

DAFTAR ISI

BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan dari penelitian sebagai berikut:.....	2
BAB II	3
PEMBAHASAN	3
2.1 Landasan Teori	3
1. Pengertian Desain Sistem.....	3
2. Tujuan desain sistem.....	5
2.2 Desain Sistem Aplikasi	5
1. Pengenalan Desain Aplikasi yang digunakan.....	5
2. Algoritma yang digunakan.....	5
3. Pembentukan Pohon Keputusan Algoritma C4.5	6
4. Prosedur Pemilahan Algoritma C4.5	8
5. Pemangkasan Pohon Keputusan.....	9
6. Pengukuran Ketepatan Hasil Klasifikasi.....	9
7. UML.....	11
8. Desain Aplikasi	11
BAB III	16
KESIMPULAN	16
DAFTAR PUSTAKA.....	17

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu efek yang dihasilkan dari adanya suatu sistem informasi adalah munculnya banyak data. Data yang ada ini berasal dari sistem operasional yang berfungsi untuk menangani transaksi yang terkait dengan proses bisnis yang ditangani oleh sistem informasi tersebut. Contoh: sistem informasi presensi mahasiswa memunculkan data jumlah kehadiran mahasiswa setiap hari dengan data yang disimpan tergantung pada apa yang dibutuhkan oleh sistem informasi tersebut.

Tersedianya data yang melimpah pada institusi pendidikan harus dimanfaatkan dengan baik. Namun sulitnya memahami dan menemukan hubungan atribut-atribut data yang mempengaruhi hasil masa studi mahasiswa yang dapat digolongkan sebagai tepat waktu dan tidak tepat waktu, menjadi kajian dalam penelitian ini. Menganalisa kinerja mahasiswa (student performance), mengidentifikasi keunikan-keunikan yang ada pada mahasiswa dan membangun suatu strategi pengembangan lebih lanjut serta tindakan-tindakan yang dapat dilakukan untuk masa mendatang, merupakan tantangan utama bagi universitas modern saat ini (Kabakchieva, 2013). Data mining dapat diusulkan sebagai salah satu pendekatan yang dapat dilakukan untuk memprediksi kinerja siswa (Osmanbegovic & Suljic, 2012). Kinerja siswa dalam hal ini adalah capaian kelulusan studi mahasiswa dimana mereka dapat lulus tepat waktu atau tidak tepat waktu.

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban dkk. 2005). Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu Knowledge discovery in databases (KDD), ekstraksi pengetahuan (knowledge extraction), Analisa data/pola (data/pattern analysis), kecerdasan bisnis (business intelligence) dan data archaeology dan data dredging (Larose, 2005). Penerapan metode data mining dalam menganalisis data yang tersedia di lembaga pendidikan didefinisikan sebagai Educational Data Mining (EDM) (Romero & Ventura, 2007). EDM merupakan suatu aliran yang relatif baru dalam penelitian data mining. EDM menggunakan beberapa teknik seperti Decision Trees, Neural Networks, Naïve Bayes, KNearest Neighbor dan lainnya (Yadav & Pal, 2012). EDM

berkaitan dengan pengembangan metode untuk mengeksplorasi jenis yang unik dari data-data pada pengelolaan pendidikan dan menggunakannya untuk lebih memahami siswa dan pengelolaannya (Baker, 2010). Hal tersebut merupakan tujuan yang ingin dicapai dalam pemanfaatan data mining di bidang pendidikan.

Penelitian Yadav dan Pal (Yadav & Pal, 2012) melakukan prediksi pada data pendidikan untuk mengidentifikasi siswa yang lemah dan membantu mereka untuk mencetak nilai yang lebih baik. Algoritma C4.5, ID3 dan CART diterapkan dan dibandingkan akurasi, hasil menunjukkan bahwa teknik C4.5 memiliki akurasi paling tinggi yaitu 67,78% dibandingkan dengan teknik lainnya. Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, algoritma C4.5 akan digunakan pada penelitian ini karena memiliki tingkat akurasi yang lebih baik dan dapat memberikan gambaran klasifikasi mahasiswa yang tepat waktu atau tidak tepat waktu berupa pohon keputusan (Decision Tree) yang bermanfaat bagi pengelola akademik Decision Tree menyerupai sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas (Han & Kamber, 2007). Pohon keputusan bekerja mulai dari akar paling atas (root node), jika diberikan sejumlah data uji, misalnya X dimana kelas dari data X belum diketahui, maka pohon keputusan akan menelusuri mulai dari akar sampai node dan setiap nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan Decision Tree, kemudian pohon keputusan akan memprediksi kelas dari tupel X.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang maka yang menjadi perumusan masalah sebagai berikut:

1. Bagaimana membuat aplikasi untuk memprediksi masa studi mahasiswa .

1.3 Tujuan dari penelitian sebagai berikut:

1. Merancang suatu perangkat lunak yang dapat membantu pihak kampus dalam memprediksi masa studi mahasiswa ,tepat waktu atau terlambat.

