

**MENENTUKAN PENJURUSAN SISWA DENGAN MENGGUNAKAN  
METODE DECISION TREE ALGORITMA C4.5  
(Studi Kasus: SMA Negeri 2 Padang)**

**PROPOSAL SKRIPSI**

*Disusun Untuk Memenuhi Syarat Menyelesaikan Pendidikan Strata 1 (S1)  
Pada Jurusan Teknik Elektronika  
Program Studi Pendidikan Teknik Informatika*



**Oleh:**

**MUHAMMAD FIBO DONYA IKHBAL  
16076065 / 2016**

**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA  
JURUSAN TEKNIK ELEKTRONIKA  
FAKULTAS TEKNIK  
UNIVERSITAS NEGERI PADANG  
2021**

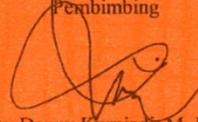
PERSETUJUAN SKRIPSI

MENENTUKAN PENJURUSANI SISWA DENGAN MENGGUNAKAN  
METODE DECISION TREE ALGORITMA C4.5  
(Studi Kasus : SMA N 2 PADANG)

Nama : Muhammad Fibo Donya Ikhbal  
NIM : 16076065  
Program Studi : Pendidikan Teknik Informatika  
Jurusan : Teknik Elektronika  
Fakultas : Teknik

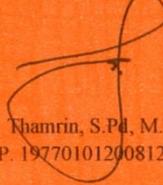
Padang, 29 Desember 2021

Dijetui Oleh,  
Pembimbing



Drs. Denny Kurniadi, M. Kom  
NIP. 196306061989031001

Mengetahui,  
Ketua Jurusan Teknik Elektronika  
Fakultas Teknik  
Universitas Negeri Padang



Rhamrin, S.Pd., M.T  
NIP. 197701012008121001

LEMBAR PENGESAHAN

Dinyatakan lulus setelah mempertahankan skripsi didepan Tim Penguji  
Program Studi Pendidikan Teknik Informatika  
Jurusan Elektronika Fakultas Teknik  
Universitas Negeri Padang

Judul : Menentukan Penjurusan Siswa dengan Menggunakan  
Metode Decision Tree Algoritma C4.5  
(Studi Kasus : SMA N 2 Padang)

Nama : Muhammad Fibo Donya Ikhbal  
NIM : 16076065  
Program Studi : Pendidikan Teknik Informatika  
Jurusan : Teknik Elektronika  
Fakultas : Teknik

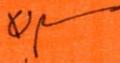
Padang, 29 Desember 2021

Tim Penguji

Nama

Tanda Tangan

1. Ketua : Khairi Budayawan, S.Pd., M. Kom.

1. 

2. Anggota : Drs, Denny Kurniadi, M. Kom.

2. 

3. Anggota : Yeka Hendriyani, S. Kom., M. Kom

3. 

### SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi / tugas akhir yang berjudul "Menentukan Penjurusan Siswa dengan Menggunakan Metode Decision Tree Algoritma C4.5 (Studi Kasus: SMA Negeri 2 Padang)" ini benar-benar karya saya sendiri. Sepanjang pengetahuan saya tidak terdapat pendapat atau karya orang lain yang ditulis dan diterbitkan kecuali sebagai acuan ilmiah atau kutipan dengan tata cara penulisan karya ilmiah yang lazim.

Padang, 29 Desember 2020



Muhammad Kibo Donya Ikhbal

(16076065)

## **ABSTRAK**

### **Muhammad Fibo Donya Ikhbal : Menentukan Penjurusan Dengan Menggunakan Metode Decision Tree Algoritma C4.5**

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 dua program penjurusan yaitu Ilmu Pengetahuan Alam dan Ilmu Pengetahuan Sosial, 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* dengan menggunakan algoritma C4.5 yang dibantu dengan aplikasi Rapidminer. Klasifikasi ini menggunakan 6 atribut sebagai atribut control 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.

## KATA PENGANTAR



Puji dan syukur hanya milik Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas Akhir ini berjudul **“MENENTUKAN PENJURUSAN SISWA DENGAN MENGGUNAKAN METODE *DECISION TREE* ALGORITMA C4.5 (Studi kasus: SMA N 2 Padang)”**. Penulisan laporan skripsi ini berguna untuk memenuhi salah satu syarat memperoleh gelar Sarjana Pendidikan Strata Satu pada Program Studi Pendidikan Teknik Informatika di Fakultas Teknik Universitas Negeri Padang. Dalam menyelesaikan skripsi ini, penulis dibantu dan dibimbing oleh berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Fahmi Rizal, M.Pd, MT selaku Dekan Fakultas Teknik Universitas Negeri Padang.
2. Bapak Thamrin, S.Pd., MT. selaku Ketua Jurusan Teknik Elektronika, Fakultas Teknik Universitas Negeri Padang.
3. Bapak Drs. Zulhendra, M.Kom selaku Dosen Penasehat Akademik
4. Bapak Drs. Denny Kurniadi, M.Kom selaku Dosen Pembimbing yang telah membimbing dalam perencanaan, pelaksanaan dan pelaporan Tugas Akhir ini.
5. Bapak Khairi Budayawan S.Pd., M.Kom. dan Ibu Yeka Hendriyani, S. Kom, M. Kom selaku Penguji yang telah memberi motivasi dalam setiap perjalanan studi penulis.

