Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

Penerapan Algoritma Decision Tree untuk Klasifikasi Prestasi Akademik Mahasiswa Berdasarkan Indeks Prestasi Semester dan Kumulatif

Nahot Marganda Simanjuntak

Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas HKBP Nommensen Pematangsiantar, Sumatera Utara, Indonesia
Email: [1,*]simanjuntaknahot76@gmail.com
(* : coressponding author)

Abstrak−Perguruan tinggi umumnya belum memanfaatkan data akademik mahasiswa secara optimal sebagai alat deteksi dini, sehingga mahasiswa yang berisiko mengalami penurunan prestasi belajar sering teridentifikasi terlambat. Penelitian ini menyelesaikan masalah tersebut dengan menerapkan algoritma Decision Tree berbasis perhitungan entropy dan information gain untuk mengklasifikasikan prestasi akademik mahasiswa ke dalam tiga kategori (Baik, Cukup, dan Kurang) menggunakan fitur Indeks Prestasi Semester (IPS1–IPS8) dan Indeks Prestasi Kumulatif (IPK) dari 518 data mahasiswa. Tujuan penelitian ini adalah membangun model klasifikasi yang akurat dan mudah diinterpretasikan, sehingga kontribusinya dapat digunakan pihak universitas sebagai alat bantu pengambilan keputusan akademik secara dini dan transparan, berbeda dengan model black-box yang sulit dijelaskan. Dataset dibagi dengan rasio 70:30 untuk data training dan testing. Hasil sementara penelitian menunjukkan bahwa atribut IPK merupakan faktor paling berpengaruh dengan nilai Information Gain tertinggi sebesar 0,8766 dan terpilih sebagai root node, sementara evaluasi model pada data testing menggunakan confusion matrix menghasilkan akurasi sebesar 96,15%, precision 94,71%, recall 95,10%, dan F1-score 94,87%. Hasil ini menunjukkan bahwa model Decision Tree yang dibangun layak digunakan sebagai alat bantu identifikasi dini mahasiswa yang memerlukan perhatian akademik.

Kata Kunci: Data Mining; Decision Tree; IPK; Klasifikasi Mahasiswa; Prestasi Akademik

Abstract−*Higher education institutions generally have not optimally utilized student academic data as an early detection tool, so students at risk of declining academic performance are often identified too late. This study addresses this problem by applying the Decision Tree algorithm, based on entropy and information gain calculations, to classify student academic performance into three categories (Good, Sufficient, and Poor) using Grade Point per Semester (IPS1–IPS8) and Cumulative Grade Point Average (IPK) features from 518 student records. The purpose of this study is to build a classification model that is both accurate and interpretable, so that its contribution can be used by universities as a transparent decision-support tool for early academic intervention, unlike black-box models that are difficult to explain. The dataset was divided using a 70:30 training-testing ratio. Preliminary results show that the IPK attribute is the most influential factor, with the highest Information Gain of 0.8766, selected as the root node, while model evaluation on the testing data using a confusion matrix yields an accuracy of 96.15%, precision 94.71%, recall 95.10%, and F1-score 94.87%. These results indicate that the constructed Decision Tree model is suitable for use as an early-detection tool for students requiring academic attention.*

Keywords: *Data Mining; Decision Tree; GPA; Student Classification; Academic Performance*

1. PENDAHULUAN

Perkembangan teknologi informasi mendorong pertumbuhan data dalam jumlah besar di berbagai sektor, termasuk pendidikan tinggi. Perguruan tinggi menyimpan data akademik mahasiswa seperti Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK) dalam jumlah besar setiap periode akademik, namun data tersebut umumnya hanya digunakan untuk keperluan administratif dan belum dimanfaatkan secara optimal sebagai alat bantu pengambilan keputusan akademik [1], [2]. Akibatnya, mahasiswa yang mengalami penurunan prestasi belajar sering kali baru teridentifikasi setelah masalah berlangsung lama, sehingga intervensi akademik dari pihak kampus menjadi terlambat dan kurang efektif. Data mining menawarkan solusi atas permasalahan tersebut melalui proses ekstraksi pola tersembunyi dari kumpulan data besar menggunakan metode statistik, machine learning, dan kecerdasan buatan [3]. Salah satu tekniknya adalah klasifikasi, yaitu pengelompokan data ke dalam kelas tertentu berdasarkan atribut yang dimiliki . Dalam bidang pendidikan, klasifikasi banyak digunakan untuk memprediksi prestasi akademik, mengidentifikasi mahasiswa berisiko, serta mendukung intervensi akademik secara dini [4], [5].

Berbagai algoritma klasifikasi telah digunakan sebagai metode pembanding dalam penelitian-penelitian sejenis, antara lain Naive Bayes, Support Vector Machine (SVM), Random Forest, dan Artificial Neural Network. Algoritma-algoritma tersebut umumnya memiliki akurasi yang kompetitif, namun bersifat black-box sehingga proses pengambilan keputusannya sulit diinterpretasikan oleh pihak akademik yang bukan berlatar belakang teknis [6], [7]. Sebaliknya, Decision Tree merupakan algoritma klasifikasi yang banyak digunakan karena mudah dipahami, mampu menangani data numerik, dan menghasilkan aturan keputusan (if-then rules) yang dapat diinterpretasikan secara langsung tanpa memerlukan keahlian teknis khusus [8], [9]. Guevara-Reyes et al. [10] menegaskan bahwa model berbasis pohon keputusan (tree-based) tetap relevan sebagai pendekatan yang mengutamakan interpretabilitas dibandingkan model ensemble atau deep learning yang lebih kompleks dalam konteks pengambilan keputusan pendidikan, sementara Chen et al. [4] menunjukkan bahwa algoritma berbasis pohon keputusan dapat dioptimalkan untuk meningkatkan akurasi prediksi performa mahasiswa. Sebagai metode pembanding, Naive Bayes bekerja berdasarkan asumsi independensi antar atribut sehingga kurang akurat ketika fitur-fitur saling berkorelasi seperti IPS antar semester, sedangkan SVM dan Artificial Neural Network mampu menangkap hubungan non-linear pada data namun memerlukan tuning parameter yang kompleks

