

IMPLEMENTASI ALGORITMA *ITERATIVE DICHOTOMISER 3* PADA PENYELEKSIAN PROGRAM MAHASISWA WIRAUSAHA UNS

Tisna Dedi Utama
Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
thama.vix@gmail.com

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

Afrizal Doewes
Jurusan Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
afrizal.doewes@gmail.com

ABSTRACT

Student Entrepreneur Program (PMW) is a program of Sebelas Maret University (UNS) in an effort to train students in the field of entrepreneurship. This program has a lot of interest which are required to make a start up bussiness. After five years of running the program, the accumulation of bussiness plan participants. This data maybe used to pile to be extracted and processed into useful information using classification. This classification process is expected to be used to consider the selection process. It is intended that the provision of venture capital funds can be targeted and there was warning or anticipation to students who could potentially fail or go without after getting capital funds.

The process classification of data later in the mining heap using iterative dechotomiser 3 (ID3) algorithm. The data in this mining process will form to the pattern that can be known linkages between participants attribute the success category. This pattern is expected to be used in the PMW for the selection process.

From the test results on the data PMW UNS, average accuracy percentage owned by 77.49% when tested with the 10% which do random testing of data as much as ten times and 96,33% for all training data.

keyword : *Data mining, Decision tree, Iterative Dichotomiser 3, Classification.*

ABSTRAK

Program Mahasiswa Wirausaha (PMW) merupakan program Universitas Sebelas Maret (UNS) dalam upaya melatih mahasiswa dalam bidang wirausaha. Program ini memiliki banyak peminat yang diharuskan membuat start up bussiness. Setelah lima tahun program berjalan, terjadi tumpukan data bussiness plan dari para peserta. Padahal tumpukan data ini dapat dimanfaatkan untuk digali dan diolah menjadi informasi yang berguna menggunakan klasifikasi. Proses klasifikasi ini diharapkan dapat digunakan untuk mempertimbangkan proses penyeleksian. Hal ini bertujuan agar pemberian dana modal usaha dapat tepat sasaran dan ada peringatan dini atauantisipasi kepada mahasiswa yang berpotensi gagal atau pergi tanpa kabar setelah mendapatkan dana modal.

Proses klasifikasi dari tumpukan data kemudian di mining dengan menggunakan algoritma iterative dechotomiser 3 (ID3). Data yang di mining ini akan membentuk pola sehingga dapat diketahui keterkaitan antara atribut dengan keberhasilan peserta. Pola

inilah yang diharapkan dapat digunakan pada PMW untuk mempertimbangkan proses penyeleksian.

Dari hasil pengujian pada data PMW UNS, persentase keakuratan rata-rata yang dimiliki sebesar 77,49% saat diuji dengan 10% data testing yang dilakukan random sebanyak sepuluh kali dan 96,33% pada seluruh data training.

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

1. PENDAHULUAN

Program Mahasiswa Wirausaha (PMW) diusung sejak tahun 2009 oleh Departemen Pendidikan Nasional (DIKNAS). Program ini bertujuan untuk menumbuhkembangkan jiwa wirausaha dan meningkatkan aktivitas kewirausahaan para calon lulusan perguruan tinggi [1]. UNS telah memulai PMW melalui Pusat Pengembangan Kewirausahaan (PPKwu). PPKwu merupakan pusat pengembangan dibawah Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM).

Lima tahun perekrutan PMW ini, terjadi penempukan data dari *business plan* dan kategori peserta, padahal dapat dimanfaatkan sebagai informasi yang strategis. Informasi ini dapat digunakan untuk mendukung keputusan atau membantu dalam menentukan kategori mahasiswa. Terbentuk empat kategori peserta dari program ini. Kategori pertama adalah diterima kemudian menjalankan usahanya. Kedua adalah diterima namun mengalami kebangkrutan. Ketiga diterima namun tidak menjalankan usahanya dan pergi tanpa kabar. Sedangkan kategori keempat adalah tidak diterima. Salah satu cara untuk menggali informasi strategis dari tumpukan data adalah dengan menggunakan *Data Mining*.

