ISSN 2597-4645 (media online) ISSN 2597-4610 (media cetak) Page: 753-760

# PENERAPAN ALGORITMA C4.5 PADA KLASIFIKASI POTENSI SISWA DROP OUT

Muhamad<sup>1</sup>, Agus Perdana Windarto<sup>2</sup>, Suhada<sup>3</sup>

<sup>1</sup>Mahasiswa Program Studi Sistem Informasi STIKOM Tunas Bangsa Pematang Siantar <sup>2,3</sup>STIKOM Tunas Bangsa Pematangsiantar, Indonesia Email: 1 muhamadd41@gmail.com, 2 agus.perdana@amiktunasbangsa.ac.id, 3 suhada.atb@gmail.com

#### Abstrak

Siswa merupakan salah satu substansi yang perlu diperhatikan dalam kaitannya dengan dunia pendidikan saat ini. Sulitnya mendapatkan siswa membuat sekolah harus mengoptimalkan sistem pembelajaran dan juga sarana prasarana untuk menarik minat calon siswa baru dan juga membuat siswa yang sudah bersekolah di sekolah tersebut tidak putus sekolah atau dropout. Salah satu faktor penyebab banyaknya jumlah siswa dropout karena kurangnya kebijakan dan tindakan dari instansi pendidikan untuk menjaga siswanya dari dropout. Tujuan Penelitian ini adalah membuat klasifikasi siswa berpotensi dropout dan tidak berpotensi dropout dengan algoritma C4.5 sebagai acuan dalam membuat kebijakan dan tindakan untuk mengurangi jumlah siswa dropout. Hasil klasifikasi dari algoritma C4.5 dievaluasi dan divalidasi dengan RapidMinerStudio untuk mengetahui tingkat akurasi Algoritma C4.5 dalam membuat klasifikasi siswa berpotensi dropout.

Kata Kunci: Siswa dropout, Klasifikasi, Algoritma C4.5

#### **Abstract**

Students are one of the substances that need to be considered in relation to the world of education today. The difficulty of getting students makes the school have to optimize the learning system and infrastructure as well to attract the interest of new students and also make students who have gone to school not drop out or drop out. One of the factors contributing to the large number of students dropping out is because of the lack of policies and actions from the education institutions to keep their students from dropping out. The purpose of this study was to classify potentially dropout students and not have the potential to drop out with the C4.5 algorithm as a reference in making policies and actions to reduce the number of students dropping out. The classification results of the C4.5 algorithm are evaluated and validated with RapidMinerStudio to determine the accuracy of the C4.5 Algorithm in classifying potential dropouts.

Keywords: Student dropout, Classification, C4.5 Algorithm

### 1. PENDAHULUAN

Siswa merupakan salah satu substansi yang perlu diperhatikan dalam kaitannya dengan dunia pendidikan saat ini, karena pada sebuah sekolah terutama sekolah swasta siswa sangat berpengaruh dalam kemajuan sekolah tersebut yang berdasarkan dari jumlah siswa yang berada di sekolah tersebut. Dikarenakan dari bantuan yang diberikan pemerintah untuk sekolah yang berdasarkan dari jumlah siswa yang berarti semakin banyak siswa maka semakin besar pula bantuan yang didapatkan oleh sekolah. Dengan hal tersebut setiap sekolah terutama sekolah swasta berusaha semaksimal mungkin untuk mendapatkan siswa baru untuk masuk ke sekolah tersebut, namun banyak sekolah lain dengan keunggulannya masing-masing untuk mendapatkan seorang siswa itu membutuhkan usaha ekstra yang diperlukan.

Cukup besarnya jumlah siswa yang putus sekolah di MTs. Al-Hikmah Marihat Bandar yang disebabkan beberapa faktor seperti masalah kehadiran siswa, masalah ekonomi dan faktor-faktor lainnya. Dengan demikian pihak sekolah harus membuat sebuah kebijakan untuk mengurangi jumlah siswa drop uot seperti dengan cara mendeteksi siswa yang berpotensi drop out yang kemudian dapat diarahkan agar siswa tersebut tidak drop out. Pihak sekolah juga harus mengoptimalkan sistem pembelajaran dan juga sarana prasarana untuk menarik minat calon siswa baru dan juga membuat siswa yang sudah bersekolah di sekolah tersebut tidak putus sekolah atau dropout (DO). Dengan mendeteksi siswa yang berpotensi untuk putus sekolah dapat membantu sekolah lebih fokus memberikan pengarahan kepada siswa yang membutuhkan. Namun belum adanya sebuah program untuk mendeteksi siswa yang berpotensi putus sekolah secara otomatis ini mempersulit sekolah untuk menentukan siswa yang berpotensi putus sekolah.

