

## Seleksi Fitur Information Gain dan Teknik Pruning untuk Memperbaiki Akurasi Algoritma C4.5 dalam Kasus Keterlambatan Biaya Kuliah

Taghfirul Azhima Yoga Siswa<sup>1)</sup> Gubtha Mahendra Putra<sup>2)</sup> Anton Prafanto<sup>3)</sup>

<sup>1)</sup>Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur  
Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur, 75124

<sup>2,3)</sup>Informatika, Fakultas Teknik, Universitas Mulawarman

Jl. Sambaliung No.7, Sempaja Selatan, Kec. Samarinda Utara, Kota Samarinda, Kalimantan Timur, 75242

E-Mail : [tay758@umkt.ac.id](mailto:tay758@umkt.ac.id)<sup>1)</sup>, [gubthamp@fkti.unmul.ac.id](mailto:gubthamp@fkti.unmul.ac.id)<sup>2)</sup>, [antonprafanto@fkti.unmul.ac.id](mailto:antonprafanto@fkti.unmul.ac.id)<sup>3)</sup>

### ABSTRAK

Penerapan biaya kuliah memiliki peranan yang sangat penting di suatu universitas untuk dapat meningkatkan mutu dan infrastruktur pendidikan khususnya di Universitas Muhammadiyah Kalimantan Timur (UMKT). Namun, dalam pelaksanaannya masih banyak mahasiswa yang terlambat dalam melakukan pembayaran biaya kuliah. Hal ini dapat mengganggu UMKT dalam sisi operasional dan pelaksanaan peningkatan mutu serta infrastruktur. Pada penelitian ini akan dilakukan penentuan fitur, penerapan algoritma C4.5, dan evaluasi kinerja algoritma C4.5 dengan menggunakan *confusion matrix* pada pembagian data 90% *data training* dan 10% *data testing*. Untuk mengoptimasi kinerja algoritma C4.5, pada penelitian ini akan diterapkan seleksi fitur menggunakan metode *information gain* dan *pruning*. Penelitian ini menggunakan data yang diperoleh dari Biro Administrasi Keuangan dan Biro Administrasi Akademik UMKT dengan jumlah data sebanyak 12.408. Hasil pengujian kinerja algoritma C4.5 tanpa menggunakan seleksi fitur *information gain* dan teknik *pruning* memperoleh nilai akurasi sebesar 61,40%. Adapun hasil pengujian kinerja algoritma C4.5 dengan menggunakan seleksi fitur *information gain* dan teknik *pruning* memperoleh hasil sebesar 64,86%. Hasil pengujian kinerja algoritma C4.5 dengan menggunakan seleksi fitur *information gain* dan teknik *pruning* terbukti mampu meningkatkan kinerja algoritma sebesar 3,45% pada kasus keterlambatan biaya kuliah.

Kata Kunci – Data mining, Klasifikasi, C4.5, information gain, pruning, optimasi, keterlambatan biaya kuliah

### 1. PENDAHULUAN

Di dalam sebuah perguruan tinggi, mahasiswa diwajibkan untuk memenuhi pembiayaan Sumbangan Pembangunan Pendidikan atau SPP yang telah disepakati ketetapanannya, khususnya di Universitas Muhammadiyah Kalimantan Timur (UMKT). SPP berperan penting bagi perguruan tinggi untuk meningkatkan kualitas dan mutu serta infrastruktur pendidikan di UMKT. Namun, jika dalam praktiknya mahasiswa terlambat dalam melakukan pembayaran SPP, UMKT harus menanggung beberapa kerugian seperti tertundanya gaji pegawai maupun dosen hingga pengembangan infrastruktur pendidikan yang terhambat. Dari permasalahan tersebut, maka dibutuhkan sebuah analisis data pembiayaan SPP mahasiswa yang dapat digunakan untuk membantu menangani pembayaran SPP mahasiswa yang terindikasi akan mengalami keterlambatan. Dengan analisis data keterlambatan pembayaran SPP mahasiswa tersebut diharapkan dapat membantu UMKT khususnya Biro Administrasi Keuangan dalam mengatur keuangan UMKT lebih baik lagi serta dapat menjadi rekomendasi positif bagi para petinggi UMKT mengevaluasi mahasiswa yang mengalami keterlambatan dalam melakukan pembayaran biaya kuliah.

Dalam penelitian Ginting, Kusri, & Taufiq (2020) telah dilakukan pembahasan penelitian mengenai prediksi keterlambatan pembayaran sumbangan pendidikan sekolah dengan menggunakan algoritma C4.5. Pada penelitian

tersebut diperoleh hasil dengan variabel penghasilan orang tua, tanggungan keluarga, pendidikan ayah, pendidikan ibu, dan keterangan dengan hasil akurasi algoritma C4.5 sebesar 73%. Selanjutnya, pada penelitian Apandi dkk. (2019), mereka menganalisis keterlambatan pembayaran sumbangan pembangunan pendidikan dengan menggunakan algoritma C4.5. Pengujian dilakukan pada data yang dikumpulkan melalui kuesioner yang diberikan kepada mahasiswa program studi Teknik Informatika pada semester 3, 5, dan 7, pada kelas reguler tahun ajaran 2015/2016. Hasil penelitian menunjukkan bahwa algoritma C4.5 mencapai tingkat akurasi tertinggi sebesar 75%. Untuk dapat mendukung UMKT dalam menganalisis data keterlambatan biaya kuliah, dalam penelitian ini akan digunakan algoritma C4.5 dengan tambahan metode *information gain* dan teknik *pruning* untuk mengoptimasi kinerja algoritma C4.5.

Pada penelitian yang dilakukan oleh Desiani, Yahdin, & Rodiah (2020), mereka membandingkan metode klasifikasi antara algoritma C4.5 dan algoritma *Naïve Bayes*. Hasil penelitian tersebut mengungkapkan bahwa kinerja algoritma C4.5 secara signifikan lebih baik daripada kinerja algoritma *Naïve Bayes*. Pada penelitian sebelumnya (Junita & Bachtiar, 2019), kombinasi algoritma C4.5 dan seleksi fitur menggunakan metode *information gain* dilakukan dengan perolehan hasil akurasi sebesar 81%. Kemudian pada penelitian Iskandar, Hiryanto, & Hendryli (2018) memperoleh hasil bahwa dengan

menggabungkan algoritma C4.5 dan teknik *pruning*, pohon keputusan dapat disederhanakan dan hasilnya dapat meningkatkan akurasi algoritma.

Dari beberapa penelitian yang telah diungkapkan menunjukkan bahwa adanya perbedaan dalam penelitian ini ketika menggunakan algoritma C4.5, yakni dengan kombinasi penggunaan seleksi fitur berdasarkan *information gain* dan teknik *pruning*, dengan tujuan meningkatkan tingkat kinerja dan akurasi algoritma.

