

PREDIKSI PRESTASI AKADEMIK SISWA TERBAIK MENGGUNAKAN ALGORITMA DECISION TREE BERBASIS DATA HISTORIS

Septiar Maulana, Agyztia Premana, Bambang Irawan
Teknik Informatika, Universitas Muhadi Setiabudi Brebes
Jl P. Diponegoro No.KM2, Kabupaten Brebes, Indonesia
septiarmaulana29@gmail.com

ABSTRAK

Kabupaten Brebes menghadapi tantangan signifikan dalam kualitas pendidikan, terutama di MTs Arrobiah Azzain Kecamatan Bulakamba, di mana proses evaluasi siswa masih manual dan berpotensi subjektif. Proses evaluasi siswa yang masih manual ini menjadi permasalahan yang perlu diatasi. Penelitian ini bertujuan untuk penerapan algoritma Decision Tree berbasis Data Mining untuk mengoptimalkan sistem prediksi lulusan terbaik melalui analisis variabel nilai akademik semester 5, prestasi akademik, keaktifan, dan kehadiran. Penelitian menggunakan metode kuantitatif dengan pengumpulan 519 data historis siswa (70% training, 30% testing) yang diolah dalam Rapidminer, didahului studi literatur, observasi, dan wawancara dengan stakeholder madrasah. Hasilnya menunjukkan akurasi prediksi mencapai 97.80% dengan precision 98.38% untuk kategori 'Baik' dan 94.55% untuk 'Terbaik', di mana prestasi akademik terbukti menjadi atribut dominan dalam pohon keputusan.

Kata kunci : *Decision Tree*, Prediksi Prestasi Akademik, Teknologi Pendidikan

1. PENDAHULUAN

Kabupaten Brebes, sebagai salah satu wilayah terluas di Provinsi Jawa Tengah, menghadapi tantangan signifikan dalam sektor pendidikan. Meskipun partisipasi pendidikan tingkat menengah pertama mencapai 85,17%, terdapat ketimpangan kualitas pendidikan antarwilayah, terutama di Kecamatan Bulakamba yang memiliki 23 lembaga pendidikan tingkat menengah pertama, baik negeri maupun swasta [1]. Ribuan siswa di jenjang ini menciptakan persaingan ketat untuk menghasilkan lulusan berkualitas yang siap menghadapi tuntutan masyarakat. Namun, proses evaluasi dan penilaian prestasi siswa saat ini belum optimal, sebagian besar masih dilakukan secara manual, memakan waktu lama, dan rentan terhadap kesalahan manusia.

Khususnya di madrasah seperti MTs Arrobiah Azzain, kebutuhan akan sistem penilaian dan prediksi prestasi siswa yang efektif dan efisien semakin mendesak. Metode konvensional yang menggunakan rapat evaluasi berulang dengan mempertimbangkan nilai akademik semester lima, prestasi, keaktifan dalam kegiatan sekolah, dan tingkat kehadiran, belum mampu memenuhi tantangan objektivitas, transparansi, dan efisiensi dalam seleksi lulusan terbaik. Hal ini menjadi semakin krusial seiring dengan meningkatnya jumlah siswa dan kompleksitas data yang harus dianalisis [2].

Kemajuan teknologi informasi memungkinkan pemanfaatan Data Mining sebagai solusi untuk permasalahan penilaian. Algoritma Decision Tree dapat mengolah data secara sistematis dan otomatis, menghasilkan pola prediksi prestasi siswa yang akurat dan adil. Algoritma ini menganalisis berbagai variabel secara simultan, mempercepat pengambilan keputusan, dan meminimalkan subjektivitas. Implementasi ini menggunakan perangkat lunak

Rapidminer untuk mengolah data historis empat angkatan, dengan tahap awal menggunakan Microsoft Excel sebagai alat preprocessing [3].

Beberapa penelitian sebelumnya telah mengkaji penggunaan algoritma Decision Tree dalam penilaian pendidikan. Misalnya, penelitian berjudul "Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi" menunjukkan bahwa algoritma ini dapat mempertimbangkan faktor akademik dan non-akademik, dengan akurasi mencapai 91,67% [4]. Penelitian lain, "Model Decision Tree Untuk Prediksi Prestasi Akademik Matematika Siswa Kelas VIII SMP Frater Don Bosco Manado," menggunakan data dari 150 siswa dan mencapai akurasi 95,56% dengan seleksi fitur, mengidentifikasi faktor kunci seperti "Durasi Belajar" dan "Dukungan Keluarga" [5]. Penelitian lain, "Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus: SMKN 2 Padang)," menggunakan data dari 321 siswa dan mencapai akurasi 90,31%, mengidentifikasi faktor kunci seperti "Motivasi Siswa" [6]. Temuan-temuan ini menegaskan perlunya integrasi teknologi informasi dalam sistem penilaian pendidikan, yang menjadi dasar bagi penelitian ini untuk mengisi gap yang ada.

