E-ISSN : 3025-1311
https://ejournal.cahayailmubangsa.institute/index.php/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 10 No 8 Tahun 2026

IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK PREDIKSIKELULUSAN MAHASISWA: STUDI EKSPERIMENTAL

Diar Sohibul Fatoni[1], Fachri Ramadhan[2], Muhammad Akbar Hasri P[3], Muhammad Ammartsaqif A[4]

Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Bekasi, Jawa Barat, Indonesia 17113

17230707@bsi.ac.id[1], 17230720@bsi.ac.id[2], 17230719@bsi.ac.id[3], 17230977@bsi.ac.id[4]
Corresponding author: 17230707@bsi.ac.id

ABSTRACT

Timely graduation is an important indicator of higher education management success. This study aims to implement and evaluate the Decision Tree C4.5 algorithm for predicting student graduation based on academic data. The research employs a quantitative approach with an experimental method using a secondary dataset from Kaggle platform consisting of 51 student records. The variables include Grade Point Average (GPA), attendance percentage, final exam scores, academic activities, and organizational involvement. Implementation was conducted using Python with scikit-learn library. Results show that the Decision Tree C4.5 model achieves 96% accuracy with GPA as the most dominant factor. This study demonstrates the feasibility of C4.5 algorithm for graduation prediction and can serve as a baseline for implementation with real institutional data.

Keywords: *Graduation prediction, Decision Tree C4.5, academic data, machine learning, experimental study*

ABSTRAK

Kelulusan tepat waktu merupakan indikator penting keberhasilan penyelenggaraan pendidikan di perguruan tinggi. Penelitian ini bertujuan mengimplementasikan dan mengevaluasi algoritma Decision Tree C4.5 untuk prediksi kelulusan mahasiswa berbasis data akademik. Metode penelitian menggunakan pendekatan kuantitatif dengan metode eksperimental menggunakan data sekunder dari platform Kaggle yang terdiri dari 51 records mahasiswa. Variabel yang digunakan meliputi Indeks Prestasi Kumulatif (IPK), persentase kehadiran, nilai ujian akhir, aktivitas akademik, dan keterlibatan organisasi. Implementasi dilakukan menggunakan Python dengan library scikit-learn. Hasil penelitian menunjukkan model Decision Tree C4.5 mencapai
akurasi 96% dengan IPK sebagai faktor paling dominan. Penelitian ini mendemonstrasikan kelayakan algoritma C4.5 untuk prediksi kelulusan dan dapat menjadi acuan untuk implementasi pada data institusi nyata.

Kata kunci: *Prediksi kelulusan, Decision Tree C4.5, data akademik, machine learning, studi eksperimental*

Article History
Received : Januari 2026
Reviewed: Januari 2026
Published: Januari 2026

COPYRIGHT : AUTHOR
PUBLISH BY : KOHESI



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

1. PENDAHULUAN

Tingkat kelulusan mahasiswa merupakan salah satu faktor penting dalam menilai kualitas pendidikan tinggi di Indonesia. Hanya sekitar 30% dari mahasiswa lulus tepat waktu, menurut data dari Badan Pusat Statistik tahun 2023. Yang tersisa akan membutuhkan waktu satu atau beberapa semester tambahan untuk lulus. mengakhiri penelitian mereka (BPS, 2023).[1]. Permasalahan ini berdampak tidak hanya pada mahasiswa secara individual, tetapi juga pada institusi pendidikan dalam hal efisiensi dan akreditasi program studi. Berbagai faktor dapat mempengaruhi kelulusan mahasiswa.

Faktor internal dapat berasal dari mahasiswa sendiri, seperti minat, motivasi, karakteristik, dan kemampuan mereka; faktor eksternal dapat berasal dari lingkungan sosial, tempat tinggal, dan sejarah sekolah mereka.[2].Identifikasi dini terhadap mahasiswa yang berpotensi mengalami kesulitan dalam menyelesaikan studi menjadi sangat penting agar institusi dapat memberikan intervensi yang tepat waktu dan efektif.

Perguruan tinggi memiliki tanggung jawab besar untuk menghasilkan lulusan yang memiliki kemampuan untuk bersaing di dunia kerja. Salah satu tantangan yang sering dihadapi adalah bagaimana memastikan mahasiswa dapat menyelesaikan studi mereka tepat waktu. Tingkat kelulusan mahasiswa berpengaruh signifikan terhadap akreditasi dan reputasi institusi pendidikan,. Di sisi lain, penurunan tingkat kelulusan akan berdampak pada akreditasi universitas. Oleh karena itu, pemantauan dan evaluasi kecenderungan mahasiswa untuk lulus tepat waktu sangatlah penting.[3].

