

APLIKASI KLASIFIKASI ALGORITMA C4.5 (STUDI KASUS MASA STUDI MAHASISWA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS MULAWARMAN ANGKATAN 2008)

Miftahul Chair¹⁾, Yuki Novia Nasution²⁾, Nanda Arista Rizki³⁾

^{1,2,3)} Program Studi Statistika, FMIPA, Universitas Mulawarman
Jl. Barong Tongkok No. 5 Kampus Unmul Gn. Kelua Sempaja Samarinda 75119
E-Mail : : iifkhair@yahoo.com¹⁾; yuki.novia.n@gmail.com²⁾; nanda.arista.r@gmail.com³⁾

ABSTRAK

Klasifikasi merupakan pengelompokan sampel berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan variabel target sebagai kategori. Pohon Keputusan adalah pohon klasifikasi yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Penelitian ini membahas tentang pohon keputusan yang dibentuk menggunakan Algoritma C4.5 untuk mengklasifikasi masa studi mahasiswa FMIPA UNMUL angkatan 2008. Algoritma C4.5 merupakan pohon klasifikasi *non* biner di mana cabang pohon bisa lebih dari dua. Dalam Algoritma C4.5, pohon keputusan dibentuk berdasarkan kriteria *entropy*. Berdasarkan hasil penelitian dalam klasifikasi masa studi mahasiswa FMIPA UNMUL angkatan 2008 (102 data) diperoleh 16 aturan yang terbentuk. Dari hasil klasifikasi yang telah dilakukan diperoleh ketepatan akurasi untuk data *training* (75 data) adalah 100 % dan untuk data *testing* (27 data) adalah 72,4 %.

Kata Kunci : algoritma C4.5, *entropy*, klasifikasi, masa studi, pohon keputusan

1. PENDAHULUAN

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk menguraikan, mengidentifikasi informasi yang bermanfaat, dan pengetahuan yang terakut dari berbagai *database* besar [3].

Klasifikasi merupakan pengelompokan sampel berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan variabel target sebagai kategori. Ada beberapa macam pengklasifikasian dalam *data mining* yaitu *decision tree*, *naive Bayes*, *svm*, dan lain-lain [5].

Decision tree (pohon keputusan) adalah pohon klasifikasi yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dikasuskan. Dalam *decision tree*, daerah pengambilan keputusan yang sebelumnya kompleks dapat diubah menjadi lebih sederhana. Banyak algoritma yang dapat dipakai dalam pembentukan *decision tree* yaitu ID3, CART, C4.5, dan lain-lain. Algoritma adalah urutan langkah-langkah yang logis untuk menyelesaikan suatu masalah [6].

Algoritma ID3 (*Iterative Dichotomiser 3*) pertama kali diperkenalkan oleh Quinlan pada Tahun 1986 yang digunakan untuk menginduksi *decision tree*. Algoritma ID3 dapat bekerja baik pada semua fitur yang mempunyai tipe data kategorik (nominal atau ordinal). Namun dalam perkembangannya, Algoritma ID3 mengalami perbaikan menjadi Algoritma C4.5. Perbaikan yang ada pada Algoritma C4.5 adalah dapat menangani fitur dengan tipe numerik (interval atau rasio),

melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set* [6]. Keunggulan algoritma C4.5 adalah tingkat akurasi klasifikasi sangat tinggi dibanding metode klasifikasi lain. Salah satu contoh dari pengklasifikasian Algoritma C4.5 yaitu tentang masa studi mahasiswa.

Masa studi adalah salah satu tolak ukur keberhasilan dalam studi seorang mahasiswa. Semakin cepat lama studi mahasiswa dalam menempuh perkuliahan, dapat mengindikasikan bahwa mahasiswa tersebut rajin dan cerdas. Masa studi juga menjadi evaluasi tersendiri bagi suatu universitas terkait kebijakan pendidikan di universitas tersebut. Salah satu kewajiban universitas adalah mengontrol masa studi mahasiswanya diantaranya dengan menentukan batasan masa studi yang ditempuh mahasiswa yaitu paling lama 7 tahun akademik untuk program sarjana dan program diploma 4 atau sarjana terapan berdasarkan Standar Nasional Pendidikan Tinggi Nomor 44 Tahun 2015.

