

PENERAPAN ALGORITMA C4.5 UNTUK PENJURUSAN SISWA SEKOLAH MENENGAH ATAS

Rina Kurniasari¹, Azizah Fatmawati²

^{1,2} Program Studi Informatika, Fakultas Komunikasi dan Informatika,
Universitas Muhammadiyah Surakarta

Jl. Ahmad Yani, Pabelan, Kartasura, Surakarta 57102, Telp: +62(271)717417

E-mail : rinakurnia768@gmail.com¹, af157@ums.ac.id²

ABSTRAK

Jurusan di Sekolah Menengah Atas (SMA) digunakan sebagai acuan untuk menyalurkan bakat dan kemampuan dari siswa. Penjurusan tersebut dilakukan ketika pendaftaran siswa baru, sebelum siswa tersebut diterima sebagai siswa kelas X. Banyaknya data siswa yang melakukan pendaftaran dan pelaksanaan penjurusan saat ini yang masih menggunakan sistem manual, mengakibatkan terjadinya kesalahan dan waktu pengolahan data yang cukup lama. Penelitian ini bertujuan mengembangkan sistem untuk mengklasifikasikan jurusan bagi siswa dengan tepat, efektif dan efisien. Metode penelitian yang digunakan dalam penelitian ini dengan menerapkan. Algoritma C4.5 untuk mengklasifikasikan siswa dalam menentukan jurusan. Algoritma C4.5 digunakan untuk mencari pola-pola aturan berdasarkan variabel pendukung berupa nilai rata-rata raport Sekolah Menengah Pertama (SMP), nilai tes akademik seperti nilai mata pelajaran Ilmu Pengetahuan Alam (IPA), nilai mata pelajaran Ilmu Pengetahuan Sosial (IPS), dan nilai Bahasa. Hasil dari penelitian ini sendiri berupa aplikasi data mining dengan Algoritma C4.5 untuk memprediksi penjurusan IPA, IPS, atau Bahasa. Tingkat akurasi yang didapatkan sebesar 97,42 %.

Kata kunci: Algoritma C4.5, Klasifikasi, Penjurusan

1. PENDAHULUAN

Kurikulum yang diterapkan di Sekolah Menengah Atas (SMA) saat ini menggunakan kurikulum 2013. Kurikulum 2013 membagi jurusan menjadi 3 yaitu IPA, IPS, dan Bahasa yang mana proses penjurusan tersebut dilakukan ketika pendaftaran siswa baru sebelum diterima sebagai siswa kelas X. Proses penjurusan tersebut masih menggunakan sistem manual dengan tulis tangan di kertas padahal jumlah data siswa yang mendaftar cukup banyak dan menyebabkan terjadinya kesalahan dan memasukkan data dan waktu yang diperlukan cukup lama dalam mengolah data. Perlu adanya klasifikasi untuk permasalahan tersebut.

Cara untuk melakukan klasifikasi yaitu dengan teknik data mining menggunakan Algoritma C4.5.

Data mining adalah suatu algoritma di dalam menggali informasi berharga yang erpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui [1]. Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan) [2]. Pohon keputusan sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal dan metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan [1]. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan [1]. Nugroho [3] dalam penelitiannya mengenai klasifikasi predikat kelulusan mahasiswa yang menggunakan Algoritma C4.5 menghasilkan akurasi sebesar 73.91%. Swastina [4], dalam penelitian Algoritma Decision Tree C4.5 akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat akurasi 93,31 % dan akurasi rekomendasi jurusan sebesar 82,64%. Maka dapat disimpulkan penggunaan Algoritma C4.5 dalam klasifikasi menghasilkan tingkat akurat yang cukup tinggi dan cocok untuk prediksi jurusan yang peneliti lakukan.

Bedasarkan latarbelakang yang dijelaskan di atas maka penelitian ini dilakukan untuk mengembangkan sistem dengan menerapkan data mining menggunakan Algoritma C4.5 guna memprediksi dan mengklasifikasi penentuan jurusan siswa SMA.

2. ISI PENELITIAN

2.1 Tinjauan Pustaka

2.1.1 Data Mining

Data mining merupakan proses pencarian informasi dari sekumpulan data yang besar dan tidak diketahui sebelumnya. Adapun Karakteristik data mining sebagai berikut [5]:

- Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi

Dalam aplikasinya, data mining sebenarnya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik [6].

Adapun proses KDD sebagai berikut [7]:

- Data Selection*: pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
- Preprocessing*: sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
- Transformation*: yaitu proses coding pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.
- Data mining: proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
- Interpretation / Evaluation*: pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

2.1.2 Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan model prediksi menggunakan struktur pohon, selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami, sehingga *Decision Tree* ini merupakan metode klasifikasi yang paling populer digunakan [5]

2.1.3 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang membangun pohon keputusan dan membentuk model pengetahuan untuk mengklasifikasikan data

dan Algoritma C4.5 memiliki performa tercepat dan memiliki akurasi yang paling tinggi [8].

