



## **Menentukan Penjurusan Siswa Dengan Menggunakan Metode *Decision Tree* Algoritma C4.5 (Studi Kasus : SMA Negeri 2 Padang)**

<sup>1,\*</sup>Muhammad Fibo Donya Ikhbal, <sup>2</sup>Denny Kurniadi

<sup>1,2</sup>Teknik Elektronika, Fakultas Teknik, Universitas Negeri Padang

\*Corresponding author e-mail: fibodonya@gmail.com

### **Abstrak**

Pendidikan menjadi salah satu hal terpenting dalam sumber daya manusia, terlebih anak-anak yang melalui pendidikan selama 9 tahun, dengan melewati jenjang Pendidikan SD, SMP, dan SMA. Setelah menyelesaikan Sekolah Menengah Pertama (SMP) selanjutnya akan ke jenjang yang lebih tinggi yaitu SMA yang mewajibkan siswa tersebut memilih satu jurusan untuk masa depan siswa itu sendiri. SMA N 2 Padang merupakan sekolah yang penyimpanan datanya telah terkomputerisasi, salah satunya data penjurusan siswa. Penjurusan siswa SMA N 2 Padang dilakukan pada awal tahun pelajaran sesuai dengan kurikulum 2013 yang memiliki 2 jurusan yaitu IPA dan IPS, yang bertujuan untuk menyalurkan kemampuan dan minat bakat siswa tersebut, sering kali siswa asal dalam memilih atau menentukan penjurusan dan untuk kedepannya siswa tersebut akan kebingungan dalam memilih jurusan di perguruan tinggi sehingga dapat menghambat cita-cita siswa tersebut, oleh karena itu diperlukan klasifikasi penjurusan yang tepat dan akurat untuk mempermudah dalam pengklasifikasian salah satunya menggunakan teknologi dibidang data mining. Dalam menentukan penjurusan siswa selama ini belum diterapkannya proses klasifikasi jurusan dengan akurasi nya juga belum diketahui, penelitian ini bertujuan untuk mengklasifikasi penjurusan siswa dan menerapkan metode *data mining* menggunakan algoritma *decision tree* C4.5 dan juga aplikasi Rapidminer. Klasifikasi ini menggunakan 6 atribut sebagai atribut kontrol yang merupakan nilai rata-rata rapor SMP yang meliputi mata pelajaran B. Indonesia, B. Inggris, Matematika dan IPA, serta rekomendasi guru BK SMP. Berdasarkan pengujian yang dilakukan terhadap siswa tahun masuk ajaran 2020/2021 diperoleh akurasi sebesar 68.42% dengan 304 sampel data dengan atribut rekomendasi guru BK SMP yang mempengaruhi keputusan penjurusan siswa. Demikian dapat disimpulkan bahwa menentukan penjurusan dengan menggunakan data mining algoritma *decision tree* C4.5 dapat mempercepat dan akurat dalam pengambilan keputusan penjurusan siswa SMA N 2 Padang.

**Kata kunci :** SMAN 2 Padang, penjurusan siswa, *data mining*, klasifikasi, algoritma C4.5, *decision tree*.



Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

## 1. Pendahuluan

Perkembangan teknologi sekarang, banyaknya data menjadi masalah dan juga peluang bagi sebuah instansi. Sementara itu data menjadi peluang jika dapat disimpan dan diolah agar lebih bermakna bagi instansi. Dengan data, seseorang dapat menemukan data dan struktur yang bisa digunakan untuk memperoleh informasi di saat tertentu.

Tujuan dari jurusan ini adalah untuk membuat pelajaran yang ditawarkan kepada siswa lebih fokus ke masa depan, karena tidak jarang siswa ceroboh dalam memilih jurusan yang akan diambil dan kendala lain datang dari dalam atau luar siswa itu sendiri. Kecendrungan yang selalu ada sekarang, siswa bingung atau ragu dalam memilih minat dan bakatnya juga memilih jurusan apa [1]. Faktor dari luar sering kali berpengaruh dalam memilih jurusan untuk siswa oleh orang tuamereka, dan paksaan itu biasanya karena harapan orang tua terhadap pekerjaan anak-anak mereka di masa depan[2].

Data mining membutuhkan *software* untuk menganalisis klasifikasi akurat dengan mencari pola dan relasi data [4].

