

## ALGORITMA C4.5 UNTUK MEMPREDIKSI MAHASISWA YANG MENGULANG MATA KULIAH (STUDI KASUS DI AMIK LABUHAN BATU)

Nurul Azwanti

Fakultas Teknik, Program Studi Sistem Informasi  
Universitas Putera Batam  
Email: nurulazw@rocketmail.com

### ABSTRAK

Penelitian ini dilakukan untuk memprediksi mahasiswa yang mengulang mata kuliah di AMIK Labuhan Batu dengan menggunakan teknik Data Mining. Algoritma C4.5 merupakan teknik Data Mining yang dapat melakukan prediksi dengan mengolah variabel Semester, IPK, Nilai, Keadaan Ekonomi dan Status. Variabel tersebut akan diklasifikasikan berdasarkan atributnya, untuk variabel keadaan ekonomi pengklasifikasian akan menggunakan rumus Sturgess agar dapat melakukan pengolahan data. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi *rule* prediksi untuk menggambarkan proses yang terkait dengan prediksi mahasiswa yang mengulang. Karakteristik data yang diklasifikasi dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun aturan sehingga dalam tahap pengujian dengan *software* WEKA dapat membantu dalam memprediksi mahasiswa yang mengulang mata kuliah. Dari dua hasil pengujian yang telah dilakukan yaitu proses secara manual dan menggunakan *software* WEKA disimpulkan bahwa hasil pengujian sangat baik karena *rule* yang dihasilkan hampir sama. Perbedaannya hanya terletak pada atribut nilai yang masuk ke dalam WEKA, namun tidak mengubah hasil keputusan. Pada hitungan manual menggunakan 34 *record* dan pada WEKA menggunakan 141 *record*.

**Kata kunci:** *data mining*, algoritma C4.5, klasifikasi, pohon keputusan.

### ABSTRACT

*This research was conducted to predict students who repeat the course in AMIK Labuhan Batu by using Data Mining technique. Algorithm C4.5 is Data Mining technique that can make prediction by processing variable Semester, GPA, Value, Economic Condition and Status. The variables will be classified based on their attributes, for the economic classification variable will use Sturgess formula in order to perform data processing. The C4.5 algorithm by the decision tree method can provide predictive rule information to describe the process associated with student predictions that repeat. Characteristics of classified data can be obtained clearly, either in the form of decision tree structures or rules so that in the testing phase with WEKA software can assist in predicting students who repeat the course. Of the two test results that have been done that is the process manually and using software WEKA concluded that the test results are very good because the rule produced almost the same. The only difference lies in the attribute value that goes into WEKA, but does not change the decision result. On manual count using 34 records and on WEKA using 141 records.*

**Keywords:** *data mining*, algorithm c4.5, classification, decision tree.

### 1. PENDAHULUAN

Pendidikan mempunyai peranan strategis dalam mempersiapkan generasi penerus yang memiliki pengetahuan dan kecerdasan tinggi sertamenguasai berbagai keahlian yang kompeten [1]. Di Indonesia sendiri program wajib belajar 12 tahun mulai dari tingkat SD, SMP dan SMA sudah diterapkan demi terciptanya sumber daya manusia yang berkualitas. Tidak sampai disitu, peminat untuk melanjutkan ke jenjang pendidikan yang lebih tinggi juga memberikan kontribusi yang besar untuk kemajuan bangsa. Perguruan tinggi merupakan penyelenggara pendidikan tingkat lanjut dari jenjang pendidikan menengah di jalur pendidikan formal.

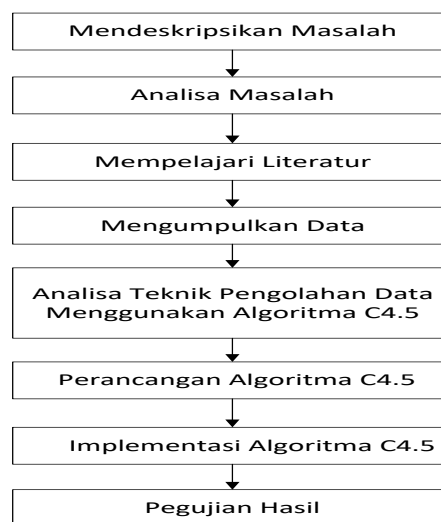
AMIK Labuhan Batu merupakan salah satu perguruan tinggi yang menarik banyak mahasiswa setiap tahunnya, terbukti dengan meningkatnya pendaftar calon mahasiswa di setiap periode. Peningkatan tersebut harus sejalan dengan kualitas dan kenyamanan bagi mahasiswa dalam menjalani masa studinya. Jika ketidakpuasan mahasiswa karena kegagalan dalam memenuhi janji untuk menyelesaikan studi tepat waktu atau kegagalan dalam memenuhi janji lainnya maka akan berakibat terjadi pengurangan retensi mahasiswa atau perpindahan mahasiswa [2]. Oleh karena itu, pihak akademik harus memperhatikan hal yang terkait dengan kenyamanan mahasiswa agar dapat menjalani perkuliahan dengan baik dengan cara membantu dalam mengelola mata kuliah.