BAB II

PEMBAHASAN

2.1 Landasan Teori

1. Pengertian Desain Sistem

Menurut John Burch dan Garry Grudnitski dalam buku Analisa dan Desain, Sistem Informasi Pendekatan Terstruktur adalah : "Desain sistem dapat didefinisikan sebagai penggambaran dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi." Dari definisi diatas, maka dapat ditarik kesimpulan bahwa desain sistem adalah tahapan berupa penggambaran, perencanaan dan pembuatan dengan menyatukan beberapa elemen terpisah ke dalam satu kesatuan yang utuh untuk memperjelas bentuk sebuah sistem.

Menurut Yavri D. Mahyuzir dalam bukunya Pengolahan Data menyebutkan beberapa langkah yang perlu dilakukan pada proses desain sistem adalah :

- Menganalisa masalah dari pemakai (user), sasarannya adalah mendapatkan pengertian yang mendalam tentang kebutuhan-kebutuhan pemakai.
- Studi kelayakan, membandingkan alternatif-alternatif pemecahan masalah untuk menentukan jalan keluar yang paling tepat.
- Rancang sistem, membuat usulan pemecahan masalah secara logika.
- Detail desain, melakukan desain sistem pemecahan masalah secara terperinci.
- Penerapannya yaitu memindahkan logika program yang telah dibuat dalam bahasa yang dipilih, menguji program, menguji data dan outputnya.
- Pemeliharaan dan evaluasi terhadap sistem yang telah diterapkan.

1. Langkah langkah dalam Membuat sistem

- Tahap Perencanaan
 - 1) Mendefinisikan Masalah
 - ✓ Sistem yang berjalan

- ✓ Sistem yang diusulkan
- 2) Menentukan tujuan sistem
- 3) Mengidentifikasi kendala sistem
- 4) Membuat studi kelayakan (TELOS)
- 5) Keputusan ditolak/diterima
- Tahap Analisis
 - 1) Membuat struktur organisasi
 - 2) Mendefinisikan kebutuhan informasi
 - 3) Mendefinisikan kriteria kinerja sistem
- Tahap Design
 - ✓ Menyiapkan rancangan
 - 1) Membuat Context Diagram
 - 2) Membuat DFD
 - 3) Membuat IOFC
 - 4) Membuat ERD
 - 5) Merancang Kamus Data
 - 6) Membuat FlowChart
 - 7) Merancang File (master, input, proses, temporary)
 - 8) Merancang Dialog Input
 - 9) Merancang Dialog Output
 - 10) Menyiapkan konfigurasi sistem
- Tahap Penerapan
 - 1) Menyiapkan hardware dan software.
 - 2) Implementasi Pemrograman
 - 3) Testing
 - 4) Cutover
- Tahap Penggunaan
 - 1) Audit Sistem
 - 2) Memelihara Sistem

2. Tujuan desain sistem

Tahap desain sistem mempunyai dua maksud atau tujuan utama, yaitu:

- Untuk memenuhi kebutuhan para pemakai sistem
- Untuk memberikan gambaran yang jelas dan rancang bangun yang lengkap kepada pemrogram komputer dan ahli-ahli teknik lainnya yang terlibat. Tujuannya lebih condong pada desain sistem yang terinci, yaitu pembuatan rancang bangun yang jelas dan lengkap untuk nantinya digunakan untuk pembuatan program komputernya.

2.2 Desain Sistem Aplikasi

1. Pengenalan Desain Aplikasi yang digunakan

Desain sistem yang digunakan pada pembuatan aplikasi ini dengan berbasis Web dengan bantuan software Sublime dan Xampp, yang mana rancangan aplikasi ini memiliki menu menu atau tools yang memudahkan user dalam penggunaannya atau sering disebut dengan user friendly. Aplikasi ini menerapkan data mining dengan menggunakan algoritma C4.5 untuk mengklasifikasikan status mahasiswa yang berpotensi mangkir. Hasil dari penelitian ini diharapkan mampu membantu Prodi SI untuk melakukan tindakan antisipatif sejak dini pada mahasiswa yang berpotensi mangkir.

