

DATA WAREHOUSE DAN DATA MINING PENDIDIKAN TINGGI: STUDI KASUS KATEGORI UNDIR DIRI DI UNIVERSITAS BINA NUSANTARA

Sulistyo Heripracoyo

Information Systems Department, School of Information Systems, Binus University
Jl. K.H. Syahdan No. 9, Palmerah, Jakarta Barat 11480
hpracoyo@binus.edu

ABSTRACT

Data warehouse and data mining is used to extract useful information and has a specific meaning and to develop a real relationship between some variables stored in the data/data warehouse. A data warehouse is appropriately designed and added a requirement to provide appropriate data and is useful in making better decisions. Hardware and software facilitate adequate access to such data, analyze and display the results interactively. Data mining software is a highly effective tool that can be used to interrogate the data contained in the data warehouse in order to find a relationship (Neary 1999). This study conducts some literature studies applies some models and case studies in a higher education institution, in terms of the benefits, functions and development. The case study conducted is objected to see the trend and prediction of the number of students who drop out (DO).

Keywords: data warehouse, data mining, prediction, drop out

ABSTRAK

Data warehouse dan data mining dipergunakan untuk mengekstraksi informasi yang bermanfaat dan mempunyai arti tertentu serta untuk mengembangkan suatu hubungan yang nyata di antara beberapa variabel yang tersimpan di dalam data. Suatu data warehouse yang dirancang dan ditambahkan dengan tepat dapat memberikan suatu kebutuhan data yang sesuai dan berguna pada pengambilan keputusan yang lebih baik. Bantuan perangkat keras dan perangkat lunak yang memadai memudahkan akses terhadap data tersebut, menganalisis dan menampilkan hasilnya secara interaktif. Perangkat lunak data mining merupakan piranti yang sangat efektif yang dapat dipergunakan untuk menginterogasi jumlah data yang terkandung dalam data warehouse agar dapat menemukan hubungan (Neary 1999). Dalam paper ini akan dilakukan beberapa studi literatur dan menerapkan beberapa model dan studi kasusnya dalam institusi pendidikan tinggi, ditinjau dari manfaat, fungsi dan pengembangannya. Studi kasus yang dilakukan adalah melihat trend dan prediksi dari jumlah mahasiswa yang melakukan undur diri atau drop out (DO).

Kata kunci: data warehouse, data mining, prediksi, drop out

PENDAHULUAN

Secara sederhana, *data warehouse* (DW) adalah kumpulan data yang dihasilkan untuk mendukung pengambilan keputusan. *Data warehouse* adalah suatu *repository* terpusat dari data histori yang menyediakan suatu platform terintegrasi yang dapat digunakan untuk menganalisis data histori. Menurut Inmon (2002, p31-34), *data warehouse* mempunyai empat karakteristik yang penting, yaitu: *subject-oriented*, *integrated*, *nonvolatile* dan *time-variant*. *Subject-Oriented* adalah data yang ada di dalam *data warehouse* berorientasi subject, bukan berorientasi pada transaksional (aplikasi). *Integrated* adalah data di dalam *data warehouse* berasal dari banyak sumber yang berbeda/terpisah untuk diintegrasikan ke dalam *data warehouse*. *Nonvolatile* adalah data di dalam *data warehouse* hanya dilakukan update secara statis dengan hanya melakukan insert data (*snapshot*) dan tidak melakukan perubahan/update terhadap data warehouse. *Time-variant* adalah data di dalam *data warehouse* hanya akan akurat pada salah satu moment waktu. Lima tahap pengambilan keputusan di dalam *data warehouse* adalah: *reporting* (pelaporan), *analyzing* (menganalisis), *predicting* (memprediksi), *operationalizing* (operasionalisasi), dan *active warehousing* (Stephen Brobst and Joe Rarey (2003) dalam artikelnya Akintola et al., 2011).