Sejauh ini kesulitan yang sering terjadi dalam mengelola mata kuliah ialah kesulitan memprediksi jumlah mahasiswa yang akan mengulang mata kuliah pada setiap semester. Hal ini berdampak pada pengaturan jadwal dan jumlah kelas yang akan diadakan bagi mahasiswa yang ingin mengulang mata kuliah tertentu. Pengaturan jadwal dilakukan pada setiap pergantian semester sesuai dengan jumlah mahasiswa yang mengambil mata kuliah tersebut. Sehingga jika terjadi kesalahan dalam memprediksi jumlah mahasiswa yang akan mengulang mata kuliah akan berdampak pada tidak tersedianya kelas dan pihak akademik harus mengatur ulang jadwal bagi mahasiswa yang tidak dapat kelas.

Penggalian informasi dari kumpulan data yang berskala besar dapat dilakukan dengan menggunakan teknik *Data Mining*. *Data Mining* mengacu pada proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data besar [3]. Definisi lain *Data Mining* adalah serangkaian proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer untuk menganalisis dan mengekstrak pengetahuan secara otomatis atau serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [4]. Salah satu teknik dari *Data Mining* yang dapat digunakan untuk melakukan prediksi pada data mahasiswa yang akan mengulang mata kuliah yaitu klasifikasi dengan algoritma C4.5. Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan [5]. Algoritma C4.5 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continue data*, dan *pruning* [6].

## 2. METODOLOGI PENELITIAN

Metodologi penelitian ini dilakukan secara sistematis yang dapat digunakan sebagai pedoman untuk peneliti dalam melaksanakan penelitian agar hasil yang dicapai tidak menyimpang dan tujuan yang diinginkan dapat terlaksana dengan baik dan sesuai dengan tujuan yang telah ditetapkan sebelumnya.



**Gambar 1. Kerangka Kerja Penelitian**

Berdasarkan kerangka kerja pada gambar 1, maka masing-masing langkahnya dapat diuraikan seperti berikut ini :

a. Mendeskripsikan Masalah

Mendeskripsikan masalah yang akan diteliti perlu ditentukan terlebih dahulu. Mendeskripsikan masalah dalam penelitian dengan menentukan dan mendefinisikan batasan masalah yang akan diteliti, sehingga membantu dalam mendapatkan suatu solusi yang terbaik dari masalah tersebut. Jadi, langkah pertama ini adalah langkah awal yang terpenting dalam penelitian ini.

- b. **Analisa Masalah**  
Langkah analisis masalah merupakan langkah untuk dapat memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisis masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik.
- c. **Mempelajari Literatur**  
Untuk mencapai tujuan, maka dipelajari beberapa literatur-literatur yang diperkirakan dapat digunakan. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian.
- d. **Mengumpulkan Data**  
Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di AMIK Labuhan Batu sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan *interview* yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan analisis terhadap data dan informasi yang didapat. Adapun data-data yang diperlukan dalam penelitian ini yang mencakup dari variabel yang akan diolah adalah :
  1. Semester. Variabel Semester berisi semester mahasiswa yakni semester 4 dan semester 6.
  2. Nilai. Nilai yang sudah ditentukan antara lain A, B, C, D, dan E.
  3. IPK. Variabel IPK dari setiap mahasiswa.
  4. Keadaan Ekonomi, merupakan variabel keadaan ekonomi mahasiswa yaitu penghasilan rata-rata orang tua per bulan.
  5. Status. Variabel ini berisi status mahasiswa apakah sudah kerja atau belum kerja.
  6. Keputusan. Variabel keputusan merupakan data yang berfungsi untuk menentukan hasil keputusan yakni MENGULANG atau TIDAK MENGULANG.
- e. **Analisa Teknik Pengolahan Data Menggunakan Algoritma C4.5**  
Data yang diperoleh dari tempat penelitian selanjutnya dilakukan analisa dan pengolahan menggunakan algoritma C4.5.
- f. **Perancangan Algoritma C4.5**  
Pada tahap ini akan dilakukan proses perancangan dari model sistem dengan algoritma C4.5 sehinggamembentuk pohon keputusan (*decision tree*) dan menghasilkan suatu *rule* prediksi mahasiswa yang mengulang mata kuliah.
- g. **Implementasi Algoritma C4.5**  
Adapun langkah-langkah dalam tahapan ini adalah :
  1. Menentukan atribut sebagai akar dan mengitung nilai informasi *gain* atribut.
  2. Menyusun *Tree* awal
  3. Menyusun *Tree* lanjutan
  4. Mengubah *Tree* menjadi *rule*
- h. **Pengujian Hasil**  
Pada tahap ini, penulis melakukan pengujian dan hasil perancangan sistem menggunakan *softwareData Mining open source WEKA*. Sistem diuji dengan prosedur-prosedur untuk melakukan eksplorasi dan permodelan dari data-data yang ada sehingga mendapatkan suatu hubungan tersembunyi dari data tersebut.

