

## DATA MINING POTENSI AKADEMIK SISWA BERBASIS ONLINE

<sup>1</sup>DIDIK SETIYADI, <sup>2</sup>ALI NURDIN

<sup>1,2</sup>Sekolah Tinggi Manajemen Informatika dan Komputer Eresha

Program Studi : Teknik Informatika

Jl. H. Samali No.51, Kalibata – Jakarta Selatan

<sup>1</sup>Email : [didik.setiyadi@eresha.ac.id](mailto:didik.setiyadi@eresha.ac.id)

<sup>2</sup>Email : [Ali\\_nurdin1120@yahoo.com](mailto:Ali_nurdin1120@yahoo.com)

***Abstract:** Problem-solving school difficulties in determining the classification of the student's academic potential can be sought through a system of student's academic potential predictive analysis. The results of predictive analysis are useful to carry out the enrichment program and the improvements program in preparation for the race of academics. Systems analysis of this prediction using data mining to determine the classification of students with academic potential classification models through techniques that form the decision tree of C4.5 algorithm. Predicted results are in the form of the rule of academic potential of students who subsequently entered into an online-based system of academic potential. The data on the student report book information academic potential of students who set and approved by the school. The rule is tested the prediction that yielding a prediction of 77.78% and then applied to other data that is testing the data as much as 20 report book data that generate 90% prediction rate. After the rules are received then put into an online-based system of academic potential.*

***Keywords:** Academic Potential Students, Data Mining, Information Systems Online, C4.5 algorithm*

### 1. PENDAHULUAN

Visi Kemendiknas 2014 adalah : “Terselenggaranya Layanan Prima Pendidikan Nasional untuk Membentuk Insan Indonesia Cerdas Komprehensif”, yang merupakan bagian dari lima tahap Rencana Pembangunan Pendidikan Nasional Jangka Panjang (RPPNJP) 2005-2025 khususnya tema pembangunan II (2010-2015) dengan fokus pada penguatan pelayanan. Untuk mencapai visi Kemendiknas 2014, maka ditetapkan misi Kemendiknas 2010-2014 dikemas dalam *Misi 5K* sebagai berikut : meningkatkan *ketersediaan* layanan pendidikan, meningkatkan *keterjangkauan* layanan pendidikan, meningkatkan *kualitas/mutu dan relevansi* layanan pendidikan, meningkatkan *kesetaraan* dalam memperoleh layanan pendidikan, dan meningkatkan *kepastian/keterjaminan* memperoleh layanan pendidikan. Salah satu *Misi 5K* Kemendiknas 2010-2014 adalah meningkatkan kualitas/mutu dan relevansi layanan pendidikan dan salah satu bagian dari isi Peraturan Mendiknas RI No. 19 Tahun 2007 tentang Standar Pengelolaan Pendidikan, sekolah dituntut untuk mengelola sistem informasi manajemen yang memadai untuk mendukung administrasi pendidikan yang efektif, efisien dan akuntabel.

Keterkaitan Peraturan Mendiknas RI No. 19 Tahun 2007, SMP Negeri 2 Tambun Selatan belum memiliki sistem kehadiran dan penilaian secara terintegrasi yang ada hanya dokumen – dokumen terpisah yang berbentuk format pengisian. Dokumen yang berbentuk format pengisian tidak tersimpan secara elektronik sehingga sering sekali memakan waktu yang lama dalam memprosesnya. Dengan bentuk format terpisah

dan tidak terintegrasi maka menyulitkan sekolah untuk memberikan informasi kehadiran dan penilaian siswa kepada orang tuanya.

Berbagai perlombaan akademis sering sekali menyulitkan sekolah mencari siswa yang berpotensi untuk mengikuti perlombaan akademis. Dengan adanya kelompok siswa yang memiliki potensi akademik maka mudah bagi sekolah untuk memberikan penguatan dan pengayaan materi untuk menghadapi berbagai perlombaan akademis. Sekolah juga mengalami kesulitan dalam memprediksi siswa yang berpotensi akademik yang nantinya akan menjadikan suatu keputusan tentang profil siswa yang berpotensi akademik. Untuk mendukung keputusan tentang profil siswa, lahirlah teknik pencari pengetahuan yang dinamakan *Knowledge Discovery in Database (KDD)* (Kusrini, 2009). Data pada rapor siswa merupakan data yang diakui oleh sekolah untuk informasi potensi akademik siswa yang terdiri dari data penilaian mata pelajaran, data ketidakhadiran, data sikap dan data keaktifan ekstrakurikuler yang coba dijadikan sebagai data yang akan diproses untuk menentukan klasifikasi siswa yang berpotensi dan belum berpotensi akademik.

