JURNAL

IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PRESTASI SISWA

IMPLEMENTATION OF DATA MINING WITH C4.5 ALGORITHM TO PREDICT STUDENT ACHIEVEMENT



Oleh:

SITI MUHIMATUL KHOIROH

NPM: 12.1.03.02.0315

Dibimbingoleh:

- 1. Dr. SURYO WIDODO, M.Pd.
- 2. RISKY ASWI RAMADHANI, M.Kom.

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI
2017





Artikel Skripsi Universitas Nusantara PGRI Kediri

SURATPERNYATAAN ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan dibawah ini:

Nama Lengkap

: SITI MUHIMATUL KHOIROH

NPM

: 12.1.03.02.0315

Telepun/HP

: 081556751867

Alamat Surel (Email)

: azka.muhim@gmail.com

Judul Artikel

: Implementasi Data Mining Dengan Algoritma C4.5

Untuk Memprediksi Prestasi Siswa

Fakultas - Program Studi

:Teknik - TeknikInformatika

Nama Perguruan Tinggi

: Universitas Nusantara PGRI Kediri

Alamat Perguruan Tinggi

: Jl. KH. Achmad Dahlan 76 Mojoroto, Kota Kediri

Dengan ini menyatakan bahwa:

- a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
- b. artikel telah diteliti dan disetujui untuk diterbitkan oleh DosenPembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila dikemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggung jawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 6 Februari 2017
Pembimbing I	Pembimbing II	Penulis,
ph	The state of the s	C. Y.J.
Dr. Suryo Widodo, M.Pd. NIDN.0002026403	Risky Aswi Ramadhani, M.Kom. NIDN. 0708049001	Siti Muhimatul Khoiroh NPM. 12.1.03.02.0315

Siti Muhimatul Khoiroh | 12.1.03.02.0315 Fakultas Teknik – Teknik Informatika simki.unpkediri.ac.id || 1||



IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PRESTASI SISWA

Siti Muhimatul Khoiroh 12.1.03.02.0315 FakultasTeknik – TeknikInformatika azka.muhim@gmail.com Dr. Suryo Widodo, M.Pd. dan Risky Aswi Ramadhani, M.Kom. UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Dalam lembaga Pendidikan khususnya di SMK Pemuda Papar memiliki banyak data namun belum dimanfaatkan secara optimal. Data yang dimiliki sekolah biasanya dalam jumlah besar dan terus bertambah setiap tahunnya. Dengan memanfaatkan metode data mining data tersebut dapat diolah menjadi informasi yang berguna khususnya bagi SMK Pemuda Papar.

Penelitian ini menerapkan algoritma C4.5 pada data mining ke dalam sebuah aplikasi yang dapat digunakan untuk memprediksi prestasi siswa SMK Pemuda Papar. Dengan menerapkan metode tersebut penelitian ini diharapkan dapat menghasilkan pola aturan prediksi prestasi siswa sehingga dapat mengantisipasi sejak dini terhadap siswa yang berpotensi mengalami hambatan dalam belajarnya.

Penelitan ini berhasil menerapkan algoritma C4.5 ke dalam aplikasi. Aplikasi tersebut dapat mengetahui pola-pola prediksi prestasi siswa di SMK Pemuda Papar. Dari pola-pola yang dihasilkan oleh aplikasi kemudian dapat diinterpretasikan menjadi sebuah informasi atau knowledge baru.

Kata kunci: data mining, algoritma C4.5, prediksi prestasi siswa.

I. LATAR BELAKANG

Sekolah Menengah Kejuruan (SMK) merupakan lembaga pendidikan yang berupaya menciptakan siswa tamatan yang terampil dan memiliki keahlian sehingga lulusannya dapat langsung terjun ke dunia kerja maupun melanjutkan ke pendidikan yang lebih tinggi [1]. Namun setiap siswa memiliki tingkat ketrampilan yang berbeda sehingga memerlukan bimbingan yang berbeda pula. Untuk itu perlu diketahui lebih dini mengenai tingkat prestasi dan ketrampilan siswa agar bisa memberikan Siti Muhimatul Khoiroh | 12.1.03.02.0315 FakultasTeknik – TeknikInformatika

pembelajaran dan bimbingan sesuai dengan kemampuan siswa tersebut.