Data Mining merupakan suatu proses di dalam menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*). Dari proses tersebut ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui[2]. Analisa data mining berjalan pada data yang cenderung terus membesar dan teknik terbaik yang digunakan. Kemudian berorientasi kepada data berukuran sangat besar untuk mendapatkan kesimpulan dan keputusan paling layak. Dengan penggunaan *data mining*, diharapkan penumpukan data yang selama ini terjadi dapat dipergunakan untuk menggali informasi yang penting. Sehingga memberikan manfaat dan pengetahuan yang sebelumnya tersembunyi. Berdasarkan fungsinya *data mining* memiliki beberapa

teknik yang dapat digunakan yaitu deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, asosiasi [3].

Salah satu metode dalam klasifikasi *Data Mining* adalah *Decision Tree* (pohon keputusan). Klasifikasi tersebut membagi data ke kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Metode *Decision Tree* dapat juga dikatakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia [4]. Konsep dasar *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*), 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*.

Ada banyak algoritma dari *Decision Tree*, salah satu diantaranya adalah algoritma ID3. Menurut penelitian yang dilakukan oleh S.Archana & Dr. K.Elangovan dengan judul “*Survey of Classification Techniques in Data Mining*” [5], algoritma ID3 merupakan algoritma yang lebih akurat dibandingkan algoritma klasifikasi yang lain yaitu C4.5, C5.0, Naïve Bayes, SVM dan KNN. Algoritma ID3 menggunakan nilai atribut dalam klasifikasi tanpa menghilangkan *value* atribut tersebut. Algoritma ID3 melakukan pencarian secara menyeluruh pada semua kemungkinan 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. Penelitian lain yang dilakukan Sofi Defiyanti yang membahas tentang Perbandingan Kinerja Algoritma ID3 dan C.45 dalam Klasifikasi Spam-Mail [6]. Pengukuran kinerja kedua algoritma yang dilakukan berdasarkan pada jumlah atribut, yang menunjukkan algoritma ID3 memiliki kinerja (*precision, recall, dan accuracy*) yang lebih baik dibandingkan dengan algoritma C4.5.

Pada penelitian ini akan dilakukan *Data Mining* untuk memudahkan mengklasifikasi kategori peserta PMW. Pengklasifikasian kategori peserta diperoleh dari data penilaian *business plan* mahasiswa yang mendaftar PMW. Penelitian ini menggunakan metode *Decision Tree* algoritma ID3.

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.[3]

2.1.2 Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [3] :

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. *Cleaning Data*

Proses *cleaning* dilakukan dengan menghilangkan *noise*, membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. Transformasi 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 proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari obyek yang label kelasnya tidak diketahui.[8]

2.3 *Decision Tree* (Pohon Keputusan)

Decision tree merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabang-cabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan [2]. Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan.

Pohon keputusan dimulai dengan sebuah *root node* (titik awal) yang diapakai oleh user untuk

mengambil tindakan. Dari *node root* ini, user memecahkan sesuai dengan algoritma *Decision Tree*.

Hasil akhirnya adalah sebuah pohon keputusan dengan setiap cabangnya menunjukkan kemungkinan skenario dari keputusan yang diambil hasilnya.

2.4 Algoritma ID3

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

2.4.1 Entropy

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. *Entropy* dapat dirumuskan seperti berikut:

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

Dimana:

S: himpunan kasus

n: jumlah partisi S

p_i : proporsi dari S_i terhadap S

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, dirumuskan 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

V: menyatakan suatu nilai yang mungkin untuk atribut A

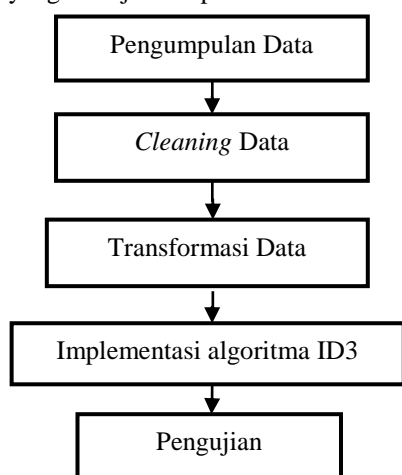
$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: 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, terdapa 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 peserta PMW di PPKWu UNS dari tahun 2009 sampai tahun 2013 dengan jumlah data 682 mahasiswa dengan limabelas atribut pada masing-masing mahasiswa. Atribut yang digunakan terdapat pada Tabel 3.1.

