

## Implementasi Algoritma C4.5 untuk Memprediksi Tingkat Ketepatan Kelulusan Mahasiswa

Imrah Sari<sup>1\*</sup>, Sarjon Defri<sup>2</sup>, Sumijan<sup>3</sup>

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Puta Indonesia “YPTK” Padang, Indonesia

E-mail: imrahsari@gmail.com<sup>1</sup>, sarjonnd@yahoo.co.uk<sup>2</sup>, sumijan@upiypk.ac.id<sup>3</sup>

### Abstract

*Timeliness of graduation not only reflects the competence of graduates but also affects the assessment of study programme accreditation. To achieve this goal, it is important to predict and classify the timeliness of graduation to support more effective academic decision making. In this research, the Knowledge Discovery in Database (KDD) process is used, which aims to find knowledge from big data. One of the main stages in KDD is data mining, which focuses on pattern extraction with various algorithms. This research uses the C4.5 algorithm, a classification method that builds a decision tree to identify attributes that affect the timeliness of student graduation. This study uses data from students in 2017, 2018, and 2019 from the Bachelor of Nursing and Bachelor of Public Health study programmes at Syedza Sainatika University, with a total sample of 46 student records. The C4.5 algorithm is applied to form a decision tree model, which produces classification rules based on attributes such as Grade Point Average (GPA), Study Programme, Gender, and Region of Origin. The results of the C4.5 algorithm implementation show a prediction accuracy of 89.13%, with GPA as the most dominant factor in influencing graduation accuracy. This research proves that the C4.5 algorithm is effective in predicting the timeliness of student graduation.*

**Keywords:** C4.5, Prediction, Classification, Graduation Accuracy, Student

### Abstrak

*Ketepatan waktu kelulusan tidak hanya mencerminkan kompetensi lulusan tetapi juga mempengaruhi penilaian akreditasi program studi. Dalam penelitian ini, digunakan proses Knowledge Discovery in Database (KDD), yang bertujuan untuk menemukan pengetahuan dari data yang besar. Salah satu tahapan utama dalam KDD adalah data mining, yang berfokus pada ekstraksi pola dengan berbagai algoritma. Penelitian ini menggunakan algoritma C4.5, sebuah metode klasifikasi yang membangun pohon keputusan untuk mengidentifikasi atribut-atribut yang berpengaruh terhadap ketepatan waktu kelulusan mahasiswa. Penelitian ini menggunakan data yang berasal dari mahasiswa angkatan 2017, 2018, dan 2019 dari program studi Sarjana Keperawatan dan Sarjana Kesehatan Masyarakat di Universitas Syedza Sainatika, dengan jumlah sampel sebanyak 46 record mahasiswa. Algoritma C4.5 diterapkan untuk membentuk model pohon keputusan, yang menghasilkan aturan-aturan klasifikasi berdasarkan atribut seperti Indeks Prestasi Kumulatif (IPK), Program Studi, Jenis Kelamin, dan Daerah Asal. Hasil implementasi algoritma C4.5 menunjukkan akurasi prediksi sebesar 89,13%, dengan IPK sebagai faktor paling dominan dalam mempengaruhi ketepatan kelulusan. Penelitian ini membuktikan bahwa algoritma C4.5 efektif dalam memprediksi ketepatan waktu kelulusan mahasiswa.*

