

PENGEMBANGAN SISTEM ANALISIS AKADEMIS MENGGUNAKAN OLAP DAN DATA CLUSTERING STUDI KASUS : AKADEMIK UNIVERSITAS SEBELAS MARET SURAKARTA

Bakharudin Yusuf Bakhtiar

Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami 36A Kertingan
Surakarta
bakharudin@student.uns.ac.id

Antonius Bima Murti Wijaya

Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami 36A Kertingan
Surakarta
bimamurti@uns.ac.id

Hasan Dwi Cahyono

Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami 36A Kertingan
Surakarta
hasandc@staff.uns.ac.id

Abstrak— Tingkat keberhasilan universitas dapat dilihat dari kualitas lulusannya. Salah satu cara meningkatkan keberhasilan tersebut adalah dengan melengkapi dan memanfaatkan sistem informasi. Universitas Sebelas Maret merupakan universitas yang memanfaatkan sistem informasi secara *departemental* karena beberapa dibangun dengan menggunakan *framework* dan *database* yang berbeda. Dimana hal ini akan menyulitkan civitas akademik dalam mengolah data dan menganalisis data-data tersebut. Aplikasi Pengembangan Sistem Analisis Akademis Menggunakan OLAP dan Data Clustering ini menjadi solusi untuk mengatasi masalah tersebut sekaligus membantu menganalisis data-data. Proses pembuatan aplikasi ini meliputi perancangan *data warehouse*, pengambilan data akademik dari sumber data, proses *extraction, transformation, loading* (ETL), pembuatan *cube*, pembuatan laporan, dan pengelompokan data menggunakan K-means clustering.

Penelitian ini berhasil mengembangkan sistem OLAP dan clustering untuk memenuhi fungsionalitas sistem yaitu sistem dapat melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah mahasiswa, jumlah calon mahasiswa baru, jumlah lulusan, mencari nilai maksimum, minimum, rata-rata dan presentase profil ipk lulusan, mencari nilai maksimum, minimum, rata-rata dan presentase profil lama studi lulusan, beban sks dosen, presentase pemakaian ruang dari berbagai dimensi.

Kata Kunci : OLAP, K-means, Clustering, ETL, Data warehouse.

1. PENDAHULUAN

Menurut Prof. Dr. Ravik Karsidi, M.S sebagai rektor dari Universitas Sebelas Maret (UNS) periode 2011-2015 menyatakan bahwa keberhasilan UNS yang lebih utama terlihat dari semakin meningkatnya kualitas lulusan yang ditandai dengan semakin meningkatnya rata-rata Indeks Prestasi Kumulatif (IPK), Angka Efisiensi Edukasi (AEE), lulusan berpredikat *cumlaude* serta semakin singkatnya masa studi. Salah satu cara yang digunakan UNS untuk meningkatkan keberhasilan tersebut adalah dengan melengkapi dan memanfaatkan sistem informasi.

Namun sistem informasi yang ada di UNS sampai saat ini masih bersifat *departemental* karena beberapa dibangun dengan menggunakan *framework* dan *database* yang berbeda. Dengan sistem yang masih bersifat *departemental* ini akan menyulitkan civitas akademik dalam mengolah data dan

menganalisis data-data tersebut karena harus dihimpun dari berbagai unit yang sulit untuk didapatkan secara cepat dan terkadang terdapat data yang tidak valid. Seperti kesulitan dalam pembuatan laporan, penyusunan borang akreditasi, menentukan segmentasi pasar promosi SMA dan lain sebagainya.

Dalam penelitian yang dilakukan oleh Kavitha [1] menyebutkan bahwa *data warehouse* merupakan dasar dari OLAP yang dapat melakukan analisis interaktif data multidimensi. *Online analytical processing* (OLAP) adalah sebuah teknik yang digunakan untuk menggabungkan data sehingga memungkinkan untuk membantu *user* memecahkan masalah bisnis dengan menggali data transaksional [2]. Dengan menggunakan OLAP, data yang bersifat *departemental* seperti yang terjadi di UNS dapat akan digabungkan menjadi sebuah database terpusat dengan konsep *multidimensi*. Kemudian dalam penelitian Usman [3] Kombinasi OLAP dan data Mining dapat digunakan untuk memvisualisasikan data yang kompleks agar dapat dianalisa secara efisien, interaktif dan bermakna. Untuk membantu melakukan analisis data-data yang terdapat dalam OLAP dapat menggunakan algoritma *clustering*. *Clustering* digunakan untuk mengelompokkan data sesuai dengan kesamaan karakteristik yang dimiliki setiap data dalam kelompok tersebut [4]. Algoritma K-means adalah salah satu metode *clustering* non hirarki yang paling populer dan banyak dikembangkan karena algoritma K-Means sederhana, efisien dan selalu konvergen [5]. Seperti halnya di UNS, algoritma K-means ini dapat digunakan untuk mengelompokkan data mahasiswa dimana data-data yang akan digunakan berasal dari *data warehouse*.

Berdasarkan permasalahan tersebut maka penulis ingin membuat sistem *clustering* K-means yang terintegrasi dengan OLAP melalui *data warehouse* sebagai sumber datanya dimana hasil dari *clustering* ini dapat membantu prodi UNS dalam menentukan segmentasi pasar SMA yang menjadi prioritas promosi program studi UNS.

2. DASAR TEORI

2.1. Data warehouse

Data warehouse merupakan database relasional yang didesain lebih kepada query dan analisis daripada proses transaksi, biasanya mengandung *history* data dari proses transaksi dan bisa juga data dari sumber lainnya. *Data warehouse* memisahkan beban kerja analisis dari beban kerja transaksi dan memungkinkan organisasi untuk menggabungkan data dari berbagai macam sumber [6] sehingga digunakan untuk membantu mendukung keputusan-keputusan manajemen [7].

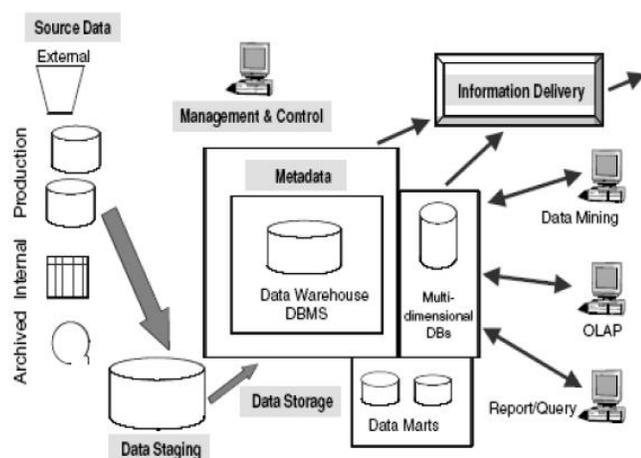
2.1.1. Karakteristik Data warehouse

Menurut Bill Inmom dalam buku Ponniah, karakteristik dari *data warehouse* [7] adalah sebagai berikut:

- a. Berorientasi subjek
Data warehouse didesain untuk menganalisis data berdasarkan subjek-subjek tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu.
- b. Terintegrasi
 Sumber data yang ada dalam *data warehouse* tidak hanya berasal dari data operasional (internal source) tetapi juga berasal dari data di luar sistem (external source).
- c. *Time-variant*
 Sistem operasional mengandung data yang bernilai sekarang sedangkan data dalam *data warehouse* mengandung data tidak hanya data terkini tetapi juga data masa lampau yang akan digunakan dalam analisis dan pengambilan keputusan.
- d. *Non-volatile*
 Data dalam database operasional akan secara berkala atau periodik dipindahkan ke dalam *data warehouse* sesuai dengan jadwal yang sudah ditentukan.
- e. *Granularity*
 Pada sistem operasional data dibuat secara real-time sehingga untuk mendapatkan informasi langsung dilakukan proses query. Granularitas menunjuk pada level perincian atau peringkasan yang ada pada unit-unit data dalam *data warehouse*.

