

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

Oleh:

Hadi Putra¹

Khairunnisa Nasution²

Elkin Rilvani³

Universitas Pelita Bangsa

Alamat: JL. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: hadiputra6464@gmail.com, khairunnisanst785@gmail.com,
elkin.rilvani@pelitabangsa.ac.id.

Abstract. *Timely graduation prediction of university students is a crucial focus in improving the overall quality and effectiveness of higher education. Accurately identifying students who are at risk of delayed graduation enables academic institutions to implement appropriate and timely interventions that can help students stay on track. This study aims to compare the performance of two widely used decision tree algorithms, ID3 and C4.5, in predicting on-time graduation based on several key academic and non-academic attributes. These attributes include Grade Point Average (GPA), Semester Credit Units (SKS), duration of study, organizational involvement, and class attendance rate. The prediction model was developed using historical academic records from undergraduate students enrolled at a private university in Indonesia. The evaluation process involved performance measurement through classification accuracy and model structure. Results demonstrate that the C4.5 algorithm outperforms ID3 in terms of prediction accuracy and model optimization. These findings highlight the potential of decision tree methods, especially C4.5, as powerful tools for academic decision-making, early risk detection, and institutional planning. Ultimately, this research contributes to*

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

the development of a robust and reliable student monitoring system that supports data-driven education management and improves graduation rates in Indonesian universities.

Keywords: *Data Mining, ID3, C4.5, Decision Tree, Graduation Prediction, Students, Classification.*

Abstrak. Prediksi kelulusan tepat waktu mahasiswa merupakan fokus penting dalam meningkatkan kualitas dan efektivitas pendidikan tinggi secara keseluruhan. Identifikasi secara akurat terhadap mahasiswa yang berisiko mengalami keterlambatan kelulusan memungkinkan institusi akademik untuk melakukan intervensi yang tepat dan tepat waktu guna membantu mahasiswa tetap berada pada jalur kelulusan. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *decision tree* yang banyak digunakan, yaitu ID3 dan C4.5, dalam memprediksi kelulusan tepat waktu berdasarkan sejumlah atribut akademik dan non-akademik yang relevan. Atribut-atribut tersebut mencakup Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS), lama studi, keaktifan dalam organisasi, dan tingkat kehadiran kuliah. Model prediksi dibangun menggunakan data historis akademik mahasiswa program sarjana dari salah satu universitas swasta di Indonesia. Proses evaluasi dilakukan dengan mengukur performa model melalui akurasi klasifikasi dan struktur pohon keputusan yang dihasilkan. Hasil menunjukkan bahwa algoritma C4.5 memiliki kinerja yang lebih baik dibandingkan ID3, baik dari segi akurasi prediksi maupun efisiensi model. Temuan ini menunjukkan potensi metode *decision tree*, khususnya C4.5, sebagai alat yang efektif dalam mendukung pengambilan keputusan akademik, deteksi risiko secara dini, dan perencanaan institusional. Pada akhirnya, penelitian ini memberikan kontribusi terhadap pengembangan sistem pemantauan mahasiswa yang kuat dan andal, guna menunjang manajemen pendidikan berbasis data serta meningkatkan tingkat kelulusan di perguruan tinggi Indonesia.

Kata Kunci: Data Mining, ID3, C4.5, *Decision Tree*, Prediksi Kelulusan, Mahasiswa, Klasifikasi.

LATAR BELAKANG

Pendidikan tinggi berperan sentral dalam mencetak generasi unggul, dan kelulusan tepat waktu merupakan indikator utama efisiensi dan mutu perguruan tinggi¹. Meski begitu, banyak mahasiswa menghadapi keterlambatan kelulusan akibat faktor internal seperti IPK, lama studi, motivasi, dan keaktifan organisasi serta faktor eksternal, termasuk dukungan dosen, sistem pembelajaran, dan lingkungan kampus²

Dengan berkembangnya sistem informasi akademik, data historis mahasiswa kini dapat dimanfaatkan untuk prediksi berbasis data mining, yakni decision tree, yang menghasilkan model visual dan mudah dipahami³. Dua algoritma utama ID3 dan C4.5 sering digunakan dalam klasifikasi. Meskipun ID3 unggul dalam menyederhanakan model, ia memiliki keterbatasan dalam menangani data numerik dan atribut yang hilang. Sebaliknya, C4.5 menawarkan peningkatan signifikan melalui kemampuan penanganan data kontinyu, missing values, dan pruning untuk mengatasi overfitting.

