

# Penerapan Algoritma C4.5 Untuk Memprediksi Prestasi Siswa Di SD Negeri 071028 Luzamanu

Ade May Luky Harefa<sup>1\*</sup>, Mhd. Ramadhan<sup>2</sup>

<sup>1</sup>Magister Teknologi Informasi, Universitas Pembangunan Panca Budi Medan

<sup>2</sup>Magister Ilmu Manajemen, STMIK Pelita Nusantara, Medan

\*ademayluky@gmail.com,

## Article History:

Received Jul 22<sup>th</sup>, 2024

Revised Aug 08<sup>th</sup>, 2024

Accepted Aug 22<sup>th</sup>, 2024

## Abstrak

SD Negeri 071028 Luzamanu merupakan instansi pendidikan yang bergerak dibidang pendidikan yang berupaya menciptakan kemampuan, keterampilan dan keahlian belajar para siswa dan siswinya. Tujuan penelitian ini adalah untuk mengidentifikasi faktor-faktor yang mempengaruhi prestasi siswa di SD Negeri 071028 Luzamanu dan mengembangkan model prediksi menggunakan algoritma C4.5. Penelitian ini juga bertujuan untuk mengevaluasi akurasi model prediksi tersebut dalam memprediksi prestasi akademik siswa. Algoritma C4.5 merupakan suatu model prediksi yang menggunakan struktur pohon keputusan. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan data record. Secara umum, algoritma C4.5 untuk membangun pohon keputusan dimulai dari pemilihan atribut sebagai akar, membuat cabang untuk tiap tiap nilai, membagi kasus dalam cabang. Dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dapat diketahui bahwa atribut dengan Gain tertinggi adalah Nilai Sikap dengan nilai 0,433593, sehingga atribut Nilai Sikap dapat menjadi Node Akar berikutnya. Penelitian ini menggunakan sistem algoritma c4.5 dapat membantu pihak sekolah untuk memprediksi prestasi siswa pada sekolah UPT SD Negeri 071028 Luzamanu. Model prediksi yang dikembangkan menggunakan algoritma C4.5 menunjukkan akurasi yang tinggi dalam memprediksi prestasi akademik siswa berdasarkan data yang tersedia.

**Kata Kunci :** Data Mining, Algoritma C4.5, Prediksi, Siswa,

## Abstract

*SD Negeri 071028 Luzamanu is an educational institution focused on creating the skills, abilities, and expertise of its students. The need for information services is very important, especially in this research on predicting student performance. Through the data mining process, it can facilitate predicting student performance at SD Negeri 071028 Luzamanu. The C4.5 algorithm is a predictive model that uses a decision tree structure. Data in decision trees is usually represented in the form of tables with data records. Generally, the C4.5 algorithm for building decision trees starts with selecting attributes as the root, creating branches for each value, splitting cases into branches, and repeating the process for each branch until all cases in the branch have the same class. It is known that the attribute with the highest Gain is Attitude Score with a value of 0.433593, so the Attitude Score attribute can be the next Root Node. This research using the C4.5 algorithm system can help the school predict student performance at UPT SD Negeri 071028 Luzamanu by selecting three (3) criteria: Knowledge Score, Attitude Score, and Exam Score.*

**Keyword :** Data Mining, C4.5 Algorithm, Prediction, Students

## 1. PENDAHULUAN

Pada era kemajuan teknologi yang sangat pesat ini terutama dalam pengetahuan dan teknologi beserta aplikasinya disemua kalangan bidang tidak bisa lepas dari yang namanya perangkat laptop ataupun komputer(MZ et al., 2020). Penggunaan komputer saat ini sudah menjangkau hampir segala bidang dalam aktivitas sehari hari dalam kehidupan manusia, baik dalam lingkungan pendidikan, organisasi, perusahaan maupun masyarakat umum dan sebagai manusia kita

harus bisa menggunakan teknologi dimasa yang akan datang(Ryanwar, 2020). Manusia juga harus dapat beradaptasi dengan teknologi baru yang sudah dikembangkan, salah satu adalah dengan dunia pendidikan(Fricticarani et al., 2023). Penggunaan komputer terbukti banyak membantu kita dalam melakukan pekerjaan dengan lebih baik dan tidak memakan banyak waktu yang cukup lama(Aeni et al., 2023).

Data mining adalah kegiatan menemukan pola yang menarik dari data yang disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya(Panggabean et al., 2020). Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan(Mustika et al., 2021).

UPT SD Negeri 071028 Luzamanu merupakan instansi pendidikan yang berupaya menciptakan kemampuan, keterampilan dan keahlian belajar para siswa dan siswinya. Prediksi prestasi tersebut dilakukan agar nantinya mata pelajaran yang diikuti oleh para siswa menjadi minat belajar sehingga para siswa dapat mengembangkan potensi yang dimilikinya

Permasalahan yang terjadi saat ini dalam prediksi prestasi tersebut adalah dalam prakteknya masih di lakukan secara manual sehingga membutuhkan waktu yang lama dengan jumlah siswa yang banyak.

Penelitian yang dilakukan oleh (Tumanggor & Hasugian, 2021) dengan judul penelitian “Penerapan Data Mining Untuk Memprediksi Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4.5 Pada SDN 105351 Bakaran Batu”, penelitian ini membahas tentang yang muncul sistem untuk mengolah tumpukan data agar dapat mengetahui mampu atau tidaknya siswa tersebut mengikuti mata pelajaran pada SDN 105351 Bakaran Batu. Proses penerapan aplikasi data mining ini, hanya membahas tentang memprediksi tingkat kemampuan anak dalam mengikuti mata pelajaran.

Penelitian ini dilakukan (Hasmin & Aisa, 2019) dengan judul penelitian “Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa”, penelitian ini membahas tentang adalah untuk memberikan solusi membangun proses rekomendasi penerimaan beasiswa untuk mahasiswa dengan menggunakan tiga kombinasi kategori penilaian yang ada, serta memproses kategori tersebut menggunakan algoritma C4.5. dalam menciptakan pohon keputusan berdasarkan dataset training dari basis data penerima beasiswa lima tahun terdahulu. Agar dihasilkan phono keputusan untuk melakukan seleksi penerima beasiswa yang tepat sasaran. Tujuan dari penelitian ini guna membangun aplikasi Penunjang keputusan yang terbentuk dari pembelajaran data lima tahun terakhir untuk menyeleksi mahasiswa STMIK Dipanegara Makassar yang berhak menerima beasiswa, dalam prosesnya menggunakan algoritma C4.5 untuk menentukan pemberian beasiswa.