2.1.2. Komponen Data warehouse

Menurut Ponniah [7] komponen-komponen *data warehouse* digambar sesuai dengan gambar 1 dimana komponen sumber data berada di sebelah kiri. kemudian komponen data *staging* sebagai blok pembangun. Komponen penyimpanan data yang mengelola *data warehouse* berada di tengah. Komponen *information delivery* yang terdiri dari semua hal menyediakan informasi dari *data warehouse* bagi pengguna berada di sebelah kanan.



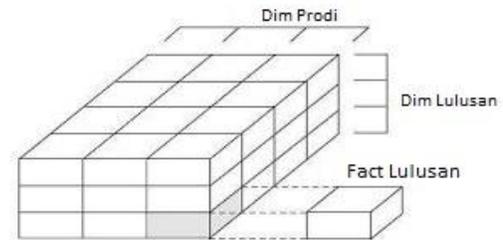
Gambar 1. Komponen Data warehouse [7]

Komponen sumber data yang digunakan dalam *data warehouse* dikelompokkan menjadi empat, yaitu: *Production Data, Internal Data, Archived Data, External Data*.

Didalam data staging akan dilakukan *Extraction, Transformation, dan Loading (ETL)* yang digunakan untuk membersihkan, mengubah, menggabungkan, mengkonversi, mencegah duplikasi data, dan menyiapkan data sumber untuk penyimpanan dan penggunaan dalam *data warehouse*.

2.2. Model Data Multidimensional

Pembuatan *data warehouse* didasarkan pada model data multidimensi yang berbentuk kubus. Data multidimensi adalah ketika sebuah data dapat dipandang dari berbagai sudut. Pusat dari objek pada multidimensional adalah *cube* atau kubus yang mengandung hubungan struktur dimensi, hirarki, level dan anggota.



Gambar 2. Data Multidimensional

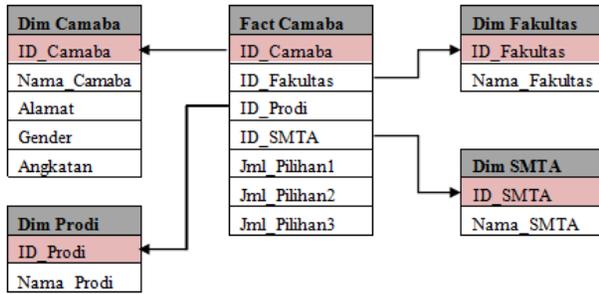
Menurut Prasetyo [8], komponen model multidimensional yang secara umum ditemukan dalam perancangan *data warehouse* adalah :

- a. Dimensi
 merupakan kategori yang independen dari multidimensional database. Tipe dari dimensi ini mengandung item yang digunakan sebagai kriteria query untuk ukuran database.
- b. Tabel fakta
 Tabel fakta merupakan pusat dari schema pada OLAP yang didalamnya mempunyai dua tipe kolom, yaitu kolom *measure* yang menyimpan nilai-nilai numerik dan kolom yang menyimpan *foreign key* yang mengacu ke tabel dimensi.
- c. Measure
Measure juga cerminan dari fakta dan juga mengandung data yang akan dianalisis. OLAP memerlukan informasi kolom bertipe numerik yang akan dijadikan *measure*
- d. Hirarki
 Hirarki merupakan bentuk kesatuan dari dimensi. Hirarki didefinisikan bagaimana hubungan antar level.
- e. Level
 Level merepresentasikan sebuah posisi pada hirarki. Level mengumpulkan data untuk agregasi dan digunakan untuk proses komputasi.
- f. Atribut
Attribute merepresentasikan informasi tambahan pada sebuah level tertentu.
- g. Cube
Cube adalah obyek OLAP yang tersusun dari *measure, dimensi dan attribute*. Sisi-sisi pada *cube* ditentukan oleh masing-masing dimensi yang terlibat dalam *cube* itu.

Menurut prasetyo [8] , model dimensional yang sering digunakan pada *data warehouse* adalah star atau snowflake skema.

a. Star schema

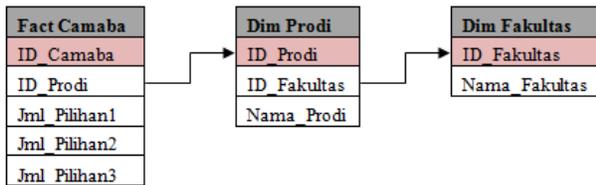
Star schema seperti yang terlihat pada gambar 3, terdiri dari satu atau lebih tabel fakta dan satu atau lebih tabel dimensi. Tabel fakta merupakan pusat dari star schema, karena fungsinya sebagai pengikat dari tabel-tabel dimensi yang terletak di sekelilingnya.



Gambar 3. Star Schema

b. Snowflake schema

Snowflake schema merupakan pengembangan dari star schema. Pada snowflake schema, untuk mengurangi nilai duplikat pada tabel tabel dimensi akan dilakukan normalisasi secara sebagian atau keseluruhan. Dengan kata lain satu atau lebih tabel dimensi tidak bergabung secara langsung kepada tabel fakta tapi pada tabel dimensi lainnya seperti pada gambar 4.



Gambar 4. Snowflake Schema

2.3. Clustering

Menurut Alfina, pengertian clustering dalam data mining adalah pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek yang berada dalam cluster lainnya [10].

