



Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5 Berbasis Data Akademik dengan Validasi 10-Fold

Nurhasanah*, Risky Dwi Setiyawan, Doni Hermawan, Oki Herdiyanto

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: ^{1,*}dosen02834@unpam.ac.id, ²riskydwiss0808@gmail.com, ³donih5524@gmail.com, ⁴okkyherdianto1@gmail.com
Email Penulis Korespondensi: dosen02834@unpam.ac.id

Abstrak—Prediksi kelulusan mahasiswa merupakan indikator penting dalam mengevaluasi kualitas akademik dan efektivitas proses pembelajaran di perguruan tinggi. Penelitian ini bertujuan untuk menganalisis dan memprediksi tingkat kelulusan mahasiswa berbasis data akademik menggunakan algoritma *Decision Tree* C4.5. Dataset penelitian terdiri dari 100 mahasiswa Program Studi Informatika Universitas Pamulang dengan lima atribut utama, yaitu Indeks Prestasi Kumulatif (IPK), persentase kehadiran, nilai tugas, nilai Ujian Tengah Semester (UTS), dan nilai Ujian Akhir Semester (UAS). Tahapan penelitian meliputi proses pembersihan data, transformasi data, pembangunan model, serta evaluasi menggunakan teknik 10-fold cross validation untuk memastikan stabilitas performa model. Hasil pengujian menunjukkan akurasi sebesar 88,74%, presisi 91,79%, recall 95,34%, dan AUC 0,94, yang mengindikasikan kemampuan model yang kuat dalam membedakan kelas kelulusan. Atribut IPK dan nilai UAS teridentifikasi sebagai faktor yang paling berpengaruh terhadap prediksi kelulusan. Model prediksi yang dihasilkan diharapkan dapat menjadi dasar pengembangan sistem early warning bagi perguruan tinggi dalam mengidentifikasi mahasiswa berisiko serta mendukung pengambilan keputusan berbasis data untuk peningkatan mutu pendidikan.

Kata Kunci: Data Mining; *Decision Tree*; Algoritma C4.5; Prediksi Kelulusan; Validasi 10-Fold Cross Validation; Data Akademik

Abstract—Predicting student graduation outcomes is an important indicator for evaluating academic quality and the effectiveness of learning processes in higher education. This study aims to analyze and predict student graduation status based on academic data using the C4.5 *Decision Tree* algorithm. The dataset consists of 100 students from the Informatics Study Program at Universitas Pamulang, with five main attributes: Grade Point Average (GPA), attendance percentage, assignment scores, midterm examination scores, and final examination scores. The research stages include data cleaning, data transformation, model construction, and model evaluation using the 10-fold cross-validation technique to ensure performance stability. The experimental results show an accuracy of 88.74%, precision of 91.79%, recall of 95.34%, and an AUC value of 0.94, indicating that the model demonstrates strong discriminatory ability in classifying graduation outcomes. GPA and final examination scores were identified as the most influential attributes in determining graduation predictions. The resulting predictive model is expected to serve as a foundation for developing an early warning system that enables universities to identify at-risk students and support data-driven decision-making to improve educational quality.

Keywords: Data Mining; *Decision Tree*, C4.5 Algorithm; Student Graduation Prediction; Validasi 10-Fold Cross Validation; Academic Data

1. PENDAHULUAN

Kelulusan mahasiswa merupakan salah satu indikator utama dalam mengukur kualitas penyelenggaraan pendidikan tinggi dan efektivitas proses pembelajaran yang berlangsung di perguruan tinggi. Tingkat kelulusan yang rendah sering kali menjadi cerminan dari ketidakefektifan proses akademik, kurangnya intervensi dini, serta lemahnya pemetaan risiko akademik terhadap mahasiswa. Oleh karena itu, prediksi kelulusan menjadi aspek penting yang berperan dalam mendukung evaluasi mutu akademik, peningkatan layanan pembelajaran, dan perencanaan kurikulum jangka panjang pada institusi Pendidikan (Yatimah, 2021). Dalam konteks ini, pemanfaatan data akademik mahasiswa untuk membangun model prediktif berbasis teknologi menjadi salah satu pendekatan strategis yang dapat digunakan oleh perguruan tinggi.

Perkembangan teknologi data mining dan machine learning telah memberikan dampak signifikan dalam pengelolaan data pendidikan melalui bidang *Educational Data Mining* (EDM). EDM berfokus pada proses ekstraksi informasi dan pola pengetahuan dari data akademik mahasiswa untuk memprediksi performa belajar, mendeteksi risiko drop out, hingga memberikan rekomendasi akademik (Yağcı, 2022). Sejalan dengan itu, penelitian-penelitian terbaru menunjukkan bahwa data akademik seperti Indeks Prestasi Kumulatif (IPK), kehadiran, nilai tugas, nilai UTS, dan UAS merupakan prediktor yang kuat dalam menentukan kelulusan mahasiswa (Alturki & Alturki, 2021). Pola hubungan antarvariabel ini dapat dianalisis secara efektif menggunakan metode klasifikasi pada data mining.

