

## ANALISIS DAN PENERAPAN ALGORITHMMA C4.5 DALAM DATA MINING UNTUK MENUNJANG STRATEGI PROMOSI PRODI INFORMATIKA UPGRIS

Khoiriya Latifah<sup>1</sup>, Setyoningsih Wibowo<sup>2</sup>, Noora Qotrun Nada<sup>3</sup>

Informatika, Fakultas Teknik dan Informatika  
latifa.upgris@gmail.com, ninink.1623@gmail.com, noora.qn@gmail.com

### ABSTRAK

Untuk menarik minat pendaftar mahasiswa baru memerlukan strategi khusus. Salah satu strategi adalah dengan melakukan analisa data dengan tujuan mengubah kumpulan data menjadi memiliki nilai bisnis melalui laporan analitik sehingga menghasilkan informasi yang akan diambil polanya menjadi pengetahuan [6]. Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya [7]. Pohon keputusan merupakan metode klasifikasi dan prediksi. pada penelitian ini algoritma yang dipakai untuk pembentukan pohon keputusan dengan menggunakan algoritma C4.5 [7]. Data yang diproses adalah data mahasiswa baru angkatan 2014 dan angkatan 2015. Hasil penelitian ini menunjukkan bahwa variabel yang paling tinggi pengaruhnya terhadap hasil registrasi mahasiswa adalah Asal Sekolah dan Jenis Kelamin. Rata-rata berasal dari Semarang dengan jurusan SMU dari IPA dan yang berasal dari luar kota rata-rata berasal dari Batang dan Pati. Dari SMU jurusan IPS dan berjenis kelamin Laki-laki berasal dari Batang dan yang berjenis kelamin Perempuan berasal dari Pati. Akurasi dari pembentukan model ini adalah sebesar 89.33 % (*Good Classification*).

**Kata Kunci:** *Klasifikasi, Prediksi, Algorithma C.45, Strategi Promosi*

### ABSTRACT

To attract new student applicants requires a special strategy. One strategy is to perform data analysis with the aim of converting the data set to have business value through analytic reports so that the information will be taken into the pattern of knowledge [6]. The classification technique is an approximate classification function in data mining used to predict information previously unknown [7]. Decision tree is a method of classification and prediction. in this study the algorithm used for the formation of decision trees using the C4.5 algorithm [7]. Processed data are new student data of class of 2014 and class of 2015. The result of this research indicates that the variable that has the highest effect on student registration result is School Origin and Gender. The average comes from Semarang with high school majors from IPA and those coming from out of town on average come from Batang and Pati. Of SMU majoring in IPS and Male sex comes from the stem and the female sex is derived from Pati. Accuracy of this model is 89.33% (*Good Classification*).

**Keywords:** *Classification, Prediction, C4.5 Algorithm, Promotion Strategy*

<http://dx.doi.org/10.15408/jti.v11i2.6706>

## I. PENDAHULUAN

Persaingan PTN maupun PTS yang semakin kompetitif merupakan tantangan yang cukup besar, satu perguruan tinggi harus memiliki celah dan strategi khusus agar mampu bersaing dengan Perguruan Tinggi lain. Salah satu strategi adalah dengan melakukan analisa data dengan tujuan mengubah kumpulan data menjadi memiliki nilai bisnis melalui laporan analitik sehingga menghasilkan informasi yang akan diambil polanya menjadi pengetahuan [6].

Data yang terkumpul dari tahun ke tahun bisa dilakukan analisis untuk menghasilkan informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen perguruan tinggi dalam membuat strategi promosi yang tepat. Salah satu contohnya adalah data mahasiswa baru.

Data mahasiswa baru yang berasal dari data pada saat pendaftaran atau registrasi misalnya berupa nama mahasiswa, jurusan SLTA, nilai UAN, kota asal mahasiswa, program studi yang dipilih dan IPK, nilai hasil tes masuk perguruan tinggi, asal sekolah, tempat tinggal serta kondisi ekonomi

Untuk mengelola data tersebut dibutuhkan sebuah metode yang bisa digunakan untuk menggali informasi tersembunyi dari data tersebut. Metode tersebut dikenal dengan *data mining*. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [6]. Teknologi *data mining* merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak digunakan pada lingkungan aplikasi bisnis seperti perbankan, *provider* telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain [1]. Dengan bantuan perangkat lunak, *data mining* akan melakukan proses analisis data untuk menemukan pola atau aturan tersembunyi dalam lingkup himpunan data tersebut.

Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya

[7]. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, *Artificial Neural Networks* (ANN), *K-Nearest Neighbor* (KNN), algoritma *Naive Bayes*, *Neural Network* serta algoritma lainnya.

Pada studi kasus ini, analisis *data mining* dilakukan dengan metode pohon keputusan.

