



Prediksi Hasil Belajar Siswa Secara Daring pada Masa Pandemi COVID-19 Menggunakan Metode C4.5

Yetti Fitriani^{1✉}, Sarjon Defit², Gunadi Widi Nurcahyo³

¹SMK Negeri 2 Padang Panjang

^{2,3}Universitas Putra Indonesia YPTK Padang

yeti.fitriani@gmail.com

Abstract

Student learning in schools has changed since the Covid-19 pandemic. Student learning in normal conditions is carried out face-to-face and turns into online or online learning. The research was conducted to predict student learning outcomes during the COVID-19 pandemic so that the results of this study can be used as a reference in policymaking in schools. The C4.5 method was used in the study to classify the data for class XII of the Multimedia Department at SMKN 2 Padang Panjang and the classification results could predict student learning outcomes during the pandemic. Processed student value data were taken from 1 (one) subject as the research data sample. Analysis of the value of student learning outcomes using the C4.5 Method to obtain new knowledge from student learning outcomes data carried out during the COVID-19 pandemic. The data analyzed consisted of attributes of attendance, assignments, daily tests, and test scores which influenced the decision criteria for student learning outcomes in online learning. The learning outcome decision criteria consist of "Satisfactory" and "Not Satisfactory" which refer to the Minimum Completion Criteria. Tests conducted on the training data of learning outcomes show that the value of the Daily Test is the most influential attribute in decision making. Implementation of the results using the RapidMiner Studio 9.2.0 software and produces an accuracy of 83.33% of the test data testing with the rules of data analysis training results. The results of the C4.5 classification testing method in this study can be used to predict student learning outcomes. The test results with an accuracy of 83.33% can be recommended to help schools in making policies.

Keywords: Online Learning, Pandemic COVID-19, Data Mining, C4.5, Rules.

Abstrak

Pembelajaran siswa di sekolah mengalami perubahan sejak pandemi Covid-19. Pembelajaran siswa pada kondisi normal dilaksanakan secara tatap muka berubah menjadi pembelajaran *Online* atau Dalam Jaringan (Daring). Penelitian dilakukan untuk memprediksi hasil belajar siswa pada masa pandemi COVID-19 sehingga hasil penelitian bisa digunakan untuk menjadi acuan dalam pengambilan kebijakan di sekolah. Metode C4.5 digunakan pada penelitian untuk mengklasifikasi data nilai siswa kelas XII Jurusan Multimedia di SMKN 2 Padang Panjang dan hasil klasifikasi dapat memprediksi hasil belajar siswa pada masa pandemi. Data nilai siswa yang diolah diambil dari 1 (satu) mata pelajaran sebagai sampel data penelitian. Nilai hasil belajar siswa dianalisa menggunakan Metode C4.5 untuk mendapatkan pengetahuan baru dari data nilai hasil belajar siswa yang dilaksanakan pada masa pandemi COVID-19. Data yang dianalisa terdiri dari atribut absensi, tugas, Ulangan Harian (UH) dan nilai ujian yang mempengaruhi kriteria keputusan hasil belajar siswa pada pelajaran yang dilaksanakan secara daring. Kriteria keputusan hasil belajar terdiri dari "Memuaskan" dan "Kurang Memuaskan".

Yang mengacu kepada Kriteria Ketuntasan Minimal (KKM). Pengujian yang dilakukan terhadap data *training* hasil belajar menunjukkan bahwa nilai Ulangan Harian (UH) merupakan atribut yang paling mempengaruhi terhadap keputusan. Implementasi hasil menggunakan *Software* Rapid Miner Studio 9.2.0 dan menghasilkan akurasi sebesar 83,33% dari pengujian data *testing* dengan *rule-rule* hasil analisis data *training*. Hasil pengujian klasifikasi metode C4.5 pada penelitian ini bisa digunakan untuk memprediksi hasil belajar siswa. Hasil pengujian dengan akurasi sebesar 83,33% sudah dapat direkomendasikan untuk membantu pihak sekolah dalam membuat kebijakan.

Kata kunci: Daring, Pandemi COVID-19, *Data Mining*, C4.5, *Rule*.

© 2021 JSisfotek

1. Pendahuluan

Pandemi Covid-19 telah membuat banyak perubahan pada pola kehidupan di masyarakat. Dunia pendidikan merupakan salah satu bidang yang mengalami perubahan sangat signifikan. Sistem-sistem yang dilaksanakan pada bidang pendidikan berubah menyesuaikan dengan kondisi yang ada ditengah pandemi Covid-19. Perubahan yang sangat drastis untuk pendidikan terjadi pada proses pembelajaran siswa yang biasanya dilaksanakan di kelas secara tatap muka beralih pada metode *online* atau dalam jaringan yang biasanya dikenal dengan sebutan daring.

SMKN 2 Padang Panjang sejak pandemi Covid-19 melaksanakan pembelajaran siswa secara daring. Pembelajaran Daring yang dilaksanakan di SMKN 2 Padang Panjang dilakukan dengan menggunakan beberapa aplikasi maupun portal untuk mendukung terlaksananya pembelajaran dalam jaringan seperti WhatsApp, Telegram, Google Classroom, dan beberapa aplikasi lain yang tersedia dan sudah melakukan kerjasama dengan Kemendikbud.

