

Eksplorasi Algoritma *Decision Tree* untuk Penentuan Siswa Berprestasi

Prestian Ramadhan^{1)*}, Yuhandri²⁾, Jhon Veri³⁾

¹⁾Universitas Putra Indonesia "YPTK" Padang
Jl. Raya Lubuk Begalung, Kota Padang, Indonesia

¹⁾prestian.ramadhan14@gmail.com

²⁾yuhandri.yunus@gmail.com

³⁾jhon080771@yahoo.co.id

Jejak Artikel:

Abstrak

Upload 03 Maret 2025;
Revisi 04 Maret 2025;
Diterima 07 Maret 2025;
Tersedia online 25 April 2025

Kata Kunci :

Data mining
Decision Tree
Klasifikasi Siswa
Prestasi Akademik
Rapid Miner

Kemajuan teknologi informasi telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk pendidikan. Salah satu tantangan utama dalam meningkatkan kualitas pembelajaran adalah identifikasi siswa berprestasi secara akurat dan objektif. Penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree C4.5* dalam menentukan siswa berprestasi di SMPN 1 Kerinci, dengan mempertimbangkan faktor akademik dan non-akademik seperti nilai disiplin, nilai tahfidz, nilai akhlak, dan nilai ujian. Metode penelitian mencakup pengumpulan data siswa, preprocessing data untuk mengatasi ketidakseimbangan data, analisis faktor-faktor yang berpengaruh, serta pembangunan model klasifikasi menggunakan perangkat lunak *RapidMiner 9.0*. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree C4.5* mampu mengklasifikasikan siswa dengan tingkat akurasi sebesar 91,67%, precision 93,33% untuk kelas "Tidak" dan 88,89% untuk kelas "Layak", serta recall masing-masing 93,33% dan 88,89%. Berdasarkan analisis *gain ratio*, nilai akhlak memiliki pengaruh terbesar dalam klasifikasi siswa, dengan nilai 0,5961, diikuti oleh nilai tahfidz dan nilai disiplin. Model klasifikasi ini dapat membantu sekolah dalam mengidentifikasi siswa secara lebih objektif, sehingga memungkinkan pengambilan keputusan berbasis data dalam memberikan intervensi pendidikan yang lebih tepat sasaran. Selain itu, hasil penelitian ini membuka peluang untuk pengembangan lebih lanjut, seperti penggunaan teknik *ensemble learning* atau optimasi model menggunakan metode *boosting* guna meningkatkan performa klasifikasi. Dengan demikian, sistem berbasis data mining ini dapat menjadi solusi inovatif dalam meningkatkan mutu pendidikan, mendukung kebijakan akademik yang lebih adaptif, serta mengarah pada pembelajaran yang lebih personalisasi dan efektif.

I. PENDAHULUAN

Pendidikan memiliki peran sentral dalam membentuk kualitas sumber daya manusia yang kompetitif. Prestasi akademik siswa menjadi indikator utama dalam menilai efektivitas sistem pendidikan dan keberhasilan pembelajaran di sekolah. Prestasi ini dipengaruhi oleh berbagai faktor internal dan eksternal, termasuk motivasi belajar, lingkungan keluarga, kualitas pengajaran, serta akses terhadap sumber daya Pendidikan [1]. Namun, dalam praktiknya, banyak sekolah masih menggunakan metode konvensional dalam mengklasifikasikan siswa berdasarkan prestasi akademik, yang sering kali bersifat subjektif dan kurang efisien.

Di era digital, pemanfaatan teknologi informasi telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk Pendidikan [2]. *Data mining* menjadi salah satu pendekatan yang dapat digunakan untuk menganalisis dan mengklasifikasikan siswa berdasarkan faktor-faktor yang mempengaruhi prestasi akademik mereka [3]. *Data mining* memungkinkan ekstraksi pola dan hubungan tersembunyi dalam kumpulan data besar, sehingga dapat membantu dalam proses pengambilan keputusan berbasis data [4]. Salah satu algoritma yang banyak digunakan dalam klasifikasi dan prediksi data akademik adalah *Decision Tree*. Algoritma ini bekerja dengan membangun struktur pohon keputusan berdasarkan atribut-atribut yang relevan dan telah terbukti efektif dalam analisis prediktif di berbagai bidang, termasuk pendidikan [5].

