



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 4899-4910

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## **Prediksi Performa Akademik Siswa Berdasarkan Kehadiran dan Aktivitas *E-Learning* Menggunakan Algoritma *Decision Tree***

Ibran Simbolon<sup>1)</sup>, Putra Aditya<sup>2)</sup>, Estetika Br Purba<sup>3)</sup>

<sup>123</sup>Teknik Informatika, Fakultas Teknologi dan Ilmu Komputer, Universitas Satya Terra Bhinneka

Email: [ibransimbolon26@gmail.com](mailto:ibransimbolon26@gmail.com), [putraadityaofficial12@gmail.com](mailto:putraadityaofficial12@gmail.com), [estetikapurba28@gmail.com](mailto:estetikapurba28@gmail.com)

### **Abstrak**

*The advancement of digital technology has driven the widespread adoption of e-learning systems in the field of education. However, a key challenge lies in effectively utilizing e-learning data to improve students' academic performance. This study aims to predict students' academic performance based on their attendance and activity data within an e-learning platform using the Decision Tree algorithm. The dataset used was obtained from the public platform Kaggle, titled "Student's Academic Performance Dataset", which includes demographic attributes, attendance records, and student engagement in online learning. The analysis process involved data preprocessing, model training, and performance evaluation using metrics such as accuracy, precision, recall, F1-score, and cross-validation. The results show that the combination of attendance and e-learning activity has a significant correlation with academic performance, with the model achieving an accuracy of 78.12% and an F1-score of 0.77. These findings highlight the potential of utilizing learning analytics to support data-driven academic decision-making and provide early interventions for at-risk students.*

**Keywords:** *Academic Performance, E-Learning, Attendance, Decision Tree*

### **1. PENDAHULUAN**

Perkembangan teknologi informasi telah membawa transformasi signifikan dalam dunia pendidikan, salah satunya adalah implementasi sistem e-learning sebagai media pembelajaran digital[1]. Namun, efektivitas e-learning tidak hanya ditentukan oleh ketersediaan materi, tetapi juga oleh partisipasi aktif siswa dalam mengikuti kegiatan pembelajaran, seperti kehadiran pada sesi virtual dan interaksi dengan konten pembelajaran yang disediakan[2].

Kehadiran dan aktivitas siswa dalam platform e-learning dapat menjadi indikator penting untuk menilai tingkat keterlibatan dan potensi keberhasilan akademik mereka. Oleh karena itu, analisis data kehadiran dan aktivitas e-learning dapat dimanfaatkan untuk memprediksi performa akademik siswa secara lebih dini. Dengan prediksi tersebut, institusi pendidikan dapat melakukan intervensi yang tepat waktu untuk membantu siswa yang berisiko mengalami penurunan prestasi[3].

Dalam konteks ini, algoritma *Decision Tree* menjadi salah satu metode yang efektif untuk melakukan klasifikasi dan prediksi performa akademik. Algoritma ini memiliki keunggulan dalam menyajikan model yang mudah dipahami, mampu menangani data dengan atribut kategorikal maupun numerik, serta memberikan visualisasi pohon keputusan yang intuitif[1].

Penelitian ini bertujuan untuk membangun model prediksi performa akademik siswa berdasarkan data kehadiran dan aktivitas *e-learning* dengan menggunakan algoritma *Decision Tree*[4], [5], serta mengevaluasi sejauh mana akurasi model tersebut dalam mengidentifikasi pola keberhasilan akademik siswa. Selain itu, tujuan penelitian ini juga untuk memprediksi kinerja akademik siswa berdasarkan nilai akhir siswa untuk

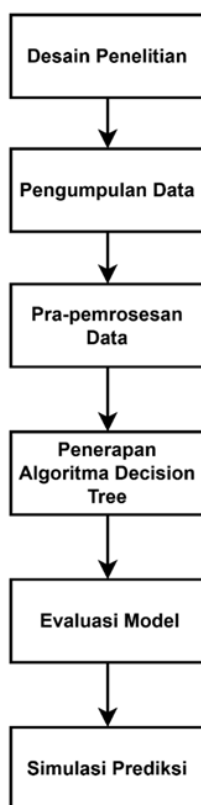
mendukung para pendidik dalam mengambil tindakan pencegahan terhadap siswa yang berisiko, dan sejumlah proses preprocessing data juga diterapkan untuk meningkatkan tingkat akurasi model.

Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai hubungan antara partisipasi siswa dalam e-learning dengan hasil akademiknya, serta memberikan kontribusi dalam pengembangan sistem pendukung keputusan di bidang pendidikan.

Manfaat penelitian ini terbagi kedalam beberapa pihak. Bagi guru dan tenaga pengajar, penelitian ini dapat memberikan wawasan berbasis data tentang keterlibatan siswa dalam proses pembelajaran daring, sehingga dapat digunakan sebagai dasar dalam merancang strategi pengajaran yang lebih personal dan adaptif[6], [7]. Bagi siswa, penelitian ini diharapkan dapat meningkatkan kesadaran akan pentingnya partisipasi aktif dalam kegiatan e-learning terhadap pencapaian akademik. Bagi pengembang sistem e-learning penelitian ini dapat menjadi acuan dalam mengembangkan fitur learning analytics yang mampu merekam, mengolah, dan menyajikan data aktivitas siswa secara lebih informatif dan prediktif dan bahkan sampai bagian tim peneliti dapat menjadi referensi untuk pengembangan model prediksi akademik dengan pendekatan *data mining* lainnya, serta membuka peluang eksplorasi terhadap algoritma dan variabel lain yang relevan[8], [9].