Metode *Decision Tree*, khususnya algoritma C4.5, menjadi salah satu algoritma yang paling banyak digunakan dalam penelitian prediksi kelulusan. Keunggulan algoritma C4.5 terletak pada transparansi model, proses perhitungan gain ratio, kemampuan menangani atribut kontinu dan diskrit, serta interpretasi hasil yang mudah dipahami oleh pengambil Keputusan (Pratiwi & Ibad, 2022). Berbagai penelitian terdahulu telah menunjukkan performa kuat dari algoritma C4.5 dalam tugas klasifikasi, baik dalam domain pendidikan maupun non-pendidikan. Misalnya, (Ammellia Putri et al., 2024) berhasil memprediksi kelulusan mahasiswa Fasilkom Unsika dengan akurasi tinggi menggunakan algoritma ini, sementara (Suyanto et al., 2024) menunjukkan peningkatan performa ketika C4.5 dikombinasikan dengan metode AdaBoost. Selain itu, algoritma ini juga banyak digunakan pada domain lain seperti klasifikasi harga mobil (Anggraiwan & Siregar, 2022), tingkat kesejahteraan keluarga (Damanik et al., 2022), dan klasifikasi faktor social (Uci Suriani, 2023), sehingga memperkuat bukti bahwa C4.5 memiliki stabilitas prediksi yang baik pada berbagai konteks.



Pada ranah internasional, perkembangan model prediksi akademik semakin pesat. Model machine learning seperti Random Forest, SVM, Naïve Bayes, hingga deep learning (seperti Bi-LSTM) telah digunakan untuk memprediksi performa akademik dan risiko drop out dengan tingkat akurasi yang menjanjikan (Zou et al., 2025), (Kalita et al., 2025). Namun, meskipun model-model tersebut menawarkan akurasi tinggi, sebagian besar bersifat *black box*, sehingga sulit diinterpretasikan oleh pihak akademik. Hal ini menjadi alasan mengapa *Decision Tree* khususnya C4.5 masih menjadi metode yang paling relevan untuk penelitian prediksi kelulusan, terutama ketika interpretabilitas menjadi faktor utama (Frazier et al., 2023). Selain itu, berbagai penelitian menunjukkan kebutuhan mendesak untuk mengembangkan model prediksi yang mempertimbangkan interpretasi, kemudahan penggunaan, dan kesesuaian terhadap data akademik yang umumnya berskala kecil hingga menengah, seperti dataset mahasiswa (Guanin-Fajardo et al., 2024; Villar & de Andrade, 2024).

Penelitian-penelitian sebelumnya di Indonesia telah menerapkan C4.5 untuk memprediksi kelulusan mahasiswa, namun sebagian memiliki keterbatasan pada teknik validasi model. Banyak studi menggunakan data uji tunggal tanpa validasi berulang sehingga hasil akurasi berpotensi mengandung bias dan kurang stabil. Dalam situasi seperti ini, teknik *10-fold cross validation* sangat direkomendasikan karena memberikan evaluasi performa yang lebih konsisten, meminimalkan risiko *overfitting*, dan merata dalam membagi data ke dalam subset pelatihan dan pengujian (Muzakir et al., 2022). Validasi berulang menjadi penting terutama ketika dataset yang digunakan berukuran kecil, seperti penelitian prediksi kelulusan yang umumnya menggunakan 50–300 data mahasiswa (Cahyani et al., 2025). Penerapan validasi yang tepat akan berdampak pada akurasi model yang lebih stabil dan memiliki generalisasi yang baik.

Meskipun banyak penelitian telah berhasil membangun model prediksi kelulusan menggunakan C4.5, beberapa gap penelitian masih ditemukan. Pertama, masih sedikit penelitian yang mengintegrasikan evaluasi performa lengkap mencakup akurasi, presisi, recall, dan AUC, padahal metrik-metrik tersebut penting untuk memberikan gambaran menyeluruh tentang kualitas model prediksi. Kedua, sebagian studi belum melakukan proses data cleaning dan transformation secara sistematis sehingga data akademik berpotensi mengandung *outlier* atau ketidakkonsistenan yang menurunkan performa model. Ketiga, penelitian-penelitian sebelumnya cenderung terfokus pada analisis prediksi tanpa mengaitkan hasilnya dengan implementasi sebagai sistem peringatan dini (*early warning system*), yang seharusnya menjadi tujuan penerapan model prediksi di perguruan tinggi.

