

## IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK PENJURUSAN SISWA (STUDI KASUS: SMA NEGERI 1 PONTIANAK)

<sup>[1]</sup>Beti Novianti, <sup>[2]</sup>Tedy Rismawan<sup>[3]</sup>Syamsul Bahri

<sup>[1][2][3]</sup>Jurusan Sistem Komputer, Fakultas MIPA Universitas Tanjungpura

Jalan Prof. Dr. H. Hadari Nawawi, Pontianak

Telp./Fax.: (0561) 577963

e-mail : <sup>[1]</sup>betinovianti@student.untan.ac.id <sup>[2]</sup>tedyrismawan@siskom.untan.ac.id,

<sup>[3]</sup>syamsul.bahri@siskom.untan.ac.id

### ABSTRAK

*Klasifikasi penjurusan siswa adalah proses pengelompokkan siswa sesuai dengan kemampuan (nilai), bakat dan minat yang relatif sama sehingga pelajaran yang akan diberikan kepada siswa akan lebih fokus dan terarah. Proses klasifikasi data siswa dapat digali polanya dalam bidang ilmu data mining, yaitu proses mendapatkan hubungan atau pola dari suatu data yang besar sehingga memberikan indikasi yang bermanfaat. SMA Negeri 1 Pontianak merupakan salah satu lembaga pendidikan yang mulai memperkenalkan penjurusan dan membaginya dalam dua pilihan jurusan, yaitu "IPA" dan "IPS". Kurikulum yang digunakan SMA Negeri 1 Pontianak saat ini ialah Kurikulum 2013, yang mengatur proses penyortiran penjurusan siswa kelas X (sepuluh) berdasarkan nilai rata-rata rapor SMP, nilai Ujian Nasional SMP, dan nilai tes MTK, IPA, dan IPS. Salah satu metode yang dapat digunakan untuk menyelesaikan masalah klasifikasi data mining ialah metode Algoritma C4.5. Algoritma C4.5 digunakan untuk membentuk pohon keputusan yang membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan untuk mengklasifikasikan data. Pada penelitian ini, penjurusan siswa diklasifikasi berdasarkan nilai tes akademik MTK, IPA, dan IPS, nilai rata-rata rapor SMP untuk mata pelajaran MTK, IPA, dan IPS, nilai Ujian Nasional SMP untuk mata pelajaran MTK dan IPA, dan minat siswa. Berdasarkan hasil penelitian, didapatkan hasil klasifikasi penjurusan siswa yang sudah diuji sesuai dengan tingkat akurasi sebesar 89.74%.*

**Kata Kunci:** Data Mining, Algoritma C4.5, Pohon Keputusan, Klasifikasi, Penjurusan Siswa

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Penjurusan adalah proses penempatan atau penyaluran dalam pemilihan program pengajaran kepada siswa. Sekolah Menengah Atas (SMA) merupakan salah satu lembaga pendidikan yang mulai memperkenalkan jurusan dan membaginya dalam beberapa pilihan jurusan. Penjurusan sangat penting untuk mengelompokkan siswa sesuai dengan kemampuan (nilai), bakat dan minat yang relatif sama agar selanjutnya, pelajaran yang diberikan kepada siswa lebih fokus dan terarah. Dalam sistem penjurusan ini, siswa diberi kesempatan memilih jurusan, baik itu jurusan "IPA" atau "IPS", sebelum nantinya di klasifikasi keputusan jurusannya menurut nilai dan minat masing-masing siswa, sehingga nanti pilihan siswa dan hasil akhir keputusan jurusan siswa dapat berbeda karena

menyesuaikan kemampuan (nilai) siswa tersebut.

SMA Negeri 1 Pontianak merupakan salah satu sekolah unggulan yang menerapkan proses penjurusan bagi siswanya, untuk dibagi menjadi 2 jurusan, yaitu "IPA" dan "IPS". Kurikulum yang digunakan SMA Negeri 1 Pontianak saat ini ialah Kurikulum 2013, yang mengatur proses penjurusan dengan menggunakan nilai rapor dan nilai UN SMP, serta nilai test penjurusan. Penjurusan mulai dilakukan pada saat murid duduk di bangku kelas X (sepuluh). Tes penjurusan diikuti oleh seluruh siswa kelas X yang kemudian akan disortir bersama dengan nilai rapor dan nilai UN SMP masing-masing siswa.

Dalam proses penjurusan, siswa diberi kesempatan memilih jurusan, baik itu jurusan IPA atau IPS, sebelum nantinya di klasifikasi keputusan jurusannya menurut nilai dan minat

masing-masing siswa, sehingga nanti pilihan siswa dan hasil akhir keputusan jurusan siswa dapat berbeda karena menyesuaikan kemampuan (nilai) siswa tersebut.

Penelitian sebelumnya yang berkaitan dengan *data mining* penjurusan siswa telah dilakukan oleh Obbie Kristianto, yaitu penjurusan siswa SMAN 6 Semarang dengan penerapan algoritma ID3. Aplikasi dibuat dengan bahasa pemrograman java dan berhasil menentukan penjurusan siswa sesuai dengan kebutuhan [1]. Penelitian lainnya dilakukan juga oleh Eka Budi Rahayu, dengan judul penelitian “Algoritma C4.5 Untuk Penjurusan Siswa SMA Negeri 3 Pati” dimana pada penelitian tersebut menggunakan perangkat lunak *RapidMiner* sebagai alat untuk permodelan menghasilkan *rule*/aturan yang akan digunakan untuk klasifikasi penjurusan. Aplikasi klasifikasi penjurusan siswa yang dibuat dapat mengklasifikasikan siswa untuk penjurusan ke kelas IPA dan IPS [2].

Berdasarkan pemaparan singkat tersebut, maka penelitian selanjutnya akan dibuat suatu aplikasi yang dirancang untuk dapat mengklasifikasikan penjurusan siswa dengan studi kasus SMA Negeri 1 Pontianak. Aplikasi penjurusan akan dibuat dengan algoritma C4.5 yang akan di proses secara otomatis di dalam aplikasi tanpa menggunakan perangkat lunak *machine learning* tertentu untuk membuat permodelan dan *rule*/aturan klasifikasi.

