

Student Graduation Analysis Using The Decision Tree Algorithm With Decision Tree Visualization

Analisis Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree dengan Visualisasi Pohon Keputusan

Muhammad Dary Nabhan ¹⁾; Gunawan ²⁾; Yovi Apridiansyah ³⁾; Muhammad Husni Rifqo ⁴⁾
^{1,2,3,4)} Study Program of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Bengkulu
Email: ¹⁾ darynabhan04@gmail.com ; ²⁾ gunawan@umb.ac.id ; ³⁾ yoviapridiansyah@umb.ac.id ; ⁴⁾ mhrifqo@umb.ac.id

How to Cite :

Nabhan, M, D., Gunawan., Apridiansyah, Y., Rifqo, M, H. (2026). Student Graduation Analysis Using The Decision Tree Algorithm With Decision Tree Visualization. Jurnal Komputer Indonesia, 5(3).

ARTICLE HISTORY

Received [07 Agustus 2026]

Revised [09 September 2026]

Accepted [11 September 2026]

KEYWORDS

Student Graduation, Decision Tree, Class Weight, Cross Validation, GUI, Python, Tkinter, Black Box Testing.

This is an open access article under the
[CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

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%.

ABSTRACT

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.

PENDAHULUAN

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.

LANDASAN TEORI

Data Mining

Data mining merupakan proses penggalian pola, hubungan, atau informasi yang bermanfaat dari sekumpulan data berukuran besar melalui pendekatan statistik, matematika, dan kecerdasan buatan. Proses data mining umumnya mencakup tahapan pembersihan data (data cleaning), integrasi data, transformasi data, penggalian pola (pattern mining), evaluasi pola, dan presentasi pengetahuan yang dihasilkan (Ridwansyah *et al.*, 2024). Dalam konteks penelitian ini, data mining diterapkan pada data kelulusan mahasiswa untuk menemukan pola yang membedakan mahasiswa yang lulus tepat waktu dengan yang mengalami keterlambatan.

Klasifikasi Dalam Data Mining

Klasifikasi merupakan salah satu teknik data mining yang termasuk dalam kategori pembelajaran terarah (supervised learning), yaitu proses membangun model dari data yang telah memiliki label/kelas target untuk kemudian digunakan memprediksi label dari data baru yang belum diketahui kelasnya. Proses klasifikasi umumnya terbagi menjadi dua tahap: tahap pembelajaran (training), di mana model mempelajari pola dari data latih (training set) yang telah diketahui labelnya dan tahap pengujian (testing), di mana model yang telah terbentuk diuji kemampuannya memprediksi label pada data uji (testing set) yang belum pernah dilihat sebelumnya, untuk mengukur kemampuan generalisasi model terhadap data baru (Rahmadeyan and Mustakim, 2023).

Decision Tree (Pohon Keputusan)

Decision Tree adalah salah satu metode klasifikasi yang merepresentasikan pengetahuan dalam bentuk struktur pohon, terdiri atas simpul akar (root node) yang merupakan atribut dengan daya pisah terbaik terhadap keseluruhan data, simpul internal (internal node) yang merepresentasikan pengujian terhadap suatu atribut, cabang (branch) yang merepresentasikan hasil pengujian tersebut, dan simpul daun (leaf node) yang merepresentasikan label kelas akhir (Nasrullah, 2021). Keunggulan utama Decision Tree dibandingkan metode klasifikasi lain adalah kemudahan interpretasi, karena aturan keputusan yang terbentuk dapat divisualisasikan dan dipahami secara langsung tanpa memerlukan pengetahuan statistik yang mendalam sebuah pertimbangan penting pada penelitian ini karena hasil analisis ditujukan untuk digunakan oleh pihak akademik yang tidak selalu memiliki latar belakang teknis.