Untuk membuat program pendeteksi siswa yang berpotensi putus sekolah ini dapat menggunakan hasil klasifikasi siswa yang berpotensi untuk putus sekolah ataupun aktif. Klasifikasi merupakan kegiatan dalam mengekstrak dan kemudian memprediksi label kategori untuk masing-masing data. Dengan kata lain, klasifikasi adalah proses pencarian model yang dapat membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu obyek yang belum diketahui kelasnya. Salah satu algoritma klasifikasi yang bisa digunakan adalah C4.5 [1]. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterprestasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik [2].

Untuk mengatasi permasalahan diatas dapat dilakukan dengan penelitian mendalam dengan membuat aplikasi data mining yang bertujuan untuk mengklasifikasi siswa yang berpotensi Drop Out atau tidak dengan menggunakan atribut yaitu absensi, penghasilan orang tua, jumlah tanggungan orang tua, dan jarak rumah ke sekolah. Dengan

Page: 753-760

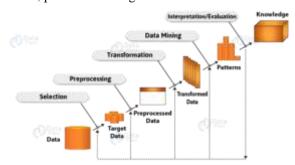
menggunakan algoritma c4.5 dikarenakan algoritma ini membuat klasifikasi lebih sederhana, mudah dan juga tingkat akurasi yang tinggi. Dan diharapkan dengan penelitian ini dapat mempermudah pihak sekolah untuk menentukan siswa yang berpotensi drop out.

#### 2. TEORITIS

### 2.1. Data Mining

Data Mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika [2][3].

Data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [4][5].



Gambar 1. Knowledge Data Discovery (KDD)

### 2.2.Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari Algoritma ID3 yang ditemukan oleh Ross Quinlan, Algoritma C4.5 digunakan untuk mengklasifikasikan data yang memiliki atribut berupa numerik ataupun kategorial, Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru [6].

Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (*decision tree*) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterprestasikan dan memiliki tingkat akurasi yang dapat diterima[7].

Menurut [2] Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- 1. Pilih atribut sebagai akar
- 2. Buat cabang untuk masing-masing nilai
- 3. Bagi kasus dalam cabang
- 4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

menghitung gain digunakan rumus seperti yang tertera berikut:
$$Gain(S,A) = Entropy(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(Si)$$
(1)

Keterangan:

S: Himpunan kasus

A: Atribut

n: Jumlah partisi atribut A

|Si|: Jumlah kasus pada partisi ke i

|S|: Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar

$$Entropy(S) = \sum_{i=1}^{n} -pi * log_2 pi$$
 (2)

S: Himpunan Kasus

n: Jumlah partisi S

pi: Proporsi dari Si terhadap S

#### 3. ANALISA DAN PEMBAHASAN

#### 3.1. Analisis Data

Dalam melakukan atau mendapatkan data, penulis menggunakan metode pengumpulan data yaitu metode porposive sampling. Dalam skripsi ini penulis mengumpulkan datanya menggunakan metode porposive sampling, yaitu metode penentuan sampel non-probabilitas dengan beberapa kriteria pertimbangan tertentu [3].

Pada penelitian ini, data yang digunakan untuk diklasifikasi adalah data siswa yang di peroleh dari data emis MTs Al-Hikmah Marihat Bandar T.P. 2016 / 2017

Tabel 1. Data Penelitian

No	NIS	Absensi	Penghasilan Orangtua	Jumlah Saudara	Jarak Rumah Ke Madrasah	Ket
1	141649	Baik	Cukup	3-4	Cukup	Tidak
2	141650	Kurang	Rendah	1-2	Jauh	Berpotensi
3	141655	Baik	Cukup	1-2	Cukup	Tidak
4	141656	Baik	Cukup	3-4	Dekat	Tidak
5	141657	Baik	Cukup	1-2	Cukup	Tidak
6	141658	Baik	Tinggi	3-4	Jauh	Tidak
7	141659	Kurang	Rendah	1-2	Jauh	Berpotensi
8	141660	Baik	Cukup	1-2	Cukup	Tidak
9	141665	Kurang	Cukup	1-2	Cukup	Berpotensi
10	141666	Cukup	Cukup	>4	Cukup	Berpotensi
					•••	
• • •	•••	•••	•••	•••	•••	•••
			•••	•••	•••	•••
170	151870	Kurang	Rendah	3-4	Cukup	Berpotensi
171	151871	Baik	Rendah	1-2	Cukup	Tidak
172	151872	Cukup	Cukup	1-2	Cukup	Tidak
173	151873	Baik	Tinggi	1-2	Cukup	Tidak

