
Aplikasi Berbasis Sistem Pakar Untuk Memprediksi Ketepatan Masa Studi Dengan Pendekatan Case Based Reasoning dan C45

EXPERT SYSTEM BASED APPLICATION TO PREDICT GRADUATION TIMELINESS BY USING CASE BASED REASONING AND C45 APPROACH

Hartini Suciati¹, Fajrian Nur Adnan²

^{1,2}Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

^{1,2}Jalan Imam Bonjol No 207, Semarang, 50131, (024) 3517261

e-mail : ¹hartinisuciati@gmail.com, ²fajrian@dsn.dinus.ac.id

Abstrak

Menjaga kualitas dan ketepatan peserta didik dalam menyelesaikan masa studi menjadi salah satu ukuran kualitas sebuah institusi pendidikan. Identifikasi dan penanganan sejak dini terhadap mahasiswa yang bermasalah dibidang akademik merupakan salah satu cara menjaga ketepatan kelulusan peserta didik tetap terjaga. Penelitian ini dilakukan dalam rangka memberikan sarana kepada pengelola di institusi pendidikan untuk dapat memonitoring dan memprediksi ketepatan masa studi setiap peserta didiknya. Dengan begitu pengelola institusi dapat memberikan penanganan lebih awal terhadap peserta didik yang teridentifikasi akan mengalami keterlambatan dalam menyelesaikan masa studinya. Aplikasi berbasis sistem pakar dikembangkan dengan pendekatan Case Based Reasoning, dimana dalam pengolahan faktanya diperoleh dari hasil pengolahan dengan menggunakan algoritma C4.5. Adapun data yang digunakan dalam pembentukan representasi pengetahuan adalah 698 data mahasiswa sistem informasi SI Universitas Dian Nuswantoro angkatan 2008 sampai dengan 2011 yang telah dinyatakan lulus, dengan menggunakan indeks prestasi semester 1 hingga semester 4 sebagai parameter. Berdasarkan pengujian yang dilakukan sebanyak 10 kali, aplikasi tersebut, dapat memprediksi ketepatan masa studi mahasiswa dengan benar sebesar 90%.

Kata kunci—Prediksi Ketepatan Masa Studi, Sistem Pakar, Case Based Reasoning, Algoritma C4.5

Abstract

Keeping the quality and students' timeliness in finishing their study period is one of educational institution quality measure. Identification and early handling towards students with academic problems is one way to keep students' graduation timeliness. This research is conducted to give medium for educational institution administrators to monitor and predict their students' study period. So that, institution administrators can give early handling towards students which are identified will go through tardiness in finishing their study period. Expert System Based Application is developed using Case Based Reasoning, where its fact processing is gained by alogarithm C4.5 processing result. The data used in the formation of knowledge representations are 698 Information System Students (Undergraduate Program) of Dian Nuswantoro University batch 2008 until 2011 who have been graduated, using semester 1 until semester 4 GPA as the parameter. Based on testing that have done in 10 times, the application can predict students' graduation timeliness 90% correct.

Keywords—Timeliness of Graduation Prediction, Expert System, Case Based Reasoning, C4.5 Algorithm

1. PENDAHULUAN

Berdasarkan ketetapan Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi tentang Sistem Pendidikan Tinggi disebutkan bahwa untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144-160 satuan kredit semester (sks) dengan masa studi selama 8-12 semester atau 4-6 tahun [1]. Namun faktanya tidak sedikit mahasiswa yang menempuh masa studi melebihi standar kelulusan yang telah ditetapkan. Minimnya jumlah mahasiswa yang lulus tepat waktu menyebabkan turunya kualitas dan mutu universitas, selain itu mahasiswa yang terlalu lama menempuh masa studi memiliki kecenderungan terkena ancaman Drop-Out (DO).

Arief Jananto (2013) pernah melakukan penelitian mengenai masa studi mahasiswa. Metode yang digunakan adalah algoritma Naïve Bayes dan atribut yang digunakan meliputi indeks prestasi semester 1, indeks prestasi semester 2, indeks prestasi semester 3, indeks prestasi semester 4, jenis kelamin, kota lahir, tipe sekolah, dan kota sekolah. Hasil yang didapat yaitu rata-rata tingkat kesalahan mencapai 20% hingga 34% [2].