Pada penelitian tahun 2004 oleh Kalles & Pierrakeas (dikutip dalam Al-Radaideh Q. et al. 2006 :h.2) menggunakan model klasifikasi dengan teknik pohon keputusan sebagai teknik klasifikasi adalah cara yang baik dan cepat untuk menghasilkan aturan – aturan yang sederhana. Pada penelitian tersebut menggunakan variasi atribut dari 9 atribut sampai 2 atribut untuk memprediksi hasil kemampuan pada siswa yang mengambil program pelajaran C++ dengan klasifikasi kemampuan A, B, C, D. Dengan demikian penelitian tersebut menggunakan variasi atribut minimal menggunakan 2 atribut dengan model klasifikasi dan teknik pohon keputusan. Teknik klasifikasi data mining meliputi : *decision tree classifier, rule-based classifier, neural network, support vector machine dan naïve bayes classifier* (Ernawati, 2008) dimana *decision tree classifier* dipilih karena banyak kelebihan – kelebihan (Moertini, 2007:18) dan algoritma pembentuk pohon keputusan C4.5 (Kusrini, 2009). Harapan yang akan dicapai dalam penelitian ini adalah mewujudkan suatu sistem analisa potensi akademik siswa yang digunakan untuk menginformasikan nilai dan kehadiran siswa secara *online*, dan mendukung pengambil keputusan dalam menyelesaikan masalah potensi akademik siswa menggunakan klasifikasi C4.5.