Ada beberapa faktor yang mempengaruhi masa studi mahasiswa yaitu IPK, jenis kelamin, asal daerah, jurusan kuliah, asal sekolah, dan pekerjaan orang tua. Penelitian ini terkait dengan penelitian sebelumnya dengan judul "Aplikasi *Classification and Regression Tree* (CART) dan Regresi Logistik Ordinal dalam bidang pendidikan dengan studi kasus Predikat Kelulusan Mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman (FMIPA UNMUL)" oleh Nurbianto (2015). Perbedaannya penelitian ini dengan sebelumnya yaitu terletak di

metode penginduksi pohon keputusan dan studi kasus.

Berdasarkan pengantar di atas, maka Penulis tertarik untuk membuat penelitian ilmiah dengan judul “Aplikasi Klasifikasi Algoritma C4.5 (Studi kasus: Masa Studi Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008)”.

2. TINJAUAN PUSTAKA

2.1. 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 [5].

2.2. Klasifikasi

Dalam pengklasifikasian Peneliti mengelompokkan sampel berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan variabel target sebagai kategori. Sebagai contoh, penggolongan masa studi dapat dipisahkan dalam dua kategori, yaitu kurang dari sama dengan 5 tahun dan lebih dari 5 tahun. Beberapa metode atau teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk memprediksi.

2.3. Decision Tree

Decision Tree (Pohon Keputusan) adalah pohon klasifikasi yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Jika semua fitur dalam data menggunakan dua macam nilai kategorik maka bentuk pohon yang didapatkan berupa pohon biner namun jika dalam fitur berisi lebih dari dua macam nilai kategorikal atau menggunakan tipe numerik maka bentuk pohon yang didapatkan biasanya tidak berupa pohon biner [6].

2.4. Kriteria Entropy

Kriteria *entropy* banyak digunakan dalam algoritma ID3, C4.5, dan C5.0 yang didasarkan pada pemilihan titik pemecahan yang memaksimalkan informasi *gain* (pengurangan *entropy* maksimal). Nilai minimal nol ketika semua data pada *node* tersebut dimiliki oleh suatu kelas yang mengimplikasikan paling informatif [6].

Karakteristik dari *decision tree*, dibentuk sejumlah elemen sebagai berikut [8]:

1. *Node*, yang menyatakan variabel. *Node* bisa berupa variabel akar, variabel cabang, dan kelas.
2. *Arm*, setiap cabang menyatakan nilai hasil pengujian di *node* bukan daun.
3. *Node* akar, tidak mempunyai *input arm* yaitu lengan masukan dan mempunyai nol atau lebih *output arm* yaitu lengan keluar.

4. *Node* internal, setiap *node* yang bukan daun (*non terminal*) yang mempunyai tepat satu *input arm* dan dua atau lebih *output arm*, *node* ini menyatakan pengujian yang didasarkan pada nilai fitur.

5. *Node* daun (*terminal*) adalah *node* yang mempunyai tepat satu *input arm* dan tidak mempunyai *output arm*. *Node* ini menyatakan label kelas (keputusan).

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [5].

2.5. Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan pada Tahun 1996 sebagai versi perbaikan dari ID3. Dalam ID3, induksi *Decision tree* hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikannya yaitu tidak hanya dapat menangani fitur bertipe kategorikal, tetapi juga dapat menangani fitur dengan tipe numerik, serta juga dapat melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Algoritma C4.5 juga menggunakan kriteria *gain* dalam menentukan fitur yang menjadi pemecah *node* pada pohon yang diinduksi.