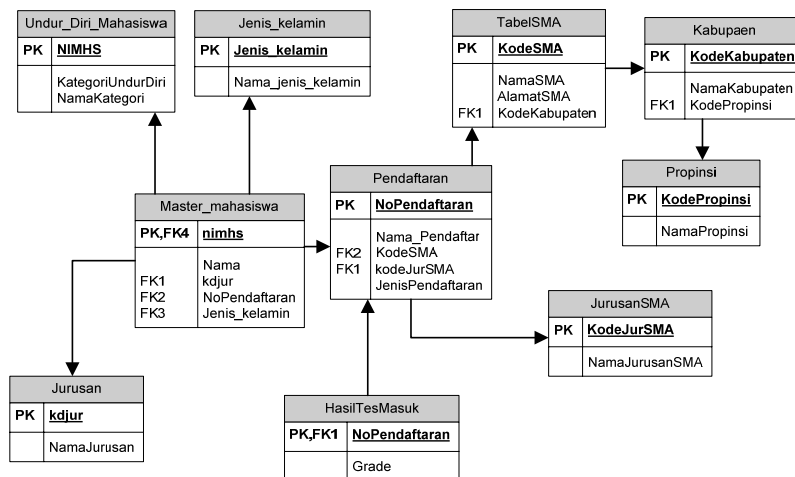
Data mining digunakan saat ini terutama oleh perusahaan dengan fokus konsumen yang kuat – ritel, organisasi keuangan, komunikasi, dan pemasaran (Akintola et al., 2011). Hal ini memungkinkan perusahaan-perusahaan untuk menentukan hubungan di antara faktor "internal" seperti harga, penempatan produk, atau keterampilan staf, dan faktor "eksternal" seperti indikator ekonomi, persaingan, dan demografi pelanggan. itu memungkinkan mereka untuk menentukan dampak pada penjualan, kepuasan pelanggan, dan keuntungan perusahaan. Dalam prakteknya, tujuan utama *data mining* cenderung untuk *prediction* (prediksi) dan *description* (deskripsi) (Kantardzic (2003, p.2). Prediksi mencakup menggunakan beberapa variabel atau *fields* dalam data set untuk memprediksi nilai yang tidak diketahui atau nilai kedepan (*future*) dari variabel yang penting. Deskripsi berfokus pada menemukan pola yang menguraikan data yang dapat diinterpretasikan oleh manusia, yang selanjutnya aktivitas *data mining* dapat dibagi dua kategori, yaitu *predictive data mining* yang menghasilkan model dari sistem yang diuraikan dengan data set yang diberikan, atau *descriptive data mining* yang menghasilkan informasi baru, *nontrivial* didasarkan pada data set yang tersedia. Tugas (*tasks*) data-mining utama dibagi menjadi beberapa, antara lain: (1) *classification* – menemukan dari suatu fungsi pembelajaran prediktif yang mengklasifikasi suatu item data ke dalam satu dari beberapa kelas yang didefinisikan sebelumnya; (2) *regression* – menemukan dari fungsi pembelajaran prediktif, yang memetakan suatu item data ke dalam suatu variable prediksi nilai-nyata; (3) *clustering* – suatu tugas deskriptif umum di mana salah satu mencari untuk mengidentifikasi suatu set kategori yang terbatas atau *cluster* untuk menguraikan data tersebut; (4) *summarization* – suatu tugas deskriptif tambahan yang mencakup metode untuk menemukannya suatu deskripsi yang erat untuk suatu set (atau subset) dari data; (5) *dependency modeling* – menemukan suatu model lokal yang menguraikan dependensi nyata di antara variabel atau di antara nilai dari fitur dalam suatu data set atau dalam suatu bagian dari data set; (6) *change and deviation detection* – menemukan perubahan paling signifikan dalam data set.

Proses sehari-hari dapat diatasi dengan sistem *online transaction processing* (OLTP) dengan baik. Akan tetapi sistem OLTP tidak cukup ketika dibutuhkan untuk mendukung keputusan strategis. Kekurangan sistem OLTP adalah bahwa data historis dari OLTP tersebut tidak secara komprehensi dapat mendukung informasi bisnisnya. Tujuan membuat *data warehouse* dalam pendidikan adalah terutama untuk mengetahui dan menjawab beberapa hal tentang: bagaimana tren penerimaan siswa, pendaftaran mahasiswa, jadwal dosen, jumlah pendaftaran tahunan, dll. Jenis pertanyaan memerlukan banyak data historis untuk menghasilkannya di mana sistem sistem OLTP tidak dapat mendukung.

METODE

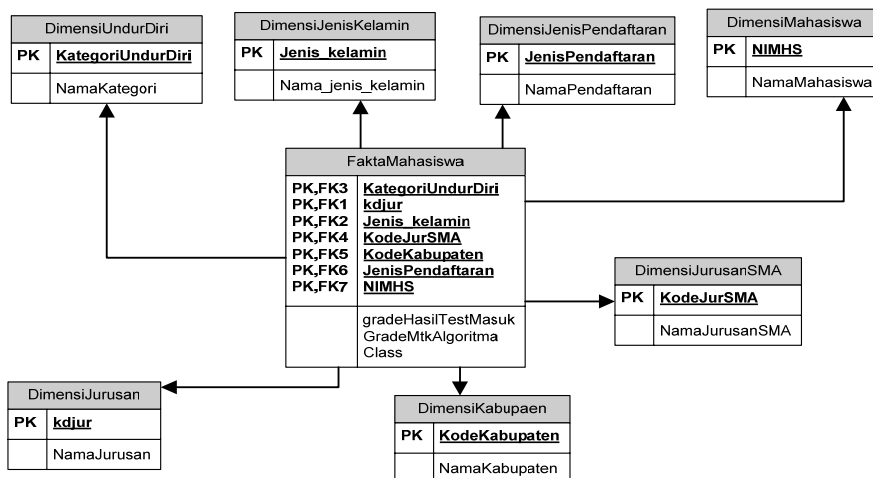
Dalam studi kasus ini akan dilakukan analisis terhadap jumlah mahasiswa yang melakukan undur diri, *drop out* (DO) dilihat dari beberapa atribut atau variabel yang terkait, yaitu jenis kelamin, jurusan yang dipilih, jurusan sma, hasil test masuk, jenis pendaftaran, kategori undur diri/DO, *grade* matakuliah algoritma.

Data operasional atau Online Transaction Processing (OLTP) adalah data transaksi operasional dari hari ke hari untuk proses secara rutin. Dalam penelitian ini data logikal dari proses OLTP dapat dilihat pada Gambar 1.



Gambar 1. Data Logikal dari proses OLTP.

Data warehouse umumnya dibuat dengan satu tabel fakta terpusat yang besar, dan beberapa tabel dimensi yang lebih kecil. Konfigurasi ini biasanya disebut dengan skema bintang. Suatu skema bintang digunakan dalam penelitian ini (Gambar 2). Di bawah ini adalah diagram hubungan di antara tabel fakta dan tabel dimensi. Data dari proses OLTP akan dilakukan ETL (Extract, Transform, Load) ke dalam tabel fakta dan dimensi.