Knowledge Discovery in Database (KDD) yang biasa disebut dengan *Data Mining* menghasilkan pola yang sangat berharga dari data yang berukuran besar. KDD

menggunakan beberapa metode untuk mendapatkan informasi baru, di antaranya yang sering digunakan metode klasifikasi dengan Algoritma C4.5.

KDD diartikan sebagai keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, di mana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti [1]. Data Mining sering dianggap sama dengan *Knowledge Discovery in Database* (KDD), padahal *Data Mining* adalah salah satu tahap dari rangkaian proses pengolahan data pada KDD. Data Mining bidang ilmu yang tidak berdiri sendiri karena Data Mining memiliki keterkaitan dengan bidang ilmu lainnya [2]. Metode-metode yang dapat digunakan dalam *Data Mining* adalah klasifikasi, regresi, seleksi variabel, market basket analisis dan *clustering* [3].

Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki *atribut-atribut* numerik dan kategorial [4]. Metode C4.5 banyak digunakan pada penelitian di bidang pendidikan. Metode C4.5 untuk memprediksi pekerjaan lulusan perguruan tinggi serta menganalisa faktor-faktor yang mempengaruhi lulusan perguruan tinggi mendapatkan pekerjaan dan kualitas dari pekerjaan tersebut [5]. Algoritma C4.5 digunakan untuk memprediksi siswa dalam memperoleh bantuan dan pendidikan [6]. Algoritma C4.5 dimanfaatkan pada penerapan *Educational Data Mining* untuk memprediksi hasil belajar siswa SMAK Ora et Labora [7].

Penelitian-penelitian sebelumnya yang menggunakan Metode C4.5 adalah implementasi *Decision Tree* untuk prediksi kelulusan mahasiswa tepat waktu [8]. Memprediksi mahasiswa yang berpotensi non aktif menggunakan *Data Mining* dalam *DecisionTree* dan Algoritma C4.5 [9]. Mengatasi permasalahan pemberian beasiswa yang kurang adil dan tidak efisien dengan membuat sistem evaluasi beasiswa mahasiswa menggunakan pohon keputusan C4.5 [10].

Penerapan Metode C4.5 pada penelitian lainnya adalah sistem prediksi dan evaluasi prestasi akademik mahasiswa menggunakan *Data Mining*, di mana klasifikasi dilakukan untuk menemukan faktor-faktor yang mempengaruhi prestasi mahasiswa [11]. Mengidentifikasi faktor-faktor yang mempengaruhi jenis kecelakaan dengan menggunakan Algoritma C4.5 [12]. Menggabungkan Algoritma C4.5 dengan Algoritma Relief-F untuk mengatasi permasalahan status stabilitas tegangan sistem agar dapat diidentifikasi secara *online* [13].

Pembelajaran daring walaupun tidak lagi pandemi namun tetap akan dipertahankan dengan dengan pola-pola tertentu. Pandemi COVID-19 menjadi tonggak perubahan pembelajaran menjadi daring, diperkirakan sebagian besar pembelajaran beberapa disiplin ilmu tetap akan dilakukan *online* bahkan setelah COVID-19 [14]. Siswa mengalami penurunan semangat belajar selama pembelajaran daring di masa pandemi COVID-

19 yang disebabkan oleh berbagai kondisi lingkungan keluarga yang tidak kondusif [15].

Pendidik dan peserta didik banyak yang belum siap menghadapi revolusi industri 4.0, tetapi dengan adanya pembelajaran daring pada masa pandemi COVID-19 memaksa semua orang untuk siap menghadapi perkembangan teknologi [16]. Beragam cara siswa dalam menyikapi dan mengikuti pembelajaran daring, di mana dapat mempengaruhi hasil belajar siswa yang bisa dilihat dari proses penilaian hasil belajar seperti penilaian harian terhadap siswa dalam hal ini bisa berupa tugas-tugas, ulangan harian dan lain-lain.

Perubahan pola pembelajaran siswa yang dilaksanakan secara daring pada masa pandemi inilah yang menjadi dasar penelitian ini, sehingga dapat dilakukan analisa terhadap hasil belajar siswa. Metode C4.5 digunakan untuk menganalisa data nilai hasil belajar siswa yang dilaksanakan pada masa pandemi, sehingga dari data nilai yang ada bisa digunakan untuk memprediksi hasil belajar siswa dan *knowledge* yang didapatkan bisa menjadi pertimbangan untuk kebijakan di sekolah dan bisa meningkatkan hasil pembelajaran.

Tujuan yang akan dicapai pada penelitian ini adalah memahami Metode C4.5 untuk memprediksi hasil belajar siswa secara daring pada masa pandemi COVID-19 di SMK Negeri 2 Padang Panjang. Kemudian menganalisa nilai hasil belajar siswa secara daring pada masa pandemi COVID-19 untuk mendapatkan *knowledge* yang dihasilkan dari data nilai hasil belajar siswa tersebut disertai variabel-variabel pembentuknya. Setelah itu menguji Metode C4.5 untuk hasil belajar siswa pada masa pandemi COVID-19 dengan menggunakan *Software* Rapid Miner Studio 9.2.0.

2. Metodologi Penelitian

Metode penelitian merupakan rancangan dari kegiatan penelitian yang dilakukan dalam mencari, merumuskan serta menganalisa hingga dapat menyusun sesuatu dengan langkah-langkah yang akan digunakan dan waktu yang dapat digunakan sebagai acuan dalam memperoleh dalam analisa data. Metode penelitian diperlukan untuk membantu penulisan dapat terarah sesuai dengan masalah yang diteliti.

