
**Pengimplementasian Tingkat Ketepatan Waktu Kelulusan Siswa (Studi Kasus Di
MTS Nur Ibarhimy) Menggunakan Algoritma C4.5**

Rizky Amansyah¹, Masrizal², Ibnu Rasyid Munthe³

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu^{1,2,3}

Email : rizkiamansyah22@gmail.com¹, masrizal120405@gmail.com²,
ibnurasyidmunthe@gmail.com³

Corresponding Author: rizkiamansyah22@gmail.com

Abstract

Education has a very important role in shaping the individual and directing the development of society. As an educational institution, MTS Nur Ibrahimi has a responsibility to improve the quality and efficiency in the implementation of Education. MTS Nur Ibrahimi is located in Rantauprapat, Rantau Selatan district, Labuhanbatu Regency. MTs Nur Ibrahimi has been established since 2000 and has produced a number of students who successfully completed their education at this school. Along with technological advances, pattern exploration can be done by using data classification techniques obtained through the data mining process. Data mining is generally done because of the large amount of data, which can be used to generate patterns and useful knowledge in the business operations of a company. One of the methods developed in data mining is a way to dig up existing data to build a model, and then use the model to recognize other data patterns that are not contained in the stored database. In this context, a classification model is created to identify data patterns related to "Passed" or "not passed" status classes, based on pattern Determination results from training data. The Decision Trees Model is an implementation of the classification model in data mining. This Model builds a decision tree from training data consisting of records in a database. The C4.5 algorithm is one of the data classification algorithms that uses decision tree techniques and is able to manage numerical (continuous) and discrete data, and can handle missing attribute values. This algorithm produces rules that are easy to interpret. C4.5 has been tested in various classification cases, including in medical, trade, personnel, and various other fields.

Keywords: Data Mining, Graduation Time, Decision Trees, C4.5 Algorithm.

I. Pendahuluan

Pendidikan memiliki peran yang sangat penting dalam membentuk individu dan mengarahkan perkembangan masyarakat. Pada tingkat pendidikan menengah, tingkat kelulusan

siswa mencerminkan mutu proses pembelajaran serta efektivitas penyelenggaraan pendidikan. Sebagai sebuah lembaga pendidikan, MTS Nur Ibrahimi memiliki tanggung jawab untuk meningkatkan kualitas dan

efisiensi dalam pelaksanaan pendidikannya. MTS Nur Ibrahimy Terletak di Rantauprapat Kecamatan Rantau Selatan, Kabupaten Labuhanbatu. MTs Nur Ibrahimy telah berdiri sejak tahun 2000 dan telah menghasilkan sejumlah siswa yang berhasil menyelesaikan pendidikan di sekolah ini. Seiring dengan kemajuan teknologi, eksplorasi pola dapat dilakukan dengan menggunakan teknik klasifikasi data yang diperoleh melalui proses data mining. Data mining umumnya dilakukan karena adanya jumlah data yang besar, yang dapat dimanfaatkan untuk menghasilkan pola dan pengetahuan yang bermanfaat dalam operasional bisnis suatu perusahaan. Salah satu metode yang dikembangkan dalam data mining adalah cara untuk menggali data yang ada guna membangun model, lalu menggunakan model tersebut untuk mengenali pola data lain yang tidak terdapat dalam basis data yang tersimpan. Dalam konteks ini, model klasifikasi dibuat untuk mengidentifikasi pola data terkait kelas status "lulus" atau "tidak lulus", berdasarkan hasil penentuan pola dari data pelatihan. Pola atau model dari data pelatihan tersebut selanjutnya diuji menggunakan set data uji. Secara sederhana, model klasifikasi dibentuk dengan menganalisis data pelatihan yang mencakup variabel-variabel yang kelasnya sudah diketahui. Model yang dihasilkan kemudian digunakan untuk memprediksi kelas dari data yang tidak diketahui (variabel-variabel yang kelasnya belum diketahui), dan set data uji digunakan untuk menguji model yang telah dibuat pada data pelatihan.

Algoritma klasifikasi yang digunakan dalam hal ini adalah algoritma C4.5, yang berfungsi untuk memprediksi atau mengklasifikasikan suatu kejadian dengan pembentukan pohon keputusan. Model *Decision Trees* adalah salah satu implementasi dari model klasifikasi dalam data mining. Model ini membangun pohon keputusan dari data pelatihan yang terdiri dari catatan-catatan dalam basis data. Algoritma C4.5 adalah salah satu algoritma klasifikasi data yang menggunakan teknik pohon keputusan dan mampu mengelola data numerik (kontinu) dan diskrit, serta dapat menangani nilai atribut yang hilang. Algoritma ini menghasilkan aturan-aturan yang mudah untuk diinterpretasikan. C4.5 telah diujicoba dalam berbagai kasus klasifikasi, termasuk dalam bidang medis, perdagangan, kepegawaian, dan berbagai bidang lainnya.

