

IMPLEMENTASI DATA MINING DALAM MEMPREDIKSI PRESTASI SISWA DENGAN ALGORITMA C4.5

Alvianti Nurjana^{1*}, Agus Perdana Windarto², Hendry Qurniawan³

^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹alviantinurjannah556@gmail.com, ²agus.perdana@amiktunasbangsa.ac.id,

³hendry.q@amiktunasbangsa.ac.id

Abstract

This study applies the data mining method with the C4.5 algorithm into an application that can be used to predict student achievement. This study aims to build a prediction system for outstanding students that can facilitate decision making. The data used in this study were taken from MTs. Al-Mukhlisin Dolok Kahean by conducting interviews and direct observation. The analysis was carried out by conducting training on data on grade 8 students for the 2021-2022 academic year, which totaled 75 students. There are 4 variables used, namely Discipline, Presence, Morals and Test Scores. The accuracy and performance obtained was 97.50%, namely there were 66 students who received the title of achievement and 9 students who were not classified as outstanding. So with this analysis it is hoped that it can help MTs schools. Al-Mukhlisin in knowing student achievement in teaching and learning process.

Keywords: Prediction, Achievement, Algorithm C4.5, MTs Schools, Students

Abstrak

Penelitian ini menerapkan metode data mining dengan algoritma C4.5 kedalam sebuah aplikasi yang dapat digunakan untuk memprediksi prestasi siswa. Penelitian ini bertujuan untuk membangun sistem prediksi siswa berprestasi yang dapat memudahkan dalam pengambilan keputusan. Data yang digunakan dalam penelitian ini diambil dari MTs. Al-Mukhlisin Dolok Kahean dengan melakukan wawancara dan observasi langsung. Analisa dilakukan dengan melakukan training pada data siswa kelas 8 tahun pelajaran 2021-2022 yang berjumlah 75 siswa. Variabel yang digunakan sebanyak 4, yaitu Disiplin, Presensi, Akhlak dan Nilai Ujian. Perolehan akurasi dan performa yang didapatkan sebanyak 97.50 % yaitu terdapat 66 siswa yang memperoleh predikat prestasi dan 9 orang siswa tidak tergolong berprestasi. Sehingga dengan analisis ini diharapkan dapat membantu pihak sekolah MTs. Al-Mukhlisin dalam mengetahui prestasi siswa dalam proses belajar mengajar.

Kata Kunci: Prediksi, Prestasi, Algoritma C4.5, Sekolah MTs, Siswa

1. PENDAHULUAN

Pendidikan merupakan upaya yang paling dasar dalam mewujudkan proses pembelajaran peserta didik secara aktif untuk mengembangkan potensi-potensi yang ada didalam dirinya agar dapat meraih prestasi [1]. Prestasi belajar yaitu kemampuan siswa untuk belajar, yakni dengan mengingat fakta dan mengkomunikasikan pengetahuannya baik secara lisan maupun tertulis. Sehingga prestasi belajar dapat dikatakan sebagai hasil yang diperoleh dari kegiatan pembelajaran di sekolah yang bersifat kognitif dan biasanya ditentukan melalui pengukuran dan penilaian. Prestasi belajar dapat digambarkan dengan nilai yang diberikan guru dari banyaknya bidang studi yang dipelajari siswa. Setiap kegiatan pembelajaran tentunya selalu mengharapkan akan menghasilkan pembelajaran yang maksimal. Dalam proses pencapaiannya, prestasi belajar sangat dipengaruhi oleh banyak factor yang berbeda. Salah satu faktor utama yang sangat berpengaruh dalam keberhasilan pembelajaran adalah keberadaan guru. Mengingat keberadaan guru dalam proses kegiatan belajar mengajar sangat berpengaruh, maka sudah semestinya kualitas guru harus diperhatikan [2]. MTs. Al-Mukhlisin merupakan sebuah lembaga pendidikan swasta di Desa Dolok Kahean Kec. Tapan Dolok Kab. Simalungun yang saat ini memiliki 192 siswa, data yang digunakan adalah data siswa kelas 8 Tahun Pelajaran 2021-2022 yang berjumlah 75 siswa dimana setiap siswa memiliki tingkat prestasi yang berbeda-beda, sehingga pihak sekolah perlu mengetahui lebih dini

mengenai tingkat prestasi dan keterampilan siswa untuk memberikan pembelajaran dan bimbingan sesuai dengan kemampuan siswa tersebut agar dapat mencegah terhadap siswa yang berpotensi mengalami hambatan dalam belajar atau kurang berprestasi [2].

