

PENERAPAN METODE *DECISION TREE*(DATA MINING) UNTUK MEMPREDIKSI TINGKAT KELULUSAN SISWA SMPN1 KINTAMANI

Putu Gede Surya Cipta Nugraha, I Wayan Aribawa, I Putu Okta Priyana, Gede Indrawan
Pasca Sarjana Universitas Pendidikan Ganesha, Program Studi Ilmu Komputer
Email : lhu.on.tho@gmail.com, side_man2003@yahoo.com

Abstract

In this paper discusses the process of research conducted to predict the passing rate of students of SMPN 1 Kintamani using decision tree method. The processed data in this study is the student data by using the criteria of daily test scores, grades midterms, final exams and the value of national test scores. The data can then be managed using Weka tool to facilitate predicting student graduation rates. The application of data mining is used to predict the graduation rate by using decision tree algorithm and C45 as well as supporters of comparison to determine students' graduation rate information SMPN1 Kintamani by both methods. Based on the decision tree can be seen that the national exam as a tree root is the most decisive criteria for graduation prediction results which can then be investigated more deeply that the daily value as a node tree also has the same effect on the predicted outcome graduation. Each category value of the daily value will make way for the new rules state that in determining students' graduation predictions in both the student pass or not pass.

Keyword: *Decision tree, data mining and C45 Algorithm*

Abstrak.

Pada jurnal ini dibahas mengenai proses penelitian yang dilakukan untuk memprediksi tingkat kelulusan siswa SMPN 1 Kintamani dengan menggunakan metode *decision tree*. Data yang diolah 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 kemudian 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 C45* sebagai pendukung sekaligus perbandingan untuk mengetahui informasi tingkat kelulusan siswa SMPN1 Kintamani berdasarkan kedua metode tersebut. Berdasarkan pohon keputusan tersebut 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.

Keyword: *Decision tree, data mining dan Algoritma C45*

PENDAHULUAN

Melalui perkembangan teknologi informasi masa ini, kebutuhan akan informasi yang jelas dan akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu hal penting dalam di masyarakat. Kadangkala kebutuhan informasi yang tinggi tidak di barengi dengan

penyajian informasi yang memadai, seringkali informasi melalui proses mining diharapkan dapat memberikan informasi yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang penting dan berharga. Sekolah-sekolah saat ini dituntut untuk mampu mencapai keunggulan bersaing dan memiliki kualitas yang baik. Untuk mengatasi hal tersebut, pihak sekolah diuntut untuk dapat

mengambil langkah yang tepat dalam memajukan kualitas pendidikan di sekolah. Hal ini dapat dilakukan dengan cara melakukan proses pengolahan terhadap data tingkat kelulusan siswa.

Proses pengolahan melalui *data mining* dapat digunakan untuk memprediksi tingkat kelulusan siswa SMPN1 Kintamani. SMPN1 Kintamani memiliki kurang lebih 1100 siswa, dan dalam setiap kelasnya memiliki kurang lebih 40 siswa. Setiap tahun sekolah harus meluluskan paling sedikit 300 siswa. Sehingga pihak sekolah setiap tahun selalu mengkhawatirkan tentang tingkat kelulusan siswanya. Berdasarkan latar belakang tersebut, penulis tertarik untuk melakukan penelitian dalam memprediksi tingkat kelulusan siswa SMPN 1 Kintamani melalui *decision tree* dan *algoritma C45*. Dalam melakukan penelitian ini, tools yang penulis gunakan adalah WEKA (*Waikato Environment for Knowledge Analysis*). Melalui pengolahan data dengan teknik *data mining*, diharapkan dapat memanfaatkan data yang sebelumnya serta membandingkan antara nilai harian, nilai semester, nilai ujian sekolah serta nilai ujian nasional untuk dapat menentukan faktor mana yang menjadi penunjang paling dominan di dalam tingkat kelulusan seorang siswa.