### 1.2 Tujuan Penelitian

Tujuan penelitian ini adalah merancang dan mengaplikasikan algoritma C4.5 dalam mengolah data siswa SMA untuk diklasifikasikan menjadi penjurusan siswa dan dapat mengetahui tingkat akurasi hasil klasifikasi penjurusan siswa.

### 1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah aplikasi dapat digunakan untuk memudahkan proses klasifikasi penjurusan siswa bagi lembaga pendidikan yang membutuhkan khususnya sekolah menengah atas.

## 2. LANDASAN TEORI

### 2.1. Data Mining

*Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan

mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar [3].

Hal penting yang terkait dengan *data mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat

Salah satu teknik *data mining* ialah klasifikasi. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, formula matematis atau *neural network*.

Proses klasifikasi biasanya dibagi menjadi dua fase : *learning* dan *test*. Pada fase *learning*, sebagian data yang telah diketahui kelas datanya diumpankan untuk membentuk model perkiraan. Kemudian pada fase *test* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

### 2.2. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk pohon keputusan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu.

Algoritma Pohon Keputusan C4.5 atau *Classification version 4.5* adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut, algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3.

Secara umum, proses algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

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 [4].

Secara khusus, algoritma C4.5 *Decision Tree* menggunakan kriteria *split* yang telah dimodifikasi yang dinamakan *Gain Ratio* dalam proses pemilihan *split* atribut. *Split* atribut merupakan proses utama dalam pembentukan pohon keputusan (*Decision Tree*) di C4.5 [5].

Tahapan dari algoritma C4.5 adalah sebagai berikut.

- a. Menghitung nilai *Entropy*,
- b. Menghitung nilai *Gain Ratio* untuk masing-masing atribut,
- c. Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi akar (*root*) dan atribut yang memiliki nilai *Gain Ratio* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*),
- d. Menghitung lagi nilai *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya,
- e. Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*),
- f. Mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa.

Untuk menghitung nilai *Entropy* dapat dihitung dengan persamaan :

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

dimana:

- S = himpunan kasus  
A = fitur  
n = jumlah partisi S  
 $p_i$  = proporsi dari  $S_i$  terhadap S

Sementara itu nilai *information gain* (*Gain*) dapat dihitung menggunakan persamaan :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

dimana:

- S = himpunan kasus  
A = atribut  
n = jumlah partisi atribut A  
 $|S_i|$  = jumlah kasus pada partisi ke-i  
 $|S|$  = jumlah kasus dalam S

Selanjutnya nilai *Split Info* dapat dihitung dengan persamaan :

$$SplitInfo(S,A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

dimana:

- S = himpunan kasus  
A = atribut.  
 $S_i$  = jumlah sampel untuk atribut i

Maka nilai *Gain Ratio* yang menentukan sebuah atribut dapat dijadikan akar maupun cabang suatu pohon keputusan dapat dihitung dengan persamaan:

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (4)$$

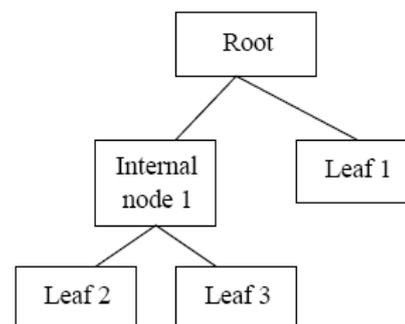
dimana:

- S = himpunan kasus  
A = atribut.  
 $Gain(S,A)$  = *info gain* pada atribut A  
 $SplitInfo(S,A)$  = *split info* pada atribut A

### 2.3. Pohon Keputusan

Pohon Keputusan (*Decision Tree*) merupakan algoritma pengklasifikasian yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan [6].

Pohon yang terbentuk menyerupai pohon terbalik, dimana akar (*root*) berada di bagian paling atas dan daun (*leaf*) berada di bagian paling bawah. Pohon keputusan merupakan model klasifikasi yang berbentuk seperti pohon, dimana pohon keputusan mudah untuk dimengerti meskipun oleh pengguna yang belum ahli sekalipun dan lebih efisien dalam menginduksi data. Pohon keputusan baik digunakan untuk klasifikasi atau prediksi [7]. Tampilan *graph* pohon keputusan dapat dilihat pada Gambar 1.



Gambar 1. Konsep Dasar *Graph* Pohon Keputusan

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* [8].

Konsep pohon keputusan secara garis besar dapat dilihat pada Gambar 2.



Gambar 2. Konsep Pohon Keputusan

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk membuat proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

#### 2.4. Confusion Matrix

Untuk permasalahan dalam klasifikasi, pengukuran yang biasa digunakan adalah *precision*, *recall* dan *accuracy*. Nilai tersebut dapat dihitung dengan *confusion matrix*. *Confusion Matrix* ialah sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar (*positive*) dan tidak benar (*negative*) oleh model klasifikasi [9]. Tabel *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

	<i>Classified Negative</i>	<i>Classified Positive</i>
<i>Actual Negative</i>	a	b
<i>Actual Positive</i>	c	d

Isian “a” berisi jumlah data negatif hasil awal yang sama dengan hasil negatif klasifikasinya. Isian “b” berisi jumlah data negatif hasil awal yang berubah menjadi hasil positif di hasil klasifikasinya. Isian “c” berisi jumlah data positif hasil awal yang berubah menjadi hasil negatif klasifikasinya. Isian “d” berisi jumlah data positif hasil awal yang sama menjadi hasil positif di hasil klasifikasinya. Total jumlah isian “b” dan “c” merupakan selisih jumlah perbedaan hasil yang dihasilkan setelah proses klasifikasi. Dari hasil selisih tersebut bisa dihitung nilai akurasi.

##### 1. Precision

*Precision* adalah bagian data yang di ambil sesuai dengan informasi yang dibutuhkan. Rumus *Precision* adalah :

$$P = \left( \frac{d}{b+d} \right) \times 100\% \quad (5)$$

##### 2. Recall

*Recall* adalah pengambilan data yang berhasil dilakukan terhadap bagian data yang relevan dengan *query*. Rumus *Recall* adalah :

$$R = \left( \frac{d}{c+d} \right) \times 100\% \quad (6)$$

##### 3. Accuracy

*Accuracy* adalah persentase dari total data ujicoba yang benar diidentifikasi. Rumus *Accuracy* adalah :

$$A = \left( \frac{a+d}{total\ sampel} \right) \times 100\% \quad (7)$$

### 3. METODOLOGI PENELITIAN

Penelitian ini menggunakan metodologi penelitian yang mencakup studi pustaka, kemudian dilakukan observasi dan pengumpulan data, selanjutnya analisa kebutuhan, kemudian dilakukan perancangan, dan terakhir dilakukan pengujian sistem.