Dengan adanya permasalahan tersebut, penulis akan mencoba menggunakan suatu metode yaitu dengan pendekatan data mining. Sementara itu, data mining adalah proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [3]. Salah satu teknik Data Mining Decision Tree yang terkenal dan dapat digunakan sebagai prediksi adalah Algoritma C4.5 [4]. Dimana Algoritma C4.5 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik (kontinyu) dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma lain [5]. Penulis memilih menggunakan Algoritma C4.5 karena dapat melakukan prediksi dengan memberikan tingkat nilai akurasi yang ideal untuk memprediksikan kemampuan siswa. Beberapa penelitian terdahulu, decision tree sering dipergunakan untuk prediksi prestasi siswa dengan performa yang sangat baik. Algoritma decision tree untuk memprediksi nilai akhir siswa dengan akurasi 85%. Optimasi decision tree dapat diterapkan untuk prediksi siswa berpotensi bermasalah dengan tingkat akurasi 99,08% . Untuk memprediksi prestasi siswa dengan decision tree, diperoleh hasil akurasi yang diperoleh sangat baik yaitu 97,22%. Sehingga dalam penelitian ini, akan menggunakan metode klasifikasi dengan Algoritma Decision Tree. [6], [7]. Beberapa kriteria yang digunakan dalam prediksi prestasi siswa meliputi: kedisiplinan siswa, presensi siswa, status ekonomi siswa dan prestasi belajar siswa. Pada penelitian sebelumnya, penelitian dengan menggunakan Algoritma C4.5 sudah pernah dilakukan oleh [8] menunjukkan bahwa Algoritma C4.5 dapat digunakan untuk memprediksi prestasi siswa berdasarkan dari pohon keputusan yang terbentuk. Penelitian yang dilakukan oleh [9] menunjukkan bahwa metode Algoritma C4.5 dapat memprediksi hasil belajar siswa secara daring pada masa pandemi covid-19. Penelitian yang dilakukan oleh [5], [10] dalam mengklasifikasikan penerimaan beasiswa, menunjukkan bahwa Algoritma C4.5 memiliki performa terbaik dengan akurasi 96.40% dibandingkan algoritma naive bayes. Selanjutnya penelitian yang dilakukan oleh [11] dalam melakukan komparasi algoritma atau metode klasifikasi algoritma naive bayes dan C4.5 penelitian ini menggunakan algoritma decision tree dan menghasilkan tingkat akurasi terbaik dalam melakukan prediksi kelulusan tepat waktu mahasiswa dengan tingkat akurasi sebesar 76.69% dibandingkan algoritma naive bayes [12].

Berdasarkan latar belakang tersebut, diharapkan hasil dari penelitian dengan judul memprediksi prestasi siswa dengan penerapan Algoritma C4.5 ini dapat membantu sekolah MTs. Al-Mukhlisin Dolok Kahean dalam menentukan siswa yang berprestasi, mampu mengetahui hasil prediksi serta mengetahui keakurasiannya. Selain itu hasil penelitian ini dapat menjadi acuan bagi penelitian lain dengan topik yang sama.

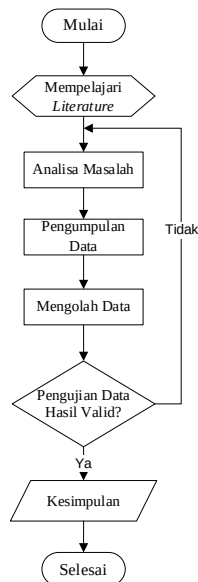
2. METODE PENELITIAN

2.1. Metode Penelitian

Metodologi merupakan sebuah proses yang dilakukan untuk memecahkan sebuah masalah yang diangkat dalam sebuah penelitian sehingga akan ditemukan hasil yang akurat dan hingga dapat diambil kesimpulan. Pada metodologi penelitian berikut ini akan dijelaskan secara rinci semua urutan dalam pelaksanaan penelitian dan menganalisisnya berdasarkan Algoritma C4.5 [13].

2.1.1. Rancangan Penelitian

Rancangan penelitian berguna untuk menjelaskan bagaimana menguraikan dan menyelesaikan masalah yang ada dalam penelitian. Rancangan penelitian yang disajikan dapat dilihat dalam flowchart pada Gambar 1.



