

Analisis dan Prediksi Kinerja Mahasiswa Menggunakan Teknik *Data Mining*

Sofi Defiyanti, M.Kom

Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Sofi.defiyanti@staff.unsika.ac.id

Abstrak - Prestasi belajar merupakan salah satu aspek yang paling penting dalam bidang pendidikan. Prestasi belajar yang tinggi selalu menjadi harapan semua pihak. Bagi pihak perguruan tinggi prestasi belajar mahasiswanya merupakan salah satu indikator efektif proses belajar mengajar, yang sekaligus dapat digunakan untuk meningkatkan citra perguruan tinggi tersebut. Di perguruan tinggi prestasi belajar yang dicapainya oleh mahasiswa menggunakan Indeks Prestasi Kumulatif (IPK). Beberapa teknik data mining diantaranya adalah *decision tree*, *naïve bayes*, dan *artificial neural network* dapat digunakan untuk menganalisa kinerja mahasiswa. Penelitian yang sudah dilakukan didapat bahwa pengujian dan validasi dengan menggunakan 10 *cross validation* dengan mengukur tingkat *accuracy* dan *ROC curve* didapat bahwa dengan menggunakan data dua semester sebelumnya untuk memprediksi kinerja mahasiswa memiliki akurasi yang paling tinggi adalah metode *decision tree* sebesar 67,63% dengan nilai *ROC curve* masuk kedalam *good classification*.

Kata Kunci: Data Mining, Prestasi Belajar, Kinerja Mahasiswa

PENDAHULUAN

Data Mining dengan menggunakan teknik klasifikasi, digunakan untuk penemuan sekumpulan model yang menggambarkan dan membedakan konsep atau kelas-kelas data, agar model yang sudah dibentuk dapat digunakan untuk memprediksi. Teknik klasifikasi dapat dikelompokkan kedalam beberapa algoritma diantaranya Pohon Keputusan, Jaringan Saraf Tiruan (JST), Naïve Bayes dan lain-lain.

Beberapa penelitian dalam bidang akademik perguruan tinggi telah banyak dilakukan dengan menggunakan teknik klasifikasi *data mining* diantaranya adalah menggunakan algoritma *Artificial Neural Network* (ANN), *naïve bayes* dan *decision tree* untuk memprediksi kinerja mahasiswa.

Di perguruan tinggi keberhasilan belajar mahasiswa ditunjukkan dengan prestasi belajar yang dicapainya berdasarkan evaluasi hasil belajar. Prestasi belajar mahasiswa merupakan salah satu ukuran dalam menghitung kinerja mahasiswa. Ukuran yang digunakan dalam menghitung kinerja mahasiswa adalah Indeks Prestasi Kumulatif (IPK). Indeks Prestasi kumulatif adalah nilai yang diperoleh mahasiswa setelah mengikuti sejumlah matakuliah. Nilainya berkisar antara 0,0 samapai dengan 4,0 dimana semakin besar nilai IPK maka semakin tinggi pula prestasi belajarnya, sebaliknya semakin kecil nilai IPK maka semakin rendah pula prestasi belajarnya. Perhitungan IPK adalah jumlah nilai tiap matakuliah dikali SKS mata kuliah tersebut lalu dibagi dengan jumlah SKS yang telah diselesaikan.

IPK merupakan salah satu persyaratan yang diperhitungkan untuk mendapatkan pekerjaan disuatu perusahaan. Selain itu prestasi belajar siswa menjadi salah satu faktor penentu kualitas dari perguruan tinggi. Terbukti dalam borang akreditasi program studi pada buku 3A standar 3 Mahasiswa dan Lulusan salah satu faktornya adalah IPK. Dengan semakin tinggi rata-rata IPK mahasiswa lebih dari 3,00 maka penilaian pada isian borang pun akan mendapat nilai empat atau nilai maksimal dari pengisian borang tersebut. Sehingga kualitas dari suatu program studi pun akan semakin baik. Rata-rata IPK tersebut didapat dari keseluruhan IPK mahasiswa sebanyak lima angkatan.

