

Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Pemahaman Siswa Pada Mata Pelajaran Matematika

Dinda Zhila Azhari^{1,*}, Irfan Sudahri Damanik², Dedi Suhendro³

^{1,2,3}Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia

Email: ¹dindazhilaazhari@gmail.com, ²irfansudahridamanik@amiktunasbangsa.ac.id, ³dedisuhendro@amiktunasbangsa.ac.id

(*: dindazhilaazhari@gmail.com)

Abstrak

Mata pelajaran matematika merupakan salah satu pelajaran yang penting bagi siswa yang mempunyai peran cukup besar dalam dunia pendidikan. Matematika merupakan ilmu yang perlu dipahami setiap orang, terutama siswa yang masih berada pada jenjang pendidikan. Kemampuan pemahaman matematika dapat mencapai tujuan pembelajarannya apabila dapat memahami konsep dengan baik. Penelitian ini bertujuan untuk mengklasifikasikan konsep pemahaman siswa pada mata pelajaran matematika siswa SMA Swasta Tamansiswa Tapian Dolok. Dalam dunia pendidikan matematika sering dianggap sebagai mata pelajaran yang sulit untuk dipahami karena matematika selalu berhubungan dengan angka, rumus, dan hitung-menghitung. Banyaknya rumus dan konsep yang harus dipahami, membuat siswa kesulitan dalam menyelesaikan persoalan pada materi yang ada. Sumber data diperoleh dari hasil kuesioner pemahaman matematika siswa SMA Swasta Tamansiswa Tapian Dolok. Metode yang digunakan dalam penelitian ini yaitu Data Mining dengan Algoritma C4.5 dan dibantu dengan software RapidMiner. Data Mining merupakan proses yang digunakan untuk mengidentifikasi informasi yang bermanfaat dari berbagai database berjumlah besar. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk decision tree berdasarkan training data. Atribut yang digunakan yaitu minat belajar siswa, komunikasi, cara belajar siswa, suasana pembelajaran, media pembelajaran, dan cara mengajar guru. Hasil Perhitungan nilai entropy dan gain diperoleh 18 rule keputusan pemahaman siswa pada mata pelajaran matematika dengan 9 rules berstatus paham dan 9 rules berstatus tidak paham. Pemodelan klasifikasi dengan Algoritma C4.5 pada Rapidminer diperoleh akurasi sebesar 95.19%.

Kata Kunci: Data Mining, Algoritma C4.5, Klasifikasi, Pemahaman Siswa, Matematika

Abstract

Mathematics is one of the important subjects for students who have a large enough role in the world of education. Mathematics is a science that needs to be understood by everyone, especially students who are still at the level of education. The ability to understand mathematics can achieve its learning objectives if it can understand the concept well. This study aims to classify the concept of students' understanding of mathematics subjects at Tamansiswa Tapian Dolok Private High School students. In the world of mathematics education is often considered a difficult subject to understand because mathematics is always associated with numbers, formulas, and arithmetic. The number of formulas and concepts that must be understood makes it difficult for students to solve problems in the existing material. The source of the data was obtained from the results of the questionnaire on the mathematical understanding of the Tamansiswa Tapian Dolok private high school students. The method used in this study is Data Mining with the C4.5 Algorithm and assisted by RapidMiner software. Data Mining is a process used to identify useful information from a large number of databases. C4.5 algorithm is an algorithm used to form a decision tree based on training data. The attributes used are student learning interest, communication, student learning methods, learning atmosphere, learning media, and teacher teaching methods. The results of the calculation of the entropy and gain values obtained 18 rules for student understanding decisions in mathematics subjects with 9 rules with the status of understanding and 9 rules of not understanding. Classification modeling with the C4.5 Algorithm on Rapidminer obtained an accuracy of 95.19%.

Keywords: Data Mining, C4.5 Algorithm, Classification, Student Understanding, Mathematics

1. PENDAHULUAN

Mata pelajaran matematika merupakan mata pelajaran yang wajib dipelajari mulai dari Sekolah Dasar sampai dengan Perguruan Tinggi. Menurut [1] 'Mata pelajaran matematika merupakan ilmu dasar serta pendukung ilmu pengetahuan di bidang lain, seperti fisika, ekonomi, biologi yang tidak terlepas dari peran matematika'. Banyak siswa yang kurang paham dalam belajar matematika. Salah satu penyebabnya, sebagian siswa masih menganggap matematika sebagai mata pelajaran yang sulit bila dibandingkan dengan mata pelajaran yang lain. Matematika merupakan ilmu pasti yang berkaitan dengan angka yang diperlu dipahami setiap orang, terutama siswa yang masih berada pada jenjang pendidikan. Pembelajaran matematika siswa diharapkan dapat menumbuhkan kemampuan pemahaman sehingga dapat berpikir kritis, logis, sistematis, cermat dan efisien dalam memecahkan masalah [2]. Pada siswa SMA Swasta Tamansiswa Tapian Dolok, matematika merupakan mata pelajaran yang dianggap sulit bila dibandingkan dengan mata pelajaran lain. Matematika sering dianggap sulit oleh siswa karena para siswa sudah berpikir bahwa matematika selalu berhubungan dengan angka, rumus, dan hitung-menghitung. Banyaknya rumus dan konsep yang harus dipahami, sehingga membuat siswa merasa kesulitan dalam menyelesaikan soal pada materi yang ada. Kebanyakan siswa hanya memahami soal yang dikerjakan guru sebagai contoh, ketika bentuk soal sedikit berbeda dari contoh, para siswa sulit untuk menyelesaikan soal pada materi yang ada. Hal ini karena para siswa kurang memahami konsep pada materi tersebut [3].