**Kata kunci:** C4.5, Prediksi, Klasifikasi, Ketepatan Kelulusan, Mahasiswa

## 1. Pendahuluan

Knowledge Discovery In Database (KDD) adalah suatu pendekatan untuk mengekstrak informasi dari database yang tersedia. Database terdiri dari tabel - tabel yang saling terhubung. pengetahuan yang dihasilkan dari proses ini dapat digunakan sebagai basis pengetahuan untuk mendukung pengambilan keputusan [1]. Knowledge Discovery in Database sering disamakan dengan Data Mining, karena keduanya bertujuan mencari pengetahuan yang bermanfaat dalam kumpulan data yang besar. Data mining merupakan salah satu tahap dalam KDD yang berfokus untuk mengambil pola atau model dari data dengan menggunakan algoritma tertentu [2]. Menurut [3] Data mining adalah proses yang melibatkan penggunaan dan pengumpulan data untuk mengidentifikasi hubungan atau pola dalam kumpulan data yang besar. Tujuan utama dari penambangan data adalah mengidentifikasi pola, hubungan, dan tren yang tidak terlihat yang sulit diungkap melalui analisis manual. Salah satu metode yang digunakan dalam penambangan data adalah klasifikasi, yang berfungsi untuk menganalisis data yang telah dikelompokkan dan mengambil contoh dari data tersebut [4]. Algoritma C4.5 adalah metode yang diterapkan untuk pengelompokan, segmentasi, atau klasifikasi data yang memiliki karakteristik prediktif [5]. Algoritma C4.5 adalah pengembangan algoritma ID3 yang dirancang oleh J. Rose Quinlan [6]. Meskipun ada pengembangan, prinsip dasar kerja C4.5 tetap sejalan dengan algoritma ID3 [7]. Algoritma C4.5 beroperasi dengan cara melakukan komputasi melalui pembacaan sampel dari penyimpanan ke memori, yang kemudian digunakan untuk membangun Decision Tree [8].

Penelitian dengan menggunakan Algoritma C4.5 sudah banyak dilakukan oleh peneliti dengan bermacam-macam masalah. Seperti dalam penelitian yang dilakukan oleh [9] dalam Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik Menggunakan Metode Algoritma Decision Tree C4.5. Akurasi yang diperoleh melalui penerapan algoritma C4.5 mencapai 94,11%, di mana faktor yang berperan sebagai root node adalah Jumlah SKS Lulus, yang berpengaruh terhadap pemilihan peminatan.

Penelitian yang dilakukan [10] dalam memprediksi masa studi mahasiswa berdasarkan data mahasiswa Universitas Muhammadiyah Prof. DR. Hamka dengan menggunakan Decision Tree Algorithm (C4.5) maka didapatkan hasil tingkat akurasi mencapai 83,87%, dengan presisi 87,50% dan recall sebesar 91,18%.

Penelitian yang dilakukan [6], dalam memprediksi ketepatan kelulusan mahasiswa berdasarkan data mahasiswa Universitas Islam Negeri Syarif Hidayatullah Jakarta menunjukkan tingkat akurasi sebesar 75,52%, dengan nilai presisi 75,50% dan recall 75,50%.

Penelitian yang dilakukan [11], dalam memprediksi status kelulusan mahasiswa di Politeknik Kampar menggunakan algoritma C4.5. Data yang dipakai dalam penelitian ini merupakan data privat yang diperoleh dari Politeknik Kampar, mencakup nilai akademik lulusan dari tahun 2015 hingga 2019 dengan total 306 catatan, serta data histori lulusan angkatan 2020 yang terdiri dari 143 catatan. Penelitian ini menunjukkan nilai akurasi sebesar 75%, dengan rata-rata presisi mencapai 74% dan rata-rata sensitivitas sebesar 67%.

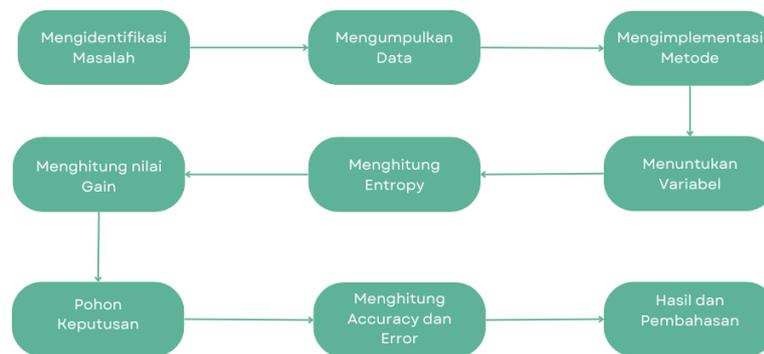
Pada Perguruan Tinggi seperti Universitas Syedza Sainika prediksi juga perlu dilakukan agar dapat menimalisir mahasiswa yang lulus tidak tepat waktu. Institusi pendidikan tinggi bertanggung jawab untuk menghasilkan lulusan yang memiliki kompetensi. Penilaian terhadap hal ini dapat dilakukan melalui tingkat kelulusan mahasiswa [12]. Kelulusan tepat waktu merupakan salah satu indikator penting dalam menilai mutu lembaga pendidikan tinggi. Indikator ini menjadi salah satu kriteria dalam proses akreditasi program studi dan lembaga yang dilaksanakan oleh Badan Akreditasi Perguruan Tinggi (BAN-PT) [6]. Lembaga pendidikan perlu memperhatikan durasi studi mahasiswa dengan mengidentifikasi dini mereka yang memiliki nilai akademis kurang memadai. Program studi harus memberikan pendampingan khusus agar mahasiswa dapat