## 2. TINJAUAN PUSTAKA

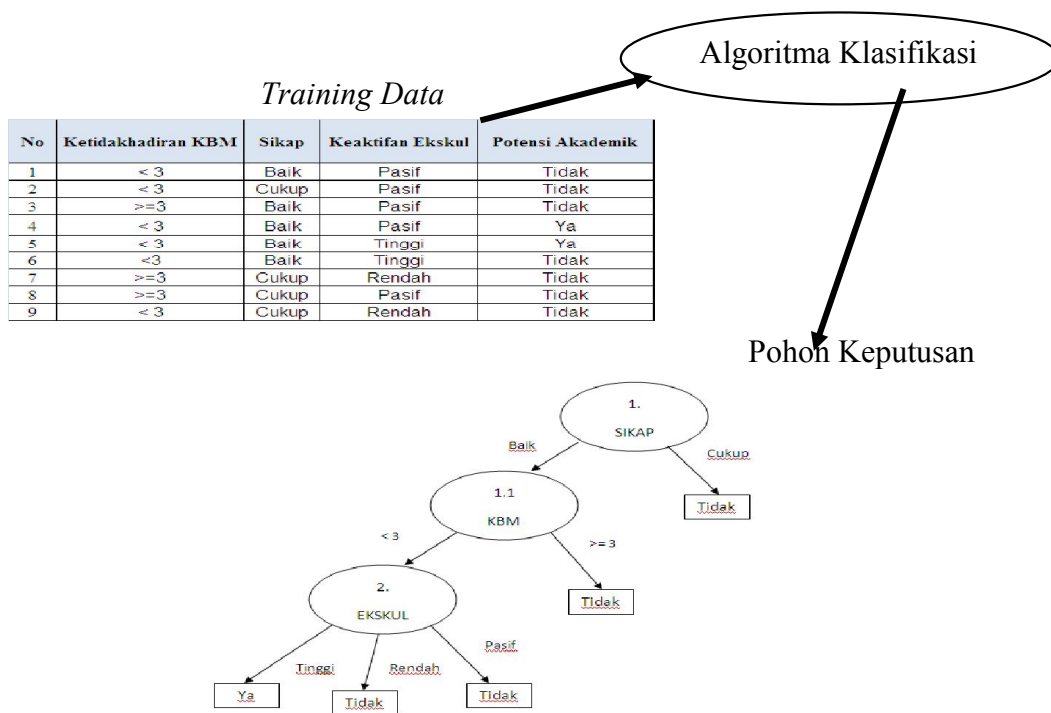
### 2.1 *Data Mining*

Pengertian *data mining* menurut Turban, dkk adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Kusrini 2009:3). Sedangkan menurut Han dan Kamber *Data mining* merupakan proses pencarian pola dan relasi – relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan melakukan klasifikasi, estimasi, prediksi, *association rule*, deskripsi dan visualisasi (Ernawati 2008:4). Dengan demikian *data mining* digunakan untuk mengekstraksi data siswa dan data nilai siswa menjadi klasifikasi informasi potensi akademik siswa yang ada potensi dan yang belum ada potensi berprestasi.

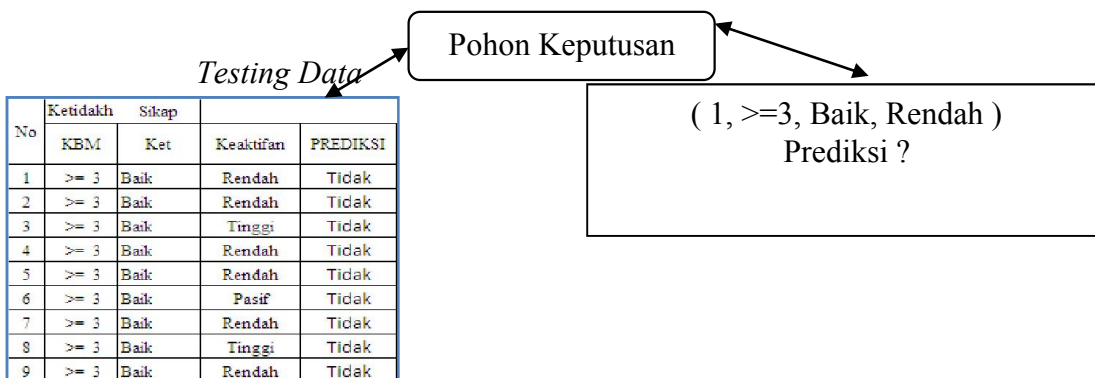
### 2.2 *Klasifikasi*

Klasifikasi dan prediksi adalah dua bentuk analisa data yang bisa digunakan untuk mengekstrak model dari data yang berisi kelas – kelas atau untuk memprediksi *trend* data yang akan datang (Ernawati, 2008:5). Klasifikasi memprediksi data yang

terdiri sejumlah *records* yang salah satu *attributes* adalah suatu kelas atau kategori. Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan potensi akademik siswa dalam dua kategori, yaitu kategori ada potensi dan kategori belum ada potensi. Sedangkan prediksi memodelkan fungsi – fungsi variabel target berdasarkan nilai variabel prediksi. Sebagai contoh, prediksi siswa berprestasi pada tiap semester. Klasifikasi merupakan proses yang terdiri dari dua tahap, yakni tahap pembelajaran dan tahap pengklasifikasian. Proses dua tahapan tersebut dikenal dengan dua model klasifikasi, yaitu model pembelajaran (*learn model*) dan model penerapan (*apply model*). Model pembelajaran dipandang sebagai tahap pembentukan fungsi  $y = f(X)$  dimana  $y$  adalah kelas hasil prediksi dan  $X$  adalah *tuple* yang ingin diprediksi kelasnya. Pengertian *tuple* adalah baris dalam *record* yang mana berasal dari *training data* untuk menghasilkan model pembelajaran.



Gambar 1. Learn Model



Gambar 2. Apply Model

Beberapa teknik klasifikasi yang digunakan adalah *decision tree classifier*, *rule-based classifier*, *neural-network*, *support vector machine*, dan *native bayes classifier* (Ernawati, 2008:7). Teknik klasifikasi yang dipilih adalah *decision tree classifier* atau teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan (Moertini, 2007:18). Proses klasifikasi data meliputi : tahap pembelajaran dan tahap pengujian dimana pada tahap pembelajaran didapatkan model pohon keputusan serta aturan – aturannya, sedangkan pada tahap pengujian diperoleh hasil keterangan variabel target dari *record – record* yang belum diketahui variabel targetnya. Bahasa algoritma yang digunakan untuk pembentuk pohon keputusan menurut Larose antara lain ID3, CART dan C4.5. Algoritma C4.5 dipilih karena algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Kusrini, 2009:14).

