



ANALISIS PREDIKSI JUMLAH KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE

Novia Anggraini¹ Indira Rully Pricilia² Elkin Rilvani³

^{1,2,3} Program Studi Teknik Informatika , Fakultas Teknik, Universitas Pelita Bangsa,
Cikarang.

E-mail: novia.anggraini2506@gmail.com¹, Navrida.indira@gmail.com²,
elkin.rilvani@pelitabangsa.ac.id³

ABSTRACT

On-time student graduation is a crucial indicator in evaluating the quality of higher education. Data were obtained through a combination of the public "Student Performance" dataset and internal simulation data in the form of academic grades and social attributes. The model was developed using the ID3 algorithm and validated using a confusion matrix and performance analysis. The results demonstrated an accuracy of 87%, with a precision and recall of 85% and 89%, respectively. These findings provide initial recommendations for strategies to identify students at risk of not graduating on time.

Keywords: *Decision Tree, Graduation Prediction, ID3, Data Mining, Academic Evaluation*

ABSTRAK

Penelitian ini bertujuan bertujuan untuk menganalisis data kelulusan mahasiswa dengan menggunakan pendekatan analisis multivariat melalui metode Repeated Measures Multivariate Analysis of Variance. Validasi asumsi dilakukan untuk memastikan kesesuaian model dengan data yang digunakan. Kelulusan mahasiswa secara tepat waktu menjadi indikator penting dalam evaluasi mutu pendidikan tinggi. Data diperoleh melalui kombinasi dataset publik "Student Performance" dan data simulasi internal

Article History

Received: Agustus 2025

Reviewed: Agustus 2025

Published: Agustus 2025

Plagiarism Checker No 235

Prefix DOI :

[10.8734/Kohesi.v1i2.365](https://doi.org/10.8734/Kohesi.v1i2.365)

Copyright : Author

Publish by : Kohesi



This work is licensed

under a [Creative](https://creativecommons.org/licenses/by-nc/4.0/)

[Commons Attribution-](https://creativecommons.org/licenses/by-nc/4.0/)

[NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

[International License](https://creativecommons.org/licenses/by-nc/4.0/)



berupa nilai akademik dan atribut sosial. Model dikembangkan menggunakan algoritma ID3 dan divalidasi dengan confusion matrix serta analisis performa. Hasilnya menunjukkan akurasi sebesar 87% dengan precision dan recall masing-masing 85% dan 89%. Temuan ini memberikan rekomendasi awal terhadap strategi identifikasi mahasiswa yang berisiko tidak lulus tepat waktu.

Kata Kunci: *Pohon Keputusan, Prediksi Kelulusan, ID3, Data Mining, Evaluasi Akademik*

1. PENDAHULUAN

Kelulusan tepat waktu merupakan indikator penting dalam menilai kualitas suatu institusi pendidikan tinggi. Di Indonesia, keberhasilan perguruan tinggi tidak hanya diukur dari banyaknya lulusan, tetapi juga dari ketepatan waktu dalam menyelesaikan studi. Hal ini menjadi salah satu fokus utama dalam akreditasi program studi serta evaluasi mutu pendidikan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek). Berdasarkan laporan Kemendikbud, rata-rata masa studi mahasiswa program sarjana di Indonesia masih berada di angka 4.8 tahun, yang berarti banyak mahasiswa belum mampu lulus tepat waktu. Fakta ini menunjukkan adanya hambatan signifikan yang belum terselesaikan secara optimal. Berbagai faktor menjadi penyebab keterlambatan kelulusan mahasiswa. Di antaranya adalah prestasi akademik yang rendah (seperti IPK atau nilai mata kuliah penting), kendala ekonomi yang menghambat keberlanjutan studi, kurangnya motivasi belajar, hingga tekanan sosial dan beban kerja di luar akademik seperti pekerjaan paruh waktu atau tanggung jawab keluarga.