2.3.1. K-means Clustering

Algoritma K-means termasuk dalam partitioning clustering yang memisahkan data ke k daerah bagian yang terpisah. K-means merupakan algoritma yang sangat terkenal karena kemudahannya dan kemampuannya untuk mengkluster data besar dan data outlier dengan sangat cepat. Sesuai dengan karakteristik partitioning clustering, setiap data harus termasuk ke dalam cluster tertentu, dan memungkinkan bagi setiap data dalam satu tahapan proses tertentu sudah masuk dalam satu cluster, pada satu tahapan berikutnya data tersebut akan berpindah ke cluster lain [11].

langkah-langkah dalam pembuatan algoritma K-Means adalah sebagai berikut [12]:

1. Menentukan k sebagai jumlah cluster yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k.
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidian

(Euclidian Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

dimana:

x_i : data kriteria

μ_j : centroid pada cluster ke-j

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Memperbaharui nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad (2)$$

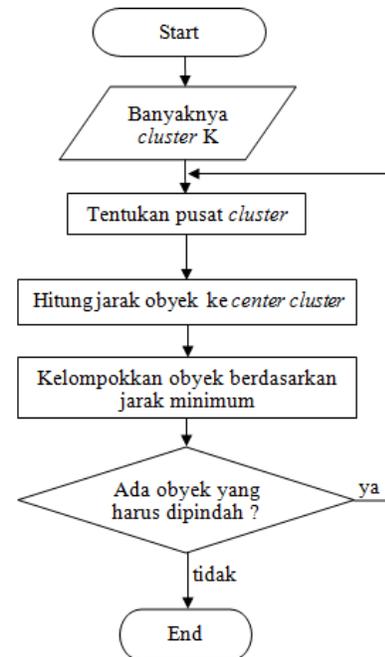
dimana:

$\mu_j(t+1)$: centroid baru pada iterasi ke (t+1),

N_{sj} : banyak data pada cluster S_j

6. Melakukan perulangan dari langkah 2 hingga 5 hingga anggota tiap cluster tidak ada yang berubah.
7. Jika langkah 6 telah terpenuhi, maka nilai pusat cluster (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

Berikut adalah gambar flowchart algoritma K-Means :



Gambar 5. Flowchart Algoritma K-means [13]

2.4. Sum Square Error (SSE)

SSE (Sum Square Error) adalah salah satu metode statistik yang dipergunakan untuk mengukur selisih total dari nilai sebenarnya terhadap nilai yang tercapai [14].

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} d^2(m_i, x) \quad (3)$$

Dimana

:

- d : jarak titik representatif ke pusat cluster
- x : pusat cluster dalam cluster C_i

- m : titik representatif untuk *cluster* Ci
 Hasil *clustering* yang baik adalah jika memiliki nilai SSE paling rendah.

3. METODOLOGI PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut :

3.1. Tahap pengumpulan data

3.1.1. Data primer

Data atau informasi yang diperoleh secara langsung dari pihak narasumber, dalam kasus ini narasumber adalah civitas akademik UNS. Setiap jawaban dari pertanyaan dan uraian hasil wawancara merupakan data yang nantinya dapat digunakan dalam menganalisis kebutuhan sistem yang akan dibuat.

3.1.2. Data sekunder

Pengumpulan data sekunder ini dapat dilakukan dengan menggunakan metode studi literatur dan telaah dokumen. Studi literatur dilakukan dengan mencari bahan materi yang berhubungan dengan permasalahan, perancangan, *data warehouse*, proses ETL, OLAP dan metode *K-Means Clustering* untuk mempermudah proses implementasi sistem. Pencarian materi dilakukan melalui pencarian di buku panduan, internet dan dokumen yang didapatkan dari UNS.

3.2. Tahap analisis

Dalam tahap analisis membuat Spesifikasi Kebutuhan Perangkat Lunak (SKPL) yang terdiri dari : menentukan tabel dan kolom pada data sumber yang akan diambil, merancang tabel dimensi dan tabel fakta *data warehouse* dengan menggunakan *snowflake skema*, membuat usecase untuk aplikasi *K-means clustering* dan membuat ERD untuk aplikasi *K-means clustering*.

3.3. Tahap desain

Dalam tahap desain ini terdiri dari dua tahapan yaitu membuat Deskripsi Perancangan Perangkat Lunak (DPPL) dan merancang algoritma *K-means clustering*.

3.4. Tahap implementasi

3.4.1. Implementasi pembersihan data (ETL)

Dalam tahapan ini, sumber data yang berasal dari database SIAKAD, SPMB, dan SIGEJE akan di *extract* ke dalam data *staging* kemudian data harus ditransformasikan sesuai dengan format desain *data warehouse* yang dibuat. Setelah data dibersihkan baru kemudian dilakukan proses *loading* ke *data warehouse*.

3.4.2. Implementasi Pembuatan OLAP

Setelah proses ETL selesai maka *data warehouse* siap digunakan untuk pembuatan *cube multidimensional* OLAP.

3.4.3. Implementasi algoritma K-means clustering

Tahap ini akan mengimplementasikan keseluruhan algoritma *k-means* yang ada di tahap proses *clustering* data.

3.5. Pengujian validasi hasil

Dalam tahapan ini dilakukan pengujian hasil *clustering* menggunakan *Sum of Square Error (SSE)* untuk menghitung kesalahan jarak dari cluster terdekat. Hasil *clustering* yang baik

adalah jika pada *iterasi* terakhir memiliki nilai SSE yang paling rendah dibanding *iterasi* sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. Pembangunan data warehouse dan OLAP

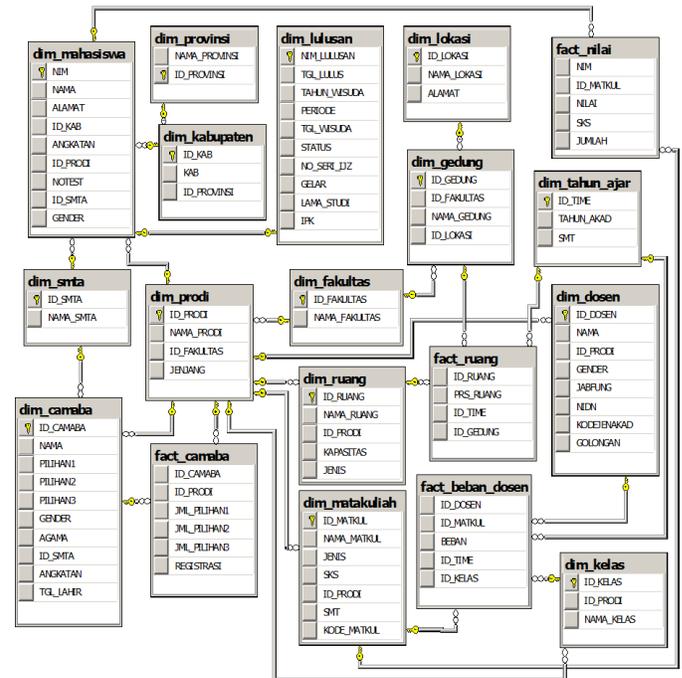
Setelah dilakukan analisis terhadap sumber data untuk mengetahui apa saja informasi yang harus disediakan oleh sistem. Fungsi-fungsi yang dapat ditangani sistem adalah :

- a. Melihat jumlah mahasiswa, jumlah lulusan, jumlah mahasiswa yang diwisuda, jumlah calon mahasiswa baru dari berbagai dimensi.
- b. Melihat jumlah calon mahasiswa baru untuk setiap prodi.
- c. Melihat profil IPK dan lama studi lulusan.
- d. Melihat profil beban sks dosen.
- e. Melihat jumlah mahasiswa tiap kelas.
- f. Melihat presentase pemakaian gedung.
- g. Melakukan pengelompokkan data

Untuk memenuhi kebutuhan fungsi-fungsi tersebut, sistem akan mengambil data-data yang terdapat dalam database SIAKAD, SPMB dan SIGEJE.