Tabel 3.1 Data Atribut

No	Jenis Atribut	Alias
1	keberadaan usaha (lama berjalannya usaha)	C1
2	pangsa pasar (potensi pasar untuk direalisasikan)	C2
3	daya saing (banyaknya persaingan usaha yang sejenis, Semakin banyak pesaing yang sudah eksis, maka peringkat rendah)	C3
4	prediksi keuntungan (prediksi keuntungan yang didapatkan tinggi maka peringkatnya juga tinggi)	C4
5	omzet yang didapat perbulan	C5
6	<i>Back Cash (BC) ratio</i>	C6
7	kebutuhan Sumber Daya Manusia(SDM) (kesesuaian SDM yang dibutuhkan untuk operasional usaha dengan omzet yang di dapat)	C7
8	kualifikasi SDM	C8
9	rasionalisme dana (perbandingan tingkat rasionalisme dana dengan jenis usaha yang diusulkan)	C9
10	keuletan	C10
11	integritas	C11
12	komitmen	C12
13	hasil dari psikotest, sebagai penilaian derajat keoptimisan	C13
14	jurusan mahasiswa peserta PMW	C14
15	lama mahasiswa kuliah di UNS	C15

3.2 Cleaning Data

Dalam kasus PMW ini, data yang dipergunakan adalah data yang telah bersih, sehingga data tidak perlu melalui proses *cleaning*.

3.3 Transformasi Data

Transformasi data merupakan proses pengubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam *data mining*.

Seringkali data yang akan digunakan dalam proses *data mining* mempunyai format yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya. Dalam penelitian ini, data yang ditransformasikan yaitu mengubah parameter psikotest, fakultas, lama study dan hasil kategori yang dapat dilihat pada tabel 3.2, tabel 3.3, tabel 3.4 dan tabel 3.5.

Tabel 3.2 Transformasi Data Atribut Psikotest

Interval Psikotest (C13)	Kategori
1	Nilai alpha<99= Rendah (Tidak Optimis)
2	Nilai 100<alpha<124 = Cukup (Optimis)
3	Nilai alpha>125= Tinggi (Sangat Optimis)

Tabel 3.3 Transformasi Data Atribut Fakultas

Fakultas (C14)	Kategori
F	Ekonomi
M	Mipa
H	Hukum
G	Kedokteran
I	Teknik
K	KIP
C	SSRS
D	FISIP
F	Pertanian

Tabel 3.4 Transformasi Data Atribut Lama Study

Interval lama study (C15)	Kategori
1	tahun ke 1
2	tahun ke 2
3	tahun ke 3
4	tahun ke 4
5	tahun ke 5

Tabel 3.5 Transformasi Data Hasil

Interval Hasil Akhir	Kategori
g1	diterima menjalankan
g2	diterima gagal (bangkrut)
g3	ditrima tidak ada kabar (pergi setelah menerima uang)
g4	tidak diterima

3.2 Implementasi algoritma ID3

Secara umum langkah algoritmaID3 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.