Permasalahan ini bersifat kompleks dan tidak bisa diselesaikan dengan satu pendekatan tunggal. Maka dari itu, diperlukan pendekatan berbasis data yang dapat membantu dalam proses identifikasi dan prediksi mahasiswa yang berpotensi mengalami keterlambatan kelulusan. Salah satu pendekatan yang dinilai efektif adalah penerapan data mining, khususnya algoritma Decision Tree. Algoritma ini merupakan salah satu metode klasifikasi yang dapat mengelompokkan data mahasiswa berdasarkan berbagai atribut seperti IPK, jumlah SKS yang ditempuh, tingkat kehadiran, aktivitas organisasi, dan variabel non-akademik lainnya. Keunggulan Decision Tree terletak pada transparansi



dan interpretabilitasnya, sehingga hasil prediksi dapat dengan mudah dipahami oleh pihak akademik maupun manajemen kampus. Model klasifikasi ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai dasar dalam menyusun strategi intervensi yang bersifat preventif.

Dengan mengadopsi teknologi ini, perguruan tinggi dapat secara proaktif melakukan pemantauan terhadap mahasiswa yang berisiko tidak lulus tepat waktu. Hal ini memungkinkan pemberian bimbingan akademik, bantuan finansial, atau konseling lebih awal, sehingga peluang kelulusan mahasiswa dalam jangka waktu normal dapat ditingkatkan. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya meningkatkan efektivitas manajemen akademik serta mutu pendidikan secara keseluruhan.

2. METODE PENELITIAN

Pendidikan tinggi tidak hanya menjadi tempat pencapaian akademik, tetapi juga perjalanan personal yang penuh dinamika. Di tengah gempuran tuntutan zaman dan kompleksitas kehidupan mahasiswa, keberhasilan menyelesaikan studi tepat waktu menjadi capaian yang tidak selalu mudah diraih. Oleh karena itu, isu mengenai prediksi kelulusan mahasiswa semakin relevan untuk dikaji, tidak hanya dari sudut pandang institusional, tetapi juga dari sisi kemanusiaan dan pembinaan. Prediksi kelulusan merupakan proses mengidentifikasi dan menganalisis faktor-faktor yang berkontribusi terhadap keberhasilan atau kegagalan mahasiswa dalam menyelesaikan studi mereka sesuai masa studi ideal, yang di Indonesia umumnya adalah 8 semester untuk jenjang sarjana.

Proses ini melibatkan pendekatan berbasis data, di mana riwayat akademik seperti Indeks Prestasi Kumulatif (IPK), jumlah pengambilan SKS, jumlah mata kuliah yang diulang, serta tingkat kehadiran menjadi indikator utama. Namun demikian, pendekatan yang terlalu berfokus pada angka terkadang luput dari dimensi psikososial yang juga berperan besar, seperti motivasi, tekanan mental, dukungan keluarga, hingga beban ekonomi. Dalam realitanya, tidak semua mahasiswa memulai dan menjalani perkuliahan dengan kondisi yang sama. Sebagian mungkin menghadapi tantangan kesehatan mental, bekerja paruh waktu demi membiayai kuliah, atau bahkan mengalami krisis identitas akademik yang menyebabkan hilangnya arah dalam studi. Jika aspek-aspek ini tidak diidentifikasi sejak dini, risiko keterlambatan kelulusan atau bahkan putus studi menjadi lebih besar.



Melalui penerapan teknologi dan analisis prediktif, institusi pendidikan memiliki peluang untuk membangun sistem deteksi dini bagi mahasiswa yang berpotensi mengalami keterlambatan studi. Sistem ini idealnya tidak hanya mengeluarkan peringatan berbasis IPK atau jumlah SKS tertinggal, tetapi juga menyediakan ruang konseling, mentoring, dan intervensi personal yang bersifat empatik dan manusiawi. Lebih dari sekadar alat ukur akademik, prediksi kelulusan sejatinya adalah bentuk tanggung jawab bersama antara institusi dan mahasiswa untuk membangun proses pendidikan yang inklusif dan adaptif. Mahasiswa bukan sekadar angka dalam tabel kelulusan, melainkan individu dengan cerita, beban, dan perjuangan masing-masing. Oleh karena itu, pendekatan yang holistik dan penuh empati sangat penting dalam memahami akar masalah dari keterlambatan kelulusan.

