Vol. 4 No. 2 – Tahun 2025

Indonesian Journal Computer Science
ISSN: 2829-3819

Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 dengan RapidMiner: Studi Kasus Data Akademik Perguruan Tinggi XYZ

Nurul Khasanah[1], Daniati Uki Eka Saputri[2], Taopik Hidayat[3], Faruq Aziz[4]
Universitas Nusa Mandiri[1,2,3,4]
nurul.nuk@nusamandiri.ac.id[1], daniati.due@nusamandiri.ac.id[2], taopik.toj@nusamandiri.ac.id[3],
faruq.fqs@nusamandiri.ac.id[4]

Diterima	Direvisi	Disetujui
(29-07-2025)	(09-09-2025)	(10-10-2025)

Abstrak - Ketepatan waktu kelulusan mahasiswa merupakan indikator penting dalam menilai kualitas dan efektivitas pendidikan tinggi. Keterlambatan kelulusan tidak hanya berdampak pada akreditasi program studi, tetapi juga pada efisiensi manajemen akademik dan kesiapan lulusan menghadapi dunia kerja. Penelitian ini bertujuan membangun model prediksi kelulusan mahasiswa menggunakan algoritma C4.5 berbasis pohon keputusan dengan dukungan perangkat lunak RapidMiner. Dataset yang digunakan terdiri atas 379 entri mahasiswa, yang mencakup atribut demografis (jenis kelamin, umur, status nikah), status mahasiswa, indeks prestasi semester (IPS 1–8), serta indeks prestasi kumulatif (IPK). Proses penelitian meliputi pengumpulan, pembersihan, transformasi data, pemodelan, dan evaluasi performa. Model diuji menggunakan pembagian data 70:30 serta validasi silang (*10-fold cross-validation*) untuk memastikan keandalan hasil. Hasil pengujian menunjukkan akurasi 97,81% dan nilai AUC 0,991, yang menegaskan kemampuan algoritma C4.5 dalam mengklasifikasikan status kelulusan secara tepat. Temuan ini menunjukkan peningkatan signifikan dibandingkan penelitian sebelumnya dengan algoritma Naïve Bayes (88,16%) dan K-NN (87,8%). Atribut yang paling berpengaruh adalah IPS3, IPS4, IPS5, status pekerjaan, dan umur mahasiswa. Penelitian ini berkontribusi pada pengembangan model prediksi yang tidak hanya akurat, tetapi juga mudah diinterpretasikan, sehingga dapat dimanfaatkan oleh institusi pendidikan dalam menyusun kebijakan intervensi akademik dini bagi mahasiswa berisiko terlambat lulus.

Kata Kunci : C4.5, prediksi kelulusan, data mining, pohon keputusan, RapidMiner

Abstract *- Timely student graduation is an important indicator in assessing the quality and effectiveness of higher education. Delayed graduation not only impacts study program accreditation, but also the efficiency of academic management and graduates' readiness to face the world of work. This study aims to build a student graduation prediction model using the decision tree-based C4.5 algorithm supported by RapidMiner software. The dataset used consists of 379 student, which includes demographic attributes (gender, age, marital status), student status, semester grade point average (IPS 1–8), and cumulative grade point average (GPA). The research process includes data collection, cleaning, transformation, modeling, and performance evaluation. The model was tested using a 70:30 data split and 10-fold cross-validation to ensure the reliability of the results. The test results showed an accuracy of 97.81% and an AUC value of 0.991, confirming the ability of the C4.5 algorithm to accurately classify graduation status. This finding shows a significant improvement compared to previous studies with the Naïve Bayes (88.16%) and K-NN (87.8%) algorithms. The most influential attributes were IPS3, IPS4, IPS5, employment status, and student age. This research contributes to the development of a predictive model that is not only accurate but also easy to interpret, so it can be used by educational institutions in developing early academic intervention policies for students at risk of late graduation.*

Keywords: C4.5, student graduation prediction, data mining, decision tree, RapidMiner

