

DATA WAREHOUSE MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK MENUNJANG ANALISIS DIVISI MARKETING DI PERUSAHAAN MULTIFINANCE

Suhari Pratama¹, Rinto Priambodo²

^{1,2}Fakultas Ilmu Komputer, Universitas Mercu Buana

Jl. Raya Meruya Selatan, Kembangan, Jakarta 11650

E-mail: pratama.arie@gmail.com ¹, rinto.priambodo@mercubuana.ac.id ²

ABSTRAK

Divisi Marketing pada sebuah perusahaan multifinance membuat laporan berkala dengan cara mengunduh beberapa laporan rinci dari sistem berjalan, kemudian diolah atau dikombinasikan untuk menghasilkan ringkasan laporan sebagai bahan analisis dalam menentukan strategi penjualan. Namun demikian laporan yang ada saat ini belum menggunakan metode tertentu untuk menghasilkan forecasting yang bertujuan untuk memprediksi penjualan berdasarkan dimensi struktur pembiayaan, jenis aset yang dibiayai dan profil debitur. Penelitian ini bertujuan untuk mengembangkan data warehouse sebagai solusi untuk pembuatan analisis forecasting dari beberapa klasifikasi dimensi data sebagai bahan analisis pada Divisi Marketing. Penelitian ini akan menghasilkan data warehouse yang akan digunakan untuk pembuatan laporan forecasting dari beberapa dimensi yang bisa dipilih untuk dibandingkan secara fleksibel dan dikalkulasi dengan algoritma Naive Bayes sehingga diperoleh prediksi data yang akurat sebagai referensi untuk Divisi Marketing dalam menentukan kriteria produk yang akan dijual berdasarkan histori data penjualan dari suatu periode. Dengan pemanfaatan data warehouse, penyajian data akan lebih cepat dan ringkas karena informasi dari beberapa sumber data sudah dikonversi secara otomatis oleh sistem berbasis OLTP (OnLine Transaction Processing) untuk menghasilkan data OLAP (OnLine Analytical Processing). Dengan pemanfaatan OLAP, data dalam jumlah besar bisa dianalisis dengan proses yang cepat dan akurat.

Kata Kunci: Laporan, data warehouse, OLTP, OLAP, Naïve Bayes, forecasting

1. PENDAHULUAN

1.1 Latar Belakang

PT. ITC Auto Multifinance di Indonesia merupakan salah satu perusahaan pembiayaan kendaraan roda empat yang memiliki beberapa cabang. Strategi bisnis terkait penjualan (*sales*) didelegasikan ke Divisi Marketing untuk menganalisis dan membuat beberapa variasi produk yang akan dipasarkan dan bisa bersaing dengan kompetitor dengan mengacu terhadap kebijakan-kebijakan yang berlaku di perusahaan.

Laporan analisis yang ada saat ini belum menggunakan metode tertentu untuk menghasilkan forecasting yang dibutuhkan oleh Divisi Marketing. Laporan juga tidak bisa diakses secara cepat menggunakan sumber data dari sistem yang berjalan.

Metode forecasting menggunakan algoritma Naïve Bayes dipilih karena lebih cocok dalam penerapan probabilitas dari beberapa dimensi data untuk bisnis perusahaan pembiayaan, karena menggunakan data penjualan pada pengalaman sebelumnya untuk memprediksi spesifik penjualan produk di masa yang akan datang.

Sementara itu data yang besar dan variatif membutuhkan waktu sekitar 1-2 hari kerja untuk diolah oleh Divisi Marketing dalam pembuatan laporan yang sesuai dengan keinginan pihak eksekutif. Dalam proses pengolahan laporan, saat ini dikerjakan oleh satu orang dengan cara mengunduh beberapa laporan detail dari sistem yang ada

kemudian laporan tersebut diolah menggunakan *tools office spread sheet* menjadi laporan ringkasan sesuai dengan kebutuhan rutin pihak eksekutif.

Laporan final berbentuk *file office (spreadsheet)* yang sudah dibuat untuk pihak eksekutif, disimpan di salah satu komputer *supervisor* Divisi Marketing dan sebagai *backup* data disimpan di *shared folder* pada *server file*.

Dalam mengambil keputusan dibutuhkan juga informasi yang jelas, mudah dimengerti, dan sesuai dengan kebutuhan dalam pengambilan keputusan. Untuk mendukung penyajian informasi seperti itu dibutuhkan *database* yang berisi data yang telah diolah dan dianalisis sesuai dengan kebutuhan pengambilan keputusan dalam pembuatan strategi pemasaran. Salah satu teknologi yang dapat memberikan solusi terhadap kebutuhan penyediaan data dan informasi adalah *data warehouse*.

Data warehouse merupakan sebuah metode pengelolaan data yang berguna untuk proses pengumpulan informasi dalam pengambilan keputusan. *Data warehouse* memberikan sebuah solusi untuk menyediakan data yang berguna untuk proses analisis dan pengambilan keputusan. *Output* dari *Data Warehouse* dapat menghasilkan informasi yang berguna dan pemakaian yang fleksibel, di mana pengguna dapat memilih jenis laporan yang ingin dihasilkan. Pemakaian teknologi *data warehouse* hampir dibutuhkan oleh setiap organisasi. *Data warehouse* memungkinkan

integrasi berbagai macam jenis data dari berbagai macam aplikasi atau sistem. Hal ini menjamin mekanisme akses menjadi cepat dan akurat bagi pihak eksekutif untuk memperoleh informasi, dan menganalisisnya sebagai bahan-bahan untuk pembuatan strategi bisnis. Manajemen *backup data* kehilangan data dan dari faktor keamanan data juga bisa lebih terjamin.