Faktanya, masih banyak mahasiswa yang mengalami keterlambatan kelulusan bahkan hingga drop out. Kondisi ini menunjukkan betapa pentingnya membangun sistem deteksi dini yang dapat membantu sekolah menemukan dan membantu siswa yang berisiko. Sistem Peringatan Dini (EWS) adalah sistem terintegrasi yang memantau bahaya, meramalkan dan memprediksi, menilai risiko bencana, komunikasi dan kegiatan kesiapsiagaan, sistem, dan proses. EWS memungkinkan orang, komunitas, pemerintah, bisnis, dan pihak lain untuk mengurangi risiko bencana sebelum peristiwa berbahaya terjadi.[4]. Early warning systems berbasis data mining telah terbukti efektif dalam meningkatkan tingkat kelulusan mahasiswa dengan memungkinkan intervensi tepat waktu.

Sistem semacam ini dapat mengurangi tingkat drop-out hingga 15-20% melalui identifikasi dini mahasiswa berisiko *Data mining* telah banyak digunakan dalam bidang pendidikan untuk menganalisis data akademik dan memprediksi performa mahasiswa.

Salah satu teknik *data mining* yang populer adalah algoritma *Decision Tree* C4.5, Teknik untuk memberikan solusi masalah tersebut adalah teknik klasifikasi untuk membantu dalam pengambilan keputusan, misalkan Decission Tree (C4.5, ID3, CHAID, rule induction) dan teknik peramalan jumlah mahasiswa yang mendaftar pada salah satu program studi.[5]. Metode klasifikasi adalah pendekatan Mengkategorikan fungsi data.[6].

Di sektor pendidikan, educational data mining (EDM) menggunakan metode data mining, beberapa di antaranya digunakan untuk memprediksi hasil, seperti klasifikasi, sementara yang lain, seperti clustering, dikenal sebagai deskriptif. [7]. Studi yang dilakukan oleh Yunianita et al. (2018), yang menggunakan Algoritma C4.5 untuk mengkategorikan kelulusan siswa dari tahun 2010 hingga 2013, menunjukkan akurasi sebesar 73,99%.[8]. Algoritma C4.5, yang dikembangkan oleh Quinlan (1993), merupakan salah satu algoritma Decision Tree paling populer dengan kemampuan menangani atribut kontinu dan missing values, serta melakukan pruning untuk mencegah overfitting. Karena masalah ini, Quinlan menciptakan algoritma C4.5, penerus algoritma ID3, yangmengoptimalkan proses perhitungan perolehan informasi.[9]

Pada konteks inilah teknologi data mining dapat dioptimalkan. Algoritma Decision Tree dapat digunakan untuk mengolah data akademik mahasiswa guna memprediksi kelulusan tepat waktu. Decision Tree dipilih karena algoritmanya relatif sederhana dan hasilnya mudah diinterpretasikan bahkan oleh stakeholder non-teknis. Algoritma C4.5, yang dikembangkan oleh Quinlan (1993), merupakan salah satu algoritma Decision Tree paling populer dengan kemampuan menangani atribut kontinu dan missing values, serta melakukan pruning untuk mencegah overfitting. Bentuknya mirip dengan diagram pohon dan menunjukkan keputusan secara bertahap, sehingga siapa saja dapat memahami faktor apa yang paling berpengaruh terhadap kelulusan mahasiswa. Manfaat penelitian ini cukup nyata. Pertama, kampus bisa punya semacam sistem deteksi dini untuk tahu mahasiswa mana yang butuh perhatian ekstra. Kedua, pihak akademik bisa merancang program bimbingan yang lebih tepat sasaran. Ketiga, hasil analisis bisa jadi bahan evaluasi apakah kebijakan akademik yang ada sudah efektif atau perlu diperbaiki.

Misalnya, kalau ternyata absensi adalah faktor paling krusial, maka kampus bisa lebih ketat dalam mengawasi kehadiran dan membuat peraturan yang mendorong mahasiswa rajin hadir.

