

PREDIKSI PRESTASI SISWA SMP NURUL JADID MENGUNAKAN ALGORITMA C4.5

KAMIL MALIK¹⁾, MOCHAMMAD FAID²⁾

^{1,2)}Sekolah Tinggi Teknologi Nurul Jadid

e-mail: nomer1001@email.com¹⁾, ahmad.faid08@email.com²⁾

ABSTRAK

Penelitian ini dilakukan untuk membangun sistem prediksi Prestasi siswa yang lebih akurat sehingga dapat memudahkan dalam pengambilan keputusan untuk pembinaan siswa lebih dalam lagi terkait prestasi yang ada pada Lembaga Smp Nurul Jadid tersebut. Pengumpulan data primer dilakukan berdasarkan wawancara langsung dengan pihak Kepala Sekolah Lembaga Nurul Jadid. Analisis dilakukan dengan melakukan training pada data beberapa tahun 2015/2016 untuk membangun model C4.5. Karenanya, dapat disimpulkan bahwa data mining dengan teknik C4.5 dapat menghasilkan prediksi prestasi yang lebih tepat sehingga dapat membantu Lembaga Smp Nurul Jadid dalam pengambilan keputusan.

Kata Kunci: C4.5, Absensi, prestasi, prediksi dan Klasifikasi

ABSTRACT

This study was conducted to establish the prediction system, student performance more accurately so as to facilitate decision-making in coaching the students more deeply related to the existing achievements of the Central Junior High School the Nurul Jadid. Primary data were collected by direct interview with the Principal Institute of Nurul Jadid. The analysis is done by training on the data several years 2015/2016 to build model C4.5. Therefore, it can be concluded that data mining techniques can produce prediction C4.5 more precise achievement that could help Junior High School Institute Nurul Jadid in decision making.

Keywords: C4.5, attendance, achievement, prediction and classification.

I. PENDAHULUAN

Dunia Pendidikan adalah dunia untuk melahirkan penurus bangsa sekaligus menjadi sebuah wahana pembentuk karakter bangsa, sekolah merupakan ruang utama dimana para “Nation Builders”. Indonesia di inginkan akan berjuang mengambil negeri beradu dikancah global. Seiring dengan derasnya tantangan global, tantandan dunia pendidikan menjadi semakin besar, aspek ini yang mendorong para peserta didik meraih prestasi paling baik.

Salah satu factor yang menunjang prestasis siswa adalah kerajinan, sarana yang memadai, nilai dibangku sekolah dan juga pendidik yang bermutu. untuk meningkatkan mutu para guru, kemendikbud bakal meningkatkan kualifikasi guru lewat beasiswa S-1 bagi guru SD dan SMP. Hamid memaparkan, jumlah guru SD disekolah Negara dan Swasta lebih kurang 1.850 ribu guru. Dari jumlah tersebut Cuma 60% guru yang telah memenuhi kualifikasi. Tiap tahunnya kemendikbud pun menyiapkan beasiswa untuk 100 ribu calon guru guna menempuh pendidikan S-1 lewat pertolongan beasiswa S-1 untuk guru SD dan SMP. Di dunia internasional, mutu pendidikan Indonesia berada di peringkat ke-64 dari 120 negeri disemua dunia berdasarkan laporan tahunan UNESCO Education For All Global Monitoring 2012. Sedangkan berdasarkan indeks perkembangan pendidikan (Education Development Index, EDI), Indonesia berada pada

peringkat ke-69 dari 127 negeri terhadap 2011.[1]

Untuk masalah kualitas pendidik yang bermutu sudah diusahakan oleh kemendikbud agar siswa nantinya bisa berprestasi dan bisa bersaing di era Global saat ini, tapi yang menjadi permasalahan lain yaitu masalah terkadang siswa yang rajin tidak berprestasi mengapa hal seperti ini bisa terjadi untuk itu dibutuhkan kajian lebih mendalam lagi tentang hal ini, dan juga sering kita jumpa Data Mining C4.5 ada siswa yang memiliki nilai raport yang diatas rata-rata tidak lulus ujian nasional, untuk mencari pola mengapa hal semacam ini bisa terjadi maka diperlukan sebuah penelitian yang focus terhadap masalah ini dengan menggunakan metode Data mining yaitu C4.5 dengan ranking selection[2].

Data mining dalam penelitian ini hanya berfungsi untuk mencari sebuah pola agar kita mengetahui apa yang menjadi kendala terhadap siswa untuk bisa berprestasi, hal ini dilakukan karena adanya data yang diIjin/Kepentingan cukup untuk menjadi sampel dalam penelitian. Sampel data yang digunakan dalam penelitian adalah data yang di SMP Nurul Jadid dari tahun 2015-2016 dari kelas 7 sampai kelas 8.