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

3.3 Pengujian

Pada tahap ini akan dilakukan pengujian data. Pengujian dilakukan dua kali,yaitu menggunakan data

testing dengan komposisi adalah 90% data training dan 10% data testing. Untuk pengujian data *testing*, dilakukan secara random dengan pengujian sebanyak sebanyak sepuluh kali, sedangkan pengujian data training, dilakukan pengujian dengan seluruh data training. Presentase keakurasian hasil output sistem dihitung menggunakan rumus sebagai berikut:

$$\text{Akurasi (\%)} = \frac{\text{jumlah data benar}}{\text{jumlah data yang diuji}} \times 100\% \quad (3)$$

Kemudian hasil akurasi dari pengujian ini adalah rata-rata akurasi setiap percobaan yang dapat digambarkan pada rumus berikut:

$$\text{Rata - rata akurasi (\%)} = \frac{\text{Total akurasi}}{\text{jumlah pengujian}} \quad (4)$$

4. PEMBAHASAN

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

4.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data seleksi PMW yang sudah dilaksanakan oleh PPKWu dari tahun 2009 sampai dengan tahun 2013. Keseluruhan data yang terkumpul sebanyak 682 data mahasiswa peserta PMW.

4.2 Cleaning Data

Dalam kasus PMW ini data yang digunakan sudah bersih, sehingga tidak memerlukan proses *cleaning* terhadap data yang terkumpul.

4.3 Transformasi Data

Data yang telah didapatkan kemudian melalui proses transformasi data. Hal ini bertujuan agar data menjadi satu format yang sama untuk mempermudah proses *mining*.

4.4 Implementasi algoritma ID3

Untuk menggambarkan pembentukan *decision tree* dengan menggunakan algoritma ID3 maka berikut ini diberikan contoh sebagai kasus menggunakan lima data PMW UNS.

Tabel 4.1 Contoh Data PMW UNS

No	Mahasiswa	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10
1	M1	5	5	4	4	4	4	4	4	5	5
2	M2	4	4	2	4	4	4	5	5	2	4
3	M3	2	1	2	4	4	4	4	2	2	5
4	M4	2	4	4	2	2	2	2	2	4	2
5	M5	1	4	2	4	1	2	2	2	4	2

Tabel 4.1 Contoh Data PMW UNS (Lanjutan)

C 11	C 12	C 13	C 14	C 15	Hasil
4	5	98	KIP	tahun ke 3	menjalankan usaha
5	4	113	Mipa	tahun ke 3	menjalankan usaha
4	4	118	Teknik	tahun ke 3	mengalami kebangkrutan
1	2	97	SSRS	tahun ke 3	tidak ada

					kabar
4	4	116	Mipa	tahun ke 4	tidak diterima

		2	0	0	0	0	0	0.000	
		1	1	0	1	0	0	0.000	

Kemudian data pada Tabel 4.1 dilakukan transformasi sesuai pada tabel 3.2, tabel 3.3, tabel 3.4 dan tabel 3.5. Hasil transformasi dat ditunjukkan pada tabel 4.2

Tabel 4.2 Hasil Transformasi Data

No	Nama Mahasiswa	C 1	C 2	C 3	C 4	C 5	C 6
1	M1	5	5	4	4	4	4
2	M2	4	4	2	4	4	4
3	M3	2	1	2	4	4	4
4	M4	2	4	4	2	2	2
5	M5	1	4	2	4	1	2

Tabel 4.2 Hasil Transformasi Data(Lanjutan)

C 7	C 8	C 9	C 10	C 11	C 12	C 13	C 14	C 15	Hasil
4	4	5	5	4	5	1	K	3	g1
5	5	2	4	5	4	2	M	3	g1
4	2	2	5	4	4	2	I	3	g2
2	2	4	2	1	2	1	C	3	g3
2	2	4	2	4	4	2	M	4	g4

Dari data-data pada tabel 4.1 akan dibangun sebuah *classifier* berdasarkan limabelas atribut dengan empat kelas yaitu g1, g2, g3, g4. Penjelasan mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma ID3 dapat dilihat melalui penjelasan berikut:

a. Menentukan *Node* Akar

Menghitung informasi *gain* dan *entropy* dari limabelas atribut data untuk menentukan *root*/akar awal dari tree. Hasil perhitungan ditunjukkan oleh Table 4.3.

