

## Analisa Profil Data Mahasiswa Baru terhadap Program Studi yang dipilih di Perguruan Tinggi Swasta Jawa Tengah dengan Menggunakan Teknik Data Mining

**Eko Nur Wahyudi, Arief Jananto dan Narwati**

Program Studi Sistem Informasi, Universitas Stikubank

email : ekowahyudi157@gmail.com, ajananto09@gmail.com, enyonge.karo73you@yahoo.co.id

### Abstrak

Jumlah PTS di lingkungan Kopertis Wilayah VI Jawa Tengah mencapai jumlah 231 buah dengan jumlah bidang ilmu sebanyak 16 buah dan program studi yang dibuka mencapai 153 buah. Dari data mahasiswa baru yang diterima selama 5 tahun terakhir, yaitu 2005 – 2009 menunjukkan adanya kenaikan dan penurunan yang cukup signifikan pada beberapa program studi.

Berdasarkan data tersebut di atas dan dengan menggunakan teknik data mining diharapkan trend mahasiswa baru dapat diketahui berdasarkan bidang ilmu dan program studi yang dipilihnya, dengan demikian orientasi penerimaan mahasiswa baru dapat diprediksi di masa yang akan datang, khususnya pada program studi yang banyak peminat dan program studi yang kurang peminat.

**Kata kunci :** Data Mining, Klustering, K-means

## PENDAHULUAN

### Latar Belakang

Jumlah Mahasiswa Baru di lingkungan Perguruan Tinggi di Jawa Tengah selama 5 (lima) tahun terakhir (2004-2009) mengalami pertumbuhan dan perkembangan yang cukup signifikan. Kenaikan jumlah mahasiswa baru terjadi di beberapa program studi yang dalam beberapa tahun terakhir sangat diminati oleh masyarakat. Namun demikian ada juga program studi yang stabil atau mengalami proses naik dan turun dalam jumlah penerimaan mahasiswa baru, bahkan ada pula program studi yang terus mengalami penurunan jumlah mahasiswa baru. Oleh karena itu perlu diketahui bagaimana sebenarnya minat calon mahasiswa terhadap program studi yang dipilihnya, serta bagaimana trend pemilihan program studi di masa yang akan datang.

Seperti diketahui bahwa jumlah perguruan tinggi swasta di wilayah Kopertis Wilayah VI Jawa Tengah mencapai 231 buah dengan rincian seperti terlihat pada tabel 1 berikut ini :

Tabel 1. Jumlah PTS di Wilayah Kopertis Wilayah VI Jawa Tengah

No.	Jenis	Jumlah
1	Universitas	34
2	Institut	2
3	Sekolah Tinggi	72
4	Akademi	105
5	Politeknik	18
	Jumlah	231

Sedangkan jumlah program studi yang ada di wilayah Kopertis Wilayah VI Jawa Tengah mencapai 153 buah, yaitu seperti terlihat pada tabel 2 berikut ini :

Tabel 2. Jumlah Program Studi di Wilayah Kopertis Wilayah VI Jawa Tengah

No.	Program Studi
1	Agribisnis
2	Agroteknologi
3	Akuntansi
4	Arsitektur
5	Bimbingan Konseling
6	Biologi
7	Budidaya Perairan

8	Desain Grafis
9	Desain Interior
10	Desain Komunikasi Visual
11	Desain Produk
12	Ekonomi Pembangunan
13	Farmasi
14	Fisika
15	Geografi
16	Hubungan Internasional
17	Ilmu Administrasi Negara
18	Ilmu Administrasi Niaga
19	Ilmu Gizi
20	Ilmu Hukum
21	Ilmu Kelautan
22	Ilmu Keperawatan
23	Ilmu Komunikasi
24	Ilmu Pemerintahan
25	Ilmu Politik
26	Ilmu Teologi
27	Kesehatan Masyarakat
28	Kimia
29	Manajemen
30	Matematika
31	Pemanfaatan Sumber Daya Perikanan
32	Pendidikan Bahasa Dan Sastra Daerah
33	Pendidikan Bahasa Dan Sastra Indonesia
34	Pendidikan Bahasa Inggris
35	Pendidikan Biologi
36	Pendidikan Dokter
37	Pendidikan Dokter Gigi
38	Pendidikan Ekonomi
39	Pendidikan Fisika
40	Pendidikan Geografi
41	Pendidikan Guru Pendidikan Anak Usia Dini
42	Pendidikan Guru Sekolah Dasar
43	Pendidikan Kepelatihan Olahraga
44	Pendidikan Matematika
45	Pendidikan Pancasila Dan Kewarganegaraan
46	Pendidikan Sejarah
47	Pendidikan Teknik Mesin
48	Pendidikan Teknik Otomotif
49	Perencanaan Wilayah Dan Kota
50	Peternakan
51	Pjj Pendidikan Guru Sekolah Dasar
52	Psikologi