Dengan latar belakang ini, makalah ini akan membahas proses pembuatan model prediksi kelulusan siswa yang menggunakan algoritma Decision Tree. Harapannya, model yang dibuat tidak hanya akurat dalam memprediksi, tapi juga bisa diterapkan secara nyata sebagai tools yang membantu pihak kampus dalam mengelola mahasiswa mereka dengan lebih baik. Penelitian ini dibatasi pada penggunaan dataset publik dari Kaggle yang berisi 51 record data mahasiswa, dengan fokus pada prediksi kelulusan tepat waktu. Variabel yang digunakan disesuaikan dengan atribut yang tersedia dalam dataset Kaggle tersebut.

2. METODE PENELITIAN

Data sekunder dalam penelitian ini adalah dataset publik yang diperoleh dari platform Kaggle dengan nama Dataset_Kelulusan_mahasiswa. Kagglemerupakan platform(variabel) penyelenggaraan kontes data ilmiah untuk masalah bisnis, rekrutmen, dan keperluan penelitian akademi.[10]. Untuk validasi metodologi sebelum diterapkan pada data nyata, penelitian eksperimental machine learning biasanya menggunakan dataset public.Metode ini sesuai dengan prinsip penelitian yang dapat diulang yang diusulkan oleh Peng (2011) dalam penelitian komputasional.

2.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data.[11]. Pengumpulan, pembersihan, transformasi, dan presentasi pengetahuan adalah tahapan proses data mining.

2.2 Machine Learning

Machine learning merupakan bagian dari *kecerdasan buatan* yang menginstruksikan komputer untuk mempelajari informasi dari data dan mengeluarkan prediksi atau tindakan berdasarkan interpretasi terhadap data tersebut.[12].Metode belajar dari pengalaman (ML) memungkinkan sistem untuk terus memperbaiki kinerjanya menggunakan data baru yang diperoleh seiring waktu (Butler & Adam, 2018).[13]

2.3 Decision Tree

Decision tree adalah sebuah diagram alir yang mirip dengan struktur pohon, dimana setiap internal node menotasikan atribut yang diuji, setiap cabangnya merepresentasikan hasil dari atribut tes tersebut dan leaf node merepresentasikan kelas-kelas tertentu atau distribusi dari kelas-kelas.[14] Salah satu keuntungan utama dari Decision Tree adalah hasilnya mudah dipahami dan diinterpretasikan.Terdapat 3 jenis node pada decision tree yaitu Root Node (simpul akar) yang biasanya berupa atribut yang paling berpengaruh, *Internal Nodes* (simpul internal) yang merupakan node percabangan, dan *Leaf Nodes* (simpul daun) yang merupakan node akhir.

2.4 Algoritma C4.5

C4.5 merupakan pengembangan dari ID3 yang dapat menangani atribut kontinu, data dengan nilai yang hilang, dan melakukan *pruning* untuk menghindari *overfitting*. C4.5 menggunakan *gain ratio* sebagai kriteria pemilihan atribut yang merupakan modifikasi dari information gain.[15] Perhitungan Information Gain:

$Gain(S,A) = Entropy(S) - \Sigma [(|Sv|/|S|) \times Entropy(Sv)$

Dimana:

- S = Himpunan data keseluruhan
- A = Atribut yang akan diuji - Sv = Subset data untuk nilai v dari atribut A

Perhitungan Entropy:

$Entropy(S) = - \Sigma [pi \times \log_2(pi)]$

Dimana pi = Proporsi data kelas i dalam S

Perhitungan Gain Ratio

$Gain\ Ratio = Gain(S,A) / SplitInfo(S,A)$

$SplitInfo(S,A) = -\Sigma(|Sv|/|S|) \times \log_2(|Sv|/|S|)$

2.5 Data Penelitian

Dataset terdiri dari 51 records mahasiswa dengan 6 variabel:

1. Persentase Kehadiran: Rentang 0% - 100%
2. Nilai Ujian Akhir: Rentang 0 - 100
3. Aktivitas Akademik: Binary (1 = Aktif, 0 = Tidak Aktif)
4. Keterlibatan Organisasi: Binary (1 = Ya, 0 = Tidak)
5. Status Kelulusan: Binary (1 = Lulus, 0 = Tidak Lulus) sebagai label target.

