

IMPLEMENTASI *DATA WAREHOUSE* UNTUK DATA JUMLAH KENDARAAN MENGGUNAKAN SOFTWARE BUSSINESS INTELEGENCE SQL SERVER 2008 PADA DISPENDA CILEGON

asyaefudin1213@gmail.com)

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 dan microsoft SQL business intelligence development. Dengan menggunakan tools microsoft SQL server 2008 dan microsoft SQL business intelligence development dapat dibangun sebuah Data Warehouse jumlah kendaraan dengan mengumpulkan data-data kendaraan meliputi jumlah kendaraan, merk Kendaraan, wilayah, jenis kendaraan, waktu, dan pajak yang dimasukan kedalam sebuah Ms. Excel kemudian hasil dari Ms. Excel dimasukan kedalam database microsoft SQL server 2008, setelah data masuk kedalam database kemudian data tersebut di load kedalam microsoft SQL business intelligence development untuk dianalisis. Hasil analisis Data Warehouse yang dibuat dapat memberikan informasi yang lebih jelas untuk pihak Dispenda dan mempermudah dalam menganalisis peningkatan jumlah data kendaraan seperti dapat melihat peningkatan jumlah kendaraan dari jumlah kendaraan, merk Kendaraan, wilayah, jenis kendaraan, waktu, dan pajak. Dengan Data mining dispenda bisa melihat pengelompokan jumlah kendaraan di suatu wilayah Dengan arti lain Data mining adalah proses untuk penggalan pola-pola dari data. Data mining menjadi alat yang semakin penting untuk mengubah data tersebut menjadi informasi

Kata kunci :

Data Warehouse, Data Mining, Expectation Maximization (EM) Clustering, microsoft SQL business intelligence development, microsoft SQL server 2008, database, pengolahan data.

Abstract : Data Warehouse is a method that is used to accommodate large amounts of data, to build a Data Warehouse required a tool microsoft microsoft SQL server 2008 and SQL business intelligence development. By using the tools microsoft microsoft SQL server 2008 and SQL development business intelligence data warehouse can be built a number of vehicles to collect vehicle data including number of vehicles, vehicle brand, region, type of vehicle, time, and taxes are incorporated into a Ms. Then the result of Ms Excel. Microsoft Excel inserted into a SQL server 2008 database, after entering the data into the database then the data is loaded into a Microsoft SQL business intelligence development for analysis. The results of the analysis of the data warehouse is created to provide clearer information to the Revenue and make it easier to analyze increasing amounts of data such vehicles may see an increase in the number of vehicles of the number of vehicles, vehicle brand, region, type of vehicle, time, and taxes. With Dispenda Data mining can see a breakdown of the number of vehicles in an area with another sense Data mining is the process of extracting patterns from data. Data mining is becoming an increasingly important tool to transform this data into information

Keywords: Data Warehouse, Data Mining, Expectation Maximization (EM) Clustering, Microsoft SQL business intelligence development, microsoft SQL server 2008, database, data processing.

1 PENDAHULUAN

Semakin berkembangnya ilmu pengetahuan dan teknologi informasi di era globalisasi ini, sehingga memberikan banyak kemudahan dalam menjalankan aktivitas baik pekerjaan yang mudah sampai pekerjaan yang sulit, sehingga seolah-olah pekerjaan pun dipermudah oleh perkembangan teknologi baik dalam lembaga pemerintahan maupun lembaga swasta. Dengan semakin banyaknya informasi data yang diterima didalam lembaga pemerintahan Dispenda setiap tahunnya maka dibutuhkan sebuah tempat untuk menampung sebuah basis data yang kemudian dapat menciptakan sebuah laporan untuk dianalisis, sehingga dibutuhkan sebuah sistem *data mining* yang dapat menjadi sarana utama dalam mendukung pengambilan keputusan manajemen dengan mengumpulkan dan mengorganisasikan data dan informasi dapat lebih maksimal. Keberhasilan suatu data yang dianalisis sangat bergantung pada informasi data yang di dapat.

2 LANDASAN TEORI

2.1 Teori Umum

2.1.1 Database (Basis Data)

Basis data (*database*) adalah sekumpulan data organisasi untuk melayani banyak aplikasi secara efisien dengan memusatkan data dan mengendalikan redundansi data. Data yang disimpan dalam *file* terpisah untuk setiap aplikasi, data disimpan supaya pada pengguna data tersebut tampak seperti disimpan hanya dalam satu lokasi. Sebuah basis data melayani banyak aplikasi sekaligus. Contohnya, alih-alih menyimpan data karyawan di dalam sistem informasi yang terpisah dan memisahkan *file* untuk personalia, pengajian, dan keuntungan, perusahaan dapat membuat sebuah basis data sumber daya manusia dan dapat digunakan bersama (Laudon 2008, h.265).

2.2 Teori Khusus

2.2.1 Data Warehouse

Data warehouse adalah Kumpulan data dari berbagai sumber yang ditempatkan menjadi b satu dalam tempat penyimpanan dalam ukuran besar lalu diproses menjadi bentuk penyimpanan *multi – dimensional* dan didesain untuk *querying* dan *reporting* (Sulianta F, Juju D, 2010).

Karakteristik *Data Warehouse* menurut Inmon:

a. *Subject-oriented* (Berorientasi Subjek)

Data Warehouse didesain untuk menganalisa data berdasarkan *subject-subject* tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu.

b. *Integrated* (Terintegrasi)

Data Warehouse dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah ke dalam suatu format yang konsisten dan saling terintegrasi satu dengan lainnya. Syarat integrasi sumber data dapat dipenuhi dengan berbagai cara seperti konsisten dalam penamaan variabel. Konsisten dalam ukuran variabel, konsisten dalam struktur pengkodean dan konsisten dalam atribut fisik dari data.

c. *Time Variant* (Rentang Waktu)

Seluruh data pada *Data Warehouse* dapat dikatakan akurat atau *valid* pada rentang waktu tertentu. Cara-cara yang digunakan untuk melihat *interval* waktu yang digunakan dalam mengukur keakuratan suatu *Data Warehouse* :

1. Cara yang paling sederhana adalah menyajikan *Data Warehouse* pada rentang waktu tertentu.
2. Menggunakan variasi/perbedaan waktu yang disajikan dalam *Data Warehouse* baik *implicit* maupun *explicit*, secara *explicit* dengan unsur waktu dalam hari, minggu, bulan dan waktu tertentu.
3. Variasi waktu yang disajikan *Data Warehouse* melalui serangkaian *snapshot* yang panjang.

d. *Nonvolatile*

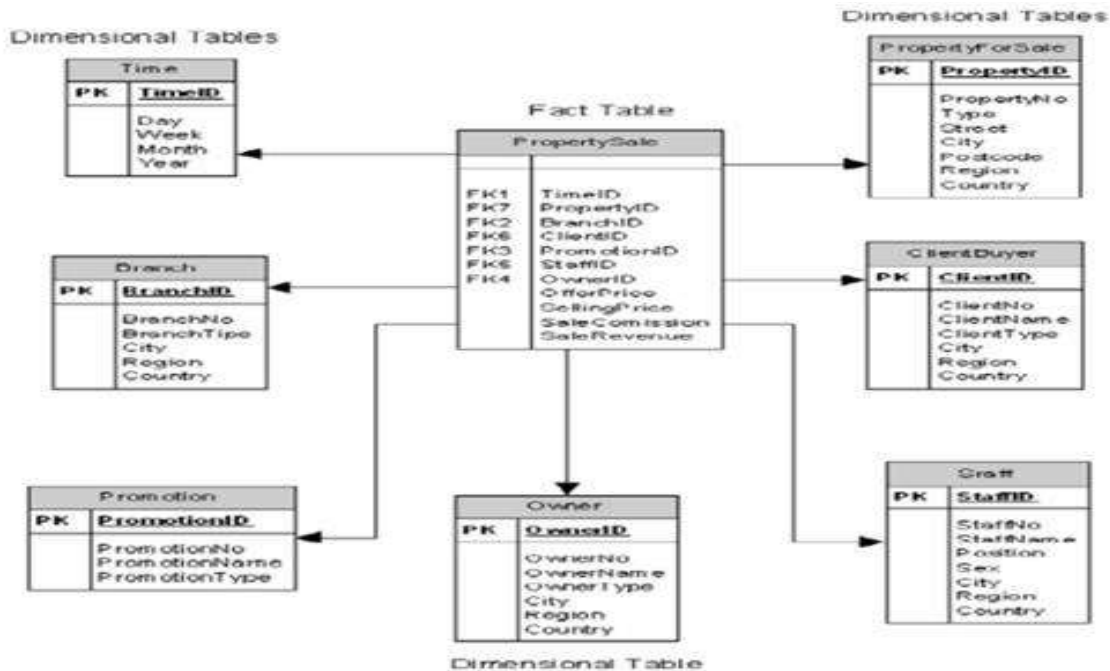
Nonvolatile maksudnya data pada *data warehouse* tidak di-*update* secara *real time* tetapi di *refresh* dari sistem operasional secara reguler. Data yang baru selalu ditambahkan bagi basis data itu sebagai sebuah perubahan. Basis data tersebut secara kontinu menyerap data baru ini, kemudian disatukan dengan data sebelumnya. Hal ini berbeda dengan basis data operasional yang dapat melakukan *update*, *insert* dan *delete* terhadap data, sedangkan pada *data warehouse* hanya ada dua kegiatan manipulasi data yaitu *loading* data (mengambil data) dan akses data