Berdasarkan permasalahan di atas penulis menggunakan teknik data mining klasifikasi dengan Algoritma C4.5 untuk mendapatkan solusi yang lebih efisien pada tingkat pemahaman siswa terhadap mata pelajaran matematika [4], [5]. Algoritma C4.5 dapat digunakan untuk meneliti berbagai macam hal, diantaranya penelitian terkait penentuan

tingkat pemahaman mahasiswa terhadap matakuliah diperoleh hasil penelitian sebanyak 14 rules dengan 9 rules berstatus paham dan 5 rules berstatus tidak paham dengan tingkat akurasi sebesar 87,10% (Rofiqo et al., 2019). Penelitian yang dilakukan oleh (Saputra et al., 2020) yaitu Klasifikasi dengan konsep data mining dengan melibatkan sebanyak 100 data siswa di SMA YPI Swasta Dharma Budi menghasilkan 20 rules dan tingkat akurasi yang dihasilkan oleh metode tersebut adalah 86.67%.

Menurut [6] 'Data mining disebut juga dengan Knowledge Discovery in Database (KDD) ataupun pattern recognition digunakan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna [7]. Pattern recognition digunakan sebagai pengolahan data untuk menemukan pola tersembunyi dari data yang diolah'. Menurut Ina, [8] 'Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk decision tree berdasarkan training data. Algoritma C4.5 mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya, sedangkan untuk samples merupakan field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data' [9], [10].

Berdasarkan latar belakang tersebut, maka diusulkan penelitian ini dengan tujuan untuk mengklasifikasikan tingkat pemahaman siswa pada mata pelajaran matematika siswa SMA Swasta Tamansiswa Tapan Dolok. Hasil dari penelitian diharapkan dapat dijadikan sebagai bahan pertimbangan khususnya bagi pihak sekolah, guru maupun siswa untuk mengambil beberapa tindakan dalam meningkatkan prestasi belajar siswa pada mata pelajaran matematika, sehingga dapat memberikan kontribusi tentang konsep pemahaman siswa pada mata pelajaran matematika yang dapat membantu dalam meningkatkan pemahaman matematika. Selain itu hasil penelitian ini dapat menjadi acuan untuk penelitian lain dengan topik yang sama.

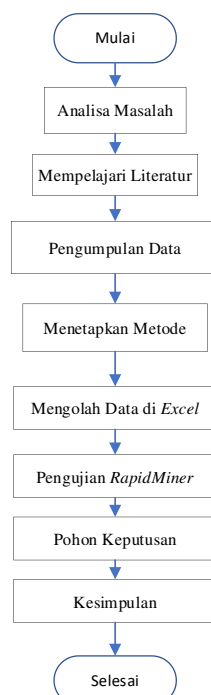
2. METODOLOGI PENELITIAN

2.1. Metode Penelitian

Penelitian ini dilakukan untuk mengimplementasikan data mining untuk klasifikasi tingkat pemahaman siswa pada mata pelajaran matematika di SMA Swasta Tamansiswa Tapan Dolok menggunakan Algoritma C4.5 agar dapat digunakan sebagai salah satu keputusan untuk meningkatkan prestasi siswa di bidang akademik. Pada metodologi penelitian ini akan diuraikan proses cara ilmiah untuk mendapatkan data yang akan digunakan dalam penelitian pemecahan suatu masalah. Metodologi penelitian ini di mulai dengan menggambarkan bagaimana mengumpulkan data yang digunakan. Gambaran terkait pengumpulan data yang dibahas meliputi lokasi penelitian, waktu pengumpulan data dan data yang akan digunakan. Alur dari proses penelitian digambarkan dalam bentuk diagram penelitian yang selanjutnya diperkuat dengan penjelasan terkait tahapan penelitian yang dilakukan.

2.1.1. Rancangan Penelitian

Rancangan penelitian digunakan untuk menguraikan dan menyelesaikan masalah dalam penelitian. Rancangan penelitian dapat dilihat pada Flowchart berikut ini [11], [12]:



Gambar 1. Rancangan Penelitian

Pada Gambar diatas menjelaskan rancangan penelitian yang dilakukan untuk mengklasifikasi tingkat pemahaman siswa pada mata pelajaran matematika dengan menggunakan Algoritma C4.5 yang terdiri dari [13], [14]:

- a) **Analisa Masalah**
Analisa masalah merupakan langkah untuk memahami masalah yang telah di tentukan pada ruang lingkup ataupun batasannya. Dengan melakukan analisis suatu masalah yang di tentukan, diharapkan agar dapat melakukan penelitian dengan lebih baik. Menganalisis masalah terkait pemahaman siswa pada mata pelajaran matematika.
- b) **Mempelajari Literatur**
Penelitian ini harus didasari rujukan yang digunakan untuk mendapatkan informasi dan teori yang akan digunakan dalam penelitian.
- c) **Pengumpulan Data**
Data dikumpulkan dengan menggunakan kuesioner yang diisi oleh siswa SMA Swasta Tamansiswa Tapian Dolok.
- d) **Menetapkan Metode**
Pada penelitian ini penulis menggunakan teknik data mining klasifikasi dengan Algoritma C4.5 untuk mencari solusi permasalahan penelitian.
- e) **Mengolah Data di Excel**
Mengelola data dengan Data Transformation Servis (DTS) yaitu dari bentuk kuesioner dipindahkan kedalam bentuk Microsoft Excel (.xls).
- f) **Pengujian RapidMiner**
Melakukan pengujian data dengan menggunakan aplikasi RapidMiner versi 5.3 dengan Algoritma C4.5.
- g) **Pohon Keputusan**
Rule dan Pohon keputusan adalah hasil dari informasi pengujian data kuesioner di RapidMiner menggunakan Algoritma C4.5.
- h) **Kesimpulan**
Kesimpulan yang diperoleh dari klasifikasi tingkat pemahaman siswa terhadap mata pelajaran matematika memberikan keputusan yang dapat dijadikan sebagai rekomendasi dalam meningkatkan kualitas pembelajaran dan prestasi siswa dibidang akademik.

