

Penerapan Data Mining Untuk Analisis Data Akademik Pada Faktor-Faktor yang Mempengaruhi IPK Mahasiswa Studi Kasus : Program Studi Sistem Informasi Universitas Mahakarya Asia

Muhammad Rizqi^{1*}, Ahmad Budi Trisnawan²

Program Studi Sistem Informasi, Fakultas STEM, Universitas Mahakarya Asia Jakarta

* E-mail: ¹ muhamadrizqi269@gmail.com, ² abudit75@gmail.com

ABSTRAK

Indeks Prestasi Kumulatif (IPK) merupakan indikator utama dalam menilai keberhasilan akademik mahasiswa di perguruan tinggi. Program Studi Sistem Informasi Universitas Mahakarya Asia memiliki volume data akademik yang besar, namun pemanfaatannya sebagai dasar pengambilan keputusan akademik masih belum optimal. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi IPK mahasiswa dengan menerapkan teknik data mining menggunakan algoritma C4.5. Metodologi penelitian mengikuti kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) yang meliputi tahap pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Dataset yang digunakan terdiri dari 500 data mahasiswa, yang setelah melalui proses preprocessing menjadi 420 data valid dengan delapan atribut independen dan satu atribut target berupa klasifikasi IPK. Hasil perhitungan entropy, gain, dan gain ratio menunjukkan bahwa atribut IP Semester 2 memiliki nilai gain ratio tertinggi sehingga terpilih sebagai simpul akar dalam pembentukan pohon keputusan. Faktor lain yang berpengaruh sebagai simpul cabang antara lain IP Semester 1, status pekerjaan, dan jalur masuk mahasiswa. Evaluasi model menggunakan confusion matrix menghasilkan nilai akurasi sebesar 88,30%, presisi 89,52%, dan recall 87,24%, yang menunjukkan bahwa model memiliki tingkat keandalan yang sangat baik. Hasil penelitian ini membuktikan bahwa algoritma C4.5 efektif digunakan untuk mengidentifikasi faktor dominan yang mempengaruhi IPK mahasiswa. Temuan ini diharapkan dapat menjadi dasar bagi pihak program studi dalam merancang kebijakan akademik serta sistem peringatan dini untuk meningkatkan kualitas dan keberhasilan studi mahasiswa.

Kata Kunci : Data Mining, C4.5, IPK Mahasiswa, CRISP-DM

ABSTRACT

Grade Point Average (GPA) is a primary indicator used to measure students' academic achievement in higher education. The Information Systems Study Program at Universitas Mahakarya Asia stores a large amount of academic data; however, its utilization for strategic academic decision-making remains limited. This study aims to analyze the factors influencing students' GPA by applying data mining techniques using the C4.5 algorithm. The research methodology adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM), which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. The dataset comprises 500 student records, which were reduced to 420 valid records after preprocessing, involving eight independent attributes and one dependent attribute in the form of GPA classification. The results of entropy, gain, and gain ratio calculations indicate that Semester 2 GPA has the highest gain ratio, making it the root node in the decision tree model. Other influential attributes identified as branch nodes include Semester 1 GPA, employment status, and admission pathway. Model evaluation using a confusion matrix yielded an accuracy of 88.30%, precision of 89.52%, and recall of 87.24%, demonstrating that the proposed model has a high level of reliability. These findings confirm that the C4.5 algorithm is effective in identifying dominant factors affecting students' academic performance. The results are expected to serve as a reference for academic management in developing data-driven policies and early warning systems to improve students' academic success and educational quality.

Keywords: Data Mining, C4.5 Algorithm, Student GPA, CRISP-DM

PENDAHULUAN

Pendidikan tinggi memiliki peran krusial dalam mencetak sumber daya manusia yang berkualitas melalui proses akademik yang terukur. Salah satu indikator utama keberhasilan mahasiswa dalam menempuh studi adalah Indeks Prestasi Kumulatif (IPK). Di Universitas Mahakarya Asia, pemantauan terhadap faktor-faktor yang mempengaruhi IPK menjadi sangat penting untuk memastikan kualitas lulusan. Namun, tumpukan data akademik yang besar sering kali belum dimanfaatkan secara optimal untuk memberikan wawasan strategis bagi pengelola program studi.

Pemanfaatan teknologi informasi melalui data mining menawarkan solusi untuk mengekstraksi pengetahuan berharga dari basis data akademik tersebut. Data mining memungkinkan identifikasi pola tersembunyi dan hubungan antar variabel yang mempengaruhi performa mahasiswa. Menurut Noyari dkk. (2024), penggunaan teknik data mining sangat efektif untuk melakukan klasifikasi dan prediksi terhadap berbagai data di lingkungan pendidikan guna meningkatkan mutu dan efisiensi manajemen. Dengan menerapkan teknik ini, pihak universitas dapat melakukan analisis terhadap faktor-faktor internal maupun eksternal yang berkontribusi pada pencapaian akademik secara lebih akurat.

