



## Klasifikasi Kelayakan Bantuan Pendidikan Menggunakan Metode *Decision Tree*

Ida Kumala Sari<sup>1\*</sup>, Arief Hermawan<sup>2</sup>, Donny Avianto<sup>3</sup>

<sup>1,2</sup>Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

<sup>3</sup>Magister Sistem Informatika, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

Email: <sup>1\*</sup>sarikumalaida@email.com, <sup>2</sup>ariefdb@uty.ac.id, <sup>3</sup>donny@uty.ac.id

### *Abstract*

*This study aims to implement the Decision Tree method to classify scholarship eligibility based on student data. The dataset used consisted of 4,424 student records with 35 numerical attributes covering academic, administrative, and socioeconomic aspects. The preprocessing stage included data quality checking, feature selection, and dataset splitting into training and testing data with a ratio of 70:30. From the initial 35 attributes, 16 main attributes were selected as the most relevant features to the target variable. The classification model was developed using the Decision Tree algorithm and evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The results showed that the model achieved an accuracy of 75.08% and a weighted average F1-score of 75.16%. Attributes such as Curricular units 2nd sem (approved), Curricular units 2nd sem (grade), and Tuition fees up to date were identified as the most influential factors in the classification process. Although the dataset had a class imbalance condition, this study maintained the original data distribution without applying oversampling techniques such as SMOTE in order to preserve the actual conditions of the scholarship selection process. In addition to providing fairly good classification performance, the Decision Tree method was also able to produce a transparent, interpretable, and easy-to-understand model through the resulting decision tree structure.*

**Keywords:** *Decision Tree, Classification, Scholarship, Data Mining, Machine Learning.*

### **Abstrak**

Penelitian ini bertujuan untuk mengimplementasikan metode Decision Tree dalam mengklasifikasikan kelayakan penerima beasiswa berdasarkan data mahasiswa. Dataset yang digunakan terdiri dari 4.424 data mahasiswa dengan 35 atribut numerik yang mencakup aspek akademik, administratif, dan sosial ekonomi. Pada tahap preprocessing dilakukan pemeriksaan kualitas data, feature selection, serta pembagian data menjadi data training dan testing dengan rasio 70:30. Dari 35 atribut awal, diperoleh 16 atribut utama yang paling relevan terhadap variabel target. Model klasifikasi dibangun menggunakan algoritma Decision Tree dan dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model memperoleh nilai accuracy sebesar 75,08% dan weighted average F1-score sebesar 75,16%. Atribut seperti Curricular units 2nd sem (approved), Curricular units 2nd sem (grade), dan Tuition fees up to date menjadi faktor yang paling berpengaruh dalam proses klasifikasi. Meskipun dataset memiliki kondisi class imbalance, penelitian ini tetap menggunakan distribusi data asli tanpa menerapkan teknik oversampling seperti SMOTE agar model tetap merepresentasikan kondisi aktual pada proses seleksi penerima beasiswa. Selain memiliki performa klasifikasi yang cukup baik, metode Decision Tree juga mampu menghasilkan model yang transparan, mudah dipahami, dan interpretatif melalui struktur pohon keputusan yang terbentuk.

**Kata Kunci:** *Decision Tree, Klasifikasi, Beasiswa, Data Mining, Machine Learning.*

## 1. PENDAHULUAN

Pendidikan berperan penting dalam meningkatkan kualitas sumber daya manusia serta mendorong pembangunan sosial dan ekonomi suatu negara. Namun, akses terhadap pendidikan tinggi yang layak masih belum merata, terutama bagi mahasiswa yang berasal dari latar belakang ekonomi kurang mampu. Hal ini sejalan dengan laporan UNESCO (2021) yang menyatakan bahwa ketimpangan akses pendidikan masih menjadi permasalahan global, khususnya akibat perbedaan kondisi sosial dan ekonomi. Oleh karena itu, berbagai program bantuan pendidikan seperti beasiswa disediakan oleh pemerintah maupun institusi pendidikan guna mendukung pemerataan akses pendidikan. Meskipun demikian, proses seleksi penerima beasiswa yang masih dilakukan secara konvensional sering kali menimbulkan permasalahan berupa ketidaktepatan sasaran akibat adanya subjektivitas dalam pengambilan keputusan (Chooses & Loses, 2021).

Seiring dengan perkembangan teknologi informasi, pemanfaatan teknik data mining menjadi salah satu solusi dalam mendukung pengambilan keputusan berbasis data. Data mining merupakan proses ekstraksi pengetahuan atau pola tersembunyi dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik dan machine learning (Bachhal, Ahuja, 2021). Dalam konteks pendidikan, data mining telah banyak digunakan untuk membantu proses analisis data mahasiswa, termasuk dalam menentukan kelayakan penerima bantuan pendidikan secara lebih objektif dan sistematis (Yupei et al., 2021).