### 4. PERANCANGAN DAN IMPLEMENTASI

#### 4.1. Data Penelitian

Data penelitian yang digunakan adalah data siswa kelas X SMA Negeri 1 Pontianak Tahun Ajaran 2015/2016. Data didapatkan dari Wakil Kepala Sekolah (WaKa) Kurikulum SMA Negeri 1 Pontianak. Data didapat kemudian dilakukan proses analisis data sehingga didapatkanlah data yang berupa tabel data kasus yang siap untuk dilakukan *data mining*. Selanjutnya data yang bersifat numerik ditransformasikan menjadi kategorikal dengan *range* klasifikasi nilai pada Tabel2.

Tabel 2. Klasifikasi Nilai

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

Data berjumlah 392 data. Data kemudian akan dibagi menjadi 2 kelompok, yaitu data testing dan data training.. Data training berfungsi untuk proses permodelan atau pembelajaran terhadap metode pohon keputusan algoritma C4.5, sedangkan data

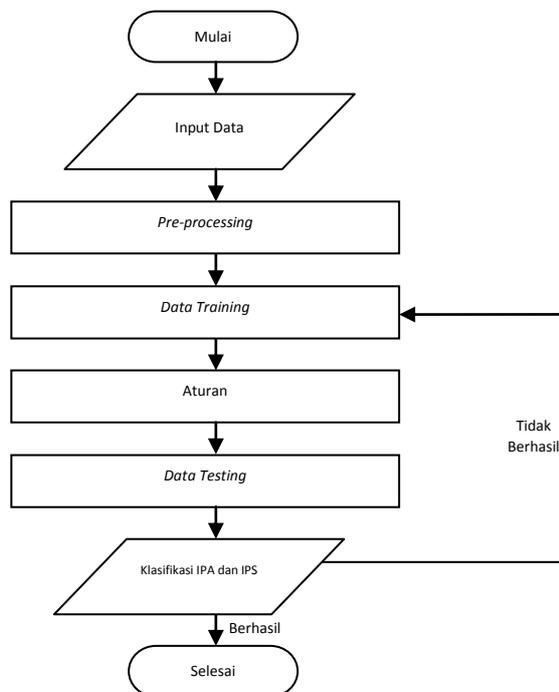
testing digunakan untuk menguji permodelan yang sudah terbentuk.

#### 4.2. Variabel Penelitian

Variabel penelitian yang akan digunakan sebagai atribut data untuk proses *data mining* klasifikasi ialah nilai tes MTK, nilai tes IPA, nilai tes IPS, nilai rata-rata rapor MTK, nilai rata-rata rapor IPA, nilai rata-rata rapor IPS, nilai UN MTK, nilai UN IPA, minat siswa, dan kelas keputusan. Kelas keputusan ialah variabel target penelitian yang berisi 2 nilai kelas, yaitu "IPA" dan "IPS".

#### 4.3. Diagram Alir Perangkat Lunak

Metode yang diusulkan untuk aplikasi penjurusan siswa ialah metode klasifikasi data menggunakan algoritma C4.5 pohon keputusan. Adapun diagram alir dari aplikasi ditunjukkan seperti pada Gambar 3.



Gambar 3. Diagram Alir Perangkat Lunak

Berikut ini penjelasan dari diagram alir pada Gambar 3:

1. Proses dimulai.
2. *Input* data siswa yang akan diproses.
3. *Pre-processing* data dilakukan dengan mentransformasikan data atribut yang bersifat numerik ke kategorikal, serta membagi data yang akan diolah menjadi 2 kelompok yaitu, data training dan data testing.
4. Data training dilakukan sebagai permodelan pembelajaran *data mining*

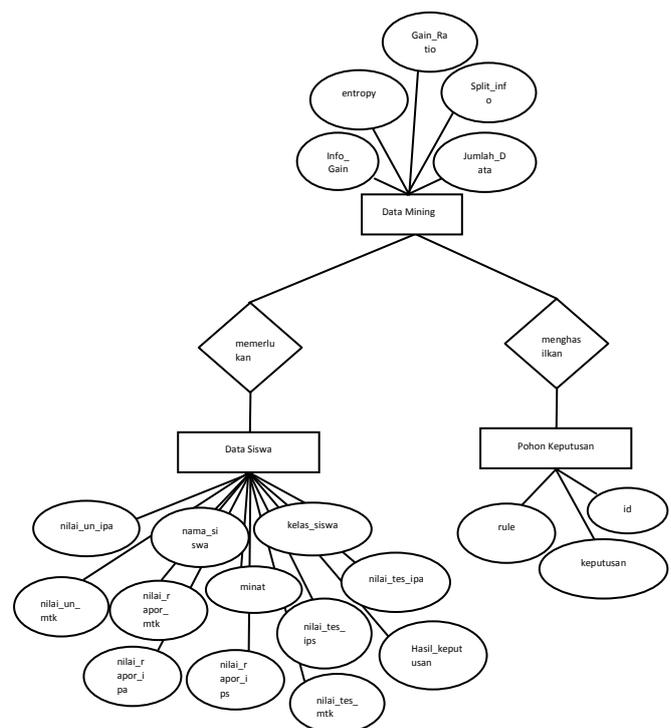
untuk mendapatkan pohon keputusan dan aturan atau *rule* terbentuk.

5. Aturan berhasil terbentuk dari proses permodelan data training yang sempurna.
6. Data testing dilakukan yaitu dengan menguji proses penjurusan siswa dengan aturan yang terbentuk oleh data testing.
7. Klasifikasi IPA dan IPS merupakan hasil yang diharapkan. Jika proses tidak berhasil, maka diulangi kembali proses Data Training. Jika berhasil, maka proses selesai.
8. Proses selesai.

#### 4.4 Entity Relationship Diagram (ERD)

Struktur logika dari basis data dapat digambarkan dalam sebuah grafik dengan menggunakan *Entity Relationship Diagram* (ERD). ERD merupakan hubungan antara entitas yang digunakan dalam sistem untuk menggambarkan hubungan antara entitas atau struktur data dan relasi antar file.

