

IMPLEMENTASI *ITERATIVE DICHOTOMISER 3* PADA DATA KELULUSAN MAHASISWA S1 DI UNIVERSITAS SEBELAS MARET

Thanh Thi Bi Dan
Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
dantina16@yahoo.com

Sari Widya Sihwi
Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
sari.widya.sihwi@gmail.com

Rini Anggrainingsih
Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
rinianggra@gmail.com

ABSTRACT

University of Sebelas Maret (UNS) selects new students every year. The selection process usually appears in three ways, known as National Selection for University Entry (SNMPTN), Local Selection for University Entry (SBMPTN), and New Students Admission Selection (SPMB). Every year UNS also passes students through four periods of graduation. From this process, UNS has graduation data that increases every period. These data has not yet been used effectively and left to accumulate unorderedly. Necessary technique is needed to process the data that can give standardized and satisfying results, thus by using data mining indeed can solve the problem.

In this research, the algorithm used is Iterative Dichotomiser 3 (ID3). ID3 algorithm is a method of learning that will build a decision tree that becomes model in finding solution of the problem. This algorithm in a greedily procedure searches every possibility of the decision tree. ID3 can be used to process graduation data of UNS bachelor's degree students that can give an significant information categorized in a 12 class structure bases on grade point average and graduation time. This group designed based on attribute gender, faculty, high school origin, enrollment process, parents' salary, and parents' job and this information can be used to determine new students acceptance in University.

From the tree model formed, then it tested by using measuring precision and recall. This testing were done by using data testing. Form the overall data 1066, used 95% for data training. Data testing used 5% from the overall data that is tested ten times randomly. From the testing result, it can be concluded that the implementation of ID3 algorithm on graduation data of UNS bachelor's degree students has average maximum precision 63.96% and average maximum recall 62.47% of data testing in a ten times testing.

Keyword : Classification, Data mining, Decision tree, Iterative Dichotomiser 3.

ABSTRAK

Setiap tahun Universitas Sebelas Maret (UNS) melakukan seleksi penerimaan mahasiswa baru. Proses seleksi tersebut biasanya melalui tiga jalur yaitu Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), Seleksi Penerimaan Mahasiswa Baru (SPMB). Setiap tahun juga UNS meluluskan mahasiswa melalui empat

kali periode wisuda. Dari proses tersebut UNS memiliki data wisuda yang setiap periode pasti bertambah. Data tersebut belum dapat dimanfaatkan secara maksimal, dan data dibiarkan menumpuk begitu saja. Diperlukan teknik untuk mengolah data tersebut agar memberi nilai tambah, yaitu salah satunya dengan memanfaatkan data mining agar mengatasi masalah tersebut.

Pada penelitian ini, algoritma yang digunakan adalah Iterative Dichotomiser 3 (ID3). Algoritma ID3 merupakan suatu metode yang akan membangun sebuah pohon keputusan yang merupakan suatu pemodelan dalam mencari solusi dari persoalan. Algoritma ini melakukan pencarian secara rakus/menyeluruh (greedy) pada semua kemungkinan pohon keputusan. Algoritma ID3 akan digunakan untuk mengolah data kelulusan mahasiswa Strata Satu (S1) UNS sehingga dapat memberikan sebuah informasi penting yang telah ditentukan yaitu berupa 12 kelas berdasarkan IPK dan lama studi. Pengelompokan ini berdasarkan atribut jenis kelamin, fakultas, asal SMA, jalur masuk, gaji orang tua, pekerjaan orang tua dan informasi ini dapat dimanfaatkan UNS untuk pengambilan keputusan penerimaan mahasiswa baru.

Dari model tree yang dibentuk maka dilakukan pengujian data dengan menggunakan pengukuran precision dan recall. Pengujian ini dilakukan menggunakan data testing. Dari data keseluruhan 1066 data, digunakan 95% untuk data training. Data testing yang digunakan 5% dari data keseluruhan yang diuji sebanyak sepuluh kali secara random. Dari hasil pengujian, dapat disimpulkan bahwa implementasi algoritma ID3 pada data kelulusan mahasiswa S1 UNS memiliki rata-rata maksimal precision 63.96% dan rata-rata maksimal recall 62.47% dari sepuluh kali pengujian saat diuji dengan data testing.