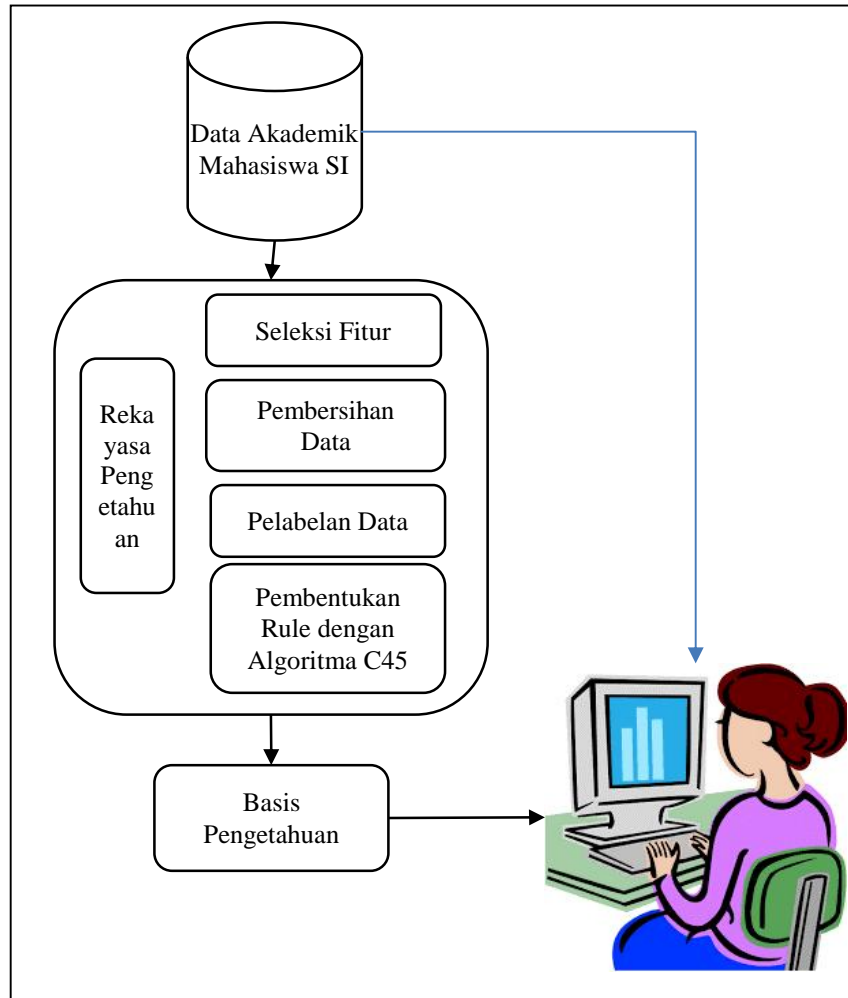
Penelitian tentang sistem pakar juga pernah dilakukan oleh Diki Andita Kusuma dan Chairani (2014) mengenai mahal biaya yang harus dikeluarkan untuk berkonsultasi ke dokter spesialis penyakit paru-paru. Metode yang digunakan adalah Case Based Reasoning. Kasus yang digunakan berjumlah 8 kasus lama dan 1 kasus baru yang akan dihitung nilai kedekatannya menggunakan algoritma Nearest Neighbor dimana kedekatan paling maksimum diperoleh pada kasus ke-6 yaitu sebesar 0.93 atau 93% [3].

Permasalahan tentang pentingnya menjaga ketepatan masa studi peserta didik mendasari peneliti untuk memprediksi ketepatan masa studi mahasiswa Sistem Informasi - S1 Universitas Dian Nuswantoro sehingga dapat meminimalisir keterlambatan kelulusan bagi mahasiswa. Penelitian ini menggunakan pendekatan Case Based Reasoning dan menggunakan algoritma C4.5 sebagai metode dalam rekayasa pengetahuan.