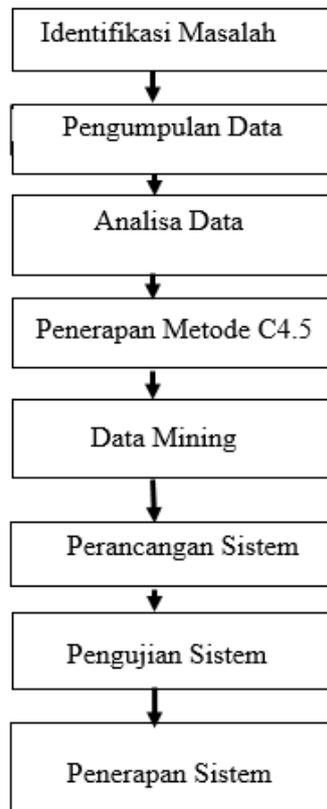
Namun, penelitian ini memiliki beberapa gap yang perlu diisi. Penelitian sebelumnya hanya memfokuskan pada kemampuan mengikuti mata pelajaran atau penentuan penerima beasiswa mahasiswa, sedangkan penelitian ini mengeksplorasi prediksi prestasi akademik secara lebih luas, mencakup berbagai aspek seperti Nilai Pengetahuan, Nilai Sikap, Dan Nilai Ujian. Selain itu, belum ada penelitian yang menerapkan algoritma C4.5 untuk memprediksi prestasi siswa di SD Negeri 071028 Luzamanu, yang memiliki karakteristik unik dan mungkin berbeda dari sekolah lain yang telah diteliti.

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi prestasi siswa di SD Negeri 071028 Luzamanu dan mengembangkan model prediksi menggunakan algoritma C4.5. Penelitian ini juga bertujuan untuk mengevaluasi akurasi model prediksi tersebut dalam memprediksi prestasi akademik siswa.

## 2. METODOLOGI PENELITIAN

### Tahapan Penelitian

Tahapan-tahapan penelitian menguraikan kerangka kerja penelitian atau tahap-tahap yang dilakukan dalam melaksanakan penelitian. Tahapan penelitian ini dilakukan agar dapat menyelesaikan masalah yang akan dibahas. Hal ini dilakukan dengan tujuan untuk mengumpulkan informasi-informasi berkaitan dengan masalah yang akan diteliti. Sehingga dapat diketahui keadaan atau kedudukan masalah tersebut baik secara teoritis maupun praktis. Pengetahuan yang diperoleh dari studi pendahuluan sangat berguna untuk menyusun kerangka teoritis tentang pemecahan masalah dalam bentuk hipotesis yang akan diuji kebenarannya melalui pelaksanaan penelitian lapangan(Yuliani & Banjarnahor, 2021). Studi pendahuluan dapat dilakukan dengan studi kelayakan, kepustakaan dan studi lapangan. Untuk menyelesaikan penelitian ini makan digambarkan tahapan-tahapan penelitian yang dilakukan seperti pada gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian

1. Identifikasi masalah  
Pada tahap ini dilakukan identifikasi masalah, yang ditemukan yaitu bagaimana cara yang tepat untuk mengambil keputusan dalam mengelompokkan prestasi siswa pada setiap kelasnya.
2. Pengumpulan Data  
Pada penelitian ini penulis mendapatkan data langsung dari objek penelitian melalui pengamatan yang telah dilakukan, pengamatan ini dilakukan secara terus menerus tentang sumber data, bahwasannya penulis ini sedang melakukan sebuah penelitian, penulis juga menggunakan pengamatan *passive particion*, yaitu penelitian datang ke tempat peneliti tapi tidak ikut terlibat dalam kegiatan ditempat penelitian untuk mendapatkan data yang sangat lengkap, hasil data yang sudah diperoleh yakni yang berupa nama siswa, NISN, tanggal lahir, kelas, dan nilai mata pelajaran.
3. Analisa Data  
Pada tahap penelitian ini dilakukan proses pemeriksaan, pembersihan, dengan tujuan menemukan informasi yang sangat berguna, menggunakan analisis data kualitatif setelah data yang sudah diperoleh yakni melalui metode wawancara maka data yang diperlukan dikelompokkan, pada tahap ini data yang didapat adalah data siswa kelas 6 semester ganjil dan semester genap tahun ajaran 2021-2022.
4. Penerapan metode C4.5  
Pada penelitian ini merupakan salah satu metode data mining untuk memprediksi prestasi siswa dilihat dari minat belajar.
5. Data Mining  
Pemakaian data yang sangatlah besar untuk menemukan suatu pola, pada penelitian ini memakai metode algoritma C4.5(Dewi, 2020).  
Berikut Tahapan-tahapannya :
  - a. *Selection*  
Pada tahap ini memilih data yang dipakai, data yang diseleksi adalah data siswa kelas 6 semester ganjil dan semester genap tahun ajaran 2021-2022.
  - b. *Transformation*  
Di tahapan ini data yang asli akan diubah ke dalam kriteria yang telah ditentukan, nilai siswa yang berupa angka akan berubah menjadi baik ataupun cukup.
  - c. *Data Mining*  
Aplikasi teknik data mining, proses mencari pola dari sebuah data yang sudah ada.
  - d. *Interpretation/Evaluation*

Pada proses ini menjadi pengetahuan yang telah dapat digunakan untuk mendukung dalam pengambilan keputusan.

### Data Mining

Data Mining adalah proses untuk mendapatkan informasi yang baru dan berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan diharapkan dapat membantu dalam pengambilan suatu keputusan (Wahyudi et al., 2022). Data mining atau Knowledge Discovery In Database (KDD) (Mutammimul Ula et al., 2022), merupakan proses pengambilan data informasi yang tersembunyi, dimana sebuah informasi tersebut sebelumnya tidak dikenal. Proses pendekatan secara teknik seperti clustering, klasifikasi dengan meliputi metode yang merupakan irisan artificial intelligence (AI), machine learning (ML) dan statistic (Winarty & Dewi, 2023).

Proses KDD ini terdiri dari tujuh tahapan yaitu (Darmansah & Wardani, 2020):

- a. Pembersih Data (*data cleaning*)  
Pembersih data ini adalah proses untuk menghilangkan data yang sudah tidak relevan.
- b. Integrasi Data (*data integration*)  
Integrasi data ini merupakan proses dalam menggabungkan sebuah data dari sejumlah database kedalam suatu database yang baru.
- c. Seleksi Data (*data selection*)  
Data yang sudah ada didatabase seringkali tidak semua data itu diperlukan.
- d. Transformasi Data (*data transformation*)  
Data ini digabung dan telah diubah sesuai dengan proses yang digunakan dalam data mining.
- e. Proses Mining  
Proses ini ialah menggali data dari sebuah database atau sekumpulan data untuk mendapatkan sebuah informasi yang tersembunyi dari sekumpulan data yang di olah.
- f. Evaluasi Pola (*pattern evaluation*)  
Proses ini adalah hasil dari teknik data mining yang berupa pola yang akan diuji pada hipotesa yang telah ada dibuat sebelumnya.
- g. Presentasi pengetahuan (*knowledge presentasion*)  
Ini termasuk salah satu langkah akhir dari data mining. Pada tahap ini saatnya untuk mempresentasikan hasil yang sudah dilakukan.