Untuk meningkatkan akurasi algoritma C4.5 yang akan dilakukan dalam penelitian ini, terdapat beberapa langkah yang akan dilakukan. Pertama, fitur yang tidak relevan dihilangkan karena tidak memiliki hubungan yang jelas dengan atribut label atau target. Hal ini membantu menghindari *overfitting* dan mengurangi jumlah data yang digunakan, sehingga waktu pelatihan menjadi lebih cepat (Daqiqil, 2021). Selain itu, teknik *pruning* digunakan dalam *decision tree* untuk memotong cabang yang terlalu panjang dan menghindari *overfitting*. Node-node pada tingkat atas diubah menjadi leaf node, yang pada akhirnya mempercepat waktu pelatihan dan memberikan hasil yang lebih akurat (Kurniawan, 2022). Demikian pula, penggunaan teknik *pruning* telah terbukti efektif dalam meningkatkan akurasi algoritma C4.5 (Iskandar, Hiryanto, & Hendryli, 2018). Oleh karena itu, dalam penelitian ini, akan digunakan kombinasi seleksi fitur *information gain* dan teknik *pruning* untuk meningkatkan akurasi algoritma C4.5. Pendekatan ini memungkinkan penggunaan atribut yang lebih sedikit namun memiliki hubungan yang jelas dengan atribut label atau target, dan menghasilkan pohon keputusan yang lebih kecil.

## 2. TINJAUAN PUSTAKA

### A. Keterlambatan Biaya Kuliah

Keterlambatan mahasiswa dilakukan dengan tujuan untuk mengidentifikasi pola klasifikasi yang mengindikasikan apakah mahasiswa akan tepat waktu atau terlambat dalam pembayaran biaya kuliah, serta dalam prosesnya indikator mana yang memiliki pengaruh paling signifikan juga dianalisis. (Apandi dkk., 2019). Rekomendasi mengenai kemungkinan mahasiswa yang akan terlambat membayar SPP dapat dilakukan dengan menggunakan teknik data mining, khususnya metode klasifikasi, dimana hasil dari proses klasifikasi ini akan menjadi dasar untuk melakukan prediksi pembayaran SPP pada semester berikutnya (Rohmayani, 2020).

### B. Data Mining

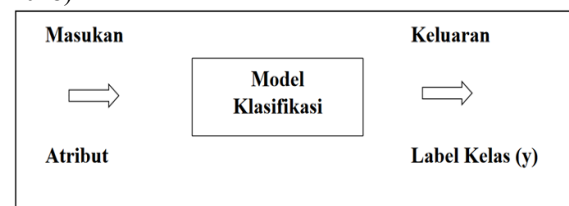
Menurut Hermawati (2013), *data mining* adalah suatu proses yang menggunakan satu atau lebih teknik *machine learning* atau pembelajaran mesin untuk secara otomatis menganalisis dan mengekstraksi pengetahuan dari data. Pendapat lain mengenai pengertian *data mining* diberikan oleh Mardi (2017), yang menyatakan bahwa data mining adalah suatu proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan

pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan terkait dari berbagai basis data besar.

Berdasarkan definisi-definisi tersebut, dapat disimpulkan bahwa *data mining* melibatkan pengambilan data dalam skala besar untuk mengidentifikasi dan mengekstraksi informasi yang bermanfaat dari kumpulan data menggunakan teknik-teknik pembelajaran komputer atau *machine learning*. Beberapa teknik data mining yang umum digunakan termasuk klusterisasi (*clustering*), regresi (*regression*), klasifikasi (*classification*), dan kaidah asosiasi (*association rule*) (Hermawati, 2013).

### C. Klasifikasi

Menurut Hermawati (2013), klasifikasi adalah suatu proses pembelajaran yang melibatkan fungsi tujuan (target)  $f$  yang memetakan setiap himpunan atribut  $x$  ke label kelas tertentu yang telah didefinisikan sebelumnya. Dalam klasifikasi, objek-objek dikelompokkan berdasarkan ciri-ciri yang dimiliki, dan proses ini dapat dilakukan secara manual atau dengan bantuan teknologi (Wibawa, 2018).



Gambar 1. Konsep Klasifikasi

### D. Algoritma C4.5

Menurut Berry & Linoff (2018) sebagaimana dikutip dalam buku Nofriansyah & Nurcahyo (2015), algoritma C4.5 merupakan salah satu solusi yang umum digunakan dalam pemecahan masalah klasifikasi. Algoritma C4.5 menghasilkan *decision tree* sebagai output, mirip dengan teknik klasifikasi lainnya. Nofriansyah & Nurcahyo (2015) menjelaskan bahwa terdapat dua elemen penting dalam penyelesaian kasus dengan algoritma C4.5, yaitu *entropy* dan *gain*. *Entropy* ( $S$ ) merupakan ukuran jumlah bit yang diperlukan untuk mengekstraksi kelas (+ atau -) dari sejumlah data acak dalam sampel  $S$ . *Entropy* dapat diartikan sebagai tingkat ketidakmurnian  $S$ . Rumus digunakan untuk menghitung nilai *entropy* adalah sebagai berikut:

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

Keterangan :

S = himpunan kasus

A = atribut

n = jumlah partisi S

$p^i$  = proporsi dari  $S^i$  terhadap S

*Gain* ( $S, A$ ) adalah ukuran informasi yang diperoleh dari atribut  $A$  dalam hubungannya dengan data output  $S$ . Informasi tersebut diperoleh dari

kelompok-kelompok output atau variabel dependen  $S$  yang diorganisasikan berdasarkan atribut  $A$ , yang dinyatakan dengan  $gain(S, A)$ . Berikut adalah rumus yang digunakan untuk menghitung nilai  $gain$ ..

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

Dimana :

A = atribut

S = sampel

n = jumlah partisi himpunan atribut a

|S<sub>i</sub>| = jumlah sampel pada partisi ke-i

|S| = jumlah sampel dalam S

### E. Information Gain

Menurut Jensen & Shen (2008), *information gain* merupakan pengurangan atau reduksi yang diharapkan dalam *entropy* (sebagai ukuran keacakan atau ketidakteraturan informasi yang sedang diproses dalam pembelajaran mesin) yang terjadi ketika objek-objek dalam kumpulan data dipartisi berdasarkan suatu fitur tertentu. Rumus perhitungan *information gain* adalah sebagai berikut:

$$IG(S, A) = Entropy(S) - \sum_{c \in \text{values}(A)} |S_v| |S| Entropy(S_v)$$

Keterangan:

S = Himpunan kasus

A = Atribut

|S<sub>i</sub>| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

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

Keterangan:

S = Himpunan Kasus

c = Jumlah Partisi

### F. Cross Validation

Menurut Nurhayati dkk. (2014), *cross validation* adalah sebuah metode yang digunakan untuk memperkirakan kesalahan prediksi sebagai evaluasi kinerja model. Dalam *cross validation*, terdapat metode yang dikenal sebagai estimasi rotasi, di mana data dibagi menjadi himpunan bagian  $k$  dengan ukuran yang hampir sama. Pada setiap iterasi, model dalam klasifikasi dilatih dan diuji sebanyak  $k$ . Dalam setiap iterasi tersebut, salah satu himpunan bagian digunakan sebagai data pengujian, sedangkan himpunan sub kelompok data yang lain berfungsi sebagai data latih.