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [4]:

1. Pemilihan variabel akar

Untuk memilih variabel sebagai akar, didasarkan pada nilai *gain* tertinggi dari variabel-variabel yang ada. Berikut adalah cara untuk menghitung nilai *gain*:

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

dengan:

- S : Himpunan kasus
- S_i : Himpunan kasus pada partisi ke i
- A : Variabel
- n : Jumlah partisi atribut A
- $|S_i|$: Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain*, dicari terlebih dahulu nilai *Entropy*. *Entropy* adalah informasi mengenai proporsi pembagian kelas, nilai *entropy* berkisar mulai dari 0 sampai dengan 1, jika nilai *entropy* = 0, maka menandakan jumlah sampel hanya berada di salah satu kelas, sedangkan jika nilai *entropy* = 1, maka menandakan jumlah sampel berada di masing-masing kelas dengan jumlah yang sama. Adapun rumus dasar dari perhitungan *Entropy* adalah sebagai berikut:

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

dengan:

- S : Himpunan Kasus
n : Jumlah partisi S
 p_i : Proporsi dari S_i terhadap S

2. Penentuan cabang untuk masing-masing nilai Untuk penentuan cabang sama seperti mencari variabel akar yaitu didasarkan pada nilai *gain* tertinggi dari variabel-variabel yang ada.
3. Kelas dibagi dalam cabang dan apabila cabang mempunyai dua kelas maka yang dipilih kelas yang terbanyak.
4. Proses diulang untuk masing-masing cabang sampai semua kelas pada cabang memiliki kelasnya masing-masing.

Dalam melakukan perbandingan hasil klasifikasi dan hasil prediksi maka dilakukan perhitungan tingkat akurasi. Untuk mendapatkan tingkat akurasi aplikasi dari hasil prediksi maka digunakan rumus berikut (David. 2014):

$$\text{Persentase} = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah Data}} \times 100\% \quad (3)$$

Algoritma C4.5 dapat menangani berbagai macam contoh kasus. Salah satu contoh dari pengklasifikasian Algoritma C4.5 yaitu tentang masa studi mahasiswa.

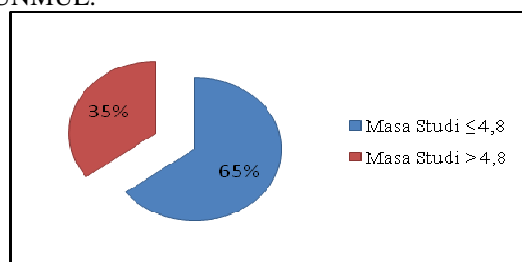
2.6. Masa Studi

Menurut KBBI (Kamus Besar Bahasa Indonesia), istilah sarjana berasal dari bahasa sansekerta, yang merupakan sebuah gelar akademik yang diberikan kepada lulusan program pendidikan strata satu atau undergraduate. Untuk program Strata Satu (S1), persyaratan penyelesaian studi ditetapkan sekurang-kurangnya telah menempuh 144 SKS dan sebanyak-banyaknya 160 SKS, dengan lama masa studi paling lama 7 tahun akademik untuk program sarjana dan program diploma 4 atau sarjana terapan (Standar Nasional Pendidikan Tinggi Nomor 44 Tahun 2015).

3. HASIL DAN PEMBAHASAN

3.1. Analisis Statistika Deskriptif

Karakteristik yang digambarkan pada analisis deskriptif adalah IPK kelulusan, Jenis Kelamin, Asal Daerah, Program Studi, Asal Sekolah, dan Pekerjaan Orang Tua. Gambar 1. merupakan persentase kelulusan masa studi mahasiswa FMIPA UNMUL.