2.2 Metodologi

Metode penelitian yang digunakan dalam penelitian ini yaitu Pengumpulan data, *Preprocessing* data, Model yang diusulkan, Implementasi sistem Pengujian model, Evaluasi dan validasi model.

2.2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan dari data siswa kelas X SMA 6 Surakarta tahun ajaran 2017/2018. Selain itu ada data yang didapatkan berdasarkan literatur *review* jurnal, buku yang berhubungan dengan permasalahan yang diteliti.

2.2.2 Preprocessing Data

Preprocessing Data merupakan proses awal pengolahan data. Dalam data mining kualitas dari data yang akan digunakan perlu diperhatikan. Ada beberapa faktor *preprocessing* data yang mempengaruhi kualitas data, antara lain keakuratan, keutuhan, konsistensi, aktualitas, dan penafsiran [9] Data yang diperoleh dari pihak SMA 6 Surakarta meliputi nama siswa, nisn, rata-rata nilai raport SMP, minat siswa, dan nilai tes akademik berupa nilai IPA, nilai IPS, dan nilai Bahasa. Kemudian jurusan yang akan dipilih ada 3 yaitu IPA, IPS, dan Bahasa.

2.2.3 Model yang diusulkan

Model yang diusulkan untuk menentukan jurusan siswa SMA yaitu menggunakan Algoritma C4.5. Secara umum Algoritma C4.5 akan membangun pohon keputusan.

Tahapan model Algoritma C4.5 sebagai berikut [4]:

- Menyiapkan data training. Data training digunakan untuk membentuk model yang akan diterapkan pada data uji.
- Menentukan akar dari pohon.
- Menentukan nilai dari *gain*. Perhitungan *gain* dirumuskan pada persamaan berikut.

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

Keterangan:

S = Himpunan kasus

A = Fitur

n = jumlah partisi atribut A

S_i = proporsi S_i terhadap S

S = jumlah kasus dalam S

Sebelum menentukan nilai *gain* terlebih dahulu menghitung nilai *entropy*. Perhitungan *entropy* adalah perhitungan untuk mencari nilai yang nantinya dapat dijadikan cabang dengan nilai 0 (nol) atau nilai yang terendag dari semua atribut [10]. Perhitungan *entropy* dirumuskan pada persamaan berikut.

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (2)$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

P_i = proporsi S_i terhadap S

- 4) Mengulangi langkah 2 hingga semua tupel terpartisi.
- 5) Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama atau tidak ada atribut di dalam tupel yang dipartisi lagi dan atau tidak ada tupel di dalam cabang yang kosong.

2.2.4 Implementasi Sistem

Pada tahap implementasi sistem dibangun sistem menggunakan bahasa pemrograman PHP (*Hypertext Preprocessor*), menggunakan *tools xampp, notepad++* sebagai text editor dan menyimpan database dalam *MySQL*. Pada implementasi sistem diterapkan Algoritma C4.5 berdasarkan rumus yang telah dijelaskan di metode model yang diusulkan. Selain menggunakan sistem, penerapan Algoritma C4.5 dilakukan dengan perhitungan melalui *Microsoft Excel*. Kemudian dibandingkan hasil perhitungan *excel* dan sistem sesuai atau tidak.

2.2.5. Pengujian

Pada tahap pengujian melakukan proses pengukuran data seperti mengukur *recall, precision, dan accuracy*. Pengujian dalam penelitian menggunakan *software pendukung yaitu Rapid Miner 7.6*. Perhitungan *accuracy, recall dan precision* dapat dirumuskan pada persamaan berikut [11].

Precision adalah perhitungan terhadap perkiraan proporsi kasus yang benar dan dirumuskan pada Persamaan 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall adalah perhitungan terhadap perkiraan proporsi kasus positif yang diidentifikasi benar dan dirumuskan pada Persamaan 4.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Accuracy adalah perhitungan terhadap proporsi dari jumlah total prediksi yang benar dan dirumuskan dalam Persamaan 5.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

Keterangan:

TP(*True Positives*) = Jumlah objek positif yang benar diklasifikasikan

TN(*True Negatives*) = Jumlah objek negatif yang salah diklasifikasikan

FP (*False Positives*) = Jumlah objek negatif yang benar diklasifikasikan

FN (*False Negatives*) = Jumlah objek positif yang salah diklasifikasikan

2.3 Hasil Dan Pembahasan

2.3.1 Perhitungan Manual Algoritma C4.5.

Perhitungan manual dilakukan menggunakan aplikasi *Microsoft excel*. Perhitungan ini dilakukan untuk menghitung nilai *gain* dan *entropy* dari tiap atribut. Kemudian dari perhitungan tersebut akan didapatkan pohon keputusan sebagai aturan guna membantu penentuan jurusan bagi siswa. Adapun atribut yang digunakan dalam penentuan jurusan yaitu:

1. Nilai Rata-rata Raport SMP, dikelompokkan dengan variabel Rendah (0-50), Sedang (51-75), Tinggi (76-100).
2. Minat siswa dengan variable IPA, IPS, dan Bahasa.
3. Nilai tes akademik IPS, dikelompokkan dengan variable Rendah (0-50), Sedang (51-75), Tinggi (76-100).
4. Nilai tes akademik IPS, dikelompokkan dengan variable Rendah (0-50), Sedang (51-75), Tinggi (76-100).
5. Nilai tes akademik Bahasa, dikelompokkan dengan variable Rendah (0-50), Sedang (51-75), Tinggi (76-100).