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

dan tidak menghasilkan aturan keputusan yang eksplisit, sehingga kurang sesuai untuk kebutuhan interpretasi cepat oleh staf akademik non-teknis.

Meskipun penelitian terdahulu telah banyak menerapkan algoritma klasifikasi untuk memprediksi performa akademik, sebagian besar penelitian tersebut menggunakan fitur berupa nilai mata kuliah, tingkat kehadiran, atau data demografis mahasiswa secara statis pada satu titik waktu, tanpa memanfaatkan pola perkembangan Indeks Prestasi Semester secara longitudinal sebagai fitur utama klasifikasi [11], [12], [13]. Selain itu, beberapa penelitian lebih berfokus pada algoritma black-box seperti Neural Network atau SVM yang sulit dijelaskan kepada pihak akademik non-teknis, sehingga hasil klasifikasinya kurang dapat ditindaklanjuti secara praktis [6], [7], [14].

Permasalahan utama yang ingin diselesaikan pada penelitian ini adalah bagaimana membangun model klasifikasi prestasi akademik yang tidak hanya akurat, tetapi juga dapat ditelusuri proses pengambilan keputusannya secara transparan oleh pihak akademik yang tidak memiliki latar belakang teknis machine learning, mengingat keterbatasan inilah yang membuat banyak model prediktif sebelumnya sulit diadopsi secara operasional di lingkungan kampus [15]. Berdasarkan uraian tersebut, terdapat gap penelitian terkait penerapan algoritma klasifikasi yang transparan dan mudah diinterpretasikan dengan memanfaatkan tren IPS per semester secara longitudinal beserta IPK sebagai fitur utama untuk mengklasifikasikan prestasi akademik mahasiswa, yang belum banyak dieksplorasi pada penelitian-penelitian sebelumnya.[16] Gap ini menjadi semakin relevan mengingat sebagian besar perguruan tinggi di Indonesia, termasuk Universitas HKBP Nommensen Pematangsiantar, telah memiliki sistem informasi akademik yang menyimpan data IPS dan IPK setiap semester, namun belum memanfaatkan data historis tersebut secara longitudinal sebagai dasar pengambilan keputusan akademik yang proaktif.

Berdasarkan gap tersebut, penelitian ini bertujuan menerapkan algoritma Decision Tree berbasis perhitungan entropy dan information gain untuk mengklasifikasikan prestasi akademik mahasiswa menggunakan data simulasi akademik yang meliputi IPS per semester (IPS1–IPS8) dan IPK. Secara lebih spesifik, tujuan penelitian ini meliputi: (1) menghitung nilai entropy dan information gain dari setiap atribut kandidat untuk menentukan struktur pohon keputusan yang optimal; (2) membangun model Decision Tree yang dapat mengklasifikasikan mahasiswa ke dalam tiga kategori prestasi (Baik, Cukup, Kurang); dan (3) mengevaluasi performa model menggunakan confusion matrix serta membandingkannya secara konseptual dengan karakteristik algoritma pembanding seperti Naive Bayes, SVM, dan Artificial Neural Network yang telah dibahas pada paragraf sebelumnya.

Kontribusi penelitian ini adalah menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga transparan dan mudah diinterpretasikan dalam bentuk aturan keputusan (rules), sehingga dapat digunakan sebagai alat bantu deteksi dini bagi pihak akademik dalam mengidentifikasi mahasiswa yang memerlukan perhatian khusus. Penelitian ini disusun dengan struktur: Pendahuluan, Metodologi Penelitian, Analisa dan Pembahasan, Implementasi, dan Kesimpulan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui tujuh tahapan utama yang digambarkan dalam bentuk bagan alur pada Gambar 1, yaitu: (1) pengumpulan data akademik mahasiswa, (2) preprocessing data berupa cleaning dan transformasi label, (3) pembagian data menjadi data training dan data testing, (4) perhitungan entropy dan information gain untuk menentukan atribut pemecah terbaik, (5) pembentukan model Decision Tree, (6) pengujian model pada data testing, dan (7) evaluasi performa model menggunakan confusion matrix. Tahapan ini disusun secara berurutan agar setiap proses dapat divalidasi sebelum dilanjutkan ke tahap berikutnya, sehingga model yang dihasilkan dapat dipertanggungjawabkan secara metodologis.

