

ALGORITMA C4.5 UNTUK PENJURUSAN SISWA SMA NEGERI 3 PATI

Eka Budi Rahayu¹

¹Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Dian Nuswantoro Semarang
Jl. Nakula 1 No 5-11 Semarang 50131
Telp : (024) 35176361, Fax : (024) 3520165
Email : ekarahayu382@gmail.com¹

Abstract

The curriculum used in High School (SMA) is currently the curriculum of 2013. In the development of the new curriculum, there are some differences with the previous curriculum. One of them is the difference in determining the direction the high school students, because majors performed in class X. Moreover majors done by hand so it takes a lot of time and not efficiently. Our of majors itself is so that later in life, lessons given students more focused. Because not infrequently also students who carelessly in determining the direction they take. Based on these problems, then do research to classify the data placement of students based on the attributes that have been determined by the school SMAN 3 Starch. The method used in the classification is the classification method by using Decision Tree (Decision Tree) with one of the algorithms used are C4.5. so that later can be used as a basis for consideration in determining the majors. Results of experiments and evaluations show using 239 data sets and split to 70% training data, testing data is 30% produces an accuracy of 81.94%

Keywords : classification, decision tree, C4.5 algorithms, placement of students

I. PENDAHULUAN

Kurikulum yang digunakan pada Sekolah Menengah Pertama (SMA) saat ini adalah kurikulum 2013. Dalam perkembangan kurikulum yang baru ini terdapat beberapa perbedaan dengan kurikulum sebelumnya. Salah satunya yaitu perbedaan dalam menentukan jurusan pada siswa-siswi SMA. Karena penjurusan dilakukan pada kelas X. Terlebihnya penjurusan dilakukan secara manual dan memerlukan banyak waktu dan tidak efisien. Faktor utama yang menentukan penjurusan pada SMA Negeri 3 Pati adalah nilai rata-rata raport matematika SMP, nilai rata-rata raport IPA SMP, nilai rata-rata raport IPS SMP, nilai Ujian Nasional matematika, nilai Ujian Nasional IPA, nilai tes akademik matematika, nilai tes akademik IPA, nilai tes akademik IPS, dan minat siswa.

Tujuan dari penjurusan itu sendiri adalah agar kelak dikemudian hari, pelajaran yang diberikan kepada siswa lebih terarah. Karena tidak jarang juga siswa-siswi yang asal-asalan dalam menentukan jurusan yang mereka ambil. Seiring dengan perkembangan teknologi hal tersebut dapat diatasi dengan teknik pengelompokan data dengan data mining. Sementara itu, data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2].

Hasil penelitian yang dilakukan oleh Liliana Swastina [2], Algoritma Decision Tree C4.5 lebih akurat daripada algoritma Naïve Bayes dalam penentuan kesesuaian jurusan dan rekomendasi jurusan mahasiswa. Dengan demikian dapat disimpulkan bahwa Algoritma Decision

Tree C4.5 akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat keakuratan 93,31% dan akurasi rekomendasi jurusan sebesar 82,84%.

Berdasarkan pemaparan singkat tentang kelebihan yang sudah peneliti jelaskan diatas, maka penulis akan membuat sebuah aplikasi yang diharapkan dapat menampilkan hasil data mining dengan Algoritma Decision Tree C4.5 dalam proses penentuan jurusan siswa SMA Negeri 3 Pati

II. TINJAUAN STUDI

2.1 Penelitian Terkait

Obbie Kristanto (2014) dalam penelitiannya akan menerapkan algoritma ID3 untuk menentukan jurusan siswa SMA N 6 Semarang. Penjurusan yang masih dilakukan secara manual akan memerlukan banyak waktu dan kurang efisien . Tujuan dari penelitian ini adalah untuk membantu SMAN 6 Semarang dalam memilih jurusan yang sesuai dengan minat dan bakat siswa serta untuk mengetahui tingkat akurasi menggunakan algoritma ID3. Metode *Iterative Dichotomiser 3* (ID3) akan membentuk sebuah pohon keputusan yang kedepannya dapat digunakan sebagai dasar pertimbangan dalam menentukan penjurusan di SMA, dan tools yang digunakan dalam pembuatan aplikasi ini adalah pemrograman java. Penelitian ini berupa aplikasi *data mining* untuk memprediksi penjurusan IPA atau IPS. Kesimpulan yang diperoleh dari penelitian ini adalah aplikasi yang dibuat sesuai dengan kebutuhan dan tujuan dari penulis, dengan menggunakan 371 dataset dan 20 data uji menghasilkan akurasi sebesar 80% [8].

