

# Klasifikasi Mahasiswa Berprestasi Berdasarkan Nilai Mahasiswa Universitas X Dengan Algoritma C4.5

Daniel Marcello\*, Petra Putri Ardika, Andri Wijaya

Universitas Katolik Musi Charitas Palembang

DOI:

<https://doi.org/10.47134/jtsi.v3i1.5213>

\*Correspondence: Petra Putri Ardika

Email: [fetraputri94@gmail.com](mailto:fetraputri94@gmail.com)

Received: 06-12-2025

Accepted: 31-12-2025

Published: 03-01-2026



**Copyright:** © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstrak:** Penelitian ini bertujuan mengklasifikasikan mahasiswa berprestasi berdasarkan nilai akademik dengan memanfaatkan algoritma C4.5 sebagai salah satu teknik klasifikasi yang populer dalam data mining. Data yang digunakan mencakup IPK semester 1–3, jumlah SKS yang telah ditempuh, status keaktifan dalam kegiatan, serta status prestasi mahasiswa. Tahapan penelitian meliputi proses *data cleaning*, seleksi atribut untuk menentukan variabel paling relevan, pembagian data menjadi training dan testing, hingga pembangunan model *decision tree* menggunakan aplikasi RapidMiner. Hasil penelitian menunjukkan bahwa atribut jumlah SKS dan IPK per semester merupakan faktor paling berpengaruh dalam menentukan kategori prestasi mahasiswa. Model C4.5 yang dihasilkan memperoleh tingkat akurasi sebesar 97%, dengan nilai precision untuk kelas “Ya” mencapai 100%, menunjukkan performa model yang sangat baik. Struktur pohon keputusan yang terbentuk mengindikasikan bahwa mahasiswa dengan jumlah SKS tinggi dan IPK yang stabil memiliki peluang lebih besar untuk diklasifikasikan sebagai mahasiswa berprestasi. Temuan ini membuktikan bahwa algoritma C4.5 efektif sebagai alat bantu dalam proses seleksi dan dapat mendukung pengambilan keputusan akademik secara objektif dan terukur.

**Keywords:** C4.5, *Decision Tree*, RapidMiner, Klasifikasi

**Abstract:** This study aims to classify outstanding students based on their academic performance using the C4.5 algorithm, one of the most widely used classification techniques in data mining. The dataset includes GPA from semesters 1–3, the total number of credits completed, participation status in student activities, and student achievement status. The research stages include data cleaning, attribute selection to determine the most relevant variables, splitting the dataset into training and testing sets, and constructing a decision tree model using RapidMiner. The results indicate that the total number of credits and semester GPA are the most influential factors in determining student achievement categories. The generated C4.5 model achieved an accuracy rate of 97%, with a precision value of 100% for the “Yes” class, demonstrating excellent model performance. The resulting decision tree structure also shows that students with a high number of completed credits and consistently strong GPAs are more likely to be classified as outstanding students. These findings confirm that the C4.5 algorithm is an effective tool to support the selection process and can assist in academic decision-making in an objective and measurable manner.

**Keywords:** C4.5, *Decision Tree*, RapidMiner, Klasifikasi

## Pendahuluan

Pendidikan tinggi merupakan salah satu pilar penting dalam pembangunan suatu bangsa (Tamimi et al., 2024), serta memiliki tanggung jawab strategis dalam menghasilkan sumber daya manusia yang unggul dan mampu bersaing di era global. Salah satu tantangan utama yang dihadapi oleh institusi pendidikan tinggi adalah bagaimana memastikan mahasiswa bisa menyelesaikan studi tepat waktu dengan hasil yang optimal

(Tahsinia et al., 2025). Untuk mencapai tujuan tersebut, universitas harus memastikan bahwa proses pembelajaran berlangsung secara efektif, terukur, dan berkelanjutan (Berkualitas et al., 2025). Tidak sedikit mahasiswa yang mengalami keterlambatan kelulusan, bahkan hingga putus studi, yang disebabkan oleh beberapa faktor seperti rendahnya capaian akademik, kondisi sosial ekonomi, hingga minimnya motivasi belajar (Tabuni et al., 2024).

Pemantauan perkembangan akademik mahasiswa secara berkelanjutan menjadi salah satu langkah penting untuk mengevaluasi kualitas pembelajaran sekaligus merumuskan strategi peningkatan mutu. Namun, seiring bertambahnya jumlah mahasiswa setiap tahunnya, volume dan kompleksitas data akademik yang dihasilkan semakin meningkat, sehingga pengelolaan manual tidak lagi efisien dan berpotensi menimbulkan kesalahan. Kondisi ini menuntut pemanfaatan teknologi informasi yang mampu mengolah data secara lebih cepat, akurat, dan bernilai bagi pengambilan keputusan akademik. Salah satu solusi yang bisa digunakan dalam mengatasi permasalahan ini adalah dengan pemanfaatan teknologi data mining untuk melakukan prediksi kelulusan berdasarkan data historis mahasiswa (Rachardian et al., 2024).

