

**SISTEM PENDUKUNG KEPUTUSAN PEMINATAN JURUSAN SISWA DI SMA  
MENGUNAKAN ALGORITMA KLASIFIKASI DATA MINING C4.5**

**THE DECISION SUPPORT SYSTEM FOR SUBJECT SPECIALIZATION STUDENTS IN  
HIGH SCHOOL USING DATA MINING CLASSIFICATION ALGORITHM C4.5**

Mutiara Shany Sinambela, RA Paramita Mayadewi, S.Kom., M.T.<sup>2</sup>, Ely Rosely, Ir., M.B.S.<sup>3</sup>

<sup>123</sup> Program Studi D3 Manajemen Informatika, Fakultas Ilmu Terapan, Universitas Telkom  
<sup>1</sup>shanymutiara@gmail.com, <sup>2</sup>paramita@tass.telkomuniversity.ac.id,  
<sup>3</sup>ely.rosely@tass.telkomuniversity.ac.id

---

**Abstrak**

Penjurusan siswa- siswi yang dilakukan di SMA bertujuan untuk membantu siswa- siswi memperoleh pengajaran yang tepat dan terarah sesuai dengan potensi masing- masing. Pada saat melakukan peminatan jurusan siswa- siswi memerlukan banyak waktu. Hal ini disebabkan banyaknya data siswa- siswi yang harus dikelola oleh sekolah sehingga ketika penentuan jurusan yang dilakukan tidak efektif. Berdasarkan permasalahan tersebut diperlukan sebuah sistem untuk membantu penentuan jurusan siswa- siswa, maka dibangun sebuah sistem yang diberi nama Sistem Pendukung Keputusan Peminatan Jurusan Siswa di SMA Menggunakan Algoritma Klasifikasi Data Mining C4.5. Sistem pendukung keputusan ini dibangun dengan *Knowledge Discovery in Database (KDD)* untuk proses data mining sedangkan untuk pembangunan aplikasinya menggunakan metode *waterfall*. Pada sistem ini dibangun untuk membantu dalam penentuan jurusan siswa- siswi dan untuk penentuan jurusan tersebut berdasarkan rule dengan proses mining yang telah dilakukan.

**Kata Kunci:** Sistem Pendukung Keputusan, Penjurusan siswa, Algoritma C4.5.

---

**Abstract**

*Majors students who performed in high school aims to help students obtain proper teaching and targeted according to the potential of each. At the time of specialization majors student need a lot of time. This is due to the amount of data students that must be managed by school so that when determining the majors are ineffective. Based on these problems required a system to help determine majors students, then built a system called Decision Support Systems Specialisation Programs in High School Students Using Data Mining C4.5 Classification Algorithm. This decision support system built by the Knowledge Discovery in Databases (KDD) to the data mining process while for application development using the waterfall method. In this system is built to assist in the determination of majors students-students and for the determination of the department based on the criteria set by the school.*

**Keywords:** Decision Support Systems, Majors students, C4.5 algorithm.

---

## 1. Pendahuluan

### 1.1 Latar Belakang

Salah satu pengambilan keputusan yang sering dijumpai pada saat ini ketika di SMA yaitu penentuan jurusan. Untuk mendukung keputusan tersebut diperlukan metode yang dapat memprediksi penjurusan tiap siswa- siswi. Metode pengelolaan data tersebut dinamakan *data mining*. *Data mining* merupakan 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 [1]. Salah satu model yang terkenal dalam *data mining* yaitu klasifikasi. Klasifikasi merupakan sebuah proses untuk menemukan suatu model atau fungsi untuk menggambarkan atau membedakan kelas data [2]. Kemudian untuk menyelesaikan model klasifikasi tersebut digunakan sebuah teknik klasifikasi yang populer yaitu pohon keputusan (*decision tree*). Dengan adanya pohon keputusan tersebut maka sebuah kumpulan data yang besar dapat menjadi himpunan- himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan [1]. Untuk pembentukan pohon keputusan tersebut maka digunakan algoritma C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [3]. Algoritma C4.5 merupakan struktur pohon di mana terdapat simpul yang mendeskripsikan atribut- atribut. Atribut- atribut tersebut merupakan parameter kriteria dalam pembentukan pohon. Sedangkan proses dalam pohon keputusan yaitu mengubah bentuk data menjadi model pohon kemudian mengubah model pohon tersebut menjadi *rule*, dan menyederhanakan *rule*. Dan dari *rule* tersebut maka didapatkan untuk menentukan keputusan penjurusan siswa- siswa. Hasil dari metode klasifikasi yang menggunakan algoritma C4.5 akan diimplementasikan pada sebuah aplikasi untuk membantu sekolah dalam menentukan peminatan jurusan siswa- siswi

