

## **PENERAPAN POHON KEPUTUSAN DALAM MEMPREDIKSI MASA STUDI MAHASISWA UIN ALAUDDIN MAKASSAR**

**IRWAN<sup>1</sup>, ADNAN SAUDDIN<sup>2</sup>, NUR IDA<sup>3</sup>**

<sup>1,2,3</sup>Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Alauddin Makassar  
Jl. H. M. Yasin Limpo No. 36

Email: <sup>1</sup>irwan@uin-alauddin.ac.id, <sup>2</sup>adnan.sauddin@uin-alauddin.ac.id,  
<sup>3</sup>60600117071@uin-alauddin.ac.id

### **ABSTRAK**

Mahasiswa merupakan subjek penting dalam menentukan keberhasilan dari suatu perguruan tinggi. Salah satu faktor atau indikator yang menentukan kualitas suatu perguruan tinggi adalah persentasi kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Adapun tujuan dari penelitian ini adalah untuk menemukan kombinasi atribut yang menentukan masa studi mahasiswa selesai tepat waktu pada Prodi Matematika Fakultas Sains dan Teknologi UIN Alauddin Makassar. Penelitian ini menerapkan Pohon Keputusan algoritma ID3 dan C4.5 menggunakan Rapid Miner 5.3. Hasil pengujian penelitian diperoleh tingkat akurasi menggunakan algoritma ID3 sebesar 90,91% dan tingkat akurasi menggunakan algoritma C4.5 sebesar 78,79%.

**Kata kunci:** Algoritma C4.5, ID3, Mahasiswa, Pohon Keputusan

### **I. PENDAHULUAN**

Mahasiswa merupakan subjek penting dalam menentukan keberhasilan dari sebuah perguruan tinggi, baik itu perguruan tinggi negeri maupun perguruan tinggi swasta. Salah satu faktor atau indikator yang menentukan kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studinya tepat waktu. Namun pada kenyataannya, tiap memasuki tahun ajaran baru mahasiswa yang diterima semakin meningkat, tetapi tidak semua mahasiswa dapat menyelesaikan studinya tepat waktu, sehingga mengakibatkan peningkatan jumlah data mahasiswa. Penerapan sebuah metode pohon keputusan dapat digunakan untuk memberikan informasi yang berguna dalam memprediksi masa studi mahasiswa. Pohon keputusan (*Decision Tree*) adalah salah satu teknik yang digunakan untuk mengklasifikasikan sekelompok objek.

Algoritma saat ini sudah banyak untuk pengoptimalan struktur pohon keputusan seperti ID3, C4.5 dan CART. Penelitian sebelumnya menggunakan

algoritma pohon keputusan algoritma ID3 pada data pendidikan adalah Priyanka Saini membahas bahwa algoritma yang sempurna untuk klasifikasi, M.J. Aitkenhead menggunakan pohon keputusan memiliki kemampuan tidak hanya untuk mengembangkan struktur keputusan sistem klasifikasi pohon, tetapi juga dapat menangani berbagai nilai dan tipe data. Selain itu, pohon keputusan klasifikasi lebih langsung bisa dimengerti.

Penelitian lain yang dilakukan Julce Adiana Sidette bahwa pohon keputusan merupakan teknik klasifikasi yang membangun model klasifikasi berdasarkan sekumpulan data masukan. Algoritma ID3 dapat menghasilkan pemodelan pohon keputusan melalui pemangkasan pohon yang lebih sederhana, sehingga cocok untuk kumpulan data yang besar, Adityas Lestari penerapan algoritma C4.5 mampu melakukan prediksi dengan baik (84%) terhadap masa studi mahasiswa, Panji Bimo Nugroho Setio, menyimpulkan bahwa algoritme C4.5 dapat mengonstruksi pohon keputusan (*decision tree*) pada algoritme ID3 dan memperbaiki atribut dengan data kosong dan pemangkasan cabang.