2.1.2. Prosedur Pengumpulan Data

Teknik pengumpulan data merupakan langkah paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data dan menghasilkan informasi baru. Prosedur pengumpulan data yang digunakan penulis dalam penelitian ini yaitu [15]:

- a) **Penelitian Kepustakaan (Library Research)** yaitu memanfaatkan perpustakaan sebagai sarana dalam mengumpulkan data dengan mempelajari karya-karya ilmiah yang berkaitan dengan metode data mining klasifikasi C4.5 seperti jurnal, artikel, buku, serta sumber ilmiah lain untuk dijadikan bahan referensi.
- b) **Penelitian Lapangan (Field Work Research)** yaitu penelitian yang dilakukan secara langsung dilapangan dengan menggunakan beberapa teknik dan cara sebagai berikut :
 - 1) **Kuesioner**
Penulis memberikan kuesioner kepada siswa SMA Swasta Tamansiswa Tapian Dolok mengenai pertanyaan yang berhubungan dengan pemahaman siswa pada mata pelajaran matematika.
 - 2) **Studi Literatur**
Penulis memperoleh informasi dengan mengumpulkan data, mempelajari data, dan mencari referensi terkait dengan kasus penelitian.

2.2. Analisis Data

Pada tahap ini menjelaskan suatu proses atau upaya pengolahan data menjadi sebuah informasi baru agar karakteristik data menjadi lebih mudah dimengerti dan berguna untuk solusi suatu permasalahan, serta dijelaskan dengan beberapa langkah atau bentuk alur agar mudah di pahami. Dalam penelitian ini data yang digunakan akan diolah dari hasil kuesioner, yang diberikan kepada siswa SMA Swasta Tamansiswa Tapian Dolok. Dalam penelitian ini digunakan perhitungan sampel menurut Rumus Slovin. Menurut Sugiyono (dalam Nurhadi, Zikri Fachrul & Mujiyanto, 2020) 'Rumus Slovin adalah sebuah rumus untuk mendapatkan sampel yang lebih pasti atau mendekati populasi yang ada. Teknik pengambilan sampel tidak memberikan peluang atau kesempatan yang sama bagi setiap anggota populasi untuk dipilih menjadi sampel'.

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Keterangan :

- n : Ukuran sampel
N : Ukuran Populasi
e : Margin of error

Berdasarkan Rumus Slovin, maka data siswa SMA Swasta Tamansiswa Tapian Dolok adalah 437 orang dalam sebuah populasi, kita bisa tentukan minimal sampel yang akan diteliti. Margin of error yang ditetapkan adalah 5% atau 0,05.

$$n = N / (1 + (N \times e^2)) \quad (2)$$

$$n = 437 / ((1 + (437 \times [0,05]^2)))$$

$$n = 437 / ((1 + (437 \times 0,0025)))$$

$$n = 437 / (1 + 1,0925)$$

$$n = 437 / 2,0925$$

$$n = 208$$

Dari hasil kuesioner yang dilakukan penulis mengambil sampel data sebanyak 208 responden dari siswa SMA Swasta Tamansiswa Tapan Dolok. Pada penelitian ini digunakan 6 atribut dalam melakukan klasifikasi terhadap tingkat pemahaman siswa pada mata pelajaran matematika yaitu :

- Minat Belajar Siswa (C1), merupakan keinginan atau ketertarikan dalam diri siswa terhadap mata pelajaran matematika.
- Komunikasi (C2), merupakan faktor antara siswa dan guru sehingga mempengaruhi siswa memahami mata pelajaran matematika.
- Cara Belajar Siswa (C3), merupakan tindakan siswa yang dilakukan dalam proses pembelajaran agar memperoleh pemahaman pada mata pelajaran matematika .
- Suasana Pembelajaran (C4), merupakan suasana kelas yang dapat memudahkan siswa memahami materi yang diberikan guru.
- Media Pembelajaran (C5), merupakan alat bantu yang digunakan guru untuk menyampaikan materi agar materi yang disampaikan lebih mudah dipahami.
- Cara Mengajar Guru (C6), merupakan proses penyampaian materi yang diberikan oleh guru yang berpengaruh pada pemahaman siswa.