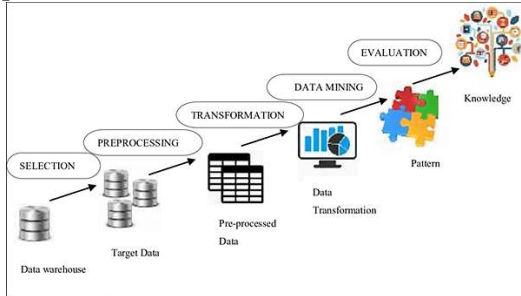
II. Landasan Teori

Knowledge discovery in Database

Knowledge discovery in Database (KDD) diartikan sebagai pengambilan informasi yang berpotensi, tersirat, dan belum diketahui dari sekelompok data. KDD merupakan suatu proses untuk menggali pola, informasi, dan pengetahuan bermanfaat dari sejumlah besar data.besar [1]. Proses ini mencakup pemanfaatan teknik-teknik data mining dan analisis data guna mengenali pola tersembunyi yang dapat memberikan wawasan atau mendukung pengambilan keputusan. *Knowledge discovery in Database* (KDD) adalah rangkaian langkah yang tidak sederhana untuk menemukan dan mengidentifikasi pola dalam data, di

mana pola yang ditemukan harus valid, baru, bermanfaat, dan dapat dimengerti.

Proses *Knowledge discovery in Database* secara detail dapat dilihat pada Gambar 1 di bawah ini



Gambar 1. Proses Knowledge In Database

Sumber:

<https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>

Data Mining

Data mining merupakan analisis data untuk menemukan hubungan yang jelas dan menyimpulkan informasi yang sebelumnya tidak diketahui, dengan menggunakan metode terkini agar dapat dipahami dan dimanfaatkan dengan mudah. Proses data mining melibatkan ekstraksi pola dan pengetahuan yang bermanfaat dari kumpulan data besar.

Decision Tree

Decision Tree adalah suatu metode klasifikasi dan prediksi yang sangat efektif dan terkenal. Pendekatan *Decision Tree* mengubah data yang besar menjadi suatu struktur pohon keputusan yang menggambarkan aturan. Aturan-aturan tersebut dapat dengan mudah dimengerti dalam bahasa alami dan juga dapat diungkapkan dalam format basis data seperti *Structure Query Language* (SQL) untuk mencari rekaman pada data tertentu.

Algoritma C4.5

ID3 (*Iterative Dichotomiser*) merupakan awal dari perkembangan algoritma *Decision Tree* dan telah membentuk dasar untuknya. Quinlan kemudian mengembangkan algoritma ID3 menjadi C4.5. C4.5 menjadi algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan populer karena memiliki keunggulan tertentu.

Entropy

Istilah *Entropy* adalah keberbedaan atau keragaman. Dalam data mining, *Entropy* didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data [6]. Semakin heterogen himpunan suatu data, semakin besar pula nilai *Entropy*-nya. Secara matematis *Entropy* dirumuskan sebagai berikut :

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

Keterangan:

- S = Himpunan Kasus
- n = Jumlah partisi S
- Pi = Proporsi dari Si terhadap S

Dimana n adalah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan Pi menyatakan porsi atau rasio antara jumlah sampel di kelas i dengan jumlah sampel pada himpunan data. Dalam konteks pohon keputusan pada data mining, entropy sering digunakan sebagai metrik untuk mengukur ketidakmurnian (*impurity*) sebuah himpunan data. Tujuan pembangunan pohon keputusan adalah mengurangi entropi, sehingga menghasilkan

pembagian kelas atau keputusan yang lebih homogen di setiap cabangnya.

Gain

Gain adalah suatu metrik yang digunakan dalam konteks pohon keputusan (decision tree) pada data mining untuk mengevaluasi kualitas pemisahan (split) pada suatu atribut. Algoritma C4.5 merupakan perbaikan dari ID3 menggunakan *Gain Ratio* untuk diperbaharui information gain [6]. Maka dengan rumus:

$$\text{Gain Ratio (S.A)} = (\text{Gain(S.A)} / (\text{Splitinfo(S.A)}))$$

Dimana :

S = Ruang/Data Sample yang digunakan untuk data training

A = Atribut

Gain (S.A) = information gain pada atribut A

Splitinfo(S.A) = split information pada atribut A

Dengan atribut yang nilai *Gain Ratio* yang paling tertinggi dipilih sebagai atribut test untuk simpul. Pendekatan ini menerapkan normalisasi pada information gain dengan menggunakan apa yang disebut dengan split information, dengan rumus [8]:

$$\text{Gain (S.A)} = \text{Entropoy (S)} = \sum_{i=1}^n \frac{|Si|}{|S|} = \text{Entropy (Si)}$$

Dimana :

S = Himpunan Kasus

A = Atribut.

n = Jumlah Partisi Atribut A

| Si | = Jumlah kasus pada partisi ke – i

| S | = Jumlah Kasus Dalam S

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atributatribut yang ada.