Pendekatan ini melibatkan pengukuran prestasi siswa tidak hanya berdasarkan nilai akademik, tetapi juga faktor penting seperti prestasi akademik, tingkat keaktifan, dan kehadiran, sehingga menghasilkan penilaian yang lebih holistik dan objektif [7]. Penelitian ini diarahkan untuk mengevaluasi keefektifan penerapan algoritma Decision Tree dalam mendukung seleksi lulusan terbaik secara transparan dan efisien, sekaligus mengidentifikasi variabel-variabel yang paling berpengaruh dalam proses penilaian prestasi siswa di MTs Arrobiah Azzain [8]. Penelitian ini diharapkan menjadi dasar digitalisasi

pengambilan keputusan pendidikan di madrasah, meningkatkan kualitas penilaian dan prediksi prestasi siswa, serta menjadi referensi bagi institusi lain dalam memanfaatkan teknologi informasi untuk mutu pendidikan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pada penelitian yang saya lakukan, ada beberapa penelitian terdahulu yang saya jadikan referensi, yaitu artikel dengan judul Eksplorasi Algoritma *Decision Tree* untuk Penentuan Siswa Berprestasi, Penelitian ini memiliki kelebihan Keunggulan penelitian algoritma *Decision Tree* untuk klasifikasi siswa berprestasi di SMPN 1 Kerinci terletak pada kemampuannya untuk mempertimbangkan faktor akademik dan non-akademik, serta menghasilkan model dengan akurasi tinggi mencapai 91,67%. Penekanan pada atribut karakter siswa, seperti akhlak dan kedisiplinan, menunjukkan relevansi dalam pengambilan keputusan Pendidikan [4]. Penelitian ini juga mereferensi pada artikel dengan judul Model *Decision Tree* Untuk Prediksi Prestasi Akademik Matematika Siswa Kelas VIII SMP Frater Don Bosco Manado, Kelebihan penelitian ini adalah penggunaan metode deskriptif kuantitatif dengan model *Decision Tree* untuk mengidentifikasi faktor-faktor yang mempengaruhi prestasi akademik siswa kelas VIII di SMP Frater Don Bosco Manado. Dengan data dari 150 siswa, model dengan seleksi fitur mencapai akurasi 95,56%, lebih tinggi dibandingkan tanpa seleksi fitur (93,33%). Faktor kunci seperti "Durasi Belajar," "Tingkat Pendidikan Ayah," dan "Dukungan Keluarga" memberikan wawasan penting untuk intervensi akademik, sementara penanganan outliers dan ketidakseimbangan kelas meningkatkan nilai analisis [5]. Penelitian ini juga mereferensi pada artikel dengan judul *Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus: SMKN 2 Padang)*. Kelebihan penelitian ini adalah penggunaan metode kuantitatif dengan teknik data mining berbasis algoritma *Decision Tree* J48, yang diimplementasikan melalui aplikasi WEKA. Penelitian dilakukan pada 450 siswa kelas XI SMKN 2 Padang, dengan mempertimbangkan faktor status sosial ekonomi orang tua, kedisiplinan, dan motivasi siswa. Setelah proses preprocessing, sebanyak 321 data siswa dianalisis, dan model yang dihasilkan mencapai akurasi 90,31% dengan nilai ROC sebesar 0,987, yang termasuk kategori *Excellent Classification*. Faktor kunci seperti "Motivasi Siswa" ditemukan sebagai variabel yang paling menentukan dalam memprediksi prestasi belajar, memberikan wawasan penting untuk strategi peningkatan kualitas pendidikan dan pengambilan keputusan yang lebih berbasis data [6].

2.2. Data Mining

Data Mining adalah proses yang memungkinkan individu untuk mengakses dan menganalisis sejumlah

besar data dengan cepat. Proses ini melibatkan pengumpulan dan pengolahan data untuk mengidentifikasi informasi penting yang mungkin tersembunyi. Tujuan utama *Data Mining* adalah mengekstrak wawasan bermanfaat dari data, baik dalam bentuk angka, pola, maupun informasi lain untuk berbagai keperluan. Penerapan *Data Mining* dapat ditemukan dalam berbagai aspek kehidupan, termasuk bisnis, penelitian, pendidikan, dan sektor lainnya [9].