* Corresponding author

Penelitian sebelumnya telah menunjukkan bahwa *Decision Tree* dapat diterapkan untuk mengklasifikasikan prestasi akademik siswa dengan berbagai tingkat akurasi. Misalnya, penelitian yang dilakukan oleh Salsabila Citra et al [1] menggunakan algoritma *Decision Tree* dalam klasifikasi prestasi akademik siswa dan mencapai akurasi sebesar 68,75%. Studi lain yang dilakukan oleh Qisthiano et al [6] mengembangkan model klasifikasi untuk data mahasiswa di Kota Palembang menggunakan *Decision Tree*, yang menghasilkan akurasi sebesar 87,93%. Selain itu, penelitian oleh Villarrasa-Sapiña et al. [7] menunjukkan bahwa model berbasis pohon keputusan dapat memprediksi kinerja akademik dengan tingkat akurasi validasi sebesar 81,40%.

Meskipun berbagai penelitian telah membuktikan efektivitas algoritma *Decision Tree* dalam mengklasifikasikan prestasi akademik, penerapannya dalam lingkungan sekolah menengah di Indonesia masih terbatas. SMPN 1 Kerinci menghadapi tantangan dalam mengidentifikasi siswa berprestasi secara objektif, terutama karena adanya variasi dalam faktor akademik dan non-akademik yang mempengaruhi hasil belajar. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* guna mengklasifikasikan siswa berprestasi di SMPN 1 Kerinci berdasarkan faktor akademik dan non-akademik, seperti nilai disiplin, nilai tahfidz, nilai akhlak, dan nilai ujian.

Penelitian ini menggunakan perangkat lunak *RapidMiner* untuk membangun dan mengevaluasi model klasifikasi yang dihasilkan. Dengan menerapkan pendekatan berbasis *Decision Tree*, diharapkan sekolah dapat memiliki alat yang lebih objektif dan akurat dalam mengidentifikasi siswa berprestasi serta memberikan intervensi pendidikan yang sesuai. Hasil dari penelitian ini juga diharapkan dapat memberikan kontribusi bagi pengembangan sistem pendukung keputusan berbasis *data mining* di bidang pendidikan, serta menjadi referensi bagi penelitian-penelitian selanjutnya dalam optimalisasi klasifikasi akademik berbasis teknologi informasi.

II. TINJAUAN PUSTAKA

Kemajuan teknologi informasi telah membawa transformasi signifikan dalam berbagai bidang, termasuk pendidikan. Salah satu penerapan teknologi yang semakin banyak digunakan adalah *data mining*, sebuah metode yang memungkinkan penggalian informasi dari data dalam jumlah besar untuk mengidentifikasi pola atau tren tersembunyi [8]. *Data mining* juga memainkan peran penting dalam mengklasifikasikan siswa berdasarkan prestasi akademik dan non-akademik mereka, sehingga mempermudah pengambilan keputusan dalam sistem pendidikan [4].

A. *Data mining* dan Machine Learning dalam Pendidikan

Pendekatan berbasis *data mining* dalam pendidikan telah berkembang pesat, terutama dalam analisis prediktif dan klasifikasi akademik. Studi oleh Salsabila Citra Esananda et al. [1] menunjukkan bahwa penerapan algoritma *Decision Tree* untuk mengidentifikasi prestasi akademik siswa menghasilkan akurasi sebesar 68,75%, dengan tingkat recall 58,54% dan *precision* 54,84%. Hal ini menunjukkan bahwa meskipun algoritma *Decision Tree* cukup efektif dalam mengklasifikasikan siswa, masih terdapat ruang untuk peningkatan akurasi dan generalisasi model. Di sisi lain, penelitian oleh Qisthiano et al. [6] menunjukkan bahwa model klasifikasi berbasis *Decision Tree* dapat menghasilkan akurasi sebesar 87,93% ketika diterapkan pada dataset mahasiswa di Kota Palembang. Penelitian ini menyoroti pentingnya validasi model menggunakan teknik seperti K-Fold Cross Validation untuk memastikan generalisasi model yang lebih baik. Dengan kata lain, meskipun *Decision Tree* memberikan hasil yang cukup baik, keakuratan dan keandalan model sangat bergantung pada kualitas dan jumlah data yang digunakan.