6. Orang Tua dan saudara-saudaraku yang telah memberikan bantuan, motivasi dan do'a sehingga dalam melaksanakan Tugas Akhir ini lancar, aman dan terkendali.
7. Teman-teman Pendidikan Teknik Informatika 2016 yang telah memberikan motivasi dalam menyelesaikan Tugas Akhir.

Semangat kepada penulis dalam proses penyelesaian laporan tugas akhir ini. Semoga bantuan dan bimbingan yang telah diberikan menjadi amal shaleh bagi Bapak dan Ibu serta mendapat balasan yang berlipat ganda dari Allah SWT. Penulis menyadari bahwa dalam laporan Tugas Akhir ini masih terdapat kekurangan dan kelemahan. Untuk itu penulis mengharapkan saran untuk menyempurnakan laporan Tugas Akhir ini. Wassallam. Padang,

## DAFTAR ISI

<b>KATA PENGANTAR</b> .....	<b>ii</b>
<b>DAFTAR ISI</b> .....	<b>iv</b>
<b>DAFTAR TABEL</b> .....	<b>vi</b>
<b>DAFTAR GAMBAR</b> .....	<b>vii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
A. Latar Belakang.....	1
B. Identifikasi Masalah.....	8
C. Batasan Masalah .....	8
D. Rumusan Masalah.....	9
E. Tujuan Penelitian .....	9
F. Manfaat Penelitian .....	9
<b>BAB II KAJIAN TEORITIS</b> .....	<b>11</b>
A. Penjurusan Siswa.....	11
B. Data Mining .....	14
1. Pengenalan pola, Data Mining dan Machine Learning.....	15
2. Tahap-tahap Data Mining .....	16
3. Metode Data Mining.....	18
4. Penerapan Metode Klasifikasi.....	20
5. Rapid Miner.....	32
C. Penelitian Relevan .....	32
D. Kerangka Berfikir .....	33
<b>BAB III METODE PENELITIAN</b> .....	<b>35</b>
A. Jenis Penelitian .....	35
B. Desain Penelitian .....	36
C. Populasi dan Sampel Penelitian.....	40

D. Teknik Pengumpulan Data.....	41
E. Teknik Analisis Data .....	42
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN .....</b>	<b>46</b>
A. Analisis Klasifikasi Penjuruan Siswa dengan Algoritma C4.5 .....	46
<b>BAB V PENUTUP.....</b>	<b>65</b>
A. Kesimpulan.....	65
B. Saran.....	66
<b>DAFTAR PUSTAKA</b>	

## DAFTAR TABEL

Tabel 1.1 Data siswa SMA N 2 Padang kelas X Tahun Ajaran 2020/2021.....	7
Tabel 2.2 Contoh Data set .....	25
Tabel 2.3 Contoh Perhitungan Nilai Gain dan Entropy .....	27
Tabel 2.4 Contoh Perhitungan Nilai Gain dan Entropy Node 1.1 .....	29
Tabel 2.5 Contoh Perhitungan Nilai Gain dan Entropy Node 1.1.2 .....	30
Tabel 4.1 Atribut yang Digunakan Pada Penelitian .....	47
Tabel 4.2 Rekap Data Training .....	48
Tabel 4.3 Rekap Data Testing .....	49
Tabel 4.4 Nilai Standar Kriteria Ketuntasan Minimal (KKM) .....	50
Tabel 4.5 Perhitungan C4.5 .....	54

