

Penerapan Algoritma Decision Tree untuk Klasifikasi Rendahnya Prestasi Siswa

Application of Decision Tree Algorithm for Classifying Low Student Academic Performance

Naila Anbar Afifah^{1*}, Syerina Arwan², Margareta Alinsa Ejung³

¹Undergraduate Program Studi Ilmu Komputer, Universitas Pancasakti, Jl. A. Mangerangi no73, Makassar 90121 Indonesia

¹Nailaanbar0@gmail.com, ²Syerinaarwan1212@gmail.com, ³Margareta@mail.com

* Corresponding author

Abstrak

Penelitian ini bertujuan untuk membangun model klasifikasi kelulusan siswa menggunakan algoritma *Decision Tree* berdasarkan tiga variabel utama, yaitu kehadiran, nilai ujian, dan status penyelesaian tugas. Data dikumpulkan melalui studi pustaka, observasi, dan wawancara. Proses *preprocessing* dilakukan dengan membersihkan dan mentransformasi data agar dapat diproses oleh algoritma. Model dikembangkan dan diuji menggunakan data uji sebanyak 20% dari dataset. Hasil evaluasi menunjukkan bahwa model memiliki akurasi, presisi, recall, dan F1-score sebesar 100%, tanpa kesalahan klasifikasi. Analisis pohon keputusan menunjukkan bahwa kehadiran merupakan variabel paling dominan dalam menentukan kelulusan, diikuti oleh penyelesaian tugas dan nilai ujian. Meskipun hasilnya sangat baik, keterbatasan jumlah data mengindikasikan perlunya penelitian lanjutan dengan dataset yang lebih besar dan validasi silang untuk meningkatkan generalisasi model. Model ini berpotensi digunakan sebagai alat bantu objektif bagi guru dalam mengidentifikasi siswa yang berisiko tidak lulus secara dini.

Kata Kunci: Pohon Keputusan; Prestasi Akademik; Kehadiran Siswa; Penambangan Data; Klasifikasi.

Abstract

This study aims to develop a student graduation classification model using the Decision Tree algorithm based on three main variables: attendance, exam scores, and assignment completion status. Data were collected through literature review, observation, and interviews. The preprocessing stage involved data cleaning and transformation to make it suitable for algorithmic processing. The model was trained and tested using 20% of the dataset. Evaluation results show that the model achieved 100% accuracy, precision, recall, and F1-score, with no classification errors. Decision tree analysis revealed that attendance was the most dominant variable in determining graduation, followed by assignment completion and exam scores. Despite the excellent performance, the limited dataset highlights the need for further studies using larger datasets and cross-validation to enhance model generalization. This model has the potential to serve as an objective decision-support tool for teachers in early identification of students at risk of not graduating.

Keywords: Decision Tree; Academic Performance; Student Attendance; Data Mining; Classification.

Pendahuluan

Prestasi belajar merupakan indikator penting dalam menilai keberhasilan proses pendidikan di sekolah. Rendahnya prestasi siswa menjadi permasalahan yang serius karena dapat berdampak pada motivasi belajar, partisipasi akademik, dan keberlanjutan pendidikan siswa [1]. Berbagai penelitian menunjukkan bahwa prestasi belajar dipengaruhi oleh berbagai faktor, baik internal maupun eksternal, seperti tingkat kehadiran, kemampuan akademik, penyelesaian tugas, kondisi sosial ekonomi, dan lingkungan belajar [2][3].

Seiring dengan perkembangan teknologi informasi, pendekatan berbasis data (data-driven) semakin banyak dimanfaatkan dalam dunia pendidikan untuk mengevaluasi dan memetakan permasalahan akademik secara lebih objektif. Salah satu metode analisis yang banyak digunakan adalah algoritma Decision Tree, yakni teknik klasifikasi yang membentuk model pohon keputusan berdasarkan atribut dalam data [4]. Keunggulan algoritma ini terletak pada kemampuannya dalam mengidentifikasi pola dan hubungan antar variabel serta menghasilkan aturan keputusan yang sudah dipahami.