2.2.2 *Dimensional Modelling*

Dimensionality modeling adalah sebuah teknik desain logis yang bertujuan untuk menghadirkan data dalam sebuah bentuk yang standar dan intuitif yang memungkinkan pengaksesan basis data dengan performa yang tinggi (Connolly dan Begg 2011, h.1182). Ada beberapa konsep pemodelan *Data Warehouse* pada *dimensionality modeling* yang dikenal umum pada saat ini, konsep-konsep tersebut antara lain adalah *star schema*, *snowflake* dan *fact constellation schema*.

2.2.2.1 *Star Schema* (Skema Bintang)

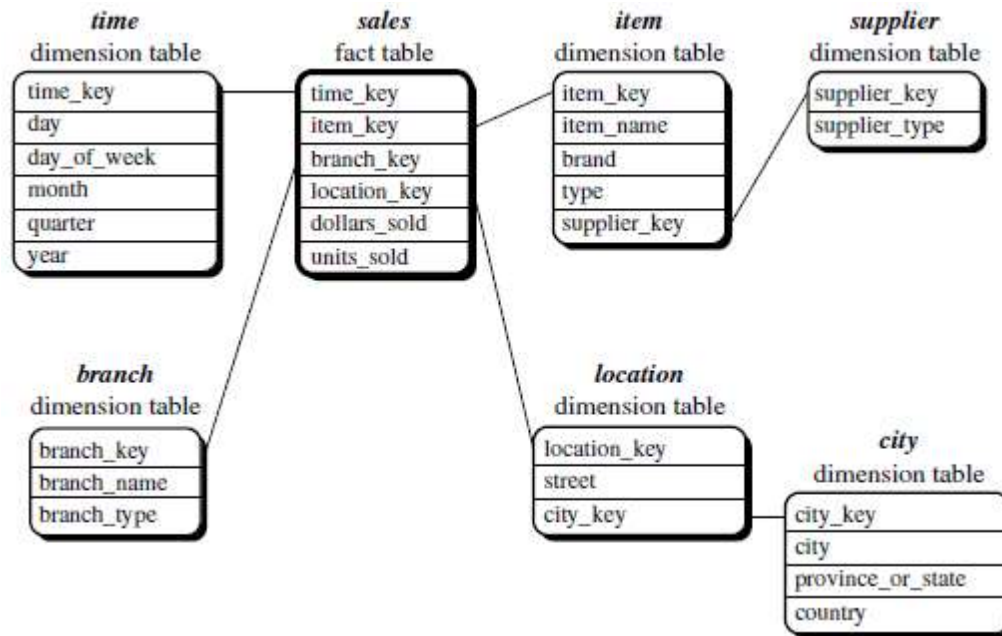
Menurut Connolly dan Begg (2010, h.1227), *star schema* adalah struktur logis yang memiliki tabel fakta yang memuat data faktual di pusat dan dikelilingi oleh tabel dimensi yang memuat data referensi (yang dapat dinormalisasi).



Gambar 1 Star Schema

2.2.2.2 Snowflake Schema

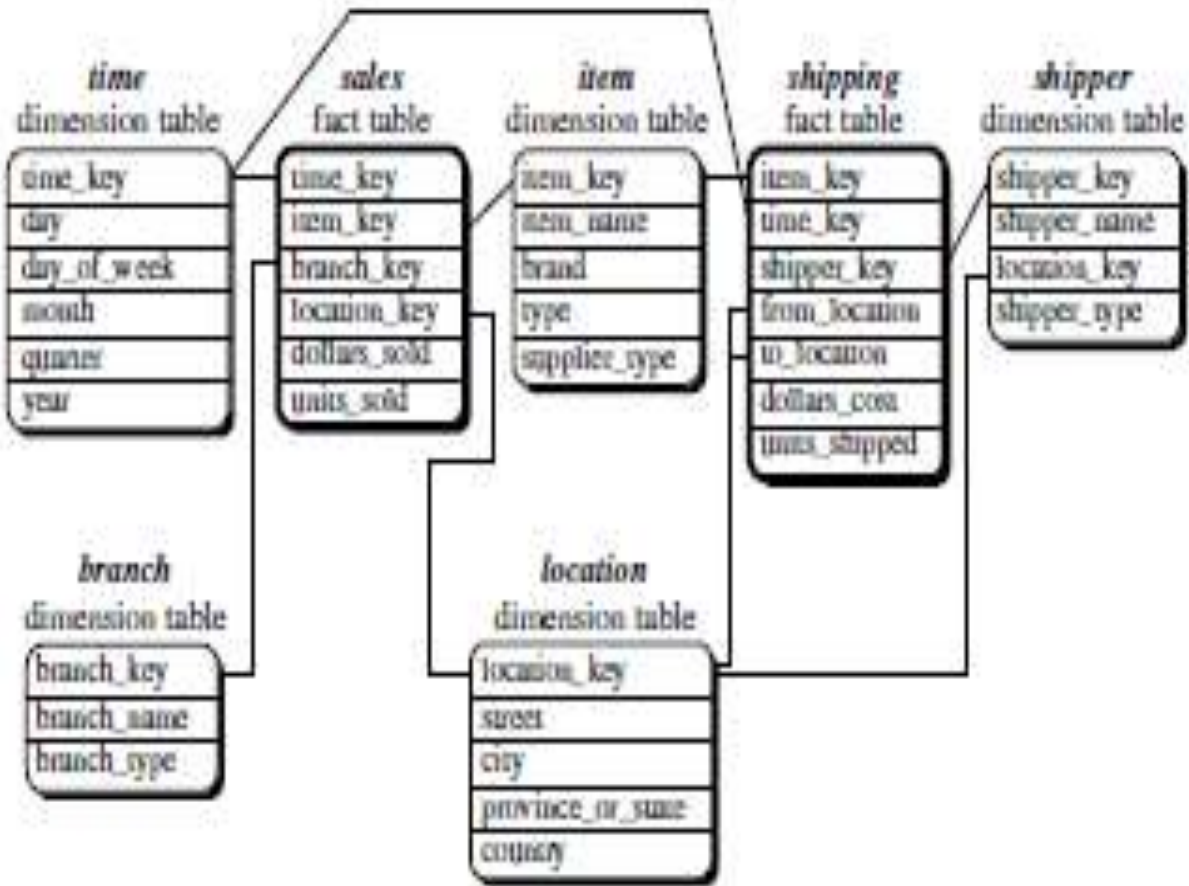
Menurut Connolly dan Begg (2010, h.1229), snowflake schema adalah sebuah variasi dari star schema dimana tabel dimensi tidak memuat data yang didenormalisasi.



Gambar 2 Snowflake Schema

2.2.2.3 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 Kusrini (2009, h.7), *Data Mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Menurut Kusrini (2009, h.10)

2.3.1 Clustering

Clustering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidak miripan dengan *record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data

menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

2.3.1 Algoritma *Expectation Maximization* (EM)

Menurut (Kusrini & Luthfi, 2009) Algoritma *Expectation Maximization* sering digunakan untuk menemukan nilai estimasi Maximum Likelihood (ML) dari parameter dalam sebuah model probabilistic, dimana model juga tergantung pada latent variabel yang belum diketahui. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu E step yang menghitung nilai ekspektasi dari likelihood termasuk latent variabel seolah-olah seperti mereka ada, dan M step menghitung nilai estimasi ML dari parameter dengan memaksimalkan nilai ekspektasi dari likelihood yang ditemukan pada E step.

