SISTEM REKOMENDASI PROSES KELULUSAN MAHASISWA BERBASIS ALGORITMA KLASIFIKASI C4.5

Mujib Ridwan [1)]
[1] Program Studi Sistem Informasi, Universitas Islam Negeri (UIN) Sunan Ampel Surabaya
email: mujibrw@uinsby.ac.id

Abstract

Graduation is a series of processing stages that must be passed by each student, one of graduation requirement must complete the courses with a pre-determined amount, carrying out field work practice, research proposal exam, final exam and must complete several requirements and other requirements which are set by the college. This process should be completed within the allotted time, if not, the student will be drop-out declared. Therefore, it needs a system that can predict and evaluate the history of student course that history has been made to optimize the learning process of the next lecture. Input of this system is the master's student, student academic data, and historical data subjects which has been pursued by the student. The input data will be processed by using the techniques of data mining with C4.5 algorithm. The outputs of this system of classification is in the form of students' academic performance that predicted their graduation and providing recommendations for graduation process timely or in the most appropriate time with the optimal result based on historical subjects that have been taken. Testing on training data student sets produced values of precision, recall, and accuracy for C4.5 mining respectively 100%, 50%, and 75%.

Keywords: *graduation, recoomendations, and c4.5 algorithm*

1. PENDAHULUAN

Pada perguruan tinggi, mahasiswa menjadi salah satu parameter penting dalam evaluasi penyelenggaraan program studi. Monitoring absensi kehadiran mahasiswa, prestasi yang dicapai mahasiswa, peningkatan kompetensi mahasiswa, rasio kelulusan terhadap jumlah total mahasiswa, dan ketercapaian profil lulusan, seharusnya mendapatkan perhatian yang serius. Kuota mahasiswa yang diterima setiap tahun semakin bertambah, namun tidak semua mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang telah ditentukan, sehingga mengakibatkan penumpukan jumlah mahasiswa yang tidak lulus sesuai dengan masa studinya.

Kelulusan merupakan rangkaian tahapan proses yang harus dilalui oleh setiap mahasiswa, diantaranya harus menyelesaikan jumlah matakuliah yang telah ditentukan, melaksanakan kerja praktik, seminar proposal penelitian, seminar tugas akhir dan harus memenuhi beberapa persyaratan serta ketentuan lain yang telah ditetapkan oleh perguruan tinggi. Proses ini harus diselesaikan dalam jangka waktu yang telah ditentukan, jika tidak maka mahasiswa tersebut dinyatakan *drop-out*. Karena itu, perlu adanya sistem yang dapat memprediksi kelulusan mahasiswa dan mengevaluasi histori matakuliah yang telah ditempuh untuk mengoptimalkan proses perkuliahan berikutnya.

Dari identifikasi masalah di atas, diperlukan sebuah sistem untuk membantu mahasiswa dalam memprediksi masa studinya dengan cara mengevaluasi kinerja pada tahun pertama dan atau tahun kedua. Dari hasil prediksi tersebut kemudian dianalisis untuk dijadikan dasar rekomendasi kepada mahasiswa dalam mengambil keputusan perencanaan proses perkuliahan yang akan dijalani.

Algoritma C4.5 adalah salah satu algoritma klasifikasi yang populer pada kelompok algoritma pohon keputusan (Moertini, 2007). Algoritma C4.5 merupakan pengembangan dari algoritma *Iteractive Dichotomiser* 3 atau Induction of Decision 3 (ID3) yang diperkenalkan dan dikembangkan pertama kali oleh J. Ross Quinlan pada tahun 1979. Dibandingkan dengan algoritma yang lain, algoritma C4.5 memiliki kelebihan utama karena dapat menghasilkan model berupa pohon atau aturan yang mudah diinterpretasikan dan dikonversi ke aturan-aturan *Structure Query Language* (SQL), memiliki tingkat akurasi yang dapat diterima, dapat menangani atribut bertipe diskret dan

numerik, dan efisien dalam menangani atribut bertipe diskret (Han dan Kamber, 2006).