Program EPSBED atau Evaluasi Program Studi Berbasis Evaluasi Diri adalah program yang digunakan untuk pengisian data untuk evaluasi penyelenggaraan program studi. Di dalam program EPSBED ini semua kegiatan akademik program studi dilaporkan, dari data mahasiswa pertama kali masuk ke dalam program studi tersebut samapai mahasiswa tersebut lulus.

Dari latar belakang tersebut maka dapat dilakukan penelitian berdasarkan data-data akademik mahasiswa yang berasal dari pelaporan EPSEBD untuk dianalisis kinerja mahasiswa dengan menggunakan beberapa algoritma data mining.

TINJAUAN PUSTAKA

1. Prestasi Belajar

Pengertian prestasi belajar menurut kamus besar bahasa Indonesia adalah penguasaan pengetahuan atau keterampilan yang dikembangkan oleh mata pelajaran, lazimnya ditunjukkan dengan nilai tes atau nilai angka yang diberikan oleh guru. Sedangkan menurut M. Suryanto prestasi belajar adalah seluruh kecakapan hasil capai yang diperoleh dari proses belajar dilembaga pendidikan, yang dinyatakan dengan nilai-nilai.

Prestasi belajar dalam perguruan tinggi ditunjukkan berupa angka yang disebut dengan IP dan IPK. IP adalah indek prestasi merupakan nilai yang diberikan setelah mahasiswa menjalani proses belajar selama satu semester. Sedangkan IPK adalah indek prestasi kumulatif merupakan nilai mahasiswa secara kumulatif yang dimulai dari semester pertama sampai akhir. Nilai IPK berkisar antara 0,0 sampai 4,0 dimana semakin nilai IPK mendekati angka 4,0 maka semakin tinggi pulalah prestasi belajarnya. Begitupun sebaliknya jika nilai IPK semakin rendah maka prestasi belajar mahasiswa tersebut rendah. Nilai IP didapat dari persamaan (1) dan nilai IPK didapat persamaan (2).

$$IP = \frac{\sum_{j=1}^n N_j \cdot K_j}{\sum_{j=1}^n K_j} \quad (1)$$

$$IPK = \frac{\sum_{j=1}^n N_j \cdot K_j \text{ seluruh semester}}{\sum_{j=1}^n K_j \text{ seluruh semester}} \quad (2)$$

Keterangan :

- IP = Indek Prestasi
- IPK = Indek Prestasi Kumulatif
- N_j = Nilai Matakuliah j
- K_j = bobot SKS Matakuliah j

2. Data Mining

Menurut Turban pada bukunya berjudul *Decision Support Systems and Intelligent Systems* Tahun 2005. *Data mining* merupakan teknologi penggalian data yang sangat banyak dengan menggunakan penggabungan metoda analisis tradisional dan algoritma semi otomatis yang menggunakan teknik statistik, matematika, kecerdasanbuatan dan *machine learning* agar dapat mengekstraksi dan mengidentifikasi informasi pengetahuan baru yang berguna yang terkubur dalam basis data yang cenderung berukuran besar. Sehingga *data mining* dapat menjawab pertanyaan-pertanyaan bisnis yang jika dengan cara tradisional memerlukan banyak waktu dan biaya yang tinggi. *Data mining* mengeksplorasi basis data untuk menemukan pola-pola tersembunyi, mencari informasi untuk memprediksi yang mungkin terlupakan oleh para pelaku bisnis.

