

Penerapan *Data Mining* Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan metode *Decision Tree*

Sriagustina¹Nyimas Sopiah, MM., M. Kom² Susan dianpurnamasari, M. Kom³

¹ Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Bina Darma Palembang, Indonesia

sriagustina.ayiek@gmail.com¹ Nyimas.sopiah@binadarma.ac.id² Susandian@binadarma.ac.id³

Abstrak. Penelitian ini dilakukan untuk memprediksi tingkat kelulusan siswa sekolah dasar Negeri 8 Merapi Barat dengan menggunakan metode *decision tree*. Data yang digunakan pada penelitian ini adalah data siswa dengan menggunakan kriteria nilai ujian harian, nilai ujian tengah semester, nilai ujian akhir semester dan nilai ujian nasional. Data akan dikelola menggunakan *tool* WEKA untuk mempermudah memprediksi tingkat kelulusan siswa. Penerapan *data mining* digunakan untuk memprediksi tingkat kelulusan dengan menggunakan metode *decision tree* dan *algoritma C4.5* sebagai pendukung sekaligus perbandingan untuk mengetahui informasi tinggi rendahnya siswa sekolah dasar negeri 8 merapi barat yang lulus berdasarkan kedua metode tersebut.

Keyword: *Decision tree*, WEKA, *data mining* dan Algoritma C4.5

1. Pendahuluan

Datamining dapat digunakan untuk memprediksi kelulusan siswa Sekolah Dasar Negeri 8 Merapi Barat dan data siswa kelas VI akan dimanfaatkan untuk memprediksi tingkat kelulusan siswa Sekolah Dasar Negeri 8 Merapi Barat. Sekolah Dasar Negeri 8 Merapi Barat memiliki kurang lebih 800 siswa dalam setiap kelasnya memiliki kurang lebih 84 siswa dalam 2 tahun. Setiap tahun sekolah harus meluluskan paling sedikit 30 siswa. Disini pihak sekolah setiap tahun selalu mengawatirkan tentang kelulusan siswanya. Maka pada penelitian ini penulis tertarik untuk memprediksi kelulusan siswa Sekolah Dasar Negeri 8 Merapi Barat melalui pohon keputusan. Pada penelitian ini tools yang akan digunakan adalah WEKA (*Waikato Environment for Knowledge Analysis*)(www.cs.waikato.ac.nz/ml/weka). Diharapkan dengan menggunakan *data mining* yaitu dengan memanfaatkan data yang sebelumnya serta membandingkan antara nilai harian, nilai semester, nilai ujian sekolah serta nilai ujian nasional diharapkan dapat menentukan faktor mana yang menjadi penunjang paling dominan di dalam tingkat kelulusan seorang siswa.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis (Gonunescu, dalam Prasetyo 2014). *Data mining* merupakan istilah yang sering dikatakan sebagai suatu cara untuk menguraikan serta mencari penemuan berupa pengetahuan didalam suatu *database*. (Gonunescu, dalam Prasetyo 2014).

2.2 KDD (*Knowledge Discovery in Databases*)

Adapun untuk menganalisis data dalam penerapan *data mining* ini menggunakan proses tahapan *knowledge discovery in databases (KDD)* yang terdiri 1). Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional. 2). perlu dilakukan proses *cleaning* pada data yang menjadi focus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). 3). Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. 4). *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu 5). Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. (Sijabat, 2015:8)

Untuk melihat atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada, Entropy merupakan ukuran kemurnian suatu atribut sedangkan Gain merupakan pengurangan entropy yang disebabkan oleh partisi atribut. Untuk menghitung gain digunakan rumus seperti tertera dalam rumus 1:
 $Gain(S,A) = Entropy$

$$\sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

Dengan:

S: Himpunan kasus

A: Atribut

n: Jumlah partisi atribut A

[S_i]: jumlah kasus pada partisi ke i

[S]: jumlah kasus S

Sedangkan perhitungan nilai entropy dapat dilihat pada rumus 2 berikut:

Entropy(S)

$$\sum_{i=1}^a -p_i * \log_2 p_i$$

Dengan:

S: Himpunan Kasus

a:Fitur
I:JumlahpartisiS
Pi :ProporsidariSiterhadap

2.3 Variabel Kelulusan Siswa

Pada Sekolah Dasar Negeri 8 Merapi Barat terdapat beberapa mata pelajaran yang menjadi acuan dalam pemberian nilai yang akan berpengaruh pada hasil kelulusan siswa nantinya. Di setiap kemungkinan nilai siswa pada akhirnya akan membentuk pola atau jalur. Nilai yang dimaksud adalah nilai harian, nilai semester, ujian sekolah dan ujian nasional. Berikut ini merupakan kriteria dari kemungkinan jalur nilai yang terjadi pada Sekolah Dasar Negeri 8 Merapi Barat beserta dengan keterangan nilai untuk setiap atributnya.