Yang harus dilakukan oleh lembaga pendidikan adalah melakukan antisipasi terhadap siswa yang berpotensi mengalami hambatan atau kurang berprestasi dalam belajarnya. Hal ini penting dilakukan karena semakin awal sekolah kejuruan mengetahui adanya potensi siswa yang kemungkinan akan mengalami hambatan dalam belajarnya, maka sekolah bisa melakukan langkah-langkah antisipatif.

simki.unpkediri.ac.id



Akibat yang mungkin muncul dari kurangnya antisipasi atau pencegahan sejak dini terhadap peserta didik yang berpotensi mengalami hambatan dalam belajarnya atau kurang berprestasi adalah banyak siswa tidak naik kelas dan sulit menerima pelajaran di kelasnya.

Untuk itu dibutuhkan sistem yang mampu memprediksi sejak dini terhadap peserta didik yang berpotensi berprestasi atau mengalami hambatan dalam belajarnya sehingga guru dapat melakukan bimbingan atau pendampingan khusus terhadap siswa tersebut. Adapun kriteria yang menjadi dasar dalam penentuan prediksi prestasi siswa adalah Tingkat Sosial Ekonomi Orang Motivasi Siswa, Kedisiplinan Siswa dan Prestasi Masa Lalu.

lembaga pendidikan Dalam khususnya di **SMK** Pemuda Papar memiliki banyak data namun belum dimanfaatkan secara optimal. Data yang dimiliki sekolah biasanya dalam jumlah besar dan akan terus bertambah setiap tahunnya antara lain data kehadiran siswa, data/ berkas penerimaan siswa baru, data tingkat sosial ekonomi orang tua yang biasa diinput pada saat penerimaan siswa baru maupun data yang dimiliki oleh BP/BK di sekolah.

Dengan banyaknya data yang ada, penulis ingin menggali suatu aturan dan pola yang terdapat pada data siswa. *Data mining*, sering 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 [2]. Algoritma C4.5 merupakan algoritma data mining yang menggunakan pohon keputusan dalam penyelesaiannya. Algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterprestasikan, memiliki tingkat akurasi yang dapat diterima.

II.METODE

A. Data Mining

Data mining adalah analisis otomatis dari data yang berjumlah besar dan kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya [3]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [3]. 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 [2].



Jadi data mining adalah suatu proses penggalian data yang memanfaatkan data yang cukup besar melalui serangkaian proses untuk mendapatkan informasi baru yang berguna.

B. Decision Tree

metodologi Ada banyak Data Mining, salah satu yang populer adalah Pohon Keputusan (Decision Tree). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabang-cabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan attribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan [4].

C. Algoritma C4.5

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

- 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.

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

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

Keterangan:

S: himpunan kasus

A : atribut

N : jumlah partisi atribut A

|S_i| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropy. Entropy digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropy adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{n} -pi * log_2pi$$

Keterangan:

S: himpunan kasus

A : fitur

n : jumlah partisi S

P_i: proporsi dari S_i terhadap S

Dari kutipan di atas Perhitungan C4.5 dimulai dengan mencari nilai entropy dari semua data. Nilai entropy digunakan sebagai dasar perhitungan *gain* tiap atribut, nilai gain tertinggi menjadi akar dari pohon keputusan perhitungan dilakukan hingga semua atribut terdefinisi.

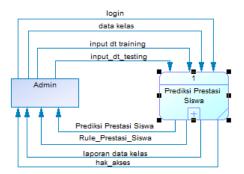
III. HASIL DAN KESIMPULAN

A. HasilImplementasiSistem

1. Desain Proses DFD



Berikutmerupakan DFD Level 0 Implementasi Data Mining dengan Algoritma C4.5 untuk memprediksi prestasi siswa.