### 2.3 Algoritma C4.5

Algoritma C4.5 adalah algoritma pembentuk struktur pohon keputusan dimana simpul – simpul dan cabang – cabang terbentuk. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

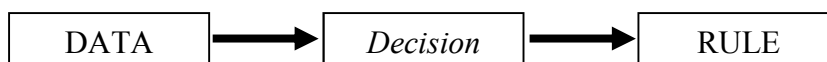
1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap – tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai simpul akar didasarkan pada nilai *gain* tertinggi dari atribut – atribut yang ada (Kusrini, 2009:15).

**Tabel 1.** Potensi Akademik Siswa

No	Ketidakhadiran KBM	Sikap	Keaktifan Ekskul	Potensi Akademik
1	< 3	Baik	Pasif	Tidak
2	< 3	Cukup	Pasif	Tidak
3	>= 3	Baik	Pasif	Tidak
4	< 3	Baik	Pasif	Ya
5	< 3	Baik	Tinggi	Ya
6	< 3	Baik	Tinggi	Tidak
7	>= 3	Cukup	Rendah	Tidak
8	>= 3	Cukup	Pasif	Tidak
9	< 3	Cukup	Rendah	Tidak

Konsep pohon keputusan adalah mengubah data menjadi pohon keputusan (*decision tree*) dan aturan – aturan keputusan (*rule*).



**Gambar 3.** Konsep Pohon Keputusan

Proses dalam *decision tree* adalah :

1. Mengubah bentuk data ( tabel ) menjadi model *tree*.
2. Mengubah model *tree* menjadi *rule*
3. Menyederhanakan *rule*

Untuk mewujudkan konsep pohon keputusan, maka harus mengubah tabel menjadi model pohon keputusan. Dalam membuat pohon keputusan dimulai dari simpul akar sampai berakhir ke simpul daun. Penentuan simpul akar dimulai dengan menghitung nilai *entropy*, *entropy (total)* dan *gain total*. Konsep *entropy(S)* adalah jumlah bit yang diperlukan untuk mengkodekan kelas (+ atau -) dari anggota S yang diambil acak (Gambetta, 2003: h.8).

Perhitungan nilai *entropy(S)* dapat dilihat pada persamaan (1), sebagai berikut :

$$Entropy(S) = -p_+ \log_2 p_+ - p_- \log_2 p_- \dots\dots\dots(1)$$

Keterangan :

$p_+$  : jumlah proporsi positif (+) di S.  
 $p_-$  : jumlah proporsi negatif (-) di S.

Sedangkan simpul akar ditentukan berdasarkan *gain* tertinggi dari atribut – atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti dalam persamaan 2 sebagai berikut (Kusrini, 2009: h.16) :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots(2)$$

Keterangan :

S : himpunan kasus  
 A : atribut  
 n : jumlah partisi atribut A  
 $|S_i|$  : jumlah kasus pada partisi ke-i  
 $|S|$  : jumlah kasus dalam S