Dalam penelitian ini menerapkan metode pohon keputusan menggunakan algoritma ID3 dan C4.5 untuk mengetahui atribut-atribut yang berpengaruh terhadap lama waktu studi mahasiswa. Adapun dalam penelitian ini atribut-atribut yang digunakan adalah jenis kelamin, kota kelahiran, jenis asal sekolah, IPK, jumlah SKS, bekerja dan aktivis.

## **II.METODE PENELITIAN**

Jenis penelitian yang digunakan adalah penelitian terapan dengan jenis data dalam penelitian ini adalah data primer diperoleh dari data kuesioner yang diisi oleh mahasiswa alumni Prodi Matematika Fakultas Sains dan Teknologi UIN Alauddin Makassar tahun angkatan 2014-2016. Penelitian ini dilakukan di Prodi Matematika mulai dari Maret-Oktober 2021. Adapun atribut yang digunakan dalam penelitian ini yaitu ketepatan masa studi mahasiswa sebagai atribut target dan untuk variabel prediktornya yaitu jenis kelamin, kota kelahiran, jenis asal sekolah, IPK, jumlah SKS, aktivis dan bekerja. Pada penelitian ini dilakukan beberapa tahapan penelitian sebagai berikut :

1. Mempersiapkan atau mengumpulkan data yang akan digunakan dalam memprediksi masa studi mahasiswa UIN Alauddin Makassar

2. Melakukan tahapan atau proses data mining
3. Menerapkan metode pohon keputusan menggunakan algoritma ID3 dan C4.5
4. Membuat kesimpulan

### III. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

##### Deskripsi Data

Data yang digunakan dalam penelitian ini untuk menentukan masa studi mahasiswa sebanyak 33 diperoleh dari data kuesioner yang diisi oleh mahasiswa alumni Prodi Matematika Fakultas Sains dan Teknologi UIN Alauddin Makassar tahun angkatan 2014-2016. Berikut ini tabel data mahasiswa alumni.

Tabel 1. Tabel Data Set Mahasiswa Alumni

		Masa studi mahasiswa			
		Tepat waktu	Tidak tepat waktu	Jumlah	Total
Jenis kelamin	Laki-laki	2	5	7	33
	Perempuan	11	15	26	
IPK	Rendah	0	0	0	33
	Sedang	4	14	18	
	Tinggi	10	5	15	
Kota kelahiran	Makassar	1	3	4	33
	Luar Makassar	12	17	29	
Asal sekolah	SMA	10	16	26	33
	SMK	1	3	4	
	MA	2	1	3	
Jumlah SKS	≤ 20	3	3	6	33
	> 20	10	17	27	
Aktivis	Ya	12	16	28	33
	Tidak	1	4	5	
Bekerja	Ya	1	5	6	33
	Tidak	12	15	27	

### 3.2 Algoritma Pohon Keputusan

Algoritma pohon keputusan yang digunakan pada penelitian ini adalah algoritma ID3 dan C4.5. Adapun langkah kerja dalam pembentukan pohon klasifikasi menggunakan algoritma ID3 yaitu dimulai dengan menghitung *information gain* dari setiap atribut, memilih atribut dengan perolehan nilai *information gain* terbesar, membentuk node yang berisi atribut tersebut, kemudian mengulangi proses penghitungan *information gain* tersebut. *Information gain* akan berlanjut hingga semua data dimasukkan dalam kategori yang sama. Atribut yang dipilih tidak lagi disertakan dalam penghitungan nilai *information gain*. Berdasarkan Tabel 4.1, maka langkah selanjutnya adalah melakukan perhitungan dengan mencari nilai *entropy* dari masing-masing atribut dan nilai *gain* untuk menentukan akar (*root*) dari pohon keputusan. Nilai *gain* tertinggi sebagai cabang pohon yang teratas, yang menjadi pohon pembuka diantara atribut lainnya.