lulus tepat waktu. Penerapan metode prediksi kelulusan dapat membantu institusi mendeteksi potensi keterlambatan dan melakukan intervensi yang efektif.

Dari uraian permasalahan di atas, penulis akan membahas bagaimana Penerapan Algoritma Decision Tree (C4.5) berfungsi untuk menyelesaikan masalah dengan cara membangun struktur pohon keputusan yang memetakan data ke berbagai kelas berdasarkan aturan-aturan yang ditemukan. (1)

## 2. Metodologi Penelitian

Metodologi penelitian berfungsi sebagai panduan terstruktur untuk mencapai tujuan penelitian secara efektif. Ini dimulai dari identifikasi masalah hingga penyusunan hasil dan pembahasan, memastikan penelitian tetap pada jalurnya. Kerangka kerja Penelitian ini diilustrasikan dalam Gambar 1.



**Gambar 1.** Kerangka Kerja

Tahap identifikasi masalah adalah langkah pertama penelitian yang bertujuan untuk mengidentifikasi isu dan batasan pada objek penelitian. Pada tahap ini, penting untuk memahami kendala ketepatan kelulusan di Universitas Syedza Saintika agar perumusan masalah dapat tepat dan fokus penelitian mahasiswa dapat ditentukan.

### a) Mengumpulkan Data

Prosedur pengumpulan data adalah langkah awal dalam penelitian, di mana data dikumpulkan untuk menjawab pertanyaan, memverifikasi hipotesis, atau mendukung keputusan. Kualitas dan akurasi data yang diperoleh sangat memengaruhi hasil analisis.

### b) Mengimplementasi Metode

Penerapan metode C4.5 merupakan langkah dalam menggunakan algoritma data mining yang berfungsi untuk prediksi pada sebuah dataset.

### c) Menentukan Variabel

Menentukan variabel adalah proses memilih elemen data yang mencerminkan karakteristik objek. Dalam data mahasiswa, variabel keputusan adalah lulus tepat waktu atau terlambat, dengan variabel penentu sebagai berikut:

1. Jenis Kelamin.
2. Daerah Asal.
3. Program Studi.
4. IPK.

### d) Menghitung Entropy

Menghitung entropi penting dalam algoritma pohon keputusan seperti C4.5 untuk menilai ketidakpastian dalam dataset dan menentukan atribut terbaik untuk memisahkan data.

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

Keterangan:

- S : himpunan kasus
- n : jumlah partisi S
- $p_i$ : proporsi dari  $S_i$  terhadap S

#### e) Menghitung nilai *Gain*

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

Keterangan:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- $|S_i|$  : jumlah kasus pada partisi ke-i
- $|S|$  : jumlah kasus dalam S

#### f) Pohon Keputusan

Pohon keputusan adalah model prediktif yang digunakan untuk klasifikasi atau pengambilan keputusan berdasarkan aturan yang dihasilkan dari data. Setiap cabang pohon mencerminkan pemisahan data berdasarkan atribut tertentu.

#### g) Menghitung *Accuracy* dan *Error*

Akurasi dan Kesalahan dari algoritma C4.5 dihitung untuk mengevaluasi kedekatan prediksi dengan nilai aktual. Akurasi tinggi dan Kesalahan rendah menunjukkan klasifikasi yang lebih akurat. Rumus perhitungannya adalah sebagai berikut:

*Accuracy* = (Jumlah yang diklasifikasikan secara benar)/(Tatal sampel testing yang di uji).

*Error* = (Jumlah yang diklasifikasikan secara salah)/(Tatal sampel testing yang di uji).