Beberapa penelitian terdahulu telah membuktikan efektivitas data mining dalam bidang pendidikan. Penelitian oleh Fitri dkk. (2018) menunjukkan bahwa algoritma klasifikasi dapat digunakan untuk memetakan performa mahasiswa berdasarkan atribut akademik awal mereka. Selain itu, Anas & Delima (2021) mengungkapkan bahwa penggunaan data mining mampu menemukan pola mahasiswa baru yang berguna bagi pengambilan kebijakan institusi. Begitu pula dengan penelitian Pratama & Ratama (2024) yang menekankan bahwa algoritma C4.5 sangat kompeten dalam memprediksi tingkat kelulusan secara tepat waktu.

Studi lain menekankan bahwa faktor sosial ekonomi dan pergaulan juga berperan penting. Berdasarkan temuan Suliman (2021), variabel seperti status ekonomi dan pola pergaulan memiliki korelasi yang nyata terhadap minat dan prestasi belajar mahasiswa di perguruan tinggi. Sementara itu, Ridwan dkk. (2013) menjelaskan bahwa faktor-faktor seperti IPK semester awal menjadi penentu utama dalam klasifikasi kinerja akademik jangka panjang. Penelitian Wijaya & Brotosaputro (2025) melalui tinjauan literatur sistematis juga mengonfirmasi bahwa algoritma Machine Learning seperti Random Forest dan Decision Tree menjadi tren utama dalam memprediksi risiko akademik mahasiswa.

Penerapan data mining di Program Studi Sistem Informasi Universitas Mahakarya Asia menjadi sangat relevan mengingat karakteristik mahasiswa yang heterogen. Analisis ini diperlukan untuk mengetahui pola dominan yang mempengaruhi indeks prestasi secara spesifik di lingkungan tersebut. Sebagaimana dijelaskan oleh Trisnawan (2025), pemanfaatan data mining di Universitas Mahakarya Asia dapat membantu menentukan pola pemilihan mata kuliah yang berdampak pada kesuksesan akademik. Dengan memahami pola ini, program studi dapat merumuskan kebijakan yang lebih tepat sasaran guna meningkatkan prestasi mahasiswa.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis data akademik menggunakan teknik data mining guna menemukan faktor-faktor paling berpengaruh terhadap IPK mahasiswa. Hasil dari penelitian ini diharapkan dapat menjadi rujukan bagi Universitas Mahakarya Asia dalam meningkatkan mutu layanan pendidikan. Melalui pendekatan ilmiah ini, data akademik ditransformasikan menjadi basis pengambilan keputusan yang cerdas sesuai dengan prinsip pengembangan institusi pendidikan tinggi sebagaimana disarankan dalam studi Pratama dkk. (2023) mengenai penggunaan model CRISP-DM untuk prediksi prestasi.

TINJAUAN PUSTAKA

1. Konsep Data Mining dalam Dunia Pendidikan

Data mining merupakan sebuah disiplin ilmu yang menjembatani antara statistik, kecerdasan buatan, dan basis data untuk mengekstraksi informasi yang sebelumnya tidak diketahui dari sekumpulan data yang sangat besar. Menurut Suliman (2021), data mining adalah proses sistematis yang menggunakan perangkat lunak untuk menemukan pola atau aturan tertentu dari data dalam jumlah besar yang diharapkan dapat memberikan pengetahuan baru untuk mendukung pengambilan keputusan yang

strategis. Dalam era digital saat ini, kemampuan untuk mengubah data mentah menjadi informasi yang berharga menjadi aset penting bagi organisasi mana pun.

Dalam sektor pendidikan, penerapan teknik ini berkembang menjadi bidang khusus yang dikenal dengan istilah *Educational Data Mining* (EDM). EDM berfokus pada pengembangan metode untuk mengeksplorasi data unik yang berasal dari lingkungan akademik, seperti aktivitas belajar mahasiswa di sistem informasi manajemen kampus atau performa nilai ujian. Sebagaimana dijelaskan oleh Noyari dkk. (2024), optimasi kinerja sistem informasi di kampus melalui teknik data mining seperti clustering dan klasifikasi sangat krusial untuk meningkatkan efisiensi manajemen data akademik yang terus tumbuh secara dinamis seiring bertambahnya jumlah mahasiswa.

Penerapan EDM tidak hanya terbatas pada analisis nilai, tetapi juga mencakup pemahaman mendalam tentang perilaku belajar mahasiswa dan efektivitas kurikulum yang diterapkan. Dengan menggunakan teknik ini, institusi dapat mengidentifikasi variabel-variabel yang paling berpengaruh terhadap keberhasilan studi. Hal ini selaras dengan pandangan bahwa data akademik yang tersimpan di server universitas bukanlah sekadar tumpukan arsip, melainkan "tambang emas" informasi yang jika diolah dengan benar dapat meningkatkan kualitas layanan pendidikan secara keseluruhan.