Pohon keputusan merupakan metode klasifikasi dan prediksi. Algoritma yang dapat dipakai untuk pembentukan pohon keputusan salah satunya menggunakan algoritma C45 [7]. Hasil yang diperoleh dari penelitian ini diarpakan dapat menampilkan profil mahasiswa, keterkaitan antara profil mahasiswa dan sebaran asal sekolahnya. Dari hasil analisis tersebut diharapkan dapat membantu tim promosi dalam menyusun strategi promosi program studi.

Dari permasalahan tersebut diatas maka untuk menunjang strategi promosi perlu menentukan metode atau strategi promosi lokasi promosi prodi Informatika UPGRIS.

## II. TINJAUAN PUSTAKA

Menurut [4] menjelaskan bahwa "*Data Mining* merupakan pemilihan atau "menggali" pengetahuan dari jumlah data yang banyak." Berbeda dengan Segall, Guha & Nonis (2008:127) menjelaskan "*Data Mining* disebut penemuan pengetahuan atau menemukan pola yang tersembunyi dalam data. *Data mining* adalah proses menganalisis data dari perspektif yang berbeda dan meringkas menjadi informasi yang berguna". Bisa disimpulkan *Data mining* adalah Proses menganalisis data yang banyak dan membuat suatu pola untuk menjadi informasi yang berguna.

Klasifikasi adalah suatu *fungsi* *data mining* yang akan menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek di dalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian.

Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis training data. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan  $Y=F(X)$  dimana Y adalah kelas hasil prediksi dan X adalah *tuple* yang ingin diprediksi kelasnya. Selanjutnya pada tahap pengklasifikasian, model yang telah dihasilkan

akan digunakan untuk melakukan klasifikasi.

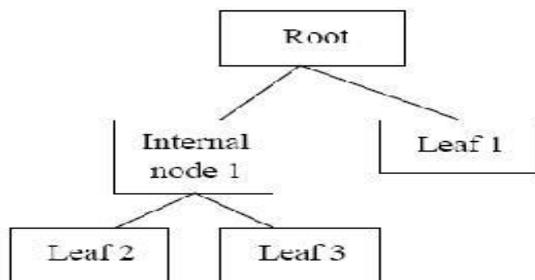
Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan/ internal (*branch/ internal node*) dan simpul daun (*leaf node*) [11].

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [11].



Gambar 1. Konsep pohon keputusan

Pohon keputusan merupakan himpunan aturan *IF...THEN*. Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, di mana premis terdiri atas sekumpulan *node-node* yang ditemui, dan kesimpulan dari aturam terdiri atas kelas yang terhubung dengan *leaf* dari *path* [12].



Gambar 2. Bentuk pohon keputusan

Bagian awal dari pohon keputusan ini adalah titik akar (*root*), sedangkan setiap cabang dari pohon keputusan merupakan pembagian berdasarkan hasil uji, dan titik akhir (*leaf*) merupakan pembagian kelas yang dihasilkan. Pohon keputusan banyak mengalami perkembangan, beberapa *algoritma* yang populer dan sering dipakai adalah C4.5 dan *CART*.

Tabel 1. Frekuensi penggunaan algoritma pohon keputusan

Algoritma Pohon Keputusan	Frekuensi
C4.5	54.55 %
CART	40.9 %
SPRINT	31.84 %
SLIQ	27.27 %
PUBLIC	13.6 %
C5.0	9 %
CLS	9 %
RANDOM FOREST	9 %
RANDOM TREE	4.5 %
ID3+	4.5 %
OCI	4.5 %
CLOUDS	4.5 %

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain [4].

Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami.

Dalam algoritma C4.5 untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiap-tiap nilai di dalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari atribut-tribut yang ada. Untuk menghitung *gain* digunakan rumus sebagai berikut [4]:

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

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

Sehingga akan diperoleh nilai gain dari atribut yang paling tertinggi. Gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu *node*.

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan:

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

Keterangan:

S : Himpunan kasus

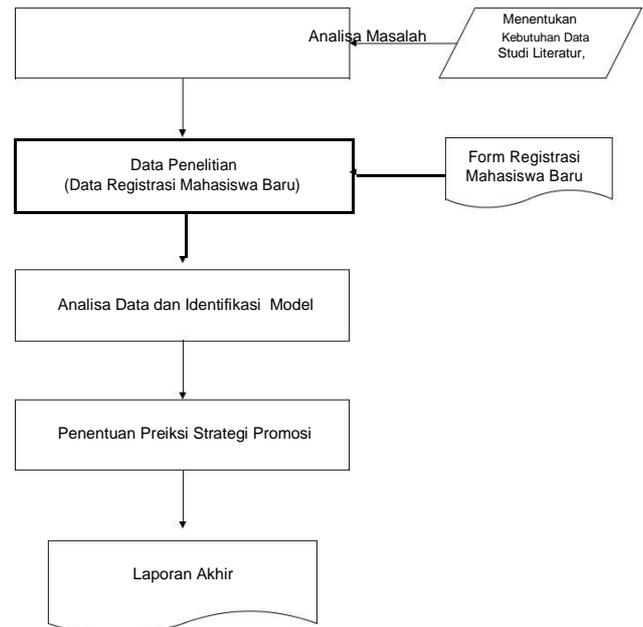
A : Atribut

N : Jumlah partisi S

P<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S

### III. METODOLOGI

Desain penelitian Analisa dan Penerapan Algoritma C45 Dalam *Data Mining* Untuk Menunjang Strategi Promosi:



Gambar 3. Desain penelitian

#### 3.1 Analisa Data dan Identifikasi Model

Menghimpun data yang bersumber dari bagian akademik untuk mahasiswa Informatika, selanjutnya diintegrasikan untuk dilakukan *data cleaning*.

Selanjutnya langkah-langkah yang dilakukan adalah menyusun data-data yang akan diolah termasuk parameter dan alternatif yang akan dirangking, merancang dan memproses perhitungan menggunakan metode C45 untuk mendapatkan nilai pengelompokan mahasiswa berdasarkan Nilai Rata-rata UN dan IPK. Hasil proses yang telah dinyatakan valid selanjutnya dilakukan analisis yang dikaitkan dengan kelompok dari asal kota sekolahnya.

#### 3.2 Penentuan Prediksi

Algoritma C45 membangun pohon keputusan adalah dengan langkah sebagai berikut:

- Pilih atribut sebagai akar
- Buat cabang untuk tiap-tiap nilai
- Bagi kasus dalam cabang
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas semua

#### 3.3 Data Selection

Data yang digunakan dalam penelitian ini adalah data mahasiswa baru prodi Informatika angkatan 2014 dan 2015, terdiri dari beberapa tabel antara lain tabel mahasiswa dan tabel registrasi mahasiswa. Tabel mahasiswa berisi tentang informasi data mahasiswa dan tabel

registrasi mahasiswa berisi tentang informasi calon mahasiswa baru. Jumlah *dataset* asli pada data registrasi mahasiswa sebanyak 116 atau selama 2 angkatan. Dari semua atribut yang ada pada tabel mahasiswa dan registrasi mahasiswa terdapat 4 atribut yang akan digunakan dalam proses *knowledge discovery in databases (KDD)*. Atribut tersebut yaitu:

- a. NPM merupakan atribut yang terdapat pada tabel mahasiswa dan tabel registrasi mahasiswa yang berperan sebagai *primary key* dalam menghubungkan tabel mahasiswa dan tabel registrasi mahasiswa.
- b. JENIS KELAMIN merupakan atribut yang terdapat pada tabel mahasiswa yang berisi informasi mengenai jenis kelamin mahasiswa.

- c. AGAMA merupakan atribut yang terdapat pada tabel mahasiswa yang berisi mengenai agama mahasiswa.
- d. JURUSAN merupakan atribut yang terdapat pada tabel mahasiswa yang berisi mengenai jurusan SMU dari mahasiswa.
- e. ASAL SEKOLAH merupakan atribut yang terdapat pada tabel mahasiswa yang berisi mengenai asal SMU dari mahasiswa apakah dari SMU atau SMK.
- f. REGISTRASI merupakan atribut yang terdapat pada tabel mahasiswa yang berisi mengenai registrasi mahasiswa baru.

Tabel 2. berikut adalah *field* yang digunakan untuk pemetaan mahasiswa berdasarkan kota asal mahasiswa.

Tabel 2. Field pemetaan mahasiswa berdasarkan kota asal mahasiswa

j_kel	agama	jurusan	asal_sekul	kota_asal	registrasi
Laki-laki	Islam	IPA	SMU	Semarang	YA
Perempuan	Kristen	IPS	SMK	Demak	TIDAK
		BHS		Purwodadi	
		Budha		Kudus	
		Hindu		Jepara	
				Pati	
				Rembang	
				Kendal	
				Batang	
				Pemalang	
				Tegal	
				Brebes	

### 3.4 Preprocessing

Pada tahapan *preprocessing* ini akan dilakukan proses integrasi data untuk menghubungkan tabel mahasiswa dan tabel registrasi mahasiswa, selanjutnya dilakukan proses data *cleaning* untuk menghasilkan *dataset* yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu *mining* dengan tujuan memperoleh pola mengenai kriteria mahasiswa untuk menentukan kriteria mahasiswa baru. Berikut merupakan penjelasan dari kedua proses di atas:

Integrasi Data, tahap integrasi data adalah tahap penggabungan data dari beberapa sumber. Pada tahapan ini dilakukan penggabungan dua tabel yaitu tabel mahasiswa dan tabel registrasi mahasiswa. Proses penggabungan dilakukan dengan merelasikan tabel mahasiswa dan tabel registrasi

mahasiswa dengan *query* seperti di bawah ini:

```
Select mahasiswa.npm, mahasiswa.nama,
mahasiswa.jkel, mahasiswa.agama,
mahasiswa.jurusan_mahasiswa.sma,mahasiswa.asal_mahasiswa.sekul,
mahasiswa.kota_asal from mahasiswa
LEFT JOIN registrasi_mahasiswa ON
mahasiswa.npm=registrasi_mahasiswa.npm;
```

*Data Cleaning*, tahap data *cleaning* merupakan tahap awal dari proses KDD.

Pada tahapan ini data yang yang tidak *relevan*, *missing value*, dan *redundant* harus di bersihkan. Hal ini dikarenakan data yang *relevan*, tidak *missing value*, dan tidak *redundant* merupakan syarat awal dalam melakukan *data mining*. Suatu data dikatakan *missing value* jika terdapat atribut dalam *dataset* yang tidak berisi nilai atau kosong, sedangkan data dikatakan *redundant* jika dalam

satu *dataset* terdapat lebih dari satu *record* yang berisi nilai yang sama.

Berdasarkan variabel-variabel yang sudah terpilih, format data sebagaimana tabel berikut:

Tabel 3. Format data calon mahasiswa setelah pemilihan variabel

nem	jk	sekolah	jurusan	.....

....	gelombang	pilihan1	pilihan2	nilai	.....
....	agama	jur_lu	nama_	nama_	registrasi
		lus	prop	kab	

Setelah data dalam format seperti tabel di atas langkah selanjutnya adalah melakukan *pre-processing* data.

a. Mengelompokkan Jenis Kelamin

Pengelompokkan Jenis Kelamin dilakukan dengan memasukkan jenis kelamin mahasiswa dalam *range* seperti tabel dibawah ini:

Tabel 4. Klasifikasi jenis kelamin

Jenis Kelamin	KLASIFIKASI
1	Perempuan
2	Laki - Laki

b. Mengelompokkan Agama

Mengelompokkan agama dilakukan dengan memasukkan agama mahasiswa dalam *range* seperti tabel di bawah ini:

Tabel 5. Klasifikasi agama

Agama	KLASIFIKASI
01	Islam
02	Protestan
03	Katolik
04	Hindu
05	Budha

c. Menerjemahkan Jurusan Sekolah

Tabel 6. Klasifikasi jurusan

Jurusan	KLASIFIKASI
01	IPA
02	IPS
03	BHS

d. Menerjemahkan Asal Sekolah

Tabel 7. Klasifikasi asal sekolah

Jurusan	KLASIFIKASI
01	SMU
02	SMK

e. Menerjemahkan Kota Asal Sekolah

Tabel 8. Tabel klasifikasi kota asal sekolah

Kota Asal Sekolah	KLASIFIKASI
1	Semarang
2	Demak
3	Purwodadi
4	Kudus
5	Jepara
6	Pati
7	Rembang
8	Kendal
9	Batang
10	Pemalang
11	Tegal
12	Brebes

### 3.5 Transformation

Tahapan *transformation* data merupakan tahap merubah data ke dalam bentuk yang sesuai untuk di-*mining*. Perubahan awal yang dilakukan yaitu menambahkan 1 atribut baru yaitu atribut registrasi. Atribut registrasi merupakan atribut yang berisi daftar registrasi mahasiswa baru.

### 3.6 Pemilihan Variabel

Data sampel awal yang digunakan dalam penelitian ini data mahasiswa Informatika angkatan tahun 2014 dan 2015. Data mahasiswa awal sebagai data *testing* dapat dilihat pada Tabel 9.

### 3.7 Pemilihan Variabel

Data sampel awal yang digunakan dalam penelitian ini data mahasiswa Informatika angkatan tahun 2014 dan 2015.

Kolom yang diambil sebagai variabel keputusannya adalah variabel Registrasi, sedangkan kolom yang diambil sebagai variabel penentu dalam pembentukan pohon keputusan adalah kolom NPM, Jenis Kelamin, Agama, Jurusan Sekolah, Asal Sekolah, Kota Asal Sekolah

Pemilihan variabel tersebut dengan pertimbangan bahwa jumlah nilai variabel variabelnya tidak banyak sehingga diharapkan calon mahasiswa yang masuk dalam klasifikasi nilai variabel tersebut cukup banyak. Data mahasiswa prodi Informatika selama 2 tahun ada 116 *record* yang bisa terbagi menjadi data *testing* dan data *training* untuk menemukan model klasifikasi C.45 yang digunakan dalam penelitian ini.