I. PENDAHULUAN

Pendidikan tinggi merupakan salah satu pilar penting dalam pembangunan suatu bangsa (Tamimi et al., 2024). Kualitas lulusan dari perguruan tinggi sangat memengaruhi daya saing sumber daya manusia di era globalisasi saat ini (Susiyati & Sihono, 2025). Salah satu tantangan utama yang dihadapi oleh institusi pendidikan tinggi adalah bagaimana memastikan mahasiswa dapat menyelesaikan studi tepat waktu dengan hasil yang optimal (Zulfa & Arifudin, 2025). Tidak sedikit mahasiswa yang mengalami keterlambatan kelulusan, bahkan hingga putus studi, yang disebabkan oleh berbagai faktor seperti rendahnya capaian akademik, kondisi sosial ekonomi, hingga kurangnya motivasi belajar (Kristen et al., 2024). Permasalahan terkait ketepatan waktu kelulusan menjadi perhatian penting bagi pengelola institusi pendidikan, karena berdampak pada akreditasi program studi, efisiensi pembelajaran, serta reputasi institusi secara keseluruhan. Oleh karena itu, diperlukan sistem yang dapat mendeteksi secara dini mahasiswa yang berpotensi tidak lulus tepat waktu. Salah satu solusi yang dapat digunakan untuk mengatasi permasalahan ini adalah pemanfaatan teknologi data mining untuk melakukan prediksi kelulusan berdasarkan data historis mahasiswa (Rachardian et al., 2024).

Data mining adalah proses menemukan pola atau informasi yang berguna dari kumpulan data besar (Syahriani, 2022). Dalam konteks pendidikan, data mining sering disebut sebagai educational data mining (EDM), yang bertujuan untuk menganalisis data akademik guna memperoleh wawasan baru yang dapat meningkatkan proses belajar mengajar. Salah satu algoritma yang banyak digunakan dalam proses klasifikasi pada data mining adalah C4.5, yaitu algoritma pohon keputusan (decision tree) yang dikembangkan oleh Ross Quinlan (Mantopani et al., 2024). C4.5 mampu menangani atribut numerik dan kategorikal, menangani data yang hilang, serta menghasilkan model pohon yang mudah ditafsirkan.

Berbagai penelitian sebelumnya telah menunjukkan efektivitas algoritma klasifikasi, khususnya C4.5, dalam berbagai domain penerapan, termasuk pendidikan dan bantuan sosial. Penelitian yang dilakukan oleh (Khasanah et al., 2022) menggunakan algoritma naïve bayes untuk memprediksi kelulusan mahasiswa berdasarkan data akademik, dan memperoleh akurasi sebesar 88,16%. Penelitian lain oleh (Ramadani et al., 2024) membandingkan kinerja algoritma C4.5 dengan Naïve Bayes dan K-Nearest Neighbor, dan hasilnya menunjukkan bahwa C4.5 memberikan akurasi tertinggi dalam klasifikasi kelayakan program keluarga harapan. Selain itu, (Putri et al., 2023) juga membandingkan algoritma K-NN, Naive Bayes and SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir Dimana hasilnya Algoritma K-NN (K-Nearest Neighbor) memiliki rata-rata lebih tinggi dibandingkan (NBC) Naïve Bayes Classifier dan SVM (Support Vector Machine) untuk prediksi kelulusan mahasiswa tingkat akhir dengan akurasi 87,8%, presisi 87,8%, dan recall 84%.

Dalam penelitian ini, digunakan data akademik sebanyak 379 mahasiswa, yang terdiri dari atribut: Nama, Jenis Kelamin, Status Mahasiswa, Umur, Status Nikah, IPS 1–8, IPK, dan Status Kelulusan. Data tersebut diolah menggunakan perangkat lunak RapidMiner, yang merupakan tools data mining berbasis GUI dan banyak digunakan dalam penelitian klasifikasi serta prediksi.

Tujuan dari penelitian ini adalah membangun model klasifikasi menggunakan algoritma C4.5 untuk memprediksi kelulusan mahasiswa berdasarkan atribut-atribut akademik dan demografis. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi pihak institusi pendidikan dalam mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu, sehingga dapat dilakukan intervensi dini untuk meningkatkan kualitas pendidikan dan manajemen akademik.

Beberapa penelitian terdahulu menggunakan algoritma Naïve Bayes, K-NN, maupun SVM untuk prediksi kelulusan, namun akurasi yang diperoleh masih berkisar 85–89%. Selain itu, sebagian penelitian lebih menekankan pada faktor akademik, tanpa mengkombinasikan variabel demografis. Oleh karena itu, penelitian ini menawarkan pendekatan berbeda dengan mengintegrasikan atribut akademik dan demografis menggunakan algoritma C4.5. Algoritma C4.5 dipilih karena mampu menangani data numerik maupun kategorikal, mengatasi missing value, serta menghasilkan model pohon keputusan yang mudah ditafsirkan oleh pihak pengambil kebijakan. Hal ini menjadikan C4.5 lebih unggul dalam konteks data akademik yang bervariasi dan membutuhkan interpretasi transparan.

