



Predicting Student Graduation Rates Using the Decision Tree Algorithm

Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree

Riska Septiani^{1*}, Listina Nadhia Ningsih², Angga Pramadjaya³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Indonesia

E-Mail: ¹dosen02912@unpam.ac.id, ²dosen02875@unpam.ac.id, ³anggapramadjaya@unpam.ac.id

Received Mar 24th 2026; Revised Apr 17th 2026; Accepted May 10th 2026; Available Online Jun 20th 2026

Corresponding Author: Riska Septiani

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Student graduation rate is one of the important indicators in assessing the quality of the educational process in higher education institutions. Educational institutions need to analyze students' academic data to predict the likelihood of graduation in order to take strategic actions to improve student success. This study aims to predict student graduation rates using the Decision Tree method. This method is chosen because it is capable of classifying data and producing an easily interpretable model in the form of decision rules. The data used in this study include several academic attributes of students, such as grade point average, length of study, total completed credits, and other supporting factors related to the academic process. The research stages include data collection, data preprocessing, model development using the Decision Tree algorithm, and evaluation of the prediction results. The results of the analysis show that the percentage of students who achieved the "With Honors" distinction in 2025 was 6.01% and increased to 6.67% in 2026, with an average (mean) value of 6.34%. However, based on the prediction results using the Decision Tree method, the percentage of students predicted to achieve the "With Honors" distinction decreased to 3.00%, representing a decline of approximately 3.34% from the previous average. This indicates that, within the analyzed data patterns, there is a tendency for a decrease in the number of students who are able to achieve the highest academic performance.

Keywords: Decision Tree, Graduation Students, Machine Learning, Prediction Level.

Abstrak

Tingkat kelulusan mahasiswa merupakan salah satu indikator penting dalam menilai kualitas proses pendidikan di perguruan tinggi. Institusi pendidikan perlu melakukan analisis terhadap data akademik mahasiswa untuk memprediksi kemungkinan kelulusan sehingga dapat mengambil langkah strategis dalam meningkatkan keberhasilan studi mahasiswa. Penelitian ini bertujuan untuk melakukan prediksi tingkat kelulusan mahasiswa menggunakan metode Decision Tree. Metode ini dipilih karena mampu mengklasifikasikan data serta menghasilkan model yang mudah dipahami dalam bentuk aturan keputusan. Data yang digunakan dalam penelitian ini meliputi beberapa atribut akademik mahasiswa, seperti indeks prestasi, lama masa studi, jumlah SKS yang telah ditempuh, serta faktor pendukung lainnya yang berkaitan dengan proses akademik. Tahapan penelitian meliputi proses pengumpulan data, pra-pemrosesan data (data preprocessing), pembangunan model menggunakan algoritma Decision Tree, serta evaluasi terhadap hasil prediksi yang dihasilkan. Hasil analisis menunjukkan bahwa persentase mahasiswa yang memperoleh predikat "Dengan Pujian" pada tahun 2025 sebesar 6,01% dan meningkat menjadi 6,67% pada tahun 2026, dengan nilai rata-rata (*mean*) sebesar 6,34%. Namun demikian, berdasarkan hasil prediksi menggunakan metode Decision Tree, persentase mahasiswa yang diperkirakan memperoleh predikat Dengan Pujian menurun menjadi 3,00%, atau mengalami penurunan sekitar 3,34% dari rata-rata sebelumnya. Hal ini mengindikasikan bahwa dalam pola data yang dianalisis terdapat kecenderungan penurunan jumlah mahasiswa yang mampu mencapai prestasi akademik tertinggi.

Kata Kunci: Decision Tree, Kelulusan Mahasiswa, Machine Learning, Tingkat Prediksi.

1. PENDAHULUAN

Pendidikan tinggi memiliki peran strategis dalam menghasilkan sumber daya manusia yang kompeten, adaptif, dan berdaya saing global. Salah satu indikator utama keberhasilan penyelenggaraan pendidikan tinggi adalah tingkat kelulusan mahasiswa, baik dari aspek kuantitas, kualitas akademik, maupun ketepatan

waktu penyelesaian studi. Tingkat kelulusan tidak hanya merefleksikan performa individual mahasiswa, tetapi juga mencerminkan efektivitas sistem pembelajaran, kurikulum, layanan akademik, serta tata kelola institusi secara keseluruhan. Dalam sistem penjaminan mutu pendidikan tinggi di Indonesia, indikator kelulusan dan masa studi menjadi komponen penting dalam proses evaluasi dan akreditasi yang dilakukan oleh Badan Akreditasi Nasional Perguruan Tinggi. Selain itu, pelaporan data akademik mahasiswa secara nasional dilakukan melalui Pangkalan Data Pendidikan Tinggi (PDDikti) yang berada di bawah koordinasi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Oleh karena itu, pengelolaan dan analisis data kelulusan mahasiswa menjadi kebutuhan strategis dalam mendukung akuntabilitas, transparansi, dan peningkatan mutu pendidikan tinggi [1]. Namun demikian, tingkat kelulusan mahasiswa dipengaruhi oleh berbagai faktor yang bersifat multidimensional dan kompleks. Faktor-faktor tersebut meliputi Indeks Prestasi Kumulatif (IPK), jumlah SKS yang ditempuh setiap semester, riwayat pengulangan mata kuliah, tingkat kehadiran, latar belakang sosial ekonomi, serta faktor non-akademik lainnya. Kompleksitas variabel ini menyebabkan sulitnya melakukan prediksi kelulusan secara akurat hanya dengan pendekatan konvensional berbasis evaluasi manual atau analisis deskriptif sederhana. Akibatnya, deteksi dini terhadap mahasiswa yang berpotensi mengalami keterlambatan studi atau risiko tidak lulus seringkali terlambat dilakukan.