Gini Index, Entropy, dan Information Gain

Gini Index dan Entropy merupakan dua ukuran impuritas (ketidakmurnian) yang umum digunakan sebagai kriteria pemilihan atribut pemisah (split) terbaik pada setiap simpul pohon keputusan. Gini Index mengukur seberapa sering suatu elemen yang dipilih secara acak akan salah

diklasifikasikan apabila diberi label secara acak sesuai distribusi kelas pada simpul tersebut, dirumuskan sebagai (Banafshah Shafa *et al.*, 2024):

$$Gini(t) = 1 - \sum_{i=1}^n (p(i|t))^2$$

Dengan $p(i|t)$ adalah proporsi data kelas i pada simpul t . Entropy mengukur tingkat ketidakteraturan (disorder) suatu kumpulan data, dirumuskan sebagai:

$$Entropy(t) = - \sum_{i=1}^n (p(i|t) \log_2 p(i|t))$$

Semakin rendah nilai Gini Index atau Entropy suatu simpul (mendekati 0), semakin murni (homogen) komposisi kelas pada simpul tersebut — nilai 0 menandakan seluruh data pada simpul berasal dari satu kelas yang sama. Untuk menentukan atribut mana yang digunakan sebagai pemisah pada suatu simpul, algoritma menghitung penurunan impuritas (impurity decrease) yang dihasilkan oleh setiap kandidat atribut, dikenal sebagai Information Gain apabila menggunakan Entropy, dirumuskan sebagai:

$$IG(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Dengan S adalah himpunan data pada simpul saat ini, A adalah atribut kandidat pemisah, dan S_v adalah subset data hasil pemisahan berdasarkan nilai v dari atribut A . Atribut dengan nilai Information Gain (atau penurunan Gini Index) terbesar dipilih sebagai pemisah pada simpul tersebut. Penelitian ini menggunakan kriteria Gini Index sebagai pengaturan bawaan (default) karena perhitungannya lebih ringan secara komputasi dibandingkan Entropy (tidak melibatkan operasi logaritma), dengan opsi Entropy juga disediakan sebagai parameter yang dapat dipilih pengguna pada antarmuka aplikasi

Feature Importance

Feature importance merupakan ukuran yang menunjukkan seberapa besar kontribusi setiap atribut (fitur) terhadap kemampuan prediksi model Decision Tree secara keseluruhan. Pada implementasi Scikit-learn, nilai feature importance suatu atribut dihitung berdasarkan total penurunan impuritas yang dihasilkan oleh atribut tersebut di seluruh simpul tempat atribut itu digunakan sebagai pemisah, dibobotkan berdasarkan proporsi jumlah sampel yang melewati simpul tersebut, dirumuskan secara umum sebagai (Nugraha, Supriadi and Setiadi, 2026):

$$Importance(A) = \sum_{t \in A} \frac{N_t}{N} \Delta Impurity(t)$$

Dengan t adalah simpul yang menggunakan atribut A sebagai pemisah, N_t adalah jumlah sampel pada simpul t , N adalah jumlah total sampel, dan $\Delta Impurity(t)$ adalah penurunan nilai impuritas (Gini/Entropy) akibat pemisahan pada simpul t . Nilai feature importance kemudian dinormalisasi sehingga jumlah seluruh atribut bernilai 1. Metrik ini digunakan pada penelitian ini untuk mengidentifikasi variabel akademik yang paling berpengaruh terhadap ketepatan waktu kelulusan mahasiswa, dan divisualisasikan dalam bentuk grafik batang pada tab Confusion Matrix aplikasi.

Ketidakseimbangan Kelas (Class Imbalance) dan Class Weight

Ketidakseimbangan kelas (class imbalance) terjadi ketika proporsi jumlah data pada masing-masing kelas target tidak seimbang secara signifikan. Salah satu solusi untuk mengatasi hal ini adalah dengan memberikan bobot lebih besar pada kelas minoritas selama proses pelatihan model (class weighting), sehingga kesalahan klasifikasi pada kelas minoritas diberi penalti lebih besar dibandingkan kesalahan pada kelas mayoritas.

Confusion Matrix dan Matrik Evaluasi

Confusion matrix adalah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model dengan label aktual, terdiri atas empat komponen: True Positive (TP), yaitu data positif yang diprediksi benar sebagai positif; True Negative (TN), yaitu data negatif yang diprediksi benar sebagai negatif; False Positive (FP), yaitu data negatif yang salah diprediksi sebagai positif; dan False Negative (FN), yaitu data positif yang salah diprediksi sebagai negative (AlShammari, 2024).