### 3. HASIL DAN PEMBAHASAN

Data yang akan diolah mempunyai beberapa kriteria yang merupakan syarat dalam pengolahan *Data Mining* dengan menggunakan teknik algoritma C4.5. Dimulai dengan perhitungan *Entropy* dan *Gain* untuk menentukan akar (*root*) dari pohon keputusan, sampai terbentuk pohon keputusan memprediksi mahasiswa yang mengulang mengambil mata kuliah.

#### 3.1 Melakukan Pra-proses

Variabel tersebut akan dikelompokkan berdasarkan atribut sebagai berikut :

- a. Mengelompokkan IPK, pengelompokan ini diambil dari IPK yang diraih oleh mahasiswa selama perkuliahan. pengelompokan IPK tersebut dapat dilihat pada tabel 1.

**Tabel 1. Klasifikasi IPK**

<i>IPK</i>	<i>Klasifikasi</i>
<1,50	Sangat Rendah
1,50 – 1,99	Rendah
2,00 – 2,49	Sedang
2,50 – 2,99	Baik
3,00 – 4,00	Baik Sekali

- b. Mengelompokkan Keadaan Ekonomi, pengelompokan ini diambil dari penghasilan rata-rata orang tua perbulan dari mahasiswa. Klasifikasi keadaan ekonomi dapat dilakukan dengan Rumus Sturgess [7]:

$$K = 1 + 3.3 \log n \quad (1)$$

Dimana :

K = jumlah kelas yang dicari

n = jumlah set data

$$K = 1 + 3.3 \log (34)$$

$$K = 1 + 3.3 (1.53147)$$

$$K = 6.0538 \text{ atau } 6$$

Untuk mendapatkan *range* dilakukan perhitungan lagi yaitu :

Gaji Tertinggi – Gaji Terendah

4.500.000 – 1.000.000 = 3.500.000, kemudian hasilnya dibagi jumlah kelas

$$3.500.000 / 6 = 583.333$$

Maka, *range* untuk klasifikasi keadaan ekonomi adalah 583.333, untuk lebih jelas dapat dilihat pada tabel 2.

**Tabel 2. Klasifikasi keadaan ekonomi**

<i>Keadaan Ekonomi</i>	<i>Klasifikasi</i>
3.916.667 – 4.500.000	A
3.333.333 – 3.916.667	B
2.750.000 – 3.333.333	C
2.166.667 – 2.750.000	D
1.583.333 – 2.166.667	E
1.000.000 – 1.583.333	F

- c. Mengelompokkan Status, pengelompokan ini diambil dari status mahasiswa, apakah sudah bekerja atau belum bekerja. Klasifikasi status dapat dilihat pada tabel 3.

**Tabel 3. Klasifikasi status**

<i>Status</i>	<i>Klasifikasi</i>
Bekerja	B
Belum Kerja	BK

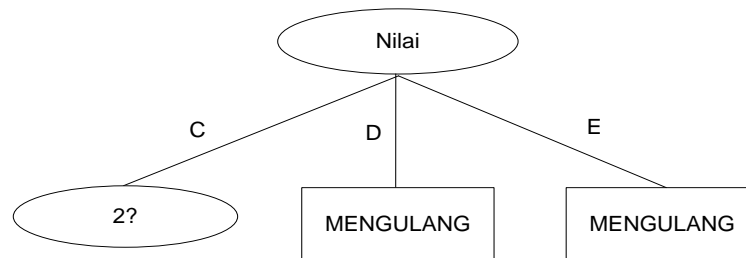
### 3.2 Pohon Keputusan

Dalam pembuatan pohon keputusan, yang harus dilakukan adalah menghitung jumlah kasus, jumlah kasus untuk keputusan “Mengulang”, jumlah kasus untuk keputusan “Tidak Mengulang” dan kasus yang dibagi berdasarkan atribut Semester, Nilai Mata Kuliah, IPK, Keadaan Ekonomi dan Status. Setelah itu, lakukan perhitungan *Gain* untuk setiap atribut. Hasil dari perhitungan *node* 1 dapat dilihat pada tabel 4.