Rapidminer

RapidMiner adalah salah satu alat yang digunakan dalam data mining, menyediakan sekitar 500 operator data mining, termasuk operator untuk input, output, preprocessing data, dan visualisasi. RapidMiner berfungsi sebagai perangkat lunak analisis data dan mesin data mining yang dapat diintegrasikan ke dalam produknya sendiri. Ditulis dalam bahasa Java, RapidMiner dapat beroperasi di berbagai sistem operasi. Awalnya dikenal sebagai YALE (*Yet Another Learning Environment*), pengembangan pertama dimulai pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Unit Kecerdasan Buatan dari *University of Dortmund*.

III. Metode Penelitian

Metode penelitian merujuk pada cara atau langkah-langkah sistematis yang digunakan untuk mengumpulkan, menganalisis, dan menafsirkan data dalam suatu studi. Metode penelitian yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Studi Literatur (*Study Of Literature*)

Studi literatur, yang juga dikenal sebagai ulasan pustaka atau tinjauan literatur, merupakan suatu proses terorganisir yang bertujuan untuk mengevaluasi dan menganalisis informasi yang sudah tersedia dalam berbagai bentuk seperti tulisan, artikel ilmiah, buku, jurnal, dan sumber lain yang berkaitan dengan topik penelitian atau bidang tertentu.

b. Wawancara (*Interview*)

Wawancara merupakan metode pengumpulan data dalam

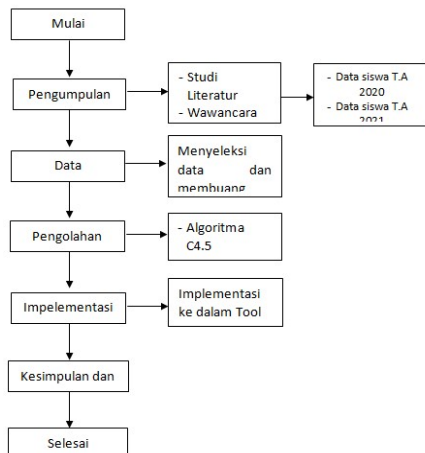
penelitian yang melibatkan komunikasi langsung antara peneliti dan responden. Fokus dari wawancara adalah untuk memperoleh informasi yang mendalam dan kontekstual mengenai pengalaman, pandangan, pengetahuan, atau sikap individu atau kelompok terkait dengan topik penelitian.

c. Observasi (*Observation*)

Observasi merupakan suatu teknik penelitian yang melibatkan pengumpulan data dengan cara mengamati dan mencatat secara sistematis perilaku, kejadian, atau fenomena yang terjadi dalam suatu konteks tertentu

Kerangka Kerja Penelitian

Langkah kinerja dalam sebuah penelitian atau tahapan penelitian sangat diperlukan untuk menentukan setiap proses yang akan dilakukan dalam perancangan sebuah sistem. Langkah kinerja yang digunakan dalam penelitian ini yaitu :



Gambar 2. Kerangka Kerja Penelitian

IV. Analisis dan Pembahasan Pengumpulan Data

Dataset yang digunakan adalah data yang diambil siswa MTS Nur Ibrahimy Rantauprapat dari Tahun Ajaran 2020-2022. Dataset keseluruhan yang digunakan adalah sebanyak 505 data siswa. Dataset dapat dilihat pada tabel berikut.