### Pohon Keputusan

*Decision Tree* atau Pohon Keputusan merupakan salah satu teknik dari Analisis Keputusan (*Decision Analysis*). Terdapat banyak definisi teknis yang bisa ditemukan dalam berbagai sumber mengenai pohon keputusan karena beragamnya aplikasi pohon keputusan ini pada berbagai jenis proses dan industri yang berbeda-beda. Definisi ini yang sangat sederhana mengenai apa itu Pohon Keputusan, adalah diagram analisis yang dapat membantu kita dalam pengambilan keputusan ketika menghadapi sebuah opsi dengan cara memproyeksikan hasil yang akan mungkin terjadi (Algoritma et al., 2022).

#### Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross Quinlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan salah satu atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja, yaitu: Membuat pohon keputusan dan membuat aturan-aturan (*rule model*). Aturan-aturan yang terbentuk dari pohon keputusan yang akan membentuk suatu kondisi dalam bentuk *if then* (Febriani & Sulistiani, 2021). Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, adalah (Setio et al., 2020):

1. Menyiapkan data *training*. Data *training* biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon akar yang diambil dari atribut yang terpilih dengan cara menghitung nilai *Gain* dari masing-masing atribut, nilai *Gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *Gain* dari atribut, hitung dahulu nilai *entropy* yaitu:

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (1)$$

Keterangan:

S : himpunan kasus.

n : jumlah partisi S.

$p_i$  : proporsi dari  $S_i$  terhadap S.

3. Kemudian hitung nilai *Gain* dengan metode *information gain*:

$$Gain(S_i) = Entropy(S) - \sum_{i=1}^n |S_i| |S| * Entropy(S_i) \quad (2)$$

**Keterangan:**

S = Himpunan Kasus.

A = Atribut.

n = Jumlah Partisi Atribut.

A | Si | = Jumlah Kasus pada partisi ke-i.

| S | = Jumlah Kasus dalam S.

4. Ulangi langkah ke-2 hingga semua semua tupel terpartisi.
5. Proses partisi pohon keputusan yang berhenti saat:
  - a. Semua tupel dalam node N mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
  - c. Tidak ada tupel di dalam cabang yang kosong.

**3. HASIL DAN PEMBAHASAN**

Data nilai dan data siswa yang dalam bentuk angka untuk mempermudah didalam pemahan dan analisa data, maka nilai-nilai dalam bentuk angka dikompersi menjadi bentuk nilai dalam bentuk keterangan atau nilai dalam bentuk yudisium dengan range nilai Nilai 1-75 nilainya “Cukup” Nilai 76-100 nilainya “Baik”. Pada data kriteria nilai yang digunakan untuk memprediksi Prestasi siswa dalam mengikuti mata pelajaran, yaitu Nilai

Berikut adalah penerapan Algoritma C4.5 yang sudah ditentukan melalui proses perhitungan adalah sebagai berikut :

Dengan menggunakan dua persamaan diatas maka akan didapatkan entropy dan gain yang digunakan sebagai akar dalam membuat pohon keputusan.

**1. Entropy Total**

$$\text{Entropy Total} = \left(-\frac{57}{60} \times \text{Log}_2 \left(\frac{57}{60}\right)\right) + \left(-\frac{3}{60} \times \text{Log}_2 \left(\frac{3}{60}\right)\right) \\ = 0,019233$$

Entropy Total adalah menghitung nilai Hasil Berprestasi (57) dan Tidak Berprestasi (3) dimana 60 adalah jumlah kasus keseluruhan.

**2. Entropy Atribut Nilai Pengetahuan**

Untuk Atribut Nilai Pengetahuans terdiri dari 2 nilai yaitu : Baik dan Cukup dimana nilai Entropy masing-masing adalah sebagai berikut :

Nilai Pengetahuan: Baik

(Jumlah Kasus = 56, Hasil Berprestasi = 55, Hasil Tidak Berprestasi = 1)

$$\text{Entropy (Baik)} = \left(-\frac{55}{56} \times \text{Log}_2 \left(\frac{55}{56}\right)\right) + \left(-\frac{1}{56} \times \text{Log}_2 \left(\frac{1}{56}\right)\right) \\ = 0,12923$$

Nilai Pengetahuan : Cukup

(Jumlah Kasus = 4, Hasil Berprestasi = 2, Hasil Tidak Berprestasi = 2)

$$\text{Entropy (Baik)} = \left(-\frac{2}{4} \times \text{Log}_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \times \text{Log}_2 \left(\frac{2}{4}\right)\right) \\ = 1$$

**3. Entropy Atribut Nilai Sikap**

Untuk Atribut Nilai Sikap terdiri dari 2 nilai yaitu : Baik dan Cukup dimana nilai Entropy masing-masing adalah sebagai berikut:

Nilai Sikap: Baik

(Jumlah Kasus = 55, Hasil Berprestasi = 54, Hasil Tidak Berprestasi = 1)

$$\text{Entropy (Baik)} = \left(-\frac{54}{55} \times \text{Log}_2 \left(\frac{54}{55}\right)\right) + \left(-\frac{1}{55} \times \text{Log}_2 \left(\frac{1}{55}\right)\right) \\ = 0,13110$$

Nilai Sikap: Cukup

(Jumlah Kasus = 5, Hasil Berprestasi = 3, Hasil Tidak Berprestasi = 2)

$$\text{Entropy (Baik)} = \left(-\frac{3}{5} \times \text{Log}_2 \left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} \times \text{Log}_2 \left(\frac{2}{5}\right)\right)$$

$$= 0,97095$$

**4. Entropy Atribut Nilai Ujian**

Untuk Atribut Nilai Ujian terdiri dari 2 nilai yaitu : Baik dan Cukup dimana nilai Entropy masing-masing adalah sebagai berikut :

Nilai Ujian: Baik

(Jumlah Kasus = 48, Hasil Berprestasi = 48, Hasil Tidak Berprestasi = 0)

$$\text{Entropy (Baik)} = \left( -\frac{48}{48} \times \text{Log}_2 \left( \frac{48}{48} \right) \right) + \left( -\frac{0}{48} \times \text{Log}_2 \left( \frac{0}{48} \right) \right) = 0$$