Berdasarkan pembahasan diatas, penerapan data mining menggunakan algoritma C4.5 terhadap siswa SMA N 2 Padang untuk menentukan penjurusan dengan atribut rata-rata nilai rapor SMP dan rekomendasi BK SMP.

Sekolah SMA N 2 Padang dapat menentukan penjurusan siswa yang tepat apabila memenuhi faktor-faktor yang telah ditentukan. Beberapa faktor yang telah ditentukan antara lain:

- a. Nilai rapor SMP, berdasarkan kurikulum 2013 penjurusan siswa dimulai pada kelas X, nilai rapor SMP siswa diharapkan menjadi pertimbangan untuk menentukan penjurusan yang tepat, beberapa siswa memiliki nilai yang rendah tapi menunjukkan grafik peningkatan setiap kenaikan semester.
- b. Rekomendasi guru bk SMP, peran dari guru bk SMP sangat penting dan berpengaruh dalam memilih jurusan yang akan diambil siswa. Karena siswa sering merasa kebingungan dalam memilih jurusan yang tepat dan kebingungan kemana arah karir setelah siswa itu tamat.

SMA N 2 Padang merupakan sekolah yang penyimpanan datanya telah terkomputerisasi, salah satunya data penjurusan siswa. Penjurusan siswa SMA N 2 Padang dilakukan pada awal tahun pelajaran sesuai dengan kurikulum 2013 yang memiliki 2 jurusan yaitu IPA dan IPS, sering kali siswa asal dalam memilih atau menentukan penjurusan dan untuk kedepannya siswa tersebut akan kebingungan dalam memilih jurusan di perguruan tinggi sehingga dapat menghambat cita-cita siswa tersebut. Penelitian ini bertujuan untuk mengklasifikasi penjurusan siswa dan menerapkan

metode *data mining* menggunakan algoritma C4.5 dan aplikasi Rapidminer. Terdapat 6 atribut sebagai atribut control untuk mengklasifikasi ini yang merupakan nilai rata-rata rapor SMP dan rekomendasi guru BK SMP. Berdasarkan pengujian yang dilakukan terhadap siswa tahun masuk ajaran 2020/2021 diperoleh akurasi sebesar 68.42% dengan 304 sampel data.

## 2. Tinjauan Pustaka

### 2.1. Decision Tree

*Decision Tree* merupakan salah satu metode klasifikasi yang paling populer, karena mudah diinterpretasikan oleh manusia. Pohon keputusan adalah model prediksi dengan menggunakan struktur pohon atau struktur hierarki. Konsep pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan keputusan. Manfaat utama menggunakan pohon keputusan adalah kemampuannya untuk memecah proses pengambilan keputusan yang kompleks menjadi proses yang lebih sederhana, sehingga pengambil keputusan akan lebih baik dalam menginterpretasikan solusi untuk masalah.

Nama lain dari pohon keputusan adalah CART (*Classification and Regression Tree*). Dimana metode ini merupakan gabungan dari dua jenis pohon yaitu pohon klasifikasi dan pohon regresi. Untuk kenyamanan, berikut adalah ilustrasi keduanya.

### 2.2. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*Decision Tree*). Pohon keputusan adalah metode klasifikasi dan prediksi yang terkenal.

Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input potensial dan variabel target.

## 3. Metode Penelitian

Pada penelitian ini alur penelitian yang akan dilakukan dengan studi pustaka, pengumpulan data, pra-proses data, klasifikasi dengan algoritma C4.5, dan membuat kesimpulan penelitian. Penelitian ini dilakukan dengan mengklasifikasi penjurusan siswa dengan beberapa tahap seperti :

1. proses analisis data
2. perhitungan algoritma C4.5
3. pengujian menggunakan *tools RapidMiner*.

### Desain Penelitian

Proses penelitian terdiri dari beberapa desain penelitian yaitu:

### 3.1. Studi Pustaka

Hal pertama yang akan dilakukan yaitu mengumpulkan data berguna untuk sampel penelitian seperti sumber beberapa buku, karya ilmiah, dan jurnal.

### 3.2. Pengumpulan Data

Penulis melakukan pengumpulan data siswa tahun ajaran 2020/2021 yang diperoleh dari data sekolah SMA N 2 Padang.

### 3.3. Praproses

Praproses data tahapan metode klasifikasi meliputi:

#### 3.3.1. Seleksi Data

Tahap ini memilih atribut yang digunakan pada penelitian ini, atribut yang akan digunakan dipilih sesuai dengan variabel yang dibutuhkan.