K-Fold Cross Validation

K-Fold Cross Validation merupakan teknik validasi model dengan cara membagi data menjadi K bagian (fold) berukuran relatif sama. Model dilatih dan diuji sebanyak K kali, di mana pada setiap putaran ke-i, fold ke-i digunakan sebagai data uji dan K-1 fold sisanya digunakan sebagai data latih. Skor evaluasi (umumnya akurasi) dari seluruh putaran kemudian dirata-ratakan untuk memperoleh estimasi performa model yang lebih stabil, dirumuskan sebagai (Wijiyanto *et al.*, 2024):

$$CV_{score} = \frac{1}{K} \sum_{i=1}^K score_i$$

dengan $score_i$ adalah skor evaluasi pada putaran (fold) ke-i. Selain nilai rata-rata, standar deviasi antar fold juga dihitung untuk menilai stabilitas model nilai standar deviasi yang kecil menunjukkan performa model relatif konsisten pada berbagai kombinasi pembagian data, sedangkan nilai yang besar mengindikasikan model sensitif terhadap data tertentu (berpotensi overfitting pada kombinasi data tertentu). Penelitian ini menggunakan skema Stratified K-Fold, yaitu varian K-Fold yang menjaga proporsi kelas target pada setiap fold tetap sebanding dengan proporsi pada keseluruhan data hal ini penting mengingat kondisi ketidakseimbangan kelas pada data penelitian, agar setiap fold tetap memuat representasi kelas Terlambat yang memadai. Penelitian ini menggunakan 10-Fold Cross Validation (K=10) sebagai metode validasi utama.

Python dan Pustaka Pendukung

Python merupakan bahasa pemrograman tingkat tinggi yang banyak digunakan dalam pengembangan aplikasi data science dan machine learning karena sintaksnya yang sederhana dan didukung oleh ekosistem pustaka yang luas (Santoso and Priyadi, 2024). Penelitian ini menggunakan beberapa pustaka pendukung sebagai berikut. Tkinter digunakan untuk membangun antarmuka grafis (GUI) desktop, termasuk komponen tab navigasi (Notebook), tabel data (Treeview), tombol interaktif, dan dropdown parameter. Pandas digunakan untuk manipulasi dan pra-pemrosesan data tabular, meliputi pembacaan berkas Excel/CSV, penggabungan data multi-sheet, pembersihan baris tidak valid, dan pembentukan variabel turunan. Scikit-learn digunakan untuk implementasi algoritma Decision Tree (DecisionTreeClassifier), pembagian data latih/uji (train_test_split), validasi silang (cross_val_score, StratifiedKFold), pengkodean label kategorikal (LabelEncoder), serta perhitungan metrik evaluasi (accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report). Matplotlib digunakan untuk visualisasi pohon keputusan (plot_tree), confusion matrix, dan grafik feature importance, yang kemudian disematkan ke dalam antarmuka Tkinter melalui FigureCanvasTkAgg. OpenPyXL digunakan secara internal oleh Pandas untuk membaca dan menulis berkas Excel (.xlsx), termasuk penataan gaya (styling) pada berkas hasil ekspor.

Pengujian Black Box

Black Box Testing merupakan metode pengujian perangkat lunak yang berfokus pada fungsionalitas sistem tanpa memeriksa struktur kode program secara internal (diperlakukan sebagai "kotak hitam" yang hanya diamati dari masukan dan keluarannya) (Ardiansyah and Sela, 2025). Pengujian dilakukan dengan memberikan input tertentu pada sistem dan memeriksa apakah keluaran (output) yang dihasilkan sesuai dengan yang diharapkan (expected result). Salah satu teknik perancangan skenario pengujian Black Box yang umum digunakan adalah equivalence partitioning, yaitu mengelompokkan kemungkinan input ke dalam kelas-kelas setara (misalnya "data valid" dan "data tidak valid") untuk memastikan cakupan pengujian yang representatif tanpa perlu menguji seluruh kemungkinan input satu per satu.

METODE PENELITIAN

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).

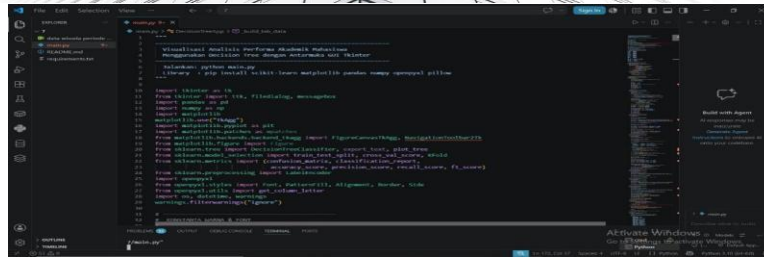
HASIL DAN PEMBAHASAN

Hasil

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%.