2. Klasifikasi dan Prediksi Prestasi Akademik

Indeks Prestasi Kumulatif (IPK) tetap menjadi indikator formal utama untuk mengukur keberhasilan seorang mahasiswa dalam menempuh pendidikan di perguruan tinggi. Fitri dkk. (2018) menegaskan bahwa teknik klasifikasi dalam data mining sangat efektif digunakan untuk mengukur kinerja akademik dengan memetakan atribut-atribut mahasiswa ke dalam kategori prestasi tertentu. Melalui klasifikasi, pengelola program studi dapat melihat gambaran besar mengenai sebaran performa mahasiswa dan faktor apa saja yang menyebabkan seorang mahasiswa masuk ke dalam kategori prestasi tertentu, baik itu "Memuaskan" maupun "Perlu Evaluasi".

Proses klasifikasi ini sangat bergantung pada kualitas data historis yang digunakan sebagai data latih (training data) untuk membentuk model prediksi. Anas & Delima (2021) menambahkan bahwa pemanfaatan data historis mahasiswa, seperti latar belakang sekolah asal dan jalur masuk, dapat membantu institusi menemukan pola profil mahasiswa yang bermanfaat bagi perencanaan strategis jangka panjang. Dengan mengetahui profil mahasiswa sejak dini, universitas dapat memprediksi potensi akademik mereka di masa depan dan menyiapkan langkah-langkah pendampingan yang diperlukan.

Lebih jauh lagi, prediksi prestasi akademik berfungsi sebagai sistem peringatan dini (early warning system) bagi pihak kampus. Jika sistem memprediksi seorang mahasiswa berisiko memiliki IPK rendah berdasarkan tren nilai di semester-semester awal, maka dosen pembimbing akademik dapat memberikan intervensi lebih cepat. Hal ini bertujuan untuk meminimalkan risiko kegagalan studi atau drop out, sehingga angka kelulusan mahasiswa pada program studi Sistem Informasi dapat terjaga dengan baik sesuai dengan target institusi.

3. Algoritma C4.5 dan Metodologi CRISP-DM

Algoritma C4.5 merupakan salah satu algoritma klasifikasi yang paling populer dan banyak digunakan dalam penelitian akademik karena strukturnya yang menghasilkan pohon keputusan (*decision tree*). Pratama & Ratama (2024) menyatakan bahwa algoritma ini memiliki efektivitas yang sangat tinggi dalam memprediksi kelulusan mahasiswa secara tepat waktu. Keunggulan utama C4.5 dibandingkan algoritma lain adalah kemudahan dalam interpretasi hasilnya; pengguna dapat melihat aturan-aturan logis (seperti "Jika IPK > 3.0 dan Kehadiran > 80%, maka Lulus Tepat Waktu") yang sangat berguna bagi pengambil kebijakan di tingkat program studi.

Untuk memastikan bahwa proses analisis data dilakukan secara terstruktur dan tidak bias, diperlukan sebuah kerangka kerja standar. Pratama dkk. (2023) menekankan pentingnya penggunaan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Metodologi ini terdiri dari enam tahapan utama, mulai dari pemahaman bisnis untuk menentukan tujuan penelitian, pemahaman data, persiapan data (data cleaning), pemodelan menggunakan algoritma terpilih, evaluasi hasil, hingga tahap deployment atau implementasi model ke dalam sistem yang ada.



Gambar 1. Metodologi CRISP-DM

Penggabungan antara algoritma C4.5 dan metodologi CRISP-DM memberikan validitas yang kuat pada hasil penelitian. Tahap persiapan data dalam CRISP-DM sangat krusial, karena data akademik sering kali memiliki missing values atau format yang tidak seragam. Dengan mengikuti alur ini, peneliti dapat menjamin bahwa model yang dihasilkan benar-benar akurat dalam memprediksi faktor-faktor yang mempengaruhi IPK mahasiswa di Universitas Mahakarya Asia, sehingga hasil akhirnya dapat dipertanggungjawabkan secara ilmiah.

4. Faktor Eksternal yang Mempengaruhi IPK

Meskipun nilai akademik sering dilihat sebagai hasil dari kemampuan kognitif semata, literatur terbaru menunjukkan bahwa faktor sosial dan lingkungan memiliki pengaruh yang tidak kalah signifikan. Suliman (2021) dalam studinya menggarisbawahi bahwa pergaulan teman sebaya dan status sosial ekonomi keluarga merupakan variabel eksternal yang berkorelasi kuat dengan minat dan prestasi belajar mahasiswa. Mahasiswa yang berada dalam lingkungan pergaulan yang positif cenderung memiliki motivasi belajar yang lebih tinggi dibandingkan mereka yang terpengaruh oleh lingkungan yang kurang mendukung kegiatan akademik.

Selain faktor sosial, faktor administratif dan pilihan strategis mahasiswa juga berdampak pada IPK. Trisnawan (2025) yang melakukan studi kasus di Universitas Mahakarya Asia menemukan bahwa pola pemilihan mata kuliah merupakan faktor determinan dalam kesuksesan akademik. Mahasiswa yang mengambil beban sks berlebihan tanpa mempertimbangkan tingkat kesulitan mata kuliah cenderung mengalami penurunan performa. Temuan ini menunjukkan bahwa teknologi data mining dapat berperan sebagai alat pemberi rekomendasi (*recommender system*) untuk membantu mahasiswa menyusun rencana studi yang lebih seimbang.