#### 3.3.2. Pembersihan Data

Adalah proses menghilangkan data yang *null* ataupun data yang ganda. Pembersihan data mempengaruhi performansi teknik data mining dikarenakan data akan berkurang jumlah.

#### 3.3.3. Transformasi Data

Untuk format data agar diproses dan diaplikasikan dalam bentuk data mining, dataset akan diperoleh setelah semua data akan dilakukan pengolahan data. Pada proses ini akan dilakukan transformasi dengan mengubah nilai atribut data kedalam bentuk data sesuai dengan menggunakan algoritma C4.5 agar bisa diproses, sehingga akan mendapatkan dataset yang digunakan pada tahap selanjutnya. Transformasi data juga dilakukan agar bisa diproses dan diaplikasikan pada *tools RapidMiner*.

#### 3.3.4. Split Validation (Validasi Terpisah)

*Split validaton* merupakan teknik membagi dua data secara terpisah dengan secara acak, dengan *data testing dan data training*. Pada pembagian data dibagi dengan *training* diambil pada tahun masuk ajaran 2019/2020 yang berjumlah 231 *record* dan *data testing* diambil pada tahun masuk ajaran 2020/2021 yang berjumlah 304 *record*.

#### 3.3.5. Data Uji

Data uji merupakan data yang disiapkan untuk dilakukan pengujian, setelah itu dilakukan pengkategorian dengan variabel dan atribut yang sudah ditentukan.

#### 3.3.6. Proses Analisis Data

Dalam analisis data peneliti menggunakan data latih (*data training*) berupa data siswa tahun masuk ajaran 2020/2021 yang diperoleh dari arsip data siswa kelas X tahun ajaran 2020/2021 SMA N 2 Padang, data tersebut diklasifikasikan menggunakan algoritma C4.5, sehingga dari proses

klasifikasinya memperoleh suatu model untuk penerapannya pada data uji.

#### 3.3.7. Perhitungan Algoritma C4.5

Algoritma ini lebih cocok untuk menentukan masalah klasifikasi pada machine learning dan data mining [6]. *Decision tree* yaitu konsep data atau struktur dari data yang besar ke data yang paling kecil dengan cara menginterpretasikannya [7]. Dalam algoritma C4.5, ada beberapa langkah dalam merancang *decision tree*, yaitu:

- Siapkan data yang akan digunakan untuk dijadikan data pelatihan (data *training*).
- menentukan akar untuk membentuk *decision tree*
- mencari nilai *entropy* dan *gain* yang sudah didapatkan menjadi dataset.

## 4. Hasil dan Pembahasan

Atribut dalam penelitian yang digunakan adalah:

Tabel 1. Atribut penelitian

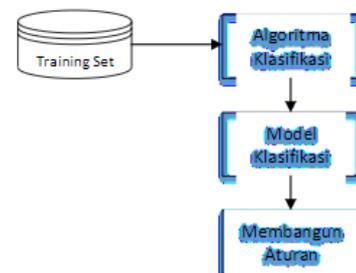
Atribut	Penjelasan
Nama	Nama Siswa
X1	Nilai B. Indonesia
X2	Nilai Matematika
X3	Nilai B. Inggris
X4	Nilai IPA
X5	Rekomendasi Guru BK SMP
Jurusan	IPA/IPS

Klasifikasi dengan Algoritma C4.5

Proses klasifikasi dibagi menjadi 2 tahapan [9], diantaranya sebagai berikut:

#### 4.1. Tahapan membangun model

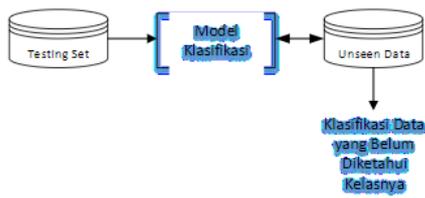
Tahap ini berdasarkan data yang sudah ditetapkan kelasnya menghasilkan model klasifikasi.



Gambar 1. Tahapan membangun model

#### 4.2. Tahapan model klasifikasi

Proses model klasifikasi dari dalam data menggunakan data *testing* atau disebut juga dengan proses deduksi seperti pada gambar 2.