Gambar 1. Rancangan Penelitian

Gambar 1. Flowchart Rancangan Penelitian

Berdasarkan gambar 1 dapat diuraikan sebagai berikut:

- Mulai**
Pada tahapan ini menunjukkan bahwa proses penelitian akan dimulai.
- Mempelajari literature**
Pada tahapan ini penelitian harus didasari pengamatan dari sumber seperti jurnal, skripsi dan buku yang berkaitan dengan penelitian ini.
- Analisa masalah**
Masalah yang terkait dengan memprediksi prestasi siswa di MTs. Al-Mukhlisin Dolok Kahean.
- Pengumpulan data**
Pada tahap ini dilakukan pengumpulan data di MTs. Al-Mukhlisin Dolok Kahean yang akan digunakan sesuai dengan masalah penelitian.
- Mengolah data**
Setelah menganalisa masalah dan mengumpulkan data maka data akan diolah dan dikelompokkan agar mempermudah dalam menganalisis data yang selanjutnya. Data yang digunakan adalah data siswa pada tahun pelajaran 2021-2022. Data yang diperoleh akan diuji dengan menggunakan Algoritma C4.5.
- Pengujian data**
Tahap ini dilakukan pengujian dengan menggunakan Software Microsoft Excel dan Software RapidMiner.
- Kesimpulan**
Pada tahapan akhir penelitian yaitu kesimpulan merupakan penentu hasil yang diperoleh dari pengujian yang telah dilakukan menggunakan Algoritma C4.5 dalam memprediksi prestasi siswa di MTs. Al-Mukhlisin Dolok Kahean.
- Selesai**
Setelah melakukan beberapa tahapan dan telah diperoleh hasil yang valid maka proses tahapan penelitian selesai.

2.1.2. Prosedur Pengumpulan Data

Dalam melakukan pengumpulan data yang diperlukan untuk penelitian terdapat beberapa metode yang digunakan yaitu [14] :

- Penelitian Lapangan (Field Work Research) yaitu penelitian yang dilakukan secara langsung ke lapangan dengan menggunakan teknik studi literature yaitu informasi yang penulis peroleh dengan mengumpulkan data, mempelajari data, validasi data dan mencari referensi terkait dengan kasus penelitian. Hasil dari studi literature ini adalah tersusun dan terkoleksinya referensi yang baik dan sesuai dengan topik penelitian.
- Penelitian Kepustakaan (Library Research) yaitu memanfaatkan perpustakaan, buku, prosiding atau jurnal sebagai sarana referensi untuk mengidentifikasi faktor, parameter dan label yang digunakan dalam penelitian.
- Data kuantitatif yang diperoleh dari sumber asal, yaitu MTs. Al-Mukhlisin Dolok Kahean untuk keperluan penelitian, sedangkan data kualitatif berisi tentang data yang dihasilkan dari pengolahan data penelitian.

2.2. Analisis Data

Analisis data merupakan sebuah proses tentang cara perancangan yang digunakan dalam membangun sistem atau melakukan proses perhitungan dari metode yang diimplementasikan. Pada penelitian ini data di ambil dari siswa kelas 8 MTs. Al-Mukhlisin Dolok Kahean tahun 2021-2022 yang berjumlah 75 siswa. Data kemudian diolah dengan melakukan perhitungan menggunakan Algoritma C4.5, setelah itu dilakukan pengujian data menggunakan aplikasi RapidMiner 5.3 [15].

Tabel 1. Data

No	Nama	NISN	Disiplin	Presensi	Akhlak	Nilai Ujian	Rata-Rata
1	Adetya Iqbal	0087038820	88	98	84	88	89,5
2	Afdiansyah Dwi	0083558471	89	98	85	90	90,5
3	Ahmad Pareji	0082422834	83	97	80	86	86,5
4	Almira Aulia	0086424855	72	70	90	92	81
5	Ananda Fasya	0083361674	88	96	83	88	88,75
6	Anggun Pratiwi	0079853600	90	99	85	88	90,5
7	Aqmal Nasrullah	0085425552	93	100	90	94	94,25
8	Arek Kurniawan	0089954380	83	95	80	87	86,25
9	Ariska Safutri	0087916006	89	99	90	90	92
10	Aufah Atha	0082707426	95	98	90	92	93,75
11	Auliya Amanda	0058610445	72	70	80	80	75,5
12	Ayulia Fadillah	0082480841	89	96	88	87	90
13	Ayu Wana Fadila	0069360352	72	70	80	78	75
14	Azay Zuhri	0073059582	84	97	83	85	87,25
15	Azmi Febriansyah	0088658144	87	100	88	88	90,75
...	...	...	...	...	...	...	...
75	Winnanda	0086017915	88	95	87	85	88,75