Kerangka kerja penelitian dibuat untuk memudahkan penulis dalam melakukan penelitian agar lebih mudah dipahami dan tidak keluar dari pokok pembahasan. Langkah-langkah yang dibuat dalam kerangka kerja penelitian disusun secara sistematis guna menyelesaikan permasalahan yang ada. Alur kerangka kerja penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada gambar dapat diuraikan masing-masing langkahnya sebagai berikut:

2.1 Mengidentifikasi Masalah

Tahap identifikasi masalah merupakan tahapan awal penelitian. Peneliti melakukan perumusan masalah dari kondisi hasil belajar siswa pada pembelajaran yang dilaksanakan secara daring masa pandemi COVID-19.

2.2 Menganalisa Masalah

Pada tahapan ini menganalisa kondisi penilaian hasil belajar pada masa pandemi COVID-19, supaya dapat diketahui variabel-variabel penilaian apa saja yang mempengaruhi hasil belajar siswa.

2.3 Mempelajari Literatur

Mempelajari literatur juga sangat penting dilakukan agar mempunyai landasan baik secara teoritis yang benar dan dijelaskan oleh para peneliti dan ahli sebelumnya. Kemudian literatur-literatur yang dipelajari tersebut diseleksi dan dipilih literatur mana yang akan digunakan dalam penelitian. Literatur yang diambil dari berbagai sumber yaitu berupa buku, artikel, jurnal ilmiah tentang Algoritma C4.5, Data mining, serta bacaan lain yang mendukung.

2.4 Mengumpulkan Data

Pengumpulan data dilakukan agar dapat memperoleh informasi sehingga tujuan dari penelitian menggunakan Metode C4.5 ini dapat tercapai. Pengumpulan data dengan melakukan observasi langsung di SMKN 2 Padang Panjang. Pengamatan secara langsung dilokasi penelitian untuk melihat data yang akan diproses dan data yang diambil merupakan data nilai siswa di SMK Negeri 2 Padang Panjang.

2.5 Menganalisa Data Menggunakan Metode C4.5

Analisa data dengan Metode C4.5 digunakan untuk memprediksi hasil belajar siswa secara daring pada masa pandemi COVID-19. Sebelum dianalisa dengan Metode C4.5, data yang dikumpulkan melalui tahap *preprocessing* (data *cleaning* dan *transformation*) terlebih dahulu, sehingga menghasilkan data *set* berupa data *training* dan data *testing*. Setelah itu dilanjutkan dengan analisa data menggunakan Metode C4.5.

2.6 Mengimplementasikan Sistem

Implementasi dilakukan untuk mendapatkan pola terbaik dalam penelitian dalam hal terhadap penentuan kriteria hasil belajar siswa yang “memuaskan” dan “tidak memuaskan”, dengan menggunakan data *training* dan dibandingkan dengan data *testing*. Metode C4.5 menggunakan *Tools Data Mining* dan untuk penelitian ini menggunakan *Software Rapid Miner Studio 9.2.0*.

2.7 Hasil dan Pembahasan

Hasil perhitungan Metode C4.5 tersebut nantinya akan terdapat *rule-rule* yang dihasilkan, kemudian dilakukan pengujian hasil yang diperoleh dari perhitungan manual tersebut dibandingkan dengan hasil dari *Software Rapid Miner Studio 9.2.0*. *Data training* dan *data testing* diuji coba terhadap *rule* untuk mendapatkan nilai akurasi. Hasil *rule* dan *knowledgenya* dapat digunakan untuk rekomendasi kebijakan di sekolah.

3. Hasil dan Pembahasan

3.1 Data

Objek yang diolah pada penelitian ini adalah nilai hasil belajar siswa pada Penilaian Tengah Semester (PTS) Semester Ganjil Tahun Pelajaran 2020/2021 yang merupakan puncak dari pandemi COVID-19 dan pembelajaran siswa semuanya dilaksanakan secara daring.

Data penelitian didapatkan dari data nilai siswa yang ada di Tata Usaha SMKN 2 Padang Panjang. Data yang digunakan pada penelitian adalah data nilai mata pelajaran Desain Media Interaktif (DMI), dengan data sampel 61 orang siswa kelas XII program Keahlian Multimedia yang terdiri dari 2 kelas yaitu 1 kelas sebagai data *training* dan 1 kelas sebagai data *testing*. Data nilai DMI menjadi sampel mata pelajaran yang akan diklasifikasi untuk dapat memprediksi nilai hasil belajar pada mata pelajaran yang pelaksanaan pembelajarannya berada di masa pandemi COVID-19.

Data nilai yang dianalisa adalah nilai hasil belajar pada Penilaian Tengah Semester (PTS) Ganjil mata pelajaran DMI disertai variabel-variabel yang mempengaruhi hasil nilai mata pelajaran tersebut. Hasil belajar yang akan diprediksi berdasarkan nilai hasil PTS yang merupakan pengolahan dari beberapa item data nilai yang ada penilaian pembelajaran siswa yang diambil pada saat pembelajaran daring pada masa pandemi. Penelitian menggunakan data nilai mata pelajaran untuk memprediksi nilai hasil belajar siswa pada masa pandemi kepada salah satu mata pelajaran sebagai sampel data dengan menganalisa data nilai menggunakan metode C4.5 dengan melihat atribut-atribut langsung atau data internal yang berpengaruh langsung terhadap penilaian hasil belajar tersebut.