Gambar 2. Ilustrasi 10-Fold Cross Validation

### G. Confusion Matrix

*Confusion matrix* adalah suatu matriks berukuran  $N \times N$  yang digunakan dalam permasalahan klasifikasi, di mana  $N$  adalah jumlah kelas yang akan diprediksi (Daqiqil, I, 2021). Pada tahap ini, diterapkan aturan yang diilustrasikan dalam Gambar 3, dengan melakukan perhitungan menggunakan tiga metrik yaitu *precision*, *recall*, dan *f-score* yang diperkenalkan oleh Baeza-Yates & Ribeiro-Neto pada tahun 1999.

Class	Actual = Yes	Actual = No
Predicted = Yes	TP	FP
Predicted = No	FN	TN

Gambar 3. Confusion Matrix

Terdapat beberapa istilah confusion matrix yang biasa digunakan pada kasus klasifikasi yaitu (Pulungan, 2019):

- True Positive (TP) confusion matrix* pada data positif yang terdeteksi benar
- False Positive (FP) confusion matrix* pada data negatif namun terdeteksi dengan benar
- False Negative (FN) confusion matrix* pada data positif yang terdeteksi sebagai data negatif
- True Negative (TN) confusion matrix* pada data negatif yang terdeteksi benar.

Beberapa perhitungan kinerja klasifikasi dapat dijelaskan pada perhitungan *confusion matrix* berikut (Sokolova & Lapalme, 2009):

#### 1. Accuracy

Nilai efektivitas keseluruhan dari proses klasifikasi

$$accuracy = \frac{TP + TN}{n}$$

#### 2. Precision

Tingkat ketepatan kelas label data dengan label positif yang diberikan oleh pengklasifikasi

$$precision = \frac{TP}{TP + FP}$$

#### 3. Recall

Tingkat efektivitas untuk mengidentifikasi label positif pada proses klasifikasi.

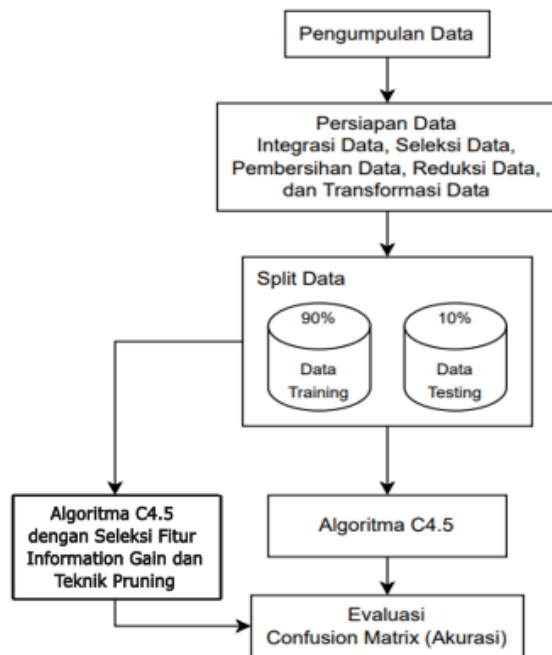
$$recall = \frac{TP}{TP + FN}$$

### 3. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen sebagai pendekatan penelitian. Menurut Adnan & Latief (2020), metode penelitian eksperimen merupakan jenis penelitian kuantitatif di mana peneliti melakukan manipulasi terhadap satu atau lebih variabel bebas (*independent variable*), mengendalikan variabel lain yang relevan, dan mengamati efek dari manipulasi tersebut pada variabel terikat (*dependent variable*). Eksperimen ini secara sengaja dan sistematis mengenalkan perubahan tertentu dan mengamati konsekuensinya.

Hanya masalah penelitian tertentu yang memungkinkan peneliti untuk memanipulasi kondisi yang spesifik dalam penelitian eksperimental.

Pada penelitian ini, fokusnya adalah pada prediksi keterlambatan pembayaran biaya kuliah menggunakan pendekatan algoritma C4.5. Selanjutnya, dilakukan peningkatan akurasi menggunakan seleksi fitur *information gain* dan teknik *pruning*. Tahapan penelitian ini dapat dilihat dalam Gambar 4.



Gambar 4. Alur Tahapan Penelitian

#### 4. HASIL DAN PEMBAHASAN

##### A. Pengumpulan Data

Data dikumpulkan melalui observasi di Biro Administrasi Keuangan, dan jumlah data yang diperoleh adalah sebanyak 39.644 data mahasiswa yang membayar biaya kuliah di Universitas Kalimantan Timur pada periode 2019 hingga 2021. Data tersebut mencakup atribut NIM, nama, dan keterangan. Selanjutnya, untuk melengkapi data mahasiswa yang membayar biaya kuliah, data juga diperoleh dari Biro Administrasi Akademik dengan jumlah 10.959 data diri mahasiswa dari angkatan 2017 hingga 2021. Data ini mencakup atribut NIM, nama, fakultas, program studi, angkatan, jenis kelamin, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu.

##### B. Persiapan Data

Dalam persiapan data, dilakukan berbagai langkah untuk mengatasi masalah yang terdapat dalam dataset. Langkah-langkah tersebut meliputi integrasi data, seleksi data, pembersihan data, reduksi data, dan transformasi data.

##### 1) Integrasi Data

Integrasi data dilakukan dengan menggabungkan data dari Biro Administrasi Keuangan dan menggunakan atribut NIM untuk mengambil data

dari Biro Administrasi Akademik. Hal ini menghasilkan data yang baru yang terdiri dari 11 atribut. Detail atribut tersebut dapat ditemukan pada Tabel 1.

Tabel 1. Hasil Seleksi dan Integrasi Data

No	Atribut	Jenis Data
1	Nim	Karakter
2	Nama	Karakter
3	Fakultas	Kategorikal
4	Prodi	Kategorikal
5	Angkatan	Kategorikal
6	Gender	Kategorikal
7	Penghasilan Ayah	Kategorikal
8	Penghasilan Ibu	Kategorikal
9	Pendidikan Ayah	Kategorikal
10	Pendidikan Ibu	Kategorikal
11	Keterangan	Kategorikal

##### 2) Seleksi Data

Proses seleksi data dilakukan dengan menghapus atribut yang tidak diperlukan, seperti NIM dan nama, karena atribut tersebut hanya digunakan saat melakukan penggabungan data.

##### 3) Pembersihan Data

Tahap pembersihan data melibatkan penghapusan data yang tidak lengkap atau bermasalah, karena penting untuk memiliki dataset yang bersih sebelum dilakukan analisis agar hasilnya akurat.

##### 4) Reduksi Data

Reduksi data dilakukan pada dataset karena ketidakseimbangan jumlah kelas target. Oleh karena itu, dilakukan proses penyeimbangan dengan mengurangi sebagian data dari kelas target yang jumlahnya lebih banyak, sehingga dataset menjadi lebih kecil namun tetap mempertahankan integritas data asli.

##### 5) Transformasi Data

Transformasi data diperlukan dalam algoritma *machine learning* karena membutuhkan data numerik sebagai data pelatihan. Pada tahap ini, atribut yang semula berjenis kategorikal diubah menjadi jenis numerik, kecuali atribut keterangan yang digunakan sebagai label atau target. Transformasi data ini diperlukan agar proses pemodelan dapat berjalan dengan baik. Rincian transformasi data dapat dilihat pada Tabel 2.