Gambar 2. Tahapan menggunakan model

#### 4.2.1. Menghitung nilai Entropy dan Gain

Berdasarkan data training yang digunakan, terdapat 2 kelas klasifikasi yang dibentuk dari probabilitas kelas tepat waktu, yang dapat dilihat di bawah ini:

#### 4.2.2. Menghitung nilai entropy

Entropy adalah parameter yang mengukur tingkat keadaan atau keberagaman dari sebuah kumpulan data, rumus untuk mencari nilai entropy adalah:

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

n: jumlah partisi S

A: fitur

S: himpunan kasus

- a. Entropy total variabel atribut jurusan adalah:

$$Entropy = \left( -\frac{203}{231} \cdot \log_2 \left( \frac{203}{231} \right) + \left( -\frac{28}{231} \cdot \log_2 \left( \frac{28}{231} \right) \right) \right) = 0.53284$$

- b. Entropy total variabel atribut B. Indonesia

$$Entropy(A) = \left( -\frac{121}{142} \cdot \log_2 \left( \frac{121}{142} \right) + \left( -\frac{21}{142} \cdot \log_2 \left( \frac{21}{142} \right) \right) \right) = 0.60453$$

$$Entropy(B) = \left( -\frac{81}{88} \cdot \log_2 \left( \frac{81}{88} \right) + \left( -\frac{7}{88} \cdot \log_2 \left( \frac{7}{88} \right) \right) \right) = 0.40058$$

$$Entropy(C) = \left( -\frac{1}{1} \cdot \log_2 \left( \frac{1}{1} \right) + \left( -\frac{0}{1} \cdot \log_2 \left( \frac{0}{1} \right) \right) \right) = 0$$

$$Entropy(D) = \left( -\frac{0}{0} \cdot \log_2 \left( \frac{0}{0} \right) + \left( -\frac{0}{0} \cdot \log_2 \left( \frac{0}{0} \right) \right) \right) = 0$$

- c. Entropy total variabel atribut Matematika adalah:

$$Entropy(A) = \left( -\frac{69}{77} \cdot \log_2 \left( \frac{69}{77} \right) + \left( -\frac{8}{77} \cdot \log_2 \left( \frac{8}{77} \right) \right) \right) = 0.48123$$

$$Entropy(B) = \left( -\frac{109}{115} \cdot \log_2 \left( \frac{109}{115} \right) + \left( -\frac{6}{115} \cdot \log_2 \left( \frac{6}{115} \right) \right) \right) = 0.29556$$

$$Entropy(C) = \left( -\frac{23}{34} \cdot \log_2 \left( \frac{23}{34} \right) + \left( -\frac{11}{34} \cdot \log_2 \left( \frac{11}{34} \right) \right) \right) = 0.90818$$

$$Entropy(D) = \left( -\frac{2}{5} \cdot \log_2 \left( \frac{2}{5} \right) + \left( -\frac{21}{5} \cdot \log_2 \left( \frac{3}{5} \right) \right) \right) = 0.97095$$

- d. Entropy total variabel atribut B. Inggris adalah:

$$Entropy(A) = \left( -\frac{64}{78} \cdot \log_2 \left( \frac{64}{78} \right) + \left( -\frac{14}{78} \cdot \log_2 \left( \frac{14}{78} \right) \right) \right) = 0.67895$$

$$Entropy(B) = \left( -\frac{89}{98} \cdot \log_2 \left( \frac{89}{98} \right) + \left( -\frac{9}{98} \cdot \log_2 \left( \frac{9}{98} \right) \right) \right) = 0.44257$$

$$Entropy(C) = \left( -\frac{48}{53} \cdot \log_2 \left( \frac{48}{53} \right) + \left( -\frac{5}{53} \cdot \log_2 \left( \frac{5}{53} \right) \right) \right) = 0.97095$$

$$Entropy(D) = \left( -\frac{2}{2} \cdot \log_2 \left( \frac{2}{2} \right) + \left( -\frac{0}{2} \cdot \log_2 \left( \frac{0}{2} \right) \right) \right) = 0$$