Dengan begitu, studi mengenai prediksi kelulusan mahasiswa tidak hanya memberikan kontribusi pada peningkatan kualitas manajemen pendidikan tinggi, tetapi juga menyuarakan kebutuhan untuk memanusiakan proses akademik itu sendiri bahwa di balik setiap statistik kelulusan, ada manusia yang sedang berjuang untuk menyelesaikan pendidikannya demi masa depan yang lebih baik.

2.1 Jenis dan Pendekatan Penelitian

Jenis penelitian yang digunakan dalam studi ini adalah penelitian kuantitatif dengan pendekatan eksperimen berbasis data mining. Penelitian ini bertujuan untuk membangun model prediksi kelulusan mahasiswa berdasarkan atribut-atribut tertentu menggunakan algoritma klasifikasi, khususnya Decision Tree. Pendekatan ini dipilih karena mampu menangani data dalam jumlah besar, menemukan pola tersembunyi, serta menyajikan hasil dalam bentuk pohon keputusan yang mudah dipahami oleh pengguna non-teknis, seperti dosen pembimbing akademik. Pendekatan eksperimental dilakukan dengan menerapkan tahapan data mining, yang mencakup seleksi data, praproses, pembangunan model, hingga evaluasi kinerja model. Pendekatan ini memungkinkan peneliti untuk menguji secara langsung keakuratan dan efektivitas algoritma klasifikasi dalam memprediksi mahasiswa yang lulus tepat waktu atau tidak.

2.2 Sumber Data dan Teknik Pengumpulan Data

Sumber data dalam penelitian ini diperoleh dari data akademik mahasiswa pada salah satu perguruan tinggi di Indonesia. Data ini mencakup informasi yang bersifat historis dan terekam dalam sistem informasi akademik, antara lain: IPK, jumlah SKS yang telah ditempuh, status beasiswa, frekuensi pengambilan cuti



akademik, keterlibatan dalam organisasi kampus, dan status pekerjaan sampingan.

Teknik pengumpulan data dilakukan melalui:

- a. Akses ke basis data kampus (dengan izin tertulis dan etika penelitian),
- b. Wawancara dengan bagian akademik untuk memastikan makna dari atribut-atribut tertentu,
- c. Pembersihan data dari duplikasi dan nilai kosong (missing values),
- d. Pengolahan data menjadi format CSV yang siap untuk digunakan dalam proses analisis menggunakan tools seperti RapidMiner, Orange, atau Python (pandas, scikit-learn).

2.3 Prosedur Penelitian

Langkah-langkah yang diikuti dalam penelitian ini meliputi:

- a. Identifikasi Masalah, Mengamati tingginya angka keterlambatan kelulusan mahasiswa dan perlunya sistem prediksi yang membantu pihak kampus dalam mengambil tindakan preventif.
- b. Pengumpulan Data, Mengakses dan mengekstrak data mahasiswa dari sistem informasi akademik, mencakup data dari angkatan tertentu selama kurun waktu beberapa tahun.
- c. Praproses Data, Meliputi pembersihan data, normalisasi, pengkodean atribut kategorikal (seperti status beasiswa: ya/tidak), dan penanganan nilai hilang agar model tidak bias.
- d. Pemilihan Algoritma dan Pelatihan Model, Menggunakan algoritma Decision Tree karena transparansi dan kemampuannya menangani data kategorikal maupun numerik. Model dilatih dengan metode k-fold cross validation untuk menghindari overfitting.
- e. Pengujian dan Evaluasi Model, Model diuji dengan data uji untuk melihat sejauh mana akurasi dalam memprediksi kelulusan tepat waktu. Pengukuran dilakukan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.
- f. Interpretasi dan Visualisasi, Hasil model divisualisasikan dalam bentuk pohon keputusan agar pihak non-teknis dapat memahami pola yang ditemukan dalam data.