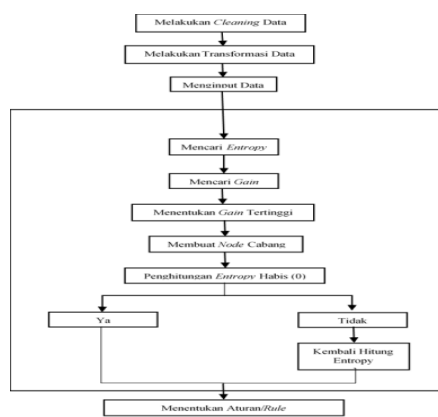
3.2 Analisa Sistem (Metode C4.5)

Metode C4.5 merupakan salah satu metode yang dapat digunakan untuk melakukan klasifikasi data dan kemudian *rule* yang didapatkan dari klasifikasi data tersebut dapat digunakan untuk melakukan prediksi.

Sistem yang akan dibangun dengan Metode C4.5 dalam penelitian ini adalah melakukan klasifikasi terhadap data nilai hasil belajar siswa sehingga menghasilkan *rule-rule* yang dapat digunakan untuk memprediksi hasil belajar siswa secara daring pada masa pandemi COVID-19.

Algoritma C4.5 menganalisa data yang ada dengan melakukan perhitungan *entropy* dan *Gain* sehingga menemukan *Node* dan *Leaf* yang akan membentuk pohon keputusan. Pohon keputusan menghasilkan *rule-rule* dapat digunakan pada data lain untuk memprediksi hasil belajar siswa pada masa pandemi COVID-19. Pengolahan data menggunakan Algoritma C4.5 untuk nilai hasil belajar siswa dimasa pandemi dan *rule-rule* yang dihasilkan dari analisa data dengan metode C4.5 dapat digunakan untuk memprediksi hasil belajar siswa.

Penerapan Algoritma C4.5 dalam penelitian digambarkan dengan diagram proses agar lebih terarah dan mudah dipahami. Diagram akan menggambarkan langkah-langkah perhitungan metode C4.5 dalam menghasilkan pohon keputusan. Diagram proses untuk Algoritma C4.5 penelitian terdapat pada Gambar 2.



Gambar 2. Diagram Proses Algoritma C4.5

3.2.1 Processing dan Transformasi Data

Sebelum dilakukan pengolahan data, data tersebut sudah dalam keadaan bersih dan siap untuk digunakan dengan melakukan *cleaning* data. Data *cleaning* dilakukan menentukan data-data mana saja yang akan dipergunakan karena pada data awal terdapat banyak data dan tidak semua *atribut* data yang akan digunakan maka dilakukan *cleaning* data, seperti pada data tersebut terdapat Nomor Induk Siswa (NIS) dan jenis kelamin.

Tabel 1 merupakan transformasi dan deskripsi untuk hasil belajar siswa yang menggunakan 1 *Atribut* sebagai *atribut* target (Keputusan=Memuaskan atau Keputusan=Kurang memuaskan). Keputusan untuk hasil belajar siswa diambil dari data Nilai DMI untuk Penilaian Tengah Semester (PTS) siswa. Data nilai siswa tersebut kemudian ditransformasi dengan memberikan *range* *atribut* keputusan. *Range* yang digunakan pada keputusan hasil belajar siswa mengacu kepada Kriteria Ketuntasan Minimal (KKM) mata pelajaran Kejuruan di SMKN 2 Padang Panjang.

Tabel 2 merupakan *atribut* penunjang yang terdiri dari 4 *atribut*. *Atribut* penunjang pada Tabel 2 memerlukan beberapa perubahan pada *atribut* dan nilai *atribut* sehingga dapat memudahkan pada saat dilakukan pengolahan data menggunakan Algoritma C4.5. *Atribut* penunjang yang digunakan pada penelitian ini merupakan *atribut-atribut* yang menentukan nilai *atribut* keputusan disajikan pada Tabel 1 sampai dengan Tabel 3.

Tabel 1. Kriteria Keputusan dan Hasil Belajar

Kriteria hasil belajar	Nilai <i>Atribut</i>	Deskripsi
Hasil Belajar	Memuaskan	Kriteria hasil belajarmemuaskan jika nilai ≥ 65
	Kurang Memuaskan	Kriteria hasil belajarkurang memuaskan jika nilai < 65

Tabel 2. *Atribut* Penunjang dan Parameter

<i>Atribut</i>	Deskripsi	Kategori	Keterangan
Absensi	Persentase kehadiran siswa	$<70\%$	Jika kehadiran siswa kurang dari 70%. minimal kehadiran siswa pada masa pandemi 70%.
		$\geq 70\%$	Jika kehadiran sama atau lebih dari 70%, minimal kehadiran siswa pada masa pandemi 70%
Tugas	Rata-rata nilai tugas siswa	A	Rata-rata nilai tugas siswa dengan range 85-100.
		B	Rata-rata nilai tugas siswa dengan range 75-84.
		C	Rata-rata nilai tugas siswa dengan range 65-74.
		D	Rata-rata nilai tugas siswa dengan range 0-64.
UH	Rata-rata nilai ulangan harian siswa	A	Rata-rata nilai ulangan harian siswa dengan range 85-100.
		B	Rata-rata nilai ulangan harian siswa dengan range 75-84.
		C	Rata-rata nilai ulangan harian siswa dengan range 65-74.
		D	Rata-rata nilai ulangan harian siswa dengan range 0-64.
Ujian	Nilai ujian tengah semester	A	Nilai ujian tengah semester dengan range 85-100.
		B	Nilai ujian tengah semester dengan range 75-84.
		C	Nilai ujian tengah semester dengan range 65-74.
		D	Nilai ujian tengah semester dengan range 0-64.