Tabel 2. Hasil Transformasi Data

No.	Atribut	Keterangan Nilai
1	Fakultas	0. Ekonomi Bisnis dan Politik
		1. Farmasi
		2. Hukum
		3. Ilmu Keperawatan
		4. Keguruan dan Ilmu Pendidikan
		5. Kesehatan Masyarakat
		6. Psikologi
		7. Sain dan Teknologi
2	Prodi	0. Farmasi
		1. Hubungan Internasional
		2. Hukum
		3. Keperawatan
		4. Kesehatan Masyarakat

		5. Manajemen
		6. Ners
		7. Pendidikan Bahasa Inggris
		8. Pendidikan Olah Raga
		9. Psikologi
		10. Teknik Informatika
		11. Teknik Mesin
		12. Teknik Sipil
3	Angkatan	0. 2017 1. 2018 2. 2019 3. 2020
4	Gender	0. L 1. P
5	Penghasilan Ayah	0. Kurang dari Rp.500,000 1. Rp. 500,000 – Rp. 999,999 2. Rp. 1,000,000 – Rp. 1.999,999 3. Rp. 2,000,000 – Rp. 4,999,999 4. Rp. 5,000,000 – Rp. 20,000,000 5. Lebih dari Rp. 20,000,000
6	Penghasilan Ibu	0. Kurang dari Rp.500,000 1. Rp. 500,000 – Rp. 999,999 2. Rp. 1,000,000 – Rp. 1.999,999 3. Rp. 2,000,000 – Rp. 4,999,999 4. Rp. 5,000,000 – Rp. 20,000,000 5. Lebih dari Rp. 20,000,000
7	Pendidikan Ayah	0. Tidak sekolah 1. TK 2. SD 3. SMP 4. SMA 5. D1 6. D2 7. D3 8. S1 9. S2 10. S3
8	Pendidikan Ibu	0. Tidak sekolah 1. TK 2. SD 3. SMP 4. SMA 5. D1 6. D2 7. D3

	8. S1
	9. S2
	10. S3
9	Keterangan Tepat dan Terlambat

### C. Split Data

Setelah melakukan persiapan data, digunakan total 12.408 data yang akan dibagi menjadi dua bagian. Sebanyak 90% data, yaitu 11.168 data, digunakan sebagai data pelatihan untuk membentuk model algoritma C4.5. Sementara itu, 10% data, yaitu 1.241 data, digunakan sebagai data pengujian untuk mengevaluasi kinerja model algoritma C4.5 yang telah terbentuk sebelumnya.

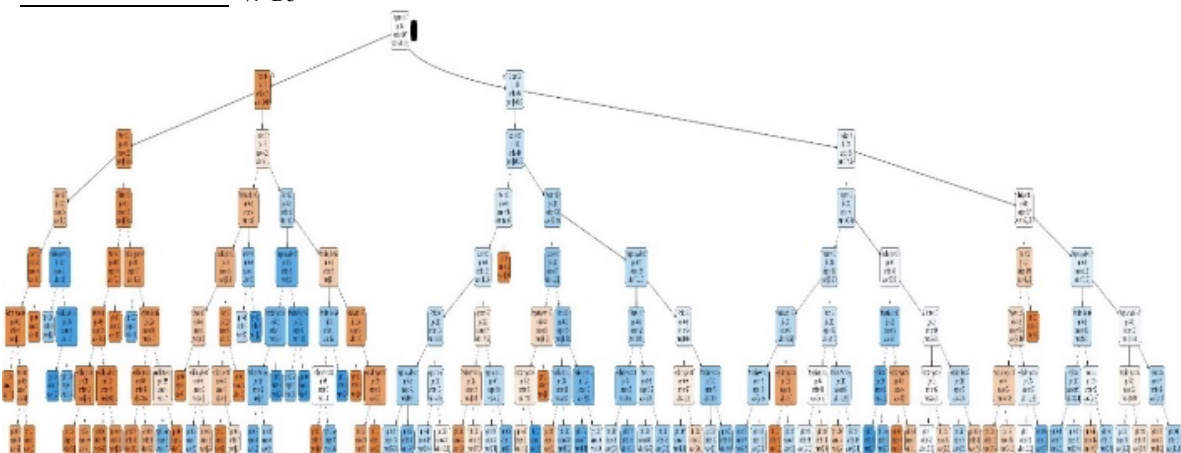
### D. Pemodelan

Proses perbandingan dan optimasi algoritma pemodelan klasifikasi dimulai dengan melakukan seleksi fitur terhadap dataset yang telah melalui persiapan data sebelumnya. Tahapan seleksi fitur dimulai dengan merankingkan pengaruh masing-masing fitur terhadap label, dengan peringkat 1 diberikan kepada fitur yang memiliki pengaruh tertinggi, sedangkan peringkat 8 diberikan kepada fitur dengan pengaruh terendah. Setelah memperoleh peringkat seleksi fitur, data kemudian dimasukkan ke dalam pemodelan klasifikasi C4.5.

Selanjutnya, algoritma C4.5 diterapkan dengan menggunakan teknik pruning untuk menghasilkan pohon keputusan yang lebih ringkas. Melalui teknik *pruning*, akurasi dari pemodelan diharapkan dapat ditingkatkan.

#### 1) Algoritma C4.5 Tanpa Seleksi Fitur dan Teknik Pruning

Hasil yang didapatkan menggunakan atribut fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu menggunakan algoritma C4.5 dalam pemrograman Python menghasilkan sebuah pohon keputusan yang terlihat pada Gambar 2.



Gambar 5. Visual Pohon Keputusan Tanpa Seleksi Fitur dan Teknik Pruning

Gambar 5 adalah hasil dari pohon keputusan yang telah terbentuk. Adapun penjelasan aturan pohon

keputusan adalah sebagai berikut:

- 1) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $\leq 1.50$ , Gender  $\leq 0.50$ ,



