

MODEL DATA MINING DALAM PENGKLASIFIKASIAN KETERTARIKAN BELAJAR MAHASISWA MENGGUNAKAN METODE CLUSTERING

Marlindawati¹⁾, Andri²⁾

^{1), 2)} Sistem Informasi UNIVERSITAS BINA DARMA Palembang
Jl. Jend. A.Yani No. 12, Palembang, 30264
Email : marlindawati@mail.binadarma.ac.id¹⁾, andri@mail.binadarma.ac.id

Abstrak

Perkembangan teknologi informasi terutama yang berhubungan dengan media penyimpanan saat ini berkembang begitu pesat. Dari waktu ke waktu data dalam sebuah perguruan tinggi, terutama yang berhubungan dengan data mahasiswa akan terus mengalami peningkatan. Hal ini tentu saja akan menyebabkan data yang tersimpan akan terus bertambah, terutama data akademik mahasiswa. Jika data yang bertumpuk tidak dimanfaatkan dengan semaksimal mungkin, maka akan terjadi kemubaziran terhadap data tersebut. Dengan memanfaatkan teknik data mining maka dari data yang banyak tadi akan didapat informasi yang berguna. Data Mining merupakan proses analisis data yang menggunakan perangkat lunak untuk menemukan pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat menemukan pengetahuan guna mendukung keputusan. Dalam penelitian ini akan dilakukan mining data yang berhubungan dengan akademik mahasiswa untuk dapat mengetahui ketertarikan belajar dari mahasiswa yang ada dalam perguruan tinggi dengan metode clustering. Data dari penelitian ini berasal dari mahasiswa Sistem Informasi angkatan 2011-2013 Universitas Bina Darma yaitu 1.000 record data dari tb_mhs, 4.1367 record data dari tb_khs dan 417 record dari tb_mk. Proses mining menggunakan tools SSAS (SQL Sever 2008 Analysis Service). Setelah melakukan beberapa tahapan dari data mining, Hasil analisis menunjukkan terdapat tiga cluster yang terbentuk berdasarkan konsentrasi matakuliah pilihan. Cluster 1 klasifikasi minat untuk konsentrasi A sebanyak 400 mahasiswa. Cluster 2 klasifikasi minat untuk matakuliah konsentrasi B sebanyak 186 mahasiswa, dan cluster 3 klasifikasi minat matakuliah konsentrasi C sebanyak 188 mahasiswa.

Kata kunci: teknologi informasi, data mining, clustering.

1. PENDAHULUAN

Sumber utama yang dapat dimanfaatkan untuk menghasilkan suatu informasi adalah data. Informasi yang dihasilkan sangatlah penting bagi sebuah perguruan

tinggi. Pengolahan data yang maksimal akan menghasilkan informasi yang berkualitas dan bermanfaat bagi perguruan tinggi. Dari waktu ke waktu, data dalam sebuah perguruan tinggi akan terus mengalami peningkatan dan menjadi tumpukan data yang menggunung. Jika data tersebut tidak dikelola dengan baik maka data tidak ada manfaatnya. Dengan memanfaatkan teknik data mining dalam pengolahan data, maka dari data tersebut akan didapatkan informasi yang berguna dan bermanfaat bagi perguruan tinggi.

Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004) [2]. Menurut Pramudiono (2007) [3] data mining dapat juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari kumpulan data. Data mining dapat dimanfaatkan untuk menggali informasi dari tumpukan data akademik yang berhubungan dengan mahasiswa untuk mendapatkan informasi atau pola dari ketertarikan belajar mahasiswa. Informasi yang didapat dari teknik mining data ini merupakan sebuah keputusan yang dapat dimanfaatkan oleh pihak yang berwenang dalam sebuah perguruan tinggi dalam mengarahkan konsentrasi mahasiswa tersebut.