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

Gambar 1. Bagan Alur Tahapan Penelitian

2.2 Kajian Pustaka Algoritma Decision Tree

Decision Tree merupakan salah satu algoritma klasifikasi dalam data mining yang membangun struktur pohon keputusan berdasarkan atribut dengan nilai information gain tertinggi pada setiap tahap pemecahan (splitting) [8]. Struktur pohon terdiri atas root node sebagai atribut pemecah pertama, internal node sebagai atribut pemecah pada cabang berikutnya, dan leaf node sebagai hasil akhir klasifikasi [4], [9]. Dibandingkan dengan algoritma klasifikasi lain seperti Naive Bayes yang berbasis probabilitas atau SVM yang berbasis hyperplane, Decision Tree memiliki keunggulan berupa kemudahan interpretasi karena hasil akhirnya dapat dituliskan dalam bentuk aturan if-then yang mudah dipahami oleh pengguna non-teknis[16] . Beberapa varian algoritma Decision Tree yang umum digunakan antara lain ID3, C4.5, dan CART, yang masing-masing berbeda dalam kriteria pemilihan atribut pemecah; penelitian ini menggunakan kriteria entropy dan information gain sebagaimana diterapkan pada algoritma ID3/C4.5 [10].

2.3 Entropy dan Information Gain

Dua formula utama dalam pembentukan pohon keputusan adalah Entropy dan Information Gain. Entropy digunakan untuk mengukur tingkat ketidakpastian data [8]:

$$Entropy(S) = -\sum p_i \times log_2(p_i) \quad (1)$$

Information Gain digunakan untuk menentukan atribut terbaik sebagai node percabangan [8], [10]:

$$Gain(S,A) = Entropy(S) - \sum (|Sv|/|S|) \times Entropy(Sv) \quad (2)$$

Dengan S = himpunan data, A = atribut yang dievaluasi, Sv = subset data dengan nilai atribut v, dan p_i = proporsi kelas ke-i.

2.4 Dataset Penelitian

Data yang digunakan merupakan data simulasi akademik mahasiswa yang mencakup IPS per semester (IPS1–IPS8) dan IPK. Variabel dataset disajikan pada Tabel 1.

Tabel 1. Variabel Dataset Penelitian

No	Variabel	Keterangan
1	IPS 1–8	Indeks Prestasi Semester per semester (0.00–4.00); nilai 0 = mahasiswa tidak aktif pada semester tersebut
2	IPK	Indeks Prestasi Kumulatif (0.00–4.00)
3	Label Asli	Predikat perolehan: A (Sangat Memuaskan), B (Memuaskan), C (Cukup), D (Kurang)

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

No	Variabel	Keterangan
4	Kategori DT	Label Decision Tree: Baik (A,B), Cukup (C), Kurang (D)

Berdasarkan Tabel 1, terdapat dua jenis atribut yang digunakan sebagai fitur klasifikasi (IPS dan IPK) serta dua jenis label, yaitu label asli (A, B, C, D) yang kemudian disederhanakan menjadi tiga kategori Decision Tree (Baik, Cukup, Kurang) agar proses klasifikasi lebih representatif terhadap kebutuhan deteksi dini akademik.
Dataset terdiri dari 518 data mahasiswa yang dibagi menjadi tiga kategori sebagaimana disajikan pada Tabel 2.

Tabel 2. Distribusi Kategori Dataset

Kategori	Label Asli	Jumlah	Persentase
Baik	A dan B	290	56,0%
Cukup	C	115	22,2%
Kurang	D	113	21,8%
Total	A, B, C, D	518	100%

Tabel 2 menunjukkan bahwa distribusi kelas pada dataset tidak seimbang sepenuhnya (imbalanced), dengan proporsi kelas Baik mendominasi sebesar 56,0%. Kondisi ini menjadi pertimbangan dalam evaluasi model, sehingga metrik precision, recall, dan F1-score macro-average digunakan agar performa model pada kelas minoritas (Cukup dan Kurang) tetap terukur secara proporsional.

2.5 Pembagian Data dan Parameter Model

Dataset dibagi dengan rasio 70:30, yaitu Training Data = 362 data dan Testing Data = 156 data (random seed = 42 untuk reproduktibilitas). Sampel data ditampilkan pada Tabel 3.

Tabel 3. Sampel Data Simulasi Mahasiswa (20 dari 518 Data)

No	NIM	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK	Kat.
1	15090032	3.00	0	0	3.90	3.87	4.00	0	0	1.85	Cukup
2	15090033	2.14	0.65	0	0	0	0	0	0	0.35	Kurang
3	15090034	3.24	3.25	3.26	3.50	3.33	3.48	3.58	3.25	3.36	Baik
4	15090035	0	0	0	0	0	0	0	0	0	Kurang
5	15090036	3.33	3.20	3.17	3.14	3.48	3.52	3.58	3.25	3.33	Baik
6	15090039	3.14	2.45	2.83	3.00	3.48	2.90	3.83	3.25	3.11	Baik
7	15090040	3.38	0	0	0	0	0	0	0	0.42	Kurang
8	15090041	3.38	3.25	3.04	3.23	3.71	3.62	3.75	4.00	3.50	Baik
9	15090042	3.48	3.25	3.26	3.36	3.62	3.43	3.67	3.25	3.42	Baik
10	15090043	3.14	2.47	3.04	3.00	3.48	3.62	3.50	2.75	3.13	Baik
11	15090044	3.33	3.15	3.30	3.55	3.14	3.10	3.58	3.25	3.30	Baik
12	15090045	3.71	3.60	3.57	3.50	3.86	3.38	4.00	3.25	3.61	Baik
13	15090046	2.38	0	0	0	0	0	0	0	0.30	Kurang
14	15090047	3.62	3.45	3.26	3.77	3.67	3.90	3.83	2.75	3.53	Baik
15	15090048	3.86	3.90	3.65	3.73	3.52	3.62	3.67	3.25	3.65	Baik
16	15090049	3.62	3.15	2.78	2.82	3.05	2.29	3.00	3.00	2.96	Cukup
17	15090050	3.43	3.50	0	0	0	0	0	0	0.87	Kurang
18	15090051	3.24	3.00	3.26	3.14	3.14	3.29	3.83	3.25	3.27	Baik
19	15090052	3.43	3.50	3.30	3.59	3.90	3.62	3.83	3.25	3.55	Baik
20	15090053	3.43	2.45	2.52	2.32	0.38	0	0	0	1.39	Cukup