Uji Coba

Sebelum masuk ke desain GUI nya atau antarmuka system ini dibangun dengan tool VS Code seperti pada gambar berikut.

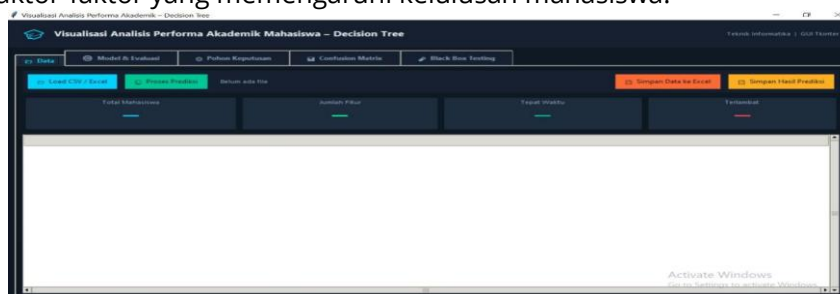


Gambar 1 Coding Analisis Performa Akademik VS Code

Gambar di atas menunjukkan cuplikan awal berkas main.py yang dikembangkan menggunakan Visual Studio Code, tempat seluruh logika aplikasi Visualisasi Analisis Performa Akademik Mahasiswa. Gambar tersebut merupakan coding yang digunakan untuk menjalankan program. Untuk menampilkan hasil cukup dengan menekan tombol Run maka tampilan akan muncul antarmuka sistem.

Tampilan Utama

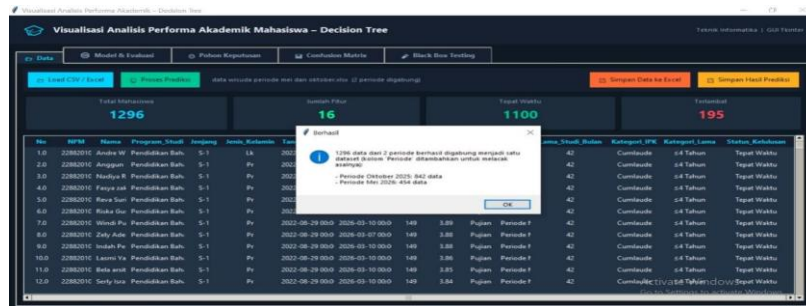
Antarmuka utama dari sistem visualisasi ini dirancang untuk memberikan kemudahan bagi pengguna dalam memahami proses analisis kelulusan mahasiswa menggunakan algoritma *Decision Tree*. Pada bagian atas, terdapat tab navigasi yang memisahkan beberapa fungsi utama, yaitu *Data*, *Model* & *Evaluasi*, *Pohon Keputusan*, *Confusion Matrix*, dan *Black Box Testing*. Tab ini memungkinkan pengguna untuk berpindah antar tahapan analisis secara terstruktur. Di area tengah, tersedia tombol aksi seperti *Load CSV / Excel* untuk mengunggah data, *Proses Prediksi* untuk menjalankan model, serta tombol *Simpan Data ke Excel* dan *Simpan Hasil Prediksi* untuk menyimpan hasil olahan. Indikator *progress bar* menampilkan ringkasan statistik seperti total mahasiswa, jumlah fitur, serta kategori kelulusan tepat waktu dan terlambat, yang memberikan gambaran cepat tentang komposisi data. Bagian bawah layar menampilkan status aktivasi Windows, yang menunjukkan bahwa antarmuka ini berjalan pada lingkungan sistem operasi Windows. Secara keseluruhan, tata letak antarmuka mendukung alur kerja analisis data yang sistematis, mulai dari input data, pemrosesan model, hingga visualisasi hasil, sehingga memudahkan pengguna dalam mengevaluasi faktor-faktor yang memengaruhi kelulusan mahasiswa.