### 1.2 Rumusan Masalah

Adapun rumusan masalah yang akan dibahas dalam proyek akhir ini adalah sebagai berikut:

1. Bagaimana membangun *rule* atau aturan menggunakan teknik klasifikasi data mining dalam memprediksi penentuan penjurusan siswa- siswi SMA?
2. Bagaimana membantu memudahkan pihak sekolah dalam melakukan prediksi terhadap keputusan penentuan penjurusan siswa- siswi SMA?

### 1.3 Tujuan

Tujuan dari proyek akhir ini adalah :

1. Mengimplementasikan metode klasifikasi *data mining* dengan algoritma C4.5 dengan tools WEKA untuk mendapatkan *rule* dalam menentukan penjurusan siswa di SMA
2. Membuat aplikasi berdasarkan *rule* yang telah dihasilkan sebelumnya dengan tools WEKA.

### 1.4 Batasan Masalah

Adapun yang menjadi batasan masalah dari Proyek Akhir ini adalah sebagai berikut:

1. Aplikasi ini menggunakan klasifikasi *data mining* algoritma C4.5
2. Pengolahan data yang telah diklasifikasikan sebelumnya menggunakan WEKA
3. Penginputan data dan pengelolaan data pada aplikasi ini terdiri dari data siswa- siswi dan data nilai yang menjadi kriteria untuk penentuan jurusan dengan data tahun pelajaran 2015/2016.

### 1.5 Metode Pengerjaan

Untuk data miningnya menggunakan metode proses *Knowledge-discovery* in Database (KDD). Proses *Knowledge-discovery* in Database dapat (KDD) dijelaskan sebagai berikut [1]:

#### a. *Data Selection*

Pada tahap ini dilakukan pemilihan data dari data keseluruhan. Data yang telah dipilih tersebut dipisah dari data keseluruhan dan data itu akan digunakan untuk proses *data mining*.

#### b. *Pre-Processing/ Cleaning*

Tahap ini memeriksa data seperti apakah ada data yang isinya tidak ada atau kesalahan cetak. Jika ada terjadi kesalahan pada data maka penulis melakukan pengurangan data atau melakukan penanganan pengisian data yang tidak ada dengan penanganan tertentu yang telah ditetapkan.

#### c. *Transformation*

Pada tahap ini telah mendapatkan data yang siap untuk dikelola untuk transformasi. Di transformasi ini data tersebut dengan tipe numerik dapat diubah dengan tipe nominal.

#### d. *Data Mining*

Tahap selanjutnya penerapan teknik atau metode data mining yang telah ditetapkan yaitu menggunakan algoritma klasifikasi

C4.5. Pengolahan data dengan algoritma klasifikasi C4.5 tersebut akan dikelola *tools* yang telah ditetapkan dengan menggunakan WEKA

e. *Interpretation/ Evaluation*

Pengolahan data dengan algoritma C4.5 tersebut menggunakan *tools* WEKA akan mendapat *rule* untuk diterapkan pada proses selanjutnya. Tahap ini juga melakukan penghitungan akurasi.

## 2. Tinjauan Pustaka

### 2.1 Data Mining

*Data mining* merupakan analisa terhadap data untuk menemukan hubungan yang berarti, pola, dan kecenderungan memeriksa sekumpulan data yang tersimpan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1].

