

## ANALISIS DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK MEMREDIKSI MASA STUDI MAHASISWA BERDASARKAN DATA NILAI AKADEMIK

Selvia Lorena Br Ginting<sup>1</sup>, Wendi Zarman<sup>2</sup>, Ida Hamidah<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Komputer, Fakultas Teknik dan Ilmu Komputer, UNIKOM Bandung  
e-mail :<sup>1</sup>selvia Lorena@yahoo.com, <sup>2</sup>wendizar@gmail.com, <sup>3</sup>iida\_hamidaaah@yahoo.com

### ABSTRACT

*Institutional of higher education certainly experience the data accumulation, this occurs because the number of students increases every year. Students data that has accumulated can be utilized to extract a potential knowledge from data to support a decision. Analysis by simple statistical methods is difficult to applied therefore it is need a classification method of data mining techniques. Data mining is the mining or the discovery of new information by looking for certain patterns of a number of large amounts of data are expected to treat the condition. The data mining techniques can be used to the colleges data in order to process and distribute information to support daily operations as well as supporting the activities of establish a strategic decision. Source of data obtained form data academic of Department of Computer Engineering UNIKOM. Based on data graduation of Undergraduate Program Class of 2007 of Department of Computer Engineering, only one student who graduated at eighth semester. It has provided that there are many students who took the course duration more than 8 semesters. If the study period can be predictable, it can perform the necessary actions of the department, however the students can graduate on time. Training data and testing data taken from the student final report who have graduated and then result of the analysis which has processed by system will be used to predict the study period of the students who are still studying. The application is implemented the C4.5 Algorithm to produce a decision tree.*

**Kata kunci :** *academic-data, classification, prediction, C4.5 Algorithm, decision-tree*

### PENDAHULUAN

Seiring dengan berkembangnya teknologi penyimpanan data, semakin berkembang pula kemampuan seseorang dalam mengumpulkan dan mengolah data. Data yang terkumpul dan berukuran besar tersebut merupakan aset yang dapat dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga untuk masa mendatang. Tidak hanya dunia bisnis, namun instansi seperti perguruan tinggi juga mengalami penumpukan data.

Jurusan Teknik Komputer Universitas Komputer Indonesia adalah program pendidikan tinggi yang memiliki beban studi sekurang-kurangnya 144 sks (satuan kredit semester) yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan paling lama 1 semester. Dari data kelulusan yang diperoleh dari sekretariat Jurusan Teknik Komputer Program Sarjana (S1) angkatan 2007 hanya 1 orang mahasiswa yang lulus dalam 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) reguler di Jurusan Teknik Komputer yang menempuh lama studi lebih dari 8.

Melihat kondisi tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Jurusan Teknik Komputer. Data yang akan dimanfaatkan disini adalah data nilai akademik mahasiswa baik yang sudah lulus (yang akan digunakan sebagai *data training* dan *data testing*) maupun yang belum lulus/yang sedang menempuh studi yang akan digunakan untuk memprediksi masa studi masing-masing mahasiswa. Penelitian ini dirasa perlu karena jika masa studi mahasiswa dapat diketahui lebih dini, maka pihak jurusan dapat melakukan tindakan-tindakan yang dirasa perlu supaya mahasiswa dapat lulus tepat waktu sekaligus meningkatkan kualitas jurusan itu sendiri.

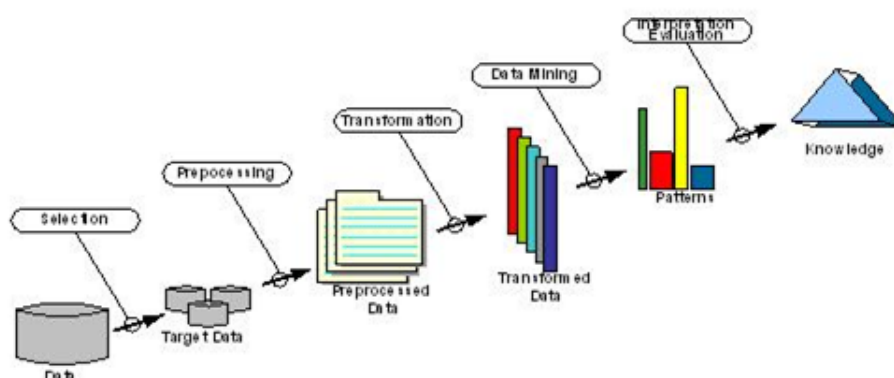
Dibutuhkan suatu teknik klasifikasi yang merupakan salah satu teknik dari *data mining* untuk menganalisis data Jurusan Teknik Komputer tersebut. Dengan menerapkan teknik ini akan dibangun pohon keputusan (*decision tree*) untuk melihat kemungkinan mahasiswa yang lulus lebih dari 8 semester. Pohon keputusan tersebut merupakan keluaran dari sebuah aplikasi yang dibangun dengan menerapkan Algoritma C4.5 untuk memprediksi masa studi mahasiswa yang sedang menempuh perkuliahan.