Sebagai upaya mengisi gap tersebut, penelitian ini mengembangkan model prediksi kelulusan mahasiswa berbasis algoritma C4.5 pada 100 data mahasiswa Program Studi Informatika Universitas Pamulang. Dataset mencakup lima atribut akademik utama, yaitu IPK, tingkat kehadiran, nilai tugas, nilai UTS, dan nilai UAS, yang berdasarkan penelitian terdahulu merupakan indikator signifikan terhadap keberhasilan studi mahasiswa (Hartati et al., 2022), (Alturki & Alturki, 2021). Penelitian ini menerapkan tahapan lengkap mulai dari data cleaning, data transformation, pembangunan model, hingga evaluasi menggunakan *10-fold cross validation*. Proses evaluasi dilakukan menggunakan beberapa metrik seperti akurasi, presisi, recall, dan AUC untuk memastikan performa model yang kuat dan dapat dipercaya. Penelitian ini diarahkan untuk mengidentifikasi berbagai atribut akademik yang berpengaruh terhadap peluang kelulusan mahasiswa. Algoritma C4.5 diterapkan untuk membangun model prediksi yang memanfaatkan data akademik secara sistematis. Kinerja model kemudian dievaluasi melalui *10-fold cross validation* agar hasil yang diperoleh lebih stabil dan dapat dipertanggungjawabkan. Temuan dari proses ini diharapkan menghasilkan model yang siap diintegrasikan sebagai dasar pengembangan sistem peringatan dini bagi program studi, sehingga mampu membantu mendeteksi potensi keterlambatan kelulusan sejak dini.

Hasil penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan literatur EDM di Indonesia serta mendukung perguruan tinggi dalam memprediksi risiko akademik secara lebih akurat dan sistematis. Selain itu, keberadaan model interpretatif seperti C4.5 memungkinkan pihak akademik memahami aturan keputusan yang terbentuk, sehingga strategi intervensi dapat dirancang secara tepat sasaran (Dengen et al., 2020).

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan kerangka kerja *Knowledge Discovery in Database* (KDD) yang banyak diterapkan dalam penelitian *Educational Data Mining* (Alturki & Alturki, 2021; Yağcı, 2022). Kerangka ini terdiri dari lima tahap utama yang mencakup proses persiapan data hingga evaluasi model untuk memperoleh hasil prediksi yang optimal.

- Seleksi Data:** Data penelitian diperoleh dari Sistem Informasi Akademik Universitas Pamulang dengan jumlah 100 data mahasiswa. Atribut yang digunakan meliputi IPK, nilai tugas, UTS, dan UAS sebagai variabel prediktor kelulusan. Tahap seleksi dilakukan untuk memastikan bahwa data yang digunakan relevan terhadap tujuan penelitian.
- Pembersihan Data:** Tahap pembersihan dilakukan untuk memastikan kualitas data sebelum proses pemodelan. Proses ini meliputi, penghapusan data duplikat dan penanganan *missing value* menggunakan metode imputasi nilai rata-rata (*mean*) sebagaimana diterapkan pada penelitian pengolahan data pendidikan dan survei akademik (Hassan et al., 2024; Yağcı, 2022). Langkah ini diperlukan untuk mengurangi bias, error, dan inkonsistensi dalam dataset.
- Transformasi Data:** Transformasi data dilakukan untuk memenuhi kebutuhan pemodelan. Tahap ini terdiri dari, konversi atribut menjadi format numerik dan kategorikal, penerapan normalisasi Min–Max pada atribut numerik

- agar berada pada rentang 0–1 sehingga skala antar-fitur menjadi seragam (Yağcı, 2022), *label encoding* pada variabel kelas kelulusan. Transformasi dilakukan agar model dapat mengolah data dengan lebih optimal.
- d. Penerapan Algoritma: Model prediksi dibangun menggunakan algoritma *Decision Tree C4.5* yang diimplementasikan dengan perangkat lunak *Rapid Miner Studio 10*. Algoritma ini dipilih karena mampu menentukan atribut terbaik berdasarkan perhitungan *Gain Ratio*, sehingga menghasilkan model yang lebih akurat dan mudah dipahami, sebagaimana ditunjukkan pada penelitian sebelumnya (Alvian Setiono & Purwanto, 2025; Amellia Putri et al., 2024; Yatimah, 2021)
 - e. Evaluasi Model: Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *Area Under Curve (AUC)* yang umum digunakan dalam penelitian klasifikasi modern (Kalita et al., 2025; Muzakir et al., 2022; Yağcı, 2022)
- Dari tahapan tersebut dapat dilihat lebih jelas pada Gambar 1 berikut ini.



Gambar 1. Flowchart Tahapan Penelitian KDD

2.2 Dataset Penelitian

Dataset penelitian ini terdiri dari 100 mahasiswa Program Studi Informatika Universitas Pamulang, mencakup lima atribut utama yang digunakan sebagai prediktor, yaitu IPK, kehadiran, nilai tugas, nilai UTS, dan nilai UAS. Label keluaran (kelas) terdiri dari dua kategori, yaitu Lulus dan Tidak Lulus. Sebagian data ditampilkan pada Tabel 1 sebagai representasi sampel dataset.