Nilai Ujian: Cukup

(Jumlah Kasus = 12, Hasil Berprestasi = 9, Hasil Tidak Berprestasi = 3)

$$\text{Entropy (Baik)} = \left( -\frac{9}{12} \times \text{Log}_2 \left( \frac{9}{12} \right) \right) + \left( -\frac{3}{12} \times \text{Log}_2 \left( \frac{3}{12} \right) \right) = 0,81127$$

**5. Gain Nilai Pengetahuan**

$$= \text{Entropy (s)} - \sum_{i=1}^n \frac{\text{Nilai Pengetahuan}}{\text{Total}} \times \text{Entropy Nilai Pengetahuan}$$

$$= 0,28639 - \left( \left( \frac{56}{60} \times 0,12923 \right) + \left( \frac{4}{60} \times 1 \right) \right)$$

$$= 0,099112$$

**6. Gain Nilai Sikap**

$$= \text{Entropy (s)} - \sum_{i=1}^n \frac{\text{Nilai Sikap}}{\text{Total}} \times \text{Entropy Nilai Sikap}$$

$$= 0,28263 - \left( \left( \frac{55}{60} \times 0,13110 \right) + \left( \frac{5}{60} \times 0,97095 \right) \right)$$

$$= 0,085303$$

**7. Gain Nilai Ujian**

$$= \text{Entropy (s)} - \sum_{i=1}^n \frac{\text{Nilai Ujian}}{\text{Total}} \times \text{Entropy Nilai Ujian}$$

$$= 0,28263 - \left( \left( \frac{48}{60} \times 0 \right) + \left( \frac{12}{60} \times 0,81127 \right) \right)$$

$$= 0,124141$$

Maka langkah selanjutnya adalah menentukan Entropi dan Gain untuk menentukan akar atau node. Berikut ini adalah hasil perhitung entropi dan gain pada node 1. selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Table :

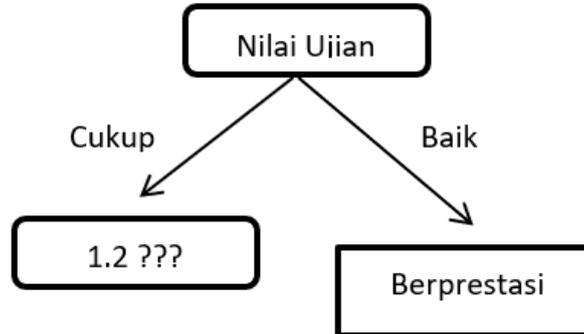
Pengetahuan,Nilai Sikap,dan Nilai Ujian. Setelah seluruh nilai Entropy dan Gain diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Nilai Akhir Node 1.

Tabel 1. Nilai akhir node 1

No	Nama Nilai	Nilai Keadaan	Jumlah Kasus (S)	Berprestasi	Tidak Berprestasi	Entropy	Gain
1	Total		60	57	3	0,28639	
2	Nilai Pengetahuan	Baik	56	55	1	0,12923	0,099114
		Cukup	4	2	2	1	
3	Nilai Sikap	Baik	55	54	1	0,13110	0,085303
		Cukup	5	3	2	0,97095	
4	Nilai Ujian	Baik	49	49	0	0	0,131415
		Cukup	11	8	3	0,845350	

Dari Tabel 4.1 nilai akhir node1 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Nilai Ujian dengan nilai 0,131415, sehingga atribut Nilai Ujian dapat menjadi Node Akar. Pada atribut Nilai Ujian yang dijadikan akar terdapat

2 nilai yaitu “Baik”, dan “Cukup”. Dari nilai atribut tersebut, nilai “Baik” mempunyai Hasil = “Baik” sehingga tidak perlu dilakukan perhitungan kembali. Sedangkan nilai “Cukup” harus dilakukan perhitungan kembali.



Gambar 2. Pohon keputusan

Dari gambar 2 bahwa yang dijadikan Node akar adalah Nilai Ujian, Nilai Ujian ini dibagi ke dalam 2 tipe yaitu “Berprestasi” dan “Tidak Berprestasi”. Untuk nilai “Baik” tidak perlu dilakukan perhitungan kembali dikarenakan hasil dari nilai tersebut adalah “Berprestasi” sedangkan untuk nilai “Cukup” harus dilakukan perhitungan kembali dikarenakan hasilnya belum bernilai “Tidak Berprestasi”. Perhitungan dilakukan untuk mencari node cabang dari nilai atribut baik. Perhitungan dilakukan dengan mencari nilai dari atribut selain yang menjadi node akar (Nilai Ujian), yaitu dengan mencari jumlah kasus untuk hasil Berprestasi dan Tidak Berprestasi, dan nilai Entropy dari semua kasus saat Nilai Ujian = Baik. Dan kemudian lakukan perhitungan untuk mencari nilai Gain, dan atribut dengan nilai Gain terbesar, maka akan menjadi node cabang dari nilai atribut Baik. Jumlah Kasus Nilai Ujian = Cukup = 11. Jumlah Kasus Nilai = Cukup, yang hasilnya Berprestasi = 8. Jumlah Kasus Nilai Ujian =Cukup, yang hasilnya Tidak Berprestasi = 3. Setelah seluruh nilai Entropy dan Gain diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 2.

Table 2. Hasil Perhitungan Entropy dan Gain Node 1

No	Nama Nilai	Nilai Keadaan	Jumlah Kasus (S)	Berpres tasi	Tidak Berpres tasi	Entropy	Gain
1	Total		11	8	3	0,845350	
2	Nilai Pengetahuan	Baik	9	8	1	0,503258	0,433593
		Cukup	2	0	2	0	
3	Nilai Sikap	Baik	9	8	1	0,503258	0,433593
		Cukup	2	0	2	0	

Pada perhitungan pada Tabel 2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Nilai Sikap dengan nilai 0,433593 , sehingga atribut Nilai Sikap dapat menjadi Node Akar berikutnya.