2.2 Penjurusan SMA

Sekolah Menengah Atas (SMA), adalah jenjang pendidikan menengah pada pendidikan formal di Indonesia setelah lulus Sekolah menengah pertama (atau sederajat). Kurikulum yang digunakan pada Sekolah Menengah Atas (SMA) saat ini adalah kurikulum 2013. Dalam perkembangan kurikulum yang baru ini terdapat perbedaan dengan kurikulum sebelumnya. Salah satu perbedaannya yaitu dalam proses penentuan jurusan pada siswa SMA, karena penjurusan dilakukan pada kelas X.

2.3 Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 akan membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun. Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, yaitu [4]:

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari kata *histori* yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :

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

Keterangan :

S = Himpunan kasus

n = jumlah partisi S

P_i= proporsi S_i terhadap S

Kemudian Hitung nilai *gain* menggunakan rumus :

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

Keterangan :

S = Himpunan kasus

A = Fitur

n = jumlah partisi atribut A

$\left| \frac{S_i}{S} \right|$ = proporsi S_i terhadap S

S = jumlah kasus dalam S

2.4 Decision Tree

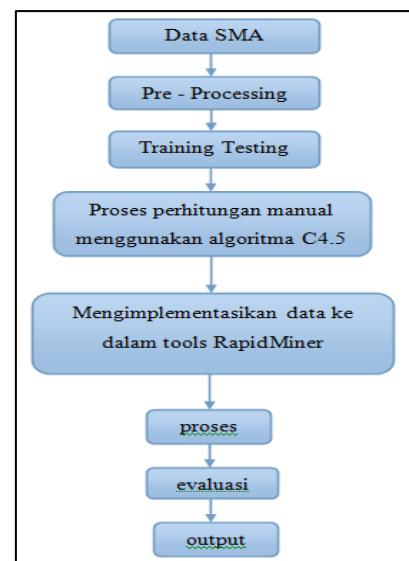
Pohon keputusan atau dikenal dengan *Decision Tree* adalah salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap *node* mempresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan

daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root* [10].

III. METODE PENELITIAN

3.1 Teknik Analisis Data

data adalah proses mencari dan menyusun data secara otomatis yang diperoleh dari hasil wawancara dan studi pustaka untuk mendapatkan sebuah kesimpulan dalam penelitian. Data yang digunakan dalam penelitian ini adalah berdasarkan kriteria yang digunakan untuk penjurusan siswa SMA Negeri 3 Pati. Metode yang diusulkan untuk proses penelitian ini adalah metode klasifikasi menggunakan algoritma C4.5.



Gambar 3.1 Alur pemodelan yang diusulkan

IV. HASIL & PEMBAHASAN

4.1 Data Uji

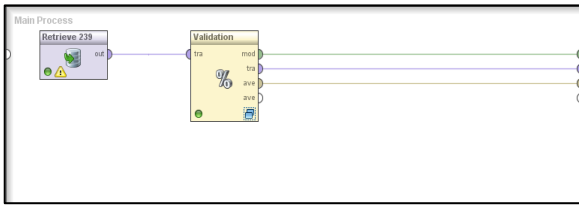
Data yang digunakan dalam tugas akhir ini merupakan data siswa kelas X SMA Negeri 3 Pati angkatan 2014/2015. Jumlah data yang digunakan sebanyak 239 data dimana 193 kelas IPA dan 46 kelas IPS. Data tersebut memiliki 10 atribut yang digunakan untuk proses penjurusan, diantaranya adalah rata-rata raport matematika, rata-rata raport IPA, rata-rata raport IPS, Ujian Nasional matematika, Ujian Nasional IPA, tes akademik matematika, tes akademik IPA, tes akademik IPS, minat siswa, hasil. Dimana atribut hasil sebagai target, dan pada atribut target ada dua kelas yang menjadi tujuan atau target yaitu IPA dan IPS.