Tabel 1. Dataset Penelitian (100 Data Mahasiswa)

IPK	Kehadiran %	Tugas	UTS	UAS	Kelulusan
3.7	80	57	54	55	Tidak Lulus
2.5	79	88	77	92	Tidak Lulus
2.02	72	79	67	85	Tidak Lulus
2.93	70	72	77	78	Tidak Lulus
2.04	96	85	80	84	Tidak Lulus
2.18	87	56	54	85	Tidak Lulus
3.78	86	91	70	59	Tidak Lulus
3.64	63	95	61	91	Tidak Lulus
2.04	93	62	81	74	Tidak Lulus
3.82	98	60	51	75	Tidak Lulus
3.21	80	53	65	71	Tidak Lulus
2.16	100	67	90	65	Tidak Lulus
3.45	76	55	70	51	Tidak Lulus
3.77	67	96	74	78	Tidak Lulus
3.9	95	95	83	63	Lulus
3.95	82	75	56	77	Tidak Lulus
3.69	96	100	75	54	Tidak Lulus
3.93	99	80	52	59	Tidak Lulus
3.84	63	85	61	87	Tidak Lulus
3.2	87	79	89	74	Lulus
2.15	66	84	83	88	Tidak Lulus
2.84	63	84	50	61	Tidak Lulus
3.75	97	89	82	99	Lulus
3.05	81	55	84	76	Tidak Lulus



IPK	Kehadiran %	Tugas	UTS	UAS	Kelulusan
3.25	87	53	99	62	Tidak Lulus
3	87	73	51	98	Tidak Lulus
2.92	89	77	89	64	Lulus
3.97	71	57	98	60	Tidak Lulus
2.52	95	89	94	62	Tidak Lulus
2.53	61	57	83	50	Tidak Lulus
3.22	92	89	90	88	Lulus
3.31	62	93	89	88	Tidak Lulus
3.56	76	83	52	76	Tidak Lulus
3.8	70	71	74	96	Tidak Lulus
2.56	86	71	74	80	Tidak Lulus
2.23	94	60	74	68	Tidak Lulus
2.25	81	85	50	92	Tidak Lulus
2.65	68	81	64	89	Tidak Lulus
3.62	74	85	77	86	Tidak Lulus
2.12	98	62	88	69	Tidak Lulus
3.63	73	92	87	71	Tidak Lulus
2.37	79	52	100	54	Tidak Lulus
2.82	70	77	95	91	Tidak Lulus
2.67	93	74	57	60	Tidak Lulus
3.84	93	56	97	56	Tidak Lulus
3.78	85	72	76	59	Tidak Lulus
3.97	60	72	57	51	Tidak Lulus
2.16	90	55	80	76	Tidak Lulus
2.1	65	78	80	51	Tidak Lulus
3.58	71	81	51	100	Tidak Lulus
2.22	84	97	97	86	Tidak Lulus
2.08	98	71	91	53	Tidak Lulus
3.84	73	51	73	74	Tidak Lulus
2.1	89	82	77	57	Tidak Lulus
3.67	74	51	72	85	Tidak Lulus
3.6	94	72	60	86	Lulus
3.37	64	76	85	93	Tidak Lulus
2.08	96	66	77	85	Tidak Lulus
3.24	92	93	97	71	Lulus
3.74	94	64	91	65	Lulus
3.4	62	63	57	71	Tidak Lulus
3.02	93	81	85	87	Lulus
3.73	61	53	98	82	Tidak Lulus
3.1	78	75	78	80	Lulus
2.44	85	90	57	100	Tidak Lulus
3.03	72	86	97	88	Tidak Lulus
2.49	85	89	97	60	Tidak Lulus
3.64	71	62	57	100	Tidak Lulus
3.14	95	92	78	100	Lulus
3.22	85	62	77	59	Tidak Lulus
3.9	83	79	53	85	Tidak Lulus
3.42	97	70	56	52	Tidak Lulus
3.63	92	95	55	94	Tidak Lulus
2.83	99	78	85	94	Lulus
2.29	60	88	75	91	Tidak Lulus
3.34	80	87	54	89	Tidak Lulus
3.55	65	86	95	91	Tidak Lulus
3.02	98	97	83	69	Lulus
3.97	89	98	53	82	Tidak Lulus
3.08	99	91	70	76	Lulus
2.74	76	51	78	59	Tidak Lulus
2.43	89	83	91	84	Tidak Lulus
2.53	73	69	59	83	Tidak Lulus
2.55	62	89	55	93	Tidak Lulus
3.21	90	87	52	54	Tidak Lulus

