

Teknik Klasifikasi untuk Melihat Kecenderungan Calon Mahasiswa Baru dalam Memilih Jenjang Pendidikan Program Studi di Perguruan Tinggi

Eko Nur Wahyudi

Program Studi Sistem Informasi, Universitas Stikubank

email : ekowahyudi157@gmail.com

Abstrak

Terdapat jumlah kenaikan yang cukup berarti pada jumlah penerimaan mahasiswa baru di PTS di lingkungan Kopertis Wilayah VI Jawa Tengah hingga tahun 2012. Bidang ilmu dan program studi tertentu mengalami kenaikan yang relatif tinggi, khususnya di bidang teknologi informasi, kependidikan dan kesehatan. Ada suatu pendapat mengenai program studi dengan jenjang pendidikan program tertentu cenderung dipilih oleh calon mahasiswa dengan gender tertentu pula, sedangkan yang lain berpendapat bidang ilmu yang lebih cenderung menjadi prioritas utama sebelum memilih program studi dengan jenjang pendidikannya.

Berdasarkan hal tersebut maka dengan menggunakan teknik klasifikasi pada data mining dapat diketahui kecenderungan calon mahasiswa dalam memilih jenjang pendidikan di perguruan tinggi, khususnya mengenai program studi tertentu yang banyak peminatnya, berdasarkan rumpun bidang ilmu dan data jenis kelamin.

Kata kunci : Data Mining, Klasifikasi, Bidang Ilmu, Jenis Kelamin

PENDAHULUAN

Jumlah Penerimaan Mahasiswa Baru atau biasa disingkat PMB di Perguruan Tinggi Swasta di lingkungan Jawa Tengah hingga tahun 2012 terus meningkat meskipun ada yang mengalami kenaikan dan penurunan. Beberapa program studi bahkan harus ditutup karena jumlah peminatnya pada beberapa tahun terakhir terus mengalami penurunan. Sementara beberapa perguruan tinggi justru mengusulkan program studi baru yang menurut trend jumlah selalu meningkat setiap tahunnya. Ada beberapa pendapat yang menyatakan adanya kecenderungan calon mahasiswa dengan gender tertentu memilih program studi dengan jenjang pendidikan tertentu pula. Namun pendapat lain menyatakan bahwa program studi dipilih oleh calon mahasiswa setelah mereka memilih bidang ilmu yang diminatinya, barulah jenjang pendidikan menjadi alternatif mereka selanjutnya. Berdasarkan hal tersebut maka perlu diketahui mana yang lebih mempengaruhi

dalam memilih jenjang pendidikan program studi antara jenis kelamin dan bidang ilmu.

Berdasarkan data yang ada, jumlah Perguruan Tinggi Swasta (PTS) di wilayah VI Jawa Tengah adalah seperti terlihat pada Tabel 1 berikut ini :

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

No.	Jenis	Jumlah
1	Akademi	111
2	Politeknik	21
3	Sekolah Tinggi	77
4	Institut	2
5	Universitas	34
	Jumlah	245

Dengan jumlah program studi sebanyak 153 buah seperti terlihat pada tabel 2 berikut ini :

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

No.	Program Studi
1	ADMINISTRASI BISNIS
2	ADMINISTRASI PERADILAN
3	ADMINISTRASI PERPAJAKAN
4	ADMINISTRASI RUMAH SAKIT
5	AGRIBISNIS
6	AGRONOMI
7	AGROTEKNOLOGI
8	AKUNTANSI
9	ANALIS FARMASI DAN MAKANAN
	...
	...
	...
149	TEKNOLOGI HASIL PERTANIAN
150	TEKNOLOGI PANGAN
151	TEKNOLOGI PENDIDIKAN
152	TEKNOLOGI PRODUKSI PERTANIAN
153	USAHA PERJALANAN WISATA

Apabila program studi seperti pada Tabel 2 dikelompokkan berdasarkan bilangan ilmu maka diperoleh 16 bidang ilmu seperti terlihat pada Tabel 3.