2. METODE PENELITIAN

Dalam penelitian ini, aplikasi yang akan dikembangkan menerapkan konsep dari Sistem Pakar, dimana terdiri dari :

1. Pengetahuan berupa fakta yang diperoleh dari data lulusan mahasiswa Sistem Informasi di Universitas Dian Nuswantoro
 2. Rekayasa Pengetahuan dengan Menggunakan Algoritma C45
 3. Basis Pengetahuan berupa Rule yang diperoleh dari hasil pengolahan fakta dengan menggunakan algoritma C4.5
 4. User Interface
-



Gambar 1. Skema Aplikasi Berbasis Sistem Pakar

Sistem Pakar yang diadopsi pada aplikasi ini menggunakan pendekatan penalaran berbasis kasus (*Case-Based Reasoning*), dimana pada bentuk ini, basis akan berisi solusi-solusi yang sudah dicapai sebelumnya. Setelah itu, dibuat sebuah solusi untuk keadaan yang terjadi sekarang (fakta yang ada). Pada kasus ini, fakta atau solusi yang digunakan adalah data akademis dan data kelulusan mahasiswa sejumlah 698 dari mahasiswa Program Studi Sistem Informasi S1 Universitas Dian Nuswantoro angkatan 2008 sampai dengan 2011 yang telah dinyatakan lulus.

Pada tahap berikutnya proses rekayasa pengetahuan dilakukan dengan C45. Pada tahap ini, terdapat beberapa sub tahapan sebelum data siap diolah dengan menggunakan algoritma C45. Adapun tahapan persiapannya adalah sebagai berikut:

1. Seleksi Fitur. Dari 698 data yang digunakan, atribut yang digunakan sebagai parameter dalam pembentukan basis pengetahuan terdiri dari 5 atribut, yaitu nomor induk mahasiswa dan indeks prestasi semester 1 hingga semester 4 sebagai parameternya.

nim	ips_1	ips_2	ips_3	ips_4
A12.2008.03281	3.1	2.84	3	3.3
A12.2008.03359	2.05	2.63	3.48	3.26
A12.2008.02958	2.4	2.42	2.4	2.1
A12.2008.03022	4	3.7	3.58	3.81
A12.2008.03063	2.4	2.79	2.86	2.86
A12.2008.03128	2.56	2.81	1.9	3.4
A12.2008.03035	3.32	3.27	3.38	3.31
A12.2008.02986	2.45	2.26	2.58	3.24
A12.2008.03243	3.22	3.32	2.6	3.13
A12.2008.03001	1.6	2.16	1.47	1.82
A12.2008.03272	2.6	2.16	1.16	1.89
A12.2008.03028	3.8	3.05	3.39	3.57
A12.2008.03253	2	2.11	0.63	1.11
A12.2008.03240	2.75	2.37	2.37	2.75
A12.2008.03067	2.55	2.32	2.47	2.37
A12.2008.03269	2.65	2.21	2.84	2.62
A12.2008.03266	1.95	0.79	2.12	2.5
A12.2008.03244	2.65	2.47	1.84	0.33
A12.2008.03145	2.1	2.53	2.62	2.37
A12.2008.02982	2.2	2.95	3	2.38
A12.2008.02965	2.3	2	0.32	1.46
A12.2008.03309	2.8	2.26	2.42	1.71
A12.2008.03123	2.25	2.84	1.95	2.83
A12.2008.02969	2.25	3.42	2.33	2.37

2. Pembersihan Data

Data yang digunakan adalah data yang memiliki nilai atribut indeks prestasi semester 1 sampai dengan indeks prestasi semester 4 lengkap atau tidak bernilai 0.