## METODE

Metode yang digunakan dalam penelitian ini meliputi lima aspek utama yaitu, studi pustaka terhadap sumber-sumber yang mendukung penelitian ini, analisis Algoritma C4.5 dalam penentuan masa studi mahasiswa, perancangan terhadap perangkat lunak (aplikasi) yang akan dibangun berdasarkan hasil yang diperoleh tahap analisis, implementasi terhadap perangkat lunak yang akan dibangun berdasarkan hasil yang diperoleh pada tahap perancangan, kemudian pengujian dan evaluasi terhadap hasil yang diperoleh berdasarkan kasus yang diteliti.

### Data Mining

*Data mining* merupakan salah satu proses inti yang terdapat dalam KDD. Banyak orang memperlakukan *data mining* sebagai sinonim dari KDD, karena sebagian besar pekerjaan dalam KDD difokuskan pada *data mining*. Namun, langkah-langkah lain merupakan proses-proses penting yang menjamin kesuksesan dari aplikasi KDD.



Gambar 1. Knowledge Data Discovery (KDD)

*Data mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola dan hubungan dalam set data berukuran besar [6]. Maksud dari pengertian ini yaitu proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data besar. Karakteristik *data mining* sebagai berikut :

- Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi [10].

Secara umum ada dua jenis metode pada *data mining*, metode *predictive* dan metode *descriptive*. Metode *predictive* adalah proses untuk menemukan pola dari data yang menggunakan beberapa variabel untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya. Teknik yang termasuk dalam *predictive mining* antara lain Klasifikasi, *Regresi* dan *Deviasi*. Metode *descriptive* adalah proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Teknik *data mining* yang termasuk dalam *descriptive mining* adalah *Clustering*, *Association* dan *Sequential Mining*.

### Klasifikasi

Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya. Teknik klasifikasi terbagi menjadi beberapa teknik yang diantaranya adalah Pohon Keputusan.

### Pohon Keputusan (Decision Tree)

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami, sehingga *Decision Tree* ini merupakan metode klasifikasi yang paling populer digunakan.

*Decision Tree* adalah *flow-chart* seperti struktur *tree*, dimana tiap *internal node* menunjukkan

sebuah *test* pada sebuah atribut, tiap cabang menunjukkan hasil dari *test* dan *leaf node* menunjukkan *class-class* atau *class distribution* [7].

#### Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data[7].

Algoritma untuk melakukan kontruksi pohon keputusan dapat dilihat pada gambar 2.

```
Algorithm : Generate_decision_tree.  
Narrative : Generate a decision tree from the given training data.  
Input      : The training samples, samples, represented by  
             discrete-valued attribute, the set of candidate attributes, attribute-list.  
Output     : A decision tree.  
  
Method  
1. Craete a node  $N$ ;  
2. If samples are all of the same class,  $C$  then  
3. Return  $N$  as a leaf node labeled with class  $C$   
4. If attribute-list is empty then  
5. Return  $N$  as leaf node labeled with the most common class in samples; // majority  
   voting  
6. Selset test-attribute, the attribute among attribute-list with the highest gain ratio;  
7. Label node  $N$  with last attribute;  
8. For each known value  $a$ , of test-attribute  
9. Grow a branch from node  $N$  for the condition test-attribute =  $a_i$   
10. Let  $S_i$  be the set of samples in samples for wichh test-attribute =  $a_i$  // a  
    partition  
11. If  $S_i$  is empty then  
12. Attach a leaf labeled with the most common class in sample;  
13. Else attach the node returned by Generate_decision_tree ( $S_i$ ,  
    attribute-list-list-attribute);
```

**Gambar 2.** Algoritma Membuat Pohon Keputusan

Pada tahap pembelajaran algoritma C4.5 memiliki 2 prinsip kerja yaitu:

1. Pembuatan pohon keputusan. Tujuan dari algoritma penginduksi pohon keputusan adalah mengkontruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau *record* baru yang belum memiliki kelas. C4.5 melakukan konstruksi pohon keputusan dengan metode *divide and conquer*. Pada awalnya hanya dibuat *node* akar dengan menerapkan algoritma *divide and conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan *gain ratio*, kemudian *node-node* yang terbentuk di level berikutnya, algoritma *divide and conquer* akan diterapkan lagi sampai terbentuk daun-daun.
2. Pembuatan aturan-aturan (*rule set*). Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan-aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*.

#### Information Gain

*Information gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan informasi gain tertinggi dipilih sebagai test atribut dari suatu node [7]. Dalam prosesnya perhitungan gain bisa terjadi atau tidak suatu *missing value*.