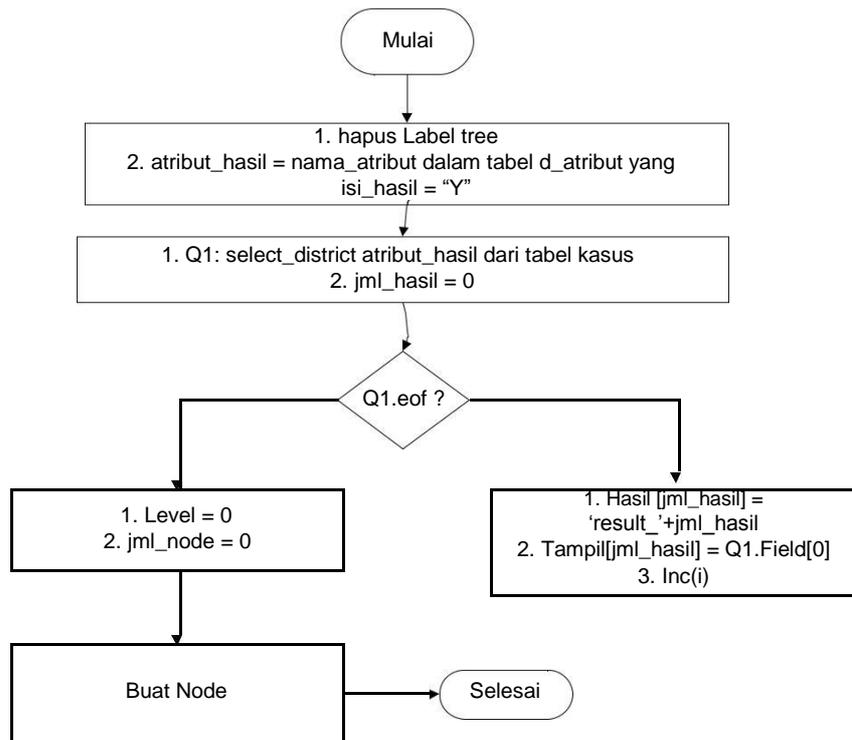
*Training* data berjumlah 76 *record* digunakan untuk pembelajaran sehingga diketahui sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang ada pada *training set* tidak boleh ada pada *testing set* untuk mengetahui apakah model *classifier* sudah “pintar” dalam melakukan klasifikasi.

Sedangkan *testing set* data berjumlah 40 *record* diambil untuk *validation set*. *Validation set* ini digunakan untuk mencari parameter yang paling baik untuk sebuah algoritma klasifikasi. Pemisahan data menjadi *training* dan *testing set* dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Model klasifikasi dapat melakukan klasifikasi data dengan sangat baik pada *training set*, tetapi sangat buruk dalam melakukan klasifikasi data yang baru dan belum pernah ada.

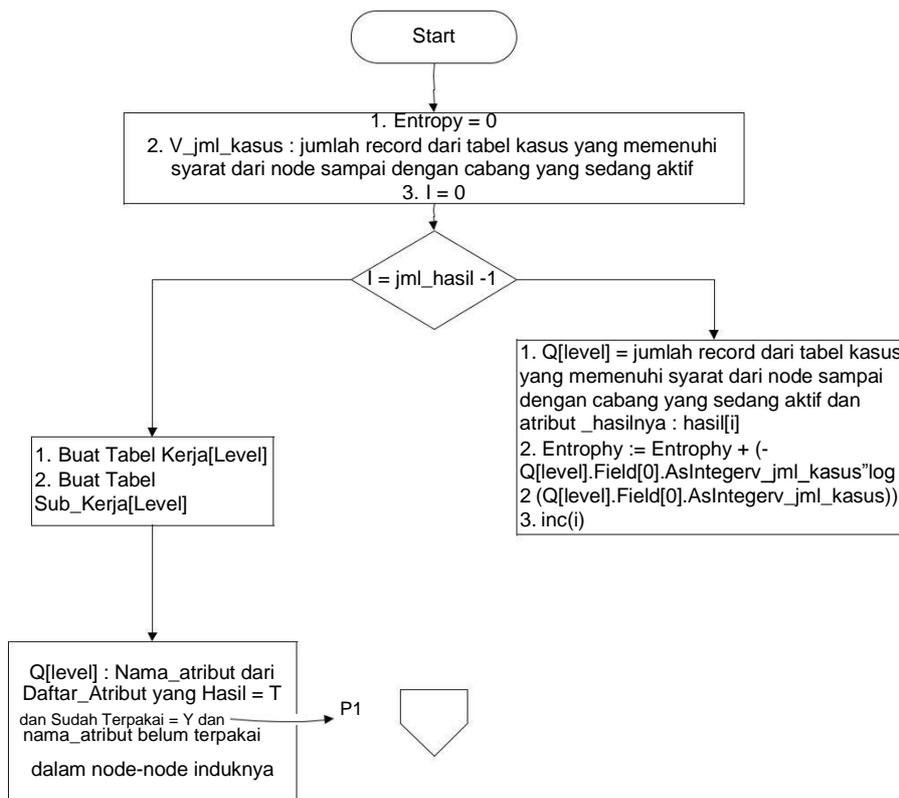
Tabel 9. Tabel Data Testing Klasifikasi C.45