Adapun contoh 10 data *training* dalam sistem yang dikembangkan ditunjukkan pada Tabel 1.

Tabel 1. Contoh Data *Training*

No	Nisn	Nama	Rata-rata Raport	Minat	Nilai IPA	Nilai IPS	Nilai Bahasa	Jurusan
1.	111	Siswa 1	83.34	IPA	76	67	89	IPA
2.	112	Siswa 2	83.87	IPA	80	49	67	IPA
3.	113	Siswa 3	84.76	IPS	76	57	86	IPA
4.	114	Siswa 4	85.12	IPS	56	87	76	IPS
5.	115	Siswa 5	86.98	IPS	47	89	88	Bahasa
6.	116	Siswa 6	86	IPA	87	54	45	IPA
7.	117	Siswa 7	80	IPA	52	83	79	IPS
8.	118	Siswa 8	79.33	Bahasa	56	93	95	Bahasa
9.	119	Siswa 9	80.76	Bahasa	67	67	94	Bahasa
10.	120	Siswa 10	78.56	Bahasa	56	87	91	IPS

Selanjutnya data *training* yang masih memiliki nilai *numeric* diubah menjadi variabel rendah, sedang atau tinggi seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Konversi Data Training

No	Nisn	Nama	Rata-rata Raport	Minat	Nilai IPA	Nilai IPS	Nilai Bahasa	Jurusan
1.	111	Siswa 1	Tinggi	IPA	Tinggi	Sedang	Tinggi	IPA
2.	112	Siswa 2	Tinggi	IPA	Tinggi	Rendah	Sedang	IPA
3.	113	Siswa 3	Tinggi	IPS	Tinggi	Sedang	Tinggi	IPA
4.	114	Siswa 4	Tinggi	IPS	Sedang	Tinggi	Tinggi	IPS
5.	115	Siswa 5	Tinggi	IPS	Rendah	Tinggi	Tinggi	Bahasa
6.	116	Siswa 6	Tinggi	IPA	Tinggi	Sedang	Rendah	IPA
7.	117	Siswa 7	Tinggi	IPA	Sedang	Tinggi	Tinggi	IPS
8.	118	Siswa 8	Tinggi	Bahasa	Sedang	Tinggi	Tinggi	Bahasa
9.	119	Siswa 9	Tinggi	Bahasa	Sedang	Sedang	Tinggi	Bahasa
10.	120	Siswa 10	Tinggi	Bahasa	Sedang	Tinggi	Tinggi	IPS

Selanjutnya melakukan perhitungan *entropy* dan *gain* masing masing atribut. Dari hasil perhitungan *node 1* pada Tabel 3 didapatkan nilai *gain* tertinggi tertinggi 0, 920081259 pada **Nilai IPA**. Nilai IPA dijadikan sebagai akar atau *node* dari pohon keputusan. Nilai IPA memiliki tiga atribut yaitu **rendah**, **sedang**, dan **tinggi**. Ketiga atribut tersebut belum mempunyai keputusan, maka selanjutnya dilakukan perhitungan *entropy* dan *gain* untuk masing-masing atribut sampai menghasilkan keputusan. Perhitungan *node 1* ditunjukkan pada Tabel 3.

Tabel 3. Perhitungan node 1

ATRIBUT	VALUE	JUMLAH KASUS	IPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		196	80	90	26	1,429850385	
MINAT	IPA	71	35	12	4	0,95264362	
	IPS	83	14	66	3	0,8691515	
	Bahasa	42	11	12	19	1,540316914	
							0,386626284
RATA-RATA NILAI	RENDAH	9	0	9	0	0	
	SEDANG	89	30	45	7	1,30544944	
	TINGGI	104	30	35	19	1,484782524	
							0,090036006
NILAI IPA	RENDAH	25	0	10	15	0,970950597	
	SEDANG	96	0	75	10	0,518560732	
	TINGGI	85	80	4	1	0,305223317	
							0,920081259
NILAI IPS	RENDAH	14	13	0	1	0,371232327	
	SEDANG	80	60	4	25	1,099716160	
	TINGGI	99	7	85	0	0,385285119	
							0,721385036
NILAI Bahasa	RENDAH	22	5	17	0	0,77322667428763	
	SEDANG	141	25	65	0	0,997059057	
	TINGGI	33	0	7	26	0,745017843	
							0,500296973

Dari hasil perhitungan *node 1.1* untuk kondisi **Nilai IPA rendah** pada Gambar 5 didapatkan nilai *gain* tertinggi tertinggi 0, 970950594 pada **Nilai IPS**. Ada tiga nilai atribut nilai IPS yaitu **rendah**, **sedang** dan **tinggi**. Nilai atribut **rendah** tidak dapat mengklasifikasikan karena semua atribut bernilai 0. Nilai atribut **sedang** mengklasifikasikan kasus data dan menghasilkan keputusan **Bahasa**. Nilai atribut **tinggi** mengklasifikasikan kasus data dan menghasilkan keputusan **IPS**. Perhitungan *node 1.1* ditunjukkan pada Tabel 4.