Menurut Tan pada buku berjudul *Introduction to Data Mining* tahun 2006, Data mining memiliki tugas yang dikategorikan menjadi dua kategori utama, yaitu:

a) Tugas prediksi

Tugas prediksi ini adalah berfokus kepada prediksi nilai dari atribut tertentu berdasarkan nilai dari atribut lainnya. Atribut yang digunakan untuk membuat prediksi disebut atribut penjelas atau *independent variable* sedangkan atribut yang diprediksi dikenal dengan atribut target atau *dependent variable*. Tugas prediksi akan membangun pemodelan prediktif yang akan membangun sebuah model untuk *dependent variable* berdasarkan *independent variable*. Pemodelan prediktif terdapat dua tipe yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk *dependent variable* yang diskrit. Sedangkan regresi digunakan untuk *dependent variable* yang kontinyu.

b) Tugas deskriptif

Tugas deskriptif berfokus pada pembuatan pola yang dapat digunakan untuk menyimpulkan hubungan di dalam data. Tugas deskriptif terdiri dari beberapa model diantaranya adalah analisa asosiasi yang digunakan untuk menentukan pola yang mendeskripsikan fitur-fitur data yang saling berhubungan, analisa cluster merupakan proses untuk mencari kelompok-kelompok data sehingga data yang berada dalam satu kelompok memiliki kemiripan dibandingkan data yang terletak pada kelompok lain dan yang terakhir adalah deteksi anomaly merupakan proses identifikasi data yang memiliki perbedaan karakteristik yang signifikan dengan data yang lain.

3. Klasifikasi

Klasifikasi merupakan salah satu teknik yang ada dalam data mining. Klasifikasi adalah suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan dan menjelaskan atau membedakan konsep atau kelas data, yang diperlukan untuk memperkirakan kelas dari suatu objek. Klasifikasi dapat berupa aturan “jika-maka”, berupa *decision tree*, formula matematis atau *neural network*.

Proses klasifikasi dibagi menjadi dua fase yaitu fase *learning* dan fase *test*. Pada fase *learning* data yang telah diketahui kelas datanya dijadikan pembelajaran untuk membentuk model perkiraan. Sedangkan pada fase *test*, model yang sudah terbentuk pada fase *learning* akan diuji tingkat akurasi. Fase pengujian inilah disebut fase *test*. Bila pada fase *test* ini tingkat akurasi sudah mencukupi maka model ini dapat dipakai untuk memprediksi kelas data yang belum diketahui.

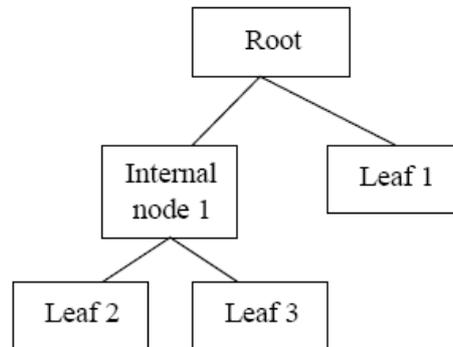
4. *Decision tree*

Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar algoritma *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*).

Decision Tree sesuai digunakan untuk kasus-kasus dimana outputnya bernilai diskrit. Walaupun banyak variasi model *Decision Tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri kasus berikut cocok untuk diterapkan *Decision Tree* :

- Data/example dinyatakan dengan pasangan atribut dan nilainya. Misalnya atribut satu *example* adalah temperatur dan nilainya adalah dingin. Biasanya untuk satu *example* nilai dari satu atribut tidak terlalu banyak jenisnya. Dalam contoh atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Sedang dalam atribut temperatur, nilainya bisa dingin, sedang atau panas. Tetapi untuk beberapa kasus bisa saja nilai temperatur berupa nilai numerik.
- Label/output data biasanya bernilai diskrit. Output ini bisa bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak. Dalam beberapa kasus mungkin saja outputnya tidak hanya dua kelas. Tetapi penerapan *Decision Tree* lebih banyak kasus binari.
- Data mempunyai *missing value*. Misalkan untuk beberapa *example*, nilai dari suatu atributnya tidak diketahui. Dalam keadaan seperti ini *Decision Tree* masih mampu memberi solusi yang baik.