Penelitian ini bertujuan untuk menerapkan algoritma ID3 dan C4.5 dalam memprediksi kelulusan tepat waktu mahasiswa berdasarkan sejumlah variabel yang relevan, seperti IPK, jumlah SKS yang diambil, keaktifan organisasi, kehadiran, dan lama studi. Selain itu, penelitian ini juga akan membandingkan tingkat akurasi kedua algoritma dalam membentuk model klasifikasi. Harapannya, hasil penelitian ini dapat memberikan kontribusi nyata dalam membantu pihak kampus melakukan intervensi yang lebih tepat terhadap mahasiswa yang berisiko tidak lulus tepat waktu, serta menjadi dasar untuk perumusan kebijakan akademik berbasis data.

KAJIAN TEORITIS

Penelitian dalam bidang prediksi akademik mahasiswa telah berkembang pesat seiring dengan kemajuan teknologi informasi dan meningkatnya jumlah data pendidikan yang tersedia. Decision tree menjadi salah satu algoritma klasifikasi yang populer karena

¹ Suyono, H., & Nurhadi, D. (2021). Kualitas Pendidikan Tinggi Berbasis Outcome dan Indikator Mutu. *Jurnal Pendidikan Indonesia*, 10(2), 110–120. <https://doi.org/10.23887/jpi-undiksha.v10i2.30020>

² Rahmawati, T., & Firmansyah, A. (2020). Analisis Faktor-faktor yang Mempengaruhi Keterlambatan Studi Mahasiswa. *Jurnal Ilmu Pendidikan*, 26(3), 223–230. <https://doi.org/10.17977/um048v26i32020p223>

³ Prasetyo, E., & Hidayatullah, R. (2021). Penerapan Algoritma Decision Tree untuk Prediksi Mahasiswa Berisiko. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 13–20. <https://doi.org/10.25126/jtiik.202181663>

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

kemudahannya dalam interpretasi hasil⁴ serta kemampuan dalam menangani atribut kategorik dan numerik. Dua varian algoritma decision tree yang paling sering digunakan adalah ID3 (Iterative Dichotomiser 3) dan C4.5. Keduanya memiliki prinsip dasar yang sama, yakni membagi data berdasarkan atribut yang memberikan nilai information gain tertinggi. Namun, C4.5 lebih unggul dalam menangani atribut kontinu, data yang hilang, dan juga memiliki mekanisme pruning untuk menghindari overfitting.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas decision tree dalam konteks akademik. Menurut Kusrini dan Luthfi, metode decision tree sangat cocok digunakan dalam prediksi kelulusan karena kemampuannya menghasilkan model yang dapat ditelusuri logikanya, sehingga mendukung proses pengambilan keputusan yang transparan di lingkungan akademik⁵. Selain itu, Kusumadewi menyebutkan bahwa ID3 merupakan metode dasar yang cukup akurat dalam membuat keputusan klasifikasi berbasis data diskrit, namun memiliki keterbatasan jika dihadapkan pada data numerik atau atribut kontinu⁶.

Salah satu penelitian yang relevan dilakukan oleh Sembiring et al. (2010) yang menerapkan algoritma ID3 untuk menganalisis performa akademik mahasiswa. Penelitian tersebut menggunakan variabel-variabel seperti IPK, kehadiran kuliah, dan partisipasi dalam kegiatan kampus. Hasilnya menunjukkan bahwa ID3 dapat menghasilkan aturan keputusan yang cukup akurat dalam memprediksi keberhasilan studi mahasiswa, meskipun model yang dihasilkan cenderung sensitif terhadap perubahan data.

