

## IMPLEMENTASI METODE DECISION TREE C4.5 UNTUK MENGANALISA MAHASISWA DROPOUT

### IMPLEMENTATION OF DECISION TREE C4.5 METHOD TO ANALYZE THE DROPOUT STUDENTS

<sup>1</sup>Sri Wahyuni, <sup>2</sup>Kana Saputra Saragih, <sup>3</sup>Mochammad Iswan Perangin-Angin

<sup>1,2,3</sup>Program Studi Teknik Komputer, Universitas Pembangunan Pancabudi Medan  
email: <sup>1</sup>sriwahyuni@dosen.pancabudi.ac.id; <sup>2</sup>kanasaputras@dosen.pancabudi.ac.id;  
<sup>3</sup>mochammadiswan@gmail.com

**Abstract.** Data Mining is the process of extracting data from large databases to find important and useful information. Classification is one of the techniques that exist in data mining. The method used is the Decision Tree and the algorithm used is the algorithm C4.5. Decision Tree is a method that alters the facts into a decision tree that represents the rules which is easy to understand. Decision Tree is useful to explore the data and find the hidden relationship between the number of input variables and targets. The Decision Tree built will obtain rules from a case. The purpose of this study is to classify the student data in Pembnagunan Panca Budi University to determine students subjected to dropout. Attributes used consisted of the previous school, student's age, parent's occupation, parent's income, and GPA. To avoid too much branching, the attributes of income, age, and GPA are grouped together. The attribute that most influence on dropout students is the previous school. The results obtained from calculation accuracy value is calculation accuracy of 59.58% and classification error of 40.42%.

**Keywords:** C4.5 algorithm, Data Mining, Decision Tree, Drop Out Students

**Abstrak.** Data Mining merupakan proses penggalian data dari database yang berukuran besar untuk menemukan informasi penting dan bermanfaat. Klasifikasi adalah salah satu teknik yang ada pada data mining. Metode yang digunakan yaitu Decision Tree (Pohon Keputusan) dan algoritma yang digunakan yaitu algoritma C4.5. Decision Tree merupakan metode yang merubah fakta menjadi sebuah pohon keputusan yang mempresentasikan aturan-aturan yang mudah dipahami. Decision Tree berguna untuk mengeksplorasi data, serta menemukan hubungan yang tersembunyi antara sejumlah variabel input dan target. Decision Tree yang dibangun akan didapat rules dari suatu kasus. Tujuan penelitian ini untuk mengklasifikasikan data mahasiswa di Universitas Pembangunan Panca Budi untuk mengetahui faktor mahasiswa yang mengalami dropout. Atribut yang digunakan terdiri dari Asal Sekolah, Umur Mahasiswa, Pekerjaan Orangtua, Pendapatan Orangtua, dan IPK. Untuk menghindari percabangan yang terlalu banyak maka atribut Pendapatan, Umur, dan IPK dikelompokkan. Atribut yang paling berpengaruh terhadap mahasiswa yang mengalami dropout adalah Asal Sekolah. Hasil perhitungan nilai akurasi yang diperoleh perhitungan akurasi sebesar 59,58% dan classification error sebesar 40,42%.

**Kata Kunci:** Algoritma C4.5, Data Mining, Decision Tree, Drop Out, Klasifikasi, Mahasiswa

## 1. Pendahuluan

*Data Mining* yang juga dikenal sebagai *Knowledge Discovery* merupakan salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari *database* dalam skala besar. Definisi sederhana dari *data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada pada *database* yang besar. Menurut Larose dalam Kusri dan Emha (2009) *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklasteran, dan Asosiasi. Salah satu teknik yang umum digunakan adalah teknik Klasifikasi.

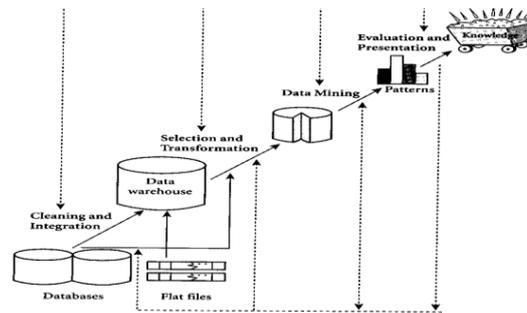
Klasifikasi adalah sebuah proses untuk menemukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Klasifikasi terdapat beberapa metode yaitu *decision tree*, *neural network*, *rough set theory*, *bayesian theory* dan *fuzzy logic* (Jantan *et al.*, 2010). Algoritma *Decision Tree C4.5* merupakan salah satu algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial.