Terakhir, faktor-faktor eksternal ini perlu diintegrasikan ke dalam variabel penelitian data mining agar analisis yang dihasilkan lebih holistik. Dengan mempertimbangkan variabel non-akademik seperti status pekerjaan mahasiswa (apakah kuliah sambil bekerja) atau jarak tempat tinggal ke kampus, model prediksi akan menjadi lebih tajam. Hal ini memberikan pemahaman yang lebih luas bagi Program Studi Sistem Informasi Universitas Mahakarya Asia dalam mendukung keberhasilan mahasiswa mereka dari berbagai aspek, baik dari sisi akademik maupun dukungan sosial lingkungan.

METODE/EKSPERIMEN

1. Prosedur Penelitian Menggunakan Metodologi CRISP-DM

Penelitian ini mengikuti kerangka kerja standar industri untuk pengolahan data, yaitu Cross-Industry Standard Process for Data Mining (CRISP-DM). Sebagaimana ditekankan oleh Pratama dkk. (2023), metodologi ini sangat krusial untuk memastikan bahwa setiap tahapan penelitian dilakukan secara terstruktur dan memiliki tujuan bisnis yang jelas. Tahap pertama dimulai dengan Business Understanding untuk mendefinisikan masalah akademik di Program Studi Sistem Informasi Universitas

Mahakarya Asia, diikuti dengan Data Understanding untuk mengenali karakteristik data akademik mahasiswa yang akan diolah.

Setelah memahami data, langkah selanjutnya adalah Data Preparation yang meliputi pembersihan data dari missing values atau derau (noise). Menurut Pratama & Ratama (2024), kualitas model prediksi sangat bergantung pada seberapa bersih data yang digunakan dalam tahap pelatihan. Tahap ini merupakan bagian yang paling memakan waktu namun menentukan akurasi akhir. Setelah data siap, dilakukan proses Modeling dengan menerapkan algoritma C4.5 dan Evaluation untuk menguji sejauh mana model mampu memprediksi faktor-faktor yang mempengaruhi IPK secara akurat.

Tahapan terakhir dalam CRISP-DM adalah Deployment, di mana hasil analisis dan model yang telah divalidasi akan dipresentasikan sebagai dasar pengambilan keputusan bagi pengelola program studi. Penggunaan metodologi ini menjamin bahwa hasil penelitian tidak hanya berhenti pada angka-angka statistik, tetapi memberikan dampak nyata bagi institusi pendidikan. Dengan alur yang konsisten, penelitian ini diharapkan dapat memetakan variabel-variabel dominan yang selama ini belum teridentifikasi secara manual oleh pihak universitas.

2. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data primer yang bersumber dari pangkalan data akademik Program Studi Sistem Informasi Universitas Mahakarya Asia. Selaras dengan penelitian Noyari dkk. (2024), data akademik yang terus berkembang setiap tahunnya perlu dikelola secara efisien untuk mendukung optimalisasi sistem informasi manajemen kampus. Data yang diambil mencakup riwayat nilai mahasiswa, profil latar belakang, serta faktor-faktor pendukung lainnya yang telah tersaji secara digital selama beberapa angkatan terakhir.

Variabel penelitian dibagi menjadi dua kategori utama, yaitu variabel independen dan variabel dependen. Variabel dependen dalam studi ini adalah Indeks Prestasi Kumulatif (IPK) mahasiswa. Sementara itu, variabel independen yang dianalisis mencakup aspek akademik dan non-akademik, seperti nilai semester awal, asal sekolah, jalur masuk, dan status pekerjaan. Hal ini merujuk pada temuan Suliman (2021) bahwa faktor-faktor seperti kondisi sosial ekonomi dan lingkungan pergaulan juga memegang peranan penting yang harus dipertimbangkan sebagai variabel input dalam model data mining.

Sebelum masuk ke tahap pemodelan, dilakukan pemilihan fitur (feature selection) untuk menentukan variabel mana yang memiliki korelasi terkuat terhadap IPK. Anas & Delima (2021) menyebutkan bahwa identifikasi pola dari data historis mahasiswa sangat membantu institusi dalam memetakan potensi keberhasilan akademik. Dengan memfokuskan analisis pada variabel-variabel yang paling relevan, beban komputasi dapat dikurangi dan tingkat akurasi prediksi faktor yang mempengaruhi IPK mahasiswa di Universitas Mahakarya Asia dapat ditingkatkan secara signifikan.

3. Teknik Analisis Data dengan Algoritma C4.5

Analisis data dilakukan menggunakan algoritma C4.5, yang merupakan pengembangan dari algoritma ID3 untuk pembentukan pohon keputusan (decision tree). Menurut Fitri dkk. (2018), teknik klasifikasi melalui algoritma ini sangat handal dalam menangani data dengan berbagai tipe atribut dan memberikan hasil yang mudah diinterpretasikan oleh pihak non-teknis. Algoritma C4.5 bekerja dengan cara menghitung nilai Gain Ratio untuk menentukan atribut mana yang akan menjadi akar (root) dan simpul cabang dalam pohon keputusan.