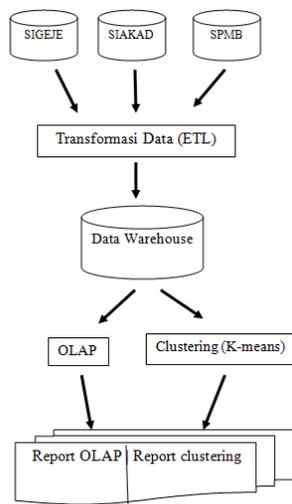
Data-data dari data sumber tersebut akan diolah sesuai dengan format yang diinginkan, seperti data yang mempunyai format “*nvarchar*” akan diubah menjadi “*varchar*” dan data-data yang akan digunakan untuk proses perhitungan, tipe datanya akan diubah sesuai dengan kebutuhan seperti dari format “*varchar*” ke format “*float*” atau dari format “*varchar*” ke format “*integer*”.

Dari analisis yang dilakukan dibuat sebuah skema model data menggunakan *snowflake schema* yang akan digunakan untuk membangun *data warehouse* seperti pada gambar 6.



Gambar 6. Rancangan *snowflake schema*

Dalam proses pengambilan data dari data sumber sampai dengan pembuatan laporan akan dilakukan seperti arsitektur perangkat lunak pada gambar 7.



Gambar 7. Arsitektur Perangkat Lunak

Pada table 1 merupakan table sumber data, *staging area* dan *data warehouse*.

Tabel 1. Daftar Tabel sumber data, staging area dan data warehouse.

Database Sumber	Tabel Sumber	Staging Area	Data warehouse (Result)	Data warehouse (Area)
SIGEJE	FAKULTAS	fakultas	dim_fakultas	-
SIGEJE	PRODI	prodi	dim_prodi	-
SIGEJE	LOKASI	lokasi	dim_lokasi	-
SIAKAD	replika_identitas_mahasiswa	r_mahasiswa	dim_provinsi	-
SIAKAD	replika_identitas_mahasiswa	r_mahasiswa	dim_kabupaten	dim_provinsi
SIGEJE	DOSEN	dosen	dim_dosen	-
SPMB	replika_camaba	r_camaba	dim_smta	-
SIAKAD	replika_identitas_mahasiswa	r_mahasiswa	dim_mahasiswa	dim_SMTA, dim_kabupaten, dim_provinsi, dim_smta
SIGEJE	MAKULPENAWARAN	mapen	dim_matakuliah	-
	MATAKULIAH	Matakuliah		
SIGEJE	RUANG	ruang	dim_ruang	dim_prodi
SIGEJE	TAS	TAS	dim_tahun_ajar	-
SPMB	replika_camaba	r_camaba	dim_camaba	dim_smta
SIAKAD	replika_alumni	r_alumni	dim_lulusan	dim_mahasiswa
SIGEJE	KELAS	kelas	dim_kelas	-
	PRODI	prodi		
SIGEJE	LOKASI	lokasi	dim_lokasi	-
SIGEJE	GEDUNG	gedung	dim_gedung	-

			fact_camaba	dim_camaba, dim_mahasiswa
SIGEJE	PLOTDOSEN	plot_dosen	fact_beban_dosen	dim_matakuliah, dim_kelas
SIGEJE	MAKULPENAWARAN	mapen		
SIAKAD	replika_khs_mipa	r_khs	fact_nilai	dim_mahasiswa, dim_matakuliah
	replika_khs_ekonomi			
	replika_khs_fk			
	replika_khs_fisip			
	replika_khs_fkip			
	replika_khs_hukum			
	replika_khs_teknik			
	replika_khs_sastra			
SIAKAD	replika_alumni	r_alumni	fact_lulusan	dim_mahasiswa
SIGEJE	RUANG	ruang	fact_ruang	-
SIGEJE	WAKTU	waktu		
SIGEJE	RUANGWAKTU	ruang_waktu		

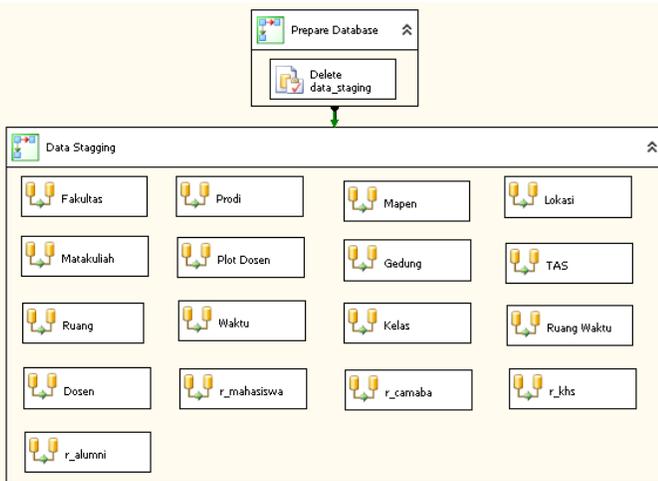
Tahapan-tahapan pembangunan yang akan dilakukan dalam membangun *data warehouse* dan pembuatan OLAP seperti tabel 2, pada tabel tersebut database yang berada pada platform berbeda (basis data *heterogen*) akan disatukan dengan platform yang sama (basis data *homogen*).

Tabel 2. Tahap-tahap pembangunan data warehouse

Proses	Sumber dan Tujuan	Tahapan Dalam Proses
Data Sumber ke Staging Area	Basis data heterogen	Membuat <i>package</i> baru
		menentukan <i>Control Flow Item</i> yang akan digunakan
		menentukan sumber metadata
		Menentukan <i>Staging Area</i> untuk metadata
Staging Area ke Data warehouse	Basis data homogen	Membersihkan, menggabungkan dan me-load tabel-tabel pada <i>staging area</i>
		<i>Execute Package</i>
		Load data ke <i>staging area</i>
		Membuat <i>package</i> baru
Pembuatan Cube	Data warehouse ke Analysis Service	menentukan <i>Control Flow Item</i> yang akan digunakan
		Menentukan sumber metadata
		Menentukan metadata untuk <i>data warehouse</i>
		Me-load data ke tabel dimensi dan tabel fakta
		<i>Execute Package</i>
		Load data ke <i>data warehouse</i>
		Menentukan <i>data source</i>
		Menentukan <i>data source views</i>
		Membuat <i>cube</i>
		Mengubah attribut, measure dan hirarki pada <i>cube</i>
		Mendefinisikan kalkulasi pada <i>cube</i>
		Deploy Analysis Services database

Langkah-langkah untuk melakukan proses ETL dalam penelitian ini pada dasarnya terbagi menjadi dua bagian yaitu :