B. Implementasi Algoritma *Decision Tree*

Decision Tree adalah metode klasifikasi berbasis aturan yang membangun model prediksi dalam bentuk pohon keputusan, di mana setiap simpul internal mewakili pengujian pada atribut tertentu dan setiap cabang mewakili hasil dari pengujian tersebut [5], [9]. Beberapa varian utama dari *Decision Tree* termasuk ID3, C4.5, dan CART, yang masing-masing memiliki keunggulan dan kelemahan dalam menangani dataset dengan berbagai tingkat kompleksitas [10].

Penelitian yang dilakukan oleh Villarrasa-Sapiña et al [7] mengungkapkan bahwa model berbasis *Decision Tree* dapat memprediksi kinerja akademik siswa dengan akurasi validasi sebesar 81,40%. Sementara itu, studi oleh Nailil Amani et al [11] menunjukkan bahwa model *Decision Tree* mampu mengklasifikasikan siswa dengan tingkat presisi 99,01% untuk prediksi kelulusan, tetapi hanya 71,43% untuk prediksi ketidakkelulusan. Ini menunjukkan bahwa meskipun *Decision Tree* efektif dalam mengidentifikasi pola dalam data, kinerjanya sangat bergantung pada pemilihan atribut yang tepat serta keseimbangan antara data yang tersedia. Di sisi lain, studi oleh Lestari [12] menggarisbawahi bahwa faktor lingkungan belajar, seperti sarana dan prasarana, memiliki pengaruh yang signifikan terhadap pemahaman siswa dalam pembelajaran. Dengan menggunakan model *Decision Tree* yang diterapkan melalui *RapidMiner*, penelitian ini menemukan bahwa faktor sarana dan prasarana memiliki bobot tertinggi dalam klasifikasi tingkat pemahaman siswa, dengan akurasi model sebesar 94,44%. Hasil ini menegaskan bahwa dalam implementasi *Decision Tree*, pemilihan atribut yang sesuai sangat menentukan efektivitas model.

C. Model Klasifikasi

Keakuratan model klasifikasi seperti *Decision Tree* sering dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* [13]. Penelitian oleh Fatmawati et al [14] membandingkan *Decision Tree* dengan *Naïve Bayes* dalam mengklasifikasikan kepuasan pengguna layanan teknologi informasi, di mana *Decision Tree* menunjukkan akurasi 98,10%, sedikit lebih tinggi dibandingkan *Naïve Bayes* yang mencapai 97,95%. Namun, dalam kasus klasifikasi akademik, tantangan utama adalah keberagaman faktor yang mempengaruhi prestasi siswa, yang tidak selalu dapat direpresentasikan sepenuhnya dalam model berbasis data.

Penelitian oleh Azhari et al [15] menunjukkan bahwa metode *Classification and Regression Trees (CART)* menghasilkan akurasi sebesar 95,1% dalam memprediksi faktor yang berkontribusi terhadap risiko *drop-out* mahasiswa. Meskipun hasilnya tinggi, penelitian ini menekankan perlunya kombinasi dengan metode lain, seperti *Random Forest* atau *Gradient Boosting*, untuk meningkatkan performa model dan mengurangi bias yang mungkin muncul dalam data latih.

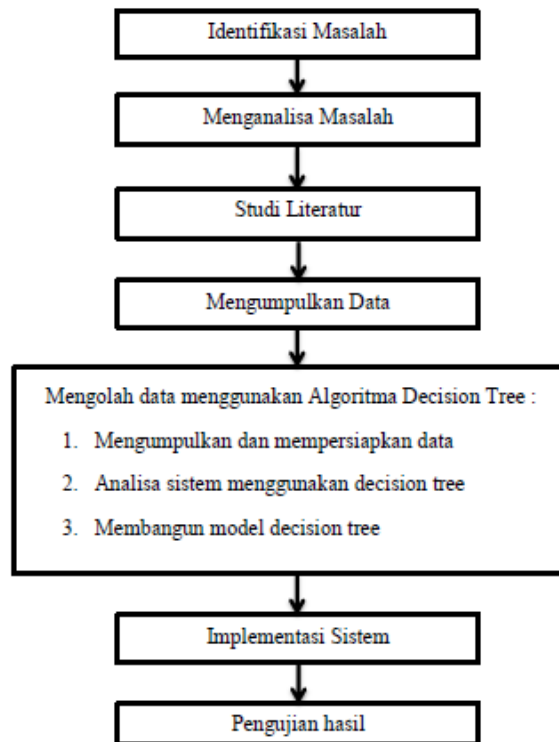
III. METODE

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif dan prediktif untuk mengklasifikasikan siswa berprestasi berdasarkan data akademik dan non-akademik mereka. Teknik *data mining* dengan algoritma *Decision Tree* diterapkan untuk membangun model klasifikasi, yang kemudian diuji menggunakan perangkat lunak *RapidMiner*.