Kata kunci : Data mining, Decision tree, Iterative Dichotomiser 3, Klasifikasi.

1. PENDAHULUAN

Untuk menerima mahasiswa, Universitas Sebelas Maret (UNS) melakukan proses penerimaan mahasiswa baru melalui tiga jalur masuk yaitu Seleksi Nasional Masuk Perguruan Tinggi (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), Seleksi Penerimaan Mahasiswa Baru (SPMB) [1]. Dalam satu tahun, UNS mengadakan empat periode wisuda untuk meluluskan mahasiswa. Setiap periode wisuda, data

kelulusan mahasiswa selalu bertambah, semakin banyak dan menumpuk. Pertumbuhan yang pesat dari akumulasi data kelulusan mahasiswa ternyata menciptakan kondisi yang sering disebut sebagai “*Rich of Data but Poor of Information*” [2] karena data yang terkumpul itu belum digunakan untuk pengambilan keputusan. Kumpulan data itu dibiarkan begitu saja seakan- akan menjadi sebuah “kuburan data (*data tombs*)” [2]. Sebenarnya dari data yang bertumpuk tersebut dapat diperoleh sebuah informasi yang dapat dimanfaatkan oleh UNS untuk mendukung keputusan atau membantu dalam menentukan strategi pendidikan yang lebih baik.

Salah satu teknik yang digunakan untuk mengolah data tersebut agar memberi nilai tambah adalah dengan menggunakan *data mining*. *Data mining* adalah suatu proses penemuan sebuah informasi baru dalam sekumpulan data yang sangat besar. Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga. Berdasarkan fungsinya, *data mining* memiliki beberapa teknik yang dapat digunakan yaitu deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, asosiasi.

Pada kasus data wisuda, dapat digunakan teknik klasifikasi karena teknik ini mampu membagi data yang sangat besar ke dalam kelas-kelas yang berbeda, misalnya klasifikasi kelulusan mahasiswa sehingga memudahkan bagi pihak UNS dalam pengambilan keputusan. Klasifikasi merupakan suatu proses yang mengklasifikasikan data ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan [3]. Salah satu metode dalam klasifikasi *data mining* adalah *decision tree* (pohon keputusan). *Decision tree* merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami [4]. *Decision tree* merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root*.

Decision tree mempunyai banyak algoritma salah satunya adalah *Iterative Dichotomiser 3* (ID3). Penelitian yang pernah dilakukan menunjukkan bahwa algoritma ID3 merupakan algoritma yang terbaik dan akurat dibandingkan algoritma klasifikasi yang lain [5]. Algoritma ID3 melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan (*decision tree*) karena algoritma ini berusaha membangun *decision tree* secara *top-down* (dari atas ke bawah), dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistik (yang banyak digunakan adalah *information gain*) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data.

Pada penelitian ini akan dilakukan *data mining* untuk membantu dan memudahkan UNS pada pengambilan keputusan dalam penerimaan mahasiswa baru dengan cara mengklasifikasikan data induk mahasiswa dan data kelulusan mahasiswa, menggunakan metode *decision tree* algoritma ID3. Tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengimplementasi algoritma ID3 pada pengolahan data kelulusan mahasiswa S1 UNS untuk pengambilan keputusan penerimaan mahasiswa baru.

2. DASAR TEORI

2.1 *Data Mining*

2.1.1 Pengertian *Data Mining*

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.[6]

2.1.2 Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [6] :

1. Deskripsi: digunakan untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi: model menggunakan *record* lengkap yang menyediakan nilai dari variabel *target* sebagai nilai prediksi.
3. Prediksi: prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi: contoh kasusnya adalah penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. Pengklusteran: merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi: Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.1.3 Tahap-tahap *Data Mining*

Berikut adalah tahapan dalam *data mining* [7] :

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional. Data hasil seleksi akan digunakan untuk proses *data mining*, dan disimpan dalam suatu berkas dan terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Proses *cleaning* dilakukan dengan menghilangkan *noise*, membuang duplikasi data, memeriksa data yang tidak konsisten, dan

memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. Transformation Data

Tahap ini merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

4. Data mining

Data mining merupakan proses mencari pola atau informasi yang sangat menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. Interpretation/Evaluation

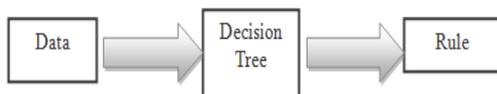
Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2 Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia.