Jika masa studi mahasiswa dapat diketahui lebih dini maka pihak jurusan Ilmu Komputer dapat melakukan tindakan-tindakan yang diperlukan supaya mahasiswa dapat lulus dalam waktu kurang dari lima tahun atau tepat lima tahun berdasarkan Peraturan Menteri Pendidikan.

2. Algoritma yang digunakan

Algoritma yang digunakan membantu pembuatan sistem tersebut ialah Algoritma C4.5. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan

C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Tahap tahap algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai 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.

3. Pembentukan Pohon Keputusan Algoritma C4.5

Menurut Ruggieri (2002), algoritma pembentukan pohon keputusan menggunakan sampel pelatihan yang terdiri atas kumpulan kasus dimana setiap kasus memiliki atribut dan kelas. Atribut-atribut dalam kasus dapat bertipe kontinu maupun diskret tetapi pada kelas setiap kasus harus bertipe diskret yang dapat dinotasikan C_1, \dots, C_n . Pohon Keputusan adalah pohon data yang terdiri atas simpul keputusan dan simpul daun yang terstruktur. Simpul daun tersebut memiliki kelas-kelas dan simpul keputusan menguji beberapa atribut sampai diperoleh atribut yang terpilih sebagai pemilah. Setiap cabang pada pengujian tersebut menghasilkan simpul anak (Child Node) di bawah simpul induknya (Parent Node).

Pembentukan pohon keputusan menggunakan prinsip membagi sampel pelatihan menjadi beberapa sub-himpunan yang berbeda. Langkah awal dari konstruksi pohon keputusan algoritma C4.5 adalah mencari simpul akar. Berikut ini langkah-langkah konstruksi pohon keputusan menggunakan Algoritma C4.5 berdasarkan Ruggieri (2002):

- 1) Misalkan T adalah himpunan kasus-kasus yang akan dibuat simpul dimana kasus-kasus tersebut memiliki kelas dan atribut-atribut. Frekuensi terboboti $\text{freq}(C_j, T)$ diperoleh dari perhitungan T dan kelas yang dihasilkan adalah C_j , untuk setiap $j \in \{1, 2, 3, \dots, n\}$

- 2) Jika semua kasus berada dalam kelas C_j yang sama maka simpul yang dihasilkan adalah simpul daun yang dilabeli dengan kelas C_j sebagai kelas terbanyak. Kesalahan klasifikasi pada simpul daun merupakan kasus-kasus dalam T yang berbeda kelas dengan kelas C_j .
- 3) Jika T berisi kasus yang memiliki dua atau lebih kelas maka dapat dihitung information gain dari setiap atribut tersebut. Untuk atribut diskret, information gain disesuaikan dengan pembagi dalam T dengan nilai atribut yang sudah diketahui sebelumnya. Untuk atribut kontinu, information gain disesuaikan dengan pembagi T ke dalam dua irisan (biner) yang dilabeli kasus dengan nilai atribut kurang dari atau sama dengan nilai ambang batas ($A \leq v$) dan nilai atribut dengan nilai atribut lebih besar dari nilai ambang batas ($A > v$).
- 4) Atribut dengan nilai information gain tertinggi terpilih sebagai pemilah dalam simpul tersebut.
- 5) Simpul keputusan memiliki cabang sebanyak s yaitu T_1, \dots, T_s dimana $s = 2$ untuk atribut kontinu dan $s = h$ untuk atribut diskret dengan nilai h yang sudah diketahui.
- 6) Untuk setiap $i = \{1, 2, \dots, s\}$, jika T_i tidak memiliki cabang lagi maka simpul tersebut secara langsung menjadi simpul daun yang diberi label kelas terbanyak di bawah simpul induknya dan kesalahan klasifikasi bernilai 0.
- 7) Apabila T_i memiliki cabang lagi maka pemilahan diproses kembali menggunakan kasus-kasus dalam T_i . Catatan khusus untuk kasus-kasus dengan nilai yang hilang pada atribut terpilih tersebut dilakukan proses pemilihan pemilah pada setiap simpul anaknya dengan pembobotan banyak kasus yang diketahui dibagi dengan banyak kasus pada simpul tersebut.
- 8) Terakhir, kesalahan klasifikasi simpul dihitung dari penjumlahan dari kesalahan-kesalahan simpul anak yang dibandingkan dengan simpul induknya.