**1. Entropy Atribut Nilai Pengetahuan**

Untuk Atribut Nilai Pengetahuan terdiri dari 2 nilai yaitu : Baik dan Cukup dimana nilai Entropy masing-masing adalah sebagai berikut :

Nilai Pengetahuan: Baik

(Jumlah Kasus = 10, Hasil Berprestasi = 9, Hasil Tidak Berprestasi = 1)

$$\text{Entropy (Baik)} = \left(-\frac{9}{10} \times \text{Log}_2 \left(\frac{9}{10}\right)\right) + \left(-\frac{1}{10} \times \text{Log}_2 \left(\frac{1}{10}\right)\right) = 0,46899$$

Nilai Pengetahuan : Cukup

(Jumlah Kasus = 2, Hasil Berprestasi = 0, Hasil Tidak Berprestasi = 2)

$$\text{Entropy (Baik)} = \left(-\frac{0}{2} \times \text{Log}_2 \left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \text{Log}_2 \left(\frac{2}{2}\right)\right) = 0$$

**2. Entropy Atribut Nilai Sikap**

Untuk Atribut Nilai Sikap terdiri dari 2 nilai yaitu : Baik dan Cukup dimana nilai Entropy masing-masing adalah sebagai berikut :

Nilai Sikap: Baik

(Jumlah Kasus = 10, Hasil Berprestasi = 9, Hasil Tidak Berprestasi = 1)

$$\text{Entropy (Baik)} = \left(-\frac{9}{10} \times \text{Log}_2 \left(\frac{9}{10}\right)\right) + \left(-\frac{1}{10} \times \text{Log}_2 \left(\frac{1}{10}\right)\right)$$

$$= 0,46899$$

Nilai Pengetahuan: Cukup

(Jumlah Kasus = 2, Hasil Berprestasi = 0, Hasil Tidak Berprestasi = 2)

$$\text{Entropy (Baik)} = \left(-\frac{0}{2} \times \text{Log}_2 \left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \text{Log}_2 \left(\frac{2}{2}\right)\right)$$

$$= 0$$

**1. Gain Nilai Pengetahuan**

$$= \text{Entropy (s)} - \sum_{i=1}^n \frac{\text{Nilai Pengetahuan}}{\text{Total}} \times \text{Entropy Nilai Pengetahuan}$$

$$= 0,81127 - \left(\left(\frac{10}{12} \times 0,46899\right) + \left(\frac{2}{12} \times 0\right)\right)$$

$$= 0,420448$$

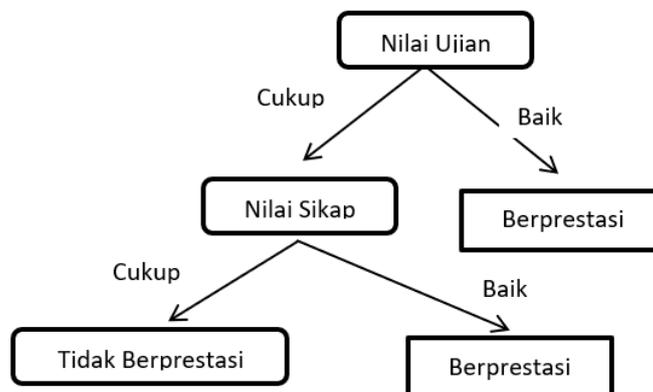
**2. Gain Nilai Sikap**

$$= \text{Entropy (s)} - \sum_{i=1}^n \frac{\text{Nilai Sikap}}{\text{Total}} \times \text{Entropy Nilai Sikap}$$

$$= 0,81127 - \left(\left(\frac{10}{12} \times 0,46899\right) + \left(\frac{2}{12} \times 0\right)\right)$$

$$= 0,420448$$

Pada perhitungan pada Tabel 4.9 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Nilai Sikap dengan nilai 0,420448 , sehingga atribut Nilai Sikap dapat menjadi Node Akar berikutnya.



Gambar 3 . Hasil Akhir Pohon Keputusan

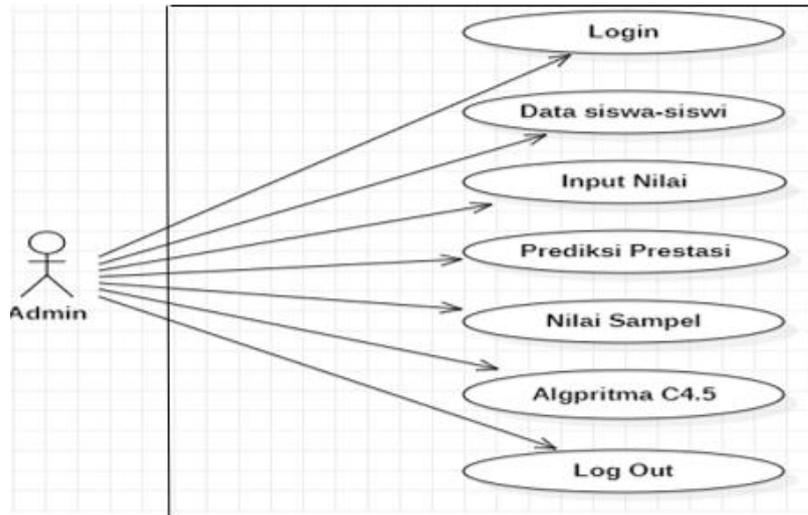
Berdasarkan dari pohon keputusan yang terbentuk seperti pada gambar 3, maka dapat dihasilkan rule sebagai berikut:

1. *If* Nilai Ujian = “Baik” maka Hasil = “Berprestasi”.
2. *If* Nilai ujian = “Cukup” dan Nilai Sikap = “Baik” maka hasil = “Berprestasi”.
3. *If* Nilai ujian = “Cukup” dan Nilai Sikap = “Cukup” = “Tidak Berprestasi”.

**3.1 Perancangan Sistem Yang Di Usulkan**

1. *Use Case Diagram*

*Use Case Diagram* merupakan gambaran dari beberapa *actor*, *use case* dan interaksi dari komponen-komponen tersebut yang memperkenalkan suatu sistem yang hendak dibangun. Berikut ini adalah model *use case diagram* dalam penelitian ini:

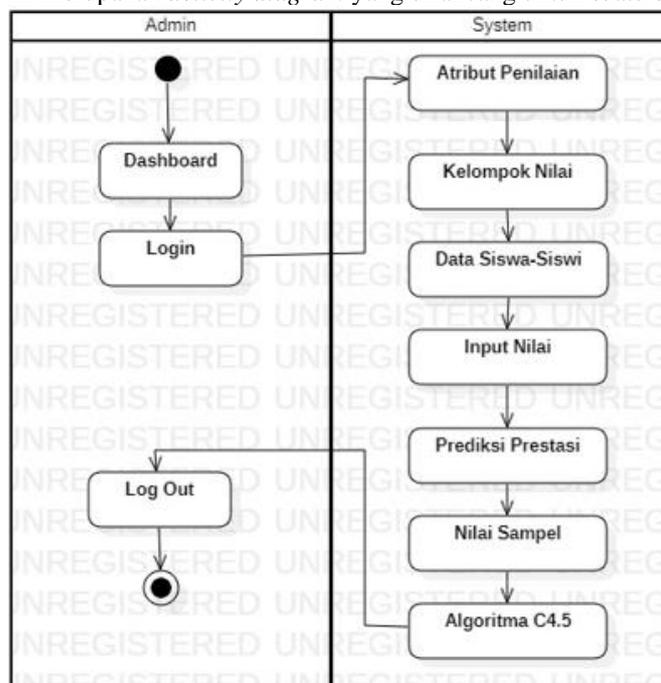


Gambar 4 Use Case Diagram

Pada gambar 4 diatas menunjukkan bahwa sistem yang akan dibuat memiliki satu aktor yaitu *user* dan 6 *use case* lainnya yaitu : login, data Siswa-siswa, Input Nilai, Prediksi Prestasi, Nilai Sampel, Algoritma C4.5, laporan pada Prediksi Prestasi Siswa di sekolah Mts Al-ihklasiyah sei buluh . Adapun hubungan aktor dengan *use case* yang ada menunjukkan hubungan satu arah antara *user* dan sistem. Oleh karena itu komunikasi dimulai dari pihak *user*.