Tabel 4.1 Klasifikasi Nilai

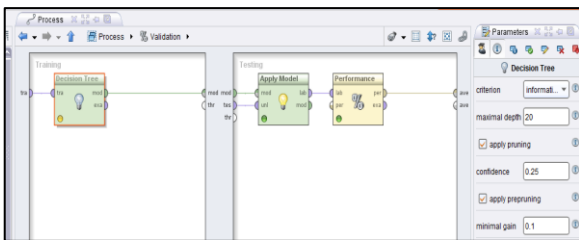
Nilai	Klasifikasi
86-100	A
71-85	B
56-70	C
41-55	D
≤ 40	E

4.2 Pengujian Model

pada penelitian ini *rules* atau model yang terbentuk diterapkan pada program yang diperoleh dengan bantuan *tools* RapidMiner. Data yang diuji yaitu sebanyak 239 record dengan 10 atribut dimana semua atribut bertipe kategorikal. Disini penulis menggunakan *Split Validation* sebagai validasi. Pada proses validasi terdapat dua kolom, yaitu *training* dan *testing*. Penulis menggunakan perbandingan 70% untuk data *training* dan 30% untuk data *testing*. Kolom *training* berisi algoritma klasifikasi yang diterapkan, yaitu algoritma C4.5 dan kolom *testing* berisi *apply model* serta *performance* untuk mengukur akurasi dari model algoritma C4.5. Berikut gambar *tools* RapidMiner untuk menghitung tingkat akurasi dengan algoritma C4.5 :



Gambar 4.1 *Split validation* dengan algoritma C4.5



Gambar 4.2 *Split validation* dengan algoritma C4.5

4.3 Evaluasi dan Validasi Hasil

Evaluasi dilakukan dengan mengembangkan hasil klasifikasi. Pengukuran data dilakukan menggunakan confusion matrix untuk mengevaluasi hasil dari algoritma Decision Tree C4.5. Confusion matrix merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. Tabel ini untuk mengukur tingkat akurasi sebuah data.

accuracy: 81.94%			
	true IPS	true IPA	class precision
pred IPS	13	11	54.17%
pred IPA	2	46	95.83%
class recall	86.67%	80.70%	

Gambar 4.3 Hasil *confusion matrix* dengan algoritma C4.5

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{46+13}{46+13+2+11} \times 100\% \\
 &= 81.94\%
 \end{aligned}$$

4.4 Halaman Input

Gambar 4.4 Halaman Input

4.5 Halaman Hasil

No.	No.Tes	Nama	Nil Raport	IPA Raport	IPS Raport	Nil UN	IPA UN	Nil Akademik	IPA Akademik	IPS Akademik	Test	Hasil	Hasil Mining
1	001	PISYITA SILESTYORNO	A	B	B	B	B	B	C	E	IPS	IPS	IPA
2	004	ALFA SETHA APRIYANNA	B	B	B	A	A	C	D	E	IPA	IPS	IPA
3	012	MELATI	B	B	A	B	B	C	C	E	IPS	IPA	IPA
4	017	LUKLIATIN NARSAH	B	B	B	C	C	E	E	E	IPA	IPA	IPA
5	020	DIMASDI DWI ARWANTO	B	B	B	A	A	B	C	E	IPA	IPA	IPA
6	021	DANI PRIBANTO	B	B	B	B	A	B	C	E	IPA	IPA	IPA
7	025	DEWA ANI PRISTIASARI	B	A	B	B	B	C	B	D	IPS	IPA	IPA
8	026	DHANS ADITIA	B	B	B	B	B	E	E	E	IPA	IPA	IPA
9	031	NALDI BRASHEN	B	B	B	A	B	E	E	E	IPA	IPA	IPA
10	036	REZY PRIRIKNO SANDE	B	B	B	B	B	E	E	E	IPA	IPS	IPA
11	038	ILHAM NIRMAL AKORDE	B	B	B	B	B	E	D	E	IPA	IPS	IPA
12	040	FARIDILLA ALFA	B	B	B	D	C	C	D	E	IPA	IPA	IPA
13	041	ANDI PRISTIASARI	A	B	B	A	B	B	C	D	IPS	IPA	IPA
14	049	LUKS SINMA SEKATI	A	A	B	B	A	D	D	E	IPS	IPA	IPA
15	051	FARA SEPTEWAL PITRE	B	B	B	A	C	B	E	E	IPA	IPA	IPA
16	053	PITRE M LILIANE	B	B	B	A	A	D	E	E	IPS	IPA	IPA
17	057	ILHAM RAMANICHANI	B	B	B	B	B	D	D	E	IPS	IPS	IPA
18	060	ALFA ALMUTADI B	B	B	B	D	B	E	D	E	IPS	IPS	IPA
19	061	SEPTIANI DEVI KERANA	B	B	B	B	B	C	C	E	IPA	IPA	IPA
20	064	HARDIATI	B	B	B	A	A	D	D	D	IPS	IPS	IPA