Menurut Davies (2004) [1], data mining memiliki beberapa karakteristik, yaitu:

- Data mining yang berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data mining biasanya menggunakan data yang sangat besar yang digunakan untuk membuat hasil yang lebih dipercaya.
- Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi

Data mining juga disebut dengan *Knowledge Discovery in Database* (KDD), yaitu kegiatan yang meliputi pengumpulan, pemakaian data untuk menemukan keteraturan, pola dan hubungan dalam set

data yang berukuran besar. Keluaran data mining dapat dipakai untuk pengambilan suatu keputusan.

Menurut Han (2006) [2] terdapat enam tahap dalam data mining:

- a. Pembersihan data (*cleaning*)
Merupakan proses untuk menghilangkan noise serta data yang tidak konsisten atau tidak relevan.
- b. Integrasi data (*integration*)
Bertujuan untuk menggabungkan data dari berbagai sumber basisdata dalam satu basisdata yang baru.
- c. Seleksi data (*selection*)
Bertujuan untuk memilih data-data apa saja yang akan diproses selanjutnya.
- d. Transformasi data (*transformation*)
Bertujuan untuk mengubah data kedalam format tertentu yang sesuai untuk proses selanjutnya.
- e. Proses mining (*data mining*)
Proses untuk menerapkan metode data mining yang bertujuan untuk menemukan pengetahuan yang baru.
- f. Evaluasi pola (*pattern evaluation*)
Bertujuan untuk mengidentifikasi pola-pola yang menarik kedalam knowledge base yang ditemukan.
- g. Presentasi pengetahuan (*knowledge presentation*)
Merupakan penyajian pengetahuan untuk memperoleh pengetahuan yang diperoleh pengguna.

Tujuan dari penelitian ini membuat model data mining untuk melakukan pengelompokan mahasiswa berdasarkan ketertarikan mereka terhadap mata kuliah yang disenanginya. Dan diharapkan dapat menghasilkan sebuah model data mining yang dapat dimanfaatkan untuk menentukan konsentrasi minat belajar mahasiswa yang ada pada program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Binadarma Palembang

Metode penelitian yang digunakan adalah metode pengembangan sistem SDLC (*system development life cycle*). Metode pengembangan sistem ini memiliki beberapa tahap, yaitu;

- a. Tahap Perencanaan
Tahap yang digunakan untuk melakukan identifikasi masalah, menentukan tujuan sistem, identifikasi kendala sistem dan membuat analisis kelayakan.
- b. Tahap Analisis
Tahapan untuk melakukan analisis terhadap permasalahan yang akan dihadapi dalam penelitian. Pada tahap ini akan dilakukan identifikasi penyebab masalah dan analisis kebutuhan sistem.
- c. Tahap Perancangan
Dalam tahap ini akan dilakukan proses perancangan berdasarkan data yang didapat dari tahap analisis kebutuhan sistem.
- d. Tahap Penerapan

Pada tahap ini dilakukan penerapan dari hasil rancangan dengan menggunakan tools yang ada.

- e. Tahap Penggunaan
Pada tahap ini dilakukan uji coba terhadap hasil rancangan yang telah dibuat.

Agar mendapatkan bahan-bahan yang relevan, akurat dan reliable dalam penelitian ini, maka penulis melakukan pengumpulan data dengan cara sebagai berikut:

- a. Observasi
Penulis melakukan pengamatan dan pencatatan data-data mahasiswa serta data-data yang berhubungan dengan nilai akademik fakultas ilmu komputer program studi Sistem Informasi angkatan 2011 s.d. 2013, Universitas Binadarma Palembang.
- b. Studi pustaka
Teknik pengumpulan data dilakukan dengan mengumpulkan bahan-bahan, referensi, dan dokumen yang berhubungan dengan penelitian ini.