Tabel 2. Data Transformasi

No	Disiplin	Presensi	Akhlak	Nilai Ujian	Rata-Rata
1	Baik	Sangat Baik	Baik	Baik	Iya
2	Sangat Baik	Sangat Baik	Baik	Sangat Baik	Iya
3	Baik	Sangat Baik	Cukup	Baik	Iya
4	Cukup	Cukup	Sangat Baik	Sangat Baik	Iya
5	Baik	Sangat Baik	Baik	Baik	Iya
6	Sangat Baik	Sangat Baik	Baik	Baik	Iya
7	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Iya
8	Baik	Sangat Baik	Cukup	Baik	Iya
9	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Iya
10	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Iya
11	Cukup	Cukup	Cukup	Cukup	Tidak

No	Disiplin	Presensi	Akhlak	Nilai Ujian	Rata-Rata
12	Sangat Baik	Sangat Baik	Baik	Baik	Iya
13	Cukup	Cukup	Cukup	Cukup	Tidak
14	Baik	Sangat Baik	Baik	Baik	Iya
15	Baik	Sangat Baik	Baik	Baik	Iya
...	...	...	...	...	...
75	Baik	Sangat Baik	Baik	Baik	Iya

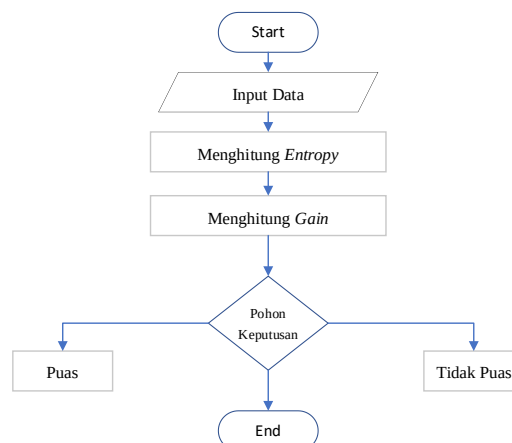
Range

>=90 & <=100	: Sangat Baik
>=80 & < 89	: Baik
>=70 & < 80	: Cukup
>=51 & < 69	: Kurang
<=0 & < 50	: Sangat Kurang

2.2.1. Alat Analisis Data

Dalam penelitian ini penulis menggunakan Software RapidMiner 5.3. Data diolah menggunakan RapidMiner 5.3 yang berguna sebagai validasi dan reabilitas data untuk mencari keakuratan data. RapidMiner adalah pemrograman perangkat lunak ilmu data yang dikembangkan oleh perusahaan bersama-sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Data yang akurat maka akan dilanjutkan kepada pengolahan data untuk mencari hasil dari masalah penelitian dengan menggunakan RapidMiner 5.3 dan mengambil keputusan.

2.2.2. Pemodelan Metode



Gambar 2. Pemodelan Metode

Gambar 2 merupakan kerangka penelitian yang dituangkan dalam diagram alir ini, menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan.

Berikut penjelasan langkah-langkah pemodelan dari gambar 3.4 sebagai berikut :

a) Menghitung nilai gain dengan menggunakan rumus sebagai berikut :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \left(\frac{s_i}{S} \right) * \text{Entropy}(s_i) \quad (2)$$

Keterangan :

S = Himpunan Kasus

A = Atribut

N = Jumlah Partisi Antribut A

(Si) = Jumlah Kasus pada partisi Ke-i

(S) = Jumlah Kasus dalam S

- b) Kemudian ulangi Langkah Kedua sampai semua record terpartisi secara sempurna.
- c) Proses partisi pohon keputusan akan berhenti jika :
 - 1) semua record pada simpul N mendapatkan kelas yang sama.
 - 2) Tidak ada atribut di dalam record yang dipartisi lagi .
 - 3) Tidak ada record didalam cabang yang kosong .

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Pada bab ini menjelaskan tentang hasil penelitian sesuai dengan kriteria dan alternatif yang digunakan pada saat proses pengumpulan data. Data mentah hasil observasi yang telah berhasil dikumpulkan untuk diolah atau ditransformasikan ke format data Microsoft Excel 2010. Data yang ditransformasikan tersebut digunakan sebagai syarat dalam pengolahan Data Mining Algoritma C4.5. Selanjutnya, peneliti menganalisis data menggunakan tools RapidMiner versi 5.3 sebagai implementasi dari pengolahan data. Hasil dari penelitian digunakan untuk menentukan keputusan untuk prediksi prestasi siswa di MTs. Al-Mukhlisin Dolok Kahean.

3.1.1. Pengelolaan Data

Pada bagian pengolahan data ini akan diuraikan tentang proses perhitungan metode dan penerapan metode terhadap permasalahan yang diangkat dalam penulisan penelitian ini.