Gambar 4.5 Hasil Mining

Gambar diatas menjelaskan hasil dari proses mining, yaitu memprediksi pohon tree yang terbentuk dari proses mining yang menentukan bahwa inputan yang dimasukan berdasarkan 10 atribut apakah target tersebut masuk IPA atau IPS dari pohon keputusan yang terbentuk dari dataset yang digunakan.

NO	No. Tes	Nama	Raport Mat	Raport IPA	Raport IPS	UN Mat	UN IPA	Aka Mat	Aka IPA	Aka IPS	Minat Siswa	Hasil IPS	Hasil Mining	Ket
1	001	PUSPITA SULISTYORENI	A	B	B	B	B	B	C	E	IPSA	IPSA	IPSA	Fail
2	004	ALIFA SHITA APRILIYANA	B	B	B	A	A	C	D	E	IPSA	IPSA	IPSA	Fail
3	012	MELATI	B	B	A	B	B	C	C	E	IPSA	IPSA	IPSA	Akurat
4	017	ULFIYATUN NAFISAH	B	B	B	C	C	E	E	E	IPSA	IPSA	IPSA	Akurat
5	020	JUNAEDI DWI ARMANTO	B	B	B	A	A	B	C	E	IPSA	IPSA	IPSA	Akurat
6	021	DWI PUTIANTO	B	B	B	B	A	B	C	E	IPSA	IPSA	IPSA	Akurat
7	025	DIHAH AYU PUSPITASARI	B	A	B	B	B	C	B	D	IPSA	IPSA	IPSA	Akurat
8	028	DHIMAS ADITYA	B	B	B	B	B	E	E	E	IPSA	IPSA	IPSA	Fail
9	031	MALIK IBRAHIM	B	B	B	A	B	E	E	E	IPSA	IPSA	IPSA	Akurat
10	036	RIZKY PURNOMO SANDI	B	B	B	B	B	E	E	E	IPSA	IPSA	IPSA	Fail
11	038	ILHAM MIFTAHUL ASRORI	B	B	B	B	B	E	D	E	IPSA	IPSA	IPSA	Fail
12	040	FARDILLA ALIFIA	B	B	B	D	C	C	D	E	IPSA	IPSA	IPSA	Akurat
13	041	ANIK PUSPITASARI	A	B	B	A	B	B	C	D	IPSA	IPSA	IPSA	Akurat
14	049	LUKI SUKMA SEJATI	A	A	B	B	A	D	D	E	IPSA	IPSA	IPSA	Akurat
15	051	FARA SEPTIYANI PUTRI	B	B	B	A	C	B	E	E	IPSA	IPSA	IPSA	Akurat
16	053	PUTRI MULYANI	B	B	B	A	A	D	E	E	IPSA	IPSA	IPSA	Akurat
17	057	ILHAM RAMANDHANU	B	B	B	B	B	D	D	E	IPSA	IPSA	IPSA	Fail
18	060	AUFA ALMUHTADU B	B	B	B	D	B	E	D	E	IPSA	IPSA	IPSA	Fail
19	061	SEPTIANI DEVI KIRANA	B	B	B	B	B	C	C	E	IPSA	IPSA	IPSA	Akurat
20	064	HARIYATI	B	B	B	A	A	D	D	D	IPSA	IPSA	IPSA	Fail