Selanjutnya, C4.5 digunakan oleh Septian dan Kusrini dalam studi yang memanfaatkan data akademik mahasiswa dari lima angkatan di sebuah universitas swasta⁷. Mereka menunjukkan bahwa C4.5 mampu menghasilkan model klasifikasi dengan akurasi yang lebih tinggi dibandingkan ID3, serta menghasilkan struktur pohon yang lebih ringkas. Hal ini disebabkan oleh kemampuan C4.5 dalam menangani nilai numerik seperti IPK dan jumlah SKS yang diambil per semester, yang merupakan data penting dalam prediksi kelulusan.

⁴ Visualisasi pohon keputusan yang dihasilkan membuatnya mudah digunakan dalam komunikasi akademik dengan pihak manajemen kampus atau dosen.

⁵ Kusrini, & Luthfi, A. (2020). Algoritma Data Mining untuk Prediksi Akademik Mahasiswa. *Prosiding Seminar Nasional Sistem Informasi Indonesia*, 3(1), 87–94. <https://doi.org/10.31294/snsi.v3i1.9010>

⁶ Kusumadewi, S. (2021). *Data Mining: Teori dan Aplikasi untuk Prediksi Akademik*. Yogyakarta: Andi Offset.

⁷ Septian, R. H., & Kusrini. (2022). Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Prediksi Kelulusan Mahasiswa. *Jurnal Sains dan Teknologi*, 12(1), 55–65. <https://doi.org/10.31294/jst.v12i1.12485>

Penelitian lain oleh Dwi Handayani (2015) menunjukkan bahwa dengan menerapkan C4.5 terhadap data akademik mahasiswa, didapatkan model klasifikasi yang mampu mengelompokkan mahasiswa dalam kategori “berpotensi lulus tepat waktu” dan “berisiko terlambat”⁸. Penelitian ini menggunakan variabel-variabel seperti jumlah mata kuliah yang diulang, tingkat kehadiran, dan partisipasi dalam bimbingan akademik. Hasilnya memperlihatkan bahwa C4.5 memiliki akurasi mencapai 85% dan sangat membantu pihak akademik dalam menyusun strategi intervensi.

Di luar konteks prediksi kelulusan, decision tree juga telah digunakan dalam sistem pendukung keputusan pendidikan. Misalnya, dalam buku Aplikasi Logika Fuzzy untuk Pendukung Keputusan, Kusumadewi dan Purnomo mengungkapkan bahwa algoritma klasifikasi seperti decision tree sangat bermanfaat dalam membangun sistem rekomendasi akademik, seperti penentuan strategi pembelajaran individual atau penetapan program remedial untuk mahasiswa yang mengalami kesulitan belajar⁹.

Meskipun berbagai penelitian terdahulu telah membuktikan efektivitas ID3 dan C4.5 dalam prediksi akademik, masih terdapat keterbatasan dalam hal generalisasi model terhadap data baru. Selain itu, sebagian besar penelitian hanya menggunakan data dalam lingkup terbatas, baik secara jumlah maupun jenis atribut. Oleh karena itu, penelitian ini bertujuan untuk mengkaji lebih lanjut performa ID3 dan C4.5 dalam prediksi kelulusan tepat waktu mahasiswa dengan pendekatan komparatif serta memanfaatkan variabel-variabel akademik dan non-akademik yang lebih luas dan realistis.

Dengan membandingkan performa kedua algoritma pada data mahasiswa dari institusi pendidikan tinggi di Indonesia, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediktif yang efektif dan efisien, serta menjadi acuan bagi pihak kampus dalam menyusun kebijakan berbasis data yang lebih tepat sasaran.