*Data Mining* merupakan suatu proses yang terjadi pada sekumpulan data yang dikelola dan dianalisis untuk menghasilkan suatu informasi dengan cara mendapatkan hubungan atau pola yang terdapat pada data tersebut yang dapat memberikan suatu manfaat bagi orang-orang yang memerlukan informasi tersebut [3].

### 2.2 Metode Knowledge-Discovery in Database

Metode *Knowledge-discovery in Database* merupakan proses yang interaktif dalam menentukan keputusan sehingga digunakan metode *Knowledge-discovery in Database* (KDD) untuk data miningnya. Proses *Knowledge-discovery in Database* (KDD) dapat dijelaskan sebagai berikut [1]:

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

2. *Pre-processing/ Cleaning*

Pada tahap proses *cleaning* melakukan pembuangan data yang terjadi duplikasi, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data misalnya kesalahan cetak (*tipografi*), proses *enrichment*.

3. *Transformation*

*Coding* merupakan proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* merupakan proses kreatif dan tergantung pada jenis atau

pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

*Data mining* merupakan proses mencari pola atau informasi pada data yang telah terpilih dengan menggunakan teknik atau metode tertentu.

5. *Interpretation/ Evaluation*

Tahap ini untuk mengevaluasi apakah hasil yang terjadi setelah proses *data mining* yang telah dilakukan sebelumnya sesuai dengan aturan atau pola yang telah dibuat sebelumnya.

### 2.3 Klasifikasi

Klasifikasi merupakan penentuan sebuah *record* data baru ke salah satu beberapa kategori dimana telah ditentukan sebelumnya.

Untuk membuat suatu klasifikasi suatu data diperlukan beberapa komponen antara lain [3]:

1. Kelas, merupakan variabel tidak bebas atau dapat dikatakan hasil dari klasifikasi.

2. Prediktor, merupakan variabel bebas yang digunakan untuk menentukan pola atau karakteristik data yang akan diklasifikasikan.

3. Set data *training*, merupakan sekumpulan data yang berisi kelas dan prediktor yang akan diuji untuk mendapat pengelompokan sesuai dengan kelas yang tepat.

4. Set Data Uji, merupakan data- data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

### 2.4 Pohon Keputusan

Pohon keputusan atau *decision tree* dapat didefinisikan juga sebagai struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan- himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan setiap rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain [1].

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Proses pohon keputusan adalah mengubah bentuk data ( tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* [1].

### 2.5 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membuat sebuah keputusan dalam bentuk pohon keputusan dengan menetapkan kriteria- kriteria yang menjadi dasar

pembentuk keputusan. Secara umum, algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [1]:

1. Memilih atribut sebagai akar
2. Membuat cabang untuk tiap- tiap nilai
3. Membagi kasus dalam cabang
5. Mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut- atribut yang ada. Adapun rumus menghitung gain tersebut adalah [1]:

$$Gain(S, A) = E(S) - \sum_{i=1}^N |S_i| / |S| * E(S_i) \quad (1)$$

Keterangan :

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A

- |S<sub>i</sub>| : jumlah kasus pada partisi ke- i
- |S| : jumlah kasus dalam S

Sementara untuk penghitungan nilai entropi dapat dilakukan sebagai berikut [1]:

$$E(S) = \sum_{i=1}^N -P_i * \log_2 P_i \quad (2)$$

Keterangan :

- S : himpunan kasus
- A : fitur
- N : jumlah partisi S
- P<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

### 2.6 Cross Validation

Cross validation merupakan metode untuk

mengevaluasi model-model dengan suatu ukuran kemampuan prediksi dan memilih satu model yang terbaik. Ukuran kemampuan prediksi dalam metode cross validation dicari dengan cara

membagi data menjadi dua bagian. Bagian pertama digunakan untuk membuat model atau training dan bagian yang kedua digunakan untuk memvalidasi atau testing agar diketahui seberapa baik kemampuan prediksi dari model tersebut [7].