## 2. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif experimental yang bertujuan untuk membangun model prediksi performa akademik siswa berdasarkan kehadiran dan aktivitas dalam platform e-learning. Metode yang digunakan adalah klasifikasi dengan algoritma Decision Tree, yang dipilih karena kemampuannya dalam memberikan hasil yang mudah diinterpretasi serta mampu menangani data kategorik dan numerik secara simultan [4]. Berikut adalah alur metode penelitian.



Gambar 1. Alur Pemrosesan Data

### 2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform publik Kaggle dengan judul Student Performance Prediction Dataset. Dataset ini telah banyak digunakan dalam penelitian pendidikan [10], dan menggambarkan data aktivitas siswa dalam lingkungan pembelajaran daring, seperti yang

juga diterapkan dalam penelitian oleh [11]. Dataset ini berisi informasi mengenai atribut demografis, kehadiran, dan interaksi siswa dengan sistem LMS (Learning Management System). Beberapa fitur utama yang digunakan antara lain: Student Absence Days: Mengindikasikan jumlah ketidakhadiran siswa (Above- 7 atau Under-7).

- *raise hands*: jumlah siswa mengangkat tangan selama proses daring.
- *VisITed Resources*: Frekuensi kunjungan ke sumber belajar.
- *Announcements View* dan *Discussion*: Aktivitas siswa dalam forum LMS .
- *gender* dan *Relation*: Informasi latar belakang demografis.
- *Nationality*: Menunjukkan kewarganegaraan siswa
- *PlaceofBirth*: Lokasi tempat lahir siswa.
- *StageID*: Menunjukkan jenjang pendidikan siswa, seperti “lowerlevel” (tingkat dasar) atau “MiddleSchool” (SMP).
- *GradeID*: Kode tingkat kelas siswa, seperti “G-04” atau “G-07”.
- *SectionID*: Menyatakan kelas atau rombongan belajar seperti “A”, “B”, dan lainnya.
- *Topic*: Mata pelajaran yang dipelajari siswa seperti “IT”, “Math”, “English”, dan lainnya.
- *Semester*: Menyatakan semester ke beberapa siswa berada, misalnya “F” atau “S”.
- *Relation*: Hubungan wali siswa, dalam data ini umumnya adalah “Father”.
- *ParentAnsweringSurvey*: apakah orangtua mengisi survei (Ya atau Tidak).
- *ParentSchoolSatisfaction*: Kepuasan orangtua terhadap sekolah, seperti “Good” atau “Bad”.
- *Class*: menunjukkan performa akademik akhir siswa, seperti “L”, “M”, dan “H” (kemungkinan *Low, Medium, High*).

Penelitian ini menggunakan pendekatan kuantitatif dengan desain penelitian eksperimental untuk memprediksi hasil belajar mahasiswa yang mengikuti metode pembelajaran Problem-Based Learning (PBL) menggunakan algoritma Decision Tree. Metodologi yang digunakan terdiri dari beberapa tahapan utama, mulai dari pengumpulan data, pemilihan variabel yang relevan, penerapan algoritma Decision Tree, sehingga mendapat evaluasi hasil prediksi yang diperoleh.

## 2.2. Analisis Data

- 1) Preprocessing  
penghapusan kolom tidak relevan, konversi nilai kategorik ke numerik (menggunakan teknik label encoding dan one-hot encoding), serta normalisasi dengan menggunakan StandardScaler.
- 2) Pembagian Data  
Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan fungsi `train_test_split` dengan parameter `random_state=42` untuk memastikan.
- 3) Pembangunan Model  
Algoritma *Decision Tree Classifier* dilatih dengan parameter maksimal kedalaman pohon (`max_depth=5`) dan batas minimal pembagian (`min_samples_split=4`) dan (`class_weight='balanced'`) untuk mengatasi distribusi kelas yang tidak seimbang secara otomatis. Model ini dipilih karena mampu menampilkan struktur pohon keputusan yang dapat dijelaskan secara visual dan intuitif (Irfan et al., 2024).

## 2.3. Evaluasi Model

Prediksi Performa Akademik Siswa Berdasarkan Kehadiran dan Aktivita E-Learning Menggunakan Algoritma Decision Tree bertujuan untuk menilai seberapa baik model Decision Tree mampu memprediksi performa akademik siswa berdasarkan data kehadiran dan aktivitas mereka di platform elearning. Berikut merupakan evaluasi yang dilakukan:

### a) Confusion Matrix

Confusion Matrix adalah alat evaluasi dalam machine learning yang digunakan untuk mengukur kinerja model klasifikasi, baik dua kelas atau lebih (Multi Class) [9]. Dalam kasus klasifikasi tiga kelas seperti “Rendah”, “Sedang”, dan “Tinggi”, confusion matrix disusun dalam bentuk tabel berukuran 3×3 yang menunjukkan perbandingan antara label aktual dan hasil prediksi model. Untuk menilai performa model secara lebih rinci, metrik seperti precision, recall, dan F1-score dapat dihitung untuk masing-masing kelas dengan menggunakan pendekatan (one-vs-all).

