



Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292
web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



PENERAPAN DATA MINING UNTUK MEMPREDIKSI PRESTASI SISWA SEKOLAH DASAR MENGGUNAKAN ALGORITMA C4.5

Telutci, Rika Harman

Universitas Putera Batam, Indonesia.

INFORMASI ARTIKEL

Diterima Redaksi: Januari 2024
Diterbitkan Online: Maret 2024

KATA KUNCI

Pendidikan,
Memprediksi,
Data Mining,
Algoritma C45

KORESPONDENSI

E-mail: Telutcioei932@gmail.com,
rika@puterabatam.ac.id

A B S T R A C T

In general, education in Indonesia is one of the most important aspects of the development of society and the state. By understanding the factors that influence student achievement, it can improve the quality of education and provide a better understanding of the factors that influence student achievement. These factors include report card grades, attendance, family support, interest in learning, and motivation to learn. These factors ensure that students achieve their goals and provide a better opportunity to improve their academic performance in primary school. Dhamma Sasana Primary School In predicting student achievement, there are many obstacles or problems that must be faced by schools in predicting student achievement, many comparisons of variables and factors to be tested, and decision making on the achievement of elementary school students. This research applies the C4.5 algorithm used to predict student achievement at Dhamma Sasana Elementary School. The C4.5 algorithm is one of the data analysis tools that is able to identify the variables that have the most influence on student achievement. Conclusions include manual calculations and found decision trees and testing conducted with data mining techniques with the C4.5 algorithm classification method, and the highest gain value resulted in 0.4628, including the C4.5 algorithm and testing conducted with the software application weka version 3.9.6.

I. Latar Belakang

Secara Umum Pendidikan Di Indonesia merupakan salah satu aspek yang sangat penting dalam pembangunan masyarakat dan negara. selaku tenaga pendidik pada sekolah dasar akan menjalankan tugasnya dengan baik bila mempunyai memahami konsep pendidikan pada sekolah dasar. Pembelajaran akan mengajak siswa untuk mengkaji secara mendalam konsep-konsep dasar pendidikan agar siswa mampu memahami dengan baik dan mempunyai wawasan terkait pendidikan di sekolah dasar.

wawasan, pemahaman dan pendalaman siswa mengenai pendidikan di sekolah dasar selalu harus mempersegar serta memperkaya diri agar pengetahuan dan konsep-konsep mengenai pendidikan di sekolah dasar tersebut merupakan hasil berpikir manusianya yang sifatnya dinamis, perubahan akibat pengaruh keadaan dan kondisi kehidupan[1]. Dari data Kementerian Pendidikan, Kebudayaan, Riset dan Informasi Teknologi (Kemendikbud Ristek) memberikan informasi jumlah siswa di Indonesia sebesar 53,14 juta orang pada semester ganjil tahun

ajaran 2023/2024. Dari jumlah tersebut, sebagian besar siswa berada di tingkat sekolah dasar (SD), yaitu sebesar 24,04 juta orang. Data pendidikan SD di Indonesia berjumlah 24.035.934 siswa. Dari hasil analisis data tersebut, terdapat data siswa yang cukup banyak untuk menganalisis siswa yang berprestasi dalam pembelajaran di sekolah.

Prestasi belajar siswa adalah hasil yang dicapai atau diperoleh siswa baik berupa pengetahuan, keterampilan maupun sikap berkat pengalaman dan latihan-latihan yang telah dilalui oleh individu. Belajar yang diperoleh anak setelah mengerjakan tugas dan ujian. Dengan memahami faktor-faktor yang memengaruhi prestasi siswa dapat meningkatkan kualitas pendidikan dan memberikan pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi prestasi siswa. Faktor tersebut diantaranya nilai rapor, kehadiran, dukungan keluarga, minat belajar, motivasi belajar. Faktor-faktor ini memastikan bahwa siswa mencapai tujuan mereka dan memberikan kesempatan yang lebih baik untuk meningkatkan prestasi akademik mereka di sekolah dasar. Oleh karena itu, pemahaman yang mendalam ini merupakan dasar penting untuk mengembangkan pendidikan yang lebih baik yang memberikan peluang yang sama untuk tumbuh dan berkembang bagi setiap siswa di sekolah dasar, [2].

Bagi Sekolah Dhamma Sasana adalah sekolah yang didedikasikan untuk menyebarkan ajaran Buddha dan pendidikan agama Buddha kepada siswa di bawah naungan yayasan Jaya Manggala yang memiliki lembaga pendidikan formal yang berdiri pada tahun 2017 sehingga terdiri dari KB.

Dalam memprediksi prestasi belajar siswa, banyak kendala atau permasalahan yang harus dihadapi sekolah dalam memprediksi prestasi belajar siswa, banyak perbandingan variabel-variabel dan faktor-faktor yang akan diuji dan pengambilan keputusan terhadap prestasi siswa sekolah dasar.

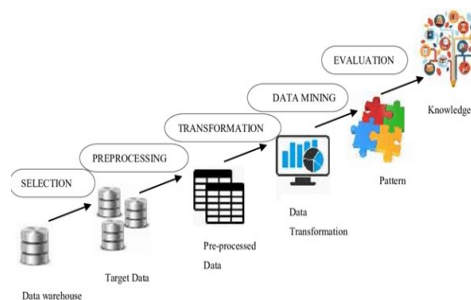
Penelitian ini menerapkan Algoritma C4.5 digunakan untuk memprediksi prestasi siswa pada Sekolah Dasar Dhamma Sasana. Salah satu tantangan dalam dunia pendidikan adalah mengidentifikasi faktor-faktor yang berpengaruh terhadap prestasi siswa. Algoritma C4.5 adalah salah satu alat analisis data yang mampu

mengidentifikasi variabel-variabel yang paling berpengaruh terhadap prestasi siswa. Dengan menerapkan algoritma ini, kita dapat lebih memahami faktor-faktor apa saja yang dapat memengaruhi prestasi siswa di Sekolah Dasar Dhamma Sasana.