II. METODOLOGI PENELITIAN

1. Pengumpulan Data

Tahap awal penelitian ini adalah pengumpulan data akademik mahasiswa yang

bersumber dari sistem informasi akademik pada salah satu perguruan tinggi. Dataset yang digunakan terdiri atas 379 entri, di mana setiap entri merepresentasikan satu individu mahasiswa dengan atribut yang meliputi: nama, jenis kelamin, status mahasiswa, umur, status pernikahan, indeks prestasi semester (IPS 1–8), indeks prestasi kumulatif (IPK), serta status kelulusan. Atribut *status kelulusan* (tepat waktu atau terlambat) ditetapkan sebagai variabel target (kelas), sedangkan atribut lainnya berfungsi sebagai variabel prediktor (fitur).

Walaupun jumlah entri relatif terbatas, dataset ini dapat dianggap representatif karena mencakup keseluruhan populasi mahasiswa pada Program Studi Informatika dalam periode tertentu, sehingga tidak bersifat sampel parsial. Dengan demikian, data yang dianalisis memiliki validitas internal yang tinggi. Selain itu, untuk mengurangi potensi bias akibat keterbatasan jumlah data, penelitian ini menggunakan metode validasi silang (*10-fold cross-validation*) yang secara luas diakui mampu meningkatkan reliabilitas serta stabilitas hasil pemodelan.

2. Pembersihan Data (Data Cleaning)

Data yang telah dikumpulkan tidak selalu dalam kondisi siap pakai. Oleh karena itu, dilakukan proses pembersihan data untuk memastikan integritas dan kualitas data sebelum dianalisis (Santoso & Priyadi, 2024). Beberapa langkah yang dilakukan pada tahap ini meliputi:

a. Penghapusan atribut irrelevan: Atribut “Nama” dihapus karena tidak memiliki nilai prediktif terhadap variabel target dan bersifat identitas pribadi (non-analitis).
b. Penanganan data kosong (missing values): Data yang memiliki nilai kosong (kosong sebagian pada atribut IPS atau IPK) diatasi dengan pendekatan imputasi rata-rata atau penghapusan entri jika jumlahnya sangat kecil dan tidak signifikan terhadap distribusi data.
c. Validasi tipe data: Dilakukan pengecekan terhadap kesesuaian tipe data, seperti memastikan bahwa atribut IPS dan IPK bernilai numerik, serta atribut kategorikal seperti “Jenis Kelamin” dan “Status Nikah” memiliki format yang konsisten.

3. Transformasi Data

Transformasi data dilakukan untuk mengubah data mentah menjadi format yang dapat diproses oleh algoritma C4.5 di RapidMiner (Bramanda, 2025). Tahapan ini meliputi:

a. Encoding variabel kategorikal: Atribut seperti “Jenis Kelamin”, “Status Mahasiswa”, dan “Status Nikah” diubah menjadi format nominal atau dikodekan (contoh: Laki-laki = 1, Perempuan = 0), agar dapat dikenali oleh algoritma.
b. Penyetaraan skala nilai (opsional): Meskipun C4.5 tidak memerlukan normalisasi, pengecekan dilakukan untuk memastikan tidak terjadi dominasi nilai oleh atribut numerik tertentu.
c. Penyusunan ulang struktur dataset: Atribut disusun ulang dengan atribut target diletakkan di bagian akhir, sesuai dengan konvensi dalam proses pemodelan klasifikasi di RapidMiner.

4. Proses Data Mining

Tahapan ini merupakan inti dari proses analisis, di mana dilakukan pembangunan model klasifikasi menggunakan algoritma C4.5. Beberapa prosedur teknis yang dilakukan adalah:

a. Pemilihan operator decision tree (C4.5) di RapidMiner, yang memungkinkan proses pemodelan berbasis entropy gain dan information gain ratio.
b. Pembagian data latih dan data uji: Dataset dibagi menjadi dua subset, yaitu 70% sebagai data latih dan 30% sebagai data uji. Selain itu, digunakan pula metode validasi silang (k-fold cross validation) untuk menghindari overfitting dan memberikan estimasi performa yang lebih stabil. Selain pembagian data latih dan data uji dengan rasio 70:30, penelitian ini menggunakan *10-fold cross-validation* untuk meminimalisir potensi overfitting.
c. Pelatihan model: Data latih digunakan untuk membangun model decision tree, di mana setiap node pada pohon keputusan merepresentasikan atribut yang membagi data berdasarkan nilai gain ratio tertinggi. Model C4.5 di RapidMiner menggunakan pengaturan default dengan *confidence factor* 0,25, *minimal gain* 0,1, dan *maximal depth* tidak dibatasi.

d. Pengujian model: Model yang telah terbentuk kemudian diuji terhadap data uji untuk menilai kemampuan prediktifnya terhadap data baru yang belum dikenali.