- Pendidikan Ibu  $\leq 9.00$ , maka mahasiswa tepat.
- 2) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $\leq 1.50$ , Gender  $\leq 0.50$ , Pendidikan Ibu  $> 9.00$ , Prodi  $\leq 0.50$ , maka mahasiswa tepat.
  - 3) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $\leq 1.50$ , Gender  $\leq 0.50$ , Pendidikan Ibu  $> 9.00$ , Prodi  $> 0.50$ , maka mahasiswa tepat.
  - 4) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $\leq 1.50$ , Gender  $> 0.50$ , maka mahasiswa tepat.
  - 5) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $> 1.50$ , Pendidikan Ayah  $\leq 5.00$ , maka mahasiswa terlambat.
  - 6) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $> 1.50$ , Pendidikan Ayah  $> 5.00$ , Pendidikan Ibu  $\leq 9.00$ , maka mahasiswa terlambat.
  - 7) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $\leq 2.50$ , Fakultas  $> 1.50$ , Pendidikan Ayah  $> 5.00$ , Pendidikan Ibu  $> 9.00$ , maka mahasiswa terlambat.
  - 8) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $\leq 4.50$ , Prodi  $\leq 4.50$ , Pendidikan Ibu  $\leq 6.50$ , maka mahasiswa tepat.
  - 9) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $\leq 4.50$ , Prodi  $\leq 4.50$ , Pendidikan Ibu  $> 6.50$ , maka mahasiswa tepat.
  - 10) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $\leq 4.50$ , Prodi  $> 4.50$ , maka mahasiswa tepat.
  - 11) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $> 4.50$ , Pendidikan Ayah  $\leq 1.00$ , maka mahasiswa terlambat.
  - 12) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $> 4.50$ , Pendidikan Ayah  $> 1.00$ , penghasilan Ibu  $\leq 4.50$ , pendidikan Ibu  $\leq 9.00$ , maka mahasiswa tepat.
  - 13) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $> 4.50$ , Pendidikan Ayah  $> 1.00$ , penghasilan Ibu  $\leq 4.50$ , Pendidikan Ibu  $> 9.00$ , maka mahasiswa tepat.
  - 14) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $> 4.50$ , Pendidikan Ayah  $> 1.00$ , penghasilan Ibu  $> 4.50$ , Pendidikan Ibu  $\leq 6.50$ , maka mahasiswa terlambat.
  - 15) Jika Angkatan  $\leq 0.50$ , Prodi  $\leq 8.50$ , Prodi  $> 2.50$ , Fakultas  $> 4.50$ , Pendidikan Ayah  $> 1.00$ , penghasilan Ibu  $> 4.50$ , Pendidikan Ibu  $> 6.50$ , maka mahasiswa tepat.
  - 16) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $\leq 7.50$ , Gender  $\leq 0.50$ , maka mahasiswa tepat.
  - 17) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $\leq 7.50$ , Gender  $> 0.50$ , Pendidikan Ayah  $\leq 6.50$ , maka mahasiswa terlambat.
  - 18) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $\leq 7.50$ , Gender  $> 0.50$ , Pendidikan Ayah  $> 6.50$ , maka mahasiswa tepat.
  - 19) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $> 7.50$ , Fakultas  $\leq 5.00$ , Pendidikan Ibu  $\leq 9.00$ , maka mahasiswa tepat.
  - 20) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $> 7.50$ , Fakultas  $\leq 5.00$ , Pendidikan Ibu  $> 9.00$ , maka mahasiswa tepat.
  - 21) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $\leq 4.00$ , Pendidikan Ibu  $> 7.50$ , Fakultas  $> 5.00$ , maka mahasiswa tepat.
  - 22) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $> 4.00$ , Prodi  $\leq 10.50$ , maka mahasiswa terlambat.
  - 23) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $\leq 11.50$ , Penghasilan Ibu  $> 4.00$ , Prodi  $> 10.50$ , maka mahasiswa terlambat.
  - 24) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $\leq 12.50$ , Penghasilan Ayah  $\leq 1.00$ , Pendidikan Ayah  $\leq 9.00$ , Pendidikan Ayah  $\leq 7.50$ , maka mahasiswa terlambat.
  - 25) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $\leq 12.50$ , Penghasilan Ayah  $\leq 1.00$ , Pendidikan Ayah  $\leq 9.00$ , Pendidikan Ayah  $> 7.50$ , maka mahasiswa terlambat.
  - 26) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $\leq 12.50$ , Penghasilan Ayah  $\leq 1.00$ , Pendidikan Ayah  $> 9.00$ , maka mahasiswa terlambat.
  - 27) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $\leq 12.50$ , Penghasilan Ayah  $> 1.00$ , Penghasilan Ayah  $\leq 3.50$ , maka mahasiswa terlambat.
  - 28) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $\leq 12.50$ , Penghasilan Ayah  $> 1.00$ , Penghasilan Ayah  $> 3.50$ , maka mahasiswa terlambat.
  - 29) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $\leq 6.50$ , Pendidikan Ibu  $\leq 5.00$ , Penghasilan Ibu  $\leq 1.00$ , maka mahasiswa terlambat.
  - 30) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $\leq 6.50$ , Pendidikan Ibu  $\leq 5.00$ , Penghasilan Ibu  $> 1.00$ , maka mahasiswa terlambat.
  - 31) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $\leq 6.50$ , Pendidikan Ibu  $> 5.00$ , maka mahasiswa terlambat.
  - 32) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $> 6.50$ , Pendidikan Ayah  $\leq 7.00$ , maka mahasiswa tepat.
  - 33) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $> 6.50$ , Pendidikan Ayah  $> 7.00$ , Penghasilan Ayah  $\leq 2.50$ , maka mahasiswa tepat.