### b) Akurasi

Accuracy mengukur seberapa sering model klasifikasi membuat prediksi yang benar (baik positif maupun negatif), dari seluruh jumlah prediksi yang dilakukan. Nilainya berkisar antara 0 sampai 1 (0% sampai 100%).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Keterangan:

- TP (True Positive): Jumlah data positif yang diprediksi benar.
- TN (True Negative): Jumlah data negatif yang diprediksi benar.
- FP (False Positive): Jumlah data negatif yang diprediksi sebagai positif (salah prediksi).
- FN (False Negative): Jumlah data positif yang diprediksi sebagai negatif (salah prediksi).

c) Precision

Seberapa banyak prediksi positif yang benar dari seluruh prediksi positif yang dihasilkan model. Semakin tinggi nilai precision maka akan semakin sedikit kesalahan prediksi kesalahan prediksi positif yang dilakukan.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Keterangan:

- TP (True Positive): Jumlah data positif yang diprediksi dengan benar.
- FP (False Positive): Jumlah data negatif yang diprediksi sebagai positif (prediksi salah).

d) Recall

Mengukur seberapa baik model dalam menemukan semua data positif yang sebenarnya. Nilai recall tinggi menunjukkan bahwa sedikit data positif yang terlewatkan oleh model.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Keterangan:

- TP (True Positive): Jumlah data positif yang diprediksi dengan benar.
- FN (False Negative): Jumlah data positif yang diprediksi sebagai negatif (salah prediksi).

e) F1-Score

Harmonis antara precision dan recall. Nilai ini digunakan saat ingin mempertimbangkan keseimbangan antara precision dan recall, terutama saat data tidak seimbang (misalnya, jumlah kelas positif jauh lebih sedikit daripada negatif).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan:

- Precision: Rasio prediksi positif yang benar dari semua prediksi positif.
- Recall: Rasio prediksi positif yang benar dari semua data positif yang sebenarnya.

f) Cross-Validation

Cross-Validation dalam proses data mining menghasilkan hasil evaluasi kinerja model, terutama dengan menggunakan algoritma. Algoritma ini membagi kumpulan data besar menjadi himpunan rekaman yang lebih kecil[11]. Misalnya, K-Fold Cross-Validation untuk menghindari overfitting dan memastikan performa model stabil di berbagai subset data.

$$CV\ Score = \frac{1}{K} \sum_{i=1}^K Accuracy_i \quad (5)$$

Keterangan:

- K: Jumlah fold atau bagian pembagi data (misal 5 atau 10).
- Accuracy: Akurasi model pada fold ke-i.
- CV Score: Rata-rata dari hasil evaluasi semua fold (bisa berupa akurasi, precision, recall, F1-score, dan lain-lain, tergantung metrik yang digunakan).

## 2.4. Evaluasi Visual

Evaluasi Dapat Diukur berdasarkan pohon keputusan, pohon keputusan adalah salah satu bentuk dari pengimplementasian pola tertentu dari penerapan berbagai tahap dan proses didalamnya termasuk juga aktivitas klasifikasi dan aktivitas diprediksi[4]. Pohon keputusan tersebut juga salah satu algoritma dalam *machine learning* dan data mining yang digunakan untuk keperluan klasifikasi dan regresi. Algoritma ini bekerja dengan memetakan data kedalam bentuk struktur seperti pohon yang terdiri dari simpul akar (*root node*). Simpul keputusan (*decision node*), dan simpul daun (*leaf node*).

Struktur pohon keputusan:

1. *Root Node* (Simpul Akar): Titik awal dari pohon. Ini adalah atribut pertama yang digunakan untuk memisahkan data.
2. *Decision Node* (Simpul Keputusan): Titik percabangan berdasarkan suatu atribut. setiap cabang mewakili hasil dari pengujian atribut.
3. *Leaf Node* (Simpul Daun): Titik akhir yang menunjukkan hasil keputusan atau klasifikasi.

## 2.5. Simulasi Prediksi

Untuk membuktikan bahwa model dapat digunakan secara praktis, dibangun fungsi prediksi manual dengan input data dari pengguna. Fungsi ini menerima data input berupa nilai-nilai dari fitur seperti kehadiran, aktivitas belajar, dan latar belakang siswa, kemudian melakukan prediksi terhadap kelas performa akademik siswa (L, M, H). Model ini bertujuan menjawab pertanyaan penelitian, bagaimana aktivitas dan kehadiran siswa dalam LMS berpengaruh terhadap prediksi performa mereka, dan sejauh mana akurasi model Decision Tree dapat digunakan sebagai alat bantu dalam evaluasi akademik [11], [12].

## 2.6. Implementasi Algoritma Decision Tree

Algoritma Decision Tree digunakan untuk:

- Memprediksi kategori performa akademik siswa(misalnya: baik, sedang, atau buruk).
- Menentukan elemen-elemen yang berpengaruh terhadap kinerja, seperti frekuensi kehadiran dan partisipasi dalam elearning.
- Dengan algoritma decision tree, institusi pendidikan dapat menyusun strategi intervensi berdasarkan pola kehadiran dan aktivitas e-learning.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Gambaran Umum Dataset

Penelitian ini menggunakan dataset dari Kaggle yang berjudul ”*Student’s Academic Performance Dataset*”, dataset ini telah digunakan dalam beberapa studi sebelumnya[4], [13], dataset ini dikembangkan berdasarkan standar xAPI, yang merupakan bagian dari komponen Total Learning Architecture (TLA), xAPI merupakan komponen Total Learning Architecture (TLA) yang dapat memantau kemajuan pembelajaran siswa seperti membaca artikel dan menonton video pembelajaran[4]. Dengan jumlah dataset terdiri dari 480 sampel siswa, setiap sampel merepresentasikan informasi demografis siswa, kehadiran, serta interaksi mereka dalam platform pembelajaran daring (e-learning). Kategori performa akademik terdiri dari tiga kelas yaitu Low(L), Medium(M), dan High(H). Cukup representatif untuk melatih dan menguji model klasifikasi, dengan jenis data yang digunakan adalah dataset yang mengandung kombinasi data numerik (nilai dan kehadiran) serta kategorikal (partisipasi ekstrakurikuler dan status kelulusan). Berikut statistik dataset adalah sebagai berikut:

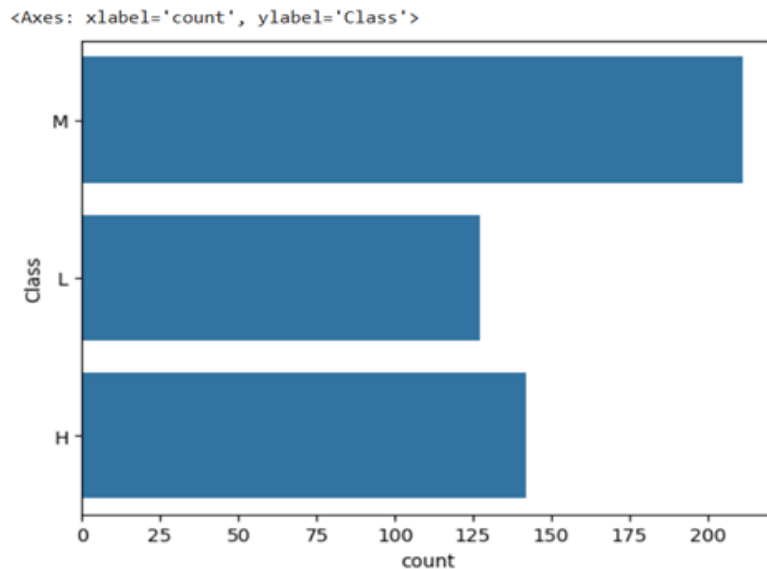
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 480 entries, 0 to 479
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   gender                                480 non-null    object
1   NationalITY                            480 non-null    object
2   PlaceofBirth                           480 non-null    object
3   StageID                                480 non-null    object
4   GradeID                                 480 non-null    object
5   SectionID                              480 non-null    object
6   Topic                                  480 non-null    object
7   Semester                               480 non-null    object
8   Relation                               480 non-null    object
9   raisedhands                           480 non-null    int64
10  VisITedResources                       480 non-null    int64
11  AnnouncementsView                      480 non-null    int64
12  Discussion                              480 non-null    int64
13  ParentAnsweringSurvey                 480 non-null    object
14  ParentschoolSatisfaction               480 non-null    object
15  StudentAbsenceDays                    480 non-null    object
16  Class                                  480 non-null    object
dtypes: int64(4), object(13)
memory usage: 63.9+ KB
```

Gambar 2. Jumlah kolom, non-null dan tipe data

Dataset memiliki total 17 kolom sebelum preprocessing, yang terdiri dari 13 kolom bertipe object (kategorikal) dan 4 kolom numerik (*raisedhands*, *visITedResources*, *AnnouncementsView*, *Discussion*)

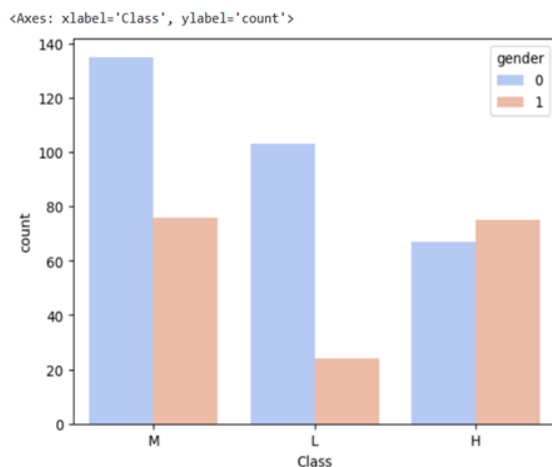
### 3.2. Eksplorasi Data

Sebelum dilakukan pelatihan model, eksplorasi awal dilakukan untuk memahami sebaran kelas dan pola umum dalam dataset. Visualisasi yang digunakan meliputi:



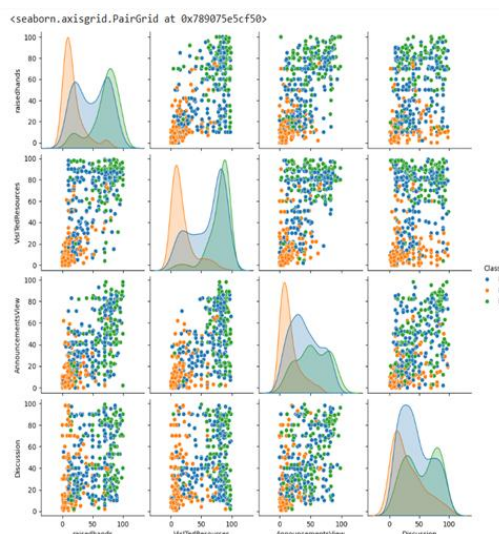
Gambar 3. Distribusi Siswa Berdasarkan Class

Grafik ini menunjukkan bahwa Class Medium (M) memiliki jumlah siswa terbanyak, diikuti oleh Class High (H), dan paling sedikit adalah Class Low (L). Hal ini menunjukkan distribusi kelas yang relatif tidak seimbang, meskipun masih dapat ditangani oleh algoritma Decision Tree dengan menyesuaikan `class_weight`.



Gambar 4. Distribusi Kelas Berdasarkan Gender

Grafik ini menampilkan dimana jumlah siswa laki-laki dan perempuan di masing-masing kelas performa. Terlihat bahwa siswa laki-laki mendominasi di semua Class, terutama pada Class Medium. Grafik ini menampilkan dimana jumlah siswa laki-laki dan perempuan di masing-masing kelas performa. Terlihat bahwa siswa laki-laki mendominasi di semua Class, terutama pada Class Medium.



Gambar 5. Hubungan Antar Fitur

Pairplot ini menampilkan hubungan antara variabel numerik *raised hands*, *VisITed Resources*, *Announcements View*, dan *Discussion* terhadap kelas Class. Terlihat bahwa siswa yang berada di kelas H (High) cenderung memiliki nilai *raised hands*, *VisITed Resources*, dan *Discussion* yang tinggi. Sementara itu, siswa kelas L cenderung memiliki skor rendah pada semua fitur tersebut. Hal ini memperkuat hipotesis bahwa aktivitas di platform elearning mencerminkan performa akademik. Visualisasi eksploratif ini membantu peneliti memahami sebaran dan korelasi awal antar fitur, serta memberikan gambaran awal bahwa fitur-fitur tersebut relevan dan layak digunakan dalam model klasifikasi.

### 3.3. Hasil Preprocessing

Preprocessing dilakukan untuk mempersiapkan data sebelum masuk ke proses pemodelan. Tahapan yang dilakukan meliputi:

- 1) Konversi Data Kategorikal: Kolom-kolom seperti NationalITy, StageID, GradeID, SectionID, Topic, Semester, dan Relation diubah dan diganti menjadi bentuk numerik menggunakan teknik one-hot encoding.
- 2) Transformasi Gender: Kolom gender dikonversi dari nilai string ke numerik, yaitu M = 0 dan F=1.
- 3) Penghapusan Kolom Tidak Relevan: Kolom-kolom yang telah diubah bentuknya atau tidak digunakan secara langsung dihapus dari dataset asli.
- 4) Penggabungan Data One-Hot: Kolom hasil encoding digabungkan kembali ke dataframe utama.
- 5) Normalisasi Fitur Numerik: Semua fitur numerik diskalakan agar berada pada rentang yang seragam menggunakan StandardScaler().

