

ABSTRAK

ANALISIS KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE DENGAN VISUALISASI POHON KEPUTUSAN

Nama : Muhammad Dary Nabhan

NPM : 2155201238

Pembimbing : Gunawan, S.Kom., M.Kom

Penelitian ini bertujuan merancang dan mengimplementasikan sistem berbasis algoritma *Decision Tree* untuk menganalisis dan memprediksi ketepatan waktu kelulusan mahasiswa Universitas Muhammadiyah Bengkulu, dibangun dalam bentuk aplikasi desktop menggunakan bahasa pemrograman Python dengan antarmuka grafis (GUI) berbasis Tkinter serta dilengkapi visualisasi pohon keputusan secara interaktif. Data yang digunakan merupakan data lulusan dari seluruh program studi pada Periode Wisuda Mei 2026 dan Oktober 2025 sebanyak 1.296 data mentah, yang selanjutnya melalui tahap pra-pemrosesan berupa pembersihan header ganda, penyeragaman format tanggal, dan pembentukan variabel turunan (Kategori IPK dan Status Kelulusan), sehingga menyisakan 1.144 data valid untuk dianalisis. Model dilatih menggunakan parameter *Criterion* Gini, *Max Depth* 6, *Min Samples Split* 2, dan pembobotan kelas (*class weight*) *balanced* untuk menangani ketidakseimbangan jumlah data antara kelas Tepat Waktu dan Terlambat, dengan proporsi data latih dan data uji sebesar 80:20 (912 data latih dan 229 data uji). Hasil pengujian menunjukkan model menghasilkan akurasi sebesar 82,1%, presisi sebesar 89,4%, *recall* sebesar 82,1%, dan *F1-score* sebesar 84,3%, serta rata-rata skor 10-Fold *Cross Validation* sebesar 76,9% dengan standar deviasi 0,0622 yang mengindikasikan model cukup stabil pada berbagai subset data. Berdasarkan struktur pohon keputusan yang terbentuk (kedalaman 6 tingkat dengan 31 node daun), faktor akademik yang paling berpengaruh terhadap ketepatan waktu kelulusan mahasiswa adalah predikat kelulusan dan kategori IPK, diikuti oleh program studi dan jenis kelamin. Aplikasi dibangun dengan struktur navigasi lima tab meliputi tab Data, Model & Evaluasi, Pohon Keputusan, *Confusion Matrix*, dan *Black Box Testing*, serta dilengkapi fitur ekspor hasil ke format Microsoft Excel. Pengujian fungsionalitas menggunakan *Black Box Testing* terhadap 15 skenario uji menghasilkan tingkat keberhasilan sebesar 100%

Kata Kunci: Kelulusan Mahasiswa, Decision Tree, Pohon Keputusan, Class Weight, Cross Validation, GUI, Python, Tkinter, Black Box Testing

ABSTRACT

STUDENT GRADUATION ANALYSIS USING THE DECISION TREE ALGORITHM WITH DECISION TREE VISUALIZATION

Name : Muhammad Dary Nabhan

NPM : 2155201238

Advisor : Gunawan, S.Kom., M.Kom

This study aims to design and implement a system based on the Decision Tree algorithm to analyze and predict the timely graduation of students at Universitas Muhammadiyah Bengkulu. The system was developed as a desktop application using Python with a Tkinter-based Graphical User Interface (GUI) and features interactive decision tree visualization. The dataset comprised 1,296 raw records of graduates from all study programs across the May 2026 and October 2025 graduation periods; following preprocessing steps—such as removing duplicate headers, standardizing date formats, and creating derived variables (GPA Category and Graduation Status)—1,144 valid records remained for analysis. The model was trained using the Gini criterion, a maximum depth of 6, a minimum sample split of 2, and "balanced" class weighting to address the data imbalance between the "Timely" and "Delayed" graduation classes, utilizing an 80:20 split for training and testing data (912 training samples and 229 test samples). Test results demonstrated an accuracy of 82.1%, precision of 89.4%, recall of 82.1%, and an F1-score of 84.3%, alongside an average 10-fold cross-validation score of 76.9% (standard deviation of 0.0622), indicating the model's stability across various data subsets. Based on the resulting decision tree structure (6 levels deep with 31 leaf nodes), the academic factors most influential on timely graduation were graduation honors and GPA category, followed by study program and gender. The application features a five-tab navigation structure—comprising Data, Model & Evaluation, Decision Tree, Confusion Matrix, and Black Box Testing tabs—and includes a function to export results to Microsoft Excel format.

