

ANALISIS KLASIFIKASI DATA MAHASISWA DALAM MENENTUKAN KELULUSAN TEPAT WAKTU MENGGUNAKAN METODE DECISION TREE

I Putu Yoga Eka Saputra¹, I Wayan Sudiarsa², Ni Made Sulistyani³, I Kadek Pica Widi Ada⁴, I Komang Nova Adiyasa⁵

yoga98099@gmail.com¹, sudiarsa@instiki.ac.id², mdsulisty05@gmail.com³,
kadekpicawidiada@instiki.ac.id⁴, komangnovaadiyasa@gmail.com⁵

Institut Bisnis Dan Teknologi Indonesia (INSTIKI)

ABSTRAK

Ketepatan waktu kelulusan mahasiswa merupakan indikator penting dalam menilai efektivitas proses akademik di perguruan tinggi. Penelitian ini bertujuan untuk menganalisis dan memprediksi kelulusan mahasiswa tepat waktu menggunakan metode Decision Tree berbasis data mining. Dataset yang digunakan berjumlah 100.000 data mahasiswa dalam format CSV dengan 10 atribut, terdiri dari 4 atribut numerik utama akademik (IPK, IPS rata-rata, IPS semester akhir, IPS tren), 4 atribut numerik pendukung (jumlah mata kuliah tidak lulus, jumlah cuti akademik, jumlah semester, status kelulusan), serta 2 atribut kategorikal (pekerjaan sambil kuliah dan kategori kehadiran). Hasil analisis statistik deskriptif menunjukkan bahwa nilai IPK rata-rata mahasiswa sebesar 2,99 dengan rentang 2,00–4,00, sementara rata-rata jumlah semester sebesar 10,5 semester. Proporsi mahasiswa yang lulus tepat waktu dan tidak tepat waktu relatif seimbang, masing-masing sekitar 49,96% dan 50,04%. Model Decision Tree yang dibangun menghasilkan akurasi sebesar 75,32%, dengan nilai recall kelas lulus tepat waktu mencapai 77%. Atribut IPK, jumlah semester, dan jumlah mata kuliah tidak lulus menjadi faktor paling dominan dalam proses klasifikasi. Hasil penelitian ini menunjukkan bahwa Decision Tree mampu memberikan interpretasi yang jelas terhadap pola kelulusan mahasiswa dan dapat digunakan sebagai alat bantu evaluasi akademik.

Kata Kunci: Data Mining, Decision Tree, ETL, Kelulusan Mahasiswa, Klasifikasi.

ABSTRACT

The punctuality of student graduation is a crucial indicator in assessing the effectiveness of the academic process in higher education. This study aims to analyze and predict on-time student graduation using a data mining-based Decision Tree method. The dataset used comprises 100,000 student data in CSV format with 10 attributes: four primary academic numeric attributes (GPA, average GPA, final semester GPA, GPA trend), four supporting numeric attributes (number of failed courses, number of academic leaves, number of semesters, graduation status), and two categorical attributes (employment while studying and attendance category). The descriptive statistical analysis results show that the average student GPA is 2.99 with a range of 2.00–4.00, while the average number of semesters is 10.5 semesters. The proportions of students graduating on time and not on time are relatively balanced, approximately 49.96% and 50.04%, respectively. The Decision Tree model produced an accuracy of 75.32%, with a recall rate of 77% for on-time graduates. The attributes of GPA, number of semesters, and number of failed courses are the most dominant factors in the classification process. The results of this study indicate that Decision Trees are capable of providing a clear interpretation of student graduation patterns and can be used as an academic evaluation tool.

Keywords: Data Mining, Decision Trees, ETL, Student Graduation, Classification.

PENDAHULUAN

Perguruan tinggi memiliki peran strategis dalam mencetak sumber daya manusia yang berkualitas dan kompetitif. Salah satu indikator keberhasilan penyelenggaraan pendidikan tinggi adalah tingkat kelulusan mahasiswa tepat waktu. Mahasiswa dikatakan lulus tepat waktu apabila mampu menyelesaikan masa studi sesuai dengan ketentuan

akademik yang berlaku. Tingginya angka keterlambatan kelulusan dapat berdampak pada akreditasi institusi, efisiensi pengelolaan akademik, serta beban biaya mahasiswa dan perguruan tinggi.