- 34) Jika Angkatan  $\leq 0.50$ , Prodi  $> 8.50$ , Prodi  $> 11.50$ , Prodi  $> 12.50$ , Pendidikan Ibu  $> 6.50$ , Pendidikan Ayah  $> 7.00$ , Penghasilan Ayah  $> 2.50$ , maka mahasiswa tepat.
- 35) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $\leq 2.50$ , Prodi  $\leq 4.00$ , Penghasilan Ayah  $\leq 4.50$ , maka mahasiswa terlambat.
- 36) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $\leq 2.50$ , Prodi  $\leq 4.00$ , Penghasilan Ayah  $> 4.50$ , maka mahasiswa terlambat.
- 37) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $\leq 2.50$ , Prodi  $> 4.00$ , Angkatan  $\leq 2.50$ , maka mahasiswa terlambat.
- 38) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $\leq 2.50$ , Prodi  $> 4.00$ , Angkatan  $> 2.50$ , maka mahasiswa tepat.
- 39) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $> 2.50$ , Angkatan  $\leq 2.50$ , Pendidikan Ayah  $\leq 3.50$ , maka mahasiswa tepat.
- 40) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $> 2.50$ , Angkatan  $\leq 2.50$ , Pendidikan Ayah  $> 3.50$ , maka mahasiswa tepat.
- 41) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $> 2.50$ , Angkatan  $> 2.50$ , Penghasilan Ibu  $\leq 2.50$ , maka mahasiswa terlambat.
- 42) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $\leq 6.50$ , Fakultas  $> 2.50$ , Angkatan  $> 2.50$ , Penghasilan Ibu  $> 2.50$ , maka mahasiswa terlambat.
- 43) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $\leq 3.50$ , Prodi  $> 6.50$ , maka mahasiswa tepat.
- 44) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $\leq 4.50$ , Penghasilan Ayah  $\leq 3.50$ , Pendidikan Ayah  $\leq 9.00$ , maka mahasiswa tepat.
- 45) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $\leq 4.50$ , Penghasilan Ayah  $\leq 3.50$ , Pendidikan Ayah  $> 9.00$ , maka mahasiswa terlambat.
- 46) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $\leq 4.50$ , Penghasilan Ayah  $> 3.50$ , maka terlambat tepat.
- 47) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $> 4.50$ , Prodi  $\leq 11.50$ , Pendidikan Ibu  $\leq 0.50$ , maka mahasiswa tepat.
- 48) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $> 4.50$ , Prodi  $\leq 11.50$ , Pendidikan Ibu  $> 0.50$ , maka mahasiswa terlambat.
- 49) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $> 4.50$ , Prodi  $> 11.50$ , Pendidikan Ayah  $\leq 7.50$ , maka mahasiswa terlambat.
- 50) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $\leq 1.50$ , Fakultas  $> 4.50$ , Prodi  $> 11.50$ , Pendidikan Ayah  $> 7.50$ , maka mahasiswa terlambat.
- 51) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $\leq 3.50$ , Fakultas  $\leq 4.50$ , Penghasilan Ayah  $\leq 1.00$ , maka mahasiswa terlambat.
- 52) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $\leq 3.50$ , Fakultas  $\leq 4.50$ , Penghasilan Ayah  $> 1.00$ , maka mahasiswa terlambat.
- 53) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $\leq 3.50$ , Fakultas  $> 4.50$ , Prodi  $\leq 11.50$ , maka mahasiswa terlambat.
- 54) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $\leq 3.50$ , Fakultas  $> 4.50$ , Prodi  $> 11.50$ , maka mahasiswa terlambat.
- 55) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $> 3.50$ , Prodi  $\leq 12.50$ , Pendidikan Ayah  $\leq 5.00$ , maka mahasiswa terlambat.
- 56) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $> 3.50$ , Prodi  $\leq 12.50$ , Pendidikan Ayah  $> 5.00$ , maka mahasiswa tepat.
- 57) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $> 3.50$ , Prodi  $> 12.50$ , Pendidikan Ayah  $\leq 6.50$ , maka mahasiswa terlambat.
- 58) Angkatan  $> 0.50$ , Gender  $\leq 0.50$ , Fakultas  $> 3.50$ , angkatan  $> 1.50$ , Penghasilan Ayah  $> 3.50$ , Prodi  $> 12.50$ , Pendidikan Ayah  $> 6.50$ , maka mahasiswa terlambat.
- 59) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $\leq 1.50$ , Pendidikan Ibu  $\leq 9.00$ , Penghasilan Ayah  $\leq 0.50$ , maka mahasiswa terlambat.
- 60) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $\leq 1.50$ , Pendidikan Ibu  $\leq 9.00$ , Penghasilan Ayah  $\leq 0.50$ , maka mahasiswa terlambat.
- 61) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $\leq 1.50$ , Pendidikan Ibu  $\leq 9.00$ , Penghasilan Ayah  $> 0.50$ , maka mahasiswa terlambat.
- 62) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $\leq 1.50$ , Pendidikan Ibu  $> 9.00$ , Penghasilan Ayah  $\leq 3.50$ , maka mahasiswa tepat.
- 63) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $\leq 1.50$ , Pendidikan Ibu  $> 9.00$ , Penghasilan Ayah  $> 3.50$ , maka mahasiswa terlambat.
- 64) Angkatan  $> 0.50$ , Gender  $> 0.50$ , Fakultas  $\leq 2.50$ , Angkatan  $\leq 2.50$ , Angkatan  $> 1.50$ , Prodi  $\leq 1.50$ , Penghasilan Ibu  $\leq 1.50$ , maka mahasiswa terlambat.

- 65) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan <= 2.50, Angkatan > 1.50, Prodi <= 1.50, Penghasilan Ibu > 1.50, maka mahasiswa tepat.
- 66) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan <= 2.50, Angkatan > 1.50, Prodi > 1.50, Pendidikan Ayah <= 2.50, maka mahasiswa tepat.
- 67) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan <= 2.50, Angkatan > 1.50, Prodi > 1.50, Pendidikan Ayah > 2.50, maka mahasiswa terlambat.
- 68) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu <= 2.50, Penghasilan Ibu <= 1.00, Prodi <= 3.50, maka mahasiswa terlambat.
- 69) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu <= 2.50, Penghasilan Ibu <= 1.00, Prodi > 3.50, maka mahasiswa terlambat.
- 70) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu <= 2.50, Penghasilan Ibu > 1.00, Penghasilan Ayah <= 2.50, maka mahasiswa tepat.
- 71) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu <= 2.50, Penghasilan Ibu > 1.00, Penghasilan Ayah > 2.50, maka mahasiswa tepat.
- 72) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu > 2.50, Fakultas <= 1.50, Pendidikan Ayah <= 9.00, maka mahasiswa terlambat.
- 73) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu > 2.50, Fakultas <= 1.50, Pendidikan Ayah > 9.00, maka mahasiswa tepat.
- 74) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu > 2.50, Fakultas > 1.50, Penghasilan Ibu <= 4.50, maka mahasiswa terlambat.
- 75) Angkatan > 0.50, Gender > 0.50, Fakultas <= 2.50, Angkatan > 2.50, Pendidikan Ibu > 2.50, Fakultas > 1.50, Penghasilan Ibu > 4.50, maka mahasiswa tepat.
- 76) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas <= 3.50, Prodi <= 5.00, Angkatan <= 2.50, Penghasilan Ayah <= 1.00, maka mahasiswa tepat.
- 77) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas <= 3.50, Prodi <= 5.00, Angkatan <= 2.50, Penghasilan Ayah > 1.00, maka mahasiswa tepat.
- 78) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas <= 3.50, Prodi <= 5.00, Angkatan > 2.50, Penghasilan Ayah <= 4.50, maka mahasiswa terlambat.
- 79) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas <= 3.50, Prodi <= 5.00, Angkatan > 2.50, Penghasilan Ayah > 4.50, maka mahasiswa terlambat.
- 80) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas <= 3.50, Prodi > 5.00, maka mahasiswa tepat.
- 81) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah <= 2.50, Pendidikan Ibu <= 9.00, Prodi <= 4.50, maka mahasiswa terlambat.
- 82) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah <= 2.50, Pendidikan Ibu <= 9.00, Prodi > 4.50, maka mahasiswa terlambat.
- 83) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah <= 2.50, Pendidikan Ibu > 9.00, Prodi <= 6.50, maka mahasiswa tepat.
- 84) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah <= 2.50, Pendidikan Ibu > 9.00, Prodi > 6.50, maka mahasiswa terlambat.
- 85) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah > 2.50, Penghasilan Ayah <= 4.50, Pendidikan Ayah <= 3.50, maka mahasiswa terlambat.
- 86) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah > 2.50, Penghasilan Ayah <= 4.50, Pendidikan Ayah > 3.50, maka mahasiswa tepat.
- 87) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah > 2.50, Penghasilan Ayah > 4.50, Fakultas <= 4.50, maka mahasiswa tepat.
- 88) Angkatan > 0.50, Gender > 0.50, Fakultas > 2.50, Fakultas > 3.50, Penghasilan Ayah > 2.50, Penghasilan Ayah > 4.50, Fakultas > 4.50, maka mahasiswa terlambat.

Untuk evaluasi kinerja model yang telah terbentuk, dapat digunakan nilai akurasi yang didapatkan melalui *confusion matrix*. Hasil dari tabel *confusion matrix* dapat ditemukan pada Tabel 3.