Keywords: *Student Graduation, Decision Tree, Class Weight, Cross Validation, GUI, Python, Tkinter, Black Box Testing*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah membuka peluang besar dalam pemanfaatan data untuk mendukung pengambilan keputusan di berbagai bidang, termasuk dunia pendidikan tinggi. Perguruan tinggi menghasilkan data akademik mahasiswa dalam jumlah yang sangat besar setiap semesternya, namun pemanfaatan data tersebut masih belum optimal. Data seperti nilai Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), lama studi, dan predikat kelulusan berpotensi besar untuk dianalisis guna mengidentifikasi pola performa akademik mahasiswa secara sistematis (Alamgir *et al.*, 2024).

Ketepatan waktu kelulusan mahasiswa merupakan salah satu indikator penting dalam mengukur kualitas penyelenggaraan pendidikan tinggi. Keterlambatan kelulusan tidak hanya berdampak pada beban biaya pendidikan yang harus ditanggung mahasiswa dan orang tua, tetapi juga memengaruhi akreditasi program studi serta citra institusi secara keseluruhan (Sutoyo and Almaarif, 2020). Oleh karena itu, pihak akademik memerlukan alat bantu yang mampu mengidentifikasi pola dan faktor-faktor yang memengaruhi ketepatan waktu kelulusan mahasiswa secara lebih cepat, objektif, dan berbasis data.

Salah satu tantangan utama yang dihadapi institusi pendidikan tinggi adalah keterlambatan kelulusan mahasiswa. Keterlambatan ini tidak hanya berdampak pada mahasiswa secara pribadi, tetapi juga berpengaruh terhadap akreditasi program studi dan reputasi institusi. Oleh karena itu, diperlukan sebuah sistem yang mampu mendeteksi sejak dini mahasiswa yang berpotensi mengalami keterlambatan, sehingga pihak akademik dapat memberikan intervensi yang tepat waktu (Alwarthan, Aslam and Khan, 2022).

Universitas Muhammadiyah Bengkulu (UM Bengkulu), sebagaimana perguruan tinggi pada umumnya, mencatat data kelulusan mahasiswa pada setiap periode wisuda dalam bentuk berkas rekapitulasi yang memuat atribut seperti Nomor Pokok Mahasiswa (NPM), program studi, jenis kelamin, tanggal masuk, tanggal kelulusan, jumlah SKS yang ditempuh, Indeks Prestasi Kumulatif (IPK), dan predikat kelulusan. Selama ini, data tersebut umumnya hanya digunakan sebagai arsip administratif dan belum dimanfaatkan secara optimal untuk keperluan analisis prediktif oleh pihak akademik.

Salah satu pendekatan yang dapat digunakan untuk menganalisis dan memprediksi ketepatan waktu kelulusan adalah data mining dengan metode klasifikasi. Algoritma Decision Tree (pohon keputusan) dipilih dalam penelitian ini karena mampu menghasilkan model prediksi yang mudah diinterpretasikan secara visual, sehingga pihak akademik yang tidak memiliki latar belakang statistik atau ilmu komputer mendalam tetap dapat memahami faktor-faktor apa saja yang memengaruhi hasil prediksi (Sari *et al.*, 2025). Penelitian ini menggunakan data lulusan dari seluruh program studi di lingkungan universitas pada dua periode

wisuda terakhir, yaitu Periode Mei 2026 dan Periode Oktober 2025, dengan total 1.296 data lulusan mentah. Penggunaan data ini bertujuan agar sistem yang dibangun dapat menganalisis pola ketepatan waktu kelulusan secara lebih menyeluruh dan representatif, serta melihat apakah pola yang terbentuk konsisten antar program studi dan antar periode wisuda.

Beberapa penelitian terdahulu telah membuktikan efektivitas Decision Tree dalam analisis performa akademik. Penelitian yang dilakukan oleh (Alvian Setiono and Purwanto, 2025) berhasil mengklasifikasikan performa mahasiswa dengan akurasi di atas 80% dengan atribut kehadiran teridentifikasi sebagai faktor prediktor paling signifikan, diikuti oleh IPK dan progres perolehan SKS. Sementara itu, penelitian oleh (Dofiyanto and Fatah, 2025) bahwa algoritma Decision Tree C4.5 handal dalam menentukan tren kelulusan mahasiswa menggunakan campuran elemen akademik dan sosial, dan dapat digunakan sebagai dasar pengambilan keputusan akademik untuk meningkatkan efektivitas proses pendidikan tinggi.