Penelitian tentang *Decision Tree C4.5* pernah dilakukan untuk memutuskan kepuasan pelanggan dengan hasil penelitian terdapat tingkat akurasi sebesar 91%, dengan nilai presisi pada prediksi puas sebesar 92.21% dan nilai presisi pada prediksi tidak puas sebesar 90,91%. Class recall untuk puas sebesar 97,71% dan class recall untuk tidak puas sebesar 75%.. (Dhika *et al.* 2016). Penelitian lain implementasi *Decision Tree C4.5* untuk prediksi terhadap pengguna jasa sebuah operator seluler akan kehadirannya pada suatu acara berdasarkan beberapa indikator, cuaca, jarak relatif terhadap lokasi acara, serta apakah pengguna jasa tersebut merupakan termasuk pelanggan pasca bayar atau tidak. Akurasi dan performansi dari mesin pembelajaran sangat tergantung dari data yang ada serta pemahaman akan metode yang diterapkan, dalam penelitian ini diperoleh akurasi prediksi dengan angka yang sama di setiap teknik yang digunakan (As'ad Bahrawi 2016). Selain itu *Decision Tree* juga digunakan untuk memprediksi kriteria nasabah kredit dengan hasil disimpulkan bahwa aplikasi yang dibangun dapat membantu Bagian Dana dalam menganalisis data nasabah untuk menentukan target pemasaran kredit sehingga diharapkan biaya operasional marketing perbankan dapat ditekan seminimal mungkin (Mabrur dan Lubis 2012).

Metode *Decision Tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Tujuan dari penelitian ini adalah menganalisa data mahasiswa yang diekstrak dengan data mining decision tree C4.5 sehingga didapat pengetahuan baru yang dapat mengetahui faktor mahasiswa *dropout*.

## 2. Kajian Literatur dan Pengembangan Hipotesis

*Data mining* merupakan pencarian *trend* atau pola yang akan dicari dalam *database* yang besar untuk pengambilan keputusan di waktu yang akan datang (Fajar Astuti Hermawati, 2013). Ciri penting dari *data mining* adalah bahwa *volume* data yang sangat besar meskipun ide-ide dari area studi yang berhubungan tadi dapat diaplikasikan pada masalah-masalah *data mining*, *scalability* yang berkaitan dengan ukuran data menjadi suatu kriteria baru yang penting. *Data mining* dapat dibagi menjadi beberapa tahapan yang diilustrasikan pada Gambar.



**Gambar 1. Tahapan *Data Mining***

Keterangan:

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining).
4. Aplikasi teknik *Data Mining*.
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik / bernilai).
6. Presentasi pengetahuan (dengan teknik visualisasi).

Pengelompokan *data mining* terdiri dari deskripsi, estimasi, prediksi, asosiasi, pengklusteran, dan klasifikasi. Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Kemudian untuk menemukan pendapatan *seorang pegawai*, dipakai cara klasifikasi dalam *data mining*. Secara umum proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus. Pada proses belajar, algoritma klasifikasi mengolah data training untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan. Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar algoritma *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rules*) (Defiyanti dan Pardede 2010).

Konsep data dalam pohon keputusan adalah sebagai berikut (Muhammad Syahril, 2011):

1. Data dinyatakan dalam bentuk tabel dengan atribut dan record.
1. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut.
2. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut berat badan mempunyai *instance* berupa *overweight*, *average*, dan *underweight*.

Algoritma *Decision Tree C4.5* mempunyai input berupa *trainingsamples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data.