Pada perancangan aplikasi dibutuhkan hubungan antara entitas *Data Mining*, *Data Siswa*, dan *Pohon Keputusan*. Hubungan yang diperlukan ialah proses *Data Mining* akan memerlukan *Data Siswa* yang selanjutnya akan menghasilkan *Pohon Keputusan*. Pada entitas *Pohon Keputusan* terdapat key atribut untuk proses klasifikasi yaitu atribut *rule*. Rancangan ERD aplikasi dapat dilihat pada Gambar 4.



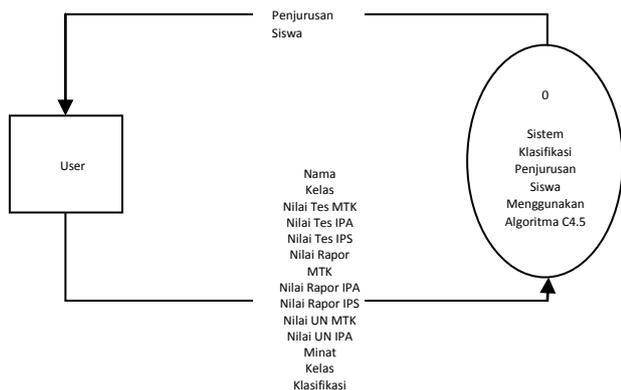
Gambar 4. Entity Relationship Diagram

#### 4.5 Data Flow Diagram

Pengembangan sistem menggunakan *Data Flow Diagram* (DFD) sebagai media untuk menjelaskan semua alur data beserta proses-proses yang terdapat di dalam sistem.

##### a. Diagram Konteks

Diagram Konteks adalah tingkatan tertinggi dalam diagram aliran data dan hanya memuat satu proses, menunjukkan sistem secara keseluruhan. Diagram konteks aplikasi dapat dilihat pada Gambar 5.

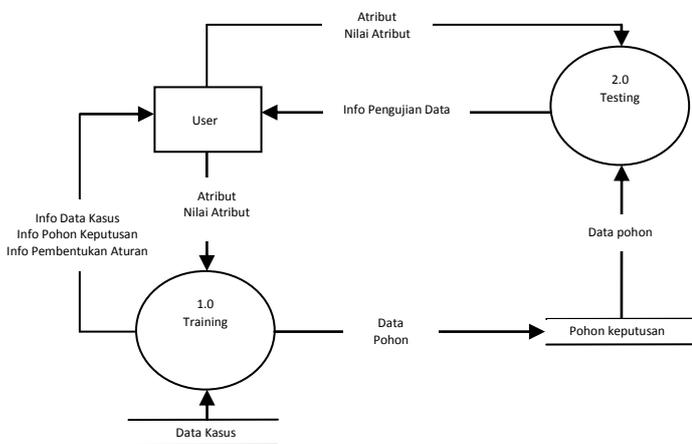


Gambar 5. Diagram Konteks

Aliran data bersumber dari masukan oleh *user* ke dalam sistem yaitu, atribut dan nilai atribut. Atribut yang diperlukan berupa nama, kelas, nilai tes MTK, nilai tes IPA, nilai tes IPS, nilai rata-rata rapor MTK, nilai rata-rata rapor IPA, nilai rata-rata rapor IPS, nilai UN MTK, nilai UN IPA, minat siswa, dan hasil keputusan. Kemudian masuk ke proses sistem dan menghasilkan keluaran penjurusan siswa.

##### b. DFD Level 1

Proses diagram konteks atau DFD level 0 akan dipecah lagi ke dalam DFD level 1, terlihat pada Gambar 6.

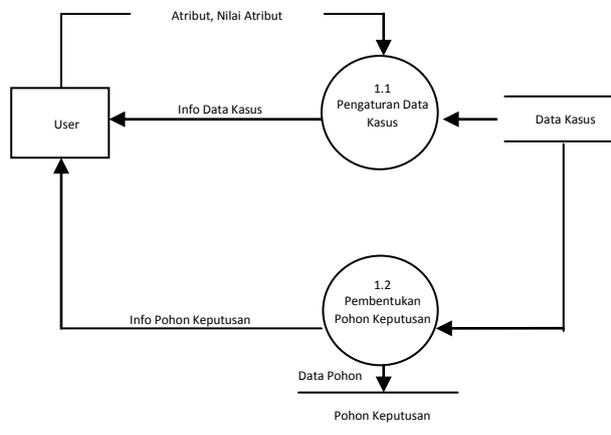


Gambar 6. DFD Level 1

Aliran data bersumber dari masukan atribut dan nilai atribut yang diperlukan ke dalam proses 1.0 untuk kemudian dilakukan training data dan menghasilkan keluaran berupa info data kasus, info pohon keputusan, info pembentukan aturan. Masukan atribut dan nilai atribut yang sama juga diperlukan untuk proses 2.0, untuk kemudian dilakukan proses testing data yang menggunakan aturan yang sudah terbentuk dari pohon keputusan. Hasil keluarannya ialah info pengujian data.

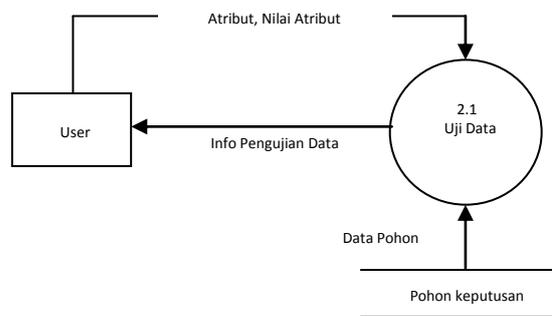
##### c. DFD Level 2

Proses DFD level 1 dipecah lagi dan dibagi dua untuk proses Training dan Testing. Untuk DFD Level 2 Proses Training dapat dilihat pada Gambar 7.



Gambar 7. DFD Level 2 Proses Training

Proses 1.1 merupakan proses pengaturan data kasus yang akan di proses *data mining* untuk dicari pola klasifikasinya sesuai algoritma C4.5. Proses 1.2 menunjukkan proses pembentukan pohon keputusan yang nantinya akan menjadi hasil keluaran akhir proses berupa pohon keputusan dan aturan/*rule*. Selanjutnya untuk DFD Level 2 Proses Testing dapat dilihat pada Gambar 8.