TINJAUAN PUSTAKA

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).

Data mining adalah sebuah bidang ilmu yang berupaya menemukan pola,

kaidah, aturan, dan informasi berharga yang menarik dan belum diketahui sebelumnya dari sekumpulan besar data. Kemunculan ilmu ini dilatarbelakangi oleh munculnya tumpukan data di berbagai bidang kehidupan. Seringkali sebuah organisasi atau kelompok kerja tertentu banyak melakukan kegiatan pengumpulan data, administrasi maupun perhitungan-perhitungan yang menghasilkan data dalam jumlah besar.

Berbagai *tools* komersial maupun non-komersial beredar dan digunakan untuk berbagai tujuan yang berkaitan dengan *data mining*. Salah satu diantaranya yang menarik dan akan dibicarakan dalam modul ini adalah WEKA, yang merupakan alat bantu *data mining*, terutama dalam penerapannya untuk menyelesaikan masalah klasifikasi.

2. Klasifikasi

Klasifikasi dapat digambarkan sebagai berikut. Data input, disebut juga *training set*, terdiri atas banyak contoh (record), yang masing-masing memiliki beberapa atribut. Selanjutnya, tiap contoh diberi sebuah label *class* khusus. Tujuannya untuk menganalisa data input dan mengembangkan deskripsi atau model akurat untuk tiap *class* menggunakan fitur-fitur pada data. Deskripsi *class* ini digunakan untuk mengklasifikasikan data pengujian lainnya dengan label *class* tidak diketahui. Deskripsi tersebut juga dapat digunakan untuk memahami tiap *class* dalam data. Aplikasi-aplikasi klasifikasi antara lain berupa *credit approval*, *target marketing*, *medical diagnosis*, *treatment effectiveness*, *store location*, dll.

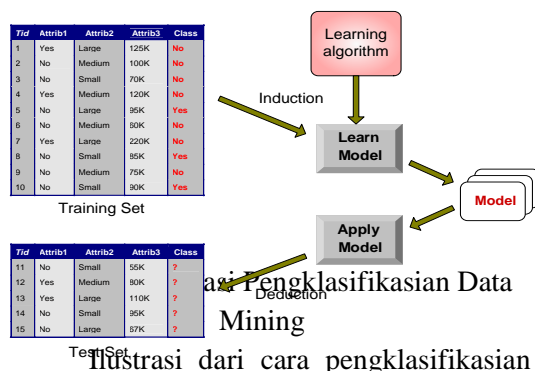
Klasifikasi dapat dilakukan dengan berbagai metode yang menghasilkan representasi model yang berbeda pula, antara lain:

1. Klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan

kriteria pembagian kelas dan kelas-kelas. Beberapa algoritma yang menggunakan metode ini adalah algoritma ID3, SLIQ, SPRINT, dan PUBLIC.

2. Klasifikasi dengan *memory-based reasoning*, yaitu metode klasifikasi yang digabungkan dengan penalaran berbasis memori.
3. *Neural network*, yaitu metode klasifikasi yang menggunakan model jaringan syaraf tiruan (JST) yang direpresentasikan dalam bentuk neuron-neuron, bobot neuron dan struktur JST.
4. *Naive Bayes* dan *Bayesian Belief Networks*
5. Klasifikasi berdasarkan kaidah (*rule*). Metode ini terbagi menjadi dua; yang pertama mengekstrak kaidah secara langsung dari data (misalnya algoritma RIPPER), sedangkan yang kedua mengekstrak data dari model klasifikasi lain (misalnya algoritma C4.5) seperti pohon keputusan, *neural network*, dll.

a. Pemodelan Klasifikasi dengan Decision Tree



Ilustrasi dari cara pengklasifikasian pada *data mining* adalah :

- a. Terdapat *training set* yang terdiri atas beberapa atribut dan *class*
- b. *Training set* tersebut kemudian diinduksikan dengan algoritma yang ada sehingga terbentuk suatu model sesuai data yang ada
- c. Dari model tersebut diaplikasikan ke dalam *test set*.

b. Sejarah WEKA

WEKA adalah sebuah paket *tools machine learning* praktis. “WEKA” merupakan singkatan dari *Waikato Environment for Knowledge Analysis*, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah *data mining* di dunia-nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan *machine learning*. Perangkat lunak ini ditulis dalam hirarki *class Java* dengan metode berorientasi objek dan dapat berjalan hampir di semua *platform*.