### 3.2. Analisa Algoritma C45

Berikut ini merupakan langkah-langkah perhitungan dalam algoritma C4.5 dalam penyelasaian kasus klasifikasi siswa berpotensi *drop* out yang akan di bagi menjadi label "iya" (berpotensi) dan "tidak" (tidak berpotensi).

### 1. Perhitungan Mencari Entropy

Proses pertama algoritma C4.5 adalah menentukan nilai *entropy*. Langkah pertama, tentukan *entropy* total kasus terlebih dahulu. Rumus untuk mencari *entropy* dari data siswa tersebut menggunakan persamaan(2) yaitu :

Entropy (Total) = 
$$\left(-\frac{53}{173} * \log_2\left(\frac{53}{173}\right)\right) + \left(-\frac{120}{173} * \log_2\left(\frac{120}{173}\right)\right) = 0.888925016$$

 Total Kasus
 Jumlah (Ya)
 Jumlah (Tidak)
 Entropy

 173
 53
 120
 0,888925016

Kemudian menghitung *Entropy* dari semua kasus yg terbagi berdasarkan atribut "Absensi", "Penghasilan Orangtua", "Jumlah Saudara", dan "Jarak rumah ke madrasah" menggunakan cara yang sama yaitu persamaan(2).

#### 2. Perhitungan Mencari Gain

Setelah mendapatkan hasil dari semua *entropy* langkah selanjutnya ialah menghitung *gain* dari setiap atribut menggunakan persamaan(1).

Gain (Total, Absensi) = 
$$0.888925016 - \left(\frac{106}{173} * 0\right) + \left(\frac{30}{173} * 0.996791632\right) + \left(\frac{37}{173} * 0\right) = 0.716070976$$
  
Gain (Total, Penghasilan Orangtua)

Gain (Total, Penghasilan Orangtua) = 
$$0.888925016 - \left(\frac{51}{173} * 0.897427193\right) + \left(\frac{101}{173} * 0.676162742\right) + \left(\frac{21}{173} * 0\right)$$

= 0,229611584

Gain (Total, Jumlah Saudara)

$$= 0,888925016 - \left(\frac{94}{173} * 0,658191266\right) + \left(\frac{65}{173} * 0,961236605\right) + \left(\frac{14}{173} * 0,591672779\right)$$

= 0,122255784

Gain (Total, Jarak rumah)

$$= 0.888925016 - \left(\frac{27}{173} * 0.764204507\right) + \left(\frac{107}{173} * 0.880145474\right) + \left(\frac{39}{173} * 0.961236605\right)$$

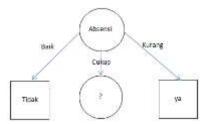
= 0.008593716

Berikut ini hasil dari perhitungan di atas :

Tabal 3	Tabal	Hacil	Perhitungar	ո 1
raber 5	. ranei	паѕп	Permuungai	11

AMDIDIA			TID A IZ (CA)	~	CATN
ATRIBUT	JML KASUS	YA (S1)	TIDAK (S2)	ENTROPY	GAIN
TOTAL	173	53	120	0,888925016	
ABSENSI					0,716070976
BAIK	106	0	106	0	
CUKUP	30	16	14	0,996791632	
KURANG	37	37	0	0	
PENGHASILAN					0,229611584
RENDAH	51	35	16	0,897427193	
CUKUP	101	18	83	0,676162742	
TINGGI	21	0	21	0	
SAUDARA					0,122255784
1-2	94	16	78	0,658191266	
3-4	65	25	40	0,961236605	
>4	14	12	2	0,591672779	
JARAK					0,008593716
DEKAT	27	6	21	0,764204507	
SEDANG	107	32	75	0,880145474	
JAUH	39	15	24	0,961236605	

Pada table di atas dapat kita lihat bahwa atribut absensi memiliki *gain* tertinggi yaitu 0,888925016, maka absensi menjadi *node* akar. Absensi memiliki 3 nilai yaitu baik, cukup dan kurang. Pada baik dan buruk telah mengklasifikasi kasus menjadi 1 yaitu keputusan "ya"dan "tidak" sedangkan untuk cukup masih diperlukan perhitungan lagi karena masih memiliki hasil antara "ya" dan "tidak", perhitungan tersebut dilakukan untuk menentukan *node* akar selanjutnya, maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut:



