

Penerapan Decision Tree CART untuk Klasifikasi Risiko Gagal Studi Mahasiswa

Chesa Saskia Rafika¹, Revano Maliq Reynanada², Anggraini Puspita Sari^{3*}
^{1,2,3} Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya
Email : corresponding anggraini.puspita.if@upnjatim.ac.id

Abstrak

Pendidikan tinggi menghadapi tantangan serius berupa masalah gagal studi mahasiswa, yang dipengaruhi oleh berbagai faktor mulai dari ketidaksesuaian minat hingga kondisi psikososial. Identifikasi dini risiko gagal studi menjadi krusial untuk mencegah tingginya angka putus studi dan merancang intervensi yang tepat. Penelitian ini memanfaatkan machine learning, khususnya algoritma Decision Tree CART, untuk mengklasifikasikan risiko gagal studi mahasiswa berdasarkan data akademik dan non-akademik, seperti IPK, masa studi, keterlibatan KKN/PKL, dan poin SKPM. Tahapan penelitian meliputi identifikasi masalah, pengumpulan dan pra-pemrosesan data, penerapan algoritma Decision Tree, pelatihan model, evaluasi, dan analisis. Data sebanyak 100 entri diproses dengan teknik data cleaning dan transformasi atribut kategorikal ke numerik, kemudian dibagi menjadi rasio 80:20 untuk pelatihan dan pengujian. Hasil evaluasi menunjukkan akurasi sebesar 85%, yang mengindikasikan bahwa model memiliki performa yang baik dalam memprediksi status kelulusan mahasiswa. Temuan ini menunjukkan bahwa model Decision Tree CART dapat menjadi solusi proaktif untuk deteksi dini risiko gagal studi dan mendukung pengambilan keputusan strategis di lingkungan perguruan tinggi.

Kata Kunci: *Decision Tree, CART, Klasifikasi, Gagal Studi, Mahasiswa*

Abstract

Higher education faces a serious challenge in the form of student study failure, influenced by various factors such as mismatched interests and psychosocial conditions. Early identification of study failure risk is crucial to prevent high dropout rates and to design appropriate interventions. This study utilizes machine learning, specifically the Decision Tree CART algorithm, to classify the risk of student study failure based on academic and non-academic data, including GPA, study duration, KKN/PKL participation, and SKPM points. The research stages include problem identification, data collection and preprocessing, algorithm implementation, model training, evaluation, and analysis. A total of 100 data entries were processed using data cleaning techniques and categorical-to-numerical transformation, then split into an 80:20 ratio for training and testing. The evaluation results show an accuracy of 85%, indicating that the model performs well in predicting students' graduation status. These findings suggest that the Decision Tree CART model offers a proactive solution for early detection of study failure risk and supports strategic decision-making in higher education institutions.

Keywords: *Decision Tree, CART, Classification, Study Failure, Students*

PENDAHULUAN

Pendidikan tinggi merupakan salah satu elemen utama dalam mencetak sumber daya manusia yang kompeten dan berdaya saing di era global. Namun demikian, fenomena gagal studi di kalangan mahasiswa masih menjadi tantangan krusial bagi perguruan tinggi, khususnya di Indonesia. Berbagai faktor seperti ketidaksesuaian minat, kendala ekonomi, beban psikososial, dan kurangnya partisipasi akademik serta non-akademik turut berperan dalam meningkatkan risiko tersebut (Kristyawan & Sumirat, 2019; Suryani & Maryanto, 2022; Aesyti et al., 2021; Suriani, 2023). Permasalahan ini menuntut adanya sistem deteksi dini yang mampu mengidentifikasi mahasiswa dengan potensi gagal studi secara lebih cepat dan akurat Noviyanto & Fauzi, 2024; Esananda et al., 2021).

Dalam beberapa tahun terakhir, pendekatan berbasis data melalui *machine learning* semakin relevan digunakan dalam bidang pendidikan karena kemampuannya dalam melakukan prediksi dan klasifikasi secara otomatis (Rismaya et al., 2023; Amani & Hayati, 2024). Salah satu metode yang paling mudah diinterpretasikan dan sesuai untuk klasifikasi dalam konteks akademik adalah algoritma Decision Tree (Gunawan et al., 2024; Widodo, 2019). Decision Tree, khususnya varian CART (*Classification and Regression Tree*), bekerja dengan membagi data berdasarkan nilai *Gini* yang menghasilkan pohon keputusan terstruktur dan logis. Keunggulan utama CART adalah kemampuannya dalam menyederhanakan proses klasifikasi kompleks menjadi alur pengambilan keputusan yang dapat divisualisasikan, sehingga hasilnya lebih mudah dipahami oleh pemangku kebijakan di institusi pendidikan (Rizalno et al., 2022).