WEKA mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia implementasi algoritma-algoritma pembelajaran *state-of-the-art* yang dapat diterapkan pada dataset dari *command line*. WEKA mengandung *tools* untuk *pre-processing* data, klasifikasi, regresi, *clustering*, aturan asosiasi, dan visualisasi. *User* dapat melakukan *preprocess* pada data, memasukkannya dalam sebuah skema pembelajaran, dan menganalisa *classifier* yang dihasilkan dan performansinya – semua itu tanpa menulis kode program sama sekali. Contoh penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisa hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performansinya untuk dipilih.

Tools yang dapat digunakan untuk *pre-processing* dataset membuat *user* dapat berfokus pada algoritma yang digunakan tanpa terlalu memperhatikan detail seperti pembacaan data dari file-file, implementasi algoritma *filtering*, dan penyediaan kode untuk evaluasi hasil.

Pengembangan WEKA mengikuti model *releases Linux*: digit kedua yang genap menunjukkan *release* yang stabil dan digit kedua yang ganjil menunjukkan *release* ‘pengembangan’ (misalnya 3.0.x

adalah *release* stabil, sedangkan 3.1.x adalah *release* yang sedang dikembangkan). Beberapa versi awal dari WEKA:

- WEKA 3.0 : “versi buku” yang sesuai dengan deskripsi buku data mining.
- WEKA 3.2 : “versi GUI” yang menambahkan GUI dari CLI awal.
- WEKA 3.3 : “versi pengembangan” dengan berbagai peningkatan.

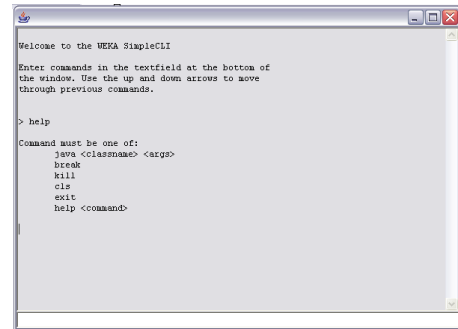
3. WEKA GUI Chooser

WEKA *GUI Chooser* adalah tampilan utama yang akan dilihat *user* pada saat pertama kali membuka perangkat lunak WEKA. Tampilan utama tersebut memberikan 4 pilihan GUI WEKA, yaitu *Simple CLI*, *Experimenter*, *Explorer*, dan *Knowledge Flow*.



Gambar 2. Weka GUI Chooser

3.1 GUI Simple CLI merupakan GUI yang memungkinkan *user* mengetikkan perintah-perintah melalui *command line* menurut standar penggunaan *classifiers* maupun *filters*. Misalnya mengeset percobaan dengan file *batch*.



Gambar 3. GUI Simple CLI

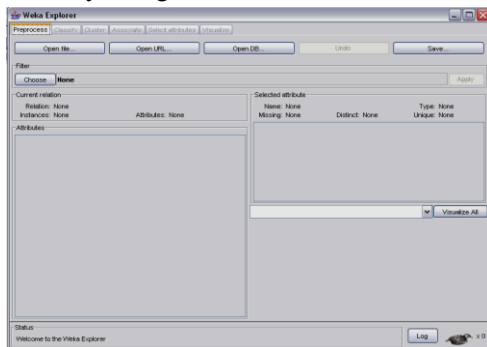
Contoh penggunaan CLI adalah dengan mengetikkan perintah:

- `java weka.classifiers.j48.J48 -t weather.arff`
Perintah ini memanggil JVM (Java Virtual Machine) dan menginstruksikannya untuk mengeksekusi algoritma J48 dari J48 package.
- `java weka.filters.unsupervised.attribute.Remove -V -R 1,4 -i trainingFile.arff -o myTrainingFile.arff`
Filter tersebut akan menghapus semua atribut kecuali yang pertama dan keempat dari sebuah dataset yang disimpan pada file *trainingFile.arff* dan menyimpan hasilnya di *myTrainingFile.arff*
- `java weka.classifiers.trees.J48 -t myTrainingFile.arff -T myTestFile.arff -U -p 1 > Results.arff`

Dari perintah ini, *decision tree J48* diterapkan pada file *myTrainingFile.arff*. File yang diuji ditunjukkan dengan *option -T*. Hasilnya *redirected* dari layar ke file *Results.arff* dan *options -U* dan *-p* menentukan bentuk output tertentu.

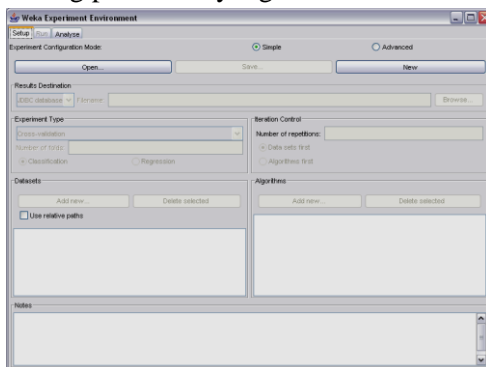
3.2 GUI Explorer adalah GUI WEKA yang paling mudah digunakan dan menyediakan semua fitur WEKA dalam bentuk tombol dan tampilan visualisasi yang menarik dan lengkap. *Preprocess*, *klasifikasi*, *asosiasi*, *clustering*, pemilihan atribut, dan visualisasi dapat

dilakukan dengan mudah dan menyenangkan di sini.



Gambar 4. GUI Explorer

3.3 GUI Experimenter memudahkan perbandingan performansi skema-skema pembelajaran yang berbeda. Experimenter biasanya digunakan untuk klasifikasi dan regresi. Hasil dari perbandingan performansi dapat dituliskan dalam file atau basis data. Pilihan evaluasi yang tersedia dalam WEKA adalah *cross-validation*, *learning curve*, *hold-out*. User juga dapat melakukan iterasi menurut beberapa setting parameter yang berbeda.



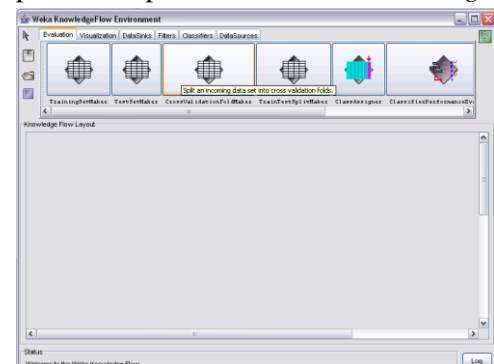
Gambar 5. GUI Experimenter

Tab *Setup* yang muncul saat user membuka *Experimenter* memungkinkan user memilih dan mengkonfigurasi eksperimen yang dilakukan. Setelah menyimpan definisi eksperimen yang dilakukan, user dapat memulai eksperimen dari tab *Run* dan meng-klik tombol *Start*. Area di bawahnya akan menunjukkan proses yang sedang dilakukan. Hasilnya disimpan dalam format CSV dan dapat dibuka dalam bentuk *spreadsheet*.