Gambar 2. Skema bintang.

Dalam studi ini digunakan satu tabel fakta dan tujuh tabel dimensi. Tabel fakta berisi data measure yang berupa grade hasil test masuk dan grade matakuliah algoritma beserta dengan *class* yang berisi data keterangan tentang proses DO. Dalam studi yang dilakukan jenis *class* yang ditampilkan adalah sebagai berikut (Tabel 2):

Tabel 2
Keterangan Pengelompokan Class

Kategori Undur Diri	Keterangan Kategori	Class
1	Diberhentikan	Worst
1	Drop Out (DO)	Worst
1	PTTKK	Worst
1	Tidak Lulus SMU	Worst
2	Pindah Jurusan/Jenjang/Renim	F-PindahJur
2	Pindah ke PT Lain	F-PT Lain
2	Terima UMPTN	F-UMPTN
2	Undur Diri	Bad-Und
2	Undur Diri karena DO	Bad-Und
2	Undur Diri karena Keuangan	Bad-Keu
2	Undur Diri Sepihak dari BINUS	Bad-Und

Dari pembagian Class dalam Tabel 1 di atas, dapat dijelaskan sebagai berikut, kode undur diri (1) merupakan undur diri dengan kategori DO, Pelanggaran (PTTKK/Peraturan Tata Tertib Kehidupan Kampus), Diberhentikan dan tidak Lulus SMA. Kategori tersebut dimasukkan/dikelompokkan sebagai *Worst*. Selanjutnya jenis undur diri (2) terbagi lagi menjadi beberapa kategori dengan keterangan karena adanya beberapa hal, antara lain Masalah keuangan (*Bad-Keu*), Karena undur Diri (*Bad-Und*), karena diterima di Perguruan Tinggi Negeri (*F-UMPTN*), pindah ke Perguruan Tinggi lain (*F-PT Lain*) dan Pindah ke jurusan lain (*F-PindahJur*).

Secara Deskriptif data yang digunakan dapat dilihat dari data di bawah ini, data yang digunakan terdiri dari 1396 record, yang meliputi data pada tahun 2005, 2006 dan 2007. Jurusan yang dipilih adalah jurusan yang merupakan mayoritas jurusan yang ada di Binus University, yaitu Jurusan Sistem Informasi, Jurusan Teknik Informatika dan Jurusan Komputerisasi Akuntansi.

Count of nimhs							Grade						Grand Total
priod	jnkel	nmjur	kategori	jenis	grade_mtk_algo	class	?	A	B	C	D		
2005	1	Komputerisasi Akuntansi	1	Diberhentikan	1	Worst		2	12	12	3	29	
					2	Worst			2	1	1	4	
				Drop Out (DO)	1	Worst		7	8	14	5	34	
				Tidak Lulus SMU	?	Worst			1		1	2	
			1 Total					9	23	27	10	69	
			2	Pindah ke PT Lain	?	F-PT Lain		1				1	
				Undur Diri	?	Bad-Und			2	4		6	

					Bad-Und	2	19	8	1	30
					Bad-Und		1			1
		Undur karena DO	Diri	1	Bad-Keu	1	1		2	4
		Undur Sepihak BINUS	Diri dari	1	Bad-Keu	1	3	4	2	10
				2	Bad-Keu		1			1
		2 Total				5	27	16	5	53
		Komputerisasi Akuntansi Total				14	50	43	15	122
	Sistem Informasi	1	Diberhentikan	?	Worst	6	7	5	1	19
			Drop Out (DO)	?	Worst	2	4		1	7
		1 Total				8	11	5	2	26
		2	Pindah ke PT Lain	?	F-PT Lain	1				1
			Terima UMPTN	?	F-UMPTN	1	1			2
			Undur Diri	?	Bad-Und	4	6	7		17
			Undur karena DO	?	Bad-Keu		3	1		4
			Undur Sepihak BINUS	?	Bad-Keu		3	5		8
		2 Total				6	13	13		32
		Sistem Informasi Total				14	24	18	2	58
	Teknik Informatika	1	Diberhentikan	?	Worst				1	1
				1	Worst	5	14	15	8	42
				2	Worst	2	1		1	4
			Drop Out (DO)	1	Worst	8	15	13	10	46
				2	Worst	1				1
		1 Total				16	30	29	19	94
		2	Pindah Jurusan/Jenjang/Renim	?	F-Pindah Jur	1	1			2
				1	F-Pindah Jur				1	1
			Pindah ke PT Lain	?	F-PT Lain	1	1			2
			Terima UMPTN	?	F-UMPTN	2	19	5		26
			Undur Diri	?	Bad-Und	10	5			15
				1	Bad-Und	10	18	11	3	42
				2	Bad-		1			1