### 3. METODOLOGI PENELITIAN

Bentuk penelitian yang dilakukan penulisan didalam melakukan penelitian adalah studi kasus pada SMP Negeri 2 Tambun Selatan. Metode penelitian yang digunakan adalah dengan metode deskriptif dengan menggambarkan keadaan yang sesungguhnya sesuai dengan apa yang dilihat pada saat penelitian dilakukan. Pendekatan yang digunakan adalah dengan. Sedangkan pendekatan yang digunakan dalam data maning ini adalah metode klasifikasi dengan teknik decision tree classifier dengan menggunakan algoritma C4.5. Teknik pengumpulan data yang digunakan dalam melakukan penelitian ini adalah dengan Wawancara, Observasi dan Studi Dokumentasi. Data yang diolah dalam keperluan data mining adalah data nilai mata pelajaran, data sikap siswa, data keaktifan ekstrakurikuler dan data ketidakhadiran siswa dalam buku rapor siswa. Siswa termasuk kategori memiliki potensi akademik jika nilai rata – rata mata pelajarannya  $\geq 80$ . Sedangkan yang nilai rata-rata mata pelajaran  $< 80$  adalah siswa dalam kategori belum memiliki potensi akademik. Variabel target adalah data nilai rata-rata mata pelajaran, sedangkan variabel kategori adalah data ketidakhadiran siswa, data sikap siswa dan data keaktifan ekstrakurikuler. *Record* siswa dan nilai yang digunakan adalah dari rapor siswa kelas 8 H Tahun pelajaran 2009 / 2010 sebanyak tiga puluh lima siswa sebagai *training data*. Sedangkan data dari rapor siswa kelas 7.8 Tahun Pelajaran 2008/2009 sebanyak dua puluh siswa sebagai *testing data*.

#### 4. HASIL PENELITIAN DAN PEMBAHASAN

Pembahasan aplikasi data mining dapat memberikan pengambil keputusan dalam penyelesaian masalah potensi akademik siswa ini didasari oleh konsep KDD (*Knowledge Discovery in Database*) yang didalamnya terdapat proses pembentukan pohon keputusan menggunakan algoritma C4.5.

##### 1. Tahapan Pertama

Dari data raport siswa sebanyak 35 ( tiga puluh lima ) siswa didapat variabel-variabel yaitu : variabel ketidakhadiran KBM, variabel sikap, variabel keaktifan ekskul sebagai variabel kategori dan variabel rata-rata nilai sebagai variabel target. Variabel ketidakhadiran KBM membentuk data kontinu yaitu :  $< 3$  hari dan  $\geq 3$  hari. Variabel sikap siswa membentuk data ordinal yaitu : Baik, Cukup, Kurang. Variabel keaktifan ekskul membentuk data ordinal yaitu : Tinggi, Rendah, Pasif. Variabel rata-rata nilai mata pelajaran membentuk data kontinu, yaitu : YA dan TIDAK.

**Tabel 2.** Hasil Seleksi Variabel dan Transformasi Variabel

No	Ket Prestasi	Ketidakhadiran KBM	Sikap Ket	Keaktifan Ekskul
1	Tidak	$< 3$	Baik	Pasif
2	Tidak	$< 3$	Cukup	Pasif
3	Tidak	$\geq 3$	Baik	Pasif
4	Ya	$< 3$	Baik	Pasif
5	Ya	$< 3$	Baik	Tinggi
6	Tidak	$\geq 3$	Baik	Pasif
7	Tidak	$\geq 3$	Baik	Pasif
8	Tidak	$< 3$	Baik	Pasif
9	Ya	$< 3$	Baik	Tinggi
10	Tidak	$< 3$	Baik	Pasif
11	Tidak	$< 3$	Baik	Pasif
12	Tidak	$\geq 3$	Baik	Pasif
13	Tidak	$< 3$	Baik	Pasif
14	Tidak	$< 3$	Baik	Pasif
15	Tidak	$< 3$	Baik	Pasif
16	Tidak	$\geq 3$	Cukup	Rendah
17	Tidak	$\geq 3$	Cukup	Pasif
18	Tidak	$\geq 3$	Baik	Pasif
19	Tidak	$< 3$	Baik	Pasif
20	Tidak	$< 3$	Cukup	Rendah
21	Tidak	$\geq 3$	Cukup	Pasif
22	Tidak	$< 3$	Cukup	Pasif
23	Ya	$< 3$	Baik	Pasif
24	Tidak	$\geq 3$	Cukup	Pasif
25	Tidak	$< 3$	Cukup	Pasif
26	Tidak	$< 3$	Baik	Pasif
27	Tidak	$\geq 3$	Cukup	Pasif
28	Tidak	$< 3$	Baik	Pasif
29	Tidak	$< 3$	Baik	Pasif
30	Ya	$< 3$	Baik	Pasif
31	Tidak	$< 3$	Baik	Pasif
32	Tidak	$< 3$	Baik	Pasif
33	Tidak	$< 3$	Cukup	Pasif
34	Tidak	$< 3$	Baik	Pasif
35	Ya	$< 3$	Baik	Pasif

##### 2. Tahapan Kedua

Data – data dari tahapan pertama yang sebanyak 35 (tiga puluh lima) siswa dimasukkan ke tahapan *pre-processing data* untuk mendapatkan data yang tidak

terduplikasi. Hasil data aplikasi *pre-processing* mendapatkan 9 ( sembilan ) data siswa yang tidak terduplikasi.