Kuesioner yang diperoleh selanjutnya dilakukan pencarian rata-rata dari setiap pertanyaan yang mewakili atribut yang digunakan. Data kuesioner yang digunakan menggunakan jenis statistik deskriptif dengan kuesioner yang diberikan menggunakan skala linker 4 yang terdiri dari Sangat Paham, Paham, Cukup Paham, dan Tidak Paham. Data hasil dari kuesioner selanjutnya diolah menggunakan RapidMiner yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data. Data yang akurat dilakukan pengolahan data untuk mencari hasil dari masalah penelitian dengan menggunakan RapidMiner versi 5.3 dan mengambil keputusan dari hasil klasifikasi dengan Algoritma C4.5. Berikut ini sampel data hasil rekapitulasi kuesioner yang diolah di Microsoft Excel pada Tabel berikut:

Tabel 1. Data Hasil Rekapitulasi Kuesioner Penelitian

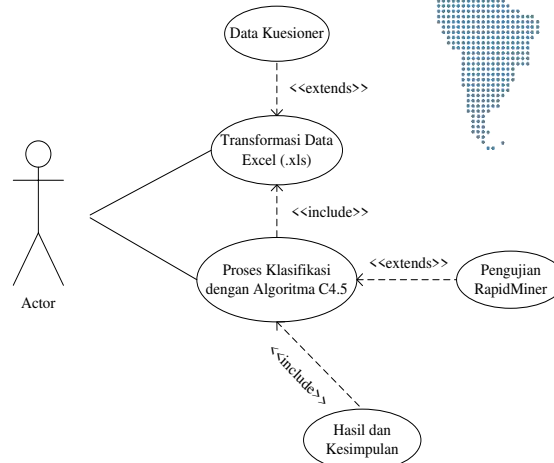
Responden	Minat Belajar Siswa	Komunikasi	Cara Belajar Siswa	Suasana Pembelajaran	Media Pembelajaran	Cara Mengajar Guru	Hasil
R1	Paham	Paham	Cukup Paham	Cukup Paham	Sangat Paham	Paham	Paham
R2	Tidak Paham	Cukup Paham	Cukup Paham	Paham	Cukup Paham	Tidak Paham	Tidak Paham
R3	Cukup Paham	Cukup Paham	Tidak Paham	Cukup Paham	Cukup Paham	Paham	Tidak Paham
R4	Paham	Cukup Paham	Paham	Paham	Cukup Paham	Paham	Paham
R5	Paham	Cukup Paham	Paham	Paham	Cukup Paham	Cukup Paham	Paham
R6	Sangat Paham	Sangat Paham	Paham	Cukup Paham	Sangat Paham	Cukup Paham	Paham
R7	Paham	Cukup Paham	Cukup Paham	Paham	Paham	Cukup Paham	Paham
R8	Cukup Paham	Cukup Paham	Cukup Paham	Paham	Paham	Paham	Paham
R9	Tidak Paham	Cukup Paham	Tidak Paham	Paham	Cukup Paham	Cukup Paham	Paham
R10	Cukup Paham	Cukup Paham	Tidak Paham	Cukup Paham	Paham	Tidak Paham	Tidak Paham
R11	Tidak Paham	Cukup Paham	Cukup Paham	Paham	Cukup Paham	Cukup Paham	Tidak Paham
R12	Sangat Paham	Paham	Cukup Paham	Paham	Cukup Paham	Cukup Paham	Paham
R13	Cukup Paham	Cukup Paham	Cukup Paham	Paham	Sangat Paham	Cukup Paham	Paham
R14	Cukup Paham	Paham	Paham	Paham	Cukup Paham	Tidak Paham	Tidak Paham
R15	Tidak Paham	Cukup Paham	Cukup Paham	Paham	Cukup Paham	Cukup Paham	Tidak Paham
R16	Sangat Paham	Paham	Cukup Paham	Paham	Cukup Paham	Cukup Paham	Paham
R17	Paham	Sangat Paham	Sangat Paham	Sangat Paham	Sangat Paham	Cukup Paham	Paham
R18	Cukup Paham	Cukup Paham	Tidak Paham	Paham	Paham	Cukup Paham	Tidak Paham
R19	Cukup Paham	Cukup Paham	Cukup Paham	Paham	Paham	Cukup Paham	Paham
R20	Paham	Paham	Cukup Paham	Paham	Cukup Paham	Cukup Paham	Paham
...
R208	Paham	Sangat Paham	Paham	Paham	Sangat Paham	Cukup Paham	Paham

2.2.1. Alat Analisis Data

Pada penelitian ini penulis menggunakan aplikasi Microsoft Excel 2010 dan RapidMiner versi 5.3 sebagai alat untuk menganalisis data. Microsoft Excel digunakan untuk membantu dalam proses pengolahan data dan melakukan proses perhitungan dengan algoritma C4.5. Penulis menggunakan aplikasi RapidMiner untuk melakukan uji validitas untuk melihat keakuratan hasil yang diperoleh dari Microsoft Excel.

2.2.2. Instrumen Penelitian

Pada penelitian ini, penulis melakukan transformasi data kuesioner dalam beberapa dataset berbentuk spreadsheet file excel 2010. Transformasi data ini diperlukan sebagai input untuk pengujian pada Rapidminer yang dilakukan dalam penelitian ini. Instrumen penelitian dapat dilihat pada Gambar berikut.