Secara umum algoritma *Decision Tree C4.5* untuk membangun pohon keputusan adalah sebagai berikut (Muhammad Syahril, 2011):

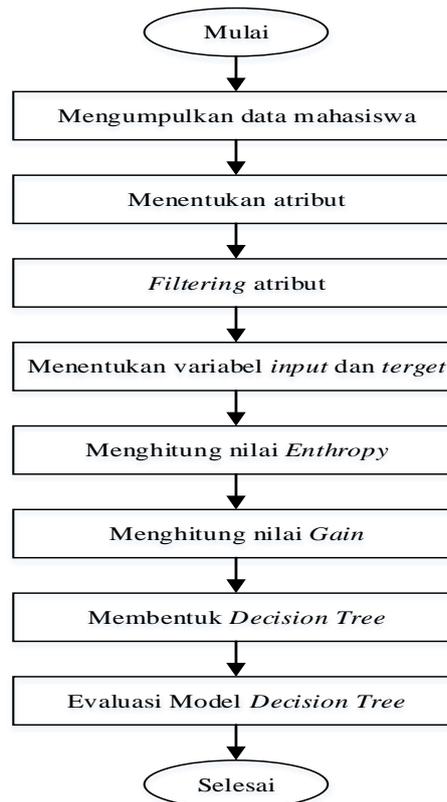
1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.

4. Ulangi proses untuk tiap cabang memiliki kelas yang sama.

### 3. Metode Penelitian

#### Tahapan Penelitian

Secara umum terdapat beberapa tahapan dalam penelitian ini dengan mengikuti pola umum penelitian ilmiah seperti yang terlihat pada Gambar 2.



**Gambar 2. Kerangka Kerja Penelitian**

#### Lokasi Penelitian

Penelitian ini dilakukan di Universitas Pembangunan Panca Budi Medan, khususnya pada bagian Biro Pelayanan Administrasi dan Akademik (BPAA). Hal ini dilakukan karena pada Universitas Pembangunan Panca Budi mengalami peningkatan jumlah mahasiswa yang signifikan, tapi sangat disayangkan jumlah mahasiswa yang masuk atau yang mendaftar dengan jumlah mahasiswa selesai atau wisuda sangat berbeda jauh, ini mengakibatkan banyaknya data jumlah mahasiswa berstatus *drop out*.

#### Pemilihan Variabel

Data mahasiswa yang digunakan dalam penelitian ini adalah mahasiswa Universitas Pembangunan Panca Budi Medan mulai tahun 2009 sampai dengan 2011. Variabel yang digunakan adalah sebagai berikut sebagai berikut:

1. Asal Sekolah
2. Umur Mahasiswa
3. Pekerjaan Orangtua

4. Pendapatan Orangtua
5. IPK

Dalam teknik *classification* keluaran dari setiap data yang dijadikan target atau *class* harus berupa bilangan bulat atau diskrit. Dari data mahasiswa di atas yang dijadikan parameter target atau variabel keputusan (*class*) adalah status mahasiswa yang berisi nilai parameter Aktif dan *Dropout*. Nilai parameter Aktif memiliki arti bahwa mahasiswa menyelesaikan studi dengan baik, sedangkan nilai parameter *dropout* adalah mahasiswa yang *dropout*. Dengan melihat keluaran atau *record* dari data pada *field* 'Status' yakni berupa bilangan bulat atau diskrit, Aktif dan Dropout, maka teknik *classification* dapat diterapkan untuk melakukan *data mining* data tersebut.

### Pengumpulan Data

Data yang digunakan untuk menganalisa mahasiswa *dropout* pada Universitas Pembangunan Panca Budi adalah data yang diperoleh dari bagian Biro Pelayanan Administrasi dan Akademik (BPAA). Data mahasiswa yang digunakan dalam penelitian ini adalah mahasiswa Universitas Pembangunan Panca Budi Medan mulai tahun 2009 sampai dengan 2011 untuk semua Fakultas dan Program Studi atau dengan kata lain diambil secara random. Kumpulan data mahasiswa dapat disimpan dalam *database* yang *cloud storage*, sehingga memudahkan dalam pengolahan data. Salah satunya aplikasi *owncloud* yang bersifat *open source* (Panjaitan Febriyanti 2017).

### Implementasi Decision Tree C4.5

Model yang digunakan untuk menentukan memprediksi mahasiswa yang berstatus *dropout* adalah algoritma *Decision Tree C4.5*. Tahapan algoritma *Decision Tree C4.5* adalah sebagai berikut (Swastina L, 2013):

1. Menyiapkan data *training*
2. Menentukan akar dari pohon
3. Hitung nilai *Gain*:

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

4. Ulangi langkah ke-2 hingga semua *tupel* terpartisi

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

5. Proses partisi pohon keputusan akan berhenti saat semua *tupel* dalam *node* N mendapat kelas yang sama dan atau tidak ada atribut di dalam *tupel* yang dipartisi lagi dan atau tidak ada *tupel* di dalam cabang yang kosong.