II. BAHAN DAN METODE

A. Bahan

Bahan yang digunakan adalah data yang ada di SMP Nurul Jadid, Dimana data diperoleh dari Database Absensi siswa dan database raport siswa, Database

absensi siswa adalah bagian dari Sistem Informasi Perizinan siswa berbasis SMS Gateway, sedangkan database raport siswa didapat dari Sistem Informasi Akademik dan sistem informasi tersebut berjalan selama 2 tahun sampai sekarang. Dari data tersebut diharapkan mendapatkan pola hubungan antara absensi dengan prestasi siswa (prestasi ini dilihat dari ranking 10 besar di dalam kelas), adapun hasil dari pengolahan anantara data absensi dengan prestasinya nampak pada gambar 1.

didalam data mining pola tersebut disebut dengan rule, selain untuk mengetahui pola penelitian ini juga mempertimbangkan akurasi dan eror dari rule yang dihasilkan, dari dalam hal ini algoritma yang digunakan adalah algoritma C4.5[3]

Software yang digunakan untuk mempecroleh rule beserta akurasinya adalah weka Library yang diimport kedalam java dengan menggunakan editor netbeans,

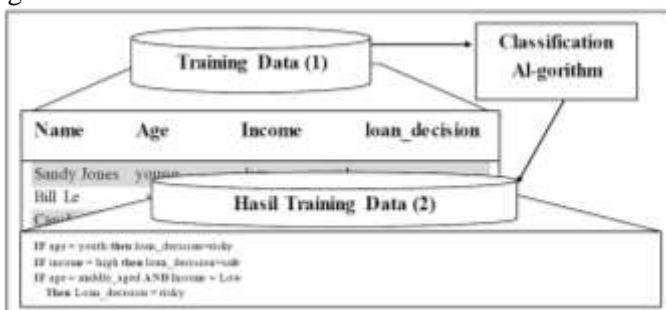
No	Nama	Kelas	Ujian	Ujian	Ujian	Ujian	Ujian
1	ABDI WAHID	VI A	4	5	19	B	TEBAK
2	ABIN HERMANGAL OTENDI	VI A	5	7	2	C	TEBAK
3	AHMAD RIZAL FIKRI	VI A	1	2	2	B	TEBAK
4	AHMAD ADAM BACHRIAN	VI A	0	2	2	B	TEBAK
5	AHMAD ALFAN ARI PAMUNGKAS	VI A	1	7	1	C	VA
7	AHMAD FANIS AEFIN	VI A	20	11	1	C	TEBAK
8	AHMAD RAMA ANALLIH	VI A	1	2	2	B	VA
9	AHMAD SAMUEL MAYARI	VI A	1	12	2	C	TEBAK
10	AHMAD YOGA PENGESTU	VI A	2	2	2	B	TEBAK
11	AHMAT MAULANA RAHMA	VI A	0	5	2	B	TEBAK
12	AHMAD ABDULKARIM NURANI	VI A	8	2	7	B	TEBAK
13	ANDRI PRATNO	VI A	1	0	0	B	VA
14	BAGUS PRINAN KURNIAWAN	VI A	1	1	0	C	VA
15	DINDA PRISTIA PERMANGSAH	VI A	0	0	0	C	TEBAK

Gambar 1 Data Siswa Berprestasi

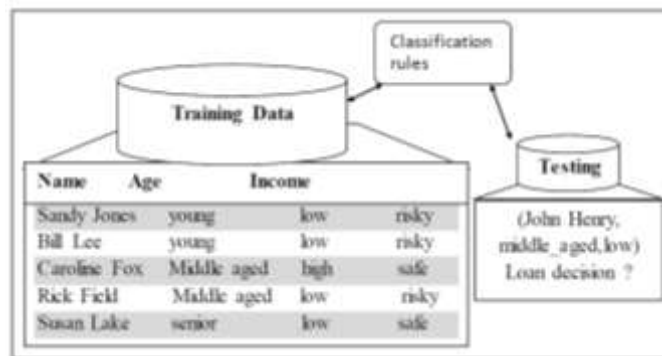
Untuk menganalisa data tersebut maka diperlukan sebuah software yang bisa menghitung akurasi dari sebuah prediksi prestasi siswa berdasarkan data raport dan absensi, adapun software yang digunakan dalam penelitian ini adalah weka 3.8.