2.3. Decision Tree

Decision Tree adalah sebuah algoritma yang digunakan dalam analisis data dan pembelajaran mesin untuk membangun model klasifikasi. Algoritma ini dirancang untuk menghasilkan struktur pohon keputusan yang dapat digunakan untuk mengklasifikasikan data berdasarkan atribut-atribut yang tersedia [10]. satu keunggulan utama dari algoritma ini adalah kemampuannya dalam menangani nilai atribut yang hilang, sehingga analisis tetap dapat dilakukan dengan akurat meskipun terdapat kekurangan data [7]. Berikut adalah langkah-langkah dalam membangun algoritma *Decision Tree*:

- Persiapan *Data Training* adalah data historis yang dikelompokkan dalam kategori berbeda, berfungsi sebagai fondasi model klasifikasi.
- Menghitung Akar Pohon dilakukan dengan memilih atribut berdasarkan nilai *gain* tertinggi sebagai akar pertama. Langkah awal adalah menghitung nilai *Entropy*.
- Pengulangan Proses adalah langkah yang diulang hingga semua rekaman dalam dataset terproses, menghasilkan cabang baru pada pohon keputusan berdasarkan atribut yang dipilih.
- Proses pembentukan *Decision Tree* akan berhenti ketika salah satu dari kondisi berikut terpenuhi: (a) semua data dalam *node* memiliki label kelas yang sama, (b) tidak ada atribut yang tersisa untuk dipilih, atau (c) jumlah data dalam *node* mencapai batas minimum yang telah ditentukan.

Pemilihan atribut sebagai akar pohon dilakukan berdasarkan atribut yang memiliki nilai *gain* tertinggi di antara atribut-atribut yang tersedia. Untuk menghitung nilai *gain*, digunakan rumus yang tercantum di bawah ini:

a. Information Gain

Information Gain adalah metrik dalam segmentasi data yang mengukur pengetahuan tentang nilai variabel acak. Terdapat hubungan berbanding terbalik antara *Information Gain* dan entropi; semakin tinggi *Information Gain*, semakin baik kualitas informasi yang diperoleh [11][12]. Pada persamaan (1), S adalah Himpunan Kasus, A adalah Atribut, n adalah Jumlah partisi atribut A , S_i adalah Jumlah kasus pada partisi ke- i , S adalah Jumlah kasus dalam S .

$$\text{Information Gain}(S, A) = \text{Entropi}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropi}(S_i) \quad (1)$$

b. Entropy

Entropi adalah ukuran yang mencerminkan tingkat ketidakpastian yang terkait dengan variabel acak. Nilai entropi akan meningkat seiring dengan bertambahnya ketidakpastian atau keacakan, dan sebaliknya akan menurun ketika ketidakpastian berkurang. Nilai entropi ini berkisar antara 0 hingga 1 [11][12]. Pada persamaan (2), S adalah Himpunan Kasus, n adalah Jumlah partisi atribut S , sedangkan P_i adalah Proporsi dari S terhadap S .

$$\text{Entropi}(S) = \sum_{i=1}^n -P_i \times \log_2(P_i) \quad (2)$$

2.4. Rapih Miner

Rapidminer adalah platform perangkat lunak yang bermanfaat dalam pengolahan data, ilmu data, dan pembelajaran mesin. Dirancang untuk mempermudah analisis data, *Rapidminer* menawarkan berbagai alat dan fitur untuk setiap tahap proses pengolahan data, dari persiapan hingga penerapan model. Keunggulan utama *Rapidminer* adalah aksesibilitas gratis bagi pengguna untuk tujuan pendidikan, menjadikannya pilihan tepat bagi pelajar dan pengajar yang ingin mendalami ilmu data tanpa biaya [13].

3. METODE PENELITIAN

3.1. Metode Pengumpulan Data

Untuk menghasilkan penelitian yang berkualitas dan akurat, diperlukan data yang valid dan relevan sebagai bahan olahan. Pengumpulan data merupakan langkah penting yang menentukan keberhasilan penelitian. Berikut adalah penjelasan mengenai ketiga metode tersebut:

a. Studi Literatur

Studi literatur adalah langkah awal untuk mengkaji teori dan temuan penelitian sebelumnya yang relevan. Metode ini melibatkan penelaahan sumber seperti buku dan jurnal ilmiah untuk membangun landasan teoritis dan mengidentifikasi celah penelitian yang dapat diisi.

b. Observasi

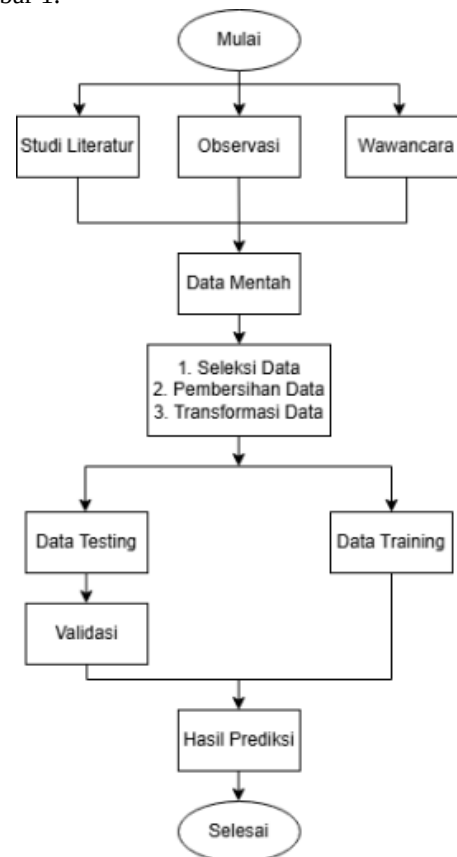
Observasi dilakukan di MTs Arrobi'ah Azzain Grinting pada siswa kelas IX untuk menentukan prediksi lulusan terbaik dengan algoritma *Decision Tree*. Metode ini memahami sistem penilaian dan faktor-faktor yang memengaruhi prediksi, sehingga data yang dikumpulkan lebih akurat.