3.1.2. Proses Perhitungan C4.5

Langkah awal yang dilakukan penulis yaitu menentukan hasil model aturan pertama dengan menggunakan Data Mining Algoritma C4.5. Penggunaan Algoritma C4.5 dilakukan untuk memberikan model aturan dari data yang digunakan. Adapun langkah-langkah dalam membentuk pohon keputusan dengan menggunakan Algoritma C4.5 adalah sebagai berikut :

- a) Pertama, melakukan perhitungan Entropy dari total data yang di proses. Adapun jumlah kasus pada data penelitian ini yaitu 75 record, diketahui label prestasi yang iya ada 66 record (data) dan status label prestasi yang tidak ada 9 record sehingga Entropy nilai akar total yang didapat, yaitu :

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-66/75 * \log_2 (66/75)) + (-9/75 * \log_2 (9/75)) \\ &= 0.5294 \end{aligned}$$

- b) Langkah kedua yaitu menghitung nilai Entropy pada atribut Disiplin dengan kategori Sangat Baik, jumlah kasus terdapat 31 record, kemudian terdapat status prestasi IYA sebanyak 31 record dan status prestasi TIDAK sebanyak 0 record.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-31/31 * \log_2 (31/31)) + 0 \\ &= 0.000 \end{aligned}$$

- c) Selanjutnya melakukan perhitungan Entropy pada atribut Disiplin sampai dengan kategori Sangat Kurang. Kemudian menghitung nilai Gain pada Atribut Disiplin dengan kategori Sangat Baik terdapat 31 record, kategori Baik terdapat 30 record, kategori Cukup terdapat 14 record, kategori Kurang terdapat 0 record dan kategori Sangat Kurang terdapat 0 record , adapun perhitungan untuk mencari nilai Gain sebagai berikut :

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0.5294 - (((31/75)*0)+((30/75)*0)+ \\ &\quad ((14/75) * 0.9403)) \\ &= 0.3538 \end{aligned}$$

- d) Kemudian menghitung SplitInfo pada atribut Disiplin berdasarkan jumlah kasus pada atribut Disiplin.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -(31/75 * \log_2 (31/75)) + (30/75 * \log_2 (30/75)) + (14/75 * \log_2 \\ &\quad (14/75)) \\ &= 1.5076 \end{aligned}$$

- e) Selanjutnya menghitung Nilai GainRatio atribut Disiplin yang merupakan nilai yang digunakan sebagai pemecahan Node pada pohon keputusan.

$$\begin{aligned}\text{RasioGain}(s,j) &= (\text{Gain}(s,j))/(\text{SplitInfo}(s,j)) \\ &= 0.3538 / 1.5076 \\ &= 0.2347\end{aligned}$$

- f) Kemudian menghitung nilai Entropy pada atribut Presensi dengan kategori Sangat Baik dengan jumlah kasus terdapat 59 record, terdapat status prestasi (Iya) sebanyak 59 record dan status prestasi (Tidak) sebanyak 0 record.

$$\begin{aligned}\text{Entropy}(S) &= \sum_{i=1}^n -p_i * \log_2 p_i \\ &= (-59/59 * \log_2 (59/59)) + 0 \\ &= 0.000\end{aligned}$$

- g) Langkah berikutnya melakukan perhitungan Entropy pada atribut Presensi sampai dengan kategori Sangat Kurang. Kemudian menghitung nilai Gain pada Atribut Presensi dengan kategori Sangat Baik terdapat 59 record, kategori Baik terdapat 2 record, kategori Cukup terdapat 14 record, kategori Kurang terdapat 0 record dan kategori Sangat Kurang terdapat 0 record :

$$\begin{aligned}\text{Gain}(S,A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \\ &= 0.5294 - (((59/75)*0) + ((2/75)*0) + ((14/75)*0.9403)) \\ &= 0.3538\end{aligned}$$

- h) Langkah berikutnya yaitu menghitung SplitInfo pada atribut Presensi berdasarkan jumlah kasus pada Presensi.

$$\begin{aligned}\text{SplitInfo} &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((59/75.\log_2(59/75)) + (2/75.\log_2(2/75)) + \\ &\quad (14/75.\log_2(14/75))) \\ &= 0.8638\end{aligned}$$

- i) Berikutnya menghitung Nilai GainRatio Presensi yang merupakan nilai yang digunakan sebagai pemecahan Node pada pohon keputusan.

$$\begin{aligned}\text{RasioGain}(s,j) &= (\text{Gain}(s,j))/(\text{SplitInfo}(s,j)) \\ &= 0.3538/ 0.8638 \\ &= 0.4096\end{aligned}$$

..... Dan seterusnya hingga data terakhir.