**Tabel 4. Hasil perhitungan node 1**

<i>Node</i>		<i>Jumlah Kasus(S)</i>	<i>Mengulang (S1)</i>	<i>Tidak (S2)</i>	<i>Entropy</i>	<i>Gain</i>
1	Total	34	19	15	0.989992	
	*SEMESTER	4	20	13	0.934068	0.03485
		6	14	6	0.98522	
	*NILAI	C	24	9	0.95443	0.31627
		D	8	8	0	
		E	2	2	0	
	*IPK	Rendah Sekali	0	0	0	0.3133
		Rendah	4	4	0	
		Sedang	3	3	0	
		Baik	8	1	0.543564	
		Baik Sekali	19	11	0.98194	
	*KEADAAN EKONOMI	A	5	4	0.72192	<b>0.06617</b>
		B	2	1	1	
		C	2	1	1	
		D	7	5	0.86312	
		E	12	5	0.97986	
	*STATUS	F	6	3	1	
		K	8	6	0.81127	0.03439
		BK	26	13	1	

Adapun pohon keputusan yang terbentuk dari pencarian *Node 1* adalah seperti pada gambar 2.



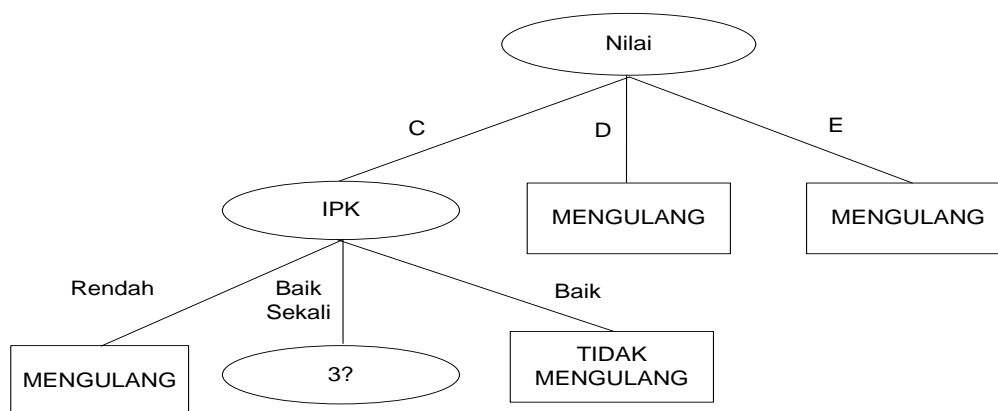
**Gambar 2. Pohon keputusan hasil perhitungan node 1**

Nilai D memiliki keputusan “Mengulang”, nilai E juga memiliki keputusan “Mengulang”, sedangkan nilai C masih perlu dilakukan perhitungan lagi. Selanjutnya, hasil dari perhitungan *node 1.1* dapat dilihat pada tabel 5.

**Tabel 5. Hasil perhitungan node 1.1**

<i>Node</i>	<i>Jumlah Kasus (S)</i>	<i>Mengulang (S1)</i>	<i>Tidak Mengulang (S2)</i>	<i>Entropy</i>	<i>Gain</i>	
1.2 Total Nilai C		24	9	15	0.9544	
*SEMESTER	4	10	2	8	0.7219	0.07029
	6	14	7	7	1	
*IPK	Rendah Sekali	0	0	0	0	<b>0.3140</b>
	Rendah	4	4	0	0	
	Sedang	0	0	0	0	
	Baik	8	1	7	0.5435	
	Baik Sekali	12	4	8	0.9182	
*EKONOMI	A	4	3	1	0.8112	0.1349
	B	1	0	1	0	
	C	1	0	1	0	
	D	3	1	2	0.9182	
	E	10	3	7	0.8812	
	F	5	2	3	0.9709	
*STATUS	K	6	4	2	0.9182	0.0855
	BK	18	5	13	0.8524	

Adapun pohon keputusan yang terbentuk dari pencarian *Node* 1.1 adalah seperti pada gambar 3.