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

Selanjutnya adalah tabel 4 yang berisi tentang kelompok bidang ilmu berdasarkan 5 (empat) terbanyak di Kopertis Wilayah VI Jawa Tengah, seperti berikut ini :

Tabel 4. Jumlah Mahasiswa Baru di Kopertis Wilayah VI Jawa Tengah berdasarkan 10 Program Studi Terbanyak

No.	Program Studi	Jenjang	Jml
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
5	PENDIDIKAN BAHASA INGGRIS	S-1	20.007
6	TEKNIK INFORMATIKA	S-1	18.859
7	KEPERAWATAN	D-3	17.165
8	PENDIDIKAN BAHASA DAN SASTRA INDONESIA	S-1	15.669
9	PENDIDIKAN MATEMATIKA	S-1	11.648
10	SISTEM	S-1	11.440

INFORMASI	
Jumlah	225.097

Dari tabel 4. dapat dikelompokkan berdasar bidang ilmu seperti terlihat pada tabel 5 di bawah ini :

Tabel 5. Jumlah Mahasiswa Baru Berdasarkan Bidang Ilmu Dari 10 Program Studi Terbanyak

No.	Bidang Ilmu	Jumlah
1	Ekonomi	79.470
2	Kependidikan	47.324
3	Kesehatan	41.128
4	Komputer	30.299
5	Hukum	26.876
Jumlah		225.097

Sedangkan berdasarkan jenjang pendidikan Strata-1 (S1) dan Diploma-III (D3), di Kopertis Wilayah VI Jawa Tengah terdistribusi seperti terlihat pada tabel 6.

Tabel 6. Jumlah Mahasiswa Baru pada Jenjang Pendidikan S1 dan D3 Berdasarkan 10 Program Studi Terbanyak

No.	Jenjang Pendidikan	Jumlah Progdi	Jumlah Mhs Baru
1	S1	8	183.969
2	D3	2	41.128
Jumlah		10	225.097

Dengan jumlah mahasiswa baru laki-laki dan perempuan seperti terlihat pada tabel 7 berikut ini :

Tabel 7. Jumlah Mahasiswa Baru Berdasarkan Jenis Kelamin Dari 10 Program Studi Terbanyak

No.	Jenis Kelamin	Jumlah Progdi
1	Laki-laki	106.260
2	Perempuan	118.837

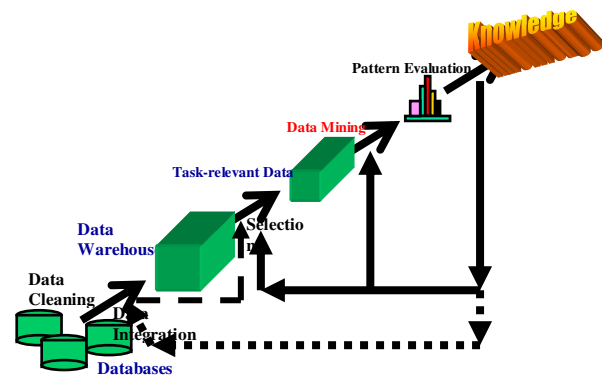
Jumlah	225.097
---------------	----------------

Dari data-data yang telah disajikan dalam tabel tampak adanya suatu trend jumlah kenaikan pada jenjang pendidikan S1 dan pada beberapa bidang ilmu, namun demikian ada juga kenaikan pada jenjang pendidikan D3 untuk program studi tertentu, sedangkan yang lain mengalami kondisi stabil dan penurunan.

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 :

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).

Association Analysis

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

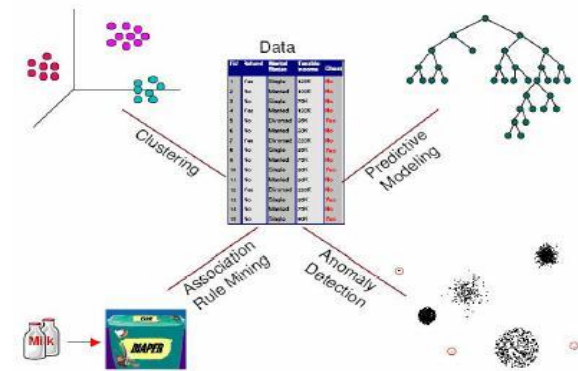
Cluster Analysis

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.