II. Kajian Literatur

2.1. KDD (Knowledge Discovery In Database)

Data mining yang merupakan singkatan dari Knowledge Discovery In Database adalah ilmu yang memanfaatkan data yang berasal dari database atau hasil dari data yang tersimpan dalam database. Dalam ilmu pengetahuan yang terakhir ini digunakan sebagai pengetahuan dasar dalam mengambil keputusan. Berikut ini adalah adapun tahap-tahap dalam proses pelaksanaan KDD, seperti berikut ini.



Gambar 1. Bentuk kdd ataupun disebut *knowledge discovery in database*

- (1) **Pemilihan Data atau *data selection*** Tahapan pertama dimulai sejak awal proses KDD dimulai dengan adanya mengumpulkan informasi dan proses pemilihan hasil datanya yang akan dipergunakan untuk dijadikan sumber data yang nantinya akan diolah dalam tahapan data mining dan disimpan dalam file yang berbeda dengan basis data operasional, [3].
- (2) ***Pre-Processing atau Cleaning Tahapan*** ini bertujuan untuk melakukan proses data mining untuk menjalankan data, kemudian akan menghilangkan beberapa duplikasi pada data [1] dengan cara mengecek data yang tidak sesuai dengan kebutuhan data yang akan diproses dalam penambahan data yang harus sesuai dengan kebutuhan penelitian, [4].

- (3) Transformation Proses transformasi data yang dapat dipilih, dari bentuk pengkodean untuk proses perubahan data dan penyesuaian pola tipe data, Dalam pemrosesan KDD, bertujuan membentuk kreatif dari tipe atau pola data sebagai informasi yang jelas dalam penyimpanan basis data yang sudah ditetapkan dalam prosedur pengolahan data mining, [3].
- (4) Data Mining Proses pemilihan tugas data mining, pada proses KDD melalui klasifikasi dalam pemilihan algoritma data mining sangat bervariasi dalam proses pencarian dan pembentukan pola aturan untuk menghasilkan informasi berupa keputusan berdasarkan tujuan penelitian dan pemrosesan KDD secara keseluruhan, [3].
- (5) Interpretation/evaluation Menampilkan hasil pola aturan dan informasi dari data mining perlu ditampilkan dalam bentuk sehingga dapat dipahami oleh pembaca, memeriksa pola dan informasi yang dapat ditemukan bertentangan dengan fakta atau hipotesis penelitian sebelumnya, [3].

2.2 Data mining adalah proses pengumpulan informasi yang dapat dimanfaatkan sebagai data penting dari bentuk data yang besar untuk kegiatan analisis yang dapat menjelaskan ilmu dan sepengetahuan agar lebih jelas dan mudah dipahami bagi para pembaca, [5].



Gambar 2. Dibidang ilmu Data mining

Bentuk proses data mining memiliki teknik pengolahan data tersendiri sehingga bentuk suatu pola berfungsi untuk mengenal pola lainnya agar tidak tersimpan dalam tempat persimpanan data atau informasi yang mirip. Dalam melakukan proses memprediksi, banyaknya peneliti yang melaksanakan

beberapa kegiatan dalam menjadikan fungsinya dari sebuah teknik data mining. Awal munculnya istilah data mining berdasarkan pada sebuah data serta informasi untuk diolah layaknya kegiatan penambangan untuk menghasilkan data yang bermanfaat, [5].

2.3 Pengelompokan data mining

Jika dibagi sesuai dengan tahap urutan tugasnya serta proses atau tahapan yang perlu dilaksanakan didalam data mining, maka bisa terbagi menjadi lebih dari enam tahap dan sebagian, sebagai berikut:

- (1) Description ataupun deskripsi
 Pada tahapan ini dimulai saat seseorang sedang melaksanakan kegiatan analisis informasi sebagai pengklasifikasian suatu tipe pola yang terkandung dalam data serta sumber informasi. Sehingga dapat tergambar ketika seseorang siswa yang berperan dalam melakukan kegiatan pendataan dan mengumpulkan suaranya saat kegiatan pemilihan untuk mengetahui ataupun memperoleh informasi secara jelas tentang baik maupun kurangnya kemampuan siswa dengan karakter tertentu yg dimiliki oleh para calon siswa yg akan mendaftarkan diri saat seleksi siswa berprestasi.
- (2) Estimation ataupun estimasi
 Sedangkan pada tahap proses alur kerja ini memiliki estimasi untuk klasifikasi, namun dapat berbeda dengan variabel target yang memiliki estimasi untuk data bersifat numeric. Pola tersebut terbuat dari perekaman data yang berisikan nilai ataupun digunakan sebagai bahan dalam memberikan nilai rapor pada siswa dalam proses prediksinya. Selanjutnya diteruskan dengan memberikan nilai rapor dengan estimasi didapat dari guru sebagai variabel yaitu berasal dari nilai rapor setiap siswa.
- (3) Prediction ataupun Prediksinya
 Untuk proses ataupun tahap prediksinya sebenarnya sama saja seperti proses klasifikasi, untuk proses estimasi, dalam memprediksi jumlah nilai tersebut digunakan untuk kedepannya. Sehingga prediksinya ini secara khusus berkaitan dengan aktivitas sekolah maupun penelitian, sebagai berikut:

a. Ketika menentukan prediksinya agar diketahui perkiraan bertahap biaya peralatan sekolah dalam setiap bulanya.

b. Ketika menyediakan prediksinya tentang kesempatan siswa berprestasi di sekolah selama satu tahun ke depan.