2.3 Decision Tree (Pohon Keputusan)

Decision tree merupakan metode yang mengubah data menjadi pohon keputusan (decision tree) dan aturan-aturan keputusan (rule) yang dijelaskan pada Gambar 2.2 [8] :



Gambar 2.2 Konsep Decision Tree [8]

Pada decision tree terdapat 3 jenis node, yaitu:

- a. Root node: merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
- b. Internal node: merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.
- c. Leaf node atau terminal node: merupakan node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.

2.4 Algoritma ID3

Algoritma ID3 merupakan algoritma decision tree learning (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara rakus/menyeluruh (greedy) pada semua kemungkinan pohon keputusan [9].

Untuk membangun decision tree, algoritma ini mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistik (yang banyak digunakan adalah information gain), untuk mendapatkan nilai information gain yang atributnya akan dipilih sebagai akar maka nilai entropy dihitung terlebih dahulu. Berikut adalah penjelasan mengenai nilai Entropy dan Information Gain.

2.4.1 Entropy

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity dan homogeneity dari kumpulan data..

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

Dimana:

S: himpunan kasus

n: jumlah partisi S

p_i : proporsi dari S_i terhadap S

Dari nilai entropy tersebut kemudian dihitung nilai information gain (IG) masing-masing atribut.

2.4.2 Information Gain

Setelah mendapat nilai entropy untuk suatu kumpulan data, maka kita dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut information gain. Secara matematis, information gain dari suatu atribut A, dituliskan sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{2}$$

Dimana:

S : himpunan kasus

A: atribut

i : menyatakan suatu nilai yang mungkin untuk atribut A

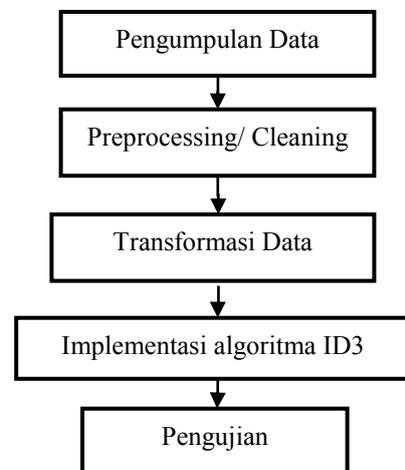
| S_i |: jumlah kasus pada partisi ke-i

|S|: jumlah kasus dalam S

Entropy (S_i): entropy untuk sampel-sampel yang memiliki nilai i

3. METODOLOGI PENELITIAN

Penelitian akan dilaksanakan berdasarkan tahap penelitian yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Proses Tahap Penelitian

3.1 Pengumpulan Data

Dalam tahapan ini, terdapat dua teknik yang digunakan:

3.1.1 Studi literatur

Metode ini dilaksanakan dengan melakukan studi kepustakaan untuk mendapatkan data literatur tambahan melalui membaca buku-buku, jurnal, karya ilmiah maupun artikel-artikel yang terkait dengan *data mining*, teknik klasifikasi, *decision tree* dan algoritma ID3.

3.1.2 Pengumpulan Data

Penelitian ini menggunakan data sekunder berupa data kelulusan mahasiswa S1 UNS dari tahun 2005 sampai tahun 2013 yang terdiri dari 19.942 data. Atribut yang digunakan dalam data kelulusan mahasiswa ini berdasarkan kemungkinan bahwa atribut tersebut memiliki pola yang berpengaruh besar terhadap kelulusan mahasiswa, atribut ini meliputi jenis kelamin, fakultas, asal SMA, jalur masuk, gaji orang tua, pekerjaan orang tua, lama studi dan IPK (Indek Prestasi Kumulatif).

3.2 Preprocessing/ Cleaning

Proses menghapus atribut yang tidak terpakai dan isian yang kosong sehingga data yang akan diolah benar-benar relevan dengan yang dibutuhkan.

3.3 Transformasi Data

Proses tranformasi data yang merupakan proses menggabungkan data yang diperoleh dan mengubahnya menjadi satu format yang sama untuk mempermudah proses *mining*.

