

Analisa Profil Data Mahasiswa Baru Universitas Stikubank (UNISBANK) Semarang Tahun 2005-2010 Dengan Teknik Data Mining

Eko Nur Wahyudi, Dwi Agus Diartono, Sulastri

Abstrak - Universitas Stikubank (Unisbank) merupakan salah satu perguruan tinggi yang sudah cukup lama berkembang dengan jumlah mahasiswa baru yang diterima setiap tahun cukup banyak. Namun demikian ternyata data mengenai mahasiswa baru belum banyak dimanfaatkan untuk kepentingan yang saling berkait, diantaranya adalah mengenai objek dan wilayah tujuan promosi.

Dengan adanya teknik data mining, salah satunya adalah metode klustering dengan K-means, diharapkan data mahasiswa baru dapat diolah menjadi suatu informasi yang lebih bermanfaat dan dapat dijadikan sebagai salah satu dasar dalam pengambilan keputusan, yaitu menentukan wilayah promosi yang tepat.

Kata kunci : Data Mining, Klustering, K-means

1. PENDAHULUAN

Pada Tahun 1993 Yayasan Pendidikan dan Penerbit Mahasiswa Indonesia (YPPMI) mendirikan Akademi Manajemen Informatika dan Komputer (AMIK) STIKUBANK. Pada waktu itu berdasarkan Surat Keputusan (SK) Menteri Pendidikan dan Kebudayaan No. 92/D/O/1993 tanggal 23 Juli 1993 jurusan yang diselenggarakan adalah Manajemen Informatika Jenjang Diploma-III.

Kemudian untuk meningkatkan keberadaannya maka AMIK STIKUBANK dikembangkan menjadi Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) STIKUBANK dengan berdasarkan SK Menteri Pendidikan dan Kebudayaan No. 086/D/O/1994. Jurusan yang diselenggarakan juga bertambah menjadi :

1. Manajemen Informatika Jenjang Program Diploma-III
2. Manajemen Informatika Jenjang Program Strata-1
3. Teknik Informatika Jenjang Program Strata-1

Selanjutnya sesuai dengan kebutuhan masyarakat akan pendidikan tingkat tinggi dengan berbagai disiplin ilmu maka sekali lagi STMIK STIKUBANK dikembangkan menjadi Universitas STIKUBANK (UNISBANK), yang merupakan penggabungan dari beberapa Sekolah Tinggi yang telah ada sebelumnya di bawah YPPMI, yaitu STMIK STIKUBANK, Sekolah Tinggi Bahasa Asing (STIBA) STIKUBANK, dan Sekolah Tinggi Ilmu Hukum (STIH) STIKUBANK.

Dengan bergabungnya ketiga Sekolah Tinggi tersebut menjadi Universitas STIKUBANK dengan SK Menteri Pendidikan Nasional Nomor 53/D/O/2001 tertanggal 5 Juli 2001 maka struktur penyelenggaraan pendidikan berubah menjadi 5 Fakultas dan 13 Program Studi. Dalam penggabungan tersebut kelompok program studi yang sebelumnya berada di bawah STMIK STIKUBANK kemudian berada di bawah Fakultas Teknologi Informasi.

Pada tanggal 1 Oktober 2003 UNISBANK Rektor UNISBANK bersama dengan Yayasan (YPPMI) menetapkan penggabungan AKPARTA STIKUBANK ke dalam UNISBANK menjadi Program Diploma Kepariwisata UNISBANK yang terdiri dari Program Studi Diploma-3 Perhotelan dan Usaha Perjalanan Wisata (UPW). Penggabungan ini ditetapkan dengan Surat Keputusan Yayasan Pendidikan dan Penerbit Mahasiswa Indonesia (YPPMI) Nomor : 130/SK/YPPMI/IX/2003.

Dengan penggabungan ini, maka UNISBANK mengelola 5 (lima) Fakultas dan sebuah program diploma kepariwisataan dengan 16 (enam belas) Program Studi seperti terlihat pada tabel 1.

Tabel 1 Daftar Fakultas dan Program Studi

No	Fakultas	Program Studi
I.	Teknologi Informasi	1. Sistem Informasi (S-1) 2. Teknik Informatika (S-1) 3. Sistem Komputer (S-1) 4. Man. Informatika (D-III) 5. Komp. Akuntansi (D-III) 6. Teknik Komputer (D-III)
II.	Bahasa dan Ilmu Budaya	7. Sastra Inggris (S-1) 8. Bahasa Inggris (D-III)
III.	Hukum	9. Ilmu Hukum (S-1)
IV.	Teknik	10. Teknik Industri (S-1) 11. Teknik Elektronika (D-III)
V.	Ekonomi	12. Akuntansi (S-1) 13. Manajemen (S-1) 14. Manajemen Industri (D-III)
VI.	Program Diploma Kepariwisata	15. Perhotelan (D-III) 16. Usaha Perjalanan Wisata (D-III)

Proses pengembangan terakhir adalah penggabungan STIE STIKUBANK ke dalam Fakultas Ekonomi UNISBANK dengan adanya SK Mendiknas tanggal 23 Agustus 2007 dengan No. SK. 160/D/O/2007.

Selama tahun 2005 sampai dengan tahun 2010 jumlah mahasiswa baru yang diterima oleh masing-masing program studi cukup bervariasi, yaitu ada yang banyak dan ada yang sedikit, namun hampir semua program studi mengalami kenaikan dan penurunan. Jumlah kenaikan

tentunya membawa kebaikan namun jumlah penurunan ternyata membawa dampak dengan adanya beberapa program studi yang akhirnya ditutup karena tidak dapat memenuhi kuota penerimaan mahasiswa baru dalam beberapa tahun.

Beberapa faktor bisa menjadi penyebab adanya penurunan jumlah penerimaan mahasiswa baru tersebut, salah satu di antaranya adalah kurang tepatnya tindakan dalam melakukan kegiatan promosi. Dinilai kurang tepat karena acuan kegiatan promosi setiap tahun hanya berdasarkan pada penerimaan mahasiswa baru tahun sebelumnya, bukan dalam kurun waktu yang cukup lama. Untuk itu maka perlu adanya suatu penelitian mengenai data mahasiswa baru yang hasilnya nanti dapat digunakan sebagai salah satu dasar pengambilan keputusan dalam melakukan kegiatan promosi.

Tabel 2 menunjukkan jumlah mahasiswa baru yang diterima oleh masing-masing program studi selama kurun waktu tahun 2005 sampai dengan tahun 2010.