2. Activity Diagram

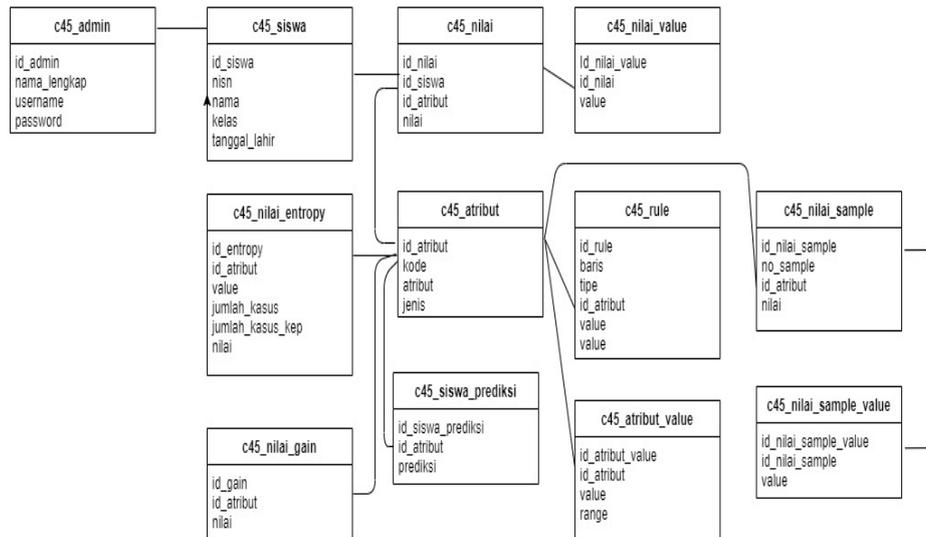
*Activity diagram* adalah salah satu memodelkan *event-event* yang terjadi di *use case*. Diagram ini sama dengan *flowchat* hanya saja dalam diagram ini lebih di jelaskan secara detail bagaimana alur yang akan dijalankan pada suatu sistem. Berikut ini merupakan *activity diagram* yang dirancang untuk suatu sistem sebagai berikut :



Gambar 5 Activity Diagram

3. Class Diagram

*Class diagram* merupakan diagram yang menampilkan beberapa kelas serta paket yang ada pada sistem. Diagram kelas memberi gambaran dan relasi-relasi yang ada didalamnya. Rancangan *class diagram* untuk sistem sebagai berikut :

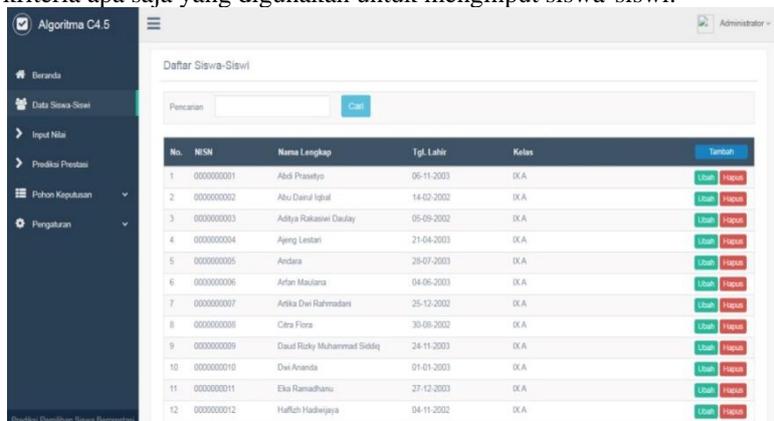


Gambar 6 Class Diagram

### 3.2 Implementasi

#### 1. Halaman Data Siswa-Siswi

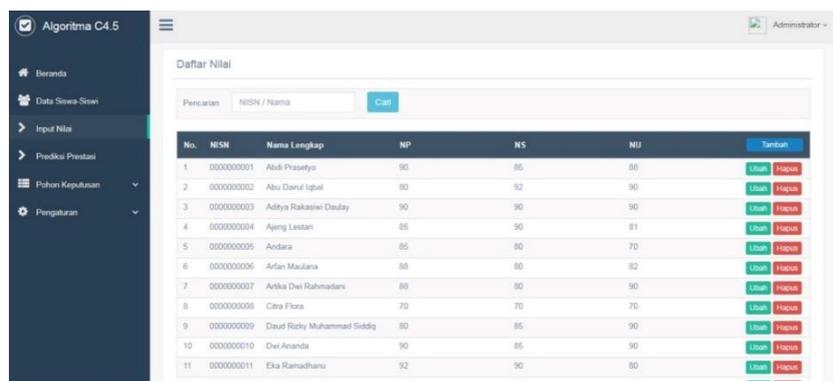
Halaman yang berisi tombol untuk menambah, memperbaharui, dan menghapus data Siswa-siswi ini berfungsi untuk melihat kriteria apa saja yang digunakan untuk menginput siswa-siswi.



Gambar 7 Tampilan halaman data siswa-siswi

#### 2. Halaman Input Data Nilai

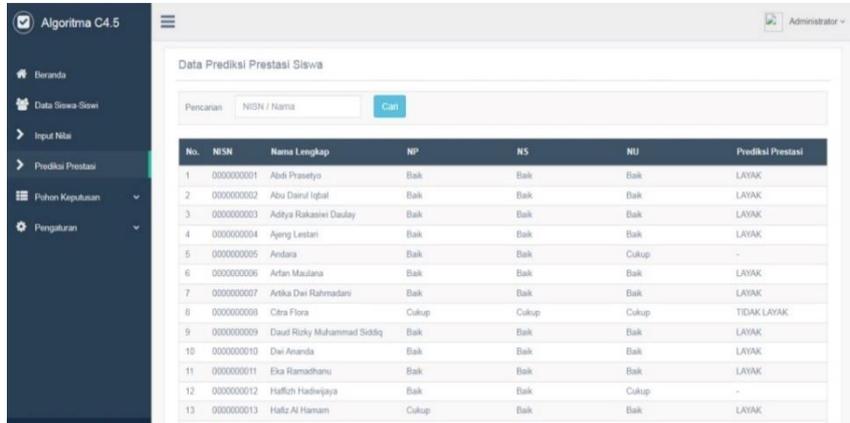
Halaman yang berisi tombol untuk menambah, memperbaharui, dan menghapus data Penginputan Nilai ini berfungsi untuk melihat subkriteria apa saja yang digunakan untuk menginput nilai siswa-siswi.