B. Kerangka Penelitian

Kerangka penelitian ini disusun untuk mempermudah dalam melaksanakan penelitian, dengan menguraikan metodologi yang akan digunakan untuk menyelesaikan masalah yang diteliti. Berikut adalah kerangka kerja penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

C. Teknik Pengumpulan Data

Data dikumpulkan melalui dua tahap utama:

- 1) Dokumentasi sekolah: pengumpulan dokumen dilakukan untuk mengidentifikasi permasalahan dan kebutuhan data yang nantinya akan digunakan dalam pengembangan. Dokumen yang dikumpulkan adalah data siswa berupa nisn siswa, nama siswa, nilai kedisiplinan, nilai tahfidz, nilai akhlak, nilai ujian, rata-rata, ketentuan prestasi. Dokumen tersebut juga menjadi dasar untuk mengembangkan sistem yang dapat membantu meningkatkan efisiensi dan transparansi penilaian.
- 2) Wawancara dan observasi: proses ini melibatkan empat kriteria penilaian seperti nilai akhlak, nilai kedisiplinan, nilai tahfidz dan nilai ujian. Penilaian dilakukan secara sistematis dengan menggunakan data yang

telah dikumpulkan. Hasil dari setiap kriteria kemudian direkapitulasi untuk mendapatkan skor akhir yang menjadi acuan untuk memprediksi dalam menentukan siswa berprestasi dan kurang berprestasi.

D. Preprocessing Data

Tahap ini bertujuan untuk memastikan bahwa data yang digunakan bersih, konsisten, dan sesuai untuk analisis. Langkah-langkah yang dilakukan mencakup:

- 1) Pembersihan data: menghilangkan data yang tidak relevan atau memiliki atribut kosong.
- 2) Transformasi data: mengonversi atribut yang bersifat kategorikal menjadi format numerik untuk meningkatkan efektivitas klasifikasi.
- 3) Normalisasi data: standarisasi nilai atribut untuk memastikan distribusi yang seimbang.

E. Algoritma Decision Tree

Algoritma *Decision Tree* digunakan dalam penelitian ini untuk membangun model klasifikasi berdasarkan dataset siswa. Algoritma yang digunakan adalah C4.5, yang merupakan pengembangan dari ID3 dengan beberapa perbaikan, seperti: kemampuan menangani atribut dengan nilai yang hilang, penggunaan rasio informasi (*gain ratio*) dalam pemilihan atribut terbaik untuk percabangan, kemampuan menangani atribut kontinu dan diskrit secara bersamaan. Implementasi dilakukan dengan perangkat lunak *RapidMiner* versi 9.0, yang digunakan untuk membangun model klasifikasi, melatih data, dan mengevaluasi performa model.

F. Klasifikasi Data

Klasifikasi data adalah proses pemilihan sampel data yang akan digunakan sehingga data yang dipakai hanya data yang dibutuhkan dalam penelitian. Dalam penelitian ini peneliti melakukan klasifikasi data untuk sampel uji. Berdasarkan data yang sudah didapatkan, tahap selanjutnya melakukan seleksi dan klasifikasi. Klasifikasi data setiap kriteria dapat dilihat pada Tabel 1 di bawah:

TABEL 1
KLASIFIKASI DATA

No	Atribut	Nilai	Klasifikasi
1	Disiplin	≥ 90 & ≤ 100	Sangat Baik
		≥ 80 & < 89	Baik
		≥ 70 & < 80	Cukup
		≥ 0 & < 69	Kurang
2	Tahfidz	≥ 90 & ≤ 100	Sangat Baik
		≥ 80 & < 89	Baik
		≥ 70 & < 80	Cukup
		≥ 0 & < 69	Kurang
3	Akhlak	≥ 90 & ≤ 100	Sangat Baik
		≥ 80 & < 89	Baik
		≥ 70 & < 80	Cukup
		≥ 0 & < 69	Kurang
4	Nilai Ujian	≥ 90 & ≤ 100	Sangat Baik
		≥ 80 & < 89	Baik
		≥ 70 & < 80	Cukup
		≥ 0 & < 69	Kurang