Salah satu metode klasifikasi yang banyak digunakan dalam data mining adalah *Decision Tree*. *Decision Tree* merupakan algoritma pembelajaran terawasi (*supervised learning*) yang membangun model prediksi dalam bentuk struktur pohon keputusan berdasarkan atribut-atribut yang tersedia dalam dataset (Nyabadza et al., 2021). Proses pembentukan pohon dilakukan melalui pemilihan atribut terbaik sebagai node dengan menggunakan ukuran tertentu, seperti information gain atau entropy, sehingga mampu memisahkan data ke dalam kelas yang lebih homogen. Metode ini banyak digunakan karena kemampuannya dalam menangani data numerik maupun kategorikal serta tidak memerlukan asumsi distribusi data tertentu dan mampu menangani berbagai tipe data secara fleksibel dalam proses klasifikasi (Pusponegoro et al., 2021). Selain itu, *Decision Tree* memiliki keunggulan utama dalam hal interpretabilitas, di mana hasil model dapat direpresentasikan dalam bentuk aturan keputusan (*decision rules*) yang mudah dipahami oleh pengguna, termasuk pihak non-teknis. Hal ini menjadikan *Decision Tree* sangat relevan dalam pengambilan keputusan yang membutuhkan transparansi dan akuntabilitas. Dalam konteks pendidikan, metode ini telah banyak diterapkan untuk berbagai keperluan analisis data mahasiswa, seperti prediksi performa akademik dan klasifikasi kelayakan bantuan pendidikan (Yupei et al., 2021). Oleh karena itu, *Decision Tree* dinilai sebagai pendekatan yang efektif dan efisien dalam menyelesaikan permasalahan klasifikasi, khususnya pada sistem pendukung keputusan yang menuntut hasil yang dapat dijelaskan secara logis. Temuan penelitian terbaru juga memperkuat hal tersebut, di mana algoritma *Decision Tree* mampu menghasilkan akurasi yang tinggi serta model yang interpretatif dalam klasifikasi penerima beasiswa (Suprapti et al., 2025).

Beberapa penelitian sebelumnya telah membahas penerapan data mining dalam penentuan penerima bantuan pendidikan atau beasiswa dengan berbagai pendekatan metode klasifikasi. Penelitian yang dilakukan oleh Stiawan (2023) dalam jurnal IJCS mengkaji perbandingan beberapa algoritma *Decision Tree*, yaitu ID3, C4.5, dan CART dalam proses klasifikasi penerima beasiswa Bank Indonesia. Penelitian ini menggunakan sejumlah atribut yang berkaitan dengan kondisi akademik dan ekonomi mahasiswa sebagai variabel penentu. Hasil analisis menunjukkan bahwa algoritma CART memiliki performa terbaik dibandingkan algoritma lainnya, dengan tingkat *accuracy* mencapai

72%. Penelitian ini juga menunjukkan bahwa model berbentuk pohon keputusan mampu menghasilkan aturan klasifikasi yang sistematis dan mudah dipahami. Dengan karakteristik tersebut, proses penentuan penerima beasiswa dapat dilakukan secara lebih objektif serta transparan, sehingga memudahkan pihak pengambil keputusan dalam menetapkan hasil yang tepat (Stiawan, 2023). Selain itu, metode Decision Tree juga telah digunakan dalam klasifikasi performa akademik mahasiswa dan menunjukkan hasil yang cukup baik dalam mengidentifikasi pola data (Rahman, 2023).

Penelitian yang dilakukan oleh Hartanto (2023) mengkaji penerapan algoritma *Decision Tree* dalam proses penentuan penerima beasiswa pada jenjang sekolah menengah. Studi tersebut memanfaatkan sejumlah kriteria sebagai dasar penilaian, antara lain prestasi akademik, kondisi ekonomi keluarga, serta jumlah tanggungan orang tua. Tahapan analisis dilakukan melalui pembentukan model pohon keputusan yang kemudian digunakan untuk mengklasifikasikan tingkat kelayakan penerima. Hasil yang diperoleh menunjukkan bahwa metode ini memiliki tingkat akurasi yang tinggi, yakni sebesar 92%. Selain itu, model yang dihasilkan dinilai mampu meningkatkan efisiensi proses seleksi dan meminimalkan subjektivitas, karena keputusan yang diambil didasarkan pada aturan-aturan yang terbentuk dari model tersebut (Hartanto, 2023).

Penelitian lain dilakukan oleh Rahimi et al. (2025) yang mengembangkan sistem pendukung keputusan berbasis algoritma *Decision Tree* C4.5 untuk klasifikasi penerima beasiswa. Penelitian ini tidak hanya berfokus pada pembangunan model klasifikasi, tetapi juga pada implementasi sistem yang dapat digunakan secara langsung oleh pihak pengelola beasiswa. Data yang digunakan diproses melalui tahapan *preprocessing*, pembentukan model, hingga evaluasi performa model. Temuan penelitian tersebut menunjukkan bahwa penerapan algoritma C4.5 dapat mempercepat proses seleksi, khususnya dalam pengolahan data yang sebelumnya masih dilakukan secara manual, sehingga menjadi lebih efisien. Selain itu, sistem yang dibangun mampu menghasilkan keputusan yang lebih konsisten, transparan, dan dapat dipertanggungjawabkan (Rahimi et al., 2025).

Penelitian yang dilakukan oleh Raya dan Arfida (2024) mengkaji penerapan algoritma *Decision Tree* C4.5 dalam sistem seleksi penerima beasiswa Kartu Indonesia Pintar (KIP) yang dikembangkan berbasis web. Studi ini menyoroti integrasi antara metode klasifikasi dengan platform berbasis website guna meningkatkan kemudahan akses serta pengelolaan data. Dalam implementasinya, sejumlah variabel seperti pendapatan orang tua, jumlah tanggungan, dan prestasi akademik digunakan sebagai parameter dalam pembentukan model. Hasil penelitian menunjukkan bahwa sistem yang dibangun mampu meningkatkan tingkat objektivitas sekaligus efisiensi dalam proses penentuan penerima bantuan pendidikan. Di samping itu, pemanfaatan sistem berbasis web turut mendukung pengolahan data dalam skala besar serta mengurangi potensi kesalahan yang umumnya terjadi pada proses manual (Raya & Arfida, 2024).