2. KAJIAN LITERATUR DAN PEGEMBANGAN HIPOTESIS

Kajian Literatur

Erdogen dan Timor (2005) dalam penelitiannya mengemukakan bahwa data mining adalah teknologi yang digunakan dalam disiplin ilmu yang berbeda-beda untuk mencari hubungan yang signifikan antara variabel dalam set data yang besar. Dalam penelitian ini, dikonsentrasikan pada penerapan data mining di lingkungan pendidikan. Hubungan antara data induk mahasiswa dan keberhasilan mereka dipelajari dengan teknik analisis klaster dan algoritma k-means. Hasil yang didapatkan adalah penggunaan teknik data mining dalam pendidikan dapat memberikan informasi tentang temuan yang lebih bervariasi dan signifikan, serta dapat meningkatkan kualitas pendidikan

Dalam jurnal IEEE tentang Model Prediksi Kinerja Akademik menggunakan Data Mining tahun 2011 yang dilakukan oleh Garcia dan Mora (2011) memberikan penjelasan secara rinci proses dan hasil penelitian tersebut dengan mengaplikasikan pada data mahasiswa di tahun pertama. Disimpulkan bahwa prediksi untuk kelompok rendah dan tinggi memiliki akurasi persentase yang signifikan dalam beberapa kasus, melebihi 70% jika Naive Bayes Classifier digunakan. Hal ini menunjukkan bahwa untuk mendapatkan model prediksi yang baik adalah sangat mungkin. Sebagai contoh, dapat digunakan untuk mendeteksi mahasiswa yang memiliki kinerja rendah dan mengambil keputusan yang tepat bahkan sebelum proses perkuliahan dimulai. Hal ini juga dapat digunakan untuk mendeteksi mahasiswa yang memiliki kinerja tinggi agar mereka memperoleh layanan pendidikan yang sesuai.

Pada Tahun 2012, Shovon Dan Haque melakukan penelitian untuk memprediksi kinerja akademik mahasiswa dengan menerapkan algoritma K-Means clustering. Dalam penelitian ini dilakukan proses evaluasi terhadap faktor penilaian mahasiswa seperti kuis, tugas, middle-test dan final-test. Setelah proses clustering, mahasiswa akan dikelompokkan ke dalam tiga kategori, yaitu kategori mahasiswa yang memiliki kinerja tinggi, sedang, dan rendah. Penelitian ini diharapkan akan menghasilkan informasi yang dapat membantu pendidik untuk mengurangi rasio drop-out dan meningkatkan kinerja mahasiswa dengan mengambil langkah yang tepat pada waktu yang tepat untuk meningkatkan kualitas pendidikan.

Kabakchieva (2012) juga melakukan penelitian untuk memprediksi kinerja mahasiswa namun menggunakan algoritma klasifikasi dalam teknik data mining yang dilakukan di salah satu universitas terkenal dan bergengsi di Bulgaria. Penelitian ini bertujuan utama untuk mengungkapkan potensi dari aplikasi data mining untuk manajemen universitas, memberikan kontribusi untuk mempromosikan universitas secara efisien, dan memperoleh calon mahasiswa sesuai dengan kategori yang diinginkan. Penelitian ini difokuskan pada pengembangan model data mining untuk memprediksi kinerja mahasiswa, berdasarkan pada data pribadi mereka, pra-universitas dan karakteristik kinerja universitas. Dataset yang digunakan untuk tujuan penelitian meliputi data tentang mahasiswa dalam tiga tahun berturut-turut. Beberapa algoritma tentang data mining klasifikasi yang terkenal, meliputi decision tree classifier, neural network, dan nearest neighbour classifier akan diterapkan pada dataset yang kemudian akan dianalisis dan dibandingan hasil kinerja masing-masing algoritma.

Ridwan, Suyono, dan Sarosa (2013) menyebutkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa yaitu jenis kelamin, Indeks Prestasi (IP) semester 1, IP semester 4, Indeks Prestasi Komulatif (IPK). Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi.

Hipotesis

Penerapan data mining menggunakan algoritma C4.5 dapat digunakan untuk mengelompokkan kinerja akademik mahasiswa ke dalam kategori “lulus tepat waktu” atau “lulus tidak tepat waktu”. Kemudian dari hasil pengelompokan tersebut, sistem akan memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat dengan nilai optimal. Sistem ini juga akan mempermudah pihak pengambil kebijakan dalam mengevaluasi kinerja akademik mahasiswa, sehingga dapat dijadikan bahan pertimbangan dalam mengambil

keputusan untuk mengoptimalkan proses kelulusan mahasiswa.