5. Evaluasi Pola (Pattern Evaluation)

Tahapan akhir adalah evaluasi model klasifikasi untuk mengukur performa dan efektivitas model dalam memprediksi status kelulusan mahasiswa (Putra et al., 2025). Evaluasi dilakukan dengan menggunakan beberapa metrik, yaitu:

a. Akurasi: Mengukur persentase klasifikasi yang benar dari total jumlah data uji (Maharani Hasan et al., 2024).
b. Precision: Mengukur ketepatan prediksi terhadap kelas "Lulus", yaitu proporsi prediksi "Lulus" yang benar dari seluruh prediksi "Lulus"(Wang et al., 2025).
c. Recall: Mengukur sensitivitas model terhadap kelas "Lulus", yaitu proporsi data "Lulus" yang berhasil terklasifikasi dengan benar (Xia et al., 2024).
d. F1-Score: Merupakan rata-rata harmonik dari precision dan recall (Khasanah et al., 2025).
e. Confusion Matrix: Digunakan untuk menggambarkan distribusi klasifikasi benar dan salah pada masing-masing kelas, baik "Lulus" maupun "Tidak Lulus" (Mega Amelia et al., 2025).

Evaluasi ini tidak hanya berfungsi sebagai tolok ukur performa model, tetapi juga memberikan gambaran mengenai atribut mana yang paling berkontribusi dalam klasifikasi status kelulusan mahasiswa. Informasi ini dapat dimanfaatkan oleh pihak akademik sebagai dasar dalam merumuskan kebijakan pembinaan dan intervensi dini terhadap mahasiswa.

III. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Tahapan awal dalam proses penelitian ini adalah pengumpulan data yang bertujuan untuk memperoleh informasi yang akurat dan representatif mengenai karakteristik mahasiswa yang berkaitan dengan kelulusan studi. Data yang digunakan dalam penelitian ini bersumber dari sistem informasi akademik, yang diklasifikasikan sebagai data sekunder. Dataset terdiri atas 379 entri mahasiswa, dengan sejumlah atribut yang mencerminkan informasi demografis dan akademik, yaitu: Nama, Jenis Kelamin, Status Mahasiswa, Umur, Status Nikah, Ips 1 Hingga Ips 8, Ipk, Dan Status Kelulusan.

Atribut Status Kelulusan berperan sebagai variabel target (kelas) dalam proses klasifikasi, sedangkan atribut lainnya berfungsi sebagai variabel prediktor. Pengumpulan data dilakukan secara sistematis dan terstandar, guna memastikan validitas serta reliabilitas data sebagai dasar analisis prediktif menggunakan algoritma C4.5. berikut adalah sample dari data yang digunakan

Tabel 1. Dataset Penelitian

Nama	Jenis Kelamin	Status Mahasiswa	Umur	Status Nikah
Anik Widayanti	Perempuan	Bekerja	28	Belum Menikah

Ips 1	Ips 2	Ips 3	Ips 4	Ips 5	Ips 6	Ips 7	Ips 8	Ipk	Status Kelulusan
2,76	2,8	3,2	3,17	2,98	3	3,03	0	3,07	Terlambat

Sumber: Hasil Penelitian (2025)

2. Pembersihan Data (Data Cleaning)

Setelah data diperoleh, tahap berikutnya adalah proses pembersihan data (data cleaning), yang bertujuan untuk menghilangkan inkonsistensi, duplikasi, serta nilai yang hilang atau tidak relevan yang dapat memengaruhi akurasi model prediksi. Tahapan ini krusial dalam mempersiapkan data agar layak untuk dianalisis secara komputasional.

Beberapa langkah yang dilakukan pada tahap ini meliputi:

a. Penghapusan atribut non-prediktif, seperti JENIS KELAMIN, yang tidak memiliki nilai informasi terhadap keluaran model.
b. Penanganan nilai hilang (missing value) dengan metode penghapusan atau imputasi sederhana sesuai kebutuhan.
c. Normalisasi dan standarisasi format penulisan untuk atribut kategorikal, seperti STATUS MAHASISWA dan JENIS KELAMIN, agar seragam dan terbaca oleh perangkat lunak RapidMiner.
d. Validasi rentang nilai pada atribut numerik, seperti IPS 1–8 dan IPK, untuk memastikan berada dalam rentang 0,00 hingga 4,00.

Proses ini dilakukan secara semi-otomatis menggunakan operator Replace Missing

Values, Filter Examples, dan Select Attributes dalam RapidMiner. Hasil pembersihan menunjukkan bahwa sebagian besar data berada dalam kondisi valid dan layak digunakan untuk proses klasifikasi, sehingga jumlah entri tetap dipertahankan sebanyak 379.