** IPS = 0 menandakan mahasiswa tidak aktif pada semester tersebut.*

Tabel 3 menunjukkan variasi pola IPS yang beragam, termasuk mahasiswa dengan IPS bernilai 0 pada beberapa semester (baris NIM 15090033, 15090035, 15090040, 15090046, dan 15090050) yang mengindikasikan periode tidak aktif kuliah. Pola inilah yang menyebabkan nilai IPK mahasiswa tersebut menjadi rendah dan terklasifikasi ke kategori Kurang, sebagaimana akan dibuktikan melalui perhitungan entropy dan information gain pada bagian Analisa dan Pembahasan. Parameter Decision Tree yang digunakan dalam penelitian ini adalah criterion = 'entropy', max_depth = None (tidak dibatasi), min_samples_split = 2, dan random_state = 42 agar hasil eksperimen dapat direproduksi pada penelitian selanjutnya.

3. ANALISA DAN PEMBAHASAN

3.1 Perhitungan Entropy Dataset Keseluruhan

Tahap pertama dalam pembentukan pohon keputusan adalah menghitung entropy dari seluruh dataset S sebagai ukuran ketidakpastian sebelum dilakukan pemecahan (splitting). Dataset total S terdiri dari 518 data dengan distribusi tiga kelas: Baik = 290 data, Cukup = 115 data, dan Kurang = 113 data, sehingga proporsi masing-masing kelas adalah p(Baik) = 290/518 = 0,5598, p(Cukup) = 115/518 = 0,2220, dan p(Kurang) = 113/518 = 0,2182. Nilai entropy total dataset dihitung menggunakan persamaan (1) sebagai berikut:

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

$$Entropy(S) = -(0{,}5598)log_2(0{,}5598) - (0{,}2220)log_2(0{,}2220) - (0{,}2182)log_2(0{,}2182)$$

Entropy(S) = 0,4694 + 0,4793 + 0,4811 = 1,4298

Nilai entropy 1,4298 yang mendekati nilai maksimum $\log_2(3) = 1{,}585$ menunjukkan bahwa ketidakpastian kelas pada dataset masih relatif tinggi sebelum dilakukan pemecahan berdasarkan atribut, sehingga diperlukan atribut pemecah yang dapat menurunkan ketidakpastian tersebut secara signifikan.

3.2 Perhitungan Information Gain Tiap Atribut

Tahap kedua adalah menghitung Information Gain untuk setiap kandidat atribut pemecah guna menentukan atribut dengan kemampuan terbaik dalam menurunkan entropy dataset. Penelitian ini mengevaluasi tiga atribut kandidat, yaitu IPK, Semester Aktif (jumlah semester dengan IPS > 0), dan IPS Rata-rata, masing-masing dipecah menggunakan ambang (threshold) 3,00 untuk IPK dan IPS Rata-rata, serta ambang 6 semester untuk Semester Aktif. Proses perhitungan menggunakan persamaan (2) dilakukan dengan langkah: (1) memecah dataset S menjadi subset Sv berdasarkan ambang atribut, (2) menghitung entropy masing-masing subset Sv, (3) menghitung entropy gabungan berbobot (weighted entropy), dan (4) mengurangkan entropy total dataset dengan weighted entropy untuk mendapatkan nilai gain. Hasil selengkapnya disajikan pada Tabel 4.

Tabel 4. Perhitungan Information Gain Tiap Atribut

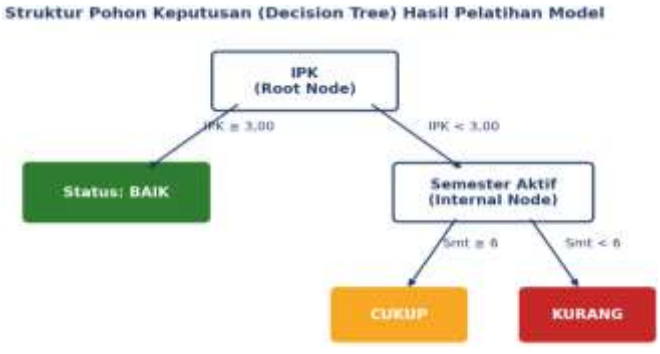
Atribut	Nilai	|Sv|	Baik	Cukup	Kurang	Entropy(Sv)	Gain(S,A)
IPK	≥ 3.00	288	285	2	1	0,0931	0,8766
	< 3.00	230	5	113	112	1,1293	
Smt. Aktif	≥ 6	399	290	103	6	0,9300	0,6051
	< 6	119	0	12	107	0,4717	
IPS Rata-rata	≥ 3.00	344	287	27	30	0,8131	0,5177
	< 3.00	174	3	88	83	1,1078	