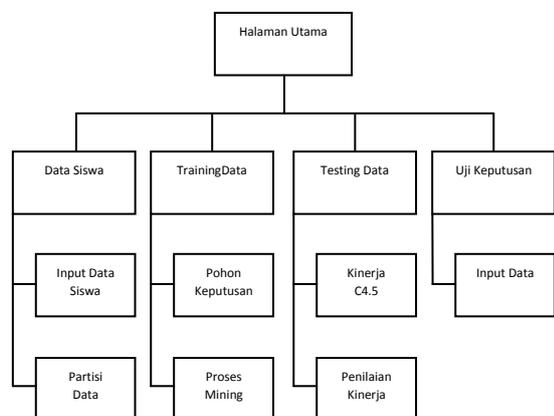


Gambar 8. DFD Level 2 Proses Testing

Selanjutnya pada Gambar 8, dijelaskan aliran proses 2.1 yaitu pengujian data testing. Masukkannya ialah atribut dan nilai atribut data dan aturan pohon keputusan yang akan menghasilkan info pengujian data berupa hasil proses klasifikasi penjurusan.

#### 4.4 Perancangan Antarmuka

Perancangan antarmuka bertujuan untuk memudahkan dalam merancang tampilan aplikasi yang akan dibuat untuk menentukan tahap pembuatan aplikasi seperti misalnya tata letak menu, tombol, tampilan warna, dan form-form yang dibutuhkan tampak lebih baik dan terstruktur sesuai dengan kebutuhan pengguna dalam menggunakan sistem aplikasi. Struktur menu yang dirancang untuk aplikasi penjurusan dapat dilihat pada Gambar 9.



Gambar 9. Struktur Menu

### 5. PENGUJIAN DAN PEMBAHASAN

#### 5.1 Pengujian Aplikasi

Pengujian aplikasi bertujuan untuk mengetahui berfungsinya aplikasi yang sudah dibuat sesuai dengan yang diharapkan. Pengujian aplikasi dilakukan pada Web Browser. Pengujian dilakukan untuk setiap menu utama pada aplikasi. Tampilan utama aplikasi dapat dilihat pada Gambar 10.



Gambar 10. Tampilan Utama Aplikasi

#### a. Menu Data Siswa

Menu data siswa adalah menu untuk memasukkan dan mengatur data siswa yang siap untuk diklasifikasi penjurusan. Data siswa yang digunakan adalah data siswa SMA Negeri 1 Pontianak kelas X Tahun 2015/2016. Setelah data dimasukkan, maka data siswa dapat ditampilkan, seperti pada Gambar 11.

No	Nama Siswa	Kelas Siswa	Nilai Tes MTK	Nilai Tes IPA	Nilai Tes IPS	Nilai Rapor MTK	Nilai Rapor IPA	Nilai Rapor IPS	Nilai UN MTK	Nilai UN IPA	Minat	Class	Status Data	Edit/Hapus
1	AFIFAH MARWAHAL-QADRIE	X-A	60	40	55	84	85.33	92.5	82.5	72.5	IPA	IPA	Data Training	🗑️   ✎️
2	ANGGUN LULUK PUJI RAHAYU	X-A	30	40	60	83.83	86.67	86.83	70	82.5	IPA	IPS	Data Training	🗑️   ✎️
3	ANNISA RIZKI RAHMANIATI	X-A	55	50	50	84.83	87.67	93.16	82.5	90	IPA	IPA	Data Training	🗑️   ✎️
4	ARIS RICKY SAPUTRO	X-A	20	30	65	78.67	83.5	78.83	85	62.5	IPA	IPS	Data Training	🗑️   ✎️
5	ARRIFQA RAUDATUL ADAWIYAH	X-A	85	60	45	89	90.5	88.67	77.5	82.5	IPA	IPA	Data Training	🗑️   ✎️
6	ASYIAH MAYANSARI	X-A	75	50	30	88.16	76.33	85.16	87.5	85	IPA	IPA	Data Training	🗑️   ✎️

Gambar 11. Tampilan Data Siswa

Selanjutnya semua data yang sudah dimasukkan dibagi menjadi 2 kelompok data, yaitu data training dan data testing. Pembagian data dilakukan dengan perbandingan 90:10. Pembagian data dilakukan di sub menu Partisi Data seperti pada Gambar 12.

Status Data	Data Training (392 Data)	Data Testing (39 Data)
IPS (82 Data)	74	8
IPA (310 Data)	279	31

Gambar 12. Tampilan Partisi Data

#### b. Menu Mining C4.5

Menu mining C4.5 adalah salah satu proses utama dari aplikasi penjurusan ini. Pada menu ini dilakukan proses *mining* seluruh data training sesuai dengan metode pohon keputusan algoritma C4.5.

Proses perhitungan dilakukan secara otomatis di aplikasi. Perhitungan mencakup nilai *entropy* setiap atribut, *information gain*, *split info*, dan *gain ratio*.

Tampilan tabel hasil perhitungan pada proses *mining* untuk menentukan akar/root pohon keputusan dapat dilihat pada Gambar 13.

NoAtribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus IPA	Jumlah Kasus IPS	Entropy	Information Gain	Split Info	Gain Ratio
1	Total	353	279	74	0.7408			0
2	nilai_tes_mtk A	30	29	1	0.2108	0.4216	2.1096	0.1998
3	nilai_tes_mtk B	125	123	2	0.1184	0.4216	2.1096	0.1998
4	nilai_tes_mtk C	101	99	2	0.1403	0.4216	2.1096	0.1998
5	nilai_tes_mtk D	66	25	41	0.9572	0.4216	2.1096	0.1998
6	nilai_tes_mtk E	31	3	28	0.4587	0.4216	2.1096	0.1998
7	nilai_tes_ipa A	0	0	0	0	0.0683	1.6169	0.0422
8	nilai_tes_ipa B	4	3	1	0.8113	0.0683	1.6169	0.0422
9	nilai_tes_ipa C	96	89	7	0.3767	0.0683	1.6169	0.0422
10	nilai_tes_ipa D	156	129	27	0.6647	0.0683	1.6169	0.0422
11	nilai_tes_ipa E	97	58	39	0.9721	0.0683	1.6169	0.0422
12	nilai_tes_ips A	0	0	0	0	0.0012	1.5949	0.0008
13	nilai_tes_ips B	7	6	1	0.5917	0.0012	1.5949	0.0008
14	nilai_tes_ips C	59	45	14	0.7905	0.0012	1.5949	0.0008
15	nilai_tes_ips D	131	103	28	0.7486	0.0012	1.5949	0.0008
16	nilai_tes_ips E	156	125	31	0.7194	0.0012	1.5949	0.0008
17	nilai_rapor_mtkA	203	186	17	0.4152	0.1008	1.0328	0.0976
18	nilai_rapor_mtkB	148	92	56	0.9569	0.1008	1.0328	0.0976
19	nilai_rapor_mtkC	1	1	0	0	0.1008	1.0328	0.0976
20	nilai_rapor_mtkD	1	0	1	0	0.1008	1.0328	0.0976
21	nilai_rapor_mtkE	0	0	0	0	0.1008	1.0328	0.0976
22	nilai_rapor_ipa A	201	177	24	0.5276	0.052	1.0541	0.0493
23	nilai_rapor_ipa B	149	101	48	0.9067	0.052	1.0541	0.0493
24	nilai_rapor_ipa C	2	1	1	1	0.052	1.0541	0.0493
25	nilai_rapor_ipa D	0	0	0	0	0.052	1.0541	0.0493
26	nilai_rapor_ipa E	1	0	1	0	0.052	1.0541	0.0493
27	nilai_rapor_ips A	184	156	28	0.6153	0.0211	1.0237	0.0206
28	nilai_rapor_ips B	168	123	45	0.8384	0.0211	1.0237	0.0206
29	nilai_rapor_ips C	0	0	0	0	0.0211	1.0237	0.0206
30	nilai_rapor_ips D	0	0	0	0	0.0211	1.0237	0.0206
31	nilai_rapor_ips E	1	0	1	0	0.0211	1.0237	0.0206
32	nilai_un_mtk A	140	126	14	0.469	0.0777	1.4924	0.0521
33	nilai_un_mtk B	176	136	40	0.7732	0.0777	1.4924	0.0521
34	nilai_un_mtk C	26	9	17	0.9306	0.0777	1.4924	0.0521
35	nilai_un_mtk D	4	2	2	1	0.0777	1.4924	0.0521
36	nilai_un_mtk E	7	6	1	0.5917	0.0777	1.4924	0.0521
37	nilai_un_ipa A	89	83	6	0.3562	0.0586	1.4129	0.0415
38	nilai_un_ipa B	207	162	45	0.7554	0.0586	1.4129	0.0415
39	nilai_un_ipa C	55	34	21	0.9593	0.0586	1.4129	0.0415
40	nilai_un_ipa D	2	0	2	0	0.0586	1.4129	0.0415
41	nilai_un_ipa E	0	0	0	0	0.0586	1.4129	0.0415
42	minat IPA	339	279	60	0.6735	0.094	0.2407	0.3905
43	minat IPS	14	0	14	0	0.094	0.2407	0.3905

Atribut minat memiliki nilai gain terbesar

```

minat = IPA (IPS = 60, IPA = 279) : ?
minat = IPS (IPS = 14, IPA = 0) : IPS
    
```

Gambar 13. Tampilan Perhitungan Node Akar

Dari proses perhitungan penentuan akar pohon keputusan, kemudian dilakukan

kembali proses perhitungan secara terus-menerus untuk menentukan cabang-cabang pohon keputusan. Proses *data mining* akan selesai jika semua atribut sudah masuk ke kelas atau variabel targetnya masing-masing. Proses akan berlangsung cukup lama sesuai dengan banyaknya data yang diolah. Pemberitahuan “Proses Mining Selesai” akan muncul pada akhir proses *mining*, menandakan proses *mining* berjalan sempurna.

Setelah proses *data mining* selesai maka akan terbentuk pohon keputusan dan serangkaian aturan/*rule* yang akan digunakan untuk proses data testing. Aturan yang dihasilkan berjumlah 162 aturan.

Aturan yang ditampilkan berupa aturan “IF” dan “THEN”. Pada aturan juga terdapat id *rule* yaitu nomor urutan pembentukan cabang pohon keputusan. Aturan yang didapat berjumlah cukup banyak dikarenakan data yang diolah cukup banyak dan untuk 1 data terdapat atribut dan nilai atribut yang banyak pula.

Berikut ditampilkan 3 aturan atau *rule* dari total 162 aturan klasifikasi penjurusan yang dapat dilihat pada Gambar 13.

```

Rule:
1. if (minat == IPA AND nilai_tes_mtk == A) then IPA (id = 2)
2. if (minat == IPA AND nilai_tes_mtk == B) then IPA (id = 3)
3. if (minat == IPA AND nilai_tes_mtk == C) then IPA (id = 4)
    
```

Gambar 13. Tampilan 3 Aturan

### c. Menu Kinerja

Menu kinerja adalah menu untuk menguji aturan atau *rule* yang sudah terbentuk dari permodelan data training menggunakan metode pohon keputusan algoritma C4.5. Data testing yang digunakan sebanyak 39 data dari keseluruhan data siswa yang digunakan untuk proses pengujian aplikasi. Pengujian data testing dapat berjalan dengan baik ditandai dengan semua data dapat diklasifikasikan ke kelas “IPA” maupun “IPS”.

Dari 39 data siswa didapatkan siswa yang diklasifikasikan ke kelas IPA berjumlah 29 siswa dan siswa yang diklasifikasikan ke kelas IPS berjumlah 10 siswa. Pada masing-masing data siswa juga dicantumkan nomor id *rule*. *Rule* inilah yang berperan untuk mengklasifikasikan data ke kelasnya sesuai dengan algoritma C4.5 pada proses data testing. Hasil pengujian data testing ditampilkan pada Gambar 14.