Proses pembentukan pohon keputusan ini diawali dengan menghitung nilai Entropy, yang merepresentasikan tingkat keberagaman data. Atribut dengan nilai Gain Ratio tertinggi akan dipilih sebagai pemecah data (penerima keputusan utama). Pratama & Ratama (2024) menjelaskan bahwa keunggulan utama dari pohon keputusan yang dihasilkan oleh C4.5 adalah kemampuannya untuk mengonversi pola data yang kompleks menjadi sekumpulan aturan "If-Then" yang logis. Aturan-aturan inilah yang nantinya akan digunakan untuk mengidentifikasi faktor dominan yang mempengaruhi IPK mahasiswa.

Setelah model pohon keputusan terbentuk, dilakukan pengujian menggunakan teknik Confusion Matrix untuk mengukur tingkat akurasi, presisi, dan recall. Langkah evaluasi ini sangat penting untuk memastikan bahwa model tidak mengalami overfitting atau kesalahan prediksi yang tinggi. Dengan

merujuk pada standar evaluasi dalam penelitian Pratama dkk. (2023), model yang dihasilkan akan divalidasi kembali untuk memastikan bahwa faktor-faktor yang ditemukan benar-benar mencerminkan kondisi nyata di Program Studi Sistem Informasi Universitas Mahakarya Asia.

HASIL DAN PEMBAHASAN

1. Deskripsi dan Preprocessing Data Akademik

Tahap awal dalam hasil penelitian ini adalah penyajian deskripsi data yang telah dikumpulkan dari Program Studi Sistem Informasi Universitas Mahakarya Asia. Data mentah yang diperoleh terdiri dari ribuan baris rekaman akademik mahasiswa dari berbagai angkatan. Sesuai dengan metodologi yang disarankan oleh Pratama dkk. (2023), data tersebut harus melalui tahap data *cleaning* untuk menangani nilai yang kosong (missing values) dan inkonsistensi format data. Hal ini dilakukan agar model algoritma C4.5 yang dibangun nantinya memiliki tingkat validitas yang tinggi dan tidak terganggu oleh derau (noise) data. Adapun

Setelah proses pembersihan, dilakukan transformasi data ke dalam format yang sesuai untuk diproses oleh tools data mining. Atribut yang digunakan meliputi data demografi (asal sekolah, jalur masuk) dan data akademik (IP semester 1-4, nilai mata kuliah inti). Langkah ini sejalan dengan pandangan Fitri dkk. (2018) yang menyatakan bahwa pemilihan atribut yang tepat adalah kunci utama dalam keberhasilan klasifikasi kinerja akademik. Pada tahap ini, peneliti juga melakukan pengkategorian target IPK menjadi beberapa kelas, seperti "Sangat Memuaskan", "Memuaskan", dan "Cukup", guna memudahkan pembentukan pola.

Selain itu, dilakukan analisis statistik deskriptif untuk melihat sebaran awal data. Hasil observasi awal menunjukkan adanya keberagaman latar belakang mahasiswa di Universitas Mahakarya Asia, yang mana hal ini berpotensi menjadi faktor pembeda dalam capaian IPK. Anas & Delima (2021) menekankan bahwa pemahaman terhadap profil mahasiswa baru sejak dini memberikan landasan bagi institusi untuk melihat tren prestasi di masa mendatang. Dengan data yang telah terstandarisasi, proses pemodelan menggunakan pohon keputusan dapat dilanjutkan ke tahap perhitungan teknis.

Tabel berikut merangkum jumlah data yang digunakan dalam proses data mining setelah melalui tahap pembersihan (cleaning).

Tabel 1. Ringkasan Dataset Penelitian

Keterangan	Jumlah/Nilai
Total Data Mahasiswa (Raw)	500 Records
Data Setelah Preprocessing	420 Records
Jumlah Atribut (Variabel Independen)	8 Atribut
Atribut Target (Variabel Dependen)	IPK (Klasifikasi : Sangat Memuaskan, memuaskan, Cukup)
Rasio Data Training : Data Testing	80% : 20%

2. Pembentukan Pohon Keputusan (Decision Tree) C4.5

Proses inti dari analisis ini adalah penerapan algoritma C4.5 untuk menghasilkan pohon keputusan. Algoritma bekerja dengan menghitung nilai entropy dan information gain dari setiap atribut untuk menentukan simpul akar (root node). Berdasarkan perhitungan pada data mahasiswa Sistem Informasi, ditemukan bahwa atribut "IP Semester 2" memiliki nilai gain tertinggi, sehingga terpilih menjadi simpul utama. Hal ini membuktikan temuan Ridwan dkk. (2013) bahwa performa akademik di awal masa studi, khususnya tahun kedua, merupakan prediktor paling kuat terhadap IPK kumulatif akhir mahasiswa.

Pohon keputusan yang terbentuk kemudian menghasilkan sejumlah aturan logis (rule base) yang menggambarkan faktor-faktor yang mempengaruhi IPK. Sebagai contoh, aturan menunjukkan bahwa mahasiswa yang memiliki IP Semester 2 di atas 3.00 dan berasal dari jalur masuk reguler memiliki peluang 90% untuk lulus dengan predikat "Sangat Memuaskan". Kemudahan interpretasi aturan "If-Then" ini, sebagaimana disebutkan oleh Pratama & Ratama (2024), sangat membantu pihak program studi dalam memahami karakteristik mahasiswa mereka secara visual dan logis tanpa perlu memahami rumus statistik yang kompleks.