**Tabel 3.** Hasil *Pre-Processing Data*

Ketidakhadiran KBM	Sikap	Keaktifan Ekskul	Potensi Akademik	Prediksi	Skor Kesesuaian
< 3	Baik	Pasif	Tidak	Tidak	1
< 3	Cukup	Pasif	Tidak	Tidak	1
≥ 3	Baik	Pasif	Tidak	Tidak	1
< 3	Baik	Pasif	Ya	Tidak	0
< 3	Baik	Tinggi	Ya	Ya	1
< 3	Baik	Tinggi	Tidak	Ya	0
≥ 3	Cukup	Rendah	Tidak	Tidak	1
≥ 3	Cukup	Pasif	Tidak	Tidak	1

Jumlah Sesuai  
Jumlah Data  
Prediksi

### 3. Tahapan Ketiga

Tahapan ketiga sebagai tahapan data mining melalui proses pembentukan pohon keputusan menggunakan algoritma C4.5 yang memiliki tahapan sebagai berikut :

- a. Data dari tahapan kedua yang sebanyak sembilan data dimasukkan ke aplikasi tabel data penentuan akar pohon. Pada tabel tersebut didapatkan data variabel target, jumlah partisi dan jumlah kasus yang akan dijadikan sebagai data masukkan aplikasi data mining. Aplikasi data mining ini menghasilkan variabel SIKAP sebagai akar pohon keputusan.
- b. Pada cabang akar variabel SIKAP yang bernilai Cukup telah membentuk kelas TIDAK, sedangkan cabang Baik dihitung kembali. Nilai variabel SIKAP dan nilai Baik dimasukkan ke aplikasi tabel data penentuan cabang pohon untuk menentukan data nilai target, jumlah partisi dan jumlah kasus yang selanjutnya digunakan sebagai data masukkan aplikasi data mining. Aplikasi data mining selanjutnya menghasilkan variabel Ketidakhadiran KBM sebagai simpul pohon keputusan yang berada pada cabang Baik. Pada cabang simpul Ketidakhadiran KBM terdapat cabang  $\geq 3$  telah membentuk kelas Tidak, sedangkan cabang  $< 3$  ditentukan selanjutnya.
- c. Pada cabang  $< 3$  dari variabel Ketidakhadiran KBM dimasukkan ke tabel penentuan cabang pohon untuk menentukan nilai target, jumlah partisi dan jumlah kasus. Berdasarkan variabel yang ada ternyata setelah variabel Ketidakhadiran KBM hanya terdapat variabel Ekskul. Penentuan kelas dari variabel Ekskul bisa ditentukan berdasarkan dua kondisi klasifikasi target YA dan TIDAK. Dengan demikian cabang Tinggi yang telah membentuk kelas YA berdasarkan konsep pembelajaran, berarti cabang lainnya membentuk kelas TIDAK.

**Tabel 4.** Hasil Data Penentuan Akar Pohon

Variabel Atribut	Nilai Atribut	Variabel Target		Jumlah Kasus pada partisi ke-i	Jumlah Kasus Dalam S
		Ya	Tidak		
Ketidakhadiran KBM	< 3	2	4	6	9
	>=3	0	3	3	
Sikap	Baik	2	3	5	9
	Cukup	0	4	4	
Keaktifan Ekskul	Tinggi	1	1	2	9
	Rendah	0	2	2	
	Pasif	1	4	5	
Variabel Target				Total	
		Ya	Tidak		
		2	7	9	

4. Tahapan Keempat

Pada tahapan ini dalam proses KDD sebagai tahapan interpretasi dari *rule – rule* yang dihasilkan. *Rule* yang dihasilkan adalah : IF Sikap = Baik DAN Ketidakhadiran KBM < 3 DAN Keaktifan Ekskul = Tinggi THEN “Ya” IF NOT THEN “Tidak”. Interpretasi dari *rule* adalah : “Jika siswa memiliki sikap yang baik dan jumlah ketidakhadiran siswa kurang dari tiga hari dan keaktifan ekskulnya tinggi maka siswa tersebut memiliki potensi akademik berprestasi”.