### Evaluasi Model Decision Tree C4.5

Evaluasi dilakukan dengan cara menghitung akurasi. Akurasi yang dihasilkan dihitung menggunakan *confusion matrix* (Andriani. 2012). Perhitungan *confusion matrix* dihitung berdasarkan prediksi positif yang benar (*True Positive*), prediksi positif yang salah (*False Positive*), prediksi negatif yang benar (*True Negative*), dan prediksi negatif yang salah (*False Negative*).

**Tabel 1. Confusion Matrix**

	Detected	
	Positive	Negative
Positive	A: True Positive	C: False Negative
Negative	A: True Positive	D: True Negative

Semakin tinggi hasil perhitungan akurasi yang dihasilkan, maka semakin baik model yang diperoleh. Rumus menghitung akurasi adalah sebagai berikut:

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

#### 4. Hasil dan Pembahasan

##### Pengumpulan Data

Penelitian ini menggunakan data mahasiswa di Universitas Pembangunan Panca Budi Medan sejumlah 2.400 mahasiswa. Data mahasiswa diambil dari tahun 2009 sampai dengan 2011 dengan status aktif dan *dropout*. Untuk Universitas Pembangunan Panca Budi tidak mengenal istilah *dropout*, tetapi *student losses*. Data yang telah diperoleh tersebut dibagi menjadi data *training* dan data *testing* dengan perbandingan 90% dan 10%, sehingga diperoleh 2.160 data *training* dan 240 data *testing*. Data *training* digunakan untuk memperoleh hasil klasifikasi mahasiswa *dropout* dalam bentuk *decisiontree*, sedangkan data *testing* digunakan untuk mengukur tingkat akurasi dari hasil klasifikasi tersebut. Contoh data sebelum preprosesing yang diperoleh dari BPAA dapat dilihat pada Gambar 3.

No	Nama Siswa	NIM	Asal Sekolah	Pekerjaan	Pendaftaran	IPK	Status Mahasiswa
1	ABDULLAH SALEH	1011210001	SMA	Wirawasta	3500000	3,00	Aktif
2	SHAMUZZAMAN	1011210002	SMA	PNIS	3500000	3,33	Aktif
3	EDY SAMPUTRA	1011210003	SMA	Lain-Lain	2400000	2,96	Aktif
4	M. REZA MANGOLATA	1011210004	SMA	Wirawasta	3000000	3,12	Aktif
5	FARIZ ABDULKADIR DALAMUNTHE	1011210005	SMA	Wirawasta	3000000	2,83	Aktif
6	MARISA PUSKASTI	1011210006	SMA	Wirawasta	2300000	3,17	Aktif
7	OTTO SANDA P. NARITURUJU	1011210007	SMA	Wirawasta	5000000	3,25	Aktif
8	MARISA TRI ANDRI	1011210008	SMA	Wirawasta	2500000	3,71	Aktif
9	DWI OCTAVIA	1011210009	SMA	PNIS	3500000	3,42	Aktif
10	DIAN PUTRI MANDASARI	1011210010	SMA	Lain-Lain	2000000	3,71	Aktif
11	MARINA DAMRANG	1011210011	SMA	Lain-Lain	2500000	3,08	Aktif
12	LINDA OCYITA	1011210012	SMA	PNIS	4000000	3,08	Aktif
13	WIKI PANJALA	1011210013	SMA	PNIS	3800000	3,71	Aktif
14	LAMBONG PERAWA AGUNG SITORU	1011210014	SMA	PNIS	4000000	3,54	Aktif
15	RENY	1011210015	SMA	PNIS	3500000	3,64	Aktif
16	MARINA MUSTARA NINGRUM	1011210016	SMA	PNIS	3500000	3,54	Aktif
17	INDAH KUNIA LESTARI	1011210017	SMA	Wirawasta	2600000	3,29	Aktif
18	DEB KARTIKA SASTRANG	1011210018	SMA	Wirawasta	2200000	3,29	Aktif
19	DENISE SAMPUTRA	1011210019	SMA	PNIS	3500000	3,54	Aktif

**Gambar 3. Contoh data mahasiswa Universitas Pembangunan Panca Budi**

##### Preprosesing Data

Preprosesing ini dilakukan untuk memilih (*filtering*) data yang tidak lengkap dan membentuk pengelompokan untuk masing-masing atribut dari data yang diperoleh dari BPAA. Jumlah data yang digunakan pada penelitian ini adalah 2.400 mahasiswa. Contoh data mahasiswa hasil preprosesing dapat dilihat pada Gambar 4.