Studi lain yang relevan dilakukan oleh Ammar dan Wiranata (2025) dengan membandingkan beberapa algoritma klasifikasi, yaitu *Naïve Bayes*, *Decision Tree*, dan *Random Forest* dalam proses analisis data. Penelitian ini difokuskan untuk mengidentifikasi metode yang paling optimal dalam menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga efisien. Hasil yang diperoleh menunjukkan bahwa algoritma *Decision Tree* memiliki kinerja yang relatif stabil dengan tingkat akurasi yang bersaing dibandingkan dengan algoritma lainnya. Selain itu, metode ini memiliki keunggulan dalam hal interpretasi, karena model yang dihasilkan mudah dipahami. Hal tersebut menjadikannya cocok digunakan pada kasus yang menuntut transparansi dalam pengambilan keputusan, seperti pada proses penentuan penerima bantuan pendidikan (Ammar & Wiranata, 2025).

Studi yang dilakukan oleh Putri et al. (2024) mengkaji penerapan algoritma *Decision Tree* C4.5 dalam analisis data pendidikan untuk memprediksi lama masa studi mahasiswa. Penelitian tersebut memanfaatkan sejumlah atribut akademik sebagai variabel masukan dalam proses klasifikasi. Temuan yang diperoleh menunjukkan bahwa metode *Decision Tree* mampu membentuk model klasifikasi dengan tingkat akurasi yang baik serta mudah untuk dipahami. Selain itu, penggunaan algoritma C4.5 dinilai dapat mendukung proses pengambilan keputusan berbasis data secara lebih efektif dan efisien (K. A. Putri et al., 2024).

Penelitian selanjutnya oleh Fitriyani et al. (2023) mengimplementasikan metode *Decision Tree* untuk memprediksi tingkat kelulusan mahasiswa berdasarkan data akademik yang tersedia. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengenali pola dalam data dengan cukup baik serta menghasilkan model klasifikasi yang optimal. Lebih lanjut, metode ini dapat dimanfaatkan sebagai alat bantu dalam proses pengambilan keputusan di bidang pendidikan, khususnya terkait pengelolaan data mahasiswa (Fitriyani, Sentosa, Desi, Riswan, 2023)

Penelitian lain dilakukan oleh Iman et al. (2024) yang memanfaatkan metode *Decision Tree* dalam proses klasifikasi data untuk mendukung pengambilan keputusan berbasis data. Dalam penelitian tersebut, algoritma C4.5 digunakan untuk mengolah data dengan berbagai atribut yang relevan. Temuan penelitian tersebut menunjukkan bahwa metode yang digunakan mampu membangun model klasifikasi dengan kinerja yang baik serta berkontribusi dalam meningkatkan ketepatan proses pengambilan keputusan. Di samping itu, metode *Decision Tree* juga memiliki kelebihan dalam hal interpretasi, karena model yang dihasilkan mudah dipahami oleh pengguna (Iman, Noor, Supriyono, 2024).

Meskipun berbagai penelitian sebelumnya menunjukkan bahwa algoritma *Decision Tree* mampu memberikan performa klasifikasi yang cukup baik dalam penentuan penerima beasiswa, sebagian besar penelitian tersebut masih menggunakan jumlah atribut yang relatif terbatas dan lebih berfokus pada peningkatan akurasi model. Dalam kasus klasifikasi kelayakan bantuan pendidikan, data mahasiswa memiliki karakteristik multidimensi yang melibatkan atribut akademik, administratif, dan sosial ekonomi yang saling berkaitan. Kondisi tersebut berpotensi meningkatkan kompleksitas proses klasifikasi serta memunculkan noise yang dapat memengaruhi performa model. Di sisi lain, algoritma ensemble seperti *Random Forest* dan *XGBoost* memang dikenal memiliki performa prediksi yang tinggi, namun model yang dihasilkan cenderung lebih sulit diinterpretasikan karena tersusun dari kombinasi banyak pohon keputusan. Dalam konteks kebijakan publik, khususnya penentuan penerima bantuan pendidikan, aspek transparansi dan *explainability* menjadi sangat penting agar hasil keputusan dapat dipahami dan dipertanggungjawabkan oleh pihak pengelola maupun penerima bantuan. Oleh karena itu, penelitian ini menggunakan metode *Decision Tree* karena mampu menghasilkan aturan keputusan yang lebih sederhana, transparan, dan mudah diinterpretasikan dengan memanfaatkan 16 atribut hasil seleksi fitur yang memiliki tingkat relevansi paling tinggi terhadap variabel klasifikasi kelayakan bantuan pendidikan.

Berdasarkan latar belakang serta hasil analisis kesenjangan yang telah diuraikan, penelitian ini diarahkan pada penerapan metode *Decision Tree* untuk mengklasifikasikan kelayakan penerima bantuan pendidikan dengan memanfaatkan data mahasiswa yang tersedia. Tahapan penelitian dilakukan melalui proses data mining yang terstruktur, dimulai dari tahap pengolahan data hingga evaluasi terhadap model yang dihasilkan. Fokus penelitian tidak hanya pada pembangunan model klasifikasi, tetapi juga pada pengujian kinerjanya secara menyeluruh menggunakan berbagai metrik evaluasi, sehingga diperoleh gambaran yang lebih jelas mengenai kemampuan model dalam menyelesaikan permasalahan klasifikasi.