#### h) Membuat Hasil dan Pembahasan

Hasil dan pembahasan menyajikan gambaran serta temuan dari penelitian ini. Melalui analisis Data Mining, narasi ini menguraikan temuan signifikan, interpretasi data, dan dampaknya terhadap tujuan penelitian.

### 3. Hasil dan Pembahasan

#### 3.1. Analisis Data Mining

Dalam prosesnya data mining mengambil informasi berharga dengan menganalisis pola atau hubungan tertentu dalam data. Proses-proses data mining yang menghasilkan informasi meliputi beberapa langkah:

- a) Pembersihan data adalah langkah penting untuk menghapus gangguan dan data yang tidak konsisten atau tidak relevan.
- b) Integrasi data melibatkan penggabungan informasi dari berbagai sumber ke dalam satu basis data yang terintegrasi.
- c) Seleksi data berfungsi untuk memilih informasi yang tepat untuk analisis dari basis data yang ada.
- d) Transformasi data adalah proses mengubah atau mengonversi data ke dalam format yang sesuai agar dapat diproses dalam data mining.
- e) Proses mining merupakan tahap kunci di mana metode digunakan untuk mengungkap pengetahuan berharga yang tersembunyi dalam data.
- f) Evaluasi pola bertujuan untuk mengidentifikasi pola-pola menarik yang dapat dimasukkan ke dalam basis pengetahuan yang telah ditemukan.

### 3.2. Analisa Teknik Pengolahan Data

Data yang akan diproses memiliki sejumlah kriteria yang menjadi syarat dalam penerapan Data Mining memanfaatkan algoritma C4.5 untuk memprediksi apakah mahasiswa lulus tepat waktu atau terlambat.

#### a) Data

Penelitian ini menggunakan data 46 mahasiswa angkatan 2017, 2018, dan 2019 dari Program Studi Sarjana Keperawatan dan Sarjana Kesehatan Masyarakat di Universitas Syedza Saintika. Data dibagi menjadi dua klasifikasi, tepat dan terlambat. Algoritma C4.5 digunakan untuk memprediksi kelulusan dalam dua kategori tersebut.

#### b) Pemilihan Variabel

Pemilihan variabel adalah proses memilih elemen data yang mencerminkan karakteristik atau fitur dari objek data. Dalam konteks data mahasiswa, variabel yang dipilih untuk keputusan adalah lulus tepat waktu dan terlambat. Sedangkan yang akan diambil sebagai variabel penentu untuk keputusan adalah sebagai berikut :

1. Jenis Kelamin.
2. Daerah Asal.
3. Program Studi.
4. IPK, yaitu nilai kumulatif mahasiswa selama menempuh studi.

#### c) Melakukan Pra-Proses

Setelah variabel dipilih, format data akan disesuaikan menjadi seperti pada tabel 1 berdasarkan variabel-variabel yang telah ditentukan.

**Tabel 1.** Format Data Mahasiswa Setelah Pemilihan Variabel

No	NIM	Jenis Kelamin	Daerah Asal	Program Studi	IPK
1	1802043	Laki-Laki	SUMBAR	Keperawatan	3,46
2	1802044	Perempuan	JAMBI	Keperawatan	3,63
3	1802046	Perempuan	LAMPUNG	Keperawatan	3,49
⋮					
44	1703022	Perempuan	SUMBAR	Kesehatan Masyarakat	3,52
45	1703033	Perempuan	SUMBAR	Kesehatan Masyarakat	3,46
46	1703038	Perempuan	SUMBAR	Kesehatan Masyarakat	3,52

Variabel yang telah dipilih selanjutnya akan dikelompokkan berdasarkan atribut yang ada:

#### 1) Mengelompokkan Jenis Kelamin

Variabel jenis kelamin terbagi menjadi dua kategori, 'Perempuan' dan 'Laki-laki'. Kategori ini disingkat menjadi 'L' untuk Laki-laki dan 'P' untuk Perempuan, seperti pada Tabel 2.