53	Sastra Inggris
54	Sastra Jepang
55	Seni Musik
56	Sistem Informasi
57	Sistem Komputer
58	Sosiologi
59	Statistika
60	Teknik Elektro
61	Teknik Industri
62	Teknik Informatika
63	Teknik Kimia
64	Teknik Lingkungan
65	Teknik Mesin
66	Teknik Sipil
67	Teknologi Hasil Pertanian
68	Teknologi Pangan
69	Teknologi Pendidikan
70	Administrasi Bisnis
71	Administrasi Peradilan
72	Administrasi Perpajakan
73	Administrasi Rumah Sakit
74	Agronomi
75	Akuntansi
76	Analisis Farmasi Dan Makanan
77	Analisis Kesehatan
78	Analisis Kimia
79	Bahasa Belanda
80	Bahasa Inggris
81	Bahasa Jepang
82	Bahasa Mandarin
83	Budidaya Perairan
84	Budidaya Pertanian
85	Desain Grafis
86	Desain Interior
87	Desain Komunikasi Visual
88	Desain Produk
89	Farmasi
90	Fisioterapi
91	Gizi
92	Hubungan Masyarakat
93	Kebidanan
94	Keperawatan
95	Kesehatan Lingkungan
96	Ketatalaksanaan Pelayaran Niaga Dan Kepelabuhan
97	Keuangan Dan Perbankan
98	Komputerisasi Akuntansi
99	Komunikasi Massa
100	Kriya Kain

101	Malariologi
102	Manajemen Administrasi
103	Manajemen Industri
104	Manajemen Informatika
105	Manajemen Pemasaran
106	Manajemen Perdagangan
107	Manajemen Perpajakan
108	Manajemen Perusahaan
109	Material
110	Mesin Dan Peralatan Pertanian
111	Mesin Industri
112	Mesin Otomotif
113	Nautika
114	Nutrisi Dan Makanan Ternak
115	Pariwisata
116	Pengecoran Logam
117	Penyiaran
118	Perawatan Dan Perbaikan Mesin
119	Perekam Dan Informasi Kesehatan
120	Perhotelan
121	Pola Pengecoran
122	Produksi Ternak
123	Radiodiagnostik Dan Radioterapi
124	Refraksi Optisi
125	Sekretari
126	Seni Pedalangan
127	Seni Tari
128	Seni Teater
129	Sosial Ekonomi Pertanian (Agrobisnis)
130	Statistika
131	Tata Boga
132	Tata Busana
133	Tata Rias
134	Teknik Bangunan Kapal
135	Teknik Elektronika
136	Teknik Industri
137	Teknik Industri Tekstil
138	Teknik Informatika
139	Teknik Kelistrikan Kapal
140	Teknik Kimia
141	Teknik Komputer
142	Teknik Lingkungan
143	Teknik Listrik
144	Teknik Manufaktur
145	Teknik Mekatronika
146	Teknik Mesin
147	Teknik Perancangan Manufaktur
148	Teknik Permesinan Kapal

149	Teknik Sipil
150	Teknik Telekomunikasi
151	Teknika
152	Teknologi Produksi Pertanian
153	Usaha Perjalanan Wisata

Berdasarkan tabel 2 tentang jumlah program studi dapat dikelompokkan menjadi bidang ilmu seperti terlihat pada tabel 3 berikut ini :

Tabel 3. Jumlah Bidang Ilmu di Wilayah Kopertis Wilayah VI Jawa Tengah

No.	Bidang Ilmu
1	Agama Dan Filsafat
2	Aneka Ilmu
3	Budaya Dan Sastra
4	Ekonomi
5	Hukum
6	Kelautan & Perikanan
7	Kependidikan
8	Kesehatan
9	Komputer
10	Mipa
11	Pertanian
12	Peternakan
13	Psikologi
14	Seni
15	Sosial
16	Teknik

Sedangkan berdasarkan jenjang program S1 dan D3, di Kopertis Wilayah VI Jawa Tengah terdistribusi seperti terlihat pada tabel 4 berikut ini :

Tabel 4. Jumlah Jenjang Program S1 dan D3 di Wilayah Kopertis Wilayah VI Jawa Tengah

No.	Jenjang Program	Jumlah
1	S1	69
2	D3	84
	Jumlah	153

Selanjutnya adalah tabel 5 yang berisi tentang jumlah mahasiswa baru yang diterima di semua perguruan tinggi mulai tahun 2005 sampai dengan 2009 berdasarkan bidang ilmu yaitu sebagai berikut :

Tabel 5. Jumlah Mahasiswa Baru di Kopertis Wilayah VI Jawa Tengah berdasarkan Bidang Ilmu mulai Tahun 2005 - 2009

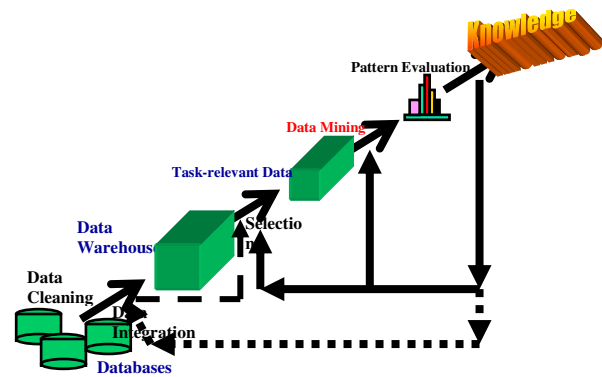
No.	Bidang Ilmu	Jumlah
1	AGAMA DAN FILSAFAT	666
2	ANEKA ILMU	6.229
3	BUDAYA DAN SASTRA	6.292
4	EKONOMI	94.093
5	HUKUM	26.954
6	KELAUTAN & PERIKANAN	607
7	KEPENDIDIKAN	95.221
8	KESEHATAN	63.956
9	KOMPUTER	54.992
10	MIPA	10.656
11	PERTANIAN	3.997
12	PETERNAKAN	515
13	PSIKOLOGI	7.546
14	SENI	2.448
15	SOSIAL	11.384
16	TEKNIK	33.273
	<b>Jumlah</b>	<b>46,339</b>

Dari tabel-tabel data yang telah disajikan nampak adanya suatu trend kenaikan pada beberapa bidang ilmu namun ada juga yang mengalami penurunan dalam hal jumlah mahasiswa baru yang diterima.