c. Wawancara

Wawancara dengan guru, operator, dan kepala sekolah dilakukan untuk menggali kriteria penilaian lulusan terbaik. Informasi ini membantu memahami proses evaluasi dan menyediakan data kualitatif yang digunakan sebagai atribut dalam model prediksi *Decision Tree*.

3.2. Tahapan Penelitian

Bagian ini memaparkan rangkaian langkah sistematis yang dijalankan dalam penelitian, dengan tujuan memberikan gambaran jelas mengenai setiap tahapan yang dilalui dalam proses studi sesuai pada gambar 1.



Gambar 1. Tahapan Penelitian

Keterangan :

a. Mulai, pada tahap ini titik awal pengolahan data.

b. Pengumpulan Data, terdiri dari tiga metode yaitu :

- Studi Literatur
Mengumpulkan data atau informasi dari sumber-sumber tertulis (jurnal, buku, artikel).
- Observasi
Pengamatan langsung terhadap objek yang diteliti.
- Wawancara
Mengumpulkan informasi secara langsung dari narasumber.

c. Data Mentah, data awal yang belum diproses hasil dari pengumpulan data.

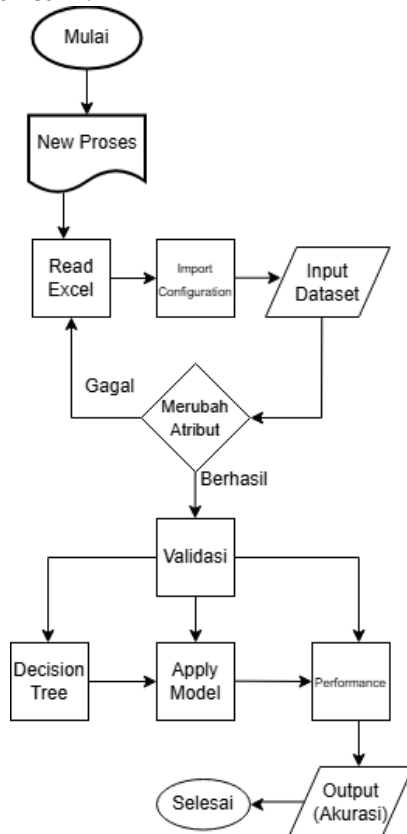
d. Pra Pemrosesan Data, terdiri dari tiga Langkah yaitu:

- Seleksi Data, Memilih data yang relevan dengan tujuan analisis.
- Pembersihan Data (*Cleaning*), menghilangkan data tidak relevan, menangani data hilang, dan memperbaiki kesalahan *input*.
- Transformasi Data
Mengubah data ke dalam format yang sesuai untuk analisis.

- e. Pembagian Data
 - Data *Training*, digunakan untuk melatih model agar dapat mempelajari pola dari data.
 - Data *Testing*, digunakan untuk menguji seberapa baik model yang sudah dilatih dalam membuat prediksi.
- f. Validasi
Validasi dilakukan pada Data *Testing*.
- g. Hasil Prediksi
Output dari sistem prediksi yang telah dilatih dan divalidasi.
- h. Selesai
Akhir dari proses pengolahan dan analisis data.

3.3. Skema Sistem

Adapun untuk memprediksi prestasi lulusan terbaik di MTs Arrobiah Azzain berdasarkan nilai akademik semester 5, prestasi *akademik*, keaktifan dalam kegiatan sekolah, dan tingkat kehadiran, digunakan Algoritma *Decision Tree*. Proses lengkapnya dapat dilihat pada flowchart berikut sesuai pada Gambar 2.