**Tabel 2.** Klasifikasi Jenis Kelamin

Jenis Kelamin	Klasifikasi
Laki-laki	L
Perempuan	P

#### 2) Mengelompokkan Daerah Asal

Variabel daerah asal dikelompokkan menjadi dua kelas: 'Dalam Provinsi' untuk mahasiswa dari Sumatera Barat dan 'Luar Provinsi' untuk mahasiswa dari luar Sumatera Barat, seperti Sumatera Utara, Riau, dan Jambi. Kelas ini disingkat sesuai dengan Tabel 3.

**Tabel 3.** Daerah Asal

Daerah Asal	Klasifikasi
Dalam Provinsi	DP
Luar Provinsi	LP

3) Mengelompokkan berdasarkan Program Studi,

**Tabel 4.** Klasifikasi Program Studi

Program Studi	Klasifikasi
Keperawatan	Kep
Kesehatan Masyarakat	IKM

4) Mengelompokkan IPK,

Data IPK saat ini belum memiliki pengklasifikasian kelas. Oleh karena itu, perlu dilakukan penyederhanaan variabel menjadi beberapa kategori. Untuk mempermudah pengklasifikasian IPK, rumus Sturges digunakan untuk membagi IPK ke dalam beberapa kelas. Berikut adalah perhitungan dari rumus *Sturges* :

$$\begin{aligned}
 \text{Jumlah Kelas} &= 1 + 3.3 \log n \text{ (jumlah data)} \\
 &= 1 + 3.3 \log (46) \\
 &= 1 + 3.3 (1,652) \\
 &= 6,487 \text{ atau } 6
 \end{aligned}$$

Untuk mendapatkan *Range*, maka dilakukan perhitungan sebagai berikut:

$$\begin{aligned}
 \text{Range} &= \text{IPK Tertinggi} - \text{IPK Terendah} \\
 &= 3,97 - 2,13 \\
 &= 0,84.
 \end{aligned}$$

Selanjutnya, hasil perhitungan *range* (selisih) dibagi jumlah kelas.

$$0,84 / 6 = 0,14$$

Maka *range* untuk klasifikasi IPK adalah 0,13, dimulai dari IPK tertinggi sampai IPK terendah pada data. Untuk lebih jelas dapat dilihat pada tabel 5.

**Tabel 5.** Klasifikasi IPK

IPK	Klasifikasi
3,13-3,27	A
3,27-3,41	B
3,41-3,55	C
3,55-3,69	D
3,69-3,83	E
3,83-3,93	F

Setelah melakukan pengelompokan atau pengklasifikasian berdasarkan atribut, maka terbentuklah *format data training* yang dapat dilihat pada tabel 6.

**Tabel 6.** *Format Data Training*

No	NIM	Jenis Kelamin	Daerah Asal	Program Studi	IPK
1	1802043	L	DP	Kep	C
2	1802044	P	LP	Kep	D
3	1802046	P	LP	Kep	C
⋮					
44	1703022	P	DP	IKM	C
45	1703033	P	DP	IKM	C
46	1703038	P	DP	IKM	C

Data training pada Tabel 6 diperoleh dari atribut yang telah diklasifikasikan. Misalnya, 'Laki-Laki' menjadi 'L', 'Jambi' menjadi 'LP' (Luar Provinsi), 'Keperawatan' menjadi 'Kep', dan IPK 3,61 menjadi 'C'. Variabel Jenis Kelamin, Daerah Asal, dan Program Studi sudah terklasifikasi, sehingga tidak memerlukan pembagian kelas lebih lanjut. Namun, variabel IPK memerlukan pengklasifikasian menggunakan rumus Sturges, yang membagi IPK menjadi 6 kelas.

### 3.3. Metode Klasifikasi

#### a) Klasifikasi Algoritma C4.5

Dalam Algoritma C4.5, langkah awal adalah menghitung nilai Entropy dan Gain untuk menentukan akar pohon keputusan. Dengan menggunakan 46 kasus dari Tabel 6 Data Training, pohon keputusan dibangun untuk memprediksi Jenis Kelamin, Daerah Asal, Program Studi, dan IPK. Atribut dengan nilai Gain tertinggi dipilih sebagai akar.