Penggunaan CART dalam penelitian ini dinilai tepat karena karakteristik data mahasiswa terdiri atas kombinasi atribut numerik dan kategorikal, seperti IPK, masa studi, status KKN/PKL,

serta penyelesaian tugas akhir. Algoritma CART mampu menangani keduanya dengan baik, tanpa perlu transformasi kompleks, serta tidak sensitif terhadap skala data (Widodo, 2019; Gunawan et al., 2024). Dalam implementasinya, CART menghasilkan pohon keputusan yang efisien dalam memisahkan data dengan tingkat kesalahan minimal. Penelitian ini menggunakan CART karena dianggap paling cocok untuk kebutuhan klasifikasi risiko kelulusan berdasarkan struktur pohon yang mendetail dan akurat (Esananda et al., 2021; Amani & Hayati, 2024). Studi lain juga memperkuat bahwa Decision Tree mampu mengklasifikasikan risiko dropout mahasiswa secara efektif berdasarkan kombinasi variabel akademik dan sosial (Gustian, 2025).

Urgensi penelitian ini terletak pada tingginya angka keterlambatan kelulusan yang belum sepenuhnya diantisipasi oleh sistem akademik konvensional. Evaluasi performa mahasiswa selama ini lebih bersifat reaktif dilakukan setelah penurunan prestasi terjadi sehingga intervensi tidak dapat dilakukan secara optimal (Aesyti et al., 2021; Noviyanto & Fauzi, 2024). Oleh karena itu, dibutuhkan pendekatan proaktif berbasis data untuk memprediksi risiko gagal studi lebih awal. Penelitian Ibu Anggraini Puspita Sari mengenai klasifikasi data spasial COVID-19 menggunakan metode K-Means juga menunjukkan pentingnya penggunaan metode berbasis clustering dan klasifikasi dalam mendukung kebijakan berbasis data (Puspita Sari, 2021). Studi kasus dalam penelitian ini berfokus pada data mahasiswa dari perguruan tinggi di Indonesia yang memiliki latar belakang heterogen, baik dari sisi akademik maupun aktivitas non-akademik. Penggunaan metode klasifikasi CART dianggap relevan karena struktur keputusannya tidak hanya memberikan hasil klasifikasi, tetapi juga mampu menjelaskan logika pengambilan keputusan berbasis atribut mahasiswa secara transparan (Rizalno et al., 2022). Studi ini menjadi penting karena belum banyak riset

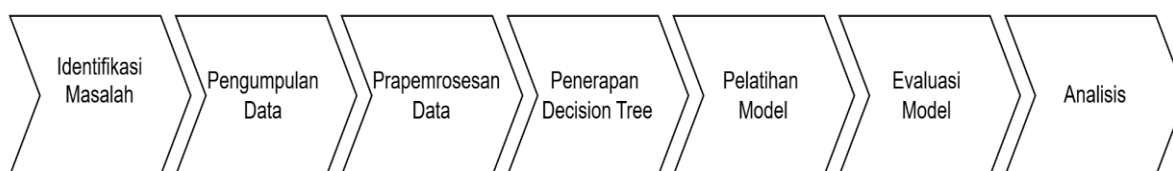
yang mengombinasikan faktor akademik dan non-akademik secara seimbang dalam model klasifikasi risiko kelulusan di lingkungan lokal. Hasil penelitian juga mendukung efektivitas CART dengan akurasi tinggi saat diterapkan dalam klasifikasi data akademik mahasiswa dari berbagai fakultas di Indonesia (Muriyatmoko et al., 2024).