2. PEMBAHASAN

Dalam pengelompokan minat mahasiswa dengan model data mining menggunakan metode Clustering, dibuatlah rancangan dari databasenya sebagai berikut :

2.1 Rancangan Database

a. Tb_mhs

Tabel 1. Tabel Mahasiswa

Nama	Type	Width
nim*	varchar	9
nama	varchar	40
jenis_kelamin	varchar	2
temp_lahir	varchar	50
tgl_lahir	date	10
kd_progdi	varchar	3
alamat	varchar	250
asal_sek	varchar	20
kota	varchar	25
pek_oru	varchar	20

b. Tb_khs

Tabel 2. Tabel Kartu Hasil Studi (KHS)

Nama	Type	Width
nim*	varchar	9
kd_mk	varchar	10
sms	varchar	1
sms_pendek	varchar	1
tahun_akademik	varchar	9
kelas	varchar	5
kd_dosen	varchar	5
kd_progdi	varchar	3
tugas	int	6
kuis	int	6
mid	int	6
semester	int	6
nilai_angka	int	6
nilai_huruf	varchar	1

c. Tb_mk

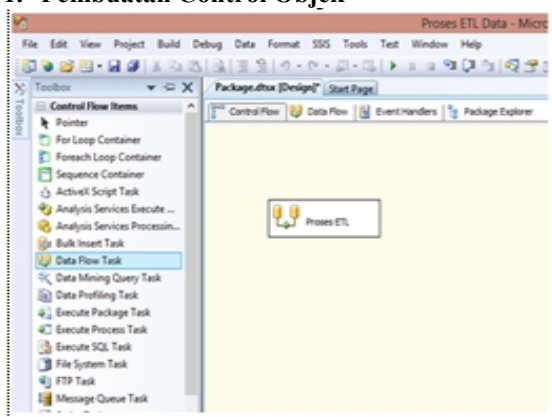
Tabel 3. Tabel Mata Kuliah

Nama	Type	Width
kd_mk*	varchar	10
nm_mk	varchar	30
sms	varchar	1
sks	int	1
kd_progdi	varchar	3

2.2 ETL (Extract, Transform and Load)

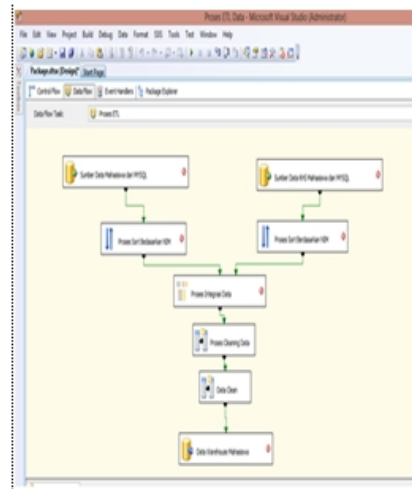
Proses ETL bertujuan melakukan pembentukan data warehouse yang berasal dari data mentah dalam format database MySQL. Data yang di ekstrak akan dilakukan proses transform dan load kedalam format baru yaitu format mdb. Proses ETL dilakukan menggunakan tools SQL Server Integration Service (SSIS) yang ada dalam Business Intelligence Development Studio (BIDS). ETL yang dilakukan dalam penelitian dapat dilihat pada gambar berikut ini:

1. Pembuatan Control Objek



Gambar 1. Pembuatan Control Objek

2. Proses ETL



Gambar 2. Proses ETL

2.3 Seleksi Data (Data Selection)

Data yang digunakan dalam penelitian berasal dari data mahasiswa Sistem Informasi Universitas Bina Darma tahun angkatan 2011 sampai dengan 2013. Selain data mahasiswa, data IPK juga digunakan untuk penelitian ini, dikarenakan informasi yang terkandung di dalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penentu dalam pengelompokan data keluaran yang diinginkan.