#### b) Pohon Keputusan

Pohon Keputusan adalah model prediksi berbentuk hierarki. Proses pembuatannya dimulai dengan menghitung jumlah kasus "Tepat" dan "Terlambat" serta mengelompokkan kasus berdasarkan atribut seperti Jenis Kelamin, Daerah Asal, Program Studi, dan IPK. Setelah itu, Nilai Gain untuk setiap atribut dihitung. Tahapan pembuatannya adalah sebagai berikut.

1) Menyipakan data *training*.

Data training merupakan data yang sudah dikelompokkan ke dalam kelas-kelas tertentu seperti pada tabel 4.6 Format Data *Training*.

2) Menentukan akar (*root*) dari pohon dan menghitung nilai *Gain* atribut.

Akar diambil dari atribut dengan nilai gain tertinggi, dihitung setelah menentukan nilai entropy. Langkah-langkahnya meliputi perhitungan Entropy dan Gain untuk setiap atribut.

Menghitung Nilai *Entropy* tiap-tiap atribut :

$$Entropy(total) = \left( -\frac{31}{46} * \log_2 \left( \frac{31}{46} \right) \right) + \left( -\frac{15}{46} * \log_2 \left( \frac{15}{46} \right) \right) = 0,9108$$

*Entropy* Jenis Kelamin

$$Entropy(L) = \left( -\frac{6}{8} * \log_2 \left( \frac{6}{8} \right) \right) + \left( -\frac{2}{8} * \log_2 \left( \frac{2}{8} \right) \right) = 0,8112$$

$$Entropy(P) = \left( -\frac{25}{38} * \log_2 \left( \frac{25}{38} \right) \right) + \left( -\frac{13}{38} * \log_2 \left( \frac{13}{38} \right) \right) = 0,9268$$

*Entropy* Daerah Asal

$$Entropy(DP) = \left( -\frac{21}{33} * \log_2 \left( \frac{21}{33} \right) \right) + \left( -\frac{12}{33} * \log_2 \left( \frac{12}{33} \right) \right) = 0,9456$$

$$Entropy(LP) = \left( -\frac{10}{13} * \log_2 \left( \frac{10}{13} \right) \right) + \left( -\frac{3}{13} * \log_2 \left( \frac{3}{13} \right) \right) = 0,7793$$

*Entropy* Program Studi

$$Entropy(Kep) = \left( -\frac{21}{26} * \log_2 \left( \frac{21}{26} \right) \right) + \left( -\frac{5}{26} * \log_2 \left( \frac{5}{26} \right) \right) = 0,7062$$

$$Entropy(IKM) = \left( -\frac{10}{13} * \log_2 \left( \frac{10}{13} \right) \right) + \left( -\frac{10}{13} * \log_2 \left( \frac{10}{13} \right) \right) = 1$$

*Entropy* IPK

$$Entropy(A) = \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) + \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) = 0$$

$$Entropy(B) = \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) + \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) = 0$$

$$Entropy(C) = \left( -\frac{23}{33} * \log_2 \left( \frac{23}{33} \right) \right) + \left( -\frac{10}{33} * \log_2 \left( \frac{10}{33} \right) \right) = 0,8849$$

$$Entropy(D) = \left( -\frac{4}{5} * \log_2 \left( \frac{4}{5} \right) \right) + \left( -\frac{1}{5} * \log_2 \left( \frac{1}{5} \right) \right) = 0,7219$$

$$Entropy(E) = \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) + \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) = 0$$

$$Entropy(F) = \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) + \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) = 0$$

Setelah melakukan perhitungan untuk mendapatkan nilai *Entropy* pada tiap-tiap atribut, selanjutnya yaitu menghitung nilai *Gain* pada masing-masing atribut :

*Gain* (Total, Jenis Kelamin)

$$= Entropy(S) - \sum_{i=1}^n \frac{|Jenis\ Kelamin_i|}{|Total|} * Entropy(Jenis\ Kelamin_i)$$

$$= 0,924819 - \left( \left( \frac{8}{46} * 0,8112 \right) + \left( \frac{38}{46} * 0,9268 \right) \right) = 0,0041$$

*Gain* (Total, Daerah Asal)

$$= Entropy(S) - \sum_{i=1}^n \frac{|Daerah\ Asal_i|}{|Total|} * Entropy(Daerah\ Asal_i)$$

$$= 0,924819 - \left( \left( \frac{33}{46} * 0,9456 \right) + \left( \frac{13}{46} * 0,7793 \right) \right) = 0,0122$$