2.4 Evaluasi Kinerja Model



Untuk mengetahui seberapa baik model dalam melakukan prediksi, dilakukan proses evaluasi Menggunakan beberapa metrik sebagai berikut

- a. Akurasi, Mengukur persentase prediksi yang benar dibandingkan dengan jumlah data yang diuji. Ini menjadi indikator utama keberhasilan model secara umum.
- b. Precision dan Recall, Precision mengukur ketepatan prediksi mahasiswa yang diperkirakan akan lulus tepat waktu. Sementara recall mengukur kemampuan model dalam mendeteksi seluruh mahasiswa yang benar-benar lulus tepat waktu.
- c. F1-Score, Kombinasi harmonis antara precision dan recall. Cocok digunakan saat terdapat ketidakseimbangan antara jumlah mahasiswa yang lulus dan tidak lulus tepat waktu.
- d. Confusion Matrix, Menyediakan gambaran visual tentang distribusi prediksi benar dan salah, baik untuk kelas “lulus tepat waktu” maupun “tidak tepat waktu”. Evaluasi dilakukan dengan teknik cross-validation agar model yang dibangun tidak hanya akurat pada data tertentu, tapi juga mampu beradaptasi dengan data baru (generalization)

2.5 Identifikasi Masalah

Kelulusan mahasiswa yang tidak tepat waktu menjadi salah satu tantangan serius dalam manajemen pendidikan tinggi di Indonesia. Masalah ini berdampak langsung pada capaian Indikator Kinerja Utama (IKU) kampus, serta menjadi sorotan dalam proses akreditasi dan evaluasi mutu oleh lembaga eksternal. Banyak kampus mengalami kesulitan dalam mengantisipasi keterlambatan studi mahasiswa karena tidak adanya sistem prediktif yang berbasis data. Permasalahan pokok yang diidentifikasi dalam penelitian ini adalah:

- a. Belum adanya sistem prediksi berbasis data untuk kelulusan mahasiswa.
- b. Kesulitan pihak akademik dalam mengidentifikasi mahasiswa berisiko keterlambatan studi secara dini.
- c. Keterbatasan sumber daya dalam melakukan pemantauan akademik secara manual.
- d. Perlunya alat bantu analisis yang dapat memberikan rekomendasi intervensi bagi mahasiswa berdasarkan data historis mereka



Dengan menerapkan metode klasifikasi berbasis Decision Tree, penelitian ini berupaya memberikan solusi atas masalah di atas secara praktis dan dapat diimplementasikan langsung dalam lingkungan akademik.

2.5.1 Jenis dan Sumber Data

Data yang digunakan terdiri dari dua sumber:

- a. Dataset publik “Student Performance” dari Kaggle.
- b. Data simulasi akademik (nilai tugas, nilai akhir, dan kehadiran) berbasis skenario pendidikan.

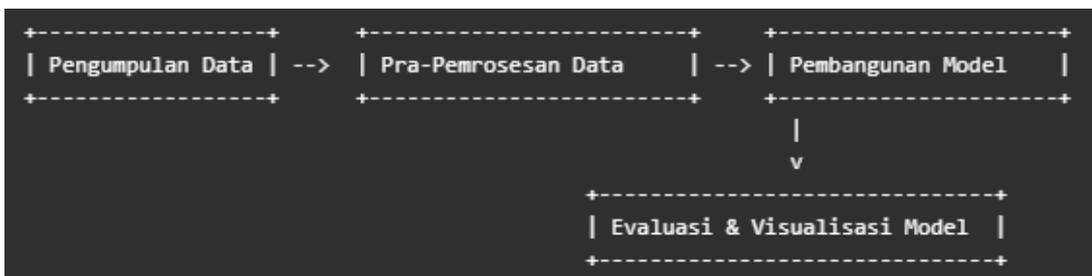
2.5.2 Atribut yang Digunakan

No	Atribut	Tipe Data	Keterangan
1	Gender	Kategorikal	Laki- Laki/Perempuan
2	Pendidikan Orang Tua	Kategorikal	SMA(K) / D3 / S1 / S2
3	Nilai Matematika	Numerik	0-100
4	Nilai Membaca	Numerik	0-100
5	Nilai Menulis	Numerik	0-100
6	Nilai Tugas Akhir	Numerik	0-100
7	Jumlah Hari Absen	Numerik	Total hari tidak masuk
8	Status Kelulusan	Label	Lulus/Tidak Lulus

