Manajemen Proyek Sistem Informasi*, Andi Offset, Yogyakarta