1) *Control Flow ETL* Bagian I

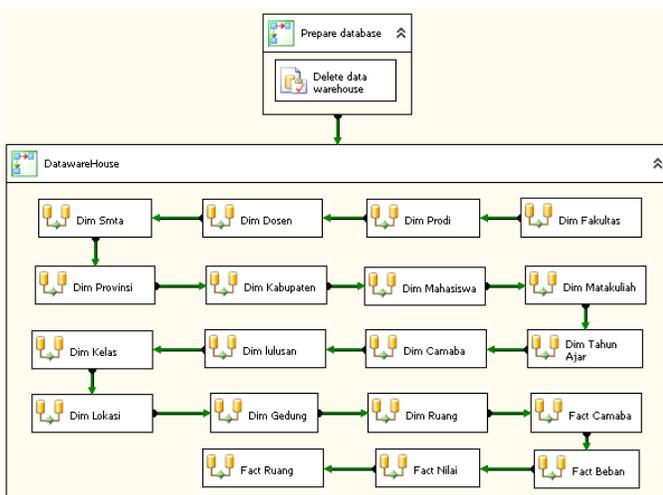


Gambar 8. Desain *Control Flow ETL* Bagian I

Pada gambar 8 merupakan desain *control flow ETL* yang pertama. Pada *container* “*Prepare Database*” merupakan proses penghapusan isi data *staging* agar siap diisi data baru.

Kemudian pada *container* “*Data Staging*” merupakan tempat dimana proses ETL yang pertama dilakukan. Dalam proses ETL data sumber yang berasal dari SPMB, SIAKAD dan SIGEJE akan di *ekstrak* kemudian dilakukan *transformasi* untuk mengisi nilai yang kosong dan mengganti tipe data agar sesuai dengan format rancangan *data warehouse*. Setelah dilakukan *transformasi* data, data tersebut disimpan dalam data *staging* sebelum diolah menjadi *data warehouse*.

2) *Control Flow ETL* Bagian II



Gambar 9. Desain *Control Flow ETL* Bagian II

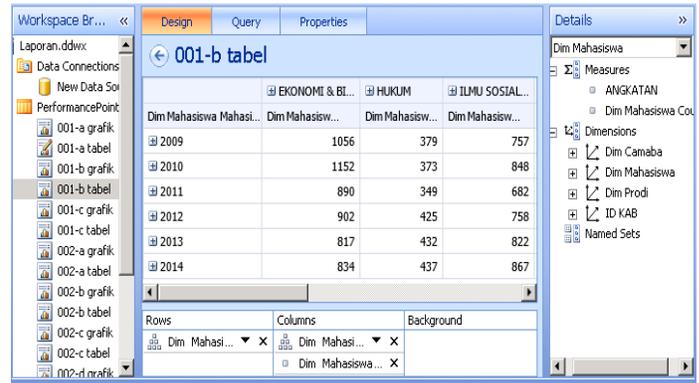
Pada gambar 9 merupakan desain *control flow ETL* yang kedua. Pada *container* “*Prepare Database*” merupakan proses penghapusan isi *data warehouse* agar siap diisi data baru.

Kemudian pada *container* “*Data warehouse*” akan dilakukan proses ETL dari data *staging* ke *data warehouse*. Data-data yang terdapat dalam data *staging* akan di *ekstrak* lalu

dilakukan transformasikan sesuai dengan kebutuhan seperti pada tabel 1, kemudian data tersebut disimpan dalam *data warehouse*.

4.2. *Dashboard OLAP*

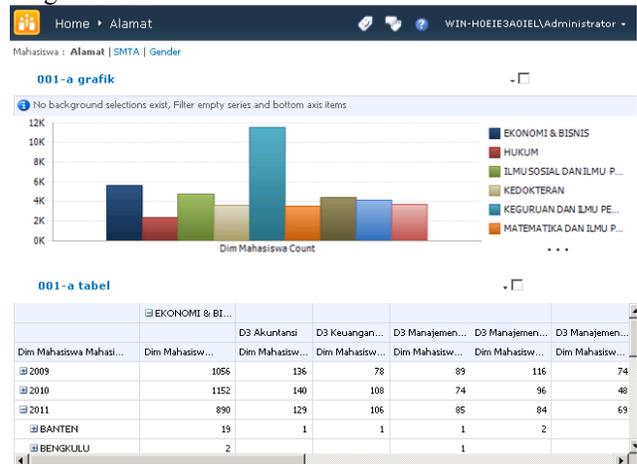
Setelah pembuatan cube OLAP selesai dilakukan, untuk menampilkan OLAP agar sesuai dengan kebutuhan yang diinginkan dapat menggunakan *sharepoint*. Pada gambar 10 merupakan contoh pembuatan laporan menggunakan *sharepoint*.



Gambar 10. Penggunaan *Sharepoint*

Berikut ini merupakan hasil dari *dashboard* laporan sesuai dengan fungsi sistem :

- a. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah mahasiswa dari dimensi fakultas, prodi, tahun masuk, provinsi dan kabupaten seperti pada gambar 11



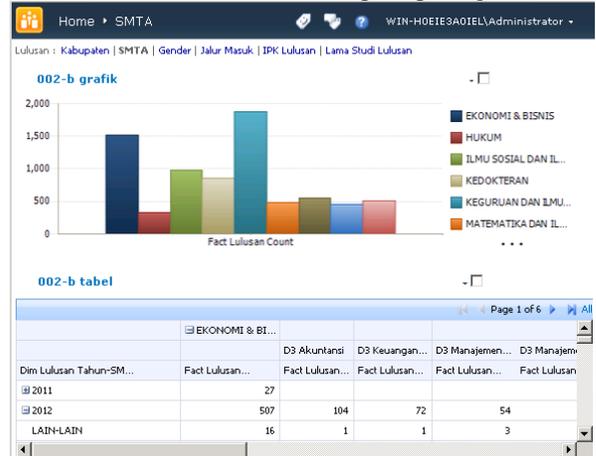
Gambar 11. *Dashboard* Laporan Alamat Mahasiswa

- b. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah mahasiswa dari dimensi fakultas, prodi, tahun masuk dan asal SMTA seperti pada gambar 12



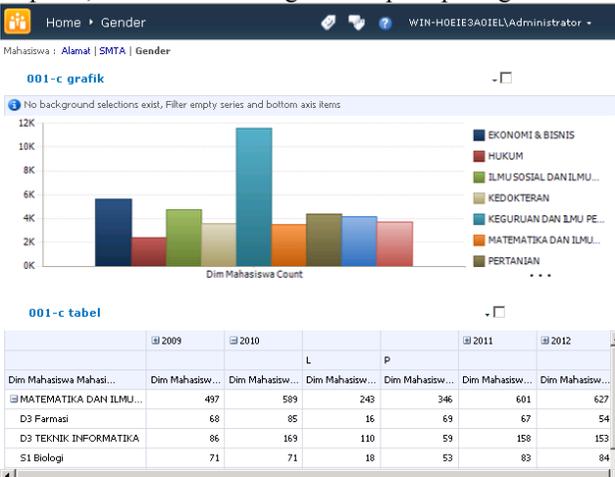
Gambar 12. Dashboard Laporan asal SMTA Mahasiswa

e. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah lulusan dari dimensi fakultas, prodi, tahun wisuda dan asal SMTA seperti pada gambar 15