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)

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).

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).

Klasifikasi

Klasifikasi merupakan koleksi dari suatu record berupa training data set, dimana setiap record berisi seperangkat atribut dan salah satu atribut adalah suatu kelas. Yang harus dilakukan adalah mencari model untuk atribut kelas sebagai fungsi dari nilai atribut yang lain. Tujuannya adalah mendapatkan suatu kelas yang seakurat mungkin dari catatan record sebelumnya yang tidak terlihat. Satu set data tes dipersiapkan untuk menentukan keakuratan model dan sekaligus validasinya.

Algoritma Teknik Klasifikasi

Proses klasifikasi menggunakan algoritma Hunt karena algoritma ini merupakan salah satu teknik klasifikasi yang menggambarkan struktur pohon keputusan dimana hasilnya dapat menunjukkan kecenderungan suatu nilai terhadap hasil atau target yang diinginkan. Proses awal dimulai dengan menentukan atribut mana dari data yang akan digunakan untuk proses klasifikasi, yaitu Bidang Ilmu (Ekonomi, Kependidikan, Kesehatan dan Komputer), Jenjang Pendidikan (S1 dan D3) dan Jenis Kelamin (Laki-laki dan Perempuan). Kemudian dari ketiga atribut tersebut diolah dengan rumus klasifikasi untuk menentukan atribut mana yang paling dominan sebagai simpul utama. Masing-masing atribut diproses untuk mengetahui *Gain* terbesar yang akan menjadi simpul utama. Adapun rumus untuk menentukan *Gain* adalah sebagai berikut :

- Ukuran informasi Info :

$$\text{Info}(S_1, S_2, \dots, S_m) = -\sum S_i/S \log_2 S_i/S \dots \dots \quad (4.1)$$

S_1, S_2, \dots, S_m adalah susunan nilai-nilai dari suatu kelas

S_i adalah nilai dari setiap kelas yang ada

S adalah jumlah nilai dari semua kelas yang ada

- Entropi dari atribut A dengan nilai-nilai $\{a_1, a_2, \dots, a_v\}$

$$\text{Entropy}(A) = \sum ((S_{i_j} + \dots + S_{m_j})/S) \text{Info}(S_{i_j}, \dots, S_{m_j}) \dots \dots \quad (4.2)$$

A adalah suatu atribut

S_{i_j}, \dots, S_{m_j} adalah nilai dari masing-masing atribut

- Informasi yang diperoleh melalui pencabangan pada atribut A

$$\text{Gain}(A) = \text{Info}(S_1, S_2, \dots, S_m) - \text{Entropy}(A) \dots \dots \quad (4.3)$$

A adalah suatu atribut

S_1, S_2, \dots, S_m adalah nilai dari masing-masing Info

(Tan, dkk, 2006)

Proses Perhitungan Gain Bidang Ilmu terhadap Jenjang Pendidikan

BIDANG ILMU

$$\text{Info}([183969, 41128]) = \text{entropy}(183969/225097, 41128/225097)$$

		pembagian	log
S1 =	183969	0,817	-0,291
D3 =	41128	0,183	-2,452
JML =	225097		
	2	hasil = $-(0,817 * -0,291) - (0,183 * -2,452)$	0,686

EKONOMI

$$\text{Info}([79470, 0]) = \text{entropy}(79470/79470, 0/79470)$$

		pembagian	log

S1 =	79470	1,000	0,000
D3 =	0	0,000	0,000
JML =	79470	225097	
	2	hasil = +- (1,000*0,000)- (0,000*0,000)	0,000
		(79470/225097) * hasil =	0,000

KEPENDIDIKAN

Info([47324,0])=entropy(47324 /47324,0/47324)

		pembagian	log
S1 =	47324	1,000	0,000
D3 =	0	0,000	0,000
JML =	47324	225097	
	2	hasil = +- (1,000*0,000)- (0,000*0,000)	0,000
		(47324/225097) * hasil =	0,000