4. Prosedur Pemilahan Algoritma C4.5

Information gain dari sebuah atribut a dari himpunan T dihitung jika a adalah diskret dan T_1, \dots, T_s adalah sub-himpunan dari T yang terdiri dari kasus-kasus yang nilainya sudah diketahui. Untuk mendapatkan information gain dari atribut a atau $\text{gain}(a)$ dibutuhkan entropy keseluruhan kelas pada himpunan T atau $\text{info}(T)$ dan entropy masing-masing atribut pada himpunan T atau $\text{info}(T_i)$. Rumus dari $\text{Gain}(a)$ adalah sebagai berikut:

$$\text{Gain}(a) = \text{info}(T) - \sum_{i=1}^s \frac{|T_i|}{|T|} \times \text{info}(T_i) \quad (1)$$

dimana

$$\text{info}(T) = - \sum_{j=1}^n \frac{\text{freq}(C_j, T)}{|T|} \times {}^2\log\left(\frac{\text{freq}(C_j, T)}{|T|}\right) \quad (2)$$

nilai *entropy* untuk setiap atribut i ,

$$\text{info}(T_i) = - \sum_{j=1}^n \frac{\text{freq}(C_j, T_i)}{|T_i|} \times {}^2\log\left(\frac{\text{freq}(C_j, T_i)}{|T_i|}\right) \quad (3)$$

keterangan:

$|T|$ = Banyak kasus dalam himpunan T

$|T_i|$ = Banyak kasus dalam sub-himpunan T_i

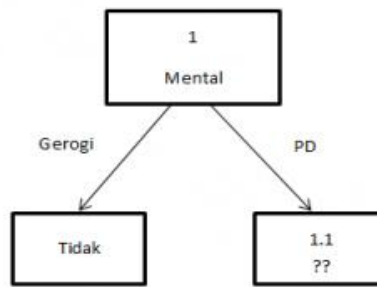
$\text{Freq}(C_j, T)$ = Banyak dari kasus-kasus dalam himpunan T yang memiliki kelas C_j

Fungsi $\text{info}(T)$ pada Persamaan (2) adalah fungsi entropy dimana entropy menurut Quinlan (1993) adalah rata-rata jumlah informasi yang dibutuhkan untuk mengidentifikasi kelas suatu kasus ke dalam himpunan T . Nilai dari setiap penghitungan entropy memiliki satuan bits atau binary digits. Jika a adalah atribut kontinu maka kasus dalam T dengan nilai atribut tersebut diurutkan dari yang terkecil sampai terbesar. Dimisalkan nilai hasil pengurutan adalah w_1, \dots, w_m , untuk $i \in \{1, 2, 3, \dots, m-1\}$ dimana nilai $v = \frac{w_i + w_{i+1}}{2}$

dan pemisahan yang terjadi untuk atribut bertipe kontinu adalah: $T_1 = \{W_j \mid W_j < v\}$ dan $T_2 = \{W_j \mid W_j \geq v\}$ (4)

Untuk setiap nilai v , nilai gain dihitung dengan mempertimbangkan prosedur pemisahan (4). Information gain untuk a didefinisikan sebagai nilai maksimum gain dari v dimana nilai v merupakan nilai ambang batas untuk atribut kontinu.

Gambar 1. Decision Tree Node 1



Gambar Decision Tree

5. Pemangkasan Pohon Keputusan

Menurut Ning Tan et al. (2006), metode pemangkasan ini menggunakan batas atas kepercayaan distribusi binomial dengan nilai α untuk Algoritma C4.5 yaitu $\alpha = 25\%$ dimana “E” adalah banyaknya kesalahan klasifikasi, “N” adalah banyaknya kasus pada simpul dan “f” adalah rasio kesalahan pada suatu simpul yaitu $f = E/N$. Proses pemangkasan terjadi apabila kesalahan terprediksi (predicted error) pada simpul anak lebih besar daripada simpul induk. Penghitungan kesalahan terprediksi merupakan hasil dari penjumlahan e N tiap cabang daun dimana e atau batas atas kepercayaan menggunakan rumus sebagai berikut:

$$e = \frac{f + \frac{z_{\alpha/2}^2}{N} + z_{\alpha/2} \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z_{\alpha/2}^2}{4N^2}}}{1 + \frac{z_{\alpha/2}^2}{N}}$$

6. Pengukuran Ketepatan Hasil Klasifikasi

Menurut Han et al. (2012), hasil klasifikasi algoritma pohon keputusan dapat diuji dengan menggunakan Matriks Konfusi. Matriks ini terdiri atas jumlah kasus yang diklasifikasikan secara tepat dan tidak tepat. Metode ini menggunakan level matriks seperti pada Tabel 1. berikut ini :