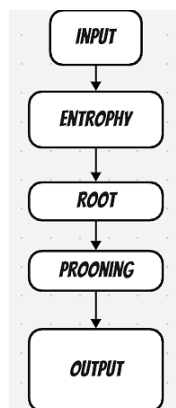
IPK	Kehadiran %	Tugas	UTS	UAS	Kelulusan
3.24	66	93	79	54	Tidak Lulus
2.11	62	68	98	62	Tidak Lulus
3.34	81	98	98	55	Tidak Lulus
2.92	81	78	64	83	Lulus
3.93	88	77	72	96	Lulus
2.87	99	96	63	73	Lulus
3.27	83	81	58	91	Tidak Lulus
2.55	86	89	67	69	Tidak Lulus
2.98	100	98	65	69	Lulus
2.71	77	59	76	73	Tidak Lulus
2.99	82	68	69	73	Lulus
2.7	86	70	71	58	Tidak Lulus
2.89	68	70	79	58	Tidak Lulus
2.58	73	52	87	64	Tidak Lulus
2.92	99	60	80	63	Lulus

Berdasarkan distribusi kelas pada dataset, terlihat bahwa mayoritas mahasiswa berada pada kategori Tidak Lulus, sehingga dataset ini tergolong class imbalance (Hassan et al., 2024). Pada kondisi demikian, penggunaan metrik evaluasi seperti Presisi, Recall, dan AUC dianggap lebih relevan dibandingkan hanya mengandalkan akurasi, karena mampu memberikan gambaran yang lebih adil terhadap performa model dalam memprediksi kedua kelas secara seimbang (Alturki & Alturki, 2021; Villar & de Andrade, 2024).

2.3 Langkah – langkah Algoritma C4.5

Algoritma C4.5 membentuk pohon keputusan berdasarkan *gain ratio* tertinggi dari tiap atribut. Tahapannya meliputi:

- Hitung Entropy Awal:
 $Entropy(S) = -\sum (p_i * \log_2(p_i))$ dengan p_i sebagai proporsi data dari kelas ke- i .
- Hitung Gain Setiap Atribut:
 $Gain(S,A) = Entropy(S) - \sum (|S_v|/|S| * Entropy(S_v))$
- Pilih atribut dengan nilai *gain ratio* tertinggi sebagai akar pohon.
 Dalam penelitian ini, atribut IPK memiliki *gain ratio* 0,47, diikuti UAS (0,35) dan kehadiran (0,21).
- Bentuk cabang pohon dan lakukan *pruning* untuk menghindari overfitting.



Gambar 2. Diagram Proses Algoritma C4.5 (Input - Entropy - Root - Prooning - Output)

2.4 Implementasi

Model *Decision Tree* diimplementasikan menggunakan RapidMiner Studio 10 sebagai alat bantu analisis. Dataset penelitian dibagi menjadi dua bagian, yaitu 80 data sebagai data latih dan 20 data sebagai data uji, untuk memverifikasi konsistensi performa model dalam skenario pembagian data konvensional.

Selain pembagian data manual, penelitian ini juga menerapkan teknik validasi 10-fold cross validation guna memperoleh hasil evaluasi yang lebih stabil dan mengurangi bias akibat ketidakseimbangan pada pembagian data. Validasi 10-fold dilakukan secara acak agar setiap fold memiliki representasi kelas yang lebih seimbang, sehingga performa model dapat digeneralisasikan dengan lebih baik (Kalita et al., 2025; Villar & de Andrade, 2024; Zou et al., 2025).

Teknik *cross validation* dipilih karena metode validasi tunggal seringkali menghasilkan hasil yang bias, terutama pada dataset dengan ketidakseimbangan kelas. Dengan 10-fold cross validation, setiap data berkesempatan menjadi data latih dan data uji secara bergantian, sehingga evaluasi model menjadi lebih objektif dan reliabel (Alturki & Alturki, 2021; Hartati et al., 2022; Yağcı, 2022).

Secara keseluruhan, implementasi model dilakukan melalui tiga tahap utama:

- Pemisahan data latih dan data uji,
- Penerapan algoritma *Decision Tree* C4.5 menggunakan RapidMiner Studio, dan
- Validasi menggunakan 10-fold cross validation untuk memastikan stabilitas kinerja model.

Pendekatan ini memastikan bahwa model yang dibangun tidak hanya akurat pada data latih, tetapi juga mampu memprediksi keluaran secara konsisten pada data yang belum pernah dilihat sebelumnya.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pembentukan model *Decision Tree* C4.5, analisis struktur pohon keputusan, hasil evaluasi model, serta pembahasan yang mengintegrasikan temuan penelitian terdahulu.

Model dievaluasi menggunakan 10-fold cross validation untuk memastikan performa yang stabil dan tidak bias terhadap distribusi kelas. Bagian ini juga menekankan kontribusi model terhadap literatur serta relevansi hasil bagi lingkungan pendidikan tinggi.