Tab ketiga, *Analyze*, dapat digunakan untuk menganalisa hasil eksperimen yang dikirim ke WEKA. Jumlah baris hasil ditunjukkan pada panel *Source*. Hasilnya dapat di-load dalam format *.ARFF* maupun dari basis data.

Antarmuka ini memungkinkan user melakukan lebih dari 1 eksperimen sekaligus, mungkin menerapkan beberapa teknik berbeda pada sebuah dataset, atau teknik yang sama dengan parameter-parameter yang berbeda.

3.4 GUI Knowledge Flow merupakan GUI baru dalam WEKA yang merupakan antarmuka *Java-Beans-based* untuk melakukan *setting* dan menjalankan percobaan-percobaan *machine learning*.



Gambar 6. GUI Knowledge Flow

Dalam *GUI Experimenter* ini, beberapa sumber data, *classifier*, dll dapat dihubungkan secara grafis. User juga dapat menggambarkan aliran data melalui komponen-komponen, misalnya:

“data source” -> “filter” ->

“classifier” -> “evaluator”

KnowledgeFlow menyediakan alternatif lain dari *Explorer* sebagai sebuah *front end* grafis untuk algoritma-algoritma inti WEKA. Karena masih dalam pengembangan, beberapa fungsionalitas dalam *Explorer* belum tersedia dalam *KnowledgeFlow*.

KnowledgeFlow menampilkan ‘aliran data’ dalam WEKA. User dapat memilih komponen-komponen WEKA dari *toolbar*, meletakkannya pada area yang

tersedia dan menghubungkannya untuk membentuk ‘aliran pengetahuan’ pemrosesan dan analisa data.

KnowledgeFlow dapat menangani data secara *incremental* maupun dalam *batches* (*Explorer* hanya menangani data *batch*). Tentunya pembelajaran dari data secara *incremental* memerlukan sebuah *classifier* yang dapat diupdate *instance* per *instance*. Dalam WEKA tersedia 5 *classifiers* yang dapat menangani data secara *incremental*: *NaiveBayesUpdateable*, *IB1*, *IBk*, *LWR* (*Locally Weighted Regression*). Tersedia pula sebuah *metadata classifier* – *RacedIncrementalLogitBoost* – yang dapat digunakan dari berbagai basis regresi untuk data *class* diskrit secara *incremental*.

4. Format Data dalam WEKA

Misalnya diketahui sekumpulan data dan ingin dibangun sebuah *decision tree* dari data tersebut, maka data tersebut harus disimpan dalam format ‘flat’, ARFF karena WEKA perlu mengetahui beberapa informasi tentang tiap atribut yang tidak dapat disimpulkan secara otomatis dari nilai-nilainya.

File ARFF (*Attribute-Relation File Format*) adalah sebuah file teks ASCII yang berisi daftar *instances* dalam sekumpulan atribut. File ARFF dikembangkan oleh *Machine Learning Project* di *Department of Computer Science of The University of Waikato* untuk digunakan dalam perangkat lunak WEKA.

Pengubahan format data ini dapat dilakukan dengan mudah. Misalkan data awal dalam format .xls, buka data tersebut dari Microsoft Excel dan simpan sebagai .csv. Selanjutnya, buka file tersebut dari Microsoft Word, notepad, atau editor teks lainnya dan data sudah berubah dalam format *comma-separated*. Lalu sesuaikan data tersebut dengan menambahkan informasi awal. Hasilnya, data tersebut sudah dapat digunakan sebagai inputan dalam WEKA.

Pastikan bahwa data dalam format .arff tersebut sudah memenuhi:

- Data dipisahkan dengan koma, dengan kelas sebagai atribut terakhir.
- Bagian *header* diawali dengan @RELATION.
- Tiap atribut ditandai dengan @ATTRIBUTE. Tipe-tipe data dalam WEKA: numerik (REAL atau INTEGER), nominal, String, dan Date.