**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 pengelompokkan 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.

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).

**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).

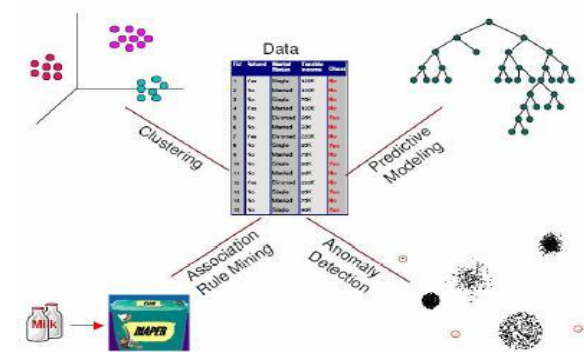
**Visualisasi**

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).

**Analisa Klaster dan K-means**

Analisa klaster 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 klastering yang



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

**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

mencoba menemukan jumlah klaster ( $K$ ), 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).

### **Penerapan Teknik Klastering dan Klasifikasi**

Teknik klastering pada jumlah data mahasiswa baru PTS di lingkungan Kopertis Wilayah VI Jawa Tengah ini bertujuan untuk menentukan kelompok-kelompok data mahasiswa baru secara parsial untuk mengetahui bidang ilmu dan program studi apa yang menjadi peminatan.

### **Data Mining dengan Teknik Klastering**

Teknik klastering yang akan digunakan adalah *K-means*. *K-means* dipilih karena merupakan salah satu teknik yang paling populer sekaligus paling mudah diimplementasikan untuk proses klastering 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 klastering 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 klastering 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

### **Menentukan klaster bidang ilmu berdasar jumlah mahasiswa baru**

Mula-mula data akan diklaster berdasarkan jumlah mahasiswa baru untuk kelompok bidang ilmu. Tujuan dari klastering ini adalah untuk menentukan kelompok bidang ilmu berdasarkan jumlah mahasiswa baru yang diterima.

Metode yang akan dipergunakan dalam klastering ini adalah *K-means*, dimana proses klastering 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 klaster yang tetap.

Penentuan jumlah kluster yang akan dibentuk disesuaikan dengan kebutuhan atau tujuan dari hasil klustering. Tiga buah kluster yang akan dibentuk berikut ini hanya merupakan suatu contoh untuk tujuan mengambil sebuah kluster dari tiga buah kluster yang terbentuk sebagai kluster yang dominan dalam hal jumlah penerimaan mahasiswa baru. Sedangkan kluster kedua mempunyai peminatan sedang dan kluster ketiga merupakan hasil kluster yang lebih sedikit dalam jumlah penerimaan mahasiswa baru.

Dari tabel rekapitulasi bidang ilmu diambil tiga buah data secara acak yaitu Agama dan Filsafat (666), Kesehatan (63.596) dan Teknik (33.273) untuk menentukan jarak awal dari tiga buah kluster yang akan dibentuk, yaitu seperti terlihat pada tabel 6.

Tabel 6. Tiga buah data awal secara acak sebagai pusat kluster sementara

No.	Bidang Ilmu	Jumlah
1	Agama dan Filsafat	666
2	Aneka Ilmu	6,229
3	Budaya dan Sastra	6,292
4	Ekonomi	94,093
5	Hukum	26,954
6	Kelautan & Perikanan	607
7	Kependidikan	95,221
8	Kesehatan	63,956
9	Komputer	54,992
10	Mipa	10,656
11	Pertanian	3,997
12	Peternakan	515
13	Psikologi	7,546
14	Seni	2,448
15	Sosial	11,384
16	Teknik	33,273

Jarak-1a adalah jarak nilai data antara masing-masing data dengan data yang dipilih pertama yaitu Agama dan Filsafat. Jarak-2a adalah jarak nilai data antara masing-masing data dengan data yang dipilih kedua yaitu Kesehatan. Sedangkan untuk Jarak-3a yaitu membandingkan masing-masing data dengan data yang dipilih ketiga yaitu Teknik. Setelah itu data jumlah dialokasikan pada posisi nilai terkecil antara Jarak-1a sampai dengan Jarak-3a untuk melihat klasternya, sehingga hasilnya seperti terlihat pada Tabel 7.

Tabel 7. Tahap 1 proses penentuan jarak dan kluster

Jarak-1a	Jarak-2a	Jarak-3a	Cluster-1a	Cluster-2a	Cluster-3a
0	602,044	632,727	666		
659,771	57,727	27,044			6,229
659,708	57,664	26,981			6,292
571,907	30,137	60,82		94,093	
639,046	37,002	6,319			26,954
59	543,044	573,727	607		
570,779	31,265	61,948		95,221	
602,044	0	30,683		63,956	
611,008	8,964	21,719		54,992	
655,344	53,3	22,617			10,656
662,003	59,959	29,276			3,997
151	451,044	481,727	515		
658,454	56,41	25,727			7,546
663,552	61,508	30,825			2,448
654,616	52,572	21,889			11,384
632,727	30,683	0			33,273
			596	77,066	12,087

Proses perhitungan jarak dan kluster tersebut kemudian diulang kembali sampai nilai rata-rata pada setiap kluster tidak berubah sehingga dapat diputuskan bahwa kluster yang terbentuk sudah optimal, seperti tampak pada tabel 8.