2.3.2 *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*

2.3.3 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 Profil Singkat Dispenda Cilegon

Dinas Pendapatan Daerah (Dispenda) Kota Cilegon merupakan instansi yang bergerak di bidang perpajakan, seperti pajak reklame, pajak restoran, pajak hotel, pajak parkir, pajak penerangan jalan dan pajak hiburan di Kota Cilegon setiap tahunnya.

Adapun visi dan misi Dispenda Kota Cilegon :

a. Visi

“Pendapatan Daerah yang Maksimal untuk Pembangunan Kota Cilegon”.

b. Misi

- 1) Meningkatkan peran serta masyarakat dalam membayar pajak.
- 2) Modernisasi pajak.

3.2 Perancangan Data Warehouse

Pada *Proses* pembuatan data *warehouse* pada Dispenda Cilegon, Metode yang digunakan menggunakan metodologi sembilan tahapan (*nine-step methodology*) atau biasa disebut dengan metode Ralph Kimbal, metode ini meliputi :

3.2.1 Pemilihan Proses

Pada Dinas Pendapatan Daerah (Dispenda) Kota Cilegon terdapat beberapa proses pendataan pendapatan daerah , antara lain pendapatan pajak kendaraan bermotor, pendataan pajak bumi dan bangunan, retribusi parkir dan banyak lagi proses yang lainnya. Mengingat waktu pengerjaan dan keterbatasan

data, adapun proses yang dipilih dalam perancangan *DataWarehouse* pada Dispenda Cilegon ialah proses pendataan pendapatan pajak kendaraan bermotor di kota Cilegon.

3.2.2 Pemilihan Grain

Grain merupakan proses untuk menentukan apa yang digambarkan oleh *record* di dalam tabel fakta. Berikut adalah *grain* yang ada dalam perancangan *Data Warehouse* pada Dispenda Cilegon.

Untuk pembayaran pajak ini, analisis yang dapat dilakukan pada yaitu jumlah kendaraan dan jumlah pembayaran pajak kendaraan bermotor (PKB) berdasarkan jenis kelamin, umur dan status pengendara, jumlah kendaraan dan pembayaran PKB berdasarkan kecamatan dan kelurahan pengendara, jumlah kendaraan dan jumlah pembayaran PKB berdasarkan jenis dan merk kendaraan, jumlah kendaraan dan pembayaran PKB berdasarkan type dan tahun keluaran kendaraan yang dapat dilihat pertahun, bulan ataupun hari.

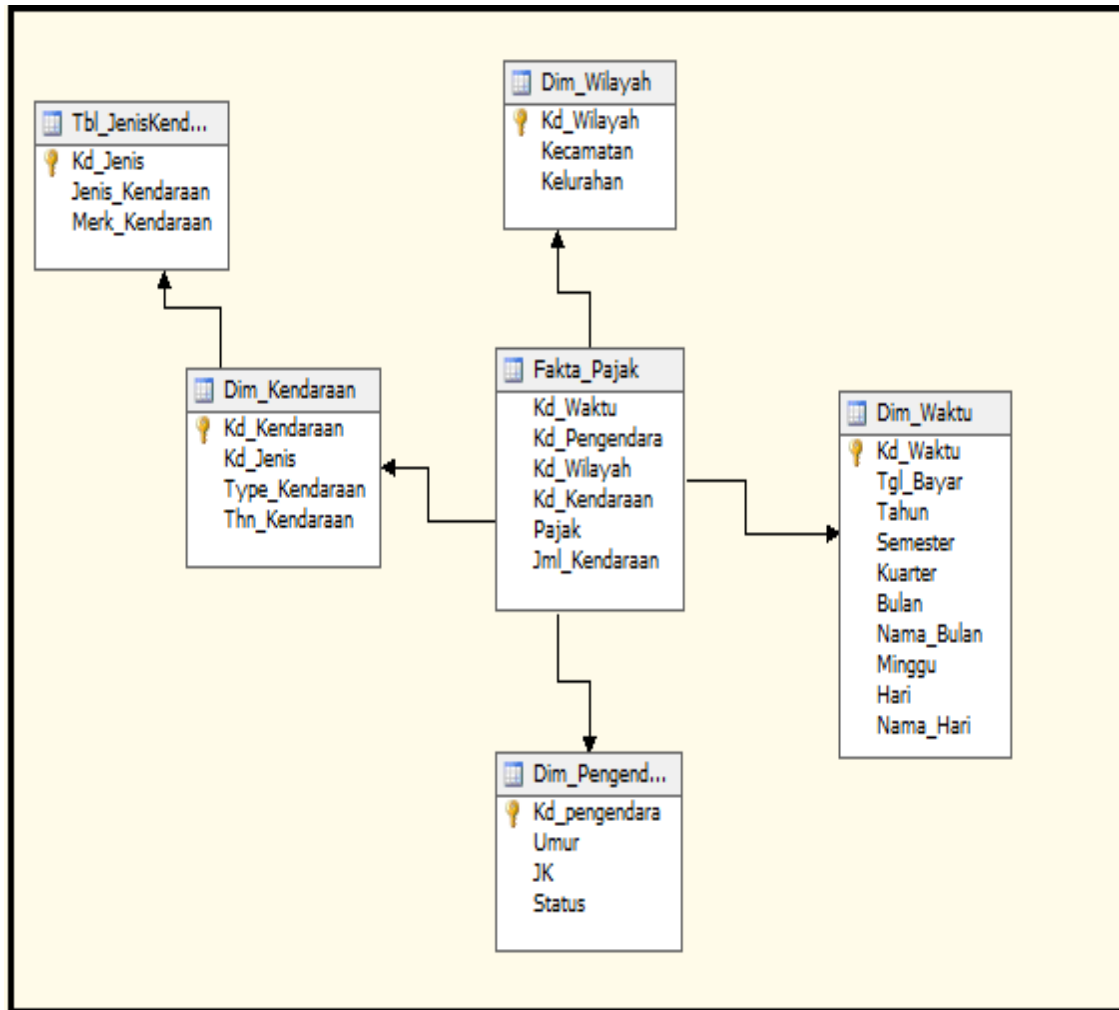
3.2.3 Identifikasi dan Penyesuaian Dimensi

Pada tahap ini dilakukan identifikasi dan penyesuaian dimensi yang terkait dengan fakta yang ditampilkan dalam bentuk matriks.

Dimensi yang digunakan dalam fakta pajak pada Dispenda Cilegon adalah dimensi waktu, dimensi pengendara, dimensi wilayah, dimensi kendaraan dan dimensi jenis kendaraan.

3.2.4 Pemilihan Fakta

Pada tahap ini dilakukan pemilihan fakta yang akan digunakan pada tabel fakta pajak. Tabel fakta pajak adalah tabel yang berisikan data-data pembayaran pajak kendaraan beromotor Kota Cilegon yang berisikan beberapa kode tabel dimensi dan perhitungan berupa pajak dan jumlah kendaraan. Adapun kolom – kolom pada tabel fakta pajak meliputi : Kd_waktu, Kd_pengendara, Kd_wilayah, Kd_kendaraan, Pajak, dan Jml_kendaraan.



Gambar 4 Fakta Skema *Snowflake*

4 Penyimpanan *Pre Calculation* Ditabel Fakta

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

a. Pajak :

Merupakan jumlah pajak kendaraan bermotor yang dibayar oleh pengendara.

b. *Jml_Kendaraan* :

Merupakan jumlah kendaraan dari setiap pembayaran pajak setiap pengendara yang bernilai 1 untuk setiap record pada tabel fakta pajak.

3.2.5 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.

3.2.6 Pemilihan Durasi *Database*

Data yang dimasukkan kedalam *Data Warehouse* ialah data pembayaran pajak kendaraan bermotor yang tercatat di dalam database Dinas Pendapatan Daerah Kota Cilegon untuk seluruh wilayah kecamatan dan kelurahan Kota Cilegon selama tiga tahun terakhir, yaitu dari tahun 2011 sampai 2013.