Keberadaan *Data Warehouse* sendiri berguna untuk meringankan kinerja dari database sistem berjalan yang digunakan untuk transaksi bisnis setiap harinya, karena beban kerja *database* yang diakses untuk mengunduh laporan dalam jumlah data yang sangat banyak oleh user akan dialihkan ke *Data Warehouse* sebagai penyedia data untuk pembuatan laporan.

1.2 Landasan Teori

1.2.1 Data Warehouse

Data warehouse adalah kumpulan data yang berorientasi subjek, terintegrasi, bervariasi waktu, dan *non-volatile*, yang mendukung proses pengambilan keputusan oleh manajemen (Triana, Y. S. & Susilo, A., 2017)

Data warehouse memiliki empat karakteristik utama (Henderi, Langgeng & Karwandi, 2016), yaitu :

1. *Subject-Oriented* adalah *data warehouse* diorganisasikan ke dalam subjek bisnis yang utama dibandingkan fungsi bisnis yang berjalan.
2. *Integrated* adalah data yang akan masuk ke dalam *data warehouse* dapat berasal dari berbagai sumber data yang berbeda dan data tersebut dapat memiliki format, nama field, dan ukuran yang berbeda.
3. *Time Variant* adalah sifat rentang waktu terjadi karena adanya perbedaan antara data operasional dan data informasional. Pada lingkungan *data warehouse*, data akurat selama periode waktu tertentu sehingga setiap *record* memiliki satuan waktu tertentu (hari, bulan atau tahunan).
4. *Non-Volatile* adalah *data warehouse* bersifat *non-volatile* artinya tidak mengalami perubahan. Dalam database, *data warehouse* hanya ada dua kegiatan manipulasi data, yaitu pemasukan (*loading*) dan pengaksesan data.

1.2.2 Metodologi Perancangan Data Warehouse

Tahapan perancangan *data warehouse* memiliki beberapa tahapan metodologi, yaitu :

1. *Choose the Process*, Pada tahap ini dilakukan pemilihan proses yang akan digunakan untuk *data warehouse*.
2. *Choose the Grain*, Memilih *grain* berarti menentukan apa yang akan diwakili atau dipresentasikan oleh sebuah tabel fakta. Setelah menentukan *grain* dari tabel fakta, selanjutnya dapat ditentukan tabel-tabel dimensi yang berhubungan dengan tabel fakta tersebut. *Grain*

pada tabel fakta juga menentukan *grain* tabel dimensi.

3. *Identify and Conform the Dimensions*, Mengidentifikasi dan menghubungkan tabel dimensi dengan tabel fakta. Dimensi merupakan kumpulan sudut pandang yang penting untuk menggambarkan fakta-fakta yang terdapat pada tabel fakta.
4. *Choose the Facts*, Pada tahap ini, tentukan *measure* yang dibutuhkan pada tabel fakta. Informasi apa saja yang ingin dibutuhkan
5. *Store Precalculations in the Fact Table*, Pada tahap ini, hasil perhitungan pada suatu atribut perlu dipertimbangkan untuk disimpan di database.
6. *Round Out the Dimension Tables*, Dari dimensi-dimensi yang telah diidentifikasi, dibuat deskripsi yang memuat informasi terstruktur mengenai atribut-atribut pada tabel dimensi dan harus diberi keterangan secara lengkap dan mudah dipahami oleh pengguna.
7. *Choose the Durations of the Database*, Durasi waktu dari data-data yang akan dimasukkan ke dalam *data warehouse* akan ditentukan pada tahap ini.
8. *Determine the Need to Track Slowly Changing Dimensions*, Dimensi dapat berubah dengan lambat dan menjadi sebuah masalah. Terdapat tiga tipe dasar dari perubahan dimensi yang lambat, yaitu Menulis ulang atribut yang berubah, Membuat *record* baru pada dimensi atau Membuat suatu atribut alternatif untuk menampung nilai yang baru.
9. *Decide the Physical Design*, media penyimpanan yang ada dalam suatu susunan secara fisik yang menggunakan sejumlah tabel untuk menggambarkan data serta hubungan antara data-data tersebut

1.2.3 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naïve Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Ada beberapa tahapan yang dilakukan dalam perhitungan probabilitas menggunakan algoritma Naïve Bayes (Putri A.N., 2017) sebagai berikut :

1. Penentuan *sample* data yang akan digunakan untuk perhitungan.
2. Penentuan klasifikasi data (dimensi)
3. Menentukan variabel probabilitas yang akan dicari dari data *pengujian*.
 - $X = (X_1, X_2, X_3, \dots, X_n)$
 - X adalah variabel probabilitas yang dicari

- $X_1, X_2, X_3, \dots, X_n$ adalah rincian kriteria dari variable probabilitas yang dicari.
- 4. Menentukan dan menghitung probabilitas dari *criteria class*.
 - $P(C_i) = C_i / \text{Total } (C)$.
 - $P(C_i)$ adalah probabilitas dari masing-masing *criteria class*
 - C_i adalah nilai dari masing-masing *criteria class*
 - $\text{Total } (C)$ adalah nilai total dari seluruh *criteria class*
- 5. Menghitung probabilitas dari *variable class* terhadap *criteria class* yang sama.