Gambar 4.6 Perbandingan Data Uji

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian pada klasifikasi penjurusan siswa dapat diambil beberapa kesimpulan sebagai berikut :

1. Klasifikasi penjurusan siswa dengan algoritma C4.5 dapat mengklasifikasikan siswa untuk penjurusan kelas IPA dan IPS.
2. Dari 239 data yang dibagi data training 70% dan data testing 30%, hasil evaluasi dan validasi dengan *confusion matrix* menunjukkan tingkat akurasi pada algoritma C4.5 sebesar 81,94%. Hasil akurasi yang dihasilkan lebih rendah dari penelitian sebelumnya yang menghasilkan tingkat akurasi sebesar 97,75%. Hal ini dikarenakan data yang digunakan berbeda.
3. Penerapan *rules* dari algoritma C4.5 selanjutnya diterapkan pada bahasa pemrograman PHP yang digunakan dalam klasifikasi jurusan IPA dan IPS di SMA Negeri 3 Pati.

5.2 Saran

Saran yang diajukan dalam penelitian ini untuk penelitian selanjutnya, yaitu menganalisa data dan atribut lain tidak hanya dari nilai rata-rata raport matematika SMP, nilai rata-rata raport IPA SMP, nilai rata-rata raport IPS SMP, nilai Ujian Nasional matematika, nilai Ujian Nasional IPA, nilai tes akademik matematika, nilai tes akademik IPA, nilai tes akademik IPS, dan minat siswa misalnya minat orang tua, tes IQ, tes psikologi agar knowledge yang dihasilkan dapat lebih baik.

DAFTAR PUSTAKA

- [1] http://id.wikipedia.org/wiki/Sekolah_menengah_atas, diakses tanggal 6 Oktober 2014
- [2] Liliana Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Gema Aktualita*, Juni 2013.
- [3] Kursini, Luthfi, and Emha, *Algoritma Data Mining..*, 2009.
- [4] Anik Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," 2012.
- [5] Wenefrida Tulit Ina, "klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5.," *Media Elektro*, April 2013.
- [6] Budanis Dwi Meilani and Achmad Fauzi Slamet, *Klasifikasi Data Karyawan Untuk Menentukan Jadwal kerja Mwnnggunakan Metode Decission Tree..*, 2011.
- [7] Holisatul Munawaroh, Bain Khusnul, and Yeni Kustiyaningsih, "Perbandingan Algoritma ID3 Dan C5.0 Dalam Identifikasi Penjurusan Siswa SMA," *Sarjana Teknik Informatika*, Juni 2013.
- [8] Obbie Kristanto, "Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang," September 2014.
- [9] Yusuf Elmande and Prabowo Pudjo Widodo, "Pemilihan Criteria Splitting dalam Algrpitma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras : Studi Kasus Perum Bulog Divre Lampung," *TELEMATIKA MKOM*, Maret 2012.
- [10] Ratih Ariadni and Isye Arieshanti, "Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur Yang Tidak Pasti".
- [11] Aa Zezen Zaenal Abidin, "Implementasi Algoritma C4.5 untuk Menentukan Tingkat Bahaya Tsunami," Juli 2011.
- [12] Ratih Ariadni and Isye Arieshanti, *implementasi metode pohon keputusan untuk klasifikasi data dengan nilai fitur*, 2011.
- [13] Fajar Astuti Hermawati, *DATA MINING*, Putri Christian, Ed. Yogyakarta: ANDI, 2013.
- [14] Nikmatul Hidayah, "Klasifikasi Penjurusan Sekolah Menengah Atas dengan Algoritma Naive Bayes Classifier pada SMAN 1 SUBAH," 2014.
- [15] Author Khoirul Mu'arif, "komparasiPemodelan Data Menggunakan C4.5 dan C4.5 Berbasis PSO untuk Memprediksi Kelulusan Mahasiswa.," 2013.