3. Transformasi Data

Setelah data dibersihkan, tahapan selanjutnya adalah transformasi data, yaitu proses konversi atau penyusunan ulang data agar sesuai dengan format dan struktur yang dibutuhkan oleh algoritma klasifikasi C4.5 pada platform RapidMiner. Transformasi diperlukan untuk memastikan bahwa setiap atribut dapat diinterpretasikan dengan benar oleh sistem, terutama atribut yang bersifat kategorikal.

Salah satu langkah transformasi yang dilakukan dalam penelitian ini adalah pengkodean ulang (re-encoding) nilai-nilai kategorikal menjadi representasi numerik. Hal ini bertujuan untuk menyederhanakan pemrosesan dan meningkatkan kompatibilitas data dengan algoritma klasifikasi. Contoh transformasi yang diterapkan adalah pada atribut STATUS NIKAH, dengan ketentuan sebagai berikut:

a) Menikah dikonversi menjadi nilai 0
b) Belum Menikah dikonversi menjadi nilai

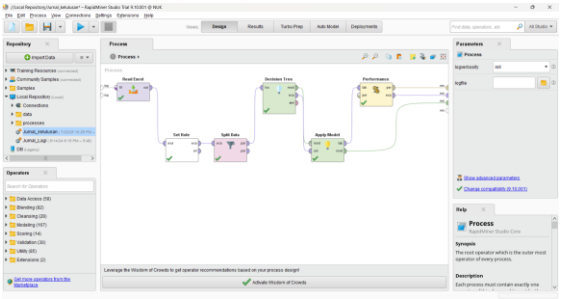
Transformasi ini dilakukan secara eksplisit menggunakan fitur Generate Attributes dan Nominal to Numerical pada RapidMiner. Selain STATUS NIKAH, atribut lain yang bersifat kategorikal, seperti JENIS KELAMIN dan STATUS MAHASISWA, juga dikonversi ke dalam bentuk nilai nominal atau biner sesuai skema klasifikasi yang relevan.

Adapun atribut numerik seperti IPS dan IPK tidak memerlukan transformasi lanjutan, karena telah memenuhi standar numerik dengan rentang nilai antara 0,00 hingga 4,00. Namun, validasi dilakukan untuk memastikan tidak terdapat outlier atau nilai ekstrem yang dapat mengganggu proses klasifikasi.

Dengan transformasi ini, data telah siap untuk digunakan dalam tahap penambangan data (data mining) menggunakan algoritma C4.5, di mana setiap atribut telah memiliki format yang dapat diproses secara otomatis dan efisien oleh sistem.

4. Proses Data Mining

Proses ini terdiri dari beberapa tahapan yang tersusun secara sistematis untuk membangun model klasifikasi kelulusan mahasiswa menggunakan algoritma C4.5 Decision Tree.



Sumber: Hasil Penelitian (2025)

Gambar 1. Proses Data Mining

Adapun penjelasan tiap operator dalam proses tersebut adalah sebagai berikut:

a. *Read Excel*
Operator ini digunakan untuk mengimpor data dari file Excel (.xlsx) yang berisi dataset mahasiswa. Dataset terdiri dari 379 entri dengan atribut seperti nama, jenis kelamin, umur, status nikah, status mahasiswa, IPS 1 hingga 8, IPK, dan status kelulusan. Pada tahap ini, sistem membaca keseluruhan data untuk selanjutnya diproses.

b. *Set Role*
Operator ini digunakan untuk menentukan atribut mana yang dijadikan sebagai label (target klasifikasi). Dalam konteks penelitian ini, atribut STATUS KELULUSAN diatur sebagai label yang akan diprediksi, sementara atribut lain berfungsi sebagai *features* atau *predictors*. Jika ada atribut identitas seperti NAMA, biasanya dikecualikan dari proses modeling karena tidak relevan terhadap prediksi.

c. Split Data
Tahapan ini bertujuan untuk membagi dataset menjadi dua bagian, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*). Umumnya, pembagian dilakukan dengan rasio 70:30 atau 80:20. Dalam penelitian ini, pembagian data memungkinkan proses pelatihan model dilakukan secara terpisah dari evaluasi performa model, sehingga hasilnya lebih valid.

d. Decision Tree
Operator ini digunakan untuk membangun model klasifikasi berdasarkan algoritma Decision Tree C4.5. Pada tahap ini, model dilatih menggunakan data training yang telah ditentukan sebelumnya untuk menemukan pola atau aturan klasifikasi dari atribut-atribut input terhadap label STATUS KELULUSAN.

e. Apply Model
Setelah model terbentuk, operator ini digunakan untuk menerapkan model terhadap data pengujian (testing data). Hasil prediksi dari model akan dibandingkan dengan nilai aktual dari data uji untuk mengevaluasi performa model.

f. Performance
Operator ini menghitung dan menampilkan metrik evaluasi model klasifikasi. Metrik yang diperoleh dalam penelitian ini meliputi akurasi (97,81%), precision, recall, dan AUC (0,991). Nilai-nilai ini menjadi indikator utama dalam mengukur efektivitas dan keandalan model yang dibangun.