Tabel 4 menunjukkan perbandingan ketiga atribut kandidat. Untuk atribut IPK dengan ambang ≥ 3,00, weighted entropy dihitung sebagai (288/518) × 0,0931 + (230/518) × 1,1293 = 0,0518 + 0,5012 = 0,5531, sehingga Gain(S, IPK) = Entropy(S) − 0,5531 = 1,4298 − 0,5531 = 0,8766. Nilai ini jauh lebih tinggi dibandingkan Gain(S, Semester Aktif) = 0,6051 dan Gain(S, IPS Rata-rata) = 0,5177, sehingga atribut IPK dipilih sebagai root node karena memberikan reduksi ketidakpastian terbesar pada langkah pemecahan pertama. Hasil ini konsisten dengan distribusi data pada Tabel 4 yang menunjukkan bahwa subset IPK ≥ 3,00 hampir seluruhnya berisi kelas Baik (285 dari 288 data), sehingga entropy subset tersebut sangat rendah (0,0931).

Berdasarkan Tabel 4, atribut IPK memiliki nilai Information Gain tertinggi sebesar 0,8766 sehingga dipilih sebagai root node pohon keputusan. Pada cabang IPK < 3,00, ketidakpastian kelas masih tinggi (entropy 1,1293) karena subset tersebut berisi campuran kelas Cukup (113 data) dan Kurang (112 data) dengan proporsi yang hampir sama, sehingga diperlukan pemecahan lanjutan. Pada cabang ini, atribut Semester Aktif dipilih sebagai node percabangan kedua karena memiliki Gain tertinggi di antara atribut yang tersisa (Gain = 0,6051), dengan subset Semester Aktif < 6 menghasilkan entropy rendah (0,4717) yang didominasi kelas Kurang (107 dari 119 data).

3.3 Proses Pembentukan Model Decision Tree

Proses pembentukan pohon keputusan dilakukan secara rekursif (top-down, greedy) dengan algoritma sebagai berikut: (1) hitung entropy node saat ini; (2) jika entropy = 0 atau seluruh data dalam node memiliki kelas yang sama, jadikan node tersebut sebagai leaf node; (3) jika tidak, hitung Information Gain seluruh atribut kandidat yang tersisa; (4) pilih atribut dengan Gain tertinggi sebagai node pemecah; (5) bagi data menjadi subset sesuai ambang atribut terpilih; dan (6) ulangi langkah 1–5 untuk setiap subset hingga seluruh data terklasifikasi atau kriteria penghentian (min_samples_split = 2) terpenuhi. Pada penelitian ini, proses tersebut menghasilkan pohon dengan kedalaman dua tingkat: IPK sebagai root node, dan Semester Aktif sebagai internal node pada cabang IPK < 3,00, sebagaimana ditampilkan pada Gambar 2.



Gambar 2. Struktur Pohon Keputusan Decision Tree Hasil Pelatihan Model

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

Sebagai ilustrasi penerapan aturan pada Gambar 2, perhatikan data mahasiswa dengan NIM 15090033 pada Tabel 3 yang memiliki IPK = 0,35. Karena IPK < 3,00, data tersebut diarahkan ke node Semester Aktif. Mahasiswa ini hanya memiliki 2 semester dengan IPS > 0 (Smt Aktif = 2 < 6), sehingga model mengklasifikasikannya ke kategori Kurang, yang sesuai dengan label aktual pada data tersebut. Sebaliknya, mahasiswa dengan NIM 15090034 memiliki IPK = 3,36 (≥ 3,00) sehingga langsung diklasifikasikan sebagai Baik tanpa perlu melalui node kedua, yang juga sesuai dengan label aktualnya.

Untuk menjelaskan secara lebih rinci bagaimana algoritma Decision Tree menyelesaikan permasalahan klasifikasi prestasi akademik pada penelitian ini, proses pembentukan model dapat diuraikan dalam empat tahapan teknis yang saling berkaitan:

a. Tahap pertama adalah perhitungan entropy dataset secara global (Bagian 3.1), yang berfungsi sebagai baseline ketidakpastian sebelum atribut apa pun dipertimbangkan; nilai entropy 1,4298 yang diperoleh menunjukkan bahwa tanpa informasi tambahan, model akan kesulitan menebak kategori prestasi seorang mahasiswa secara acak.
b. Tahap kedua adalah evaluasi seluruh atribut kandidat (IPK, Semester Aktif, dan IPS Rata-rata) secara paralel menggunakan persamaan Information Gain, di mana algoritma menghitung seberapa besar reduksi ketidakpastian yang dihasilkan oleh setiap kemungkinan ambang pemecahan pada masing-masing atribut; proses pencarian ambang terbaik ini dilakukan secara greedy, yaitu mengevaluasi seluruh nilai unik atribut sebagai kandidat threshold dan memilih yang menghasilkan Gain maksimum.
c. Tahap ketiga adalah pemilihan atribut pemecah dan pembentukan cabang, di mana algoritma secara rekursif menerapkan langkah yang sama pada setiap subset data hasil pemecahan hingga kriteria penghentian terpenuhi (entropy = 0 atau seluruh data dalam subset homogen).
d. Tahap keempat adalah penyusunan aturan keputusan akhir (Rule 1–3 pada Bagian 3.4), yang merupakan representasi eksplisit dari jalur (path) yang dilalui data mulai dari root node hingga leaf node pada struktur pohon yang telah terbentuk.