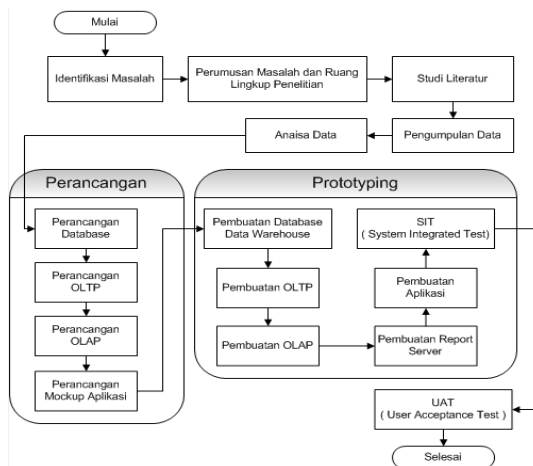
$$P(X|C_i) = X|C_i / \text{Total } (X|C_i).$$
 - $P(X|C_i)$ adalah probabilitas *variable class* terhadap *criteria class* yang sama
 - $X|C_i$ adalah nilai dari *variable class* terhadap *criteria class* yang sama
 - $\text{Total } (X|C_i)$ adalah nilai total *variable class* terhadap *criteria class* yang sama
- 6. Menghitung masing-masing probabilitas *criteria class* dengan probabilitas dari *criteria class* yang sama.

$$P(X_i|C_i) = P(X|C_i) * P(C_i).$$
- 7. Membandingkan perhitungan nilai probabilitas dari setiap *criteria class* dan *variable class* dan ambil nilai probabilitas terbesar terhadap kriteria yang dicari.

$$\text{Max} \rightarrow P(X|C_i) * P(C_i).$$

1.3 Diagram Alir Penelitian

Alir Penelitian yang dilakukan bisa dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

2. PEMBAHASAN

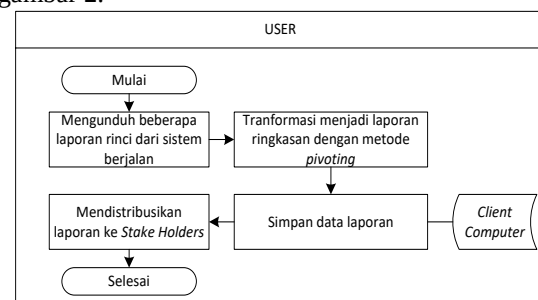
2.1 Sistem Berjalan

Dalam proses pengolahan laporan pada Divisi Marketing, saat ini dikerjakan oleh satu orang dengan cara mengunduh beberapa laporan rinci terkait penjualan dari sistem berjalan, kemudian laporan tersebut diolah menggunakan *tools office spread sheet* menjadi laporan ringkasan sesuai dengan kebutuhan rutin pihak eksekutif maupun

membuat laporan analisis untuk pembuatan paket produk pembiayaan yang akan di pasarkan ke konsumen. Proses pengolahan laporan ini membutuhkan waktu sekitar satu sampai dua hari kerja.

Laporan Final berbentuk *file office (spread sheet)* yang sudah dibuat untuk pihak eksekutif, disimpan di salah satu komputer supervisor divisi marketing dan sebagai *backup data* disimpan di *shared folder* pada *server file*.

Proses pengunduhan data detail yang dilakukan user langsung dari sistem berjalan membutuhkan *resource* memori yang cukup besar baik dari sisi komputer *client* maupun dari *server core system* sendiri jika periode data *sales* yang diunduh cukup banyak. Proses pada sistem berjalan bisa dilihat pada gambar 2.



Gambar 2. Sistem Berjalan

2.2 Kebutuhan data dan informasi

Setelah dilakukan analisis dari proses pembuatan laporan di Divisi Marketing mengenai informasi yang dibutuhkan untuk penyusunan laporan ke pihak eksekutif maupun laporan untuk kebutuhan analisis penjualan, maka didapatkan kebutuhan tersebut adalah :

- a) Laporan analisis *sales* berdasarkan periode penjualan
Laporan ringkasan yang bisa diturunkan untuk menampilkan data penjualan berdasarkan periode *sales* (tahun, kuartal, bulan, minggu dan hari)
- b) Laporan analisis *sales* berdasarkan kategori aset
Laporan ringkasan yang bisa diturunkan untuk menampilkan data penjualan berdasarkan kategori aset (kategori aset , merk aset dan tahun aset)
- c) Laporan analisis *sales* berdasarkan struktur kredit
Laporan ringkasan yang bisa diturunkan untuk menampilkan data penjualan berdasarkan struktur kredit (tenor kredit, harga OTR, *Down Payment*, pokok Hutang dan Bunga)
- d) Laporan analisis *sales* berdasarkan profil debitur
Laporan ringkasan yang bisa diturunkan untuk menampilkan data penjualan berdasarkan profil debitur (tipe debitur, jenis kelamin, usia, lama bekerja, lama tinggal dan status kepemilikan rumah)
- e) Laporan *forecasting* berbasis algoritma *Naive Bayes*

Laporan pembuatan prediksi kriteria dari satu Class dimensi terhadap beberapa variabel dimensi dari history data sales dalam periode tertentu. Laporan ini yang nantinya akan digunakan sebagai bahan analisis pembuatan segmen produk yang akan dijual.

2.3 Perancangan Forecasting menggunakan Algoritma Naïve Bayes

Untuk perancangan perhitungan forecasting menggunakan algoritma Naïve Bayes terhadap data training penjualan dari Divisi Marketing akan dilakukan beberapa tahap sebagai berikut :

1) Penentuan data training

Data Training yang digunakan sebagai contoh perancangan perhitungan Naïve Bayes adalah data sales selama dua tahun (tahun 2014 sampai dengan 2016) dan data yang digunakan sebanyak 7,556 rows. Pada Gambar 3 dijelaskan untuk data training diunduh dari sistem berjalan dengan beberapa informasi klasifikasi data seperti kategori aset (Asset Category), harga aset (On The Road Price), uang muka (Down Payment) dan jangka waktu kredit (Tenor).