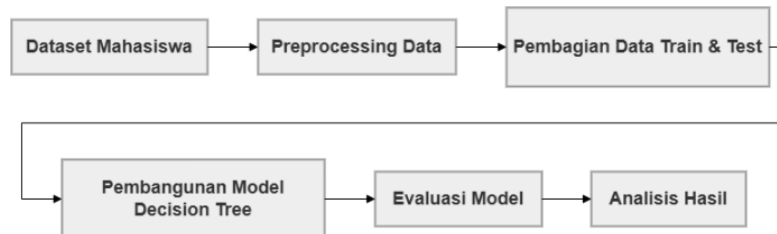
Tujuan dari penelitian ini adalah mengimplementasikan metode *Decision Tree* dalam menentukan kelayakan penerima bantuan pendidikan sekaligus menganalisis performa model yang dihasilkan. Selain itu, penelitian ini juga bertujuan untuk mengkaji sejauh mana metode tersebut dapat mendukung proses pengambilan keputusan yang lebih objektif dan berbasis pada data.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang sistematis untuk menghasilkan model klasifikasi kelayakan bantuan pendidikan yang optimal. Proses penelitian ini meliputi serangkaian tahapan yang dimulai dari pengolahan data hingga tahap evaluasi model, sehingga hasil yang dihasilkan dapat dipertanggungjawabkan secara ilmiah.

Alur tahapan penelitian dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Alur Proses Penelitian

Mengacu pada Gambar 1, setiap tahapan dalam penelitian ini dapat diuraikan sebagai berikut.

1. Tahap pertama adalah penggunaan dataset mahasiswa sebagai sumber data utama dalam penelitian. Dataset ini berisi informasi yang berkaitan dengan kondisi akademik maupun ekonomi mahasiswa yang digunakan sebagai dasar dalam menentukan kelayakan penerimaan bantuan pendidikan.
2. Tahap kedua adalah preprocessing data. Pada tahap ini dilakukan pengecekan terhadap kualitas data yang digunakan. Karena seluruh atribut pada dataset telah berbentuk numerik dan tidak terdapat missing value, maka tidak dilakukan proses normalisasi data lebih lanjut
3. Tahap ketiga adalah dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*) dengan perbandingan 70:30. Data latih dimanfaatkan dalam proses pembentukan model, sedangkan data uji digunakan untuk menguji kemampuan model dalam mengklasifikasikan data yang belum pernah diproses sebelumnya.
4. Tahap keempat adalah pembangunan model klasifikasi menggunakan metode *Decision Tree*. Pada tahap ini dilakukan proses pembentukan pohon keputusan berdasarkan atribut-atribut yang terdapat dalam dataset hingga menghasilkan model klasifikasi.
5. Tahap kelima adalah evaluasi model. Setelah model terbentuk, tahap selanjutnya adalah pengujian menggunakan data uji untuk menilai kinerjanya. Proses evaluasi dilakukan dengan menggunakan beberapa metrik, di antaranya *accuracy*, *precision*, *recall*, dan *F1-score*.
6. Tahap terakhir adalah analisis hasil, yaitu menginterpretasikan hasil evaluasi model untuk mengetahui kelebihan dan kekurangan model dalam mengklasifikasikan data kelayakan bantuan pendidikan.

## 2.2 Dataset

Data yang digunakan dalam penelitian ini berupa data mahasiswa terkait klasifikasi kelayakan penerimaan beasiswa. Dataset terdiri dari 4.424 data mahasiswa dengan 35 atribut numerik yang mencakup aspek akademik, administratif, dan sosial ekonomi. Variabel target terdiri atas dua kelas, yaitu “Layak” dan “Tidak Layak”. Dari total data, sebanyak 1.099 mahasiswa termasuk kategori layak menerima beasiswa dan 3.325 mahasiswa termasuk kategori tidak layak. Distribusi data tersebut menunjukkan kondisi class imbalance karena jumlah data tidak layak lebih dominan dibandingkan data layak.

Karakteristik dataset meliputi 4.424 data mahasiswa, 35 atribut numerik awal yang kemudian direduksi menjadi 16 atribut melalui proses feature selection, dua kelas target (“Layak” dan “Tidak Layak”), serta tidak terdapat missing value. Penggunaan atribut numerik mempermudah proses pengolahan data dan pembangunan model menggunakan algoritma *Decision Tree*.