Gambar 2 Tampilan Antarmuka Analisis Kelulusan Mahasiswa

Tampilan antarmuka sistem ini memperlihatkan panel kontrol utama yang terdiri dari tab navigasi, tombol aksi, dan ringkasan statistik data mahasiswa. Tab *Pohon Keputusan* menampilkan visualisasi grafis dari struktur pohon keputusan yang dihasilkan, sedangkan tab *Confusion Matrix* menyajikan matriks evaluasi untuk mengukur akurasi model. *Progress bar* di bagian tengah memberikan informasi kuantitatif tentang jumlah data yang diproses, yang mencerminkan distribusi mahasiswa berdasarkan status kelulusan. Dengan adanya tombol *Simpan Hasil Prediksi*, pengguna dapat mendokumentasikan output prediksi untuk keperluan pelaporan atau tindak lanjut akademik. Desain antarmuka yang informatif dan interaktif ini bertujuan untuk mendukung pengambilan keputusan berbasis data dalam upaya meningkatkan angka kelulusan mahasiswa tepat waktu.

Proses selanjutnya yaitu input data kelulusan mahasiswa untuk mengetahui hasil prediksi, proses tersebut terlihat pada gambar berikut :



Gambar 3 Tampilan Antarmuka Input Data Kelulusan

Tampilan antarmuka sistem ini menunjukkan panel utama aplikasi berbasis GUI Tkinter dengan judul "Teknik Informatika | GUI Tkinter". Pada bagian atas terdapat lima tab navigasi, yaitu *Data*, *Model & Evaluasi*, *Pohon Keputusan*, *Confusion Matrix*, dan *Black Box Testing*, yang memungkinkan pengguna untuk mengakses setiap tahapan analisis secara terpisah. Di area tengah, tombol *Load CSV / Excel* dan *Proses Prediksi* menjadi fitur utama untuk memasukkan data dan menjalankan model prediksi. Tabel data menampilkan 12 sampel mahasiswa dengan atribut seperti NPM, Nama, Program Studi, Jenjang, Jenis Kelamin, dan Tahun Masuk, yang mencerminkan struktur dataset yang digunakan. Pada bagian ringkasan statistik, terlihat bahwa total mahasiswa mencapai 1.296 data dari dua periode wisuda, yaitu Oktober 2025 (842 data) dan Mei 2026 (454 data), dengan 16 fitur yang dianalisis, serta rincian kelulusan tepat waktu sebanyak 1.100 mahasiswa dan terlambat sebanyak 195 mahasiswa. Informasi ini memberikan gambaran komprehensif tentang komposisi data yang telah berhasil digabungkan, dengan penambahan kolom 'Periode' sebagai pelacak asal data. Secara keseluruhan, antarmuka ini dirancang untuk memudahkan pengguna dalam mengelola data, menjalankan algoritma Decision Tree, serta mengevaluasi hasil prediksi kelulusan mahasiswa secara visual dan interaktif.

KESIMPULAN DAN SARAN

Kesimpulan

1. Berdasarkan Sistem berbasis algoritma Decision Tree berhasil dirancang dan diimplementasikan dalam bentuk aplikasi desktop Python-Tkinter yang mampu memuat data kelulusan mahasiswa dari berkas Excel multi-periode (multi-sheet) secara otomatis, membersihkan data yang tidak konsisten (header ganda, format tanggal campuran), serta melatih model klasifikasi tanpa memerlukan intervensi manual pada tahap pra-pemrosesan.
2. Berdasarkan struktur pohon keputusan yang terbentuk, faktor akademik yang paling berpengaruh terhadap ketepatan waktu kelulusan mahasiswa adalah predikat kelulusan dan kategori IPK, yang secara konsisten muncul sebagai pemisah pada tingkat-tingkat atas pohon keputusan, diikuti oleh program studi dan jenis kelamin.
3. Model Decision Tree akhir (Criterion Gini, Max Depth 6, Class Weight Balanced) menghasilkan akurasi sebesar 82,1%, presisi sebesar 89,4%, recall sebesar 82,1%, F1-score tertimbang sebesar 84,3%. Nilai 6 dipilih melalui pertimbangan antara akurasi model dan risiko overfitting, biasanya melalui eksperimen (trial and error) atau teknik tuning seperti grid search.
4. Pengujian Black Box Testing terhadap 15 skenario fungsional sistem menunjukkan seluruh skenario berhasil (PASS), yang mengindikasikan bahwa sistem berjalan sesuai dengan spesifikasi fungsional yang dirancang, mencakup pemuatan data, pelatihan model, visualisasi, prediksi, dan ekspor hasil.