al											
			2	Terima UMPTN	?	F- UMPT N	5	1			6
				Undur Diri	?	Bad- Und	3		1		4
					1	Bad- Und	8	2	1	1	12
					2	Bad- Und			1		1
				Undur Diri karena DO	1	Bad- Keu			1		1
				Undur Sepihak BINUS	1	Bad- Keu	1			2	3
			2 Total				17	5	2	3	27
				Teknik Informatika Total			19	6	5	4	34
			2 Total				1	38	29	23	8
2005							12	16	13		
Total							3	3	9	6	53
2006	1	Komputerisasi Akuntansi	1	Diberhentikan	?	Worst	1	1	1		3
					1	Worst	2	13	11	3	29
				Tidak Lulus SMU	?	Worst			1		1
			1 Total				3	14	13	3	33
			2	Terima UMPTN	?	F- UMPT N	1				1
				Undur Diri	?	Bad- Und	3	2	1		6
					1	Bad- Und	2	2	8	2	14
				Undur Diri karena DO	1	Bad- Keu				1	1
				Undur Sepihak BINUS	1	Bad- Keu	1		1		2
			2 Total				7	4	10	3	24
				Komputerisasi Akuntansi Total			10	18	23	6	57
				Sistem Informasi	1	Diberhentikan	?	Worst	12	43	33
				Tidak Lulus SMU	?	Worst			1		1
			1 Total				12	43	34	6	95
			2	Pindah Jurusan/Jenjan g/Renim	?	F- Pindah Jur		1		1	2
				Terima UMPTN	?	F- UMPT N	8	3			11
				Undur Diri	?	Bad-	5	20	26	5	56

					Und					
		Undur	Diri		Bad-					
		karena DO	?		Keu	2	3	2	1	8
		Undur	Diri		Bad-					
		Sepihak	dari		Keu		2	1	1	4
		BINUS	?							
	2									
	Tot					15	29	29	8	81
	al									
	Sistem									
	Informasi									
	Total					27	72	63	14	176
	Teknik									
	Informatika	1	Diberhentikan	1	Worst	6	21	20	10	57
				2	Worst		3			3
				3	Worst	1	1			2
		Tidak	Lulus		Worst					
		SMU	?				2		1	3
	1									
	Tot					7	27	20	11	65
	al									
		Pindah			F-					
		Jurusan/Jenjan			Pindah					
		g/Renim	?		Jur				1	1
	2									
		Terima			F-					
		UMPTN	?		UMPT	5	38	12	2	57
					N					
				1	F-			2		2
					UMPT					
					N					
		Undur	Diri		Bad-					
		Diri	?		Und	1	1		1	3
				1	Bad-	2	17	26	19	8
					Und					72
				2	Bad-	1				1
					Und					
				3	Bad-	1				1
					Und					
		Undur	Diri		Bad-					
		karena DO	1		Keu	1	2	4	1	8
				2	Bad-	1			1	2
					Keu					
		Undur	Diri		Bad-					
		Sepihak	dari		Keu			1	1	2
		BINUS	1							
	2									
	Tot					7	60	44	26	12
	al									149
	Teknik									
	Informatika					7	67	71	46	23
	Total									214
	1						10	16	13	
	Total					7	4	1	2	43
										447
	Komputerisasi									
	Akuntansi	1	Diberhentikan	1	Worst	1	2	8		11
				2	Worst	1				1
		Tidak	Lulus		Worst				1	1
		SMU	?							
	1									
	Tot					2	2	9		13
	al									
		Pindah			F-					
		Jurusan/Jenjan			Pindah					
		g/Renim	?		Jur				1	1
	2									

		Terima UMPTN	?	F-UMPTN	2	3	1		6
		Undur Diri	?	Bad-Und		1	2	2	1
			1	Bad-Und		1	2	1	4
			2	Bad-Und				1	1
			3	Bad-Und		1			1
		Undur Diri karena DO	1	Bad-Keu				1	1
		Undur Diri Sepihak BINUS	1	Bad-Keu					1
	2	Total			2	6	5	5	3
		Komputerisasi Akuntansi Total			2	8	7	14	3
	1	Sistem Informasi	Diberhentikan	?	Worst	3	3	3	9
			Tidak Lulus SMU	?	Worst			1	1
	1	Total				3	4	3	10
	2	Pindah Jurusan/Jenjang/Renim	?	F-Pindah Jur			1		1
		Terima UMPTN	?	F-UMPTN		2	2		4
		Undur Diri	?	Bad-Und	2		3	4	3
		Undur Diri karena DO	?	Bad-Keu				1	1
	2	Total			2	2	6	5	3
		Sistem Informasi Total			2	5	10	8	3
	1	Teknik Informatika	Diberhentikan	1	Worst	1	2	2	2
			Tidak Lulus SMU	?	Worst			1	1
	1	Total				1	2	3	2
	2	Pindah Jurusan/Jenjang/Renim	?	F-Pindah Jur				1	1
			1	F-Pindah Jur			1		1
		Terima UMPTN	?	F-UMPTN	4	17			21
		Undur Diri	?	Bad-Und		1			1
			1	Bad-Und	1	4	6	2	1
		Undur Diri	2	Bad-			1		1