$$\begin{aligned} Entropy (Total) &= \sum_{i=1}^k -p_i \log_2 p_i \\ &= \left(-\frac{13}{33} \log_2 \left(\frac{13}{33}\right) + \left(-\frac{20}{33} \log_2 \left(\frac{20}{33}\right)\right)\right) \\ &= 0,529436 + 0,437858 \\ &= 0,967294 \end{aligned}$$

$$\begin{aligned} Entropy (IPK Sedang) &= \sum_{i=1}^k -p_i \log_2 p_i \\ &= \left(-\frac{4}{18} \log_2 \left(\frac{4}{18}\right) + \left(-\frac{14}{18} \log_2 \left(\frac{14}{18}\right)\right)\right) \\ &= 0,482205 + 0,281998 \\ &= 0,764203 \end{aligned}$$

$$\begin{aligned} Entropy (IPK Tinggi) &= \sum_{i=1}^k -p_i \log_2 p_i \\ &= \left(-\frac{10}{15} \log_2 \left(\frac{10}{15}\right) + \left(-\frac{5}{15} \log_2 \left(\frac{5}{15}\right)\right)\right) \\ &= 0,389975 + 0,528320 \\ &= 0,918295 \end{aligned}$$

Menghitung nilai *gain* pada atribut IPK

$$\begin{aligned} Gain (S, A) &= Entropy (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy (S_i) \\ &= 0,967294 - \left(\left(\frac{18}{33} * 0,764203\right) + \left(\frac{15}{33} * 0,918295\right)\right) \end{aligned}$$

$$\begin{aligned}
 &= 0,967294 - (0,416838 + 0,417406) \\
 &= 0,967294 - 0,834244 \\
 &= 0,13305
 \end{aligned}$$

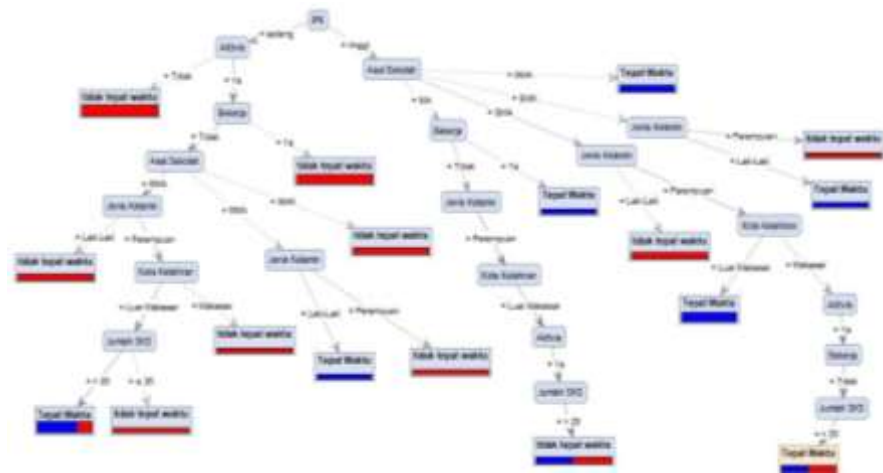
Berdasarkan perhitungan nilai *entropy* dan *gain* pada masing-masing atribut disajikan dalam sebuah tabel berikut ini.

Tabel 2. Hasil Perhitungan Nilai *Entropy* dan *Gain*

Atribut	Kelas	Jumlah kasus	Tepat waktu	Tidak tepat waktu	<i>Entropy</i>	<i>Gain</i>
		33	13	20	0,967294	
IPK	Rendah	0	0	0	0	0,13305
	Sedang	18	4	14	0,764203	
	Tinggi	15	10	5	0,918295	
Jenis kelamin	Laki-laki	7	2	5	0,86312	0,009836
	Perempuan	26	11	15	0,982857	
Kota kelahiran	Makassar	4	1	3	0,811278	0,009109
	Luar Makassar	29	12	17	0,978449	
Asal sekolah	SMA	26	10	16	0,961235	0,028141
	SMK	4	1	3	0,811278	
	MA	3	2	1	0,918295	
Jumlah SKS	≤ 20	6	3	3	1	0,007422
	>20	27	10	17	0,950955	
Aktivis	Ya	28	12	16	0,985227	0,021962
	Tidak	5	1	4	0,721927	
Bekerja	Ya	6	1	5	0,650022	0,03823
	Tidak	27	12	15	0,991075	

Berdasarkan hasil perhitungan *gain* dari masing-masing atribut terlihat pada Tabel 2. bahwa atribut IPK memiliki nilai *gain* tertinggi yaitu 0,13305, sehingga atribut IPK dijadikan sebagai *root node* atau simpul akar dari pohon keputusan.