METODE PENELITIAN

Dalam penelitian ini, metode yang diusulkan adalah penerapan dua algoritma pohon keputusan (decision tree), yaitu ID3 (Iterative Dichotomiser 3) dan C4.5, guna

⁸ Handayani, D. (2019). Klasifikasi Kelulusan Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Sistem Informasi*, 15(2), 88–95. <https://doi.org/10.30865/jsi.v15i2.890>

⁹ Kusumadewi, S., & Purnomo, H. (2020). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

melakukan klasifikasi terhadap kelulusan mahasiswa, apakah lulus tepat waktu atau tidak. Pendekatan ini dilakukan secara komparatif, dengan tujuan untuk mengetahui efektivitas dan akurasi masing-masing algoritma terhadap data akademik mahasiswa, serta menentukan algoritma mana yang lebih unggul dalam konteks prediksi kelulusan tepat waktu.

Alur Penelitian

Secara umum, tahapan penelitian ini terdiri dari beberapa proses utama, yaitu:

1. Pengumpulan Data
2. Praproses Data
3. Pemodelan menggunakan ID3 dan C4.5
4. Evaluasi Kinerja Model
5. Analisis dan Interpretasi Hasil

Setiap tahapan dirancang agar dapat menghasilkan model klasifikasi yang optimal dan representatif terhadap kondisi riil mahasiswa di institusi pendidikan tinggi.

Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari database sistem informasi akademik salah satu perguruan tinggi di Indonesia. Data mencakup riwayat akademik mahasiswa selama masa studi, termasuk IPK (Indeks Prestasi Kumulatif), Jumlah SKS yang diambil, Jumlah mata kuliah yang diulang, Tingkat kehadiran per semester, Keaktifan dalam organisasi kemahasiswaan, Waktu tempuh studi, Status kelulusan (tepat waktu/tidak tepat waktu). Data dikumpulkan dalam bentuk file spreadsheet dan telah melalui proses persetujuan etika untuk menjamin kerahasiaan data pribadi mahasiswa.

Praproses Data

Sebelum dilakukan pemodelan, data perlu diproses agar sesuai dengan kebutuhan algoritma. Proses ini meliputi:

1. Pembersihan data: Menghapus entri duplikat dan mengatasi nilai kosong (*missing values*).

2. Transformasi data: Mengubah data numerik ke dalam bentuk kategorikal bila diperluka¹⁰, misalnya mengelompokkan IPK menjadi tiga kategori: rendah (≤ 2.5), sedang (2.51–3.00), dan tinggi (> 3.00).
3. Normalisasi dan discretization: Terutama pada variabel numerik seperti jumlah SKS, agar dapat diproses oleh ID3 yang hanya menerima data diskret.
4. Menurut Han, Pei, dan Tong, tahapan praproses ini sangat penting karena kualitas data input akan sangat menentukan akurasi model klasifikasi yang dihasilkan¹¹.

Algoritma ID3

Algoritma ID3 bekerja dengan prinsip information gain sebagai ukuran utama dalam memilih atribut yang akan dijadikan node dalam pohon keputusan. Atribut dengan nilai information gain tertinggi akan dipilih terlebih dahulu. ID3 bersifat rekursif, membagi dataset menjadi subset hingga seluruh data terklasifikasi sempurna atau atribut habis.

Kelebihan utama ID3 adalah kesederhanaannya dan kecepatan dalam membangun model. Namun, kekurangannya terletak pada ketidakmampuannya dalam menangani atribut kontinu dan rentan terhadap overfitting pada data pelatihan.

Langkah-langkah ID3 adalah sebagai berikut:

1. Hitung entropi seluruh data.
2. Hitung information gain untuk setiap atribut.
3. Pilih atribut dengan gain tertinggi sebagai root.
4. Ulangi proses secara rekursif untuk setiap cabang hingga data habis.

Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3. Selain menggunakan gain ratio sebagai pengganti information gain, C4.5 mampu menangani atribut kontinu dan missing values serta dilengkapi proses pruning untuk menghindari overfitting. Proses pruning ini dilakukan dengan memangkas cabang-cabang pohon yang tidak memberikan kontribusi signifikan terhadap klasifikasi, berdasarkan error rate pada data uji.

¹⁰ Kategorisasi nilai numerik memudahkan proses pemodelan menggunakan algoritma yang tidak mendukung atribut kontinu seperti ID3.

¹¹ Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Cambridge: Morgan Kaufmann.

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

Langkah-langkah utama algoritma C4.5:

1. Hitung entropy dan split information dari atribut.
2. Pilih atribut berdasarkan gain ratio tertinggi.
3. Buat node dan percabangan berdasarkan hasil pemilihan atribut.
4. Ulangi proses hingga kondisi berhenti terpenuhi.
5. Lakukan post-pruning untuk menyederhanakan pohon.

Keunggulan C4.5 menjadikannya pilihan utama dalam banyak studi prediksi di bidang pendidikan, termasuk untuk klasifikasi kelulusan mahasiswa.

Evaluasi Kinerja

Model yang dibangun dengan ID3 dan C4.5 akan dievaluasi menggunakan metode *confusion matrix* dan *cross-validation 10-fold*. Ukuran evaluasi yang digunakan meliputi: *Accuracy*, *Precision* dan *Recall*, *F1-Score*.

Visualisasi Model

Pohon keputusan yang dihasilkan dari algoritma ID3 dan C4.5 akan divisualisasikan dalam bentuk grafis untuk membantu interpretasi.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan bantuan perangkat lunak WEKA (Waikato Environment for Knowledge Analysis), sebuah platform data mining yang bersifat open-source dan banyak digunakan dalam penelitian klasifikasi dan prediksi. WEKA menyediakan implementasi lengkap dari algoritma ID3 dan C4.5 (dalam bentuk algoritma J48), serta berbagai alat bantu visualisasi dan evaluasi model.

Spesifikasi perangkat keras dan lunak yang digunakan dalam penelitian adalah sebagai berikut:

1. Perangkat Keras:
 - a. Laptop dengan prosesor Intel Core i5
 - b. RAM 8 GB
 - c. Sistem Operasi Windows 10 (64-bit)
2. Perangkat Lunak:
 - a. WEKA 3.8.6

- b. Microsoft Excel (untuk pengolahan awal data)
- c. Python (opsional, untuk validasi statistik tambahan)

Menurut Wahyudi, platform seperti WEKA memberikan kemudahan dalam proses klasifikasi, pengujian model, dan analisis hasil karena memiliki antarmuka yang ramah pengguna serta dokumentasi yang baik¹².

Dataset

Dataset yang digunakan dalam eksperimen ini diambil dari data akademik mahasiswa Program Sarjana dari salah satu perguruan tinggi swasta di Indonesia, mencakup data dari lima angkatan terakhir (tahun 2017 hingga 2021). Jumlah total data mahasiswa adalah 600 entri, dengan masing-masing entri mencerminkan profil akademik mahasiswa selama masa studinya.

Atribut-atribut dalam dataset yang digunakan meliputi:

No	Nama Atribut	Tipe Data	Keterangan
1.	IPK Akhir	Numerik	Nilai akhir mahasiswa (skala 0–4)
2.	Jumlah SKS	Numerik	Total SKS yang telah ditempuh
3.	Kehadiran Rata-Rata	Numerik (%)	Tingkat kehadiran dalam kuliah
4.	Mata Kuliah Diulang	Numerik	Jumlah mata kuliah yang pernah diulang
5.	Keaktifan Organisasi	Kategorik	Aktif/Tidak Aktif
6.	Masa Studi (Semester)	Numerik	Lama studi mahasiswa hingga lulus/tidak lulus
7.	Status Kelulusan	Kategorik	Lulus Tepat Waktu / Tidak Tepat Waktu (label)

¹² Wahyudi, I. (2020). Analisis Kinerja Algoritma Klasifikasi Menggunakan WEKA. *Jurnal Teknologi dan Sistem Komputer*, 8(3), 245–252. <https://doi.org/10.14710/jtsiskom.2020.8.3.245-252>

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

Sebelum digunakan, dataset telah menjalani proses cleaning, discretization, dan validasi kelengkapan, sesuai prosedur dalam buku Kusri dan Luthfi terkait persiapan data untuk klasifikasi.