### 2.7 Confusion Matrix

Confusion matrix merupakan sebuah metode untuk evaluasi pada data mining [8]. Dalam confusion matrix dalam evaluasi menggunakan tabel matrix terdiri dari banyak baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi tersebut [1]. Tabel tersebut diperlukan untuk mengukur kinerja suatu model klasifikasi. Berikut tabel pada confusion matrix [9]:

Tabel 2-1

Confusion Matrix

Predication	Actual Class	
	Positif	Negatif

False	FP	TN

Keterangan

- TP : True Positif
- TN : True Negatif
- FP : False Positif
- FN : False Negatif

Dari tabel confusion matrix tersebut akan

menghasilkan nilai accuracy, precision, dan recall. Nilai precision atau confidence merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya [2]. Nilai precision dapat dihitung [9]:

$$Precision(p) = \frac{TP}{TP+FP} \quad (3)$$

Dan nilai dari recall atau sensitivity merupakan jumlah record yang diklasifikasikan benar dan

hasil prediksi juga benar [2]. Nilai recall dapat dihitung [9]:

$$Recall(r) = \frac{TP}{TP+FN} \quad (4)$$

Nilai accuracy merupakan persentase jumlah record data yang diklasifikasikan secara benar oleh algoritma dapat membuat klasifikasi setelah dilakuakn pengujian pada hasil klasifikasi [2]. Nilai accuracy dapat dihitung [9]:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

### 2.8 WEKA

Weka merupakan perangkat lunak data mining yang memiliki sekumpulan algoritma standar data mining yang dapat melakukan generalisasi atau formulasi dari sekumpulan data.

## 3. Analisis dan Perancangan Sistem

### 3.1 Proses Data mining

Dalam penentuan jurusan siswa di SMA dalam metode pengerjaanya menggunakan Knowledge Discovery in Database (KDD) dapat dilakukan sebagai berikut :

- a. Data Selection

Pada penentuan jurusan siswa di SMA ini menggunakan data siswa- siswa SMA N 8 Bandung sebanyak 465 record dengan mempunyai atribut antara lain nama, nilai UN ( nilai UN matematika, nilai UN IPA) SMP, nilai rerata rapor (nilai rerata raport matematika, nilai rerata raport IPA, nilai rerata

raport IPS) SMP, nilai TPA (nilai TPA matematika, nilai TPA IPA, nilai TPA IPS ).

b. *Pre-processing/Cleaning*

Pada tahap pre- processing ini untuk mengatasi data *missing value* digunakan mencari nilai dengan frekuensi yang paling sering muncul (modus). Penentuan jurusan siswa di SMA N 8 Bandung, salah satu kriterianya menggunakan nilai rerata raport. Untuk mendapatkan nilai rerata raport tersebut digunakan nilai semester 1,2,3,4, dan 5 siswa. Setelah dilakukan pengamatan ternyata nilai semester tersebut banyak yang kosong dan perlu adanya penanganan data yang kosong tersebut. Akibat banyaknya data yang kosong maka dilakukan penangan yaitu dengan pengurangan data sedangkan untuk data yang frekuensinya lebih banyak terisi nilai daripada nilai kosong maka dilakukan penanganan *missing value*. Penanganan data yang *missing value* juga dilakukan terhadap data yang bernilai 0 dan nilai UN yang berada di bawah 44. Daya yang *missing value* dilakukan penanganannya dengan mencari nilai frekuensinya paling tinggi (modus) sehingga didapatkan data dengan jumlah 442 record.