		karena DO			Keu												
		2 Total															
					5	22	8	3	2	40							
		Teknik															
		Informatika Total			5	23	10	6	4	48							
		2 Total															
					9	36	27	28	10	110							
2006 Total					16	14 0	18 8	16 0	53	557							
2007		Komputerisasi Akuntansi															
1	1	Diberhentikan	?	Worst					1	1							
		PTTKK			1	Worst				1	1						
		1 Total															
									1	1	2						
		2 Terima UMPTN			?	F- UMPT N				1	1						
		Undur Diri			?	Bad- Und				2	3	5					
					1	Bad- Und				1	2	2	5				
		Undur Diri karena DO			1	Bad- Keu				1		1					
		Undur Diri karena Keuangan			?	Bad- Keu				1	1	2					
		Undur Diri Sepihak BINUS			1	Bad- Keu				2		2					
		2 Total															
									4	8	2	2	16				
		Komputerisasi Akuntansi Total															
					5	8	2	3	18								
		Sistem Informasi			1	Diberhentikan				?	Worst	2	1	3			
		1 Total															
									2	1	3						
		2 Pindah Jurusan/Jenjang/ Renim			?	F- Pindah Jur				1	1						
		Terima UMPTN			?	F- UMPT N				5	3	8					
		Undur Diri			?	Bad- Und				14	16	9	2	41			
		Undur Diri karena DO			?	Bad- Keu				2	5	4	1	12			
		Undur Diri karena Keuangan			?	Bad- Keu				1	5	3	1	10			
		Undur Diri Sepihak BINUS			?	Bad- Keu				4	14	3	1	22			
		2 Total															
									26	44	19	5	94				
		Sistem Informasi Total															
					26	46	20	5	97								
		Teknik			1	Diberhentikan				1	Worst	2	2	1	2	1	8

Informatika										
		1	Tot							8
		2	Terima	?	F-					
			UMPTN		UMPT	4	30	11	2	48
			Undur Diri	?	Bad-	2	3	7	2	14
				1	Und	1	23	13	9	51
				2	Und		4	3		7
				3	Und		1			1
			Undur Diri	1	Bad-					
			karena DO		Keu		2	5	2	10
			Undur Diri	1	Bad-					
			karena		Keu		6	8	2	19
			Keuangan							
			Undur Diri	1	Bad-					
			Sepihak		Keu		7	7	1	18
			BINUS							
		2	Tot			7	76	54	18	168
			Teknik							
			Informatika Total			9	78	55	20	176
1							10	10		
Total						9	9	9	42	291
		2	Terima	?	F-					
			UMPTN		UMPT		1			1
			Undur Diri	?	Bad-	1	2	3	3	9
				1	Und				2	2
			Undur Diri	?	Bad-					
			karena		Keu		1	1	3	5
			Keuangan							
		2	Tot			1	4	4	8	17
			Komputerisasi							
			Akuntansi Total			1	4	4	8	17
		1	Diberhentikan	?	Worst					1
		1	Tot							1
		2	Terima	?	F-					
			UMPTN		UMPT		5			5
			Undur Diri	?	Bad-		1	4	1	7
			Undur Diri	?	Und					
			karena		Keu		1			1
			Keuangan							
			Undur Diri	?	Bad-					
			Sepihak		Keu	1			1	2
			BINUS							
		2	Tot			1	7	4	2	15

Sistem Informasi Total					1	7	4	2	2	16
Teknik Informatika					2					
Terima UMPTN					?					
Undur Diri					?					
					1			1	1	20
					2					
Undur Sepihak BINUS					1					
2 Total					4	16	6	4	1	31
Teknik Informatika Total					4	16	6	4	1	31
2 Total					6	27	14	14	3	64
2007 Total					15	6	3	56	25	355
Grand Total					34	9	0	2	1	6

Metode yang digunakan dalam paper ini adalah dengan case study dan analisis menggunakan *data mining*. *Tool* yang digunakan untuk analisis *data mining* ini adalah menggunakan WEKA, di mana data diproses dengan *decision tree*. Berikut adalah tampilan data yang digunakan (Gambar 3).