## DAFTAR GAMBAR

Gambar 2.1 <i>Data mining</i> merupakan bagian pada berbagai disiplin.....	16
Gambar 2.2 Konsep Dasar Graph Pohon Keputusan.....	22
Gambar 2.3 Konsep Pohon Keputusan Secara Garis Besar .....	23
Gambar 2.4 Pohon Keputusan Hasil Perhitungan Node 1 .....	28
Gambar 2.5 Pohon Keputusan Hasil Perhitungan Node 1.1 .....	30
Gambar 2.6 Pohon Keputusan Hasil Perhitungan Node 1.1 .....	31
Gambar 3.1 Proses Analisis Data .....	42
Gambar 4.1 Tampilan Awal RapidMiner.....	55
Gambar 4.2 Tampilan Menu Utama Sebelum Proses.....	56
Gambar 4.3 Tampilan Process Memilih Data Training dan Data Testing.....	56
Gambar 4.4 Tampilan Process Memilih Decision tree .....	57
Gambar 4.5 Tampilan Process Memilih Apply Model.....	58
Gambar 4.6 Tampilan Process Memilih Performance .....	58
Gambar 4.7 Tampilan Process Run dan Hasil Pohon Keputusan.....	59
Gambar 4.8 Tampilan Performance Vector .....	60
Gambar 4.9 Hasil Proses Accurasy.....	61
Gambar 4.10 Hasil Proses Precision.....	62

Gambar 4.11 Hasil Proses recall.....	63
--------------------------------------	----

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Dalam perkembangan teknologi masa kini, banyaknya data menjadi sebuah permasalahan sekaligus kesempatan bagi sebuah instansi. Data menjadi permasalahan apabila tidak dapat disimpan, dikelola, maupun diproses dengan baik. Sedangkan data menjadi sebuah kesempatan apabila dapat disimpan, dikelola dan diproses menjadi lebih berarti untuk instansi tersebut. Dengan adanya data, maka dapat ditemukan sebuah *trend* maupun struktur yang nantinya dapat dipergunakan untuk mendapatkan informasi di masa mendatang.

Pengelolaan data yang sangat besar akan melibatkan proses data mining. Roiger (2017) menyatakan *data mining* adalah proses maupun tahapan dalam menemukan sebuah struktur data. Struktur data tersebut dapat mengambil banyak bentuk, termasuk aturan, grafik atau jaringan, pohon (tree) maupun persamaan, serta beberapa yang lain. Dengan menggunakan *data mining*, maka sebuah kasus dapat dilihat *trend*, struktur maupun prediksinya di masa mendatang. *Data mining* sendiri memiliki banyak tahapan dan teknik yang dapat di implementasikan dalam kehidupan nyata.

Data mining merupakan proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan (Mc Leod, Jr.,R. 2007). Perlunya data mining karna adanya sejumlah data yang besar yang dapat digunakan sehingga dapat menghasilkan informasi dan *knowledge* yang berguna. Dalam kasus dunia nyata, banyak teknik dalam *data mining* yang dapat digunakan, salah satunya adalah teknik klasifikasi. Klasifikasi sendiri merupakan bentuk dasar dari analisis data. Bansar Sharma dan Goel (2017) menyatakan klasifikasi adalah sebuah teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula.

Dengan menerapkan teknik klasifikasi *data mining*, dapat digali informasi yang menjadi salah satu bahan rekomendasi penjurusan siswa SMA. Informasi ini dapat digunakan sebagai penunjang data yang dimiliki oleh pihak sekolah. Terdapat beberapa algoritma yang dapat membantu proses klasifikasi data siswa untuk rekomendasi penjurusan, misalnya C4.5, *Naïve Bayes*, *K-NN*, *Rule Induction*, dan lain-lain.

Dalam pengambilan keputusan untuk memilih bidang studi siswa harus sesuai keahliannya. Hal ini sejalan dengan Kristianto, O., (2014), yang menyatakan bahwa penjurusan siswa adalah suatu proses pengambilan keputusan dalam memilih keahlian bidang studi berdasarkan

kemampuan dan potensi diri yang ada. Secara formal pemilihan jurusan merupakan ketentuan yang sudah ditetapkan oleh pemerintah melalui kurikulum 2013 yang berlaku saat ini. Untuk menentukan penjurusan siswa dilakukan pada awal masuk, yaitu pada kelas X Sekolah Menengah Atas (SMA). Penerapan kurikulum 2013 bertujuan untuk penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah sekolah siswa. Tujuan dari penjurusan itu sendiri adalah kelak dikemudian hari pelajaran yang diberikan kepada siswa lebih terarah. Karena tidak jarang juga siswa siswi yang asal asalan dalam menentukan jurusan yang mereka ambil. Seiring perkembangan teknologi hal tersebut dapat diatasi dengan teknik pengelompokan data dengan *data mining* (Liliana Swastina, 2013).