Tabel 1. Ringkasan Atribut Dataset

No	Atribut	No	Atribut	No	Atribut	No	Atribut	No	Atribut
1	Marital status	8	Nationality	15	Educational special needs	22	Curricular units 1st sem (credited)	29	Curricular units 2nd sem (enrolled)
2	Application mode	9	Mother’s qualification	16	Debtor	23	Curricular units 1st sem (enrolled)	30	Curricular units 2nd sem (evaluations)
3	Application order	10	Father’s qualification	17	Tuition fees up to date	24	Curricular units 1st sem (evaluations)	31	Curricular units 2nd sem (approved)
4	Course	11	Mother’s occupation	18	Gender	25	Curricular units 1st sem (approved)	32	Curricular units 2nd sem (grade)
5	Daytime/evening attendance	12	Father’s occupation	19	Scholarship holder	26	Curricular units 1st sem (grade)	33	Curricular units 2nd sem (without evaluations)
6	Previous qualification	13	Admission grade	20	Age at enrollment	27	Curricular units 1st sem (without evaluations)	34	Unemployment rate
7	Previous qualification grade	14	Displaced	21	International	28	Curricular units 2nd sem (credited)	35	Inflation rate

Tabel 2. Atribut Hasil Feature Selection

No	Atribut Terpilih	No	Atribut Terpilih	No	Atribut Terpilih
1	Curricular units 2nd sem (approved)	7	Tuition fees up to date	13	Curricular units 1st sem (approved)
2	Curricular units 2nd sem (grade)	8	Scholarship holder	14	Curricular units 1st sem (grade)
3	Curricular units 2nd sem (evaluations)	9	Debtor	15	Curricular units 1st sem (evaluations)
4	Curricular units 2nd sem (enrolled)	10	Admission grade	16	Curricular units 1st sem (enrolled)
5	Age at enrollment	11	Previous qualification grade		
6	Gender	12	Application mode		

### 2.3 Preprocessing

Pada tahap preprocessing, dilakukan pemeriksaan kualitas data untuk memastikan data siap digunakan dalam proses klasifikasi. Proses ini bertujuan untuk meningkatkan kualitas data sehingga dapat menghasilkan model klasifikasi yang lebih akurat dan optimal (Buysse et al., 2022; A. Putri & Purnama, 2025). Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki missing value sehingga seluruh data dapat digunakan pada tahap analisis.

Selanjutnya, dilakukan feature selection berdasarkan nilai *feature importance* untuk mengurangi atribut yang kurang relevan terhadap variabel target, sehingga jumlah atribut berkurang dari 35 atribut menjadi 16 atribut utama. Dataset juga memiliki distribusi kelas yang tidak seimbang (*class imbalance*), di mana jumlah data “Tidak Layak” lebih dominan dibandingkan “Layak”. Namun, penelitian ini tidak menerapkan teknik *oversampling* seperti *SMOTE* agar distribusi data tetap merepresentasikan kondisi aktual pada proses seleksi penerima beasiswa. Setelah preprocessing selesai, dataset dibagi menjadi data *training* sebesar 70% dan data *testing* sebesar 30% sebelum dilakukan proses klasifikasi menggunakan algoritma *Decision Tree*.

### 2.4 Pembangunan Model *Decision Tree*

Dalam penelitian ini, proses pembentukan model dilakukan dengan menerapkan algoritma *Decision Tree* sebagai salah satu metode klasifikasi pada data mining. Pendekatan ini menghasilkan model berupa struktur pohon yang dibangun berdasarkan atribut-atribut yang terdapat dalam dataset.

Proses pembangunan *Decision Tree* dimulai dengan pemilihan atribut terbaik sebagai akar (*root node*). Pemilihan atribut ini didasarkan pada nilai *information gain*, yaitu ukuran yang digunakan untuk menentukan seberapa baik suatu atribut dalam memisahkan data berdasarkan kelasnya.

Entropy dimanfaatkan untuk mengukur tingkat ketidakpastian atau keragaman data dalam suatu dataset, yang dinyatakan melalui rumus sebagai berikut :

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

Selanjutnya, nilai *information gain* dihitung untuk menentukan atribut terbaik menggunakan rumus :

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

Atribut yang memiliki nilai *information gain* paling tinggi akan diprioritaskan sebagai node dalam pembentukan pohon keputusan. Proses ini berlangsung secara iteratif hingga seluruh data berhasil diklasifikasikan atau ketika tidak tersedia lagi atribut yang dapat digunakan sebagai pemisah.

Hasil dari proses ini adalah model pohon keputusan yang dapat digunakan untuk mengklasifikasikan data baru ke dalam kategori “Layak” atau “Tidak Layak”.

Keunggulan dari metode *Decision Tree* adalah kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan. Hal ini sangat penting dalam konteks pengambilan keputusan, terutama dalam penentuan penerima beasiswa yang membutuhkan transparansi.

## 2.5 Evaluasi Model

Tahap evaluasi dilakukan untuk menilai performa model *Decision Tree* yang telah dikembangkan. Proses ini bertujuan untuk mengidentifikasi tingkat kemampuan model dalam melakukan klasifikasi secara tepat terhadap data yang diuji.

Penelitian ini, pengukuran kinerja model dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi berdasarkan perbandingan antara nilai prediksi dan nilai aktual (Firdaus & Setiadi, 2023). Metode ini terdiri dari empat elemen utama, yaitu:

- *True Positive (TP)*
- *True Negative (TN)*
- *False Positive (FP)*
- *False Negative (FN)*

Berdasarkan nilai tersebut, beberapa metrik evaluasi dihitung, yaitu :  
Akurasi :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision :

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

Recall :

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

F1-Score :

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Penerapan beberapa metrik evaluasi bertujuan untuk memperoleh gambaran kinerja model yang lebih komprehensif. Penilaian tidak hanya difokuskan pada tingkat akurasi, tetapi juga pada kemampuan model dalam mengklasifikasikan setiap kelas secara tepat. Pendekatan ini penting untuk menghindari bias terhadap kelas tertentu serta memberikan evaluasi yang lebih representatif terhadap performa model klasifikasi (Attamami, Agung & Rima, 2023; Ramadhan & Ashari, 2024)

Selain itu, penerapan metode *Decision Tree* dalam berbagai penelitian sebelumnya juga menunjukkan bahwa penggunaan *confusion matrix* dan metrik evaluasi turunannya mampu memberikan gambaran performa model yang lebih jelas dan mudah dipahami, terutama dalam kasus klasifikasi data pendidikan (K. A. Putri et al., 2024).