Saran

1. Melakukan koordinasi dengan bagian akademik untuk melengkapi data IPK yang kosong pada berkas sumber, mengingat sebanyak 151 dari 1.296 data (11,7%) pada penelitian ini tidak dapat digunakan akibat data IPK yang tidak terisi.

2. Menambahkan variabel akademik lain yang berpotensi relevan, seperti nilai per semester, status pekerjaan mahasiswa, atau jalur masuk perkuliahan, apabila data tersebut tersedia, untuk meningkatkan akurasi dan kedalaman analisis model.
3. Membandingkan performa algoritma Decision Tree dengan algoritma klasifikasi lain seperti Random Forest atau Gradient Boosting, yang berpotensi menghasilkan akurasi lebih tinggi meskipun dengan interpretabilitas yang lebih rendah dibandingkan Decision Tree tunggal.
4. Menambah cakupan data pada periode-periode wisuda berikutnya secara berkelanjutan, agar model dapat dilatih ulang secara berkala (retraining) dan tetap relevan terhadap pola kelulusan mahasiswa yang mungkin berubah dari waktu ke waktu.

DAFTAR PUSTAKA

- Alamgir, Z. *et al.* (2024) "Enhancing Student Performance Prediction via Educational Data Mining on Academic Data," *Informatics in Education*, 23(1), pp. 1–24. Available at: <https://doi.org/10.15388/infedu.2024.04>.
- Alfiani Abdilah *et al.* (2025) "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, 7(2), pp. 401–406. Available at: <https://doi.org/10.47701/4q3z9j41>.
- AlShammari, A.F. (2024) "Implementation of Model Evaluation Using Confusion Matrix in Python," *International Journal of Computer Applications*, 186(50), pp. 42–48. Available at: <https://doi.org/10.5120/ijca2024924236>.
- Alvian Setiono, S. and Purwanto, E. (2025) "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, pp. 401–406. Available at: <https://doi.org/10.47701/4q3z9j41>.
- Alwarthan, S.A., Aslam, N. and Khan, I.U. (2022) "Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review," *Applied Computational Intelligence and Soft Computing*, 2022. Available at: <https://doi.org/10.1155/2022/8924028>.
- Ardiansyah, B. and Sela, E.I. (2025) "Evaluasi dan Desain Antarmuka Pengguna Aplikasi Kesehatan Berbasis Kebutuhan Pengguna Informatika , Fakultas Ilmu Komputer , Universitas Teknologi Yogyakarta , Indonesia EVALUATION AND DESIGN OF HEALTH APPLICATION USER INTERFACE," *JPTI*, 5(5), pp. 1497–1511. Available at: <https://doi.org/https://doi.org/10.52436/1.jpti.630>.
- Banafshah Shafa *et al.* (2024) "Prediksi Kanker Paru dengan Normalisasi menggunakan Perbandingan Algoritma Random Forest, Decision Tree dan Naïve Bayes," *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3), pp. 1057–1070. Available at: <https://doi.org/10.51454/decode.v4i3.779>.
- Dofiyanto and Fatah, Z. (2025) "Penerapan Algoritma Decision Tree untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Faktor Akademik dan Sosial," *JISCO: Journal of Information System and Computing*, 3(2), pp. 66–76. Available at: <https://doi.org/10.30631/jisco.v3i2.4030>.
- Irwan *et al.* (2022) "Penerapan Pohon Keputusan Dalam Memprediksi Masa Studi Mahasiswa Uin Alauddin Makassar," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, 7(2), pp. 201–210. Available at: <https://doi.org/10.24252/instek.v7i2.31390>.
- Jevintya, N.R., Darussalam, U. and Abdullah, S. (2024) "Application of the K-Means and Decision Tree Algorithms in Determining Student Achievement," *JIKO (Jurnal Informatika dan Komputer)*, 7(1), pp. 13–18. Available at: <https://doi.org/10.33387/jiko.v7i1.7580>.
- Nasrullah, A.H. (2021) "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer*, 7(2), pp. 45–51. Available at: <https://doi.org/10.35329/jiik.v7i2.203>.
- Nuari Anisa Sivi, Rudi Hartono and Putra Hanafi (2023) "Penerapan Algoritma C4.5 untuk Prediksi Kelulusan Mahasiswa berdasarkan Data Akademik," *Polygon: Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam*, 1(5), pp. 01–17. Available at: <https://doi.org/10.62383/polygon.v1i5.855>.