5. Evaluasi Pola (Pattern Evaluation)

Algoritma C4.5 telah diterapkan untuk melakukan klasifikasi terhadap status kelulusan mahasiswa, yang dikategorikan ke dalam dua kelas yaitu TEPAT (lulus tepat waktu) dan TERLAMBAT (lulus tidak tepat waktu). Evaluasi model dilakukan menggunakan confusion matrix yang dihasilkan melalui perangkat lunak RapidMiner, dengan hasil sebagai berikut:

accuracy: 97.81%

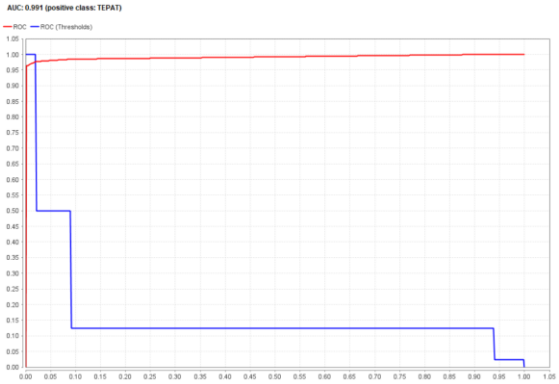
	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	98	5	95.15%
pred. TEPAT	0	125	100.00%
class recall	100.00%	96.15%	

Sumber: Hasil Penelitian (2025)

Gambar 2. Confusion Matrix

Model berhasil mengklasifikasikan 98 data mahasiswa yang tergolong TERLAMBAT secara benar, serta 125 data mahasiswa yang tergolong TEPAT juga diklasifikasikan dengan tepat. Terdapat 5 data mahasiswa yang sebenarnya TEPAT, namun diklasifikasikan sebagai TERLAMBAT, sedangkan tidak ditemukan kesalahan klasifikasi terhadap mahasiswa yang TERLAMBAT. Dengan demikian, model tidak menghasilkan nilai false negative untuk kelas TERLAMBAT.

Akurasi keseluruhan dari model mencapai 97,81%, yang menunjukkan bahwa algoritma C4.5 memiliki kemampuan klasifikasi yang sangat baik terhadap data yang digunakan. Nilai precision untuk kelas TERLAMBAT adalah sebesar 95,15%, sedangkan untuk kelas TEPAT mencapai 100,00%, yang berarti bahwa seluruh prediksi mahasiswa TEPAT oleh model bersifat benar (tanpa false positive). Sementara itu, nilai recall untuk kelas TERLAMBAT adalah 100,00%, menunjukkan bahwa semua mahasiswa yang benar-benar TERLAMBAT berhasil diidentifikasi oleh model. Recall untuk kelas TEPAT adalah 96,15%, yang mengindikasikan bahwa sebagian besar mahasiswa yang seharusnya TEPAT berhasil dikenali dengan benar.

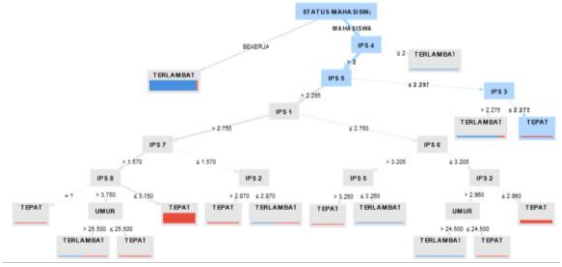


Sumber: Hasil Penelitian (2025)

Gambar 3. Kurva ROC

Selain evaluasi menggunakan confusion matrix, dilakukan pula analisis performa model dengan memanfaatkan kurva Receiver Operating Characteristic (ROC). Kurva ROC digunakan untuk mengevaluasi kemampuan model dalam membedakan antara dua kelas, yaitu TEPAT dan TERLAMBAT. Pada Gambar 4, tampak bahwa kurva ROC yang dihasilkan oleh model C4.5 sangat mendekati sisi kiri atas dari grafik, yang menandakan performa klasifikasi yang sangat baik.

Nilai Area Under Curve (AUC) yang diperoleh adalah sebesar 0,991, yang menunjukkan bahwa model memiliki tingkat diskriminasi yang sangat tinggi terhadap kelas positif, yaitu TEPAT. Secara umum, AUC berada pada rentang 0 hingga 1, dengan nilai mendekati 1 menunjukkan kemampuan klasifikasi yang hampir sempurna. Nilai AUC sebesar 0,991 mengindikasikan bahwa probabilitas model dalam membedakan mahasiswa yang lulus tepat waktu dan tidak tepat waktu sangat tinggi.