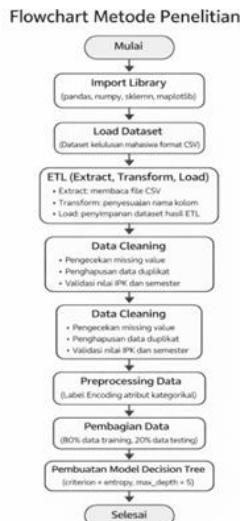
Permasalahan kelulusan mahasiswa tidak hanya dipengaruhi oleh faktor akademik seperti Indeks Prestasi Kumulatif (IPK) dan jumlah mata kuliah tidak lulus, tetapi juga oleh faktor non-akademik, seperti pekerjaan sambil kuliah, kehadiran, serta cuti akademik. Oleh karena itu, diperlukan suatu pendekatan yang mampu menganalisis data mahasiswa dalam jumlah besar untuk menemukan pola-pola yang berpengaruh terhadap kelulusan.

Data mining merupakan salah satu pendekatan yang dapat digunakan untuk menggali informasi tersembunyi dari dataset berukuran besar. Teknik klasifikasi dalam data mining bertujuan untuk memetakan data ke dalam kelas tertentu berdasarkan atribut yang dimiliki. Salah satu algoritma klasifikasi yang populer dan mudah diinterpretasikan adalah Decision Tree, karena mampu menangani data numerik dan kategorikal serta menghasilkan model berbentuk pohon keputusan yang mudah dipahami. Penelitian sebelumnya menunjukkan bahwa Decision Tree efektif digunakan dalam memprediksi kelulusan mahasiswa serta mengidentifikasi faktor-faktor dominan yang memengaruhi keberhasilan studi. Selain itu, penggunaan perangkat berbasis cloud seperti Google Colab mempermudah proses pengolahan data, pemodelan, dan visualisasi tanpa memerlukan spesifikasi perangkat keras yang tinggi. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis klasifikasi data mahasiswa dalam menentukan kelulusan tepat waktu menggunakan metode Decision Tree dengan tahapan ETL dan implementasi Google Colab sebagai media pengolahan data.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan algoritma Decision Tree yang diimplementasikan menggunakan Google Colab. Penelitian ini menggunakan pendekatan data mining dengan algoritma Decision Tree yang diimplementasikan menggunakan Google Colab. Tahapan penelitian meliputi proses Extract, Transform, dan Load (ETL) untuk menyiapkan data, dilanjutkan dengan data cleaning untuk menghilangkan data tidak valid dan duplikasi. Selanjutnya dilakukan preprocessing data dengan mengubah atribut kategorikal menjadi numerik, pembagian data menjadi data latih dan data uji, pembuatan model Decision Tree, evaluasi performa model, serta visualisasi hasil berupa confusion matrix dan pohon keputusan.

Flowchart Metode Penelitian



Gambar diatas Adalah flowchart dari tahapan Metode Penelitian yang kami lakukan.

Data Penelitian

Dataset penelitian terdiri dari 100.000 baris data mahasiswa tanpa missing value setelah proses data cleaning. Statistik deskriptif utama dari dataset ditunjukkan sebagai berikut:

- IPK: rata-rata 2,99 dengan standar deviasi 0,57
- Jumlah mata kuliah tidak lulus: rata-rata 2,49 mata kuliah
- Jumlah cuti akademik: rata-rata 0,99 kali
- Jumlah semester: minimum 7 semester dan maksimum 14 semester
- IPS rata-rata: rata-rata 3,00
- IPS semester akhir: rata-rata 2,99

Distribusi status kelulusan menunjukkan bahwa dari 100.000 mahasiswa, sekitar 49.962 mahasiswa lulus tepat waktu dan 50.038 mahasiswa tidak lulus tepat waktu, sehingga dataset berada pada kondisi relatif seimbang. Atribut yang digunakan dalam penelitian ini adalah:

Atribut	Keterangan
IPK	Indeks Prestasi Kumulatif
Mata Kuliah Tidak Lulus	Jumlah mata kuliah yang tidak lulus
Jumlah Cuti Akademik	Total cuti akademik
Pekerjaan Sambil Kuliah	Status bekerja saat kuliah
Jumlah Semester	Lama masa studi
IPS Rata-rata	Rata-rata Indeks Prestasi Semester
IPS Semester Akhir	IPS pada semester terakhir
IPS Tren	Perubahan IPS
Kategori Kehadiran	Tingkat kehadiran mahasiswa
Status Kelulusan	Label kelulusan (tepat/tidak tepat)

Pengumpulan Data

Pengumpulan data pada penelitian dilakukan dengan memanfaatkan dataset mahasiswa yang tersimpan dalam bentuk file Comma Separated Values (CSV). Dataset tersebut berisi data akademik dan non-akademik mahasiswa yang digunakan sebagai bahan analisis dalam proses klasifikasi kelulusan tepat waktu. Data yang digunakan merupakan data sekunder yang diperoleh dari sistem pengelolaan akademik dan telah melalui proses anonimisasi untuk menjaga kerahasiaan identitas mahasiswa. Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, sebuah repositori dataset dan komunitas data science yang menyediakan berbagai dataset untuk keperluan riset dan pembelajaran. Dataset yang digunakan berjudul “Dataset Kelulusan Mahasiswa”, dipublikasikan oleh pengguna Kaggle dengan nama akun afitoindrapermana dan dapat diakses melalui <https://www.kaggle.com/datasets/afitoindrapermana/dataset-kelulusan-mahasiswa>. Dataset ini memuat data mahasiswa yang mencakup atribut-atribut akademik dan non-akademik yang relevan untuk analisis klasifikasi kelulusan tepat waktu.

Tahapan pengumpulan data dalam penelitian ini meliputi beberapa langkah sebagai berikut:

1. Identifikasi Kebutuhan Data Peneliti menentukan atribut-atribut yang relevan dengan tujuan penelitian, yaitu untuk memprediksi kelulusan mahasiswa tepat waktu. Atribut yang dipilih mencakup Indeks Prestasi Kumulatif (IPK), jumlah mata kuliah tidak lulus, jumlah cuti akademik, status pekerjaan sambil kuliah, jumlah semester, IPS rata-rata, IPS semester akhir, tren IPS, serta kategori kehadiran mahasiswa.

2. Pengambilan Dataset Dataset diambil dalam format CSV yang berisi 100.000 data mahasiswa. Format CSV dipilih karena mudah diolah dan kompatibel dengan berbagai perangkat lunak analisis data, termasuk perangkat lunak data mining.
3. Seleksi Data Pada tahap ini dilakukan pemilihan data yang sesuai dengan kebutuhan penelitian. Data yang tidak memiliki keterkaitan dengan proses penentuan kelulusan mahasiswa, seperti informasi identitas pribadi, dihilangkan dari dataset.
4. Pengumpulan data dilakukan dengan mengakses data, yang kemudian dijadikan sebagai dasar acuan dalam menganalisis data mahasiswa aktif, dengan mempertimbangkan kesesuaian terhadap atribut-atribut yang digunakan didalam penelitian(Iqbal et al., 2024)

HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian menggunakan data uji sebanyak 10.000 data (20%), model Decision Tree menghasilkan nilai akurasi sebesar 75,32%. Confusion matrix menunjukkan bahwa model berhasil mengklasifikasikan 4.291 mahasiswa lulus tepat waktu dengan benar, namun masih terdapat 4.568 mahasiswa tidak tepat waktu yang diklasifikasikan sebagai tidak lulus tepat waktu.