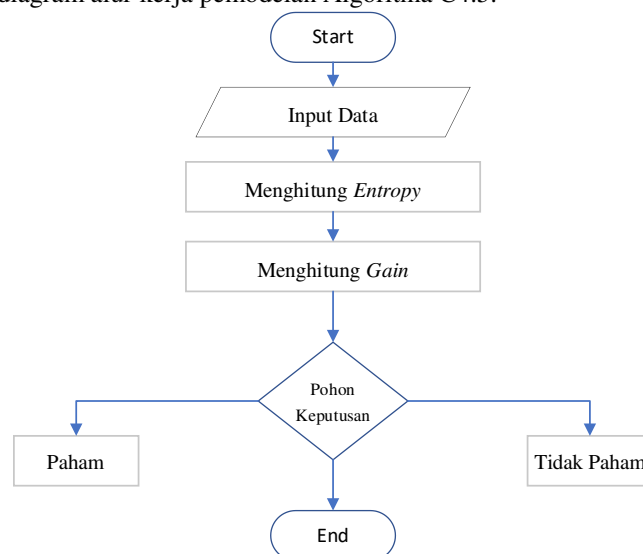


Gambar 2. Instrumen Penelitian

2.2.3. Pemodelan Metode Moora

Dalam penelitian ini penulis menggunakan teknik klasifikasi data mining dengan Algoritma C4.5 dan melakukan pengujian di RapidMiner. Berikut ini pemodelan yang penulis gunakan dalam penelitian ini.

Metode yang digunakan pada penelitian ini adalah algoritma C4.5. Dalam pemodelan algoritma C4.5 dilakukan pada sampel dataset, kemudian menghitung entropy (S) dari keseluruhan atribut, setelah entropy (S) ditemukan kemudian perhitungan Gain tertinggi dari seluruh atribut, atribut yang memiliki Gain tertinggi yang akan digunakan sebagai akar/node. Selanjutnya buat cabang untuk tiap nilai, bagi kasus dalam cabang, ulangi perhitungan Gain sampai semua data masuk kedalam kelas yang sama. Atribut yang telah dipilih tidak lagi dibutuhkan dalam perhitungan, proses pembentukan pohon keputusan terhenti jika sudah tidak ada lagi atribut yang dipartisi dan semua tuple dalam node N telah memiliki kelas yang sama. Dalam kasus dataset pada penelitian ini terdiri dari 2 kelas keterangan yaitu paham dan tidak paham. Berikut adalah diagram alur kerja pemodelan Algoritma C4.5.



Gambar 3. Flowchart Metode C4.5

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Hasil akhir dari penerapan Algoritma C4.5 dibagi ke dalam dua tahap yaitu proses perhitungan manual menggunakan Microsoft Excel dan penyesuaian hasil perhitungan melalui pengujian data menggunakan software RapidMiner 5.3. Data yang digunakan adalah data hasil kuesioner sebagai input untuk membuat model aturan menggunakan Algoritma C4.5. Pohon keputusan digunakan untuk membuat model aturan yang akan dipilih dalam pengambilan keputusan. Berikut ini tahapan pengolahan data dengan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan tingkat pemahaman siswa pada mata pelajaran matematika sesuai data hasil kuesioner yang diperoleh.

a) Proses Perhitungan Algoritma C4.5

Perhitungan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan dapat diuraikan sebagai berikut:

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan Paham, jumlah kasus untuk keputusan Tidak Paham.

Langkah 2: Menghitung *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan kelas atribut. Selanjutnya dilakukan perhitungan *Gain* untuk masing-masing atribut. Berikut ini adalah perhitungan nilai *entropy* dan *gain*.

Menghitung *entropy* total :

$$\begin{aligned} \text{Entropy [Total]} &= \left(-\frac{123}{208} \times \log_2 \left(\frac{123}{208} \right) \right) + \left(-\frac{85}{208} \times \log_2 \left(\frac{85}{208} \right) \right) \\ \text{Entropy [Total]} &= 0,97578821 \end{aligned}$$

Menghitung *entropy* dan *gain* Minat Belajar Siswa :

$$\text{Entropy [Minat Belajar Siswa – Sangat Paham]} = 0$$

$$\text{Entropy [Minat Belajar Siswa – Paham]} = 0$$

$$\text{Entropy [Minat Belajar Siswa – Cukup Paham]} =$$

$$\left(-\frac{58}{84} \times \log_2 \left(\frac{58}{84} \right) \right) + \left(-\frac{26}{84} \times \log_2 \left(\frac{26}{84} \right) \right) = 0,892623013$$

$$\text{Entropy [Minat Belajar Siswa – Tidak Paham]} =$$

$$\left(-\frac{7}{66} \times \log_2 \left(\frac{7}{66} \right) \right) + \left(-\frac{59}{66} \times \log_2 \left(\frac{59}{66} \right) \right) = 0,487917993$$

$$\text{Gain [Total, Minat Belajar Siswa]} =$$

$$0,97578821 - \left(\left(\frac{6}{208} \times 0 \right) + \left(\frac{52}{208} \times 0 \right) + \left(\frac{84}{208} \times 0,892623013 \right) + \left(\frac{66}{208} \times 0,487917993 \right) \right) = 0,460485707$$

..... Hingga Perhitungan *entropy* dan *gain* Cara Men :

Berikut ini hasil perhitungan nilai *entropy* dan *gain* yang diuraikan pada tabel berikut.