Data yang diperoleh adalah sebanyak 1.000 record data dari tb_mhs, 4.1367 record data dari tb_khs dan 417 record dari tb_mk. Dataset mahasiswa terdiri dari 23 atribut yang menjelaskan identitas diri mahasiswa dan informasi tentang keadaan mahasiswa yang bersangkutan. Atribut data mahasiswa diambil dari penggabungan beberapa tabel yang terdiri dari tb_mhs, tb_khs, dan tb_mk. Atribut tersebut diantaranya adalah nim, nama, jenis_kelamin, temp_lahir, tgl_lahir, kd_progdi, alamat, asal_sek, kota, pek_oru, kd_mk, sms, sms_pendek, tahun_akademik, kelas, kd_dosen, kd_progdi, tugas, kuis, mid, semester, nilai_angka, nilai_huruf, dan sks. Sedangkan dataset IPK hanya terdiri dari 5 atribut yang memberikan informasi mengenai prestasi akademik dan beban studi yang diambil mahasiswa yang bersangkutan. Atribut tersebut diantaranya adalah nim, nama, sum(sks), total_nilai, dan IPK.

Tahap-tahap yang akan digunakan dalam penelitian ini :

1. Proposes Data

Data yang digunakan dalam penelitian ini terdiri dari tabel mahasiswa, tabel mata kuliah, dan tabel khs.

a. Tb_mhs

Gambar 3. Record Tbel Mahasiswa

Gambar 6. Record Tabel Praproses Data

b. Tb_khs

Gambar 4. Record Tabel Kartu Hasil Studi

c. Tb_mk

Gambar 5. Record Tabel Mata Kuliah

2. Pembersihan Data (Data Cleaning)

Tahap berikutnya pada proses *data mining* adalah *cleaning data* yaitu melakukan pembersihan data terhadap *noise* yang ditemukan berupa *missing value*, *inkonsisten data*, dan *redundant data*.

3. Integrasi Data

Tahap ketiga adalah tahap integrasi data yang merupakan penggabungan data bertujuan memindahkan seluruh data yang telah di-*cleaning* ke dalam satu tabel. Pada tahap ini, ketiga tabel mhs, khs, dan mk akan di integrasikan untuk mendapatkan *data source* mahasiswa dan *data source* IPK. Untuk mendapatkan *data source* mahasiswa proses penggabungan data dilakukan dengan merelasikan *tbl_khs* dan *tbl_mk* dengan *join* antar table.

Proses selanjutnya merelasikan *tbl_mhs* dan tabel hasil integrasi atau *join* dari *tbl_khs* dan *tbl_mk*, dengan *join* antar tabel dengan *key nim*. Hasil dari integrasi atau *join* antar tabel diatas maka didapat *data source* mahasiswa, seperti pada gambar berikut 7 berikut.

Gambar 7. Tabel Data Source Mahasiswa

4. Transformasi Data

Proses selanjutnya adalah tahap tranformasi data yaitu data diubah menjadi bentuk yang sesuai untuk

diproses dalam *data mining*. Dalam penelitian ini data yang akan diproses dari *database mysql* untuk digunakan di *tool SSIS* maka format tersebut diubah menjadi *mdf*.

Setelah melakukan proses transformasi data, maka akan dilanjutkan ke proses selanjutnya yaitu proses *mining data*.

Data mining merupakan proses mencari pola atau informasi menarik dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan teknik dan algoritma yang tepat sangat bergantung pada proses KDD secara keseluruhan. Pada penelitian ini penerapan data mining menggunakan teknik *clustering* dan algoritma K-Means.

2.4 Analisis Data Mining

Analisis data mining yang digunakan dalam penelitian ini adalah teknik *clustering* atau pengelompokan yang bertujuan untuk mengelompokkan mahasiswa berdasarkan konsentrasi matakuliah yang telah dipilih dengan acuan nilai ipk yang didapat oleh mahasiswa tersebut.

Data yang digunakan untuk proses analisis data mining adalah data warehouse yang didapat dari proses awal dalam tahap data mining. Data warehouse berisikan data tentang nilai ipk dan konsentrasi pilihan matakuliah yang diambil oleh mahasiswa program studi sistem informasi.