No	Nama Siswa	Kelas Siswa	Nilai Tes MTK	Nilai Tes IPA	Nilai Tes IPS	Nilai Raport MTK	Nilai Raport IPA	Nilai Raport IPS	Nilai UN MTK	Nilai UN IPA	Minat	Keputusan Asli	Keputusan C4.5 ID Rule
1	ARINTO MANGKRU SAPUTRO	X-I	B	C	B	A	A	A	A	A	IPA	IPA	IPA 3
2	ASTARI RAFITA GALANTTIARA	X-I	B	E	C	A	A	A	A	B	IPA	IPA	IPA 3
3	AULIIPA NURULIANNAH	X-I	B	D	D	A	B	B	A	C	IPA	IPA	IPA 3
4	AURA ANUN INTANY	X-I	B	C	E	A	B	B	A	B	IPA	IPA	IPA 3
5	CLEMENTINEKE CERELIA AURELIA	X-I	B	E	D	B	B	B	A	B	IPA	IPA	IPA 3
6	CUT VANIA ISMARIZA	X-I	C	C	E	B	B	B	A	B	IPA	IPA	IPA 4
7	DEVINA WANDA ANGELITA	X-I	E	E	E	B	B	A	D	C	IPA	IPA	IPA 200
8	ERIKA PRATIWI	X-I	A	D	D	A	A	A	A	B	IPA	IPA	IPA 2
9	FADILA	X-I	C	E	E	B	A	B	B	B	IPA	IPA	IPA 4
10	FAKHRI RAMADHAN ISHAN	X-I	B	D	D	A	A	A	A	B	IPA	IPA	IPA 3
11	FARDHI LIRA RAMADHAN	X-I	A	C	D	A	A	A	A	B	IPA	IPA	IPA 2
12	HESTI AYUNINGTYAS	X-I	B	D	D	B	B	B	B	C	IPA	IPA	IPA 3
13	HIKMAHWAN MUHAMMAD AKMAL SESAR	X-I	D	C	E	B	B	B	B	C	IPA	IPA	IPS 84
14	HIMA MAHDA NADIA	X-I	D	E	E	B	B	B	B	B	IPA	IPS	IPS 118
15	INGGIT SEPTANIA AURELIA BAKRI	X-I	B	D	E	B	B	A	A	A	IPA	IPA	IPA 3
16	JUMITA PRATIWI	X-I	A	C	D	A	A	A	A	A	IPA	IPA	IPA 2
17	MUHAMMAD MARUF	X-I	D	D	D	B	B	B	B	A	IPA	IPS	IPA 32
18	MUHAMMAD AZHAR ZAENI	X-I	C	D	E	A	B	A	B	B	IPA	IPA	IPA 4
19	MUHAMMAD FARHAN	X-I	C	D	E	A	A	A	A	B	IPA	IPA	IPA 4
20	MUHAMMAD LIO RADIFAN	X-I	A	E	E	A	B	B	A	B	IPA	IPA	IPA 2
21	MUHAMMAD RIFQI HARYADI	X-I	D	E	E	B	B	B	D	B	IPA	IPS	IPS 144
22	NURAINI	X-I	A	C	E	A	A	B	B	A	IPA	IPA	IPA 2
23	NURTANIA AFIFAH	X-I	D	E	E	A	A	A	B	B	IPA	IPS	IPS 56
24	PUTRI KUSUMA WIARDANI	X-I	A	C	C	A	A	B	A	B	IPA	IPA	IPA 2
25	QUEENTERA CANTIKA ARASANDA	X-I	D	D	E	B	A	A	C	B	IPA	IPA	IPS 133
26	RADINDA FARIDZIQA FITRIYANTO	X-I	A	C	E	A	A	A	C	B	IPA	IPA	IPA 2
27	REDHA DHIKA MAHFUSZAH	X-I	C	C	D	A	A	A	B	A	IPA	IPA	IPA 4
28	RETNY AYU CHOIR RASYIDAH	X-I	B	C	D	A	A	A	B	A	IPA	IPA	IPA 3
29	RIDHO BAGASKORO	X-I	E	E	E	B	B	B	B	B	IPA	IPS	IPS 193
30	RIFKI CHARULLAH	X-I	B	E	D	A	B	B	B	B	IPA	IPA	IPA 3
31	RIZKA RAMADHANTI	X-I	C	C	C	B	B	B	B	B	IPA	IPA	IPA 4
32	SADDAM HUSEIN SURAJ MUHAMMAD	X-I	E	E	E	B	B	A	B	C	IPA	IPS	IPS 193
33	SALSABILA AWANIS ZHARFA	X-I	C	D	B	A	A	A	A	B	IPA	IPA	IPA 4
34	SANDRA SRI PUSPA	X-I	E	E	E	B	B	B	B	B	IPA	IPS	IPS 193
35	SARAH SALSABILA	X-I	D	D	D	B	A	A	C	A	IPA	IPA	IPS 133
36	SEPTIANTO	X-I	C	E	D	A	A	B	B	A	IPA	IPA	IPA 4
37	SYAHWAL DESPRIADI	X-I	A	E	E	A	A	A	A	B	IPA	IPA	IPA 2
38	SYARIF MUHAMMAD FARHAN AL IDRUS	X-I	E	D	E	B	B	B	B	C	IPA	IPS	IPS 193
39	YOGAMA WISNU RIKTYANDITO	X-I	C	D	E	A	A	A	B	A	IPA	IPA	IPA 4

Gambar 14. Tampilan Pengujian Data Testing

#### d. Menu Penentu Keputusan

Menu penentu keputusan digunakan untuk menguji cobakan1 data tunggal akan diklasifikasikan ke salah satu kelas penjurusan IPA atau IPS.Data siswa dimasukkan sesuai dengan variabel atribut yang diminta ke dalam form *input* data, kemudian data tersebut akan mendapat hasil klasifikasi penjurusannya.

Telah dilakukan 3 pengujian data pada menu penentu keputusan. Hasilnya dapat dilihat pada Gambar 15.

No	Nama Siswa	Kelas Siswa	Nilai Tes MTK	Nilai Tes IPA	Nilai Tes IPS	Nilai Raport MTK	Nilai Raport IPA	Nilai Raport IPS	Nilai UN MTK	Nilai UN IPA	Minat	Keputusan C4.5 ID Rule	Hapus
1	Satu	X-A	55 D	60 C	75 B	80 B	80 B	80 B	80 B	90 A	IPA	IPS 84	
2	Lian	X-A	35 E	55 D	60 C	80 B	81.5 B	84.33 B	87.5 A	80 B	IPA	IPS 159	
3	Yaya	X-H	90 A	80 B	70 C	60 C	78 B	88 A	87 A	88 A	IPA	IPA 2	

Gambar 15. Hasil Pengujian Penentu Keputusan

Data yang sudah dimasukkan dapat diklasifikasi dengan baik. Pada pengujian data pertama hasil yang dihasilkan ialah “IPS” dengan id *rule*: 84. Pada pengujian kedua, hasil yang didapatkan ialah “IPS”, dengan id *rule*: 159. Pada pengujian ketiga, hasil yang didapatkan ialah “IPA”, dengan id *rule*: 2.