	Asset_Category	OTR	Down_Payment	Manufacturing_Year	Tenor
1					
2	Passenger Car+Double Cabin	100.000.001 - 120.000.000	25.00% - 30.00%	2003 - 2006	3 Years
3	Passenger Car+Double Cabin	< 70.000.000	> 30.00%	< 1998	2 Years
4	Passenger Car+Double Cabin	70.000.001 - 100.000.000	> 30.00%	2001 - 2002	1 Years
5	Passenger Car+Double Cabin	100.000.001 - 120.000.000	> 30.00%	> 2006	2 Years
6	Passenger Car+Double Cabin	< 70.000.000	25.00% - 30.00%	< 1998	3 Years
7	Passenger Car+Double Cabin	> 140.000.001	> 30.00%	> 2006	3 Years
8	Passenger Car+Double Cabin	120.000.001 - 140.000.000	25.00% - 30.00%	> 2006	3 Years
9	PickUp	> 140.000.001	> 30.00%	> 2006	3 Years
10	Passenger Car+Double Cabin	70.000.001 - 100.000.000	25.00% - 30.00%	1998 - 2000	3 Years
11	Passenger Car+Double Cabin	< 70.000.000	25.00% - 30.00%	< 1998	3 Years
12	Passenger Car+Double Cabin	> 140.000.001	> 30.00%	> 2006	3 Years
13	Passenger Car+Double Cabin	> 140.000.001	25.00% - 30.00%	> 2006	2 Years
14	Passenger Car+Double Cabin	120.000.001 - 140.000.000	25.00% - 30.00%	> 2006	2 Years
15	Passenger Car+Double Cabin	70.000.001 - 100.000.000	25.00% - 30.00%	> 2006	2 Years
16	Passenger Car+Double Cabin	70.000.001 - 100.000.000	25.00% - 30.00%	2003 - 2006	3 Years
17	Passenger Car+Double Cabin	< 70.000.000	25.00% - 30.00%	1998 - 2000	3 Years
18	Passenger Car+Double Cabin	70.000.001 - 100.000.000	> 30.00%	1998 - 2000	1 Years
19	Passenger Car+Double Cabin	< 70.000.000	> 30.00%	< 1998	3 Years
7556	Passenger Car+Double Cabin	70.000.001 - 100.000.000	> 30.00%	> 2006	3 Years
7557	Passenger Car+Double Cabin	100.000.001 - 120.000.000	> 30.00%	> 2006	3 Years

Gambar 3. Data Training Sales

2) Penentuan klasifikasi data (dimensi)

Klasifikasi data diambil dari beberapa field pada laporan sistem berjalan. klasifikasi data (dimensi) yang akan digunakan dalam perhitungan ini adalah sebagai berikut :

a) Kategori aset (Asset Category)

- Passenger Car + Double Cabin
- Pickup
- Truck

b) Harga aset (On The Road price)

- < 70.000.000
- 70.000.001 - 100.000.000
- 100.000.001 - 120.000.000
- 120.000.001 - 140.000.000
- > 140.000.001

c) Uang muka (Down payment)

- <20.00%
- 20.00% - 24.99%
- 25.00% - 30.00%
- > 30.00%

d) Tahun pembuatan aset (Manufacturing Year)

- < 1998
- 1998 - 2000
- 2001 - 2002
- 2003 - 2006
- > 2006

e) Jangka waktu Kredit (Tenor)

- 1 Years
- 2 Years
- 3 Years
- 4 Years

3) Menentukan variable probabilitas yang akan dicari dari data training

Mencari prediksi (class) jangka waktu kredit (tenor) yang banyak digunakan oleh konsumen berdasarkan Variabel (criteria) sebagai berikut :

- Asset Category : Pessenger Car + Double Cabin (X_1)
- On The Road Price : < 70 Juta Rupiah (X_2)
- Down Payment : > 30% (X_3)
- Manufacturing year : > 2006 (X_4)

4) Menentukan dan menghitung probabilitas dari criteria class

Criteria class yang akan dicari probabilitasnya adalah jangka waktu kredit (Tenor) dan untuk perhitungannya bisa dilihat pada Gambar 4.

Probabilitas Class – Tenor								
Jumlah Kejadian "Tenor"					Probabilitas "Tenor"			
1 Years	2 Years	3 Years	4 Years	Total	1 Years	2 Years	3 Years	4 Years
187	1,054	4,309	2,006	7,556	0.02475	0.13949	0.57028	0.26548

Gambar 4. Probabilitas Class Tenor

Keterangan : dari jumlah data (Total (C_i)) sebanyak 7,556, didapat rincian untuk mencari probabilitas dari masing-masing criteria class ($P(C_i)$) sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$P(C_1) = \frac{187}{7,556} = 0.02475$$

- Probabilitas untuk Tenor 2 Tahun

$$P(C_2) = \frac{1,054}{7,556} = 0.13949$$

- Probabilitas untuk Tenor 3 Tahun

$$P(C_3) = \frac{4,309}{7,556} = 0.57028$$

- Probabilitas untuk Tenor 4 Tahun

$$P(C_4) = \frac{2,006}{7,556} = 0.26548$$

5) Menghitung probabilitas dari variable class terhadap criteria class yang sama.

Perhitungan masing-masing variabel dimensi adalah sebagai berikut :

a) Probabilitas dari Asset Category Passenger Car + Double Cabin.

Untuk mencari probabilitas Asset Category = Passenger Car + Double Cabin $P(X_i|C_i)$ dari setiap kriteria tenor adalah sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$P(X_1|C_1) : \frac{159}{187} = 0.85027$$

- Probabilitas untuk Tenor 2 Tahun

$$P(X_1|C_2) : \frac{921}{1,054} = 0.87381$$

- Probabilitas untuk Tenor 3 Tahun

$$P(X_1|C_3) : \frac{3,763}{4,309} = 0.87329$$

- Probabilitas untuk Tenor 4 Tahun

$$P(X_1|C_4) : \frac{1,920}{2,006} = 0.95713$$

Perhitungan tersebut bisa dilihat pada gambar 5.