Gambar 2. Skema Sistem

Keterangan :

- a. Mulai
Titik awal untuk menjalankan proses sistem.
- b. New Proses
Memulai proses baru, menandai inisialisasi sistem untuk pemrosesan data.
- c. Read Excel
Membaca file excel yang berisi data untuk dianalisis. Jika format tidak sesuai maka proses akan gagal dan Kembali ke konfigurasi.

d. Input Configuration

Tahap konfigurasi pengaturan atribut agar sesuai dengan kebutuhan model.

e. Input Dataset

Memasukan dataset ke dalam sistem.

f. Merubah Atribut

Jika proses *input* dataset gagal, sistem meminta pengguna untuk mengubah atribut, seperti nama kolom, tipe data, atau format lain yang dibutuhkan agar bisa dikenali sistem.

g. Validasi

Mengecek data sudah sesuai dengan syarat sistem. Jika validasi berhasil maka akan lanjut ke pemodelan.

h. Decision Tree

Algoritma yang digunakan untuk membangun model prediksi berdasarkan data yang telah divalidasi.

i. Apply Model

Model *Decision Tree* yang sudah dibuat dan diaplikasikan pada data.

j. Performance

Mengukur performa model berdasarkan metrik tertentu, dalam hal ini adalah akurasi.

k. Output (Akurasi)

Hasil akhir berupa Tingkat akurasi model dalam memprediksi data.

l. Selesai

Akhir dari proses prediksi dan evaluasi model.

4. HASIL DAN PEMBAHASAN

4.1. Proses Pengembangan Model Decision Tree

Penelitian ini menggunakan teknik *Decision Tree* untuk memprediksi lulusan terbaik di MTS Arrobi'ah Azzain, mempertimbangkan faktor seperti nilai akademik, prestasi akademik, partisipasi ekstrakurikuler, dan kehadiran. *Decision Tree* mengidentifikasi faktor yang memengaruhi prestasi siswa dan memprediksi hasil berdasarkan faktor tersebut. Metode ini menyajikan data dalam bentuk diagram pohon dengan atribut sebagai kriteria pembentukan pohon, termasuk target atribut yang menunjukkan solusi untuk setiap data. Proses pengembangan model melibatkan beberapa langkah penting.

4.2. Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan informasi dari sumber terpercaya, termasuk nilai akademik, prestasi akademik, keaktifan, dan kehadiran. Metode ini melibatkan observasi untuk memahami sistem penilaian dan wawancara dengan pihak sekolah, seperti kepala sekolah dan guru bimbingan konseling. Penulis mengumpulkan 519 data siswa dari angkatan 2022 hingga 2025, yang dibagi menjadi dua kategori: 363 siswa (70%) untuk data *Training* dan 156 siswa (30%) untuk data *Testing*.

4.3. Data Training

Data *Training* adalah kumpulan data yang digunakan untuk membangun model *Decision Tree*, terdiri dari berbagai variabel yang memengaruhi prediksi lulusan terbaik. Setiap variabel dalam Data

Training diwakili oleh nilai dari masing-masing individu. Tabel 1 di bawah ini menyajikan sampel dari 363 Data *Training* beserta beberapa variabel yang digunakan dalam analisis ini dan nilai-nilainya.

Tabel 1. Data Training

NO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	JL
1	B	B	A	B	B	B	B	B	A	B	B	B	A	A	A	T
2	B	B	A	B	B	B	A	A	A	B	B	B	A	A	A	T
3	B	B	A	B	A	B	B	A	A	B	B	B	A	A	A	T
4	B	B	A	B	B	B	B	B	A	B	B	B	A	A	A	T
5	B	B	A	B	B	B	A	B	A	B	B	A	A	A	A	T
6	B	B	A	B	A	B	B	A	A	B	B	B	A	A	A	T
7	B	B	B	B	B	B	B	B	A	B	B	B	A	A	A	T
8	B	B	B	B	B	B	B	B	B	B	B	B	C	A	A	B
9	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
10	B	B	A	B	B	B	B	B	A	B	B	B	C	A	B	B
11	B	B	B	B	B	B	B	B	A	B	B	B	B	A	A	B
12	B	B	A	B	B	B	B	B	A	B	B	B	B	A	B	B
13	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
363	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B

Dari beberapa atribut atau variabel penilaian data *traininig* diatas, berikut pada Tabel 2 dijelaskan secara rinci makna dari setiap variabel :

Tabel 2. Variabel dan Makna Data

Variabel	Makna
P1	Pendidikan Pancasila dan Kewarganegaraan
P2	Bahasa Indonesia
P3	Bahasa Arab
P4	Matematika
P5	Ilmu Pengetahuan Alam
P6	Ilmu Pengetahuan Sosial
P7	Bahasa Inggris
P8	Seni Budaya
P9	Pendidikan Jasmani Olahraga dan Kesehatan
P10	Prakarya
P11	Rata-Rata Nilai Mata Pelajaran Agama
P12	Rata-Rata Nilai Mata Pelajaran Muatan Lokal

Variabel	Makna
P13	Prestasi Akademik
P14	Ekstrakurikuler
P15	Kedisiplinan
T	Terbaik
B	Baik

4.4. Data Testing

Setelah model *Decision Tree* dibangun dengan menggunakan Data *Training*, langkah berikutnya adalah menguji model tersebut dengan Data *Testing*. Data *Testing* berperan dalam mengevaluasi seberapa efektif model dalam memprediksi lulusan terbaik berdasarkan variabel yang sama. Tabel 3 di bawah ini menyajikan sampel dari 156 Data *Testing* beserta beberapa variabel yang digunakan dalam analisis ini dan nilai-nilainya.