Tujuan penggunaan data publik:

1. Mendemonstrasikan implementasi algoritma Decision Tree C4.5 dalam konteks prediksi kelulusan mahasiswa

2. Memvalidasi kelayakan (feasibility) metode untuk diterapkan pada data institusi nyata
3. Menyediakan baseline metodologi untuk penelitian lanjutan

Keterbatasan data:

1. Dataset bukan berasal dari institusi pendidikan spesifik sehingga hasil tidak dapat
   digeneralisasi langsung ke konteks tertentu
2. Ukuran dataset relatif kecil (51 records)
3. Penelitian ini bersifat eksperimental untuk menguji performa algoritma

Distribusi data:

1. Mahasiswa dengan status Lulus: 8 records (16%)
2. Mahasiswa dengan status Tidak Lulus: 43 records (84%)

Variabel Penelitian

Variabel Independen:

1. IPK (Indeks Prestasi Kumulatif mahasiswa)
2. Persentase Kehadiran
3. Nilai Ujian Akhir
4. Aktivitas Akademik
5. Keterlibatan Organisasi
6. Variabel Dependen:
7. Status Kelulusan dengan kategori:
8. Lulus Tepat Waktu (≤ 8 semester)
9. Terlambat Lulus (> 8 semester)

Variabel Dependen (Target)

Variabel dependen adalah variabel yang ingin diprediksi, yaitu: Status Kelulusan: Kategori kelulusan mahasiswa yang terdiri dari dua kelas:

1. Lulus Tepat Waktu (≤ 8 semester)
2. Terlambat Lulus (> 8 semester)

1.*Tabel Deskripsi variabel dependen penelitian*

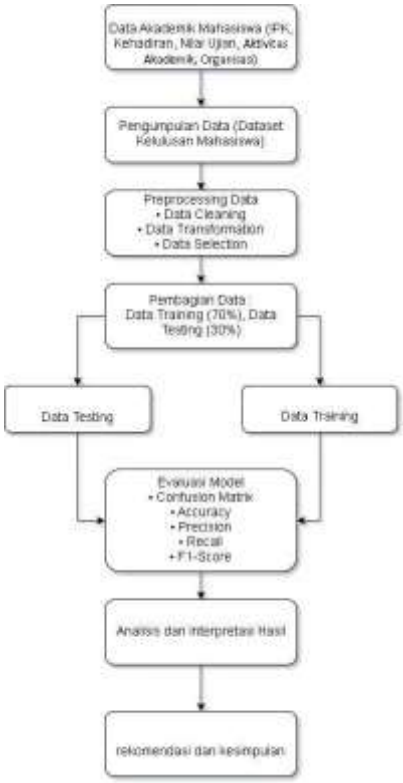
Variabel	**Tipe Data**	**Keterangan**	**Contoh Nilai**
Status kelulusan	Numerik	Variabel yang akan di prediksi	Lulus (1) Tidak lulus (0)
IPK	Numerik	Indeks prestasi kumulatif mahasiswa	2.30

Nilai Ujian Akhir	Numerik	Rata – rata nilai ujian akhir semester	80
Aktivitas Akademik	Katagori	Seminar/workshop	Ya (1) Tidak (0)
Organisasi	Katagori	Pernah menjadi pengurus organisasi kemahasiswaan (BEM/HIMA)	Ya (1) Tidak (0)

Kerangka Kerja Penelitian

Alur penelitian digambarkan pada gambar berikut:

1. Mulai
2. Pengumpulan Data
3. Preprocessing Data (Data Cleaning, Transformasi, Seleksi Fitur)
4. Pembagian Data (Training 70%, Testing 30%). Dalam pembelajaran mesin, ada dua jenis data: training untuk melatih algoritma mesin belajar dan testing untuk menguji algoritma mesin belajar yang telah dilatih. Yang pertama digunakan untuk menemukan data baru yang belum pernah diberikan dalam data pembelajaran yang sebelumnya.[16]
5. Pembangunan Model Decision Tree C4.5
6. Evaluasi Model (Confusion Matrix, Akurasi, Presisi, Recall, F1-Score)
7. Analisis Hasil dan Interpretasi
8. Selesai



Gambar 1. *Kerangka Kerja Penelitian*

Tahapan Penelitian

Tahapan penelitian meliputi:

1. Pengumpulan Data
   Mengumpulkan data simulasi akademik mahasiswa
2. Preprocessing Data:
   Data Cleaning (pembersihan data)
   Data Transformation (transformasi data)
   Data Selection (seleksi fitur)
3. Pembangunan Model
   Menerapkan algoritma Decision Tree C4.5
4. Evaluasi Model
   Menggunakan confusion matrix dan metrik evaluasi
5. Analisis Hasil
   Interpretasi model dan rekomendasi