Tabel 1: Durasi Database

<u>Nama Data Warehouse</u>	<u>Database</u>	<u>Database ada sejak tahun</u>	<u>Data yang masuk ke Data Warehouse</u>	<u>Total Record</u>
DW_ <u>Dispenda</u>	<u>Dispenda</u>	2006	2011-2013	613123 <u>baris data</u>

3.2.7 Melacak Perubahan dari Dimensi Secara Perlahan

Sebagian data yang berada dalam tabel dimensi memiliki nilai yang tetap, misalnya jenis kelamin pengendara, waktu dan lain sebagainya. Tetapi ada beberapa kemungkinan atribut atau nilai dari data tersebut akan berubah dalam waktu yang cukup lama, misalnya nama suatu kecamatan ataupun kelurahan suatu saat mengalami perubahan. Oleh karena itu data yang sudah lama harus dilakukan perubahan atau *updating* untuk tetap menjaga keakuratan data

3.2.8 Penentuan Prioritas dan Model Query

Data warehouse yang dibuat akan menjadi dasar laporan analisis dispenda, baik dari segi pembuatan dokumentasi atau pencarian informasi baru (*data mining*). Pihak dispenda dapat membuat laporan analisis dalam bentuk tabel ataupun grafik dari berbagai dimensi.

3.3 Arsitektur Data Warehouse

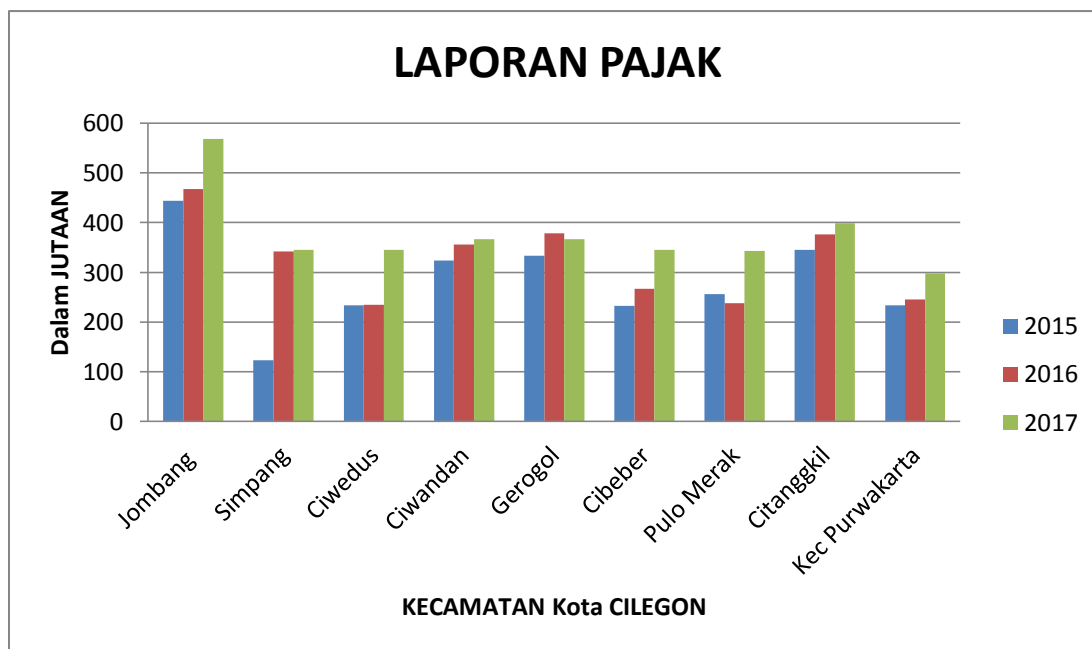
Perancangan *Data Warehouse* pada Dispenda Kota Cilegon ini menggunakan *Enterprise Data Warehouse Architecture*. Arsitektur ini merupakan tahapan proses pengumpulan beberapa sumber yang terpisah kemudian disatukan kedalam satu tempat yang bertujuan untuk memudahkan proses *loading* data ke dalam tabel fakta. Penggunaan *enterprise data warehouse* juga bertujuan untuk menghindari penggandaan data dan mempermudah dalam pemeliharaan data. Terdapat tiga langkah utama pada arsitektur ini yaitu proses ETL (*Extract, Transform, Loading*). Proses ETL dijalankan pada *Integration Service Project* di *Visual Studio : Business Intelligence*.

3.3.1 Extracting Data

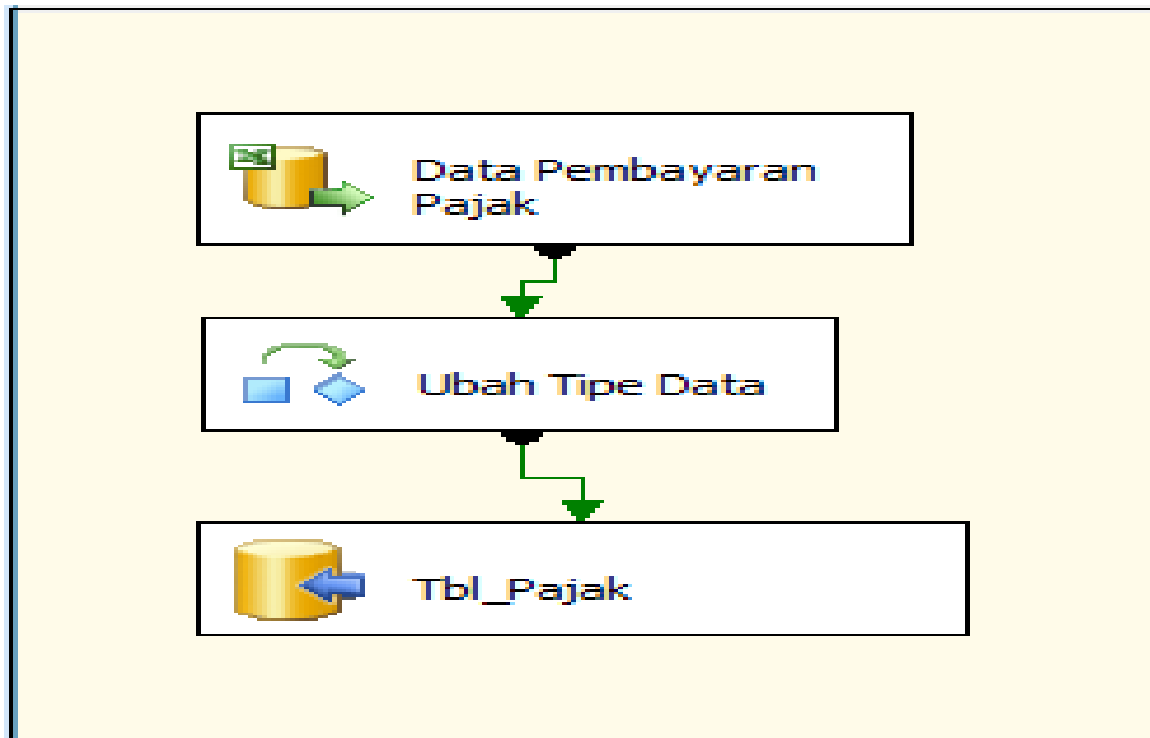
Sumber data yang digunakan dalam perancangan *data warehouse* dan sumber data yang digunakan dalam perancangan *data warehouse* pada Dispenda Cilegon ialah data *MS. Excel*.

Tabel 2 Sumber Data Pembayaran Pajak Kendaraan Bermotor dalam Bentuk *Ms. Excel*

Keluahan/Kecamatan		2015	2016	2017
Jombang		444	467	568
Simpang		123	342	345
Ciwedus		234	235	345
Ciwandan		324	356	367
Gerogol		333	378	367
Cibeber		232	267	345
Pulo Merak		256	238	343
Citangkil		345	376	399
Kec Purwakarta		234	245	298