Tabel 1. Dataset Siswa MTS Nur Ibrahimy

NO	Ujian Sekolah	Ujian Nasional	Nilai Harian	Nilai Semester	Hasil
1	Tercapai	KKM	Atas	Baik	LULUS
2	Tercapai	KKM	Atas	Baik	LULUS
3	Tercapai	KKM	Atas	Baik	LULUS
4	Tercapai	KKM	Atas	Baik	LULUS
5	Tercapai	KKM	Atas	Baik	LULUS
6	Tercapai	KKM	Atas	Baik	LULUS
7	Tercapai	KKM	Atas	Cukup Baik	LULUS
8	Tercapai	KKM	Atas	Cukup Baik	LULUS
9	Tercapai	KKM	Atas	Cukup Baik	LULUS
10	Tercapai	KKM	Atas	Cukup Baik	LULUS
11	Tercapai	KKM	Atas	Cukup Baik	LULUS
12	Tercapai	KKM	Atas	Cukup Baik	LULUS
13	Tercapai	KKM	Atas	Cukup Baik	LULUS
14	Tercapai	KKM	Atas	Cukup Baik	LULUS
15	Tercapai	KKM	Atas	Cukup Baik	LULUS
16	Tercapai	KKM	Atas	Cukup Baik	LULUS
17	Tercapai	KKM	Atas	Cukup Baik	LULUS
18	Tercapai	KKM	Menengah	Cukup Baik	LULUS
19	Tercapai	KKM	Bawah	Cukup Baik	LULUS
20	Tercapai	KKM	Menengah	Cukup Baik	LULUS
21	Tercapai	KKM	Bawah	Cukup Baik	LULUS
22	Tercapai	KKM	Menengah	Cukup Baik	LULUS
23	Tercapai	KKM	Bawah	Cukup Baik	LULUS
24	Tercapai	KKM	Menengah	Cukup Baik	LULUS
25	Tercapai	KKM	Bawah	Cukup Baik	LULUS
26	Tercapai	KKM	Bawah	Cukup Baik	LULUS
27	Tercapai	KKM	Menengah	Cukup Baik	LULUS
28	Tercapai	KKM	Menengah	Cukup Baik	MENGULANG
29	Tercapai	KKM	Menengah	Cukup Baik	MENGULANG
30	Tercapai	KKM	Menengah	Baik	MENGULANG
31	Tercapai	KKM	Menengah	Baik	MENGULANG
32	Tercapai	KKM	Menengah	Baik	MENGULANG
33	Tercapai	KKM	Menengah	Baik	MENGULANG
34	Tercapai	KKM	Menengah	Baik	MENGULANG
35	Tercapai	KKM	Menengah	Baik	MENGULANG
36	Tercapai	KKM	Menengah	Baik	MENGULANG
....
505	Eltum tercapai	Dibawah kkm	Menengah	Baik	MENGULANG

Data Transformation

Pada tahapan ini data akan di transformasikan ke dalam bentuk yang lebih sederhana untuk diolah. Transformasi data dapat dilihat pada tabel 3.1 berikut

Tabel 3. Data Transformation

Atribut	Kategori	Keterangan
Jenis kelamin	Laki – laki	235
	Perempuan	270
Ujian Sekolah	Belum Tercapai	300
	Tercapai	205
Ujian Nasional	Dibawah KKM	276
	Memenuhi KKM	229
Nilai Harian	Bawah	349
	Menengah	119
	Atas	37

Pengolahan Data

Data dikumpulkan kemudian dikelola, pada tahap ini dilakukan perhitungan atribut-atribut. Tahap selanjutnya perhitungan Entrophy beserta gain untuk melakukannya dilihat dari jumlah data pada siswa lulus dan siswa tidak lulus. Setelah mendapatkan hasil langsung ke pohon keputusan pada setiap tabel. Data dibagi menjadi 2 bagian dengan perbandingan 80 : 20. Data training sebesar 80 % dengan jumlah data 404 dan data testing sebesar 20 % dengan jumlah data 101.

Algoritma C4.5

Dalam algoritma C4.5 kita akan menghitung terlebih dahulu nilai Entrophy. Setelah nilai Entrophy pada masing – masing kategori ditemukan, maka selanjutnya adalah menghitung nilai gain nya. Hasil dari perhitungan node 1 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 4.

Tabel 4. Perhitungan Node 1

node 1.1	jumlah (S)	lulus (S1)	tidak lulus (S2)	entrophy	gain
total	404	380	24	0.325082198	
Ujian Sekolah					-0.13523271
Tercapai	249	216	33	0.564336444	
Belum Tercapai	155	147	8	0.293209343	
Ujian Nasional					-0.1450856
KKM	256	226	30	0.521222018	
Dibawah KKM	148	137	11	0.381857804	
Nilai Harian					-0.14786176
Atas	156	141	15	0.456683632	
menengah	127	108	19	0.608854191	
bawah	121	113	8	0.351258091	
Nilai Semester					-0.14025099
Baik	170	156	14	0.410416142	
Cukup Baik	126	116	10	0.399938047	
Sangat Baik	108	91	17	0.628071016	

Rumus yang digunakan pada pencarian nilai Entrophy pada atribut yang dipakai adalah sebagai berikut:

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

Nilai entrophy total

$$= ((-380/404)*\text{IMLOG2}(380/404))+((-24/404)*\text{IMLOG2}(24/404))$$

$$= 0.325082198$$

Perhitungan yang sama dilakukan untuk mencari semua nilai entrophy. Setelah semua nilai Entrophy pada kategori sudah didapatkan, selanjutnya adalah menghitung nilai gain nya menggunakan rumus seperti berikut :

$$\text{Gain (S.A)} = Entrophy (S) = \sum_{i=1}^n \frac{|S_i|}{|S|} =$$

$$Entropy (S_i)$$

a. Nilai Gain ujian sekolah adalah

$$= (0.325082198)-$$

$$((249/404)*0.564336444)-$$

$$((155/404*0.293209343))$$

$$= -0.13523271$$

Dari hasil perhitungan nilai Entrophy dan nilai gain pada tabel perhitungan kelulusan maka dapat disimpulkan nilai hasil gain tertinggi yaitu gain Ujian Sekolah yang mendapatkan hasil -0.13523271 untuk menentukan nilai akarnya. Lakukan node sampai selesai, hingga hasil

perhitungan = 0 atau tidak dapat dihitung kembali.