- Nugraha, V.P., Supriadi, F. and Setiadi, D. (2026) "Decision Tree-Based Early Warning System for Academic Failure: Comparative Analysis with Random Forest and Logistic Regression," *JASMINE: Journal of Intelligent Systems and Machine Learning*, 1(1), pp. 11–21. Available at: <https://journals.telkomuniversity.ac.id/jasmine/article/view/10123>.
- Nur Sofiyanto, I. et al. (2025) "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (KNN) dengan Klasifikasi Biner," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, pp. 40–45. Available at: <https://doi.org/10.47701/gpqsm502>.
- Purnama, J.J. et al. (2020) "Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Classification of Her Students Based on Svm Algorithm and," *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(6), pp. 1253–1260. Available at: <https://doi.org/10.25126/jtiik.202073080>.
- Rafi Jonathan Siger, Muhammad Naufal and Farikh Alzami (2026) "Analisis Performa Class Weight Dan Focal Loss Pada Model Indobert Untuk Klasifikasi Teks Depresi Berbahasa Indonesia," *JURIKOM (Jurnal Riset Komputer)*, 13(2), pp. 558–568. Available at: <https://doi.org/10.30865/jurikom.v13i2.9620>.
- Rahmadeyan, A. and Mustakim, M. (2023) "Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19," *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(1), pp. 21–32. Available at: <https://doi.org/10.25077/teknosi.v9i1.2023.21-32>.
- Ridwansyah, R. et al. (2024) "Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan," *SemanTIK: Teknik Informasi*, 10(2), pp. 1–10. Available at: <https://doi.org/10.55679/semantik.v10i2.67>.
- Rizalno, M.F., Johar, A. and Coastera, F.F. (2022) "Analisis Prediksi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritme Cart (Classification and Regression Trees) (Studi Kasus Data Alumni Fakultas Teknik Universitas Bengkulu)," *Rekursif: Jurnal Informatika*, 10(1), pp. 96–106. Available at: <https://doi.org/10.33369/rekursif.v10i1.21362>.
- Santoso, L. and Priyadi (2024) "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," *Jurnal Elektronika Dan Komputer*, 17(2), pp. 657–664. Available at: <https://doi.org/https://doi.org/10.51903/elkom.v17i2.2311>.
- Sari, S.N. et al. (2025) "Teknik Data Mining Dengan Menggunakan Algoritma Decision Tree Untuk Mengetahui Pola Pemahaman Mahasiswa Pada Matakuliah Pemrograman," *Bulletin of Information Technology (BIT)*, 6(4), pp. 417–431. Available at: <https://doi.org/10.47065/bit.v5i2.2339>.
- Setio, P.B.N., Saputro, D.R.S. and Bowo Winarno (2020) "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," *PRISMA, Prosiding Seminar Nasional Matematika*, 3, pp. 64–71.
- Sutoyo, E. and Almaarif, A. (2020) "Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *Masa Berlaku Mulai*, 4(1), pp. 95–101. Available at: <https://doi.org/https://doi.org/10.29207/resti.v4i1.1502>.
- Taufik Agus Haryanto and Eka Rini Yulia (2025) "Analisis Usability Aplikasi Streaming Musik Menggunakan Metode System Usability Scale (Sus)," *Jurnal Mahasiswa Sistem Informasi (JMSI)*, 7(1), pp. 247–254. Available at: <https://doi.org/10.24127/jmsi.v7i1.9809>.
- Wicandi, N.P.W.C., Utami, N.W. and Dewi, P.A.C. (2026) "Application of data mining for analyzing book borrowing patterns in SMP Negeri 2 Denpasar library using the Apriori algorithm," *Jurnal Ilmiah Teknologi Informasi Asia*, 20(1). Available at: <https://doi.org/https://doi.org/10.32815/jitika.1232>.
- Wijiyanto, W. et al. (2024) "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, 21(1), pp. 239–248. Available at: <https://doi.org/10.33364/algoritma/v.21-1.1618>.