Tabel 4. Perhitungan node 1.1

ATRIBUT	VALUE	JUMLAH KASUS	MIPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		25	0	10	15	0,970950594	
MINAT	IPA	1	0	1	0	0	
	IPS	8	0	7	1	0,543564443	
	BB	16	0	2	14	0,543564443	
							0,449128729
RATA-RATA NILAI	RENDAH	4	0	4	0	0	
	SEDANG	1	0	1	0	0	
	TINGGI	20	0	5	15	0,811278124	
							0,321928095
NILAI IPS	RENDAH	0	0	0	0	0	
	SEDANG	15	0	0	15	0	
	TINGGI	10	0	10	0	0	
							0,970950594
NILAI Bahasa	RENDAH	0	0	0	0	0	
	SEDANG	8	0	8	0	0	
	TINGGI	17	0	2	15	0,522559375	
							0,61561022

Dari hasil perhitungan *node 1.2* untuk kondisi Nilai IPA sedang pada Tabel 5 didapatkan nilai *gain* tertinggi tertinggi 0,3896053 pada **Nilai IPS**. Nilai IPS mempunyai atribut **rendah**, **sedang**, dan **tinggi**. Atribut rendah mengklasifikasikan kasus data dan menghasilkan keputusan **Bahasa**. Atribut tinggi mengklasifikasikan kasus data dan menghasilkan keputusan **IPS**. Sedangkan atribut sedang belum menghasilkan keputusan sehingga perlu dilakukan perhitungan *entropy* dan *gain* pada kondisi Nilai IPA sedang dan nilai IPS sedang. Perhitungan *node 1.2* ditunjukkan Pada Tabel 5.

Tabel 5. Perhitungan node 1.2

ATRIBUT	VALUE	JUMLAH KASUS	IPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		85	0	76	10	0,318969752	
MINAT	MIPA	15	0	11	4	0,336640742	
	IPS	57	0	55	2	0,219797308	
	Bahasa	14	0	10	4	0,365120569	
							0,080767806
RATA-RATA NILAI	RENDAH	4	0	4	0	0	
	SEDANG	49	0	42	7	0,391672779	
	TINGGI	35	0	30	5	0,438495987	
							0,072870654
NILAI IPS	RENDAH	1	0	0	1	0	
	SEDANG	13	0	4	9	0,80042164	
	TINGGI	72	0	72	0	0	
							0,3896053
NILAI Bahasa	RENDAH	15	0	16	0	0	
	SEDANG	55	0	55	0	0	
	TINGGI	15	0	5	10	0,918293854	
							0,358401854

Dari hasil perhitungan *node 1.2.1* untuk kondisi Nilai IPA sedang dan Nilai IPS sedang pada Gambar 7 didapatkan nilai *gain* tertinggi tertinggi 0,890491640 pada **Nilai Bahasa**. Ada tiga nilai atribut nilai Bahasa yaitu **rendah**, **sedang** dan **tinggi**. Nilai atribut **rendah** tidak dapat mengklasifikasikan karena semua atribut bernilai 0. Nilai atribut **sedang** mengklasifikasikan kasus data dan menghasilkan keputusan **IPS**. Nilai atribut **tinggi** mengklasifikasikan kasus data dan menghasilkan

keputusan **Bahasa**. Perhitungan *node* 1.2.1 ditunjukkan pada Tabel 6.

Tabel 6. Perhitungan *node* 1.2.1

ATRIBUT	VALUE	JUMLAH KASUS	MIPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		13	0	4	9	0,89049164	
MINAT	MIPA	4	0	0	4	0	
	IPS	5	0	3	2	0,570950594	
	Bahasa	4	0	1	3	0,811278124	
							0,267425066
RATA-RATA NILAI	RENDAH	0	0	0	0	0	
	SEDANG	11	0	4	7	0,530702107	
	TINGGI	2	0	0	2	0	
							0,441435754
NILAI Bahasa	RENDAH	0	0	0	0	0	
	SEDANG	4	0	4	0	0	
	TINGGI	9	0	0	9	0	
							0,890491640

Dari hasil perhitungan *node* 1.3 untuk kondisi Nilai IPA tinggi pada Gambar 5 didapatkan nilai *gain* tertinggi tertinggi 0,156237374 pada Nilai IPS. Ada tiga nilai atribut nilai IPS yaitu rendah, sedang dan tinggi. Nilai atribut rendah mengklasifikasikan kasus data dan menghasilkan keputusan IPA. Sedangkan atribut sedang dan atribut tinggi belum dapat mengklasifikasikan kasus data dan belum menghasilkan keputusan. Untuk itu dilanjutkan perhitungan *entropy* dan *gain* pada kondisi Nilai IPA tinggi dan Nilai IPS sedang, dan kondisi Nilai IPA tinggi dan Nilai IPS tinggi. Perhitungan *node* 1.3 ditunjukkan pada Tabel 7.