Tabel 2. Jumlah mahasiswa baru UNISBANK tahun 2005 - 2010

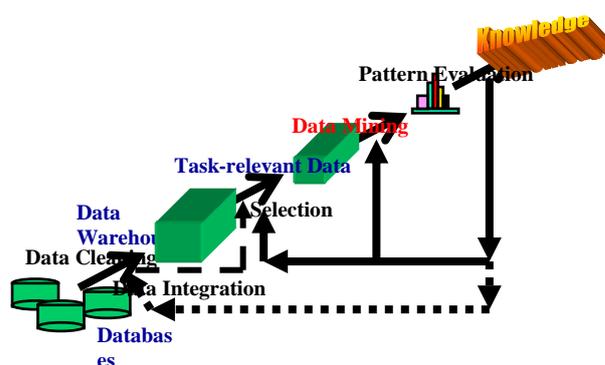
No.	Mahasiswa Baru	Jumlah
1	Akuntansi S-1	812
2	Bahasa Inggris D-3	83
3	Ilmu Hukum S-1	220
4	Keuangan dan Perbankan D-3	55
5	Komputerisasi Akuntansi D-3	155
6	Manajemen S-1	671
7	Manajemen Industri D-3	1
8	Manajemen Informatika D-3	249
9	Perhotelan D-3	134
10	Sastra Inggris S-1	347
11	Sistem Informasi S-1	917
12	Teknik Elektronika D-3	10
13	Teknik Industri S-1	82
14	Teknik Informatika S-1	1.051
15	Teknik Komputer D-3	81
16	Usaha Perjalanan Wisata D-3	3
	Jumlah	4.871

Dari jumlah mahasiswa baru yang diterima seperti tercantum dalam tabel 2 belum diketahui secara pasti dari mana mahasiswa tersebut berasal. Untuk itu diperlukan suatu penelitian guna mengetahui asal mahasiswa baru tersebut dengan metode data mining agar dapat diperoleh profil mahasiswa baru yang tepat serta dapat dilakukan analisa yang lebih akurat sebagai bahan pengambilan keputusan dalam melakukan kegiatan promosi.

2. Data Mining

Data mining merupakan suatu metode menemukan suatu pengetahuan dalam suatu database yang cukup besar. Data mining adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru, sangat bermanfaat dan akhirnya dapat dimengerti suatu corak atau pola dalam data tersebut (Han & Kamber, 2006).

Data mining adalah bagian integral dari penemuan pengetahuan dalam database (KDD), yang merupakan proses keseluruhan mengubah data mentah menjadi informasi yang bermanfaat, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Proses penemuan pengetahuan dalam database / KDD (Han & Kamber, 2006)

Proses KDD ini terdiri dari serangkaian langkah-langkah transformasi, dari proses data *preprocessing* dan proses data *postprocessing* dari data yang merupakan hasil penggalian. Input data dapat disimpan dalam berbagai format (flat file, spreadsheet, atau relasional tabel) dan mungkin berada dalam penyimpanan data terpusat atau didistribusikan di beberapa alamat. Tujuan dari proses data *preprocessing* adalah untuk mengubah data input mentah menjadi format yang sesuai untuk analisis selanjutnya. Langkah-langkah yang dilakukan antara lain dengan memperbaiki data yang kotor atau ganda, dan memilih catatan dan fitur yang relevan dengan proses pengelolaan data selanjutnya. Karena banyak cara data dapat dikumpulkan dan disimpan, maka proses pengolahan data mungkin akan melelahkan dan memakan waktu yang lama dalam keseluruhan proses penemuan pengetahuan (Tan, dkk, 2006).

Terdapat empat tugas utama data mining seperti terlihat pada gambar 2, yaitu :

1. Predictive Modelling

Predictive modelling digunakan untuk membangun sebuah model untuk target variable sebagai fungsi dari explanatory variable.

Explanatory variable dalam hal ini merupakan semua atribut yang digunakan untuk melakukan prediksi, sedangkan variabel target merupakan atribut yang akan diprediksi nilainya. Predictive modelling dibagi menjadi dua tipe yaitu : Classification digunakan untuk memprediksi nilai dari target variable yang discrete (diskrit) dan regression digunakan untuk memprediksi nilai dari target variable yang continue (berkelanjutan).

2. Association Analysis

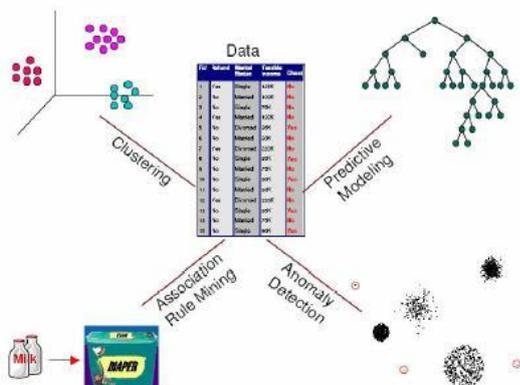
Association analysis digunakan untuk menemukan aturan asosiasi yang memperlihatkan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

3. Cluster Analysis

Tidak seperti klasifikasi yang menganalisa kelas data obyek yang mengandung label. Clustering menganalisa objek data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data training. Karena belum diketahui sebelumnya. Clustering merupakan proses pengelompokan sekumpulan objek yang sangat mirip.

4. Anomaly Detection

Anomaly detection merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. Anomaly dapat di deteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.



Gambar 2. Tugas utama data mining (Tan, dkk, 2006)

2.1. Jenis Data

Sebuah kumpulan data sering kali dapat dikatakan sebagai kumpulan objek data. Nama lain untuk objek data adalah catatan, titik, vektor, pola, acara, kasus, contoh, pengamatan, atau entitas. Selanjutnya objek data digambarkan oleh sejumlah atribut yang memiliki karakteristik dasar suatu objek, misalnya bentuk sebuah objek secara fisik atau waktu yang menunjukkan di mana sebuah kegiatan terjadi. Nama lain untuk sebuah atribut adalah variabel, karakteristik, bidang, fitur, atau dimensi (Tan, dkk, 2006).

2.2. Atribut dan Skala Pengukuran

Sebuah atribut adalah properti atau karakteristik dari sebuah objek yang mungkin berbeda-beda. Sebagai contoh, warna mata berbeda antara orang yang satu dengan orang yang lain, atau contoh lain adalah suhu suatu benda yang bervariasi dari waktu ke waktu. Warna mata adalah atribut simbolis dengan sejumlah kecil kemungkinan nilai (coklat, hitam, biru, hijau, dll), sedangkan suhu adalah atribut numerik yang berpotensi dengan nilai-nilai dalam jumlah tidak terbatas.