Membangun *tree* dimulai dengan data pada simpul akar (*root node*) kemudian pilih sebuah *atribut* dan formulasi sebuah *logical test* pada *atribut* tersebut lakukan percabangan pada setiap hasil dari *test*, dan terus bergerak ke subset ke contoh yang memenuhi hasil dari simpul anak cabang (*internal node*) yang sesuai lakukan proses rekursif pada setiap simpul anak cabang. Ulangi hingga dahan-dahan dari *tree* memiliki contoh dari satu kelas tertentu. Contoh dari sebuah *decision tree* (Gambar 1).



Gambar 1 *Decision Tree*

Beberapa model *decision tree* yang sudah dikembangkan antara lain IDS, ID3, C4.5, CHAID dan CART.

Kelebihan yang dimiliki oleh *Decision Tree* adalah karena mudah dipahami, *Decision Tree* juga memiliki kemudahan jika dikonversi menjadi seperangkat aturan produksi, *Decision Tree* dapat mengklasifikasikan data kategorikal dan numeric tetapi *Decision Tree* hanya dapat memberikan output berupa data kategori.

TEKNIK PENGUJIAN

1. *Ten Folds Cross Validation*

Ten Fold Cross Validation adalah salah satu metode pengujian untuk pembelajaran yang terlatih (*supervised learning*). Dimana dalam setiap *fold* dibagi menjadi beberapa *subset* sebanyak 10 *subset* dengan ukuran tiap *subset* sama banyaknya. Setelah itu maka akan dilakukan training sebanyak 10 kali dengan menggunakan Sembilan fold untuk training set dan satu fold digunakan sebagai test set. Kemudian dari setiap iterasi rata-rata error dari keseluruhan iterasi. Dengan ilustrasi seperti yang digambarkan pada Table 1 sebagai berikut:

Tabel 1 contoh 10-Folds Cross Validation

Iterasi ke-	Fold yang digunakan sebagai training	Fold yang digunakan sebagai tes
1	2,3,4,5,6,7,8,9,10	1
2	1,3,4,5,6,7,8,9,10	2
3	1,2,4,5,6,7,8,9,10	3
4	1,2,3,5,6,7,8,9,10	4
5	1,2,3,4,6,7,8,9,10	5
6	1,2,3,4,5,7,8,9,10	6
7	1,2,3,4,5,6,8,9,10	7
8	1,2,3,4,5,6,7,9,10	8
9	1,2,3,4,5,6,7,8,10	9
10	1,2,3,4,5,6,7,8,9	10

2. *Confusion Matrix*

Confusion matrix merupakan metode yang menggunakan tabel matriks seperti pada Tabel 2, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Olson, 2008).

Tabel 2 Model *Confusion Matrix* (Olson, 2008)

		True Class	
		Positive	Negative
Predicted Class	Positive	true positives count (TP)	false negatives count (FP)
	Negative	false positives count (FN)	true negatives count (TN)

True positives adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*. Setelah data-data telah masuk ke dalam *confusion matrix* maka dapat dihitung nilai-nilai *sensitivity (recall)*, *specificity*, *precision* dan *accuracy*. Untuk menghitung digunakan persamaan di bawah ini (Olson, 2008):

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (3)$$

Ketika klasifikasi tidak menghasilkan binary class, maka *Confusion Matrix* akan berubah menjadi lebih kompleks. Contoh untuk klasifikasi non binary diperlihatkan pada Table.3 untuk tiga class dibawah ini :

Tabel 3 *Confusion Matrix untuk Tiga Class*: (a) Actual dan (b) Expected (Witten, 2011)

		Predicted Class						Predicted Class			
		a	b	c	Total			a	b	c	Total
Actual Class	a	88	10	2	100	Actual Class	a	60	30	10	100
	b	14	40	6	60		b	36	18	6	60
	c	18	10	12	40		c	24	12	4	40
	Total	120	60	20			Total	120	60	20	
(a)					(b)						