<b>j_kel</b>	<b>agama</b>	<b>jurusan</b>	<b>asal_sekul</b>	<b>kota_asal</b>	<b>registrasi</b>
LAKI-LAKI	Islam	IPA	SMK	Demak	YA
PEREMPUAN	Islam	IPA	SMAN	Demak	YA
PEREMPUAN	Islam	IPS	SMAN	Purwodadi	YA
PEREMPUAN	Katholik	IPA	SMK	Semarang	YA
LAKI-LAKI	Islam	IPA	SMA	Rembang	TIDAK
LAKI-LAKI	Islam	IPA	SMKN	Rembang	YA
LAKI-LAKI	Islam	IPS	SMAN	Batang	YA
LAKI-LAKI	Islam	IPA	SMKN	Semarang	YA
LAKI-LAKI	Islam	IPA	SMK	Semarang	TIDAK
LAKI-LAKI	Islam	IPA	SMA	Batang	YA
LAKI-LAKI	Islam	BHS	SMKN	Semarang	YA
LAKI-LAKI	Islam	IPS	SMAN	Purwodadi	YA
LAKI-LAKI	Islam	BHS	SMK	Kudus	YA
PEREMPUAN	Islam	IPA	SMA	Kendal	YA
LAKI-LAKI	Islam	IPS	SMAN	Brebes	YA
PEREMPUAN	Islam	IPA	SMA	Semarang	YA
PEREMPUAN	Islam	IPS	MAN	Pati	YA
LAKI-LAKI	Islam	BHS	SMKN	Purwodadi	YA
PEREMPUAN	Islam	IPA	SMK	Ungaran	YA
PEREMPUAN	Islam	IPS	SMAN	Blora	YA
LAKI-LAKI	Islam	BHS	SMAN	Pemalang	YA
LAKI-LAKI	Islam	IPA	SMK	Pemalang	YA
PEREMPUAN	Islam	IPS	SMK	Brebes	YA
LAKI-LAKI	Islam	IPA	SMKN	Kendal	YA
PEREMPUAN	Islam	IPA	SMAN	Kudus	YA
LAKI-LAKI	Islam	IPA	SMKN	Semarang	YA
LAKI-LAKI	Islam	IPA	SMKN	Semarang	YA
LAKI-LAKI	Islam	IPA	SMAN	Semarang	YA
PEREMPUAN	Islam	IPS	SMAN	Pati	YA
LAKI-LAKI	Islam	IPA	SMKN	Semarang	YA
LAKI-LAKI	Kristen	BHS	MAN	Demak	YA
LAKI-LAKI	Islam	IPS	SMAN	Pemalang	YA
LAKI-LAKI	Islam	IPA	SMAN	Jepara	YA
LAKI-LAKI	Islam	IPA	SMA	Semarang	YA
LAKI-LAKI	Islam	IPA	SMKN	Semarang	TIDAK
LAKI-LAKI	Islam	IPA	SMA	Pati	TIDAK
LAKI-LAKI	Islam	IPA	SMAN	Cirebon	YA
LAKI-LAKI	Islam	IPA	SMK	Semarang	YA
PEREMPUAN	Islam	IPA	SMAN	Semarang	YA