Gambar 1. Karakteristik mahasiswa berdasarkan persentase masa studi

Tabel 1. merupakan tabulasi silang IPK kelulusan dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 1. Tabulasi Silang IPK Lulusan Terhadap Masa Studi

Masa Studi	IPK Lulusan (X_1)			Total
	<3,00	3,00≤IPK<3,50	≥3,50	
Masa Studi ≤ 4,8 tahun	0	32	34	66
Masa Studi > 4,8 tahun	8	19	9	36
Total	8	51	43	102

Tabel 2. merupakan tabulasi silang jenis kelamin dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 2. Tabulasi Silang Jenis Kelamin Terhadap Masa Studi

Masa Studi	Jenis Kelamin (X_2)		Total
	Perempuan	Laki-laki	
Masa Studi ≤ 4,8 tahun	45	21	66
Masa Studi > 4,8 tahun	22	14	36
Total	67	35	102

Tabel 3. merupakan tabulasi silang asal daerah dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 3. Tabulasi Silang Asal Daerah Terhadap Masa Studi

Masa Studi	Asal Daerah (X_3)		Total
	Samarinda	Luar Samarinda	
Masa Studi ≤ 4,8 tahun	27	39	66
Masa Studi > 4,8 tahun	15	21	36
Total	42	60	102

Tabel 4. merupakan tabulasi silang program studi dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 4. Tabulasi Silang Program Studi Terhadap Masa Studi

Masa Studi	Program Studi (X_4)				Total
	Statistika	Kimia	Fisika	Biologi	
Masa Studi ≤ 4,8 tahun	24	33	4	5	66
Masa Studi > 4,8 tahun	7	0	6	23	36
Total	31	33	10	28	102

Tabel 5. merupakan tabulasi silang asal sekolah dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 5. Tabulasi Silang Asal Sekolah Terhadap Masa Studi

Masa Studi	Asal Sekolah (X_5)		Total
	SMA	SMK	
Masa Studi ≤ 4,8 tahun	59	7	66
Masa Studi > 4,8 tahun	31	5	36
Total	90	12	102

Tabel 6. merupakan tabulasi silang pekerjaan orang tua dengan masa studi mahasiswa FMIPA UNMUL.

Tabel 6. Tabulasi Silang Asal Sekolah Terhadap Masa Studi

Masa Studi	Pekerjaan Orang Tua				Total
	PNS	PS	Wiraswasta	Lainnya	
Masa Studi ≤ 4,8 tahun	17	24	11	14	66
Masa Studi > 4,8 tahun	11	11	7	7	36
Total	28	35	18	21	102

3.2. Pembentukan Pohon Klasifikasi Dengan Menggunakan Algoritma C4.5

Dalam proses pembentukan pohon klasifikasi, terhadap empat alur yaitu pemilihan variabel akar, penentuan cabang, kasus dibagi dalam cabang dan proses diulang sampai setiap cabang memiliki kelas yang sama. Pohon klasifikasi yang telah terbentuk akan memunculkan beberapa aturan (*rule*) sebanyak kelas yang terbentuk. Adapun data yang

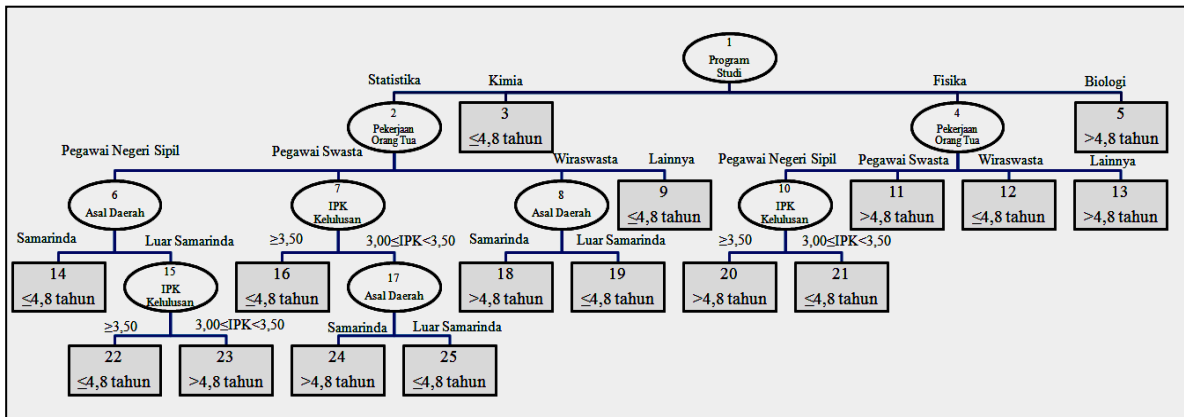
digunakan untuk proses pembentukan pohon klasifikasi ada 75 sampel, sedangkan 27 sampel sisanya untuk data *testing* pohon klasifikasi yang terbentuk.