Penelitian ini bertujuan untuk menganalisis dan memprediksi tingkat kelulusan mahasiswa, khususnya pada kategori kelulusan Dengan Pujian, dengan menggunakan metode Decision Tree. Seiring dengan berkembangnya sistem informasi akademik, perguruan tinggi kini memiliki basis data mahasiswa yang besar dan terstruktur. Data tersebut mencakup riwayat nilai, beban studi, IP semester, status akademik, serta data administratif lainnya [2]. Volume data yang terus bertambah setiap semester membuka peluang penerapan teknik *data mining* untuk menggali pola dan hubungan antarvariabel yang tidak terlihat secara langsung. Pendekatan ini memungkinkan institusi menerapkan sistem pengambilan keputusan berbasis data (*data driven decision making*) [3]. Salah satu metode klasifikasi dalam *data mining* yang banyak digunakan adalah Decision Tree [4]. Metode ini membangun model prediksi dalam bentuk struktur pohon keputusan berdasarkan atribut-atribut yang paling berpengaruh terhadap variabel target. Algoritma seperti ID3, C4.5, maupun CART mampu menentukan atribut dengan nilai information gain atau gain ratio tertinggi untuk membentuk aturan klasifikasi yang mudah diinterpretasikan. Keunggulan utama Decision Tree terletak pada kemampuannya menghasilkan model yang transparan dan mudah dipahami oleh pengambil kebijakan, sehingga hasil analisis dapat langsung diterjemahkan menjadi strategi intervensi akademik [5].

Dalam konteks prediksi tingkat kelulusan mahasiswa, Decision Tree dapat digunakan untuk mengklasifikasikan mahasiswa ke dalam kategori tertentu, seperti lulus tepat waktu, lulus tidak tepat waktu, atau berisiko tidak lulus [6]. Penelitian terkait prediksi tingkat kelulusan mahasiswa menggunakan metode Decision Tree ini menjadi penting karena dapat berfungsi sebagai sistem peringatan dini (*early warning system*) bagi pihak akademik. Penelitian ini juga memiliki urgensi dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*) di lingkungan perguruan tinggi. Model yang dihasilkan dari metode Decision Tree bersifat interpretable, sehingga memudahkan pihak manajemen dalam memahami faktor-faktor dominan yang memengaruhi kelulusan mahasiswa.

Model yang dihasilkan dari metode Decision Tree bersifat interpretable, sehingga memudahkan pihak manajemen dalam memahami faktor-faktor dominan yang memengaruhi kelulusan mahasiswa. Model ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai sistem peringatan dini (*early warning system*) yang membantu pihak program studi atau fakultas dalam merancang kebijakan pembinaan akademik [7]. Misalnya, apabila atribut IP semester awal terbukti menjadi faktor dominan dalam menentukan kelulusan, maka intervensi akademik dapat difokuskan pada mahasiswa tahun pertama melalui program pendampingan atau tutorial tambahan [8]. Meskipun demikian, keberhasilan penerapan metode Decision Tree sangat bergantung pada kualitas data yang digunakan. Data yang tidak lengkap, inkonsisten, atau mengandung noise dapat menurunkan akurasi model prediksi [9]. Oleh karena itu, tahapan pra-pemrosesan data seperti pembersihan data (*data cleaning*), seleksi atribut, transformasi data, dan validasi model menjadi bagian penting dalam penelitian ini [10]. Evaluasi model menggunakan parameter seperti *accuracy*, *precision*, *recall*, dan *confusion matrix* diperlukan untuk memastikan keandalan dan performa sistem prediksi yang dihasilkan [11]. Urgensi penelitian ini semakin relevan dalam konteks transformasi digital pendidikan tinggi dan implementasi kebijakan kampus berdampak, di mana fleksibilitas kurikulum dan mobilitas akademik mahasiswa semakin meningkat [12]. Dinamika tersebut berpotensi memengaruhi pola kelulusan mahasiswa, sehingga dibutuhkan sistem analitik yang adaptif dan responsif terhadap perubahan [13]. Model prediksi berbasis Decision Tree dapat menjadi salah satu solusi inovatif dalam mendukung manajemen akademik yang lebih proaktif dan strategis [14].

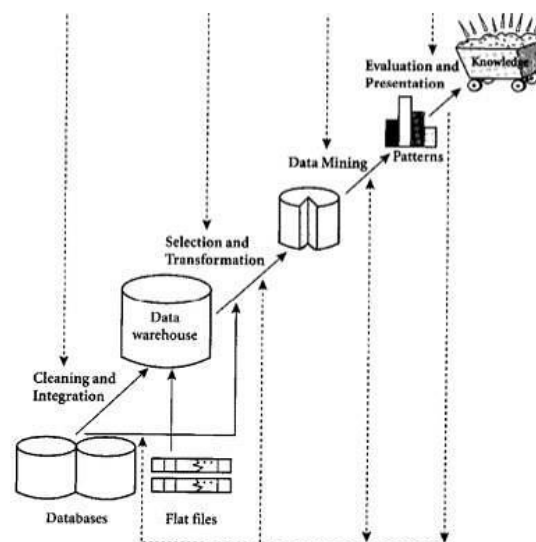
Penelitian sebelumnya membahas metode decision tree yang memiliki performa mahasiswa karena kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan. Salah satu metode yang banyak digunakan dalam prediksi kelulusan mahasiswa adalah Decision Tree. Metode ini termasuk dalam teknik klasifikasi yang membangun model berbentuk struktur pohon untuk menggambarkan hubungan antara variabel *input* dan *output* [4]. Penelitian lain menunjukkan bahwa algoritma Decision Tree mampu menghasilkan tingkat akurasi yang tinggi dalam memprediksi kelulusan mahasiswa [8]. Penelitian oleh Jurnal