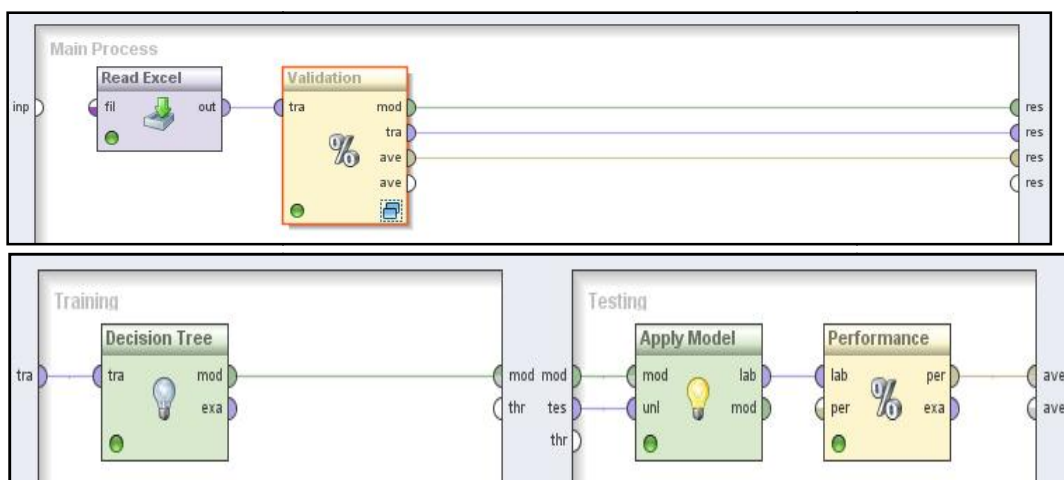
3. Pelabelan Data. Data yang telah diperoleh akan dibedakan menjadi 2 kategori yaitu "TEPAT WAKTU" dan "TIDAK TEPAT WAKTU". Pengkategorian dilakukan dengan mempertimbangkan Masa Studi yang diperoleh dari tahun masuk dan tahun kelulusan.

a. Data dikategorikan "TEPAT WAKTU" apabila Masa Studi = 4 Tahun

b. Data dikategorikan "TIDAK TEPAT WAKTU" apabila Masa Studi > 4 Tahun

nim	ips_1	ips_2	ips_3	ips_4	label
A12.2008.03281	3.1	2.84	3	3.3	TEPAT WAKTU
A12.2008.03359	2.05	2.63	3.48	3.26	TEPAT WAKTU
A12.2008.02958	2.4	2.42	2.4	2.1	TIDAK TEPAT WAKTU
A12.2008.03022	4	3.7	3.58	3.81	TEPAT WAKTU
A12.2008.03063	2.4	2.79	2.86	2.86	TEPAT WAKTU
A12.2008.03128	2.56	2.81	1.9	3.4	TIDAK TEPAT WAKTU
A12.2008.03035	3.32	3.27	3.38	3.31	TEPAT WAKTU
A12.2008.02986	2.45	2.26	2.58	3.24	TIDAK TEPAT WAKTU
A12.2008.03243	3.22	3.32	2.6	3.13	TIDAK TEPAT WAKTU
A12.2008.03001	1.6	2.16	1.47	1.82	TEPAT WAKTU
A12.2008.03272	2.6	2.16	1.16	1.89	TIDAK TEPAT WAKTU
A12.2008.03028	3.8	3.05	3.39	3.57	TIDAK TEPAT WAKTU
A12.2008.03253	2	2.11	0.63	1.11	TIDAK TEPAT WAKTU
A12.2008.03240	2.75	2.37	2.37	2.75	TEPAT WAKTU
A12.2008.03067	2.55	2.32	2.47	2.37	TIDAK TEPAT WAKTU
A12.2008.03269	2.65	2.21	2.84	2.62	TEPAT WAKTU
A12.2008.03266	1.95	0.79	2.12	2.5	TEPAT WAKTU
A12.2008.03244	2.65	2.47	1.84	0.33	TIDAK TEPAT WAKTU
A12.2008.03145	2.1	2.53	2.62	2.37	TIDAK TEPAT WAKTU
A12.2008.02982	2.2	2.95	3	2.38	TIDAK TEPAT WAKTU
A12.2008.02965	2.3	2	0.32	1.46	TIDAK TEPAT WAKTU
A12.2008.03309	2.8	2.26	2.42	1.71	TIDAK TEPAT WAKTU
A12.2008.03123	2.25	2.84	1.95	2.83	TIDAK TEPAT WAKTU
A12.2008.02969	2.25	3.42	2.33	2.37	TIDAK TEPAT WAKTU

4. Pembentukan Rule dengan C4.5 dengan operator Decision Tree, dengan menggunakan RapidMiner



Gambar 4. Pemodelan Decision Tree dengan RapidMiner - Algoritma C45

Adapun perhitungan manual dari algoritma C45 dapat dilakukan dengan menggunakan tahapan sebagai berikut: [6]

1. Menyiapkan data training.
2. Mencari nilai *entropy* menggunakan rumus *entropy*. Setelah *entropy* didapat kemudian mencari nilai *gain* dari setiap atribut yang ada. Atribut yang memiliki nilai *gain* terbesar akan digunakan sebagai akar pohon.