Langkah selanjutnya adalah menghitung kembali nilai *entropy* dan *gain* untuk menentukan simpul berikutnya dengan cara yang sama seperti pengerjaan di atas, nilai yang dihitung berdasarkan *entropy* kelas IPK rendah dan IPK tinggi. Setelah semua nilai *entropy* dan *gain* dihitung, maka diperoleh hasil pohon keputusan sebagai berikut :



Gambar 1. Pohon Keputusan Masa Studi Mahasiswa

Berdasarkan Gambar 1. di atas, pohon yang dihasilkan ada 16 aturan (*rule*) yang dapat disimpulkan dan dinyatakan dalam bentuk jika maka, sebagai contoh Jika IPK [sedang] dan Aktivis [tidak] maka mahasiswa tidak tepat waktu. Hasil perhitungan pada pohon keputusan algoritma ID3 menggunakan software Rapid Miner diperoleh nilai akurasi sebesar 90,91%

Tabel 3. Akurasi Prediksi Masa Studi Mahasiswa

accuracy: 90.91%			
	true Tepat Waktu	true tidak tepat waktu	class precision
pred. Tepat Waktu	12	2	85.71%
pred. tidak tepat waktu	1	18	94.74%
class recall	92.31%	90.00%	

Berdasarkan Tabel 2. Hasil Perhitungan Nilai *Entropy* dan *Gain*, maka langkah selanjutnya pada algoritma C4.5 adalah menghitung nilai *split info* dan *gain ratio* dari masing-masing atribut. Nilai *gain ratio* ini nantinya yang akan mejadi akar (*root*) dari pohon keputusan.

$$\begin{aligned}
 \text{Split Info (IPK)} &= - \sum_{i=1}^k \frac{s_i}{S} \log_2 \frac{s_i}{S} \\
 &= - \left( \left( \frac{18}{33} * \log_2 \frac{18}{33} \right) + \left( \frac{15}{33} * \log_2 \frac{15}{33} \right) \right) \\
 &= - \left( -0,476983 + (-0,517047) \right) \\
 &= - \left( -0,476983 - 0,517047 \right) \\
 &= 0,99403
 \end{aligned}$$

Menghitung *gain ratio* atribut IPK

$$\begin{aligned}
 \text{Gain Ratio} &= \frac{\text{Gain (A)}}{\text{Split Info(S, A)}} \\
 &= \frac{0,13305}{0,99403} \\
 &= 0,133849
 \end{aligned}$$

Berdasarkan perhitungan nilai *entropy*, *gain*, *split info*, dan *gain ratio* pada masing-masing atribut disajikan dalam sebuah tabel berikut ini.

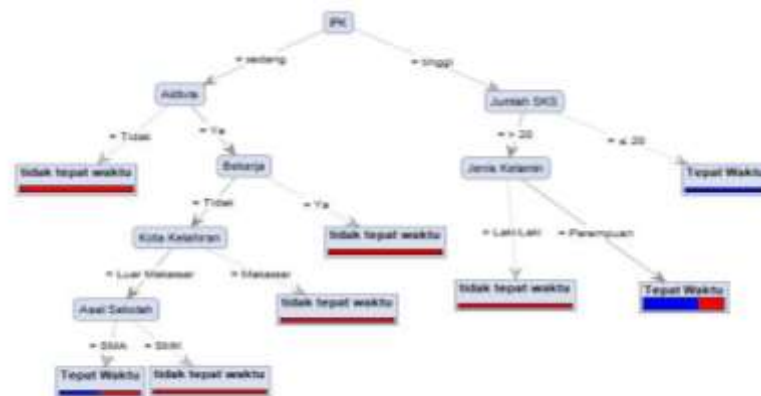
Tabel 4. Hasil Perhitungan Nilai *Entropy*, *Gain*, *Split Info*, dan *Gain Ratio*