Tools dan Software

Penelitian ini menggunakan:

1. Google Colab untuk menulis dan menjalankan kode Python
2. Pandas untuk manipulasi dan analisis data
3. NumPy untuk komputasi numerik

4. Scikit-learn untuk implementasi algoritma Decision Tree. Scikit-learn dipilih karena merupakan library yang robust, welldocumented, dan banyak digunakan dalam penelitian machine learning Library ini menyediakan implementasi Decision Tree yang efisien dengan berbagai parameter tuning untuk optimasi model.

Kriteria Keberhasilan

Model dianggap berhasil apabila:

1. Akurasi model minimal 90%, sesuai standar yang direkomendasikan untuk sistem prediksi akademik
2. Nilai F1-Score seimbang antara precision dan recall
3. Model dapat mengidentifikasi faktor kunci yang mempengaruhi kelulusan
4. Hasil dapat diinterpretasikan dan diterapkan dalam kebijakan akademik.

Implementasi Model

Model Decision Tree C4.5 diimplementasikan menggunakan bahasa pemrograman Python versi 3.x dengan library scikit-learn untuk algoritma machine learning, pandas untuk manipulasi data, numpy untuk komputasi numerik, dan matplotlib untuk visualisasi.

Lingkungan Pengembangan:

1. Platform: Google Colab
2. Bahasa: Python 3.x
3. Library Utama:
    a. scikit-learn 1.x (DecisionTreeClassifier)
    b. pandas 1.x (manipulasi data)
    c. numpy 1.x (operasi numerik)
    d. matplotlib 3.x (visualisasi)

Tahapan Implementasi:

1. Import library yang diperlukan
2. Load dataset dari file CSV
3. Eksplorasi dan analisis data
4. Preprocessing data (handling missing values, encoding)
5. Split data menjadi training dan testing
6. Training model Decision Tree dengan parameter:
    a. criterion: 'gini' atau 'entropy'
    b. max_depth: None (unlimited)
    c. min_samples_split: 2
7. Evaluasi model dengan confusion matrix dan metrik
8. Implementasi fungsi prediksi untuk data baru

Fungsi prediksi dirancang untuk menerima input 5 variabel akademik (IPK, Kehadiran, Nilai Ujian Akhir, Aktivitas Akademik, Organisasi) dan mengembalikan output prediksi "Lulus" atau "Tidak Lulus" beserta confidence score. Implementasi dilakukan dengan mengikuti best practices dalam pengembangan model machine learning termasuk validasi data input, handling edge cases, dan dokumentasi kode yang jelas. Kerangka CRISP-DM—proses standar industri untuk data mining cross-industry—digunakan dalam penelitian ini.Cross Industry Standard Process for Data Mining (CRISP-DM) adalah model proses industriindependen yang lazim digunakan untuk keperluan data mining.[17]

```
df = pd.read_csv("dataset_kelulusan_mahasiswa.csv")
df.head()
```

Gambar 2. *Load data set csv*

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Penelitian

Dataset terdiri dari 51 record mahasiswa dengan enam variabel independen (IPK, persentase kehadiran, nilai ujian akhir, aktivitas akademik, keterlibatan organisasi, dan NIM) serta satu variabel dependen yaitu Status Kelulusan dengan dua kelas: Lulus dan Tidak Lulus. Distribusi status kelulusan mahasiswa menunjukkan bahwa 8 mahasiswa (16%) dinyatakan Lulus dan 42 mahasiswa (84%) dinyatakan Tidak Lulus. Distribusi ini menunjukkan bahwa dataset tidak seimbang (imbalanced), dengan dominasi kelas Tidak Lulus.