Bagian data diawali dengan @DATA

5. KDD (*Knowledge Discovery in Databases*)

Dalam proses melakukan analisis data pada penerapan data *mining*, digunakan beberapa tahapan *knowledge discovery in databases* (KDD) yang terdiri :

- a. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.
- b. 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*).
- c. Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.
- d. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu
- e. Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

Memilih atribut yang akan digunakan 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 berikut :

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

Keterangan :

S:Himpunan kasus

A:Atribut

n :Jumlah partisi atribut A

[Si]:jumlah kasus pada partisi ke-i

[S]:jumlah kasus S

Sedangkan perhitungan nilai entropy:

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

Keterangan :

S:Himpunan Kasus

n :Fitur

i :Jumlah partisi S

pi :Proporsi dari S terhadap S

6. Variabel Kelulusan Siswa

Di SMPN1 Kintamani 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 di SMPN 1 Kintamani beserta dengan keterangan nilai untuk setiap atributnya.

Keterangan Atribut :

a. Nilai Harian (NH)

1.) Bawah = 0 – 64

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini akan dilakukan remedial dari masing-masing guru untuk melakukan perbaikan nilai yang lebih baik.

2.) Menengah = 65 – 74

Siswa SMPN1 Kintamani yang tergolong *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.

3.) Atas = 75 – 100

Siswa SMPN1 Kintamani yang tergolong *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.

b. Nilai Semester (NS)

1.) Cukup Baik = 0 – 64

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini maka siswa tersebut akan melakukan remedial dari masing-masing mata pelajaran 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 mata pelajaran terutama untuk mata pelajaran yang akan menjadi mata pelajaran yang termasuk ke dalam ujian nasional.

2.) Baik = 65 – 74

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini maka akan dianggap sudah memenuhi standar nilai semester yang ditetapkan oleh sekolah dan dapat membantu untuk menutupi kekurangan nilai pada mata pelajaran yang lainnya.

3.) Sangat Baik = 75 – 100

Siswa SMPN1 Kintamani yang tergolong *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.

c. Ujian Sekolah (US)

1.) Belum Tercapai = 0 – 59,5

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini maka akan dianggap belum tuntas dalam penguasaan materi dan memenuhi standar kelulusan sekolah.

2.) Tercapai = 59,6 – 100

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini maka akan dianggap telah menguasai materi dan lulus dalam memenuhi syarat kelulusan nilai ujian dari sekolah.

d. Ujian Nasional (UN)

1.) Di bawah standar : 0 – 55

Siswa SMPN1 Kintamani yang tergolong *range* ini sudah bisa dipastikan bahwa siswa tersebut gagal dalam ujian nasional dan harus mengulang tahun depan untuk mengikuti ujian yang sama

2.) Memenuhi standar : 56 – 100

Siswa SMPN1 Kintamani yang tergolong *range* nilai ini maka sudah bisa dipastikan bahwa siswa tersebut telah lulus dalam mengikuti ujian

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 1. Perhitungan Entropi Node Pertama

Atribut		Jumlah	Lulus	Tidak Lulus	Entropi
Total		294	210	84	0,8631206
Nilai Semester	Cukup	84	42	42	1
	Baik	126	84	42	0,9182958
	Baik	84	84	0	0
Nilai Harian	Sangat Baik	105	42	63	0,9709506
	Bawah	105	84	21	0,7219281
	Menengah	84	84	0	0

Ujian Sekolah	Atas	168	126	42	0,8112782
	Tercapai	126	84	42	0,9182958
Ujian Nasional	Belum tercapai	147	63	84	0,9852281
	Dibawah standar	147	147	0	0
	Memenuhi standar				

Gain (Total - Nilai Semester)

$$\begin{aligned}
 &= 0,863120 - \left(\frac{84}{294} \times 1 \right) \\
 &\quad + \left(\frac{126}{294} \times 0,9182958 \right) \\
 &\quad + \frac{84}{294} \times 0 \\
 &= 0,863120 - (0,2857 + 0,3936 + 0) \\
 &= 0,863120 - 0,6793 \\
 &= 0,1839
 \end{aligned}$$