Gambar 2. Pohon Keputusan 1

Langkah selanjutnya ialah menghitung node akar cukup, caranya sebagai berikut :

Pilih *node* akar cukup kemudian hitung jumlah kasus, jumlah kasus keputusan "ya", jumlah kasus keputusan "tidak" dan *entropy* pada atribut penghasilan orangtua, jumlah saudara, dan jarak rumah ke madrasah seperti dibawah.

$$Entropy \text{ (Penghasilan, rendah)} = \left(-\frac{10}{10} * \log_2\left(\frac{10}{10}\right)\right) + \left(-\frac{0}{10} * \log_2\left(\frac{0}{10}\right)\right) = 0$$

$$Entropy \text{ (Penghasilan, cukup)} = \left(-\frac{6}{17} * \log_2\left(\frac{6}{17}\right)\right) + \left(-\frac{11}{17} * \log_2\left(\frac{11}{17}\right)\right) = 0,936667382$$

$$Entropy \text{ (Penghasilan, tinggi)} = \left(-\frac{0}{3} * \log_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3} * \log_2\left(\frac{3}{3}\right)\right) = 0$$

$$Entropy \text{ (Saudara, 1 - 2)} = \left(-\frac{0}{8} * \log_2\left(\frac{0}{8}\right)\right) + \left(-\frac{8}{8} * \log_2\left(\frac{8}{8}\right)\right) = 0$$

$$Entropy \text{ (Saudara, 3 - 4)} = \left(-\frac{8}{13} * \log_2\left(\frac{8}{13}\right)\right) + \left(-\frac{5}{13} * \log_2\left(\frac{5}{13}\right)\right) = 0,961236605$$

$$Entropy \text{ (Saudara, > 4)} = \left(-\frac{8}{9} * \log_2\left(\frac{8}{9}\right)\right) + \left(-\frac{1}{9} * \log_2\left(\frac{1}{9}\right)\right) = 0,503258335$$

$$Entropy \text{ (Jarak rumah, dekat)} = \left(-\frac{0}{5} * \log_2\left(\frac{9}{5}\right)\right) + \left(-\frac{5}{5} * \log_2\left(\frac{5}{5}\right)\right) = 0$$

$$Entropy \text{ (Jarak rumah, cukup)} = \left(-\frac{11}{16} * \log_2\left(\frac{11}{16}\right)\right) + \left(-\frac{5}{16} * \log_2\left(\frac{5}{16}\right)\right) = 0,8960382332$$

$$Entropy \text{ (Jarak rumah, jauh)} = \left(-\frac{5}{9} * \log_2\left(\frac{5}{9}\right)\right) + \left(-\frac{4}{9} * \log_2\left(\frac{4}{9}\right)\right) = 0,99107606$$

$$Gain \text{ (Absensi} = \text{Cukup, Penghasilan orangtua)} = 0,996791632 - \left(\frac{10}{30} * 0\right) + \left(\frac{17}{30} * 0,936667382\right) + \left(\frac{3}{30} * 0\right) = 0,466013449$$

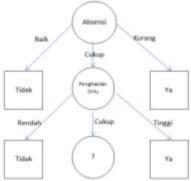
$$Gain \text{ (Absensi} = \text{Cukup, Jumlah Saudara)} = 0,996791632 - \left(\frac{8}{30} * 0\right) + \left(\frac{13}{30} * 0,961236605\right) + \left(\frac{9}{30} * 0,503258335\right)$$

Gain (Absensi = Cukup, Jarak Rumah) = 0,996791632 -  $\left(\frac{5}{30} * 0\right)$  +  $\left(\frac{16}{30} * 0,896038233\right)$  +  $\left(\frac{9}{30} * 0,99107606\right)$ = 0.221581757

Tabel 4. Tabel Hasil Perhitungan 2

Atribut	Jml Kasus	Ya (S1)	Tidak (S2)	Entropy	Gain
ABSENSI =CUKUP	30	16	14	0,996791632	
PENGHASILAN					0,304864554
Rendah	11	10	1	0,439496987	
Cukup	17	6	11	0,936667382	
Tinggi	2	0	2	0	
SAUDARA					0,58025577
1-2	9	0	9	0	
3-4	13	8	5	0,961236605	
>4	8	8	0	0	
JARAK					0,221581757
Dekat	5	0	5	0	
Cukup	16	11	5	0,896038233	
Jauh	9	5	4	0,99107606	