#### Sumber Data

Dari data yang ada, kolom yang diambil sebagai atribut atau variabel keputusan adalah kolom masa studi mahasiswa. Atribut-atribut yang digunakan sebagai variabel penentu dalam pembentukan pohon keputusan adalah :

1. Algoritma Pemrograman 1
2. Fisika Dasar 1
3. Fisika Dasar 2
4. Kalkulus 1
5. Kalkulus 2
6. Pengantar Sistem Komputer
7. Masa Studi

Dari proses di atas dihasilkan tabel yang siap untuk proses klasifikasi dengan atribut dan nilai atribut yang ada pada tabel 1.

**Tabel 1.** Atribut

Algoritma Pemrograman 1	Fisika Dasar 1	Fisika Dasar 2	Kalkulus 1	Kalkulus 2	Pengantar Sistem Komputer	Masa Studi
A	A	A	A	A	A	<= 5 TAHUN
B	B	B	B	B	B	> 5 TAHUN
C	C	C	C	C	C	
D	D	D	D	D	D	
E	E	E	E	E	E	

## PEMBAHASAN

### Analisis Algoritma C4.5

Perhitungan berdasarkan kasus data nilai mata kuliah mahasiswa untuk menentukan masa studi dilakukan dengan mengimplementasikan Algoritma C4.5 terlihat pada tabel di bawah ini. Perhitungan akan dilakukan berdasarkan 3 nilai mata kuliah yaitu : Algoritma Pemrograman 1, Fisika 1 dan Fisika 2. Dengan nilai masing-masing mata kuliah adalah A, B, C, D dan E. Mata kuliah ini dijadikan atribut awal dalam implementasi Algoritma C4.5 dan atribut tujuannya adalah masa studi dengan nilai dari atribut masa studi adalah “kurang” dan “lebih”. Kurang artinya lulus kurang dari lima tahun dan lebih artinya lulus lebih dari lima tahun.

**Tabel 2.** Contoh Kasus Algoritma C4.5

NODE	No	Algoritma	Fisika			Masa	E(x)	Gain				
			Jlh_Kasus	Kurang	Lebih				info(x)			
1	jml_semua	Algoritma Pemrograman 1	15	5	10	0,9183	0,551	0,3673				
			A	3	2	1			0,9183			
			B	3	1	2			0,9183			
			C	3	2	1			0,9183			
			D	3	0	3			0			
			E	3	0	3			0			
			Fisika 1	A	3	2			1	0,9183	0,7264	0,1919
				B	2	1			1	1		
				C	7	1			6	0,59167		
				D	2	1			1	1		
				E	1	0			1	0		
			Fisika 2	A	3	2			1	0,9183	0,4	0,5183
				B	4	3			1	0,81128		
				C	3	0			3	0		
				D	4	0			4	0		
E	1	0		1	0							

**Tabel 3.** Perhitungan Mencari Root

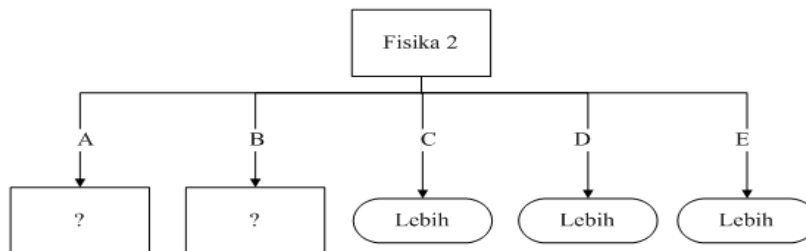
Dari tabel di atas terdapat gain untuk masing-masing mata kuliah. Dimana hasil gain dari masing-masing mata kuliah tersebut diantaranya adalah: Algoritma Pemrograman 1 = 0,3673, Fisika 1 = 0,1919 dan Fisika 2 = 0,5183

Sesuai dengan ketentuan dalam algoritma C4.5, setiap atribut yang memiliki gain tertinggi akan menjadi *root* atau *node*. Dari ketiga mata kuliah tersebut, matakuliah yang memiliki gain tertinggi adalah Fisika 2, maka mata kuliah Fisika 2 menjadi *root*. Setelah diperoleh *root* maka akan dilihat  $info(x)$  dari masing nilai atribut mata kuliah yang jadi *root* yaitu Fisika 2. Dari tabel di atas

dapat dilihat nilai  $infoa(x)$  untuk nilai mata kuliah Fisika 2 diantaranya adalah:

1. Fisika 2 dengan nilai A : 0,9183
2. Fisika 2 dengan nilai B : 0,81128
3. Fisika 2 dengan nilai C : 0
4. Fisika 2 dengan nilai D : 0
5. Fisika 2 dengan nilai E : 0

Hasil *tree* dari perhitungan yang pertama adalah sebagai berikut:



**Gambar 3.** Root dan Cabangnya

Karena nilai untuk atribut Fisika 2 adalah A dan Fisika 2 adalah B, maka harus dilakukan perhitungan kembali untuk menentukan node cabang Fisika 2 dengan nilai A dan Fisika 2 dengan nilai B. Tabel dibawah ini merupakan perhitungan untuk cabang Fisika 2 dengan nilai A.