Keempat tahapan tersebut secara langsung menjawab permasalahan penelitian yang dirumuskan pada Bagian 1, yaitu kebutuhan akan model klasifikasi yang transparan dan dapat ditelusuri proses pengambilan keputusannya. Berbeda dengan algoritma black-box seperti Artificial Neural Network yang menyimpan pengetahuan dalam bentuk bobot numerik pada ratusan hingga ribuan koneksi antar neuron yang tidak dapat ditafsirkan secara langsung, setiap keputusan yang dihasilkan Decision Tree pada penelitian ini dapat ditelusuri kembali ke nilai entropy dan information gain pada setiap node yang dilalui. Sebagai contoh, ketika model mengklasifikasikan seorang mahasiswa ke kategori Kurang, pihak akademik dapat langsung melihat bahwa keputusan tersebut diambil karena IPK mahasiswa berada di bawah 3,00 dan jumlah semester aktifnya kurang dari 6, tanpa perlu memahami detail komputasi internal algoritma. Karakteristik inilah yang menjadi nilai tambah utama penelitian ini dibandingkan algoritma pembanding yang telah disebutkan pada Bagian 1, sekaligus menjadi jawaban atas gap penelitian terkait kebutuhan model yang transparan dan mudah diinterpretasikan oleh pihak non-teknis.

3.4 Aturan Klasifikasi (Rules)

Berdasarkan pohon keputusan yang terbentuk, diperoleh tiga aturan klasifikasi yang dapat diinterpretasikan secara langsung oleh pihak akademik:

a. *Rule 1: IF IPK ≥ 3,00 THEN Status = Baik*
b. *Rule 2: IF IPK < 3,00 AND Smt_Aktif ≥ 6 THEN Status = Cukup*
c. *Rule 3: IF IPK < 3,00 AND Smt_Aktif < 6 THEN Status = Kurang*

3.5 Confusion Matrix dan Evaluasi Model

Evaluasi model dilakukan pada 156 data testing yang tidak digunakan selama proses pelatihan. Hasil confusion matrix 3 kelas disajikan pada Tabel 5.

Tabel 5. Confusion Matrix (Data Testing = 156)

Aktual \ Prediksi	Pred. Baik	Pred. Cukup	Pred. Kurang
Aktual Baik	90	2	0
Aktual Cukup	1	31	1
Aktual Kurang	0	2	29

Tabel 5 menunjukkan bahwa dari 92 data aktual berkategori Baik, model berhasil memprediksi 90 data dengan benar dan hanya 2 data yang salah diklasifikasikan sebagai Cukup. Pada kelas Cukup, dari 33 data aktual, 31 data terklasifikasi benar, sedangkan pada kelas Kurang, 29 dari 31 data terklasifikasi benar. [19]Kesalahan klasifikasi (misclassification) terjadi pada total 6 dari 156 data testing (3,85%), yang seluruhnya berada pada kelas-kelas dengan nilai IPK mendekati ambang batas 3,00 atau Semester Aktif mendekati ambang 6, sehingga kesalahan tersebut wajar terjadi pada algoritma berbasis ambang batas tunggal (single threshold split). Perhitungan metrik evaluasi berdasarkan Tabel 5:

a. Accuracy: (90+31+29)/156 × 100% = 96,15%
b. Precision Macro Avg: (98,90% + 88,57% + 96,67%) / 3 = 94,71%

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

c. Recall Macro Avg: (97,83% + 93,94% + 93,55%) / 3 = 95,10%
d. F1-Score Macro Avg: 2 × (94,71% × 95,10%) / (94,71% + 95,10%) = 94,87%

Tabel 6. Rekapitulasi Evaluasi Model per Kelas

Kelas	Precision	Recall	F1-Score	Support
Baik	98,90%	97,83%	98,36%	92
Cukup	88,57%	93,94%	91,18%	33
Kurang	96,67%	93,55%	95,08%	31
Macro Avg	94,71%	95,10%	94,87%	156

Tabel 6 menunjukkan bahwa kelas Baik memiliki performa tertinggi (F1-score 98,36%) karena karakteristik datanya paling konsisten (IPK tinggi cenderung selalu berkategori Baik), [20]sedangkan kelas Cukup memiliki precision terendah (88,57%) karena posisinya berada di antara dua kelas lain pada struktur pohon, sehingga lebih rentan tertukar dengan kelas Kurang maupun Baik pada data dengan nilai atribut yang berada tepat di sekitar ambang batas.

3.5.1 Analisis Hasil dan Perbandingan dengan Penelitian Terkait

Model Decision Tree yang dibangun pada penelitian ini berhasil mengklasifikasikan prestasi mahasiswa dengan akurasi 96,15% pada data testing. Nilai ini menunjukkan performa yang sangat baik dan sejalan dengan penelitian Hidayat [13] yang melaporkan bahwa Decision Tree memiliki performa kompetitif dibandingkan Random Forest pada prediksi performa akademik, serta penelitian Guevara-Reyes et al. [15] yang menegaskan bahwa interpretabilitas model menjadi syarat penting agar prediksi performa akademik dapat ditindaklanjuti secara nyata oleh pengelola institusi pendidikan. Dibandingkan dengan penelitian Gufroni et al. [16] dan Gul et al. [17] yang menggunakan algoritma machine learning yang lebih kompleks (ensemble dan neural-based) namun bersifat kurang transparan, model pada penelitian ini menawarkan keunggulan berupa kemudahan interpretasi tanpa mengorbankan akurasi secara signifikan.