Sumber: Hasil Penelitian (2025)

Gambar 4. Decision Tree

Gambar 5 menunjukkan struktur pohon keputusan yang dihasilkan oleh algoritma C4.5 untuk memprediksi kelulusan mahasiswa ke dalam dua kategori, yaitu *TEPAT* (tepat waktu) dan *TERLAMBAT* (tidak tepat waktu). Pohon keputusan ini dibangun berdasarkan sejumlah atribut prediktor, antara lain status mahasiswa (bekerja/tidak bekerja), IP semester (IPS1–IPS8), dan umur mahasiswa.

Node akar (root node) dimulai dari atribut *status mahasiswa*. Jika mahasiswa diketahui bekerja, maka model langsung memprediksi bahwa mahasiswa tersebut cenderung lulus *terlambat*. Jika tidak bekerja, maka pohon melanjutkan pemisahan berdasarkan nilai IPS4. Apabila *IPS4* ≤ 2, maka mahasiswa juga diprediksi lulus *terlambat*. Namun, jika *IPS4* > 2, maka klasifikasi lebih lanjut dilakukan berdasarkan atribut *IPS5* dan seterusnya.

Beberapa pola yang dapat diamati dari pohon keputusan ini antara lain:

a. Mahasiswa yang memiliki IPS3 > 2,275 cenderung lulus *tepat waktu*, sedangkan nilai di bawahnya mengarah pada prediksi *terlambat*.
b. Atribut IPS1 dan IPS2 menjadi faktor pembeda penting pada cabang lain, di mana nilai *IPS1 > 2,750* cenderung mengarah ke prediksi *tepat waktu*.
c. Di bagian bawah pohon, atribut umur juga memengaruhi hasil klasifikasi. Misalnya, pada salah satu cabang, mahasiswa dengan *umur ≤ 24,500* dan *IPS2 > 2,960* diprediksi akan lulus *tepat waktu*.

Pohon ini menyajikan struktur klasifikasi yang transparan dan dapat ditelusuri kembali logika keputusannya. Warna biru menunjukkan dominasi prediksi kelas *TERLAMBAT*, sedangkan warna merah menunjukkan kelas *TEPAT* pada setiap daun (leaf node).

Dari struktur tersebut dapat disimpulkan bahwa nilai IP per semester (khususnya IPS3, IPS4, dan IPS5) merupakan atribut yang sangat berpengaruh dalam menentukan kelulusan, disusul oleh status pekerjaan dan umur mahasiswa.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma C4.5 efektif dan dapat diandalkan dalam memodelkan dan memprediksi faktor-faktor yang memengaruhi kelulusan mahasiswa. Jika dibandingkan dengan penelitian sebelumnya, akurasi model C4.5 (97,81%) lebih tinggi daripada Naïve Bayes (88,16%) dan K-NN (87,8%). Implementasi model semacam ini dapat dimanfaatkan oleh pihak institusi pendidikan untuk melakukan identifikasi dini terhadap mahasiswa yang berisiko terlambat lulus, serta merancang intervensi akademik secara tepat sasaran.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa algoritma C4.5 menunjukkan performa yang sangat baik dalam memprediksi kelulusan mahasiswa berdasarkan data akademik dan demografis. Model klasifikasi yang dibangun menggunakan perangkat lunak RapidMiner mampu menghasilkan akurasi sebesar 97,81% dan Area Under Curve (AUC) sebesar 0,991, yang mengindikasikan tingkat akurasi dan kemampuan diskriminasi model yang sangat tinggi.

Atribut yang paling berpengaruh dalam klasifikasi kelulusan mahasiswa adalah nilai Indeks Prestasi Semester (IPS), khususnya pada semester ke-3 hingga ke-5, disertai dengan status pekerjaan mahasiswa dan umur. Model mampu mengidentifikasi dengan tepat mahasiswa yang berpotensi mengalami keterlambatan kelulusan, dan hasil ini menunjukkan keunggulan dibandingkan penelitian sebelumnya yang hanya mencapai akurasi 87,2%.

Secara keseluruhan, penelitian ini membuktikan bahwa metode C4.5 efektif dan layak diterapkan sebagai alat bantu pengambilan keputusan dalam bidang akademik, terutama untuk keperluan monitoring kelulusan dan pemberian intervensi secara tepat waktu.