Penelitian oleh Nugroho dan Basuki (2016) menunjukkan bahwa Decision Tree efektif dalam klasifikasi data pendidikan serta dalam mengidentifikasi faktor dominan yang memengaruhi prestasi siswa. Pendekatan ini

memungkinkan pihak sekolah untuk memperoleh informasi yang lebih sistematis dan berbasis bukti dalam pengambilan keputusan. [5]

Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree dalam menganalisis dan mengklasifikasikan data siswa untuk mengidentifikasi faktor utama yang berkontribusi terhadap rendahnya prestasi belajar. Fokus penelitian terletak pada atribut seperti kehadiran, nilai ujian akhir, dan status penyelesaian tugas. Hasil dari penelitian ini diharapkan dapat digunakan sebagai dasar dalam merumuskan strategi peningkatan prestasi belajar siswa secara lebih efektif dan terarah.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif analitis. Tujuan utamanya adalah membangun dan mengevaluasi model klasifikasi kelulusan siswa berdasarkan data kehadiran, nilai ujian, dan status penyelesaian tugas menggunakan algoritma Decision Tree. Penelitian dilaksanakan melalui beberapa tahapan, mulai dari analisis kebutuhan, pengumpulan data, pra-pemrosesan data, hingga pengujian dan evaluasi model.

A. Analisis Kebutuhan

Analisis kebutuhan dilakukan untuk merumuskan variabel-variabel yang relevan dan memahami konteks permasalahan yang ingin diselesaikan. Tiga pendekatan utama yang digunakan adalah:

1. Studi Pustaka : Studi ini mencakup penelaahan literatur terkait faktor-faktor yang memengaruhi prestasi belajar siswa, serta studi terdahulu yang mengaplikasikan algoritma *Decision Tree* dalam sistem pendidikan. Literatur diperoleh dari jurnal ilmiah, buku ajar, artikel online terverifikasi, serta laporan penelitian sebelumnya.
2. Observasi Lapangan : Pengamatan dilakukan terhadap proses belajar-mengajar, terutama dalam hal penilaian siswa. Kegiatan observasi berfokus pada bagaimana guru menilai kehadiran, penyelesaian tugas, dan nilai ujian dalam menentukan kelulusan siswa.
3. Wawancara Terstruktur : Wawancara dilakukan terhadap guru dan staf pengajar untuk mengumpulkan informasi kualitatif mengenai penyebab utama kegagalan siswa, serta ekspektasi mereka terhadap sistem klasifikasi berbasis teknologi. Hasil wawancara memberikan wawasan kontekstual yang memperkuat perumusan variabel penelitian.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang dikumpulkan secara manual dari lingkungan sekolah dan disimpan dalam format *Comma Separated Values* (CSV). Dataset terdiri atas 20 entri siswa, dengan setiap entri memuat informasi sebagai berikut:

1. Kehadiran (%): Persentase kehadiran selama satu periode pembelajaran.
2. Nilai Ujian: Skor ujian akhir yang diperoleh siswa.
3. Status Tugas: Status penyelesaian tugas, dikodekan sebagai “Selesai” atau “Tidak”.
4. Status Kelulusan: Target klasifikasi, yaitu “Lulus” atau “Tidak Lulus”.

C. Pra-pemrosesan Data

Tahap ini bertujuan untuk memastikan data siap digunakan dalam pemodelan algoritma *machine learning*. Proses mencakup:

1. Pembersihan Data (Data Cleaning) : Seluruh entri diperiksa untuk memastikan tidak ada nilai kosong, data ganda, atau anomali input. Karena data diinput secara manual, validasi dilakukan secara ketat, dan tidak ditemukan nilai yang hilang.
2. Transformasi Data (Data Transformation) : Variabel kategorikal diubah menjadi numerik agar dapat diproses oleh algoritma *Decision Tree*, dengan rincian sebagai berikut:
 - o “Tugas”: Selesai = 1; Tidak Selesai = 0
 - o “Lulus”: Ya = 1; Tidak = 0

D. Perancangan dan Implementasi Model

Model klasifikasi dibangun menggunakan algoritma *Decision Tree* dengan pendekatan supervised learning. Proses implementasi meliputi:

1. Pemisahan data menjadi data latih (80%) dan data uji (20%).
2. Pelatihan model menggunakan data latih.

3. Evaluasi performa model menggunakan data uji.
4. Visualisasi struktur pohon keputusan dan interpretasi variabel dominan.

Perangkat lunak yang digunakan dalam implementasi adalah Python dengan pustaka *scikit-learn* untuk membangun model klasifikasi dan *matplotlib* serta *seaborn* untuk visualisasi data.

E. Evaluasi Model

Evaluasi dilakukan menggunakan empat metrik utama dalam klasifikasi biner, yaitu:

- Akurasi (Accuracy): Proporsi prediksi yang benar dari seluruh data uji.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Presisi (Precision): Proporsi prediksi positif yang benar-benar positif.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall (Sensitivity): Kemampuan model dalam mendeteksi semua kasus positif.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1-Score: Rata-rata harmonik antara presisi dan recall.

$$F1 - Score = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

Selain itu, confusion matrix digunakan untuk memvisualisasikan performa klasifikasi terhadap kelas positif dan negatif.