Pendidikan Progresif melaporkan bahwa model Decision Tree mampu mencapai akurasi sebesar 88,9% dalam memprediksi kelulusan tepat waktu berdasarkan data performa akademik mahasiswa[15]. Beberapa penelitian juga membandingkan metode Decision Tree dengan algoritma lain seperti Support Vector Machine (SVM), Naïve Bayes, dan Random Forest. Hasil penelitian menunjukkan bahwa meskipun metode lain seperti SVM atau ensemble learning dapat memberikan akurasi lebih tinggi dalam beberapa kasus, Decision Tree tetap unggul dari segi interpretabilitas dan kemudahan implementasi[16].

Berdasarkan uraian tersebut, penelitian mengenai prediksi tingkat kelulusan mahasiswa menggunakan metode Decision Tree menjadi relevan dan penting untuk dilakukan. Penelitian ini diharapkan mampu menghasilkan model klasifikasi yang akurat, interpretatif, dan aplikatif dalam mendukung kebijakan akademik [17]. Dengan pendekatan berbasis analitik data, perguruan tinggi dapat meningkatkan efektivitas pengelolaan mahasiswa, menekan angka keterlambatan kelulusan, serta memperkuat sistem penjaminan mutu pendidikan secara berkelanjutan [18].

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis *data mining*. Pendekatan ini bertujuan untuk membangun model klasifikasi yang mampu memprediksi tingkat kelulusan mahasiswa berdasarkan data historis akademik [19]. Desain penelitian bersifat *predictive modeling*, yaitu membangun model prediksi menggunakan algoritma Decision Tree dan mengevaluasi performanya melalui pengujian statistik[20]. Penelitian ini termasuk dalam kategori penelitian terapan (*applied research*) karena hasil model yang dibangun diharapkan dapat diimplementasikan sebagai sistem pendukung keputusan akademik. Data yang digunakan merupakan data sekunder yang diperoleh dari sistem informasi akademik perguruan tinggi yang terintegrasi dengan pelaporan nasional melalui Pangkalan Data Pendidikan Tinggi di bawah koordinasi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi [21].

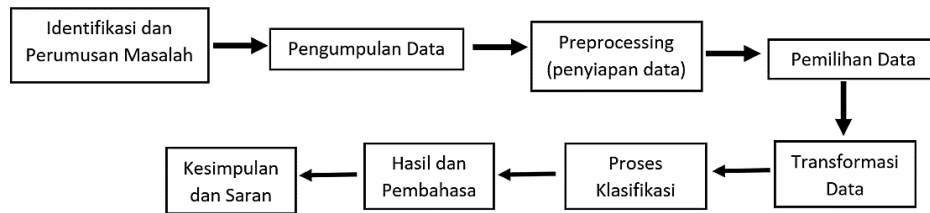


Gambar 1 Metode Analisis

Pada Gambar 1 sebelum penggalian informasi di mulai, perlu adanya penyelesaian data dari sekumpulan *database* yang dilakukan untuk menyelesaikan suatu masalah, supaya masalah tersebut dapat terselesaikan, karena dari *database* yang dikumpulkan dalam *data warehouse* sangat berpotensi untuk mencapai informasi yang sangat berguna [22]. Sebagai *database* yang berada di kampus menjadi *data warehouse* yang sangat penting. Sehingga dapat memberikan informasi yang sangat berguna, yaitu bila diselesaikan dengan akar permasalahan yang sedang dihadapinya [23]. Dalam hal pemilihan variabel yang penting atau signifikan sering digunakan istilah seleksi variabel atau *variable/ feature selection*. Seringkali variabel yang penting untuk pengklasifikasian tidak meliputi semua variabel yang tersedia [16]. Data yang menjadi focus KDD, sebelumnya perlu dilakukan *cleaning* yaitu memeriksa data yang tidak konsisten, memperbaiki kesalahan data, data yang tidak terpakai atau kesalahan penulisan harus di buang atau di hilangkan karena data tersebut digunakan dalam proses *data mining* [24]. Misalnya saja atribut yang tidak di gunakan atau tidak terpilih perlu di buang sedangkan data yang terseleksi perlu disimpan [25]. Pengolahan data awal dari atribut-atribut yang ada dari data mahasiswa aktif, dan data kelulusan. Semua data itu di dapatkan dari bagian RPU. Data set tersebut di olah, yang tidak dipakai Atau tidak digunakan akan di hilangkan. Setiap atribut yang ada dalam *database* bisa diganti dari atribut yang menggunakan teks menjadi numerik[26]. Karena dengan menggunakan numerik maka pengolahan data lebih mudah [27].