Gambar2. Confusion Matrix Model Decision Tree

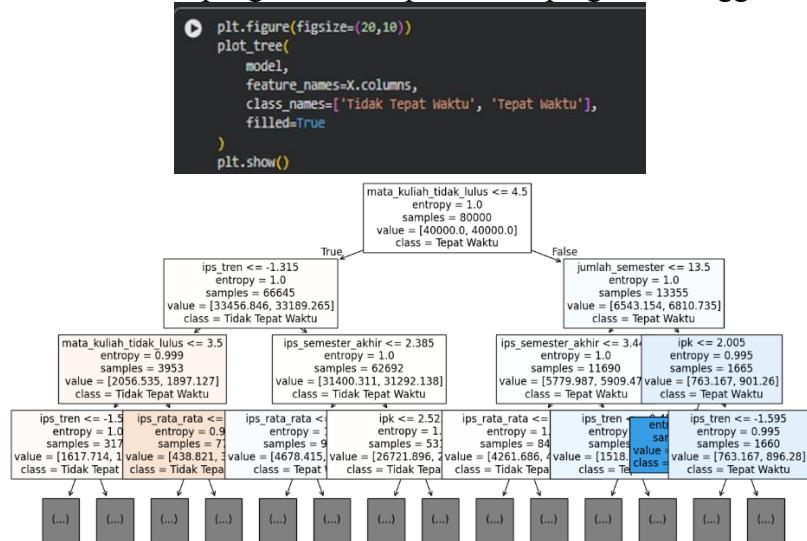
Evaluasi model menggunakan confusion matrix menunjukkan bahwa model mampu mengklasifikasikan sejumlah besar mahasiswa dengan benar baik pada kelas lulus tepat waktu maupun tidak lulus tepat waktu. Confusion matrix memberikan gambaran distribusi hasil klasifikasi prediksi terhadap data aktual, dan menegaskan bahwa tingkat kesalahan klasifikasi menurun dibandingkan dengan model awal.

Visualisasi pohon keputusan memperlihatkan bahwa IPK dengan ambang sekitar 3,0, jumlah semester lebih dari 10, serta jumlah mata kuliah tidak lulus lebih dari 2 menjadi aturan keputusan utama. Mahasiswa dengan $IPK \geq 3,0$ dan jumlah semester ≤ 10 memiliki probabilitas lebih tinggi untuk lulus tepat waktu dibandingkan mahasiswa dengan IPK rendah dan masa studi yang lebih panjang.

Peningkatan akurasi secara nyata ini menunjukkan bahwa optimasi parameter model Decision Tree dapat memperbaiki performa klasifikasi secara substansial. Meskipun model memiliki keterbatasan inheren, hasil eksperimen ini menunjukkan bahwa Decision Tree dengan pengaturan yang tepat memiliki potensi untuk menjadi alat evaluasi yang efektif dalam konteks akademik.

Hasil visualisasi pohon keputusan menunjukkan bahwa atribut IPK, jumlah semester, dan jumlah mata kuliah tidak lulus merupakan faktor dominan dalam menentukan kelulusan mahasiswa. Mahasiswa dengan IPK tinggi dan jumlah semester yang sesuai cenderung diklasifikasikan sebagai lulus tepat waktu. Sebaliknya, mahasiswa

dengan IPK rendah dan banyak mata kuliah tidak lulus memiliki kecenderungan tidak lulus tepat waktu. Meskipun akurasi model tergolong sedang, hasil ini menunjukkan bahwa Decision Tree mampu memberikan interpretasi yang jelas mengenai faktor-faktor akademik yang memengaruhi kelulusan mahasiswa. Model ini dapat digunakan sebagai dasar evaluasi akademik dan pengambilan keputusan di perguruan tinggi.



Gambar 3. Visualisasi Pohon Keputusan Decision

KESIMPULAN

Peningkatan akurasi secara nyata ini menunjukkan bahwa optimasi parameter model Decision Tree dapat memperbaiki performa klasifikasi secara substansial. Meskipun model memiliki keterbatasan inheren, hasil eksperimen ini menunjukkan bahwa Decision Tree dengan pengaturan yang tepat memiliki potensi untuk menjadi alat evaluasi yang efektif dalam konteks akademik.

Dengan akurasi yang telah dioptimasi sebesar 75,32%, model Decision Tree tidak hanya memberikan hasil klasifikasi yang lebih baik, tetapi juga menyediakan interpretasi yang jelas mengenai faktor-faktor yang memengaruhi kelulusan mahasiswa. Oleh karena itu, model ini dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan akademik di perguruan tinggi.