Tabel 7. Perhitungan *node* 1.3

ATRIBUT	VALUE	JUMLAH KASUS	MIPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		55	30	4	1	0,265223317	
MINAT	MIPA	55	55	0	0	0	
	IPS	18	14	4	0	0,761204507	
	Bahasa	12	11	0	1	0,41381685	
							0,144970572
RATA-RATA NILAI	RENDAH	1	0	1	0	0	
	SEDANG	53	30	3	0	0,439196987	
	TINGGI	51	50	0	1	0,139232998	
							0,111055276
NILAI IPS	RENDAH	13	13	0	0	0	
	SEDANG	51	60	0	1	0,120581014	
	TINGGI	11	7	4	0	0,915560305	
							0,156237374
NILAI Bahasa	RENDAH	6	5	1	0	0,650022422	
	SEDANG	78	75	3	0	0,235193387	
	TINGGI	1	0	0	1	0	
							0,103514867

Dari hasil perhitungan *node* 1.3.1 untuk kondisi Nilai IPA tinggi dan Nilai IPS sedang pada Gambar 9 didapatkan nilai *gain* tertinggi 0,120681014 pada Nilai Bahasa. Ada tiga nilai atribut nilai Bahasa yaitu rendah, sedang dan tinggi. Nilai atribut rendah mengklasifikasikan kasus data dan menghasilkan keputusan IPA. Nilai atribut sedang mengklasifikasikan kasus data dan menghasilkan keputusan IPA. Nilai atribut tinggi mengklasifikasikan kasus data dan menghasilkan

keputusan **Bahasa**. Perhitungan *node* 1.3.1 ditunjukkan pada Tabel 8.

Tabel 8. Perhitungan *node* 1.3.1

ATRIBUT	VALUE	JUMLAH KASUS	MIPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		61	50	0	1	0,120581014	
MINAT	MIPA	57	37	0	0	0	
	IPS	17	17	0	0	0	
	Bahasa	12	11	0	1	0,41381685	
							0,039274421
RATA-RATA NILAI	RENDAH	0	0	0	0	0	
	SEDANG	29	29	0	0	0	
	TINGGI	32	31	0	1	0,200522324	
							0,015436516
NILAI Bahasa	RENDAH	5	5	0	0	0	
	SEDANG	55	55	0	0	0	
	TINGGI	1	0	0	1	0	
							0,120581014

Dari hasil perhitungan *node* 1.3.2 untuk kondisi Nilai IPA tinggi dan Nilai IPS tinggi pada Gambar 10 didapatkan nilai *gain* tertinggi 0, pada Minat. Ada tiga nilai atribut nilai Minat yaitu IPA, IPS, dan Bahasa. Minat dengan atribut IPA mengklasifikasikan kasus data dan menghasilkan keputusan IPA. Minat dengan atribut IPS mengklasifikasikan kasus data dan menghasilkan keputusan IPS. Sedangkan minat dengan atribut Bahasa tidak mengklasifikasikan kasus data dan tidak menghasilkan keputusan karena mempunyai nilai 0 semua. Perhitungan *node* 1.3.2 ditunjukkan pada Tabel 9.

Tabel 9. Perhitungan *node* 1.3.2

ATRIBUT	VALUE	JUMLAH KASUS	MIPA	IPS	Bahasa	ENTROPY	INFORMATION GAIN
TOTAL		11	7	4	0	0	
MINAT	MIPA	7	7	0	0	0	
	IPS	4	0	4	0	0	
	Bahasa	0	0	0	0	0	
							0,34586030460354
RATA-RATA NILAI	RENDAH	1	0	1	0	0	
	SEDANG	4	1	3	0	0,81127812445913	
	TINGGI	6	6	0	0	0	
							0,53055007732459
NILAI Bahasa	RENDAH	1	0	1	0	0	
	SEDANG	10	7	3	0	0,88125089923069	
	TINGGI	0	0	0	0	0	
							0,1448675834547

Dari semua perhitungan *entropy* dan *gain* tiap atribut dapat dibuat *rule* pohon keputusan sebagai berikut:

1. Jika Nilai IPA rendah dan Nilai IPS sedang maka jurusan Bahasa.
2. Jika Nilai IPA rendah dan Nilai IPS tinggi maka jurusan IPS.
3. Jika Nilai IPA=sedang dan Nilai IPS=rendah maka jurusan Bahasa.
4. Jika Nilai IPA=sedang dan Nilai IPS=sedang dan Nilai Bahasa=sedang maka jurusan IPS.