2.1. Tahapan Penelitian

Desain penelitian ini dimulai dari beberapa tahapan yang akan di jalankan pada penelitian ini. Dalam melakukan proses kalsifikasi data prediksi terhadap tingkat kelulusan mahasiswa menggunakan pendekatan penelitian Decision Tree. Tools yang di gunakan adalah RpidMiner sebagai alat untuk mengelola data. Tahapan penelitian ini dapat dilihat pada Gambar 2.

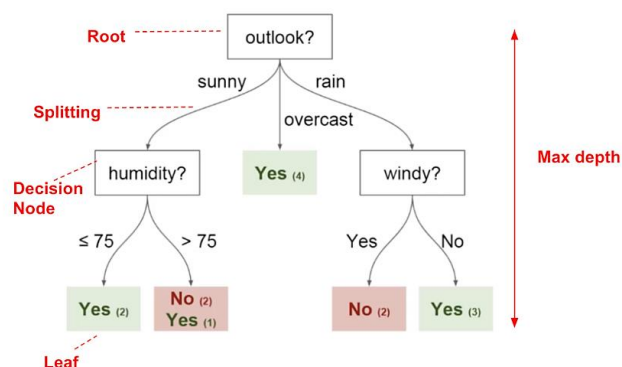


Gambar 2 Tahapan Penelitian

Tahap penelitian yang pertama yaitu menentukan identifikasi dan perumusan masalah pada penelitian [28]. Proses selanjutnya pengumpulan data. Selanjutnya *preprocessing* (penyiapan data) pada proses ini data yang di kumpulkan diolah menggunakan teknik pengolahan data yg lebih efektif dengan menggunakan algoritma decision tree. Pemilihan data : memilih data yang relevan untuk penelitian, pada data ini peneliti menggunakan data berupa Nilai, UTS dan UAS sebagai penentu kelulusan mahasiswa. Proses selanjutnya yaitu proses transformasi data yakni mengubah data kedalam format yang cocok untuk di analisis. Kemudian proses klasifikasi dengan RapidMiner dengan langkah ini melibatkan penggunaan perangkat lunak Rapid miner Untuk menerapkan algoritma Pohon Keputusan untuk mengklasifikasikan data sehingga dapat membantu mengidentifikasi pola dan memprediksi hasil kelulusan mahasiswa. Hasil dan Pembahasan, merupakan tahap pembahasan hasil dari analisis yang di peroleh dari klasifikasi Pohon keputusan. Keakuratan model di bahas dalam kaitannya dengan pernyataan penelitian. Kesimpulan dan saran, proses dimana peneliti menarik kesimpulan dari temuan peringkasan wawasan utama dari penelitian, tahap ini juga dapat memberikan rekomendasi untuk penelitian mendata atau aplikasi potensial dari temuan.

2.2. Algoritma Decision Tree

Decision Tree pada Gambar 3 merupakan salah satu metode klasifikasi dalam bidang *data mining* dan *machine learning* yang digunakan untuk membangun model prediksi berbasis struktur pohon[29]. Metode ini termasuk dalam kategori *supervised learning*, karena proses pembentukan model dilakukan menggunakan dataset yang telah memiliki label atau kelas target [30]. Model Decision Tree direpresentasikan dalam bentuk struktur hierarkis yang terdiri atas *Root Testing* (Akar Pohon) dengan atribut pertama yang digunakan untuk membagi dataset. *Internal Testing* (Testing Cabang) yaitu atribut lanjutan yang digunakan untuk membagi subset data dan *Leaf Testing* (Daun) hasil klasifikasi atau keputusan akhir. Adapun tahapan dalam metode decision tree ialah dengan memilih atribut sebagai akar, buat cabang untuk masing-masing nilai, bagi kasus dalam cabang, ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.



Gambar 3 Metode Decision Tree

Pemilihan atribut sebagai akar didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera pada persamaan 1.

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

Keterangan:

S : Himpunan kasus A : Atribut
 n : Jumlah partisi atribut
 $|S_i|$: Jumlah kasus pada partisi ke i
 $|S|$: Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar ditunjukkan pada Persamaan 2.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

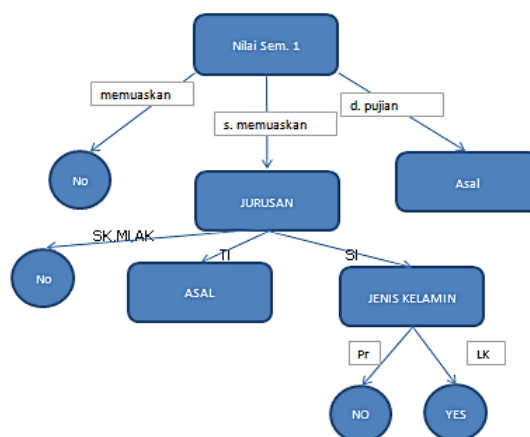
S : Himpunan Kasus
 n : Jumlah partisi S
 p_i : Proporsi dari S_i terhadap

2. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk memprediksi tingkat kelulusan mahasiswa menggunakan metode Decision Tree berdasarkan beberapa variabel akademik yang mempengaruhi keberhasilan studi mahasiswa. Data yang digunakan dalam penelitian ini berasal dari data akademik mahasiswa yang meliputi beberapa atribut, seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS yang ditempuh, tingkat kehadiran, nilai mata kuliah inti, serta lama masa studi mahasiswa. Tahapan awal penelitian dilakukan melalui proses *preprocessing* data yang meliputi pembersihan data, seleksi atribut, serta transformasi data agar sesuai dengan kebutuhan analisis. Selanjutnya data dibagi menjadi dua bagian, yaitu data *training* dan *testing*, dengan tujuan untuk membangun model prediksi dan menguji tingkat akurasi model yang dihasilkan.