Tabel 3. Data Testing

NO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
1	B	B	B	C	B	B	B	B	A	B	B	B	A	A	A
2	B	B	B	C	B	B	B	B	A	B	B	B	A	A	A
3	B	B	B	B	B	B	B	B	A	B	B	B	A	A	A
4	B	B	B	B	B	B	B	B	A	B	B	B	A	A	A
5	A	B	A	A	A	B	A	B	A	B	B	B	A	A	A
6	A	B	A	B	B	A	B	B	A	B	B	B	A	A	A
7	A	B	A	C	B	A	B	B	A	B	B	B	A	A	A
8	B	B	B	C	B	B	B	B	A	B	B	B	A	A	A
9	A	B	A	B	A	A	A	B	A	B	B	B	A	A	A
10	B	B	B	A	B	B	B	B	A	B	B	B	A	A	A
11	A	B	A	B	A	B	B	B	A	B	B	B	A	A	A
12	B	B	B	B	B	B	B	B	A	B	B	B	A	A	A
13	B	B	B	B	B	B	B	B	A	B	B	B	A	A	A
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
156	A	B	A	B	A	B	B	B	A	B	B	B	A	A	A

4.5. Proses Perhitungan Decision Tree

Dalam perhitungan Algoritma *Decision Tree*, penulis menerapkan studi kasus dengan menggunakan 14 data sampel sebagai contoh perhitungan. Data

sampel ini disertakan untuk memudahkan pemahaman mengenai algoritma yang akan dijelaskan. Tabel 4 menyajikan sampel perhitungan yang digunakan sebagai acuan dalam analisis ini.

Tabel 4. Sampel Perhitungan Decision Tree

NO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	JL
1	B	B	B	B	B	B	B	B	B	B	B	B	C	A	A	B
2	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
3	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
4	B	B	B	B	B	B	B	B	A	B	B	B	B	A	A	B
5	B	B	A	B	B	B	B	B	A	B	B	B	B	A	B	B
6	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
7	B	B	A	B	B	B	B	B	A	B	B	B	A	A	A	T
8	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
9	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B
10	B	B	A	B	B	B	A	A	A	B	B	B	A	A	A	T
11	B	B	A	B	B	B	B	A	A	B	B	B	C	A	A	B
12	B	B	A	B	A	B	B	A	A	B	B	B	C	A	A	B
13	B	B	B	B	B	B	B	B	A	B	B	B	C	A	A	B
14	B	B	A	B	B	B	B	B	A	B	B	B	C	A	A	B

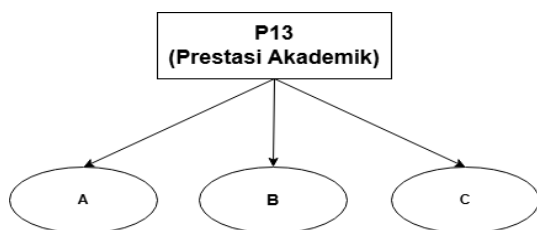
Dalam contoh kasus pada Tabel 5, akan dibangun *Decision Tree* dengan mempertimbangkan variabel atau atribut yang digunakan. Langkah pertama adalah memilih atribut sebagai akar berdasarkan nilai *node* tersebut. Dalam studi kasus ini, penulis akan melihat 2 perhitungan *node* yang mendapatkan 2

nilai *gain* tertinggi, yaitu *Node 1* untuk memberikan gambaran alur kerja *Decision Tree*. Berikut adalah perhitungan Algoritma *Decision Tree* untuk menentukan *Node 1*, sebagaimana tercantum dalam Tabel 8.

Tabel 5. Hasil Perhitungan Nilai Gain Tertinggi

1	2	3	TOTAL	TERBAIK	BAIK	7	8
Node		Keterangan	JK (S)	LT (Si)	LB (Si)	Entropy	Gain
i	Total		14	2	12	0,5917	
	P8 (SB)						
		A	3	1	2	0,9183	0,0496
		B	11	1	10	0,4395	
		C	0	0	0	0	
		D	0	0	0	0	
	P13 (PA)						
		A	2	2	0	0	0,5917
		B	2	0	2	0	
		C	10	0	10	0	
		D	0	0	0	0	

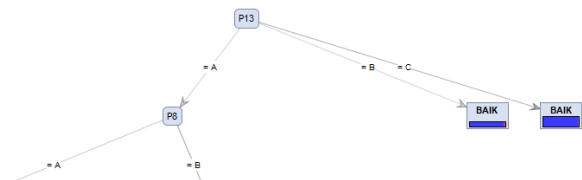
Berdasarkan analisis terhadap 14 sampel data *Training* dalam Tabel 5, atribut dengan nilai *gain* tertinggi adalah P13 (Prestasi Akademik) dengan nilai 0,5917. Oleh karena itu, P13 dipilih sebagai *node* akar dalam pembentukan pohon keputusan. Dari *node* P13, terdapat tiga cabang atribut utama, yaitu A, B, dan C, yang menjadi dasar pengelompokan selanjutnya dalam struktur pohon sesuai pada Gambar 3.