Tabel 8. Tahap 2 proses penentuan jarak dan kluster

Jarak-1b	Jarak-2b	Jarak-3b	Cluster-1b	Cluster-2b	Cluster-3b
70	588,9345	653,9134	666		
589,771	70,8365	5,857556			6,229
589,708	70,7735	5,794556			6,292
501,907	17,0275	82,00644		94,093	
569,046	50,1115	14,86744			26,954
11	529,9345	594,9134	607		
500,779	18,1555	83,13444		95,221	
532,044	13,1095	51,86944		63,956	
541,008	22,0735	42,90544		54,992	
585,344	66,4095	1,430556			10,656
592,003	73,0685	8,089556			3,997
81	437,9345	502,9134	515		
588,454	69,5195	4,540556			7,546
593,552	74,6175	9,638556			2,448

584,616	65,6815	0,702556			<b>11,384</b>
562,727	43,7925	21,18644			<b>33,273</b>
			<b>596</b>	<b>77,066</b>	<b>12,087</b>

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. Proses seperti ini dilakukan beberapa kali dengan data acak yang berbeda untuk menguji apakah hasil kluster yang diperoleh sudah benar. Dengan demikian dapat disimpulkan hasilnya bahwa dalam tahun 2005-2009 berdasarkan jumlah mahasiswa baru yang diterima di PTS di lingkungan Kopertis Wilayah VI Jawa Tengah sesuai dengan bidang ilmunya adalah seperti terlihat pada tabel 9. sebagai berikut :

Tabel 9. Kesimpulan hasil kluster berdasarkan bidang ilmu

Kluster		Jumlah
<b>Pertama</b>		
1	Kependidikan	95.221
2	Ekonomi	94.093
3	Kesehatan	63.956
4	Komputer	54.992
<b>Kedua</b>		
1	Teknik	33.273
2	Hukum	26.954
3	Sosial	11.384
4	MIPA	10.656
5	Psikologi	7.546
6	Budaya dan Sastra	6.292
7	Aneka Ilmu	6.229
8	Pertanian	3.997
9	Seni	2.448
<b>Ketiga</b>		
1	Agama dan Filsafat	667
2	Kelautan dan Perikanan	607
3	Peternakan	515

Proses selanjutnya adalah membuat kluster berdasarkan Program Studi, dengan tujuan mengetahui minat mahasiswa terhadap program studi yang dipilihnya. Pertama diambil empat buah data acak sebagai titik awal kluster sementara, yaitu Akuntansi (27.586), Pendidikan Ekonomi (3.708), Administrasi Rumah Sakit (220) dan Manajemen Perusahaan (2.687), yaitu seperti tampak pada tabel 10 sebagai berikut :

Tabel 10. Empat buah data awal secara acak sebagai pusat kluster sementara

No.	Program Studi	Jenjang	Jml
1	Agribisnis	S-1	1.539
2	Agroteknologi	S-1	1.723
<b>3</b>	<b>Akuntansi</b>	<b>S-1</b>	<b>27.586</b>
4	Arsitektur	S-1	1.788
5	Bimbingan Konseling	S-1	10.884
6	Biologi	S-1	244
7	Budidaya Perairan	S-1	246
8	Desain Grafis	S-1	100
9	Desain Interior	S-1	50
10	Desain Komunikasi Visual	S-1	1.024
11	Desain Produk	S-1	128
12	Ekonomi Pembangunan	S-1	1.407
13	Farmasi	S-1	6.340
14	Fisika	S-1	57
15	Geografi	S-1	349
16	Hubungan Internasional	S-1	277
17	Ilmu Administrasi Negara	S-1	4.270
18	Ilmu Administrasi Niaga	S-1	518
19	Ilmu Gizi	S-1	396
20	Ilmu Hukum	S-1	26.876
21	Ilmu Kelautan	S-1	9
22	Ilmu Keperawatan	S-1	11.349
23	Ilmu Komunikasi	S-1	3.003
24	Ilmu Pemerintahan	S-1	649
25	Ilmu Politik	S-1	264
26	Ilmu Teologi	S-1	666
27	Kesehatan Masyarakat	S-1	3.350
28	Kimia	S-1	300
29	Manajemen	S-1	51.884
30	Matematika	S-1	511
31	Pemanfaatan Sumber Daya Perikanan	S-1	137
32	Pendidikan Bahasa Dan Sastra Daerah	S-1	2.386
33	Pendidikan Bahasa Dan Sastra Indonesia	S-1	15.669
34	Pendidikan Bahasa Inggris	S-1	20.007
35	Pendidikan Biologi	S-1	3.900
36	Pendidikan Dokter	S-1	2.439
37	Pendidikan Dokter Gigi	S-1	105
<b>38</b>	<b>Pendidikan Ekonomi</b>	<b>S-1</b>	<b>3.708</b>
39	Pendidikan Fisika	S-1	2.332