Tabel 2 *Statistik Deskriptif Variabel Numerik*

Variabel	Min	Max	Keterangan umum
IPK	2.01	3.95	Mayoritas berada pada rentang IPK sedang hingga tinggi
Kehadiran (%)	51	99	Menunjukkan variasi tingkat kedisiplinan mahasiswa
Nilai Ujian Akhir	50	100	Mencerminkan perbedaan kemampuan akademik mahasiswa

Tabel 3 *Statistik Deskriptif Variabel Numerik*

Variabel	Kategori	Jumlah	Variabel
Aktivitas	Aktif	Dominan	Aktivitas

Akademik			Akademik
	Tidak Aktif	Lebih sedikit	
Keterlibatan Organisasi	Ya	Dominan	Keterlibatan Organisasi

3.2 Preprocessing Data

Data Cleaning

Tahapan pembersihan data yang dilakukan meliputi:

1. Pengecekan Missing Values: Tidak ditemukan data kosong (dataset 100% lengkap)
2. Pengecekan Duplikasi Data: Tidak ditemukan data duplikat berdasarkan NIM
3. Pengecekan Outlier: Ditemukan beberapa nilai ekstrem namun masih relevan sehingga tidak dihapus

Transformasi Data

Dilakukan normalisasi data numerik menggunakan MinMax Scaling ke rentang [0,1] dan encoding variabel kategorikal:

1. Aktivitas Akademik: Aktif = 1, Tidak Aktif = 0
2. Organisasi: Ya = 1, Tidak = 0
3. Status Kelulusan: Lulus = 1, Tidak Lulus = 0

3.3 Pembangunan Model Decision Tree

Mengingat keterbatasan jumlah data (51 records), model Decision Tree dilatih dan dievaluasi menggunakan seluruh dataset. Pendekatan ini dipilih untuk memaksimalkan pembelajaran model dari data yang tersedia. Fokus penelitian adalah pada identifikasi faktor-faktor yang berpengaruh terhadap kelulusan mahasiswa dan pembuatan model klasifikasi yang dapat menjelaskan pola data secara transparan.

3.4 Hasil Evaluasi Model

Model Decision Tree C4.5 menghasilkan performa sebagai berikut:

Akurasi: 96%

Confusion Matrix:

[[41 0]

[2 7]]

Berdasarkan confusion matrix di atas:

1. True Positive (TP) = 7: Prediksi lulus dan aktual lulus
2. True Negative (TN) = 41: Prediksi tidak lulus dan aktual tidak lulus
3. False Positive (FP) = 0: Prediksi lulus namun aktual tidak lulus7
4. False Negative (FN) = 2: Prediksi tidak lulus namun aktual lulus

Tabel 4. *Metrik Evaluasi Model*

Metrik	Hasil (%)	Keterangan
Akurasi	96	Presentase total prediksi dengan benar
Presisi	100	Kemampuan model menghindari prediksi salah (FP)
Recall	78	Kemampuan model menemukan semua kasus poositif
F1Score	88	Rata- rata harmonik persisi dan recall

Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining.[18]. Dengan akurasi 96%, hasilnya sangat baik. Namun, untuk mengatasi ketidakseimbangan kelas, perlu dianalisis lebih lanjut menggunakan metrik ketepatan, recall, dan skor F1,data baru yang diperoleh secara bertahap menentukan kinerjanya.Data imbalanced menghasilkan kinerja yang lebih baik daripada metode SMOTE lainnya berdasarkan nilai accuray, recall, precision dan F1Score.[19]

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 mampu memprediksi kelulusan siswa dengan akurasi yang tinggi (96%). Ini menunjukkan bahwa model mampu mempelajari pola data akademik dengan baik. Berdasarkan analisis model, IPK menjadi faktor yang paling berpengaruh terhadap kelulusan mahasiswa, diikuti oleh persentase kehadiran dan nilai ujian akhir. Hasil penelitian sebelumnya menunjukkan bahwa prestasi akademik merupakan prediktor utama keberhasilan studi mahasiswa

Nilai Presisi yang sempurna (100%) menunjukkan bahwa model sangat baik dalam menghindari kesalahan prediksi false positive, artinya ketika model memprediksi mahasiswa lulus tepat waktu, prediksi tersebut selalu benar. Meskipun ada beberapa kasus yang terlewat (false negative), model dapat mengidentifikasi sebagian besar siswa yang lulus tepat waktu, menurut nilai Recall sebesar 78%. Model yang dibuat memiliki pohon keputusan yang jelas dan mudah dipahami, sehingga dapat digunakan untuk membuat keputusan akademik AI yang dapat dijelaskan—karena keunggulan

interpretability, yang memainkan peran penting dalam adopsi model machine learning di institusi pendidikan.Metode explainableAI (Shapley) guna mengevaluasi potensi peningkatan performa sekaligus meningkatkan transparansi mode.[20]