Tabel 1. Matriks Konfusi Tiga Kelas

Aktual	Prediksi (+)	Prediksi (-)	Total
(+)	A	B	P
(-)	C	D	N
Total	P'	N'	P+N

Untuk mengetahui hasil kinerja dari klasifikasi yang dikonstruksikan dapat diukur menggunakan akurasi. Akurasi adalah prosentase dari keseluruhan set yang diklasifikasikan secara tepat. Rumusnya:

$$\text{Akurasi} = \frac{A+D}{A+B+C+D} \times 100\%$$

Konstruksi pohon keputusan menghasilkan pohon dengan banyak simpul mencapai 51 simpul yang terdiri dari simpul akar, simpul keputusan dan simpul daun. Banyak simpul daun sendiri mencapai 31 simpul yang dilabeli dengan kelasnya masing-masing. Simpul akar merupakan simpul yang terletak paling atas. Dalam penelitian ini atribut jurusan terpilih sebagai pemilah pada simpul akar. Atribut Jurusan terpilih karena nilai information gain terbesar diantara atribut lainnya. Berikut ini proses penghitungan manual mendapatkan information gain atribut jurusan yang menjadi pemilah pada simpul akar:

- 1) Hitung frekuensi tiap kelas (tepat waktu dan tidak tepat waktu)

Kelas	Frek($C_j, T $)
Tepat Waktu	61
Tidak Tepat Waktu	99
Total ($ T $)	160

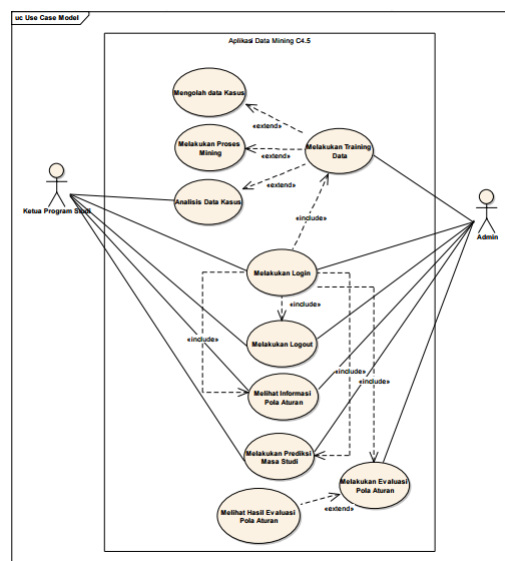
- 2) Hitung nilai entropy kelas yang disimbolkan Info(T)

$$\begin{aligned} \text{info}(T) &= - \sum_{j=1}^n \frac{\text{freq}(C_j, T)}{|T|} \times {}^2\log\left(\frac{\text{freq}(C_j, T)}{|T|}\right) \\ &= - \sum_{j=1}^2 \frac{\text{freq}(C_j, T)}{|T|} \times {}^2\log\left(\frac{\text{freq}(C_j, T)}{|T|}\right) \end{aligned}$$

$$\begin{aligned}
&= -\left(\frac{61}{160} \times {}^2\log\left(\frac{61}{160}\right) + \frac{99}{160} \times {}^2\log\left(\frac{99}{160}\right)\right) \\
&= -(0,38 \times {}^2\log 0,38 + 0,62 \times {}^2\log 0,62) \\
&= -((-0,53) + (-0,43)) \\
info(T) &= 0,96 \text{ bits}
\end{aligned}$$

7. UML

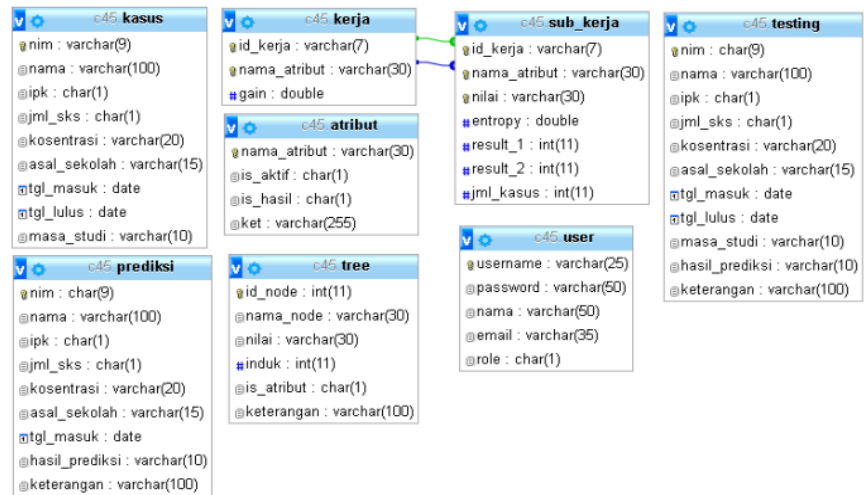
Perancangan sistem dibuat menggunakan UML (Unified Modeling Language). merupakan use case diagram dari aplikasi data minig. Diagram memiliki dua actor yang berhubungan dengan fungsi sistem, yaitu Admin dan Ketua Program Studi.