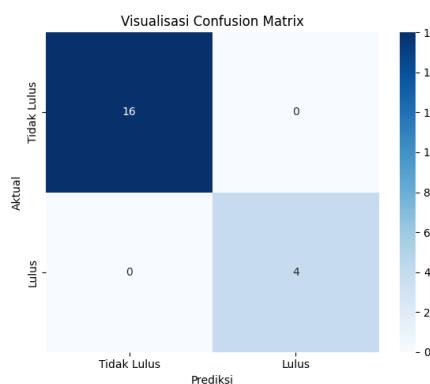
Hasil dan Diskusi

Evaluasi kinerja model dilakukan untuk menilai sejauh mana algoritma Decision Tree mampu mengklasifikasikan kelulusan siswa berdasarkan variabel masukan: kehadiran, nilai ujian, dan status penyelesaian tugas. Penilaian dilakukan dengan membandingkan hasil prediksi terhadap data aktual pada data uji, yang terdiri atas 20% dari total data.

Visualisasi confusion matrix disajikan pada Tabel 1 dan Gambar 1. Hasilnya menunjukkan bahwa seluruh prediksi yang dihasilkan model sesuai dengan kelas aktual.

Tabel 1. Visualisasi Confusion Matrix

<i>Logistic Regresion</i>	<i>Prediksi Siswa</i>	<i>Lulus</i>	<i>Tidak Lulus</i>
Kelas Aktual	Lulus	4	0
	Tidak Lulus	0	16



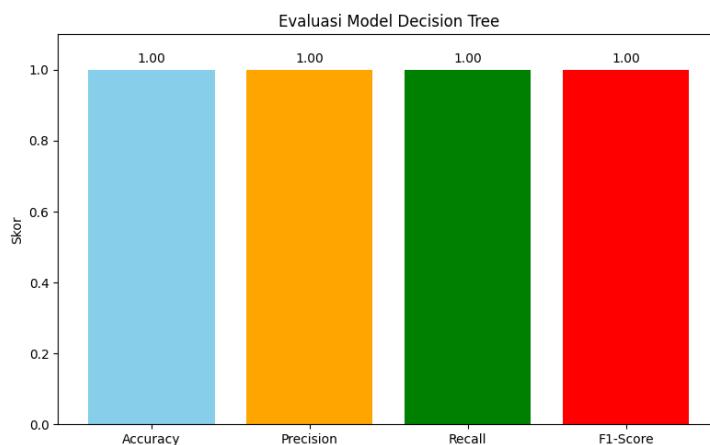
Gambar 1. Visualisasi Confusion Matrix

Model berhasil memprediksi seluruh siswa yang lulus (True Positive = 4) dan yang tidak lulus (True Negative = 16) dengan tepat. Tidak terdapat kesalahan klasifikasi, baik berupa false positive maupun false negative, sehingga performa klasifikasi tergolong sempurna.

Hasil evaluasi metrik klasifikasi disajikan pada Gambar 2 dan dirangkum dalam Tabel 2 berikut:

Gambar 2. Evaluasi Model Decision Tree

Metrik Evaluasi	Nilai
Akurasi	1.00
Presisi	1.00
Recall	1.00
F1-Score	1.00



Gambar 2. Evaluasi Model Decision Tree

Gambar 2 memperlihatkan visualisasi grafik batang dari metrik evaluasi model. Semua metrik utama—yakni akurasi, presisi, recall, dan F1-score—mencapai nilai 1.00, yang setara dengan 100%, menunjukkan bahwa model berhasil mengklasifikasikan seluruh data uji dengan sangat baik.

Analisis Pohon Keputusan

Struktur pohon keputusan yang dihasilkan oleh algoritma *Decision Tree* disajikan pada **Gambar 2**. Analisis terhadap struktur tersebut menunjukkan bahwa:

1. **Kehadiran (%)** menjadi *root node* dan merupakan variabel yang paling dominan dalam menentukan kelulusan siswa.
2. Siswa dengan kehadiran $\leq 75\%$ hampir selalu diklasifikasikan sebagai **tidak lulus**, tanpa memperhatikan faktor lainnya.
3. Untuk siswa dengan kehadiran tinggi, model melanjutkan pengambilan keputusan berdasarkan:
 - o **Status Tugas:** Siswa yang tidak menyelesaikan tugas cenderung diklasifikasikan sebagai tidak lulus.
 - o **Nilai Ujian:** Siswa dengan tugas selesai dan kehadiran penuh tetapi memiliki nilai ujian ≤ 65 juga berisiko tidak lulus.