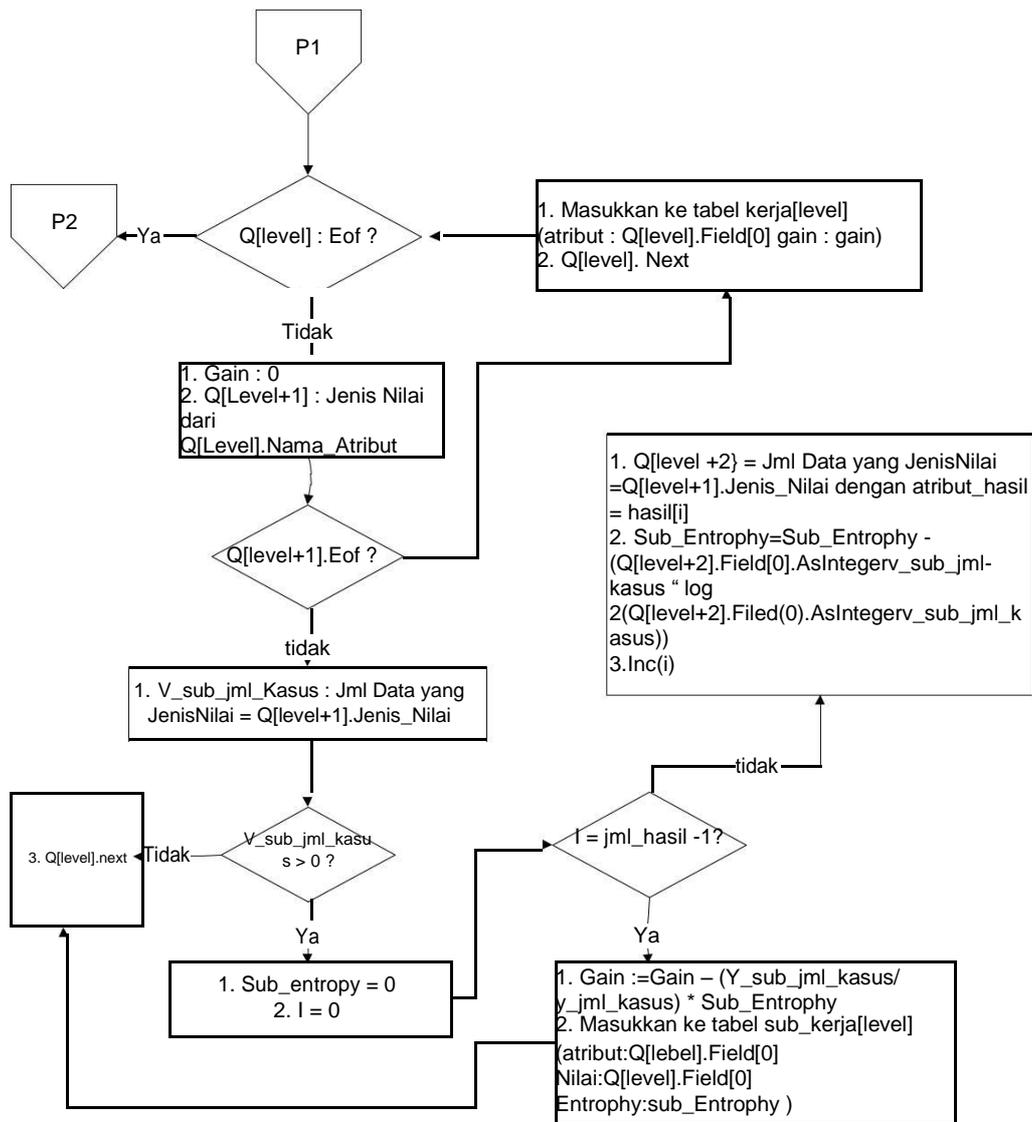
IV. HASIL DAN PEMBAHASAN



Gambar 4. Algoritma pembentukan node 1



Gambar 5. Algoritma pembentukan node 2



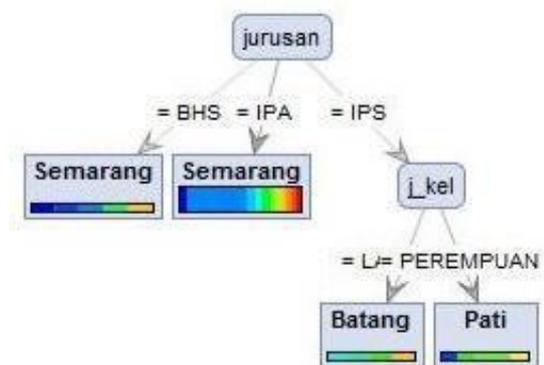
Gambar 6. Algoritma pembentukan node 3

**4.1 Pembentukan Model**

Data yang digunakan untuk proses klasifikasi penentuan registrasi mahasiswa baru menggunakan *decision tree* sebanyak 116' data untuk mengetahui pola atau tren mahasiswa Informatika yang akhirnya mendaftarkan dan teregistrasi sebagai mahasiswa Informatika UPGRIS berdasarkan variabel-variabel yang diajukan.

Hasil proses klasifikasi dengan metode *Decision Tree* ditunjukkan pada Gambar 7. Berdasarkan hasil pohon keputusan pada Gambar 7, dapat dilihat bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan klasifikasi mahasiswa adalah yang berasal dari SMU dan berjenis kelamin laki-laki serta berasal dari Semarang. Hal ini ditunjukkan dengan simpul akar (*root node*).