40	Pendidikan Geografi	S-1	2.003	77	Analisis Kesehatan	D-3	1.561
	Pendidikan Guru			78	Analisis Kimia	D-3	136
41	Pendidikan Anak Usia Dini	S-1	3.374	79	Bahasa Belanda	D-3	66
42	Pendidikan Guru Sekolah Dasar	S-1	3.724	80	Bahasa Inggris	D-3	2.624
43	Pendidikan Kepelatihan Olahraga	S-1	5.189	81	Bahasa Jepang	D-3	130
44	Pendidikan Matematika	S-1	11.648	82	Bahasa Mandarin	D-3	135
45	Pendidikan Pancasila Dan Kewarganegaraan	S-1	8.500	83	Budidaya Perairan	D-3	215
46	Pendidikan Sejarah	S-1	948	84	Budidaya Pertanian	D-3	198
47	Pendidikan Teknik Mesin	S-1	289	85	Desain Grafis	D-3	407
48	Pendidikan Teknik Otomotif	S-1	100	86	Desain Interior	D-3	55
49	Perencanaan Wilayah Dan Kota	S-1	149		Desain Komunikasi Visual	D-3	158
50	Peternakan	S-1	268	87	Desain Produk	D-3	166
51	Pjj Pendidikan Guru Sekolah Dasar	S-1	414	88	Farmasi	D-3	1.845
52	Psikologi	S-1	7.546	89	Fisioterapi	D-3	524
53	Sastra Inggris	S-1	3.126	90	Gizi	D-3	366
54	Sastra Jepang	S-1	211	91	Hubungan Masyarakat	D-3	214
55	Seni Musik	S-1	229	92	Kebidanan	D-3	23.963
56	Sistem Informasi	S-1	11.440	93	Keperawatan	D-3	17.165
57	Sistem Komputer	S-1	2.388	94	Kesehatan Lingkungan	D-3	61
58	Sosiologi	S-1	149		Ketatalaksanaan Pelayaran Niaga Dan Kepelabuhan	D-3	1.477
59	Statistika	S-1	56	95	Keuangan Dan Perbankan	D-3	2.087
60	Teknik Elektro	S-1	4.473		Komputerisasi Akuntansi	D-3	3.659
61	Teknik Industri	S-1	2.155	96	Komunikasi Massa	D-3	119
62	Teknik Informatika	S-1	18.859	97	Kriya Kain	D-3	65
63	Teknik Kimia	S-1	808	98	Malariologi	D-3	89
64	Teknik Lingkungan	S-1	65		Manajemen Administrasi	D-3	487
65	Teknik Mesin	S-1	3.163	99	Manajemen Industri	D-3	231
66	Teknik Sipil	S-1	5.634	100	Manajemen Informatika	D-3	9.728
67	Teknologi Hasil Pertanian	S-1	720	101	Manajemen Pemasaran	D-3	267
68	Teknologi Pangan	S-1	1.083		Manajemen Perdagangan	D-3	96
69	Teknologi Pendidikan	S-1	146	102	Manajemen Perpajakan	D-3	657
70	Administrasi Bisnis	D-3	96	<b>108</b>	<b>Manajemen Perusahaan</b>	<b>D-3</b>	<b>2.687</b>
71	Administrasi Peradilan	D-3	78	109	Material	D-3	5
72	Administrasi Perpajakan	D-3	187		Mesin Dan Peralatan Pertanian	D-3	138
<b>73</b>	<b>Administrasi Rumah Sakit</b>	<b>D-3</b>	<b>220</b>	110	Mesin Industri	D-3	1.154
74	Agronomi	D-3	102	111	Mesin Otomotif	D-3	1.096
75	Akuntansi	D-3	6.935	112	Nautika	D-3	1.322
76	Analisis Farmasi Dan Makanan	D-3	408		Nutrisi Dan Makanan Ternak	D-3	55
				113	Pariwisata	D-3	658
				114			
				115			

116	Pengecoran Logam	D-3	113
117	Penyiaran	D-3	200
118	Perawatan Dan Perbaikan Mesin	D-3	14
119	Perekam Dan Informasi Kesehatan	D-3	1.650
120	Perhotelan	D-3	983
121	Pola Pengecoran	D-3	13
122	Produksi Ternak	D-3	192
123	Radiodiagnostik Dan Radioterapi	D-3	314
124	Refraksi Optisi	D-3	404
125	Sekretari	D-3	1.438
126	Seni Pedalangan	D-3	32
127	Seni Tari	D-3	14
128	Seni Teater	D-3	20
129	Sosial Ekonomi Pertanian (Agrobisnis)	D-3	435
130	Statistika	D-3	546
131	Tata Boga	D-3	82
132	Tata Busana	D-3	114
133	Tata Rias	D-3	76
134	Teknik Bangunan Kapal	D-3	47
135	Teknik Elektronika	D-3	2.573
136	Teknik Industri	D-3	89
137	Teknik Industri Tekstil	D-3	100
138	Teknik Informatika	D-3	4.858
139	Teknik Kelistrikan Kapal	D-3	73
140	Teknik Kimia	D-3	270
141	Teknik Komputer	D-3	4.060
142	Teknik Lingkungan	D-3	272
143	Teknik Listrik	D-3	120
144	Teknik Manufaktur	D-3	101
145	Teknik Mekatronika	D-3	175
146	Teknik Mesin	D-3	4.853
147	Teknik Perancangan Manufaktur	D-3	178
148	Teknik Permesinan Kapal	D-3	160
149	Teknik Sipil	D-3	484
150	Teknik Telekomunikasi	D-3	821
151	Teknika	D-3	1.232
152	Teknologi Produksi Pertanian	D-3	19
153	Usaha Perjalanan Wisata	D-3	285
		<b>Jumlah</b>	<b>418.829</b>