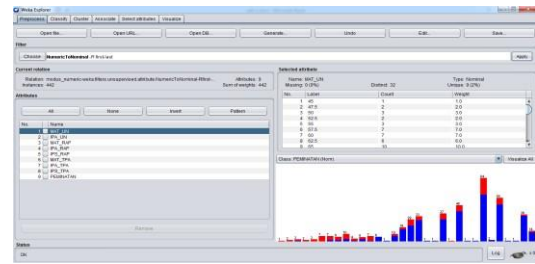
c. *Transformation*

Salah satu yang dapat dilakukan dalam proses *transformation* ini adalah dengan melakukan pengkategorian terhadap data yaitu mengubah data dengan tipe *numeric* ke data dengan tipe nominal. Pada tools WEKA dengan algoritma tertentu dapat melakukan transformasi sebab WEKA telah menyediakan fungsionalitas untuk transformasi yaitu mengubah tipe *numeric* ke nominal. Peminatan jurusan siswa ini menggunakan data dengan tipe *numeric* dan tanpa harus melakukan transformasi secara sendiri dapat menggunakan fungsionalitas yang disediakan oleh WEKA. Berikut sebagian data siswa yang telah melakukan proses seleksi dan *pre-processing* dengan tipe *numeric* :

MAT_UN	IPA_UN	MAT_RAP	IPA_RAP	IPS_RAP	MAT_TPA	IPA_TPA	IPS_TPA	PEMINATAN
86	87	83,2	82	80,8	67,5	55	47,5	A
77,5	90	85	81,6	85,6	70	75	67,5	A
85	82,5	91,8	89	86	55	77,5	50	A
95	90	87	85,6	86,4	85	85	52,5	A
80	77,5	87,2	88	88,2	62,5	45	42,5	A
97,5	92,5	91,6	92,6	91,2	95	82,5	60	A
90	87,5	75,2	77,2	83	50	50	45	A
70	85	87,2	87,6	88,6	62,5	55	47,5	A
37,5	52,5	77,4	81	77	15	32,5	52,5	A
80	85	83	83	87,2	80	77,5	57,5	A
95	97,5	92	91	90,6	90	77,5	55	A
97,5	82,5	90,6	89	92,8	85	80	65	A
90	80	92,4	90	85,2	75	55	57,5	A
90	90	92,6	89,4	85	90	80	47,5	A
95	95	90,4	91,6	90,2	72,5	77,5	72,5	A
27,5	70	81,4	82	82,4	45	55	45	A
76	80	82,4	80	79,8	82,5	67,5	42,5	A
92,5	77,5	91,2	90	90,8	80	70	62,5	A
92,5	92,5	80,6	83,4	82,8	72,5	55	40	A

Gambar 3-1  
Data Siswa Siswi

Berikut hasil transformasi dari tipe *numeric* ke nominal menggunakan WEKA.



Gambar 3-2  
Hasil Transformasi Dengan WEKA

Proses transformasi dapat kita lakukan secara sendiri tanpa menggunakan fungsi di WEKA dengan menentukan interval nilai lalu lakukan pengkategorian dengan interval tersebut ke tipe nominal. Untuk mendapatkan interval tersebut kita menentukan panjang kelas, banyak kelas interval. Sebelumnya kita menentukan jangkauan pada data. Pada data tersebut mempunyai nilai terbesar 100 dan nilai terkecil 100 maka jangkauan dapat ditentukan dengan:

$$R = X_{max} - X_{min} \quad (6)$$

Keterangan:

$X_{max}$  = nilai terbesar

$X_{min}$  = nilai terkecil

Maka didapatkan:

$$R = X_{max} - X_{min}$$

$$R = 100 - 12,5 = 87,5$$

Untuk menentukan banyaknya kelas interval dapat diberikan:

$$C = 1 + 3,3 \log n$$

Keterangan:

$n$  = jumlah data

sehingga banyak kelas dapat ditentukan:

$$C = 1 + 3,3 \log n \quad (7)$$

$$C = 1 + 3,3 \log 442 = 9,72 \text{ (bisa dibulatkan menjadi 10)}$$

Langkah selanjutnya menentukan batas kelas interval, untuk batas bawah dapat ditentukan dengan nilai terendah dari data atau nilai yang terdekat dengan nilai terendah. Maka nilai intervalnya dapat ditentukan:

$$P = \frac{R}{C} \quad (8)$$

$$P = \frac{87,5}{9,72}$$

$$P = 9,002$$

Maka dapat ditentukan nilainya dengan mempunyai interval nilai 9.