Seiring perkembangan teknologi, analisis berbasis data menjadi alat yang semakin penting untuk mengatasi tantangan ini (Sumual et al., n.d.). *Data mining* adalah serangkaian proses yang secara manual mengekstraksi nilai tambahan yang berasal dari informasi yang sebelumnya tidak diketahui dari sekumpulan data (Sari et al., 2021). Dalam *data mining* terdapat beberapa fungsi, salah satu fungsi utama data mining adalah klasifikasi, yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu. *Data mining* memungkinkan institusi mengekstraksi pola tersembunyi dari data sehingga dapat digunakan untuk memahami karakteristik mahasiswa, memprediksi performa belajar, serta mengidentifikasi faktor-faktor yang mempengaruhi prestasi (Muhammad Hamdan Sukri & Handrianto, 2024). Universitas X menghadapi tantangan serupa, terutama dalam pengelolaan data nilai mahasiswa yang dibutuhkan untuk proses seleksi mahasiswa berprestasi. *Data mining* adalah proses menemukan pola atau informasi yang berguna dari sekumpulan data besar (Syahriani, 2022). Proses seleksi yang melibatkan banyak indikator seperti nilai akademik, prestasi individu, dan aktivitas organisasi sering kali dilakukan secara manual, sehingga rawan subjektivitas dan kurang efektif. Peningkatan prestasi mahasiswa akan berdampak dalam peningkatan kualitas sumber daya manusia (Meliawati, 2020).

Untuk mengatasi permasalahan tersebut, diperlukan metode analisis yang mampu mengelompokkan data secara objektif dan akurat. 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). Algoritma C4.5 merupakan algoritma klasifikasi yang efektif dalam menghasilkan pohon keputusan yang mudah dipahami serta mampu menangani berbagai jenis atribut, termasuk atribut kontinu. Konsep dari *decision tree* adalah mengubah data menjadi aturan-aturan keputusan (Steffany et al., 2023). Cara untuk membuat model *Decision Tree* adalah dengan memecah data ke dalam kelompok yang jauh lebih kecil berdasarkan

dengan atribut di dalam data (Ramadhon et al., 2024) atau secara umum tahapan dalam algoritma C4.5 untuk membangun sebuah pohon keputusan dengan memilih atribut sebagai akar, kemudian cabang untuk tiap-tiap nilai, membagi kasus dalam cabang, dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama dan membentuk pohon keputusan (Noverta Effendi\*1, Witrihan Ramadhani2, Fitri Farida3, 2024). Melalui pendekatan ini, universitas dapat mengidentifikasi pola-pola yang menentukan prestasi mahasiswa berdasarkan data akademik yang tersedia, sehingga proses seleksi dapat dilakukan secara lebih terukur dan minim subjektivitas (Sihombing et al., 2021). Berbagai penelitian sebelumnya telah menunjukkan efektivitas algoritma klasifikasi, khususnya C4.5, dalam berbagai domain penerapan, termasuk pendidikan dan bantuan social (Khasanah et al., 2025)

Penerapan algoritma C4.5 dalam klasifikasi mahasiswa berprestasi di Universitas X diharapkan mampu membantu proses seleksi agar lebih cepat, efisien, dan akurat. Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas atau konsep data (Ismanto & Novalia, 2021). Dengan memanfaatkan data historis mahasiswa, algoritma ini dapat menghasilkan model klasifikasi yang menunjukkan apakah seorang mahasiswa termasuk kategori berprestasi atau tidak. Dalam penerapannya, pada umumnya pengukuran akurasi bersumber dari kerangka kebingungan, juga disebut *confusion matrix* untuk mengetahui tingkat akurasi klasifikasi dan memastikan kelayakan model sebagai pendukung keputusan akademik (Hasibuan & Mahdiana, 2023). Akurasi yang tinggi menjadi indikator penting bahwa model tersebut dapat diandalkan dalam proses seleksi mahasiswa berprestasi.

Rapidminer adalah platform perangkat lunak yang menawarkan berbagai alat mulai dari persiapan data, pemodelan, evaluasi, dan penerapan. RapidMiner dirancang agar pengguna dapat dengan mudah membuat dan menguji berbagai model tanpa memerlukan keterampilan pemrograman. RapidMiner memiliki antarmuka yang memungkinkan pengguna secara mudah merancang alur kerja untuk pengolahan dan analisis data. Perangkat lunak ini juga menyediakan banyak operator yang telah siap pakai, yang menjadi komponen alur kerja, mencakup semua tahap dalam proses penambangan data, seperti pembersihan data, pemilihan fitur, dan pemodelan (Nahjan et al., 2023).