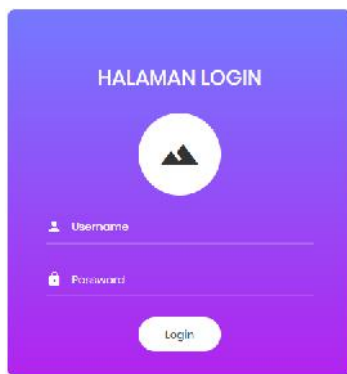
5. Jika Nilai IPA=sedang dan Nilai IPS=sedang dan Nilai Bahasa=sedang maka jurusan Bahasa.
6. Jika Nilai IPA=sedang dan Nilai IPS=tinggi maka jurusan IPS.
7. Jika Nilai IPA=tinggi dan Nilai IPS=rendah maka jurusan IPA.
8. Jika Nilai IPA=tinggi dan Nilai IPS=sedang dan Nilai Bahasa rendah maka jurusan IPA.
9. Jika Nilai IPA=tinggi dan Nilai IPS=sedang dan Nilai Bahasa sedang maka jurusan IPA.
10. Jika Nilai IPA=tinggi dan Nilai IPS=sedang dan Nilai Bahasa tinggi maka jurusan Bahasa.
11. Jika Nilai IPA=tinggi dan Nilai IPS=tinggi dan minat=IPA maka jurusan IPA.
12. Jika Nilai IPA=tinggi dan Nilai IPS=tinggi dan minat=IPS maka jurusan IPS.

2.3.2 Implementasi Sistem

Implementasi sistem merupakan bentuk pelaksanaan atau penerapan dari model Algoritma C4.5 dalam bentuk sistem berbasis *website*. Hasil yang didapatkan dalam penelitian ini berupa sebuah sistem penentuan jurusan siswa menggunakan Algoritma C4.5. Sistem yang dikembangkan mempunyai beberapa menu seperti menu *Home*, menu *Data Training*, menu *Data Testing*, menu *Perhitungan C4.5* dan menu *Prediksi*. Sebelum admin dapat mengakses meu, sebelumnya admin melakukan *login* dengan memasukkan *username* dan *password*.

a. Halaman *Login*

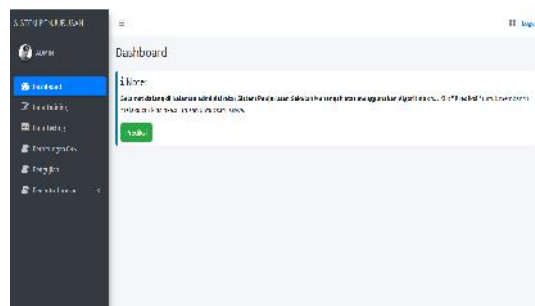
Halaman *Login* digunakan sebagai langkah awal admin agar dapat masuk ke menu yang disediakan dalam sistem. Halaman *Login* dapat dilihat pada Gambar 1.



Gambar 1. Halaman *Login*

b. Halaman *Home*

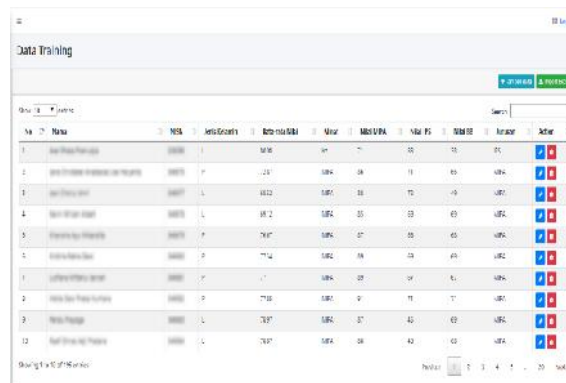
Halaman *Home* merupakan halaman awal setelah admin berhasil melakukan *login*. Pada halaman ini ada menu seperti menu *Data Training*, menu *Data Testing*, menu *Perhitungan C4.5*, menu *Prediksi* dan tombol *logout*. Halaman *Home* dapat dilihat pada Gambar 2.



Gambar 2. Halaman *Home*

c. Halaman *Data Training*

Halaman data *training* berfungsi menampilkan daftar data *training* yang telah ditambahkan. Pada halaman ini dapat menambahkan data dengan mengisi *form* yang disediakan ketika mengakses pada tombol *tambah data* atau dapat menambahkan data melalui tombol *import excel*. Selain itu data yang telah ditambahkan dapat diubah dan dihapus jika terdapat kesalahan dalam penulisannya. Halaman *Data Training* dapat dilihat pada Gambar 3.