(4) Classification ataupun klasifikasi

Didalam proses klasifikasi, menggunakan sejumlah sasaran dan variabel sesuai kebutuhan indikator yang telah ada. Di sini bisa diilustrasikan ketika mengelompokkan prestasi yang diraih siswa di sekolah. Pengelompokan dapat dilakukan dengan tiga kategori, Berdasarkan siswa yang berprestasi sangat baik, siswa dengan kategori tingkat sedang atau menengah dikatakan cukup, kategori siswa yang tidak berprestasi dikatakan kurang. Kemudian jika klasifikasi tersebut dijelaskan dalam kaitannya dengan kegiatan penelitian, sebagai berikut:

a. Ketika melakukannya proses klasifikasi maka ada motivasi belajar untuk mencapai hasil tersebut baik ataupun kurang.

b. Ketika melakukannya peramalan perencanaan prestasi belajar untuk meningkatkan pengetahuan siswa berdasarkan baik dan kurang.

(5) Clustering ataupun pengklusteran

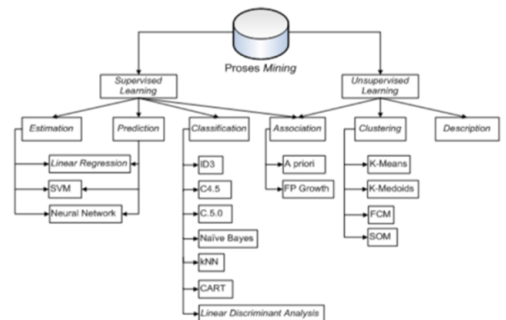
Tahap proses ini dilakukannya pada clustering yaitu aktivitas ataupun proses pengelompokkan hasilnya merekam datanya lalu disusun berdasar datanya itu mempunyai pola yang serupa. Dalam proses tahapan klasifikasi melalui variabel-variabel yang menjadi target. Tahapan dan proses dalam melakukan estimasi, prediksi pada proses clustering berfokus pada pengelompokan data yang sama dalam variabel dan target.

(6) Association ataupun asosiasi

Asosiasi menjalankan proses dengan tujuan untuk memperoleh kelengkapan ataupun keterangan dukungan atas data yang ada. Tentunya hal ini tergambar dalam bentuk aktivitas seperti yang dilakukannya, sebagai berikut:

a. Ketika melakukannya penelitian ataupun menganalisis jumlahnya siswa berprestasi pada setiap sekolah dengan tanggapan positif kepada orang tua atas prestasi yang diberikan sekolah sebagai siswa berprestasi, [6].

Penjelasan pada gambarnya tentang hasilnya pengelompokan sejumlah proses ataupun bertahap berdasar tugas ketika seorang meneliti untuk melaksanakan penelitian memakai teknik data mining, sebagai berikut.



Gambar 3. Teknik data mining

2.4 Decision Tree (Pohon Keputusan)

Menurut [7] Pohon Keputusan adalah data yang menunjukkan variabel dan tujuan untuk pembagian kumpulan data yang besar sebagai tumpukan catatan yang lebih mini. Kita dapat memperkirakan pendapatan seseorang mengenai hal-hal yang tidak diketahui yang dapat ditentukan dengan menggunakan data siswa dan data nilai rapor sebagai pilihan dan sejumlah faktor lainnya.

Menurut [8], Dalam tindakan manusia diawali dalam berdasarkan lingkungan sekolah dalam pengembangan suatu pola dalam sistem untuk membantu mereka dalam menyelesaikan hal dan masalah yang terjadi dan salah satu hasil adalah pohon keputusan ataupun yang biasa disebut dengan *decision tree*. Pohon keputusan adalah salah satu bentuk dari pengimplementasian pola tertentu dari penerapan berbagai tahap dan proses didalamnya termasuk pula aktivitas klasifikasi dan aktivitas memprediksi. Pohon keputusan menghasilkan alternatif untuk menyelesaikan permasalahan dalam hasilnya yang kuat dan bermanfaat untuk pohon keputusan agar dikenal dan banyak dipergunakan oleh para peneliti ketika melaksanakan penelitian dalam teknik *data mining*. Pohon keputusan, yang dapat dijelaskan sebagai berikut:

1. Simpul bagian akar

Simpul dari bagian akar terletak pada bagian paling atas dan pertama dari struktur yang terdapat pada *decision tree*, [9].

2. Simpul *internal*

Simpul dari bagian *internal* ini adalah percabangan lanjut dari simpul bagian akar.

3. Simpul daun

Simpul merupakan simpul yang berada pada ujung bagian dari pohon.

Selanjutnya, Demikian penjelasan dari kelebihan dan kekurangan yang dimiliki oleh *decision tree*, yaitu:

1. Kelebihan dari *decision tree*

- Dalam menggunakan pohon keputusan, terdapat cakupan data yang sulit, sehingga dirubah menjadi sesederhana mungkin untuk digunakan sebagai dasar pengambilan keputusan.
- Keuntungan menggunakan pohon keputusan adalah dengan menghilangkan hitung-hitungan tersebut hingga proses pengujian cukup memadai indikator yang cocok sebagai kebutuhan dan pengujian.

2. Kekurang dari *Decision Tree*

- Dalam penggunaan pohon keputusan dalam jumlah hasil keputusan untuk menghasilkan keputusan dan indikator yang dipergunakan mempunyai jumlahnya sangatlah banyaknya, makanya waktu yang diperlukan untuk menghasilkan sebuah keputusan akan semakin bertambah dan membutuhkan memori akan meningkat.
- Pada pohon keputusan terkandung perhitungan jumlahnya kesalahan yang muncul pada tiap bagian strukturnya didalam pohon keputusan, [10].