Dengan demikian, penerapan algoritma C4.5 tidak hanya mendukung proses seleksi mahasiswa berprestasi, tetapi juga memberikan manfaat lebih luas bagi pengembangan kebijakan akademik dan peningkatan kualitas pembelajaran. Informasi yang diperoleh melalui proses klasifikasi membantu universitas memahami faktor-faktor yang mempengaruhi prestasi mahasiswa, sehingga dapat dijadikan dasar dalam merumuskan strategi pembinaan dan pengembangan akademik yang lebih efektif. Oleh karena itu, penelitian terkait klasifikasi mahasiswa berprestasi menggunakan algoritma C4.5 menjadi penting dilakukan sebagai bagian dari transformasi digital di lingkungan pendidikan tinggi. Tujuan dari penelitian ini adalah membangun model klasifikasi menggunakan algoritma C4.5 untuk memprediksi kelulusan mahasiswa berdasarkan atribut-atribut akademik dan demografis

## Metodologi

Penelitian ini memanfaatkan pendekatan kualitatif dengan menggunakan metode *data mining* berbasis algoritma C4.5 (Azzahra et al., 2023), untuk mengklasifikasikan mahasiswa berprestasi berdasarkan data nilai akademik mahasiswa Universitas X. Proses penelitian terdiri dari empat tahapan utama, yaitu pengumpulan data, *preprocessing* data, pembangunan model klasifikasi menggunakan algoritma C4.5, serta evaluasi model melalui *confusion matrix*.

### 1. Pengumpulan Data

Data penelitian diperoleh dari dataset nilai akademik mahasiswa Universitas X. Atribut yang digunakan meliputi nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), serta status prestasi mahasiswa. Seluruh atribut ini dijadikan variabel input untuk menentukan kategori mahasiswa berprestasi dan tidak berprestasi. Pemilihan atribut dilakukan untuk memastikan bahwa variabel yang digunakan relevan dengan tujuan klasifikasi.

### 2. *Preprocessing*

Proses *preprocessing* dilakukan mengacu pada tahapan KDD, meliputi:

- a. *Data cleaning*, untuk menghapus data ganda, hilang, atau tidak lengkap.
- b. *Data selection*, guna memilih atribut yang relevan dengan proses klasifikasi.
- c. *Data transformation*, yaitu mengubah format data ke dalam bentuk sesuai dengan pemodelan di *RapidMiner*, seperti pengubahan tipe data, normalisasi, atau pengelompokan nilai. Tahapan *preprocessing* ini dilakukan untuk memastikan bahwa data berkualitas dan siap digunakan dalam proses pemodelan, karena kualitas data sangat berpengaruh terhadap hasil klasifikasi (Azhari et al., 2021).

### 3. Penerapan Algoritma C4.5

Setelah data siap, algoritma C4.5 diterapkan menggunakan aplikasi *RapidMiner*. Proses pemodelan diawali dengan perhitungan *entropy* dan *gain* dari setiap atribut untuk menentukan atribut dengan pengaruh terbesar sebagai root (akar) pohon keputusan (Rofiani et al., 2024). Selanjutnya, pembentukan cabang dilakukan berdasarkan nilai atribut lain hingga menghasilkan *leaf node* yang mewakili kelas akhir, yaitu kategori mahasiswa berprestasi atau tidak berprestasi. Pohon keputusan yang dihasilkan memberikan gambaran visual mengenai pola penentuan prestasi mahasiswa.

### 4. Evaluasi Model

Model yang telah dibangun dievaluasi menggunakan *confusion matrix* untuk mengukur tingkat akurasi, melalui perbandingan antara label aktual dan hasil prediksi model. *Confusion matrix* menghasilkan nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative*, yang kemudian digunakan untuk menghitung akurasi model (Widiastuti et al., 2022). Nilai akurasi menjadi indikator tingkat keandalan model dalam mengklasifikasikan mahasiswa berprestasi. Semakin tinggi nilai akurasi, semakin baik kemampuan model dalam memberikan hasil klasifikasi yang tepat bagi Universitas X.

## Hasil dan Pembahasan

### Pengumpulan Data

Tahapan awal yang digunakan pada penelitian ini adalah proses pengumpulan data dengan tujuan untuk memperoleh dataset yang akurat dan dapat merepresentasikan kriteria yang berkaitan dengan terpilihnya mahasiswa berprestasi. Data yang digunakan sebanyak 1.000 data mahasiswa Universitas X dengan format file .xlsx yang digunakan untuk perhitungan alternatif tertinggi penentuan mahasiswa yang akan dipilih menjadi mahasiswa berprestasi. Metode yang diusulkan untuk proses seperti yang telah dijelaskan diatas yaitu metode klasifikasi dengan algoritma yang digunakan adalah algoritma C4.5 dengan kriteria yang digunakan sebagai berikut: NIM Mahasiswa, Nama Mahasiswa, Status Kegiatan, IPK Semester 1, IPK Semester 2, IPK Semester 3, Jumlah SKS, dan Status Mahasiswa Berprestasi. Berikut adalah sampel dari data yang digunakan:

**Tabel 1.**Data Mahasiswa Universitas X

NIM	Nama	Status	IPK	IPK	IPK	Jumlah SKS	Status
		Kegiatan	Semester 1	Semester 2	Semester 3		Mahasiswa Berprestasi
202500001	Adi Santoso	Tidak Aktif	3.02	3.14	3.91	58	Tidak
202500002	Joko Lestari	Tidak Aktif	2.3	3.2	2.14	72	Tidak
202500003	Dewi Lestari	Aktif	3.35	3.18	3.58	51	Tidak
202500004	Oki Sari	Aktif	3.39	2.79	2.93	43	Tidak
202500005	Adi Saputra	Tidak Aktif	2.61	3.18	2.47	67	Tidak
202500006	Gita Santoso	Tidak Aktif	3.65	3.82	2.62	47	Tidak
202500007	Kiki Kusuma	Tidak Aktif	2.29	2.68	3.33	48	Ya
202500008	Gita Santoso	Aktif	2.55	3.31	3.47	51	Tidak
202500009	Dewi Pratama	Tidak Aktif	3.84	2.56	2.51	41	Tidak
202500010	Qori Lestari	Tidak Aktif	3.52	02.05	2.19	57	Tidak
202500011	Oki Sari	Aktif	2.42	2.44	2.97	44	Tidak
202500012	Tono Santoso	Tidak Aktif	2.33	3.44	3.76	62	Tidak
202500013	Budi Wijaya	Aktif	3.43	3.3	2.99	72	Tidak
202500014	Joko Sari	Aktif	3.39	3.95	2.47	49	Tidak

### Pembersihan Data (*Data Cleaning*)

Setelah data terkumpul, tahap yang selanjutnya dilakukan adalah pembersihan data atau data cleaning. Data cleaning adalah tahap pembersihan data yang tidak konsisten, dengan tujuan untuk menghilangkan nilai yang tidak mempengaruhi akurasi model atau tidak relevan yang akan mempengaruhi akurasi dari model prediksi (Diki Wahyudi, 2023). Tahap ini sangat penting untuk mempersiapkan data agar dapat dianalisis secara komputasional.

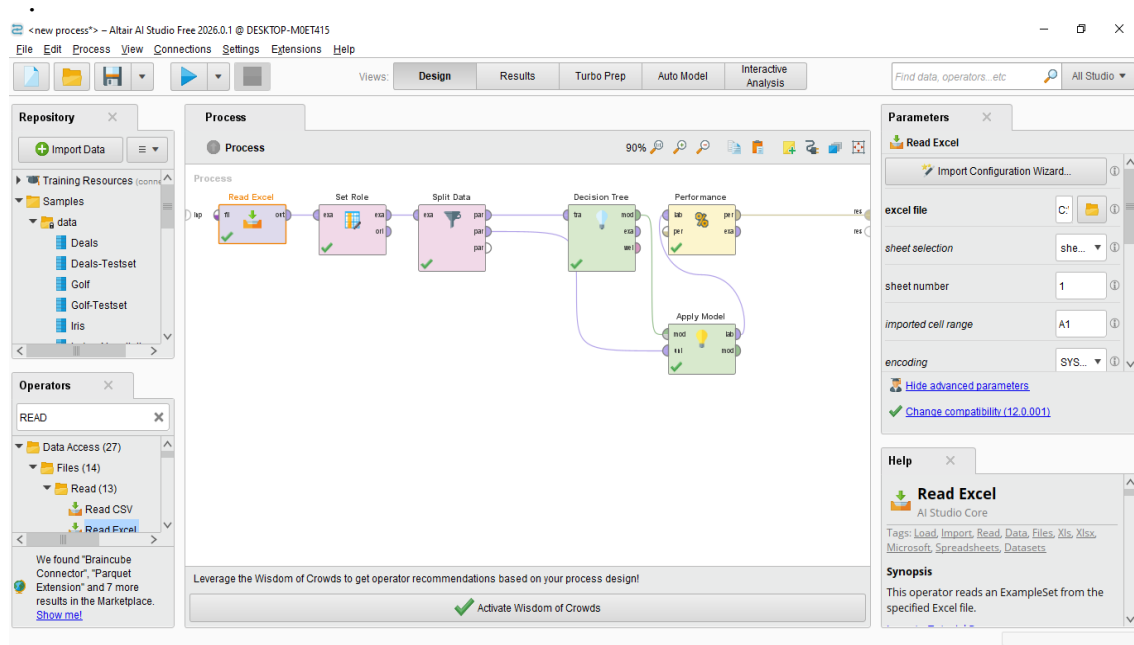
Berikut adalah beberapa langkah yang dilakukan pada tahap pembersihan data atau *data cleaning* meliputi:

1. Melakukan penghapusan pada atribut yang nonpredikatif, seperti atribut NIM, yang tidak mempengaruhi nilai akurasi terhadap model prediksi.
2. Melakukan penggantian nilai hilang (missing value) dengan menggunakan metode penghapusan sederhana yang sesuai dengan kebutuhan.
3. Melakukan validasi rentang nilai pada atribut numerik, seperti IPK, untuk memastikan agar berada dalam rentang 0,00 sampai dengan 4,00.