3.4 Implementasi algoritma ID3

Secara umum langkah algoritma ID3 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari masing-masing atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti dalam persamaan (1) . Sementara itu, perhitungan nilai *entropy* dapat dilihat pada persamaan (2).

3.5 Pengujian

Setelah melakukan proses *mining* data maka akan dihasilkan model pohon keputusan. Model yang telah terbentuk itu akan diuji coba dengan data *testing* yang telah ditentukan. Pengujian ini menggunakan data *testing* dengan komposisinya 95% data *training* dan 5% data *testing* lalu diuji sebanyak sepuluh kali dengan data yang diambil secara *random*. Randomisasi dilakukan secara proporsional yaitu jumlah *records* data dari setiap kelas akan diacak 5% data *testing*, yang sisanya dijadikan data *training*. Pengujian ini menggunakan

pengukuran *precision* dan *recall*. Penjelasan tentang pengukuran *precision* dan *recall* dapat di lihat pada Tabel 3.1 dan perhitungan di berikut:

Tabel 3.1 Perhitungan Precision dan Recall

Actual Class	Predicted Class		
		Yes	No
	Yes	a (TP)	b (FN)
No	c (FP)	d (TN)	

Keterangan:

- a: TP (true positive)
- b: FN (false negative)
- c: FP (false positive)
- d: TN (true negative)

Precision dihitung dengan rumus:

$$Precision (p) = \frac{a}{a + c} \quad (3)$$

Recall dapat dihitung dengan rumus:

$$Recall(r) = \frac{a}{a + b} \quad (4)$$

4. PEMBAHASAN

Pada implementasi algoritma ID3 ini melalui beberapa tahapan. Dimulai dari Deskripsi Data, Transformasi Data, Implementasi Algoritma ID3, dan Pengujian.

4.1 Deskripsi Data

Penelitian ini menggunakan data sekunder berupa data kelulusan mahasiswa S1 UNS dari tahun 2005 sampai tahun 2013 dengan jumlah data 19.942 mahasiswa. Setelah melakukan proses pembersihan data, memperoleh sejumlah 1066 data yang terdiri dari berbagai atribut yaitu lama studi, IPK, jenis kelamin, jalur masuk, fakultas, asal SMA, pekerjaan orang tua, gaji orang tua.

4.2 Transformasi Data

Setelah proses *cleaning* data, dilakukan proses transformasi data, sehingga data tersebut sesuai untuk proses *data mining*. Tabel tranformasi data atribut prediktor yang disajikan pada Tabel 4.1

No	Atribut Prediktor	Kategori	Alias
1.	Jenis Kelamin	Laki-laki	1
		Perempuan	2
2.	Fakultas	Kedokteran	1
		Mipa	2
		Ekonomi	3
		Hukum	4
		Pertanian	5
		FSSR	6
		FISIP	7
		FKIP	8
		Teknik	9

Tabel 4.1 Transformasi Data Atribut Prediktor

Tabel 4.1 Transformasi Data Atribut Predikator (Lanjutan)

No	Atribut Predikator	Kategori	Alias
3.	Asal SMA	SKA (Surakarta)	1
		Solo raya	2
		Jawa tengah	3
		Jawa	4
		Luar jawa	5
		Luar negeri	6
4.	Jalur masuk	SPMB	1
		Swadana	2
		PMDK	3
		Transfer	4
		PBUS	5
		Tugas Belajar	6
5.	Gaji orang tua	kurang dari Rp. 500.000	1
		Rp. 500.000 - Rp. 1.000.000	2
		Rp. 1.000.000 - Rp. 2.500.000	3
		Rp. 2.500.000 - Rp. 5.000.000	4
		Rp. 5.000.000 - Rp. 7.500.000	5
		Rp. 7.500.000 - Rp. 10.000.000	6
		Lebih dari Rp. 10.000.000	7
6.	Pekerjaan orang tua	Pensiunan Pegawai Negeri/ABRI	1
		Pensiun Pegawai Swasta	2
		Guru/Dosen Negeri	3
		Guru/Dosen Swasta	4
		Pegawai Swasta bukan Guru/Dosen	5
		Pegawai Negeri bukan Guru/Dosen	6
		Buruh/orang yang bekerja dengan tenaga fisik saja	7
		Ahli profesional yang hanya bekerja secara perorangan	8
		Abri	9
		Pedagang/Wiraswasta	10
		Petani/Nelayan	11
		Lain-lain	12