Pengolahan data menggunakan *tools Rapid Miner* sehingga model yang terbentuk:



Gambar 7. *Decision tree* C.45

Sesuai model di atas dapat kami simpulkan bahwa pemetaan mahasiswa prodi Informatika dalam 2 tahun angkatan 2014 dan 2015 berdasarkan kota asal rata-rata berasal dari Semarang dengan jurusan SMU dari IPA dan yang berasal dari luar kota rata-rata berasal dari Batang dan Pati. Dari SMU jurusan IPS dan berjenis kelamin Laki-laki berasal dari Batang dan yang berjenis kelamin Perempuan berasal dari Pati. Model tersebut diatas mempunyai *Accuracy*: 89.33% seperti pada tabel berikut:

Tabel 10. Tabel accuracy

	True YA	True TIDAK	Class Precision
pred. YA	67	8	89.33%
pred. TIDAK	0	0	0.00%
Class Recall	100.00%	0.00%	

Pada penelitian ini *accuracy* sebesar 89,33 % termasuk dalam klasifikasi yang bagus [3], adapun kriteria menurut [3] adalah sebagai berikut:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

## V. PENUTUP

Hasil penelitian dapat disimpulkan bahwa telah diperoleh klasifikasi mahasiswa yang terregistrasi pada program studi Informatika Universitas PGRI Semarang. Variabel yang paling tinggi pengaruhnya terhadap hasil registrasi mahasiswa adalah asal sekolah dan jenis kelamin. Interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi Prodi Informatika dalam melakukan strategi promosi lebih banyak ke sekolah umum. Seorang mahasiswa dari SMU jika berasal dari jurusan IPA semasa sekolah menengah atas memiliki probabilitas terregistrasi ke prodi Informatika lebih tinggi dibandingkan dengan mahasiswa dari jurusan lainnya. Rata-rata berasal dari Semarang dengan jurusan SMU dari IPA dan yang berasal dari luar kota rata-rata berasal dari Batang dan Pati. Dari SMU jurusan IPS dan berjenis kelamin Laki-laki berasal dari Batang dan yang berjenis kelamin Perempuan berasal dari Pati. *Accuracy* dari pembentukan model ini adalah sebesar 89.33% termasuk dalam klasifikasi yang bagus.

Berdasarkan hasil penelitian penerapan algoritma C.45 dapat membantu panitia PMB melakukan ekspansi dalam promosi. Dari segi algoritma untuk penelitian yang akan datang perlu menambah variabel penelitian dengan melakukan komparasi antara metode klasifikasi C.45 dengan metode klasifikasi yang lainnya dan membandingkan hasil klasifikasi dari bermacam-macam model sehingga bisa mengambil model mana yang paling sesuai dengan karakteristik data yang kita punya.

## DAFTAR PUSTAKA

- [1] Arief Jananto. 2010. *Memprediksi kinerja mahasiswa menggunakan teknik data mining (studi kasus data akademik mahasiswa Unisbank)*. Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada.
- [2] Efraim T. 2005. *Decision support systems and intelligent systems*. Edisi ke-7. Yogyakarta: Penerbit Andi
- [3] Florin Gorunescu, *Data Mining: Concepts, Models and Techniques*, Springer, 2011
- [4] Han, J. Kamber, M. 2001. *Data Mining: Concepts and Technique*. San Fransisco: Morgan Kaufman Publisher.
- [5] Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M., 2012, *Analytics: The Widening Divide*. MIT Sloan Management Review, 53(2), 1-22.
- [6] Kusriani. 2009. *Algoritma Data Mining*. Yogyakarta, Andi Publisher
- [7] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- [8] Lesmana, Dody Putu, 2012, Perbandingan Kinerja Decision Tree J48 dan ID3 Dalam Pengklasifikasian Diagnosis Penyakit Diabetes Mellitus, *Jurnal Teknologi dan Informatika*, Vol. 2, no. 2.
- [9] Luan, J., 2002, *Data Mining and Knowledge Management in Higher Education Applications*, Paper presented at the Annual Forum for the Association for Institutional Research, Toronto, Ontario, Canada. <http://eric.ed.gov/ERICWebPortal/detail?accno=ED474143>
- [10] MacLeman, J dan Tang, ZH. 2005. *Data Mining with SQL Server 2005*, United States of America: Wiley Publishing.

- [11] Hermawati, F. A. 2013. Data Mining. Yogyakarta: Penerbit Andi
- [12] Wibowo, Bagus Ari. 2011. Perancangan dan Implementasi Sistem Pendukung Keputusan untuk Jalan Menggunakan Metode ID3 (Studi Kasus BAPPEDA Kota Salatiga). Universitas Kristen Satya Wacana: Jawa Tengah.