Dengan demikian, melalui penggunaan *confusion matrix* dan berbagai metrik evaluasi tersebut, diharapkan model yang dihasilkan dalam penelitian ini tidak hanya memiliki tingkat akurasi yang baik, tetapi juga mampu memberikan keseimbangan performa dalam mengklasifikasikan data kelayakan penerima beasiswa

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil

Bagian ini menyajikan hasil eksperimen dan visualisasi model secara objektif, tanpa interpretasi mendalam. Seluruh tabel dan *figure* ditampilkan sebagai output langsung dari proses pengolahan data dan pemodelan.



Gambar 2. Distribusi Kelas Status Penerima Beasiswa

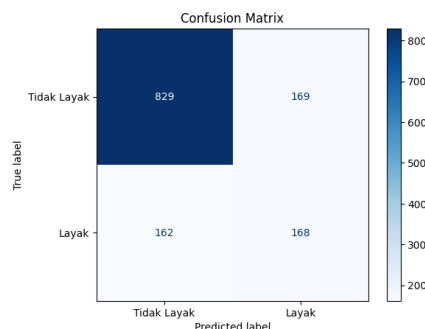
Gambar 2 menunjukkan distribusi kelas pada variabel target *Scholarship holder* yang memperlihatkan perbandingan jumlah mahasiswa layak dan tidak layak menerima beasiswa. Berdasarkan visualisasi tersebut, jumlah mahasiswa tidak layak menerima beasiswa (label 0) lebih dominan dibandingkan mahasiswa layak menerima beasiswa (label 1). Distribusi data tersebut menunjukkan bahwa dataset memiliki karakteristik *class imbalance*.

Tabel 1. *Classification Report*

Kelas	Precision	Recall	F1-Score	Support
Tidak Layak	0.836529	0.830661	0.833585	998
Layak	0.498516	0.509091	0.503748	330
Accuracy			0.750753	1328
Macro Avg	0.667523	0.669876	0.668666	1328
Weighted Avg	0.752535	0.750753	0.751622	1328

Hasil evaluasi model Decision Tree menggunakan data testing ditunjukkan pada Tabel 1. Evaluasi dilakukan menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy* untuk mengukur performa model dalam mengklasifikasikan kelayakan penerima beasiswa.

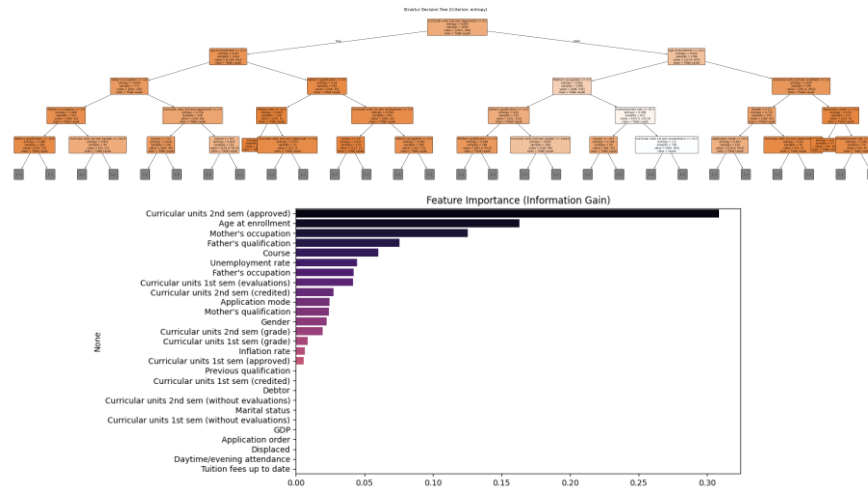
Berdasarkan Tabel 1, model memperoleh nilai *accuracy* sebesar 75,08%. Nilai *precision*, *recall*, dan *F1-score* pada kelas “Tidak Layak” menunjukkan hasil yang lebih tinggi dibandingkan kelas “Layak”.



Gambar 3. Confusion Matrix Heatmap

Gambar 3 menunjukkan *confusion matrix* hasil klasifikasi model *Decision Tree* pada data *testing*. *Confusion matrix* digunakan untuk memperlihatkan perbandingan antara hasil prediksi model dengan kondisi aktual pada masing-masing kelas.

Berdasarkan *confusion matrix*, diperoleh nilai *True Negative* sebanyak 829 data dan *True Positive* sebanyak 168 data. Selain itu, terdapat *False Positive* sebanyak 169 data dan *False Negative* sebanyak 162 data.