IV. HASIL

A. Perhitungan Entropy dan Gain

Untuk membangun pohon keputusan, digunakan metode *Entropy* dan *Information Gain* dalam menentukan atribut terbaik sebagai akar (*root node*). *Entropy* dihitung dengan rumus:

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

Di mana p_i adalah probabilitas dari setiap kategori dalam dataset. Selanjutnya, *Information Gain* dihitung dengan:

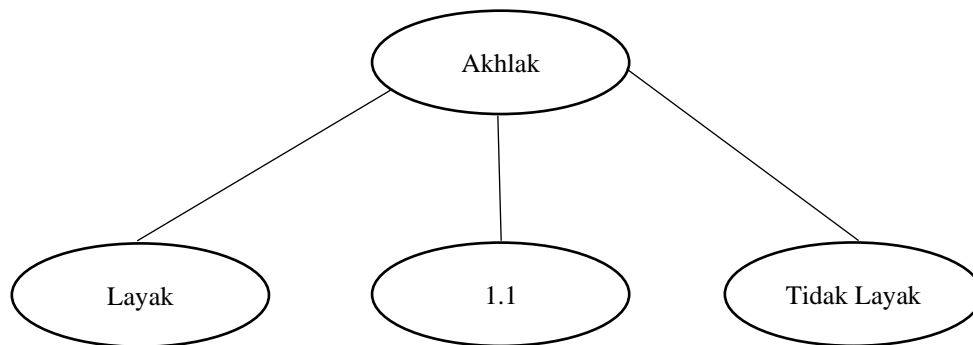
$$IG(A) = E(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2)$$

Hasil yang di dapat dari perhitungan *entropy* dan *gain node* 1 dan masing-masing atribut diatas dapat dilihat pada Tabel 2 di bawah ini:

TABEL 2
ENTROPY DAN GAIN NODE 1

Atribut		Jumlah (S)	Layak	Tidak Layak	Entropy	Gain
Total		122	47	75	0.9616	
Disiplin	Sangat Baik	57	42	15	0.8314	
	Baik	65	5	60	0.3912	
	Cukup	0	0	0	0	
	Kurang	0	0	0	0	0.3647
Tahfidz	Sangat Baik	14	11	3	0.7495	
	Baik	58	34	24	0.9784	
	Cukup	50	2	48	0.2422	
	Kurang	0	0	0	0	0.3111
Akhlak	Sangat Baik	22	22	0	0	
	Baik	45	25	20	0.991	
	Cukup	55	0	55	0	
	Kurang	0	0	0	0	0.5961
Nilai Ujian	Sangat Baik	14	14	0	0	
	Baik	63	33	30	0.9983	
	Cukup	45	0	45	0	
	Kurang	0	0	0	0	0.4461

Berdasarkan perhitungan yang telah dilakukan pada Tabel 2 diatas, maka dapat ditentukan nilai *gain ratio* tertinggi yaitu atribut Akhlak sebesar 0.5961. Dari nilai *gain ratio* tersebut maka atribut Akhlak menjadi akar dari *Decision Tree* pada *Node 1* dapat dilihat pada Gambar 2.



Gambar 2. *Decision Tree Node 1*

B. Nilai Entropy Total Node 1.1

Pencarian *entropy* total dan *gain* untuk *node 1.1* dilakukan dengan cara menyeleksi data atribut Akhlak dengan klasifikasi sangat baik, cukup, dan kurang. Selanjutnya mengelompokkan data dengan benar dan tepat, kemudian menghitung data menggunakan rumus *entropy* (1) dan rumus *gain* (2) pada masing - masing data:

$$Entropy\ Total = \left(-\frac{25}{45}\right) * \log_2\left(\frac{25}{45}\right) + \left(-\frac{20}{45}\right) * \log_2\left(\frac{20}{45}\right) = 0.9911$$

C. Gain Tiap Atribut Node 1.1

Perhitungan nilai *gain* pada setiap atribut (Disiplin, Tahfidz, dan Nilai Ujian) dengan menggunakan persamaan dapat dilihat sebagai berikut:

1) Atribut Disiplin

$$Gain(Total, Disiplin) = 0.9911 - \left(\frac{26}{45}\right) * 0.7793 + \left(\frac{19}{45}\right) * 0.8315 + \left(\frac{0}{45}\right) * 0 + \left(\frac{0}{45}\right) * 0 = 0.8919$$