Atribut	Kelas	Jumlah kasus	Tepat waktu	Tidak tepat waktu	<i>Entropy</i>	<i>Gain</i>	<i>Split info</i>	<i>Gain ratio</i>
		33	13	20	0,967294			
IPK	Rendah	0	0	0	0	0,13305	0,99403	0,133849
	Sedang	18	4	14	0,764203			
	Tinggi	15	10	5	0,918295			
Jenis kelamin	Laki-laki	7	2	5	0,86312	0,009836	0,745517	0,013193
	Perempuan	26	11	15	0,982857			
Kota kelahiran	Makassar	4	1	3	0,811278	0,009109	0,532834	0,017095
	Luar Makassar	29	12	17	0,978449			
Asal sekolah	SMA	26	10	16	0,961235	0,028141	0,954504	0,029482
	SMK	4	1	3	0,811278			
	MA	3	2	1	0,918295			
Jumlah SKS	≤ 20	6	3	3	1	0,007422	0,684038	0,010850



	>20	27	10	17	0,950955			
Aktivis	Ya	28	12	16	0,985227	0,021962	0,613618	0,035790
	Tidak	5	1	4	0,721927			
Bekerja	Ya	6	1	5	0,650022	0,03823	0,684038	0,055888
	Tidak	27	12	15	0,991075			

Berdasarkan hasil perhitungan *gain ratio* dari masing-masing atribut terlihat pada Tabel 4. bahwa atribut IPK memiliki nilai *gain ratio* tertinggi yaitu 0,133849 sehingga atribut IPK dijadikan sebagai *root node* atau simpul akar dari pohon keputusan. Langkah selanjutnya adalah menghitung kembali nilai *entropy* dan *gain*, *split info* dan *gain ratio* untuk menentukan simpul berikutnya dengan cara yang sama seperti pengerjaan di atas, nilai yang dihitung berdasarkan *entropy* kelas IPK rendah dan IPK tinggi. Setelah semua nilai dihitung, maka diperoleh hasil pohon keputusan sebagai berikut :



Gambar 2. Pohon Keputusan Masa Studi Mahasiswa

Berdasarkan Gambar 2. di atas, pohon yang dihasilkan ada 8 aturan (*rule*) dapat disimpulkan dan dinyatakan dalam bentuk jika maka seperti pada lagoritma ID3. Hasil perhitungan pada pohon keputusan algoritma C4.5 menggunakan software Rapid Miner diperoleh nilai akurasi sebesar 78,79% seperti pada tabel berikut :



Tabel 5. Akurasi Prediksi Masa Studi M

accuracy: 78.79%			
	true Tepat Waktu	true tidak tepat waktu	class precision
pred. Tepat Waktu	13	7	65.00%
pred. tidak tepat waktu	0	13	100.00%
class recall	100.00%	65.00%	

### 3.3 Pembahasan

Berdasarkan data penelitian yang diperoleh pada Tabel 1. dengan menggunakan metode pohon keputusan algoritma ID3 pada penelitian ini, pembentukan pohon keputusan dilakukan perhitungan nilai *entropy* total dan nilai *entropy* dari setiap kelas atribut yang digunakan. *Entropy* digunakan untuk mengukur seberapa informatifnya atau seberapa baiknya sebuah node. Setelah memperoleh nilai *entropy*, selanjutnya adalah menghitung nilai *gain* dari setiap atribut, setelah memperoleh nilai *gain*, selanjutnya pemilihan atribut yang dilakukan dengan nilai *information gain* tertinggi, seperti yang ditunjukkan pada Tabel 2. Hasil Perhitungan Nilai *Entropy* dan *Gain*.