Hasil dari perhitungan node 2 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 5.

Tabel 5. Perhitungan Node 1.1

Node 1.1	Tercapai	249	216	33	0.564336444	
Ujian Nasional						0.008116855
	KKM	150	123	27	0.680077046	
	Dibawah KKM	99	92	7	0.368556778	
Nilai Harian						0.00177417
	Atas	97	88	9	0.445693178	
	menengah	99	80	19	0.705469041	
	bawah	53	47	6	0.509515718	
Nilai Semester						0.004860351
	Baik	87	72	15	0.66319684	
	Cukup Baik	57	54	3	0.297472249	
	Sangat Baik	105	89	16	0.615766705	

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel diatas maka dapat disimpulkan nilai hasil gain tertinggi yaitu gain Ujian Nasional yang mendapatkan hasil 0.008116855.

Langkah selanjutnya adalah menghitung node 1.2 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya. Hasil dari perhitungan node 3 yang telah dihitung menggunakan rumus dapat dilihat pada tabel tabel 6.

Tabel 6. Perhitungan Node 1.2

Node 1.2	Belum Tercapai	155	147	8	0.293209	
Ujian Nasional						0.015846
	KKM	106	103	3	0.185803	
	Dibawah KKM	49	44	5	0.475432	
Nilai Harian						0.112652
	Atas	59	53	6	0.474345	
	menengah	28	28	0	0	
	bawah	68	66	2	0.191433	
Nilai Semester						0.209226
	Baik	84	84	0	0	
	Cukup Baik	68	61	7	0.47825	
	Sangat Baik	3	2	1	0.918296	

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel perhitungan kelulusan maka dapat disimpulkan nilai hasil gain tertinggi

yaitu gain Nilai Semester yang mendapatkan hasil 0.209226.

Langkah selanjutnya adalah menghitung node 1.1.1 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya. Hasil dari perhitungan node 4 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 7.

Tabel 7. Perhitungan Node 1.1.1

Ujian Nasional						
	KKM	150	123	27	0.680077046	
Nilai Harian						0.036862389
	Atas	39	31	8	0.73206669	
	menengah	93	74	19	0.730447286	
	bawah	18	18	0	0	
Nilai Semester						0.019874622
	Baik	84	70	14	0.650022422	
	Cukup Baik	31	28	3	0.458685816	
	Sangat Baik	35	25	10	0.863120569	

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel perhitungan kelulusan maka dapat disimpulkan nilai hasil gain tertinggi yaitu gain Nilai Harian yang mendapatkan hasil 0.036862389.

Langkah selanjutnya adalah menghitung node 1.1.2 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 8.

Tabel 8. Perhitungan Node 1.1.2

Node 1.1.2	Dibawah KKM	99	92	7	0.368557	
Nilai Harian						0.061265
	Atas	58	57	1	0.125658	
	menengah	6	6	0	0	
	bawah	35	29	6	0.660962	
Nilai Semester						-0.0712
	Baik	3	2	1	0.918296	
	Cukup Baik	26	26	0	0	
	Sangat Baik	70	64	14	0.582587	

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel perhitungan kelulusan maka dapat

disimpulkan nilai hasil gain tertinggi yaitu gain Nilai Harian yang mendapatkan hasil 0.061265.

Langkah selanjutnya adalah menghitung node 1.2.1 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 9.

Tabel 9. Perhitungan Node 1.2.1

Node 1.2.1					
	Cukup Baik	68	61	7	0.47
Ujian Nasional					
	KKM	28	26	2	0.371
	Dibawah KKM	40	35	5	0.543
Nilai Harian					
	Atas	42	37	5	0.526
	bawah	26	24	2	0.391

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel perhitungan kelulusan maka dapat disimpulkan nilai hasil gain tertinggi yaitu gain Ujian Nasional yang mendapatkan hasil 0.005646.

Langkah selanjutnya adalah menghitung node 1.2.2 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 10.

Tabel 10. Perhitungan Node 1.2.2

Node 1.2.2					
	Sangat Baik	3	2	1	0.918296
Ujian Nasional					0
	KKM	3	2	1	0.918296
	Dibawah KKM	0	0	0	0
Nilai Harian					0
	Atas	3	2	1	0.918296
	bawah	0	0	0	0

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel

diatas dan hasilnya adalah 0 maka perhitungan dihentikan.

Langkah selanjutnya adalah menghitung node 1.1.1.1 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 11.