Tujuan dari penelitian ini adalah mengembangkan model klasifikasi berbasis Decision Tree CART yang mampu memprediksi risiko gagal studi mahasiswa secara akurat berdasarkan atribut gabungan. Penelitian ini menawarkan kebaruan dari sisi pemilihan variabel prediktor yang tidak hanya mencakup IPK dan masa studi, tetapi juga keterlibatan KKN/PKL, poin SKPM, dan penyelesaian tugas akhir. Selain itu, pendekatan yang digunakan dievaluasi dengan dua rasio pembagian data (70:30 dan 80:20), yang menunjukkan bahwa proporsi pelatihan data dapat mempengaruhi performa klasifikasi secara

signifikan. Temuan ini diharapkan dapat menjadi landasan bagi institusi pendidikan tinggi untuk membangun sistem pendukung keputusan akademik yang adaptif dan prediktif. Sebagai perbandingan, penelitian oleh Nasution & Sriani (2025) menggunakan algoritma Decision Tree C5.0 untuk menganalisis faktor penyebab putus sekolah pada tingkat SMA Negeri 6 Padangsidimpuan dan berhasil mencapai akurasi sebesar 87%, yang menunjukkan efektivitas algoritma pohon keputusan dalam konteks pendidikan di Indonesia.

METODE

Bagian ini menjelaskan tahapan sistematis yang dilakukan dalam penelitian Penerapan Decision Tree Untuk Klasifikasi Risiko Gagal Studi Mahasiswa. Proses penelitian dirancang secara terstruktur mulai dari identifikasi masalah hingga analisis hasil model klasifikasi. Diagram alur proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Identifikasi Masalah

Tingginya angka keterlambatan kelulusan dan kasus gagal studi di perguruan tinggi masih menjadi persoalan serius yang belum tertangani secara optimal. Selama ini, proses evaluasi terhadap performa mahasiswa cenderung bersifat reaktif, baru dilakukan setelah prestasi menurun, sehingga intervensi seringkali terlambat dan kurang efektif. Padahal, mahasiswa yang berisiko sebenarnya bisa dikenali lebih awal jika data akademik dan non-akademik dimanfaatkan secara menyeluruh. Informasi seperti IPK, masa studi, keikutsertaan dalam KKN atau PKL, serta poin SKPM menyimpan potensi besar untuk mengungkap pola-pola risiko yang

selama ini luput dari perhatian. Dengan pendekatan yang tepat, data tersebut dapat dijadikan dasar dalam membangun sistem prediksi yang lebih akurat dan proaktif. Oleh karena itu, perlu dikembangkan model klasifikasi yang mampu mengidentifikasi mahasiswa dengan risiko gagal studi sejak dini, sehingga perguruan tinggi dapat mengambil langkah intervensi secara cepat dan tepat sasaran.

2. Pengumpulan Data

Dalam penelitian ini, data yang digunakan berasal dari kumpulan data mahasiswa di Universitas Surabaya, dengan fokus khusus pada mahasiswa Institut Teknologi Sepuluh Nopember (ITS) yang berada pada rentang semester 7 hingga 11. Dataset ini memuat 100 entri mahasiswa, yang masing-masing merepresentasikan

satu individu dengan informasi akademik dan non-akademik yang dianggap relevan dalam menentukan status kelulusan. Atribut utama yang dianalisis mencakup IPK (Indeks Prestasi Kumulatif), masa studi, status keikutsertaan dalam KKN atau PKL, jumlah poin SKPM (Satuan Kredit Partisipasi Mahasiswa), serta status penyelesaian tugas akhir atau skripsi. Seluruh data tersebut dipersiapkan sebagai fitur dalam proses klasifikasi dengan tujuan untuk memetakan potensi risiko gagal studi mahasiswa secara lebih sistematis. Sementara itu, status kelulusan menjadi variabel target yang diklasifikasikan menjadi dua kategori, yaitu Lulus dan Risiko Tidak Lulus. Analisis deskriptif dilakukan untuk memahami distribusi dan karakteristik dari masing-masing variabel sebelum masuk ke tahap pelatihan model.