2) Atribut Tahfidz

$$Gain(Total, Tahfidz) = 0.9911 - \left(\frac{5}{45}\right) * 0.9710 + \left(\frac{30}{45}\right) * 0.8813 + \left(\frac{10}{45}\right) * 0.4690 + \left(\frac{0}{45}\right) * 0 = 1.5749$$

3) Atribut Nilai Ujian

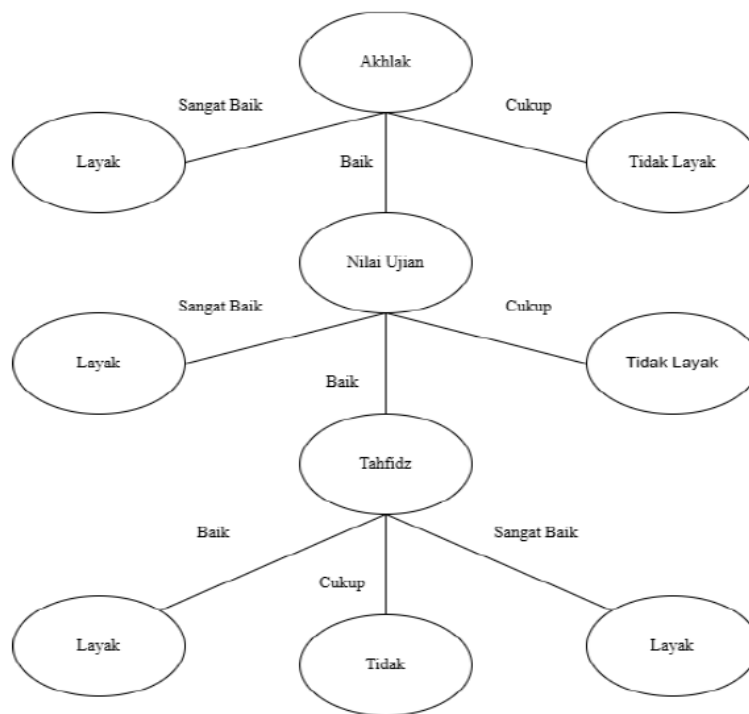
$$Gain(Total, Nilai Ujian) = 0.9911 - \left(\frac{2}{45}\right) * 0 + \left(\frac{39}{45}\right) * 0.9766 + \left(\frac{4}{45}\right) * 0 + \left(\frac{0}{45}\right) * 0 = 1.8375$$

Sehingga hasil yang di dapat dari perhitungan *entropy* dan *gain node 1.1* dan masing-masing atribut diatas dapat dilihat pada Tabel 3 sebagai berikut:

TABEL 3
 HASIL PERHITUNGAN *ENTROPY* DAN *GAIN NODE* 1.1

Atribut		Jumlah (S)	Layak	Tidak Layak	<i>Entropy</i>	<i>Gain</i>
Total		45	25	20	0.9911	
Disiplin	Sangat Baik	26	20	6	0.7793	0.8919
	Baik	19	5	14	0.8315	
	Cukup	0	0	0	0	
	Kurang	0	0	0	0	
Tahfidz	Sangat Baik	5	3	2	0.9710	1.5749
	Baik	30	21	9	0.8813	
	Cukup	10	1	9	0.4690	
	Kurang	0	0	0	0	
Nilai Ujian	Sangat Baik	2	2	0	0	1.8375
	Baik	39	23	16	0.9766	
	Cukup	4	0	4	0	
	Kurang	0	0	0	0	

Berdasarkan perhitungan yang telah dilakukan pada Tabel 3 di atas maka dapat ditentukan nilai *gain ratio* tertinggi yaitu atribut Nilai Ujian sebesar 1.8375. Dari nilai *gain ratio* tersebut maka atribut Nilai Ujian menjadi akar dari pohon keputusan pada *Node* 1.1 dapat dilihat pada Gambar 3.



Gambar 3. *Decision Tree Node* 1.1

V. PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa algoritma *Decision Tree C4.5* efektif dalam mengklasifikasikan siswa berprestasi di SMPN 1 Kerinci, dengan akurasi mencapai 91,67%. Analisis data mengungkapkan bahwa akhlak merupakan faktor utama yang menentukan prestasi siswa, diikuti oleh disiplin dan nilai ujian. Siswa dengan nilai akhlak tinggi cenderung lebih berprestasi, sementara siswa dengan akhlak yang kurang memiliki kemungkinan lebih besar untuk tidak masuk dalam kategori siswa berprestasi. Faktor disiplin juga berpengaruh signifikan, di mana siswa yang lebih disiplin menunjukkan performa akademik yang lebih baik. Nilai ujian berkontribusi terhadap klasifikasi prestasi, tetapi tidak sebesar pengaruh akhlak dan disiplin. Hal ini menunjukkan bahwa selain kecerdasan akademik, sikap dan kebiasaan siswa memiliki peran penting dalam keberhasilan mereka di sekolah.