Pada tabel diatas *gain* tertinggi pada absensi = cukup adalah penghasilan = 0,466013449 maka menghasilkan pohon keputusan sebagai berikut :



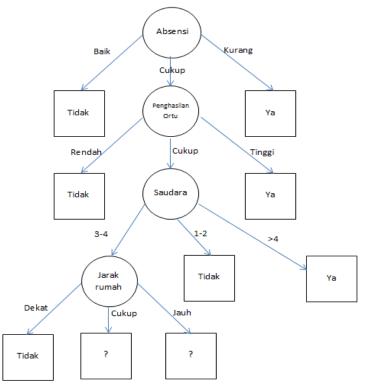
Gambar 3. Pohon Keputusan 2

Karena cukup masih beri Ya 'dan ''Tidak'' maka diperlul Tidak ngan kembali.  $Entropy \text{ (saudara, } 1-2) = \left(-\frac{0}{8}*\log_2\left(\frac{0}{8}\right)\right) + \left(-\frac{8}{8}*\log_2\left(\frac{8}{8}\right)\right) = 0$   $Entropy \text{ (saudara, } 3-4) = \left(-\frac{3}{6}*\log_2\left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6}*\log_2\left(\frac{3}{6}\right)\right) = 1$   $Entropy \text{ (saudara, } > 4) = \left(-\frac{3}{3}*\log_2\left(\frac{3}{3}\right)\right) + \left(-\frac{0}{3}*\log_2\left(\frac{0}{3}\right)\right) = 0$   $Entropy \text{ (Jarak, dekat)} = \left(-\frac{0}{4}*\log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4}*\log_2\left(\frac{4}{4}\right)\right) = 0$   $Entropy \text{ (Jarak, cukup)} = \left(-\frac{4}{8}*\log_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8}*\log_2\left(\frac{4}{8}\right)\right) = 1$   $Entropy \text{ (Jarak, jauh)} = \left(-\frac{2}{5}*\log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5}*\log_2\left(\frac{3}{5}\right)\right) = 0,970950594$  Gain (absensi = cukup &penghasilan orangtua = cukup, saudara  $= 0,936667382 - \left(\frac{8}{17}*0\right) + \left(\frac{6}{17}*1\right) + \left(\frac{3}{17}*0\right) = 0,583726205$  Gain (absensi = cukup &penghasilan orangtua = cukup, jarak  $= 0,936667382 - \left(\frac{4}{17}*0\right) + \left(\frac{8}{17}*1\right) + \left(\frac{5}{17}*0,970950594\right) = 0,180505442$ 

Tabel 5. Tabel Hasil Perhitungan 3

Atribut	Jml Kasus	Ya (S1)	Tidak (S2)	Entropy	Gain
Absensi=cukun Penghasilan orangtua	17	6	11	0.936667382	

= Cukup						
SAUDARA	A					0,583726205
	1-2	8	0	8	0	
	3-4	6	3	3	1	
	>4	3	3	0	0	
JARAK						0,180505442
	Dekat	4	0	4	0	
	Cukup	8	4	4	1	
	Jauh	5	2	3	0,970950594	



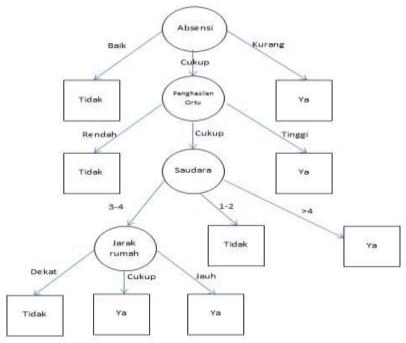
Gambar 4. Pohon Keputusan 3

Karena cukup dan jauh pada jarak rumah masih terdapat nilai "ya" dan "tidak" maka diperlukan perhitungan lagi dengan memfilter data sebagai berikut :

Tabel 6. Tabel Hasil Perhitungan 5

ruser of ruser ruser remittinguit s					
Atribut	Jml Kasus	Ya (S1)	Tidak (S2)		
Absensi = Cukup ,Penghasilan Orangtua = Cukup,	3	3	0		
Saudara= 3-4	3	3	V		
Jarak Rumah					
Cukup	3	3	0		
Jauh	0	0	0		

Dari tabel diatas dapat kita ketahui bahwa nilai dari cukup = ya dan jauh tidak memiliki data maka diambil kesimpulan bahwa jarak rumah yang bernilai jauh = ya dikarenakan nilai atribut jauh lebih besar dari "cukup", dengan demikian seluruh cabang sudah memiliki nilai masing-masing, maka perhitungan pohon keputusan sudah selesai.