Tabel 3. Hasil Transformasi Data Training

Absensi	Tugas	UH	Ujian	Hasil Belajar
$<70\%$	B	A	A	Memuaskan
$\geq 70\%$	A	A	B	Memuaskan
$\geq 70\%$	D	B	D	Kurang Memuaskan
$\geq 70\%$	A	A	C	Memuaskan
$<70\%$	D	B	B	Kurang Memuaskan
$\geq 70\%$	D	A	C	Memuaskan
$\geq 70\%$	A	A	B	Memuaskan
$\geq 70\%$	D	A	D	Memuaskan
$\geq 70\%$	A	A	D	Memuaskan
$\geq 70\%$	D	D	D	Kurang Memuaskan

Hasil dari transformasi data siswa berupa data *training* dengan perhitungan *Gain* semua *atribut*, yang dapat dilihat pada Tabel 3. *Data training* terdiri dari 31 data nilai siswa, yang ditampilkan hanya 10 data.

3.2.2 Pengolahan data

a. Mencari Node 1

Langkah awal dalam membentuk pohon keputusan dengan Algoritma C4.5 adalah dengan melakukan beberapa perhitungan untuk membentuk *node* 1, yaitu menghitung *Entropy* Total, *Entropy* masing-masing *atribut* dan *Gain* untuk semua *atribut*. Hasil perhitungannya ditampilkan pada Tabel 4.

Tabel 4. Perhitungan Node 1

No de		S	S1	S2	E	G
1	Total	31	23	8	0,8238	
	Absensi					0,2982
	<70%	8	2	6	0,8113	
	>=70%	23	21	2	0,4262	
	Tugas					0,2804
	A	11	11	0	0,0000	
	B	6	5	1	0,6500	
	C	1	1	0	0,0000	
	D	13	6	7	0,9957	
	UH					0,4335
	A	18	18	0	0,0000	
	B	4	1	3	0,8113	
	C	4	2	2	1,0000	
	D	5	2	3	0,9710	
	Ujian					0,1914
	A	4	4	0	0,0000	
	B	7	6	1	0,5917	
	C	9	8	1	0,5033	
	D	11	5	6	0,9940	

Perhitungan tiap *nodes* sama tahapannya sama seperti dimulai dengan perhitungan *Entropy* dan *Gain*. Berikut rumus perhitungan *Entropy*.

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

Dimana S adalah Jumlah sampel data (sampling), n adalah Jumlah partisi S , dan p_i merupakan Proporsi dari S_i terhadap S .

Entropy (Total)

$$= \left(-\frac{23}{31} * \log_2 \left(\frac{23}{31} \right) \right) + \left(-\frac{8}{31} * \log_2 \left(\frac{8}{31} \right) \right) = 0,8238$$

Entropy (< 70%)

$$= \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8} \right) \right) + \left(-\frac{6}{8} * \log_2 \left(\frac{6}{8} \right) \right) = 0,8113$$

Entropy (>= 70%)

$$= \left(-\frac{21}{23} * \log_2 \left(\frac{21}{23} \right) \right) + \left(-\frac{2}{23} * \log_2 \left(\frac{2}{23} \right) \right) = 0,4262$$

Perhitungan *entropy* tiap nilai *atribut* (variabel *atribut*) berlaku untuk semua *atribut*. Perhitungan yang ditampilkan hanya 1 *atribut* saja. Kemudian dilanjutkan

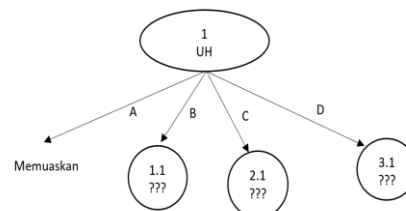
Gain (Total, Absensi)

$$= Entropy(Total) - \sum_{i=1}^n \frac{Absensi}{Total}$$

Gain (Total, Absensi)

$$= 0,8238 - \left(\left(\frac{8}{31} * 0,8113 \right) + \left(\frac{23}{31} * 0,4262 \right) \right) = 0,2982$$

Tabel 4 perhitungan Node 1 memperlihatkan *Gain* tertinggi adalah pada *atribut* UH dengan nilai = 0,4335 dan menjadi *node* akar untuk penghitungan algoritma C 4.5 berikutnya. *Atribut* UH dengan nilai *atribut* "A" mempunyai keputusan "Memuaskan". *Atribut* UH dengan nilai *atribut* "B", "C" dan "D" masih mempunyai nilai yang akan dihitung, dan menjadi cabang untuk perhitungan pohon keputusan selanjutnya. Pohon keputusan awal ditampilkan pada Gambar 3.