*Gain* (Total, Program Studi)

$$= Entropy(S) - \sum_{i=1}^n \frac{|Program\ Studi_i|}{|Total|} * Entropy(Program\ Studi_i)$$

$$= 0,9945 - \left( \left( \frac{26}{46} * 0,7062 \right) + \left( \frac{20}{46} * 1 \right) \right) = 0,0769$$

*Gain* (Total, IPK)

$$= Entropy(S) - \sum_{i=1}^n \frac{|IPK_i|}{|Total|} * Entropy(IPK_i)$$

$$= 0,9945 - \left( \left( \frac{0}{46} * 0 \right) + \left( \frac{0}{46} * 0 \right) + \left( \frac{33}{46} * 0,8849 \right) + \left( \frac{5}{46} * 0,7219 \right) \right. \\ \left. + \left( \frac{2}{46} * 0 \right) + \left( \frac{2}{46} * 0 \right) \right) = 0,1975$$

Setelah melakukan perhitungan untuk mendapatkan nilai *Entropy* dan *Gain* dari masing-masing atribut. Langkah berikutnya adalah melanjutkan proses tersebut hingga semua variabel terlibat dalam pembuatan pohon keputusan proses ini dilakukan berulang hingga pohon keputusan terbentuk sepenuhnya.

### c) Rule Hasil Prediksi Algoritma C4.5

Rule hasil prediksi yang diperoleh berdasarkan pohon keputusan Node 4 yang merupakan pohon terakhir yang terbentuk, diperoleh 9 aturan (*rule*) dalam memprediksi ketepatan waktu kelulusan mahasiswa. Aturan atau kaidah yang diperoleh dari pohon keputusan adalah sebagai berikut :

1. IF IPK = (3,13 - 3,26), THEN Mahasiswa Lulus = Terlambat.
2. IF IPK = (3,27 - 3,40), THEN Mahasiswa Lulus = Terlambat.
3. IF IPK = (3,41 - 3,54), AND Program Studi = Kep, THEN Mahasiswa Lulus = Tepat Waktu.
4. IF IPK = (3,41 - 3,54), AND Program Studi = IKM, AND Jenis Kelamin = Laki-Laki, THEN Mahasiswa Lulus = Tepat Waktu.
5. IF IPK = (3,41 - 3,54), AND Program Studi = IKM, AND Jenis Kelamin = Perempuan, AND Daerah Asal = Dalam Provinsi, THEN Mahasiswa Lulus = Terlambat.
6. IF IPK = (3,41 - 3,54), AND Program Studi = IKM, AND Jenis Kelamin = Perempuan, AND Daerah Asal = Luar Provinsi, THEN Mahasiswa Lulus = Tepat Waktu.
7. IF IPK = (3,55 - 3,68), THEN Mahasiswa Lulus = Tepat Waktu.
8. IF IPK = (3,69 - 3,82), THEN Mahasiswa Lulus = Tepat Waktu.
9. IF IPK = (2,83 - 3,97), THEN Mahasiswa Lulus = Tepat Waktu.

### 3.4. Menghitung Accuracy

Pada perhitungan akurasi Algoritma C4.5, diperoleh akurasi 89,13% dengan 41 data diklasifikasikan benar dari 46 data.

## 4. Kesimpulan

Algoritma C4.5 mencapai akurasi 89,13% dalam memprediksi kelulusan mahasiswa tepat waktu di Universitas Syedza Saintika, menjadikannya metode yang andal. Indeks Prestasi Kumulatif (IPK) muncul sebagai faktor paling dominan dan menjadi akar pohon keputusan, menunjukkan pentingnya pencapaian akademik dalam menentukan kelulusan. Selain IPK, atribut lain seperti program studi, jenis kelamin, dan daerah asal juga memengaruhi prediksi, meski dengan tingkat yang lebih rendah. Pohon keputusan menghasilkan 9 aturan, yang semuanya bermula dari IPK sebagai faktor penentu utama dalam analisis kelulusan.