**Tabel 4.** Perhitungan Cabang Fisika 2 dengan Nilai A

NODE		Jlh_Kasus	Kurang	Lebih	$infoa(x)$	$E(x)$	Gain
2	Fisika 2(A)	3	2	1	0,9183		
	Algoritma Pemrograman 1					0	0,9183
	A	1	1	0	0		
	B	1	0	1	0		
	C	1	1	0	0		
	D	0	0	0	0		
	E	0	0	0	0		
	Fisika 1					0,6667	0,2516
	A	2	1	1	1		
	B	0	0	0	0		
	C	0	0	0	0		
	D	1	1	0	0		
	E	0	0	0	0		

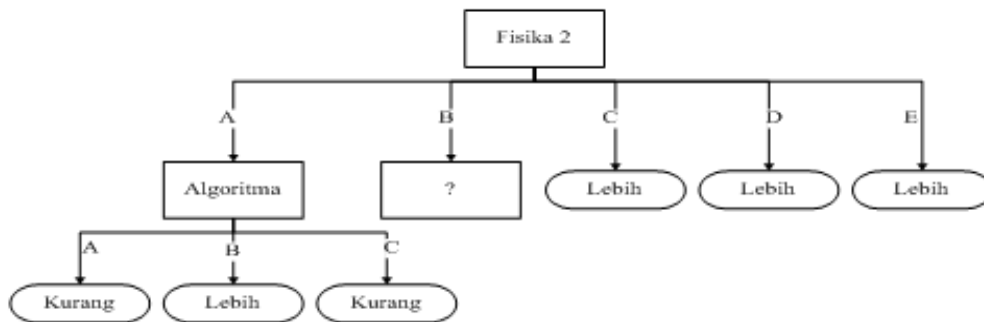
Dari tabel di atas terdapat gain untuk masing-masing mata kuliah, dimana hasil gain dari masing-masing mata kuliah tersebut diantaranya adalah:

1. Algoritma Pemrograman 1 = 0,9183
2. Fisika 1 = 0,2516

Dari kedua mata kuliah tersebut, yang memiliki gain tertinggi adalah Algoritma Pemrograman 1, maka Algoritma Pemrograman 1 yang akan menjadi *node* cabang dari Fisika 2 dengan nilai A. Setelah diperoleh *node* cabang Fisika 2 dengan nilai A maka akan dilihat  $infoa(x)$  dari masing nilai atribut mata kuliah yang jadi *node* cabang Fisika 2 dengan nilai A yaitu Algoritma Pemrograman 1. Dari tabel di atas, nilai  $infoa(x)$  untuk nilai mata kuliah Algoritma Pemrograman 1 diantaranya adalah:

1. Algoritma Pemrograman 1 dengan nilai A : 0
2. Algoritma Pemrograman 1 dengan nilai B : 0
3. Algoritma Pemrograman 1 dengan nilai C : 0
4. Algoritma Pemrograman 1 dengan nilai D : 0
5. Algoritma Pemrograman 1 dengan nilai E : 0

Dari hasil infoa(x) tersebut semua bernilai 0, artinya semua cabang dari nilai mata kuliah Algoritma Pemrograman 1 sudah mencapai keputusan akhir atau mencapai *leaf*. Dari data tabel di atas terlihat untuk nilai atribut Algoritma Pemrograman 1 dengan nilai A, B, dan C memiliki jumlah data, namun untuk nilai atribut Algoritma Pemrograman 1 yang D dan E bernilai 0. Ini artinya dalam *data training* tidak ada data yang menunjukkan Fisika dasarnya A dan Algoritmanya D dan E. Sehingga untuk cabang Algoritma D dan E tidak ada dalam *tree*. Sedangkan untuk cabang Algoritma yang A, B dan C sudah mencapai *leaf* atau keputusan akhir, dimana keputusan akhirnya adalah untuk Algoritma Pemrograman 1 yang nilainya A adalah “kurang”, Algoritma Pemrograman 1 yang nilainya B adalah “lebih” dan Algoritma Pemrograman 1 yang C adalah kurang”. Hasil dari perhitungan tabel di atas:



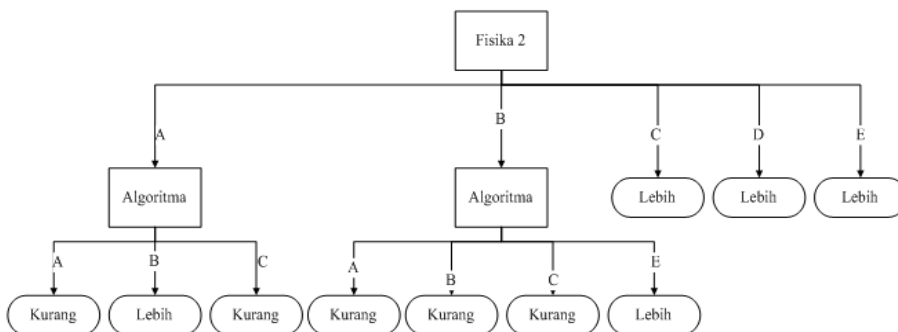
Gambar 4. Cabang Fisika 2 dengan Nilai A

Pada tabel 5 dibawah ini adalah perhitungan untuk Cabang Fisika 2 dengan nilai B.

Tabel 5. Perhitungan Cabang Fisika 2 dengan Nilai B

NODE	Jlh_Kasus	Kurang	Lebih	infoa(x)	E(x)	Gain
2 Fisika 2 (B)	4	3	1	0,81128		
Algoritma Pemrograman 1					0	0,8113
A	1	1	0	0		
B	1	1	0	0		
C	1	1	0	0		
D	0	0	0	0		
E	1	0	1	0		
Fisika 1					0,5	0,3113
A	1	1	0	0		
B	1	1	0	0		
C	2	1	1	1		
D	0	0	0	0		
E	0	0	0	0		

Hasil perhitungan dari tabel di atas dapat diperoleh strukturnya sebagai berikut:



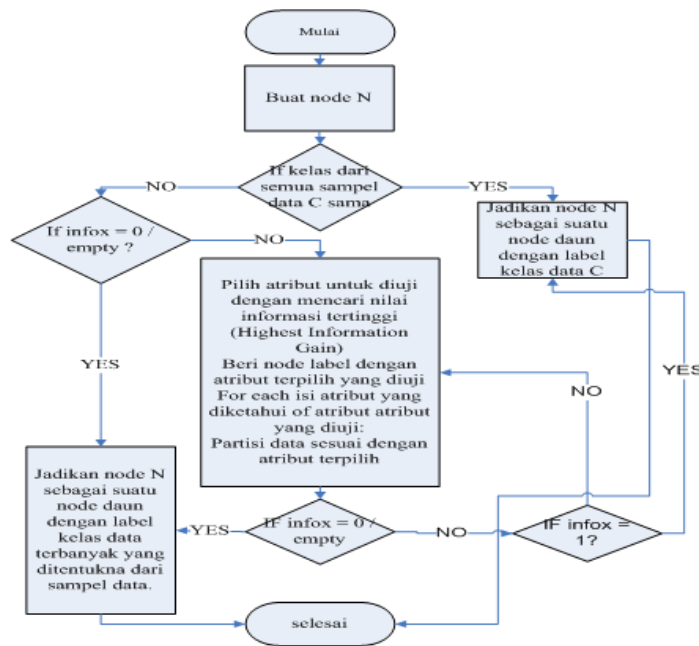
Gambar 5. Cabang Fisika 2 dengan Nilai B

Gambar di atas merupakan hasil akhir dari proses pembentukan *tree*. Setelah selesai pembentukan *tree* merupakan proses pembuatan aturan berdasarkan *tree* yang terbentuk. Aturan dari *tree* di atas adalah sebagai berikut:

1. Jika Fisika 2 nilai A dan Algoritma Pemrograman 1 nilai A maka masa studinya adalah kurang

2. Jika Fisika 2 nilai A dan Algoritma Pemrograman 1 nilai B maka masa studinya adalah lebih
3. Jika Fisika 2 nilai A dan Algoritma Pemrograman 1 nilai C maka masa studinya adalah kurang
4. Jika Fisika 2 nilai B dan Algoritma Pemrograman 1 nilai A maka masa studinya adalah kurang
5. Jika Fisika 2 nilai B dan Algoritma Pemrograman 1 nilai B maka masa studinya adalah kurang
6. Jika Fisika 2 nilai B dan Algoritma Pemrograman 1 nilai C maka masa studinya adalah kurang
7. Jika Fisika 2 nilai B dan Algoritma Pemrograman 1 nilai E maka masa studinya adalah lebih
8. Jika Fisika 2 nilai C maka masa studinya adalah lebih
9. Jika Fisika 2 nilai D maka masa studinya adalah lebih
10. Jika Fisika 2 nilai E maka masa studinya adalah lebih

**Flowchart Sistem**



**Gambar 6.** Flowchart Sistem

**Data Training**

Data training yang digunakan adalah sebanyak 112. Berikut ini adalah cuplikan data training yang akan digunakan dalam aplikasi yang mengimplementasikan Algoritma C4.5.