3. PraPemrosesan Data

Tahap pengelolaan dan normalisasi data merupakan langkah krusial dalam memastikan bahwa data yang digunakan dalam proses pembelajaran model memiliki kualitas yang layak, representatif, dan sesuai dengan format yang dibutuhkan oleh algoritma machine learning. Proses ini terdiri dari beberapa tahapan utama, yaitu:



Gambar 2. Tahapan PraPemrosesan Data

1. Data Cleaning
Langkah awal dalam prapemrosesan adalah normalisasi nama kolom dengan menghapus spasi dan mengubah semua

huruf menjadi kapital agar konsisten saat pemanggilan kolom. Selanjutnya dilakukan penghapusan kolom bertipe non-numerik, yaitu kolom yang bertipe *object* (umumnya berupa teks atau kategori), karena model Decision Tree hanya dapat memproses data numerik secara langsung. Kolom-kolom tersebut dicari menggunakan fungsi `select_dtypes(include=["object"])` dan kemudian di-*drop* dari `DataFrame`. Tidak terdapat proses imputasi atau pengisian nilai kosong secara eksplisit dalam script, karena data dianggap sudah bersih atau telah dipersiapkan sebelumnya.

2. Transformasi Data
Proses transformasi data dilakukan dengan memisahkan fitur prediktor (X) dari target klasifikasi (y), yaitu atribut `STATUS_KELULUSAN`, di mana diasumsikan bahwa data kategorikal seperti “KKN” dan “PKL” sudah dalam bentuk numerik (misalnya 1 untuk “Sudah”, 0 untuk “Belum”) sehingga model Decision Tree dapat langsung melatihnya tanpa encoding eksplisit. Untuk melengkapi persiapan ini dan menjaga konsistensi serta interpretabilitas, dilakukan juga standarisasi atribut dengan menyesuaikan rentang nilai pada fitur-fitur tertentu, seperti Indeks Prestasi Semester (IPS) ke skala [0.0 - 4.0], serta menyesuaikan Jumlah SKS Kumulatif dan Lama Studi ke dalam rentang nilai yang telah ditetapkan. Adapun ketentuan masing-masing atribut adalah sebagai berikut:

Tabel 1. Tabel Standarisasi Atribut

Atribut	Tipe Data	Rentang Nilai	Standarisasi
IPK	Numerik	0-4	Syarat kelulusan minimal IPK > 2
Masa Studi	Numerik	7-10 semester	Masa studi normal: 8 semester; mendekati 14 semester indikasi resiko
KKN/PKL	Kategorikal	0 (belum), 1 (sudah)	Wajib diikuti untuk kelulusan
Point SKPM	Numerik	0-150+	Minimal poin kelulusan: 150; <150 → risiko tidak lulus
Tugas Akhir	Numerik	>0(sudah). 0(belum)	Penyelesaian menjadi syarat mutlak kelulusan

4. Penerapan Decision Tree

Penelitian ini menggunakan model klasifikasi berbasis *Decision Tree* dengan algoritma CART (*Classification and Regression Tree*) yang diimplementasikan menggunakan pustaka Scikit-Learn pada bahasa pemrograman Python. Algoritma ini dipilih karena mampu membentuk struktur pohon keputusan yang intuitif dan mudah ditafsirkan. Proses pembentukan pohon dilakukan dengan mengevaluasi atribut-atribut input berdasarkan nilai *Gini impurity*, yaitu ukuran yang digunakan untuk menentukan tingkat ketidakhomogenan suatu node. Semakin rendah nilai Gini, semakin baik atribut tersebut dalam memisahkan data. Algoritma akan memilih atribut yang paling efektif dalam membagi data menjadi kelompok yang homogen. Setiap node pada pohon berisi pertanyaan atau kondisi berdasarkan salah satu atribut, seperti lama masa studi, nilai IPK, status penyelesaian tugas akhir, partisipasi dalam KKN atau PKL, dan jumlah poin SKPM. Data akan dialirkan ke cabang-cabang berdasarkan jawaban dari kondisi tersebut. Proses ini berlangsung secara rekursif hingga pohon mencapai kedalaman tertentu, atau hingga tidak ada lagi pemisahan yang berarti. Pada bagian ujung pohon (*leaf node*), model akan memberikan prediksi berupa dua kelas: mahasiswa yang

diperkirakan lulus dan mahasiswa yang termasuk dalam risiko tidak lulus.