Algoritma Klasifikasi C4.5

Algoritma C4.5 merupakan salah satu algoritma yang dapat digunakan dalam pembentukan pohon keputusan. Algoritma C4.5 merupakan pengembangan dari algoritma Iteractive Dichotomiser 3 atau Induction of Decision 3 (ID3) yang diperkenalkan dan dikembangkan pertama kali oleh J. Ross Quinlan pada tahun 1979. Dalam prosedur algoritma ID3, input berupa sampel training, label training, dan atribut. Idenya adalah membuat pohon dengan percabangan awal adalah atribut yang signifikan. Maksud signifikan adalah yang paling bisa mempartisi antara keputusan ya atau tidak.

Algoritma ID3 dikembangkan lagi menjadi algoritma C4.5. Sebagai bentuk pengembangan dari ID3, algoritma C4.5 memiliki beberapa keunggulan antara lain:

- Mampu menangani atribut dengan tipe diskret atau kontinu.
- Mampu menangani atribut yang kososng (missing value).
- Dapat memangkas pohon.

Algoritma ID3 dan C4.5 dikembangkan lagi oleh J. Ross Quinlan menjadi algoritma C5.0. Dibandingkan dengan algoritma pendahulunya, algoritma C5.0 memiliki beberapa keunggulan fitur di antaranya:

- Algoritma C5.0 dirancang untuk menganalisis basis data subtansial yang berisi puluhan sampai ratusan record dan satuan hingga ratusan field numerik dan nominal.
- Untuk memaksimumkan tingkat penafsiran pengguna terhadap hasil yang disajikan, maka klasifikasi algoritma C5.0 disajikan dalam dua bentuk, menggunakan pohon keputusan dan sekumpulan aturan IF-THEN yang lebih mudah untuk dimengerti.
- Algoritma C5.0 mudah digunakan dan tidak membutuhkan pengetahuan tinggi tentang statistik atau *machine learning*.

Meskipun sebagai bentuk pengembangan dari algoritma C4.5, penggunaan algoritma C5.0 masih memiliki frekuensi yang lebih kecil dibandingkan dengan algoritma C4.5. Menurut Murtini (2007) Algoritma C4.5 adalah salah satu algoritma klasifikasi yang populer pada kelompok algoritma pohon keputusan.

Dibandingkan dengan algoritma yang lain, algoritma C4.5 memiliki kelebihan utama karena dapat menghasilkan model berupa pohon atau aturan yang mudah diinterpretasikan dan dikonversi ke aturan-aturan Structure Query Language (SQL), memiliki tingkat akurasi yang dapat diterima, dapat menangani atribut bertipe diskret dan numerik, dan efisien dalam menangani atribut bertipe diskret (Han dan Kamber, 2006)

Menurut Kusrini dan Luthfi (2009), secara umum algoritma C4.5 untuk membangun pohon keputusan dalam data mining adalah sebagai berikut:

- Pilih atribut sebagai simpul akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai simpul, baik simpul akar atau simpul internal didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai Gain digunakan rumus seperti tertera dalam Persamaan di bawah ini. Nilai Gain berguna untuk mengukur efektivitas atributnya dalam proses klasifikasi.

$$\text{Gain}(S,A)=\text{Entropy}(S)-\sum_{i=1}^{n}\frac{|S_i|}{|S|}*\text{Entropy}(S_i)$$

dengan:
S : himpunan kasus
A : atribut
n : jumlah partisi atribut A
$|S_i|$: jumlah kasus pada partisi ke-i
$|S|$: jumlah kasus dalam S.

Untuk menghitung nilai Entropy dapat dilihat pada Persamaan berikut ini. Entropy digunakan sebagai suatu parameter untuk mengukur keragaman dari suatu sampel.

$$\text{Entropy}(S)=\sum_{i=1}^{n}-pi*\log_2 pi$$

dengan:
n : jumlah partisi S
pi : proporsi dari S_i terhadap S

3. METODOLOGI PENELITIAN

Data

Pada penelitian ini akan digunakan data sekunder, yaitu data yang berasal dari referensi atau data dummy. Data tersebut meliputi:

1. Data training dan data testing
   Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orangtua, IP semester 1-4, IPK semester 1-4, dan keterangan lulus.
2. Data riwayat akademik
   Data ini memiliki atribut NIM, kode matakuliah, nama matakuliah, sks matakuliah, tahun akademik, semester akademik, nilai matakuliah, dan keterangan