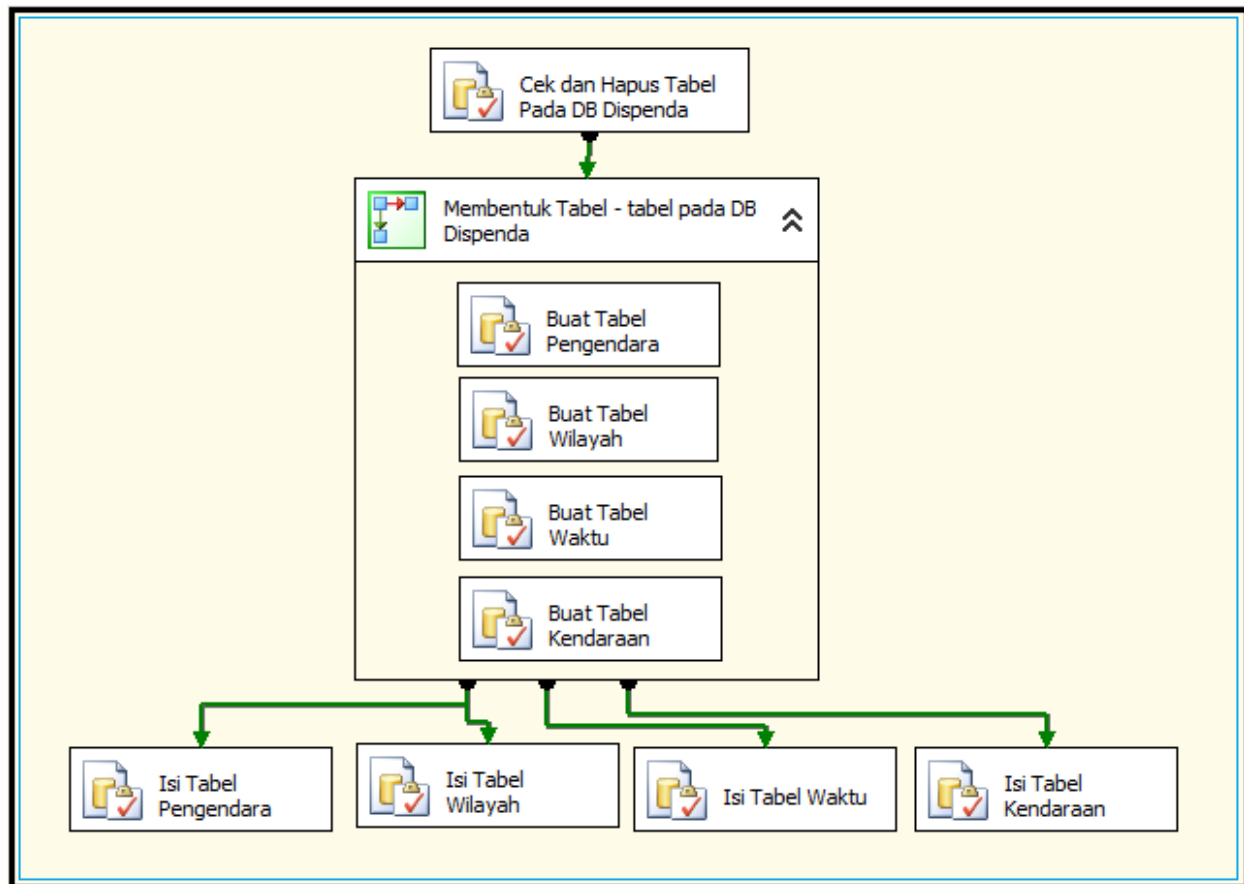
Kemudian pada sumber data tersebut di lakukan proses Ekstraksi atau perubahan format sumber data ke menjadi format *data warehouse* yang dalam hal ini merupakan proses *input* data dari *MS. Excel* ke dalam tabel yang disediakan pada database Dispenda di *MS. SQL Server*.



Gambar 5 Proses Ekstrak Data *Ms. Excel*

3.3.2 Transformation

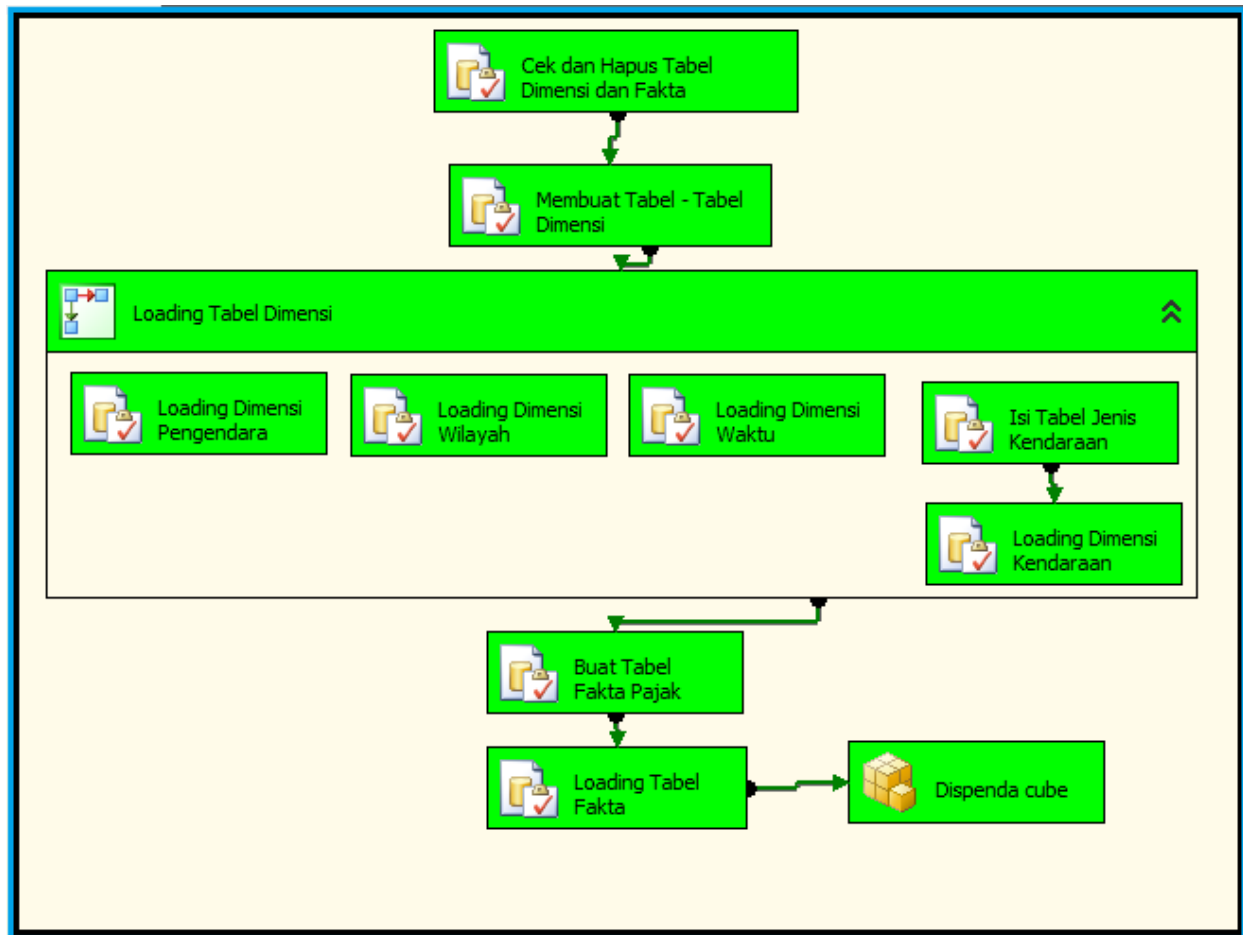
Setelah data-data berhasil dimasukkan ke dalam tabel pajak pada *database* Dispenda, isi dari tabel pajak akan diisi ke beberapa tabel baru yang dibuat di dalam *database* Dispenda yaitu table waktu, pengendara, kendaraan dan tabel wilayah.



Gambar 6 Paket Transformasi pada SSIS

3.3.3 Loading Tabel Dimensi dan Tabel Fakta

Proses ini merupakan tahapan pembuatan tabel-tabel dimensi pada *database* DW_Dispenda yang bertujuan untuk mengisi tiap tabel-tabel dimensi, setiap isi tabel dimensi berasal dari tabel *database* Dispenda 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* Dispenda yang berisikan data-data pembayaran pajak kendaraan bermotor selama 3 tahun terakhir dari tahun 2011 sampai 2013 untuk setiap atribut pada dimensi waktu, antara lain :

- Jumlah kendaraan dan jumlah pembayaran pajak kendaraan bermotor (PKB) berdasarkan jenis kelamin, umur dan status pengendara.
- Jumlah kendaraan dan pembayaran PKB berdasarkan kecamatan dan kelurahan pengendara.
- Jumlah kendaraan dan jumlah pembayaran PKB berdasarkan jenis dan merk kendaraan.
- Jumlah kendaraan dan pembayaran PKB berdasarkan type dan tahun keluaran kendaraan.