Sebuah skala pengukuran adalah aturan (fungsi) yang mengaitkan nilai numerik atau simbolis dengan atribut dari sebuah objek. Secara formal, proses pengukuran adalah aplikasi dari suatu skala pengukuran yang mengasosiasikan sebuah nilai dengan atribut tertentu dari suatu objek tertentu (Tan, dkk, 2006).

2.3. Visualisasi

Visualisasi data adalah tampilan informasi dalam format grafik atau tabel. Tujuan visualisasi adalah representasi dari informasi yang disampaikan kepada pihak-pihak yang melihat agar mudah memahami informasi yang disampaikan tersebut (Tan, dkk, 2006).

2.4. Analisa Kluster dan K-means

Analisa kluster adalah kelompok yang berdasar pada objek data hanya pada informasi yang ditemukan dalam data yang menggambarkan objek dan hubungannya. Tujuannya adalah bahwa objek di dalam suatu kelompok yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak terkait dengan) objek di kelompok lain. Yang lebih besar kesamaan (atau homogenitas) dalam suatu kelompok dan lebih besar perbedaan antara kelompok lainnya, yang lebih baik atau lebih berbeda dengan kelompoknya (Tan, dkk, 2006).

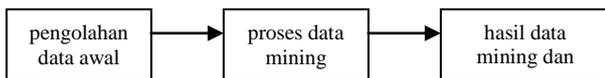
Dasar dari K-means ini adalah prototipe, yaitu bagian dari teknik klustering yang mencoba menemukan jumlah kluster (K), yang diwakili oleh pusat kelompok data (Tan, dkk, 2006).

Prototipe berbasis teknik *clustering* ini membuat satu tingkat pemisahan objek data. Ada beberapa teknik yang dapat digunakan, tapi dua teknik yang paling menonjol adalah *K-means* dan *K-medoid*. K-means mendefinisikan sebuah prototipe dalam suatu pusat kelompok data, yang biasanya merupakan titik-tengah dari sekelompok titik data, dan biasanya diterapkan pada objek yang merupakan bentuk n-dimensi (Tan, dkk, 2006).

Dasar algoritma K-means adalah sebagai berikut :

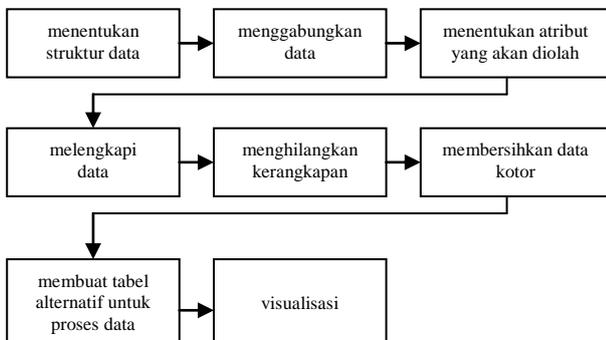
1. Tentukan K sebagai titik tengah awal dari sekumpulan objek
2. Kemudian hitung jarak masing-masing objek dengan titik K tersebut
3. Bandingkan nilai jarak rata-rata masing-masing objek dengan titik K
4. Jika terdapat perbedaan maka titik K akan mengalami pergeseran letak
5. Ulangi langkah kedua sampai dengan keempat sehingga titik K tidak bergeser lagi dan diperoleh sekumpulan objek yang saling berdekatan dengan jarak pemisah paling pendek (Tan, dkk, 2006).

Penelitian diawali dengan mempersiapkan data melalui beberapa tahapan kegiatan, kemudian setelah data tersusun dilakukan proses pengolahan data menggunakan teknik data mining dan tahap terakhir adalah hasil dari data mining tersebut dianalisa dan disimpulkan, seperti terlihat pada gambar 3.1.

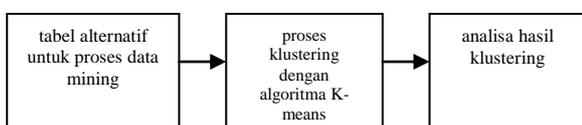


Gambar 3. Kegiatan utama penelitian

Pengolahan data awal merupakan bagian dari persiapan data di mana langkah-langkah yang dilakukan antara lain meliputi menentukan struktur data, menggabungkan data, menentukan atribut yang akan diolah, melengkapi data, menghilangkan kerangkapan data dan membersihkan data kotor. Hasil proses persiapan data disajikan dalam bentuk tabel dan grafik. Pengolahan data dilakukan dengan menggunakan rumus-rumus yang ada pada aplikasi Microsoft Excel 2003 dan Macro Visual Basic yang telah diintegrasikan, seperti terlihat pada gambar 4.



Gambar 3. Proses pengolahan data awal dan visualisasi



Gambar 4. Proses data mining dan analisa hasil

Langkah berikutnya setelah membentuk tabel alternatif untuk proses data mining adalah melakukan proses klustering dan klasifikasi. Langkah ini merupakan bagian dari pengolahan data dimana dilakukan proses pengolahan data menggunakan algoritma K-means untuk teknik klustering dan algoritma Hunt untuk klasifikasi, seperti terlihat pada gambar 3. Proses data mining dengan teknik klustering dijelaskan lebih lanjut pada bab 4.

3. Sumber Data

Sumber data yang akan diolah dan divisualisasikan berasal dari data mahasiswa baru UNISBANK Semarang mulai tahun angkatan 2005 sampai dengan tahun angkatan 2010. Data mahasiswa baru tersebut diperoleh dari data yang tersimpan dalam database sistem informasi penerimaan mahasiswa baru dengan cara melakukan proses *query* dan disimpan kembali dalam bentuk tabel 2 (dua) dimensi menggunakan Microsoft Excel 2003.

4. Pengolahan Data

Pertama-tama data mahasiswa baru diurutkan berdasarkan Tahun Angkatan, Fakultas, Program Studi dan NIM. Dengan demikian diperoleh hasil tabel mahasiswa baru seperti terlihat pada gambar 5.