Terdapat 700 data mahasiswa yang tersedia dalam data warehouse yang selanjutnyadigunakan sebagai pembentukan model data mining dalam penelitian ini. Informasi yang terdapat dalam data warehouse yaitu informasi tentang nim, nama, ipk dan konsentrasi yang diambil oleh mahasiswa tersebut.

2.5 Implementasi Tools SSAS

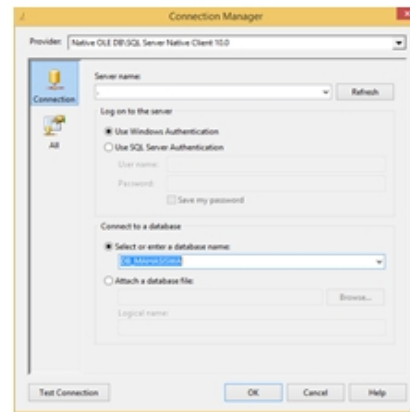
Implementasi merupakan tahapan penerapan hasil perancangan data warehouse yang telah dilakukan menggunakan SSAS (*SQL Server 2008 Analysis Service*). Hasil dari perancangan datawarehouse menggunakan tools SSIS dapat dilihat pada gambar berikut ini:

NIM	NAMA	SUM_SKS	TOTAL NILAI	IPK	THN_AKDMK	JENIS_K	NONSENTRASI
1	141001	DAHERLY ANDUSTINA	176	476	2.70	2002/2003	W
2	141002	META ANDIKA SARI	274	830	3.03	2002/2003	P
3	141003	DEDDY WILAYYA	268	814	3.04	2002/2003	P
4	141005	AFRIANTI	105	289	2.85	2002/2003	W
5	00141006P	RITA ZAHARA	99	303	3.06	2002/2003	W
6	141007	YUSDIEN	132	408	3.09	2002/2003	P
7	141008	RISNANI MASDALIPAH	122	311	2.55	2002/2003	W
8	141009	KIKI PRADIKA	105	292	2.78	2002/2003	P
9	141010	YASNI	116	274	2.36	2002/2003	P
10	141011	ACHMAD ZULKARNAIN	129	373	2.89	2002/2003	P
11	141012	RIMA APRIYANTI	134	389	2.90	2002/2003	W
12	141013	YUSDI RIZAL	117	356	2.19	2002/2003	P
13	141015	NURHAYANI	132	384	2.98	2002/2003	W
14	141016	NURHAYATI	128	367	2.79	2002/2003	W
15	141017	LUPI FEBRINTHA	133	410	3.08	2002/2003	W
16	141018	YUSHEL DAMI SARI	115	360	3.13	2002/2003	P
17	141019	IMA SRIHARTATI	125	347	2.78	2002/2003	W
18	141020	MARWATI	117	271	2.32	2002/2003	W
19	141021	EGA APRIYANI	20	56	2.75	2005/2011	W
20	00141022P	TRIANIA	134	413	3.08	2002/2003	W
21	141023	NURRIALTA	128	362	2.79	2002/2003	W
22	141024	DWI PRISTIANI	137	445	3.25	2002/2003	W
23	141026	DEDI PUTRA	110	231	2.10	2002/2003	P
24	141027	BUDI HERNO	102	203	1.99	2002/2003	P
25	141028	PETRA RAHMANHAYANI	130	444	3.29	2002/2003	W
750	7141116	APRINALDO	136	288	2.12	2011/2012	P
751	7141117	MUHAMMAD HERBONI	152	306	1.99	2011/2012	P
752	7141126	LEO CHANDRA HANISI	42	106	2.52	2007/2008	P
753	7141127	DESE SURENDRA	184	420	2.16	2011/2012	P
754	7141128	BIMO ADWOSO	184	440	2.39	2010/2011	P
755	7141133	ANDIKA PRATAMA	22	66	3.00	2007/2008	P
756	7141137	AHMAD ARDIYAN PU...	118	120	1.02	2010/2011	P
757	7141139	RIZKI ANALIA	22	6	0.27	2007/2008	P
758	7141149	FAHRI ALAN SYAH...	68	78	1.15	2008/2009	P
759	7141153	WAHYU ZULPRATAM...	22	56	2.55	2007/2008	P
760	7141160	RADANINDHA	58	248	2.52	2008/2009	P
761	7141165	JULIAN PRAMANA PU...	120	258	2.15	2008/2010	P
762	7141173	WAHYU MUSTIKA PU...	132	102	0.77	2010/2011	P
763	7141174	ARI FEBRIANZI	210	360	1.71	2011/2012	P
764	7141185	KES FAJRI	156	114	0.73	2011/2012	P
765	7141184	HENDRI PUTRA WUA...	142	202	1.59	2010/2011	P
766	7141196	ADE RINALDI	62	104	1.68	2007/2008	P
767	7141199	FAJRI HAZRIN	62	94	1.52	2007/2008	P
768	714201	GENEVA ALAM SYAH...	96	210	2.19	2008/2009	P
771	7141208	ARIES RAVO	110	288	2.62	2008/2010	P
776	7141210	IRFAN SEPTA USATA	58	120	2.07	2007/2008	P
777	7141211	JAYE EGI BUDI SAPU...	214	266	1.85	2011/2012	P
772	7141212	EDWARD SARWONO	88	130	2.24	2007/2008	P
773	7141215	MELISA	92	180	1.96	2009/2010	P
774	7141216	DENI SAFUTRA	180	342	1.90	2011/2012	P