Atribut Respon	Kategori	Interval Kelas
IPK dan Lama Studi	Lama studi 3,5 – 4 tahun dan IPK 3,51 – 4,00	A1
	Lama studi 3,5 – 4 tahun dan IPK 3,01 – 3,50	A2
	Lama studi 3,5 – 4 tahun dan IPK 2,51 – 3,00	A3
	Lama studi 3,5 – 4 tahun dan IPK 2,00 – 2,50	A4
	Lama studi >4 – 5 tahun dan IPK 3,51 – 4,00	B1

Tabel tranformasi data atribut respon yang disajikan pada Tabel 4.2

Tabel 4.2 Transformasi Data Atribut Respon (Lanjutan)

Atribut Respon	Kategori	Interval Kelas
IPK dan Lama Studi	Lama studi >4 – 5 tahun dan IPK 3,01 – 3,50	B2
	Lama studi >4 – 5 tahun dan IPK 2,51 – 3,00	B3
	Lama studi >4 – 5 tahun dan IPK 2,00 – 2,50	B4
	Lama studi > 5 tahun dan IPK 3,51 – 4,00	C1
	Lama studi > 5 tahun dan IPK 3,01 – 3,50	C2
	Lama studi > 5 tahun dan IPK 2,51 – 3,00	C3
	Lama studi > 5 tahun dan IPK 2,00 – 2,50	C4

4.3 Implementasi algoritma ID3

Untuk menggambarkan pembentukan *decision tree* dengan menggunakan algoritma ID3, berikut ini adalah langkah-langkahnya:

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

Berikut ini diberikan contoh sebagai kasus menggunakan lima data kelulusan mahasiswa S1 UNS yang disajikan pada Tabel 4.3

Tabel 4.3 Contoh Data Kelulusan Mahasiswa S1

No	1	2	3	4	5
JK	P	P	L	P	L
Fakultas	Kedokt-eran	Kedokt-eran	FISIP	FSSR	Ekonomi
Asal SMA	Ska	Ska	Jawa tengah	Ska	Ska
Jalur masuk	Swadana	spmb	spmb	spmb	spmb
Gaji ortu	1.000.000 0 2.500.000	1.000.000 0 2.500.000	1.000.000 0 2.500.000	1.000.000 0 2.500.000	Kurang dari Rp.500.000
Pekerjaan ortu	Guru/Dosen Negeri	Pegawai Negeri Bukan Guru/Dosen	Buruh	Guru/Dosen Negeri	Lain-lain
IPK	3.51	3.01	3.51	3.62	3.02
Lama studi	3.5	4	4.5	5.5	5.5

Node	Atribut	Nilai atribut	Jml kasus						Entropy	Gain
				A 1	A 2	B 1	C 1	C 2		
1	Total		5	1	1	1	1	1	2.32	
	A									0.97
		1	2	0	0	1	0	1	1	
		2	3	1	1	0	1	0	1.58496	
	B									1.92
		1	2	1	1	0	0	0	1	
		2	1	0	0	0	0	1	0	
		6	1	0	0	0	1	0	0	
		7	1	0	0	1	0	0	0	
	C									0.72
		1	4	1	1	0	1	1	2	
		3	1	0	0	1	0	0	0	
	D									0.72
		1	4	0	1	1	1	1	2	
		2	1	1	0	0	0	0	0	
	E									0.72
		1	1	0	0	0	0	1	0	
		3	4	1	1	1	1	0	2	
	F									1.92
		3	2	1	0	0	1	0	1	
		6	1	0	0	1	0	0	0	
		7	1	0	1	0	0	0	0	
		12	1	0	0	0	0	1	0	

Kemudian data-data pada Tabel 4.3 dilakukan transformasi sesuai dengan Tabel 4.1 dan 4.2, dan data hasil transformasi ditunjukkan pada Tabel 4.4

Tabel 4.4 Hasil Data Transformasi

A	B	C	D	E	F	G
2	1	1	2	3	3	A1
2	1	1	1	3	6	A2
1	7	3	1	3	7	B1
2	6	1	1	3	3	C1
1	2	1	1	1	12	C2

- Keterangan :
- A: Jenis Kelamin
 - B: Fakultas
 - C: Asal SMA
 - D: Jalur Masuk
 - E: Gaji orang tua
 - F: Pekerjaan orang tua
 - G: Kelas (Target)

Dari data-data pada Tabel 4.4 akan dibangun sebuah *classifier* berdasarkan atribut jenis kelamin, fakultas, asal SMA, jalur masuk, gaji orang tua, dan pekerjaan orang tua. Ada lima kelas yaitu A1, A2, B1, C1, C2.