5. Pelatihan Model

Model ini bekerja dengan membentuk struktur pohon keputusan, di mana setiap node pada pohon berisi aturan pemisahan data berdasarkan atribut yang paling berkontribusi dalam memisahkan kelas target, yaitu status kelulusan mahasiswa. Algoritma memilih atribut secara bertahap dengan mempertimbangkan nilai *Gini impurity*, yang mencerminkan ketidakteraturan data dalam setiap percabangan. Dengan demikian, atribut seperti IPK, masa studi, status KKN atau PKL, penyelesaian tugas akhir, dan jumlah poin SKPM dapat dievaluasi dalam menentukan pola klasifikasi. Untuk mengetahui pengaruh perbandingan data latih dan data uji terhadap performa model, data dibagi ke dalam dua skenario: skenario pertama menggunakan 70% data latih dan 30% data uji, sedangkan skenario kedua menggunakan 80% data latih dan 20% data uji. Pembagian ini dilakukan menggunakan fungsi `train_test_split` dari pustaka Scikit-Learn dengan parameter `random_state=42` untuk menjaga replikasi hasil. Setelah pembagian data, proses pelatihan dilakukan dengan memanggil metode `.fit()` pada objek model terhadap set data latih. Model mempelajari pola-pola dari data latih untuk kemudian diuji kemampuannya dalam

memprediksi data yang belum dikenali sebelumnya (data uji).

6. Evaluasi Model

Setelah model Decision Tree selesai dilatih pada kedua skenario data, langkah selanjutnya adalah melakukan evaluasi performa terhadap data uji. Evaluasi dilakukan untuk mengetahui seberapa baik model dapat memprediksi status kelulusan mahasiswa pada data yang belum pernah digunakan dalam pelatihan. Metode evaluasi yang digunakan mencakup empat metrik utama:

1. **Akurasi:** Mengukur proporsi prediksi yang benar terhadap total keseluruhan data.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. **Precision:** Menilai ketepatan model dalam memprediksi kelas positif (*LULUS*), yakni seberapa banyak dari prediksi “LULUS” yang benar-benar “LULUS”.

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

3. **Recall:** Mengukur sejauh mana model mampu menemukan seluruh data yang benar-benar termasuk dalam kelas “LULUS”.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. **F1-Score:** Merupakan rata-rata harmonik dari precision dan recall. Metrik ini sangat penting, terutama ketika distribusi data tidak seimbang.

$$F1 - Score = \frac{2.Presisi.Recall}{Presisi+Recall} \quad (4)$$

Penjelasan:

True Positive(TP) =Jumlah kasus positif yang diprediksi benar

False Positive(FP) =Jumlah kasus negatif yang diprediksi positif secara salah

False Negative (FN)=Jumlah kasus positif yang diprediksi negatif secara salah

True Negative (TN)=Jumlah kasus negatif yang diprediksi benar

Evaluasi dilakukan secara terpisah untuk kedua rasio pembagian data (70:30 dan 80:20). Hasil dari masing-masing skenario dibandingkan untuk menentukan rasio mana yang menghasilkan performa paling optimal dan andal untuk diterapkan dalam konteks prediksi risiko gagal studi mahasiswa.

7. Analisis

Berdasarkan hasil evaluasi, dilakukan analisis terhadap kemampuan model dalam mengklasifikasikan mahasiswa ke dalam kategori “LULUS” atau “RESIKO TIDAK LULUS”. Fokus utama dari analisis ini adalah mengidentifikasi seberapa akurat model dalam memisahkan dua kelas tersebut serta memahami fitur-fitur mana yang paling berkontribusi terhadap keputusan model. Dari struktur pohon keputusan yang dihasilkan, terlihat bahwa atribut-atribut seperti IPK, masa studi, status penyelesaian tugas akhir, serta keterlibatan dalam KKN dan PKL merupakan faktor yang paling dominan dalam mempengaruhi hasil prediksi.

HASIL DAN PEMBAHASAN

1. Hasil PraPemrosesan Data

Proses pra-pemrosesan data menghasilkan dataset yang bersih, terstruktur, dan siap digunakan untuk pelatihan model. Data yang semula mengandung nilai kosong dan atribut kategorikal berhasil dibersihkan dan ditransformasi menjadi bentuk numerik. Nilai kosong ditangani melalui teknik drop dan imputasi, sementara atribut seperti “KKN”, “PKL”, dan “Tugas Akhir” dikodekan menjadi nilai biner (1 untuk “sudah”, 0 untuk “belum”). Label target “STATUS_KELULUSAN” juga dikonversi menjadi biner, yaitu 1 untuk “LULUS” dan 0 untuk “RESIKO TIDAK LULUS”. Selain itu, atribut numerik seperti IPK, masa studi, dan poin SKPM telah distandarisasi sesuai rentang dan batas kelulusan yang ditentukan. Hasil akhir pra-pemrosesan ini menjadi dasar yang kuat