No	Nama Siswa	NIM	Asal Sekolah	Umur	Pekerjaan	Prependikan	IPK	Status Mahasiswa
1	KABIRNA ZAULS	00121007	1004	20	Wiraswasta	A	A	ADIST
2	ZOHARUDZAN	00424003	1004	22	PGS	B	C	ADIST
3	EDD LAMPURITA	00421002	1004	21	Lain-Lain	B	B	ADIST
4	A. RA. BIDA MANGELITA	00421001	1004	21	Wiraswasta	C	A	ADIST
5	FARIZ ARDIANDHIAN DAURMANTHE	00177003	1004	20	Wiraswasta	C	B	ADIST
6	PANJALI MANGUNTIT	00177002	1004	21	Wiraswasta	C	A	ADIST
7	OTTO GANDA P. NARTIPULU	00100003	1004	20	Wiraswasta	A	A	ADIST
8	FRANSKA TRI ANTONI	00100002	1004	21	Wiraswasta	C	A	ADIST
9	DINA OCTAVI	00100001	1004	21	PGS	A	A	ADIST
10	DIANA PUTRI WANDAGARI	00100000	1004	21	Lain-Lain	C	A	ADIST
11	BANGLAN DAMASRI	00100009	1004	21	Lain-Lain	A	A	ADIST
12	DIANA OCTAVIA	00100008	1004	21	PGS	B	A	ADIST
13	IRWAN PANGILA	00100007	1004	21	PGS	A	A	ADIST
14	LARONDY FERIANA ADANG SITORUS	00100006	1004	21	PGS	B	A	ADIST
15	RIZKI	00100005	1004	21	PGS	B	A	ADIST
16	NURMA NURHANA NINGRUM	00100004	1004	21	PGS	B	A	ADIST
17	INDAH KURNIA LESTARI	00100003	1004	20	Wiraswasta	B	A	ADIST
18	DIANA KARTIKA NOSTANDA	00100002	1004	20	Wiraswasta	A	A	ADIST
19	DENISA LAMPURITA	00100001	1004	20	PGS	C	A	ADIST
20	DIANA SARAGAN	00100000	1004	21	Lain-Lain	B	A	ADIST
21	MANGOLFOLOP MANULANG	00100004	1004	21	Lain-Lain	C	B	ADIST
22	ELIZAVY	00100003	1004	21	PGS	A	A	ADIST

Gambar 4. Contoh data mahasiswa hasil preprocessing

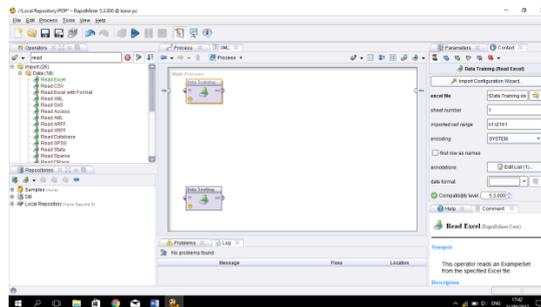
Keterangan:

- Pendapatan
  - A : > 3.500.000
  - B : 2.100.000 – 3.500.000
  - C : < 2.000.000
- IPK
  - A : 3,01 – 4,00
  - B : 2,51 – 3,00
  - C : < 2,50
- Umur
  - A : 23
  - B : 22
  - C : 21
  - D : 20
  - E : 19

### Implementasi Decision Tree C4.5 Menggunakan RapidMiner

Berikut tahapan untuk memperoleh model *Decision Tree* menggunakan *RapidMiner*:

- Tambahkan operator “*read excel*” ke dalam halaman kerja seperti yang terlihat pada Gambar 5. Terdapat 2 operator “*read excel*” yang digunakan untuk menyimpan Data *Training* dan Data *Testing*.