Tabel 11. Perhitungan Node 1.1.1.1

Node 1.1.1.1					
	Atas	39	31	8	0.73206669
Nilai Semester					0.085274193
	Baik	23	16	7	0.886540893
	Cukup Baik	11	10	1	0.439496987
	Sangat Baik	5	5	0	0

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel diatas mendapatkan hasil akhir yaitu entropy baik 0.886540893, entropy Cukup Baik 0.439496987, dan entropy Sangat Baik = 0 maka perhitungan dihentikan.

Langkah selanjutnya adalah menghitung node 1.1.1.2 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 12.

Tabel 12. Perhitungan Node 1.1.1.2

Node 1.1.1.2					
	menengah	93	74	19	0.730447
Nilai Semester					0.097023
	Baik	61	54	7	0.514091
	Cukup Baik	2	0	2	0
	Sangat Baik	30	20	10	0.918296

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel diatas mendapatkan hasil akhir yaitu entropy baik 0.514091, entropy Cukup Baik 0, dan entropy Sangat Baik = 0.918296 maka perhitungan dihentikan.

Langkah selanjutnya adalah menghitung node 1.1.2.1 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 13.

Tabel 13. Perhitungan Node 1.1.2.1

node 1.1.2.1					
	Atas	58	57	1	0.125658
Nilai Semester					
	Baik	3	2	1	0.918296
	Cukup Baik	26	26	0	0
	Sangat Baik	70	64	14	0.582587

Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel diatas mendapatkan hasil akhir yaitu entropy baik 0.918296, entropy Cukup Baik0, dan entropy Sangat Baik = 0.582587maka perhitungan dihentikan.

Langkah selanjutnya adalah menghitung node 1.1.2.2 dengan menggunakan rumus entropy dan gain yang sama dengan sebelumnya Hasil dari perhitungan node 5 yang telah dihitung menggunakan rumus dapat dilihat pada tabel 14.

Tabel 14. Perhitungan Node 1.1.2.2

node 1.1.2.2					
	bawah	35	29	6	0.660962
Nilai Semester					
	Baik	3	2	1	0.918296
	Cukup Baik	26	26	0	0
	Sangat Baik	70	64	14	0.582587

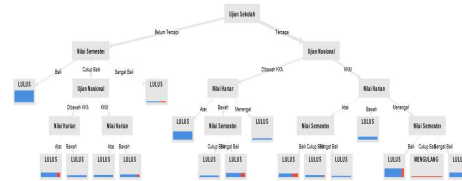
Dari hasil perhitungan nilai Entropy dan nilai gain pada tabel diatas mendapatkan hasil akhir yaitu entropy baik 0.918296, entropy Cukup Baik0, dan entropy Sangat Baik = 0.582587maka perhitungan dihentikan.

Selanjutnya adalah membuat pohon keputusan (*Decision Tree*) dari

nilai entropy dan gain yang sudah dihitung sebelumnya.

Decision Tree

Salah satu teknik klasifikasi yang paling populer digunakan dalam proses data mining adalah Decision Tree. Decision Tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas.

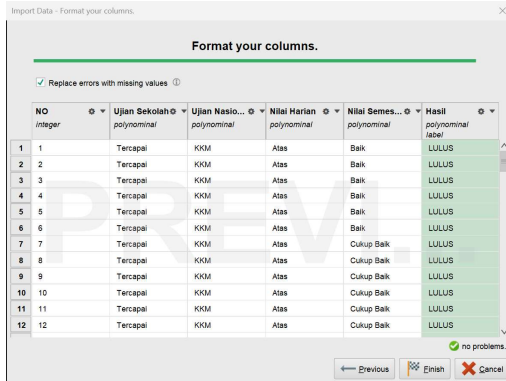


Gambar 2. Decision Tree Kelulusan Siswa

Dari gambar diatas diketahui bahwa node 1 yaitu Ujian Sekolah merupakan akar dari pohon keputusan (*root node*) dikarenakan pada perhitungan keseleruhan yang mendapatkan nilai terbesar ada pada atribut Ujian Sekolah. Jika hasil akhir mendapat nilai 0 maka perhitungan tidak akan dilanjutkan dan diberhentikan.

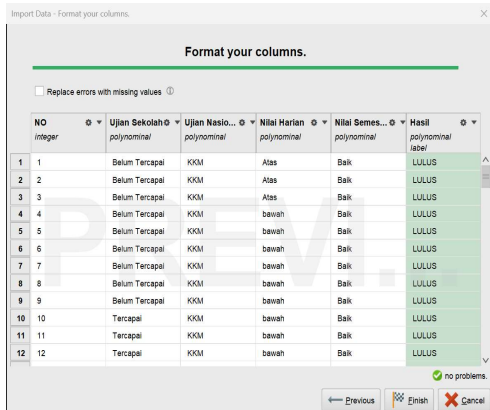
Implementasi Pengolahan Data dan Teknik Pengujian

Berikut merupakan import data training yang akan diproses oleh aplikasi rapidminer dengan algoritma C4.5 dan menggunakan operator “Read Excel”. Data training dapat dilihat pada gambar 3.