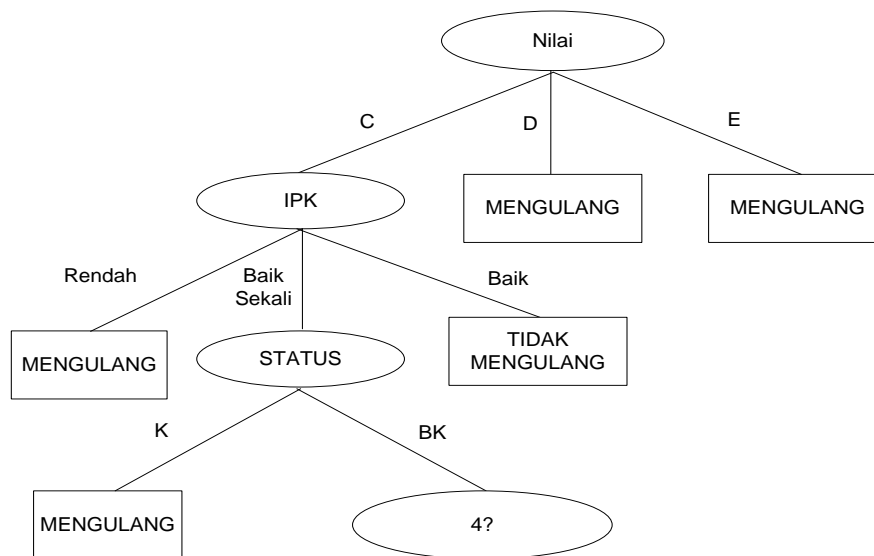
**Gambar 3. Pohon Keputusan Hasil Perhitungan Node 1.1**

IPK Rendah memiliki keputusan “Mengulang”, IPK Baik memiliki keputusan “Tidak Mengulang”, sedangkan IPK Baik Sekali masih perlu dilakukan perhitungan lagi. Selanjutnya, hasil dari perhitungan *node* 1.2 dapat dilihat pada tabel 6.

**Tabel 6. Perhitungan Node 1.2**

Node		Jumlah Kasus (S)	Mengulang (S1)	Tidak Mengulang (S2)	Entropy	Gain
1.3	Total Nilai C, IPK BS	12	4	8	0.9182	
	*SEMESTER	4	6	1	0.6500	0.0932
		6	6	3	1	
	*EKONOMI	A	3	2	0.91829	0.1174
		B	0	0	0	
		C	0	0	0	
		D	0	0	0	
		E	5	1	0.72192	
		F	4	1	0.81127	
	*STATUS	K	3	3	0	<b>0.5408</b>
		BK	9	1	0.50325	

Adapun pohon keputusan yang terbentuk dari pencarian *Node 1.2* adalah seperti pada gambar 4.



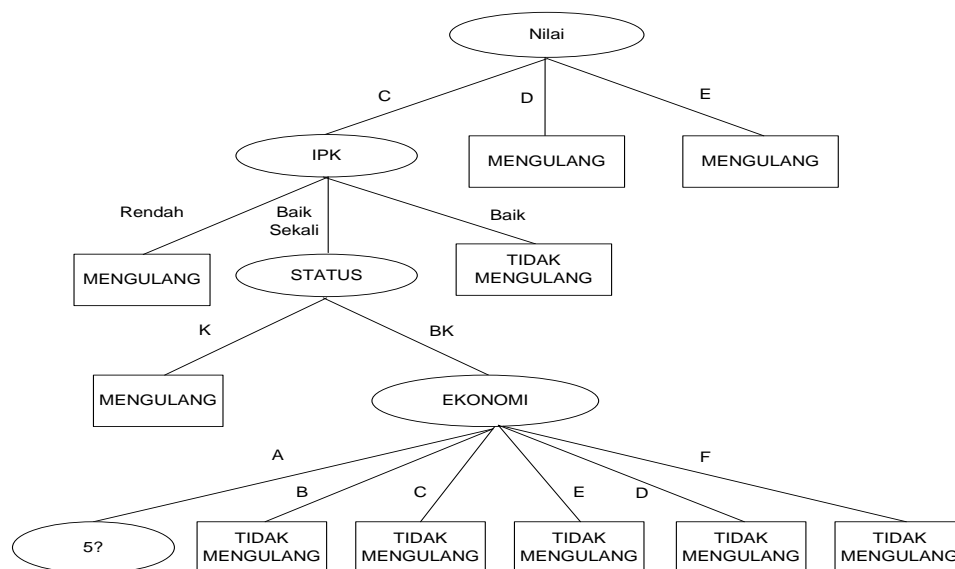
**Gambar 4. Pohon Keputusan Hasil Perhitungan Node 1.2**

Status Kerja memiliki keputusan “Mengulang”, sedangkan Status Belum Kerja masih perlu dilakukan perhitungan lagi. Selanjutnya, hasil dari perhitungan *node 1.3* dapat dilihat pada tabel 7.

**Tabel 7. Perhitungan Node 1.3**

<i>Node</i>	<i>Jumlah Kasus (S)</i>	<i>Mengulang (S1)</i>	<i>Tidak Mengulang (S2)</i>	<i>Entropy</i>	<i>Gain</i>
1.4 Total Nilai C, IPK BS, Status BK	9	1	8	0.50325	
*SEMESTER	4	5	0	0	0.14269
	6	4	1	0.81127	
*EKONOMI	A	2	1	1	<b>0.28103</b>
	B	0	0	0	
	C	0	0	0	
	D	0	0	0	
	E	4	4	0	
	F	3	0	0	

Adapun pohon keputusan yang terbentuk dari pencarian *Node 1.3* adalah seperti pada gambar 5.