Rumus mencari nilai *entropy* :

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

Keterangan:

S : Himpunan kasus

n : Jumlah partisi S

Pi : Proporsi Si kepada S

Rumus mencari nilai *gain* :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

S : Himpunan kasus

A : fitur

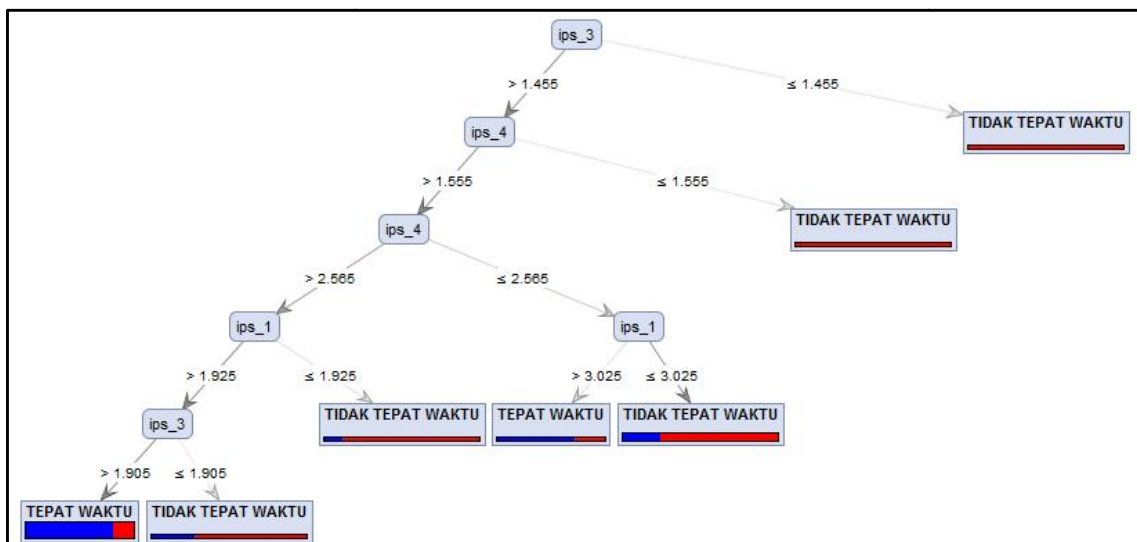
n : Jumlah partisi atribut A

|Si| : Proporsi Si terhadap S

|S| : Jumlah kasus dalam S

3. Ulanglangkah ke-2 sampai seluruh record terpartisi.
4. Langkah-langkah ini akan berakhir apabila seluruh record pada simpul N memiliki kelas sama serta tidak ada atribut cabang yang koso

Tahap selanjutnya diperoleh sebuah basis pengetahuan baru berupa rule hasil dari rekayasa pengetahuan dengan menggunakan algoritma C4.5. Adapun basis data yang terbetuka dalah sebagai berikut:



Gambar Pohon Keputusan Hasil Rekayasa Pengetahuan Dengan Algoritma C45

```

ips_3 > 1.455
|
|   ips_4 > 1.555
|   |
|   |   ips_4 > 2.565
|   |   |
|   |   |   ips_1 > 1.925
|   |   |   |
|   |   |   |   ips_3 > 1.905: TEPAT WAKTU {TEPAT
WAKTU=384, TIDAK TEPAT WAKTU=90}
|   |   |   |   ips_3 1.905: TIDAK TEPAT WAKTU {TEPAT
WAKTU=2, TIDAK TEPAT WAKTU=5}
|   |   |   |   ips_1 1.925: TIDAK TEPAT WAKTU {TEPAT
WAKTU=1, TIDAK TEPAT WAKTU=7}
|   |   |   |   ips_4 2.565
|   |   |   |   |
|   |   |   |   |   ips_1 > 3.025: TEPAT WAKTU {TEPAT WAKTU=8,
TIDAK TEPAT WAKTU=3}
|   |   |   |   |   ips_1 3.025: TIDAK TEPAT WAKTU {TEPAT
WAKTU=40, TIDAK TEPAT WAKTU=121}
|   |   |   |   |   ips_4 1.555: TIDAK TEPAT WAKTU {TEPAT WAKTU=0,
TIDAK TEPAT WAKTU=10}
|   |   |   |   |   ips_3 1.455: TIDAK TEPAT WAKTU {TEPAT WAKTU=0, TIDAK
TEPAT WAKTU=27}

```

Gambar Basis Data berupa rule hasil Rekayasa Pengetahuan Dengan Algoritma C45

Tahap yang terakhir adalah pengembangan aplikasi berbasis Sistem Pakar dengan menerapkan basis pengetahuan yang telah diperoleh. Aplikasi yang dikembangkan dapat di integrasikan dengan basis data yang menampung data akademik mahasiswa. Adapun data yang diakses langsung dari basis data yang ada adalah Nim. Nama. IP Semester 1 hingga Semester 4 sebagai bahan prediksi ketepatan masa studi. Dan untuk memfasilitasi pengguna dalam mengakses dan melihat hasil prediksi tersebut, maka telah dikembangkan antarmuka sebagai berikut:



Gambar Antarmuka Menu

prediku

Nim :

**Format penulisan Nim (A12.20xx.xxxxx)*
(Tekan Tab untuk mulai prediksi)

Data Mahasiswa

Nama :

IPS 1 : **IPS 3 :**

IPS 2 : **IPS 4 :**

Berdasarkan hasil prediksi, waktu kelulusan mahasiswa tersebut adalah :

Clear

Gambar Antarmuka Prediksi Berdasarkan NIM

prediku

Prediksi Ketepatan Waktu Kelulusan Mahasiswa Berdasarkan Tahun Angkatan

Pilih data mahasiswa yang akan diprediksi

All

Tahun Angkatan

Nim	Nama	IPS 1	IPS 2	IPS 3	IPS 4	Waktu

Cancel

Gambar Antarmuka Prediksi Seluruh Mahasiswa atau Berdasarkan Tahun Angkatan

3. HASIL DAN PEMBAHASAN

Dari aplikasi yang telah dikembangkan, pengujian hasil prediksi dilakukan dengan menggunakan 10 data mahasiswa dari angkatan 2008 sampai dengan 2011 sebagai berikut:

NIM	NAMA	IPS 1	IPS 2	IPS 3	IPS 3
A12.2008.03037	WAHYU SETYO DWI UTOMO	2.1	2	0.74	1.06
A12.2008.03136	HUSNI MUBAROK	2.05	1.84	2.06	2.35
A12.2008.03236	BAGUS ARDY SUSANTO	2.75	2.42	3.26	2.7

A12.2009.03486	MAYCHAEL ANDRIANTO	3.1	2.74	2.71	2.86
A12.2009.03490	NUFUS WIRASTAMA	2.3	2.37	3.05	2.65
A12.2009.03798	RIZKI CHANDRA KURNIAWAN	3.05	3.05	3.1	2.92
A12.2010.04028	ABDUL GHOFUR SIDIK	1.9	0.94	0.73	2.47
A12.2010.04030	RACHMATDITYA FITRA PRATAMA	1.65	2.06	3.21	2.74
A12.2010.04109	PUTRA SATRIYA SUTRISNA	2.5	2.24	2.84	2.48
A12.2011.04203	ADRIANA DINA	3.85	3.92	3.9	3.91

Pengujian dilakukan per mahasiswa (dengan Menu Pertama) maupun dengan memprediksi seluruh data secara langsung (dengan Menu Kedua). Dari hasil uji coba tersebut diperoleh hasil sebagai berikut:

Gambar X Contoh Hasil Pengujian per Nim

Nim	Nama	IPC 1	IPC 2	IPC 3	IPC 4	Prediksi
A12.2008.03037	WAFI, SETYO DWI UTOMO	2.1	2	2.74	1.06	NO (100%)
A12.2008.03135	HUSENI MUDAROK	2.35	1.34	2.06	2.35	NO (75%)
A12.2008.03236	EAGUS ARDYUGHANTO	2.75	2.42	2.26	2.7	YES (61%)
A12.2008.03488	MAYCHARI ANDRIANTO	3.1	2.74	2.71	2.88	YES (61%)
A12.2008.03490	NUR, SURASTAMA	2.3	2.37	3.05	2.68	YES (61%)
A12.2008.03798	RIZKI CHANDRA, SRI, AWAN	3.35	3.35	3.3	2.93	YES (61%)
A12.2010.04028	ANHUI, BEKHAR, HIK	1.3	1.14	1.73	2.41	NO (100%)
A12.2010.04109	KHO-WAI, SYA HIMA PRATIWA	1.85	2.36	3.21	2.74	NO (67%)
A12.2011.04203	FUTRA SARIYA SUIHENA	2.5	2.24	2.81	2.48	NO (65%)
A12.2011.04203	ADRIANA JUNA	3.85	3.42	3.9	3.91	YES (61%)

Gambar Hasil Pengujian kepada Seluruh Data Uji

Tabel 1 Perbandingan Hasil Prediksi dengan Masa Studi Asli

Nim	Hasil Prediksi	Masa Studi	Keterangan
A12.2008.03037	NO	TIDAK TEPAT WAKTU	Sesuai
A12.2008.03136	NO	TIDAK TEPAT WAKTU	Sesuai
A12.2008.03236	YES	TIDAK TEPAT WAKTU	Tidak Sesuai
A12.2009.03486	YES	TEPAT WAKTU	Sesuai
A12.2009.03490	YES	TEPAT WAKTU	Sesuai
A12.2009.03798	YES	TEPAT WAKTU	Sesuai
A12.2010.04028	NO	TIDAK TEPAT WAKTU	Sesuai
A12.2010.04030	NO	TIDAK TEPAT WAKTU	Sesuai
A12.2010.04109	NO	TIDAK TEPAT WAKTU	Sesuai
A12.2011.04203	YES	TEPAT WAKTU	Sesuai

Dari table # diatas diketahui bahwa jumlah hasil prediksi yang sesuai dengan fakta yang ada adalah sejumlah 9 dari 10 data uji yang digunakan. Maka nilai akurasi yang diperoleh dari aplikasi ini adalah :

$$Akurasi = \frac{\text{Jumlah Pengujian Benar}}{\text{Jumlah Pengujian}} \times 100\% \quad (3)$$

$$Akurasi = \frac{9}{10} \times 100\%$$

$$Akurasi = 90\%$$

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, aplikasi berbasis sistem pakar untuk memprediksi ketepatan masa studi telah berhasil dikembangkan dengan keakuratan mencapai 90%. Dengan adanya aplikasi ini, maka pihak manajemen di perguruan tinggi dapat mengetahui kemungkinan keterlambatan penyelesaian masa studi mahasiswa, sehingga dapat menentukan tindakan antisipatif lebih awal.

5. SARAN

Adapun untuk penelitian selanjutnya, beberapa atribut seperti status bekerja dan status kelas malam dapat ditambahkan untuk dapat memprediksi dengan lebih tepat. Namun sebelumnya, penyimpanan data tersebut juga perlu dipersiapkan pada sistem akademik di institusi terkait agar penelitian selanjutnya dapat dilakukan.

DAFTAR PUSTAKA

- [1] Direktorat Jenderal Perguruan Tinggi. (2014) Direktorat Jenderal Perguruan Tinggi. [Online]. <http://dikti.go.id/profil-dikti/sistem-pendidikan-tinggi/>
 - [2] Arief Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *ISSN : 0854-9524*, vol. 18, januari 2013.
 - [3] Diki Andita Kusuma and Chairani, "RancangBangun Sistem Pakar Pendiagnosa Penyakit Paru-Paru Menggunakan Metode Case Based Reasoning," *Jurnal Infotel*, vol. 6, november 2014.
-