No.	prod	jnlk	rimhs	kdjnr	nmjnr	jrsnu	Jurusan_SMU	ktsnu	Grade	gradednganko	tpe_snu	nama_tpe	jenis_snu	nama_jenis	jnpdf	nmpgf	kdkab	nmkab	kdpro	nmpro	kode	jenis	deskripsi	kategori	ketkatego
227	2007.0	2.0	1.100...	23.0	Kompu...	120.0	SMU JURUS...	Kab. S...		0.0	1.0	SMA	1.0	Swasta	8.0	Pindah...	332.0	Kab. S...	29.0	Banten	D	Undur ...	Pengun...	2.0	Undur Diri
428	2007.0	1.0	1.100...	24.0	Teknik...	110.0	SMU JURUS...	KAB. S...		0.0	1.0	SMA	1.0	Swasta	8.0	Pindah...	55.0	Kab. S...	3.0	Jawa ...	D	Undur ...	Pengun...	2.0	Undur Diri
1	2007.0	1.0	1.100...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...D		1.0	1.0	SMA	2.0	Negeri	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	A	PITTK	Dikelu...	1.0	Berhenti
2	2005.0	1.0	9.007...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	2.0	Jakar...	1.0	DKI Ja...	B	Pindah...	Pindah...	2.0	Undur Diri
3	2006.0	1.0	1.000...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	1.0	Jakar...	1.0	DKI Ja...	B	Pindah...	Pindah...	2.0	Undur Diri
4	2006.0	2.0	1.000...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...D		1.0	1.0	SMA	1.0	Swasta	10.0	Penda...	1.0	Jakar...	1.0	DKI Ja...	B	Pindah...	Pindah...	2.0	Undur Diri
5	2006.0	2.0	1.000...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	B	Pindah...	Pindah...	2.0	Undur Diri
6	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	KAB T...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	335.0	Kab. T...	29.0	Banten	B	Pindah...	Pindah...	2.0	Undur Diri
7	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	1.0	Jakar...	1.0	DKI Ja...	B	Pindah...	Pindah...	2.0	Undur Diri
8	2006.0	2.0	1.000...	24.0	Teknik...	110.0	SMU JURUS...	TANJU...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	346.0	Kota T...	34.0	Kepula...	B	Pindah...	Pindah...	2.0	Undur Diri
9	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	2.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
10	2005.0	1.0	9.007...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...D		1.0	1.0	SMA	1.0	Swasta	10.0	Penda...	3.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
11	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	Kota J...D		1.0					10.0	Penda...	167.0	Kota J...	10.0	Jambi	C	Drop ...	DO kar...	1.0	Berhenti
12	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	Kota P...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	121.0	Kota P...	7.0	Sumat...	C	Drop ...	DO kar...	1.0	Berhenti
13	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...B		3.0	1.0	SMA	2.0	Negeri	10.0	Penda...	3.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
14	2005.0	2.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	1.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
15	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	KAB. S...D		1.0					10.0	Penda...	327.0	Kab. S...	33.0	Jnan J...	C	Drop ...	DO kar...	1.0	Berhenti
16	2005.0	1.0	9.008...	23.0	Kompu...	110.0	SMU JURUS...	KAB. B...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	27.0	Kab. B...	2.0	Jawa ...	C	Drop ...	DO kar...	1.0	Berhenti
17	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	KAB. B...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	159.0	Kab. B...	9.0	Riau	C	Drop ...	DO kar...	1.0	Berhenti
18	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	3.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
19	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	Kota T...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	330.0	Kota T...	29.0	Banten	C	Drop ...	DO kar...	1.0	Berhenti
20	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	KOTA ...B		3.0					10.0	Penda...	330.0	Kota T...	29.0	Banten	C	Drop ...	DO kar...	1.0	Berhenti
21	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...B		3.0	1.0	SMA	2.0	Negeri	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
22	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	KAB. L...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	181.0	Kab. L...	11.0	Sumat...	C	Drop ...	DO kar...	1.0	Berhenti
23	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	5.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
24	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
25	2005.0	2.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...A		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
26	2005.0	2.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	TEBIN...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	119.0	Kota B...	7.0	Sumat...	C	Drop ...	DO kar...	1.0	Berhenti
27	2005.0	2.0	9.007...	23.0	Kompu...	120.0	SMU JURUS...	JAKAR...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	3.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
28	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	Kab. T...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	335.0	Kab. T...	29.0	Banten	C	Drop ...	DO kar...	1.0	Berhenti
29	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	KAB. B...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	27.0	Kab. B...	2.0	Jawa ...	C	Drop ...	DO kar...	1.0	Berhenti
30	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	KAB. B...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	27.0	Kab. B...	2.0	Jawa ...	C	Drop ...	DO kar...	1.0	Berhenti
31	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...B		3.0	1.0	SMA	2.0	Negeri	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
32	2005.0	2.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	Kota T...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	330.0	Kota T...	29.0	Banten	C	Drop ...	DO kar...	1.0	Berhenti
33	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...B		3.0					10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
34	2005.0	1.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	KAB. K...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	26.0	Kab. K...	2.0	Jawa ...	C	Drop ...	DO kar...	1.0	Berhenti
35	2005.0	1.0	9.008...	24.0	Teknik...	250.0	STM LISTRIK...	JAKAR...D		1.0	1.0	SMA	1.0	Swasta	10.0	Penda...	5.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
36	2005.0	2.0	9.008...	23.0	Kompu...	120.0	SMU JURUS...	Kab. B...C		2.0	1.0	SMA	1.0	Swasta	10.0	Penda...	337.0	Kab. B...	30.0	Kepula...	C	Drop ...	DO kar...	1.0	Berhenti
37	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	JAKAR...A		4.0	1.0	SMA	1.0	Swasta	10.0	Penda...	4.0	Jakar...	1.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti
38	2005.0	1.0	9.008...	24.0	Teknik...	110.0	SMU JURUS...	Kota ...B		3.0	1.0	SMA	1.0	Swasta	10.0	Penda...	790.0	Kota ...	19.0	DKI Ja...	C	Drop ...	DO kar...	1.0	Berhenti

Gambar 3. Daftar data yang digunakan.

HASIL DAN PEMBAHASAN

Dari data pada Gambar 3 dipilih attribute (9 attribute) untuk melakukan analisis terhadap jenis kategori undur diri yang ada. Data diproses dengan menggunakan aplikasi WEKA dengan melakukan klasifikasi terhadap data tersebut. Attribute yang digunakan adalah Jenis Kelamin, Jenis Pendaftaran, Jurusan yang diminati, Jurusan SMA, Propinsi Asal, Nilai Algoritma (semester awal), Kategori Undur Diri, Grade Hasil tes Masuk dan Kelas dari Jenis undur diri tersebut. Data diproses dengan pilihan *Tree* dan jenis J48 dengan *cross-validation* 10 Fold.