#### 5.2 Evaluasi Hasil

Dari hasil pengujian aplikasi, semua dataset berhasil diklasifikasikan ke kelas keputusan “IPA” dan “IPS”. Didapatkan jumlah hasil selisih perbedaan untuk keputusan awal sebelum proses mining dan sesudah proses mining C4.5. Selanjutnya akan dilakukan evaluasi hasil.

Evaluasi dilakukan dengan mengembangkan hasil klasifikasi *data mining*. Penilaian hasil kinerja proses diukur dengan *confusion matrix*. Tabel penilaian *confusion matrix* diisi berdasarkan hasil yang didapat dari data testing. Hasil awal atau hasil asli klasifikasi penjurusan siswa yang didapat dari sekolah dibandingkan dengan hasil klasifikasi menggunakan metode algoritma C4.5.

Jumlah seluruh data yang dinilai ialah sebanyak 39 data. Untuk kolom negatif diisi sebagai nilai “IPS”, sedangkan untuk kolom positif diisi sebagai nilai “IPA”. Hasil keputusan awal sebelum dilakukan proses C4.5 diperoleh keputusan kelas IPS sebanyak 8 siswa sedangkan keputusan kelas IPA sebanyak 31 siswa. Untuk hasil setelah proses C4.5 ialah diperoleh keputusan kelas IPS

sebanyak 10 siswa dan keputusan kelas IPA sebanyak 29 siswa.

Hasil yang tetap sama dari hasil keputusan awal "IPS" menjadi hasil klasifikasi C4.5 "IPS" berjumlah 7 data, sedangkan hasil yang berubah dari hasil keputusan awal "IPS" menjadi hasil klasifikasi C4.5 "IPA" berjumlah 1 data. Hasil yang tetap sama dari hasil keputusan awal "IPA" menjadi hasil klasifikasi C4.5 "IPA" berjumlah 28 data, sedangkan hasil yang berubah dari hasil keputusan awal "IPA" menjadi hasil klasifikasi C4.5 "IPS" berjumlah 3 data. Sehingga kemudian didapatkan jumlah selisih perbedaan hasil dataset sebelum dan sesudah di proses mining dan dapat dihitung tingkat akurasi klasifikasi. Tabel penilaian aplikasi klasifikasi penjurusan dapat dilihat pada Tabel 3.

Tabel 3. Tabel Penilaian

		Hasil Klasifikasi C4.5	
		IPS	IPA
Hasil Awal	IPS	7	1
	IPA	3	28

$$\text{Precision: } P = \frac{28}{1+28} \times 100\% = 96.55\%$$

$$\text{Recall: } R = \frac{28}{3+28} \times 100\% = 90.32\%$$

$$\text{Accuracy: } A = \frac{7+28}{8+31} \times 100\% = 89.74\%$$

Dari tabel penilaian didapatkan selisih hasil klasifikasi berjumlah 4 data. Dari selisih ini dapat dihitung tingkat akurasi sebesar 89.74%.

## 6. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian dapat diambil beberapa kesimpulan sebagai berikut :

1. Aplikasi klasifikasi penjurusan siswa dengan algoritma pohon keputusan C4.5 dapat mengklasifikasi penjurusan siswa .
2. Hasil penilaian tingkat kinerja hasil klasifikasi penjurusan siswa dengan menggunakan tabel penilaian *confusion matrix* menghasilkan nilai *accuracy* sebesar 89.74%, nilai *precision* sebesar 96.55%, dan nilai *recall* sebesar 90.32%.

## 7. SARAN

Saran yang diajukan untuk penelitian selanjutnya, ialah sebagai berikut:

1. Dilakukan penambahan atribut untuk klasifikasi penjurusan SMA, seperti contohnya nilai IQ.
2. Pengembangan aplikasi selanjutnya dapat lebih diperbaiki khususnya dalam efisiensi waktu pemrosesan *data mining* agar tidak memakan waktu yang banyak.
3. Dapat menggunakan data uji yang memiliki *missing value* dan menambahkan fungsi *import* data pada aplikasi selanjutnya agar data tidak perlu dimasukkan satu per satu oleh admin.

## DAFTAR PUSTAKA

- [1] Kristianto, O., (2014). *Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang*.<http://eprints.dinus.ac.id/5398/1/14005.pdf> diakses tanggal 16 Maret 2016.
- [2] Rahayu, E.B., (2015). *Algoritma C4.5 Untuk Penjurusan Siswa SMA Negeri 3 Pati*.[http://eprints.dinus.ac.id/15291/2/abstrak\\_15319.pdf](http://eprints.dinus.ac.id/15291/2/abstrak_15319.pdf) diakses tanggal 16 Maret 2016.
- [3] Turban, E., dkk. (2005) *Decision Support Systems and Intelligent Systems*. Yogyakarta: ANDI
- [4] Kusriani., &Luthfi., M.T. (2009). *Algoritma Data Mining*. Yogyakarta: ANDI
- [5] Quinlan, J.R. (1986). *Induction of Decision Trees*. Machine Learning, 81-106
- [6] Mantas, C. J., & Abellán, J. (2014). Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications*, 41(10), 4625–4637. doi:10.1016/j.eswa.2014.01.017
- [7] Sammut, G.W.C. (2011). *Encyclopedia of Machine Learning*. (C. Sammut & G. I. Webb, Eds.). Boston, MA: Springer US.
- [8] Basuki, A., & Syarif, I. (2003). *Pohon Keputusan*. <http://lecturer.eepis-its.edu/~basuki/lecture/DecisionTree.pdf>, diakses tanggal 16 Maret 2016.
- [9] Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process*. [robotics.stanford.edu/~ronnyk/glossary.pdf](http://robotics.stanford.edu/~ronnyk/glossary.pdf) diakses tanggal 1 Agustus 2016