Tabel 4.3 Perhitungan Node 1

Node		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1	Total	5	2	1	1	1	0.980	
	C1							0.580
		5	1	1	0	0	0.000	
		4	1	1	0	0	0.000	
		3	0	0	0	0	0.000	
		2	2	0	1	1	1.000	
		1	1	0	0	0	1.000	
	C2							0.030
		5	1	1	0	0	0.000	
		4	3	1	0	1	1.585	
		3	0	0	0	0	0.000	

Tabel 4.3 Perhitungan Node 1 (Lanjutan)

No de		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1	Total	5	2	1	1	1	0.980	
	C3							-0.370
		5	0	0	0	0	0.000	
		4	2	1	0	1	1.000	
		3	0	0	0	0	0.000	
		2	3	1	1	0	1.585	
		1	0	0	0	0	0.000	
	C4							-0.220
		5	0	0	0	0	0.000	
		4	4	2	1	0	1.500	
		3	0	0	0	0	0.000	
		2	1	0	0	1	0.000	
		1	0	0	0	0	0.000	
	C5							0.430
		5	0	0	0	0	0.000	
		4	3	2	1	0	0.918	
		3	0	0	0	0	0.000	
		2	1	0	0	1	0.000	
		1	1	0	0	0	1.000	
	C6							0.030
		5	0	0	0	0	0.000	
		4	3	2	1	0	0.918	
		3	0	0	0	0	0.000	
		2	2	0	0	1	1.000	
		1	0	0	0	0	0.000	
	C7							0.180
		5	1	1	0	0	0.000	
		4	2	1	1	0	1.000	
		3	0	0	0	0	0.000	
		2	2	0	0	1	1.000	
		1	0	0	0	0	0.000	
	C8							0.030
		5	1	1	0	0	0.000	
		4	1	1	0	0	0.000	
		3	0	0	0	0	0.000	
		2	3	0	1	1	1.585	
		1	0	0	0	0	0.000	
	C9							0.180
		5	1	1	0	0	0.000	

		4	2	0	0	1	1	1.000	
		3	0	0	0	0	0	0.000	
		2	2	1	1	0	0	1.000	
		1	0	0	0	0	0	0.000	

		2	0	0	0	0	0	0.000	
		1	0	0	0	0	0	0.000	

Tabel 4.3 Perhitungan Node 1 (Lanjutan)

Node		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1	Total	5	2	1	1	1	0.980	
	C10							0.180
		5	2	1	1	0	0	1.000
		4	1	1	0	0	0	0.000
		3	0	0	0	0	0	0.000
		2	2	0	0	1	1	1.000
		1	0	0	0	0	0	0.000
	C11							0.030
		5	1	1	0	0	0	0.000
		4	3	1	1	0	1	1.585
		3	0	0	0	0	0	0.000
		2	0	0	0	0	0	0.000
		1	1	0	0	1	0	0.000
	C12							0.030
		5	1	1	0	0	0	0.000
		4	3	1	1	0	1	1.585
		3	0	0	0	0	0	0.000
		2	1	0	0	1	0	0.000
		1	0	0	0	0	0	0.000
	C13							-0.37
		3	0	0	0	0	0	0
		2	3	1	1	0	1	1.585
		1	2	1	0	1	0	1
	C14							0.580
		F	0	0	0	0	0	0.000
		M	2	1	0	0	1	0.000
		E	0	0	0	0	0	0.000
		G	0	0	0	0	0	0.000
		I	1	0	1	0	0	0.000
		K	1	1	0	0	0	0.000
		C	1	0	0	1	0	0.000
		D	0	0	0	0	0	0.000
		H	0	0	0	0	0	0.000
	C15							-0.220
		5	0	0	0	0	0	0.000
		4	1	0	0	0	1	0.000
		3	4	2	1	1	0	1.500

Baris Total kolom Entropy pada Tabel 4.3 dihitung dengan persamaan (2) sebagai berikut:

$$Entropy(Total) = (-\frac{2}{5} \log_2(\frac{2}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5})) + (-\frac{1}{5} \log_2(\frac{1}{5}))$$

$$Entropy(Total) = 0.980$$

Sementara itu, nilai Gain pada baris C1 dihitung dengan menggunakan persamaan (1) sebagai berikut:

$$Gain(Total, C1) = Entropy(Total) \sum_{i=1}^n \frac{|C1|}{|Total|} * Entropy(C1)$$