Berdasarkan hasil dari perhitungan nilai Entropy, nilai Gain, SplitInfo dan GainRatio dalam tiap atribut dapat dilihat pada Tabel 3., dengan keterangan yaitu SB = Sangat Baik, B = Baik, C = Cukup, K = Kurang dan SK = Sangat Kurang yakni sebagai berikut :

Tabel 3. PerhitunganGain Node 1

No.	Atribut	Jumlah Kasus	Iya	Tidak	Entropy	Gain	Split Info	Gain Ratio
1	Total	75	66	9	0.5294	0.3538	1.5076	0.2347
	DISIPLIN							
	Sangat Baik	31	31	0	0.0000			
	Baik	30	30	0	0.0000			
	Cukup	14	5	9	0.9403			
	Kurang	0	0	0	0.0000			
2	Sangat Kurang	0	0	0	0.0000	0.3538	0.8638	0.4096
	PRESENSI							
	Sangat Baik	59	59	0	0.0000			
	Baik	2	2	0	0.0000			
	Cukup	14	5	9	0.9403			
	Kurang	0	0	0	0.0000			
3	Sangat Kurang	0	0	0	0.0000	0.4471	1.5000	0.2981
	AKHLAK							
	Sangat Baik	34	34	0	0.0000			
	Baik	27	26	1	0.2285			
	Cukup	14	6	0	0.0000			
	Kurang	0	0	0	0.0000			
4	Sangat Kurang	0	0	0	0.0000	0.5294	1.4000	0.3781
	NILAI UJIAN							

No.	Atribut	Jumlah Kasus	Iya	Tidak	Entropy	Gain	Split Info	Gain Ratio
	Sangat Baik	29	29	0	0.0000			
	Baik	37	37	0	0.0000			
	Cukup	9	0	9	0.0000			
	Kurang	0	0	0	0.0000			
	Sangat Kurang	0	0	0				

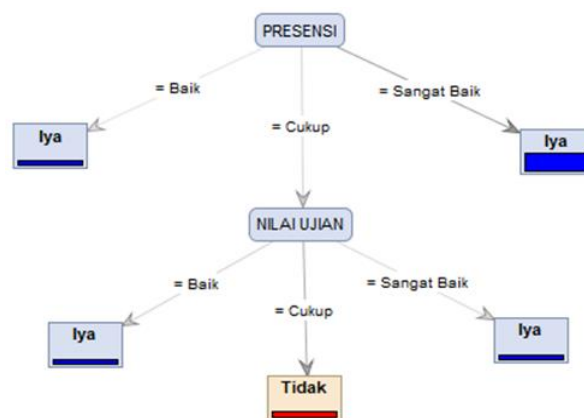
Berdasarkan pada Tabel 3. diketahui nilai GainRatio terbesar adalah atribut Presensi dengan nilai 0.4096, kategori Cukup dengan nilai 0.9403 yang digunakan sebagai simpul akar untuk mengulangi perhitungan nilai Entropy dan nilai Gain selanjutnya hingga node terakhir.

3.1.3. Proses Pengujian dengan RapidMiner

Pada tahapan ini penulis akan menguraikan beberapa proses dan hasil penelitian yang telah dilakukan yang berisikan tampilan eksekusi atau implementasi dari sistem, validasi data dan solusi pemecahan masalah yang telah dibuat.

Pada bagian ini menjelaskan tentang tampilan sistem dengan menggunakan Algoritma C4.5 dengan menggunakan software RapidMiner 5.3.

Hasil Pohon Keputusan setelah melakukan running process atau menjalankan desain model dari Algoritma C4.5, maka dapat dihasilkan gambar berupa pohon keputusan yang tampil pada software RapidMiner. Adapun gambar pohon keputusan dapat dilihat pada Gambar 3., sebagai berikut :



Gambar 3. Tampilan Pohon Keputusan

Berdasarkan pada tampilan Gambar 3., dapat dilihat bahwa Node tertinggi dari pohon keputusan adalah atribut Presensi dengan kategori Sangat baik, Baik dan Cukup. Selanjutnya Node Nilai Ujian dengan kategori Sangat Baik , Baik dan Cukup.