Berdasarkan hasil perhitungan tersebut diperoleh nilai *entropy* total yaitu 0,967294 dan nilai *gain* tertinggi terdapat pada atribut IPK yaitu 0,13305 sehingga berdasarkan nilai *gain* tertinggi, atribut IPK dijadikan sebagai *root node* (simpul akar) dari pohon keputusan. Langkah selanjutnya adalah menghitung kembali nilai *entropy* dan *gain* untuk menentukan simpul 1.1 dan 1.2, nilai yang dihitung berdasarkan *entropy* kelas IPK sedang dan tinggi. Setelah semua perhitungan nilai *entropy* dan nilai *gain* pada atribut selesai, pohon keputusan yang dihasilkan seperti terlihat pada Gambar 1.

Berdasarkan data penelitian yang diperoleh pada Tabel 1. dengan menggunakan metode pohon keputusan algoritma C4.5, pembentukan pohon keputusan dilakukan dengan perhitungan nilai *entropy* total dan nilai *entropy* dari setiap kelas atribut yang digunakan, seperti pada langkah-langkah perhitungan pohon keputusan menggunakan algoritma ID3. Hanya saja pada algoritma C4.5 nilai *split info* dan *gain ratio* harus dihitung, karena nilai *gain ratio* yang akan digunakan untuk menentukan akar (*root*) dari pohon keputusan. Setelah semua

perhitungan nilai pada atribut selesai, pohon keputusan yang dihasilkan seperti terlihat pada Gambar 2.

#### IV.KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa pohon keputusan dengan algoritma ID3 dan C4.5 dapat diimplementasikan untuk memprediksi masa studi mahasiswa Prodi Matematika dengan dua kategori yaitu lulus tepat waktu dan lulus tidak tepat waktu. Penggunaan algoritma ID3 dan C4.5 dengan menggunakan software rapidminer menggunakan 33 data training, menghasilkan 7 *rule* lulus tepat waktu, 9 *rule* lulus terlambat dengan nilai akurasi sebesar 90,91% dan algoritma C4.5 menghasilkan 3 *rule* lulus tepat waktu, 5 *rule* lulus terlambat dengan nilai akurasi sebesar 78,79%.

#### DAFTAR PUSTAKA

- Aguilar-Chinea, Rosa Maria, Ivan Castilla Rodriguez, Christopher Exposito, Belen Melian-Batista, Jose Marcos Moreno-Vega. 2019. *Using a decision tree algorithm to predict the robustness of a transshipment schedule. Procedia Computer Science*, 149.
- Aitkenhead, M.J. 2008. "A co-evolving decision tree classification method". *Expert Systems with Applications* 34, (2008): h. 19.
- Ashraf Aysha, Sajid Anwer, Muhammad Gufran Khan. 2018. A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*. vol 44(1):122-136.
- Lestari NA. 2019. Penerapan *Data Mining* menggunakan Metode *Decision Tree C4.5* untuk Prediksi Tingkat Kelulusan Mahasiswa (Studi Kasus : STMIK WIT). *Jurnal Web Informatika Teknologi*. Vol 4(1) (Mei 2019): hal. 1-8.
- Saini P, Sweta Rai, Ajit Kumar Jain. 2014. Decision Tree Algorithm Implementation Using Educational Data. (*International Journal of Computer-Aided technologies (IJCAx)*). Vol 1(1):31
- Sauddin Adnan, Wahidah Alwi, A. Ningsih AN. 2019. Klasifikasi Tingkat Partisipasi Angkatan Kerja Kota Makassar Menggunakan Metode CART. *Jurnal Matematika dan Statistika serta Aplikasinya*. vol 7(2) (Ed. Juli-Des 2019): hal. 20-36.
- Setio PBN, Dewi Retno Sari Saputro, Bowo Winarno. 2020. Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *Prisma* vol 3. hal 64-71.