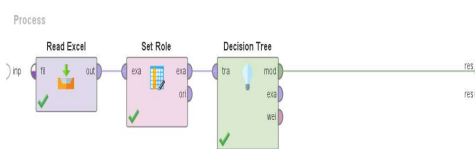
Gambar 3. Import Data Training

Total keseluruhan data adalah sebanyak 505. Perbandingan data yang digunakan pada data training dan data testing adalah 80 : 20. Proses Import Data Testing Dapat Dilihat Pada Gambar 4 di bawah ini



Gambar 4. Import Data Testing

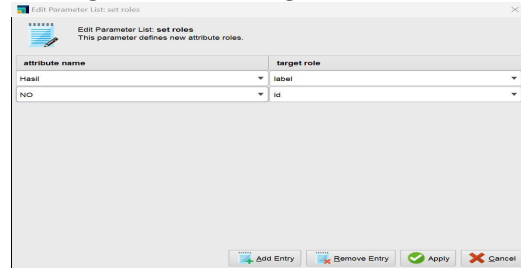
Gambar berikut ini adalah proses untuk menampilkan decision tree.



Gambar 5. Proses Pembentukan Decision Tree

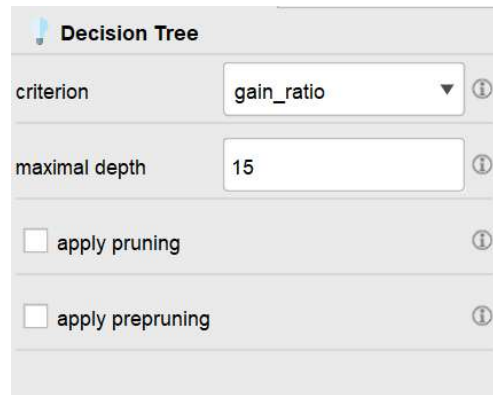
Berikut merupakan proses setting set role yang dilakukan dalam

menentukan role label pada data training dan data testing.



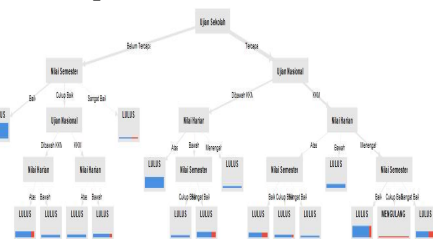
Gambar 6. Set Role Data Training dan Data Testing

Berikut merupakan cara setting operator decision tree



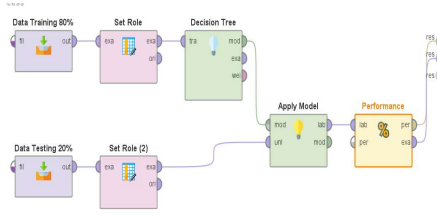
Gambar 7. Setting Operator Decision Tree

Berikut merupakan hasil decision tree yang dihasilkan oleh aplikasi rapidminer



Gambar 8. Decision Tree

Secara keseluruhan proses yang dilakukan dengan menggunakan RapidMiner adalah sebagai berikut :



Gambar 9. Proses Perhitungan Menggunakan Algoritma C4.5

Ada beberapa operator yang digunakan dalam proses perhitungan menggunakan algoritma C4.5. beberapa operator yang digunakan adalah read excel, set role, decision tree, apply model dan performance. Berikut merupakan hasil yang diperoleh menggunakan algoritma C4.5 dengan alat bantu pengujian data rapidminer.

```

Ujian Sekolah = Belum Tercapai
| Nilai Semester = Baik: LULUS {LULUS=84, MENGULANG=0}
| Nilai Semester = Cukup Baik
| | Ujian Nasional = Dibawah KKM
| | | Nilai Harian = Atas: LULUS {LULUS=25, MENGULANG=5}
| | | Nilai Harian = Bawah: LULUS {LULUS=10, MENGULANG=0}
| | Ujian Nasional = KKM
| | | Nilai Harian = Atas: LULUS {LULUS=12, MENGULANG=0}
| | | Nilai Harian = Bawah: LULUS {LULUS=14, MENGULANG=2}
| Nilai Semester = Sangat Baik: LULUS {LULUS=2, MENGULANG=1}
Ujian Sekolah = Tercapai
| Ujian Nasional = Dibawah KKM
| | Nilai Semester = BAIK: MENGULANG {LULUS=0, MENGULANG=1}
| | Nilai Semester = Baik: LULUS {LULUS=2, MENGULANG=0}
| | Nilai Semester = Cukup Baik: LULUS {LULUS=26, MENGULANG=0}
| | Nilai Semester = Sangat Baik
| | | Nilai Harian = Atas: LULUS {LULUS=43, MENGULANG=0}
| | | Nilai Harian = Bawah: LULUS {LULUS=21, MENGULANG=6}
| Ujian Nasional = KKM
| | Nilai Harian = Atas
| | | Nilai Semester = Baik: LULUS {LULUS=16, MENGULANG=7}
| | | Nilai Semester = Cukup Baik: LULUS {LULUS=10, MENGULANG=1}
| | | Nilai Semester = Sangat Baik: LULUS {LULUS=5, MENGULANG=0}
| | Nilai Harian = Bawah: LULUS {LULUS=18, MENGULANG=0}
| | Nilai Harian = Menengah
| | | Nilai Semester = Baik: LULUS {LULUS=54, MENGULANG=7}
| | | Nilai Semester = Cukup Baik: MENGULANG {LULUS=0, MENGULANG=2}
| | | Nilai Semester = Sangat Baik: LULUS {LULUS=20, MENGULANG=10}
    
```