KESEHATAN

Info([0,41128])=entropy(0/41128,41128/41128)

		pembagian	log
S1 =	0	0,000	0,000
D3 =	41128	1,000	0,000
JML =	41128	225097	
	2	hasil = +- (0,000*0,000)- (1,000*0,000)	0,000
		(41128/225097) * hasil =	0,000

KOMPUTER

Info([30299,0])=entropy(30299 /30299,0/30299)

		pembagian	log
S1 =	30299	1,000	0,000
D3 =	0	0,000	0,000
JML =	30299	225097	
	2	hasil = +- (1,000*0,000)- (0,000*0,000)	0,000
		(30299/225097) * hasil =	0,000

HUKUM

Info([26876,0])=entropy(26876 /26876,0/26876)

		pembagian	log
S1 =	26876	1,000	0,000
D3 =	0	0,000	0,000
JML =	26876	225097	
	2	hasil = +- (1,000*0,000)- (0,000*0,000)	0,000
		(26876/225097) * hasil =	0,000

JUMLAH INFO **0,000**

GAIN(BIDANG ILMU) 0,686

Proses Perhitungan Gain Jenis Kelamin Terhadap Jenjang Pendidikan

JENIS KELAMIN

Info([183969,41128])=entropy(183969/225097,41128/225097)

		pembagian	log
S1 =	183969	0,817	-0,291

D3 =	41128	0,183	-2,452
JML =	225097		
	2	hasil = $\pm(0,817*0,291)-(0,183*2,452)$	0,686

LAKI-LAKI

$Info([101194,5066])=entropy(101194/106260,5066/106260)$

		pembagian	log
S1 =	101194	0,952	-0,070
D3 =	5066	0,048	-4,391
JML =	106260	225097	
	2	hasil = $\pm(0,952*0,070)-(0,048*4,391)$	0,276
		$(101194/225097) * \text{hasil} =$	0,130

PEREMPUAN

$Info([82775,36062])=entropy(82775/118837,36062/118837)$

		pembagian	log
S1 =	82775	0,697	-0,522
D3 =	36062	0,303	-1,720
JML =	118837	225097	
	2	hasil = $\pm(0,697*0,522)-(0,303*1,720)$	0,885
		$(118837/225097) * \text{hasil} =$	0,467

JUMLAH INFO **0,598**

GAIN(JENIS KELAMIN) 0,088

Karena Gain Bidang Ilmu dengan nilai 0,686 lebih besar dari Gain Jenis Kelamin

dengan nilai 0,088 maka Bidang Ilmu dipilih sebagai faktor yang lebih menentukan daripada Jenis Kelamin pada calon mahasiswa dalam memilih Jenjang Pendidikan S1 atau D3.

PEMBAHASAN DAN VISUALISASI HASIL KLASIFIKASI

Berdasarkan hasil klasifikasi pada Bab III sebelumnya yaitu berdasarkan bidang ilmu dan jenis kelamin terhadap jenjang pendidikan program studi nampak bahwa calon mahasiswa lebih menentukan pilihan bidang ilmu terlebih dahulu baru kepada pilihan jenjang pendidikan, hanya kemudian jenis kelamin juga menentukan kecenderungan berikutnya seperti laki-laki lebih banyak di bidang ilmu ekonomi, komputer dan hukum sementara perempuan lebih banyak di bidang ilmu kependidikan dan kesehatan.

Berikut ini adalah visualisasi hasil klasifikasi bidang ilmu dan jenis kelamin terhadap jenjang pendidikan seperti terlihat pada gambar 3.

Gambar 3. Visualisasi klasifikasi bidang ilmu dan jenis kelamin terhadap jenjang pendidikan

BIDANG ILMU				
10,46%	15,57%	20,41%	22,64%	30,92%
HUKUM	KOMPUTER	KESEHATAN	KEPENDIDIKAN	EKONOMI
26.876	30.299	41.128	47.324	79.47

JENIS KELAMIN									
20,71%	79,29%	31,90%	68,10%	82,31%	17,69%	65,94%	34,06%	47,44%	52,56%
P	L	P	L	P	L	P	L	P	L
5.565	21.311	8.299	22	36.062	5.066	31.207	16.117	37.704	41.766