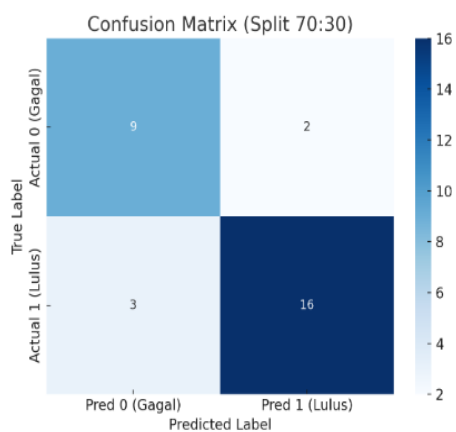
untuk proses pelatihan model menggunakan algoritma Decision Tree.

Gambar 3. Hasil Tabel Excel setelah PraPemrosesan

NO	LAMA_KULIAH	IPK	TUGAS_AKHIR	KKN	PKL	POINT_SKPM	STATUS_KELULUSAN
1	8	2.59	2	1	1	150	1
2	11	3.01	2.3	1	1	180	0
3	12	2.98	4	1	1	150	0
4	11	2.46	2.17	1	1	170	0
5	10	2.94	3	0	0	150	0
96	8	3.18	3	1	1	150	1
97	8	1.97	2.1	1	1	150	0
98	9	3.47	4	1	1	120	0
99	12	3.31	4	1	1	150	0
100	8	2.40	0	1	1	160	0

2. Hasil Pelatihan dan Evaluasi Model

Setelah proses pelatihan menggunakan algoritma Decision Tree Classifier pada data latih sebanyak 70 entri (70%) dan data uji sebanyak 30 entri (30%) dari total 100 data, model dievaluasi untuk mengukur kemampuannya dalam memprediksi status kelulusan mahasiswa. Evaluasi dilakukan dengan menghitung beberapa metrik kinerja seperti akurasi, precision, recall, dan F1-score berdasarkan hasil prediksi pada data uji. Berikut adalah gambar hasil evaluasi performa model pada rasio 70:30 ini:

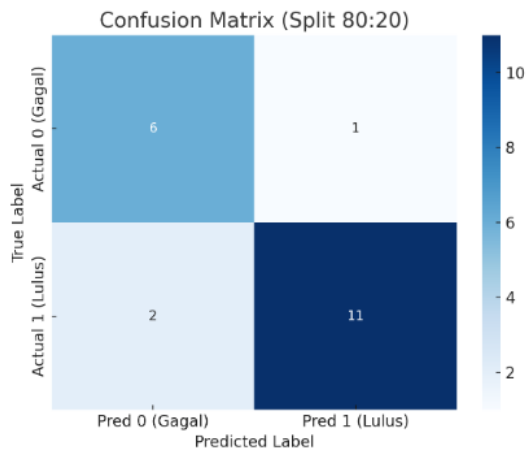


Gambar 4. Confusion Matrix (Split 70:30)

	precision	recall	f1-score	support
0	0.75	0.82	0.78	11
1	0.89	0.84	0.86	19
accuracy			0.83	30
macro avg	0.82	0.83	0.82	30
weighted avg	0.84	0.83	0.83	30

Gambar 5. Laporan Klasifikasi Rasio 70:30

Selanjutnya, model dilatih kembali dengan rasio data yang berbeda, yaitu 80 entri (80%) untuk data latih dan 20 entri (20%) untuk data uji. Langkah evaluasi dilakukan dengan cara yang sama, yaitu menggunakan metrik akurasi, precision, recall, dan F1-score berdasarkan hasil prediksi pada data uji. Evaluasi ini bertujuan untuk membandingkan kinerja model terhadap variasi proporsi data latih dan data uji.