Gambar 8 Tampilan halaman input data nilai

3. Halaman Prediksi Prestasi

Halaman yang berisi tombol untuk menambah, memperbaharui, dan menghapus Data Prediksi Prestasi ini berfungsi untuk melihat Prestasi siswa yang masuk dalam kategori perhitungan.

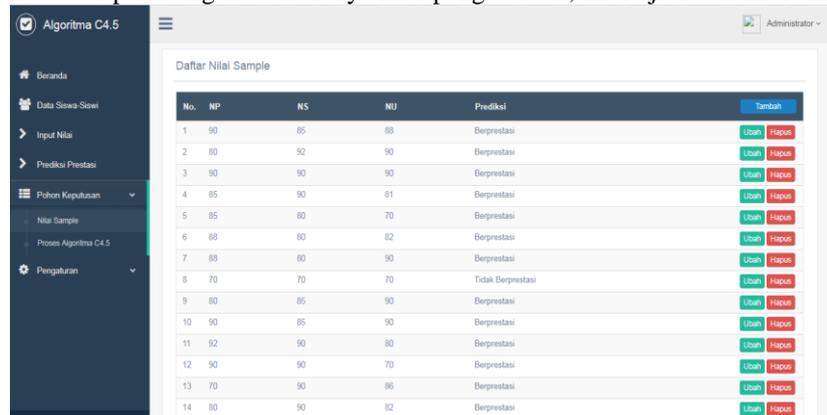


No.	NISN	Nama Lengkap	NP	NS	NU	Prediksi Prestasi
1	000000001	Abdi Prasetyo	Baik	Baik	Baik	LAYAK
2	000000002	Abu Darul Iqbal	Baik	Baik	Baik	LAYAK
3	000000003	Aditya Rakasari Dauly	Baik	Baik	Baik	LAYAK
4	000000004	Ayeng Lestari	Baik	Baik	Baik	LAYAK
5	000000005	Andara	Baik	Baik	Cukup	-
6	000000006	Artan Maulana	Baik	Baik	Baik	LAYAK
7	000000007	Artika Dei Rahmadani	Baik	Baik	Baik	LAYAK
8	000000008	Citra Flora	Cukup	Cukup	Cukup	TIDAK LAYAK
9	000000009	David Rizky Muhammad Siddiq	Baik	Baik	Baik	LAYAK
10	000000010	Dei Ananda	Baik	Baik	Baik	LAYAK
11	000000011	Eka Ramadhani	Baik	Baik	Baik	LAYAK
12	000000012	Halifah Hadwijaya	Baik	Baik	Cukup	-
13	000000013	Haliz Al Hamam	Cukup	Baik	Baik	LAYAK

Gambar 9 Halaman Prediksi Prestasi

4. Nilai Sample

Pada halaman ini adalah untuk melihat data dari nilai sampel yang mana atributnya berbentuk nilai nilai yang menjadi dasar sebuah perhitungan diantaranya nilai pengetahuan, nilai ujian dan nilai sikap.

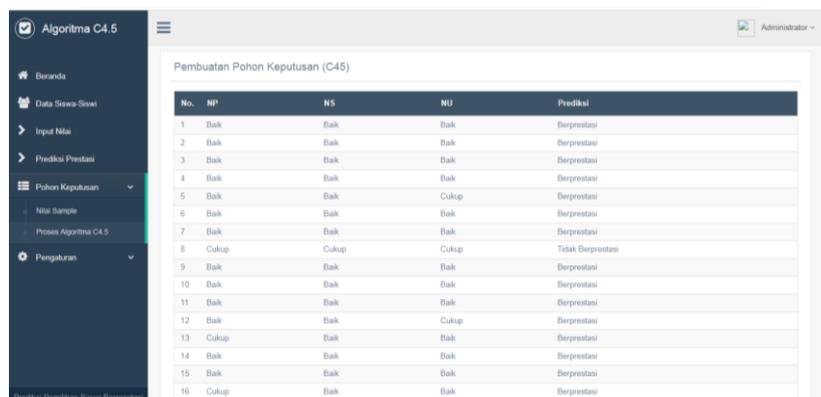


No.	NP	NS	NU	Prediksi
1	90	85	88	Berprestasi
2	80	92	90	Berprestasi
3	90	90	90	Berprestasi
4	85	90	81	Berprestasi
5	85	80	70	Berprestasi
6	88	80	82	Berprestasi
7	88	80	90	Berprestasi
8	70	70	70	Tidak Berprestasi
9	80	85	90	Berprestasi
10	90	85	90	Berprestasi
11	92	90	80	Berprestasi
12	90	90	70	Berprestasi
13	70	90	86	Berprestasi
14	80	90	82	Berprestasi

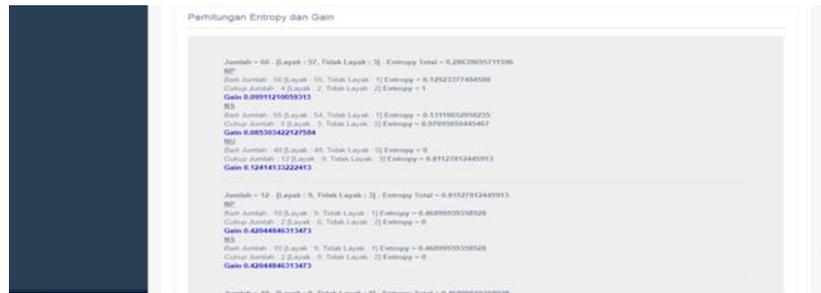
Gambar 10 Atribut Penilaian

5. Rancangan Halaman Keluaran Algoritma C4.5

Halaman yang berisi menampilkan hasil dari perhitungan siswa-siswi yang berprestasi serta menampilkan hasil dari perhitungan berbentuk nilai entropy dan gain serta menampilkan hasil dari pohon keputusannya .