Tabel 2. Perhitungan Node 1

Node 1	Jumlah (S)	Paham	Tidak Paham	Entropy	Information Gain
Total	208	123	85	0,97578821	
Minat Belajar Siswa					0,460485707
Sangat Paham	6	6	0	0	
Paham	52	52	0	0	
Cukup Paham	84	58	26	0,89262301	
Tidak Paham	66	7	59	0,48791799	
Komunikasi					0,125733777
Sangat Paham	18	18	0	0	
Paham	67	49	18	0,8395305	
Cukup Paham	103	50	53	0,99938797	
Tidak Paham	20	6	14	0,8812909	
Cara Belajar Siswa					0,156520273
Sangat Paham	3	3	0	0	
Paham	78	55	23	0,87493159	
Cukup Paham	79	56	23	0,87018834	
Tidak Paham	48	9	39	0,69621226	
Suasana Pembelajaran					0,170242097
Sangat Paham	22	22	0	0	
Paham	107	71	36	0,92137965	
Cukup Paham	70	30	40	0,98522814	
Tidak Paham	9	0	9	0	
Media Pembelajaran					0,142297315
Sangat Paham	24	23	1	0,24988229	
Paham	94	54	40	0,9839394	
Cukup Paham	77	46	31	0,97244972	
Tidak Paham	13	0	13	0	
Cara Mengajar Guru					0,218903536
Sangat Paham	11	8	3	0,84535094	
Paham	77	57	20	0,82635417	
Cukup Paham	90	58	32	0,93893201	

Langkah 3: Dari hasil Perhitungan pada Tabel 2. diperoleh nilai atribut tertinggi adalah Minat Belajar Siswa dengan gain sebesar 0,460485707. Maka atribut Minat Belajar Siswa dipilih sebagai node akar. Nilai kelas atribut Sangat Paham dan Paham mengklasifikasikan kasus menjadi satu keputusan yaitu Paham. Untuk kelas atribut Cukup Paham dan Tidak Paham belum diperoleh hasil antara keputusan Paham atau Tidak Paham, maka perlu dilakukan perhitungan lebih lanjut. Berikut ini hasil perhitungan dari kelas atribut Minat Belajar Siswa = Cukup Paham yang ditunjukkan pada Tabel berikut ini:

Tabel 3. Perhitungan Node 1.1

Node 1.1	Jumlah (S)	Paham	Tidak Paham	Entropy	Information Gain
Minat Belajar Siswa-Cukup Paham	84	58	26	0,89262361	0,074590243
Komunikasi					
Sangat Paham	7	7	0	0	
Paham	23	18	5	0,75537541	
Cukup Paham	53	32	21	0,96870034	
Tidak Paham	1	1	0	0	0,518431849
Cara Belajar Siswa					
Sangat Paham	0	0	0	0	
Paham	35	24	11	0,89805879	
Cukup Paham	34	34	0	0	
Tidak Paham	15	0	15	0	0,140980148
Suasana Pembelajaran					
Sangat Paham	13	13	0	0	
Paham	36	26	10	0,85240518	
Cukup Paham	33	19	14	0,98337619	
Tidak Paham	2	0	2	0	0,081751399
Media Pembelajaran					
Sangat Paham	13	12	1	0,39124356	
Paham	34	24	10	0,87398105	
Cukup Paham	35	22	13	0,95176268	
Tidak Paham	2	0	2	0	0,29243704
Cara Mengajar Guru					
Sangat Paham	6	4	2	0,91829583	
Paham	31	28	3	0,45868582	
Cukup Paham	36	26	10	0,85240518	
Tidak Paham	11	0	11	0	

Perhitungan dilakukan hingga hasil keputusan akhir yang terbentuk pada node terakhir dengan atribut Komunikasi sebagai node cabang. Dari hasil tersebut diketahui semua kasus sudah masuk kedalam kelas. Dari perhitungan node terkahir nnti akan didapat terdapat 18 (delapan belas) rules yang dapat dijadikan sebagai referensi dalam menentukan tingkat pemahaman siswa pada mata pelajaran matematika. Adapun aturan atau rule yang terbentuk berdasarkan pohon keputusan pada Gambar 4.9 diatas yaitu 9 (sembilan) rules keputusan Paham dan 9 (sembilan) rules keputusan Tidak Paham dijelaskan melalui teks narasi sebagai berikut:

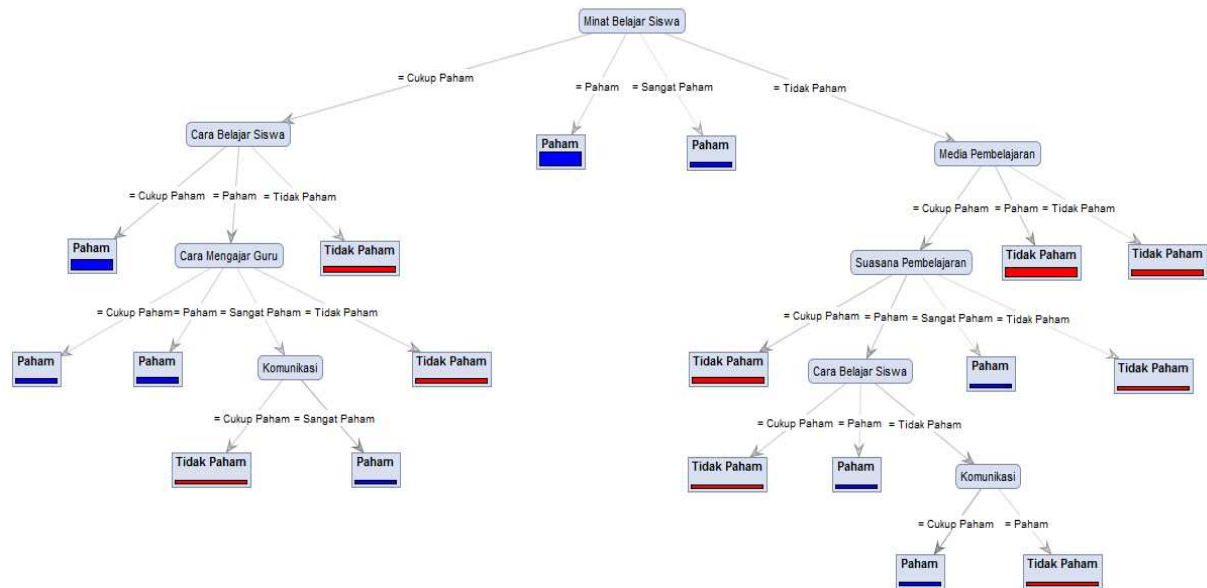
- Jika Minat Belajar Siswa = Cukup Paham dan Cara Belajar Siswa = Cukup Paham, maka hasilnya Paham {Paham = 34, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Cukup Paham, Cara Belajar Siswa = Paham dan Cara Mengajar Guru = Cukup Paham, maka hasilnya Paham {Paham = 8, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Cukup Paham, Cara Belajar Siswa = Paham dan Cara Mengajar Guru = Paham, maka hasilnya Paham {Paham = 12, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Cukup Paham, Cara Belajar Siswa = Paham, Cara Mengajar Guru = Sangat Paham dan Komunikasi = Cukup Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 2}
- Jika Minat Belajar Siswa = Cukup Paham, Cara Belajar Siswa = Paham, Cara Mengajar Guru = Sangat Paham dan Komunikasi = Sangat Paham, maka hasilnya Paham {Paham = 4, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Cukup Paham, Cara Belajar Siswa = Paham dan Cara Mengajar Guru = Tidak Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 9}
- Jika Minat Belajar Siswa = Cukup Paham dan Cara Belajar Siswa = Tidak Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 15}
- Jika Minat Belajar Siswa = Paham, maka hasilnya Paham {Paham = 52, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Sangat Paham, maka hasilnya Paham {Paham = 6, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham dan Suasana Pembelajaran = Cukup Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 11}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham, Suasana Pembelajaran = Paham dan Cara Belajar Siswa = Cukup Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 4}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham, Suasana Pembelajaran = Paham dan Cara Belajar Siswa = Paham, maka hasilnya Paham {Paham = 2, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham, Suasana Pembelajaran = Paham, Cara Belajar Siswa = Tidak Paham dan Komunikasi = Cukup Paham, maka hasilnya Paham {Paham = 4, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham, Suasana Pembelajaran = Paham, Cara Belajar Siswa = Tidak Paham dan Komunikasi = Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 1}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham dan Suasana Pembelajaran = Sangat Paham, maka hasilnya Paham {Paham = 1, Tidak Paham = 0}
- Jika Minat Belajar Siswa = Tidak Paham, Media Pembelajaran = Cukup Paham dan Suasana Pembelajaran = Tidak Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 2}

- q) Jika Minat Belajar Siswa = Tidak Paham dan Media Pembelajaran = Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 30}
- r) Jika Minat Belajar Siswa = Tidak Paham dan Media Pembelajaran = Tidak Paham, maka hasilnya Tidak Paham {Paham = 0, Tidak Paham = 11}

b) Proses Pengujian dengan RapidMiner

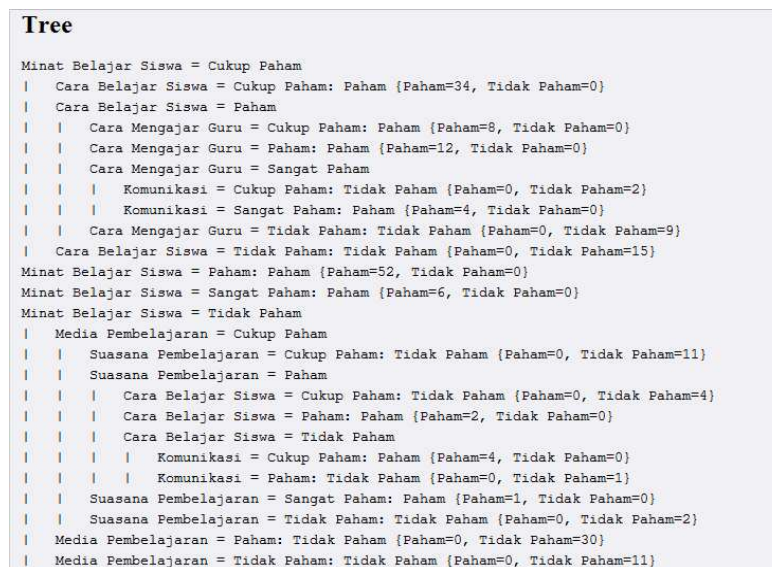
Pada tahap akhir penerapan Algoritma C4.5 dilakukan penyesuaian hasil perhitungan manual melalui pengujian menggunakan software RapidMiner 5.3. Pengujian terhadap hasil perhitungan manual menggunakan software RapidMiner dilakukan melalui beberapa tahapan proses seperti berikut ini:

Hasil pengolahan data dengan model pohon keputusan sesuai dengan software RapidMiner, dapat dilihat pada Gambar 4. sebagai berikut :



Gambar 4. Decision Tree Pada Rapidminer

Gambar diatas merupakan pohon keputusan yang dihasilkan pada *Rapidminer* dengan aturan atau *rule* yang dapat dilihat pada text view pada Gambar berikut:



Gambar 5. Rule Decision Tree Pada Rapidminer

3.2. Pembahasan

Hasil penerapan Algoritma C4.5 menggunakan software *RapidMiner* dengan operator *Split Validation* diperoleh nilai akurasi yaitu sebesar 95.19%. Hasil akurasi tersebut diperoleh dengan pengaturan pada operator *split validation* dengan nilai *split ratio* = 0,5 dan *sampling type* = *linear sampling*. Berikut ini adalah hasil akurasi yang diperoleh.

accuracy: 95.19%			
	true Paham	true Tidak Paham	class precision
pred. Paham	61	1	98.39%
pred. Tidak Paham	4	38	90.48%
class recall	93.85%	97.44%	

Gambar 6. Nilai Akurasi Algoritma C4.5

PerformanceVector

```

PerformanceVector:
accuracy: 95.19%
ConfusionMatrix:
True:   Paham   Tidak Paham
Paham:  61      1
Tidak Paham:  4      38

```

Gambar 7. Nilai Performance Vector Algoritma C4.5

Dari gambar diatas *Accuracy* sebesar 95.19%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana untuk *Class Precision* pada prediksi label Paham sebesar 98.39% dan prediksi label Tidak Paham sebesar 90.48%. Hasil yang dilakukan peneliti dalam perhitungan Algoritma C4.5 diperoleh 18 model aturan atau rule tingkat pemahaman siswa pada mata pelajaran matematika dengan 9 rules berstatus Paham dan 9 rules berstatus Tidak Paham. Model aturan dalam bentuk pohon keputusan yang diperoleh peneliti dapat dilihat pada Tabel. diatas. Berdasarkan pengolahan data menggunakan aplikasi RapidMiner didapat nilai akurasi sebesar 95.19%, artinya bahwa rule yang dihasilkan tingkat kebenarannya mendekati 100%. Dimana untuk Class Precision pada prediksi label Paham sebesar 98.39% dan prediksi label Tidak Paham sebesar 90.48%. Parameters yang digunakan pada decision tree telah dilakukan penyesuaian terhadap kriteria decision tree yang digunakan pada RapidMiner yaitu information gain, maksimal depth= 20, confidance= 0,25, minimal gain= 0,01, minimal leaf size= 1, minimal size for split= 2, number of pruning alt= 3. Sesuai dengan ketentuan tersebut maka hasil perhitungan manual dengan pengujian RapidMiner menghasilkan sebanyak 18 model aturan atau rule tingkat pemahaman siswa pada mata pelajaran matematika. Model aturan dalam bentuk pohon keputusan yang dihasilkan dari Rapidminer dapat dilihat pada Gambar diatas.

Hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan Rapidminer diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan RapidMiner dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus tingkat pemahaman siswa pada mata pelajaran matematika.

4. KESIMPULAN

Berdasarkan hasil penelitian yang diperoleh dapat disimpulkan bahwa , penerapan Algoritma C4.5 dapat diterapkan untuk mengklasifikasi tingkat pemahaman mata pelajaran matematika siswa SMA Swasta Tamansiswa Tapian Dolok dan telah menghasilkan 18 (delapan belas) model aturan atau rule keputusan yaitu 9 (sembilan) keputusan Paham dan 9 (sembilan) keputusan Tidak Paham. Hasil penerapan Algoritma C4.5 dapat diuji dengan aplikasi Rapidminer dan diperoleh hasil yang sama dengan perhitungan Algoritma C4.5 dimana diperoleh tingkat akurasi yaitu 95.19%.

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. Prasetyo, "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] 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.
- [7] P. Alkhairi, P. P. P. A. N. . F. I. R.H.Zer, E. R. Batubara, F. N. Tambunan, And R. Rosnelly, "Pengenalan Pola Kemampuan Pelanggan Dalam Membayar Air Pdam Menggunakan Algoritma Naive Bayes," *Jurnaltimes*, Vol. X, No. 2, Pp. 29–38,

- 2022.
- [8] Hendrianto, "Manajemen Strategi Pengelolaan Pasar Dalam Meningkatkan Pendapatan Pedagang Perspektif Ekonomi Islam (Studi Di Pasar Segamas Purbalingga)," Pp. 1–94, 2018.
- [9] I. Mulia And M. Muanas, "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.
- [10] F. R. S. U. P. Eka Irawan, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes," Vol. 2, No. 1, Pp. 39–46.
- [11] A. D. I. Suradi, "Penerapan Data Mining Untuk Menentukan Rekomendasi Beasiswa Dengan Metode Algoritma C4.5," 2018.
- [12] 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.
- [13] M. N. P. Pamulang, M. N. Aini, And U. Enri3, "Komparasi Distance Measure Pada K-Medoids Clustering Untuk Pengelompokan Penyakit Ispa," *Edumatic J. Pendidik. Inform.*, Vol. 5, No. 1, Pp. 99–107, 2021, Doi: 10.29408/Edumatic.V5i1.3359.
- [14] F. Fajriana, "Analisis Algoritma K-Medoids Pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara," *J. Edukasi Dan Penelit. Inform.*, Vol. 7, No. 2, P. 263, 2021, Doi: 10.26418/Jp.V7i2.47795.
- [15] A. Aditya, B. N. Sari, And T. N. Padilah, "Comparison Analysis Of Euclidean And Gower Distance Measures On K-Medoids Cluster," *J. Teknol. Dan Sist. Komput.*, Vol. 9, No. 1, Pp. 1–7, 2021, Doi: 10.14710/Jtsiskom.2020.13747.