Gambar 3. Hasil Pohon Keputusan Setelah Perhitungan

4.1. Hasil Decision Tree

Berikut ditunjukkan hasil Pohon Keputusan sesuai pada Gambar 4 dan deskripsinya di Gambar 5.



Gambar 4. Hasil Proses Decision Tree

Tree

```

P13 = A
|   P8 = A: TERBAIK {BAIK=0, TERBAIK=26}
|   P8 = B

```

Gambar 5. Deskripsi Hasil Decision Tree

4.2. Hasil Akurasi

Hasil akurasi yang diujikan dengan operator performance adalah 97,80%, sesuai dengan Gambar 6.

	True Data	True Terbak	Class Precision
pred Baik	303	5	98.16%
pred Terbak	3	52	94.12%
Class Recall	99.12%	97.23%	

Gambar 6. Hasil Akurasi

Berikut ini adalah pembahasan lengkap mengenai perhitungan metrik-metrik tersebut berdasarkan *confusion matrix* hasil prediksi model *Decision Tree* di *RapidMiner*.

4.3. Akurasi

Akurasi adalah metrik paling umum yang menunjukkan proporsi prediksi yang benar dibandingkan dengan total jumlah data. Berikut adalah perhitungannya :

$$\text{Akurasi} = \frac{TP + TP2}{TP1 + FP1 + FN1 + TP2}$$

$$\text{Akurasi} = \frac{303 + 52}{303 + 5 + 3 + 52}$$

$$\text{Akurasi} = \frac{355}{363}$$

$$\text{Akurasi} = 97,80 \%$$

4.4. Precision

Precision menunjukkan seberapa banyak prediksi positif yang benar-benar positif. *Precision* sangat berguna ketika kesalahan positif (*false positive*) perlu diminimalkan. Berikut perhitungan untuk kelas baik dan terbaik :

Precision Baik :

$$\text{Precision Baik} = \frac{TP1}{TP1 + FP1}$$

$$\text{Precision Baik} = \frac{303}{303 + 5}$$

$$\text{Precision Baik} = \frac{303}{308}$$

$$\text{Precision Baik} = 98,38\%$$

Precision Terbaik :

$$\text{Precision Terbaik} = \frac{TP2}{TP2 + FN1}$$

$$\text{Precision Terbaik} = \frac{52}{52 + 3}$$

$$\text{Precision Terbaik} = \frac{52}{55}$$

$$\text{Precision Terbaik} = 94,55\%$$

4.5. Recall

Recal menunjukkan seberapa banyak data positif yang berhasil dikenali dengan benar oleh model. Metrik ini penting ketika kesalahan negatif (*false negative*) harus diminimalkan. Berikut perhitungan untuk kelas baik dan terbaik :

Recall Baik :

$$\text{Recall Baik} = \frac{TP1}{TP1 + FN1}$$

$$\text{Recall Baik} = \frac{303}{303 + 3}$$

$$\text{Recall Baik} = \frac{303}{306}$$

$$\text{Recall Baik} = 99,02\%$$

Recall Terbaik :

$$\text{Recall Terbaik} = \frac{TP2}{TP2 + FP2}$$

$$\text{Recall Terbaik} = \frac{52}{52 + 5}$$

$$\text{Recall Terbaik} = \frac{52}{57}$$

$$\text{Recall Terbaik} = 91,23\%$$

4.6. F1-Score

F1-Score adalah harmonic mean dari *precision* dan *recall*. *F1-score* memberikan ukuran keseimbangan antara keduanya, dan sangat berguna ketika kita ingin mempertimbangkan kesalahan positif dan negatif secara bersamaan.