Gambar 3. Pohon Keputusan Awal

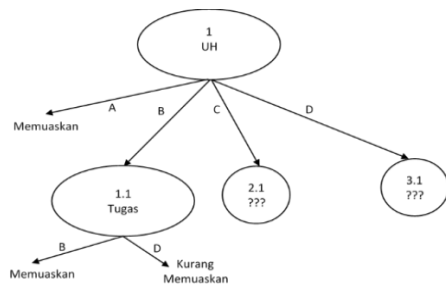
b. Mencari Node 1.1

Proses pencarian *node* 1.1 tahapannya sama dengan *node* 1 dan hasil perhitungan *Entropy* dan *Gain*nya berada pada Tabel 5.

Tabel 5. Perhitungan Node 1.1

No de		S	S1	S2	E	G
1.1	UH	B	4	1	3	0,8113
	Absensi					0,3113
	<70%	2	0	2	0,0000	
	>=70%	2	1	1	1,0000	
	Tugas					0,8133
	A	0	0	0	0,0000	
	B	1	1	0	0,0000	
	C	0	0	0	0,0000	
	D	3	0	3	0,0000	
	Ujian					0,1226
	A	0	0	0	0,0000	
	B	1	0	1	0,0000	
	C	0	0	0	0,0000	
	D	3	1	2	0,9183	

Tabel 5 memperlihatkan *Gain* tertinggi adalah *atribut* Tugas yang akan menjadi cabang akar untuk pohon keputusan selanjutnya, dengan nilai *Gain* adalah 0.8113. “Tugas_B” sudah mempunyai keputusan “Memuaskan”, dan “Tugas_D” mempunyai keputusan “Tidak Memuaskan”.



Gambar 4. Pohon Keputusan Node 1.1

c. Mencari Node 2.1

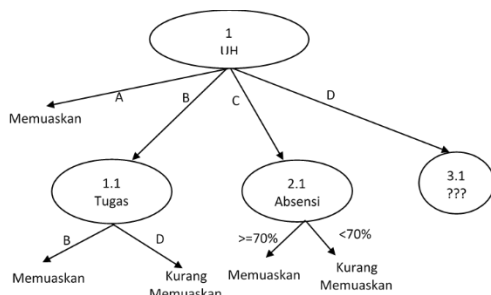
Perhitungan untuk *node* 2.1 merupakan perulangan dengan pola yang sama seperti *node* 1 dan *node* 1.1.

Tabel 6. Perhitungan Node 2.1

No de		S	S1	S2	E	G
2.1	UH	C	4	2	2	1,0000
	Absensi					1,0000
		<70%	2	0	2	0,0000
		>=70%	2	2	0	0,0000
	Tugas					
		A	0	0	0	0,0000
		B	1	1	0	0,0000
		C	0	0	0	0,0000
		D	3	1	2	0,9183
	Ujian					0,5000
		A	0	0	0	0,0000
		B	1	1	0	0,0000
		C	1	0	1	0,0000
		D	2	1	1	1,0000

Hasil perhitungan pada Tabel 6 merupakan perhitungan pencarian *Entropy*, *Gain* dan memilih *gain* tertinggi sebagai *node* akar sama halnya seperti pada *node* 1 dan *node* 1.1.

Proses perhitungan *node* 2.1 untuk *Entropy* dan *Gain* pada Tabel 6 yang mempunyai nilai *Gain* tertinggi adalah Absensi dengan nilai *gain* sebesar 1,0000. Selanjutnya *atribut* Absensi dengan nilai *atribut* <70% mempunyai keputusan “Kurang Memuaskan” dan *atribut* Absensi dengan nilai *atribut* >=70% mempunyai keputusan “Memuaskan” akan menjadi cabang akar untuk pohon keputusan selanjutnya, maka pohon keputusan yang terbentuk adalah seperti Gambar 5.



Gambar 5. Pohon Keputusan Node 2.1

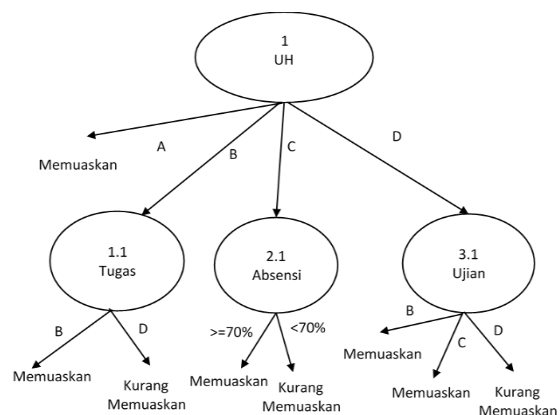
d. Mencari Node 3.1

Perhitungan selanjutnya adalah untuk *atribut* “UH_D”, karena cabang akarnya untuk *node* 3.1 masih memerlukan keputusan yang tepat.