**Gambar 5. Pohon Keputusan Hasil Perhitungan Node 1.3**

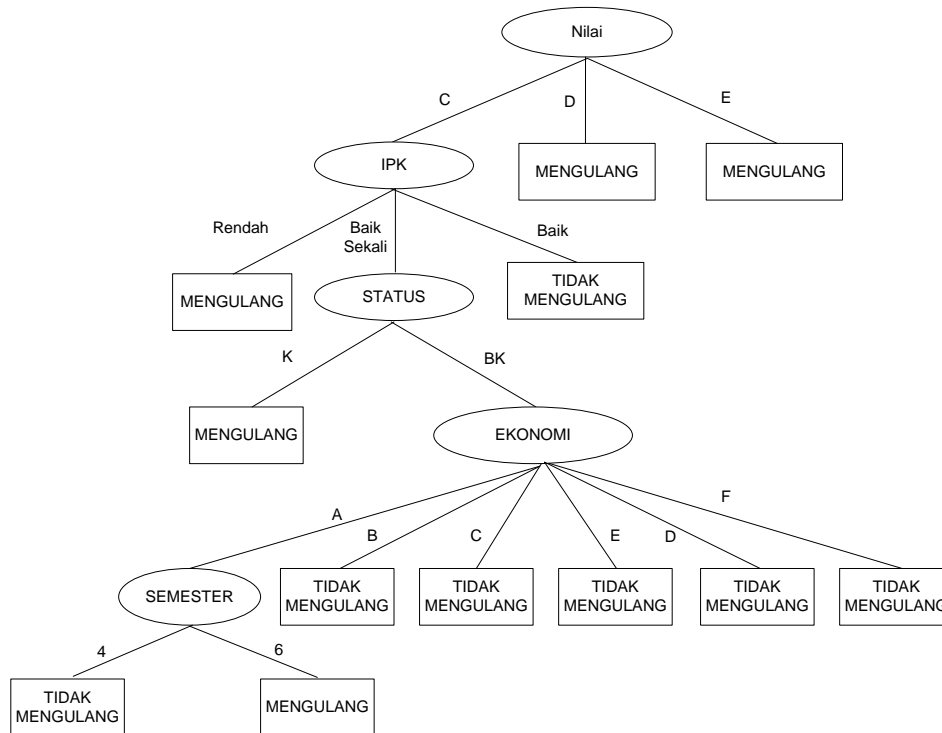
Ekonomi kategori B memiliki keputusan “Tidak Mengulang”, ekonomi kategori C memiliki keputusan “Tidak Mengulang”, ekonomi kategori D memiliki keputusan “Tidak Mengulang”, Ekonomi kategori E memiliki keputusan “Tidak Mengulang”, Ekonomi kategori F memiliki keputusan “Tidak Mengulang”, sedangkan Ekonomi kategori A masih perlu dilakukan perhitungan lagi. Selanjutnya, hasil dari perhitungan *node 1.4* dapat dilihat pada tabel 8.

**Tabel 8. Perhitungan Node 1.4**

<i>Node</i>	<i>Jumlah Kasus(S)</i>	<i>Mengulang (S1)</i>	<i>Tidak Mengulang (S2)</i>	<i>Entropy</i>	<i>Gain</i>
1.5 Total Nilai C, IPK BS, Status BK	2	1	1	1	
*SEMESTER	4	1	0	0	1
	6	1	1	0	0



Dari tabel 8 dapat dilihat bahwa tidak ada perbandingan *Gain* tertinggi karena yang dihitung hanya kasus pada Semester 4 dan 6. Atribut 4 memiliki nilai “Tidak Mengulang”, sedangkan 6 memiliki nilai “Mengulang”. Perhitungan akan selesai sampai pada tahap ini dan pohon keputusan akhir akan terbentuk. Pohon keputusan yang terbentuk sampai dengan *Node* 1.4 seperti gambar 6.