Proses *data cleaning* dilakukan dengan menggunakan operator *Replace Missing Values*, *Set Role* dan *Select Attributes* yang ada pada *RapidMiner*. Hasil *data cleaning* menghasilkan data yang berada dalam kondisi valid serta layak digunakan untuk melakukan proses klasifikasi (Andarista & Jananto, 2022).

### Proses Data Mining

Proses yang dilakukan selanjutnya adalah data mining, Proses ini terdiri dari beberapa tahapan yang tersusun dengan sistematis agar dapat membangun model klasifikasi untuk memilih mahasiswa berprestasi menggunakan algoritma *C4.5 Decision Tree*.



Gambar 1. Proses Data Mining

#### 1. Read Excel

Operator Read Excel digunakan untuk mengimport data dari file dengan format *.xlsx* yang berisikan data set mahasiswa. Data yang terdiri dari 1.000

mahasiswa dengan atribut NIM, Nama Mahasiswa, Status Kegiatan, IPK Semester 1, IPK Semester 2, IPK Semester 3, Jumlah SKS, dan Status Mahasiswa Berprestasi. Pada tahap ini, operator membantu untuk membaca semua data agar dapat diproses pada tahap selanjutnya.

## 2. *Set Role*

Operator *Set Role* digunakan untuk menentukan atribut yang akan diberikan label atau target klasifikasinya. Dalam penelitian ini, atribut Status Mahasiswa Berprestasi digunakan dan di setting sebagai label yang akan diprediksi, sementara itu atribut lain berfungsi sebagai *feature* atau *predictors*.

## 3. *Split Data*

Operator *Split Data* digunakan untuk membagi dataset menjadi dua bagian, yaitu data training dan data testing. Pada penelitian ini, pembagian dilakukan dengan menggunakan rasio 80:20. Pembagian data ini dapat memungkinkan proses pelatihan model dilakukan secara terpisah dari evaluasi performance, sehingga hasil prediksi yang dihasilkan menjadi lebih valid.

## 4. *Decision Tree*

Operator *Decision Tree* digunakan untuk membangun model klasifikasi yang berdasarkan dengan algoritma *Decision Tree* C4.5. Pada tahap ini, model akan dilatih menggunakan data training yang telah ditentukan dengan *Split Data* agar dapat menemukan pola klasifikasi dari atribut atribut input terhadap label Status Mahasiswa Berprestasi.

## 5. *Apply Model*

Operator *Apply Model* digunakan setelah model klasifikasi terbentuk, untuk menerapkan model dengan data testing. Hasil Prediksi dari model akan digunakan untuk dilakukan perbandingan dengan nilai aktual dari data testing untuk mengevaluasi model klasifikasi.

## 6. *Performance*

Operator *Performance* digunakan dalam menghitung dan menampilkan metrik evaluasi dari model klasifikasi yang telah dibuat. Metrik yang didapatkan dalam penelitian ini meliputi akurasi (85,00%). Nilai ini akan menjadi indikator utama dalam mengukur efektivitas, kemahiran, dan kecakapan model yang telah dibangun.

## Evaluasi Pola

Algoritma C4.5 telah diterapkan untuk melakukan klasifikasi terhadap status mahasiswa berprestasi, yang dikategorikan ke dalam dua kelas yaitu YA (lulus menjadi mahasiswa berprestasi) dan TIDAK (tidak lulus menjadi mahasiswa berprestasi). Evaluasi model ini dilakukan dengan *confusion matrix* yang ditampilkan melalui *RapidMiner*, dengan hasil sebagai berikut:

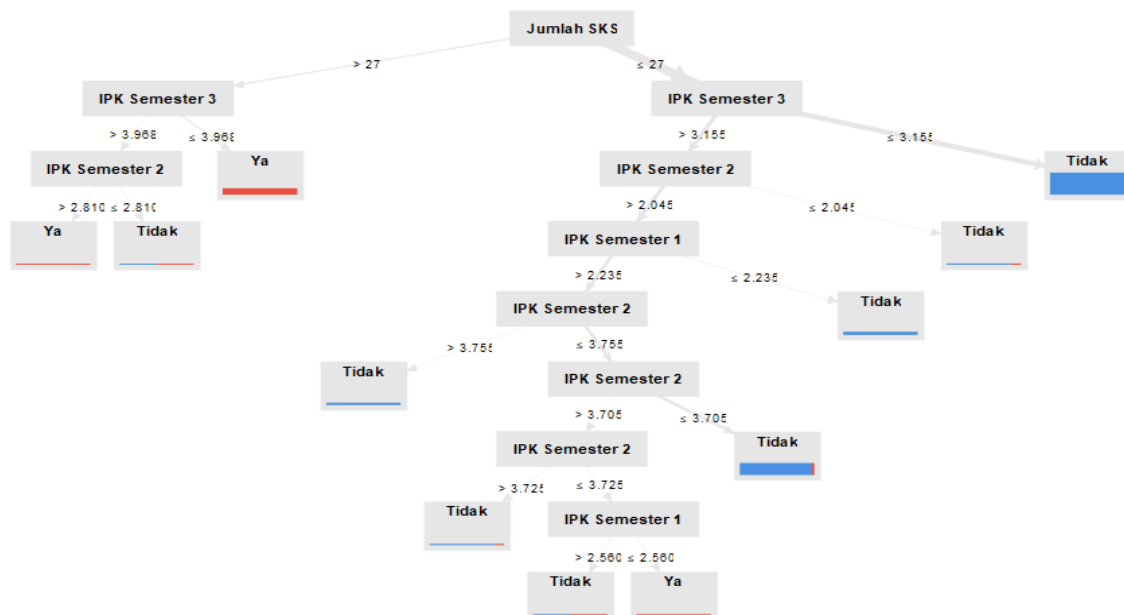
accuracy: 97.00%

	true Tidak	true Ya	class precision
pred. Tidak	171	6	96.61%
pred. Ya	0	23	100.00%
class recall	100.00%	79.31%	

Gambar 2. Hasil Confusional Matrix

Model berhasil mengklasifikasikan 171 data mahasiswa yang tergolong tidak lulus menjadi mahasiswa berprestasi, serta 23 data mahasiswa yang tergolong lulus menjadi mahasiswa berprestasi. Terdapat 6 data mahasiswa yang sebenarnya lulus menjadi mahasiswa berprestasi, namun diklasifikasikan sebagai tidak lulus menjadi mahasiswa berprestasi, sedangkan tidak ditemukan kesalahan klasifikasi terhadap mahasiswa yang seharusnya tidak lulus, namun diklasifikasikan lulus. Dengan demikian, model tidak menghasilkan nilai *false negative* untuk kelas tidak lulus.

Akurasi dari keseluruhan model mencapai hingga 97,00%, yang artinya bahwa algoritma C4.5 memiliki kemampuan klasifikasi yang sangat baik terhadap dataset yang digunakan. Nilai precision untuk kelas Tidak adalah sebesar 96,61%, sedangkan untuk kelas Ya mencapai 100,00%, sehingga seluruh prediksi mahasiswa Ya oleh model bersifat benar. Sementara itu, nilai *recall* untuk kelas Tidak adalah 100,00%, menunjukkan bahwa semua mahasiswa yang benar-benar Tidak terpilih menjadi mahasiswa berprestasi berhasil diidentifikasi oleh model. *Recall* untuk kelas Ya adalah 79,31%, yang mengindikasikan bahwa sebagian besar mahasiswa yang seharusnya YA berhasil dikenali dengan benar, sehingga mahasiswa tersebut terpilih menjadi mahasiswa berprestasi.



Gambar 3. Hasil Decision Tree

Gambar diatas menunjukkan struktur pohon keputusan (*Decision Tree*) yang dihasilkan oleh algoritma C4.5 untuk memprediksi terpilihnya mahasiswa berprestasi ke dalam dua kategori, yaitu Ya (terpilih) dan Tidak (tidak terpilih). Pohon keputusan ini

dibangun berdasarkan sejumlah atribut prediktor, antara lain jumlah sks dan IPK semester 1-3. Node awal pada pohon keputusan menunjukkan bahwa mahasiswa dengan jumlah SKS kurang atau sama dengan 27 cenderung diklasifikasikan sebagai “Tidak Berprestasi”. Sebaliknya, mahasiswa yang mengambil jumlah SKS lebih dari 27 diarahkan pada evaluasi IPK Semester untuk penentuan klasifikasi berikutnya. Pada cabang kiri pohon keputusan, mahasiswa dengan IPK Semester lebih dari 3,968 langsung dikategorikan sebagai “Berprestasi”. Sementara itu, mahasiswa dengan IPK Semester berada di bawah 3,968 namun di atas 2,810 tetap diklasifikasikan sebagai “Berprestasi”, sedangkan IPK Semester kurang dari atau sama dengan 2,810 mengarah pada kategori “Tidak Berprestasi”.

Adapun cabang kanan dari pohon menunjukkan bahwa meskipun mahasiswa mengambil lebih dari 27 SKS, mereka akan dikategorikan sebagai “Tidak Berprestasi” jika memiliki IPK Semester kurang dari atau sama dengan 3,155. Bahkan ketika IPK Semester lebih besar dari 3,155, klasifikasi tetap berlanjut melalui sejumlah batas nilai IPK yang menunjukkan bahwa ketidakstabilan atau penurunan nilai IPK dari semester ke semester menyebabkan mahasiswa cenderung tidak termasuk dalam kategori berprestasi. Mahasiswa baru akan dikategorikan sebagai “Berprestasi” apabila memiliki rentang nilai IPK Semester yang stabil tinggi, terutama lebih besar dari 3,725 dan IPK akhir lebih besar dari 2,580. Secara umum, hasil ini memperlihatkan bahwa mahasiswa berprestasi cenderung memiliki jumlah SKS yang tinggi, disertai nilai IPK yang konsisten tinggi dan tidak mengalami penurunan signifikan, sedangkan mahasiswa dengan jumlah SKS rendah atau IPK yang fluktuatif cenderung diklasifikasikan sebagai tidak berprestasi. Beberapa pola yang dapat diamati dari pohon keputusan (*Decision Tree*) antara lain sebagai berikut:

1. Mahasiswa dengan  $SKS \leq 27$  akan diprediksi tidak lulus menjadi mahasiswa berprestasi.
2. Mahasiswa dengan  $SKS > 27$  &  $IPK > 3.968$  akan diprediksi lulus menjadi mahasiswa berprestasi.
3. Mahasiswa dengan  $SKS > 27$  &  $IPK$  antara 2.810 – 3.968 akan diprediksi lulus menjadi mahasiswa berprestasi.
4. Mahasiswa dengan  $SKS > 27$  &  $IPK$  tidak stabil / cenderung menurun akan diprediksi tidak lulus menjadi mahasiswa berprestasi.
5. Mahasiswa dengan  $SKS > 27$  &  $IPK$  tinggi konsisten (terutama  $> 3.725$ ) akan diprediksi lulus menjadi mahasiswa berprestasi.

Pohon ini menyajikan struktur klasifikasi yang transparan dan dapat ditelusuri kembali logika keputusannya. Warna biru menunjukkan dominasi prediksi kelas Ya, sedangkan warna merah menunjukkan kelas Tidak pada setiap daun (*leaf node*). Dari struktur tersebut dapat disimpulkan bahwa nilai IPK per semester (1-3) dan jumlah SKS merupakan atribut yang sangat berpengaruh dalam menentukan terpilihnya seorang mahasiswa menjadi mahasiswa berprestasi.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma C4.5 efektif dan dapat digunakan dalam memodelkan dan memprediksi faktor-faktor dengan akurasi model C4.5 (97,00%), yang mempengaruhi terpilihnya seorang mahasiswa menjadi mahasiswa berprestasi. Implementasi model semacam ini dapat dimanfaatkan oleh pihak universitas atau institusi perguruan tinggi untuk melakukan identifikasi dini terhadap mahasiswa yang memiliki potensi menjadi mahasiswa yang berprestasi kedepannya, serta merancang intervensi akademik secara tepat sasaran.

## Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa algoritma C4.5 dapat digunakan untuk pengembangan sistem pendukung keputusan di lingkungan akademik, khususnya dalam proses identifikasi dan pembinaan mahasiswa berprestasi. Dengan tingkat akurasi sebesar 97%, model klasifikasi berbasis algoritma C4.5 memberikan bukti kuat bahwa data akademik seperti IPK pada semester awal dan jumlah SKS dapat menjadi indikator yang sangat efektif dalam memprediksi potensi prestasi mahasiswa. Hal ini membuka peluang bagi institusi pendidikan untuk menerapkan sistem deteksi dini yang mampu memonitor perkembangan mahasiswa sejak semester awal, sehingga intervensi seperti bimbingan akademik, pendampingan belajar, atau program pengembangan diri dapat diberikan lebih cepat dan tepat sasaran. Selain itu, pemanfaatan algoritma ini dapat membantu perguruan tinggi dalam menyusun strategi peningkatan mutu akademik berdasarkan data yang objektif dan terukur.

Untuk penelitian selanjutnya, dapat disarankan agar menggunakan cakupan variabel yang lebih luas, tidak hanya berfokus pada data akademik, tetapi juga mencakup faktor non-akademik seperti motivasi belajar, keaktifan organisasi, prestasi non-formal, maupun kondisi psikologis mahasiswa, sehingga model prediksi dapat menjadi lebih komprehensif. Penggunaan dataset yang lebih besar serta pengujian dengan metode data mining lain seperti Random Forest, Gradient Boosting, atau Neural Network juga dapat memberikan perbandingan performa yang lebih luas. Selain itu, implementasi model dalam bentuk aplikasi atau dashboard interaktif akan memberikan manfaat praktis bagi pihak kampus untuk memantau dan mengambil keputusan secara real time. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademik, tetapi juga dapat dikembangkan menjadi solusi praktis yang berkelanjutan bagi manajemen pendidikan tinggi.