A	B	C	D	E	G	H	I	J	K	L	M				
1	No	Tahun	Mesuk	Fakultas	Prodi	Jenjang	NIM	Nama	TmpLahir	TglLahir	Kota	Alamat	JK	Kelas	AsaBek
2	1	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0177	ANWAR	BANDUNG	07/03/1983	BANDUNG	01	1	SIMN	MP	
3	2	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0100	DWILUSTYARNI	BATANG	26/01/1983	BATANG	1	1	SIMN	PG	
4	3	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0114	AGUSTINAWINGSIH	BATANG	22/10/1983	BATANG	1	1	SIMN	PG	
5	4	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0115	HANA MARJUNA	BATANG	18/03/1984	BATANG	1	1	SIMN	I	
6	5	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0084	NORA ROSIANA	SUBAH	30/05/1982	SUBAH	0	0	SIMN	I	
7	6	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0043	WARLINA RILYA	BLORA	18/03/1982	BLORA	1	1	SIMN	KA	
8	7	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0004	WINRYANTI UTAMI	BLORA	11/08/1979	BLORA	1	1	SIMN	I	
9	8	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0082	WUJI WIYASTUTI	BOYOLALI	01/08/1982	BOYOLALI	1	1	SIMN	BH	
10	9	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0007	ZULYAN	BOYOLALI	18/07/1980	BOYOLALI	1	1	SIMN	BH	
11	10	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0130	LITA SUGIANTI	JAKARTA	04/06/1981	BOYOLALI	1	1	SIMN	I	
12	11	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0133	SITI ZUMMERCH	BREBES	08/03/1982	BREBES	1	1	QMAN	I	

Gambar 5. Tampilan tabel awal mahasiswa baru

4.1. Menentukan struktur data

Data mahasiswa baru yang di *query* dan disimpan dalam tabel Microsoft Excel 2003 memiliki struktur awal sebagai berikut :

- Nomor Tes (numerik)
- NIM (karakter)
- Nama (karakter)
- Alamat (karakter)
- Kota (karakter)
- Tempat Lahir (karakter)
- Tanggal Lahir (tanggal)

- Jenis Kelamin (numerik)
- Kelas (numerik)
- Program (karakter)
- Sekolah Asal (karakter)
- Alamat Sekolah Asal (karakter)
- Kota Sekolah Asal (karakter)

Berdasarkan struktur data tersebut dibuat tambahan kolom atau atribut untuk melengkapi kebutuhan data yang akan diolah. Proses menambah kolom atau atribut dan menentukan nilai datanya dilakukan secara manual menggunakan rumus yang telah disediakan oleh Microsoft Excel 2003. Tambahan kolom atau atribut tersebut antara lain adalah sebagai berikut :

- TahunAngkatan : diperoleh dari 2 (dua) digit pertama NIM
- Fakultas : diperoleh dari digit ketiga dan keempat NIM
- ProgramStudi : diperoleh dari digit kelima NIM
- Jenjang : diperoleh dari digit keenam NIM
- Usia : diperoleh dari tanggal masuk dikurangi TanggalLahir

4.2. Menggabungkan data

Data mahasiswa baru tersimpan dalam beberapa folder yang masing-masing berisikan data mahasiswa baru untuk setiap angkatan. Menggabungkan data dari beberapa file di dalam folder yang berbeda harus dilakukan dengan teliti karena beberapa nama dan jumlah atribut tidak sama. Penggabungan file dilakukan dengan menyimpan data terlebih dahulu ke dalam masing-masing sheet untuk setiap angkatan mahasiswa baru. Setelah itu dibuka *sheet* baru yang akan menampung semua sheet dari tahun angkatan 2005 sampai dengan angkatan 2010. Proses ini dilakukan secara manual setelah dipilih atribut-atribut yang sama dan dapat digabungkan. Hasil dari proses pencarian dan penggabungan adalah seperti terlihat pada gambar 6 dimana masing-masing data mahasiswa baru masih tersimpan di dalam *sheet* untuk tiap tahun angkatan, sedangkan untuk hasil penggabungan semua angkatan disimpan dalam *sheet* dengan nama gabung.

54	DINA INDRU NURHARIYANTI	SEMARANG	17-01-1984	P	SMU 10 SMG
55	ISTIYADI	SEMARANG	10-06-1984	L	SMU MUHAMMADIYAH 01
56	GANUNG HASTANANDARU	SEMARANG	16-05-1977	L	MAN 1 SMG
57	TAUFAN MOCHAMMAD HESA	PATI	17-04-1982	L	SMU 2PATI
58	TRIYANTO	SEMARANG	30-01-1983	L	SMUN 16 SMG
59	YOLANDA MULYANINGRUM	SEMARANG	27-11-1984	P	SMU INUSAPUTRA
60	INDRA SUSILOWATI	SEMARANG	09-02-1983	P	SMU MASEHI 1
61	DYAH AYU TRI KUMALASARI	GROBOGAN	07-10-1984	P	SMU 1 GODONG
62	MOH YUNUS	SEMARANG	04-10-1970	L	SMU MARDI SISWO

Gambar 6. Tampilan proses penggabungan data

4.3. Memilih atribut yang akan diolah

Hasil penggabungan data mahasiswa baru dengan beberapa format struktur data yang tidak sama perlu disesuaikan agar susunan data menjadi lebih baik. Berdasarkan tabel mahasiswa baru yang diperoleh dari hasil query sebelumnya, langkah berikutnya adalah memilih atribut-atribut yang mempunyai kemungkinan akan berpengaruh dalam pengambilan keputusan terhadap kegiatan promosi setelah mengalami proses data mining, seperti terlihat pada gambar 7 beberapa atribut yang telah dipilih disertai dengan hasil proses dari NIM yaitu ThnMasuk, Fakultas, ProgDi dan Jenjang.

	A	B	C	D	E	G	H	I
1	No	ThnMasuk	Fakultas	ProgDi	Jenjang	NIM	Nama	TmpLahir
2	1	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0177	ANWAR	BANDUNG
3	2	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0100	DWILISTIYARNI	BATANG
4	3	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0114	AGUSTINAININGSIH	BATANG
5	4	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0115	HANA MARIANA	BATANG
6	5	2001	Teknologi Informasi	Komputerisasi Akuntansi	D-3	01.01.32.0094	NORA ROSIANA	SUBAH

Gambar 7. Tampilan hasil proses pemilihan atribut

Tabel awal perlu diurutkan berdasarkan ThnMasuk, Fakultas, ProgDi dan NIM untuk mempermudah proses persiapan data sebelum diolah lebih lanjut menggunakan teknik data mining.