Menurut (kementrian pendidikan dan kebudayaan, 2013), adanya kurikulum 2013 memberikan kesempatan kepada siswa khususnya tingkat SMA/MA memilih mata pelajaran yang diminati dan sesuai dengan kemampuan belajarnya. Selain itu memberikan peluang dalam mengembangkan potensi yang dimiliki siswa sesuai dengan kemampuan dasar umum, bakat, minat, dan karakteristik kepribadian. Hal ini juga sejalan dengan (Sopianti, Lilis & Nurdin Bahtiar, 2015), penelitian sebelumnya tentang peminatan peserta didik pernah dilakukan dengan menggunakan kriteria nilai rapor, nilai ujian nasional, nilai ujian akhir sekolah, catatan prestasi, pilihan peminatan dan hasil psikotest. Banyak

kasus dijumpai bahwa pemilihan jurusan yang tidak sesuai dengan kemampuan, kepribadian, minat dan bakat dapat mempengaruhi siswa dalam mengikuti pelajaran. Dalam beberapa penelitian psikologi pendidikan, minat dan bakat siswa diketahui cukup terkait dengan prestasi akademiknya (Musrofi M, 2010).

Penjurusan di SMA Negeri 2 Padang dilakukan pada awal tahun pelajaran sesuai dengan kurikulum 2013 yang memiliki dua program penjurusan, yaitu jurusan Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Penjurusan di SMA Negeri 2 Padang dilakukan agar para siswa bisa menyalurkan minat dan bakat serta kemampuan jurusan yang ada. Sehingga siswa dapat mengembangkan potensi yang dimilikinya. Maka sekolah harus mampu memfasilitasi siswa agar sesuai dengan kemampuan yang dimiliki sehingga siswa di tempatkan di jurusan yang tepat. Penjurusan ini sangat membantu siswa dalam mempersiapkan diri memasuki perguruan tinggi, dimana saat itu siswa harus memiliki spesialisasi program studi tertentu. Sering kali siswa asal dalam memilih jurusan sehingga nantinya setelah lulus SMA siswa akan kebingungan memilih jurusan apa di perguruan tinggi sehingga dapat menghambat cita-citanya. Setelah observasi dengan pihak sekolah pada tanggal 15 Februari 2021 , penjurusan di SMA Negeri 2 Padang, baik IPA dan IPS dipertimbangkan dalam beberapa faktor, yaitu nilai rapor Sekolah Menengah Pertama (SMP) dan rekomendasi guru BK SMP. Setelah

melakukan wawancara secara langsung terhadap sebagian siswa SMA N 2 Padang, dapat disimpulkan bahwa masih ada siswa yang merasa dirinya tidak dalam jurusan yang tepat.

Menurut M. Nur (2014) pengambilan keputusan penjurusan sesuai kurikulum 2013 yaitu menentukan jurusan saat siswa duduk dibangku kelas X dilakukan oleh pihak sekolah dengan melihat beberapa faktor diantaranya yaitu nilai Rapor SMP dan rekomendasi guru BK SMP.

Nilai rapor SMP siswa diharapkan menjadi bahan pertimbangan untuk menentukan penjurusan yang tepat, beberapa siswa memiliki nilai yang rendah namun setiap kenaikan semester menunjukkan grafik yang meningkat. Psikotest sebenarnya bukanlah hasil ujian, karna sebenarnya tidak benar kalau dikatakan tidak lulus ujian psikotes karena yang dicari adalah meminta respond atas pertanyaan/pernyataan yang diberikan kepada siswa agar membantu pendidik untuk mengetahui minat bakat siswa. Oleh karena itu penjurusan siswa berdasarkan kurikulum 2013 mempertimbangkan banyak hal, peminatan pada kelas X juga bertujuan untuk membentuk mental siswa yang siap bersaing sesuai penjurusan yang telah ditentukan.

Setelah melakukan konsultasi dengan pihak di sekolah SMA Negeri 2 Padang tersebut, dalam menentukan penjurusan dengan memperhatikan banyak faktor yang kompleks, data mining lebih efisiensi jika diterapkan didalam ilmu data mining di penjurusan, karena kegunaan

dari data mining yaitu untuk menggali informasi dari banyak nya data. Selain itu, cara seperti ini memungkinkan terjadinya kesalahan baik yang manusiawai maupun yang disengaja. Oleh karna itu diperlukan klasifikasi penjurusan yang tepat dan akurat, salah satunya menggunakan teknologi dalam bidang data mining.

Penggunaan perangkat lunak aplikasi *data mining* banyak diterapkan di berbagai bidang, diharapkan aplikasi data mining dapat digunakan untuk dapat mengubah data menjadi suatu pengetahuan dan membantu pekerjaan. Perangkat lunak aplikasi *data mining* yaitu seperti Orange, *Rapid miner* dan Weka. Aplikasi yang sering digunakan saat ini yaitu Weka dan *Rapid miner*. Aplikasi Weka memiliki tampilan yang sederhana berbasis java, memiliki banyak *tools* untuk pengolahan data, mulai dari *Pre-processing*, *classification*, *association rules*, dan *visualization*. Sedangkan *Rapid miner* mampu menganalisa data yang berukuran yang sangat besar sampai ke level big data, *Rapid miner* merupakan sebuah *tools* yang digunakan dalam teknik yang berada dilingkungan *machine learning*, *data mining*, *text mining* dan *predictive analytics* (Muslehatin & Ibnu, 2017).