## Daftar Pustaka

- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29. <https://doi.org/10.33365/jtk.v16i2.1525>
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan

- Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Azzahra, S. P., Apriyanto, Y. A., & Wijaya, A. (2023). *Analisis Dan Perancangan Data Warehouse Untuk Pengelolaan Stok Barang Pada Cv Aneka Artha Niaga*. 01(03), 103–112.
- Berkualitas, P., Berdaya, D. A. N., & Global, S. (2025). 1, 2 1,2. 10.
- Diki Wahyudi, M. (2023). *Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan Buku Article History*. 1(1), 1–6.
- Hasibuan, T. H., & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *Skanika*, 6(1), 61–74. <https://doi.org/10.36080/skanika.v6i1.2976>
- Ismanto, E., & Novalia, M. (2021). Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification. *Techno.COM*, 20(3), 400–410.
- Khasanah, N., Saputri, D. U. E., Hidayat, T., & Aziz, F. (2025). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 dengan RapidMiner: Studi Kasus Data Akademik Perguruan Tinggi XYZ. *Indonesian Journal Computer Science*, 4(2), 100–107. <https://doi.org/10.31294/ijcs.v4i2.9647>
- Mantopani, R., Ramadhan, I. L., Samsara, R. P., & Samsinar, R. (2024). *Analisis Perbandingan Decision Tree Dan Algoritma C4 . 5 Untuk Mengklasifikasikan Penerimaan Mahasiswa Elektro*. 295–302.
- Meliawati, K. (2020). Kolerasi Antara Self Esteem Dengan Prestasi Mahasiswa Program Studi Pendidikan Bahasa Inggris. 25(3), 422–430.
- Muhammad Hamdan Sukri, & Handrianto, Y. (2024). Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta. *Informatics and Computer Engineering Journal*, 4(1), 11–24. <https://doi.org/10.31294/icej.v4i1.2582>
- Nahjan, M. R., Heryana, N., Voutama, A., Komputer, F. I., Karawang, U. S., & Miner, R. (2023). *IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS*

UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL. 7(1), 101–104.

- Noverta Effendi\*1, Witrihan Ramadhani2, Fitri Farida3, M. D. (2024). Jurnal Computer Science and Information Technology ( CoSciTech ) things. *Jurnal Computer Science and Information Technology (CoSciTech)*, 5(2), 358–366.
- Rachardian, S., Sedyono, E., Informasi, F. T., Kristen, U., Wacana, S., Tengah, J., & Artikel, R. (2024). *Prediksi kelulusan tepat waktu mahasiswa untuk pemantauan program studi menggunakan metode data mining*. 21(2), 168–182.
- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, 3(2), 1860–1874. <https://doi.org/10.30997/karimahtauhid.v3i2.11952>
- Rofiani, Riska, Luluk Oktaviani2, D. V., & Hendriawan, T. (2024). Penerapan Metode Klasifikasi Decision Tree dalam Prediksi. *Jurnal TEKNO KOMPAK*, 18(1), 126–139.
- Sari, M., Windarto, A. P., & Okprana, H. (2021). *BEES : Bulletin of Electrical and Electronics Engineering Penerapan Data Mining Klasifikasi C4 . 5 Pada Penerima Beasiswa di*. 1(3), 115–121.
- Sihombing, I. A., Hartama, D., Parlina, I., Gunawan, I., & Kirana, I. O. (2021). *Analisis Keberhasilan Pembelajaran Daring pada Masa Pandemi Covid-19 menggunakan Algoritma C4 . 5 dan Naive Bayes*. 3(November), 89–96.
- Steffany, A., Aldzikri, K. A., Tricahya, M. A. R., & Arfiandono, M. I. (2023). Penerapan Algoritma C4.5 Untuk Menganalisis Kepuasan E-Commerce Shopee. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 6(6), 700–706. <https://doi.org/10.32672/jnkti.v6i6.7112>
- Sumual, I. M., Supriadi, J., Effendy, E., & Wijaya, A. (n.d.). *KLASIFIKASI KATEGORI PRODUK TERLARIS PADA E-COMMERCE*. 5(2), 51–60.
- Syahriani, S. (2022). *Penerapan Data Mining Untuk Menentukan Pola Penjualan Sepatu Menggunakan Metode Algoritma Apriori*. 9(1), 43–52.
- Tabuni, N., Kristen, U., Wacana, S., Satyawati, S. T., Kristen, U., Wacana, S., Ismanto, B.,

- 
- Kristen, U., & Wacana, S. (2024). *Evaluasi Program Beasiswa Bagi Mahasiswa Kabupaten Pegunungan Bintang Menggunakan Model Goal Oriented Evaluation*.
- Tahsinia, J., Zulfa, A. A., Ibrahim, T., & Arifudin, O. (2025). *PERAN SISTEM INFORMASI AKADEMIK BERBASIS WEB*. 6(1), 115–134.
- Tamimi, A., Octafriyanda, D., Hutagaol, R., & Yuhdi, A. (2024). *Merdeka Belajar : Mewujudkan Pendidikan Berkualitas Melalui Kampus Merdeka di Universitas Negeri Medan*. *Merdeka Belajar : Realizing Quality Education Through the Merdeka Campus at Medan State University*. 3183–3190.
- Widiastuti, T., Karsa, K., & Juliane, C. (2022). *Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5*. *Technomedia Journal*, 7(3), 364–380. <https://doi.org/10.33050/tmj.v7i3.1932>
- Widiastuti, T., Karsa, K., & Juliane, C. (2022). *Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5*. *Technomedia Journal*, 7(3), 364–380. <https://doi.org/10.33050/tmj.v7i3.1932>