4.4. Melengkapi Data

Setelah data digabung dan dipilih atribut yang akan diolah, langkah selanjutnya adalah melengkapi data-data yang masih kosong atau kurang. Data yang masih kosong atau kurang ini karena beberapa hal, diantaranya karena belum diisi secara lengkap pada saat memasukkan data, atau karena sumber data yang akan dimasukkan tidak ada. Proses mencari data yang kosong atau kurang dilakukan secara manual pada setiap record secara urut dari data awal hingga akhir. Melengkapi data memerlukan waktu yang cukup lama karena harus mencari dokumen mahasiswa baru tersebut di bagian arsip, bahkan beberapa dokumen diantaranya sudah digudangkan, kemudian dimasukkan ke bagian tabel yang masih

kosong atau kurang lengkap. Data yang harus dilengkapi tentu saja merupakan data dari atribut yang sudah dipilih untuk diolah. Gambar 8 berikut ini mewakili tampilan data yang harus dilengkapi dimana masih terdapat sel yang kosong belum terisi data.

J	L	M	N	O	Q
KOTA	KODE KULIAH	TEMPAT LAHIR	TGL LAHIR	SMU	KOTA SMU
SEMARANG					
SEMARANG		SEMARANG	07/04/1986 0:00	SMAN 2	SEMARANG
SEMARANG		SEMARANG	14/03/1987 0:00	SMAN 2 MARANGEN	DEMAK
SEMARANG		SEMARANG	22/08/1988 0:00	SMK PANGUDI LUHUR TAROSUS	SEMARANG
SEMARANG		SEMARANG	05/11/1981 0:00	SMAN 4	SEMARANG
SEMARANG		SEMARANG	29/03/1987 0:00	KESATRIAN 1	SEMARANG
SEMARANG		DIURI	04/01/1986 0:00	SMA NEGERI 2	TANJUNGPINANG

Gambar 8. Contoh data yang tidak lengkap

4.5. Menghilangkan kerangkapan data

Setelah semua data yang kosong atau kurang dilengkapi dan semua sel data telah terisi maka langkah berikutnya adalah melakukan pencarian terhadap kemungkinan terdapat data yang sama atau rangkap dan menghapus salah satu dari data yang rangkap tersebut. Hal ini dilakukan agar setiap sel data hanya memiliki nilai data yang tunggal. Proses pencarian data yang rangkap ini dilakukan secara manual dengan melihat isi sel dari setiap record. Kerangkapan data misalnya terdapat pada data nama asal sekolah dan kota asal sekolah, seperti terlihat pada gambar 9 dimana kolom nama asal sekolah dan kota asal sekolah sama-sama mencantumkan kota asal sekolah yaitu "AMBARAWA", yang semestinya nama asal sekolah cukup dituliskan dengan "SMA ISLAM SUDIRMAN" saja.

SEMARANG	20/03/1987 0:00	SMU KARTINI SEMARANG	SEMARANG
BREBES	02/05/1986 0:00	SMK BAHARI	TEGAL
PALANGKARAYA	30/08/1987 0:00	SMAN 1 KURUN	PALANGKARAYA
SEMARANG	27/08/1988 0:00	SMA INSTITUT INDONESIA	SEMARANG
SEMARANG	25/01/1986 0:00	SMA ISLAM SUDIRMAN, AMBARAWA	AMBARAWA
SEMARANG	20/05/1988 0:00	SMA ISLAM SUDIRMAN, AMBARAWA	AMBARAWA
SEMARANG	05/05/1988 0:00	SMA ISLAM SUDIRMAN, AMBARAWA	AMBARAWA
BANYUWANGI	15/04/1982 0:00	SMK TELKOM MALANG	MALANG

Gambar 9 Tampilan contoh data yang rangkap

4.6. Membersihkan data kotor

Data kotor yang dimaksud adalah data yang masih tidak sesuai dengan nilai data yang sebenarnya ada pada dokumen. Data kotor ini antara lain berupa data yang salah, yaitu tidak sesuai dengan dokumen aslinya, atau data yang belum lengkap sehingga masih perlu disesuaikan dengan dokumen aslinya. Data tersebut dapat

ditunjukkan antara lain seperti pada gambar 10 dimana penulisan nama sekolah yang berbeda padahal yang dimaksud adalah sekolah yang sama yaitu antara "SMA NEGERI 1 TUNJUNGAN" dan "SMAN 1 TUNJUNGAN", maka salah satu nama sekolah perlu disesuaikan.

FODEWO V	AMBARAWA	SMA VIRGO FIDELIS	BAWEN
INGAH IV N	SEMARANG	SMA MUHAMMADIYAH 1	SEMARANG
IO. 263	SEMARANG	SMA KSATRIAN 1	SEMARANG
IO. 409	SEMARANG	SMA N 14	SEMARANG
YA No.44	SEMARANG	SMA KESATRIAN 2	SEMARANG
01/RW.02	SEMARANG	SMA WALISONGO	SEMARANG
	BLORA	SMA NEGERI 1 TUNJUNGAN	BLORA
UK RT 05	SEMARANG	SMA GITA BAHARI	SEMARANG
OYOSO	PATI	SMA NASIONAL	PATI
MBUL, DK.	BLORA	SMAN 1 TUNJUNGAN	BLORA
AT RT 01	SEMARANG	IPI KARANG PANAS	SEMARANG
RA IX	SEMARANG	SMA MASEHI 3 PSAK	SEMARANG
O. 21	SEMARANG	SMA NEGERI 6	SEMARANG

Gambar 10. Contoh data yang kotor

4.7. Membuat Tabel Alternatif untuk Proses Data Mining

Langkah berikutnya adalah menentukan tabel-tabel secara parsial yang diturunkan dari tabel data utama yang telah bersih sebagai alternatif kemungkinan proses data mining. Tabel-tabel ini dimungkinkan digunakan untuk proses data mining, meskipun tidak dilakukan pada semua tabel. Tentu saja meskipun tidak semua tabel dipilih untuk dilakukan proses data mining tetapi cukup memberikan informasi yang dapat dijadikan gambaran secara umum dari data mahasiswa baru.

4.8. Penerapan Teknik Klustering Pada Data Mahasiswa Baru

Teknik klustering dalam pengolahan data mahasiswa baru ini bertujuan untuk menentukan kelompok-kelompok data mahasiswa baru secara parsial untuk mengetahui potensi dari mana asal mahasiswa baru tersebut berasal, diantaranya adalah kota asal sekolah dan nama asal sekolah. Sedangkan teknik *decision tree* (pohon keputusan) yang akan diterapkan bertujuan untuk menentukan kecenderungan asal mahasiswa baru tersebut berasal terhadap program studi yang dipilih, yaitu seperti jenis sekolah, kota dan status sekolah.