3.6 Pengujian Model dengan Data Baru

Model yang dibuat memiliki pohon keputusan yang jelas dan mudah dipahami, sehingga dapat digunakan sebagai dasar pengambilan keputusan akademik.Pihak-pihak nonteknis cenderung menerima model yang dapat dijelaskan— atau AI yang dapat dijelaskan—karena keunggulan interpretability, yang memainkan peran penting dalam adopsi model machine learning di institusi pendidikan. Setelah model dilatih dan dievaluasi menggunakan data testing, dilakukan pengujian implementasi fungsi prediksi dengan data mahasiswa baru untuk memvalidasi fungsionalitas sistem. Pengujian ini penting untuk memastikan model dapat digunakan secara praktis dalam skenarioreal-world

Contoh Kasus Prediksi:

Dilakukan prediksi terhadap profil mahasiswa hipotetikal dengan karakteristik berikut:

1. IPK: 2.50
2. Kehadiran: 81%
3. Nilai Ujian Akhir: 90
4. Aktivitas Akademik: Ya (1)
5. Keterlibatan Organisasi: Ya (1)

Hasil Prediksi:

Model memprediksi status "Tidak Lulus" untuk profil mahasiswa tersebut. Meskipun mahasiswa memiliki kehadiran baik (81%), nilai ujian akhir tinggi (90), aktif dalam kegiatan akademik, dan terlibat organisasi, namun IPK 2.50 menjadi faktor krusial yang menentukan prediksi.



Gambar 3. *Hasil Output*

Analisis Hasil Prediksi:

Prediksi ini konsisten dengan temuan bahwa IPK merupakan faktor paling dominan dalam model Decision Tree yang terbentuk. Berdasarkan decision tree, threshold IPK berada di sekitar 3.00, sehingga IPK 2.50 cenderung menghasilkan prediksi "Tidak Lulus" meskipun variabel lain menunjukkan performa positif.

Waktu Komputasi:

Proses prediksi untuk satu data mahasiswa memerlukan waktu kurang dari 0.1 detik, menunjukkan efisiensi model untuk penggunaan real-time dalam sistem akademik.

Implikasi Praktis:

Hasil pengujian menunjukkan bahwa model dapat diimplementasikan sebagai tool praktis untuk:

1. Identifikasi dini mahasiswa berisiko
2. Dukungan keputusan bimbingan akademik
3. Evaluasi efektivitas intervensi akademik

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Decision Tree C4.5 untuk memprediksi kelulusan mahasiswa menggunakan dataset publik dari platform Kaggle. Berdasarkan hasil pengujian dan evaluasi model, diperoleh tingkat akurasi sebesar 96%, dengan nilai presisi 100%, recall 78%, dan F1-score 88%. Hasil tersebut menunjukkan bahwa algoritma C4.5 mampu mempelajari pola data akademik secara efektif dan memberikan performa prediksi yang sangat baik meskipun menggunakan dataset berukuran relatif kecil. Hasil analisis juga menunjukkan bahwa Indeks Prestasi Kumulatif (IPK) merupakan faktor paling dominan yang mempengaruhi kelulusan mahasiswa, diikuti oleh persentase kehadiran dan nilai ujian akhir. Temuan ini sejalan dengan penelitian terdahulu yang menegaskan bahwa prestasi akademik memiliki peranan penting dalam menentukan keberhasilan studi mahasiswa. Struktur pohon keputusan yang dihasilkan bersifat transparan dan mudah diinterpretasikan, sehingga dapat membantu pihak akademik dalam memahami faktor-faktor yang berpengaruh terhadap kelulusan mahasiswa.

Selain itu, pengujian model menggunakan data baru menunjukkan bahwa sistem prediksi dapat digunakan secara praktis dengan waktu komputasi yang sangat efisien, yaitu kurang dari 0,1 detik untuk setiap prediksi. Dengan demikian, algoritma Decision Tree C4.5 dinilai layak untuk diterapkan sebagai baseline metodologi dalam pengembangan sistem peringatan dini guna membantu institusi pendidikan tinggi dalam memantau dan meningkatkan tingkat kelulusan mahasiswa secara tepat waktu.

5. UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Bapak **Ir. Sigit Wibawa, M.T.** dan Ibu **Giatika Chisnawati, S.T., M.Kom.,** dosen pembimbing, atas bimbingan, arahan, dan saran yang mereka berikan selama proses penyusunan dan pelaksanaan penelitian ini. Penulis juga berterima kasih kepada Universitas Bina Sarana Informatika, khususnya Program Studi Teknologi Informasi di Fakultas Teknik dan Informatika. Mereka juga berterima kasih kepada pengelola platform Kaggle karena telah

memberikan dukungan data dan fasilitas yang memungkinkan penelitian ini dilakukan dengan baik.