Hasil running melalui aplikasi WEKA dapat dilihat dari data di bawah ini:

==== Run information ====

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: DO_2-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35-weka.filters.unsupervised.attribute.Remove-R1,3-5,7,9-10,12-20,22-24,26-28,30-33

Instances: 1396

Attributes: 9

- jinkel
- nmjur
- Jurusan_SMU
- gradedgnangka
- nmjpf
- nmpro
- kategori
- grade_mtk_algo
- class

Test mode:10-fold cross-validation

==== Classifier model (full training set) ====

J48 pruned tree

kategori <= 1: Worst (463.0)

kategori > 1

| gradedgnangka <= 3

| | gradedgnangka <= 0

| | | nmjur = Komputerisasi Akuntansi: F-UMPTN (3.0/1.0)

| | | nmjur = Teknik Informatika: F-UMPTN (25.0/6.0)

| | | nmjur = Sistem Informasi: Bad-Und (3.0/1.0)

| | gradedgnangka > 0: Bad-Und (578.0/224.0)

| gradedgnangka > 3

| | nmpro = DKI Jakarta

| | | Jurusan_SMU = SMU JURUSAN IPS

| | | | nmjur = Komputerisasi Akuntansi: Bad-Und (13.0/1.0)

| | | | nmjur = Teknik Informatika: Bad-Und (0.0)

| | | | nmjur = Sistem Informasi: F-UMPTN (24.0/15.0)

| | | Jurusan_SMU = SMU JURUSAN IPA: F-UMPTN (165.0/73.0)

```

| | | Jurusan_SMU = SMEA/SMK: F-UMPTN (0.0)
| | | Jurusan_SMU = STM LISTRIK/ELEKTRONIKA: F-UMPTN (0.0)
| | | Jurusan_SMU = STM MESIN: F-UMPTN (0.0)
| | | Jurusan_SMU = SMA JURUSAN A4: F-UMPTN (0.0)
| | nmpro = Banten
| | | jnkel <= 1: Bad-Und (12.0/5.0)
| | | jnkel > 1: F-UMPTN (4.0/1.0)
| | nmpro = Kalimantan Barat: Bad-Und (3.0/1.0)
| | nmpro = Kepulauan Riau: Bad-Und (2.0/1.0)
| | nmpro = Kepulauan Bangka Belitung: Bad-Und (3.0/2.0)
| | nmpro = Jawa Barat
| | | jnkel <= 1: Bad-Und (30.0/16.0)
| | | jnkel > 1
| | | | nmjur = Komputerisasi Akuntansi: Bad-Und (3.0)
| | | | nmjur = Teknik Informatika: F-UMPTN (7.0/1.0)
| | | | nmjur = Sistem Informasi: F-UMPTN (1.0)
| | nmpro = Sumatera Utara: Bad-Und (7.0/1.0)
| | nmpro = Sulawesi Selatan: Bad-Und (2.0)
| | nmpro = Riau
| | | Jurusan_SMU = SMU JURUSAN IPS: Bad-Keu (2.0)
| | | Jurusan_SMU = SMU JURUSAN IPA: Bad-Und (7.0)
| | | Jurusan_SMU = SMEA/SMK: Bad-Und (0.0)
| | | Jurusan_SMU = STM LISTRIK/ELEKTRONIKA: Bad-Und (0.0)
| | | Jurusan_SMU = STM MESIN: Bad-Und (0.0)
| | | Jurusan_SMU = SMA JURUSAN A4: Bad-Und (0.0)
| | nmpro = Irian Jaya Barat: F-UMPTN (0.0)
| | nmpro = Jambi: F-UMPTN (5.0/1.0)
| | nmpro = Sumatera Selatan: F-UMPTN (4.0/1.0)
| | nmpro = Jawa Tengah: F-UMPTN (10.0/4.0)
| | nmpro = Sumatera Barat: Bad-Und (2.0/1.0)
| | nmpro = DI Yogyakarta: Bad-Und (1.0)
| | nmpro = Jawa Timur: Bad-Und (2.0)
| | nmpro = Lain-lain: Bad-Und (4.0/1.0)
| | nmpro = Lampung: Bad-Und (2.0/1.0)
| | nmpro = Sulawesi Utara: F-UMPTN (1.0)
| | nmpro = Bali: F-UMPTN (0.0)
| | nmpro = Bengkulu: F-UMPTN (0.0)
| | nmpro = Luar Negeri: Bad-Und (2.0/1.0)
| | nmpro = Kalimantan Selatan: F-UMPTN (0.0)
| | nmpro = Kalimantan Timur: F-UMPTN (3.0/1.0)
| | nmpro = Nusa Tenggara Timur: F-UMPTN (0.0)
| | nmpro = Papua (Irian Jaya): F-UMPTN (0.0)
| | nmpro = Nanggroe Aceh Darussalam: F-UMPTN (1.0)
| | nmpro = Gorontalo: F-UMPTN (0.0)
| | nmpro = Sulawesi Tengah: Bad-Keu (1.0)
| | nmpro = Kalimantan Tengah: F-UMPTN (0.0)
| | nmpro = Nusa Tenggara Barat: F-PT Lain (1.0)

```

Number of Leaves : 52
Size of the tree: 63
Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	996	71.3467 %
Incorrectly Classified Instances	400	28.6533 %
Kappa statistic	0.5848	
Mean absolute error	0.1221	
Root mean squared error	0.2524	
Relative absolute error	51.0743 %	
Root relative squared error	73.0225 %	
Total Number of Instances	1396	