4.9. Data Mining Mahasiswa Baru dengan Teknik Klustering

Teknik klustering yang akan digunakan dalam pengolahan data mahasiswa baru ini adalah K-means. K-means dipilih karena merupakan salah satu teknik yang paling populer sekaligus paling mudah diimplementasikan untuk proses klustering pada suatu kelompok objek namun

sudah dapat menunjukkan hasil yang cukup optimal. Dengan teknik tersebut data mahasiswa baru akan dikelompokkan berdasarkan urutan jarak terdekat dari masing-masing kelompok data. Berbeda dengan model pengelompokan secara manual, pada proses klustering pengelompokan data dihitung berdasarkan suatu rumus jarak sehingga jelas dan tegas proses pemisahan antar kelompok tersebut, sedangkan pada pengelompokan secara manual tidak jelas batas atau jarak antar kelompok yang dibentuk tersebut.

Langkah yang dilakukan dalam proses klustering dengan metode K-means adalah sebagai berikut :

1. Tentukan sembarang data awal sebagai titik pusat sementara
2. Membandingkan semua data dengan data awal yang telah ditentukan sebagai titik pusat sementara untuk dihitung masing-masing jaraknya
3. Setelah diketahui nilai dari masing-masing jarak kemudian dihitung nilai rata-ratanya
4. Hasil nilai rata-rata kemudian dibandingkan dengan jarak terpendek dari titik pusat sementara
5. Titik pusat sementara kemudian bergeser sesuai dengan nilai selisih dari perbandingan tersebut
6. Langkah kedua sampai dengan kelima diulang sehingga tidak diperoleh pergeseran titik pusat lagi

4.10. Menentukan kluster program studi berdasar jumlah mahasiswa baru

Mula-mula data akan dikluster berdasarkan jumlah mahasiswa baru untuk masing-masing program studi. Tujuan dari klustering ini adalah untuk menentukan kelompok program studi berdasarkan jumlah mahasiswa baru yang diterima.

Metode yang akan dipergunakan dalam klustering ini adalah K-means, dimana proses klustering didasarkan atas jarak terdekat dari suatu pusat yang telah ditentukan sebelumnya secara acak. Nilai dari pusat dapat berubah jika nilai perhitungan jarak antara pusat dengan data yang ada didekatnya juga mengalami perubahan, sehingga akhirnya nilai pusat tidak mengalami perubahan dan terbentuklah kluster yang tetap.

Penentuan jumlah kluster yang akan dibentuk disesuaikan dengan kebutuhan atau tujuan dari hasil klustering. Empat buah kluster yang akan dibentuk berikut ini hanya merupakan suatu contoh untuk tujuan mengambil tiga buah kluster dari empat buah kluster yang terbentuk sebagai kluster yang paling dominan dalam hal jumlah penerimaan mahasiswa baru. Sedangkan

kluster terakhir yaitu kluster keempat merupakan hasil kluster yang perlu untuk mendapatkan perhatian agar jumlah penerimaan mahasiswa baru lebih ditingkatkan.

Dari tabel rekapitulasi program studi diambil empat buah data secara acak yaitu Sistem Informasi, Teknik Komputer, Teknik Industri dan Perhotelan untuk menentukan jarak awal dari empat buah kluster yang akan dibentuk, yaitu seperti terlihat pada tabel 3.

Tabel 3. Empat buah data awal secara acak sebagai pusat kluster sementara

Tabel Program Studi

No.	Program Studi	Jumlah
1	Akuntansi	812
2	Bahasa Inggris	83
3	Ilmu Hukum	220
4	Keuangan dan Perbankan	55
5	Komputerisasi Akuntansi	155
6	Manajemen	671
7	Manajemen Industri	1
8	Manajemen Informatika	249
9	Perhotelan	134
10	Sastra Inggris	347
11	Sistem Informasi	917
12	Teknik Elektronika	10
13	Teknik Industri	82
14	Teknik Informatika	1051
15	Teknik Komputer	81
16	Usaha Perjalanan Wisata	3
	Jumlah	4871

Jarak1 adalah jarak nilai data antara masing-masing data dengan data yang dipilih pertama yaitu Akuntansi. Jarak2 adalah jarak nilai data antara masing-masing data dengan data yang dipilih kedua yaitu Komputerisasi Akuntansi. Untuk Jarak3 dan Jarak4 berlaku ketentuan yang sama, yaitu membandingkan masing-masing data dengan data yang dipilih ketiga dan keempat yaitu Manajemen Informatika dan Teknik Industri. Hasilnya seperti terlihat pada Tabel 4.

Tabel 4. Hasil tahap 1 proses penentuan jarak

Klust1	Klust2	Klust3	Klust4
812	-	-	-
-	-	-	83
-	-	220	-
-	-	-	55

-	155	-	-
671	-	-	-
-	-	-	1
-	-	249	-
-	134	-	-
-	-	347	-
917	-	-	-
-	-	-	10
-	-	-	82
1051	-	-	-
-	-	-	81
-	-	-	3
862,75	144,50	272,00	45,00

Setelah itu data jumlah dialokasikan pada posisi nilai terkecil antara jarak1 sampai dengan jarak4, sehingga hasilnya nampak seperti pada tabel 5.

Tabel 5. Hasil akhir proses penentuan kluster

Klust1	Klust2	Klust3	Klust4
812	-	-	-
-	-	-	83
-	-	220	-
-	-	-	55
-	155	-	-
671	-	-	-
-	-	-	1
-	-	249	-
-	134	-	-
-	-	347	-
917	-	-	-
-	-	-	10
-	-	-	82
1051	-	-	-
-	-	-	81
-	-	-	3
862,75	144,50	272,00	45,00

Dari proses tahap 1 dan tahap 2 terlihat bahwa nilai rata-rata kluster sudah tidak mengalami perubahan, artinya hanya dengan tiga langkah proses klustering telah diperoleh kluster yang optimal.

Selanjutnya dengan mencoba memilih kembali empat set data lain, yaitu Ilmu Hukum, Manajemen, Perhotelan dan Teknik Komputer, untuk mendapatkan empat buah kluster dengan proses yang sama, yaitu seperti terlihat pada tabel 6. Tujuan dari proses klustering yang kedua ini adalah membandingkan dengan hasil klustering pada proses yang pertama. Hasil proses klustering yang kedua adalah seperti terlihat pada tabel 6.