Tabel 3
Tampilan Data Warehouse Dispenda
Berdasarkan Dimensi Pengendara

Drop Filter Fields Here									
			Tahun ▾	Nama Bulan ▾					
			2011						
			Januari		Februari		Maret		April
Status ▾	Umur ▾	JK ▾	Jml Kendaraan	Pajak	Jml Kendaraan	Pajak	Jml Kendaraan	Pajak	Jml Kendaraan
Pelajar/Mhs.	17		69	24842475	45	21964500	68	32493975	54
	18		77	26516650	50	18583000	62	28209775	63
	19		76	29282650	48	19750500	43	22930275	50
	20	L	35	15036150	32	22079150	29	14919900	33
		P	40	20829800	27	8532000	33	21710150	22
		Total	75	35865950	59	30611150	62	36630050	55
	21		71	34222650	51	25415400	59	35814925	66
	22		84	31082775	53	29965300	65	41934425	44
	23		105	31191900	52	22770800	59	28189400	70
	24		83	36102775	56	24096000	55	25956450	45
	25		69	26987825	49	12983250	48	16183900	62
	26		70	30569250	49	18864275	63	43243750	51
	27	L	28	8699000	18	5207000	22	7109225	33
		P	32	8139775	27	7957000	28	15044325	22
		Total	60	16838775	45	13164000	50	22153550	55
	Total		839	323503675	557	238168175	634	333740475	615
PNS	19		83	31391775	56	21013700	54	22652475	67
	20		72	25618875	59	20982200	63	19357700	55
	21		72	27029825	49	16333300	53	26028575	55
	22		73	27940150	56	23151700	51	19388000	50
	23		87	35881975	55	19869125	51	24538075	52
	24		72	22523600	49	18783900	68	25842250	53
	25		67	21420575	41	13738200	56	26307800	50
	26		91	39045200	53	18029900	52	23186325	67
	27		77	33582000	60	19083750	66	36345400	67
	28								

Pada tabel 3 dispenda dapat dilihat jumlah kendaraan untuk pengendara dengan status pelajar atau mahasiswa yang berumur 20 tahun dan berjenis kelamin laki-laki sejumlah 35 kendaraan dengan jumlah pembayaran pajak sebesar Rp. 15.036.150,00 untuk bulan Januari 2011.

4.1.2 Analisis Pertumbuhan Jumlah Kendaraan dan Pembayaran Pajak Kendaraan

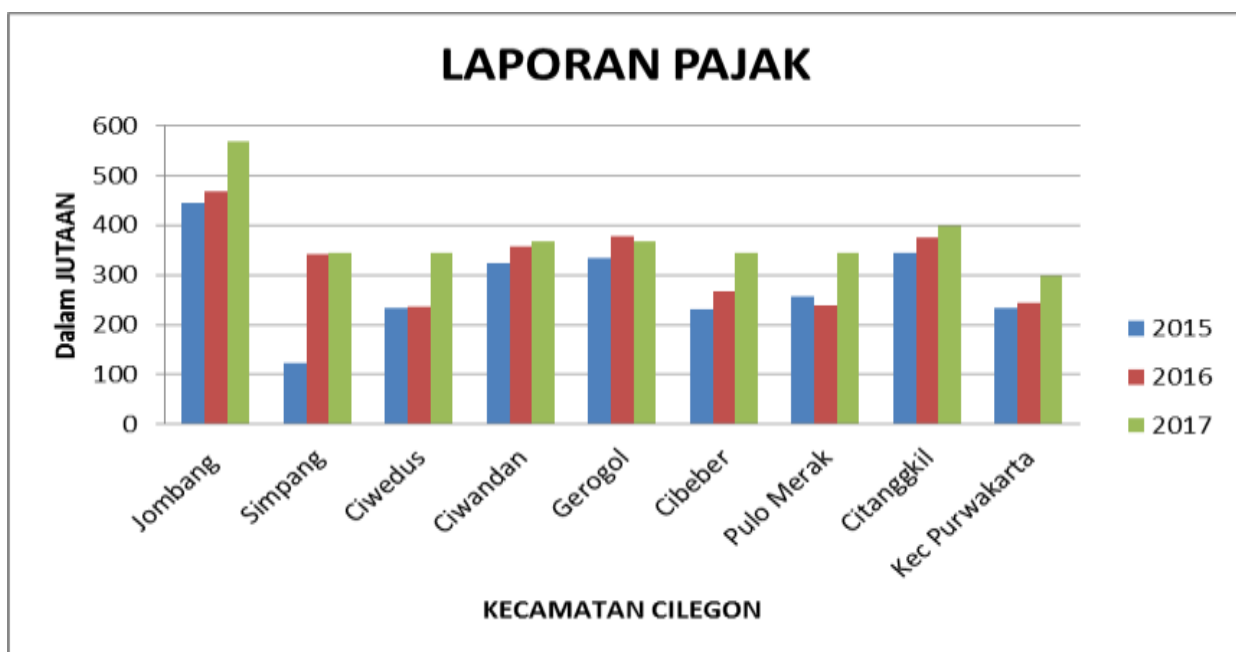
Salah satu tujuan dibangunnya *data warehouse* pada Dispenda Kota Cilegon yaitu agar dapat dilakukan analisis pertumbuhan jumlah kendaraan dan pembayaran untuk setiap periode waktu. Analisis pertumbuhan dapat dilihat baik pertahun, pertiga bulan, perbulan dan perminggu bahkan perhari. Tampilan analisis pertumbuhan jumlah kendaraan dan pembayaran pajak bermotor pertahun yang tercatat dalam database Dispenda Kota Cilegon, dapat dilihat pada tabel 4.

Tabel 4 Pertumbuhan Data Pertahun

Tahun ▼ Tahun Dim Waktu Calculations ▼					
2011		2012			
Current Dim Waktu		Current Dim Waktu		Year Over Year Growth %	
Jml Kendaraan	Pajak	Jml Kendaraan	Pajak	Jml Kendaraan	Pajak
190315	85476219345	235984	109033456490	24.00%	27.56%

Berikut ini tampilan beberapa laporan dalam bentuk grafik batang, *surface* dan *pai*. Laporan dalam bentuk grafik batang dapat dilihat pada gambar 8 yang menunjukkan

jumlah pembayaran pajak untuk setiap kecamatan selama 3 tahun

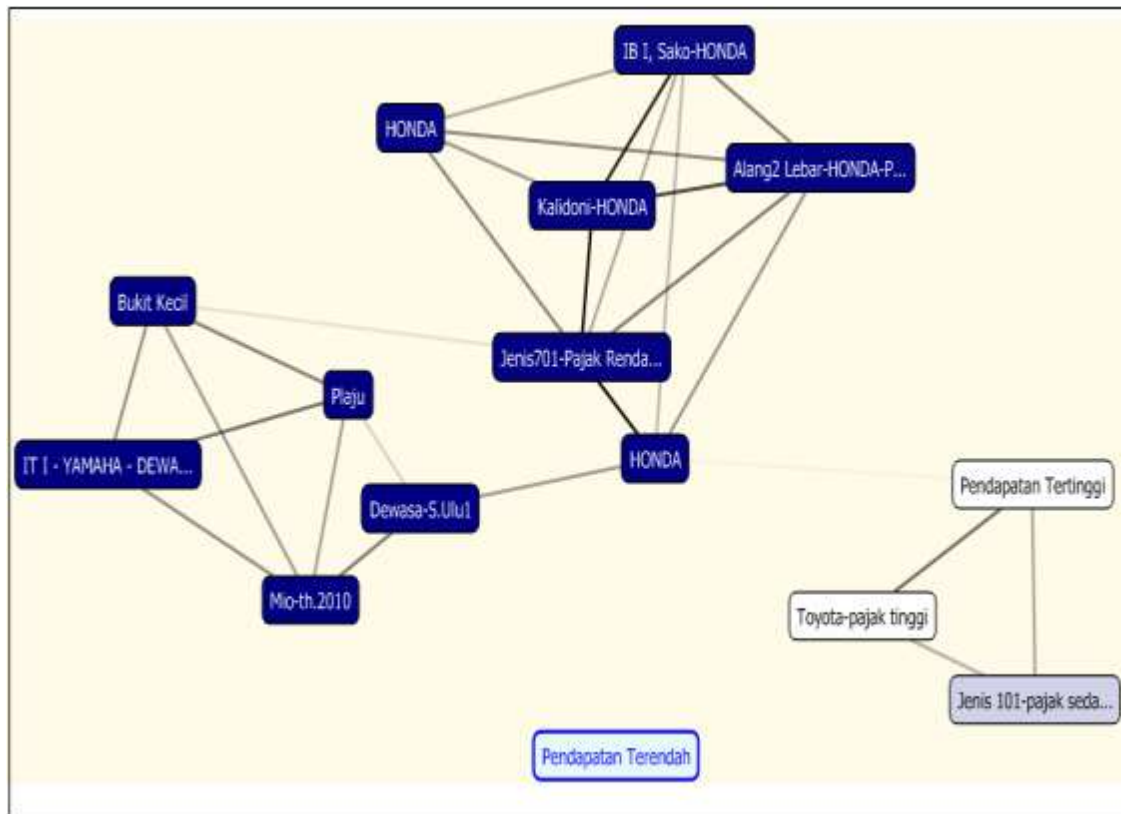


Gambar 8 Laporan Jumlah Pembayaran Pajak Per Kecamatan dalam Bentuk Grafik Batang