Keterangan Atribut :

1) Nilai Harian (NH)

- a.) Rendah = 0 – 6,4
Bagi siswa yang mendapatkan *range* nilai ini maka siswa tersebut akan melakukan remedial dari masing-masing guru untuk perbaikan nilai yang lebih baik.
- b.) Menengah = 6,5 – 7,4
Bagi siswa yang mendapatkan *range* nilai ini maka akan dianggap sudah memenuhi standar nilai harian yang ditetapkan oleh sekolah dan mempunyai nilai yang cukup untuk melengkapi nilai kelulusan lainnya.
- c.) Tinggi = 7,5 – 10
Bagi siswa yang mendapatkan *range* nilai ini maka dianggap melebihi standar nilai harian yang telah ditetapkan sekolah dan mempunyai kemungkinan lulus yang lebih tinggi dibandingkan siswa yang mendapatkan nilai di bawah *range* nilai ini.

2.) Nilai Semester (NS)

- a.) Cukup Baik = 0 – 6,4
Bagi siswa yang mendapatkan *range* nilai ini maka siswa tersebut akan melakukan remedial dari masing-masing matapelajaran yang mendapatkan nilai yang masih di bawah rata-rata untuk perbaikan nilai yang lebih baik serta akan dilakukan pemanggilan orangtua siswa jika nilai ini didapatkan lebih dari satu matapelajaran terutama untuk matapelajaran yang akan menjadi matapelajaran yang termasuk ke dalam ujian nasional.
- a.) Baik = 6,5 – 7,4
Bagi siswa yang mendapatkan *range* nilai ini maka akan dianggap sudah memenuhi standar nilai semester yang ditetapkan oleh sekolah dan dapat membantu untuk menutupi kekurangan nilai pada matapelajaran yang lainnya.
- b.) Sangat Baik = 7,5 – 10
Bagi siswa yang mendapatkan *range* nilai ini maka dianggap melebihi standar nilai semester yang telah ditetapkan sekolah dan mempunyai kemungkinan tingkat kelulusan yang lebih tinggi dibandingkan siswa yang mendapatkan nilai di bawah *range* nilai ini.

3.) Ujian Sekolah (US)

a.) Belum Tercapai = 0 – 5,95
Bagi siswa yang mendapatkan *range* nilai ini maka akan dianggap belum tuntas dalam penguasaan materi dan memenuhi standar kelulusan sekolah.

b.) Tercapai = 5,96 – 10
Bagi siswa yang mendapatkan *range* nilai ini maka akan dianggap telah menguasai materi dan lulus dalam memenuhi syarat kelulusan nilai ujian dari sekolah.

4.) Ujian Nasional (UN)

a.) Di bawah standar : 0 – 5,5
Bagi siswa yang mendapatkan *range* nilai ini maka sudah bisa dipastikan bahwa siswa tersebut gagal dalam ujian nasional dan harus mengulang tahun depan untuk mengikuti ujian yang sama

b.) Memenuhi standar : 5,6 – 10
Bagi siswa yang mendapatkan *range* nilai ini maka sudah bisa dipastikan bahwa siswa tersebut telah lulus dalam mengikuti ujian

3. Hasil Dan Pembahasan

Langkah selanjutnya dilakukan perhitungan entropi total beserta gain dari setiap atribut untuk menentukan *node* pertama berdasarkan tabel data sebelumnya berdasarkan ketentuan dasar entropi sebagai berikut :

Tabel 4.3 Perhitungan Entropi Node Pertama

Atribut		Jumlah	Lulus	Tidak Lulus	Entropi
Total Nilai Semester		14	10	4	0,8632
	Cukup Baik	4	2	2	1
	Baik	6	4	2	0,9183
	Sangat Baik	4	4	0	0
Nilai Harian					
	Rendah	5	2	3	0,9710
	Menengah	5	4	1	0,7219
	Tinggi	4	4	0	0
Ujian Sekolah					
	Belum Tercapai	8	6	2	2,4152
	Tercapai	6	4	2	0,9183
Ujian Nasional					
	Dibawah Standar	7	3	4	0,9853
	Memenuhi Standar	7	7	0	0

Gain (Total - Nilai Semester

$$= 0,8632 - \left(\frac{4}{14} * 1\right) + \left(\frac{6}{14} * 0,9183\right) + \left(\frac{4}{14} * 0\right)$$

$$\begin{aligned} &= 0,8632 - \left(\frac{4}{14}\right) + \left(\frac{5,5098}{14}\right) + (0) \\ &= 0,2692 - (0,2857 + 0,3936 + 0) \\ &= 0,2692 - 0,6793 \\ &= 0,1839 \end{aligned}$$

Gain (Total – Nilai Harian)

$$\begin{aligned} &= 0,8632 - \left(\frac{5}{14} * 0,9710\right) + \left(\frac{5}{14} * 0,7219\right) + \left(\frac{4}{14} * 0\right) \\ &= 0,8632 - \left(\frac{4,855}{14}\right) + \left(\frac{3,6095}{14}\right) + (0) \\ &= 0,8632 - (0,3468 + 0,2578 + 0) \\ &= 0,8632 - 0,6046 \\ &= 0,2586 \end{aligned}$$