2.1. Proses Decision Tree

Gambar 4 merupakan visualisasi model klasifikasi menggunakan metode Decision Tree yang digunakan untuk memprediksi tingkat kelulusan mahasiswa berdasarkan beberapa atribut akademik dan karakteristik mahasiswa. Model ini menggambarkan proses pengambilan keputusan secara hierarkis melalui serangkaian *testing* yang merepresentasikan atribut, sedangkan cabang menunjukkan kondisi atau nilai dari atribut tersebut yang mengarah pada hasil klasifikasi akhir.

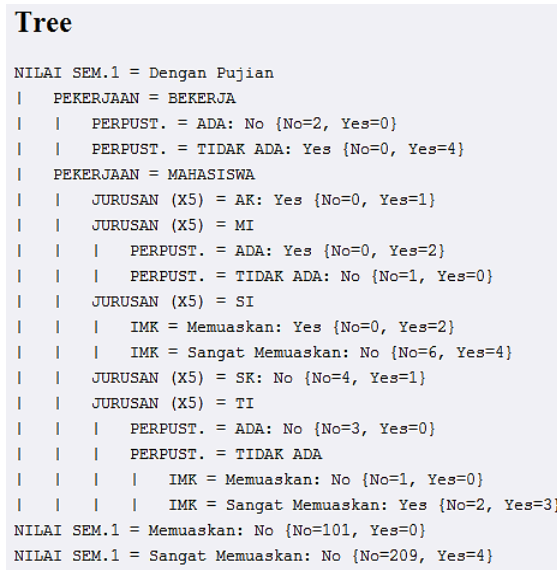


Gambar 4 Pohon Keputusan Hasil Perhitungan

Testing paling atas pada pohon keputusan adalah Nilai Semester 1 (Nilai Sem. 1). Posisi ini menunjukkan bahwa variabel tersebut merupakan atribut yang paling dominan dalam menentukan prediksi kelulusan mahasiswa. Hal ini mengindikasikan bahwa performa akademik mahasiswa pada semester pertama memiliki pengaruh besar terhadap kemungkinan keberhasilan studi mahasiswa hingga lulus. Pada *testing* ini terdapat tiga kategori nilai, yaitu memuaskan, sangat memuaskan dengan pujian. Adapun *rule* hasil yang di implementasikan ialah :

1. Jika nilai Semester 1 memuaskan maka *No*
2. Jika nilai Semester 1 sangat memuaskan maka *No*
3. Jika nilai semester 1 dengan pujian, pekerjaan bekerja dan perpustakaan ada maka *No*
4. Jika nilai semester 1 dengan pujian, pekerjaan bekerja dan perpustakaan tidak ada maka *Yes*
5. Jika nilai semester 1 dengan pujian, pekerjaan mahasiswa dan jurusan AK maka *Yes*

6. Jika nilai semester 1 dengan pujian, pekerjaan Mahasiswa, perpustakaan ada, jurusan MI maka *Yes*



Gambar 5. Alur Proses Decision Tree

Hasil dari pemrosesan data yang ada pada Gambar 5 yaitu tentang mahasiswa yang dinyatakan dengan pujian maka dapat disimpulkan bahwa *rule* yang pertama dalam mahasiswa yang dinyatakan dengan pujian yaitu nilai semester 1 sebagai akar pertama. Selanjutnya di susul dengan atribut yang lainnya, yaitu Asal atau Kota Kelahiran, Jenis Kelamin, Pekerjaan, Jurusan, Perpustakaan, dan IMK.

Berdasarkan dari pola yang sudah didapat dari data historis, maka *rule* yang digunakan ada 13 *rule* yang menghasilkan predikat dengan pujian atau *yes* = 15 atau 3% atau sebanyak 15 orang dari 500 sampel mahasiswa aktif dan Predikat Memuaskan atau *No* = 485 atau 97.00% dari 500 sampel mahasiswa aktif. Adapun alur proses Decision Tree yang digunakan ialah:

1. Pengumpulan Data (*Data Collection*) Tahap pertama adalah mengumpulkan data akademik mahasiswa dari sistem informasi akademik atau *database* perguruan tinggi. Data yang dikumpulkan biasanya meliputi:
 - a. NIM / Identitas Mahasiswa
 - b. Jenis Kelamin
 - c. Indeks Prestasi Semester (IPS)
 - d. Indeks Prestasi Kumulatif (IPK)
 - e. Jumlah SKS yang ditempuh
 - f. Status keaktifan mahasiswa
 - g. Lama masa studi
 - h. Status kelulusan (tepat waktu / tidak tepat waktu)