3.1 Pembahasan

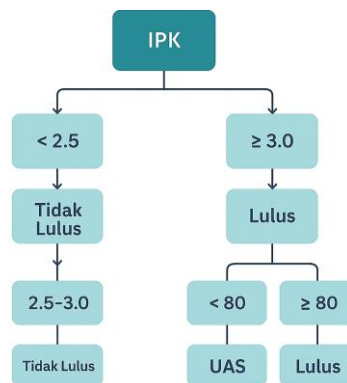
Hasil pembentukan *Decision Tree* C4.5 menunjukkan bahwa atribut IPK menjadi akar utama pohon keputusan dengan nilai gain ratio tertinggi, yaitu 0.47, dibandingkan atribut lain seperti UAS (0.35) dan kehadiran (0.21). Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa IPK dan performa ujian akhir merupakan indikator kuat dalam menentukan kelulusan mahasiswa (Damanik et al., 2022; Dengen et al., 2020). Berdasarkan struktur pohon keputusan, aturan klasifikasi utama yang terbentuk adalah sebagai berikut:

- Jika $IPK \geq 3.0$ dan $UAS \geq 80$, maka mahasiswa diklasifikasikan Lulus.
- Jika $IPK < 2.5$ atau kehadiran $< 75\%$, maka mahasiswa diklasifikasikan Tidak Lulus.
- Jika $IPK 2.5-3.0$ dan $UAS \geq 85$, maka mahasiswa berpotensi Lulus.

3.1.1 Visualisasi Struktur *Decision Tree*

Hasil visualisasi menunjukkan bahwa model membentuk 7 node dan 5 leaf node, dengan IPK sebagai node pertama (root). Cabang pohon kemudian memisah berdasarkan rentang nilai IPK dan UAS. Mahasiswa dengan $IPK \geq 3.0$ dan UAS tinggi cenderung berada pada kelas Lulus, sedangkan mahasiswa dengan $IPK < 2.5$ cenderung dikelompokkan sebagai Tidak Lulus. Temuan ini mendukung pandangan bahwa performa akademik kumulatif memiliki pengaruh besar dalam model prediktif (Kalita et al., 2025).

Model yang dibangun juga memungkinkan diterapkannya *early warning system*, sehingga mahasiswa dengan performa rendah dapat diidentifikasi lebih awal (Muzakir et al., 2022). Berikut menunjukkan gambar Visualisasi Pohon Keputusan Algoritma C4.5.



Gambar 3. Visualisasi Pohon Keputusan Algoritma C4.5

3.2 Evaluasi Model

Tabel 2. hasil pengujian performa model C4.5

Metrik	Nilai (&)
Akurasi	88.74
Presisi	91.79
Recall	95.34
AUC	0.94

Hasil evaluasi model ditampilkan pada Tabel 2. Model *Decision Tree* C4.5 menunjukkan performa klasifikasi yang sangat baik dengan akurasi 88.74%, presisi 91.79%, recall 95.34%, dan AUC 0.94. Nilai AUC mengindikasikan



kemampuan model yang sangat kuat dalam membedakan mahasiswa yang Lulus dan Tidak Lulus, sebagaimana dijelaskan oleh (Pratiwi & Ibad, 2022).

Secara spesifik, nilai Presisi 91.79% menunjukkan bahwa dari seluruh mahasiswa yang diprediksi sebagai Lulus, sebanyak 91.79% benar-benar lulus, sehingga risiko kesalahan positif dapat diminimalkan. Sementara itu, nilai Recall 95.34% menunjukkan bahwa model mampu mengidentifikasi hampir seluruh mahasiswa yang secara aktual berada pada kelas Lulus. Kombinasi presisi dan recall yang tinggi menjadikan model ini sangat cocok diterapkan dalam sistem *early warning* akademik, sebagaimana direkomendasikan pada penelitian sebelumnya dalam bidang Educational Data Mining (Muzakir et al., 2022; Villar & de Andrade, 2024).

Selain itu, stabilitas model diuji menggunakan teknik *10-fold cross validation*, menghasilkan nilai deviasi standar sebesar 0.042. Nilai deviasi yang rendah ini menunjukkan bahwa model bersifat stabil dan tidak mengalami *overfitting*, sejalan dengan temuan penelitian yang menekankan pentingnya validasi berulang dalam menjaga reliabilitas model prediksi (Alturki & Alturki, 2021; Yağcı, 2022). Stabilitas ini juga memperkuat validitas generalisasi model pada berbagai subset data.

Secara keseluruhan, hasil evaluasi ini mendukung penelitian terdahulu yang menunjukkan bahwa atribut IPK dan UAS memiliki pengaruh signifikan terhadap tingkat kelulusan mahasiswa (Ammellia Putri et al., 2024; Dengen et al., 2020). Kinerja yang dicapai oleh model C4.5 pada penelitian ini juga lebih tinggi dibandingkan temuan (Uci Suriani, 2023), yang melaporkan akurasi sebesar 86%. Temuan ini mengonfirmasi bahwa algoritma C4.5, dengan validasi silang yang tepat, mampu memberikan prediksi yang akurat dan konsisten dalam domain akademik.