Berdasarkan uraian masalah diatas dan sesuai hasil observasi di sekolah, maka akan dibuat pengklasifikasian penjurusan siswa pada SMA Negeri 2 Padang. Penelitian ini menggunakan teknologi *data mining* dengan menggunakan algoritma C4.5. Diharapkan algoritma C4.5 tersebut

mampu jadi solusi pendukung dalam menentukan penjurusan siswa di SMA Negeri 2 Padang.

Tabel 1.1 Data siswa SMA N 2 Padang kelas X Tahun Ajaran 2020/2021

NO	Jurusan	Jumlah Kelas	Jumlah Siswa
1	IPA	7	280
2	IPS	3	120
	TOTAL	10	400

(Sumber: Arsip Data Siswa Kelas X Tahun Ajaran 2020/2021)

Salah satu cara untuk mengolah data siswa ini dengan menggunakan *data mining*. Dengan proses data mining dapat ditemukan pola atau aturan yang dapat digunakan untuk menentukan penjurusan siswa berdasarkan variable tersebut. Metode dalam *data mining* diantaranya yaitu klasifikasi, prediksi, klustering, estimasi, dan asosiasi.

Dari beberapa algoritma klasifikasi, dapat digunakan sesuai kebutuhan klasifikasi data. Penelitian ini bertujuan untuk menentukan penjurusan siswa SMA N 2 Padang menggunakan algoritma *Decision Tree* C4.5. Algoritma *Decision Tree* menjadi algoritma yang memiliki keakuratan paling bagus di antara algoritma yang lainnya (Nugroho, S. Yusuf. 2015). Algoritma klasifikasi C4.5 dapat digunakan sebagai

algoritma klasifikasi dalam *data mining*, yang menghasilkan pohon keputusan (Dhika, Harry, 2015).

## **B. Identifikasi Masalah**

Berdasarkan latar belakang masalah yang ditemukan, dapat diidentifikasi masalah sebagai berikut :

1. Pihak sekolah belum akurat atau belum tepat dalam menentukan penjurusan siswa karna masih banyak siswa yang tidak didalam jurusan yang tepat.
2. Belum adanya semua aspek-aspek atau ketentuan yang harus dipakai dalam menentukan penjurusan siswa di SMAN 2 Padang.
3. Pihak sekolah masih menggunakan *microsoft excel* yang menyita banyak waktu dan membutuhkan tenaga ekstrim untuk menentukan penjurusan siswa.

## **C. Batasan Masalah**

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Ruang lingkup penelitian dibatasi pada data siswa kelas X SMA Negeri 2 Padang tahun ajaran 2020.
2. Metode yang digunakan adalah metode klasifikasi dengan menggunakan algoritma *Decision Tree C4.5*.
3. Variabel yang digunakan dalam penelitian ini adalah nilai rapor Sekolah Menengah Pertama (SMP) dan rekomendasi guru BK SMP

4. Program analisis yang digunakan adalah *Rapid Miner*.

#### **D. Rumusan Masalah**

Adapun rumusan masalah yang menjadi pokok penelitian adalah bagaimana menentukan penjurusan siswa di SMA Negeri 2 Padang dengan lebih akurat dan tepat dengan menggunakan metode klasifikasi algoritma C4.5.

#### **E. Tujuan Penelitian**

Dari rumusan masalah yang telah diuraikan di atas, ada beberapa tujuan dari penelitian ini, yaitu :

1. Untuk menentukan penjurusan siswa dengan hasil yang akurat dan tepat dengan menerapkan metode *data mining* dengan menggunakan algoritma *Decision Tree C4.5*.
2. Untuk mempercepat waktu dan tidak memakan banyak tenaga dalam menentukan penjurusan siswa.

#### **F. Manfaat Penelitian**

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoritis
  - a. Menambah hasanah penelitian dalam pemahaman dalam data mining.
  - b. Menjadi *best practice* dalam penelitian menggunakan metode algoritma C4.5.