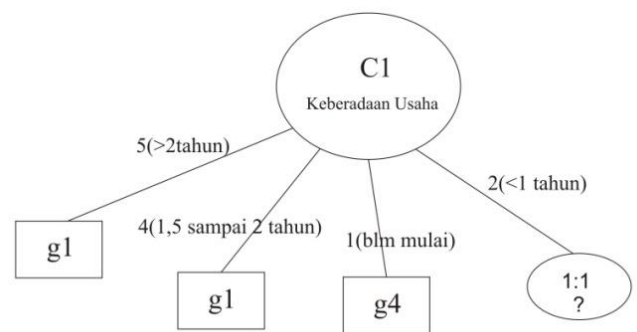
$$Gain(Total, C1) = 0.980 - ((\frac{1}{5} * 0) + (\frac{1}{5} * 0) + (\frac{2}{5} * 1) + (\frac{1}{5} * 0)) = 0.580$$

Untuk nilai Gain atribut yang lain juga dihitung seperti nilai Gain atribut C1.

Dari hasil pada Tabel 4.3 dapat diketahui bahwa atribut dengan Gain tertinggi adalah C1(Keberadaan Usaha), yaitu sebesar 0.580. Dengan demikian, C1 dapat menjadi node akar.

Ada empat nilai rating atribut dari C1, yaitu 1, 2, 4, dan 5. Dari keempat rating atribut tersebut, rating nilai atribut 1, 4, dan 5 sudah mengklasifikasi kasus menjadi rating 1 kelas g4, rating 4 kelas g1 dan rating 5 kelas g1. Dari hasil klasifikasi tersebut maka rating 1, 4, dan 5 dari atribut C1 tidak perlu dilakukan perhitungan lebih lanjut. Akan tetapi untuk nilai rating 2 di atribut C1 masih perlu dilakukan perhitungan lagi.

Dari hasil tersebut digambarkan pohon keputusan semmentaranya tampak seperti Gambar 4.1.



Gambar 4.1 Pohon Keputusan Perhitungan Node 1

b. Menentukan Node Daun

Setelah mendapatkan atribut C1 sebagai *node* akar, kemudian dilakukan perhitungan untuk mencari *node* daun. *Node* daun ini dihitung dari rating C1 yang sama-sama mempunyai rating 2, namun kategori hasilnya masih berbeda. Hasil perhitungan *node* ditunjukkan oleh Tabel 4.4.

		5	0	0	0	0	0	0	
		4	0	0	0	0	0	0	
		3	0	0	0	0	0	0	
		2	2	0	1	1	0	1	
		1	0	0	0	0	0	0	

Tabel 4.4 Perhitungan Node 1.1

No de		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1.1	C1 - 2	2	0	1	1	0	1	
	C2							1
		5 0	0 0	0 0	0 0	0 0		
		4 1	0 0	1 0	0 0			
		3 0	0 0	0 0	0 0			
		2 0	0 0	0 0	0 0			
		1 1	0 1	0 0	0 0			
	C3							1
		5 0	0 0	0 0	0 0			
		4 1	0 0	1 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 1	0 0	0 0			
		1 0	0 0	0 0	0 0			
	C4							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C5							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C6							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C7							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C8							0

Tabel 4.4 Perhitungan Node 1.1 (lanjutan)