## Daftar Pustaka

- [1] I. W. M. Putri, R. Rusdah, L. Suryadi, and D. Anubhakti, "Prediksi Kelulusan Mahasiswa Fakultas Teknologi Informasi ISB Atma Luhur Menggunakan Algoritma C4.5," *J. SISFOKOM (Sistem Inf. dan Komputer)*, vol. 12, no. 03, pp. 363–369, 2023.
- [2] F. M. Almufqi and A. Voutama, "Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa," *J. Tek.*, vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [3] J. Donga, A. Sarungalo, N. Koru, and G. Lante, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- [4] S. Lestari, Yulmaini, Aswin, Sylvia, Y. A. Pratama, and Sulyono, "Implementation of the C4.5 algorithm for micro, small, and medium enterprises classification," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 6, pp. 6707–6715, 2022, doi: 10.11591/ijece.v12i6.pp6707-6715.
- [5] N. H. Purnomo, B. Pamungkas, and C. Juliane, "Penerapan Algoritma C4 . 5 Untuk Klasifikasi Tren Pelanggaran Kendaraan Angkutan Barang dengan Metode CRISP-DM," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 30–40, 2023, doi: 10.30865/mib.v7i1.5247.
- [6] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *Skanika*, vol. 6, no. 1, pp. 61–74, 2023, doi: 10.36080/skanika.v6i1.2976.
- [7] E. Prasetyaningrum and P. Susanti, "Analisa Tingkat Kepuasan Pelanggan Pada Percetakan Cv. Mega Media Menggunakan Algoritma C4.5," *Sisfotenika*, vol. 13, no. 1, pp. 65–75, 2023.
- [8] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 193–199, 2023, doi: 10.37859/coscitech.v4i1.4755.
- [9] N. A. Prahastiwi, R. Andreswari, and R. Fauzi, "Students Graduation Prediction Based on Academic Data Record Using the Decision Tree Algorithm C4.5 Method," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 8, no. 3, pp. 295–304, 2022, doi: 10.33330/jurteks.v8i3.1680.
- [10] K. A. Putri, D. Febriawan, and F. N. Hasan, "Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm ( C4 . 5 )," vol. 13, pp. 39–47, 2024.
- [11] S. Sains, P. Kelulusan, M. Di, P. Kampar, A. Saputra, and T. A. Fitri, "Penerapan Data Mining Algoritma C4 . 5 Dalam Memprediksi," 2023.

- [12] N. Khasanah and A. Salim, "Rachman Komarudin 4) , Yana Iqbal Maulana 5) 1) Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 2,3) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika 4) Sistem Informasi, Fakultas Teknologi Inf," *Fak. Teknol. Inf.*, vol. 13, no. 3, pp. 207–214, 2022.
- [13] A. F. Firdaus, R. Saedudin, R. Andeswari, and U. Telkom, "Implementasi Metode Klasifikasi Naive Bayes Implementation of Naive Bayes Classification Method in Predicting," vol. 8, no. 5, pp. 9274–9279, 2021.
- [14] P. Pangestu, R. Novita, and M. Mustakim, "Systematic Literature Review: Perbandingan Algoritma Klasifikasi," *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 2, p. 431, 2023, doi: 10.35314/isi.v8i2.3698.
- [15] N. Widiastuti, A. Hermawan, and D. Avianto, "Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Minat Pencari Kerja," *J. Teknoinfo*, vol. 17, no. 1, pp. 1–9, 2023.
- [16] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [17] A. Purwanto and H. Widi Nugroho, "Jurnal Teknoinfo," *Tong Sampah Pint. Dengan Perintah Suara Guna Menghilangkan Perilaku Siswa Membuang Sampah Sembarangan Di Sekol.*, vol. 14, pp. 48–58, 2023, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/article/view/336/329>
- [18] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, and Shelly Janu Setyaning Tyas, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *Tematik*, vol. 8, no. 1, pp. 96–103, 2021, doi: 10.38204/tematik.v8i1.576.
- [19] L. Tastilia, D. A. Megawaty, and A. Sulistiyawati, "Sistem Informasi Administrasi Akademik Untuk Meningkatkan Pelayanan Terhadap Siswa (Study Kasus : Sma Pgri Katibung)," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 63–69, 2022.