## 2. Manfaat Praktis

- a. Memudahkan pihak akademik, khususnya sekolah SMA Negeri 2 Padang dalam proses pengklasifikasi penjurusan siswa berdasarkan nilai rapor SMP dan rekomendasi guru BK SMP.
- b. Bagi pendidik memudahkan secara langsung dalam proses pembelajaran ataupun mendidik supaya lebih terarah untuk kedepannya bagi siswa.
- c. Memudahkan siswa secara akademik maupun non akademik, secara akademik siswa lebih nyaman dalam mengikuti kegiatan pembelajaran di sekolah dan secara non akademik terbentuknya mental siswa yang siap bersaing sesuai penjurusan yang telah ditentukan.
- d. Memudahkan bagi orang tua dalam mengarahkan anaknya saat dirumah untuk mempelajari yang sesuai dengan peminatan yang telah ditentukan
- e. Bagi penulis merupakan salah satu syarat untuk menyelesaikan pendidikan program strata satu, program studi Pendidikan Teknik Informatika.

## **BAB II**

### **KAJIAN TEORISTIS**

#### **A. Penjurusan Siswa**

Pada kurikulum baru, penjurusan dilakukan pada awal masuk sekolah yaitu kelas X. Perubahan kurikulum dimaksudkan agar program pendidikan dapat disesuaikan dengan satuan pendidikan dengan kondisi dan kekhasan potensial dilingkungan siswa (Peraturan Kemendikbud).

(Bahar, 2011) Dalam proses pendidikan di sekolah, perbedaan pada setiap siswa harus diperhatikan karena dapat menentukan baik buruknya prestasi belajar siswa. Tujuan dasar sekolah adalah untuk mengembangkan semua bakat dan kemampuan siswa selama proses pendidikan. Perbedaan individu antar siswa di sekolah perbedaan kemampuan kognitif, motivasi berprestasi, minat, dan kreativitas. Adanya perbedaan individu tersebut maka fungsi pendidikan tidak hanya dalam proses belajar mengajar, tetapi juga penyuluhan, sehingga pemilihan dan penempatan jurusan bagi peserta didik harus didasarkan pada kapasitas individu sebagai mahasiswa. (Musrofi, 2010) Dalam beberapa penelitian psikologi pendidikan, minat dan bakat siswa diketahui cukup terkait dengan prestasi akademiknya.

#### **1. Faktor-faktor yang Mempengaruhi Penjurusan Siswa**

Pihak sekolah biasanya menetapkan jurusan berdasarkan nilai yang diperoleh oleh siswa atas mata pelajaran yang diberikan. Siswa

sendiri biasanya menentukan jurusan yang di inginkan berdasarkan kemampuan sendiri. Menurut (Maya, 2007) kecendrungan pemilihan dari teman dekat atau pun arahan dari orang tua biasanya berdasarkan dari pengalaman pekerjaan dan gengsi maupun trend yang sedang terjadi.

Perkembangan karier yang pernah dirumuskan oleh *the national vocational guidance association* (1973) sebagai gabungan faktor-faktor psikologis, sosiologis, pendidikan, fisik, ekonomi, dan kesempatan yang bersama-sama membentuk jabatan seseorang. Gabungan ini mencakup banyak faktor internal dan eksternal (Winkel, 2007).

#### a. **Faktor Internal**

Faktor internal dapat dibedakan satu dengan yang lain karena sama-sama membentuk keunikan kepribadian seseorang. Faktor internal yang mempengaruhi siswa dalam memilih jurusan antara lain: minat, cita-cita, bakat, hasil prestasi belajar dan motivasi (Winkel, 2007).

Menurut Dinn Wahyudin (2008) faktor internal meliputi keadaan fisik, keadaan emosi, intelegensi, bakat khusus, perhatian dan kebiasaan belajar.

#### b. **Faktor Eksternal**

Faktor eksternal dibedakan satu dengan yang lain, tetapi tidak dapat dipisahkan satu sama lain karena bersama-sama menciptakan keseluruhan ruang gerak hidup.

Menurut (Winkel, 1997) faktor eksternal yang mempengaruhi siswa dalam pemilihan jurusan antara lain: pengaruh sosial, pengaruh teman sebaya dan pendidikan sekolah.

Menurut (Siahan & Ronaldi, 2013) faktor eksternal terdiri dari perhatian orang tua dan peran guru. Semakin tinggi perhatian orang tua terhadap prestasi belajar anaknya maka akan semakin tinggi pula prestasi anak tersebut.

## **2. Tujuan Penjurusan**

Sedangkan menurut (Darminto & Erika Ristiyani, 2016) faktor yang paling dominan yang mempengaruhi keberhasilan pembelajaran salah satunya adalah kualitas guru. Peran guru dapat mempengaruhi hasil belajar, bisa dilihat dari cara mengajar kepada siswa, hal ini sangat menentukan dalam keberhasilan belajar.