No de		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1.1	C1-2	2	0	1	1	0	1	
	C9							1
		5 0	0 0	0 0	0 0			
		4 1	0 0	1 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 1	0 0	0 0			
		1 0	0 0	0 0	0 0			
	C10							1
		5 1	0 1	0 0	0 0			
		4 0	0 0	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C11							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 0	0 0	0 0	0 0			
		1 1	0 0	1 0	0 0			
	C12							1
		5 0	0 0	0 0	0 0			
		4 1	0 1	0 0	0 0			
		3 0	0 0	0 0	0 0			
		2 1	0 0	1 0	0 0			
		1 0	0 0	0 0	0 0			
	C13							1
		3 0	0 0	0 0	0 0			
		2 1	0 1	0 0	0 0			
		1 1	0 0	1 0	0 0			
	C14							1
		F 0	0 0	0 0	0 0			
		M 0	0 0	0 0	0 0			
		E 0	0 0	0 0	0 0			
		G 0	0 0	0 0	0 0			
		I 1	0 1	0 0	0 0			
		K 0	0 0	0 0	0 0			
		C 1	0 0	1 0	0 0			

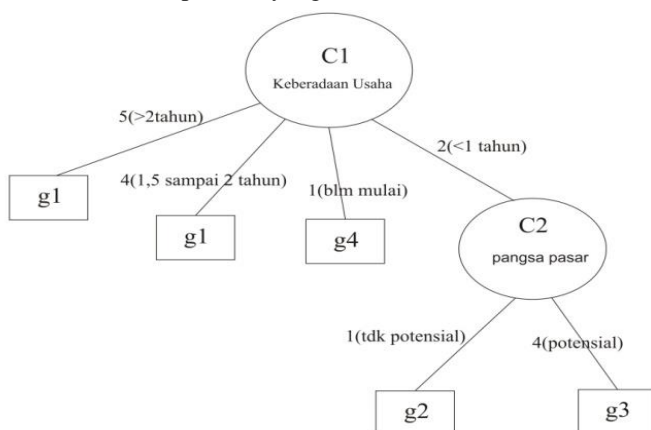
		D	0	0	0	0	0	0	
		H	0	0	0	0	0	0	

Tabel 4.4 Perhitungan Node 1.1 (lanjutan)

No de		Jml (S)	g 1	g 2	g 3	g 4	Entropy	Gain
1.1	C1- 2	2	0	1	1	0	1	
	C15							0
		5	0	0	0	0	0	
		4	0	0	0	0	0	
		3	2	0	1	1	1	
		2	0	0	0	0	0	
		1	0	0	0	0	0	

Dari hasil pada Tabel 4.3 dapat diketahui bahwa atribut dengan *Gain* tertinggi bernilai 1. C2 (pangsa pasar) dapat menjadi *node* cabang dari rating 2 atribut C1 (keberadaan usaha) karena atribut tersebut terletak lebih awal dari atribut lain yang memiliki nilai sama sebesar 1. Ada dua nilai rating dari atribut C2, yaitu 1 dan 4. Dari kedua rating tersebut, rating 1 dan 4 sudah mengklasifikasikan kasus menjadi rating 1 kelas g2 dan rating 4 kelas g3, sehingga tidak perlu dilakukan perhitungan lebih lanjut.

Pohon keputusan yang terbentuk dari kasus ini:



Gambar 4.2 Pohon Keputusan Hasil Perhitungan Node 1.1

Gambar 4.2 di atas merupakan hasil akhir dari proses pembentukan *tree*. Setelah pembentukan *tree* selesai, kemudian dapat dilakukan proses pembuatan aturan berdasarkan *tree* yang terbentuk. Aturan dari *tree* yang terbentuk adalah sebagai berikut:

- Jika Keberadaan Usaha (C1) lebih dari 2 tahun maka mahasiswa tersebut menjalankan usahanya (g1).

- Jika Keberadaan Usaha (C1) 1,5 sampai 2 tahun maka mahasiswa tersebut menjalankan usahanya (g1).
- Jika Keberadaan Usaha (C1) belum mulai maka mahasiswa tersebut tidak mendapatkan dana/tidak diterima (g4).
- Jika Keberadaan Usaha (C1) kurang dari 1 tahun dan pangsa pasar (C2) tidak potensial maka mahasiswa tersebut mengalami kegagalan atau kebangkrutan (g2).
- Jika Keberadaan Usaha (C1) kurang dari 1 tahun dan pangsa pasar (C2) potensial maka mahasiswa tersebut di indikasikan tidak ada kabar (pergi setelah menerima uang) (g3).