Probabilitas Asset Category									
Asset Category	Jumlah Kejadian "Tenor"					Probabilitas "Tenor"			
	1 Years	2 Years	3 Years	4 Years	Total	1 Years	2 Years	3 Years	4 Years
Passenger Car + Double Cabin	159	921	3,763	1,920	6,763	0.85027	0.87381	0.87329	0.95713
PickUp	25	120	511	1	657	0.13369	0.11385	0.11859	0.00050
Truck	3	13	35	85	136	0.01604	0.01233	0.00812	0.04237
Total	187	1,054	4,309	2,006	7,556	0.02475	0.13949	0.57028	0.26548

Gambar 5. Probabilitas Asset Category
Passenger Car + Double Cabin

b) Probabilitas dari On The Road Price : < 70 Juta Rupiah

Untuk mencari probabilitas "On The Road Price : < 70 Juta Rupiah" $P(X|C_i)$ dari setiap kriteria tenor adalah sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$P(X_2|C_1) : \frac{27}{187} = 0.14439$$

- Probabilitas untuk Tenor 2 Tahun

$$P(X_2|C_2) : \frac{226}{1,054} = 0.21442$$

- Probabilitas untuk Tenor 3 Tahun

$$P(X_2|C_3) : \frac{802}{4,309} = 0.18612$$

- Probabilitas untuk Tenor 4 Tahun

$$P(X_2|C_4) : \frac{66}{2,006} = 0.03290$$

Perhitungan tersebut bisa dilihat pada gambar 6.

Probabilitas OTR (Harga on the Road)									
OTR	Jumlah Kejadian "Tenor"					Probabilitas "Tenor"			
	1 Years	2 Years	3 Years	4 Years	Total	1 Years	2 Years	3 Years	4 Years
< 70.000.000	27	226	802	66	1,121	0.14439	0.21442	0.18612	0.03290
70.000.001 - 100.000.000	47	338	1,590	479	2,454	0.25134	0.32068	0.36900	0.23878
100.000.001 - 120.000.000	31	192	910	598	1,731	0.16578	0.18216	0.21119	0.29811
120.000.001 - 140.000.000	31	118	533	404	1,086	0.16578	0.11195	0.12369	0.20140
> 140.000.001	51	180	474	459	1,164	0.27273	0.17078	0.11000	0.22881
Grand Total	187	1,054	4,309	2,006	7,556	0.02475	0.13949	0.57028	0.26548

Gambar 6. Probabilitas On The Road Price: <70
Juta Rupiah

c) Probabilitas dari Down Payment : > 30%

Untuk mencari probabilitas "Down Payment : >30%" $P(X|C_i)$ dari setiap kriteria tenor adalah sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$P(X_3|C_1) : \frac{122}{187} = 0.65241$$

- Probabilitas untuk Tenor 2 Tahun

$$P(X_3|C_2) : \frac{661}{1,054} = 0.62713$$

- Probabilitas untuk Tenor 3 Tahun

$$P(X_3|C_3) : \frac{1,488}{4,309} = 0.34532$$

- Probabilitas untuk Tenor 4 Tahun

$$P(X_3|C_4) : \frac{389}{2,006} = 0.19392$$

Perhitungan tersebut bisa dilihat pada gambar 7.

Probabilitas Down Payment									
Down Payment	Jumlah Kejadian "Tenor"					Probabilitas "Tenor"			
	1 Years	2 Years	3 Years	4 Years	Total	1 Years	2 Years	3 Years	4 Years
< 20%	-	-	-	-	-	-	-	-	-
20.00% - 24.99%	2	-	1	7	10	0.01070	-	0.00023	0.00349
25.00% - 30.00%	63	393	2,820	1,610	4,886	0.33690	0.37287	0.65444	0.80259
> 30.00%	122	661	1,488	389	2,660	0.65241	0.62713	0.34532	0.19392
Grand Total	187	1,054	4,309	2,006	7,556	0.02475	0.13949	0.57028	0.26548

Gambar 7. Probabilitas Down Payment: >30%

d) Probabilitas dari Manufacturing Year >2006

Untuk mencari probabilitas "Manufacturing Year >2006" $P(X|C_i)$ dari setiap kriteria tenor adalah sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$P(X_4|C_1) : \frac{144}{187} = 0.60963$$

- Probabilitas untuk Tenor 2 Tahun

$$P(X_4|C_2) : \frac{532}{1,054} = 0.50474$$

- Probabilitas untuk Tenor 3 Tahun

$$P(X_4|C_3) : \frac{1,932}{4,309} = 0.44836$$

- Probabilitas untuk Tenor 4 Tahun

$$P(X_4|C_4) : \frac{1,142}{2,006} = 0.56929$$

Perhitungan tersebut bisa dilihat pada gambar 8.