Teknik Pengumpulan Data

Bahwa untuk memperoleh data yang menunjang penelitian ini, maka akan digunakan teknik dalam pengumpulan data, yaitu :

1. Teknik Observasi
   Observasi mengamati dengan panca indera manusia (penglihatan dan pendengaran) untuk menangkap gejala yang diamati. Apa yang ditangkap tadi, dicatat dan selanjutnya catatan tersebut dianalisis. Observasi bertujuan untuk menjawab masalah penelitian, peneliti terlebih dahulu melakukan pengamatan/observasi, yakni mengamati gejala yang diteliti.
2. Teknik Wawancara
   Teknik wawancara yang dipergunakan peneliti adalah teknik wawancara tidak terstruktur, artinya pedoman wawancara yang hanya memuat garis besar yang akan ditanyakan. Tentu saja kreativitas pewawancara sangat diperlukan, bahkan hasil wawancara dengan jenis pedoman ini lebih banyak bergantung dari pewawancara. Pewawancaralah sebagai pengemudi jawaban responden (Arikunto, 2006).
   Teknik wawancara ini bertujuan untuk mendapatkan data tentang kegiatan percakapan antara pewawancara dan yang diwawancarai dengan maksud untuk mendapatkan informasi mengenai hal yang berkaitan dengan proses kelulusan dan evaluasi kinerja akademik mahasiswa.
3. Teknik Dokumentasi
   Teknik dokumentasi adalah mencari data mengenai hal-hal atau variabel yang berupa catatan, transkrip, buku, surat kabar, majalah, prasasti, notulen rapat, agenda, dan sebagainya (Arikunto, 2006). Hal ini dilakukan untuk memperoleh hasil dokumentasi yang ada pada proses kelulusan dan evaluasi kinerja akademik mahasiswa.
   Dokumen dan *record* digunakan untuk keperluan penelitian karena alasan-alasan yang dapat dipertanggungjawabkan, yaitu: Dokumen dan *record* digunakan karena merupakan sumber yang stabil dan mendukung; Berguna sebagai bukti untuk pengujian.

Definisi Operasional Variabel

Secara garis besar variabel yang digunakan dalam penelitian meliputi data pribadi, data pra-universitas, data nilai ujian, dan data indeks prestasi. Maka dari itu, pada penelitian ini ditetapkan variabel-variabel yang akan digunakan adalah sebagai berikut:

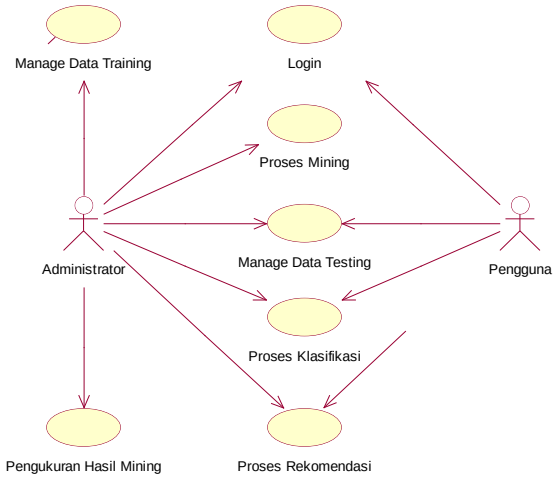
- Jenis kelamin: bernilai Laki-laki atau Perempuan.
- Asal sekolah: bernilai SMA, SMAN, SMK, SMKN, MA, atau MAN.
- Jalur masuk: bernilai SNMPTN, Mandiri atau Undangan.
- Nilai UNAS: bernilai rendah, sedang, atau tinggi.
- Gaji orang tua: bernilai, rendah, sedang, atau tinggi.
- IPS1, IPS2, IPS3, dan IPS4: bernilai A, B, C, atau D
- IPK semester 1 sampai 4: bernilai A, B, C, atau D.
- Keterangan lulus: bernilai lulus tepat waktu atau lulus tidak tepat waktu
- Nilai matakuliah: bernilai A, B, C, D, atau E.