Gambar 15. Dashboard Laporan asal SMTA Lulusan

c. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah mahasiswa dari dimensi fakultas, prodi, tahun masuk dan gender seperti pada gambar 13



Gambar 13. Dashboard Laporan Gender Mahasiswa

f. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah lulusan dari dimensi fakultas, prodi, tahun wisuda dan gender seperti pada gambar 16



Gambar 16. Dashboard Laporan Gender Lulusan

d. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah lulusan dari dimensi fakultas, prodi, tahun wisuda, provinsi dan kabupaten seperti pada gambar 14



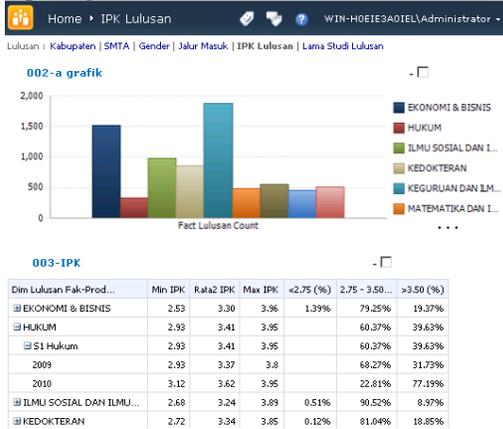
Gambar 14. Dashboard Laporan Kabupaten Lulusan

g. Melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah lulusan dari dimensi fakultas, prodi, tahun wisuda dan jalur masuk seperti pada gambar 17



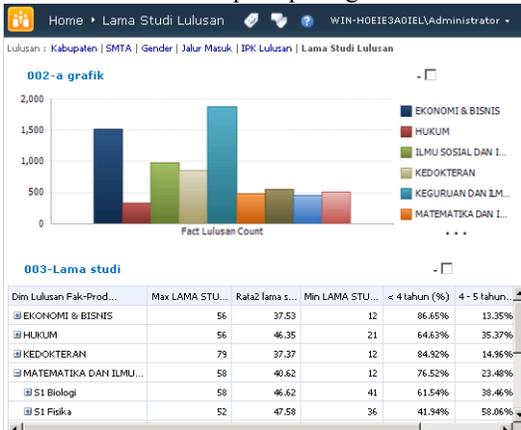
Gambar 17. Dashboard Laporan Jalur Masuk Lulusan

- h. Melakukan *drill up* dan *drill down* untuk membantu merekap profil IPK lulusan yang terdiri dari nilai ipk minimal, ipk maksimal, rata-rata IPK dan persentase IPK yang dilihat dari dimensi fakultas, prodi dan tahun wisuda seperti pada gambar 18



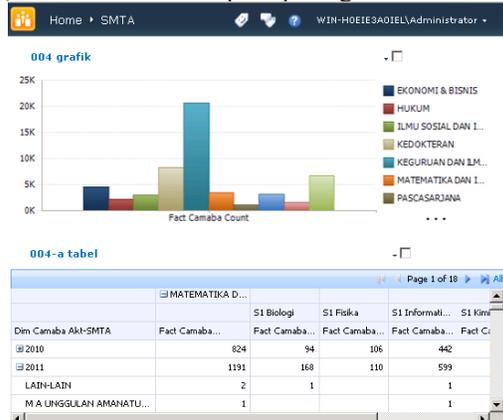
Gambar 18. Dashboard Laporan IPK Lulusan

- i. Melakukan *drill up* dan *drill down* untuk membantu merekap profil lama studi lulusan yang terdiri dari lama studi terlama, lama studi tercepat, rata-rata lama studi dan persentase lama studi yang dilihat dari dimensi fakultas, prodi dan tahun wisuda seperti pada gambar 19



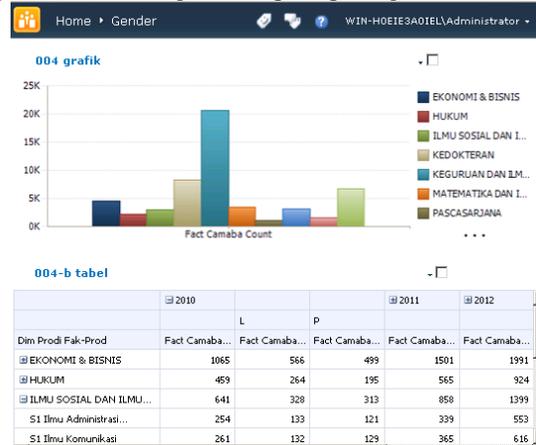
Gambar 19. Dashboard Laporan Lama Studi Lulusan

- j. Melakukan *drill up* dan *drill down* untuk membantu merekap data calon mahasiswa baru dari dimensi fakultas, prodi, tahun dan SMTA seperti pada gambar 20



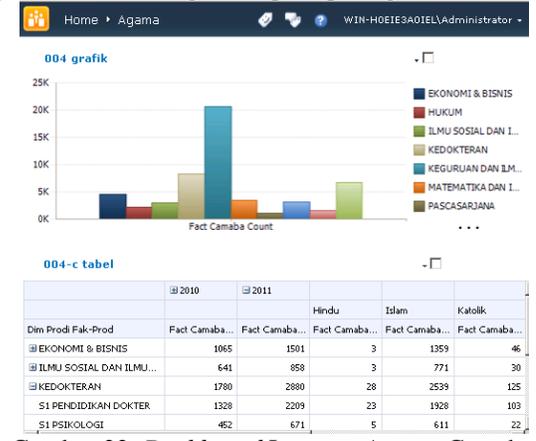
Gambar 20. Dashboard Laporan asal SMTA Camaba

- k. Melakukan *drill up* dan *drill down* untuk membantu merekap data calon mahasiswa baru dari dimensi fakultas, prodi, tahun dan gender seperti pada gambar 21



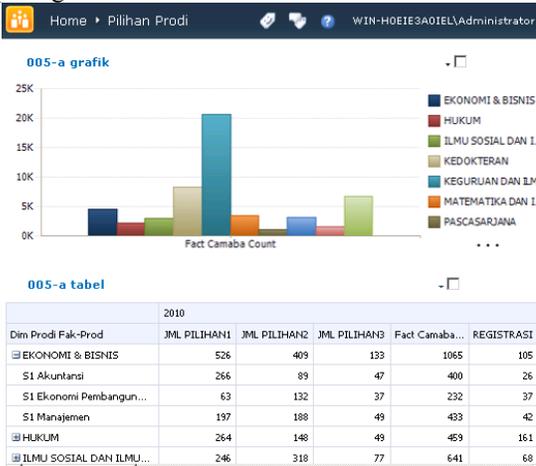
Gambar 21. Dashboard Laporan Gender Camaba