Tabel 7. Perhitungan Node 3.1

No de		S	S1	S2	E	G
3.1	UH	D	5	2	3	0,9710
	Absensi					0,0200
		<70%	3	2	1	0,9183
		>=70%	2	1	1	1,0000
	Tugas					0,5710
		A	0	0	0	0,0000
		B	2	1	1	1,0000
		C	1	1	0	0,0000
		D	2	0	2	0,0000
	Ujian					0,9710
		A	0	0	0	0,0000
		B	1	1	0	0,0000
		C	1	1	0	0,0000
		D	3	0	3	0,0000

Proses perhitungan *node* 3.1 untuk *entropy* dan *gain* terlihat pada Tabel 7, yang mempunyai nilai *Gain* tertinggi adalah *atribut* Ujian dengan nilai *gain* sebesar 0,9710. Selanjutnya *atribut* Ujian dengan nilai *atribut* B mempunyai keputusan “Memuaskan”, *atribut* Ujian dengan nilai *atribut* C mempunyai keputusan “Memuaskan” dan *atribut* Absensi dengan nilai *atribut* D mempunyai keputusan “Kurang Memuaskan” akan menjadi cabang akar untuk pohon keputusan selanjutnya, maka pohon keputusan yang terbentuk adalah seperti Gambar 6.



Gambar 6. Pohon Keputusan Node 3.1

Gambar 6 merupakan pohon keputusan dengan akar cabang terakhir karena setiap *node* cabang sudah habis atau mempunyai keputusan masing-masing. Pohon keputusan tersebut sudah dapat menghasilkan *rule* dan *knowledge*.

Berdasarkan pohon keputusan terakhir yang terbentuk pada Gambar 6, maka *rule-rule* keputusan yang dihasilkan dijabarkan pada Tabel 8.

Tabel 8. Daftar Aturan (Rule) Pohon Keputusan Akhir

No	Aturan (Rule)
1	IF UH = A THEN Keputusan Memuaskan
2	IF UH = B AND Tugas = B THEN Keputusan Memuaskan
3	IF UH = B AND Tugas = D THEN Keputusan Kurang Memuaskan
4	IF UH = C AND Absensi = $\geq 70\%$ THEN Keputusan Memuaskan
5	IF UH = C AND Absensi = $< 70\%$ THEN Keputusan Kurang Memuaskan
6	IF UH = D AND Ujian = B THEN Keputusan Memuaskan
7	IF UH = D AND Ujian = C THEN Keputusan Memuaskan
8	IF UH = D AND Ujian = D THEN Keputusan Kurang Memuaskan

Penelitian yang dilakukan dengan Metode C4.5 terhadap data nilai hasil belajar siswa di masa pandemi menghasilkan *rule-rule* keputusan yang ditampilkan pada Tabel 8. *Rule-rule* keputusan tersebut bisa digunakan untuk memprediksi hasil belajar siswa pembelajaran daring suatu mata pelajaran pada masa pandemi COVID-19 di SMKN 2 Padang Panjang.

Selanjutnya dilakukan pengujian terhadap data nilai hasil belajar siswa menggunakan *software* Rapid Miner Studio 9.2.0. Pengujian dilakukan untuk mengetahui kecocokan hasil klasifikasi *output* dan pohon keputusannya.

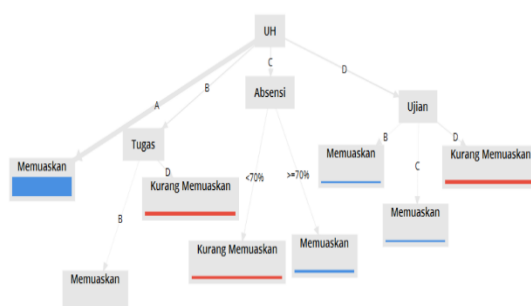
Tree

```

UH = A: Memuaskan {Memuaskan=18, Kurang Memuaskan=0}
UH = B
| Tugas = B: Memuaskan {Memuaskan=1, Kurang Memuaskan=0}
| Tugas = D: Kurang Memuaskan {Memuaskan=0, Kurang Memuaskan=3}
UH = C
| Absensi = <70%: Kurang Memuaskan {Memuaskan=0, Kurang Memuaskan=2}
| Absensi = >=70%: Memuaskan {Memuaskan=2, Kurang Memuaskan=0}
UH = D
| Ujian = B: Memuaskan {Memuaskan=1, Kurang Memuaskan=0}
| Ujian = C: Memuaskan {Memuaskan=1, Kurang Memuaskan=0}
| Ujian = D: Kurang Memuaskan {Memuaskan=0, Kurang Memuaskan=3}

```

Gambar 7. Klasifikasi *Output* Pengujian RapidMiner



Gambar 8. *DecisionTree* hasil RapidMiner 9.2.0

Rule-rule hasil pengujian RapidMiner Studio 9.2.0:

1. IF UH = A, THEN Keputusan Memuaskan

2. IF UH = B AND Tugas = B, THEN Keputusan Memuaskan
3. IF UH = B AND Tugas = D, THEN Keputusan Kurang Memuaskan
4. IF UH = C AND Absensi = $\geq 70\%$, THEN Keputusan Memuaskan
5. IF UH = C AND Absensi = $< 70\%$, THEN Keputusan Kurang Memuaskan
6. IF UH = D AND Ujian = B, THEN Keputusan Memuaskan
7. IF UH = D AND Ujian = C, THEN Keputusan Memuaskan
8. IF UH = D AND Ujian = D, THEN Keputusan Kurang Memuaskan

Pengujian dengan menggunakan RapidMiner Studio 9.2.0 menghasilkan klasifikasi *outputrule* dan pohon keputusan yang sama dengan hasil perhitungan manual C4.5. Klasifikasi *output* dan *rule* ditampilkan pada Gambar 7, sedangkan hasil pohon keputusan (*DecisionTree*) pada Gambar 8.