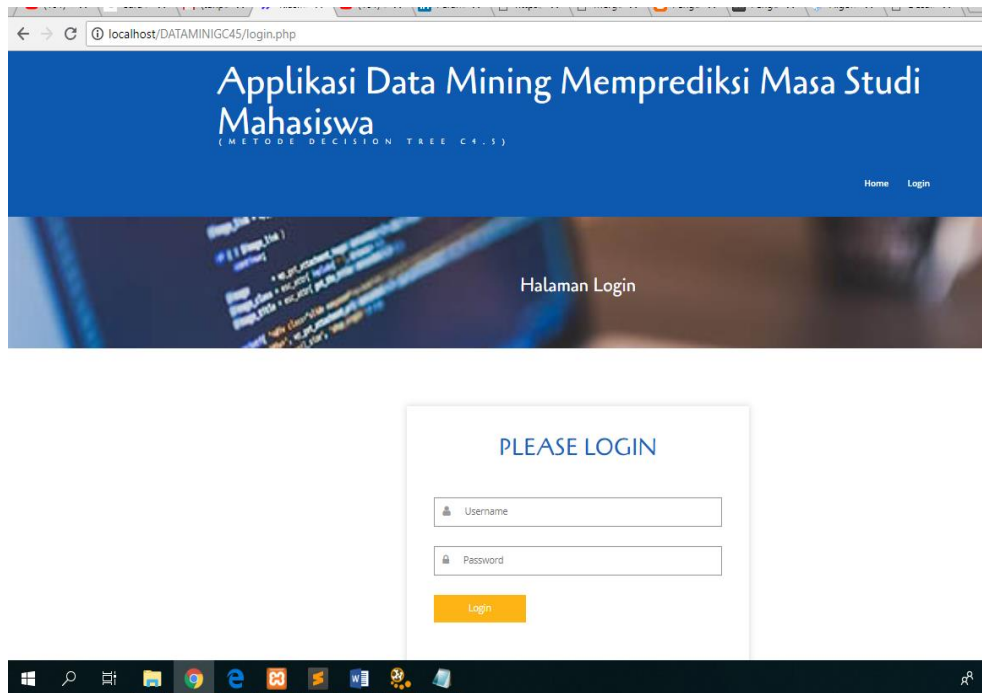
Gambar Use Case Diagram Aplikasi Data Mining C4.5

8. Desain Aplikasi

Pada preancangan database terdapat 8 tabel yaitu kasus, predikis, kerja, atribut, tree, user, sub_kerja, testing. Perancangan database aplikasi data mining ini ditunjukkan pada gambar dibawah ini.



Gambar Rancangan Aplikasi Data Mining C45



Gambar 1 Tampilan Login



Gambar 2. Tampil Halaman Utama

Data Mahasiswa

NIM

Username

Jenis Kelamin
☐ Laki-laki
☐ Perempuan

IPK

PENJURUSAN
☐ JARINGAN KOMPUTER
☐ SISTEM INFORMASI

Gambar 3. Tampilan inputan Data mahasiswa

NIM	IPK	Jumlah SKS	Konsentrasi	Asal Sekolah	Masa Studi
08052871	1	1	Jaringan Komputer	Urmam	> 4.6 th
08052872	1	2	Jaringan Komputer	Kejarman	> 4.6 th
08052873	1	2	Jaringan Komputer	Urmam	> 4.6 th
08052874	2	2	Sistem Informasi	Urmam	<= 4.6 th
08052875	1	1	Jaringan Komputer	Urmam	> 4.6 th
08052876	3	2	Sistem Informasi	Urmam	> 4.6 th
08052880	3	2	Jaringan Komputer	Urmam	<= 4.6 th

Gambar 4 Data kasus

Show 10 entries

Search:

#	IPK	Asal Sekolah	Jumlah SKS	Konsentrasi	Masa Studi	Ket
1	3,01 s/d 3,50	Kejuruan	>= 135 SKS	Jaringan Komputer	> 4,5 th	44.44% : <= 4,5 th, 55.56% : > 4,5 th
2	3,01 s/d 3,50	Umum	>= 135 SKS	Jaringan Komputer	> 4,5 th	42.50% : <= 4,5 th, 57.50% : > 4,5 th
3	3,01 s/d 3,50	Kejuruan	>= 135 SKS	Sistem Informasi	> 4,5 th	38.89% : <= 4,5 th, 61.11% : > 4,5 th
4	3,01 s/d 3,50	Umum	>= 135 SKS	Sistem Informasi	> 4,5 th	31.43% : <= 4,5 th, 68.57% : > 4,5 th
5	3,01 s/d 3,50	...	< 135 SKS	...	> 4,5 th	...

Applikasi Data Mining Memprediksi Masa Studi Mahasiswa

(METODE DECISION TREE C4.5)

Home Data Siswa C4.5+ Hasil Logout

Mining
Pohon Keputusan

Detail Proses

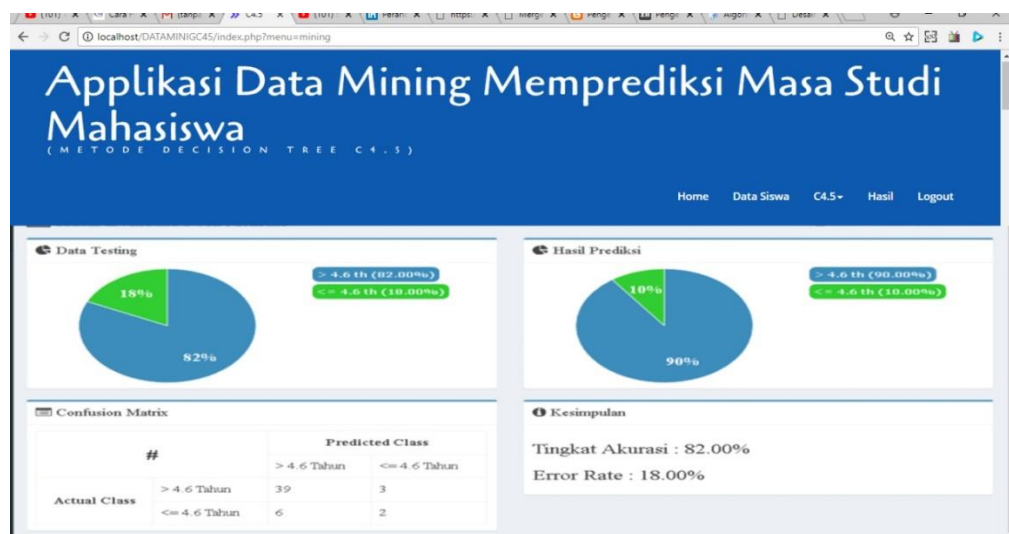
```

Menghapus Data Dari Tabel Tree
Mencari Atribut Tujuan/Hasil
=> Atribut Hasil : masa_studi
> 4.6 th
<= 4.6 th
Masuk Perulangan Pada Node
-> Jumlah Kasus : 139
=> Perulangan Pada masa_studi = > 4.6 th
=> Jumlah Kasus = 91
=> Perulangan Pada masa_studi = <= 4.6 th
=> Jumlah Kasus = 48
Entropy Level 0 = 0.92982227925097
Daftar Atribut Aktif