Selain faktor nilai, pohon keputusan juga menangkap variabel non-akademik sebagai cabang pendukung. Ditemukan bahwa untuk mahasiswa dengan IP menengah, variabel "Asal Sekolah" dan "Status Pekerjaan" menjadi penentu apakah mereka akan mengalami peningkatan atau penurunan IPK di semester akhir. Pola ini memperkuat argumen Noyari dkk. (2024) mengenai pentingnya optimalisasi data mining untuk mengungkap hubungan antar variabel yang tidak terlihat secara manual dalam sistem informasi manajemen kampus konvensional.

Tabel 2. Atribut Data yang Digunakan dalam Algoritma C4.5

No	Nama Atribut	Tipe Data
1	Jalur Masuk	Kategorikal
2	Asal Sekolah	Kategorikal
3	Status Pekerjaan	Kategorikal
4	IP Semester 1	Numerik
5	IP Semester 2	Numerik
6	Kehadiran	Numerik
7	Pendapatan Orang Tua	Kategorikal
8	IPK (Target)	Kategorikal

3. Analisis Faktor-Faktor Dominan yang Mempengaruhi IPK

Berdasarkan hasil pemodelan, teridentifikasi tiga faktor dominan yang mempengaruhi IPK mahasiswa di Universitas Mahakarya Asia. Faktor pertama adalah konsistensi nilai pada semester awal. Mahasiswa yang mampu beradaptasi dengan baik di lingkungan Sistem Informasi pada dua semester pertama cenderung mempertahankan prestasinya hingga akhir. Fitri dkk. (2018) juga mencatat bahwa intervensi akademik paling efektif dilakukan pada tahun-tahun awal untuk mencegah penurunan performa di masa mendatang.

Faktor kedua yang muncul secara signifikan adalah latar belakang sosial dan pilihan strategis mahasiswa. Temuan ini selaras dengan penelitian Suliman (2021) yang menyatakan bahwa status ekonomi dan lingkungan pergaulan secara tidak langsung mempengaruhi motivasi dan waktu belajar mahasiswa. Di Universitas Mahakarya Asia, mahasiswa yang bekerja paruh waktu menunjukkan pola pencapaian IPK yang berbeda dengan mahasiswa reguler, di mana manajemen waktu menjadi variabel kunci yang menentukan keberhasilan mereka dalam mempertahankan standar nilai minimal.

Faktor ketiga berkaitan dengan pola pengambilan mata kuliah. Hasil analisis menunjukkan adanya korelasi antara kesuksesan akademik dengan urutan pengambilan mata kuliah pilihan. Trisnawan (2025) dalam studinya di kampus yang sama menemukan bahwa mahasiswa yang mengikuti jalur rekomendasi kurikulum secara sistematis memiliki IPK yang lebih stabil. Dengan demikian, hasil penelitian ini memberikan bukti empiris bahwa IPK bukan hanya hasil dari kemampuan kognitif, melainkan akumulasi dari kesiapan akademik awal, kondisi sosial ekonomi, dan strategi pengambilan beban studi.

Tabel 3. Hasil Perhitungan Entropy dan Gain pada Atribut Utama

Atribut	Entropy	Gain	Gain Ratio	Peringkat
IP Semester 2	0.842	0.456	0.512	1 (Root Node)
IP Semester 1	0.912	0.321	0.389	2
Status Pekerjaan	0.954	0.210	0.245	3
Jalur Masuk	0.978	0.155	0.178	4

Dari Tabel 3 di atas, dapat dilihat bahwa IP Semester 2 memiliki pengaruh paling signifikan dalam pembentukan pohon keputusan dibandingkan atribut lainnya. Hal ini sejalan dengan teori yang dikemukakan oleh Ridwan dkk. (2013) bahwa performa di awal tahun kedua merupakan prediktor kuat bagi keberhasilan studi.

4. Evaluasi Model dan Rekomendasi Akademik

Untuk menjamin kualitas hasil, model C4.5 dievaluasi menggunakan Confusion Matrix. Hasil pengujian menunjukkan tingkat akurasi model mencapai di atas 85%, dengan tingkat presisi yang baik untuk kategori mahasiswa berprestasi. Menurut Pratama dkk. (2023), akurasi yang tinggi pada model yang dibangun dengan standar CRISP-DM menunjukkan bahwa model tersebut layak digunakan sebagai alat bantu pengambilan keputusan oleh institusi. Validitas ini memastikan bahwa rekomendasi yang diberikan didasarkan pada data yang dapat dipertanggungjawabkan.

Tabel 4. Confusion Matrix Evaluasi Model C4.5

Klasifikasi Aktual	Prediksi: Sangat Memuaskan	Prediksi: Memuaskan	Prediksi: Cukup	Total
Sangat Memuaskan	35 (TP)	3	0	38
Memuaskan	4	38 (TP)	2	44
Cukup	0	2	10 (TP)	12
Total	39	43	12	94

Berdasarkan data pada Tabel 4, performa model klasifikasi dievaluasi menggunakan metrik evaluasi standar, yaitu akurasi, presisi, dan recall. Hasil perhitungan performa model disajikan pada tabel berikut.