F1-Score Baik :

$$F1 \text{ Baik } 2x = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$F1 \text{ Baik } 2x = \frac{0,9838 \times 0,9902}{0,9838 + 0,9902}$$

$$F1 \text{ Baik } 2x = \frac{0,9740}{1,9740}$$

$$F1 \text{ Baik} = 98,70\%$$

F1-Score Terbaik :

$$\text{Recall Terbaik } 2x = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$F1 \text{ Terbaik } 2x = \frac{0,9455 \times 0,9123}{0,9455 + 0,9123}$$

$$F1 \text{ Terbaik } 2x = \frac{0,8626}{1,8578}$$

$$F1 \text{ Terbaik} = 92,86\%$$

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian mengenai prediksi prestasi akademik lulusan terbaik di MTs Arrobi'ah Azzain menggunakan algoritma *Decision Tree*, dapat disimpulkan bahwa penerapan algoritma ini efektif dalam meningkatkan efisiensi dan akurasi penentuan lulusan terbaik. Dengan perangkat lunak *Rapidminer*, analisis data yang sebelumnya manual kini lebih cepat dan sederhana. Algoritma ini mengolah variabel seperti nilai akademik, prestasi akademik, keaktifan, dan tingkat kehadiran secara otomatis, menghasilkan keputusan yang lebih cepat dan akurat. Tingkat akurasi prediksi mencapai 97,80%, dengan kriteria penting yang mencakup Nilai Akademik, Prestasi Akademik, Keaktifan, dan Tingkat Kehadiran untuk penilaian yang objektif dan transparan. Penelitian ini juga menunjukkan bahwa prestasi akademik adalah

variabel paling berpengaruh dalam menentukan prestasi siswa, menegaskan pentingnya nilai akademik dalam evaluasi di MTs Arrobi'ah Azzain.

Berdasarkan temuan penelitian, disarankan agar sistem yang dikembangkan diimplementasikan secara berkelanjutan dan diperbarui untuk mengikuti perkembangan pendidikan dan kebutuhan siswa. Penelitian ini juga dapat menjadi acuan bagi studi selanjutnya yang mengeksplorasi penerapan teknologi dalam memprediksi prestasi akademik siswa. Hasil ini diharapkan mendorong inovasi yang lebih luas dalam meningkatkan mutu pendidikan.

DAFTAR PUSTAKA

- [1] S. K. B. Dinas Komunikasi, Informatika, "Brebes Dalam Data Tahun 2023," 2024.
- [2] M. Roza, Lufri, and Asrizal, "Model Matematika Pengambilan Keputusan Mahasiswa Dalam Memilih Jurusan:Dampak Minat dan Bakat," *J. Pendidik. Mat. Dan Ipa*, vol. 11, no. 1, pp. 13–21, 2020.
- [3] H. A. Setyadi, G. S. Nurohim, and W. Nugroho, "Penentuan Kelayakan dan Jenis Lembaga Keuangan dalam Pemberian Modal UMKM Menggunakan Metode AHP dan Decision Tree," *J. Ilm. FIFO*, vol. 16, no. 1, p. 64, 2024, doi: 10.22441/fifo.2024.v16i1.007.
- [4] P. Ramadhan and J. Veri, "Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2210.
- [5] M. T. Gunawan *et al.*, "Model decision tree untuk prediksi prestasi akademik matematika siswa kelas VIII SMP Frater Don Bosco Manado Monica," *J. Pendidik. Inform. dan Sains*, vol. 13, no. 2, pp. 141–153, 2024, doi: 10.31571/saintek.v13i2.7696.
- [6] R. Ramadhani, Y. Hendriyani, and I. Pendahuluan, "Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus : SMKN 2 Padang) (Studi Kasus : SMKN 2 Padang) (Font menggunakan Times New Roman berukuran 12 point spasi tunggal) ABSTRAK Pengujian ini bertujuan untuk me," *J. Vocat. Tek. Elektron. dan Inform.*, vol. 9, no. 3, 2021.
- [7] L. Yuningsih, I. R. Setiawan, and A. A. Sunarto, "Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5," *Progresif J. Ilm. Komput.*, vol. 16, no. 2, p. 121, 2020, doi: 10.35889/progresif.v16i2.517.
- [8] H. Rifa'i, Ryan Hamonangan, Dian Ade Kurnia, Kaslani, and Mulyawan, "Implementasi Algoritma Decision Tree Dalam Klasifikasi Kompetensi Siswa," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 6, no. 1, pp. 15–20, 2022, doi: 10.32485/kopertip.v6i1.131.
- [9] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Portal Data*, vol. 2, no. 3, pp. 1–12, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>.
- [10] N. Nazifah, C. Prianto, and C. A. Id, "Decision Tree Algoritma C4.5 dengan algoritma lainnya: Sitematic Literature Review," *J. Inform. dan Teknol. Komput.*, vol. 04, no. 57–64, 2023, [Online]. Available: <https://ejurnalunsam.id/index.php/jicom/>.
- [11] A. C. Darmawan, "Pengembangan Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5," *J. Pendiidikan Konseling*, vol. 4, no. 5, pp. 5351–5362, 2022.
- [12] N. Tulus Ujianto and N. A. Ramdhan, "Implementasi Data Mining C4.5 Dalam Mengukur Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 01, pp. 105–111, 2022.
- [13] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.