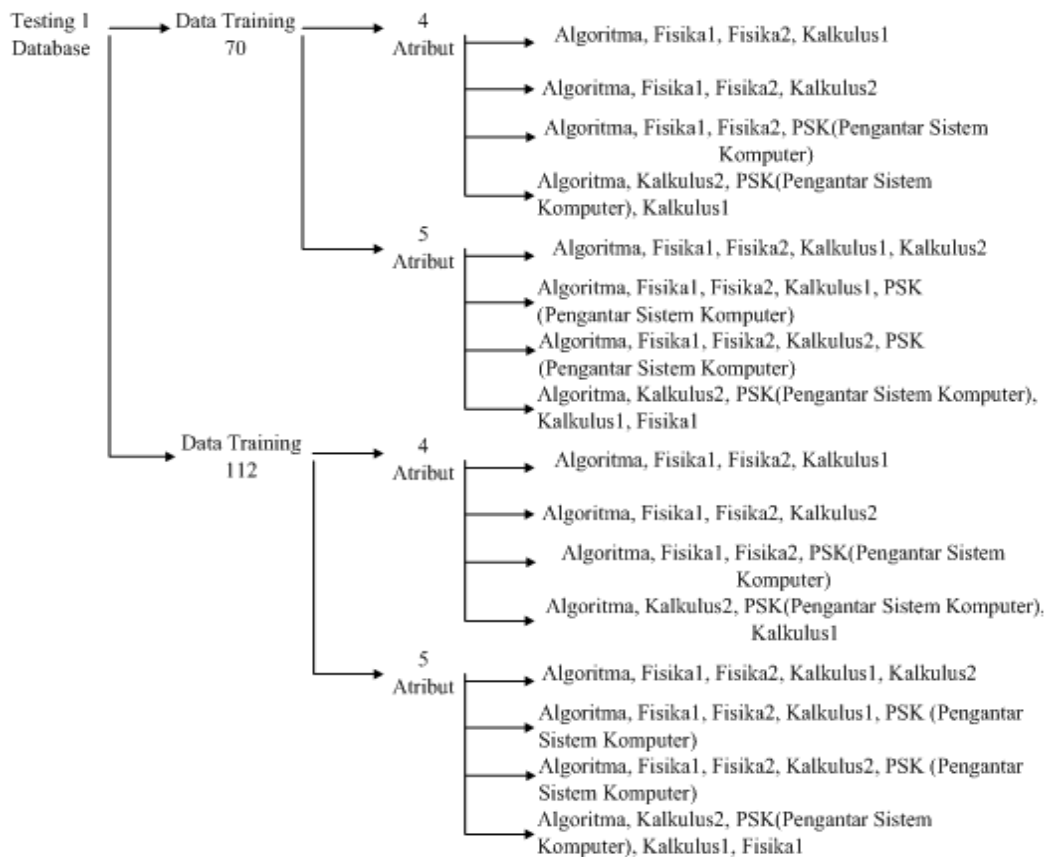
ALGO	FISIKA1	FISIKA2	KAL1	KAL2	PSK	MASA_STUDI
A	A	B	B	B	B	Prediksi Masa Studi <= 5 Tahun
B	B	B	C	C	B	Prediksi Masa Studi <= 5 Tahun
C	C	B	D	E	E	Prediksi Masa Studi > 5 Tahun
B	A	A	A	A	C	Prediksi Masa Studi <= 5 Tahun
D	C	C	D	C	D	Prediksi Masa Studi > 5 Tahun
A	A	A	C	B	C	Prediksi Masa Studi <= 5 Tahun
A	A	A	C	B	C	Prediksi Masa Studi <= 5 Tahun
E	C	B	C	D	E	Prediksi Masa Studi > 5 Tahun
D	C	D	C	C	B	Prediksi Masa Studi > 5 Tahun
D	C	C	C	C	C	Prediksi Masa Studi > 5 Tahun
A	B	D	C	C	D	Prediksi Masa Studi > 5 Tahun
E	D	C	C	D	C	Prediksi Masa Studi > 5 Tahun
B	C	D	C	E	B	Prediksi Masa Studi > 5 Tahun
E	C	D	C	D	D	Prediksi Masa Studi > 5 Tahun
C	E	E	D	E	C	Prediksi Masa Studi > 5 Tahun
C	D	A	B	C	A	Prediksi Masa Studi <= 5 Tahun

**Gambar 7.** Cuplikan Data Training



**Pengujian**

Program/aplikasi diujikan pada sebuah *database* dengan ketentuan sebagai berikut:



**Gambar 8.** Skenario Pengujian

Berikut disajikan gambar hasil pengujian program terhadap 6 atribut penentu untuk memprediksi masa studi mahasiswa yang mengimplementasikan Algoritma C4.5

The screenshot shows a query result in MS Access. The title of the query is 'Persentase Kecocokan = 45,71 %'. The table has columns for 'no', 'fiskal', 'kalkulus1', 'psk', 'fisika2', 'kalkulus2', 'algoritma', 'masa\_studi', and 'hasil\_1st'. The data rows show various combinations of these attributes and their corresponding predicted study durations.

no	fiskal	kalkulus1	psk	fisika2	kalkulus2	algoritma	masa_studi	hasil_1st
16	D	C	D	C	C	A	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
17	D	C	B	B	C	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi <= 5 Tahun
18	C	B	C	B	C	C	Prediksi Masa Studi > 5 Tahun	Tidak Ada
19	C	C	D	A	B	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
20	D	C	D	B	C	B	Prediksi Masa Studi > 5 Tahun	tidak teridentifikasi
21	C	C	B	A	B	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
22	A	D	A	A	C	B	Prediksi Masa Studi > 5 Tahun	Tidak Ada
23	B	B	B	A	B	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
24	A	B	A	A	B	A	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi <= 5 Tahun
25	C	C	A	A	C	B	Prediksi Masa Studi > 5 Tahun	tidak teridentifikasi
26	C	E	D	C	C	B	Prediksi Masa Studi > 5 Tahun	Tidak Ada
27	B	C	B	B	C	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi <= 5 Tahun
28	D	E	E	D	E	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
29	E	D	E	C	D	C	Prediksi Masa Studi > 5 Tahun	Tidak Ada
30	E	E	D	D	C	C	Prediksi Masa Studi > 5 Tahun	Tidak Ada
31	C	D	E	B	E	D	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
32	C	E	E	D	D	D	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun
33	C	C	B	B	C	A	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi <= 5 Tahun
34	D	D	C	D	D	B	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi <= 5 Tahun
35	F	D	C	B	C	D	Prediksi Masa Studi > 5 Tahun	Prediksi Masa Studi > 5 Tahun

**Gambar 9.** Hasil Pengujian Mata Kuliah Algoritma Pemrograman 2, Kalkulus 2, Pengantar Sistem Komputer, Kalkulus 1, Fisika 1 dan Fisika

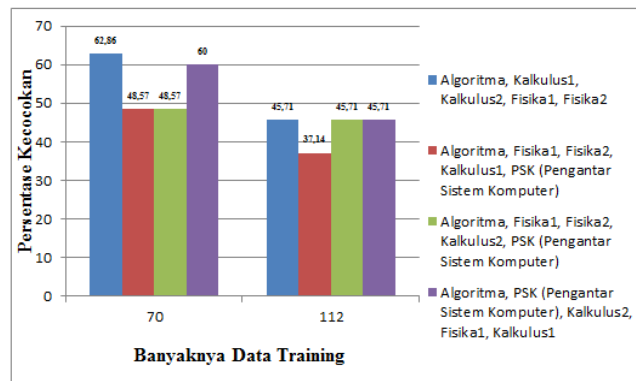


**Tabel 6.** Hasil Pengujian Mata Kuliah Algoritma Pemrograman 2, Kalkulus 2, Pengantar Sistem Komputer, Kalkulus 1, Fisika 1 dan Fisika 2.