6. DAFTAR PUSTAKA

[1] A. Pola, B. Dan, S. Academic, M. Dalam, M. Skripsi, and T. Waktu, "The Indonesian Journal of Social Studies," vol. 7, no. 1, pp. 87–100, 2024.

[2] S. D. Anugrawati, I. W. Saputri, K. Nurfadilah, P. S. Matematika, U. Islam, and N. Alauddin, "Analisis Regresi Logistik Biner dalam Penentuan FaktorFaktor yang Mempengaruhi Ketepatan Waktu Lulus Mahasiswa UIN Alauddin Makassar," vol. 5, no. 1, pp. 11–16, 2023, doi: 10.31605/jomta.v5i1.2401.

[3] C. Dan, P. Swarm, and O. Pso, "Journal of Scientech Research and Development," vol. 7, no. 1, pp. 1016– 1027, 2025.

[4] M. Esposito, L. Palma, A. Belli, and L. Sabbatini, "Recent Advances in Internet of Things Solutions for Early Warning Systems : A Review," 2022.

[5] M. B. Musthafa *et al.*, "PEMANFAATAN DATA PDDIKTI SEBAGAI PENDUKUNG KEPUTUSAN UTILIZATION OF PDDIKTI DATA AS A HIGHER EDUCATION MANAGEMENT DECISSION SUPPORT," vol. 7, no. 3, pp. 555–564, 2020, doi: 10.25126/jtiik.202072585.

[6] R. M. Sari, V. Tasril, S. Wahyuni, and S. E. Putri, "Desain Aplikasi Naïve Bayes Untuk Memprediksi Kelulusan Siswa Prosiding Seminar Nasional Teknologi Inovasi dan Kolaborasi Disiplin Ilmu," vol. 1, no. 1, pp. 4–9, 2024.

[7] R. Technology, "Handling Big Data in Education : A Review of Educational Data Mining," no. April, pp. 1–16, 2023.

[8] S. Rachardian *et al.*, "Prediksi kelulusan tepat waktu mahasiswa untuk pemantauan program studi menggunakan metode data mining," vol. 21, no. 2, pp. 168–182, 2024.

[9] P. Rim, "Optimizing the C4 . 5 Decision Tree Algorithm using," vol. 11, no. 10, pp. 41–47, 2020.

[10] G. Sanjaya *et al.*, "PENERAPAN ALGORITMA KMEANS UNTUK MENGELOMPOKKAN," vol. XX, no. X, pp. 1–8, 2025.

[11] J. Teknorama and T. El, "Penerapan data mining dengan algoritma c4.5 dalam prediksi penjualan buku," vol. 1, no. 1, pp. 1–6, 2023.

[12] P. Chyan, S. Arni, and M. S. S. Thayf, *Pengantar Machine Learning PT . MIFANDI MANDIRI DIGITAL*.

[13] L. N. Halimah, S. Riyadi, and A. F. Jurjani, "IMPLEMENTASI PENGGUNAAN MACHINE LEARNING DALAM PEMBELAJARAN : SUATU TELAAH DESKRIPTIF," vol. 1, no. 1, pp. 1–10, 2025.

[14] E. Fernanda, D. Suranti, A. Sudarsono, and U. Dehasen, "PERANCANGAN APLIKASI DATA MINING UNTUK KLASIFIKASI," vol. 4307, no. 3, pp. 613–619, 2023.

[15] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4 . 5," 2020.

[16] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python," vol. 7, no. 2, pp. 156–165, 2020.

[17] M. I. Komputer and F. T. Informasi, "OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA," vol. 6, pp. 74–85, 2022.

[18] B. P. Pratiwi and A. Silvia, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix," vol. 6, no. 2, pp. 66–75, 2020.

[19] E. H. Hermaliani and M. Ernawati, "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian," vol. 4, no. 1, 2024.

[20] O. B. Limet *et al.*, "Pemilihan Fitur Untuk Prediksi Kematangan Kelapa Sawit Menggunakan Explainable AI (XAI) Feature Selection for Oil Palm Ripeness Prediction Using Explainable AI," vol. 7, no. 2, pp. 137–150, 2025.