4.2 Hasil Data Mining

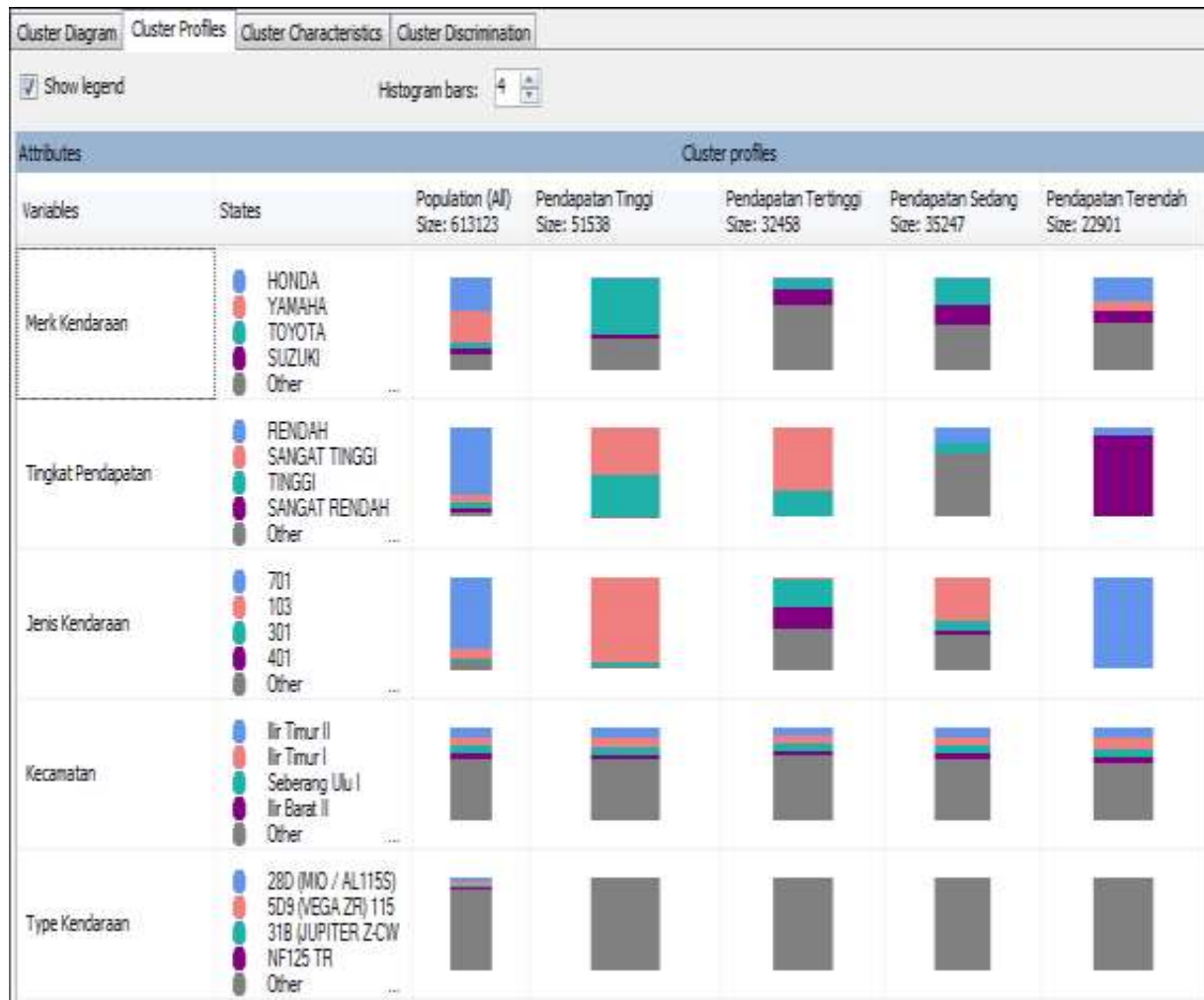
Untuk mendapatkan tujuan yang diinginkan, akan dibentuk 15 cluster dengan menggunakan algoritma EM. Berikut pengaturan untuk mining seluruh data pada Dispenda Kota Cilegon.

Dari gambar 9 dapat dilihat hasil cluster terbagi menjadi tiga area yaitu cluster – cluster yang memiliki nilai pendapatan pajak yang rendah, cluster yang memiliki nilai pendapatan terendah dan cluster – cluster yang memiliki nilai pendapatan menengah ke atas.



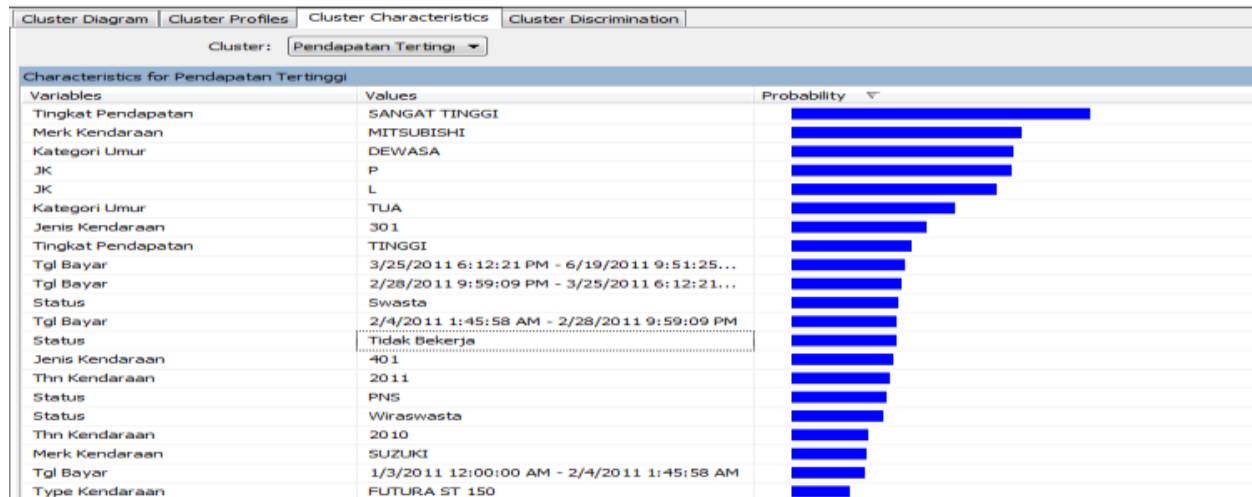
Gambar 9 Diagram Cluster pada Model Mining