Tabel 3. Hasil *confusion matrix* tanpa seleksi fitur *information gain* dan *pruning*

	Actual = Yes	Actual = No.
Predicted = Yes	393	223
Predicted = No	256	369

Tabel 3 memuat hasil dari *confusion matrix* yang dapat dijelaskan sebagai berikut:

- 1) *True Positive* (TP) adalah jumlah mahasiswa (393) yang diprediksi dengan benar sebagai membayar biaya kuliah tepat waktu dan memang benar-benar membayar biaya kuliah tepat waktu.
- 2) *True Negative* (TN) adalah jumlah mahasiswa (396) yang diprediksi dengan benar sebagai mahasiswa yang terlambat membayar biaya kuliah dan memang benar-benar terlambat membayar biaya kuliah tepat waktu.
- 3) *False Positive* (FP) adalah jumlah mahasiswa (223) yang diprediksi benar tepat



membayar biaya kuliah, namun ternyata benar-benar terlambat membayar biaya kuliah tepat waktu.

4) *False Negative* (FN) adalah jumlah mahasiswa (256) yang diprediksi dengan terlambat membayar biaya kuliah, namun ternyata benar-benar tepat membayar biaya kuliah tepat waktu.

Hasil *confusion matrix* diatas dapat dihitung akurasinya dengan menggunakan persamaan berikut.

$$\begin{aligned}
 accuracy &= \frac{393 + 369}{393 + 369 + 223 + 256} * 100\% \\
 &= \frac{762}{1241} * 100\% \\
 &= 61,40\%
 \end{aligned}$$

Berdasarkan hasil tersebut, model algoritma C4.5 untuk memprediksi keterlambatan pembayaran biaya kuliah mencapai tingkat akurasi sebesar 61,40%. Selanjutnya, akan dilakukan upaya peningkatan akurasi algoritma C4.5 dengan menerapkan seleksi fitur menggunakan metode *information gain* dan melakukan teknik *pruning*.

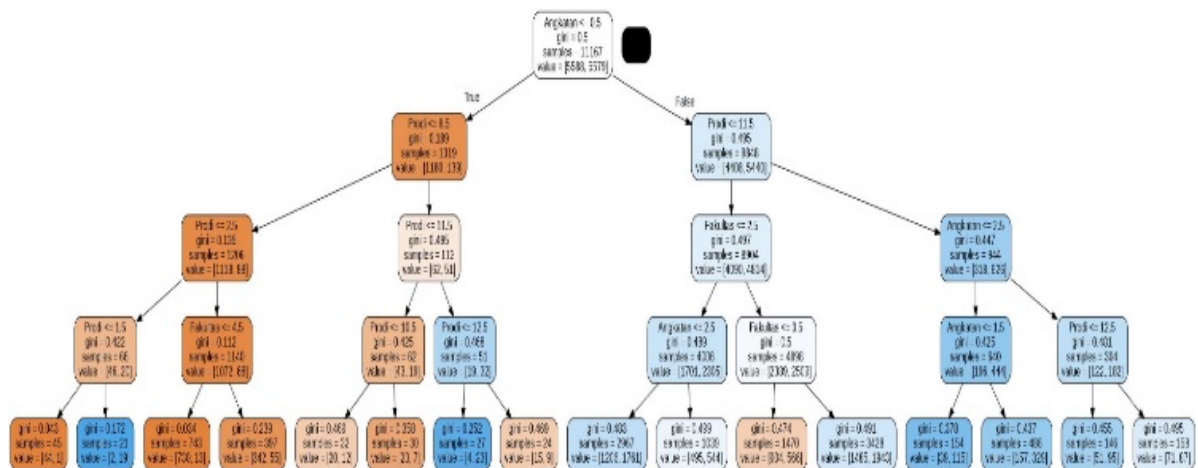
## 2) Algoritma C4.5 Menggunakan Seleksi Fitur *Information Gain* dan Teknik *Pruning*

Tabel VII menampilkan hasil dari seleksi fitur menggunakan metode *information gain*. Dalam seleksi fitur ini, digunakan threshold sebesar 0.01. Oleh karena itu, atribut yang dipilih untuk pemodelan algoritma C4.5 dengan seleksi fitur *information gain* adalah angkatan, prodi, dan fakultas.

Tabel 4. Hasil seleksi fitur *information gain*

No.	Atribut	Nilai <i>Information Gain</i>
1	Angkatan	0.070
2	Prodi	0.039
3	Fakultas	0.030
4	Gender	0.006
5	Penghasilan_Ayah	0.002
6	Penghasilan_Ibu	0.001
7	Pendidikan_Ayah	0.001
8	Pendidikan_Ibu	0.000

Berikut adalah hasil dari pohon keputusan dengan menggunakan seleksi fitur *information gain*.



Gambar 6: Visualisasi Pohon Keputusan

Menggunakan Seleksi Fitur *information gain* dan Teknik *pruning*

Gambar merupakan visualisasi pohon keputusan dengan penjelasan sebagai berikut.

- 1) Angkatan <= 0.50, Prodi <= 8.50, Prodi <= 2.50, Prodi <= 1.50, maka mahasiswa tepat.
- 2) Angkatan <= 0.50, Prodi <= 8.50, Prodi <= 2.50, Prodi > 1.50, maka mahasiswa terlambat.
- 3) Angkatan <= 0.50, Prodi <= 8.50, Prodi > 2.50, Fakultas <= 4.50, maka mahasiswa tepat.
- 4) Angkatan <= 0.50, Prodi <= 8.50, Prodi > 2.50, Fakultas > 4.50, maka mahasiswa tepat.
- 5) Angkatan <= 0.50, Prodi > 8.50, Prodi <= 11.50, Prodi <= 10.50, maka mahasiswa tepat.
- 6) Angkatan <= 0.50, Prodi > 8.50, Prodi <= 11.50, Prodi > 10.50, maka mahasiswa tepat.

- 7) Angkatan <= 0.50, Prodi > 8.50, Prodi > 11.50, Prodi <= 12.50, maka mahasiswa terlambat.
- 8) Angkatan <= 0.50, Prodi > 8.50, Prodi > 11.50, Prodi > 12.50, maka mahasiswa tepat.
- 9) Angkatan > 0.50, Prodi <= 11.50, Fakultas <= 2.50, Angkatan <= 2.50, maka mahasiswa terlambat.
- 10) Angkatan > 0.50, Prodi <= 11.50, Fakultas <= 2.50, Angkatan > 2.50, maka mahasiswa tepat.
- 11) Angkatan > 0.50, Prodi <= 11.50, Fakultas > 2.50, Fakultas <= 3.50, maka mahasiswa tepat.
- 12) Angkatan > 0.50, Prodi <= 11.50, Fakultas > 2.50, Fakultas > 3.50, maka mahasiswa terlambat.

- 13) Angkatan > 0.50, Prodi > 11.50, Angkatan <= 2.50, Angkatan <= 1.50, maka mahasiswa terlambat.
- 14) Angkatan > 0.50, Prodi > 11.50, Angkatan <= 2.50, Angkatan > 1.50, maka mahasiswa terlambat.
- 15) Angkatan > 0.50, Prodi > 11.50, Angkatan > 2.50, Prodi <= 12.50, maka mahasiswa terlambat.
- 16) Angkatan > 0.50, Prodi > 11.50, Angkatan > 2.50, Prodi > 12.50, maka mahasiswa terlambat.