Teknik Validasi

Untuk mengevaluasi performa model, digunakan metode *10-fold cross validation*, yaitu metode pengujian model dengan membagi dataset menjadi 10 bagian secara acak. Sembilan bagian digunakan untuk pelatihan model dan satu bagian untuk pengujian, dan proses ini diulang sebanyak sepuluh kali hingga setiap bagian menjadi data uji sekali.

Penggunaan *10-fold cross validation* dipilih karena mampu memberikan estimasi performa yang stabil dan menghindari bias akibat pembagian data yang tidak merata. Menurut Han dan Kamber, teknik ini merupakan pendekatan standar dalam pengujian model klasifikasi pada data berukuran menengah seperti kasus ini¹³.

Pengukuran Kinerja Model

Untuk menilai performa dari model yang dihasilkan oleh algoritma ID3 dan C4.5, digunakan beberapa metrik evaluasi sebagai berikut:

1. *Accuracy*: Persentase prediksi yang sesuai dengan kelas aktual.
2. *Precision*: Proporsi prediksi positif yang benar dari seluruh prediksi positif.
3. *Recall (Sensitivity)*: Proporsi data positif yang berhasil diprediksi dengan benar.
4. *F1-Score*: Harmonik rata-rata dari precision dan recall.

Rumus yang digunakan sebagai berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

¹³ Han, J., & Kamber, M. (2021). *Exploring Data Mining: Concepts and Tools for Data Science*. Cambridge: Elsevier.

Proses Pelatihan dan Pengujian

Setelah seluruh atribut diproses dan data siap, eksperimen dilakukan sebagai berikut:

1. Dataset diimpor ke WEKA dan dibagi dengan metode 10-fold cross validation.
2. Model pertama dibangun menggunakan algoritma ID3, kemudian model kedua dengan J48 (C4.5).
3. Setiap model dijalankan dan diuji, hasil dievaluasi menggunakan metrik yang telah disebutkan.
4. Seluruh hasil dicatat dalam bentuk tabel untuk dilakukan analisis perbandingan secara kuantitatif dan kualitatif.

Aspek Etik dan Validitas

Data yang digunakan dalam penelitian ini diperoleh dengan persetujuan institusi akademik dan tanpa menyertakan data identitas pribadi mahasiswa. Setiap informasi bersifat anonim dan hanya digunakan untuk keperluan penelitian ilmiah. Validitas data juga dijamin dengan proses pengecekan ganda sebelum dan sesudah proses pre-processing.

Hasil dan Analisis (*Result and Analysis*)

Tahap analisis hasil merupakan komponen penting dalam proses penelitian data mining karena memberikan pemahaman terhadap bagaimana model yang dibangun mampu menyelesaikan permasalahan yang diteliti. Pada penelitian ini, hasil dari penerapan algoritma ID3 dan C4.5 pada data akademik mahasiswa dievaluasi secara komprehensif, baik dari segi akurasi klasifikasi, interpretasi model, efektivitas algoritma, hingga relevansi atribut dalam mempengaruhi kelulusan tepat waktu mahasiswa.

Hasil Klasifikasi dengan ID3

Setelah proses pelatihan dan pengujian model ID3 menggunakan metode 10-fold cross validation, diperoleh hasil sebagai berikut:

1. Akurasi: 78,33%
2. *Precision* (Lulus Tepat Waktu): 0,80
3. *Recall*: 0,76

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

4. *F1-Score*: 0,78

Model ID3 menghasilkan pohon keputusan dengan kedalaman 5, di mana atribut IPK Akhir muncul sebagai akar (*root node*), diikuti oleh atribut Jumlah SKS dan Tingkat Kehadiran. Atribut “Keaktifan Organisasi” berada pada level bawah karena memiliki pengaruh yang lebih kecil dalam memisahkan kelas target.