Berikut ini penjelasan mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma ID3 untuk menyelesaikan permasalahan pada Tabel 4.3

- a. Menghitung jumlah kasus, jumlah kasus untuk kelas A1, kelas A2, kelas B1, kelas C1, kelas C2, dan *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut A, B, C, D, E, F. Setelah itu, lakukan perhitungan *gain* untuk

tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh Tabel 4.5

Tabel 4.5 Perhitungan Node 1

Baris **TOTAL** kolom *entropy* pada Tabel 4.5 dihitung dengan persamaan (2) sebagai berikut:

$$Entropy(Total) = (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5}))$$

$$Entropy(Total) = 2.32$$

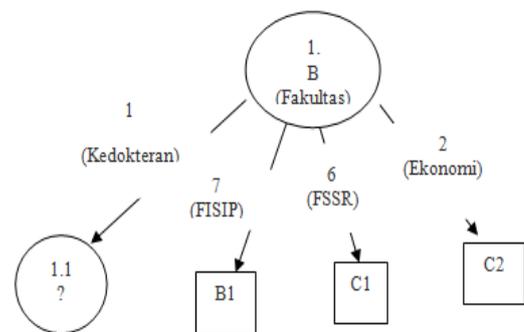
Sementara itu, nilai *gain* pada baris A dihitung dengan menggunakan persamaan (1) sebagai berikut:

$$Gain(Total, A) = Entropy(Total) \sum_{i=1}^n \frac{|A_i|}{|Total|} * Entropy(A_i)$$

$$Gain(Total, A) = 2.32 - ((\frac{2}{5} * 1) + (\frac{3}{5} * 1.58496)) = 0.97$$

Untuk nilai *gain* atribut yang lain juga dihitung seperti nilai *gain* atribut A.

Dari hasil pada Tabel 4.5 dapat diketahui bahwa ada dua atribut memiliki *gain* tertinggi yaitu B (Fakultas) dan F (Pekerjaan ortu) yaitu sebesar 1.92. Pada kasus ini, memilih atribut B untuk menjadi *node* akar. Ada empat nilai atribut dari B, yaitu 1, 2, 6, 7. Dari keempat nilai atribut tersebut, nilai atribut 2, 6, 7 sudah mengklasifikasi kasus menjadi kelas C2, C1, dan B1, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut 1 masih perlu dilakukan perhitungan lagi. Dari hasil tersebut digambarkan pohon keputusan pada *node* 1 tampak seperti Gambar 4.1



Gambar 4.1 Pohon Keputusan Hasil Perhitungan Node 1

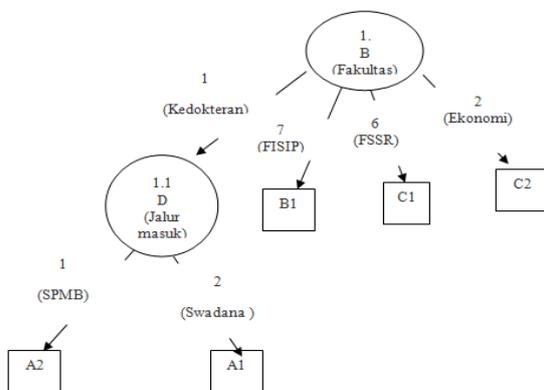
- b. Menghitung jumlah kasus, jumlah kasus untuk kelas A1, kelas A2, kelas B1, kelas C1, kelas C2, dan *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut A, C, D, E, F yang dapat dijadikan sebagai *node* akar dari nilai atribut Setelah itu, dilakukan perhitungan *gain*

untuk tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh Tabel 4.6

Node	Atribut	Nilai atribut	Jml Kasus	A1	A2	Entropy	Gain	
1.1	A		2	1	1	1.00	0	
		1	0	0	0	0		
		2	2	1	1	1		
	C							0
		1	2	1	1	1		
		3	0	0	0	0		
	D							1
		1	1	0	1	0		
		2	1	1	0	0		
	E							1
		1	0	0	0	0		
		3	2	1	1	1		
F							1	
	3	1	1	0	0			
	7	0	0	0	0			
	6	1	0	1	0			
	12	0	0	0	0			