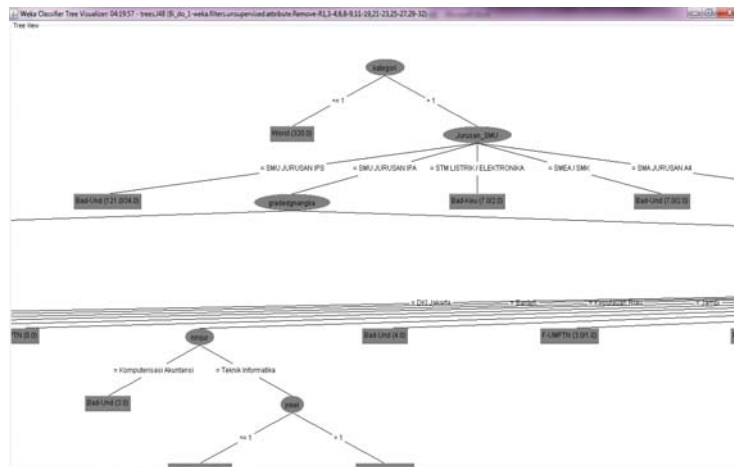
=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	Worst	
0	0	0	0	0.702	F-PindahJur	
0.792	0.294	0.601	0.792	0.683	0.8	Bad-Und
0	0.008	0	0	0.752	Bad-Keu	
0	0	0	0	0.56	F-PT Lain	
0.623	0.109	0.517	0.623	0.565	0.833	F-UMPTN
Weighted Avg.	0.713	0.123	0.628	0.713	0.665	0.863

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
463	0	0	0	0	0	a = Worst
0	0	10	0	0	1	b = F-PindahJur
0	0	396	5	0	99	c = Bad-Und
0	0	173	0	0	25	d = Bad-Keu
0	0	1	0	0	3	e = F-PT Lain
0	0	79	4	0	137	f = F-UMPTN

Tree Visualization dari proses tersebut dapat terlihat di bawah ini (Gambar 4), tapi tidak semua ditampilkan.



Gambar 4. Tree Visualization.

Dari hasil proses menggunakan aplikasi WEKA terhadap data tersebut, dapat diambil beberapa hal antara lain: dengan menggunakan J48, angkanya menunjukkan untuk Correctly Classified Instances = 996 (71.3467 %) dan Incorrectly Classified Instances= 400 (28.6533 %). Angka false positif nya juga rendah. Dengan menggunakan Tree-J48, Kategori undur diri ditentukan oleh parameter antara lain: Jurusan SMA, Grade Tes Masuk, Propinsi, Jurusan yang dipilih dan Jenis Kelamin. Dari data hasil analisis dengan Tree-J48 melalui aplikasi WEKA, dapat diklasifikasikan sebagai berikut.

Untuk kategori Undur diri (1) yaitu Diberhentikan, Drop Out (DO), PTTKK, Tidak Lulus SMU tidak ada variabel yang mempengaruhi. Sedangkan untuk Kategori undur diri lainnya (2), dan dengan grade test masuk bukan grade A dan gradenya kosong dan jurusan Teknik Informatika mempunyai nilai terbanyak karena lulus UMPTN. Kategori 2 (undur diri) dan gradenya bukan grade A dan grade berada di antara B, C dan D mempunyai nilai undur diri yang cukup banyak. Untuk kategori 2 (undur diri) dan grade test masuknya adalah Grade A, dan propinsinya berasal dari Jakarta serta Jurusan SMA nya IPS dengan pilihan jurusan di binus adalah Komputerisasi Akuntansi dan Teknik Informatika mempunyai jumlah undur diri karena diterima di perguruan tinggi Negeri dan benar-benar mengundurkan diri. Sedangkan untuk Jurusan IPA untuk kategori ini pengunduran diri disebabkan oleh banyaknya yang diterima di perguruan tinggi Negeri. Jika dianalisis berdasarkan propinsi, DKI Jakarta mempunyai jumlah tertinggi disusul Banten. Tetapi jika dilihat dari propinsi diluar JABODETABEK, jumlah tertinggi adalah Sumatra Utara, Riau dan Jawa Tengah. Dengan model ini, dapat diperkirakan (diprediksi) tingkat undur diri dari mahasiswa dengan berbagai kategori undur diri, jika diberikan atribut-atribut seperti yang ditunjukkan oleh perhitungan di atas.

PENUTUP

Dari hasil analisis proses yang dilakukan dapat dijelaskan bahwa undur diri yang dilakukan mahasiswa terkait dengan beberapa parameter/attribute antara lain Jurusan SMA, Grade Tes Masuk, Propinsi, Jurusan yang dipilih dan Jenis Kelamin. Dari data di atas dapat diprediksi jumlah undur diri dan kategorinya sesuai dengan proses yang telah dilakukan dengan menggunakan *data mining* dari parameter tersebut. Prediksi dapat dilakukan terhadap data sesungguhnya dari proses yang berjalan. Proses analisis ini hanya menggunakan beberapa parameter/atribut, untuk analisis yang lebih lengkap, dapat ditambahkan parameter-parameter lainnya yang dapat mendukung keakuratan proses analisis misalnya nilai ujian Nasional, tingkat penghasilan orang tua, pekerjaan orang tua dan sebagainya.

DAFTAR PUSTAKA

- Akintola K.G., Adetunmbi A.O., Adeola O.S. (2011). Building data warehousing and data mining from course management systems: a case study of futa course management information systems. *International Journal of Database Theory and Application* 4(3).
- Inmon, W. H.(2005). *Building the Data Warehouse* (4th edition). New Jersey: John Wiley and Sons.
- Kantardzic, Mehmet. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New Jersey: IEEE Press.
- Neary, R. (1999). Building a data warehouse and data mining for a strategic advantage. *Journal of Information Technology Theory and Application (JITTA)*, 1(1), 7 – 21.