3. *ROC Curve*

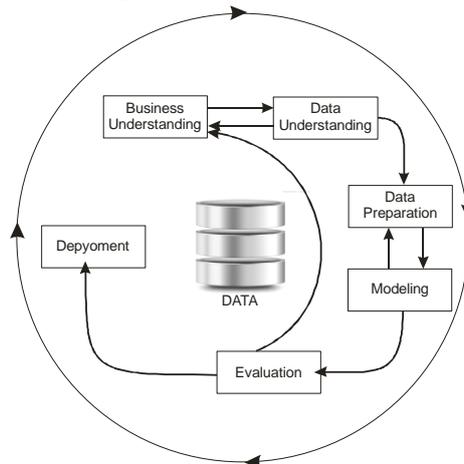
ROC (Receiver Operating Characteristics) curve adalah teknik untuk menampilkan, mengorganisasi dan memilik teknik klasifikasi berdasarkan performanya. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal (Vercellis, 2009). *The area under curve (AUC)* dihitung untuk mengukur perbedaan performansi metode yang digunakan. Nilai dari *ROC curve* hanya terdiri dari 0 sampai 1. Semakin nilai *ROC curve* mendekati 1 maka akan semakin baik seperti diperlihatkan pada Tabel 4 menurut (Gorunescu, 2011)

Tabel 4 *Classifying the accuracy of diagnostic tes* (Gorunescu, 2011)

0,90-1,00	<i>Excellent Classification</i>
0,80-0,90	<i>Good Classification</i>
0,70-0,08	<i>Fair Classsification</i>
0,60-0,70	<i>Poor Classification</i>
0,50-0,60	<i>Failure</i>

METODE PENELITIAN

Metodologi yang digunakan dalam mengembangkan *data mining* adalah CRISP-DM, CRISP-DM adalah metodologi yang dibentuk oleh komisi Eropa pada tahun 1996 yang menerapkan standar dalam proses *data mining*. Dalam CRISP-DM terdapat enam fase yang akan dilakukan dalam penelitian pengembangan *data mining* sesuai dengan ilustrasi pada Gambar 2.



Gambar 2 Siklus Hidup CRISP-DM

Berikut enam fase siklus hidup pengembangan *data mining* menggunakan CRISP-DM (<http://www.crisp-dm.org>):

1. Fase *Business Understanding* (Pemahaman Bisnis)
Pada fase pemahaman bisnis ini adalah tahapan pertama dalam pengembangan *data mining*. Pada fase ini adalah fase dimana memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pemahaman tersebut kedalam pendefinisian *data mining*, yang kemudian akan diterjemahkan menjadi rencana dan strategi untuk mencapai tujuan tersebut.
2. Fase *Data understanding* (Pemahaman Data)
Pemahaman data dilakukan pada tahap kedua setelah pemahaman bisnis selesai dilakukan. Dalam tahap pemahaman data dimulai dengan pengumpulan data awal dan hasil yang menarik dari data yang dapat digunakan. Kemudian akan dilanjutkan dengan proses mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa informasi yang tersembunyi.
3. Fase *Data Preparation* (Persiapan Data)
Tahap persiapan data dilakukan setelah tahap pemahaman data selesai dilakukan. Tahap persiapan data adalah tahapan yang mencakup semua kegiatan yang diperlukan untuk membangun dataset akhir atau data yang akan dimasukkan kedalam alat pemodelan yang berasal dari data mentah awal. Tahap persiapan data kemungkinan akan dilakukan beberapa kali dan tidak dalam urutan yang ditentukan. Pada tahap ini mencakup pemilihan tabel, *record*, atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk tahap pemodelan.
4. Fase *Modeling* (Pemodelan)
Tahapan Pemodelan dilakukan setelah tahap persiapan data selesai dilakukan. Pada tahap pemodelan dilakukan pemilihan dan penerapan teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Terdapat beberapa pemodelan untuk menyelesaikan masalah yang sama dalam data mining. Beberapa pemodelan tersebut memiliki spesifikasi dan format data yang khusus. Oleh karena itu pada tahap pemodelan akan sangat memungkinkan untuk mebalik ke tahap sebelumnya yaitu tahap persiapan data.
5. Fase *Evaluation* (Evaluasi)
Setelah tahap pemodelan selesai dilakukan maka sebuah model atau beberapa model sudah terbentuk. Sebelum masuk kedalam tahap penyebaran akhir dari model, maka perlu diadakan evaluasi dan meninjau