Tahap pertama dalam pembentukan pohon klasifikasi adalah pemilihan *node* akar. Perhitungan untuk menentukan *node* akar menggunakan Persamaan (1) untuk menentukan nilai *Gain* dan Persamaan (2) untuk menentukan nilai *Entropy*.

Tabel 7. Hasil Perhitungan *Gain* dan *Entropy* untuk *Node* Akar

Variabel	Kategori	Jumlah	Masa Studi ≤4,8 tahun	Masa Studi >4,8 tahun	<i>Entropy</i>	<i>Gain</i>
Total		75	43	32	0,9844	
IPK						0,1735
	≥3,50	29	22	7	0,7973	
	3,00≤IPK<3,50	38	21	17	0,9919	
	<3,00	8	0	8	0	
Jenis						0,0049
Kelamin	Perempuan	48	29	19	0,9684	
	Laki-laki	27	14	13	0,9990	
Asal						0,0032
Daerah	Samarinda	31	19	12	0,9629	
	Luar Samarinda	44	24	20	0,9940	
Program						0,6499
Studi	Statistika	23	17	6	0,8280	
	Kimia	24	24	0	0	
	Fisika	7	2	5	0,8631	
	Biologi	21	0	21	0	
Asal						0,0001
Sekolah	SMA	66	38	28	0,9833	
	SMK	9	5	4	0,9910	
Pekerjaan						0,0097
Orang	PNS	21	12	9	0,9852	
Tua	Pegawai Swasta	27	17	10	0,9509	
	Wiraswasta	13	6	7	0,9957	
	Lainnya	14	8	6	0,9852	

Perhitungan *entropy* dan *gain* dilakukan terus sampai setiap cabang memiliki kelasnya masing-masing. Maka terbentuk pohon klasifikasi sebagai berikut:



Gambar 2. Pohon Klasifikasi

Adapun bentuk aturan *If* dan *Then* untuk *decision tree* sebagai berikut:

- a. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = PNS dan Asal Daerah = Samarinda maka Masa Studi $\leq 4,8$ tahun
- b. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = PNS dan Asal Daerah = Luar Samarinda dan IPK Kelulusan $\geq 3,50$ maka Masa Studi $\leq 4,8$ tahun
- c. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = PNS dan Asal Daerah = Luar Samarinda dan IPK Kelulusan $3,00 \leq IPK < 3,50$ maka Masa Studi $> 4,8$ tahun
- d. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Pegawai Swasta dan IPK Kelulusan $\geq 3,50$ maka Masa Studi $\leq 4,8$ tahun
- e. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Pegawai Swasta dan IPK Kelulusan $3,00 \leq IPK < 3,50$ dan Asal Daerah = Samarinda maka Masa Studi $> 4,8$ tahun
- f. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Pegawai Swasta dan IPK Kelulusan $3,00 \leq IPK < 3,50$ dan Asal Daerah = Luar Samarinda maka Masa Studi $\leq 4,8$ tahun
- g. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Wiraswasta dan Asal Daerah = Samarinda maka Masa Studi $> 4,8$ tahun
- h. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Wiraswasta dan Asal