Probabilitas Manufacturing Year									
Manufacturing Year	Jumlah Kejadian "Tenor"					Probabilitas "Tenor"			
	1 Years	2 Years	3 Years	4 Years	Total	1 Years	2 Years	3 Years	4 Years
< 1998	4	53	130	-	187	0.02139	0.05028	0.03017	-
1998 - 2000	13	112	511	15	651	0.06952	0.10626	0.11859	0.00748
2001 - 2002	11	102	515	165	793	0.05882	0.09677	0.11952	0.08225
2003 - 2006	45	255	1,221	684	2,205	0.24064	0.24194	0.28336	0.34098
> 2006	114	532	1,932	1,142	3,720	0.60963	0.50474	0.44836	0.56929
Grand Total	187	1,054	4,309	2,006	7,556	0.02475	0.13949	0.57028	0.26548

Gambar 8. Probabilitas Manufacturing Year
>2006

6) Menghitung masing-masing probabilitas criteria class dengan variable class yang sama

Untuk mencari probabilitas $P(X|C_i)$ dari setiap kriteria tenor adalah sebagai berikut :

- Probabilitas untuk Tenor 1 Tahun

$$= P(X_1|C_1) * P(X_2|C_1) * P(X_3|C_1) * P(X_4|C_1) * P(C_1) \\ = 0.85027 * 0.14439 * 0.65241 * 0.60963 * 0.02475 \\ = 0.00121$$

- Probabilitas untuk Tenor 2 Tahun

$$= P(X_1|C_2) * P(X_2|C_2) * P(X_3|C_2) * P(X_4|C_2) * P(C_2) \\ = 0.87381 * 0.21442 * 0.62713 * 0.50474 * 0.13949 \\ = 0.00827$$

- Probabilitas untuk Tenor 3 Tahun

$$= P(X_1|C_3) * P(X_2|C_3) * P(X_3|C_3) * P(X_4|C_3) * P(C_3) \\ = 0.87329 * 0.18612 * 0.34532 * 0.44836 * 0.57028 \\ = 0.01435$$

- Probabilitas untuk Tenor 4 Tahun

$$\begin{aligned}
 &= P(X_1|C_4) * P(X_2|C_4) * P(X_3|C_4) * P(X_4|C_4) * \\
 &\quad P(C_4) \\
 &= 0.95713 * 0.03290 * 0.19392 * 0.56929 * \\
 &\quad 0.26548 \\
 &= 0.00092
 \end{aligned}$$

Perhitungan tersebut bisa dilihat pada gambar 9.

variable				Tenor			
Asset Category	OTR	Down Payment	Manufacturing Year	1 Years	2 Years	3 Years	4 Years
Passenger Car + Double Cabin	< 70.000.000	> 30.00%	> 2006	0.00121	0.00827	0.01435	0.00092
persentase				4.88%	33.42%	57.97%	3.73%

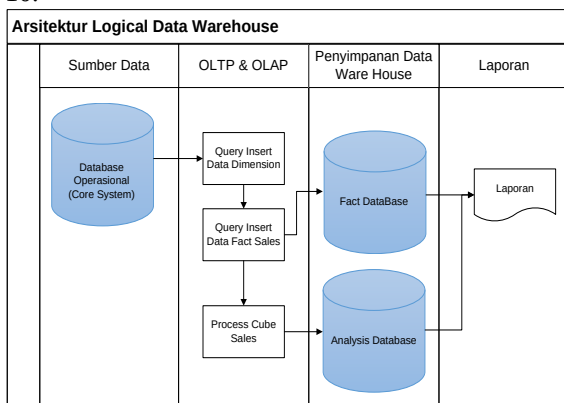
Gambar 9. Probabilitas $P(X_i|C_i)$

7) Membandingkan perhitungan nilai probabilitas dari setiap criteria class dan variable class dan ambil nilai probabilitas terbesar terhadap kriteria yang dicari.

Probabilitas tenor (*criteria class*) terbesar terhadap segmen (*criteria variable*) : *Asset Category* : Passenger Car + Double Cabin , *On The Road Price* : < 70 Juta Rupiah, *Down Payment* : > 30% dan *Manufacturing year* : > 2006 adalah di “Tenor 3 Tahun” dengan nilai probabilitas = 0.01435 atau dipilih sebanyak 57.97% dari populasi data *training*. Sehingga divisi marketing bisa membuat promosi menarik untuk paket produk tenor 3 tahun terhadap segmen Kendaraan Passenger Car + Double Cabin, Harga OTR < 70 Juta dengan uang muka > 30% dan tahun kendaran > 2006.

2.4 Arsitektur Data Warehouse

Sumber data dari sistem berjalan dilakukan transformasi menggunakan *query* dan menghasilkan data *Fact Sales* yang disimpan di *Fact Database*. Kemudian data-data dari beberapa tabel dimensi dan tabel fakta di proses untuk menghasilkan *Cube Sales* dan disimpan *Analysis Database*. Data dari *Fact Database* dan *Analysis Database* selanjutnya akan digunakan untuk pembuatan laporan analisis. Arsitektur *Data Warehouse* bias dilihat pada gambar 10.



Gambar 10. Arsitektur Data Warehouse

2.5 Perancangan Skema Data Warehouse

Perancangan skema *Data Warehouse* ini terdiri dari 1 tabel fakta dan 17 tabel dimensi yang

mengelilinginya. Skema yang digunakan adalah skema bintang (*star scheme*). Model ini memiliki satu table sebagai tabel fakta (*fact table*) dengan beberapa tabel dimensi yang mengelilinginya dan berhubungan langsung ke table fakta. kalkulasi awal yang disimpan pada table fakta penjualan (*FactSales*) adalah:

- Unit, menunjukan jumlah unit (aset) yang dibiayai.
- OTR, menunjukan jumlah harga unit (aset) / OTR (*on the road*) yang dibiayai.
- NTF, menunjukan jumlah nilai pokok pembiayaan.

Tabel dimensi yang dibuat ada 17 tabel yang digunakan untuk pengklasifikasian.