- e. Entropy total variabel atribut IPA adalah:

$$Entropy(A) = \left( -\frac{22}{25} \cdot \log_2 \left( \frac{22}{25} \right) + \left( -\frac{3}{25} \cdot \log_2 \left( \frac{3}{25} \right) \right) \right) = 0.52936$$

$$Entropy(B) = \left( -\frac{142}{159} \cdot \log_2 \left( \frac{142}{159} \right) + \left( -\frac{17}{159} \cdot \log_2 \left( \frac{17}{159} \right) \right) \right) = 0.49055$$

$$Entropy(C) = \left( -\frac{39}{47} \cdot \log_2 \left( \frac{39}{47} \right) + \left( -\frac{8}{47} \cdot \log_2 \left( \frac{8}{47} \right) \right) \right) = 0.65819$$

$$Entropy(D) = \left( -\frac{0}{0} \cdot \log_2 \left( \frac{0}{0} \right) + \left( -\frac{0}{0} \cdot \log_2 \left( \frac{0}{0} \right) \right) \right) = 0$$

- f. Entropy total variabel atribut Rekomendasi guru BK SMP adalah:

$$Entropy(A) = \left( -\frac{195}{198} \cdot \log_2 \left( \frac{195}{198} \right) + \left( -\frac{3}{198} \cdot \log_2 \left( \frac{3}{198} \right) \right) \right) = 0.11327$$

$$Entropy(B) = \left( -\frac{8}{32} \cdot \log_2 \left( \frac{8}{32} \right) + \left( -\frac{25}{32} \cdot \log_2 \left( \frac{25}{32} \right) \right) \right) = 0.79905$$

#### 4.2.3. Menghitung nilai Gain

Nilai gain berfungsi untuk mengukur efektifitas suatu atribut dengan mengklasifikasikan data, rumusnya yaitu:

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

- a. Gain total variabel atribut B. Indonesia adalah:

$$Gain(\text{total}) = 0.53484 - \left( \left( \frac{142}{231} \cdot 0.60453 \right) - \left( \left( \frac{88}{231} \cdot 0.40058 \right) - \left( \left( \frac{1}{231} \cdot 0 \right) - \left( \frac{0}{0} \cdot 0 \right) \right) \right) \right) = 0.00862$$

- b. Gain total variabel atribut Matematika adalah:

$$Gain(\text{total}) = 0.53484 - \left( \left( \frac{77}{231} \cdot 0.48123 \right) - \left( \left( \frac{115}{231} \cdot 0.29556 \right) - \left( \left( \frac{34}{231} \cdot 0.90818 \right) - \left( \frac{5}{231} \cdot 0.97095 \right) \right) \right) \right) = 0.0706$$

- c. Gain total variabel atribut B. Inggris adalah:

$$\begin{aligned} \text{Gain (total)} &= 0.53284 - \left( \left( \frac{78}{231} * 0.67895 \right) \right. \\ &- \left( \left( \frac{98}{231} * 0.44257 \right) - \left( \frac{53}{231} * 0.45079 \right) - \right. \\ &\left. \left. \left( \frac{2}{231} * 0 \right) \right) \right) = 0.01239 \end{aligned}$$

d. *Gain* total variabel atribut IPA adalah:

$$\begin{aligned} \text{Gain (total)} &= 0.53284 - \left( \left( \frac{25}{231} * 0.52936 \right) - \right. \\ &\left( \left( \frac{159}{231} * 0.49055 \right) - \left( \frac{47}{231} * 0.65819 \right) - \left( \frac{0}{231} * 0 \right) \right) = 0.00398 \end{aligned}$$

e. *Gain* total variabel atribut Rekomendasi Guru BK adalah:

$$\begin{aligned} \text{Gain (total)} &= 0.53284 - \left( \left( \frac{198}{231} * 0.11327 \right) - \right. \\ &\left( \left( \frac{33}{231} * 0.79905 \right) \right) = 0.32159 \end{aligned}$$

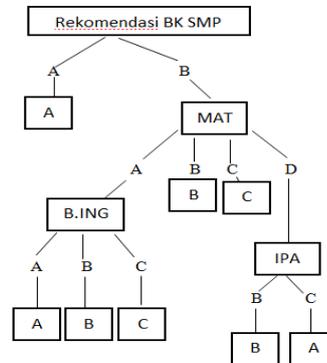
Setelah nilai atribut dicari, dan mendapatkan nilai atribut rekomendasi guru BK SMP menjadi nilai yang paling tinggi diantara nilai atribut lainnya dibawah ini adalah tabel hasil dari perhitungan entropy dan gain.