Pertimbangan dalam penetapan pemilihan variabel-variabel tersebut didasarkan bahwa nilai untuk masing-masing variabel tidak banyak bervariasi. Karena itu, diharapkan mahasiswa yang masuk dalam satu klasifikasi nilai variabel tersebut cukup banyak. Misalnya variabel "Asal Sekolah" jumlah nilai variabelnya ada enam, yaitu SMA, SMAN, SMK, SMKN, MA, atau MAN, sehingga jumlah mahasiswa dengan nilai variabel tersebut cukup banyak. Maka dari itu, variabel seperti "Nama Mahasiswa" dan "Alamat Asal" tidak dipilih karena jumlah nilai variabelnya cukup banyak.

Pemodelan Sistem

Langkah yang harus dilakukan untuk menangkap kebutuhan pengguna adalah memodelkan sistem dengan menggunakan use

case diagram. Diagram ini menggambarkan tingkah laku sistem yang akan dibuat. Use case diagram mendeskripsikan sebuah interaksi satu atau lebih aktor dengan sistem yang akan dibuat. Secara garis besar, use case diagram dalam sistem ini digunakan untuk mengetahui fungsi apa saja yang ada di dalam sistem dan siapa saja yang memiliki hak akses menggunakan fungsi-fungsi itu.



Gambar 1 Use Case Diagram

Dari Gambar 1 dapat diketahui use case diagram dari sistem ini terdiri atas:

- Actor, merupakan pengguna sistem yang dapat berupa orang, proses, hardware, atau sistem lain yang berinteraksi dengan sistem yang akan dibuat dan berada diluar sistem yang akan dibuat itu sendiri. Aktor dapat memasukan informasi ke sistem, menerima informasi dari sistem, atau juga keduanya. Dalam sistem ini, aktor meliputi administrator dan pengguna.
- Use case, merupakan fungsionalitas atau perilaku yang disediakan sistem, termasuk interaksi antara aktor dengan sistem tersebut. Dalam sistem ini, use case yang berhubungan dengan aktor administrator meliputi login, manage data training, proses mining, pengukuran hasil mining, manage data testing, proses klasifikasi, dan proses rekomendasi.

Sedangkan use case yang berhubungan dengan aktor pengguna meliputi login, manage data testing, proses klasifikasi, dan proses rekomendasi.

4. HASIL DAN PEMBAHASAN

Implementasi Sistem

Pada tahap implementasi ini, data yang akan digunakan sudah dibersihkan dan ditranformasikan dalam bentuk kategori, misalkan untuk data Indeks Prestasi (IP) sudah dikategorikan menjadi nilai A, B, C, dan D dimana:

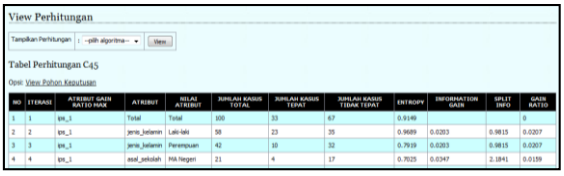
A : untuk kategori IP 3.50 – 4.00
B : untuk kategori IP 2.75 – 3.49
C : untuk kategori IP 2.00 – 2.74
D : untuk kategori IP yang kurang dari 2.00.

Data ditransformasikan menjadi bentuk kategori dengan tujuan untuk menghindari adanya missing value pada data. Dalam pengujian ini digunakan data sampel mahasiswa angkatan 2005-2011 yang sudah dinyatakan lulus dari perguruan tinggi. Jumlah data yang digunakan adalah 100 data dengan kelas LTW (Lulus Tepat Waktu) dan LTTW (Lulus Tidak Tepat Waktu) masing-masing berjumlah 33 dan 67.

Dalam proses pengujian, akan digunakan dua jenis data, yaitu data *training* dan data *testing*. Pada algoritma C4.5, data *training* digunakan untuk membentuk pohon keputusan, dan data *testing* digunakan untuk menguji aturan yang terbentuk pada *decision tree*.