B. Metode C4.5

Data yang diperoleh akan dianalisa oleh algoritma klasifikasi dan menghasilkan sejumlah aturan, kemudian tahap kedua yaitu *training* adalah menguji akurasi dari algoritma klasifikasi dengan memasukkan sejumlah data. Kedua tahap tersebut terlihat pada gambar 2.



Gambar 2 Tahap *learning* Data Training Dianalisa oleh Algoritma Klasifikasi[3]



Gambar 3 Tahap Klasifikasi Menguji Akurasi dari Algoritma[3]

C4.5 diperkenalkan Quinlan (1996) sebagai versi terbaru dari ID3. Dalam induksi tree hanya bisa dilakukan pada fitur bertipe katagorikal, sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan decision tree[2][3].

Algoritma C4.5 juga menggunakan kreteria dalam menentukan fitur yang menjadi pemecah pada pohon yang diinduksi[3][4], Algoritma untuk C4.5 Bisa dilihat dibawah ini:

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

$$Entropy(A) = \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

Keterangan :

- S : Himpunan Kasus
- N : Jumlah Partisi atribut A
- |S| : Jumlah Kasus Dalam S
- A : Atribut
- |Si| : Jumlah Kasus pada partisi ke i

III. PEMBAHASAN

Pada penelitian, peringkasan abstraksi yang dibuat merupakan sistem peringkasan berupa sebuah dokumen dan secara otomatis menghasilkan ringkasan.[4],

1. Membangun Decision Tree dengan C4.5

Dalam kasus yang tertera pada tabel 3, akan dibuat Decision Tree untuk menentukan Prediksi prestasi siswa dengan melihat keadaan Sakit, Ijin, keadaan Sikap. langkah-langka dalam membangun Decision Tree dengan C4.5[4] Perhatikan Tabel 3[5].

Tabel 3.1 Sampel Data Siswa

No	Sakit	Ijin	Alpha	Sikap	Prestasi
1	4	5	19	B	TIDAK
2	5	7	1	C	TIDAK
3	1	0	1	B	TIDAK
4	0	1	1	B	TIDAK
5	1	7	1	C	YA
6	10	11	6	C	TIDAK
7	1	0	0	B	YA
8	1	12	1	C	TIDAK

Secara umum algoritma c4.5 untuk membangun pohon keputusan adalah sebagai berikut :[3]

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.

Tabel 3.2 Hasil Perhitungan Node 1

No	Atribut	Value		Prestasi	Tidak Prestasi	E	Gain
1	Total	1006	:	298	708	0.87	0,80
	Sakit	>2	:	43	117	0.83	
		<=2	:	255	591	0.88	
2	Ijin	>0	:	115	266	0.88	0,80
		<=0	:	183	469	0.85	
3	Alphs	>0	:	98	356	0.75	0,81
		<=0	:	200	352	0.94	
4	Sikap	A	:	110	64	1,17	0,82
		B	:	152	386	0.94	
		C	:	36	258	1.04	

Baris Total kolom entropy pada tabel 2.8 dihitung dengan rumus 2-3,

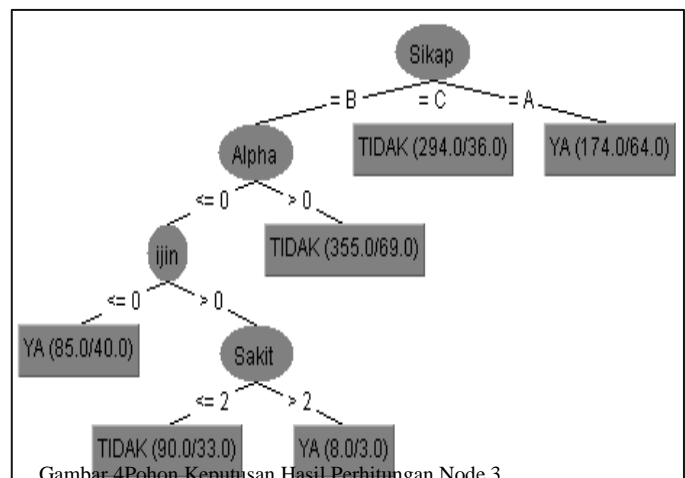
$$Entropy(Total) = \left(-\frac{298}{1006} \log_2\left(\frac{298}{1006}\right)\right) + \left(-\frac{708}{1006} \log_2\left(\frac{708}{1006}\right)\right) = 0,87$$

Sedangkan nilai gain pada baris Sakit dihitung dengan rumus C4.5 [3][4], sebagai:

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

$$Gain(Total, Sakit) = 0,87 - \left(\frac{160}{1006} * 0.83\right) + \left(\frac{846}{1006} * 0.83\right) = 0,809441131$$

dari hasil pada tabel 3.2 dapat diketahui bahwa atribut dengan gain tertinggi adalah Sikap yaitu sebesar 0.82. dengan demikian Sikap dapat menjadi Node akar. Dari pohon keputusan, terus lakukan perhitungan sampai yang terbentuk seperti pada gambar 4.[4]



Gambar 4 Pohon Keputusan Hasil Perhitungan Node 3

IV. HASIL

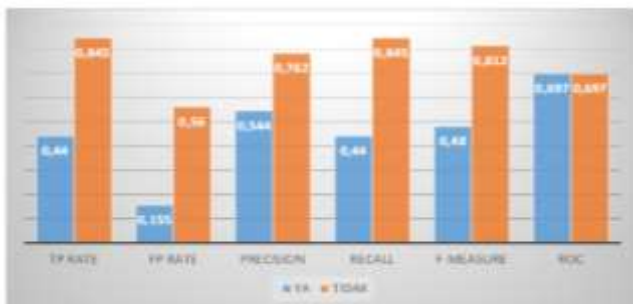
Penelitian ini melakukan uji coba dataset absensi dan raport siswa sebanyak 1006 sebagai sample untuk dimodelkan. Proses pemodelan diawali dengan melakukan feature selection[6]. Tahap feature selection dilakukan dengan dua tahap yaitu rank-ing selection. Atribut yang di-filter diperoleh dari proses ranking selection[4] dimana 2 buah atribut paling bawah yang tidak berpengaruh akan diabaikan dalam proses selanjutnya, Proses feature selection yang dilakukan pada tools Weka dengan 10 cross-validation diperoleh 2 atribut terbawah yang tidak terlalu berpengaruh terhadap akurasi perhitungan. Dalam kasus ini, 2 atribut yang diabaikan, yaitu: Nomer dan kelas

Dengan begitu atribut yang digunakan dalam penelitian ini mencoba untuk mengurangi dengan 2 atribut dan membandingkan hasilnya.

Tabel 7 Attribute selection 10 fold

Average merit	Average rank	Rank	Attribute
0.067	2.4	7	Sikap
0.058	3.4	5	Sakit
0.055	3.4	4	Ijin
0.045	4.8	6	Alpha
0.0004	6	3	Kelas
0	7	1	No

Nilai *accuracy*, *precision*, dan *recall* dari data training dapat dihitung dengan menggunakan Weka 3.6.9, Setelah diuji coba dengan metode *cross-validation*, didapatkan hasil pengukuran terhadap data training yaitu *accuracy*= 72.4652% pada hal sebelum atribut dikurangi tingkat akurasi adalah sebesar 70.377%.



Gambar 5 Detail akurasi tanpa attribute Kelas dan Nomer

V. KESIMPULAN

Penelitian ini telah menyelesaikan:

1. Menentukan Prediksi Siswa, Metode yang digunakan C4.5/J48 dan menghasilkan akurasi sebesar 72.4652% namun ketika attribute yang kurang berpengaruh diabaikan tingkat akurasi menurun sebesar 2%.
2. Atribut Sikap memiliki pengaruh yang sangat signifikan dalam hal prestasi di lembaga SMP Nurul jadid ini dibuktikan bahwa setiap siswa yang bersikap baik dan rajin masuk sekolah ada kemungkinan siswa tersebut akan berprestasi
3. Kehadiran juga memiliki pengaruh setelah attribute sikap, hal ini dibuktikan dengan Tree yang terbentuk sikap yang baik belum tentu berprestasi

jika absensi dan presensinya melebihi batas kewajarannya.

VI. DAFTAR PUSTAKA

- [1] P. I. a. 08.22, "Potret Dunia Pendidikan di Indonesia," in *PendidikanIndonesia.Com*, 2015.
- [2] M. K. J. P. Jiawei Han, *Data Mining Concepts and Techniques*, San Francisco: Morgan Kaufmann, 2006.
- [3] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.
- [4] A. B. V. T. Miranda Lakshmi, *An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data*, Coimbatore, India.: MECS, 2013.
- [5] F. D. M. Mambang, *PREDIKSI CALON MAHASISWA BARU MENGGUNAKAN METODE KLASIFIKASI DECISION TREE*, Banjarmasin: LPPM Universitas Potensi Utama, 2015.
- [6] A. Azar, *A model for personnel selection with a data mining*, Iran: AOSIS, 2013.