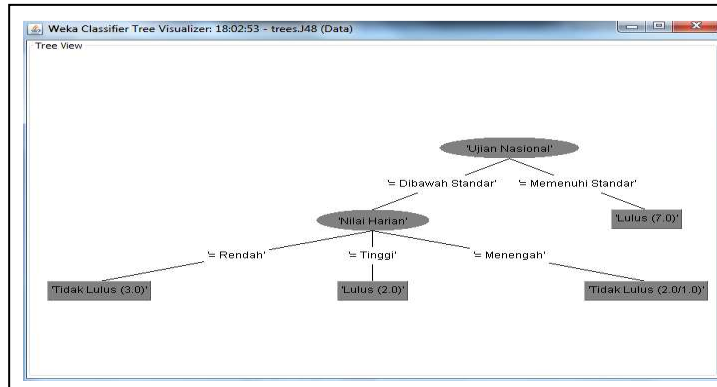
Gain (Total – Ujian Sekolah)

$$\begin{aligned} &= 0,8632 - \left(\frac{8}{14} * 2,4152\right) + \left(\frac{6}{14} * 0,9183\right) \\ &= 0,8632 - \left(\frac{19,3216}{14}\right) + \left(\frac{5,5098}{14}\right) \\ &= 0,8632 - (1,3801 + 0,3936) \\ &= 0,8632 - 1,7737 \\ &= -0,9105 \text{ (Tidak mendukung kriteria kelulusan)} \end{aligned}$$

Gain (Total – Ujian Nasional)

$$\begin{aligned} &= 0,8632 - \left(\frac{7}{14} * 0,9853\right) + \left(\frac{7}{14} * 0\right) \\ &= 0,8632 - \left(\frac{6,8971}{14}\right) + (0) \\ &= 0,8632 - (0,4927 + 0) \\ &= 0,8632 - 0,4927 \\ &= 0,3705 \end{aligned}$$

Pada pohon keputusan yang terbentuk nilai semester dan ujian sekolah tidak termasuk ke dalam pohon yang datanya diolah dari *tool* WEKA. Hasil ini sejalan antara pohon yang terbentuk dari perhitungan manual algoritma C4.5 dengan pohon yang terbentuk dari *tool* WEKA. Nilai ujian nasional dan nilai harian terpilih oleh algoritma C4.5 sebagai kriteria yang berpengaruh sehingga saling membentuk daftar aturan pembacaan pohon keputusan. Berdasarkan pohon keputusan inilah dapat diketahui bahwa ujian nasional sebagai *root* pohon merupakan kriteria yang paling menentukan hasil prediksi kelulusan siswa yang kemudian dapat di selidiki lagi lebih dalam bahwa nilai harian sebagai node pohon juga ikut mempunyai pengaruh yang sama terhadap hasil prediksi kelulusan siswa. Setiap kategori nilai dari nilai harian akan membentuk jalur aturan kondisi yang baru dalam menentukan prediksi hasil kelulusan siswa baik itu siswa yang lulus maupun yang tidak lulus.



Gambar 2. Pohon Keputusan Hasil Dari WEKA

4. Simpulan

Berdasarkan hasil dari penelitian yang dilakukan oleh penulis, maka dapat disimpulkan bahwa

1. Pengolahan *data mining* dengan menggunakan metode *decision tree* dan algoritma C4.5 menghasilkan sebuah pohon keputusan beserta dengan tingkat kelulusan masing-masing siswa lewat jalur pohon yang terbentuk di mana tingkat kelulusan siswa kelas VI Sekolah Dasar Negeri 8 Merapi Barat ini dibagi menjadi dua kriteria yaitu nilai ujian nasional dan nilai harian berdasarkan hasil pengolahan *data mining* pada *tool* WEKA. Hal ini dikarenakan ujian semester dan nilai harian pada perhitungan manual algoritma C4.5 dan menggunakan *tool* WEKA menyatakan bahwa kedua nilai tersebut tidak berpengaruh terhadap prediksi kelulusan siswa
2. Dari hasil pengolahan *data mining* maka didapatlah informasi bahwa siswa yang nilai ujian nasionalnya dibawah standar dan nilai ulangan hariannya rendah ataupun menengah mempunyai kemungkinan lebih tinggi untuk mendapatkan hasil tidak lulus dengan kondisi yang terbentuk dari algoritma C4.5, yaitu jika nilai ujian nasional memenuhi standar maka hasilnya lulus, jika ujian nasional dibawah standar dan nilai harian rendah maka hasilnya tidak lulus, jika ujian nasional dibawah standar dan nilai harian menengah maka hasilnya tidak lulus dan jika ujian nasional dibawah standar dan nilai harian tinggi maka hasilnya lulus. Jalur ini terbentuk dari algoritma C4.5 berdasarkan data dari Sekolah Dasar Negeri 8 Merapi Barat.

5. Daftar Rujukan

- [1]. Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta : CV. Andi Offset
- [2]. Sijabat, Alimancon. (2015). *Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Trees (Studi Kasus : Yayasan Perguruan Kristen Andreas)* Medan : STMK Budi Darma Medan
- [3.] www.cs.waikato.ac.nz/ml/weka

Student Colloquium Sistem Informasi & Teknik Informatika (SC-SITI)
Palembang, 11 Maret 2016