Klasifikasi dengan Algoritma C4.5

Uji coba mining C4.5 betujuan untuk mengetahui unjuk kerja dari algoritma C4.5 dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Pada pengujian ini, diberikan data *training* untuk membentuk pohon keputusan dan rule. Langkah selanjutnya akan diberikan data *testing* untuk menguji rule atau aturan yang sudah terbentuk dari proses *mining*. Unjuk kerja diperoleh dengan memberikan nilai pada confusion matrix untuk menghitung nilai precision, recall, dan accuracy dari hasil pengujian. Berikut hasil pengujian Menggunakan data latih sebanyak 20 data sampel 2005 dengan data uji sebanyak 20 set data. Setelah proses *import* data training, berikut hasil perhitungan dari 20 data training pada aplikasi:



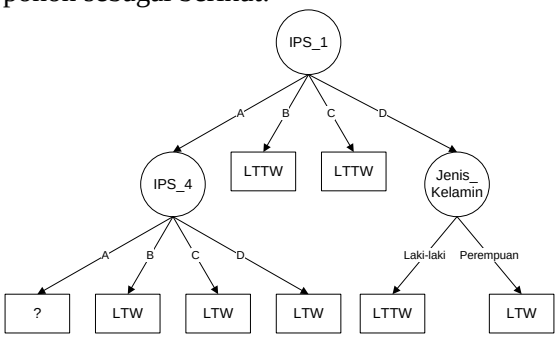
Gambar 2 View Hasil Testing

Pada percobaan ini diperoleh pohon keputusan berikut:

```
ips_1 = A (Tepat = 11, Tidak Tepat = 5) : ?
|   ips_4 = A (Tepat = 1, Tidak Tepat = 3) : ?
|   ips_4 = B (Tepat = 10, Tidak Tepat = 2) : LTW
|   ips_4 = C (Tepat = 0, Tidak Tepat = 0) : LTW
|   ips_4 = D (Tepat = 0, Tidak Tepat = 0) : LTW
ips_1 = B (Tepat = 17, Tidak Tepat = 39) : LTTW
ips_1 = C (Tepat = 4, Tidak Tepat = 22) : LTTW
ips_1 = D (Tepat = 1, Tidak Tepat = 1) : ?
|   jenis_kelamin = Laki-laki (Tepat = 0, Tidak Tepat = 1) :
LTTW
|   jenis_kelamin = Perempuan (Tepat = 1, Tidak Tepat = 0) :
LTW
```

Gambar 3 Pohon Keputusan

Pada percobaan ini juga didapat diagram pohon sebagai berikut:



Gambar 4 Diagram Pohon

Sehingga aturan yang terbentuk dari percobaan ini adalah sebagai berikut:

- jika ips_1 = B, maka : LTTW
- jika ips_1 = C, maka : LTTW
- jika ips_1 = A AND ips_4 = B, maka : LTW
- jika ips_1 = A AND ips_4 = C, maka : LTW
- jika ips_1 = A AND ips_4 = D, maka : LTW
- jika ips_1 = D AND jenis_kelamin = Laki-laki, maka : LTTW
- jika ips_1 = D AND jenis_kelamin = Perempuan, maka : LTW

Perhitungan *confusion matrix* berikut:

		Predicted class	
		LTW	LTTW
Actual class	LTW	5	5
	LTTW	0	10

Precision: P = 5 / (5 + 0) x 100% = 100.00%
Recall: R = 5 / (5 + 5) x 100% = 50.00%
Accuracy: A = (5 + 10) / 20 x 100% = 75.00%

Proses Rekomendasi

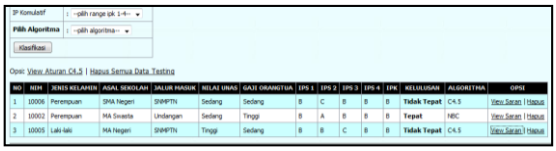
Setelah proses klasifikasi, selanjutnya akan diberikan saran atau rekomendasi untuk proses kelulusan. Rekomendasi ini difokuskan untuk data target yang diklasifikasikan ke dalam kelas "lulus tidak tepat waktu" karena diharapkan dari rekomendasi yang diberikan, maka mahasiswa dengan data target tersebut dalam lulus dalam waktu yang tepat dengan nilai optimal.