Atribut IPK terbukti sebagai faktor dominan (Gain = 0,8766), konsisten dengan temuan Chen et al. [5] yang menyatakan bahwa IPK merupakan prediktor terkuat dalam model prediksi prestasi mahasiswa dibandingkan atribut-atribut lain seperti kehadiran atau nilai tugas individual. Semester Aktif sebagai node kedua (Gain = 0,6051) menunjukkan bahwa keaktifan mahasiswa dalam mengikuti perkuliahan setiap semester berpengaruh signifikan terhadap kategori prestasinya, sebuah temuan yang melengkapi penelitian Rismaya et al. [6] yang berfokus pada prediksi performa tanpa mempertimbangkan pola keaktifan akademik secara longitudinal.

Kelebihan model pada penelitian ini terletak pada interpretabilitasnya: aturan klasifikasi yang dihasilkan (Rule 1–3 pada Bagian 3.4) dapat langsung digunakan oleh staf akademik tanpa perlu memahami detail teknis algoritma machine learning. Namun, model ini juga memiliki keterbatasan, yaitu kerentanan terhadap overfitting pada data dengan dimensi fitur lebih banyak serta sensitivitas terhadap data pada sekitar ambang batas pemecahan, sebagaimana terlihat dari 6 data yang salah klasifikasi pada Tabel 5. Keterbatasan ini dapat diatasi pada penelitian selanjutnya melalui teknik pruning atau kombinasi dengan algoritma ensemble seperti Random Forest.

4. IMPLEMENTASI

Implementasi model Decision Tree pada penelitian ini menggunakan bahasa pemrograman Python dengan library scikit-learn. Dataset sebanyak 518 data simulasi dimuat dari berkas CSV, dilakukan preprocessing berupa penggantian nilai IPS nol dengan penanda khusus tidak aktif dan transformasi label asli (A, B, C, D) menjadi tiga kategori Decision Tree (Baik, Cukup, Kurang), kemudian dibagi menjadi data training (362 data) dan data testing (156 data) menggunakan fungsi train_test_split dengan parameter stratify untuk menjaga proporsi kelas tetap seimbang pada kedua subset. Parameter Decision Tree yang digunakan: criterion = ‘entropy’, max_depth = None (auto), min_samples_split = 2, random_state = 42. Model dilatih menggunakan data training kemudian dievaluasi pada data testing untuk menghasilkan confusion matrix dan laporan klasifikasi sebagaimana disajikan pada Tabel 5 dan Tabel 6.

Hasil pengujian menunjukkan bahwa model dapat memproses dan mengklasifikasikan 156 data testing dalam waktu komputasi kurang dari satu detik, yang menunjukkan efisiensi algoritma Decision Tree dibandingkan algoritma berbasis ensemble atau neural network yang umumnya memerlukan waktu pelatihan dan inferensi lebih lama. Sebagai contoh pengujian, ketika model diberikan data baru dengan IPK = 2,80 dan Semester Aktif = 7, model akan menyusuri Rule 2 (IPK < 3,00 AND Smt_Aktif ≥ 6) dan menghasilkan prediksi Status = Cukup, sesuai dengan pola yang telah dipelajari selama pelatihan. Pengujian tambahan terhadap 6 data yang sebelumnya salah klasifikasi (lihat Tabel 5) menunjukkan bahwa kesalahan tersebut konsisten terjadi pada data dengan nilai IPK pada rentang 2,90–3,00 atau Semester Aktif tepat bernilai 6, yang menegaskan bahwa kesalahan model bersumber dari sensitivitas terhadap data yang berada tepat di sekitar ambang batas pemecahan (decision boundary), bukan dari kesalahan implementasi algoritma.

5. KESIMPULAN

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

Algoritma Decision Tree berbasis perhitungan entropy dan information gain terbukti mampu mengklasifikasikan prestasi akademik 518 mahasiswa ke dalam tiga kategori dengan tingkat akurasi 96,15%, precision makro 94,71%, recall makro 95,10%, dan F1-score makro 94,87% pada data testing, dengan struktur pohon yang hanya melibatkan dua atribut utama, yaitu IPK sebagai root node (Information Gain 0,8766) dan Semester Aktif sebagai node percabangan kedua (Information Gain 0,6051), sehingga menghasilkan tiga aturan klasifikasi yang ringkas dan dapat langsung diterapkan oleh pihak akademik tanpa memerlukan keahlian teknis machine learning. Capaian ini menegaskan bahwa model berbasis Decision Tree dapat menjadi alternatif yang seimbang antara akurasi dan interpretabilitas dibandingkan algoritma black-box seperti SVM dan Artificial Neural Network, sehingga layak diadopsi sebagai alat bantu deteksi dini mahasiswa yang berisiko mengalami penurunan prestasi. Hasil ini sekaligus menjawab tujuan penelitian yang dirumuskan pada Bagian 1, yaitu membangun model klasifikasi yang akurat sekaligus transparan dalam proses pengambilan keputusannya. Keterbatasan penelitian ini terletak pada penggunaan data simulasi serta sensitivitas model terhadap data di sekitar ambang batas pemecahan, sehingga penelitian selanjutnya disarankan menggunakan data riil dari sistem informasi akademik kampus, membandingkan performa dengan algoritma Random Forest atau Gradient Boosting untuk menekan risiko overfitting, serta menambahkan fitur tambahan seperti tingkat kehadiran dan nilai tugas guna meningkatkan akurasi model pada penelitian mendatang. Secara praktis, pihak akademik UHKBP Nommensen Pematangsiantar maupun perguruan tinggi lain dapat memanfaatkan aturan klasifikasi yang dihasilkan sebagai dasar penyusunan kebijakan bimbingan akademik dini, terutama bagi mahasiswa yang teridentifikasi pada kategori Cukup dan Kurang.