Seperti halnya pada proses klastering sesuai bidang ilmu, pada program studi kali ini juga dilakukan beberapa tahapan penentuan jarak dan klaster sehingga diperoleh hasil seperti pada tabel 11 berikut ini :

**Tabel 11. Hasil klastering program studi berdasarkan Jumlah mahasiswa baru**

klaster-1i	klaster-2i	klaster-3i	klaster-4i
		1539	
		1723	
27586			
		1788	
	10884		
		244	
		246	
		100	
		50	
		1024	
		128	
		1407	
			6340
		57	
		349	
		277	
			4270
		518	
26876		396	
		9	
	11349		
			3003
		649	
		264	
		666	
			3350
		300	
51884			
		511	
		137	
			2386
	15669		
	20007		
			3900
			2439
		105	
			3708
			2332
		2003	

			3374
			3724
			5189
	11648		
			8500
		948	
		289	
		100	
		149	
		268	
		414	
			7546
			3126
		211	
		229	
	11440		
			2388
		149	
		56	
			4473
		2155	
	18859		
		808	
		65	
			3163
			5634
		720	
		1083	
		146	
		96	
		78	
		187	
		220	
		102	
			6935
		408	
		1561	
		136	
		66	
			2624
		130	
		135	
		215	
		198	
		407	
		55	
		158	
		166	
		1845	

		524	
		366	
		214	
23963			
	17165		
		61	
		1477	
		2087	
			3659
		119	
		65	
		89	
		487	
		231	
	9728		
		267	
		96	
		657	
			2687
		5	
		138	
		1154	
		1096	
		1322	
		55	
		658	
		113	
		200	
		14	
		1650	
		983	
		13	
		192	
		314	
		404	
		1438	
		32	
		14	
		20	
		435	
		546	
		82	
		114	
		76	
		47	
			2573
		89	
		100	
			4858

		73	
		270	
			4060
		272	
		120	
		101	
		175	
			4853
		178	
		160	
		484	
		821	
		1232	
		19	
		285	
<b>32577,25</b>	<b>14083,22</b>	<b>448,469</b>	<b>4114,593</b>

Proses klastering program studi dilakukan sebanyak tiga kali dengan menggunakan susunan data acak yang berbeda, namun demikian ternyata hasilnya tetap sama dengan hasil klastering proses pertama, yaitu seperti terlihat pada tabel 12 sebagai berikut :

Tabel 12. Hasil klasifikasi program studi

Klaster 1 :

No.	Program Studi	Jenjang	Jumlah
1	Manajemen	S-1	51.884
2	Akuntansi	S-1	27.586
3	Ilmu Hukum	S-1	26.876
4	Kebidanan	D-3	23.963

Klaster 2 :

No.	Program Studi	Jenjang	Jumlah
1	Pendidikan Bahasa Inggris	S-1	20.007
2	Teknik Informatika	S-1	18.859
3	Keperawatan	D-3	17.165
4	Pendidikan Bahasa dan Sastra Indonesia	S-1	15.669
5	Pendidikan Matematika	S-1	11.648
6	Sistem Informasi	S-1	11.440
7	Ilmu Keperawatan	S-1	11.349
8	Bimbingan Konseling	S-1	10.884
9	Manajemen Informatika	D-3	9.728

Klaster 3 :

No.	Program Studi	Jenjang	Jumlah
1	Pendidikan Pancasila dan Kewarganegaraan	S-1	8.500
2	Psikologi	S-1	7.546

3	Akuntansi	D-3	6.935
4	Farmasi	S-1	6.340
5	Teknik Sipil	S-1	5.634
6	Pendidikan Keperawatan Olahraga	S-1	5.189
7	Teknik Informatika	D-3	4.858
8	Teknik Mesin	D-3	4.853
9	Teknik Elektro	S-1	4.473
10	Ilmu Administrasi Negara	S-1	4.270
11	Teknik Komputer	D-3	4.060
12	Pendidikan Biologi	S-1	3.900
13	Pendidikan Guru Sekolah Dasar	S-1	3.724
14	Pendidikan Ekonomi	S-1	3.708
15	Komputerisasi Akuntansi	D-3	3.659
16	Pendidikan Guru Pendidikan Anak Usia Dini	S-1	3.374
17	Kesehatan Masyarakat	S-1	3.350
18	Teknik Mesin	S-1	3.163
19	Sastra Inggris	S-1	3.126
20	Ilmu Komunikasi	S-1	3.003
21	Manajemen Perusahaan	D-3	2.687
22	Bahasa Inggris	D-3	2.624
23	Teknik Elektronika	D-3	2.573
24	Pendidikan Dokter	S-1	2.439
25	Sistem Komputer	S-1	2.388
26	Pendidikan Bahasa Dan Sastra Daerah	S-1	2.386
27	Pendidikan Fisika	S-1	2.332