- l. Melakukan *drill up* dan *drill down* untuk membantu merekap data calon mahasiswa baru dari dimensi fakultas, prodi, tahun dan agama seperti pada gambar 22



Gambar 22. Dashboard Laporan Agama Camaba

- m. Melakukan *drill up* dan *drill down* untuk membantu merekap data calon mahasiswa baru dari dimensi fakultas, prodi, tahun, jumlah pilihan 1, jumlah pilihan 2, jumlah pilihan 3, total camaba yang melakukan registrasi seperti pada gambar 23



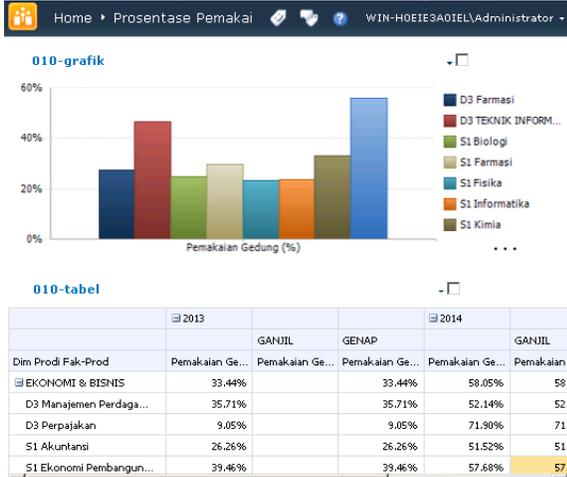
Gambar 23. Dashboard Laporan Pilihan Camaba

- n. Melakukan *drill up* dan *drill down* untuk membantu merekap data beban sks dosen dari dimensi fakultas, prodi, dosen dan tahun ajaran seperti pada gambar 24



Gambar 24. Dashboard Laporan Bebas SKS Dosen

- o. Melakukan *drill up* dan *drill down* untuk membantu merekap data pemakaian gedung dari dimensi fakultas, gedung, tahun ajaran dan semester seperti gambar 25



Gambar 11. Dashboard Laporan Pemakaian Gedung

4.3. K-means Clustering

K-means *clustering* digunakan untuk mengelompokkan data-data alumni UNS sesuai dengan *attribute* yang digunakan. Dalam penggunaan K-means ini *attribute* yang digunakan adalah lama studi dan nilai IPK setiap alumni UNS.

Dalam penelitian ini, data yang digunakan untuk melakukan *clustering* adalah data alumni jurusan S1 Informatika Fakultas MIPA dari tahun lulus 2013 sampai 2014 dengan jumlah *cluster* sebanyak 3 buah.

4.3.1. Perhitungan algoritma K-means

Berikut ini adalah perhitungan K-Means menggunakan sample data pada titik :

Tabel 3. Sample data k-means clustering

NIM_LULUSAN	LAMA_STUDI	IPK
M0509006	4.5	3.22

- 1. Langkah pertama dalam proses perhitungan ini adalah menentukan banyaknya “k” sebanyak jumlah *cluster* yang diinginkan yaitu 3 buah.

- 2. Langkah kedua adalah menentukan *centroid* setiap *cluster*.
 - a. Pusat cluster 1 (4.48 ; 3.17)
 - b. Pusat cluster 2 (4.29 ; 3.43)
 - c. Pusat cluster 3 (4.10 ; 3.69)
- 3. Langkah ketiga adalah menghitung jarak obyek ke pusat *cluster* (*centroid*).

$$P1(x_i, \mu_j) = \sqrt{(4.5 - 4.48)^2 + (3.22 - 3.17)^2}$$

$$= \sqrt{0.0004 + 0.0025}$$

$$= 0.053852$$

Tabel 4. Hasil perhitungan jarak obyek ke centroid

Pusat	Lama Studi	IPK	Jarak
P1	4.48	3.17	0.053852
P2	4.29	3.43	0.296985
P3	4.10	3.69	0.617171

- 4. Kemudian data dikelompokkan berdasarkan jarak minimum obyek ke pusat *cluster*. Sesuai dengan hasil perhitungan jarak obyek ke pusat *cluster* diketahui bahwa jarak terdekat data dengan pusat *cluster* 1 (P1) sehingga *sample* data masuk kedalam kelompok *cluster* 1. Karena jarak P1 < P3 < P2.

- 5. Selanjutnya dilakukan pengecekan kelompok data *sample* terhadap kelompok data sebelumnya, apabila data mengalami perubahan *cluster* maka nilai *centroid* akan diperbarui dimana nilai *centroid* yang baru diperoleh dari rata-rata kelompok *cluster* yang sama. Kemudian kembali ke langkah nomer 3.

$$P1_{Lama\ studi} = \frac{Jumlah\ lama\ studi\ di\ centroid\ 1}{Jumlah\ data\ pada\ centroid\ 1}$$

$$= \frac{4.5 + 4.58 + 4.58 + 4.58}{4}$$

$$= \frac{18.24}{4}$$

$$= 4.56$$

$$P1_{IPK} = \frac{Jumlah\ IPK\ pada\ centroid\ 1}{Jumlah\ data\ pada\ centroid\ 1}$$

$$= \frac{3.22 + 3.06 + 3.05 + 3.36}{4}$$

$$= \frac{12.69}{4}$$

$$= 3.173$$

Tabel 5. Nilai centroid lama dan centroid baru

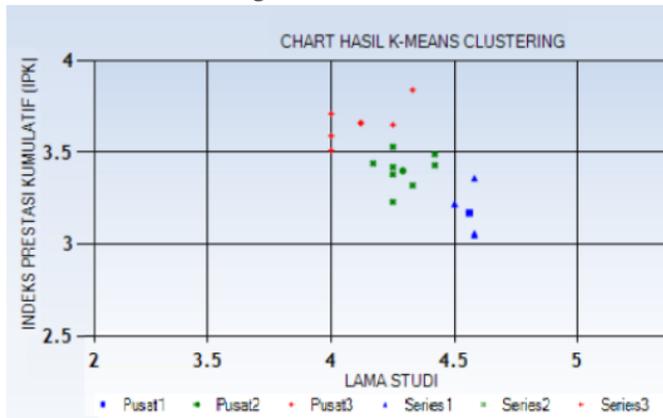
Pusat	Lama	Baru
P1	(4.48 ; 3.17)	(4.56 ; 3.173)
P2	(4.29 ; 3.43)	(4.29 ; 3.405)
P3	(4.10 ; 3.69)	(4.11 ; 3.660)

- 6. Jika sudah tidak ada lagi data yang berpindah kelompok pada masing-masing *cluster* maka proses dinyatakan selesai.

4.3.2. Hasil algoritma K-means

Hasil pengelompokan ini akan ditampilkan dalam bentuk *chart* dan table dimensi yang terdapat dalam table database *dim_lulusan* dan *dim_mahasiswa* seperti :

a. *Chart* hasil *clustering*



Gambar 25. *Chart* Hasil *Clustering*

b. Nilai IPK (<2.75, 2.75-3.50, dan >3.50).