2.5 Algoritma C4.5

Menurut [11], Algoritma C4.5 adalah sesuatu metode yang berguna dalam melakukan proses klasifikasi data dengan menentukan tingkat keakuratan dalam bentuk angka atau menggunakan kategori setelah tahapan proses klasifikasi dapat dilakukan dengan hasil dalam aturan pola yang digunakan dalam melakukan tahapan proses prediksi kemampuan siswa sekolah dasar, prediksi nilai rapor yang berasal dari perekaman data lama menjadi baru [6]. Algoritma c4.5 memiliki faktor-faktor yang dapat mengatasi masalah kemampuan siswa sekolah dasar dalam bidang pembelajaran untuk mencapai nilai tinggi pada mata pelajaran yang bersifat pemecahan masalah,[12].

Tahap pertama dalam pohon keputusan, menentukan pada waktu seleksi kelengkapan ataupun deskripsi data yang perlu dilakukannya dengan mencari nilai *gain* dalam mencari hasilnya yang tinggi dalam semua nilai yang dimiliki oleh kelengkapan ataupun informasi terdapat, maka beberapa rumus dalam mencari nilai *gain* yang tertinggi dan nilai yang sesuai, sebagai berikut:

$$Gain(A) = Entropi(S) - \sum_i^n \frac{|S_i|}{|S|} \times Entropi(S_i)$$

Rumus 1. Perhitungan *Gain*

Keterangan simbolnya, yaitu:

S adalah himpunan

A adalah keterangan atau kelengkapan yang digunakan.

n adalah jumlah partisi yang terdapat pada bagian keterangan atau kelengkapan A

$|S_i|$ adalah jumlah kasus yang terdapat pada partisi ke- i .

$|S|$ adalah jumlah kasus yang terdapat pada S

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

Rumus 2. Perhitungan *Entropy*

Keterangan simbolnya adalah:

S adalah himpunan

A adalah keterangan atau kelengkapan yang digunakan.

n adalah jumlah partisi yang terdapat pada keterangan atau kelengkapan S

p_i adalah proporsi dari S_i terhadap S [13].

2.6 *Waikato Environment for Knowledge Analysis (WEKA)*

Menurut [14], Weka atau kata lain *Waikato Environment for Knowledge Analysis* aplikasi sumber daya yang dikembangkan oleh *University of Waikato*, terdiri dari berbagai algoritma dan alat pembelajaran mesin yang digunakan untuk melakukan pemrosesan data. Weka Versi 3.9.6 juga melakukan tugas-tugas lain seperti mengekstraksi data, melakukan klasifikasi, regresi, pengelompokan, dan penambangan aturan hubungan.

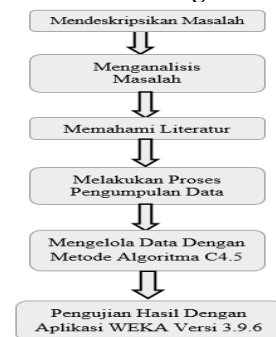
2.7 Microsoft Excel

Microsoft Excel adalah program aplikasi spreadsheet Microsoft yang digunakan untuk memproses data yang dikumpulkan dalam bentuk tabel. Microsoft Excel juga berfungsi sebagai lembar kerja elektronik untuk input dan pemrosesan data secara otomatis. Program ini memungkinkan penghitungan, analisis, pembuatan grafik, dan visualisasi data dari program ini untuk pemrosesan data yang lebih kompleks [15].

2.8 Teori Khusus

Prestasi siswa adalah penguasaan pengetahuan dan keterampilan yang berkaitan dengan mata pelajaran dan sikap yang diperoleh setiap individu melalui pengalaman dan pendidikan matematika, serta hasil atau usaha yang dilakukan untuk mencapai prestasi bagi diri sendiri dikenal dengan prestasi belajar siswa juga merupakan hasil dari upaya sungguh-sungguh yang mereka lakukan untuk mencapainya, [16].

III. Metodologi



Gambar 3.1 Desain Penelitian

Gambar 4. Desain Penelitian

Berdasarkan desain penelitian di atas, model dan gambar alur penelitian dapat dilihat penjelasan sebagai berikut:

- (1) Mendeskripsikan Masalah
Pada awalnya, penulis akan menjelaskan masalah yang akan diteliti dan memprediksi prestasi siswa dengan menggunakan algoritma C4.5.
- (2) Menganalisis Masalah
Pada tahap kedua, penulis ingin melakukan analisis data menggunakan algoritma c4.5 untuk mengetahui seberapa baik siswa berprestasi dan memberikan deskripsi masalah. Penulis juga ingin menganalisis beberapa

pertanyaan yang ingin penulis angkat dan diteliti.

(3) Memahami Literatur

Pada tahap ketiga, penulis menjadikan jurnal penelitian sebagai sumber panduan bagi penulis dalam melakukan penelitian ini, penulis berusaha semaksimal mungkin untuk melakukan penelitian dan memahami beberapa sumber teori.

(4) Melakukan Proses Pengumpulan Data

Pada tahap keempat, peneliti turun ke lapangan untuk melakukan wawancara dengan guru dan ketua yayasan SD Dhamma Sasana. Penulis juga meminta informasi tambahan. Dalam hal ini, menguraikan sumber informasi dari tempat penelitian, termasuk informasi yang sangat penting. Informasi opsional berasal dari tujuan penelitian yang jelas dalam informasi hasil penelitian.

(5) Mengelola Data Dengan Metode Algoritma C4.5

Pada tahap kelima, penulis akan mengolah data menggunakan aturan-aturan dalam perhitungan algoritma C4.5 pada saat proses pembuatan struktur pohon keputusan. Untuk menghasilkan pola-pola yang menunjukkan prestasi siswa dalam bentuk aturan. Selanjutnya akan didapatkan nilai gain tertinggi dari setiap komponen data dengan memakai nilai gain termaksimum untuk proses terbentuknya struktur pohon keputusan.

6. Pengujian Hasil Dengan Aplikasi WEKA

Dalam tahap terakhir, peneliti lakukan pengujian untuk melengkapi hasil hitung nilai gain tertingginya dan pembuatan decision tree dalam algoritma c45 memakai software Weka.