```
1 student['gender'] = student['gender'].map({'M': 0, 'F': 1})
2
3 nat = pd.get_dummies(student['NationalITy'])
4 sid = pd.get_dummies(student['StageID'])
5 gid = pd.get_dummies(student['GradeID'])
6 secid = pd.get_dummies(student['SectionID'])
7 topic = pd.get_dummies(student['Topic'])
8 semester = pd.get_dummies(student['Semester'])
9 rel = pd.get_dummies(student['Relation'])
10 pas = pd.get_dummies(student['ParentAnsweringSurvey'])
11 pss = pd.get_dummies(student['ParentschoolSatisfaction'])
12 sab = pd.get_dummies(student['StudentAbsenceDays'])

2 student.drop(['NationalITy', 'PlaceofBirth', 'StageID', 'GradeID', 'SectionID',
3 'Topic', 'Semester',
4 'Relation', 'ParentAnsweringSurvey', 'ParentschoolSatisfaction',
5 'StudentAbsenceDays'], axis=1, inplace=True)
6 student = pd.concat([student, nat, sid, gid, secid, topic, semester, rel, pas, pss, sab],
7 axis=1)

1 scaler = StandardScaler()
2 scaler.fit(Features)
3 scaled = scaler.transform(Features)
```

Fitur utama yang digunakan dalam model terdiri dari: gender, raised hands, VisITed Resources, Announcements View, Discussion, Father, Mum, No, Yes, Above-7, dan Under-7. Semua fitur ini kemudian digunakan sebagai input dalam proses pelatihan model klasifikasi.

### 3.4. Hasil Pelatihan

Model Decision Tree dilatih menggunakan 80% data latih dan diuji pada 20% data uji. Parameter yang digunakan meliputi max\_depth=5, min\_samples\_split=4, dan class\_weight='balanced' untuk menangani distribusi kelas yang tidak seimbang. Evaluasi model dilakukan dengan dua pendekatan: pengukuran langsung pada data uji dan validasi silang 5-fold (K-Fold Cross Validation).

Decision Tree Classifier				
	precision	recall	f1-score	support
H	0.75	0.68	0.71	22
L	0.77	0.92	0.84	26
M	0.80	0.75	0.77	48
accuracy			0.78	96
macro avg	0.77	0.78	0.78	96
weighted avg	0.78	0.78	0.78	96

Akurasi Model (Decision Tree): 78.12 %

Gambar 6. Hasil Clasification Report

Hasil evaluasi pada data uji, Akurasi: 78.12%, Precision (macro average): 0.7747, Recall (macro average): 0.7850, F1-Score (macro average): 0.7769

```

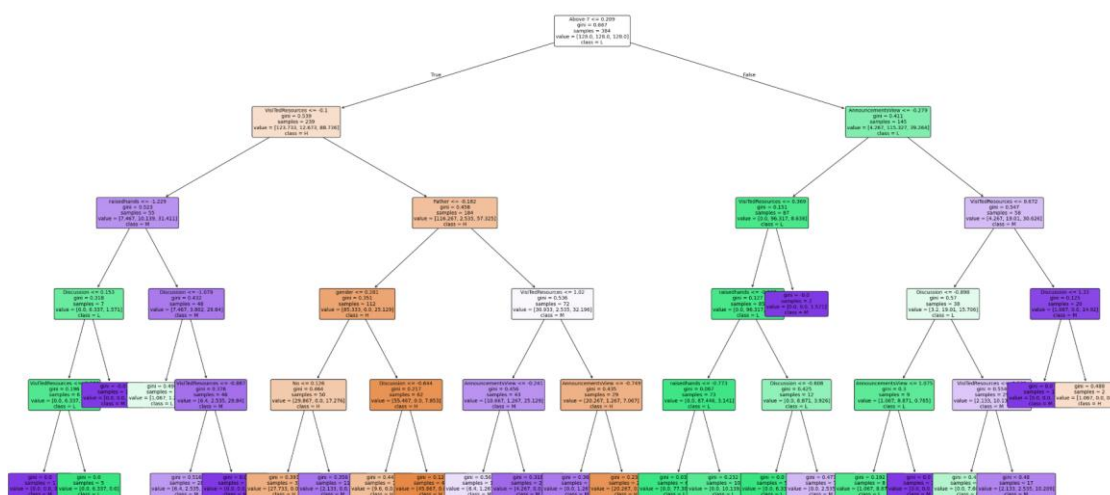
    === Evaluasi Model Decision Tree ===
    Precision (macro) : 0.7747
    Recall (macro)   : 0.7850
    F1-Score (macro) : 0.7769