Untuk mengukur performa atau kinerja model yang telah dibentuk dengan menggunakan seleksi fitur *information gain* dalam prediksi keterlambatan pembayaran biaya kuliah, dilakukan evaluasi melalui nilai akurasi menggunakan *confusion matrix*. Tabel IX menunjukkan hasil dari *confusion matrix* untuk evaluasi tersebut.

Tabel 5. Hasil *confusion matrix* dengan menggunakan seleksi fitur *information gain* dan *pruning*

	Actual = Yes	Actual = No.
Predicted = Yes	245	371
Predicted = No	65	560

Tabel 5 memuat hasil dari *confusion matrix* yang dapat dijelaskan sebagai berikut.

- 1) *True Positive* (TP) adalah jumlah mahasiswa (245) yang diprediksi dengan benar sebagai membayar biaya kuliah tepat waktu dan memang benar-benar membayar biaya kuliah tepat waktu.
- 2) *True Negative* (TN) adalah jumlah mahasiswa (560) yang diprediksi dengan benar sebagai mahasiswa yang terlambat membayar biaya kuliah dan memang benar-benar terlambat membayar biaya kuliah tepat waktu.
- 3) *False Positive* (FP) adalah jumlah mahasiswa (371) yang diprediksi benar tepat membayar biaya kuliah, namun ternyata benar-benar terlambat membayar biaya kuliah tepat waktu.
- 4) *False Negative* (FN) adalah jumlah mahasiswa (65) yang diprediksi dengan terlambat membayar biaya kuliah, namun ternyata benar-benar tepat membayar biaya kuliah tepat waktu.

Hasil *confusion matrix* diatas dapat dihitung akurasi dengan menggunakan persamaan berikut.

$$\begin{aligned}
 \text{Akurasi} &= \frac{245 + 560}{245 + 560 + 371 + 560} * 100\% \\
 &= \frac{805}{1241} * 100\% \\
 &= 64,86\%
 \end{aligned}$$

Hasil dari algoritma C4.5 dengan menggunakan seleksi fitur *information gain* dan teknik *pruning* memperoleh nilai akurasi sebesar 64,86%.

## E. Evaluasi

Pada pengujian pemodelan algoritma C4.5, terjadi peningkatan yang signifikan dalam kinerja saat menggunakan seleksi fitur *information gain* dan teknik *pruning* terbukti mampu meningkatkan kinerja algoritma sebesar 3,45%. Hasil ini menunjukkan bahwa penggunaan seleksi fitur *information gain* berhasil meningkatkan akurasi dibandingkan dengan menggunakan algoritma C4.5 tanpa modifikasi. Perbandingan nilai akurasi dapat dilihat pada Tabel 6.

Tabel 6. Hasil perbandingan akurasi algoritma tanpa dan dengan seleksi fitur *information gain* dan *pruning*

	Algoritma C4.5	Algoritma C4.5 + <i>Information Gain</i> + <i>Pruning</i>
Akurasi	61,40%	64,86%

## 5. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa prediksi keterlambatan pembayaran biaya kuliah dengan menggunakan algoritma C4.5 menghasilkan akurasi tertinggi ketika menggunakan seleksi fitur *information gain* dan teknik *pruning*, dengan nilai akurasi mencapai 64,86%. Beberapa atribut yang digunakan dalam prediksi tersebut antara lain fakultas, prodi, angkatan, gender, dan penghasilan ibu.

Dalam penelitian ini, ditemukan bahwa fitur yang digunakan memiliki pengaruh signifikan terhadap akurasi prediksi. Oleh karena itu, untuk mencapai hasil yang lebih baik, disarankan untuk menggunakan metode seleksi fitur lainnya seperti forward selection atau particle swarm optimization, serta menjalankan metode klasifikasi data mining lainnya seperti Naïve Bayes, Random Forest, dan K-Nearest Neighbors. Dengan demikian, dapat meningkatkan kualitas prediksi keterlambatan pembayaran biaya kuliah.

## 6. DAFTAR PUSTAKA

- Adnan, G., & Latief, M. A. (2020). Metode Penelitian Pendidikan: Penelitian Kuantitatif, Penelitian Kualitatif, Penelitian Tindakan Kelas. Erhaka Utama, Yogyakarta. ISBN 978-602-5715-34-1.
- Apandi, T. H., Maulana, R. B., Piarna, R., & Vernanda, D. (2019). Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung). *Jurnal Techno Nusa Mandiri*, 16(2), 93–98.
- Desiani, A., Yahdin, S., & Rodiah, D. (2020). Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa dengan Menggunakan Teknik Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1237-1244.
- Daqiqil, I. (2021). *Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python*.

- (Versi 1). UNRI Press. <https://doi.org/10.5281/zenodo.5113507>.
- Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4. 5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 36-44.
- Hermawati, F.A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi
- Iskandar, I., Hiryanto, L., & Hendryli, J. (2018). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4. 5 Dengan Teknik Pruning. *Jurnal Ilmu Komputer dan Sistem Informasi*, 6(1), 64-64.
- Jensen, R., & Shen, Q. (2008). *Computational Intelligence and Feature Selection - Rough and Fuzzy Approaches*. IEEE Press series on computational intelligence.
- Johnson, K. J., & Synovec, R. E. (2002). Pattern recognition of jet fuels: comprehensive GC× GC with ANOVA-based feature selection and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 60(1-2), 225-237.
- Junita, V., & Bachtiar, F. A. (2019). Klasifikasi Aktivitas Manusia menggunakan Algoritma Decision Tree C4. 5 dan Information Gain untuk Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Kurniawan, D. (2022). *Pengenalan Machine Learning dengan Python*. Jakarta: PT Elex Media Komputindo.
- Mardi, Y. (2017). *Data Mining: Klasifikasi Menggunakan Algoritma C4. 5*. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.
- Nurhayati, Soekarno, I., Hadihardaja, I. K., & Cahyono, M. (2015). IEEE. A study of hold-out and k-fold cross validation for accuracy of groundwater modelling in tidal lowland reclamation using extreme learning machine. 10.1109/TIME-E.2014.7011623.
- Nofriansyah, D., & Nurcahyo, G. (2015). *Algoritma Data Mining dan Pengujian*. Yogyakarta: DEEPUBLISH.
- Rohmayani, D. (2020). Analysis of Student Tuition Fee Pay Delay Prediction Using Naive Bayes Algorithm With Particle Swarm Optimization Optimazation (Case Study : Politeknik Tedc Bandung). *Jurnal Teknologi Informasi dan Pendidikan*, 13(2), 1-8.
- Pulungan, A. F. (2019). Analisis Kinerja Bray Curtis Distance, Canberra Distance dan Euclidean Distance pada Algoritma K-Nearest Neighbour. Tersedia di: <https://repositori.usu.ac.id/handle/123456789/15051> [Diakses 28 Januari 2022]
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4), 427-437.
- Wibawa, A. P. (2018). Metode-metode Klasifikasi. In *Prosiding SAKTI (Seminar Ilmu Komputer dan Teknologi Informasi)* (Vol. 3, No. 1, pp. 134-138).