Tabel 3-2  
Interval Nilai

Interval Nilai
13 – 21
22 – 30
31 – 39
40 – 48
49 – 57
58 – 66
67 – 75

76 – 84
85 – 100

Kemudian nilai nilai tersebut dikategorikan menjadi 3 bagian yaitu rendah, tinggi, sedang, rendah. Maka nilai tersebut ditentukan untuk:

1. 13 – 39 kategori rendah
2. 39,5 – 66,4 kategori sedang
3. 66,5 – 100 kategori tinggi

Nilai setelah dikategorikan dapat dilihat pada gambar berikut dan data yang ditampilkan hanya sebagian:

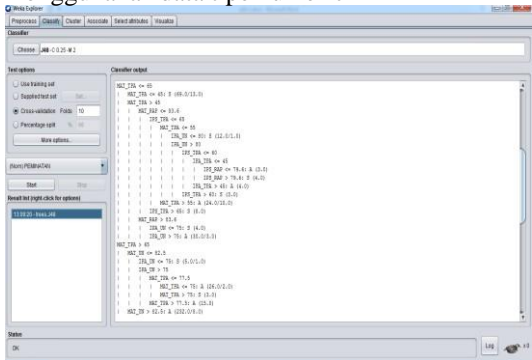
MAT_IPA	IPA_IPA	MAT_IPA	IPA_IPA	IPS_IPA	MAT_IPA	IPA_IPA	IPS_IPA	PERBATAN
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A
tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	tinggi	A

**Gambar 3-3**  
Nilai Dengan Tipe Nominal

d. Data Mining

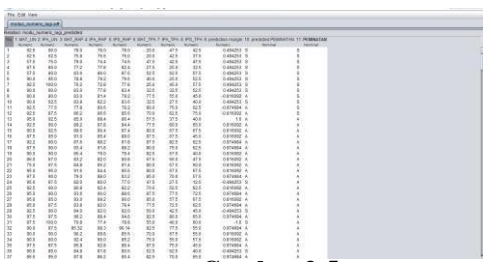
Tahap ini dipilih teknik data mining menggunakan klasifikasi dengan jenis algoritma C4.5. Penerapan algoritma C4.5 menggunakan perangkat lunak WEKA. Pada tools WEKA untuk algoritma C4.5 digunakan J48 sebab dalam tools WEKA menggunakan J48 merupakan implementasi algoritma C4.5.

Penerapan teknik klasifikasi data mining dengan algoritma C4.5 akan menghasilkan rule yang akan diterapkan pada pembangunan aplikasi. Berikut rule yang dihasilkan oleh WEKA dengan menggunakan data tipe numeric



**Gambar 3-4**  
Hasil Rule Pada WEKA Dengan Tipe Numeric

Untuk penerapan algoritma klasifikasi data mining c4.5 menggunakan data dengan bertipe numeric. Pengolahan data menggunakan algoritma klasifikasi data mining C4.5 mendapatkan hasil prediksi penentuan jurusan siswa. Berikut hasil prediksi penentuan menggunakan WEKA.



**Gambar 3-5**  
Hasil Prediksi Menggunakan WEKA

e. Interpretation/Evaluation

Test option pada tools WEKA menggunakan cross validation dengan nilai fold 10. Setelah melakukan penerapan algoritma C4.5 dengan menggunakan WEKA maka didapatkan rule. Sebelum rule tersebut diterapkan dilakukan tahap evaluasi. Adapun evaluasi terhadap rule tersebut dilakukan dengan confusion matrix.