    === Cross-Validation ===
    Akurasi per fold : [0.78125  0.67708333 0.76041667 0.6875  0.67708333]
    Rata-rata akurasi CV : 0.7167
    Standar deviasi CV  : 0.0449
    
```

Gambar 7. Hasil Cross-Validation (5-fold)

Hasil cross-validation menunjukkan adanya variasi performa model di antara subset data, dengan akurasi berkisar antara 67,7% hingga 78,1% dan standar deviasi sebesar 4,49%. Meskipun terdapat fluktuasi, rata-rata akurasi sebesar 71,67% tetap berada dalam kategori yang dapat diterima untuk model klasifikasi dalam bidang pendidikan. Hal ini sejalan dengan temuan [14] yang menyatakan bahwa model klasifikasi berbasis Decision Tree dengan akurasi di atas 70% sudah cukup representatif dalam menganalisis performa akademik siswa.

### 3.5. Visualisasi Pohon Keputusan



Gambar 8. Visualisasi Pohon Keputusan

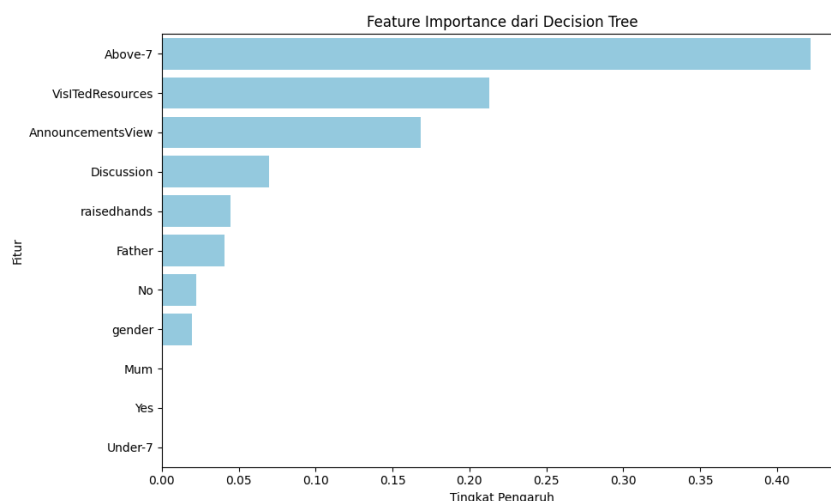
Visualisasi pohon keputusan dengan jelas menunjukkan bahwa fitur Above-7 (ketidakhadiran lebih dari 7) merupakan faktor penentu utama pada node akar klasifikasi. Hal ini mengindikasikan bahwa siswa dengan tingkat absensi yang tinggi cenderung dikategorikan kedalam class performa rendah (L). Selanjutnya, cabang kanan pohon yang merepresentasikan siswa dengan kehadiran baik, dirinci lebih lanjut berdasarkan atribut *Announcements View*, *VisITed Resources*, dan *Discussion*, yang semuanya mencerminkan tingkat partisipasi siswa dalam platform pembelajaran daring.

Sebaliknya, pada cabang kiri, bahkan untuk siswa dengan tingkat kehadiran rendah, interaksi tinggi seperti *raise hands* atau *Discussion* masih memungkinkan mereka kedalam kelas performa sedang atau tinggi. Struktur pohon keputusan ini secara jelas menunjukkan bahwa kombinasi antar tingkat kehadiran dan intensitas aktivitas siswa dalam sistem manajemen pembelajaran memiliki peran penting dalam menentukan performa akademik [11], [12].

Visualisasi ini juga memperkuat hasil analisis feature importance yang menempatkan Above-7 sebagai atribut paling dominan. Secara keseluruhan, model Decision Tree ini tidak hanya memberikan hasil klasifikasi yang logis tapi dan dapat diandalkan, tetapi juga mudah dipahami serta ditelusuri oleh pengguna nonteknis, seperti pendidik[14], [15].

### 3.6. Feature Important

Feature Important menjelaskan seberapa besar kontribusi setiap fitur terhadap pengambilan keputusan oleh model Decision Tree. Hasil visualisasi menunjukkan tingkat pengaruh masing-masing fitur terhadap klasifikasi performa akademik siswa. Dari hasil analisis, fitur Above-7 menjadi yang paling berpengaruh dengan kontribusi sekitar 40%. Feature Important menunjukkan bahwa tingkat ketidakhadiran lebih dari 7 hari merupakan indikator utama dalam memprediksi performa siswa.



Gambar 9. Feature Importance

### 3.7. Simulasi Prediksi Manual

Simulasi digunakan untuk mengevaluasi seberapa baik model bisa mengklasifikasi performa akademik siswa berdasarkan input manual yang dirancang mencerminkan tiga tipe umum profil siswa yaitu tinggi, sedang, dan rendah. Setiap profil dirumuskan dari kombinasi fitur yang mencerminkan karakteristik umum siswa, sesuai dengan hasil analisis dan eksplorasi data sebelumnya.