Gambar 8. Data Warehouse DB_Mhs

1. Koneksi Sumber Data

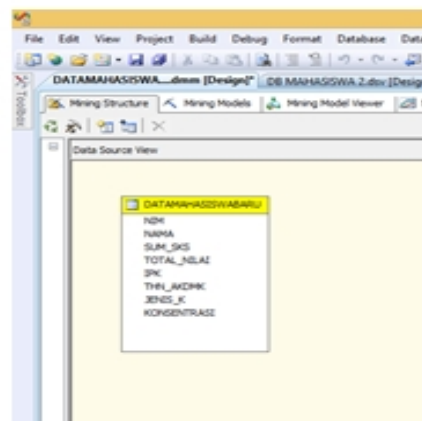
Langkah awal dalam analisis data mining menggunakan tools SSAS adalah dengan mendefinisikan sumber data yang digunakan sebagai data source untuk proses data mining. yang disimpan dalam database SQL Server 2008 dengan nama dbMahasiswa. Dalam db_Mahasiswa terdapat sebuah tabel yang diberi nama table_mhs yang berisi field-field seperti; Nim, Nama, Jumlah SKS, Jumlah Nilai, Jenis Kelamin, Angkatan, IPK dan Konsentrasi Matakuliah yang dipilih. Proses koneksi ke sumber data dalam penelitian ini dapat dilihat pada gambar 9 berikut.



Gambar 9. Koneksi Sumber Data

2. Pembentukan Datasource View

Langkah berikutnya dalam proses SSAS adalah menentukan *Data Source View* dari sumber data yang telah dikoneksikan. *Data source view* database DB_Mahasiswa dapat dilihat pada gambar 10.