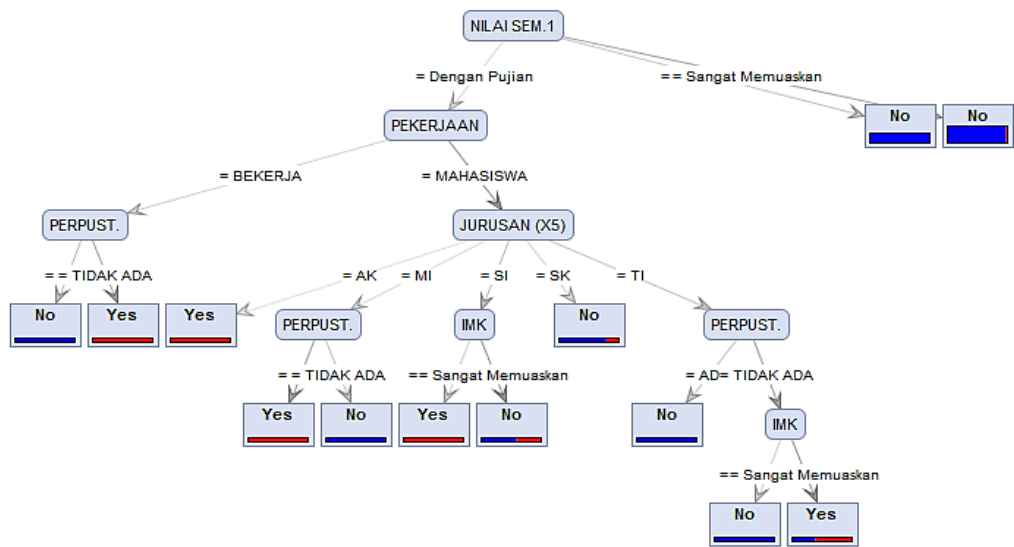
Data historis ini akan menjadi dataset utama dalam proses pembelajaran model.

2. *Preprocessing* Data Tahap *preprocessing* bertujuan untuk mempersiapkan data agar dapat diproses oleh algoritma decision tree. Proses ini meliputi:
 - a. *Data cleaning* Menghapus data yang tidak lengkap (*missing value*) atau tidak konsisten.
 - b. *Data Transformation* Mengubah data numerik menjadi kategori jika diperlukan, misalnya: $IPK \geq 3.50 \rightarrow$ Sangat Baik $IPK 3.00 - 3.49 \rightarrow$ Baik $IPK < 3.00 \rightarrow$ Cukup
 - c. *Data Selection* Memilih atribut yang paling relevan untuk prediksi kelulusan.
3. Pembagian *Dataset* (*Data Splitting*) *Dataset* kemudian dibagi menjadi dua bagian utama: *Training Data* (sekitar 70–80%) Digunakan untuk membangun model decision tree. *Testing Data* (sekitar 20–30%) Digunakan untuk menguji tingkat akurasi model.

3.1. Pengolahan Data

Berdasarkan diagram Decision Tree pada Gambar 6 yang Anda kirimkan, model tersebut menunjukkan pola keputusan untuk memprediksi status kelulusan mahasiswa (*Yes / No*) berdasarkan beberapa atribut utama seperti Nilai Semester 1, Pekerjaan, Jurusan, Nilai IMK, dan status Perpustakaan. Faktor Utama Penentu Prediksi Atribut yang menjadi *Root Testing* (akar keputusan) adalah NILAI SEMESTER 1 (NILAI SEM.1). Hal ini menunjukkan bahwa nilai pada semester pertama merupakan faktor

paling berpengaruh dalam memprediksi kelulusan mahasiswa. Secara keseluruhan, model Decision Tree menunjukkan bahwa prediksi tingkat kelulusan mahasiswa dipengaruhi oleh kombinasi faktor akademik, program studi, serta faktor administratif. Variabel yang paling dominan adalah Nilai Semester 1, yang kemudian diikuti oleh status pekerjaan, jurusan, nilai mata kuliah tertentu, dan status perpustakaan. *Presentase Confidence* dapat dilihat pada Tabel 1.



Gambar 6 Hasil Pengolahan Data Histori

Tabel 1. *Presentase Confidence*

No	Nilai Sem.1	Pekerjaan	Perpust	Jurusan	IMK	DP
1	memuaskan					No
2	sangat memuaskan					No
3	dengan pujian	bekerja	ada			No
4	dengan pujian	Bekerja	Tidak ada			Yes
5	dengan pujian	mahasiswi		Ak		Yes
6	dengan pujian	mahasiswi	ada	mi		Yes
7	dengan pujian	mahasiswa	Tidak ada	Mi		No
8	dengan pujian	mahasiswa		Sk		No
9	dengan pujian	mahasiswa		Si	memuaskan	Yes
10	dengan pujian	mahasiswa	Ada	Ti	sangat memuaskan	No
11	dengan pujian	mahasiswa	Ada	Ti	memuaskan	No
12	dengan pujian	mahasiswa	Tidak ada	Ti	memuaskan	No
13	dengan Pujian	mahasiswa	Tidak ada	Ti	sangat memuaskan	Yes

Data uji adalah data saat ini yang akan dikenai proses prediksi berdasarkan pola yang telah diperoleh dari data histori. Untuk menguji data yang dimiliki dalam penelitian ini akan menggunakan *rule* yang telah dibentuk oleh Rapidminer. Diperoleh hasil prediksi terhadap sejumlah mahasiswa aktif Hasil *knowledge* ini selanjutnya digunakan sebagai model untuk membentuk pohon keputusan terhadap data tes, dalam hal ini berjumlah 395 sampel. Setah dilakukan analisis dan pengolahan data pada data tes terdapat perbedaan yang signifikan. Tabel 2 merupakan perbandingan antara data *history* dengan data *test*.