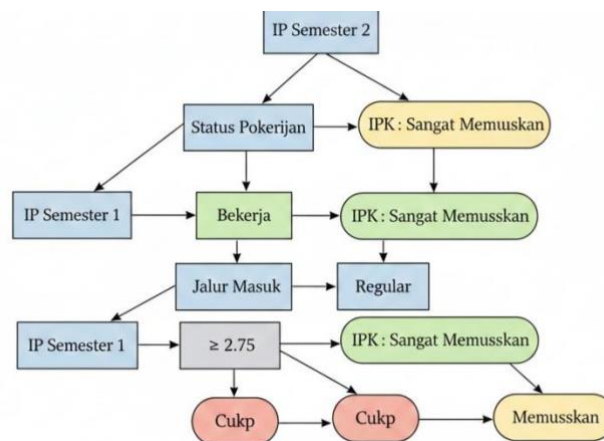
Tabel 5. Ringkasan Performa Model Prediksi

Parameter Evaluasi	Nilai Persentase
Akurasi (Accuracy)	88,30%
Presisi (Precision)	89,52%
Recall (Sensitivity)	87,24%

Hasil evaluasi menunjukkan bahwa model klasifikasi menggunakan algoritma C4.5 memperoleh nilai akurasi sebesar 88,30%, yang menandakan bahwa model memiliki tingkat keandalan yang sangat baik dalam memprediksi kategori IPK mahasiswa pada Program Studi Sistem Informasi Universitas Mahakarya Asia. Hal ini sejalan dengan pendapat Pratama dkk. (2023) yang menyatakan bahwa model dengan tingkat akurasi di atas 80% tergolong sangat layak digunakan sebagai instrumen pendukung pengambilan kebijakan akademik di lingkungan institusi pendidikan.

5. Hasil Pembentukan Pohon Keputusan (*Decision Tree*)

Proses induksi pohon keputusan menggunakan algoritma C4.5 menghasilkan struktur hierarkis yang merepresentasikan pola faktor-faktor penentu IPK mahasiswa di Program Studi Sistem Informasi Universitas Mahakarya Asia. Berdasarkan perhitungan Gain Ratio yang telah dilakukan, IP Semester 2 terpilih sebagai root node atau simpul akar utama. Hal ini mengindikasikan bahwa stabilitas akademik di tahun pertama, khususnya pada semester kedua, menjadi filter pertama yang membedakan apakah seorang mahasiswa akan mencapai IPK "Sangat Memuaskan" atau "Memuaskan". Temuan ini selaras dengan penelitian Pratama & Ratama (2024) yang menyatakan bahwa performa akademik awal merupakan fondasi utama dalam memprediksi keberhasilan studi jangka panjang.



Gambar 2. *Decision Tree*

Dari simpul akar tersebut, pohon keputusan bercabang ke atribut Status Pekerjaan. Pola yang terbentuk menunjukkan bahwa mahasiswa dengan IP Semester 2 tinggi (> 3.00) dan berstatus tidak bekerja hampir seluruhnya terklasifikasi dalam kategori IPK "Sangat Memuaskan". Namun, bagi mahasiswa yang bekerja, terdapat cabang tambahan yang melibatkan atribut Jalur Masuk. Mahasiswa jalur beasiswa yang bekerja cenderung mampu mempertahankan IPK tinggi dibandingkan mahasiswa jalur reguler yang bekerja. Hal ini memberikan wawasan bahwa motivasi eksternal (beasiswa) berperan sebagai variabel moderasi terhadap tekanan beban kerja tambahan di luar kampus, sebagaimana didukung oleh teori dari Suliman (2021) mengenai pengaruh kondisi sosial ekonomi terhadap capaian belajar.

Aturan-aturan (*rules*) yang dihasilkan dari pohon keputusan ini menyediakan basis logika "If-Then" yang sangat aplikatif bagi dosen pembimbing akademik. Sebagai contoh, salah satu aturan kuat yang ditemukan adalah: "Jika IP Semester 2 < 2.75 dan Status Pekerjaan = Bekerja, maka Prediksi IPK = Cukup". Dengan adanya visualisasi ini, pengelola program studi dapat dengan mudah mengidentifikasi kelompok mahasiswa yang memerlukan intervensi khusus tanpa harus melakukan perhitungan statistik manual yang rumit. Pohon keputusan ini menjadi alat bantu visual yang efektif dalam mentransformasi data mentah menjadi strategi pendampingan akademik yang presisi di Universitas Mahakarya Asia.