No.	NP	NS	NU	Prediksi
1	Baik	Baik	Baik	Berprestasi
2	Baik	Baik	Baik	Berprestasi
3	Baik	Baik	Baik	Berprestasi
4	Baik	Baik	Baik	Berprestasi
5	Baik	Baik	Cukup	Berprestasi
6	Baik	Baik	Baik	Berprestasi
7	Baik	Baik	Baik	Berprestasi
8	Cukup	Cukup	Cukup	Tidak Berprestasi
9	Baik	Baik	Baik	Berprestasi
10	Baik	Baik	Baik	Berprestasi
11	Baik	Baik	Baik	Berprestasi
12	Baik	Baik	Cukup	Berprestasi
13	Cukup	Baik	Baik	Berprestasi
14	Baik	Baik	Baik	Berprestasi
15	Baik	Baik	Baik	Berprestasi
16	Cukup	Baik	Baik	Berprestasi



Gambar 11 Pohon Keputusan

#### 4. KESIMPULAN

Penelitian ini memberikan kontribusi signifikan dalam bidang pendidikan, khususnya dalam prediksi prestasi siswa menggunakan Algoritma C4.5. Dengan menerapkan algoritma ini pada data siswa UPT SD Negeri 071028 Luzamanu, penelitian ini membuktikan bahwa teknik data mining dapat membantu pihak akademik dalam memprediksi prestasi siswa berdasarkan kriteria yang ditentukan, yaitu Nilai Pengetahuan, Nilai Sikap, dan Nilai Ujian. Hasil penelitian menunjukkan bahwa dari 60 data siswa yang diuji, terdapat 3 siswa yang tidak berprestasi, yakni siswa dengan nomor urut data 8, 27, dan 56. Penelitian ini memberikan dasar untuk pengembangan sistem prediksi yang lebih canggih, yang dapat digunakan oleh pihak sekolah untuk mengidentifikasi siswa yang memerlukan bantuan lebih awal.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah data yang digunakan masih terbatas, hanya 60 data siswa, sehingga hasilnya mungkin tidak sepenuhnya representatif untuk populasi yang lebih besar. Kedua, penelitian ini hanya mempertimbangkan tiga kriteria dalam proses prediksi, yang mungkin belum mencakup semua faktor yang mempengaruhi prestasi siswa. Ketiga, penelitian ini belum menguji keefektifan sistem prediksi ini dalam jangka panjang.

Untuk penelitian ke depan, disarankan agar jumlah data yang digunakan diperbanyak agar hasilnya lebih representatif. Selain itu, perlu dipertimbangkan untuk menambah lebih banyak kriteria yang relevan, seperti faktor lingkungan, keterlibatan orang tua, dan kegiatan ekstrakurikuler, untuk mendapatkan prediksi yang lebih akurat. Penelitian lanjutan juga perlu menguji implementasi sistem prediksi ini dalam jangka waktu yang lebih lama untuk menilai keefektifannya dalam membantu meningkatkan prestasi siswa secara berkelanjutan.

#### DAFTAR PUSTAKA

- Aeni, R. N., Firmansyah, A., & Hadikristanto, W. (2023). Sistem Informasi Kasir Pada Cv Tunas Suka Nyata Berbasis Desktop Menggunakan Metode Waterfall. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 7(3), 657–671.
- Algoritma, I. C., Sistem Prediksi Prestasi Siswa Di SMK Al-Ishlah Berbasis Web, U., Martantoh, E., & Putri Haryanto, A. (2022). Implementation of The C4.5 Algorithm for A Web-Based Student Achievement Prediction System in SMK Al-Ishlah. In *JTSI* (Vol. 3, Issue 2).
- Anzalia, A., & Hasugian, P. S. (2020). Application of Data Mining in Determining Sales Patterns at 212 Mart Supermarkets in Lubuk Pakam Using the Apriori Algorithm. *Journal of Intelligent Decision ...*, 3(4).
- Darmansah, D., & Wardani, N. W. (2020). Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 126–134.
- Dewi, A. O. P. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *Anuva: Jurnal Kajian Budaya, Perpustakaan, Dan Informasi*, 4(2), 223–230.
- Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *89Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(4), 89–95.
- Frictarani, A., Hayati, A., Ramdani, R., Hoirunisa, I., & Rosdalina, G. M. (2023). Strategi pendidikan untuk sukses di era teknologi 5.0. *Jurnal Inovasi Pendidikan Dan Teknologi Informasi (JIPTI)*, 4(1), 56–68.
- Hasmin, E., & Aisa, S. (2019). Penerapan algoritma c4. 5 untuk penentuan penerima beasiswa mahasiswa. *Cogito Smart Journal*, 5(2), 308–320.
- Jijon R. Sagala, Penda Sudarto Hasugian, & Sulindawaty. (2022). 258-Article Text-1210-1-10-Penerapan Sistem Informasi Profil Desa Berbasis Web Di Desa Situnggaling. *Jurnal Pengabdian Kepada Masyarakat Nusantara (JPKMN)*, 2(2), 91–96.
- Mustika, Ardila, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, Manuhutu, M. A., Ridwan, M., Hozairi, Wardhani, A. K., Alim, S., Romli, I., Religia, Y., Octafian, D. T., Sufandi, U. U., & Ernawati, I. (2021). *Data Mining dan Aplikasinya*.

- Mutammimul Ula, Ananda Faridhatul Ulva, Mauliza, Muhammad Abdullah Ali, & Yumna Rilasmi Said. (2022). Application Of Machine Learning In Predicting Children's Nutritional Status With Multiple Linear Regression Models. *Multica Science And Technology (MST) JOURNAL*, 2(2), 124–130. <https://doi.org/10.47002/mst.v2i2.363>
- MZ, S. P. H. S., Marzuki, & Asslia Johar Latipah. (2020). Penerapan Teknologi Informasi Dalam Pengembangan Pariwisata Berbasis Smart Village Desa Aikdewa. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, 1(1), 8–17. <https://doi.org/10.46764/teknimedia.v1i1.8>
- Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56–62.
- Ryanwar, R. (2020). *Penerapan Metode Algoritma C4. 5 Untuk Memprediksi Loyalitas Karyawan Pada Pt. Xyz Berbasis Web*. Kodeuniversitas041060# UniversitasBuddhiDharma.
- Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4. 5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Tumanggor, A., & Hasugian, P. S. (2021). Penerapan Data Mining Untuk Memprediksi Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4. 5 Pada SDN 105351 Bakaran Batu. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 4(1), 57–63.
- Wahyudi, A. K., Azizah, N., & Saputro, H. (2022). Data Mining Klasifikasi Kepribadian Siswa SMP Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4. 5. *Journal of Information System and Computer*, 2(2), 8–13.
- Winarty, D., & Dewi, Y. N. (2023). Penerapan K-Means Dalam Pengelompokkan Pemantauan Kasus Terkonfirmasi Covid-19. *INTI Nusa Mandiri*, 17(2), 78–83.
- Yuliani, W., & Banjarnahor, N. (2021). Metode penelitian pengembangan (rnd) dalam bimbingan dan konseling. *Quanta Journal*, 5(3), 111–118.