Adapun karekteristik masing – masing *cluster* dapat dilihat pada gambar 10



Gambar 10 Diagram Cluster pada Model Mining

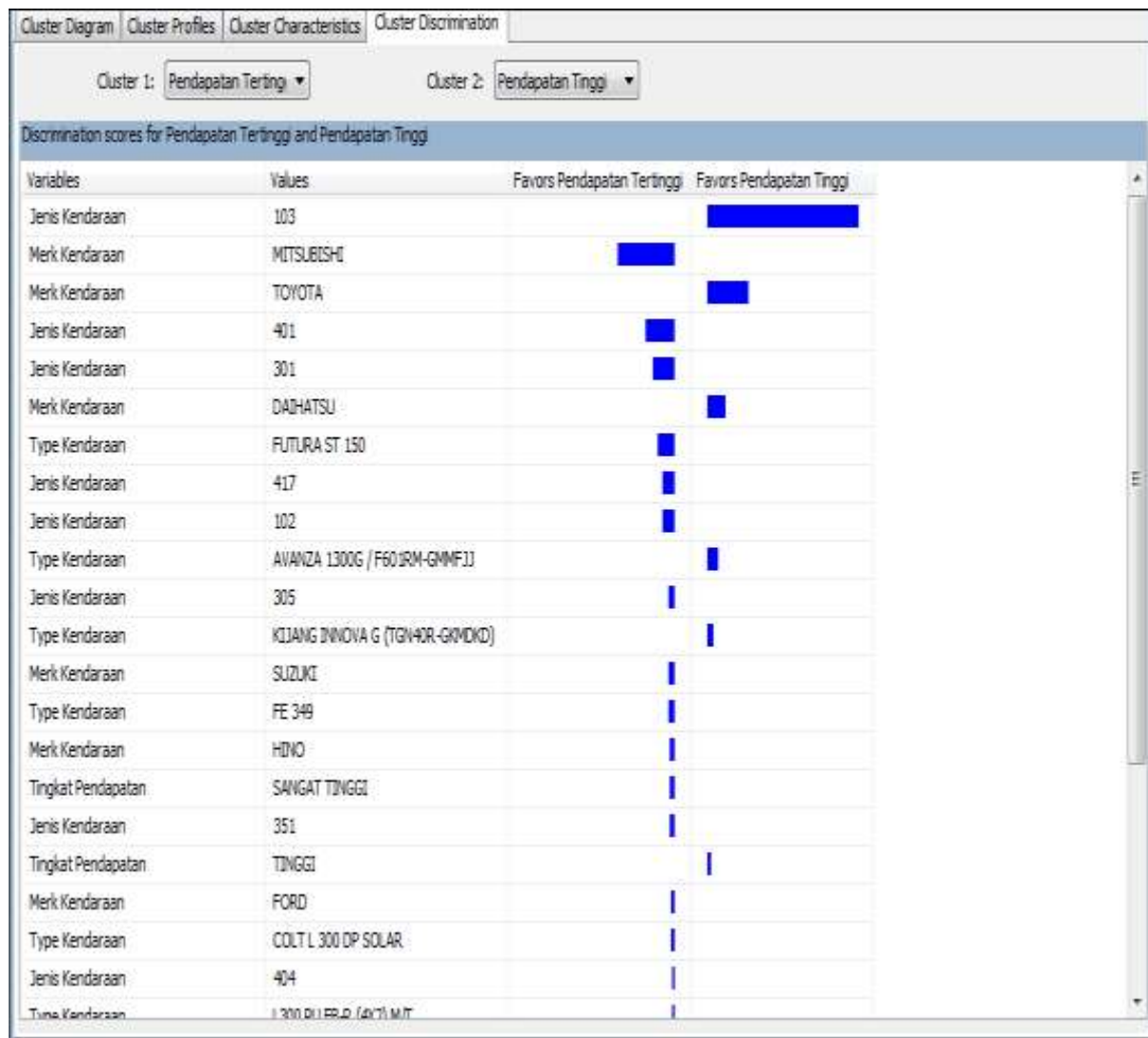
Dari gambar 10 dapat dilihat karakteristik pada kluster pendapatan tertinggi, sedang dan terendah. Dapat dilihat kluster yang memiliki pendapatan menengah keatas memiliki didominasi oleh jenis kendaraan 103, 301 dan 401, sedangkan pada kluster dengan nilai pendapatan sangat rendah memiliki seluruh atribut jenis kendaraan 701.



Gambar 11 Karakteristik Kluster Pendapatan Tertinggi pada Model Mining

Dari gambar 11 dapat dilihat karakteristik dari kluster dengan nilai pendapatan sangat tinggi, dengan nilai atribut merk kendaraan yang paling tinggi yaitu „MITSUBISHI“, jenis kendaraan „301“ dan status pengemudi karyawan swasta.

Pada *Microsoft Clustering* dapat juga dilihat nilai perbedaan diantara dua kluster. Tampilan *cluster discrimination* dapat dilihat pada gambar 12.



Gambar 1
Diskriminasi Cluster Pendapatan Tertinggi dengan
Cluster Pendapatan Terendah pada Model Mining

Dapat disimpulkan dari gambar 12, perbedaan antara kluster pendapatan tertinggi dengan pendapatan tinggi, yaitu banyak anggota kluster pendapatan tinggi yang memiliki jenis kendaraan „103“ yang tidak dimiliki oleh kluster pendapatan tertinggi. Dan banyak lagi hal yang dapat disimpulkan dari keterangan di atas.

Dapat disimpulkan dari beberapa penjelasan di atas, hasil analisis dari *data mining* dengan menggunakan algoritma EM pada dispenda ini antara lain :

1. Dari 15 cluster yang dibuat dapat terbagi menjadi tiga area yaitu 12 cluster yang memiliki nilai pendapatan pajak yang rendah, dan masing-masing 1 cluster yang memiliki nilai pendapatan terendah, sedang, dan tinggi .
2. Perbedaan yang signifikan antara cluster pendapatan tertinggi dengan cluster pendapatan terendah yaitu pada cluster tertinggi terdapat banyak jenis kendaraan „301“ sedangkan pada pendapatan terendah semua populasi cluster merupakan tipe kendaraan „701“ dan pada cluster terendah banyak terdapat kendaraan

dengan merk kendaraan „Honda“ yang pada cluster tertinggi terdapat banyak kendaraan dengan merk kendaraan „Suzuki“ dan „Toyota“.

5.PENUTUP

5.1 Kesimpulan

1. Perancangan *data warehouse* pembayaran pajak kendaraan bermotor pada Dispenda Kota Cilegon, dirancang dengan menggunakan metodologi *nine step* dengan skema *snowflake* yang memiliki satu tabel fakta pajak dan empat tabel dimensi, yaitu dimensi waktu, pengendara, wilayah dan kendaraan dimana dimensi kendaraan memiliki relasi dengan tabel jenis kendaraan.
2. Penerapan *data mining* pada Dispenda Kota Cilegon dirancang menurut metodologi CRISP-DM, dengan membentuk model mining *clustering* menggunakan algoritma EM(*Expectation Maximation*) pada *Business Intelligent* sehingga terbentuk 15 *cluster* dari data pembayaran pajak kendaraan bermotor selama 3 tahun, yang setiap *cluster* memiliki karakteristik yang unik.

5.2 Saran

1. Diharapkan *data warehouse* ini dapat dikembangkan lagi seperti dapat melihat jumlah pengendara yang membayar pajak tepat waktu ataupun yang terkena denda.
2. Pihak Dispenda Kota Cilegon dapat menambah data-data yang digunakan untuk proses *data mining*, sehingga informasi yang dihasilkan lebih akurat.
3. Diterapkan Teknik data mining *clustering* dengan menggunakan *algoritma* selain EM(*Expectation Maximation*) misalnya dengan menggunakan *K-Means* untuk mendapatkan hasil *cluster* yang baru.
4. Dibuat *mining model* yang baru selain *Clustering* pada *data warehouse* Dispenda, misalnya model *mining* menggunakan *Microsoft time series* untuk memprediksi jumlah kendaraan dan pembayaran pajak pada periode ke depan.

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*, Springer-Verlag Berlin Heidelberg, London.
- [3] Kimball, R, Margy R, Warren T, Joy M and Bob B 2008, *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing Inc, Canada.
- [4] Indrajani 2009, *Sistem Basis Data Dalam Paket Five In One*, PT.Elex Media Komputindo, Jakarta.
- [5] Nuraida, Ida 2008, *Manajemen Administrasi Perkantoran*, Kanisius, Yogyakarta.
- [6] Ponniah, Paulraj 2010, *Data Warehouseing*. Canada : John Wiley & Sons Inc.
- [7] Rivero, L, Jorge D and Vivian F 2006, *Encyclopedia of Database Technologies and Applications* : Idea Group Inc.
- [8] Rainaldi, Vincent 2008, *Building A Data Warehouse With Examples In SQL Server*, Appress.
- [9] Tantra, Rudi 2012, *Manajemen Proyek Sistem Informasi*, Andi Offset, Yogyakarta.
- [10] Jiawei Han and Micheline Kambar 2010, *Data Mining Concepts and Techniques*, Verlag Berlin.
- [11] Oded Maimon and Lior Rokach 2010, *Data Mining And Knowledge Discovery Hanbook*, Springer Science.