Gambar 3. Plot Visualisasi *Decision Tree*

Gambar 4 memperlihatkan visualisasi struktur pohon keputusan yang dihasilkan dari proses pelatihan model *Decision Tree* menggunakan data latih. Visualisasi ini menunjukkan hubungan antar atribut serta aturan keputusan (*decision rules*) yang digunakan model dalam mengklasifikasikan kelayakan penerima beasiswa. Berdasarkan hasil *feature importance*, dari 35 atribut awal terpilih 16 atribut utama yang digunakan dalam proses klasifikasi. Fitur-fitur akademik seperti *Curricular units 2nd sem (approved)* dan *Curricular units 2nd sem (grade)*, serta atribut administrasi seperti *Tuition fees up to date*, terlihat memiliki peran dominan pada node awal pohon keputusan. Hal ini menunjukkan bahwa atribut-atribut tersebut memiliki kontribusi paling tinggi terhadap variabel target. Struktur pohon keputusan juga memperlihatkan bahwa proses klasifikasi dilakukan secara hierarkis dimulai dari atribut yang memiliki tingkat informasi paling tinggi hingga menghasilkan keputusan klasifikasi pada node akhir (*leaf node*).

### 3.2. Pembahasan

Berdasarkan hasil evaluasi model, algoritma *Decision Tree* memperoleh nilai *accuracy* sebesar 75,08% dalam mengklasifikasikan kelayakan penerima beasiswa. Hasil tersebut menunjukkan bahwa model mampu mengenali pola klasifikasi dengan cukup baik pada data mahasiswa yang digunakan dalam penelitian. Selain *accuracy*, nilai *weighted average F1-score* sebesar 75,16% juga menunjukkan bahwa performa model secara keseluruhan relatif stabil dalam mengklasifikasikan data pada kedua kelas.

Pada kelas “Tidak Layak”, model memperoleh nilai *precision* sebesar 83,65%, *recall* sebesar 83,07%, dan *F1-score* sebesar 83,36%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali data pada kelas mayoritas. Sementara itu, pada kelas “Layak”, nilai *precision* sebesar 49,85%, *recall* sebesar 50,91%, dan *F1-score* sebesar 50,37% menunjukkan bahwa model masih mengalami keterbatasan dalam mengenali data pada kelas minoritas.

Perbedaan performa antar kelas dipengaruhi oleh distribusi dataset yang tidak seimbang (*class imbalance*), di mana jumlah data “Tidak Layak” lebih dominan dibandingkan “Layak”. Kondisi tersebut menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas sehingga kemampuan klasifikasi terhadap kelas minoritas menjadi lebih rendah. Meskipun demikian, hasil penelitian menunjukkan bahwa algoritma *Decision Tree* tetap mampu memberikan performa klasifikasi yang cukup baik serta menghasilkan model yang mudah dipahami dan diinterpretasikan.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa algoritma *Decision Tree* memiliki kemampuan yang cukup baik dalam proses klasifikasi data pendidikan dan penentuan penerima beasiswa (Hartanto, 2023; Stiawan,

2023). Selain menghasilkan performa klasifikasi yang kompetitif, metode ini juga memiliki keunggulan dalam transparansi model karena aturan keputusan yang dihasilkan dapat dijelaskan secara lebih sederhana dibandingkan metode ensemble seperti Random Forest maupun XGBoost.

#### 4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode Decision Tree untuk mengklasifikasikan kelayakan penerima beasiswa menggunakan dataset mahasiswa sebanyak 4.424 data. Dari 35 atribut awal, dilakukan feature selection sehingga diperoleh 16 atribut utama yang paling relevan terhadap variabel target. Proses penelitian meliputi preprocessing data, pembagian data training dan testing dengan rasio 70:30, pembangunan model, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score.

Hasil evaluasi menunjukkan bahwa model Decision Tree memperoleh accuracy sebesar 75,08% dan weighted average F1-score sebesar 75,16%, sehingga model dinilai mampu mengklasifikasikan kelayakan penerima beasiswa dengan cukup baik. Atribut akademik dan administratif seperti *Curricular units 2nd sem (approved)*, *Curricular units 2nd sem (grade)*, dan *Tuition fees up to date* menjadi faktor yang paling berpengaruh dalam proses klasifikasi. Meskipun dataset memiliki kondisi *class imbalance* yang menyebabkan performa kelas minoritas lebih rendah dibandingkan kelas mayoritas, penelitian ini tetap menggunakan distribusi data asli tanpa menerapkan teknik oversampling seperti *SMOTE* agar pola klasifikasi yang dihasilkan tetap merepresentasikan kondisi aktual pada proses seleksi penerima beasiswa. Selain itu, metode *Decision Tree* mampu menghasilkan model yang transparan, mudah dipahami, dan interpretatif melalui struktur pohon keputusan yang terbentuk.

Penelitian ini masih memiliki keterbatasan, terutama pada penggunaan satu metode klasifikasi dan belum dilakukannya perbandingan performa dengan algoritma lain seperti *Random Forest*, *XGBoost*, atau metode penanganan *class imbalance*. Oleh karena itu, penelitian selanjutnya diharapkan dapat mengembangkan model menggunakan metode klasifikasi yang lebih beragam, menerapkan teknik penyeimbangan data, serta mengeksplorasi atribut tambahan agar diperoleh performa klasifikasi yang lebih optimal dan komprehensif.