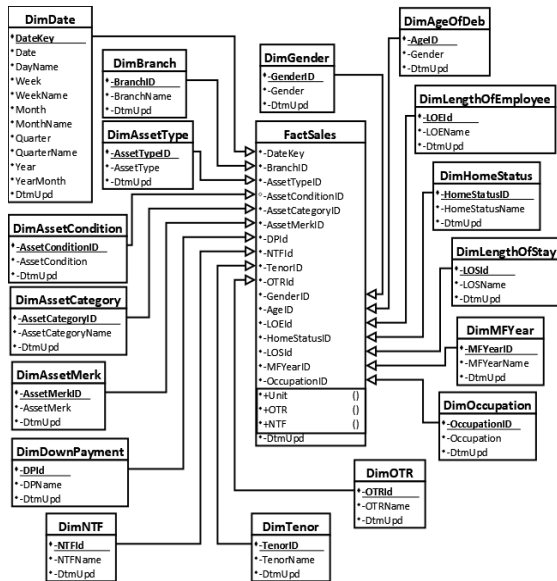
Terdapat 5 tabel dimensi klasifikasi yang bersumber dari tabel master pada sistem berjalan adalah sebagai berikut:

- DimBranch, berisi data semua kantor cabang yang tersebar di berbagai lokasi.
- DimAssetType, berisi data jenis aset yang dibiayai.
- DimAssetCategory, berisi data kategori aset yang dibiayai.
- DimAssetMerk, berisi data merk aset yang dibiayai.
- DimHomeStatus, berisi data status kepemilikan rumah nasabah.

Terdapat 12 tabel dimensi klasifikasi yang tidak bersumber dari tabel master pada sistem berjalan dan didapatkan dari klasifikasi data dilakukan berdasarkan kebutuhan analisis. Tabel-tabel dimensi tersebut adalah:

- DimDate, berisi klasifikasi data dari periode waktu penjualan dan memiliki hirarki waktu (tahun, kuartal, bulan dan minggu).
- DimGender, berisi data klasifikasi jenis kelamin nasabah.
- DimAssetCondition, berisi data klasifikasi kondisi aset yang dibiayai.
- DimDownPayment, berisi data klasifikasi uang muka pembiayaan.
- DimNTF, berisi data klasifikasi nilai pokok hutang pembiayaan.
- DimTenor, berisi data klasifikasi jangka waktu pembiayaan.
- DimOTR, berisi data klasifikasi harga aset yang dibiayai.
- DimAgeOfDeb, berisi data klasifikasi usia nasabah.
- DimLengthOfEmployee, berisi data klasifikasi lama kerja nasabah.
- DimLengthOfStay, berisi klasifikasi lama tinggal nasabah.
- DimMFYear, berisi klasifikasi tahun aset yang dibiayai.
- DimOccupation, berisi klasifikasi tipe nasabah.

Skema *Data Warehouse* bisa dilihat pada gambar 11.



Gambar 11. Skema Data Warehouse

2.6 ETL (Extract, Tranform and Load)

Proses transformasi tabel dimensi yang bersumber dari tabel master sistem berjalan bisa dilihat pada gambar 12.



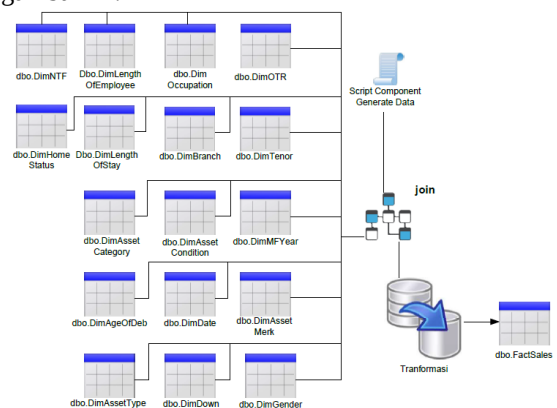
Gambar 12. ETL Dimensi dari tabel master

Proses transformasi tabel dimensi yang bukan bersumber dari tabel master sistem berjalan bisa dilihat pada gambar 13.



Gambar 13. ETL Dimensi manual

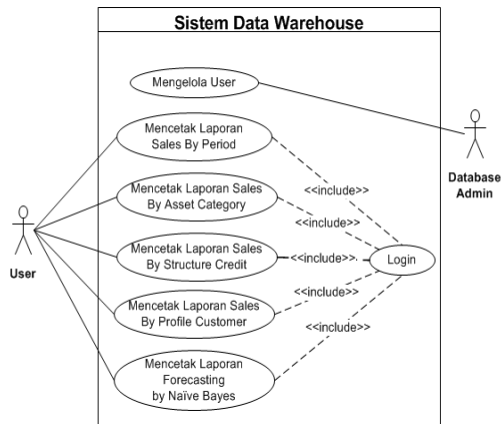
Proses transformasi tabel fakta bisa dilihat pada gambar 14.



Gambar 14. ETL Tabel Fakta

2.7 Use Case Diagram

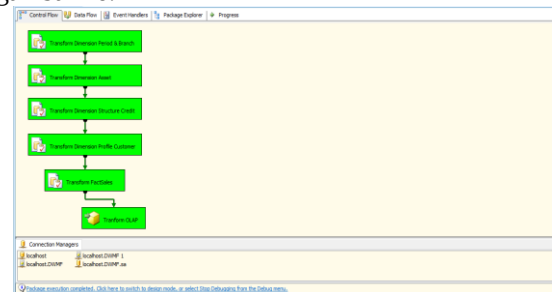
Use case diagram pada gambar 15 menjelaskan terdapat dua aktor yaitu Database Admin yang mengelola akses user yang digunakan untuk aplikasi data warehouse dan user dari divisi marketing yang akan mengakses beberapa laporan analisis sales dari aplikasi data warehouse.