Penelitian ini menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan kelulusan mahasiswa tepat waktu menggunakan dataset berukuran besar sebanyak 100.000 data mahasiswa. Meskipun nilai akurasi model masih tergolong sedang, analisis numerik menunjukkan bahwa atribut IPK, jumlah semester, dan mata kuliah tidak lulus memiliki pengaruh signifikan terhadap kelulusan. Penambahan analisis statistik deskriptif memperkuat validitas hasil penelitian dan memberikan gambaran kuantitatif yang lebih jelas terkait kondisi akademik mahasiswa.

DAFTAR PUSTAKA

- Atyyatullatifah dan M. Kamayani, "Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa Teknik Informatika dengan Orange Data Mining," *Indonesian J. Comput. Sci.*, vol. 13, no. 2, pp. –, 2024. ijcs.net
- D. A. Wulandari, B. N. Sari, dan T. N. Padilah, "Prediction of Student Graduation Accuracy Using C45 Algorithm (Case Study: Fasilkom Unsika)," *SYSTEMATICS*, vol. 4, no. 1, 2025. *Jurnal UNSIKA*
- D. Destiani dan E. Rahmawati, "Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa," *J. Algoritma*, vol. 18, no. 2, 2025. *Jurnal*

- D. Kurniawan, A. Anggrawan, dan H. Hairani, “Graduation Prediction System on Students Using C4.5 Algorithm,” *J. MATRIK*, vol. 19, no. 2, 2025. *Jurnal Universitas Bumigora*
- E. P. Rohmawan, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Decision Tree dan Artificial Neural Network,” *J. Ilmiah Matrik*, 2025. *journal.binadarma.ac.id*
- G. Shevchenko, W. Maharani, dan A. Wijaya, “Penerapan Data Mining untuk Memprediksi Kelulusan Mahasiswa Menggunakan Decision Tree,” *J. Sains Teknol.*, 2025. *Jurnal Global Scientis*
- K. A. Putri, D. Febriawan, dan F. N. Hasan, “Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4.5),” *J. Sisfokom*, vol. 13, no. 1, 2025. *jurnal.atmaluhur.ac.id*
- M. B. Hartanto, T. Destanto, Y. Yuniarthe, dan T. Winarko, “Implementation of Data Mining for Classifying Student Graduation Levels Using Naive Bayes, Decision Tree, Random Forest, SVM and Neural Networks Methods,” *CCIT J.*, vol. 18, no. 1, pp. 80–87, 2025. *ejournal.raharja.ac.id*
- M. Riski Qisthiano, P. A. Prayesy, dan I. Ruswita, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa,” *G-Tech: J. Teknol. Terapan*, vol. 7, no. 1, 2025. *E- Journal Unira Malang*
- Mulia dan M. Muanas, “Prediction Model of Student Graduation Using Decision Tree C4.5 and Weka,” *Proc. Int. Conf. Global Optimiz. Its Applic.*, 2021. *Jurnal IBIK*
- Q. Ameliatus dan Z. Fatah, “Klasifikasi Kelulusan Mahasiswa Menggunakan Decision Tree,” *J. Riset Teknik Komputer*, 2025. *Journal Denasya*
- R. Lutunani dan A. Nugroho, “Analisis Prediksi Mahasiswa Terhadap Kelulusan Tepat Waktu Menggunakan Decision Tree,” *J. Teknol. Inform. Komunik.*, 2023. *Lembaga KITA*
- S. A. Setiono dan E. Purwanto, “Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree,” dalam *Proc. Seminar Nasional Teknol. Informasi dan Bisnis*, 2025. *Duta Bangsa University Journal*
- U. Suriani, “Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5,” *J. Comput. Inf. Syst. Ampera*, vol. 4, no. 2, 2025. *journal-computing.org*
- W. Putri, R. Rusdah, L. Suryadi, dan D. Anubhakti, “Prediction of Graduation for Students Using C4.5 Algorithm,” *J. Sisfokom*, vol. 12, no. 3, 2025. *jurnal.atmaluhur.ac.id*