#### REFERENCES

- Ammar, H., & Wiranata, A. D. (2025). Perbandingan Tingkat Akurasi Algoritma Naïve Bayes, Decision Tree dan Random Forest dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional. *IJAI (Indonesian Journal of Applied Informatics)*, 10(1), 59–73. [https://jurnal.uns.ac.id/ijai/article/view/111010?utm\\_source=chatgpt.com](https://jurnal.uns.ac.id/ijai/article/view/111010?utm_source=chatgpt.com)
- Attamami, Agung, & Rima. (2023). Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(2). <https://doi.org/10.35870/jtik.v7i2.756>
- Bachhal, Ahuja, G. (2021). *Educational Data Mining : A Review Educational Data Mining : A Review*. <https://doi.org/10.1088/1742-6596/1950/1/012022>
- Buysse, Kevin, & Steven. (2022). *Hierarchical Reinforcement Learning : A Survey and Open Research Challenges*. 172–221.
- Chooses, W. H. O., & Loses, W. H. O. (2021). *Non-state actors in education Non-state actors in education*.

- Firdaus, L., & Setiadi, T. (2023). Perbandingan Algoritma Naive Bayes , Decision Tree , dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix. *Jurnal Sistem Komputer Informatika (JSON)*, 5, 185–195. <https://doi.org/10.30865/json.v5i2.6124>
- Fitriyani, Sentosa, Desi, Riswan, B. (2023). Data Mining untuk Memprediksi Kelulusan Mahasiswa Jurusan DIII Kebidanan Menggunakan Metode Klasifikasi C4 . 5 *JURNAL MEDIA INFORMATIKA [JUMIN ]*. *JURNAL MEDIA INFORMATIKA [JUMIN]*, 5(1), 100–105. <https://doi.org/10.55338/jumin.v5i1.6082>
- Hartanto. (2023). PENERAPAN ALGORITMA DECISION TREE UNTUK SELEKSI PENERIMA BEASISWA (STUDI KASUS: SMPN 1 SOREANG). *Journal of Comprehensive Science*, 2(7), 1294–1302. <https://jcs.greenpublisher.id/index.php/jcs/article/view/452/440>
- Iman, Noor, Supriyono, N. (2024). Pemanfaatan Metode Decision Tree dengan Algoritma C4 . 5 Untuk Prediksi Potensi Kunjungan Wisatawan. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 5(3), 611–619. <https://doi.org/10.30865/json.v5i3.6684>
- Nyabadza, A., Mercedes, V., Coyle, S., Fitzpatrick, B., & Brabazon, D. (2021). *applied sciences Review of Materials and Fabrication Methods for Flexible Nano and Micro-Scale Physical and Chemical Property Sensors*.
- Pusponegoro, N. H., Kurnia, A., Notodiputro, K. A., Soleh, A. M., & Astuti, E. T. (2021). ScienceDirect ScienceDirect Small Area Estimation of Sub- District ' s Per Capita Expenditure through Area Effects Selection using LASSO Method. *Procedia Computer Science*, 179, 754–761. <https://doi.org/10.1016/j.procs.2021.01.064>
- Putri, A., & Purnama, B. (2025). Classification of Scholarship Eligibility Using Naïve Bayes with Attribute Optimization Based on K-Means Clustering Klasifikasi Kelayakan Penerima Beasiswa Menggunakan Naive Bayes dengan Optimasi Atribut Berbasis K-Means Clustering. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(October), 1450–1462. <https://doi.org/10.57152/malcom.v5i4.2312>
- Putri, K. A., Febriawan, D., & Hasan, F. N. (2024). Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm ( C4 . 5 ). *Jurnal SISFOKOM (Sistem Informasi Dan Komputer)*, 13(1), 39–47. <https://doi.org/10.32736/sisfokom.v13i1.1943>
- Rahimi, H. Z., Defit, S., & Veri, J. (2025). IMPLEMENTASI DECISION TREE DALAM PENGAMBILAN KEPUTUSAN UNTUK PEMBERIAN BEASISWA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4175–4180. <https://ejournal.itn.ac.id/jati/article/view/13580/7556>
- Rahman, A. (2023). Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes. *Jurnal Sainstekom*, 9, 22–31. <https://doi.org/10.33020/sainstekom.v13i1.349>
- Ramadhan, R. F., & Ashari, W. M. (2024). Performance Comparison of Random Forest and Decision Tree Algorithms for Anomaly Detection in Networks. *Journal of Applied Informatics and Computing (JAIC)*, 8(2), 367–375. <https://doi.org/doi.org/10.30871/jaic.v8i2.8492>
- Raya, Y. C., & Arfida, S. (2024). Penerapan Algoritma Decision Tree C4.5 Untuk Penerimaan Beasiswa Kip Bagi Mahasiswa Baru Berbasis Website. *JURNAL TEKNIKA*, 18(2), 377–388.
- Stiawan, Y. (2023). Perbandingan Performa Algoritma Decision Tree untuk Klasifikasi Penerima Beasiswa Bank Indonesia. *Indonesian Journal of Computer Science*, 12(1), 2108–2123. <https://www.ijcs.net/ijcs/index.php/ijcs/article/view/3339/226>
- Suprpti, T., Nurhakim, B., Warni, B., Hermina, A., & Syahputra, V. A. (2025). A Decision Tree Model with Grid Search Optimization for Scholarship Recipient Classification. *Jurnal Teknik Informatika (JUTIF)*, 6(5), 3800–3813. <https://doi.org/10.52436/1.jutif.2025.6.5.5235>
- Yupei, Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). *Educational Data Mining Techniques for Student Performance Prediction : Method Review and Comparison Analysis*. 12(December), 1–19. <https://doi.org/10.3389/fpsyg.2021.698490>