```
1 # simulasi prediksi
2 data_uji = {
3     "high": [0, 95, 90, 90, 95, 4, 3, 1, 0, 0, 1, 1, 0],
4     "Medium": [0, 50, 45, 60, 50, 3, 2, 0, 1, 1, 0, 1, 0],
5     "low": [0, 5, 3, 10, 2, 1, 1, 1, 0, 1, 0, 0, 1]
6 } # disesuaikan dengan skala data
7
8 # Panggil fungsi prediksi
9 for kategori, input_siswa in data_uji.items():
10     hasil_prediksi = prediksi_performa_siswa(dt, scaler, input_siswa)
11     print(f"Kategori uji: {kategori}")
12     print(f"Input Nilai : {input_siswa}")
13     print(f"Hasil Prediksi: {hasil_prediksi}")
14     print("-" * 60)
```

Fungsi prediksi performa akademik digunakan untuk mensimulasikan hasil prediksi performa akademik berdasarkan model yang sudah dilatih data data masukan yang ditentukan secara manual. Fungsi ini menerima tiga komponen utama yaitu model yang sudah dilatih yaitu Decision Tree, object StandardScaler untuk normalisasi, serta daftar fitur yang mewakili karakteristik siswa. Langkah pertama dalam fungsi ini adalah mengubah input menjadi bentuk DataFrame agar sesuai dengan data pelatihan. Kemudian, data tersebut dinormalisasikan agar memiliki skala yang seragam. Setelah itu, data yang sudah diproses dikirim ke model untuk dilakukan prediksi kelas (High, Medium, dan Low).