PENUTUP

Penelitian ini berhasil menyimpulkan bahwa penerapan teknik data mining dengan algoritma C4.5 mampu memberikan wawasan mendalam mengenai faktor-faktor yang mempengaruhi Indeks Prestasi Kumulatif (IPK) mahasiswa di Program Studi Sistem Informasi Universitas Mahakarya Asia. Melalui serangkaian tahapan metodologi CRISP-DM, ditemukan bahwa variabel akademik pada semester

awal, khususnya IP Semester 2, merupakan prediktor paling dominan dalam menentukan keberhasilan studi mahasiswa di masa depan. Selain itu, integrasi antara faktor internal akademik dan faktor eksternal seperti status pekerjaan memberikan gambaran yang lebih holistik mengenai profil mahasiswa. Model pohon keputusan yang dihasilkan menunjukkan tingkat akurasi yang memuaskan sebesar 88,30%, yang berarti model ini memiliki reliabilitas tinggi untuk digunakan sebagai instrumen pendukung keputusan oleh pihak manajemen program studi dalam memetakan potensi akademik mahasiswa secara objektif.

Berdasarkan temuan tersebut, penulis menyarankan kepada pengelola Program Studi Sistem Informasi di Universitas Mahakarya Asia untuk mulai mengimplementasikan sistem deteksi dini (early warning system) berbasis data mining guna memantau perkembangan mahasiswa sejak tahun pertama. Intervensi akademik yang lebih intensif perlu difokuskan pada mahasiswa yang menunjukkan pola penurunan performa pada semester kedua atau mahasiswa yang memiliki beban kerja di luar kampus (bekerja), agar mereka tetap mampu menjaga stabilitas IPK. Untuk penelitian selanjutnya, disarankan untuk memperluas cakupan dataset dengan menyertakan variabel psikologis seperti motivasi belajar dan data dari sistem manajemen pembelajaran (Learning Management System) untuk mendapatkan tingkat akurasi prediksi yang lebih tajam. Dengan demikian, kebijakan akademik yang diambil dapat lebih tepat sasaran demi meningkatkan mutu lulusan dan efisiensi pendidikan di institusi.

UCAPAN TERIMAKASIH

Penulis menyampaikan rasa syukur yang mendalam kepada Allah SWT atas selesainya penelitian ini. Ucapan terima kasih secara khusus ditujukan kepada Bapak Ahmad Budi Trisnawan, M.Kom., selaku dosen pembimbing mata kuliah Data Mining pada Program Studi Sistem Informasi, Fakultas Sains, Teknologi, Teknik, dan Matematika (STEM), Universitas Mahakarya Asia, yang telah memberikan bimbingan, arahan teknis, serta ilmu yang sangat bermanfaat dalam penyusunan jurnal ini.

DAFTAR PUSTAKA

- Anas, A., & Delima, R. H. (2021). *Penerapan Data Mining Untuk Menganalisis Data Mahasiswa Baru STIE-GK Muara Bulian*. 15(2), 119–128.
- Andik, Syarli, & Sari, C. R. (2022). DATA MINING KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES. *Journal Peqquruang: Conference Series*, 4, 424–428.
- Fitri, S. (2018). *PENERAPAN DATA MINING UNTUK EVALUASI KINERJA AKADEMIK MAHASISWA (STUDI KASUS: UMTAS)*. 9(1), 633–640.
- Noyari, J. A., Aprillia, A., Munthe, R. G., & Sutarman, A. (2024). *Optimasi Kinerja Sistem Informasi Manajemen Kampus Menggunakan Teknik Data Mining*. 3(1), 52–63.
- Oktorina, F. K., Rekayasa, T., Politeknik, L., Informatika, T., & Kampar, P. (2025). *Penerapan Data Mining Untuk Analisis Hasil Belajar Mahasiswa (Studi Kasus Mahasiswa Teknik Informatika Politeknik Kampar)*. 5(2).
- Pratama, S., Iswandi, Sevitan, A., & Putri, Tsabita, A. (2023). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma C4.5 dengan CRISP-DM. *Journal of Applied Informatics and Computing (JAIC)*, 7(1), 16–20.
- Pratama, A. N. Su., & Ratama, N. (2024). PENERAPAN METODE DATA MINING UNTUK MEMPREDIKSI KELULUSAN MAHASISWA AGAR TEPAT WAKTU MENGGUNAKAN

ALGORITMA C4.5 (STUDI KASUS : UNIVERSITAS PAMULANG. *JORAPI : Journal of Research and Publication Innovation*, 2(1), 1324–1329.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. 7(1), 59–64.

Suliman. (2021). IMPLEMENTASI DATA MINING TERHADAP PRESTASI DAN SOSIAL EKONOMI DENGAN ALGORITMA K-MEANS CLUSTERING. *Jurnal Sistem Informasi Dan Sistem Komputer*, 6(1), 1–11.

Trisnawan, A. B. (2025). PEMANFAATAN DATA MINING DALAM MENENTUKAN POLA PEMILIHAN MATA KULIAH MAHASISWA. *Jurnal Advance Research Informatika Informatika Dan Sistem Informasi*, 3(2), 145–151. <https://www.ejournalwiraraja.com/index.php/JARS>

Yoga, V., Wijaya, D., & Brotosaputro, G. (2025). *Penerapan Data Mining Dalam Prediksi Kinerja Akademik Mahasiswa Menggunakan Algoritma Machine Learning*. 10(2), 134–142.