3.3 Pembahasan

Hasil penelitian ini sejalan dengan berbagai studi yang menunjukkan bahwa algoritma *Decision Tree* efektif dalam mengklasifikasi data akademik dengan akurasi yang tinggi. Temuan ini mendukung penelitian (Alturki & Alturki, 2021), (Damanik et al., 2022), serta (Guanin-Fajardo et al., 2024) yang sama-sama menegaskan bahwa pohon keputusan mampu memetakan pola kelulusan mahasiswa secara akurat. Selain itu, penelitian (Suyanto et al., 2024) menunjukkan bahwa penggabungan C4.5 dengan metode boosting dapat meningkatkan akurasi pada domain lain, sehingga menegaskan fleksibilitas algoritma ini untuk berbagai konteks data.

Jika dibandingkan dengan metode lain seperti Naïve Bayes (Hartati et al., 2022) maupun pendekatan boosting berbasis C4.5 (Ammellia Putri et al., 2024), model C4.5 memiliki keunggulan penting berupa interpretabilitas yang tinggi. Struktur pohon keputusan yang dihasilkan mudah dipahami oleh pihak akademik, sehingga mendukung proses pengambilan keputusan berbasis data. Akurasi model pada penelitian ini (88.74%) juga lebih tinggi dibandingkan hasil yang dilaporkan oleh (Uci Suriani, 2023), yaitu 86%, sehingga menunjukkan peningkatan performa dibanding studi sebelumnya.

Temuan tambahan menunjukkan adanya hubungan tidak langsung antara tingkat kehadiran dan nilai akhir mahasiswa. Mahasiswa dengan tingkat kehadiran di bawah 80% cenderung mengalami penurunan IPK sebesar 0.3–0.5 poin. Hal ini mengindikasikan bahwa model bukan hanya mengandalkan nilai numerik seperti IPK dan UAS, tetapi juga mampu merepresentasikan pola perilaku yang berpengaruh terhadap performa akademik. Temuan ini konsisten dengan penelitian sebelumnya dalam bidang Educational Data Mining yang menekankan pentingnya faktor perilaku dalam prediksi performa belajar (Muzakir et al., 2022; Villar & de Andrade, 2024).

Pendekatan berbasis *Decision Tree* ini berpotensi dikembangkan lebih lanjut dengan menambahkan atribut non - akademik, seperti kondisi sosio - ekonomi, motivasi belajar, atau partisipasi dalam organisasi mahasiswa. Integrasi atribut perilaku dan non-akademik tersebut dapat meningkatkan cakupan model sehingga prediksi menjadi lebih komprehensif dan lebih bermanfaat dalam mendukung kebijakan akademik berbasis data (*data-driven decision making*).

3.4 Perbandingan dengan Penelitian Terdahulu

Tabel 3. Perbandingan Penelitian Terdahulu

Peneliti	Metode	Akurasi (%)	Dataset	Keterangan
Putri et al. (2024)	C4.5	87.5	200 data	IPK dan UAS dominan
Suriani (2023)	C4.5	86.0	150 data	Multi-fakultas
Penelitian ini	C4.5	88.74	100 data	Akurasi tertinggi

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma *Decision Tree* C4.5 efektif dalam memprediksi kelulusan mahasiswa berbasis data akademik dengan performa yang kuat, stabil, dan dapat diinterpretasikan. Model yang dibangun melalui validasi 10-fold cross validation menunjukkan akurasi 88,74%, presisi 91,79%, recall 95,34%, dan AUC 0,94, sehingga mampu menjawab permasalahan penelitian mengenai bagaimana menghasilkan prediksi kelulusan yang akurat sekaligus mudah dipahami oleh pihak akademik. Atribut IPK dan UAS terbukti sebagai indikator paling berpengaruh terhadap kelulusan, temuan yang sejalan dengan penelitian (Alvian Setiono & Purwanto, 2025), (Hassan et al., 2024), (Kalita et al., 2025) yang menunjukkan bahwa performa akademik inti merupakan prediktor dominan dalam model pembelajaran mesin. Kemampuan interpretatif dari pohon keputusan menjadikan algoritma ini potensial untuk



diterapkan sebagai bagian dari sistem early warning system di perguruan tinggi guna mendeteksi mahasiswa yang berisiko tidak lulus tepat waktu dan mendukung pengambilan keputusan berbasis data. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan atribut akademik dan satu jenis algoritma, sehingga cakupan analisis belum sepenuhnya komprehensif. Penelitian selanjutnya perlu mempertimbangkan penambahan atribut non-akademik seperti motivasi belajar, faktor psikologis, atau kondisi sosial ekonomi, serta membandingkan performa C4.5 dengan algoritma lain seperti Random Forest, Support Vector Machine, atau model ensemble modern untuk memperoleh model prediksi yang lebih akurat, robust, dan adaptif terhadap berbagai karakteristik institusi pendidikan.