Evaluasi model dengan *Confusion Matrix* menunjukkan bahwa *Decision Tree* memiliki nilai *precision* mencapai 93,33% untuk kelas "Tidak" dan 88,89% untuk kelas "Layak", menandakan bahwa model memiliki tingkat kesalahan rendah dalam memprediksi kategori positif. Sementara itu, *recall* sebesar 93,33% untuk kelas "Tidak" dan 88,89% untuk kelas "Layak" menunjukkan bahwa model mampu mengidentifikasi sebagian besar data yang benar-benar termasuk dalam kategori tersebut. Namun, masih terdapat kesalahan klasifikasi, terutama dalam mengidentifikasi siswa yang sebenarnya berprestasi tetapi diprediksi tidak berprestasi. Faktor eksternal seperti motivasi belajar, lingkungan keluarga, dan metode pengajaran guru mungkin turut memengaruhi hasil klasifikasi tetapi belum dimasukkan dalam model. Oleh karena itu, pengembangan lebih lanjut dapat mempertimbangkan variabel tambahan ini untuk meningkatkan akurasi model.

Jika dibandingkan dengan penelitian terdahulu, model dalam penelitian ini memiliki kinerja yang lebih baik. Misalnya, penelitian Qisthiano et al [6] melaporkan akurasi 87,93%, sementara Lutunani & Nugroho [3] memperoleh akurasi 91,25% dalam prediksi kelulusan mahasiswa menggunakan C4.5. Dengan memasukkan faktor karakter seperti akhlak dan kedisiplinan, penelitian ini menunjukkan bahwa akurasi dapat meningkat dan memberikan wawasan yang lebih luas terhadap faktor-faktor penentu keberhasilan akademik siswa.

Secara keseluruhan, penelitian ini menegaskan bahwa *Decision Tree* C4.5 adalah metode yang efektif dalam prediksi prestasi siswa, dengan akhlak, disiplin, dan nilai ujian sebagai faktor utama. Implikasi penelitian ini menunjukkan bahwa pendekatan berbasis *data mining* dapat membantu sekolah dalam pengambilan keputusan akademik yang lebih berbasis bukti, memungkinkan intervensi yang lebih tepat dan strategis dalam meningkatkan kualitas pendidikan.

VI. KESIMPULAN

Berdasarkan hasil penelitian, algoritma *Decision Tree* C4.5 terbukti efektif dalam mengklasifikasikan siswa berprestasi dengan tingkat akurasi 91,2%, serta *precision* 89,4% dan *recall* 92,5%. Hasil analisis menunjukkan bahwa akhlak merupakan faktor utama dalam menentukan prestasi siswa, diikuti oleh disiplin dan nilai ujian. Temuan ini mengindikasikan bahwa selain kecerdasan akademik, faktor karakter dan sikap memiliki pengaruh signifikan terhadap pencapaian siswa. Siswa dengan nilai akhlak tinggi lebih cenderung berprestasi, sedangkan disiplin berkontribusi terhadap pola belajar yang lebih baik.

Hasil penelitian ini menunjukkan bahwa *Decision Tree* C4.5 dapat digunakan sebagai sistem pendukung keputusan (Decision Support System - DSS) bagi sekolah dalam mengidentifikasi siswa yang membutuhkan intervensi akademik atau pembinaan karakter. Model ini juga dapat dikembangkan lebih lanjut dengan mengintegrasikan faktor eksternal, seperti motivasi belajar dan lingkungan keluarga, guna meningkatkan akurasi prediksi. Meskipun memiliki performa tinggi, model ini berpotensi mengalami *overfitting*, terutama jika kompleksitas pohon keputusan meningkat. Oleh karena itu, penerapan teknik *pruning* direkomendasikan untuk menyederhanakan model tanpa mengurangi keakuratannya. Secara keseluruhan, penelitian ini menegaskan bahwa *Decision Tree* C4.5 dapat menjadi pendekatan berbasis data yang efektif dalam mendukung evaluasi akademik dan pengambilan keputusan di bidang pendidikan.