Klaster 4 :

No.	Program Studi	Jenjang	Jumlah
1	Teknik Industri	S-1	2.155
2	Keuangan Dan Perbankan	D-3	2.087
3	Pendidikan Geografi	S-1	2.003
4	Farmasi	D-3	1.845
5	Arsitektur	S-1	1.788
6	Agroteknologi	S-1	1.723
7	Perekam Dan Informasi Kesehatan	D-3	1.650
8	Analisis Kesehatan	D-3	1.561
9	Agribisnis	S-1	1.539
10	Ketatalaksanaan Pelayaran Niaga Dan Kepelabuhan	D-3	1.477

11	Sekretari	D-3	1.438
12	Ekonomi Pembangunan	S-1	1.407
13	Nautika	D-3	1.322
14	Teknika	D-3	1.232
15	Mesin Industri	D-3	1.154
16	Mesin Otomotif	D-3	1.096
17	Teknologi Pangan	S-1	1.083
18	Desain Komunikasi Visual	S-1	1.024
19	Perhotelan	D-3	983
20	Pendidikan Sejarah	S-1	948
21	Teknik Telekomunikasi	D-3	821
22	Teknik Kimia	S-1	808
23	Teknologi Hasil Pertanian	S-1	720
24	Ilmu Teologi	S-1	666
25	Pariwisata	D-3	658
26	Manajemen Perpajakan	D-3	657
27	Ilmu Pemerintahan	S-1	649
28	Statistika	D-3	546
29	Fisioterapi	D-3	524
30	Ilmu Administrasi Niaga	S-1	518
31	Matematika	S-1	511
32	Manajemen Administrasi	D-3	487
33	Teknik Sipil	D-3	484
34	Sosial Ekonomi Pertanian (Agrobisnis)	D-3	435
35	Pjj Pendidikan Guru Sekolah Dasar	S-1	414
36	Analisis Farmasi Dan Makanan	D-3	408
37	Desain Grafis	D-3	407
38	Refraksi Optisi	D-3	404
39	Ilmu Gizi	S-1	396
40	Gizi	D-3	366
41	Geografi	S-1	349
42	Radiodiagnostik Dan Radioterapi	D-3	314
43	Kimia	S-1	300
44	Pendidikan Teknik Mesin	S-1	289
45	Usaha Perjalanan Wisata	D-3	285
46	Hubungan Internasional	S-1	277
47	Teknik Lingkungan	D-3	272
48	Teknik Kimia	D-3	270

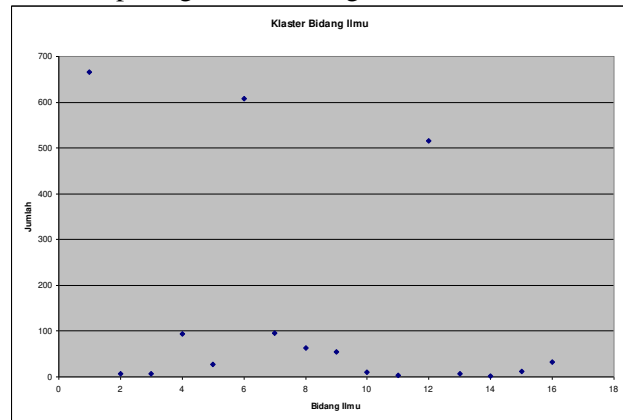
49	Peternakan	S-1	268
50	Manajemen Pemasaran	D-3	267
51	Ilmu Politik	S-1	264
52	Budidaya Perairan	S-1	246
53	Biologi	S-1	244
54	Manajemen Industri	D-3	231
55	Seni Musik	S-1	229
56	Administrasi Rumah Sakit	D-3	220
57	Budidaya Perairan	D-3	215
58	Hubungan Masyarakat	D-3	214
59	Sastra Jepang	S-1	211
60	Penyiaran	D-3	200
61	Budidaya Pertanian	D-3	198
62	Produksi Ternak	D-3	192
63	Administrasi Perpajakan	D-3	187
64	Teknik Perancangan Manufaktur	D-3	178
65	Teknik Mekatronika	D-3	175
66	Desain Produk	D-3	166
67	Teknik Permesinan Kapal	D-3	160
68	Desain Komunikasi Visual	D-3	158
69	Perencanaan Wilayah Dan Kota	S-1	149
70	Sosiologi	S-1	149
71	Teknologi Pendidikan	S-1	146
72	Mesin Dan Peralatan Pertanian	D-3	138
73	Pemanfaatan Sumber Daya Perikanan	S-1	137
74	Analisis Kimia	D-3	136
75	Bahasa Mandarin	D-3	135
76	Bahasa Jepang	D-3	130
77	Desain Produk	S-1	128
78	Teknik Listrik	D-3	120
79	Komunikasi Massa	D-3	119
80	Tata Busana	D-3	114
81	Pengecoran Logam	D-3	113
82	Pendidikan Dokter Gigi	S-1	105
83	Agronomi	D-3	102
84	Teknik Manufaktur	D-3	101
85	Desain Grafis	S-1	100
86	Pendidikan Teknik Otomotif	S-1	100
87	Teknik Industri Tekstil	D-3	100
88	Administrasi Bisnis	D-3	96