Gambar 5. Input Data *Training* dan *Testing*

- Gambar 6 merupakan proses penginputan data dengan cara hilangkan cek list pada atribut nama mahasiswa dan IPK (angka) karena nama mahasiswa dan IPK (angka) tidak digunakan. Atribut yang digunakan NIM (sebagai *id*), Asal Sekolah (sebagai *attributes*), Umur Grade (sebagai *attributes*), Pekerjaan Orangtua (sebagai *attributes*), IPK Grade (sebagai *attributes*), dan Status Mahasiswa (sebagai *label*).



### Model Decision Tree C4.5

Proses *Decision Tree* yang dilakukan menggunakan *software RapidMiner* diperoleh *tree*. Untuk mendapatkan informasi mengenai atribut yang paling berpengaruh dapat dilihat pada *tree* yang terbentuk. Berdasarkan proses *running* dari *Data Training* diperoleh hasil atribut yang paling berpengaruh adalah asal sekolah. Mahasiswa yang mengalami *dropout* berasal dari sekolah SMA dengan jumlah 759, SMK berjumlah 384, dan MA berjumlah 143.

### Pengujian Model Decision Tree C4.5

Untuk mengetahui akurasi dari model *Decision Tree* yang terbentuk, maka dilakukan pengujian menggunakan *Confusion Matrix* seperti yang terlihat pada Tabel 2.

**Tabel 2. Confusion Matrix**

	<i>pred. Aktif</i>	<i>pred. Dropout</i>	<i>class precision</i>
<i>pred. Aktif</i>	32	26	55.17%
<i>pred. Dropout</i>	71	111	60.99%
<i>class recall</i>	31.07%	81.02%	

Berdasarkan Tabel 3 diperoleh perhitungan akurasi sebesar 59,58% dan *classificationerror* sebesar 40,42%. Akurasi yang dihasilkan tidak begitu baik dikarenakan data yang digunakan tidak cukup representatif (bervariasi).

## 5. Kesimpulan

Data yang diperoleh berjumlah 2.400 mahasiswa yang dibagi menjadi 2.160 data *training* dan 240 data *testing*. Atribut yang digunakan terdiri dari Asal Sekolah, Umur Mahasiswa, Pekerjaan Orangtua, Pendapatan Orangtua, dan IPK. Untuk menghindari percabangan yang terlalu banyak maka atribut Pendapatan, Umur, dan IPK dikelompokkan. Atribut yang paling berpengaruh terhadap mahasiswa yang mengalami *dropout* adalah Asal Sekolah. Hasil perhitungan nilai akurasi yang diperoleh perhitungan akurasi sebesar 59,58% dan *classificationerror* sebesar 40,42%.

## Daftar Pustaka

- Andriani Anik. 2012. Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout. *Seminar Nasional Matematika*.
- As'ad Bahrawi. 2016. Prediksi Keputusan Menggunakan Metode Klasifikasi *Naïve Bayes*, *One-R*, dan *Decision Tree*. *Jurnal Penelitian Komunikasi dan Opini Publik*. 20(1): 1-10.
- Defiyanti S, Pardede DLC. 2010. Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi *Spam-Mail*. Gunadarma.
- Dhika H, Destiwati F, Fitriansyah A. 2016. Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan. *Prosiding SNaPP2016 Sains dan Teknologi*. 6(1): 16-22.
- Fajar Astuti Hermawati. 2013. *Data Mining*. Penerbit: CV. ANDI OFFSET.

- Jantan H, Hamdan AR, Othman ZA. 2010. *Human Talent Prediction in HRM Using C4.5 Classification Algorithm. International Journal on Computer Science and Engineering*. 2(8): 2526-2534.
- Kusrini, Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Penerbit: CV. ANDI OFFSET.
- Mabrur AG, Lubis R. 2012. Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit. *Jurnal Komputer dan Informatika*. 1(1): 53-57.
- Muhammad Syahril. 2011. Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan Dengan Teknik Klasifikasi Menggunakan Tools Rapid Miner 4.1. *Jurnal SAINTIKOM*.
- Panjaitan Febriyanti. 2017. Penerapan *Cloud Computing* Pada SMKN 1 Indralaya Selatan. Prosiding SNaPP2017 Sains dan Teknologi. 7(1):
- Swatina L. 2013. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. *Jurnal GEMA AKTUALITA*. 2(1): 93-98.