Hasil Model Aturan setelah mendapatkan hasil pohon keputusan pada software RapidMiner, maka dapat dilihat pula model aturan berdasarkan pohon keputusan yang tampil pada RapidMiner. Adapun hasil model aturan dari RapidMiner dapat dilihat pada Gambar 4 berikut :

Tree

```

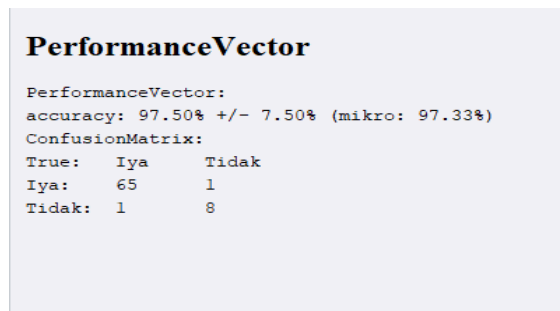
PRESENSI = Baik: Iya {Iya=2, Tidak=0}
PRESENSI = Cukup
|  NILAI UJIAN = Baik: Iya {Iya=3, Tidak=0}
|  NILAI UJIAN = Cukup: Tidak {Iya=0, Tidak=9}
|  NILAI UJIAN = Sangat Baik: Iya {Iya=2, Tidak=0}
PRESENSI = Sangat Baik: Iya {Iya=59, Tidak=0}
  
```

Gambar 4. Tampilan Model Aturan

Berdasarkan pada Gambar 4., dapat diketahui bahwa model aturan dengan prestasi Iya tertinggi adalah Jika Presensi = Sangat Baik, maka hasilnya Iya {Iya=59 dan Tidak=0}. Sedangkan model aturan dengan prestasi Tidak tertinggi adalah Jika Presensi = Cukup dan Nilai Ujian = Cukup, Maka hasilnya Tidak {Iya=0 dan Tidak=9}.

3.2. Pembahasan

Pada bagian ini menjelaskan tentang proses pencarian akurasi ataupun performansi data yang telah diuji. Hasil Validasi Data Setelah design proses validasi data dirancang maka selanjutnya running atau jalankan program. Adapun hasil validasi data yang tampil pada software RapidMiner dapat dilihat pada Gambar 5.



PerformanceVector

PerformanceVector:
accuracy: 97.50% +/- 7.50% (mikro: 97.33%)
ConfusionMatrix:
True: Iya Tidak
Iya: 65 1
Tidak: 1 8

Gambar 5. Tampilan Hasil Validasi Data

Berdasarkan Gambar 5., dapat diketahui bahwa pada performance vector diperoleh akurasi pengujian data sebanyak 97.50% dengan 66 siswa yang memperoleh predikat prestasi dan tidak tergolong berprestasi adalah 9 orang siswa. Siswa yang tergolong prestasi berarti siswa tersebut harus memiliki nilai Disiplin, Presensi, Akhlak dan Nilai Ujian minimal 80 dengan predikat Baik disetiap atribut nilainya. Sedangkan siswa yang tidak tergolong prestasi berarti siswa tersebut memiliki nilai rendah atau nilai dibawah 80 dari setiap atributnya atau terdapat predikat nilai Cukup maupun predikat nilai Kurang pada salah satu nilai atribut Disiplin, Presensi, Akhlak maupun Nilai Ujian. Adapun hasil prestasi (Iya) tertinggi adalah Jika Presensi = Sangat Baik, Maka hasilnya Iya {Iya=59 dan Tidak=0}. Sedangkan model aturan dengan prestasi (Tidak) tertinggi adalah Jika Presensi = Cukup dan Nilai = Cukup, Maka hasilnya Tidak {Iya=0 dan Tidak=9}. Artinya hasil yang dilakukan penulis dalam perhitungan manual Algoritma C4.5 menggunakan Microsoft Excel 2010 dengan pengujian RapidMiner 5.3 diperoleh hasil yang sama atau valid. Sehingga dari hasil tersebut dapat disimpulkan bahwa Algoritma C4.5 dapat dijadikan sebagai rekomendasi untuk pengambil keputusan.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan sebelumnya, maka dapat disimpulkan bahwa Data Mining Algoritma C4.5 dapat diterapkan sebagai pengambil keputusan dalam memprediksi prestasi siswa pada MTs. Al-Mukhlisin Dolok Kahean dengan menggunakan data nilai 75 siswa dengan variabel disiplin, presensi, akhlak, dan nilai ujian yang merupakan hasil observasi langsung pada tahun 2022. Berdasarkan hasil perhitungan manual Algoritma C4.5 menggunakan Microsoft Excel 2010 dengan pengujian RapidMiner 5.3 diperoleh hasil yang sama atau valid dengan perolehan hasil akurasi dan performansi data sebanyak 97.50% yaitu terdapat 66 siswa yang memperoleh predikat prestasi dan 9 orang siswa tidak tergolong berprestasi. Berdasarkan hasil penelitian yang diperoleh tersebut dapat mempermudah pihak MTs. Al-Mukhlisin Dolok Kahean dalam melakukan proses penilaian siswa berprestasi dan tidak berprestasi sehingga hasil pemberian nilai dapat diperoleh dengan lebih akurat serta efektif dan efisien.