- i. Jika Program Studi = Statistika dan Pekerjaan Orang Tua = Lainnya maka Masa Studi $\leq 4,8$ tahun
- j. Jika Program Studi = Kimia maka Masa Studi $\leq 4,8$ tahun
- k. Jika Program Studi = Fisika dan Pekerjaan Orang Tua = PNS dan IPK Kelulusan $\geq 3,50$ maka Masa Studi $> 4,8$ tahun
- l. Jika Program Studi = Fisika dan Pekerjaan Orang Tua = PNS dan IPK Kelulusan $3,00 \leq IPK < 3,50$ maka Masa Studi $\leq 4,8$ tahun
- m. Jika Program Studi = Fisika dan Pekerjaan Orang Tua = Pegawai Swasta maka Masa Studi $> 4,8$ tahun
- n. Jika Program Studi = Fisika dan Pekerjaan Orang Tua = Wiraswasta maka Masa Studi $\leq 4,8$ tahun
- o. Jika Program Studi = Fisika dan Pekerjaan Orang Tua = Lainnya maka Masa Studi $> 4,8$ tahun
- p. Jika Program Studi = Biologi maka Masa Studi $> 4,8$ tahun

3.3. Uji akurasi Algoritma C4.5

Untuk mendapatkan tingkat akurasi aplikasi dari hasil prediksi maka digunakan Persamaan (3). Setelah diperoleh hasil akurasi untuk jumlah data *training* 75, dilakukan juga perhitungan analisis untuk jumlah data *training* 50 dan 25, sehingga diperoleh hasil seperti pada Tabel 8.

Tabel 8. Hasil Akurasi Data *Training* dan Data *Testing*

Jenis Data	Percobaan 1		Percobaan 2		Percobaan 3	
	Data <i>Training</i>	Data <i>Testing</i>	Data <i>Training</i>	Data <i>Testing</i>	Data <i>Training</i>	Data <i>Testing</i>
Jumlah Data	75	27	50	52	25	77
Akurasi	100%	77,7%	100%	73%	100%	76,6%

Berdasarkan Tabel 4.14 menunjukkan bahwa untuk setiap data *training* pada setiap percobaan data memiliki tingkat akurasi 100 persen. Sementara itu untuk data *testing* yang memiliki tingkat akurasi tertinggi ada pada percobaan 1 dengan jumlah data *testing* 27 data yaitu 77,7 persen, sedangkan untuk data *testing* yang memiliki tingkat akurasi terendah ada pada percobaan 2 dengan jumlah data *testing* 52 data yaitu 73 persen.

4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan, kesimpulan yang diperoleh dari penelitian mengenai masa studi mahasiswa FMIPA UNMUL yaitu:

1. Hasil pengklasifikasian masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2008 dengan menggunakan metode Algoritma C4.5 didapat pohon klasifikasi dengan 16 aturan yang terbentuk.
2. Ketepatan hasil prediksi terhadap hasil pengklasifikasian masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2008 untuk data *training* dengan jumlah 75 data didapat tingkat akurasi sebesar 100 persen dan data *testing* dengan jumlah data 27 didapat tingkat akurasi sebesar 77,7 persen.

5. DAFTAR PUSTAKA

- [1]. David, N., 2015. Aplikasi Classification and Regresion Tree (CART) dan Regresi Logistik Ordinal dalam Bidang Ilmu Pendidikan dengan Studi Kasus Predikat Kelulusan Mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman. Skripsi, Prodi Statistika. Samarinda: FMIPA UNMUL.
- [2]. David, H., 2014. *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa*. Jurnal yang dipublikasikan. Tangerang: Universitas Multimedia Nusantara Tangerang.
- [3]. Efraim, Turban, J. E. Aronson., dan T. P. Liang, 2005. *Decision Support Systems and Intelligent System*. Yogyakarta : Andi Offset.
- [4]. Kusriani, dan E. T. Luthfy, 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [5]. Larose, 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey : John Willey & Sons.

- [6]. Prasetyo, E., 2012. *Data Mining : Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta : Andi Offset.
- [7]. Quinlan J., 1986. "Introduction of Decision Tree", *Machine Learning*, vol. 1, pp. 81-106.
- [8]. Tan, P. N., M. Steinbach., dan V. Kumar, 2006. *Introduction to Data Mining*. New York : Pearson Education.