IV. Pembahasan

4.1 Hasil Penelitian

Dalam melaksanakan pengujian data menggunakan ms excel dengan aplikasi yang digunakan weka versi 3.9.6 sebagai software yang akan mengolah data sebagai berikut:

- (1) Kriteria pertama yang mempengaruhi prestasi belajar siswa adalah Motivasi Belajar dengan nilai gainnya 0,4628, dengan rincian siswa yang tidak termotivasi berjumlah 32 orang siswa sedangkan yang termotivasi 30 orang siswa.

2. Kriteria Kedua yang mempengaruhi prestasi belajar siswa setelah motivasi belajar adalah Kehadiran dari siswa dengan nilai gain 0,4900, dengan rincian siswa yang kurang kehadirannya berjumlah 7 orang siswa sebaliknya yang bagus kehadirannya berjumlah 23 orang siswa.
3. Tahap Ketiga yang mempengaruhi prestasi belajar siswa setelah motivasi belajar dan Kehadiran siswa adalah minat belajar dengan nilai gain 0,5586, dengan rincian siswa yang rajin berjumlah 20 orang siswa sedangkan yang kurang 3 orang siswa.
4. Terakhir dapat disimpulkan bawah variable yang mempengaruhi prestasi siswa adalah motivasi belajar setelah itu dapat dilihat kehadiran dan yang terakhir adalah minat belajar dengan nilai gain dan rincian jumlah siswa tertera diatas.

4.2 Pembahasan

Penelitian ini akan memperkirakan bagaimana siswa berprestasi untuk mencapai prestasi belajar yang sempurna, di sekolah dasar dhamma sasana. Dalam penelitian ini menggunakan data dari objek penelitian ini yaitu data siswa Sekolah Dasar Dhamma Sasana dalam periode 1 tahun terakhir, yaitu tahun 2022 semester ganjil dan 2023 semester genap, dengan Indikator Penilaian Prestasi siswa yaitu Nilai rapor, Kehadiran, Dukungan Keluarga, Minat Belajar, Motivasi belajar. Ada dua kategori variabel keputusan: iya berprestasi untuk siswa yang mencapai prestasi belajar dan tidak berprestasi untuk siswa yang tidak mencapai prestasi belajar yang sempurna untuk mendapatkan prestasi belajar siswa. Peneliti menggunakan teknik data mining untuk melakukan klasifikasi metode algoritma C4.5 untuk penelitian ini.

4.3 Melaksanakan Pra Proses Data

Berdasarkan data yang ada, variabel dapat disesuaikan dengan kebutuhan penelitian dan Format data yang sudah ditentukan agar data bisa ditemukan dalam tabel 1 dibawah ini sebagai berikut:

Tabel 1. Format data Prestasi Belajar Siswa

NO	Nilai Rapor	Kehadiran	Dukungan Keluarga	Minat Belajar	Motivasi Belajar	Prestasi Siswa
1	Kurang	Kurang	Ada	Rajin	Tidak Termotivasi	Tidak Berprestasi
2	Baik	Bagus	Ada	Rajin	Termotivasi	Iya Berprestasi

3	Kurang	Bagus	Tidak Ada	Rajin	Tidak Termotivasi	Tidak Berprestasi
4	Kurang	Kurang	Ada	Kurang	Tidak Termotivasi	Tidak Berprestasi
5	Kurang	Bagus	Tidak Ada	Rajin	Tidak Termotivasi	Tidak Berprestasi
6	Baik	Bagus	Ada	Rajin	Termotivasi	Iya Berprestasi
7	Kurang	Kurang	Tidak Ada	Rajin	Tidak Termotivasi	Tidak Berprestasi
8	Kurang	Kurang	Tidak Ada	Kurang	Tidak Termotivasi	Tidak Berprestasi
9	Kurang	Bagus	Ada	Rajin	Termotivasi	Iya Berprestasi
10	Baik	Bagus	Tidak Ada	Kurang	Termotivasi	Tidak Berprestasi
11	Kurang	Bagus	Ada	Rajin	Termotivasi	Iya Berprestasi
62	Baik	Bagus	Ada	Rajin	Termotivasi	Iya Berprestasi

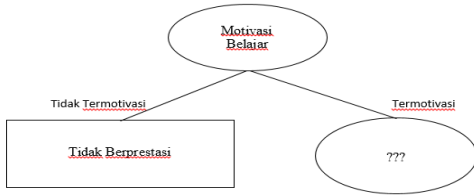
Setelah di hitung nilai entropy ataupun gain dihitung, dengan hasil hitungannya dimasuk ke dalam table 1, yaitu sebagai berikut:

Tabel 2. Perhitungan node 1 (manual)

	Prestasi Siswa	Iya Berprestasi	Tidak Berprestasi	Entropy	Gain
Total		62	42	0,9071	
Nilai Rapor					
Baik	BK	22	12	0,9940	0,0886
Kurang	KR	40	8	0,7219	
		62			
Kehadiran					
Bagus	BG	33	20	0,9672	0,3923
Kurang	KR	29	0	0	
		62			
Dukungan Keluarga					
Ada	AD	34	17	1	0,1369
Tidak Ada	TA	28	3	0,4912	
		62			
Minat Belajar					
Rajin	RJN	39	20	0,9995	0,2784
Kurang	KRG	23	0	0	
		62			
Motivasi Belajar					
Tidak Termotivasi	TTM	32	0	0	0,4628
Termotivasi	TM	30	20	0,9182	
		62			

Dimulai dari perhitungan manual, motivasi belajar ditemukan memiliki nilai gain tertinggi yaitu 0.4628, baik yang termotivasi maupun yang tidak termotivasi merupakan atribut dengan gain tertinggi, seperti yang ditunjukkan oleh perhitungan pada tabel 2. Karena memiliki nilai perhitungan gain tertinggi, maka atribut motivasi belajar dapat dianggap sebagai poin utama. Atribut motivasi belajar dibagi menjadi dua kategori, yaitu tidak termotivasi dan termotivasi. Kategori pertama berfungsi untuk memprediksi prestasi siswa dan melihat akar

dan pohonnya, sedangkan kategori kedua akan digunakan sebagai proses hitung lebih menlanjut untuk mengambil keputusan agar bisa memahami dalam decision tree lebih besar pada gambar 5 di bawah ini:



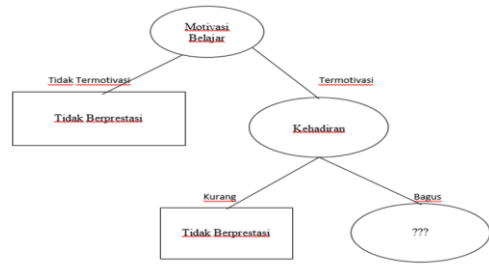
Gambar 5. Pohon Keputusan Hasil Perhitungan Node 1.

Setelah di hitung nilai entropy ataupun gain dihitung, maka hasil hitungannya di masuk ke dalam tabel 3, yaitu sebagai berikut:

Tabel 3. Perhitungan Node 2 (manual)

	Prestasi Siswa	Iya Berprestasi	Tidak Berprestasi	Entropy	Gain
Total	30	20	10	0,9182	
Nilai Rapor					
Baik	BK 17	12	5	0,8739	0,0065
Kurang	KR 13	8	5	0,9612	
	30				
Kehadiran					
Bagus	BG 23	20	3	0,5586	0,4900
Kurang	KR 7	0	7	0	
	30				
Dukungan Keluarga					
Ada	AD 19	17	2	0,4854	
Tidak Ada	TA 11	3	8	0,8453	0,3008
	30				
Minat Belajar					
Rajin	RJN 24	20	4	0,6500	
Kurang	KRG 6	0	6	0	0,3982
	30				

Dari perhitungan yang ditunjukkan pada tabel 3, nilai perhitungan gain tertinggi adalah atribut kehadiran dengan atribut kurang dan bagus, dengan nilai 0.4900. Karena memiliki nilai perhitungan gain tertinggi, maka kehadiran dapat menjadi root node. Selain itu, untuk atribut yang kurang, kategori ini tidak mencapai prestasi belajar, sedangkan untuk atribut yang bagus kategori ini mencapai prestasi belajar yang sempurna.



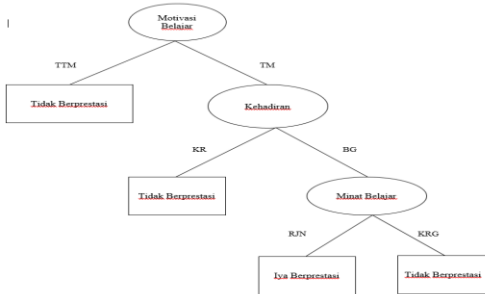
Gambar 6. Pohon Keputusan Hasil Perhitungan Node 2

Proses pada tahap awal ini sangat penting untuk memprediksi keputusan pada tahap selanjutnya, yaitu menghitung Node 2 sebagai akar. Proses ini serupa dengan proses sebelumnya yaitu menghitung nilai entropi dari atribut-atribut yang tersisa, seperti nilai rapor, dukungan keluarga dan minat belajar. Setelah di hitung nilai entropy ataupun gain dihitung, diberikan hasilnya dihitungnya di masuk ke dalam tabel 4, yaitu sebagai berikut:

Tabel 4. Perhitungan node 3 (manual)

	Prestasi Siswa	Iya Berprestasi	Tidak Berprestasi	Entropy	Gain
Total	23	20	3	0,5586	
Nilai Rapor					
Baik	BK 13	12	1	0,3912	0,0236
Kurang	KR 10	8	2	0,7219	
	23				
Dukungan Keluarga					
Ada	AD 17	17	0	0	
Tidak Ada	TA 6	3	3	1	0,2977
	23				
Minat Belajar					
Rajin	RJN 20	20	0	0	0,5586
Kurang	KRG 3	0	3	0	
	23				

Dalam perhitungan table 4. terlihat hasil dari minat belajar, dengan nilai atribut 0,5586, adalah atribut dengan gain tertinggi, seperti yang ditunjukkan oleh perhitungan yang disertakan dalam tabel 4. Berdasarkan nilai gain tersebut, minat belajar adalah node akar dari proses sebelumnya, yaitu kehadiran. Secara garis besar, node akar di sini adalah tiga: motivasi belajar, kehadiran, dan minat belajar. Selain itu, atribut rajin (RJN) dianggap iya berprestasi dengan jumlah prestasi siswa 20 sedangkan atribut kurang (KRG) dianggap tidak berprestasi dengan jumlah prestasi siswa 3 orang siswa.



Gambar 7. Pohon Keputusan Hasil Perhitungan Node 3.

4.4 Pengujian melaksanakan software weka.

Setelah penghitungan manual dilaksanakan untuk menghasilkan pohon keputusan, sangatlah penting menguji hasil perhitungan untuk memastikan bahwa hasilnya benar. Peneliti melakukan penelitian ini untuk menguji benar tidaknya hasil hitung ini dengan memakai program WEKA versi 3.9.6, melakukan hal-hal berikut:

1. Buat file Microsoft Excel dengan nama Dataset Prestasi Siswa.xls, untuk mengumpulkan data dari semua variabel yang digunakan untuk menilai tingkat prestasi belajar siswa.
2. Selanjutnya, buat file.csv. Untuk penelitian ini, peneliti menggunakan program notepad, berikut.
3. Untuk membuka program WEKA 3.9.6, klik dua kali pada shortcut.
4. Klik satu kali pada Explorer untuk membuka kotak dialog, seperti yang ditunjukkan pada gambar 8 di bawah ini.