Tabel 2 Hasil Prediksi Data Test

Data	Jumlah	Persen
Sampling	500	100%
Dengan Pujian	15	3.00%
Bukan Dengan Pujian	485	97.00%

Tabel 2 menunjukkan bahwa jumlah total data (sampling) yang digunakan dalam penelitian adalah 500 mahasiswa (100%). Dari jumlah tersebut, terdapat dua kategori kelulusan, yaitu mahasiswa Lulus Dengan Pujian dengan Jumlah 15 mahasiswa dengan persentase 3.00%. Untuk Mahasiswa Lulus Bukan Dengan Pujian dengan Jumlah 485 mahasiswa dan persentasenya ialah 97.00%. Sebagian besar di data

tersebut mahasiswa lulus tanpa predikat pujian. Faktor-faktor yang memengaruhi kelulusan dengan pujian perlu dianalisis lebih dalam model decision tree dapat digunakan untuk mengidentifikasi pola atau aturan (*rules*) yang membedakan kedua kategori tersebut, seperti IPK, lama studi, atau nilai mata kuliah tertentu.

Tabel 3 Evaluasi Predikat Kelulusan Berdasar Prediksi

Predikat	2025	2026	Mean	(Prediksi)	Kesimpulan
Dengan Pujian	6.01%	6.67%	6.34%	3.00%	Turun 3.34%
Bukan Dengan Pujian	93,99%	93.33%	93.33%	97.00%	Naik 3.67%

Pada Tabel 3 prediksi menunjukkan penurunan mahasiswa dengan pujian sebesar 3.34% . Sebaliknya, terjadi kenaikan mahasiswa tanpa pujian sebesar 3.67%. Decision Tree berhasil mengidentifikasi pola perubahan distribusi kelulusan.

4. KESIMPULAN

Hasil analisis menunjukkan bahwa persentase mahasiswa yang memperoleh predikat “Dengan Pujian” pada tahun 2025 sebesar 6,01% dan meningkat menjadi 6,67% pada tahun 2026, dengan nilai rata-rata (*mean*) sebesar 6,34%. Namun demikian, berdasarkan hasil prediksi menggunakan metode Decision Tree, persentase mahasiswa yang diperkirakan memperoleh predikat Dengan Pujian menurun menjadi 3,00%, atau mengalami penurunan sekitar 3,34% dari rata-rata sebelumnya. Hal ini mengindikasikan bahwa dalam pola data yang dianalisis terdapat kecenderungan penurunan jumlah mahasiswa yang mampu mencapai prestasi akademik tertinggi.

REFERENSI

- [1] D. D. S. Fatimah and E. Rahmawati, “Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa,” *J. Algoritma*, vol. 18, no. 2, pp. 553–561, 2022, doi: 10.33364/algoritma/v.18-2.932.
- [2] N. Drydakis, “The formation of AI Capital in higher education: Enhancing students’ academic performance and employment rates,” *Comput. Educ. Artif. Intell.*, vol. 9, no. August, p. 100476, 2025, doi: 10.1016/j.caeai.2025.100476.
- [3] W. Rao, “Design and implementation of college students’ physical education teaching information management system by *data mining* technology,” *Heliyon*, vol. 10, no. 16, p. e36393, 2024, doi: 10.1016/j.heliyon.2024.e36393.
- [4] A. K. F. Loder, “*Machine learning* for university management: Micro Cluster Learning to predict ‘active’ students,” *Stud. Educ. Eval.*, vol. 85, no. February, p. 101463, 2025, doi: 10.1016/j.stueduc.2025.101463.
- [5] M. Vaarma and H. Li, “Predicting student dropouts with *machine learning*: An empirical study in Finnish higher education,” *Technol. Soc.*, vol. 76, no. September 2023, p. 102474, 2024, doi: 10.1016/j.techsoc.2024.102474.
- [6] Y. Chen, X. Sun, W. Wei, Y. Dong, and C. J. Liang, “A Prediction and Visual Analysis Method for Graduation Destination of Undergraduates Based on LambdaMART Model,” *Int. J. Inf. Commun. Technol. Educ.*, vol. 18, no. 2, pp. 1–19, 2022, doi: 10.4018/IJICTE.315010.
- [7] M. M. Islam, F. H. Sojib, M. F. H. Mihad, M. Hasan, and M. Rahman, “The integration of explainable AI in Educational *Data mining* for student academic performance prediction and support system,” *Telemat. Informatics Reports*, vol. 18, no. May, p. 100203, 2025, doi: 10.1016/j.teler.2025.100203.
- [8] A. M. Rabelo and L. E. Zárate, “A model for predicting dropout of higher education students,” *Data Sci. Manag.*, vol. 8, no. 1, pp. 72–85, 2025, doi: 10.1016/j.dsm.2024.07.001.
- [9] A. K. F. Loder, “Master’s programs’ dropout and graduation clusters in a university system with a multiple enrollment policy,” *Int. J. Educ. Res. Open*, vol. 8, no. December 2024, p. 100423, 2025, doi: 10.1016/j.ijedro.2024.100423.
- [10] I. Vega-Rebolledo, A. J. Sánchez-García, J. J. Muñoz León, J. O. Ocharán-Hernández, and K. Cortés-Verdín, “Applying survival analysis and Explainable Artificial Intelligence to understand academic success in software engineering students,” *Array*, vol. 28, no. July, p. 100540, 2025, doi: 10.1016/j.array.2025.100540.
- [11] E. Y. Seo, J. Yang, J. E. Lee, and G. So, “Predictive modelling of student dropout risk: Practical insights from a South Korean distance university,” *Heliyon*, vol. 10, no. 11, p. e30960, 2024, doi: 10.1016/j.heliyon.2024.e30960.
- [12] A. Muresan, M. Cardei, and I. Cardei, “Predicting Student Success with Heterogeneous Graph Deep Learning and *Machine learning* Models,” *Proc. Int. Conf. Educ. Data Min.*, pp. 265–275, 2025, doi: 10.5281/zenodo.15870191.
- [13] R. Ordoñez-Avila, N. Salgado Reyes, J. Meza, and S. Ventura, “*Data mining* techniques for predicting teacher evaluation in higher education: A systematic literature review,” *Heliyon*, vol. 9, no. 3, 2023,