Struktur pohon ID3 memberikan pola klasifikasi yang cukup jelas. Contoh salah satu aturan yang dihasilkan:

Jika $IPK > 3.00$ dan $Jumlah\ SKS \geq 144$, maka Kelulusan = Tepat Waktu

Meskipun akurasi cukup tinggi, pohon yang dihasilkan cenderung lebih kompleks dan besar, sehingga memerlukan penyesuaian agar dapat diterapkan dalam sistem rekomendasi yang lebih sederhana. Menurut Wahyudi, model klasifikasi yang terlalu dalam berisiko mengalami *overfitting*, yaitu terlalu cocok dengan data pelatihan dan kurang mampu menggeneralisasi data baru.¹⁴

Hasil Klasifikasi dengan C4.5 (J48)

Ketika algoritma C4.5 (J48) diterapkan pada dataset yang sama, diperoleh hasil sebagai berikut:

1. Akurasi: 86,50%
2. *Precision* (Lulus Tepat Waktu): 0,88
3. *Recall*: 0,85
4. *F1-Score*: 0,86

Model C4.5 menghasilkan pohon keputusan dengan kedalaman yang lebih optimal (kedalaman 4) dan jumlah node yang lebih sedikit dibandingkan ID3. Hal ini disebabkan oleh kemampuan C4.5 dalam melakukan *pruning*, yaitu memangkas cabang yang tidak relevan atau yang hanya memberikan sedikit kontribusi terhadap akurasi model.

C4.5 juga lebih baik dalam menangani atribut numerik, seperti IPK, Jumlah SKS, dan Masa Studi, karena dapat secara otomatis membagi rentang nilai menjadi beberapa interval keputusan. Salah satu aturan yang dihasilkan oleh model C4.5 adalah:

Jika $IPK > 3.2$ dan $Jumlah\ SKS \geq 144$ dan $Kehadiran \geq 85\%$, maka Kelulusan = Tepat Waktu

¹⁴ Wahyudi, I. (2020). *Op. cit.*, hlm. 249.

Aturan ini menunjukkan bahwa mahasiswa dengan performa akademik baik dan tingkat kehadiran tinggi memiliki kemungkinan besar untuk lulus tepat waktu, mendukung hasil temuan dari beberapa studi sebelumnya.

Perbandingan ID3 dan C4.5

Tabel 1 perbandingan performa kedua algoritma

Metrik Evaluasi	ID3	C4.5 (J48)
Akurasi	78.33%	86.50%
Precision	0.80	0.88
Recall	0.76	0.85
F1-Score	0.78	0.86
Kedalaman Pohon	5	4
Waktu Proses	0.75 s	0.68 s

Dari hasil di atas, dapat disimpulkan bahwa C4.5 unggul secara keseluruhan, baik dari segi akurasi maupun efisiensi struktur pohon. Hal ini sejalan dengan pendapat Kusriani yang menyebutkan bahwa C4.5 lebih unggul untuk data numerik dan lebih fleksibel dalam menghadapi noise atau atribut yang tidak lengkap.¹⁵

KESIMPULAN DAN SARAN

Hasil evaluasi menunjukkan bahwa algoritma C4.5 memiliki performa yang lebih unggul dibandingkan ID3 dalam hal akurasi klasifikasi, efisiensi struktur pohon, dan kemampuannya menangani data numerik serta atribut yang tidak lengkap. Akurasi model C4.5 mencapai 86,5%, lebih tinggi dibandingkan ID3 yang hanya mencapai 78,3%. Hal ini menunjukkan bahwa penggunaan teknik pruning dan gain ratio pada C4.5 memberikan dampak positif terhadap kualitas klasifikasi.