apakah model yang tercipta sudah benar-benar mencapai tujuan bisnis yang telah ditetapkan pada tahap pemahaman bisnis.

6. Fase *Deployment* (penempatan)

Setelah fase evaluasi selesai dilakukan dan didapat hasil yang memuaskan dalam tahap evaluasi, bahwa semua tujuan bisnis sudah tercapai. Maka masuk kedalam fase *deployment*. Pada fase *deployment* ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna.

EKSPERIMEN DAN PEMBAHASAN

a. Fase Business understanding

Bagian akademik adalah bagian yang dibawahi oleh wakil rektor bidang II. Semua aktifitas mahasiswa dari rekrutmen mahasiswa baru sampai mahasiswa tersebut lulus. Salah satu tugas bagian akademik adalah menyiapkan bahan kebijakan akreditasi perguruan tinggi dari fakultas, pascasarjana dan program studi dan menyiapkan pelaporan EPSBED yang disampaikan ke Kopertis dan Dikti. Bagian akademik merupakan salah satu bagian yang terpenting dalam menyiapkan bahan kebijakan akreditasi program studi, karena di dalam proses akreditasi ada bidang akademik yang berisi seputar data mahasiswa, dari mahasiswa yang daftar, mahasiswa diterima, mahasiswa lulus, mahasiswa DO dan data IPK rata-rata mahasiswa lima tahun terakhir.

Dengan memanfaatkan data-data akademik mahasiswa terdahulu yang sudah ada maka dapat dianalisis dan diprediksi IPK mahasiswa perangkatan dengan menggunakan teknik data mining, dengan mengetahui prediksi IPK mahasiswa baru, maka pihak prodi mendapat peringatan dini agar lebih perhatian terhadap mahasiswanya, sehingga kualitas dari prodi tersebut tetap baik atau bahkan lebih baik

b. Fase Data Understanding

Sesuai dengan yang diatur dalam SK Dirjen DIKTI No. 34/DIKTI/Kep/2001, pelaporan program studi pada setiap semester harus menggunakan data elektronik. Program EPSBED atau Evaluasi Program Studi Berbasis Evaluasi Diri adalah program yang digunakan untuk pengisian data untuk evaluasi penyelenggaraan program studi.

Maka untuk dapat melakukan penelitian ini penulis meminta data EPSBED Universitas Singaperbangsa Karawang (UNSIKA) ke bagian akademik dari pelaporan 2006 sampai pelaporan 2009.

c. Fase Data Preparation

Pemilihan data dilakukan dari sebuah sistem pelaporan yang diberi nama EPSBED. Dari 24 tabel yang ada di program EPSBED akan digunakan tiga buah tabel yaitu tabel master mahasiswa, tabel transaksi dan tabel master program studi.