Gambar 6. Confusion Matrix (Split 80:20)

Laporan Klasifikasi:				
	precision	recall	f1-score	support
0	0.75	0.86	0.80	7
1	0.92	0.85	0.88	13
accuracy			0.85	20
macro avg	0.83	0.85	0.84	20
weighted avg	0.86	0.85	0.85	20

Gambar 7. Laporan Klasifikasi Rasio 80:20

Tabel 2. Tabel Perbandingan Hasil Rasio

Rasio	Akurasi	Precision	Recall	F1-Score
70:30	0.83	0.75	0.82	0.78
80:20	0.85	0.75	0.86	0.80

Model dilatih menggunakan dua skenario pembagian data: 70:30 dan 80:20. Perbandingan hasil menunjukkan bahwa pembagian data latih dan uji mempengaruhi kinerja model:

1. Akurasi meningkat dari 83% pada rasio 70:30 menjadi 85% pada rasio 80:20. Ini menunjukkan bahwa dengan lebih banyak data latih (80%), model memiliki pemahaman yang lebih baik terhadap pola data sehingga meningkatkan ketepatan prediksi secara keseluruhan.
2. Precision tetap sama pada kedua rasio yaitu 0.75. Artinya, dari seluruh prediksi yang diklasifikasikan sebagai "Lulus" (positif), sebesar 75% benar-

benar lulus. Nilai ini stabil, menunjukkan bahwa perubahan rasio tidak terlalu mempengaruhi jumlah prediksi positif yang benar.

3. Recall meningkat dari 0.82 pada rasio 70:30 menjadi 0.86 pada 80:20. Artinya, model dengan lebih banyak data latih mampu mendeteksi lebih banyak kasus "Lulus" yang benar secara proporsional, sehingga menurunkan jumlah kesalahan tipe II (False Negative).
4. F1-Score adalah harmonisasi antara precision dan recall. Nilai F1 meningkat dari 0.78 menjadi 0.80, mencerminkan peningkatan kinerja model dalam menyeimbangkan presisi dan sensitivitas.

SIMPULAN (PENUTUP)

Penelitian ini berhasil menerapkan algoritma Decision Tree Classifier dengan metode CART untuk mengklasifikasikan risiko gagal studi mahasiswa berdasarkan data akademik dan non-akademik. Model yang dikembangkan mempertimbangkan atribut-atribut seperti IPK, masa studi, penyelesaian tugas akhir, status KKN/PKL, dan poin SKPM, yang terbukti menjadi faktor dominan dalam mempengaruhi prediksi kelulusan. Setelah melalui tahap pra-pemrosesan data yang meliputi pembersihan dan transformasi, model dievaluasi menggunakan dua skenario rasio pembagian data, yaitu 70:30 dan 80:20. Hasil evaluasi menunjukkan bahwa rasio 80:20 memberikan performa yang sedikit lebih baik dengan akurasi 85%, precision 0.75, recall 0.86, dan F1-score 0.80, dibandingkan rasio 70:30 yang mencapai akurasi 83%. Peningkatan akurasi dan recall pada rasio 80:20 menunjukkan bahwa model memiliki pemahaman yang lebih baik terhadap pola data dengan lebih banyak data latih, sehingga mampu mendeteksi lebih banyak kasus "Lulus" yang benar dan mengurangi kesalahan tipe II. Meskipun Decision Tree mudah diinterpretasikan dan mampu menangani fitur kategorikal dan numerik, penelitian ini juga mengidentifikasi potensi kerentanan

terhadap overfitting dan sensitivitas terhadap perubahan data. Secara keseluruhan, model ini memberikan kontribusi signifikan sebagai sistem deteksi dini proaktif untuk membantu institusi pendidikan merancang intervensi yang tepat sasaran guna mencegah tingginya angka gagal studi.