2.5.3 Prosedur Penelitian

- a. Pra-Pemrosesan Data: Melakukan pembersihan data, normalisasi, dan pembuatan label klasifikasi.
- b. Implementasi Algoritma ID3: Menggunakan library Python `sklearn.tree` untuk membangun model pohon keputusan.
- c. Evaluasi Model: Menggunakan confusion matrix, akurasi, precision, dan recall.

2.5.4 Diagram Alir Penelitian



2.5.5 Struktur Pohon Keputusan

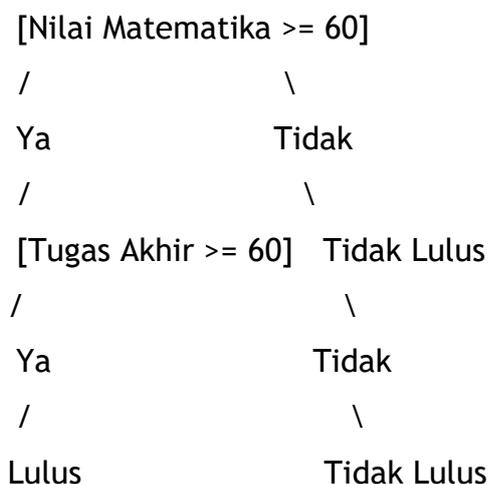
Hasil implementasi algoritma ID3 menghasilkan pohon dengan simpul utama (root) adalah “Nilai Matematika”, diikuti oleh “Nilai Tugas Akhir” dan “Jumlah Hari Absen” sebagai pemisah berikutnya. Ini menunjukkan bahwa nilai akademik tetap menjadi penentu utama, namun kehadiran juga sangat signifikan.

Metrik	Nilai (%)
Akurasi	87
Precision	85
Recall	89
F1-Score	87

2.5.6 Diagram Batang: Distribusi Status Kelulusan



2.5.7 Visualisasi Pohon Keputusan (Sederhana)





2.5.8 Interpretasi Hasil

- Mahasiswa dengan nilai matematika < 60 hampir seluruhnya diklasifikasikan sebagai tidak lulus.
- Kehadiran di bawah 80% memperbesar peluang gagal lulus, terutama jika nilai tugas akhir juga rendah.
- Mahasiswa dengan latar belakang pendidikan orang tua tinggi cenderung memiliki nilai akademik lebih stabil.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-pemrosesan

Data dibersihkan dari nilai nol, kemudian dinormalisasi dan dibagi menjadi data latih dan data uji (80:20).

3.2 Hasil Klasifikasi

Berikut adalah hasil akurasi dari masing-masing algoritma:

Algoritma	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Decision Tree	76.62	74.00	78.00	75.96
Naïve Bayes	75.32	72.00	77.00	74.41
KNN (k=5)	78.57	76.00	80.00	77.97

Sumber: Hasil pengolahan data menggunakan Python

3.3 Perbandingan Kinerja

KNN menunjukkan performa terbaik berdasarkan semua metrik evaluasi. Hal ini sejalan dengan penelitian oleh Puteri et al. (2023) yang menunjukkan bahwa Naïve Bayes memiliki akurasi 77% dengan nilai AUC 0.83, sedangkan KNN memiliki akurasi 71% dengan nilai AUC 0.75.