Agar data-data ini dapat diterapkan dalam fase modeling maka perlu di transformasikan. Model yang akan digunakan adalah *Decision Tree*. Setelah dilakukan analisis terhadap model *Decision Tree* tersebut maka dataset yang dihasilkan menjadi Tabel 5 di bawah ini:

Tabel 5 Transformasi Data

FIELD NAME	DESCRIPTION	VALUE
Jenjang	Jenjang akademik yang diselenggarakan oleh program studi	A=S3 F=D2 B=S2 G=D1 C=S1 H=Sp-1 D=D4 I=Sp-2 E=D3 J=Profesi
Kluster	Pengklasifikasian berdasarkan jurusan program studi	sains, social dan kesehatan
Asal	Tempat tanggal lahir	karawang dan luar karawang
Umur	Umur saat masuk kuliah	Umur <=18 kategori 1 Umur antara 19-21 kategori 2 Umur >=22 kategori 3
JK	Jenis kelamin	Laki-laki = L Perempuan = P
SKS1	SKS yang diambil disatu semester sebelumnya	SKS <=12 kategori 0 SKS antara 13-15 kategori 1
SKS2	SKS yang diambil didua semester sebelumnya	SKS antara 16-18 kategori 2 SKS antara 19-21 kategori 3 SKS antara 22-24 kategori 4
IPK	Indeks prestasi kumulatif disemester yang bersangkutan	Dan ditransffromasikan menjadi lima kategori
IPS1	Indeks prestasi semester disatu semester sebelumnya	IPK antara 0-2,25 kategori 0 IPK antara 2,26-2,50 kategori 1
IPS2	Indeks Prestasi semester didua semester sebelumnya	IPK antara 2,51 – 2,75 kategori 2 IPK antara 2,76-3,00 kategori 3
IPK_out	Indeks prestasi kumulatif di semester selanjutnya. Field ini adalah field output dari prediksi kinerja mahasiswa	IPK >3.00 kategori 4

d. Fase Modeling

Setelah melewati fase data preparation maka masuk ketahap selanjutnya yaitu fase modeling dengan menggunakan algoritma *Decision Tree* yaitu C4.5.

e. Fase Evaluasi

Fase evaluasi akan dilakukan dengan menggunakan teknik *confution matrix* dan *ROC curve*. Didapat hasil seperti table 6 untuk *confution matrik* dengan akurasi sebesar 65,8 % dan table 7 untuk *ROC curve* dengan rata-rata sebesar 0,84.

Tabel 6 Evaluasi Confusion matrix

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
614	83	33	36	2	a = 0
154	261	101	46	4	b = 1
87	73	438	204	46	c = 2
63	47	180	980	360	d = 3
12	18	67	280	1354	e = 4

Tabel 7 evaluasi ROC Curve

ROC Area	Class
0.901	0
0.835	1
0.798	2
0.783	3
0.891	4
Weighted Avg.	0.841

f. Fase Deployment

Dari hasil fase pemodelan dan fase evaluasi yang sudah didapat maka dapat digunakan untuk memprediksi kinerja mahasiswa (IPK) yaitu dengan data yang didapat dari data EPSBED, dari data EPSBED diambil atribut-atribut yang sudah ditentukan pada Table 5.

KESIMPULAN

Data akademik dapat digunakan untuk menganalisis kinerja mahasiswa dengan menggunakan teknik-teknik data mining salah satunya adalah *Decision Tree C4.5*. Dengan variabel-variabel penentu adalah jenjang,

kluster jurusan, asal mahasiswa, umur, jenis kelamin, SKS satu semester sebelumnya, SKS dua semester sebelumnya, IPK sebelumnya, IPS satu semester sebelumnya, IPS dua semester sebelumnya, dan IPK akhir yang menjadi hasil dari prediksi. Dengan hasil evaluasi dengan menggunakan dua buah metode yaitu dengan menggunakan *confusion matrix* dan *ROC curve*. Dengan menggunakan *confusion matrix* didapat angka akurasi sebesar 65,8% sedangkan dengan menggunakan *ROC curve* didapat rata-rata sebesar 0.84. maka didapat kesimpulan bahwa teknik *Decision tree* termasuk *good classification* seperti pada table 4.