Setelah proses data mining dilakukan selanjutnya adalah menguji *rule-rule* melalui konsep *learn model* dan konsep *apply model*. Pada konsep *learn model* adalah menghasilkan *rule-rule* dari tabel data *pre-processing* atau dikenal dengan *training data*. Sedangkan konsep *apply model* adalah menerapkan *rule-rule* yang dihasilkan *training data* ke tabel data lain yakni *testing data*. Pada *testing data* dimasukkan kolom yang bernama kolom prediksi yang berisi rumus *rule - rule* contoh : =IF(AND(C5="A",D5="Baik",E5="Tinggi"),"YA","TIDAK). Berdasarkan *testing data* yang dilakukan prediksinya menghasilkan 90%.

Tingkat prediksi didapatkan dari fungsi rumus *rule-rule* pohon keputusan sebagai berikut : =IF(AND(C5="A",D5="Baik",E5="Tinggi"),"Ya","Tidak")

No	Ketidakhadiran KBM	Sikap	Keaktifan Ekskul	Potensi Akademik	Prediksi	Skor Kesesuaian
1	< 3	Baik	Pasif	Tidak	Tidak	1
2	< 3	Cukup	Pasif	Tidak	Tidak	1
3	>=3	Baik	Pasif	Tidak	Tidak	1
4	< 3	Baik	Pasif	Ya	Tidak	0
5	< 3	Baik	Tinggi	Ya	Ya	1
6	< 3	Baik	Tinggi	Tidak	Ya	0
7	>=3	Cukup	Rendah	Tidak	Tidak	1
8	>=3	Cukup	Pasif	Tidak	Tidak	1
9	< 3	Cukup	Rendah	Tidak	Tidak	1
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
30					Jumlah Sesuai	7
31					Jumlah Data	9
32					Prediksi	77,78%
33						

**Gambar 4.** Hasil Tingkat Prediksi Penerapan Rule-Rule Pohon Keputusan



**Mencari Simpul Pohon Keputusan Dengan Algoritma C4.5**

Posisi Baris (maks 12)

Posisi Kolom (maks 2)

Nilai Data

1. Harus Diawali Dengan Mengisi Data Nilai Variabel Target  
2. Isilah Data Nilai Variabel Kategori Berdasarkan Baris dan Kolom

Variabel Kategori		Jml Nilai Atribut	Entropy Nilai Atribut	Entropy Terpilih	Gain (Total)
Yes	No				
0	4	0	0,318235818426175	0	KBM
0	3	0	0,97095057690894	0	<input type="button" value="Hitung Gain Total 1"/>
0	4	0	0	0	
0	1	0	0	0	Sikap
0	4	0	0,721928084147098	0	<input type="button" value="Hitung Gain Total 2"/>
0	0	0	0	0	
0	0	0	0	0	Ekskul
0	0	0	0	0,721928084147098	<input type="button" value="Hitung Gain Total 3"/>
0	0	0	0	0	
0	0	0	0	0	<input type="button" value="Hitung Gain Total 4"/>
0	0	0	0	0	
0	0	0	0	0	<input type="button" value="Hitung Gain Total 5"/>
0	0	0	0	0	

**Variabel Target**

Total Yes  Total No

**Sikap**

ALI NURDIN Konsentrasi : SIM

NPM : 341208710

Gambar 5. Hasil Proses Data Mining Untuk Memilih Akar Pohon

**Mencari Simpul Pohon Keputusan Dengan Algoritma C4.5**

Posisi Baris (maks 12)

Posisi Kolom (maks 2)

Nilai Data

1. Harus Diawali Dengan Mengisi Data Nilai Variabel Target  
2. Isilah Data Nilai Variabel Kategori Berdasarkan Baris dan Kolom