Gambar 5. Hasil Pohon Keputusan

## 3.3.Penerapan Algoritma C4.5 Pada RapidMiner Studio 9.1

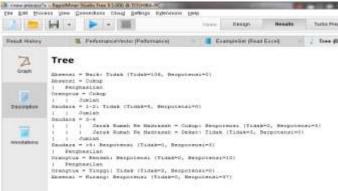
1. Tampilan decision tree



Gambar 6. hasil decision tree

Setelah dilakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir.

2. Tampilan deskripsi decision tree



Gambar 7. Deskripsi decision tree

Gambar di atas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (decision tree) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa penggunaan data mining algoritma C4.5 baik digunakan dalam proses menggali data (data mining process) untuk menarik beberapa

ISSN 2597-4645 (media online) ISSN 2597-4610 (media cetak) Page: 753-760

kesimpulan yang divisualisasikan dengan pohon keputusan (decision tree). Berikut ini rules yang dihasilkan dari pohon keputusan tersebut:

#### Tabel 7. *Rule* yang dihasilkan

#### Rule

- Jika Absensi Baik Maka Tidak Berpotensi
- Jika Absensi Kurang Maka Berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Tinggi Maka Tidak Berpotensi 3.
- Jika Absensi Cukup dan Penghasilan Orangtua Rendah Maka Berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Cukup dan Jumlah Saudara 1-2 Maka tidak berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Cukup dan Jumlah Saudara >4 Maka berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Cukup dan Jumlah Saudara 3-4 dan Jarak Rumah Jauh Maka berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Cukup dan Jumlah Saudara 3-4 dan Jarak Rumah Cukup Maka berpotensi
- Jika Absensi Cukup dan Penghasilan Orangtua Cukup dan Jumlah Saudara 3-4 dan Jarak Rumah Dekat Maka tidak berpotensi

### 4. KESIMPULAN

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada Penerapan Algoritma C4.5 Pada Klasifikasi Siswa Berpotensi Drop Out untuk dapat disimpulkan sebagai berikut.

- 1. Permasalahan menentukan siswa berpotensi drop out dapat diselesaikan menggunakan teknik data mining, yaitu dengan Algoritma C4.5. menghasilkan 9 (sembilan) rules dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 96,15% dan nilai AUC (Area Under the ROC Curve) 0,998.
- 2. Dengan adanya penerapan data mining algoritma C4.5 diharapkan mampu memberikan solusi dalam menentukan siswa yang berpotensi *drop out* di MTs. Al-Hikmah Marihat Bandar.

### 5. REFERENCES

- [1] A. H. Nasrullah, Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out, Ilk. J. Ilm., Vol. 10, Pp. 244-
- [2] D. H. Kamagi And S. Hansun, Implementasi Data Mining Dengan Algoritma C4 . 5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa, Ultimatics, Vol. Vi, No. 1, Pp. 15-20, 2014.
- A. P. Windarto, Implementation Of Data Mining On Rice Imports By Major Country Of Origin Using Algorithm Using K-Means Clustering Method, Int. J. Artif. Intell. Res., Vol. 1, No. 2, Pp. 26–33, 2017.
- [4] M. Ridwan, H. Suyono, And M. Sarosa, Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, J. Eeccis, Vol. 7, No. 1, Pp. 59-64, 2013.
- [5] A. P. Windarto, Implementation Of Data Mining On Rice Imports By Major Country Of Origin Implementation Of Data Mining On Rice Imports By Major Country Of Origin Using Algorithm Using K-Means Clustering Method, Int. J. Artif. Intelegence Res., Vol. 1, No. 2, Pp. 26-33, 2017.
- [6] E. Elisa, Analisa Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi Pt. Arupadhatu Adisesanti, J. Online Inform., Vol. 2, No. 1, Pp. 36–41, 2017.
- [7] Y. S. Luvia, D. Hartama, A. P. Windarto, And Solikhun, Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa, Phys. Rev. Mater., Vol. 1, No. Issn 2527-5771, P. 7, 2016.