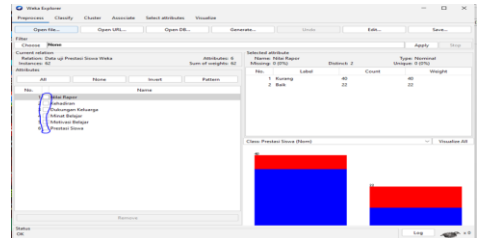


Gambar 8. Menu Awal WEKA 3.9.6

5. Buka WEKA Explorer dan cari lokasi file Uji Data weka.csv. Kemudian,

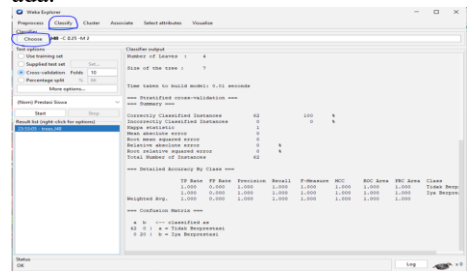
seperti yang ditunjukkan pada gambar 8, pilih tombol "Buka" di bagian bawah kanan kotak dialog.

6. Pilih variabel yang mempengaruhi penelitian, seperti yang ditunjukkan pada gambar 9 di bawah ini.



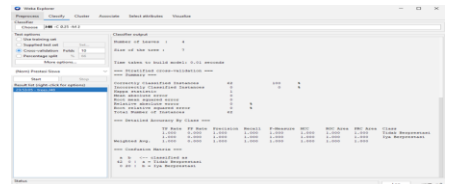
Gambar 9. Pemilihan Variabel pada WEKA

7. Selanjutnya, seperti yang ditunjukkan pada gambar 10 di bawah ini, pergi ke tab Classify dan klik pilih gambar yang ada.



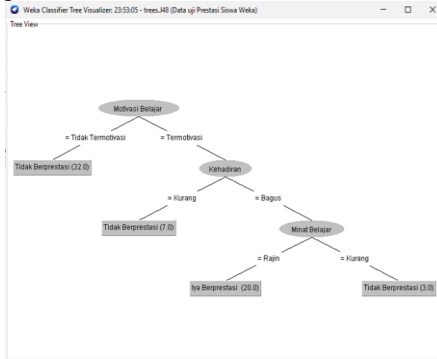
Gambar 10. Tab Classify pada WEKA

8. Seperti yang ditunjukkan pada gambar 10, pilihan akan muncul; pilih Trees, lalu J48
9. Setelah Anda klik start, hasil pengujian WEKA akan muncul di area kanan. Klasifikasi output menampilkan jumlah kasus dan keputusan, bersama dengan banyak cabang dari pohon keputusan. Gambaran yang lebih jelas tentang masalah ini disajikan pada Gambar 11 di bawah ini.



Gambar 11. Visualize Tree Pada WEKA

10. Seperti yang ditunjukkan pada gambar 12 berikut, pilih pohon visualisasi. Untuk melihat pohon keputusan, klik kanan pada daftar hasil.
11. Pohon keputusan WEKA ditampilkan pada Gambar 12 di bawah ini.



Gambar 12. Hasil Visualize Tree Weka

4.5 Keputusan Hasil Pengujian

Setelah pengujian pertama selesai, dapat ditarik kesimpulan, baik melalui pengujian secara manual berupa hitungan matematis atau melalui pengujian perangkat lunak dengan bantuan weka. Kemudian, setelah hasilnya kedua pengujian sesuai, dapat diambil kesimpulan dan diambil kesimpulan tentang motivasi. Untuk memastikan bahwa siswa di sekolah dasar dhamma sasana mencapai prestasi belajar yang optimal, pohon keputusan terakhir digunakan sebagai dasar untuk pembuatan aturan, seperti yang ditunjukkan pada Gambar 7 di atas:

- (1) IF motivasi belajar = tidak termotivasi, maka 32 orang siswa Sekolah Dasar Dhamma sasana tidak Berprestasi tidak mencapai prestasi belajar yang sempurna.
- (2) IF motivasi belajar = termotivasi, maka langkah selanjutnya terlihat kehadiran siswa di Sekolah Dasar Dhamma Sasana.
- (3) IF Kehadiran = Kurang, maka 7 orang siswa Sekolah Dasar Dhamma Sasana tidak berprestasi tidak mencapai prestasi belajar yang sempurna.
- (4) IF Kehadiran = Bagus, maka langkah selanjutnya terlihat minat belajar siswa sekolah dhamma sasana.
- (5) IF Minat Belajar = Kurang, maka 3 orang siswa sekolah dasar dhamma sasana tidak

berprestasi tidak mencapai prestasi belajar yang sempurna sebaliknya.

- (6) IF Minat Belajar = Rajin, maka 20 orang siswa sekolah dasar dhamma sasana iya berprestasi dan mencapai prestasi belajar yang sempurna.
- (7) Untuk memprediksi Prestasi siswa pada sekolah dasar dhamma sasana adalah siswa yang mencapai prestasi belajar sudah termotivasi dan Iya berprestasi siswa di sekolah dasar dhamma sasana, memiliki motivasi untuk belajar dan Jumlah kehadiran yang bagus, dan minat belajar yang rajin.
- (8) Sebaliknya, siswa yang tidak berprestasi tidak mencapai prestasi belajar dari sekolah dasar dhamma sasana adalah siswa yang tidak termotivasi di sekolah dasar dhamma sasana, tidak memiliki jumlah kehadiran atau dikatakan kurangnya jumlah kehadiran, dan memiliki minat belajar yang sangat kurang sehingga tidak bisa mencapai prestasi belajar yang sempurna.