Gambar 3. Halaman *Data Training*

d. Halaman *Data Testing*

Halaman data *testing* berfungsi menampilkan daftar data *testing* yang telah ditambahkan. Pada halaman ini dapat menambahkan data dengan mengisi *form* yang disediakan ketika mengakses pada tombol *tambah data* atau dapat menambahkan data melalui tombol *import excel*. Selain itu data yang telah ditambahkan dapat diubah dan dihapus jika terdapat kesalahan dalam penulisannya. Halaman *Data Testing* dapat dilihat pada Gambar 4.

No	Nama	Nilai	Jenis Kelamin	Nilai Ujian	Minat	Nilai IPA	Nilai IPS	Nilai SB	Amatur	Prediksi	Aktif
1	Andriani	8000	F	27	4000	80	20	8000	80	20	20
2	Andriani	8000	L	42	4000	70	20	20	8000	80	20
3	Andriani	8000	L	42	4000	70	20	20	8000	80	20
4	Andriani	8000	F	25	4000	80	20	8000	80	20	20
5	Andriani	8000	F	42	4000	70	20	20	8000	80	20
6	Andriani	8000	L	42	4000	70	20	20	8000	80	20
7	Andriani	8000	F	22	4000	80	20	8000	80	20	20
8	Andriani	8000	F	28	4000	80	20	8000	80	20	20
9	Andriani	8000	F	42	4000	70	20	20	8000	80	20
10	Andriani	8000	F	15	4000	80	20	8000	80	20	20

Gambar 4. Halaman Data Testing

e. Halaman Perhitungan C4.5.

Halaman Perhitungan C4.5 menampilkan hasil perhitungan data *training* menggunakan perhitungan Algoritma C4.5. Perhitungan algoritma ini meliputi perhitungan *gain* dan *entropy* dari setiap atribut untuk membentuk pohon keputusan. Kemudian setelah semua atribut memiliki pohon keputusan maka dibentuk *rule* pohon keputusan guna diterapkan dalam data *testing* untuk melakukan proses penentuan jurusan. Halaman Perhitungan C4.5 dapat dilihat pada Gambar 5. Sedangkan halaman *rule* pohon keputusan dapat dilihat pada Gambar 6.

Nama Atribut	Entropi	Gain	Entropi	Gain
Nilai Ujian	0.9183	0.0817	0.9183	0.0817
Minat	0.9183	0.0817	0.9183	0.0817
Nilai IPA	0.9183	0.0817	0.9183	0.0817
Nilai IPS	0.9183	0.0817	0.9183	0.0817
Nilai SB	0.9183	0.0817	0.9183	0.0817
Amatur	0.9183	0.0817	0.9183	0.0817

Gambar 5. Halaman Perhitungan C4.5

No	Aturan	Prediksi
1	IF (Nilai Ujian <= 27) THEN Jurusan = IPS	IPS
2	IF (Nilai Ujian > 27) THEN Jurusan = IPS	IPS
3	IF (Minat <= 42) THEN Jurusan = IPS	IPS
4	IF (Minat > 42) THEN Jurusan = IPS	IPS
5	IF (Nilai IPA <= 80) THEN Jurusan = IPS	IPS
6	IF (Nilai IPA > 80) THEN Jurusan = IPS	IPS
7	IF (Nilai IPS <= 70) THEN Jurusan = IPS	IPS
8	IF (Nilai IPS > 70) THEN Jurusan = IPS	IPS
9	IF (Nilai SB <= 20) THEN Jurusan = IPS	IPS
10	IF (Nilai SB > 20) THEN Jurusan = IPS	IPS
11	IF (Amatur <= 20) THEN Jurusan = IPS	IPS
12	IF (Amatur > 20) THEN Jurusan = IPS	IPS

Gambar 6. Halaman Rule Pohon Keputusan

f. Halaman Prediksi

Halaman prediksi berfungsi untuk melakukan prediksi jurusan terhadap satu siswa dimana perlu adanya *input* data seperti nama, nisn, minat, nilai rata-rata raport, nilai IPA, nilai IPS, dan nilai Bahasa. Setelah adanya *input* data maka diklasifikasikan jurusan berdasarkan model pengetahuan *rule* yang didapat. Halaman *input* prediksi dapat dilihat pada Gambar 7 dan halaman prediksi jurusan ditunjukkan pada Gambar 8. Sedangkan untuk melihat daftar data hasil prediksi ditunjukkan pada Gambar 18.