Gambar 15. Use Case Diagram sistem Data Warehouse

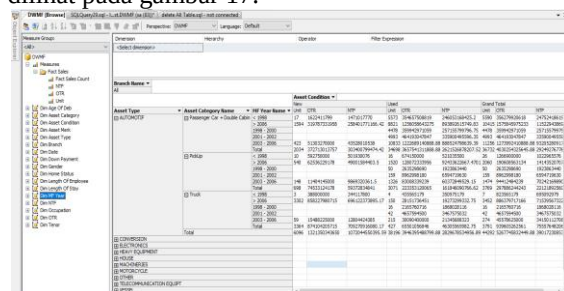
2.8 Implementasi OLTP dan OLAP

Proses OLTP (*OnLine Transaction Processing*) berisikan proses tranformasi beberapa data dimensi dan data fakta sales. Proses tersebut bisa dilihat pada gambar 16.



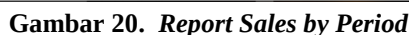
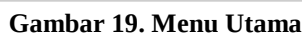
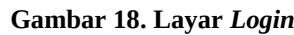
Gambar 16. Rangkaian OLTP

Proses OLAP (*OnLine Analytical Processing*) menghasilkan Cubes Sales dan beberapa dimensi sesuai dengan rancangan. Analysis Database yang berupa *Measures* dari data fakta sales (*factSales*) dan dikombinasikan dengan beberapa data dimensi untuk menghasilkan analisis data. Hasil OLAP bisa dilihat pada gambar 17.



Gambar 17. Tampilan OLAP dengan SQL Server Management Studio

Implementasi antar muka menghasilkan beberapa *screen* terutama untuk mengundun beberapa laporan analisis yang salah satunya sudah menggunakan metode forecasting dengan algoritma naïve bayes. implementasi tersebut bisa dilihat pada gambar 18 sampai dengan gambar 24.



Pengujian sistem terhadap hasil yang ditampilkan dengan kesesuaian fungsi yang diharapkan oleh *user* dengan metode *black box testing*.

Berdasarkan hasil penelitian dan analisis yang telah dilakukan pada Divisi Marketing di PT. ITC Auto Multifinance, maka bisa disimpulkan sebagai berikut :

1. Algoritma Naïve Bayes dapat digunakan untuk membuat analisis *forecasting* terhadap beberapa dimensi data untuk mengetahui segmen produk dari sesuai dan banyak diminati oleh konsumen.
2. Penyusunan laporan untuk menunjang analisis divisi marketing bisa dilakukan lebih cepat dan terstruktur dengan pemanfaatan OLTP (*OnLine Transaction Processing*) dan OLAP (*OnLine Analytical Processing*).

3. Pengelolaan *backup* data yang digunakan untuk menunjang laporan analisis divisi marketing bisa lebih efisien dan terorganisir di dalam *database*.

PUSTAKA

- Ardista, N., Taufik & Purbandini. 2017. *Rancang Bangun Data Warehouse Untuk Pembuatan Laporan dan Analisis pada Data Kunjungan Pasien Rawat Jalan Rumah Sakit Universitas Airlangga Berbasis Online Analytical Processing (OLAP)*. Jurnal of Information Systems Engineering and Business Intelligence. Vol. 3, No. 1, April 2017, e-ISSN: 2443-2555.
- Aristo Hintarsyah, P., Christy, J. & Warnars, H.L.H.S. 2018. *Forecasting Sebagai Decision Support Systems Aplikasi Dan Penerapannya Untuk Mendukung Proses Pengambilan Keputusan*. Jurnal Sistem Komputer – Vol.8 No.1 Tahun 2018, ISSN: 2087-4685, EISSN: 2252-3456.
- Ayuliana & Yosieto, F. 2015. *Pemanfaatan Data Warehouse Proses Penjualan Dan Pembelian Untuk Dukungan Pengambilan Keputusan*. Jurnal Ilmiah FIFO Volume VII/No. 2/November/2015, P-ISSN: 2085-4315 / E-ISSN: 2502-8332.
- Dores, A., Masya, F. & Prastiawan, H. 2018. *Indonesian Text News Classification Using the Naïve Bayes Algorithm*. International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 7, Issue. 8, August 2018, pg.159 – 169.
- Fithri, D.L. 2016. *Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naïve Bayes*. Jurnal SIMETRIS. Vol 7 No 2 November 2016, ISSN: 2252-4983.
- Henderi, Langgeng & Karwandi. 2016. *Rancangan Data Warehouse Sistem Evaluasi Pemasaran*. Seminar Nasional Teknologi Informasi dan Multimedia 2016, STMIK AMIKOM Yogyakarta. 6-7 Februari 2016, ISSN: 2302-3805.
- Mujiono & Musdholifah, A. 2016. *Pengembangan Data Warehouse Menggunakan Pendekatan Data-Driven untuk Membantu Pengelolaan SDM*. IJCCS, Vol.10, No.1. January 2016, pp. 1~10 ISSN: 1978-1520.
- Pramudia, H. & Nugroho, A. 2017. *Sistem Informasi Kerusakan Laptop Menggunakan Metode Naïve Bayes*. Jurnal Teknologi Elektro, Universitas Mercu Buana ISSN: 2086-9479, Vol. 8 No. 3 September 2017.
- Putri, A.N. 2017. *Penerapan Naive Bayesian Untuk Perankingan Kegiatan Di Fakultas TIK Universitas Semarang*. Jurnal SIMETRIS. Vol 8 No 2 November 2017, ISSN: 2252-4983.
- Saleh, A. 2015. *Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. Citec Jurnal. Vol. 2, No. 3, Mei 2015 – Juli 2015, ISSN: 2354-5771.
- Triana, Y.S. & Susilo, A. 2017. *Aplikasi Data Warehouse Untuk Menunjang Standar 3 Borang Akreditasi Prodi Informatika Universitas Mercu Buana*. JURNAL ILMIAH FIFO P-ISSN 2085-4315 / E-ISSN 2502-8332, Volume IX/No.2/November/2017.