Tabel 6. Menentukan empat data awal tahap kedua secara acak sebagai pusat kluster sementara

No.	Program Studi	Jumlah
1	Akuntansi	812
2	Bahasa Inggris	83
3	Ilmu Hukum	220
4	Keuangan dan Perbankan	55
5	Komputerisasi Akuntansi	155
6	Manajemen	671
7	Manajemen Industri	1
8	Manajemen Informatika	249
9	Perhotelan	134
10	Sastra Inggris	347
11	Sistem Informasi	917
12	Teknik Elektronika	10
13	Teknik Industri	82
14	Teknik Informatika	1051
15	Teknik Komputer	81
16	Usaha Perjalanan Wisata	3
	Jumlah	4871

Setelah melalui dua langkah proses penentuan jarak dan kluster maka diperoleh hasil seperti pada tabel 7.

Tabel 7. Hasil proses klustering pada empat data tahap kedua

Klust1	Klust2	Klust3	Klust4
-	812	-	-
-	-	-	83
220	-	-	-
-	-	-	55
-	-	155	-
-	671	-	-
-	-	-	1
249	-	-	-
-	-	134	-
347	-	-	-
-	917	-	-
-	-	-	10
-	-	-	82
-	1051	-	-
-	-	-	81
-	-	-	3
272,00	862,75	144,50	45,00

Dengan demikian dapat disimpulkan bahwa setelah melalui dua proses klustering dengan menggunakan empat set data yang berbeda secara acak, masing-masing memiliki jumlah tahapan yang berbeda pula yaitu 3 tahap dan 4 tahap, hasilnya adalah sebagai berikut :

Kesimpulan 1 :

Kluster Pertama	
1 Teknik Informatika	1051
2 Sistem Informasi	917
3 Akuntansi	812
4 Manajemen	671
Kluster Kedua	
1 Sastra Inggris	347
2 Manajemen Informatika	249
3 Ilmu Hukum	220
Kluster Ketiga	
1 Komputerisasi Akuntansi	155
2 Perhotelan	134
Kluster Keempat	
1 Bahasa Inggris	83
2 Teknik Industri	82
3 Teknik Komputer	81
4 Keuangan dan Perbankan	55
5 Teknik Elektronika	10
6 Usaha Perjalanan Wisata	3
7 Manajemen Industri	1

Kluster Pertama	
1 Semarang	2817
Kluster Kedua	
1 Kendal	245
2 Pati	160
3 Demak	139
4 Jepara	116
5 Kudus	103
6 Purwodadi	90
7 Tegal	88
8 Ungaran	80
9 Rembang	70
Kluster Ketiga	
1 Grobogan	59
2 Pemalang	57
3 Pekalongan	46
...

4.11. Menentukan kluster kota asal sekolah berdasarkan jumlah mahasiswa baru

Seperti halnya proses pengklusteran sebelumnya, proses kluster kali ini ditujukan untuk mengetahui kelompok dari kota-kota asal sekolah mahasiswa baru yang paling mendominasi jumlah penerimaan mahasiswa baru, yaitu seperti terlihat pada tabel 8. Sekali lagi bahwa pemilihan jumlah kluster yang akan dibentuk adalah bebas sesuai dengan kebutuhan maupun keinginan.

Tabel 8. Empat buah data awal sebagai pusat kluster sementara pada proses kluster tahap pertama

KotaAsalSekolah	Jumlah
ACEH TENGGARA	1
ACEH UTARA	1
AMBARAWA	44
AMBON	1
...	
BLITAR	2
BLORA	40
BOGOR	4
...	...
WONOGIRI	20
WONOSOBO	48
YOGYAKARTA	40

Dengan mengambil tiga buah data secara acak, yaitu pada data kota Ambarawa, Blora, dan Jepara, dan setelah melalui 8 langkah proses kluster, maka diperoleh hasil empat kluster sebagai berikut :

Agar hasil kluster yang telah diperoleh bisa lebih optimal maka perlu dilakukan proses kluster yang kedua dengan mengambil tiga buah data secara acak yang berbeda yaitu untuk kota Batang, Kendal dan Wonosobo, seperti terlihat pada tabel 9.

Tabel 9. Memilih secara acak empat buah data awal sebagai pusat kluster sementara pada proses kluster tahap kedua

Asal	Jml
ACEH UTARA	1
AMBARAWA	18
AMBON	1
...	
BATAM	1
BATANG	35
BATURAJA	1
...	
WONOGIRI	20
WONOSOBO	48
YOGYAKARTA	40

Terbukti pada proses kluster tahap kedua dengan mengambil tiga buah set data yang berbeda hasilnya masih sama seperti pada proses kluster yang pertama, meskipun dengan jumlah langkah yang lebih banyak yaitu 8 langkah.

4.12. Hasil Kluster

Berdasarkan proses data mining dengan teknik kluster yang diterapkan pada data mahasiswa baru yang dilakukan pada Bab 4 diperoleh informasi mengenai beberapa kelompok data berdasarkan program studi dan kota asal. Dengan adanya informasi mengenai kelompok program studi dimana sebagai dasar perhitungan adalah jumlah mahasiswa baru yang diterima,

maka dapat disimpulkan adanya beberapa program studi yang cukup berpotensi untuk dikembangkan sebagai program studi unggulan. Namun demikian terdapat juga kelompok program studi yang menunjukkan adanya potensi yang menurun dan harus ditingkatkan mengingat jumlah mahasiswa baru yang diterima relatif sedikit setiap tahun, seperti terlihat pada tabel 11. Beberapa program studi yang masuk dalam kelompok jumlah mahasiswa yang banyak seperti Sistem Informasi dan Teknik Informatika belum tentu juga selalu mendapatkan jumlah mahasiswa baru yang meningkat setiap tahun, hal tersebut dikarenakan jumlah mahasiswa baru yang dihitung merupakan rekapitulasi dari tahun angkatan 2005-2010. Secara umum jumlah penerimaan mahasiswa baru setiap tahun mengalami penurunan mulai tahun 2005, namun demikian pada beberapa program studi ada yang mengalami penurunan secara drastis seperti Manajemen Informatika dan Komputerisasi Akuntansi sedangkan pada beberapa program studi yang lain mengalami sedikit kenaikan seperti Teknik Informatika. Beberapa program studi dengan jumlah sangat sedikit disebabkan karena program studi tersebut sudah tidak menerima mahasiswa baru lagi dan dalam proses passing out atau ditutup, sedangkan alasan lain adalah karena program studi tersebut baru masuk atau bergabung pada tahun-tahun terakhir. Hasil klustering seperti terlihat pada gambar 11.