Gamabr 7. Halaman input prediksi

Gambar 8. Halaman Prediksi Jurusan

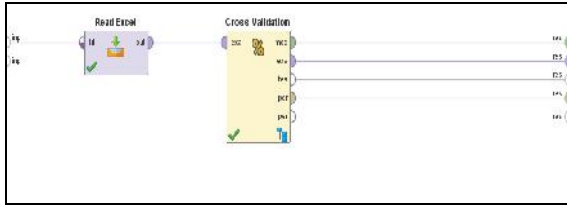
No	Nama	Nilai	Jenis Kelamin	Nilai Ujian	Minat	Nilai IPA	Nilai IPS	Nilai SB	Prediksi	Action
1	Andriani	8000	F	27	4000	80	20	8000	IPS	20
2	Andriani	8000	L	42	4000	70	20	20	IPS	20

Gambar 9. Halaman Daftar Data Prediksi

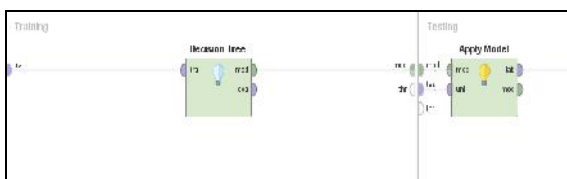
2.4 Pengujian

Pengujian dilakukan dengan *software rapid miner* untuk mengetahui nilai *accuracy*, *recall* dan *precision* dengan Algoritma C4.5. Data yang digunakan dalam pengujian adalah data *testing* yang terdiri dari 196 *record* dan 5

atribut. Disini penulis menggunakan *cross validation* sebagai proses validasi. Kemudian akan diperoleh hasil *accuracy*, *recall* dan *precision*.



Gambar 10. Cross Validation dengan Algoritma C4.5



Gambar 11. Cross Validation dengan Algoritma C4.5

accuracy: 97,42% # 3,45% (mikro: 97,45%)				
	luc BB	luc PS	luc MPA	luc total
pred: LU	21	0	0	100,00%
pred: PS	4	30	0	92,74%
pred: MPA	1	0	30	99,77%
class recall	80,77%	100,00%	100,00%	

Gambar 12. Perhitungan Accuracy, Recall, dan Precision

Dari perhitungan menggunakan *Rapid Miner* menghasilkan nilai *accuracy* sebesar 97.42%.

$$\text{Nilai recall} = \frac{\text{recall Bahasa} + \text{recall IPS} + \text{recall IPA}}{3} = \frac{80.77 + 100 + 100}{3} = 93.59\%$$

$$\text{Nilai precision}(P) = \frac{P.\text{Bahasa} + P.\text{IPS} + P.\text{IPA}}{3} = \frac{100 + 95.74 + 98.77}{3} = 98.17\%$$

3 PENUTUP

Dari hasil penilitan yang dilakukan dapat ditarik kesimpulan sebagai berikut:

- Algoritma C4.5 dapat memudahkan dalam pengambilan keputusan dalam penentuan jurusan siswa dan dapat mengklasifikasikan 3 kelas jurusan yaitu IPA, IPS, dan Bahasa.

- Perhitungan Algoritma C4.5 menggunakan *Excel* dan menggunakan aplikasi sistem menghasilkan *rule* pohon keputusan yang sama.
- Rule* yang didapatkan dari perhitungan Algoritma C4.5 berjumlah 12 buah dan digunakan untuk proses penentuan jurusan siswa.
- Penerapan Algoritma C4.5 menghasilkan Nilai Accuracy dengan *Rapid Miner* sebesar 97,42 %, menghasilkan *precision* sebesar 98,17% dan *recall* sebesar 93,59%.

Adapun saran dalam pengembangan sistem untuk hasil prediksi yang lebih akurat dibutuhkan data dalam jumlah yang besar agar semakin akurat dan pengetahuan yang dihasilkan menjadi lebih baik.

DAFTAR PUSTAKA

- I. Sujai, Purwanto, and H. Himawan, "Prediksi Hasil Jurusan Siswa Sekolah Menengah Atas Dengan Menggunakan Algoritma Decision Tree C4.5," vol. 12, no. April, pp. 42–53, 2016.
- N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018.
- Y. S. Nugroho, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta," no. November, pp. 1–6, 2014.
- L. Swastina, "Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa," vol. 2, no. 1, 2013.
- S. L. B. Ginting, W. Zarman, and I. Hamidah, "Analisis dan Penerapan Algoritma C4.5 dalam Data Mining untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik," *Pros. Semin. Nas. Apl. Sains Teknol.*, vol. 3, no. November, pp. 211–216, 2014.
- A. G. Maburur and L. Riani, "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 1, no. 1, pp. 53–57, 2012.
- M. Agushinta, D., Irfan, "Perancangan Aplikasi Data Mining Untuk Memprediksi Permintaan Customer Pada Perusahaan Persewaan Mobil, Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT)," *Univ. Gunadarma*, no. Kommit, pp. 20–21, 2008.
- R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan*

- Inform.*, vol. 4, no. 1, p. 1, 2018.
- [9] D. Dharmayanti, A. M. Bachtiar, A. C. Prasetyo, T. Informatika, U. Komputer, and J. D. Bandung, "Penerapan Metode Clustering untuk Membentuk Kelompok Belajar Menggunakan di SMPN 19 Bandung," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 6, no. 2, 2017.
- [10] S. M. Khoiroh, "Implementasi Data Mining dengan Algoritma C4.5 untuk memprediksi Prestasi Siswa," no. 6, pp. 67–72, 2017.
- [11] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," vol. 1–9, p. 55, 2015.