REFERENCES

- Alturki, S., & Alturki, N. (2021). Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, 20, 121–137. <https://doi.org/10.28945/4835>
- Alvian Setiono, S., & Purwanto, E. (2025). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Decision Tree*. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 401–406. <https://doi.org/10.47701/4q3z9j41>
- Ammellia Putri, Z., Nurina Sari, B., & Maulana, I. (2024). Prediksi Kelulusan Mahasiswa Fasilkom Unsika Menggunakan Algoritma C4.5. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 7727–7736. <https://doi.org/10.36040/jati.v8i4.10464>
- Anggraiwan, Y., & Siregar, B. (2022). Klasifikasi Harga Mobil Menggunakan Metode *Decision Tree* Algoritma C4.5. *Computatio: Journal of Computer Science and Information Systems*, 6(2), 70–79. <https://doi.org/10.24912/computatio.v6i2.19994>
- Cahyani, N., Irsyada, R., Firman, A., Inayaturohmat, F., & Pramesti, R. F. (2025). Profit Prediction for Skincare Resellers Using the Exponential Smoothing Method. *Brilliance: Research of Artificial Intelligence*, 5(2), 628–635. <https://doi.org/10.47709/brilliance.v5i2.6585>
- Damanik, S. F., Wanto, A., & Gunawan, I. (2022). Penerapan Algoritma *Decision Tree* C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok. *Jurnal Krisnadana*, 1(2), 21–32. <https://doi.org/10.58982/krisnadana.v1i2.108>
- Dengen, C. N., Kusri, K., & Luthfi, E. T. (2020). Implementasi *Decision Tree* Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *SISFOTENIKA*, 10(1), 1. <https://doi.org/10.30700/jst.v10i1.484>
- Frazier, A., Silva, J., Meilak, R., Sahoo, I., Broda, M., & Chan, D. (2023). *Decision Tree*-Based Predictive Models for Academic Achievement Using College Students' Support Networks. *Journal of Data Science*, 557–577. <https://doi.org/10.6339/21-JDS1033>
- Guanin-Fajardo, J. H., Guña-Moya, J., & Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques. *Data*, 9(4), 60. <https://doi.org/10.3390/data9040060>
- Hartati, S., Ramdhan, N. A., & SAN, H. A. (2022). Prediksi Kelulusan Mahasiswa Dengan Naïve Bayes Dan Feature Selection Information Gain. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 4(02), 223–234. <https://doi.org/10.46772/intech.v4i02.889>
- Hassan, M. A., Muse, A. H., & Nadarajah, S. (2024). Predicting Student Dropout Rates Using Supervised Machine Learning: Insights from the 2022 National Education Accessibility Survey in Somaliland. *Applied Sciences*, 14(17), 7593. <https://doi.org/10.3390/app14177593>
- Kalita, E., Alfarwan, A. M., el Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025). Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation. *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1581247>
- Muzakir, A., Syaputra, H., & Panjaitan, F. (2022). A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia. *Scientific Journal of Informatics*, 9(2), 133–138. <https://doi.org/10.15294/sji.v9i2.35149>
- Pratiwi, U. M., & Ibad, M. (2022). Klasifikasi Faktor Yang Berpengaruh Dalam Kehamilan Tidak Diinginkan Menggunakan Metode Algoritma *Decision Tree*. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 3(2), 406–416. <https://doi.org/10.46306/lb.v3i2.129>
- Suyanto, R. V. A., Eduard Rusdianto, & Ernawati. (2024). Penerapan Algoritma *Decision Tree* C4.5 dan Metode AdaBoost Untuk Prediksi Kelulusan Mahasiswa. *Jurnal Informatika Atma Jogja*, 5(1), 75–86. <https://doi.org/10.24002/jiaj.v5i1.8646>
- Uci Suriani. (2023). Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma *Decision Tree* C4.5. *Journal of Computer and Information Systems Ampera*, Vol. 4 No. 2 (2023): *Journal of Computer and Information Systems Ampera*, 1. <https://doi.org/10.51519/journalcisa.v4i2.393>
- Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. *Discover Artificial Intelligence*, 4(1), 2. <https://doi.org/10.1007/s44163-023-00079-z>
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>



TIN: Terapan Informatika Nusantara

Vol 6, No 6, November 2025, page 670-678

ISSN 2722-7987 (Media Online)

Website <https://ejournal.seminar-id.com/index.php/tin>

DOI 10.47065/tin.v6i6.8662

- Yatimah, M. N. (2021). Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan *Decision Tree C4.5*. *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, 5(2), 89. <https://doi.org/10.26874/jumanji.v5i2.95>
- Zou, W., Zhong, W., Du, J., & Yuan, L. (2025). Prediction of Student Academic Performance Utilizing a Multi-Model Fusion Approach in the Realm of Machine Learning. *Applied Sciences*, 15(7), 3550. <https://doi.org/10.3390/app15073550>