Setelah pengujian untuk *rule* pohon keputusan dilanjutkan dengan pengujian data *testing* menggunakan *Software* Rapid Miner. Data *testing* yang diuji sebanyak 30 data nilai siswa.

accuracy: 83.33%

	true Kurang Memuaskan	true Memuaskan	class precision
pred. Kurang Memuaskan	8	5	61.54%
pred. Memuaskan	0	17	100.00%
class recall	100.00%	77.27%	

Gambar 9. Hasil Pengujian Data *Testing*

Pengujian data *testing* dengan Rapid Miner Studio 9.2.0 berada pada Gambar 9.

1. *Accuracy*: 83,33%.
2. *Precision*: Prediksi “Kurang Memuaskan” 61,54%, Prediksi “Memuaskan” 100%.
3. *Class Recall*: “Kurang Memuaskan” 100%, “Memuaskan” 77,27%.

4. Kesimpulan

Berdasarkan penelitian dan pembahasan yang dilakukan dapat diambil kesimpulan yaitu: Klasifikasi Metode C4.5 terhadap nilai hasil belajar siswa menghasilkan *rule-rule* dan *knowledge* yang dapat digunakan untuk memprediksi hasil belajar siswa secara daring pada masa pandemi COVID-19. Pengujian yang dilakukan dengan *Software* RapidMiner Studio.9.2.0 mendapatkan akurasi sebesar 83,33%, dengan nilai akurasi ini maka sudah dapat digunakan sebagai rekomendasi kebijakan di sekolah.

Penelitian selanjutnya dapat dilakukan dengan data yang lebih banyak untuk mendapatkan *rule-rule* yang lebih bervariasi sehingga mendapatkan tingkat akurasi yang lebih baik.

Daftar Rujukan

- [1] Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Klik-Kumpulan Jurnal Ilmu Komputer*, 5(1), 27. DOI: <https://doi.org/10.20527/klik.v5i1.100> .
- [2] Buulolo, Efori. (2020). *Data Mining untuk Perguruan Tinggi*. Yogyakarta: Deepublish. 6.
- [3] Febianto, N. I., & Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 8(2), 130. DOI: <https://doi.org/10.32736/sisfokom.v8i2.653>
- [4] Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Konstruksi PT. Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1). DOI: <https://doi.org/10.15575/join.v2i1.71> .
- [5] Zhou, F., Xue, L., Yan, Z., & Wen, Y. (2020). Research on College Graduates Employment Prediction Model Based On C4.5 Algorithm. *Journal of Physics: Conference Series*, 1453. DOI: <https://doi.org/10.1088/1742-6596/1453/1/012033> .
- [6] Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3). DOI: <https://doi.org/10.30998/faktorexacta.v11i3.2777> .
- [7] David, D. (2019). Penerapan Educational Data Mining Untuk Memprediksi Hasil Belajar Siswa SMAK Ora et Labora. *Jurnal Ilmu Komputer*, 12(2). DOI: <https://doi.org/10.24843/jik.2019.v12i02.p02> .
- [8] Dengen, C. N., Kusriani, K., & Luthfi, E. T. (2020). Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah SISFOTENIKA*, 10(1). DOI: <https://doi.org/10.30700/jst.v10i1.484> .
- [9] Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5. *Jurnal Informasi dan Teknologi*, 2(1), 23–29. DOI: <https://doi.org/10.37034/jidt.v2i1.22> .
- [10] Wang, X., Zhou, C., & Xu, X. (2019). Application of C4.5 Decision Tree for Scholarship Evaluations. *Procedia Computer Science*, 151, 179–184. DOI: <https://doi.org/10.1016/j.procs.2019.04.027> .
- [11] Rolansa, F., Yunita, Y., & Suheri, S. (2020). Sistem Prediksi dan Evaluasi Prestasi Akademik Mahasiswa di Program Studi Teknik Informatika Menggunakan Data Mining. *Jurnal Pendidikan Informatika dan Sains*, 9(1). DOI: <https://doi.org/10.31571/saintek.v9i1.1696> .
- [12] Li, J., He, J., Liu, Z., Zhang, H., & Zhang, C. (2019). Traffic Accident Analysis Based on C4.5 Algorithm in WEKA. *MATEC Web of Conferences*, 272. DOI: <https://doi.org/10.1051/mateconf/201927201035> .
- [13] Meng, X., Zhang, P., & Zhang, D. (2020). Decision Tree for Online Voltage Stability Margin Assessment Using C4.5 and Relief-F Algorithms. *Energies*, 13(15). DOI: <https://doi.org/10.3390/en13153824> .
- [14] Kumar, R. (2020). Assessing Higher Education in COVID-19 Era. *Brock Education Journal*, 29(2). DOI: <https://doi.org/10.26522/brocked.v29i2.841> .
- [15] Cahyani, A., Listiana, I. D., & Larasati, S. P. D. (2020). Motivasi Belajar Siswa SMA pada Pembelajaran Daring di Masa Pandemi Covid-19. *IQ (Ilmu Al-Qur'an): Jurnal Pendidikan Islam*, 3(01), 123–140. DOI: <https://doi.org/10.37542/iq.v3i01.57> .
- [16] Siahaan, M. (2020). Dampak Pandemi Covid-19 Terhadap Dunia Pendidikan. *Jurnal Kajian Ilmiah*, 1(1), 73–80. DOI: <https://doi.org/10.31599/jki.v1i1.265> .