Dalam proses rekomendasi, dibutuhkan analisis terhadap data riwayat akademik dari mahasiswa yang bersangkutan. Berikut adalah beberapa hal yang diolah pada proses rekomendasi oleh sistem:

- Informasi total sks yang sedang ditempuh, total sks yang sudah ditempuh, total sks yang belum ditempuh, dan IPK sementara.
- Saran mengulang matakuliah atau mengikuti Semester Pendek (SP), jika dalam riwayat akademik ditemukan matakuliah yang bernilai D atau E.
- Saran melaksanakan Praktik Kerja Lapangan (PKL), jika jumlah sks yang ditempuh sudah lebih dari 80 sks dengan nilai ≥ C.
- Saran seminar proposal tugas akhir, jika mahasiswa tersebut sudah melaksanakan PKL dan nilainya ≥ C.
- Saran ujian komprehensif, jika semua matakuliah kecuali tugas akhir sudah ditempuh semua dengan nilai ≥ C.
- Saran ujian tugas akhir, jika semua matakuliah kecuali tugas akhir sudah ditempuh semua dengan nilai ≥ C dan mahasiswa yang bersangkutan memprogram matakuliah Skripsi.

Gambar 5 dan Gambar 6 adalah tampilan dari proses rekomendasi:

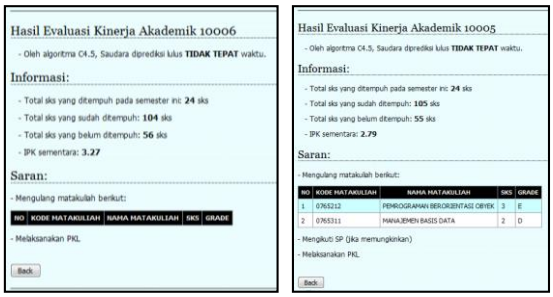
1. Gambar 5 adalah Hasil klasifikasi data target, klasifikasi ini diperoleh dari proses memasukkan data target dalam form testing kemudian gunakan algoritma C4.5 sebagai classifier-nya.



Gambar 5 Hasil Klasifikasi Data Target

2. Gambar 6 adalah Hasil evaluasi kinerja akademik mahasiswa, dalam form ini berisi prediksi kelulusan, informasi sks, dan saran untuk proses kelulusan. Informasi dan saran akan muncul jika data target diklasifikasi "lulus tidak tepat waktu". Tampilan informasi dan saran dapat dilihat seperti pada Gambar 7.

Gambar 6 Hasil Evaluasi "Lulus Tepat Waktu"



Gambar 7 Hasil Evaluasi "Lulus Tidak Tepat Waktu"

5. KESIMPULAN

Hasil dari pengujian klasifikasi menggunakan algoritma C4.5 menunjukkan nilai akurasi dalam kategori *fair classification*, hal ini disebabkan karena faktor penentu kelulusan mahasiswa tepat atau tidak tepat waktu pada kenyataannya memiliki nilai yang tidak konsisten. Dari hasil klasifikasi, jika kelas data target adalah "lulus tidak tepat waktu", maka saran untuk proses perkuliahan berikutnya akan diberikan. Saran-saran tersebut diberikan dengan menganalisis riwayat akademik mahasiswa yang menjadi data target tersebut.

6. REFERENSI

[1] Arikunto, S. 2006. *Prosedur Penelitian: Suatu Pendekatan Praktik Edisi Revisi VI*. Jakarta: PT Rineka Cipta.

[2] Erdogan, S.Z. dan M. Timor. 2005. "A Data Mining Application In A Student Database". Journal Of Aeronautics And Space Technologies. Volume 2 Number 2: 53-57.

[3] Garcia, E.P.I. dan P.M. Mora. 2011. "Model Prediction of Academic Performance for First Year Students". *IEEE Computer Society*.

[4] Han, J. dan M. Kamber. 2006. *Data Mining: Concepts and Techniques, Second Edition*. San Francisco: Morgan Kaufmann Publishers.

[5] Kabakchieva, D. 2012. "Student Performance Prediction by Using Data Mining Classification Algorithms". *IJCSMR*. Vol 1 Issue 4: 686-690.

[6] Kusrini dan E.M. Luthfi. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit ANDI.

[7] Moertini, V.S. 2007. "Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 dengan Pendekatan Konsep Operator Relasi (Studi Kasus: Pra-Pengolahan dan Klasifikasi Citra Batik)". http://home.unpar.ac.id/~moertini/DisertasiVeronicaSM.pdf. 30 Januari 2013.

[8] Shovon, Md.H.I. dan M. Haque. 2012. "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm". IJARCSSE. Volume 2 Issue 7: 353-355.

[9] Ridwan, M., Suyono, Hadi dan Sarosa, M. 2013. "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma NBC". Jurnal EECCIS Vol.7, No. 1.