Tabel 4.6 Perhitungan Node 1.1 untuk B-1

Dari hasil pada Tabel 4.6 dapat diketahui bahwa ada tiga atribut dengan gain tertinggi adalah D (Jalur masuk), E (Gaji ortu) dan F (Pekerjaan ortu) yaitu sebesar 1.00. Sehingga dipilih salah satu atribut tersebut untuk menjadi node akar dan dikasus ini memilih D menjadi node cabang dari nilai atribut 1 (Kedokteran). Ada dua nilai atribut dari D, yaitu 1 dan 2. Dari kedua nilai atribut tersebut, nilai atribut 1 dan 2 sudah mengklasifikasikan kasus menjadi kelas A2 dan A1, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Pohon keputusan yang terbentuk dari kasus ini disajikan pada Gambar 4.3

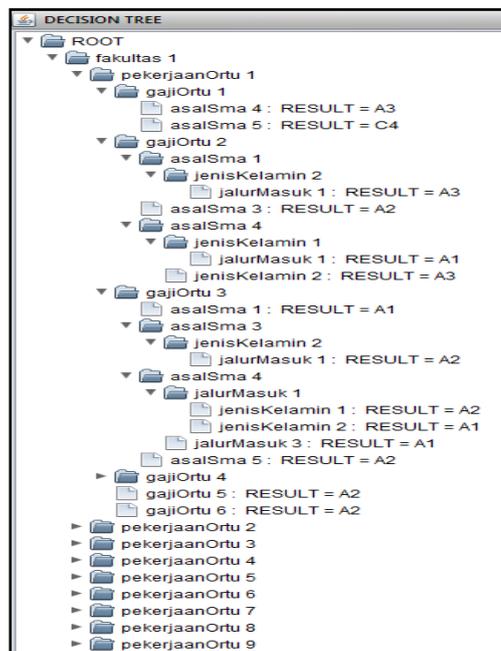


Gambar 4.2 Pohon Keputusan Hasil Perhitungan Node 1.1

Gambar 4.2 menunjukkan hasil akhir dari proses pembentukan tree. Setelah selesai pembentukan tree, selanjutnya dilakukan proses pembuatan aturan berdasarkan tree yang terbentuk. Dan aturan dari tree di atas adalah sebagai berikut:

- Jika fakultas Kedokteran , lalu masuk SPMB maka lama studi 3,5 – 4 tahun dan IPK 3,01 – 3,50.
- Jika fakultas Kedokteran, jalur masuk Swadana maka lama studi 3,5 – 4 tahun dan IPK 3,51 – 4,00.
- Jika fakultas Ekonomi maka lama studi > 5 tahun dan IPK 3,01 – 3,50.
- Jika fakultas FSSR maka lama studi > 5 tahun dan IPK 3,51 – 4,00.
- Jika fakultas FISIP maka lama studi > 4 – 5 tahun dan IPK 3,51 – 4,00.

Dari langkah-langkah pada tahap ini, akan diimplementasikannya ke dalam sistem sehingga menghasilkan model pohon keputusan. Implementasi ini menggunakan bahasa pemrograman Java jdk 6 dengan editor Netbean 7.1, database MySql. Gambar 4.3 tampilan potongan tree yang terbentuk dari sistem setelah menjalankan proses mining.



Gambar 4.3 Potongan Pohon Keputusan

4.4 Pengujian Model

Model pohon keputusan yang telah terbentuk diuji dengan menggunakan data testing. Pengujian model dilakukan dengan membandingkan hasil model pohon keputusan yang terbentuk dengan data yang diuji. Jika salah satu atau seluruh nilai atribut dan kelas diperoleh sama, maka hasil klasifikasi dinyatakan sesuai. Sebaliknya, jika salah satu atau semua nilai atribut sama tetapi kelas berbeda maka hasil klasifikasi dinyatakan tidak sesuai. Dari seluruh pengujian yang telah dilakukan menunjukkan bahwa hasil yang didapatkan tidak memiliki perbedaan yang signifikan. Berikut ini contoh hasil pengujian dengan menggunakan data testing yang diuji sebanyak sepuluh kali yang disajikan pada Tabel 4.7.