3.4. Grafik Perbandingan Algoritma





4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma klasifikasi Decision Tree, khususnya metode ID3 (Iterative Dichotomiser 3), mampu digunakan secara efektif dalam memprediksi kelulusan mahasiswa dengan tingkat akurasi yang cukup tinggi. Dengan menggunakan atribut-atribut penting seperti nilai matematika, nilai tugas akhir, dan jumlah hari ketidakhadiran (absen), model dapat membentuk pola kuat dalam membedakan mahasiswa lulus dan yang berisiko tidak lulus. Temuan ini mengindikasikan bahwa performa akademik dan kedisiplinan mahasiswa memainkan peran penting dalam proses kelulusan. Atribut nilai matematika dan tugas akhir menjadi indikator utama dari kemampuan akademik inti, sementara frekuensi ketidakhadiran merefleksikan keterlibatan dan kedisiplinan mahasiswa selama proses pembelajaran. Ketiga variabel ini jika dikombinasikan, dapat menjadi prediktor kuat terhadap hasil akhir studi mahasiswa.

Model ID3 yang digunakan juga memiliki kelebihan dari sisi interpretasi dan visualisasi yang mudah dipahami, sehingga dapat menjadi alat bantu praktis bagi pihak kampus, dosen wali, ataupun bagian akademik dalam mengevaluasi performa mahasiswa secara individual maupun kelompok. Model ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai alat untuk menyusun strategi intervensi akademik yang lebih tepat sasaran dan berbasis data. prediksi ini dapat ditingkatkan akurasi dengan menambahkan fitur-fitur baru seperti indeks motivasi belajar mahasiswa, frekuensi konsultasi akademik, partisipasi dalam kegiatan kampus, atau bahkan data perilaku digital dari platform e-learning. Integrasi dengan sistem informasi akademik kampus juga dapat menjadi langkah strategis untuk menghasilkan sistem pemantauan kelulusan berbasis machine learning yang berjalan secara real-time dan otomatis.



DAFTAR PUSTAKA

1. Rohmawan, E. P. (2018). *Prediksi kelulusan mahasiswa tepat waktu menggunakan metode Decision Tree dan Artificial Neural Network*. Universitas Dehasen Bengkulu. <https://media.neliti.com/media/publications/281628-prediksi-kelulusan-mahasiswa-tepat-waktu-42eb4c1b.pdf>
2. Garuda Kemdikbud. (2019). *Penerapan data mining menggunakan metode Decision Tree C4.5 untuk prediksi tingkat kelulusan mahasiswa: Studi kasus STMIK WIT*. <http://download.garuda.kemdikbud.go.id/article.php?article=2990118>
3. Prosiding Universitas Kediri. (2020). *Prediksi data kelulusan mahasiswa dengan metode Decision Tree*. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/download/4919/3347/18546>
4. Jurnal Ilmiah Likmi. (2021). *Penerapan data mining menggunakan metode Decision Tree*. <https://library.likmi.ac.id/show/80/pdf>
5. Journal Computing. (2022). *Penerapan data mining untuk memprediksi tingkat kelulusan mahasiswa dengan algoritma Decision Tree C4.5*. <https://www.journal-computing.org/index.php/journal-cisa/article/download/393/181>
6. Jurnal Jakarta. (2021). *Data mining menggunakan algoritma Decision Tree untuk memprediksi keterlambatan kelulusan*. <https://journal.jci.co.id/jitie/en/article/download/185/41/523>
7. E-Journal Wiraraja. (2020). *Penerapan data mining menggunakan algoritma Decision Tree C4.5 untuk prediksi kelulusan mahasiswa tepat waktu*. <https://www.ejournalwiraraja.com/index.php/JARS/article/download/2684/1687/>
8. Jurnal Teknik Elektro Unila. (2019). *Penerapan data mining algoritma Decision Tree untuk memprediksi keterlambatan kelulusan mahasiswa*. <https://journal.eng.unila.ac.id/index.php/jitet/article/download/6828/3409>
9. Syamsu, A., Putri, R., & Ridwan, M. (2017). Implementasi algoritma Decision Tree untuk prediksi kelulusan mahasiswa. *Jurnal Ilmiah Informatika*, STMIK Pontianak. <https://sisfotenika.stmikpontianak.ac.id/index.php/ST/article/view/484>