```

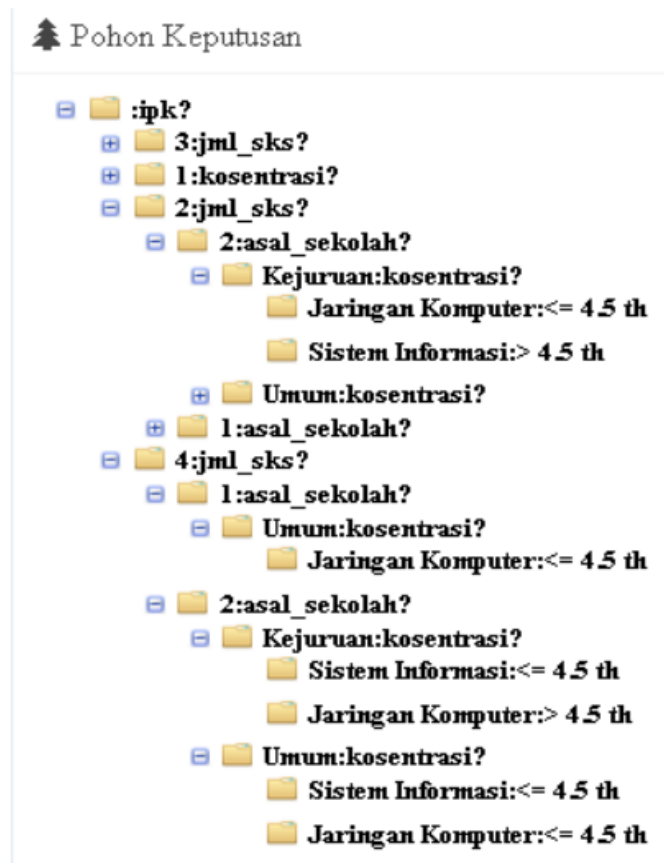
Gambar 5 Halaman Proses Mining

Hasil implementasi perancangan antarmuka halaman proses mining ditunjukkan pada Gambar 4. Pada halaman ini admin dapat memulai proses mining terhadap data kasus.



Gambar 6 Halaman Hasil Evaluasi Pola Aturan

Berdasarkan hasil evaluasi yang dilakukan terhadap perangkat lunak yang dibangun, dapat ditarik kesimpulan bahwa pola aturan yang dibangun algoritme C4.5 mampu mengklasifikasikan mahasiswa ke dalam kelas-kelas yang telah ditentukan sebelumnya dengan nilai error sebesar 27.75% dan nilai akurasi diatas 50% yaitu 72.25%. Dalam penelitian ini, pola aturan yang berhasil diekstrak sebanyak 22 pola aturan, dengan 17 pola aturan mengelompokkan data ke kelas '> 4.6 th', dan 5 pola aturan mengelompokkan data ke kelas '<= 4.6 th'. Sesuai data hasil evaluasi, maka tingkat akurasi yang berhasil dicapai oleh algoritme c4.5 sebesar 72.25%. Hasil Pola aturan terlihat pada pohon keputusan pada Gambar 7



Gambar 7 Pohon Keputusan Algoritme C4.5

BAB III

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka kesimpulan yang dapat diambil adalah:

1. Dalam penelitian ini, pola aturan yang berhasil diekstrak sebanyak 22 pola aturan, dengan 17 pola aturan mengelompokkan data ke kelas '> 4.6 th', dan 5 pola aturan mengelompokkan data ke kelas '<= 4.6 th'
2. Hasil evaluasi pola aturan menggunakan k-fold cross validation menunjukkan tingkat akurasi 72.25% dengan tingkat error 27.75% sehingga dapat disimpulkan bahwa data mining dengan algoritme C4.5 dapat diimplementasikan untuk memprediksi masa studi mahasiswa dengan empat atribut yaitu data jumlah SKS yang sudah diselesaikan saat semester 7, IPK mahasiswa saat semester 7, mata kuliah konsentrasi yang diambil mahasiswa dan asal sekolah mahasiswa.
3. Pohon keputusan yang dihasilkan dari algoritme c4.5 menunjukkan bahwa atribut IPK dan asal sekolah merupakan atribut yang paling dominan diantara atribut yang lain dalam menentukan masa studi mahasiswa. Hal ini terlihat dari root node pohon keputusan yang merupakan atribut IPK dan internal node ke-2 yang memiliki terminal node '<= 4.6 th' merupakan atribut asal sekolah.

DAFTAR PUSTAKA

- [1] RiadiMuchlisin,2017 , ”Pengertian, Fungsi, Proses dan Tahapan Data Mining”
diakses dari <https://www.kajianpustaka.com/2017/09/data-mining.html> pada tanggal
27 Juni 2018
- [2] ChendrawiraJennifer, 2016 ,”Makalah Data Mining “ diakses dari
https://www.academia.edu/29553365/MAKALAH_DATA_MINING-SEM_3_PSI
pada tanggal 27 juni 2018
- [3] Quinlan, J.R. 1993. C4.5 : Programs For Machine Learning. Morgan Kaufmann
Publisher, Inc. San Mateo
- [4] Ilmu skripsi ,2016, “Algoritma C45” diakses dari
<https://www.ilmuskripsi.com/2016/07/algoritma-c45.html> pada tanggal 27 juni2018
- [5] Pramudiono. (2017, February 26). iko-datamining. Retrieved from ilmukomputer:
<http://www.ilmukomputer.org/wpcontent/uploads/2006/08/iko-datamining.zip>
- [6] Rokach, L and Maimon, O. 2008. Data Mining With Decision Trees : Theory and
Applications. World Scientific Publishing Co. Pte.Ltd. Singapura.