REFERENSI

- Aher, Sunita,. And Lobo. (2011). *Data Mining in Educational System using WEKA*. International Conference on Emerging Technology Trends (ICETT) 2011.
- Aher, Sunita,. And Lobo. (2011). *Data Mining in Educational System using WEKA*. International Conference on Emerging Technology Trends (ICETT) 2011.
- Baradwaj, Brijesh,. And Pal, Saurabh. (2011). *Mining Educational Data to Analyze Students' Performance*. International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
- Chandra, E., dan Nandhini, K.. (2005). *Predicting Student Performance using Classification Techniques*. Proceedings of SPIT – IEEE Colloquium and International Conference. Mumbai. India Volumen 5, 83.
- Cripps, Al. (1996). *Using Artificial Neural Nets to Predict Academic Performance*. ACM Symposium on Applied Computing
- Gorunescu, Florin. (2011). *Data Mining Concepts, Models and Techniques*. Springer
- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Proc. of the 14th International Joint Conference on Artificial Intelligence: IJCAI-95
- Kovacic, Zlatko. (2010). *Early Prediction of Student Success: Mining Students Enrolment Data*. Proceedings of Informing Science & IT Education Conference (InSITE).
- Lin, J. Imbrie P.K., and Reid Kenneth J. (2009). *Student Retention Modelling: An Evaluation of Different Methods and their Impact on Prediction Results*. Proceedings of the Research in Engineering Education Symposium.
- Maucary, Chady. (2011). *Data Mining for Engineering Schools Predicting Students' Performance and Enrollment in Masters Programs*. (IJACSA) International Journal of Advanced Computer Science and Applications.
- Oladokum, V.O., Adebajo, A.T., and Charles-Owaba. (2008). *Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course*. The Pacific Journal of Science and Technology.
- Olson, David, and Yong, Shi. (2008). *Pengantar Ilmu Penggalan Data Bisnis* (Chriswan Sungkono, Penerjemah). Jakarta : Salemba empat.
- Olson, David and Delen Dursun. (2008). *Advanced Data Mining Techniques*. Springer.
- Pusat Bahasa. (2008). *Kamus besar bahasa Indonesia*. Gramedia Pustaka Utama. Jakarta.
- Shannaq, Boumedyen., Rafael, Yusuppov., and Alexandro. (2010). *Student Relationship in Higher Education Using Data Mining Techniques*. Global Journal of Computer Science and Technology Vol. 10 Issue 11 (Ver. 1.0) October 2010.
- Slameto. (2010). *Belajar dan Faktor-Faktor yang Mempengaruhinya*. Rineka Cipta.
- Surjeet Kumar, Brijesh Bharadwaj, and Saurabh Pal. (2012). *Mining Education Data to Predict Student's Retention: A comparative Study*. (IJCSIS) International Journal of Computer Science and Information Security.
- Tan, P., M. Steibach, and V. Kumar. (2006). *Introduction to Data Mining*. Pearson Addison Wesley Boston.
- Turban, E., Aronson, J.E., & Liang, T. (2005). *Decision Support Systems and Intelligent Systems (7th ed.)*. Upper Saddle River, NJ: Pearson Education Inc.
- Weng, Fumei. (2010). *Modelling IT Student Retention at Taiwanese Higher Education Institutions*.thesis School of Business Information Technology and Logistics College of Business RMIT University 2010.
- Witten, Ian H. Eibe, Frank. And Hall, Mark A. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*. The Morgan Kaufmann series in data management systems
- Yadav, Surjeet Kumar. Bharadwaj, Brijesh, and Pal, Saurabh. (2012). *Mining Education Data to Predict Student's Retention: A comparative Study*. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 10, No. 2, 2012
- IBM, IBM SPSS Modeler CRISP-DM Guide, http://www.crisp-dm.org/software/analytics/spss/documentation/modeler/14.2/en/CRISP_DM.pdf (diunduh tanggal 03 Juli 2012).