Gain (Total - Nilai Harian)

$$\begin{aligned}
 &= 0,863120 - \left(\frac{105}{294} \times 0,9709506 \right) \\
 &\quad + \left(\frac{105}{294} \times 0,7219281 \right) \\
 &\quad + \frac{84}{294} \times 0 \\
 &= 0,8632 - (0,3468 + 0,2578 + 0) \\
 &= 0,8632 - 0,6046 \\
 &= 0,2586
 \end{aligned}$$

Gain (Total - Ujian Sekolah)

$$\begin{aligned}
 &= 0,863120 - \left(\frac{168}{294} \times 0,8112782 \right) \\
 &\quad + \left(\frac{126}{294} \times 0,9182958 \right) \\
 &= 0,8632 - (0,4636 + 0,3936) \\
 &= 0,8632 - 0,8572 \\
 &= 0,006
 \end{aligned}$$

Gain (Total - Ujian Nasional)

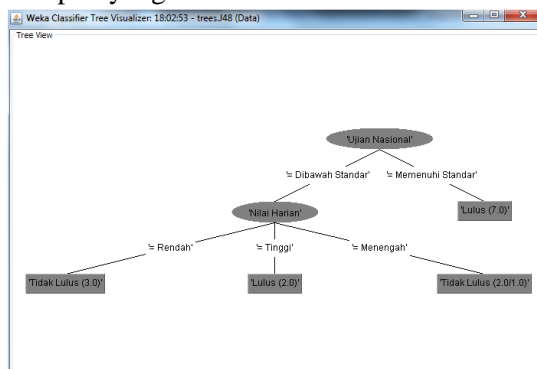
$$\begin{aligned}
 &= 0,863120 - \left(\frac{147}{294} \times 0,9852281 \right) \\
 &\quad + \left(\frac{147}{294} \times 0 \right) \\
 &= 0,8632 - (0,4927 + 0)
 \end{aligned}$$

$$= 0,8632 - 0,4927$$
$$= 0,3705$$

Setelah melalui proses pengolahan data mining dengan WEKA, diperoleh gambar pohon keputusan. Untuk nilai semester dan ujian sekolah tidak termasuk ke dalam pohon yang datanya diolah dari *tool* WEKA. Hasil dari proses tersebut menggambarkan hasil yang sama dengan pohon yang terbentuk dari perhitungan manual algoritma C45 dan pohon yang terbentuk dari *tool* WEKA. Nilai ujian nasional dan nilai harian terpilih oleh algoritma C45 sebagai kriteria yang berpengaruh sehingga saling membentuk daftar aturan pembacaan pohon keputusan.

Sehingga berdasarkan pohon keputusan tersebut 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.



Gambar7.Pohon Keputusan Hasil dari Weka

SIMPULAN

Berdasarkan proses analisis untuk mengetahui tingkat kelulusan siswa di SMPN

1 Kintamani dengan menggunakan teknik data mining, maka didapatkan hasil kesimpulan sebagai berikut:

1. Pengolahan *data mining* dengan menggunakan metode *decision tree* dan algoritma C45 menghasilkan sebuah pohon keputusan beserta dengan tingkat kelulusan masing-masing siswa lewat jalur pohon yang terbentuk di mana tingkat kelulusan siswa kelas IXSMPN 1 Kintamani 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 C45 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 C45, 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 C45 berdasarkan data dari SMPN1 Kintamani.

DAFTAR RUJUKAN

Ariana Azimah, Yudho Giri Sucahyo, *Penggunaan Data Warehouse Dan Data Mining Untuk Data Akademik*



Sebuah Studi Kasus Pada Universitas Nasional, 2007

Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta : CV. Andi Offset

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

www.cs.waikato.ac.nz/ml/weka