Gambar 10. Data Source View Database DB_Mahasiswa

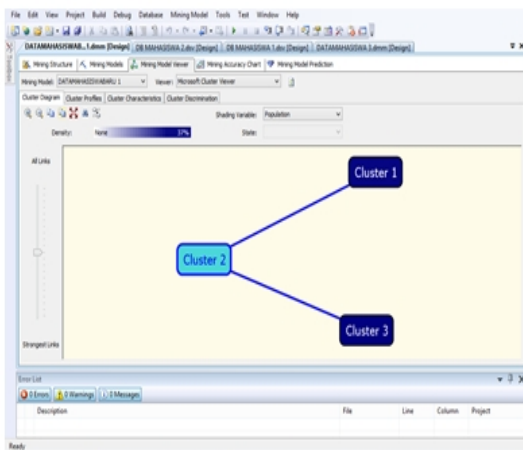
3. Data Mining Structure

Langkah terakhir dari proses SSAS dalam adalah proses analisis data mining, yang merupakan tahapan untuk melakukan analisis terhadap data akademik mahasiswa untuk menghasilkan sebuah model yang dapat digunakan untuk melakukan

prediksi minat atau ketertarikan matakuliah pilihan terhadap mahasiswa Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma.

Teknik data mining yang digunakan dalam penelitian menggunakan teknik clustering (pengelompokan). Dalam teknik clustering dilakukan proses pengelompokan sejumlah objek yang memiliki kemiripan dalam suatu kelompok yang sama.

Dari hasil proses mining menggunakan tools SSAS didapatkan hasil pengelompokan data berdasarkan minat konsentrasi matakuliah pilihan yang ada pada Program Studi Sistem Informasi menjadi tiga kelompok (cluster). Hasil pengelompokan (cluster) dapat dilihat pada gambar 11 berikut.



Gambar 11. Cluster Hasil Analisis

Berdasarkan hasil analisis terdapat tiga cluster yang terbentuk berdasarkan konsentrasi matakuliah pilihan. Cluster 1 merupakan pengelompokan minat untuk konsentrasi A. Cluster 2 merupakan kelompok minat untuk matakuliah konsentrasi B, dan cluster 3 merupakan pengelompokan untuk minat matakuliah konsentrasi C.

Predicted	A (Actual)	C (Actual)	B (Actual)
A	0	0	0
C	400	188	188
B	0	0	0

Gambar 12. Pengelompokan Data Mahasiswa Berdasarkan Konsentrasi

Dari data hasil analisis dapat dihitung jumlah mahasiswa yang mengambil konsentrasi A sebanyak 400 mahasiswa, konsentrasi B sebanyak 186 mahasiswa dan konsentrasi C sebanyak 188 mahasiswa. **KESIMPULAN**

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan:

1. Penelitian yang dilakukan telah menghasilkan sebuah model data mining pengelompokan minat matakuliah konsentrasi mahasiswa.
2. Model data mining yang dihasilkan dapat digunakan untuk memprediksi minat matakuliah konsentrasi mahasiswa program

Dan selanjutnya pada penelitian ini diharapkan data yang digunakan untuk proses analisis data mining menggunakan jumlah data mahasiswa yang cukup besar untuk menghasilkan keakuratan dalam prediksi, dan tidak saja dari program studi Sistem Informasi tapi juga dari semua program studi yang terdapat pada fakultas ilmu komputer.

Daftar Pustaka

[1] Davies, and Paul Beynon, 2004, “Database System Third Edition”, Palgrave macmillan, New York.
 [2] Han, J. And Kamber, M, 2006, “Data Mining Concept and Techniques Second Edition”, Morgan Kauffma, San Fransisco
 [3] Pramudiono, I. 2007. Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data. <http://www.ilmukomputer.org/wp-content/upload/2006/iko-datamining.zip> Diakses pada tanggal 5 September 2014.

Biodata Penulis

Marlindawati, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Manajemen Informatika Universitas Gunadarma Jakarta, lulus tahun 1997. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas Bina Darma Palembang, lulus tahun 2011. Saat ini menjadi Dosen di Universitas Bina Darma Palembang.

Andri, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi di Uiversitas BinaDarma Palembang, lulus tahun 2002. Memperoleh gelar M. Cs. Program Pasca Sarjana, Ilmu Komputer pada Universitas Gajah Mada Yogyakarta, lulus tahun 2011. Saat ini menjadi Dosen di Universitas Bina Darma Palembang.