REFERENCES

- [1] K. F. Irnanda, D. Hartama, And A. P. Windarto, "Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi," *J. Media Inform. Budidarma*, Vol. 5, No. 1, P. 327, 2021, Doi: 10.30865/Mib.V5i1.2763.

- [2] A. Prasetio, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *J. Nas. Komputasi Dan Teknol. Inf.*, Vol. 4, No. 3, Pp. 209–214, 2021, Doi: 10.32672/Jnkti.V4i3.2983.
- [3] V. Anestiviya, A. Ferico, And O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar)," *J. Teknol. Dan Sist. Inf.*, Vol. 2, No. 1, Pp. 80–85, 2021, [Online]. Available: [Http://jim.Teknokrat.Ac.Id/Index.Php/Jtsi](http://jim.teknokrat.ac.id/index.php/jtsi)
- [4] P. Alkhairi And Z. Situmorang, "Penerapan Data Mining Untuk Menganalisis Kepuasan Pegawai Terhadap Pelayanan Bidang Sdm Dengan Algoritma C4.5," *Jurasik (Jurnal Ris. Sist. Inf. Dan Tek. Inform.)*, Vol. 7, No. 1, P. 40, 2022, Doi: 10.30645/Jurasik.V7i1.414.
- [5] R. Putra Eshardiansyah, N. Sulistiyowati, M. Jajuli, T. Informatika, U. H. Singaperbangsa Karawang Jl Ronggowaluyo Kel Puseurjaya Kec Telukjambe Timur Kab Karawang, And P. Jawa, "Algoritma C4.5 Untuk Klasifikasi Jenis Kekerasan Pada Anak (Kasus Dp3a Kabupaten Karawang)," *J. Sains Komput. Inform. (J-Sakti)*, Vol. 5, No. September, Pp. 687–696, 2021.
- [6] P. Alkhairi, I. S. Damanik, And A. P. Windarto, "Penerapan Jaringan Saraf Tiruan Untuk Mengukur Korelasi Beban Kerja Dosen Terhadap Peningkatan Jumlah Publikasi," *Pros. Semin. Nas. Ris. Inf. Sci.*, Vol. 1, No. September, P. 581, 2019, Doi: 10.30645/Senaris.V1i0.65.
- [7] D. Y. Hakim Tanjung, "Optimalisasi Algoritma C4.5 Untuk Prediksi Kerusakan Mesin Atm," *Infosys (Information Syst. J.)*, Vol. 6, No. 1, P. 12, 2021, Doi: 10.22303/Infosys.6.1.2021.12-21.
- [8] I. Mulia And M. Muana, "Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 Dan Software Weka," *Jas-Pt (Jurnal Anal. Sist. Pendidik. Tinggi Indones.)*, Vol. 5, No. 1, P. 71, 2021, Doi: 10.36339/Jaspt.V5i1.417.
- [9] H. Hariati, M. Wati, And B. Cahyono, "Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah Di Kabupaten Kutai Kartanegara," *J. Rekayasa Teknol. Inf.*, Vol. 2, No. 2, P. 106, 2018, Doi: 10.30872/Jurti.V2i2.1861.
- [10] T. B. Tusarwenda, "Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan Botol Pada Cv. Seribukilo," 2018.
- [11] Hendrianto, "Manajemen Strategi Pengelolaan Pasar Dalam Meningkatkan Pendapatan Pedagang Perspektif Ekonomi Islam (Studi Di Pasar Segamas Purbalingga)," Pp. 1–94, 2018.
- [12] F. Fitriana, E. Utami, And H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 Pada Media Sosial Twitter Menggunakan Support Vector Machine Dan Naive Bayes," *J. Komtika (Komputasi Dan Inform.)*, Vol. 5, No. 1, Pp. 19–25, 2021, Doi: 10.31603/Komtika.V5i1.5185.
- [13] A. D. I. Suradi, "Penerapan Data Mining Untuk Menentukan Rekomendasi Beasiswa Dengan Metode Algoritma C4.5," 2018.
- [14] L. Y. Lumban Gaol, M. Safii, And D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, Vol. 2, No. 2, Pp. 97–106, 2021, Doi: 10.30645/Brahmana.V2i2.71.
- [15] Y. Yuliana, P. Paradise, And K. Kusrini, "Sistem Pakar Diagnosa Penyakit Ispa Menggunakan Metode Naive Bayes Classifier Berbasis Web," *Csrid (Computer Sci. Res. Its Dev. Journal)*, Vol. 10, No. 3, P. 127, 2021, Doi: 10.22303/Csrid.10.3.2018.127-138.