JENJANG PENDIDIKAN																			
0%	100%	0%	100%	35,00%	65,00%	19,30%	80,70%	83,49%	16,51%	54,57%	45,43%	0%	100%	0%	100%	0%	100%	0%	100%
D3	S1	D3	S1	D3	S1	D3	S1	D3	S1	D3	S1	D3	S1	D3	S1	D3	S1	D3	S1
0	5.565	0	21.311	0	8.299	0	22	36.062	0	5.066	0	0	31.207	0	16.117	0	37.704	0	41.766

KESIMPULAN

1. Teknik klasifikasi pada data mahasiswa baru pada PTS di lingkungan Kopertis Wilayah VI Jawa Tengah berdasarkan jumlah yang registrasi menghasilkan informasi bahwa calon mahasiswa baru lebih dipengaruhi oleh bidang ilmu daripada jenis kelamin dalam memilih program studi yang diminatinya
2. Hasil klasifikasi menunjukkan bahwa beberapa bidang ilmu dan program studi tertentu memang didominasi oleh jenis kelamin tertentu, seperti laki-laki pada bidang ilmu ekonomi, komputer dan hukum, sedangkan perempuan mendominasi pada bidang ilmu kependidikan dan kesehatan
3. Telah terjadi pergeseran peminatan antara pilihan jenjang pendidikan program S1 dan D3 yang bisa saja dipengaruhi oleh beberapa faktor seperti dunia kerja, wilayah serta kebijakan dalam dunia pendidikan tinggi itu sendiri

SARAN

1. Penelitian ini akan menjadi lebih baik jika bisa diperoleh data tahun akademik terbaru disertai dengan beberapa atribut lain yang lebih lengkap
2. Cakupan data pada periode 5 tahun terakhir tampaknya lebih bisa mewakili apa yang terjadi pada masa sekarang dan sebagai penggambaran di masa yang akan datang daripada kurun waktu sebelumnya
3. Perlu beberapa penelitian dengan beberapa metode lain untuk lebih melengkapi hasil penelitian sebelumnya dalam upaya mendapatkan hasil penelitian terpadu yang optimal sebagai dasar pengambilan keputusan

DAFTAR PUSTAKA

- Berry, M., J., A., & Linoff, G., S., (2000). *Mastering Data Mining*. New York: Wiley.
- Budiarti Andina, Suchayo Yudho Giri dan Ruldeviyani Yova (2006). *Studi Karakteristik Kelulusan Peserta Didik*

dengan Teknik Clustering, Universitas Indonesia

- Chong Ho Yu, Samuel Digangi, Angel Kay Jannasch-Pennell & Charles Kaprolet (2008). *Profiling Students Who Take Online Courses Using Data Mining Methods*, Arizona State University
- Cristobal Romero, Sebastian Ventura, Pedro G. Expejo dan Cesar Hervas (2008). *Data Mining Algorithms to Classify Students*, Cordoba University
- Edelstein, H., A. (1999). *Introduction to Data Mining and Knowledge Discovery (3rd ed)*. Potomac, MD: Two Crows Corp.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery & Data Mining*. Cambridge, MA: MIT Press.
- Han, J., Kamber, M. (2000). *Data mining: Concepts and Techniques*. New York: Morgan-Kaufman.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. New York: Springer.
- James Otto, Douglas Sanford & William Wagner (2005). *Analisis Of Online Student Ratings Of University Faculty*, Towson University & Villanova University
- Pregibon, D. (1997). *Data Mining*. Statistical Computing and Graphics, 7, 8.
- Qasem A. Al-Radaideh, Emad M. Al-Shawakfa dan Mustafa I. Al-Najjar (2006). *Mining Student Data Using Decision Trees*, Yarmouk University
- Senol Zafer Erdogan & Mehpare Timor (2005). *A Data Mining Application In A Student Database*, Maltepe University & Istanbul University.
- Weiss, S. M., & Indurkha, N. (1997). *Predictive Data Mining: A practical guide*. New York: Morgan-Kaufman.

Westphal, C., Blaxton, T. (1998). *Data Mining Solutions*. New York: Wiley.

Witten, I. H., & Frank, E. (2000). *Data Mining*. New York: Morgan-Kaufmann.