REFERENCES

[1] S. Citra Esananda, B. Nugroho, and F. Tri Anggraeny, "PENERAPAN ALGORITMA DECISION TREE DALAM MENENTUKAN PRESTASI AKADEMIK SISWA," 2021.

[2] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Educational Data Mining on Student Academic Performance Prediction: A Survey Educational Data Mining Pada Prediksi Kinerja Akademik Mahasiswa : Sebuah Survey," 2022. [Online]. Available: https://ieeexplore.ieee.org/

[3] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 22–31, Mar. 2023, doi: 10.33020/saintekom.v13i1.349.

[4] M. Chen and Z. Liu, "Predicting performance of students by optimizing tree components of random forest using genetic algorithm," *Heliyon*, vol. 10, no. 12, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32570.

[5] Riska Rismaya, Dwi Yuniarto, and David Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," *Router : Jurnal Teknik Informatika dan Terapan*, vol. 3, no. 1, pp. 15–23, Feb. 2025, doi: 10.62951/router.v3i1.389.

[6] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 14, pp. 92–104, 2019, doi: 10.3991/ijet.v14i14.10310.

[7] M. N. Gul, W. Abbasi, and M. Y. Wani, "Revolutionizing educational decision-making: a robust machine learning mechanism for predicting student performance," *Journal of Electrical Systems and Information Technology*, vol. 12, no. 1, Jun. 2025, doi: 10.1186/s43067-025-00230-z.

[8] T. Suprapti, B. Nurhakim, B. Warni Ayu Hermina, and V. A. Syahputra Simbolon, "A Decision Tree Model with Grid Search Optimization for Scholarship Recipient Classification," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3800–3813, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5235.

[9] M. M. Hussain, S. Akbar, S. A. Hassan, M. W. Aziz, and F. Urooj, "Prediction of Student's Academic Performance through Data Mining Approach," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 241–251, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.16.

[10] "15".

[11] Z. Fatah and S. Bella, "Klasifikasi Status Akademik Mahasiswa Menggunakan Decision Tree."

[12] R. Hidayat, H. Gultom, and Y. Samudra, "Predicting Student Academic Performance Using Learning Activity Data: A Comparative Study of Random Forest and Decision Tree Models," vol. 4, no. 3, pp. 2829–2529, 2025, doi: 10.52158/rjti.v4i3.4032.

[13] M. B. Handoko, Z. W. Fauzi, H. Agung, and D. Sofia, "Prediction of Science and Social Science Students Interests Using Decision Tree Algorithm with CRISP-DM," *bit-Tech*, vol. 8, no. 1, pp. 437–447, Aug. 2025, doi: 10.32877/bt.v8i1.2577.

[14] S. Malik *et al.*, "Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-92324-x.

[15] U. Al Faruq, M. Ainun Naja Fauzi, I. Fatayasya, E. Daniati, A. Ristyawan, and N. PGRI Kediri, "Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 131 Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer," Online, 2023.

[16] T. Trisna, A. Putri, R. Rahmadani, R. Siregar, and H. Hasan, "Academic Performance Prediction of PTIK Students through Machine Learning Models at Universitas Negeri Medan," *Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)*, vol. 7, no. 1, pp. 1128–1134, 2026, doi: 10.30596/jcositte.v7i1.29570.

[17] N. Nailil Amani and U. Hayati, "PENGGUNAAN ALGORITMA DECISION TREE UNTUK PREDIKSI PRESTASI SISWA DI SEKOLAH DASAR NEGERI 3 BAYALANGU KIDUL," 2024.

[18] N. Khasanah, D. Uki, E. Saputri, T. Hidayat, F. Aziz, and U. N. Mandiri, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 dengan RapidMiner: Studi Kasus Data Akademik Perguruan Tinggi XYZ," *Journal Computer Science*, vol. 4, no. 2, 2025.

Jurnal Sains Dan Teknologi Informasi
Vol 5, No 3, Juni 2026, Hal 146-154
ISSN 2809-610X (Media Online)
DOI 10.47065/jussi.v5i3.10223
Website https://ejurnal.seminar-id.com/index.php/jussi

[19] A. Hidayahtulloh and D. Prasetyo, "Penerapan Algoritma Decision Tree Untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik." [Online]. Available: https://jurnalmahasiswa.com/index.php/teknobis

[20] Dofiyanto and Z. Fatah, "Penerapan Algoritma Decision Tree untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Faktor Akademik dan Sosial," *JISCO : Journal of Information System and Computing*, vol. 3, no. 2, pp. 66–76, Dec. 2025, doi: 10.30631/jisco.v3i2.4030.