Tabel 4.7 Hasil Pengujian Menggunakan Data Testing yang Diuji Sebanyak Sepuluh Kali

Percobaan ke-	Jumlah data	Sesuai	Tidak sesuai	Maksimal Precision	Maksimal Recall
1	54	12	32	50.0	53.9
2	54	16	27	87.5	58.3
3	54	16	26	63.6	58.3
4	54	14	34	50.0	50.0
5	54	19	22	63.6	63.6
6	54	14	27	78.6	68.75
7	54	16	24	56.25	69.2
8	54	17	25	60.0	64.3
9	54	15	31	50.0	58.3
10	54	16	23	80.0	80.0
Hasil rata-rata				63.96 %	62.47%

Dari Tabel 4.8 diperoleh hasil pengujian menggunakan data *testing* yang telah diuji sebanyak sepuluh kali. Nilai maksimal *precision* dan maksimal *recall* yang diperoleh dari pengujian pertama sampai pengujian kesepuluh terdapat dalam kelas A2 dan B3, karena pada jumlah data *testing* sebanyak 54 data yang berasal dari kelas A2 dan B3 memiliki jumlah *records* yang paling tinggi dibandingkan kelas lain yaitu A2 yang memiliki 17 *records* dan B2 yang memiliki 16 *records*. Sehingga, nilai maksimal *precision* dan maksimal *recall* yang tertinggi ada di kelas A2 dan B2. Hasil rata-rata maksimal *precision* memiliki 63.96% dan rata-rata maksimal *recall* 62.47%.

5. KESIMPULAN

Berdasarkan hasil eksperimen yang telah dilakukan pada penelitian ini, maka penulis menarik kesimpulan, antara lain: [1] Model pohon keputusan dengan algoritma ID3 dapat digunakan untuk memberikan informasi kepada UNS khususnya terkait pengambilan keputusan pada saat penerimaan mahasiswa baru. [2] Penentuan data *training* sangat menentukan tingkat akurasi *tree* yang dibuat. Besar persentase kebenaran *tree* sangat dipengaruhi oleh data *training* yang digunakan untuk membangun model *tree* tersebut. [3] Hasil pengujian data *testing* yang diuji menggunakan algoritma ID3 memiliki nilai rata-rata maksimal *precision* 63.96% dan rata-rata maksimal *recall* 62.47% dari sepuluh kali pengujian dengan data yang diperoleh secara *random*.

Penulis menyarankan sebaiknya penelitian selanjutnya dapat menggunakan algoritma lain atau menggabungkan algoritma ID3 dengan algoritma lain, kemudian hasil pengujian yang diperoleh dibandingkan demi mendapatkan hasil penelitian yang lebih baik.

6. DAFTAR PUSTAKA

- [1] Sistem Informasi SPMB. Online Universitas Sebelas Maret. Akses pada tanggal 1 Juni, 2014. <http://www.spmb.uns.ac.id>
- [2] Pramudiono, I. (2003). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. Ilmu Komputer.com.

- [3] Hamidah, I. (2012). *Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5*. Bandung.
- [4] Kusriani & Luthfi, E. (2009). *Algoritma Data Mining*. Andi: Yogyakarta.
- [5] Wawan & Hendayun, M. (2013). *Kinerja Algoritma Data Mining Decision Tree (ID3, C4.5, C5.0), Naïve Bayes, SVM dan kNN untuk Klasifikasi Email Spam dan Non-Spam*. Bandung.
- [6] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data mining*. John Wiley and Sons, Inc.
- [7] Fayyad, U. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [8] Basuki, A. & Syarif, I. (2003). *Decision Tree*. Surabaya.
- [9] Wahyudin. (2009). Metode Iterative Dichotomizer 3 (ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru. *Jurnal Pendidikan Teknologi Informasi Dan Komunikasi (Ptik)*, Vol.2, No.2.