**Tabel 3-3**  
Confusion Matrix

Accuracy =85,06%	Predicted Class		Precision
	True IPA	True IPS	
Prediksi IPA	290	36	90,6%
Prediksi IPS	30	86	70,5%
Recall	89%	74,1%	

Untuk perhitungan precision ditentukan dari:

$$Precision(p)_{IPA} = \frac{290}{30+290} = 0,906$$

$$Precision(p)_{IPS} = \frac{86}{36+86} = 0,7049$$

Nilai recall ditentukan dari,

$$Recall(r)_{IPA} = \frac{290}{290+36} = 0,889$$

$$Recall(r)_{IPS} = \frac{86}{86+30} = 0,741$$

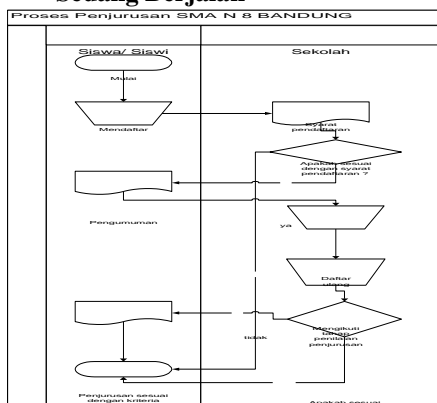
Nilai accuracy ditentukan dari,

$$Accuracy = \frac{290+86}{290+86+30+36}$$

$$Accuracy = 0,8506$$

$$Accuracy = 85,06\%$$

### 3.2 Proses Penjurusan Siswa Yang Sedang Berjalan

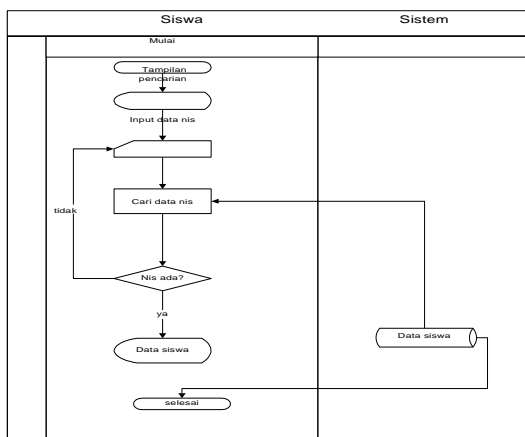


Gambar 3-6

### Proses Penjurusan Siswa Di SMA N 8 Bandung

### 3.3 Proses Penjurusan Siswa Yang Diusulkan

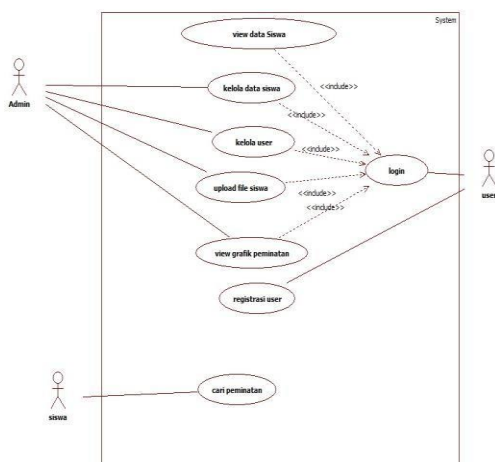
Proses pencarian jurusan oleh siswa



Gambar 3-7

### Proses Pencarian Penjurusan Oleh Siswa

### 3.4 Use Case Diagram



Gambar 3-8  
Use Case Diagram

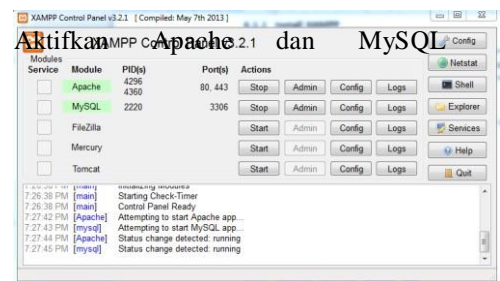
## 4. Implementasi

### 4.1 Install XAMPP

XAMPP merupakan aplikasi *web server* yang berisi Apache, interpreter PHP, dan basis data MySQL. Berikut ini adalah proses instalasi XAMPP.

1. Jalankan XAMPP yang telah diunduh
2. Pilih direktori untuk menyimpan hasil proses instalasi (C:\XAMPP)
3. Tunggu proses instalasi hingga selesai
4. Setelah selesai, jalankan XAMPP

### 5. Aktifkan Apache dan MySQL



Gambar 4-2  
Aplikasi XAMPP

### 4.2 Pembuatan Database

Pada tahap ini dilakukan pembuatan *database* melalui XAMPP yang telah diinstall dan dijalankan. Berikut adalah proses pembuatan *database* aplikasi.