Gambar 10. Deskripsi Decision Tree
Berikut merupakan hasil yang dihasilkan oleh apply model.

Gambar 11. Hasil Apply Model

Berikut merupakan hasil akurasi yaitu sebesar 66,34 %

accuracy: 66.34%

	true LULUS	true MENGULANG	class precision
pred LULUS	67	34	66.34%
pred MENGULANG	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 12. Hasil Akurasi

Berikut merupakan hasil dari performance dan confusion matrix

PerformanceVector

```

PerformanceVector:
accuracy: 66.34%
ConfusionMatrix:
True:   LULUS   MENGULANG
LULUS:  67      34
MENGULANG:  0      0
    
```

Gambar 13. Performance Confusion Matrix

V. Kesimpulan dan Saran

Kesimpulan

Berdasarkan hasil penelitian diatas maka kesimpulan yang didapat adalah sebagai berikut :

1. Pohon keputusan yang dihasilkan menggunakan perhitungan manual dengan rapidminer hasilnya berbeda.
2. Rapidminer sangat akurat untuk digunakan dalam memprediksi tingkat kelulusan siswa, sehingga pihak sekolah dapat lebih memperhatikan hal – hal yang sangat berpengaruh pada proses pendidikan.

Saran

Saran yang dapat disampaikan adalah

1. Perhitungan dan Pengujian data dapat digunakan menggunakan

aplikasi lain selain dari rapidminer seperti excel.

2. Kepada peneliti untuk lebih menjelaskan secara detail proses keseluruhan agar lebih mudah dipahami oleh banyak orang.
3. Kepada pihak sekolah untuk lebih memahami factor – factor yang dapat mempengaruhi proses pendidikan.

VI. Daftar Pustaka

- S. Andayani, “Formation of clusters in Knowledge Discovery in Databases by Algorithm K-Means,” SEMNAS Mat. dan Pendidik. Mat. 2007, 2007.
- C. N. Dengen, K. Kusriani, and E. T. Luthfi, “Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu,” Sisfotek, vol. 10, no. 1, p. 1, 2020, doi: 10.30700/jst.v10i1.484.
- Hendry, “Data mining Prediksi Data Tingkat Malas Siswa Di Sekolah SMA,” Apl. dan Anal. Lit. Fasilkom UI, vol. m, no. 1998, pp. 7–34, 2021, [Online]. Available: <http://elib.unikom.ac.id/files/disk1/655/jbptunikompp-gdl-supriadini-32740-6-12.unik-i.pdf>
- L. Yuningsih, I. R. Setiawan, and A. A. Sunarto, “Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5,” Progresif J. Ilm. Komput., vol. 16, no. 2, p. 121, 2020, doi: 10.35889/progresif.v16i2.517.
- Y. Partogi and A. Pasaribu, “Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan STMIK Kuwera,” J. Sist. Inf. dan Teknol., vol. 1, no. 2, pp. 20–25, 2022, doi: 10.56995/sintek.v1i2.4.
- I. Romli and A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5,” J. Sains Komput. Inform. (J-SAKTI), vol. 4, no. 2, pp. 694–702, 2020.
- A. Ulfa, D. Winarso MKom, and E. Arribe MMSi, “SISTEM REKOMENDASI JURUSAN KULIAH BAGI CALON MAHASISWA BARU MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus : Universitas Muhammadiyah Riau),” J. Fasilkom, vol. 10, no. 1, pp. 61–65, 2020.
- P. S. Akuntansi, “1* , 2 1,2,” vol. 20, no. 1, pp. 105–123, 2022.
- M. Kamil and W. Cholil, “Analisis Perbandingan Algoritma C4.5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang,” J. Inform., vol. 7, no. 2, pp. 97–106, 2020, doi: 10.31294/ji.v7i2.7723.
- M. M. Effendi and A. Setiawan, “Menentukan Prediksi Kelulusan Siswa Dengan Membandingkan Algoritma C4.5 Dan Naive Bayes Studi Kasus Smkn. 1 Cikarang Selatan,” SIGMA - J. Teknol. Pelita Bangsa, vol. 11, no. 3, pp. 143–148, 2020.