Dari langkah-langkah pada tahap tersebut, kemudian akan diimplementasikan ke dalam sistem sehingga menghasilkan model pohon keputusan. Gambar 4.3 adalah tampilan *tree* yang terbentuk dari sistem setelah menjalankan proses klasifikasi.

where clause	goal
where c1 = '1'	g4
where c1 = '2' AND c2 = '1'	g2
where c1 = '2' AND c2 = '4'	g3
where c1 = '4'	g1
where c1 = '5'	g1

Gambar 4.3 Pohon Keputusan sistem

4.5 Pengujian

Setelah mendapatkan model pohon keputusan dari data training, kemudian dilakukan pengujian terhadap hasil model pohon keputusan data training tersebut. Tabel 4.5 merupakan pengujian data *testing*, dilakukan secara random dengan pengujian sebanyak sebanyak sepuluh kali, sedangkan Tabel 4.6 pengujian data training, dilakukan pengujian dengan seluruh data training.

Tabel 4.5 Hasil Pengujian Menggunakan Data Testing dengan 10 kali Pengujian

Percobaan ke-	Hasil Akurasi (%)
1	72.05
2	80.88
3	77.94
4	77.94
5	80.88
6	77.94
7	73.52
8	76.47
9	85.29
10	72.05
Rata-rata Akurasi = 77,49%	

Tabel 4.6 Hasil Pengujian Menggunakan Data Training Keseluruhan

Jenis Data	Jumlah Data	Klasifikasi Benar	Klasifikasi Salah	Akurasi
Data Training	682	657	25	96,33%

Dari Tabel 4.5 di atas dapat disimpulkan bahwa dengan 10% data testing random porposional yang dilakukan sepuluh kali percobaan hasil rata-rata akurasi adalah 77,49%, Sedangkan Tabel 4.6 akurasi pengujian menggunakan seluruh data training sebesar 96,33%.

5. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa pohon keputusan dengan algoritma ID3 dapat dipertimbangkan untuk digunakan dalam hal penyeleksian penerimaan Program Mahasiswa Wirausaha. Hal ini ditunjang dari hasil pengujian sistem yang menggunakan algoritma ID3 memiliki rata-rata akurasi 77,49% dengan 10% data testing random porposional yang dilakukan sepuluh kali dan persentase akurasi 96,33% untuk pengujian menggunakan seluruh data training.

Saran yang dapat diberikan dari penelitian ini yaitu adanya penambahan algoritma lain untuk membandingkan apakah algoritma ID3 memiliki keakuratan yang tinggi dibandingkan algoritma lain dalam kasus penelitian ini.

6. DAFTAR PUSTAKA

- [1] Dikti. 2009. *Pedoman Program Mahasiswa Wirausaha (PMW)*. Kementrian Pendidikan Nasional Direktorat Jendral Pendidikan Tinggi.
- [2] Han, J and Kamber, M. 2006. *Data Mining Concepts and Techniques Second Edition*. San Francisco : Morgan Kauffman
- [3] Larose, D. T. 2005. *Discovering Knowledge in Data: An Introduction to Data mining*. John Willey and Sons, Inc.
- [4] Kusriani & Luthfi, E.T. 2009. *Algoritma Data Mining*. Andi: Yogyakarta
- [5] S. Archana & Dr. K.Elangovan. 2014. *Survey of Classification Techniques in Data Mining*. International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, pg. 65-71
- [6] Sofi Defiyanti. 2008. Perbandingan Kinerja Algoritma ID3 dan C.45 dalam Klasifikasi Spam-Mail.
- [7] Fayyad, U. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [8] Hamidah, I. 2012. *Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5*. Bandung.

- [9] Wahyudin. 2009. *Metode Iterative Dichotomizer 3 (ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru*. Jurnal Pendidikan Teknologi Informasi Dan Komunikasi (Ptik). Vol.2, No.2