DFD level 0 mempunyai 1 entitas yaitu admin. Entitas admin mempunyai aktivitas untuk mengolah sistem dan menggunakan aplikasi. Relasi dari admin ke proses yaitu admin mengirim data login admin, menginputkan data training, menginputkan data rule dan menginputkan data testing. Relasi dari proses ke admin mendapatkan hak akses login admin, mendapat laporan data training, mendapat laporan data rule dan mendapat laporan data testing

2. *Input* Data Training



Halaman *input* data training terletak pada menu master, pilih training, lalu pilih simbol tambah. Halaman ini digunakan untuk memasukkan data training yang ingin ditambahkan.

3. *OutputTraining*



Pada halaman *output* training merupakan halaman yang menampilkan hasil dari data training yang sudah diinputkan sebelumnya.

4. *Output* Hasil Perhitungan



Pada halaman *output*hasil perhitungan merupakan halaman yang menampilkan hasil perhitungan berdasarkan data training yang sudah diinputkan.

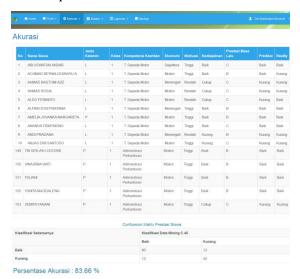
5. InputTesting



∯ A Home ■ Profil • ◆ Metode •	Master ▼	
Data Testing		
Nama Siswa		
Jenis Kelamin	OL OP	
Kelas	¥	
Kompetensi Keahlian	○ Administrasi Perkantoran ○ T. Komputer dan Jaringan ○ T. Sepeda Motor	
Ekonomi	W	
Motivasi	○ Tinggi ○ Rendah	
Kedisiplinan	W	
Prestasi Masa Lalu	W	
Reality	V	
	proses	

Halaman *input*data testing terletak pada menu metode, lalu pilih data testing. Halaman ini digunakan admin untuk melakukan testing data siswa sesuai atribut yang dimasukkan dan dibandingkan dengan data real untuk masing-masing siswa.

6. OutputAkurasi



Pada halaman *output* akurasi merupakan halaman yang menampilkan hasil inputan data testing yang dibandingkan dengan keadaan sebenarnya.

B. KESIMPULAN

Berikutmerupakankesimpulan yang diperolehdarihasilpenelitian:

 Telah dihasilkan rancangan sistem dari perhitungan Implementasi Data

- mining dengan Algoritma C4.5 untuk prediksi prestasi siswa, yang sudah dijelaskan pada bab 4. Penelitian ini menggunakan data siswa kelas X, XI, XII Tahun pelajaran 2015/2016. Dari data siswa tersebut dijadikan data training, untuk dijadikan acuan dalam perhitungan Algoritma C4.5.
- Data testing yang digunakan adalah data kelas X Tahun Pelajaran 2016/2017 yang berjumlah 153 siswa, dari hasil testing di atas ditemukan Hasil akurasi sebesar = 83,66%.
- 3. Program aplikasi prediksi prestasi siswa menggunakan Algoritma C4.5 yang telah dihasilkan ini merupakan aplikasi berbasis web menggunakan bahasa pemrograman PHP yang memiliki tampilan menarik dan mudah diimplementasikan.

IV. DAFTAR PUSTAKA

- [1] Susanto, Heri, dan Sudiyatno. 2014.

 Data Mining untuk Memprediksi
 Prestasi Siswa Berdasarkan Sosial
 Ekonomi, Motivasi, Kedisiplinan dan
 Prestasi Masa Lalu. Yogyakarta.
- [2] Santosa, Budi. 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- [3] Pramudiono, I. 2007. Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data. http://www.ilmukomputer.org/wpcontent/uploads/2006/08/ikodatamining.zip Diakses pada tanggal 31 Desember 2015 jam 11.30



- [4] Pramudiono, I. 2007. Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data. http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip Diakses pada tanggal 31 Desember 2015 jam 11.30
- [5] Kusrini, dan Emha Taufik Luthfi.2009. Algoritma Data Mining.Yogyakarta: Penerbit Andi.