DAFTAR PUSTAKA

- Aesy, U. S., Lahitani, A. R., & Diwangkara, T. W. (2021). Deteksi dini mahasiswa drop out menggunakan C5.0. *JISKA: Jurnal Informatika Sunan Kalijaga*, 6(2), 113–119. from <https://doi.org/10.14421/jiska.2021.6.2.113-119>
- Amani, N. N., & Hayati, U. (2024). Penggunaan algoritma decision tree untuk prediksi prestasi siswa di Sekolah Dasar Negeri 3 Bayalangu Kidul. *Jurnal Mahasiswa Teknik Informatika*, 8(1), 1–7. from <https://journal.upgripnk.ac.id/index.php/saintek/article/view/7696>
- Anggraini, P. S. (2021). Implementasi metode K-Means untuk pengelompokan kasus COVID-19 tingkat provinsi di Indonesia. *Jurnal Informatika dan Fisika (JIFOSI)*, 3(1), 82–90. from <http://dx.doi.org/10.33005/jifosi.v3i1.472>
- Esananda, S. C., Nugroho, B., & Anggraeny, F. (2021). Penerapan algoritma decision tree dalam menentukan prestasi akademik siswa. *Router: Jurnal Informatika dan Teknologi Informasi*, 2(2), 1–9. from <https://journal.aptii.or.id/index.php/Router/article/view/389>
- Gunawan, M. T., Tine, J. Y., & Murwaningtyas, C. E. (2024). Model decision tree untuk prediksi prestasi akademik matematika siswa kelas VIII SMP Frater Don Bosco Manado. *Jurnal Pendidikan Informatika dan Sains*, 13(2), 141–153. from <https://ejournal.itn.ac.id/index.php/jati/article/view/8355>
- Gustian, A. S. (2025). Analisis klasifikasi risiko dropout mahasiswa menggunakan algoritma Decision Tree dan Random Forest. *Jurnal Teknologi Pendidikan*, 11(1), 12–22. from https://www.researchgate.net/publication/391857423_Analisis_Klasifikasi_Risiko_Dropout_Mahasiswa_Menggunakan_Algoritma_Decision_Tree_dan_Random_Forest
- Kristyawan, A., & Sumirat, E. (2019). Analisis terhadap faktor-faktor yang mempengaruhi mahasiswa gagal studi menggunakan teknik klasifikasi. *Jurnal Sistem Informasi dan Bisnis Cerdas*, 9(1), 1–10. from <http://ejournal.upnjatim.ac.id/index.php/sibc/article/view/12025>
- Nasution, W. S., & Sriani. (2025). Analysis of student dropouts at State Senior High School 6 Padangsidempuan City using the C5.0 decision tree algorithm. *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, 22(1), 64–75. from <https://doi.org/10.33751/komputasi.v22i1.18>
- Noviyanto, H., & Fauzi, A. (2024). Penerapan metode data mining untuk prediksi dini kegagalan mahasiswa menggunakan algoritma Decision Tree. *SPIRIT*, 16(2), 334–339. from <https://jurnal.stmik-yadika.ac.id/index.php/spirit/article/view/367>
- Rismaya, R., Yuniarto, D., & Setiadi, D. (2023). Penerapan algoritma machine learning dalam prediksi prestasi akademik mahasiswa. *Router: Jurnal Teknik Informatika dan Terapan*, 3(1), 45–54. from <https://journal.aptii.or.id/index.php/Router/article/view/389>
- Rizalno, M. F., Johar, A., & Coastera, F. F. (2022). Analisis prediksi masa studi mahasiswa

- menggunakan metode decision tree dengan penerapan algoritme CART (Classification and Regression Trees). *Rekursif: Jurnal Informatika*, 10(1), 96–106. from <https://doi.org/10.33369/rekursif.v10i1.21362>
- Suriani, U. (2023). Penerapan data mining untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma decision tree C4.5. *Journal of Computer and Information Systems Ampera*, 4(2), 55–61. from <https://www.journal-computing.org/index.php/journal-cisa/article/view/393>
- Suryani, N., & Maryanto, H. (2022). Faktor-faktor penyebab mahasiswa putus studi pada masa pandemi COVID-19. *Jurnal Kajian Pendidikan dan Pembelajaran*, 8(1), 59–68. from <https://journal.unesa.ac.id/index.php/jkpp/article/view/17893>
- Muriyatmoko, D., Musthafa, A., & Salman, S. A. I. (2024). *Klasifikasi profil mahasiswa pada pola nilai AKPAM dengan metode Decision Tree CART*. *Jurnal Sistem Informasi UNIDA Gontor*, 5(2), 55–67. from <https://corisindo.utb-univ.ac.id/index.php/penelitian/article/view/158>
- Widodo, M. (2019). Kinerja algoritma Classification and Regression Tree (CART) dalam mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. *Jurnal Pinter*, 3(2), 139–145. from <https://doi.org/10.21009/pinter.3.2.9>
- Hafizh Athallah. (2024). Dataset Kelulusan Mahasiswa. *Kaggle*. from <https://www.kaggle.com/datasets/hafizhathallah/kelulusan-mahasiswa>