Penjurusan atau Course yang ditawarkan di level pendidikan menengah diterapkan di Indonesia sejak zaman Belanda. Sekolah HBS yang merupakan Sekolah Menengah untuk anak-anak Eropa, dan SMA yang merupakan Sekolah Menengah Atas untuk anak-anak pribumi pertama kalinya dibagi atas 2 course yaitu Budaya (Kelompok A) dan Sains

(Kelompok B). Pada masa selanjutnya sistem penjurusan di Indonesia diterapkan sejak SMP, yang kemudian dihapuskan pada tahun 1962.

Penjurusan diperkenalkan sebagai upaya untuk lebih mengarahkan siswa berdasarkan minat dan kemampuan akademiknya. Siswa yang mempunyai kemampuan sains dan ilmu eksakta yang baik biasanya akan memilih jurusan IPA, dan yang memiliki minat pada sosial ekonomi akan memilih jurusan IPS, lalu yang gemar berbahasa akan memilih jurusan Bahasa.

### **3. System Penjurusan**

Sekolah SMAN 2 Padang dapat menentukan penjurusan siswa yang tepat apabila telah memenuhi faktor-faktor yang telah ditentukan. Beberapa faktor yang telah ditentukan ialah :

- a. Nilai rapor SMP, berdasarkan kurikulum 2013 pejurusan siswa siswa SMA dimulai pada waktu kelas X. Nilai rapor SMP siswa diharapkan menjadi pertimbangan untuk menentukan penjurusan yang tepat, beberapa siswa memiliki nilai yang rendah tetapi setiap kenaikan semester menunjukkan grafik peningkatan.
- b. Rekomendasi BK SMP

Menurut Gibson & Mitchell (2011:99) menyatakan fungsi terpenting guru BK sebagai berikut:

- 1) penyediaan konseling individu
- 2) menawarkan bantuan, perencanaan, dan informasi karir.

- 3) melakukan tes dan menafsirkan hasilnya.
- 4) membantu menempatkan siswa di jurusan studi dan arah pendidikan lebih lanjut.

Terdapat beberapa hal beberapa hal yang merujuk kea arah kajian teori tentang penjurusan di sekolah, antara lain sebagian siswa cenderung memiliki komitmen mengenai jurusan yang tepat untuk dirinya, kebingungan untuk menentukan jurusan tertentu, dan bingung mengenai arah karir setelah tamat sekolah. Oleh karena itu, mereka mempertimbangkan rekomendasi dari guru BK, dan hal ini menunjukkan pentingnya dukungan guru BK terhadap minat siswa pada jurusan tertentu.

## **B. Data Mining**

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. *Data mining* adalah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengestrak dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai *database* besar (Turban, dkk. 2005).

*Data mining* adalah suatu proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (Angga Ginanjar Maburur, dkk. 2012).

Menurut Han (2006) “*data mining* merupakan kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data yang disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain seperti, *database system*, *data warehousing*, *statistic*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*.”

Dari definisi-definisi yang telah disampaikan, hal terpenting yang terkait dengan *data mining* menurut (Bramer, Max. 2007).

- a. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
- b. Data yang akan diproses merupakan data yang sangat besar
- c. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Berdasarkan beberapa pengertian pengertian tersebut, maka dapat disimpulkan bahwa data mining merupakan suatu proses teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu data dan menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting. Kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karna itu data mining sebenarnya memiliki akar yang panjang dari

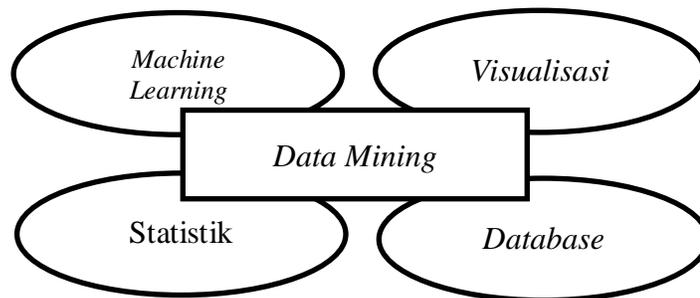
bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistika dan *database*.

### **1. Pengenalan pola, *Data Mining* dan *Machine Learning***

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasi obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohonan kredit, *image*, atau *signal* atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya (Santoso, 2007).

Data mining sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, histori untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).

*Machine Learning* adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, *data mining* dan *machine learning* sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi. *Machine learning* menjadi alat analisis dalam *data mining*. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam gambar (Santoso, 2007)



**Gambar 2.1.** *Data mining* merupakan bagian pada berbagai disiplin.

## 2. Tahap-tahap *data mining*

Menurut Mujib Ridwan, dkk, (2013), sebagai suatu rangkaian proses, data mining juga dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakaian terlibat langsung atau dengan perantaraan *knowledge base*.