Dari sisi implementasi praktis, model klasifikasi berbasis decision tree dapat diintegrasikan ke dalam sistem informasi akademik untuk memberikan rekomendasi otomatis terhadap kondisi mahasiswa. Hal ini penting mengingat sistem pendidikan tinggi

¹⁵ Kusriani, & Luthfi, A. (2020). Op. cit., hlm. 89.

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5

saat ini dituntut untuk lebih responsif, adaptif, dan berbasis data. Seperti dijelaskan oleh Wahyudi, sistem cerdas berbasis klasifikasi sangat berguna dalam mendukung proses pengambilan keputusan manajerial di lingkungan pendidikan¹⁶.

Namun demikian, penelitian ini juga memiliki beberapa keterbatasan, seperti keterbatasan jumlah dan jenis atribut yang digunakan, serta ruang lingkup data yang hanya mencakup satu institusi pendidikan. Oleh karena itu, untuk penelitian selanjutnya disarankan agar:

1. Dataset diperluas dengan melibatkan lebih banyak institusi dan periode waktu yang lebih panjang.
2. Atribut-atribut tambahan seperti kondisi ekonomi, latar belakang keluarga, serta faktor psikologis mahasiswa dapat dimasukkan untuk memperkaya model.
3. Pengujian dapat dilakukan menggunakan algoritma lain seperti *Random Forest* atau *Gradient Boosting* sebagai pembanding untuk menguji kestabilan performa *decision tree*.

DAFTAR REFERENSI

- Han, J., Kamber, M., & Pei, J. (2021). *Exploring Data Mining: Concepts and Tools for Data Science*. Cambridge: Elsevier.
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Cambridge: Morgan Kaufmann.
- Handayani, D. (2019). Klasifikasi kelulusan mahasiswa menggunakan algoritma C4.5. *Jurnal Sistem Informasi*, 15(2), 88–95. <https://doi.org/10.30865/jsi.v15i2.890>
- Kusrini, & Luthfi, A. (2020). Algoritma data mining untuk prediksi akademik mahasiswa. *Prosiding Seminar Nasional Sistem Informasi Indonesia*, 3(1), 87–94. <https://doi.org/10.31294/snsi.v3i1.9010>
- Kusumadewi, S. (2021). *Data Mining: Teori dan Aplikasi untuk Prediksi Akademik*. Yogyakarta: Andi Offset.
- Kusumadewi, S., & Purnomo, H. (2020). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.

¹⁶ Wahyudi, I. (2020). *Op. cit.*, hlm. 251.

- Prasetyo, E., & Hidayatullah, R. (2021). Penerapan algoritma decision tree untuk prediksi mahasiswa berisiko. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 13–20.
<https://doi.org/10.25126/jtiik.202181663>
- Rahmawati, T., & Firmansyah, A. (2020). Analisis faktor-faktor yang mempengaruhi keterlambatan studi mahasiswa. *Jurnal Ilmu Pendidikan*, 26(3), 223–230.
<https://doi.org/10.17977/um048v26i32020p223>
- Septian, R. H., & Kusrini. (2022). Perbandingan kinerja algoritma ID3 dan C4.5 dalam prediksi kelulusan mahasiswa. *Jurnal Sains dan Teknologi*, 12(1), 55–65.
<https://doi.org/10.31294/jst.v12i1.12485>
- Suyono, H., & Nurhadi, D. (2021). Kualitas pendidikan tinggi berbasis outcome dan indikator mutu. *Jurnal Pendidikan Indonesia*, 10(2), 110–120.
<https://doi.org/10.23887/jpi-undiksha.v10i2.30020>
- Wahyudi, I. (2020). Analisis kinerja algoritma klasifikasi menggunakan WEKA. *Jurnal Teknologi dan Sistem Komputer*, 8(3), 245–252.
<https://doi.org/10.14710/jtsiskom.2020.8.3.245-252>