V. REFERENSI

Bramanda, J. (2025). Klasifikasi Masyarakat Penerima Bantuan Sosial dari Pemerintah dengan Metode Algoritma C4.5. *Jurnal Komputer Antartika*, *3*(1), 34–41. https://doi.org/10.70052/jka.v3i1.234

Khasanah, N., Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES. In *Fakultas Teknologi Informasi* (Vol. 13, Issue 3).

Khasanah, N., Uki Eka Saputri, D., Aziz, F., Hidayat, T., Nusa Mandiri Jl, U., & Author, C. (2025). Studi Perbandingan Algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur. In *Cipinang Melayu, Kec. Makasar* (Vol. 5, Issue 1). http://jurnal.bsi.ac.id/index.php/co-science

Kristen, U., Wacana, S., Tabuni, N., Tri, S., Universitas, S., Wacana, K. S., & Ismanto, B. (2024). *Evaluasi Program Beasiswa Bagi Mahasiswa Kabupaten Pegunungan*

Bintang Menggunakan Model Goal Oriented Evaluation. *11*, 245–259.

Maharani Hasan, P., Ayu, N., & Adi Saputra, R. (2024). KLASIFIKASI KERINGANAN UKT MAHASISWA UHO MENGGUNAKAN K-NEAREST NEIGHBOR (KNN). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 6).

Mantopani, R., Ramadhan, I. L., Samsara, R. P., Samsinar, R., & Elektro, T. (2024). *Analisis Perbandingan Decision Tree Dan Algoritma C4.5 Untuk Mengklasifikasikan Penerimaan Mahasiswa Elektro*.

Mega Amelia, M., Maulana Fazrin, B., Yosua Panjaitan, Y., Dicky Kurniawan, M., & Khasanah, N. (2025). Implementasi Naive Bayes Untuk Klasifikasi Gangguan Tidur. *Journal Computer Science*, *4*(1).

Putra, H., Nasution, K., & Rilvani, E. (2025). PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5. *JMA)*, *3*(7), 3031–5220. https://doi.org/10.23887/jpi-undiksha.v10i2.30020

Putri, A., Syaficha Hardiana, C., Novfuja, E., Try Puspa Siregar, F., Fatma, Y., & Wahyuni, R. (2023). Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir. *Institut Riset Dan Publikasi Indonesia (IRPI) MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage*, *3*(1), 20–26.

Rachardian, S., Sediyono, E., Teknologi Informasi, F., Kristen Satya Wacana Jl Notohamidjojo, U. O., & Tengah, J. (2024). Prediksi kelulusan tepat waktu mahasiswa untuk pemantauan program studi menggunakan metode data mining. *AITI: Jurnal Teknologi Informasi*, *21*(2), 168–182.

Ramadani, P., Fadillah, R., Adawiyah, Q., & Restu Al Ghazali, B. (2024). *Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan*.

Santoso, L., & Priyadi. (2024). *Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI*. *17*(2). https://doi.org/10.51903/elkom.v17i2.2311

Susiyati, & Sihono. (2025). *REFORMASI PERGURUAN TINGGI KEAGAMAAN ISLAM NEGERI MENUJU PENDIDIKAN BERKUALITAS DAN BERDAYA SAING GLOBAL*.

Syahriani, S. (2022). Penerapan Data Mining Untuk Menentukan Pola Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Bina Insani ICT Journal*, *9*(1), 43–52.

Tamimi, A., Octafriyanda, D., Hutagaol, R., Yuhdi, A., Puteri, A., Bahasa, P., Indonesia, S., Bahasa, F., & Seni, D. (2024). *Merdeka Belajar: Mewujudkan Pendidikan Berkualitas Melalui Kampus Merdeka di Universitas Negeri Medan*. https://jicnusantara.com/index.php/jicn

Wang, B., Liang, Z., Zhang, J., & Wu, D. (2025). Two-Stage High-Precision Visual 6-D Pose Measurement Method of Thin Parts With Irregular Shape for Precision Assembly. *IEEE Transactions on Instrumentation and Measurement*, *74*, 1–13. https://doi.org/10.1109/TIM.2025.3560727

Xia, P., Liu, Z., Wang, W., Shi, W., & Van Nieuwenhuyse, I. (2024). Effects of Firm Innovation and Consumer eWOM on Product Recalls in an Emerging Market: The Moderating Role of Ownership Structure. *IEEE Transactions on Engineering Management*, *71*, 2121–2137. https://doi.org/10.1109/TEM.2022.3168429

Zulfa, A. A., & Arifudin, O. (2025). PERAN SISTEM INFORMASI AKADEMIK BERBASIS WEB DALAM UPAYA MENINGKATKAN EFEKTIVITAS DAN EFISIENSI PENGELOLAAN AKADEMIK DI PERGURUAN TINGGI. In *Jurnal Tahsinia* (Vol. 6, Issue 1).