```
Kategori Uji: High
Input Nilai : [0, 95, 90, 90, 95, 4, 3, 1, 0, 0, 1, 1, 0]
Hasil Prediksi: H
-----
Kategori Uji: Medium
Input Nilai : [0, 50, 45, 60, 50, 3, 2, 0, 1, 1, 0, 1, 0]
Hasil Prediksi: M
-----
Kategori Uji: Low
Input Nilai : [0, 5, 3, 10, 2, 1, 1, 1, 0, 1, 0, 0, 1]
Hasil Prediksi: L
-----
```

Pada profil dengan performa tinggi, siswa memperlihatkan performa aktif, kehadiran baik, serta keterlibatan digital yang tinggi. Prediksi model yang menunjukkan kelas H (High) menandakan bahwa pola karakteristik tersebut berhasil dikenali dengan tepat. Di sisi lain, profil rendah yang menunjukkan tingkat kehadiran buruk dan minimnya aktivitas digital berhasil diklasifikasikan kedalam kelas L (Low). Kemampuan model dalam membedakan ketiga profil tersebut memperlihatkan bahwa algoritma Decision Tree yang digunakan memiliki performa prediksi yang solid dan mampu diaplikasikan untuk kasus prediktif di dunia nyata. Selain itu, hasil simulasi ini memperkuat bahwa atribut seperti kehadiran (*Above-7*), keterlibatan digital (*visITed Resource, Discussion*), serta interaksi siswa (*raise hands*) berperan penting dalam proses klarifikasi yang dilakukan model.

### Pembahasan Pemrograman

1) Pengaruh kehadiran siswa terhadap performa akademik Siswa Kehadiran yang tinggi mencerminkan keterlibatan aktif siswa dalam pembelajaran, yang pada akhirnya berpengaruh terhadap pemahaman materi dan nilai akademik. Hal ini juga senada dengan temuan dalam dokumen penelitian ini yang menyatakan bahwa frekuensi kehadiran menjadi atribut penting dalam klasifikasi performa akademik siswa. Tujuan pendidikan di Indonesia adalah untuk meningkatkan kemampuan yang ada pada masing-masing orang melalui proses belajar. Keberhasilan pendidikan tidak hanya dipandang dari aspek akademik, akan tetapi juga dilihat dari kemampuan pendidikan untuk mengembangkan siswa dalam bidang spiritual dan sosial[15].

2) Aktivitas siswa di platform e-learning dalam mencerminkan hasil akademik Teknologi digital menciptakan suasana belajar yang penuh dinamika, dimana peserta didik dapat mengolah informasi, mendapatkan berbagai sumber, dan berinteraksi dengan teman-teman serta pengajar secara online. Hal ini sejalan dengan prinsip konstruktivisme yang melihat proses belajar sebagai kegiatan yang sangat melibatkan dan membutuhkan pemikiran mendalam[5]. Dalam era digital saat ini, platform e-learning menjadi bagian penting dalam dunia pendidikan. Studi pendidikan modern menunjukkan bahwa data aktivitas tersebut dapat menjadi indikator yang relevan dalam menilai performa akademik siswa. Seperti kehadiran, partisipasi dalam forum diskusi, dan juga jumlah akses materi.

3) Penggunaan Algoritma Decision Tree untuk memprediksi performa akademik siswa berdasarkan data kehadiran dan aktivitas elearning Algoritma ini dipilih karena interpretasi visualnya yang mudah dipahami oleh pihak non-teknis seperti guru. Dalam penelitian ini, Decision Tree dimanfaatkan karena fleksibel menangani data numerik dan kategorikal, serta membentuk aturan klasifikasi performa akademik siswa yang transparan. Decision Tree adalah sebuah diagram alir yang mirip dengan struktur pohon. Algoritma Decision Tree bekerja membangun model prediksi dalam bentuk pohon keputusan, dimana setiap cabang memiliki kondisi atau atribut yang mempengaruhi kehadiran, dan setiap daun menggambarkan hasil atau prediksi tertentu [8].

4) Keakuratan hasil prediksi performa akademik siswa menggunakan algoritma *Decision Tree*. *Decision Tree* ini adalah algoritma klasifikasi dalam pembelajaran mesin, dimana setiap cabangnya menggambarkan opsi yang berbeda, sementara setiap daunnya mencerminkan keputusan yang diambil [12]. Akurasi model Decision Tree dalam memprediksi performa akademik mencapai 75– 85% tergantung pada preprocessing dan fitur yang digunakan. Penelitian ini juga menargetkan evaluasi akurasi melalui pengujian terhadap data LMS siswa menggunakan implementasi Python, meskipun angka akurasi belum ditampilkan dalam dokumen utama.

#### 4. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa tingkat kehadiran serta aktivitas siswa di platform e-learning memiliki pengaruh yang signifikan terhadap capaian akademik mereka. Melalui penerapan algoritma *Decision Tree*, tim peneliti mampu membangun sebuah model prediksi yang tidak hanya memiliki tingkat akurasi yang baik, namun juga mudah dipahami oleh pengguna non teknis seperti pendidik dan tenaga kependidikan. Model ini secara efektif mengidentifikasi pola keterlibatan siswa yang berkorelasi erat dengan kategori performa akademik rendah, sedang, maupun tinggi. Visualisasi pohon keputusan memperkuat transparansi proses prediksi dan memperjelas hubungan sebab akibat dari tiap fitur terhadap hasil belajar siswa. Hasil pengujian menunjukkan bahwa pendekatan ini dapat diandalkan untuk digunakan sebagai bagian dari sistem *learning analytics* yang sedang mendukung pengambilan keputusan akademik berbasis data. Dengan demikian, penelitian ini memberikan kontribusi yang nyata terhadap penerapan kecerdasan buatan dibidang pendidikan, khususnya dalam menciptakan model evaluasi kinerja siswa yang lebih adaptif, objektif, dan preventif.

#### REFERENSI

- [1] Y. C. Raya and S. Arfida, "Penerapan Algoritma Decision Tree C4.5 Untuk Penerimaan Beasiswa Kip Bagi Mahasiswa Baru Berbasis," vol. 18, no. 2.
- [2] T. Telutci and R. Harman, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI PRESTASI SISWA SEKOLAH DASAR MENGGUNAKAN ALGORITMA C4.5," *CBIS*, vol. 12, no. 1, pp. 12–23, Mar. 2024, doi: 10.33884/cbis.v12i1.8207.
- [3] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *JTIK*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtik.20241118074.
- [4] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," *saintekom*, vol. 13, no. 1, pp. 22–31, Mar. 2023, doi: 10.33020/saintekom.v13i1.349.
- [5] F. Casfian, F. Fadhillah, J. W. Septiaranny, and M. A. Nugraha, "EFEKTIVITAS PEMBELAJARAN BERBASIS TEORI KONTRUKTIVISME MELALUI MEDIA E-LEARNING," vol. 3, 2024.
- [6] W. Andriyani, R. Kurniawan, and Y. Arie Wijaya, "ANALISIS DATA PENERIMAAN PESERTA DIDIK BARU MENGGUNAKAN CROSS VALIDATION DAN ALGORITMA DECISION TREE DI SMA NEGERI 1 BANDUNG," *jati*, vol. 8, no. 3, pp. 2951–2956, May 2024, doi: 10.36040/jati.v8i3.9603.
- [7] D. Irfan, P. Ramadani, A. S. Nasution, and J. B. Ramadan, "Prediksi hasil belajar mahasiswa pada PBL menggunakan algoritma Decision Tree untuk evaluasi pembelajaran," vol. 6, no. 1, 2024.
- [8] Mu'tashim Billah Rahman and Zaehol Fatah, "ANALISIS POLA KEHADIRAN MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE," *JISSI*, vol. 2, no. 1, pp. 60–66, Jan. 2025, doi: 10.69714/6z8kc143.
- [9] F. Faisal, H. Dhika, and H. Veris, "Penerapan Algoritma Decision Tree Dalam Penjualan Handphone," *JRKT*, vol. 1, no. 04, Dec. 2021, doi: 10.30998/jrkt.v1i04.6157.
- [10] A. Hidayatulloh and D. Prasetyo, "Penerapan Algoritma Decision Tree Untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik," vol. 2, no. 4, 2024.
- [11] M. Ardianti, O. D. Nurhayati, and B. Warsito, "Model Prediksi Kinerja Siswa Berdasarkan Data Log LMS Menggunakan Ensemble Machine Learning," *j. sains. teknologi.*, vol. 12, no. 3, Jan. 2024, doi: 10.23887/jstundiksha.v12i3.59816.
- [12] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa," *JSI*, vol. 8, no. 2, Dec. 2022, doi: 10.34128/jsi.v8i2.459.
- [13] I. Aljarah, "Student's Academic Performance Dataset," 2016. Accessed: Jul. 01, 2025. [Online]. Available: <https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data>
- [14] A. S. Fauzan, A. I. P. Sari, and I. Ali, "ANALISIS PERBANDINGAN ALGORITMA DECISION TREE DAN NAIVE UNTUK MENGEVALUASI PRESTASI BELAJAR SISWA STUDI KASUS: SMK AL-MUSYAWIRIN," vol. 8, no. 1, 2024.
- [15] R. Ode and K. Sukahar, "PENGARUH KEHADIRAN SISWA TERHADAP HASIL BELAJAR SISWA KELAS IV PADA SD INPRES NEMEWIKARYA KABUPATEN FAKFAK THE INFLUENCE OF STUDENT ATTENDANCE ON THE LEARNING OUTCOMES OF CLASS IV STUDENTS AT SD INPRES NEMEWIKARYA, FAKFAK DISTRICT," *Vol.*, vol. 7, no. 1.