- doi: 10.1016/j.heliyon.2023.e13939.
- [14] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. Abou Naaj, "Student course grade prediction using the random forest algorithm: Analysis of predictors' importance," *Trends Neurosci. Educ.*, vol. 33, p. 100214, 2023, doi: 10.1016/j.tine.2023.100214.
- [15] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using *machine learning* techniques," *Educ. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/educsci11090552.
- [16] R. D. Deleña *et al.*, "Predicting student retention: A comparative study of *machine learning* approach utilizing sociodemographic and academic factors," *Syst. Soft Comput.*, vol. 7, no. July, 2025, doi: 10.1016/j.sasc.2025.200352.
- [17] M. H. Mehta, N. C. Chauhan, and A. Gokhale, "Predicting Institute Graduation Rate using Evolutionary Computing and *Machine learning*," *Procedia Comput. Sci.*, vol. 252, pp. 758–767, 2025, doi: 10.1016/j.procs.2025.01.036.
- [18] R. D. Deleña *et al.*, "Predicting student retention: A comparative study of *machine learning* approach utilizing sociodemographic and academic factors," *Syst. Soft Comput.*, vol. 7, no. June, 2025, doi: 10.1016/j.sasc.2025.200352.
- [19] H. Kotaka and K. Misue, "Predictive automatic selection of guidance methods for reducing student dropout," *Procedia Comput. Sci.*, vol. 246, no. C, pp. 1720–1729, 2024, doi: 10.1016/j.procs.2024.09.668.
- [20] Y. E. Yuspita, R. Okra, and M. Rezeki, "Penerapan Algoritma Klasifikasi Untuk Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Rappidminer," *Djtechno J. Teknol. Inf.*, vol. 6, no. 1, pp. 376–388, 2025, doi: 10.46576/djtechno.v6i1.6169.
- [21] A. Izbassar, M. Muratbekova, D. Amangeldi, N. Oryngozha, A. Ogorodova, and P. Shamoi, "Intelligent System for Assessing University Student Personality Development and Career Readiness," *Procedia Comput. Sci.*, vol. 231, no. 2018, pp. 779–785, 2024, doi: 10.1016/j.procs.2023.12.138.
- [22] Dofiyanto and Z. Fatah, "Penerapan Algoritma Decision Tree untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Faktor Akademik dan Sosial," *JISCO J. Inf. Syst. Comput.*, vol. 3, no. 2, pp. 66–76, 2025, doi: 10.30631/jisco.v3i2.4030.
- [23] R. V. A. Suyanto, Eduard Rusdianto, and Ernawati, "Penerapan Algoritma Decision Tree C4.5 dan Metode AdaBoost Untuk Prediksi Kelulusan Mahasiswa," *J. Inform. Atma Jogja*, vol. 5, no. 1, pp. 75–86, 2024, doi: 10.24002/jiaj.v5i1.8646.
- [24] M. Yin, W. Xu, and Y. Wang, "Controlling or directing? Text mining to decode supervisor-graduate student relationship," *Acta Psychol. (Amst.)*, vol. 262, no. November 2025, p. 106030, 2026, doi: 10.1016/j.actpsy.2025.106030.
- [25] K. Okoye, J. T. Nganji, J. Escamilla, and S. Hosseini, "*Machine learning* model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education," *Comput. Educ. Artif. Intell.*, vol. 6, no. January, p. 100205, 2024, doi: 10.1016/j.caeai.2024.100205.
- [26] R. Setiawan, E. Noersasongko, A. Syukur, F. Budiman, and D. Kurniadi, "Imbalanced Multi-Class Prediction Of Student Drop-Out And Graduation :," vol. 19, no. 1, pp. 72–90, 2025.
- [27] S. Boujmiraz, H. Darhmaoui, and A. Drissi el maliani, "Predicting student performance: A comprehensive review of *machine learning*, deep learning, and explainable AI approaches," *Comput. Educ. Artif. Intell.*, vol. 10, no. January, p. 100548, 2026, doi: 10.1016/j.caeai.2026.100548.
- [28] S. Boujmiraz, H. Darhmaoui, and A. Drissi el maliani, "Predicting student performance: A comprehensive review of *machine learning*, deep learning, and explainable AI approaches," *Comput. Educ. Artif. Intell.*, vol. 10, no. July 2025, p. 100548, 2026, doi: 10.1016/j.caeai.2026.100548.
- [29] S. Sarker, M. K. Paul, S. T. H. Thasin, and M. A. M. Hasan, "Analyzing students' academic performance using educational *data mining*," *Comput. Educ. Artif. Intell.*, vol. 7, no. December 2023, p. 100263, 2024, doi: 10.1016/j.caeai.2024.100263.
- [30] D. Coelho, E. Papenhausen, and K. Mueller, "Evolutionary design of a visual analytics interface to study predictive patterns in high dimensional data," *Vis. Informatics*, p. 100303, 2025, doi: 10.1016/j.visinf.2025.100303.