Variabel Kategori		Jml Nilai Atribut	Entropy Nilai Atribut	Entropy Terpilih	Gain (Total)
Yes	No				
2	2	4	1	1	KBM
0	1	1	0	0	<input type="button" value="Hitung Gain Total 1"/>
1	1	2	1	0	
0	0	0	0	0	Ekskul
1	2	3	0,918295818426175	0,918295818426175	<input type="button" value="Hitung Gain Total 2"/>
0	0	0	0	0	
0	0	0	0	0	<input type="button" value="Hitung Gain Total 3"/>
0	0	0	0	0	
0	0	0	0	0	<input type="button" value="Hitung Gain Total 4"/>
0	0	0	0	0	
0	0	0	0	0	<input type="button" value="Hitung Gain Total 5"/>
0	0	0	0	0	

**Variabel Target**

Total Yes  Total No

**KBM**

ALI NURDIN Konsentrasi : SIM

NPM : 341208710

Gambar 6. Hasil Proses Data Mining Untuk Memilih Simpul 1

## 5. KESIMPULAN

Berdasarkan analisa data, pengujian data dan pembahasan hasil penelitian maka dapat ditarik kesimpulan, yaitu aplikasi data mining melalui metode klasifikasi dengan teknik pohon keputusan dan algoritma C4.5 dapat mendukung dalam mengelompokkan siswa yang memiliki potensi akademik, melalui : a). Data siswa yang berupa nilai rapor, ketidakhadiran, sikap dan keaktifan dalam kegiatan ekstrakurikuler; dan b) Penyusunan *rule - rule* pohon keputusan sesuai dengan persyaratan siswa yang

berpotensi akademik. Sistem informasi data mining untuk siswa berpotensi akademik berbasis online dapat memberikan dukungan bagi orang tua dalam mengevaluasi potensi putra dan putrinya, dikarenakan Kemudahan dalam melakukan akses informasi dan akurasi yang tinggi dari algoritma C4.5.

#### DAFTAR PUSTAKA

- Kusrini & Luthfi, E. 2009, *Algoritma Data Mining*. Andi Offset, Yogyakarta.
- Hamzah, B. 2007, *Teori Motivasi & Pengukurannya*. Bumi Aksara, Jakarta.
- Ernawati, Iin. 2008, *Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 dan K-Nearest Neighbor*. Tesis Magister Sains, Institut Pertanian Bogor.
- Moertini, S. 2007, *Pengembangan Skalabilitas Algoritma klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi*, Disertasi Doctoral, Institut Teknologi Bandung.
- Simarmata, J. 2010, *Rekayasa WEB*. Andi Offset, Yogyakarta.
- Riduwan. 2009, *Pengantar Statistika Sosial*. Alfabeta, Bandung.
- Puwanto, N. 1995, *Psikologi Pendidikan*. Rosdakarya, Bandung.
- Arikunto, S. 2002, *Prosedur Penelitian*. Rineka Cipta, Jakarta.
- Lee, F & Santana, J 2010, *Data Mining : Meramalkan Bisnis Perusahaan*. Elex Media Komputindo, Jakarta.
- Moertini, S 2002, *Data Mining Sebagai Solusi*. Integral. Vol. 7 (1), hh. 44-56.
- Al-Radaideh, Q., Al-Shawakfa M., & Al-Najjar M. 2006, *Mining Student Data Using Decision Trees*, ACIT'2006, pp. 1-5.
- Varsha N., Singh A., Singh D.,& Jain 2010, *Result Analysis Using Classification Techniques*. International Journal of Computer Applications. Vol. 1 No. 22. pp. 22-26.
- Moertini, S.,2003, *Towards The Use Of C4.5 Algorithm For Classifying Banking Dataset.*, Integral. Vol. 8 (2), pp. 105-115.
- Gambetta, W. 2003, *Pohon Keputusan*, Materi Kuliah Informatika, IF5032, Institut Teknologi Bandung.
- Tan, Stenbach & Kumar 2004, *Pengantar Data Mining*, Catatan Kuliah Untuk Bab 4, Universitas Gunadarma.