Interpretasi dan Keterbatasan

Hasil evaluasi menunjukkan bahwa model *Decision Tree* memiliki performa klasifikasi yang sangat tinggi pada data uji. Namun, perlu dicatat bahwa dataset yang digunakan relatif kecil dan bersifat homogen (hanya terdiri dari **20 entri siswa**). Hal ini berpotensi menyebabkan **overfitting**, yaitu kondisi di mana model sangat cocok terhadap data pelatihan namun memiliki generalisasi yang rendah pada data yang lebih luas. Sebagai tindak lanjut, penelitian selanjutnya disarankan untuk:

1. Menggunakan **dataset yang lebih besar dan beragam** untuk meningkatkan representasi data.

2. Menerapkan teknik **validasi silang (cross-validation)** guna memastikan bahwa performa model dapat digeneralisasikan ke populasi siswa yang berbeda.

Kesimpulan

Penelitian ini berhasil merancang dan mengevaluasi model klasifikasi kelulusan siswa menggunakan algoritma Decision Tree berbasis data kehadiran, nilai ujian, dan status penyelesaian tugas. Hasil evaluasi menunjukkan bahwa model mampu melakukan klasifikasi dengan akurasi sempurna, ditandai dengan nilai akurasi, presisi, recall, dan F1-score sebesar 1.00 atau 100%. Visualisasi confusion matrix juga mengonfirmasi bahwa tidak terdapat kesalahan klasifikasi pada data uji, dengan seluruh prediksi sesuai dengan kelas aktual. Struktur pohon keputusan yang dihasilkan mengungkap bahwa variabel kehadiran siswa merupakan faktor paling dominan dalam menentukan kelulusan. Siswa dengan kehadiran $\leq 75\%$ hampir selalu diklasifikasikan sebagai tidak lulus, sedangkan untuk siswa yang hadir penuh, faktor penyelesaian tugas dan nilai ujian menjadi penentu tambahan. Meskipun performa model sangat tinggi, perlu diakui bahwa ukuran dataset yang kecil (20 entri siswa) menjadi keterbatasan utama dalam penelitian ini. Oleh karena itu, untuk meningkatkan validitas dan generalisasi model, disarankan agar penelitian lanjutan menggunakan dataset yang lebih besar dan beragam, serta menerapkan teknik validasi silang (cross-validation) guna menghindari overfitting. Secara keseluruhan, model klasifikasi ini tidak hanya menunjukkan efektivitas teknis, tetapi juga berpotensi menjadi alat bantu objektif bagi guru dan pihak sekolah dalam mengidentifikasi siswa yang berisiko tidak lulus, sehingga intervensi pendidikan dapat dilakukan lebih awal dan tepat sasaran.

Daftar Pustaka

- [1] S. Arikunto, Dasar-dasar Evaluasi Pendidikan, Jakarta: Bumi Aksara, 2013.
- [2] Slameto, Belajar dan Faktor-Faktor yang Mempengaruhinya, Jakarta: Rineka Cipta, 2010.
- [3] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, 2002.
- [4] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., San Francisco: Morgan Kaufmann, 2011.
- [5] E. R. Nugroho and A. Basuki, "Penerapan algoritma Decision Tree C4.5 untuk menentukan faktor yang mempengaruhi prestasi belajar siswa," *J. Teknol. dan Sist. Komput.*, vol. 4, no. 2, pp. 234–240, 2016, doi: 10.14710/jtsiskom.4.2.2016.234-240.
- [6] N. W. Safitri and L. E. Nugroho, "Implementasi Decision Tree dalam klasifikasi prestasi belajar siswa," *J. Tek. ITS*, vol. 7, no. 2, pp. D221–D224, 2018, doi: 10.12962/j23373539.v7i2.38433.
- [7] D. Indriani, "Analisis pengaruh kehadiran dan tugas terhadap hasil belajar siswa menggunakan regresi linier," *J. Pendidik. Mat.*, vol. 11, no. 1, pp. 45–52, 2017.
- [8] A. R. Putra and E. Prasetyo, "Klasifikasi kelulusan mahasiswa dengan algoritma Decision Tree dan Naive Bayes," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 523–530, 2020, doi: 10.25126/jtiik.2020733417.
- [9] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed., Burlington, MA: Morgan Kaufmann, 2017.
- [10] R. Setiawan and N. Maftukhah, "Penerapan algoritma C4.5 untuk prediksi prestasi belajar siswa," *J. Sains dan Informatika*, vol. 5, no. 1, pp. 12–18, 2019.