Tabel 10. Hasil klustering program studi

No.	Program Studi	Klust1	Klust2	Klust3	Klust4
1	Akuntansi	-	812	-	-
2	Bahasa Inggris	-	-	-	83
3	Ilmu Hukum	220	-	-	-
4	Keuangan dan Perbankan	-	-	-	55
5	Komputerisasi Akuntansi	-	-	155	-
6	Manajemen	-	671	-	-
7	Manajemen Industri	-	-	-	1
8	Manajemen Informatika	249	-	-	-
9	Perhotelan	-	-	134	-
10	Sastra Inggris	347	-	-	-
11	Sistem Informasi	-	917	-	-
12	Teknik Elektronika	-	-	-	10
13	Teknik Industri	-	-	-	82
14	Teknik Informatika	-	1051	-	-
15	Teknik Komputer	-	-	-	81
16	Usaha Perjalanan Wisata	-	-	-	3
Jumlah		272,00	862,75	144,50	45,00

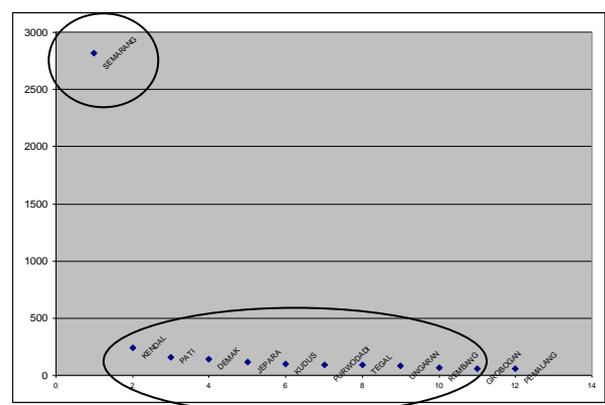
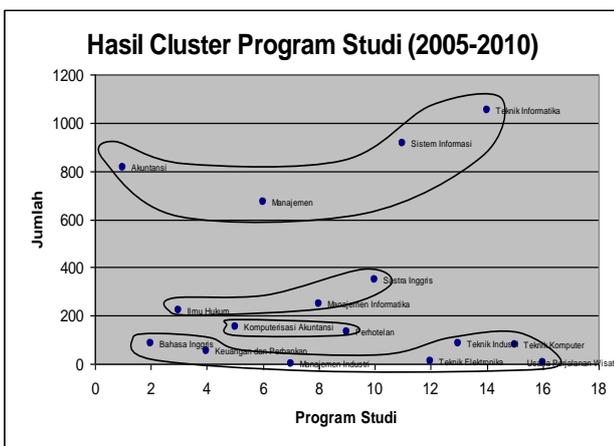
Gambar 11. Hasil klustering program studi

Hasil berikutnya dari proses klustering adalah kelompok kota asal mahasiswa baru, dimana dengan sengaja telah dipilih sebanyak empat buah kluster untuk mendapatkan hasil kluster yang maksimal dengan beberapa kali proses menggunakan set data yang berbeda. Dari sejumlah 173 kota asal mahasiswa baru, setelah melalui proses klustering diperoleh tiga buah kluster dengan jumlah mahasiswa terbanyak dan pada 10 kota. Hal ini tentunya akan sangat bermanfaat dalam membantu tim penerimaan mahasiswa baru dan promosi dalam menentukan kota tujuan promosi. Tabel 11 dan gambar 12 menunjukkan hasil klustering kota asal sekolah mahasiswa baru dengan mengambil dua hasil kluster terbesar yang menghasilkan 10 kota asal sekolah.

Tabel 11. Hasil klustering kota asal sekolah

No.	Kota	Kluster1	Kluster2	Kluster3
1	SEMARANG	2817		
2	KENDAL		245	
3	PATI		160	
4	DEMAK		139	
5	JEPARA		116	
6	KUDUS		103	
7	PURWODADI		90	
8	TEGAL		88	
9	UNGARAN		80	
10	REMBANG		70	
11	GROBOGAN			59
12	PEMALANG			57
	...			

Hasil Cluster Program Studi (2005-2010)





Gambar 12. Hasil klustering kota asal sekolah

5. KESIMPULAN

1. Data mining dengan teknik klustering pada data mahasiswa baru berdasarkan jumlah mahasiswa yang melakukan registrasi menghasilkan informasi mengenai kelompok fakultas, kelompok program studi dan kelompok kota asal sekolah, mulai dari jumlah yang paling banyak hingga jumlah yang paling sedikit
2. Hasil klustering menunjukkan bahwa beberapa program studi yang ada di Fakultas Teknologi Informasi seperti Sistem Informasi dan Teknik Informatika masih menjadi program studi pilihan dengan jumlah mahasiswa baru yang cukup banyak, selanjutnya disusul oleh Fakultas Ekonomi, Bahasa dan Ilmu Budaya, Hukum, Teknik dan terakhir adalah Program Diploma Kepariwisata

6. SARAN

1. Data mahasiswa baru sebaiknya dikelola dan disimpan dengan baik karena dapat membawa manfaat yang sangat besar untuk kegiatan pengelolaan akademik di masa yang akan datang
2. Setelah penelitian ini perlu adanya penambahan kelengkapan data mahasiswa baru yang dibutuhkan seperti status sosial dan referensi darimana calon mahasiswa baru mendapatkan informasi mengenai Universitas STIKUBANK
3. Perlu adanya penelitian lebih lanjut dalam rangka mengembangkan informasi yang lebih banyak dan mendalam mengenai asal mahasiswa baru berdasarkan atribut-atribut lain yang belum dimanfaatkan pada proses data mining dalam penelitian ini

DAFTAR PUSTAKA

- [1] Al-Radaideh, Q., A., Al-Shawakfa, E., M., & Al-Najjar, M., I., (2006). *Mining Student Data Using Decision Trees*, Yarmouk University
- [2] Budiarti A., Giri, S., Y, Yova, R., (2006). *Studi Karakteristik Kelulusan Peserta*

- Didik dengan Teknik Clustering*, Universitas Indonesia
- [3] Erdogan, S., Z., & Timor, M., (2005). *A Data Mining Application In A Student Database*, Maltepe University & Istanbul University
 - [4] Han, J. & Kamber, M. (2006). *Data mining: Concepts and Techniques*. New York: Morgan-Kaufman
 - [5] Romero, C., Ventura, S., Expejo, P., G., & Hervas, C., (2008). *Data Mining Algorithms to Classify Students*, Cordoba University
 - [6] Tan, P., Steinbach, M., & Kumar, V., (2006). *Introduction to Data Mining*, Pearson Education.