V. Kesimpulan

Peneliti mendapatkan beberapa kesimpulan yang dapat membantu pengambilan keputusan dari hasil pembahasan bab IV, yang mencakup perhitungan manual, penemuan pohon keputusan, dan pengujian metode data mining dan klasifikasi, termasuk algoritma c4.5 dan pengujian dengan software WEKA versi 3.9.6. Hasilnya menunjukkan hal-hal berikut:

- (1) Untuk menilai prestasi siswa untuk mencapai prestasi belajar yang sempurna di sekolah dasar dhamma sasana, teknik datamining dengan metode klasifikasi algoritma C4.5 dapat menemukan aturan atau pengambilan keputusan untuk digunakan sebagai standar atau acuan dalam menilai prestasi siswa untuk mencapai prestasi belajar. Aturan yang dibuat adalah sebagai berikut:
 - (2) Dalam menentukan pohon keputusan yang terdapat pada perhitungan algoritma c4.5 memberikan variabel dan menghasilkan hasil yang terbaik untuk faktor utama dalam mempengaruhi prestasi siswa di sekolah dhamma sasana adalah motivasi belajar, kehadiran, dan minat belajar.
 - (3) Pengujian yang dilakukan dengan software WEKA juga menunjukkan hasil yang sama

untuk perhitungan data mining manual algoritma C4.5, perhitungan dengan software Weka juga menunjukkan hasil yang sama dalam bentuk pohon keputusan dan aturan yang dibuat dengan nilai klasifikasi 100% yang diperoleh, yang menunjukkan bahwa perhitungan manual yang dilakukan oleh aplikasi adalah benar dan memiliki hasil yang memprediksi prestasi.

Ucapan Terima Kasih

Penulis mengucapkan banyak terima kasih kepada Dosen Pembimbing dan seluruh pihak yang telah memberi kontribusi positif dalam penyelesaian penelitian ini.

Daftar Pustaka

- [1] A. P. Barus and N. Sapitri, "Penerapan Data Mining Untuk Memprediksi Prestasi Siswa Menggunakan Algoritma J48," *J. Sist. Inf. dan ...*, vol. 1, pp. 20–27, 2020, [Online]. Available: <http://ejournal.pustakatimur.org/index.php/sisfotekjar/article/view/11>
- [2] A. Noviriandini and N. Nurajijah, "Analisis Kinerja Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Prestasi Siswa Sekolah Menengah Kejuruan," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 23–28, 2019, doi: 10.33480/jitk.v5i1.607.
- [3] R. Harman, "Penerapan Data Mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada Pt Remicon Widyaprima," *J. Comasie*, vol. 3, no. 3, pp. 84–93, 2020.
- [4] T. Tzelepis, G. Matlis, N. Dimokas, P. Karvelis, P. Malliou, and A. Beneka, "An Intelligent Injury Rehabilitation Guidance System for Recreational Runners Using Data Mining Algorithms," *Algorithms*, vol. 16, no. 11, p. 523, Nov. 2023, doi: 10.3390/a16110523.
- [5] E. Elisa, "PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PENJUALAN BARANG PADA PT BATAM BANGUN PRATHAMA," *J. Comasie*, vol. 7, no. 1, pp. 127–133, 2022.
- [6] S. Anastassia Amellia Kharis and A. Haqqi Anna Zili, "Learning Analytics dan Educational Data Mining pada Data Pendidikan," *J. Ris. Pembelajaran Mat. Sekol.*, vol. 6, pp. 12–20, 2022.
- [7] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, 2021, doi: 10.3390/a14110318.
- [8] S. Dewi and Oktaviawati, "Penerapan Algoritma C4.5 untuk Pehamanan Siswa SMK Pada Pelajaran Kompetensi Keahlian," *Intern. (Information Syst. Journal)*, vol. 5, no. 2, pp. 116–125, 2022.
- [9] R. Harman and E. Rosiska, "Prediksi Durasi Kuliah dengan Algoritma Classifier," *InfoTekJar J. Nas. Inform. ...*, vol. 1, 2020, [Online]. Available: <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/1641>
- [10] S. F. Utami, "Penerapan Data Mining Algoritma Decision Tree Berbasis PSO," *Semin. Nas. Teknol. Komput. Sains*, pp. 677–681, 2020.
- [11] E. E. Koko Handoko, Pastima Simanjuntak, Ellbert Hutabri, "Penerapan algoritma c4.5 untuk penentuan jurusan siswa sekolah menengah atas," *J. TEKINKOM*, vol. 6, pp. 153–157, 2023, doi: 10.37600/tekinkom.v6i1.788.
- [12] A. Y. Ramdan and P. Y. Fauziah, "Peran orang tua dan guru dalam mengembangkan nilai-nilai karakter anak usia sekolah dasar," *Prem. Educ. J. Pendidik. Dasar dan Pembelajaran*, vol. 9, no. 2, p. 100, Dec. 2019, doi: 10.25273/pe.v9i2.4501.
- [13] JASMIR and YUDI, "PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI KECOCOKAN GAYA BELAJAR BAGI SISWA SISWI SEKOLAH DASAR (STUDI KASUS :

- SD SARIPUTRA Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar (Studi Kasus: SD Sariputra Jambi),” *J. Process.*, vol. 14, no. 2, pp. 141–152, 2019, doi: 10.33998/processor.2019.14.2.637.
- [14] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [15] F. Fatimah and F. Maisa, “KLASIFIKASI MINAT SISWA DALAM KEGIATAN E-KSTRAKURIKULER DI SEKOLAH DASAR MENGGUNAKAN C45,” *J. Bisnis Digit. dan Sist. Inf. Vol.*, vol. 4, no. 1, pp. 6–17, 2023.
- [16] D. Yuliatika, R. Rusdinal, and N. Gistituati, “Kepemimpinan Kepala Sekolah dalam Pengambilan Keputusan Di Sekolah Dasar,” *Edukatif J. Ilmu Pendidik.*, vol. 3, no. 5, pp. 2944–2951, 2021, [Online]. Available: <https://edukatif.org/index.php/edukatif/article/view/972>