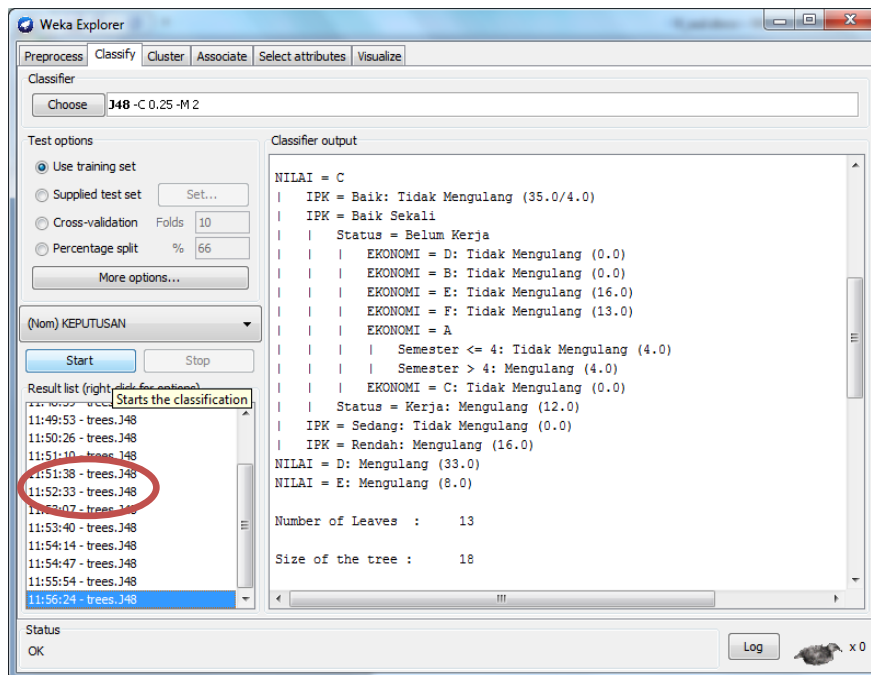
**Gambar 6. Pohon Keputusan Hasil Perhitungan *Node* 1.4**

Aturan atau *rule* yang terbentuk dari perhitungan *Gain* dan *Entropy* adalah sebagai berikut :

- Jika Nilai = D, maka Mahasiswa = Mengulang
- Jika Nilai = E, maka Mahasiswa = Mengulang
- Jika Nilai = C dan IPK = Rendah, maka Mahasiswa = Mengulang
- Jika Nilai = C dan IPK = Baik, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Kerja, maka Mahasiswa = Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = F, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = E, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = D, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = C, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = B, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = A dan Semester = 4, maka Mahasiswa = Tidak Mengulang
- Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi = A dan Semester = 6, maka Mahasiswa = Mengulang

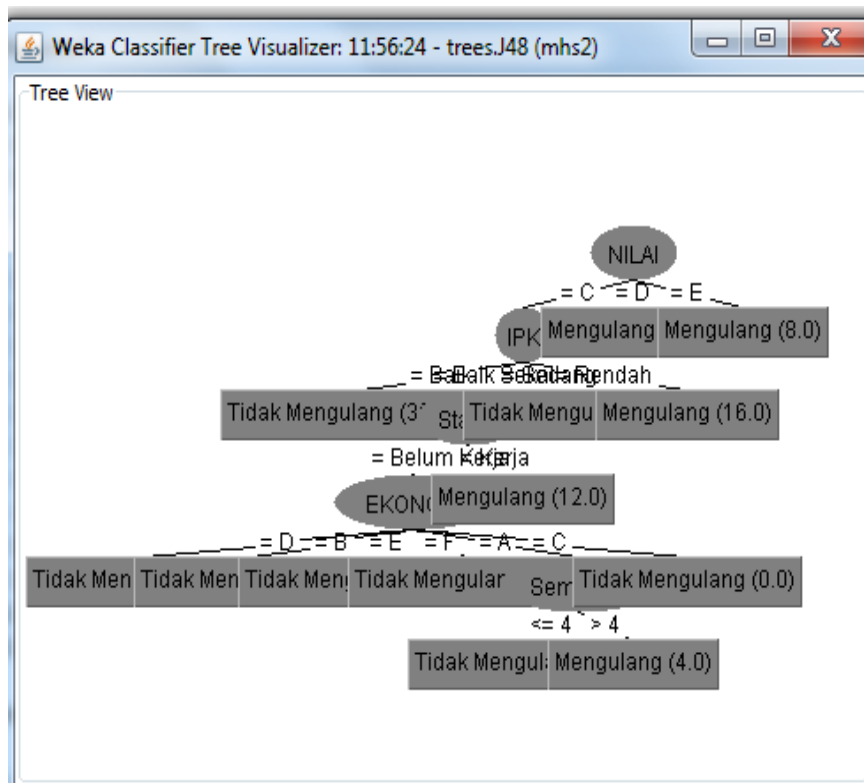
### 3.3 Pengujian

Untuk menguji kebenaran dari hasil pengolahan data yang dikerjakan secara manual dilakukan perbandingan hasil dengan menggunakan *software WEKA*. *Classifier output* berisi *running* data yang diinputkan yang merupakan atribut-atribut yang akan membentuk pohon keputusan. Informasi yang ditampilkan berupa jumlah kasus dan keputusannya dan banyaknya cabang dari pohon keputusan. *Classifier output* dapat dilihat pada gambar 7.