Tabel 2. Perhitungan entropy gain

	Jumlah (S)	IPA	IP S	Entropy	Gain
Total	231	203	28	0.532835	
<b>BINDO</b>					
A	142	121	21	0.604528	0.00862
B	88	81	7	0.400576	
C	1	1	0	0	
D	0	0	0	0	
<b>MTK</b>					
A	77	69	8	0.481226	0.070598
B	115	109	6	0.295561	
C	34	23	11	0.908178	
D	5	2	3	0.970951	
<b>BING</b>					
A	78	64	14	0.678954	0.012392
B	98	89	9	0.442571	
C	53	48	5	0.450791	
D	2	2	0	0	
<b>IPA</b>					
A	25	22	3	0.529361	0.003976
B	159	142	17	0.49055	
C	47	39	8	0.658191	
D	0	0	0	0	
<b>Rekomen dasi BK</b>					

A	198	195	3	0.113274	0.321593
B	33	8	25	0.799049	

Berdasarkan hasil diatas, dapat digambarkan pohon keputusan sebagai berikut:

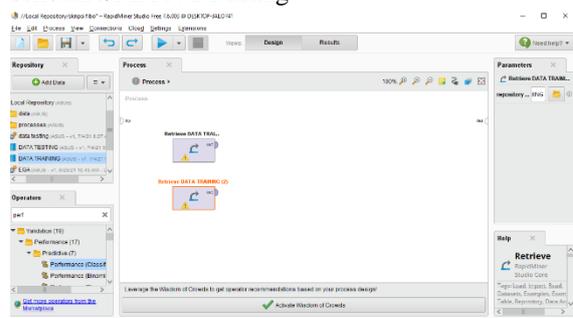


Gambar 3. Proses implementasi data *training* dan data *testing*

Gambar diatas adalah proses implementasi data *training* dan data testing untuk klasifikasi melalui aplikasi RapidMiner.

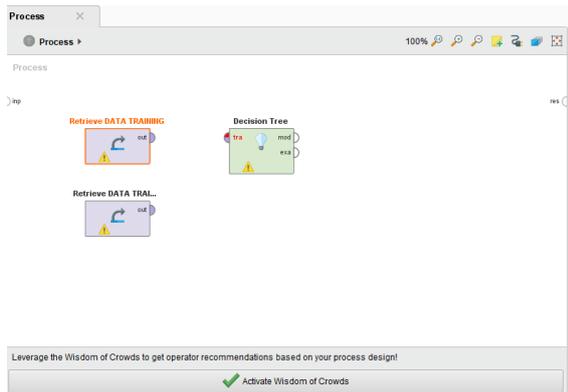
#### 4.2.4. Pengujian dengan Tools RapidMiner

Pengujian dengan menggunakan aplikasi RapidMiner ini bertujuan untuk menguji keakuratan lasifikasi penjurusan siswa. Data yang dianalisis yaitu data *training* dari data siswa TA. 2019/2020 dan untuk data *testing* digunakan data siswa tahun ajaran 2020/2021, data diambil di sekolah SMA N 2 Padang.

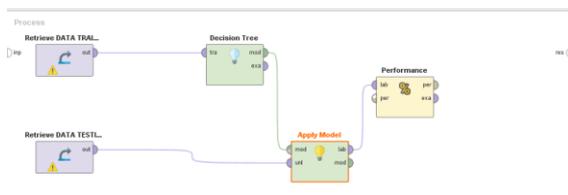


Gambar 4. Memasukan data *training* dan data *testing*

Gambar dibawah ini proses memasukan *decision tree* ke data *training* dan data melalui aplikasi RapidMiner.

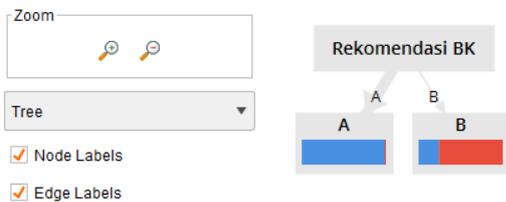


Gambar 5. Menggunakan *decision tree* terhadap *data training* dan *data testing*



Gambar 6. Menghubungkan semua atribut dengan menggunakan *apply model*.

Gambar diatas merupakan proses menghubungkan semua atribut dengan menggunakan *apply model* dan setelah itu kita *run*



Gambar 7. Hasil pohon keputusan

Gambar di atas menunjukkan pohon keputusan dengan atribut *Rekomendasi BK* menjadi atribut yang sangat berpengaruh terhadap penjurusan siswa.