89	Manajemen Perdagangan	D-3	96
90	Malariologi	D-3	89
91	Teknik Industri	D-3	89
92	Tata Boga	D-3	82
93	Administrasi Peradilan	D-3	78
94	Tata Rias	D-3	76
95	Teknik Kelistrikan Kapal	D-3	73
96	Bahasa Belanda	D-3	66
97	Teknik Lingkungan	S-1	65
98	Kriya Kain	D-3	65
99	Kesehatan Lingkungan	D-3	61
100	Fisika	S-1	57
101	Statistika	S-1	56
102	Desain Interior	D-3	55
103	Nutrisi Dan Makanan Ternak	D-3	55
104	Desain Interior	S-1	50
105	Teknik Bangunan Kapal	D-3	47
106	Seni Pedalangan	D-3	32
107	Seni Teater	D-3	20
108	Teknologi Produksi Pertanian	D-3	19
109	Perawatan Dan Perbaikan Mesin	D-3	14
110	Seni Tari	D-3	14
111	Pola Pengecoran	D-3	13
112	Ilmu Kelautan	S-1	9
113	Material	D-3	5

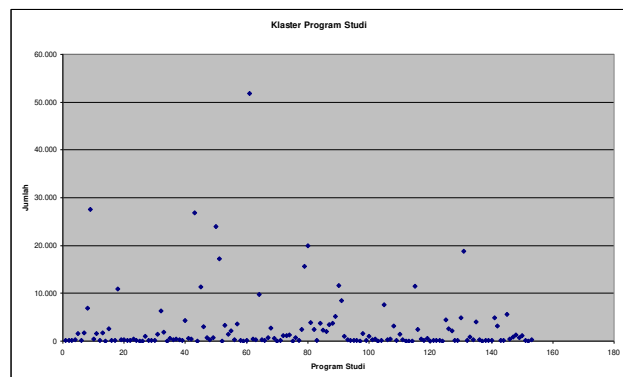
**Pembahasan dan Visualisasi Hasil Klastering**

Berdasarkan hasil klastering pada pembahasan sebelumnya yaitu berdasarkan bidang ilmu dan program studi nampak bahwa tidak semua bidang ilmu yang dominan pada klaster pertama seperti kependidikan menjadi minat utama pada calon mahasiswa karena hanya beberapa program studi kependidikan saja yang jumlahnya tinggi sementara kependidikan lainnya tetap pada jumlah sedang dan rendah. Demikian pula pada bidang ilmu komputer dimana hanya program studi Teknik Informatika, Sistem Informasi dan Manajemen Informatika yang cukup tinggi peminatnya sedangkan untuk program studi lainnya berada pada hasil klaster berikutnya.

Berikut ini adalah visualisasi hasil klastering bidang ilmu dan program studi seperti terlihat pada gambar 3 dan gambar 4.



Gambar 3. Visualisasi hasil klastering bidang ilmu



Gambar 4. Visualisasi hasil klastering program studi

**Kesimpulan**

Berdasarkan analisa profil data mahasiswa baru terhadap program studi yang dipilih di perguruan tinggi swasta Jawa Tengah dengan menggunakan teknik data mining maka dapat disimpulkan bahwa :

1. Data mining dengan teknik klustering pada data mahasiswa baru pada PTS di lingkungan Kopertis Wilayah VI Jawa Tengah berdasarkan jumlah mahasiswa yang melakukan registrasi menghasilkan informasi mengenai kelompok bidang ilmu dan program studi mulai dari jumlah yang paling banyak hingga jumlah yang paling sedikit
2. Hasil klastering menunjukkan bahwa beberapa bidang ilmu mempunyai dominasi

yang cukup tinggi terhadap minat masuk mahasiswa baru, namun demikian tidak semua program studi yang ada pada bidang ilmu yang dominan meraih jumlah mahasiswa yang banyak, hanya beberapa program studi saja yang memiliki dominasi yang cukup tinggi sesuai hasil klaster pada bidang ilmu

3. Trend minat mahasiswa terhadap bidang ilmu cukup signifikan terhadap program studi yang dipilihnya

### Saran

Dari hasil kesimpulan yang di dapat maka dapat diberikan saran sebagai berikut :

1. Data mahasiswa baru sebaiknya dikelola dan disimpan dengan baik karena dapat membawa manfaat yang sangat besar untuk kegiatan pengelolaan institusi di masa yang akan datang
2. Perlu adanya penambahan kelengkapan data mahasiswa baru yang dibutuhkan seperti status sosial dan referensi darimana calon mahasiswa baru berasal
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

- Al-Radaideh, Q., A., Al-Shawakfa, E., M., & Al-Najjar, M., I., (2006). *Mining Student Data Using Decision Trees*, Yarmouk University
- Budiarti A., Giri, S., Y, Yova, R., (2006). *Studi Karakteristik Kelulusan Peserta Didik dengan Teknik Clustering*, Universitas Indonesia.
- Erdogan, S., Z., & Timor, M., (2005). *A Data Mining Application In A Student Database*, Maltepe University & Istanbul University
- Han, J. & Kamber, M. (2006). *Data mining: Concepts and Techniques*. New York: Morgan-Kaufman

Romero, C., Ventura, S., Expejo, P., G., & Hervas, C., (2008). *Data Mining Algorithms to Classify Students*, Cordoba University

Tan, P., Steinbach, M., & Kumar, V., (2006). *Introduction to Data Mining*, Pearson Education.