Uji									
no	algo1	fisika1	fisika2	kalkulus1	kalkulus2	psk	masa_studi	Hasil_Uji	Ket
1	B	C	A	D	B	A	<= 5 Tahun	<= 5 Tahun	Sama
2	B	C	B	D	C	A	<= 5 Tahun	> 5 Tahun	Beda
3	B	A	B	B	C	B	<= 5 Tahun	<= 5 Tahun	Sama
4	A	A	A	A	B	A	<= 5 Tahun	<= 5 Tahun	Sama
5	B	C	A	B	B	A	<= 5 Tahun	<= 5 Tahun	Sama
6	A	A	A	B	B	A	<= 5 Tahun	<= 5 Tahun	Sama
7	B	C	C	D	C	C	<= 5 Tahun	> 5 Tahun	Beda
8	A	A	B	B	B	A	<= 5 Tahun	<= 5 Tahun	Sama
9	A	D	A	B	B	A	<= 5 Tahun	<= 5 Tahun	Sama
10	B	C	B	C	C	A	<= 5 Tahun	> 5 Tahun	Beda
11	B	C	A	C	B	E	> 5 Tahun	<= 5 Tahun	Beda
12	B	C	C	C	B	C	> 5 Tahun	> 5 Tahun	Sama
13	D	E	C	C	B	D	> 5 Tahun	<= 5 Tahun	Beda
14	A	C	A	B	C	A	> 5 Tahun	<= 5 Tahun	Beda
15	C	C	A	B	B	C	> 5 Tahun	<= 5 Tahun	Beda
16	A	D	C	C	C	D	> 5 Tahun	<= 5 Tahun	Beda
17	B	D	B	C	C	B	> 5 Tahun	Tidak Ada	Beda
18	C	C	B	B	C	C	> 5 Tahun	Tidak Ada	Beda
19	B	C	A	C	B	D	> 5 Tahun	<= 5 Tahun	Beda
20	B	D	B	C	C	D	> 5 Tahun	Tidak Ada	Beda
21	B	C	A	C	B	B	> 5 Tahun	<= 5 Tahun	Beda
22	B	A	A	D	C	A	> 5 Tahun	Tidak Ada	Beda
23	B	B	A	B	B	B	> 5 Tahun	<= 5 Tahun	Beda
24	A	A	A	B	B	A	> 5 Tahun	<= 5 Tahun	Beda
25	B	C	A	C	C	A	> 5 Tahun	<= 5 Tahun	Beda
26	B	C	C	E	C	D	> 5 Tahun	> 5 Tahun	Sama
27	B	B	B	C	C	B	> 5 Tahun	<= 5 Tahun	Beda
28	B	D	D	E	E	E	> 5 Tahun	Tidak Ada	Beda
29	C	E	C	D	D	E	> 5 Tahun	> 5 Tahun	Sama
30	C	E	D	E	C	D	> 5 Tahun	Tidak Ada	Beda
31	D	C	B	D	E	E	> 5 Tahun	> 5 Tahun	Sama
32	D	C	D	E	D	E	> 5 Tahun	> 5 Tahun	Sama
33	A	C	B	C	C	B	> 5 Tahun	<= 5 Tahun	Beda
34	B	D	D	D	D	C	> 5 Tahun	Tidak Ada	Beda

Jumlah *data testing* adalah 35, jumlah data yang sama dengan *data training* sebanyak 16 dan persentase kecocokan dengan *data training* adalah  $16/35 \times 100\% = 47,71\%$

### Grafik Pengujian Database



Gambar 10. Grafik Pengujian

Dari grafik pengujian dapat disimpulkan bahwa, semakin banyak *data training*, persentase kecocokan semakin kecil dengan *data testing*. Dari hasil pengujian terhadap berbagai kombinasi atribut, konfigurasi atribut yang paling kecil persentasenya adalah konfigurasi atribut Algoritma Pemrograman 1, Kalkulus 1, Kalkulus 2, Fisika 1, Fisika 2 dan Pengantar Sistem Komputer.

### KESIMPULAN

Aplikasi untuk memprediksi masa studi mahasiswa berhasil dirancang dan diimplementasikan menggunakan algoritma C4.5. Aplikasi tersebut dapat memprediksi masa studi mahasiswa.

Berdasarkan training dan pengujian kemudian dilakukan analisis maka dapat diambil beberapa kesimpulan bahwa jumlah *data training* mempengaruhi persentase kecocokan atau keakuratan. Persentase kecocokan hasil pengujian yang didapatkan dari *data training* 112 lebih kecil di bandingkan dengan *data training* 70 untuk berapapun kombinasi atribut. Demikian pula untuk semua konfigurasi atribut, persentase kecocokan hasil pengujian dengan *data training* 112 lebih kecil di bandingkan dengan *data training* 70.

### UCAPAN TERIMA KASIH

Terima kasih sebesar-besarnya kepada pihak-pihak yang membantu penelitian ini, terutama Ketua Jurusan Teknik Komputer dan sekretariat, juga rekan-rekan sepenelitian.

### DAFTAR PUSTAKA

- Han, J., & Kamber, M., 2006. Data mining Concepts and Techniques. San Fransisco: Morgan Kaufmann.
- Kadir, A., 2003. MySQL Delphi. Yogyakarta: ANDI.
- Kusrini, & Lutfhi, E. T., 2009. Algoritma Data Mining. Yogyakarta: Andi.
- Pranata, A., 1997. Pemrograman Borlan Delphi. Yogyakarta: ANDI.
- Santosa, B., 2007. Data mining (Teknik Pemanfaatan Data untuk Keperluan Bisnis). Surabaya: Graha Ilmu.
- Santosa, B., 2007. Data mining Terapan. Surabaya: Graha Ilmu.
- Sunjana, 2010. Seminar Nasional Aplikasi Teknologi Informasi 2010. Snati 2010. Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree , 24-29.
- Tan, P. N., Steinbach, M., & Kumar, V., 2005. DATA MINING. New York: Addison Wesley.