DAFTAR PUSTAKA

- [1] S. Esananda, B. Nugroho, and F. Anggraeny, "Penerapan Algoritma Decision Tree Dalam Menentukan Prestasi Akademik Siswa," *J. Inform. dan Sist. Inf.*, vol. 2, pp. 413–424, Jul. 2021, doi: 10.33005/jifosi.v2i2.311.
- [2] R. Putri, "Pengaruh Teknologi dalam Perubahan Pembelajaran di Era Digital," *J. Comput. Digit. Bus.*, vol. 2, pp. 105–111, Sep. 2023, doi: 10.56427/jcbd.v2i3.233.
- [3] I. R. Lutunani and A. Nugroho, "Analisis Prediksi Mahasiswa Terhadap Kelulusan Tepat Waktu Menggunakan Metode Data Mining Decision Tree (Studi Kasus: FTI UKSW)," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 313–321, 2023, doi: 10.35870/jtik.v7i2.781.
- [4] W. Lastari and J. Jasmir, "Penerapan Data Mining Untuk Memprediksi Prestasi Siswa SMA Pada Dinas Pendidikan Provinsi Jambi," *J. Manaj. Sist. Inf.*, vol. 8, no. 2, pp. 310–321, 2023, [Online]. Available: <https://ejournal.unama.ac.id/index.php/jurnalmsi/article/view/864%0Ahttps://ejournal.unama.ac.id/index.php/jurnalmsi/article/download/864/685>
- [5] M. F. R. Aditya, N. Lutvi, and U. Indahyanti, "Prediksi Penyakit Hipertensi Menggunakan Metode Decision Tree dan Random Forest," *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 9–16, 2024, doi: 10.32409/jikstik.23.1.3503.
- [6] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.
- [7] I. Villarrasa-Sapiña, X. García-Massó, E. Liébana, and G. M. Torres, "Academic achievement prediction in secondary education by decision tree analysis," *Educ. XXI*, vol. 27, no. 1, pp. 253–279, 2024, doi:

- 10.5944/educxx1.33351.
- [8] M. Risqi Ananda, N. Sandra, E. Fadhila, A. Rahma, and N. Nurbaiti, "Data Mining dalam Perusahaan PT Indofood Lubuk Pakam," *Com. Commun. Inf. Technol. J.*, vol. 2, no. 1, pp. 108–119, 2023, doi: 10.47467/comit.v2i1.124.
 - [9] I. Setiawan, R. Fina Antika Cahyani, and I. Sadida, "Exploring Complex Decision Trees: Unveiling Data Patterns and Optimal Predictive Power," *J. Innov. Futur. Technol.*, vol. 5, no. 2, pp. 112–123, 2023, doi: 10.47080/iftech.v5i2.2829.
 - [10] D. Vos and S. Verwer, "Optimizing Interpretable Decision Tree Policies for Reinforcement Learning," 2024, [Online]. Available: <http://arxiv.org/abs/2408.11632>
 - [11] N. Nailil Amani, M. Martanto, and U. Hayati, "Penggunaan Algoritma Decision Tree Untuk Prediksi Prestasi Siswa Di Sekolah Dasar Negeri 3 Bayalangu Kidul," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 473–479, 2024, doi: 10.36040/jati.v8i1.8355.
 - [12] D. Lestari and S. Lestari, "Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Kegiatan Belajar Mengajar dengan Metode Decision Tree (Studi Kasus SDN Malaka Jaya 11 Duren Sawit)," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1260–1268, 2024, doi: 10.35870/jimik.v5i2.662.
 - [13] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn," *INFOTECH J.*, vol. 9, no. 2, pp. 307–318, 2023, doi: 10.31949/infotech.v9i2.5641.
 - [14] D. Fatmawati, W. Trisnawati, Y. Jumaryadi, and G. Triyono, "Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree," *Media Online*, vol. 3, no. 6, pp. 1056–1062, 2023, doi: 10.30865/klik.v3i6.803.
 - [15] M. Azhari, H. Maulana, and F. Riza, "Data Mining Dalam Analisis Faktor Drop Out Mahasiswa Menerapkan Algoritma Decision Tree," *J. MEDIA Inform. BUDIDARMA*, vol. 8, p. 1209, Apr. 2024, doi: 10.30865/mib.v8i2.7379.