1. Buka *browser* yang diinginkan
2. Pada kotak *address* ketik 'localhost'
3. Pilih menu *phpmyadmin*
4. Pilih menu *database* dan ketik nama *database* 'proyek\_akhir' dan klik *create*.

### 4.3 Aplikasi Instalasi WEKA

WEKA merupakan salah satu *tools* untuk yang dapat digunakan untuk mengelola data *mining*. Berikut ini adalah proses instalasi XAMPP.

1. Jalankan WEKA yang telah diunduh
2. Pilih direktori untuk menyimpan hasil proses instalasi
3. Tunggu proses instalasi hingga selesai
4. Setelah selesai, jalankan WEKA





**Gambar 4-3**  
**Tampilan WEKA**

## 5. Penutup

### 5.1 Kesimpulan

Berdasarkan aplikasi penentuan peminatan jurusan siswa ini maka dapat disimpulkan bahwa :

1. Implementasi dari metode klasifikasi data mining dengan algoritma C4.5 menggunakan *tools* WEKA menghasilkan *rule* untuk penentuan jurusan siswa
2. *Rule* yang didapatkan dari *tools* WEKA diterapkan pada aplikasi ini.

### 5.2 Saran

Setelah pembangunan aplikasi penentuan jurusan ini selesai, diharapkan dapat:

1. Mengembangkan aplikasi ini dengan tidak menggunakan teknik data mining klasifikasi dengan algoritma C4.5.
2. Pengolahan data yang telah dilakukan dapat menggunakan *tools* lain selain WEKA.
3. Menggunakan data yang tidak sama dalam aplikasi ini.

## Daftar Pustaka:

- [1] K. and E. T. Luthfi, Algoritma Data Mining, Yogyakarta: C.V ANDI OFFSET, 2009.
- [2] A. A. "Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa," *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA 2013)*, pp. 164-166, 9 Maret 2013.
- [3] P. P. Widodo and R. T. Handayanto, Penerapan Data Mining Menggunakan Matlab, Bandung: Rekayasa Sains, 2013.
- [4] I. S. Rekayasa Perangkat Lunak, Jakarta: Erlangga, 2003.
- [5] E. P. DATA MINING-Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: C.V ANDI OFFSET, 2014.
- [6] D. M. Dasar-Dasar Statistika, Bandung: Alfabeta, 2014.
- [7] N. A. S. Saucha Diwandari, "Perbandingan Algoritma J48 dan NBTree Untuk Klasifikasi Diagnosa Penyakit Pada Soybean," *Seminar Nasional Teknologi Informasi dan Komunikasi 2015 (SENTIKA 2015)*, p. 208, 2015.
- [8] P. M. and E. R. , "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining," *Seminar Nasional Sistem Informasi Indonesia, 2-3 November 2015*, pp. 330-332, 2015.
- [9] S. A. S. and C. S. , "Klasifikasi Spam Email Menggunakan Algoritma C4.5 Dengan Seleksi Fitur," *Teknologi Informasi*, vol. X, no. 1, pp. 23-24, 2014.
- [10] R. A.S. and M. Shalahuddin, Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek, Bandung: INFORMATIKA BANDUNG, 2013.
- [11] A.-. B. b. L. Analisa Dan Desain Sistem Informasi, Yogyakarta: Graha Ilmu, 2005.
- [12] F. L. S and J. S. , Data Mining: Meramalkan Bisnis Perusahaan, Jakarta: PT Elex Media Komputindo, 2010.
- [13] P. H. and J. K. Kawistara, Pemrograman Web, Bandung: Informatika Bandung, 2014.
- [14] Arbie, Manajemen Database dengan MySQL, Yogyakarta: Penerbit Andi, 2014.
- [15] R. A. and M. S. , Modul Pembelajaran Rekayasa Perangkat Lunak, Bandung: Modula, 2011.