Gambar 7. Classifier Output

Untuk melihat pohon keputusan yang terbentuk dari hasil pengolahan WEKA adalah klik kanan pada *Resultlist*, pilih *Visualizetree*. Hasilnya akan terlihat pada gambar 8 berikut ini.



Gambar 8. Tree View oleh WEKA

Dari *tree view* oleh WEKA, adapun *rule* atau aturan yang terbentuk adalah sebagai berikut:

- Jika Nilai = D, maka mahasiswa = Mengulang.
- Jika Nilai = E, maka mahasiswa = Mengulang.
- Jika Nilai = C dan IPK = Baik, maka mahasiswa = Tidak Mengulang.

- d. Jika Nilai = C dan IPK = Rendah, maka mahasiswa = Mengulang.
- e. Jika Nilai = C dan IPK = Baik Sekali dan Status = Kerja, maka mahasiswa = Mengulang.
- f. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= F, maka mahasiswa = Tidak Mengulang.
- g. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= E, maka mahasiswa = Tidak Mengulang.
- h. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= D, maka mahasiswa = Tidak Mengulang.
- i. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= C, maka mahasiswa = Tidak Mengulang.
- j. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= B, maka mahasiswa = Tidak Mengulang.
- k. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= A dan Semester = 4, maka mahasiswa = Tidak Mengulang.
- l. Jika Nilai = C dan IPK = Baik Sekali dan Status = Belum Kerja dan Ekonomi= A dan Semester = 6, maka mahasiswa = Mengulang.

### 3.4 Pembahasan

Dari dua hasil pengujian yang telah dilakukan yaitu proses secara manual dan menggunakan *software WEKA 3.7* dapat kita ambil sebuah kesimpulan bahwa hasil pengujian sangat baik karena *rule* yang dihasilkan hampir sama. Perbedaannya hanya terletak pada atribut nilai yang masuk ke dalam *WEKA*, namun tidak mengubah hasil keputusan. Pada hitungan manual menggunakan 34 *record* dan pada *WEKA* menggunakan 141 *record*.

Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasi dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun aturan *if-then*, sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan. Pada pemilihan atribut juga sangat mempengaruhi dalam pengolahan Algoritma C4.5 karena keputusan sangat bergantung pada atribut yang dipilih.

## 4. KESIMPULAN

Berdasarkan dari penelitian yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

- a. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang digunakan dalam memprediksi sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan.
- b. *Data Mining* dengan algoritma C4.5 dapat membantu dalam melakukan prediksi mahasiswa yang mengulang mata kuliah.
- c. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi *rule* prediksi untuk menggambarkan proses yang terkait dengan prediksi mahasiswa yang mengulang.
- d. Pengujian dengan *WEKA* lebih efektif dan fleksibel karena pohon keputusan yang terbentuk hampir sama dengan perhitungan yang dilakukan dengan manual. Oleh karena itu, penggunaan *WEKA* juga sangat membantu dalam pengujian yang dilakukan.

## DAFTAR PUSTAKA

- [1] Utari, Putri, Sukma. 2015. *Implementasi Metode C4.5 Untuk Menentukan Guru Terbaik Pada SMK 1 Percut Sei Tuan Medan*. Pelita Informatika Budi Darma, Volume : IX, Nomor: 3, April 2015 ISSN : 2301-9425.
- [2] Foster, Bob. 2012. *Kualitas Jasa Pelayanan Di Perguruan Tinggi Swasta Indonesia (Sebuah Kajian Pustaka)*. Jurnal Ekono Insentif Kopwil4, Volume 6 No. 2, Oktober 2012.
- [3] Lorena, Silvia. 2014. *Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik*. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 ISSN: 1979-911X.
- [4] Sijabat, Alimancon. 2015. *Penerapan Data Mining untuk Pengolahan Data Siswa dengan Menggunakan Metode Decision Tree*. Jurnal Informasi dan Teknologi Ilmiah. Volume 5 No 3. ISSN : 2339-210X.
- [5] Elmande, Yusuf. 2012. *Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras : Studi Kasus Pada Perum Bulog Divre Lampung*. Jurnal TELEMATIKA MKOM.

- [6] Faradillah, Sarah. 2013. *Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer Dengan Menggunakan Algoritma C4.5*. Pelita Informatika Budi Darma, Volume : V, Nomor: 3.
- [7] Wismarini, T. D., & Anis, Y. 2014. *Metode Klasifikasi Spasial sebagai Pendukung Informasi Kelas pada Data Indikator Banjir*, 19(2), 120–136