Dimensi IPK			
IPK	kelompok1	kelompok2	kelompok3
2.00 - 2.75	0	0	0
2.76 - 3.50	4	7	0
3.51 - 4.00	0	1	5

Gambar 26. Hasil *Clustering* Dimensi IPK

c. Lama Studi (< 4 tahun, 4-5 tahun dan, > 5 tahun)

Dimensi Lama Studi			
Masa Studi	kelompok1	kelompok2	kelompok3
< 4 tahun	4	8	5
4 - 5 tahun	0	0	0
> 5 tahun	0	0	0

Gambar 27. Hasil *Clustering* Dimensi Lama Studi

d. Gender

Dimensi GENDER				
GENDER	kelompok1	kelompok2	kelompok3	Jumlah
P	2	4	2	8
L	2	4	3	9

Gambar 28. Hasil *Clustering* Dimensi Gender

e. Asal SMTA

Dimensi SMTA				
Nama SMTA	kelompok1	kelompok2	kelompok3	Jumlah
SMAN 2 KOTA MANNA KAB BENGKULU SELATA	0	1	0	1
SMA NEG 2 BENGKULU	0	1	0	1
SMA NEG 2 CIREBON	0	1	0	1
MAS COKROAMINOTO KARANGKOBAR KAB BANJ	0	0	1	1
SMA NEG 1 PURWOREJO	0	1	0	1
SMAN 1 GEMOLONG KAB SRAGEN	0	1	0	1
SMAN 1 KOTA SURAKARTA	0	0	1	1
SMA NEG 1 SURAKARTA	0	1	1	2
SMA NEG 3 SURAKARTA	2	1	1	4
SMA WARGA, SURAKARTA	1	0	0	1
SMA NEG 1 TEGAL	0	0	1	1
SMA NEG 1 YOGYAKARTA	0	1	0	1
SMA NEG 1 PONOROGO	1	0	0	1

Gambar 29. Hasil *Clustering* Dimensi SMTA

f. Provinsi

Dimensi Provinsi				
Nama Provinsi	kelompok1	kelompok2	kelompok3	Jumlah
BENGKULU	0	1	0	1
JAWA TENGAH	3	7	5	15
JAWA TIMUR	1	0	0	1

Gambar 30. Hasil *Clustering* Dimensi Provinsi

4.4. Pengujian K-means Clustering

Pada gambar 18 merupakan *chart* dan tabel hasil pengujian K-means *clustering* menggunakan metode SSE (*Sum Square Error*).



Gambar 31. Hasil Pengujian K-means menggunakan SSE

Dari gambar 31 diketahui bahwa pengujian memiliki hasil *clustering* yang baik karena nilai SSE selalu berkurang setiap kali iterasi dilakukan.

5. KESIMPULAN DAN SARAN

Pada penelitian ini berhasil mengembangkan sistem OLAP dan *clustering* untuk memenuhi fungsionalitas sistem yaitu sistem dapat melakukan *drill up* dan *drill down* untuk membantu merekap data jumlah mahasiswa, jumlah calon mahasiswa baru, jumlah lulusan, mencari nilai maksimum, minimum, rata-rata dan presentase profil ipk lulusan, mencari nilai maksimum, minimum, rata-rata dan presentase profil lama studi lulusan, beban sks dosen, presentase pemakaian ruang dari berbagai dimensi. Pada pengujian SSE *clustering* iterasi 1 adalah 0.398 sedangkan pada iterasi 2 nilai SSE *clustering* adalah 0.357. Penurunan nilai SSE ini menunjukkan bahwa jumlah selisih jarak setiap data ke pusat *cluster* semakin sedikit yang berarti semakin bagus. Hasil dari pengelompokkan ini, setiap data lulusan akan dikelompokkan berdasarkan kualitas lulusan yang dinilai berdasarkan IPK dan lama studi.

Adapun saran yang dipertimbangkan untuk pengembangan penelitian ini adalah mengembangkan proses ETL sehingga data dapat diperbarui secara otomatis jika terdapat perubahan pada data sumber. Kemudian menggabungkan algoritma k-means dengan algoritma pengambil keputusan seperti algoritma genetika agar sistem dapat melakukan pengambilan keputusan.

6. DAFTAR PUSTAKA

- Kavitha, P., 2013. A Survey of *Data warehouse* and OLAP Technology. *International Journal of Latest Trends in Engineering and Technology*, 3(1), pp.387-90.
- Janus, P. & Fouché, G., 2010. *Pro SQL Server 2008 Analysis Services*. United States, United States of America: Paul Manning.
- Usman, M., Asghar, S. & Fong, S., 2009. A Conceptual Model for Combining Enhanced OLAP and Data Mining Systems. In *INC, IMS and IDC. Fifth International Joint Conference on*. Seoul, 2009.
- Vipin Kumar, H.C.D.P., 2013. K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(4), pp.1-4.
- Kumar, Y. & G., S., 2014. A New Initialization Method to Originate Initial Cluster Centers for K-Means Algorithm. *International Journal of Advanced Science and Technology*, 62, pp.43-54.
- Lane, P., 2012. *Oracle OLAP Application Developer's Guide 10g Release 2 (10.2)*. California: Oracle Corporation.
- Ponniah, P., 2010. *Data Warehousing Fundamentals For IT Professionals*. 2nd ed. New Jersey, Canada: John Wiley & Sons, Inc.
- Prasetyo, M.A..S.A.&S., 2010. *Pembuatan Aplikasi OLAP Untuk Pelaporan pada PT. Aneka Tuna Indonesia Menggunakan SQL Server 2005*. [Online] Available at: <http://digilib.its.ac.id/public/ITS-Undergraduate-9803-Paper.pdf> [Accessed 22 February 2014].
- Adithama, S.P., Wisnubhadra, I. & Sinaga, B.L., 2013. Analisis Dan Desain Real-Time Business Intelligence Untuk Subjek Kegiatan Akademik Pada Universitas Menggunakan Change Data Capture. *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp.87-95.
- Alfina, T., Santosa, B. & Barakbah, A.R., 2012. Analisis Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS). *JURNAL TEKNIK ITS*, 1, pp.521-25.
- Noor, M.H. & Hariadi, M., 2009. Image Cluster Berdasarkan Warna untuk Identifikasi Kematangan Buah Tomat dengan Metode Valley Tracing. *Seminar Nasional Informatika*, pp.15-24.
- Sarwono, Y.T., 2010. Aplikasi Model Jaringan Syaraf Tiruan Dengan Radial Basis Function Untuk Mendeteksi Kelainan Otak (Stroke Infark). *Sekolah Tinggi Manajemen Informatika & Teknik Komputer Surabaya*.
- Andayani, S., 2007. Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means. *SEMNAS Matematika dan Pendidikan Matematika*.
- Sari, I.P., 2013. Perancangan Sistem Pengenalan Wajah Manusia Menggunakan Web Camera dengan Metode Summary Squared Error (SSE). *Repository Universitas Andalas*.