## PerformanceVector

PerformanceVector:

accuracy: 68.42%

ConfusionMatrix:

True:	B	A
B:	50	66
A:	30	158

Gambar 8. *Performance vector*

Gambar diatas adalah deskripsi yang dihasilkan melalui pemodelan *decision tree* atau pohon keputusan.

Tabel 2. Hasil pengujian dengan menggunakan RapidMiner

Accuracy	Precision	Recall
68,42 %	43,10 %	62,50 %

## 4.2.5. Pembahasan Hasil Pengolahan Data Menggunakan Aplikasi RapidMiner.

### 4.2.5.1. Proses *Accuracy*

*Accuracy* merupakan tingkat kedekatan klasifikasi dengan hasil fakta, perhitungan algoritma C4.5 dilakukan melalui jumlah TP+TN dibagi jumlah total *data testing* yang duji untuk melihat akurasi. Berikut adalah hasil dari pengukuran *data accuracy*:

accuracy: 68.42%

	true B	true A	class precision
pred. B	50	66	43.10%
pred. A	30	158	84.04%
class recall	62.50%	70.54%	

Gambar 9. Hasil proses *accuracy*

### 4.2.5.2. Proses *Precision*

*Precision* merupakan tingkat akurasi antara informasi yang diminta dan jawaban yang diberikan oleh sistem. Ini dihitung dengan membagi jumlah data positif benar dengan data salah dan kemudian dibagi dengan jumlah data positif salah. Berikut adalah hasil pengukuran *data precision*.

accuracy: 68.42%

	true B	true A	class precision
pred. B	50	66	43.10%
pred. A	30	158	84.04%
class recall	62.50%	70.54%	

Gambar 10. Hasil proses *Precision*

### 4.2.5.3. Proses *Recall*

*Recall* merupakan tingkat keberhasilan dengan menemukan sebuah informasi. Menghitung proses recall dengan membagi data *True Positive* dengan menjumlahkan hasil dari data *True Positive* dan data salah bernilai negatif (*False Negative*). Berikut adalah hasil dari proses *recall*

accuracy: 68.42%

	true B	true A	class precision
pred. B	50	66	43.10%
pred. A	30	158	84.04%
class recall	62.50%	70.54%	

Gambar 11. Hasil proses *recall*

## 5. Kesimpulan

Berdasarkan pengujian data siswa SMA N 2 Padang tahun masuk ajaran 2020/2021 diperoleh kesimpulan sebagai berikut:

1. Penerapan metode data mining dengan menggunakan algoritma C4.5 dapat mempercepat pengambilan keputusan klasifikasi

profesi siswa saat proses penerimaan mahasiswa baru.

2. Tingkat akurasi dalam proses klasifikasi menggunakan algoritma C4.5 sangat ditentukan dengan adanya data *training* dan atribut yang dipilih dan juga berpengaruh terhadap factor internal dan eksternal.

Dari pengujian penelitian yang dilakukan algoritma C4.5 menghasilkan tingkat akurasi sebesar 68,42% dan atribut rekomendasi guru BK SMP menjadi atribut yang paling berpengaruh dalam menentukan penjurusan siswa.

## 6. Daftar Rujukan

- [1] E. B. Sambani and F. Nuraeni, "Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 9, no. 3, p. 144, 2018.
- [2] I. M. P. dan Subiyanto, "Sistem Rekomendasi Penjurusan Sekolah Menengah Kejuruan Dengan Algoritma C4.5," *J. Kependidikan*, vol. 1, no. 1, pp. 139–149, 2017.
- [3] A. Z. Mafakhir and A. Solichin, "Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta," *Fountain Informatics J.*, vol. 5, no. 1, p. 21, 2020
- [4] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online].
- [5] Y. S. Nugroho, "Klasifikasi dan Klustering Penjurusan Siswa SMA Negeri 3 Boyolali," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, p. 1, 2015
- [6] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Kelayakan Kredit Pada Bank," *J. Tek. Komput. AMIK BSI*, vol. II, no. 1, pp. 132–137, 2016.
- [7] I. P. Astuti, "Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5," *Fountain Informatics J.*, vol. 2, no. 2, p. 5, 2017, doi: 10.21111/fij.v2i2.1067.
- [8] 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] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019
- [10] B. Novianti, T. Rismawan, and S. Bahri, "Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)," *J. Coding, Sist. Komput. Untan*, vol. 04, no. 3, pp. 75–84, 2016.