Berdasarkan latar belakang di atas, penelitian ini bertujuan untuk mengembangkan aplikasi desktop berbasis Python Tkinter yang mengintegrasikan algoritma Decision Tree untuk menganalisis kelulusan mahasiswa UM Bengkulu disertai visualisasi pohon keputusan secara interaktif

1.2 Pertanyaan Penelitian

1. Bagaimana merancang dan membangun sistem berbasis algoritma Decision Tree yang mampu memuat, membersihkan, dan memproses data kelulusan mahasiswa dari berkas Excel multi-periode secara otomatis?

2. Faktor akademik apa saja yang paling berpengaruh terhadap ketepatan waktu kelulusan mahasiswa berdasarkan struktur pohon keputusan yang terbentuk?
3. Bagaimana performa model Decision Tree dalam memprediksi ketepatan waktu kelulusan mahasiswa, khususnya dalam menangani kondisi data yang tidak seimbang antar kelas (class imbalance)?
4. Bagaimana hasil pengujian fungsionalitas (Black Box Testing) terhadap sistem yang telah dibangun?

1.3 Tujuan Penelitian

1. Merancang dan mengimplementasikan sistem/aplikasi desktop berbasis algoritma Decision Tree yang dapat memuat dan memproses data kelulusan mahasiswa multi-periode secara otomatis.
2. Mengidentifikasi variabel akademik (jenis kelamin, kategori IPK, predikat kelulusan, dan program studi) yang paling berpengaruh terhadap ketepatan waktu kelulusan melalui visualisasi pohon keputusan.
3. Mengevaluasi performa model Decision Tree menggunakan metrik akurasi, presisi, recall, F1-score, dan K-Fold Cross Validation, termasuk penanganan ketidakseimbangan kelas menggunakan pembobotan kelas (class weight).
4. Melakukan pengujian Black Box terhadap seluruh fungsi sistem untuk memastikan sistem berjalan sesuai dengan spesifikasi yang dirancang.

1.4 Kerangka Kerja Penelitian (*Research Framework*)

Framework penelitian ini menggambarkan alur tahapan pelaksanaan secara keseluruhan dari tahap awal hingga menghasilkan luaran akhir. Berikut adalah kerangka penelitian yang disajikan dalam bentuk tabel:

Tabel 1. Framework Penelitian

No	Tahapan	Deskripsi
1	Identifikasi Masalah	Mengidentifikasi kebutuhan analisis ketepatan waktu kelulusan mahasiswa dan permasalahan pada pengelolaan data secara manual.
2	Studi Literatur	Mengkaji penelitian terdahulu serta teori terkait data mining, algoritma Decision Tree, dan metrik evaluasi model klasifikasi.
3	Pengumpulan Data	Mengumpulkan data kelulusan mahasiswa Periode Mei 2026 dan Periode Oktober 2025 dari bagian akademik universitas.
4	Pra-pemrosesan Data	Membersihkan data (baris header ganda, format tanggal tidak konsisten), menormalisasi nama kolom, dan membentuk variabel turunan (Kategori IPK, Status Kelulusan).
5	Perancangan Sistem	Merancang arsitektur sistem, alur data (data flow), dan antarmuka pengguna (GUI) aplikasi.
6	Implementasi	Mengimplementasikan sistem menggunakan bahasa pemrograman Python beserta pustaka Tkinter, Pandas, Scikit-learn, dan Matplotlib.
7	Pengujian dan Evaluasi	Menguji model menggunakan confusion matrix dan K-Fold Cross Validation, serta menguji fungsionalitas sistem menggunakan Black Box Testing.
8	Penarikan Kesimpulan	Menyimpulkan hasil penelitian serta memberikan saran untuk pengembangan lebih lanjut.

Framework di atas menggambarkan alur penelitian yang dimulai dari identifikasi masalah hingga dokumentasi akhir. Setiap tahapan menghasilkan luaran yang menjadi masukan bagi tahapan berikutnya sehingga membentuk proses penelitian yang sistematis dan terukur. Tahapan integrasi sistem merupakan tahap kritis di mana seluruh komponen model Decision Tree, visualisasi matplotlib, dan antarmuka Tkinter digabungkan menjadi satu aplikasi desktop yang fungsional