Tahap-tahap *data mining* adalah sebagai berikut :

### a. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembuangan data juga akan

mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

b. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

c. Seleksi Data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi Data (*data transformation*)

Data diubah atau digabungkan ke dalam format yang sesuai untuk diproses dalam *data mining*. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal.

e. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk mencari pola atau pengetahuan berharga yang tersembunyi dari data dengan menggunakan teknik, metode dan algoritma tertentu.

f. Evaluasi pola (*pattern evaluation*)

Mengidentifikasi pola-pola yang menarik dari hasil *data mining*. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai atau tidak.

g. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

### 3. Metode data mining

Dengan definisi data mining yang luas, ada banyak jenis metode analisis yang dapat digolongkan dalam data mining.

a. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable targetnya *numeric* dari pada kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variable target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variable target ditentukan berdasarkan nilai variable prediksi.

Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana

b. Prediksi

Prediksi adalah menentukan jumlah kebutuhan bulan mendatang terkait dengan dukungan data historis atau serangkaian waktu/periode yang dianalisis sehingga dapat diperhitungkan untuk memprediksi jumlah kebutuhan pada bulan mendatang. Prediksi juga dapat digunakan dalam pengklasifikasi, tidak hanya untuk memprediksi *time series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan atribut yang ada

Prediksi juga hampir sama dengan klasifikasi dan estimasi, akan tetapi dalam prediksi, hasil terjadi di masa datang. Contoh prediksi dalam bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang
- b. Prediksi tingkat pengangguran lima tahun yang akan datang
- c. Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

c. Klasifikasi

Variabel target dalam klasifikasi merupakan kategorikal. Model *data mining* memeriksa set *record* yang besar, tiap *record*

mempunyai informasi variable target dan set input atau variabel prediksi

Pada persoalan klasifikasi, kita memiliki sejumlah kasus (sampel data) dan ingin memprediksi beberapa *class* yang ada pada sampel data tersebut. Tiap instan data berisi banyak atribut, dimana masing-masing atribut memiliki satu dari beberapa kemungkinan nilai. Hanya satu atribut diantara banyak atribut tersebut yang disebut dengan atribut target, sedangkan atribut yang lain disebut sebagai atribut prediktor. Tiap kemungkinannilai yang dimiliki oleh atribut target menunjukkan *class* yang diprediksi berdasarkan nilai-nilai dari atribut prediktor .

Klasifikasi digunakan untuk *segmentasi costumer*, pemodelan bisnis, analisa kartu kredit, dan banyak aplikasi yang lain. Sebagai contoh, perusahaan kartu kredit ingin memprediksi *costumer* berdasarkan tipe pembayaran.

d. *Clustering*

Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam klusteran tidak mencoba untuk melakukan

klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal. Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok konsumen untuk target pemasaran suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar
  - b. Untuk tujuan audit akuntansi, yaitu melakukan pemeriksaan terhadap perilaku finansial dalam baik dan mencurigakan.
  - c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku gen dalam jumlah besar.
- e. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Salah satu implementasi dari asosiasi adalah analisis keranjang belanja. Teknik analisa keranjang pasar merupakan teknik yang mengadaptasi ilmu data mining. Teknik ini digunakan untuk merancang suatu strategi penjualan dan pemasaran barang melalui proses pencarian asosiasi atau hubungan antar item data dari suatu basis data relasional. Pencarian asosiasi berawal

daripengolahan data transaksi pembelian barang dari setiap pembeli, kemudian dicari hubungan antar barang-barang yang dibeli. Pencarian informasi ini hampir sama dengan mencari peluang kemunculan barang yang dibeli sesuai dengan kebiasaanberbelanja masyarakat dan jumlah transaksi yang ada.

#### **4. Penerapan Metode Klasifikasi**

##### **a. Klasifikasi**

Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*. Berikut beberapa definisi klasifikasi yaitu menurut Kusri dan Lutfi dalam jurnal Dicky Nofriansyah et al. (2016 : 87), “Klasifikasi adalah fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah unsur (item) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan”.Dwi Sartika dan Dana Indra Sensuse (2017 : 153) menyebutkan dalam jurnalnya yang berjudul Perbandingan Algoritma Klasifikasi *Naïve Bayes*, *Nearest Neighbour*, dan *Decision Tree* pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian yaitu : “Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas”.

Dari beberapa definisi diatas maka dapat disimpulkan bahwa klasifikasi merupakan pengelompokan sampel berdasarkan ciri-ciri

persamaan dan perbedaan dengan menggunakan variabel target sebagai kategori. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik dan merupakan variabel bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data.

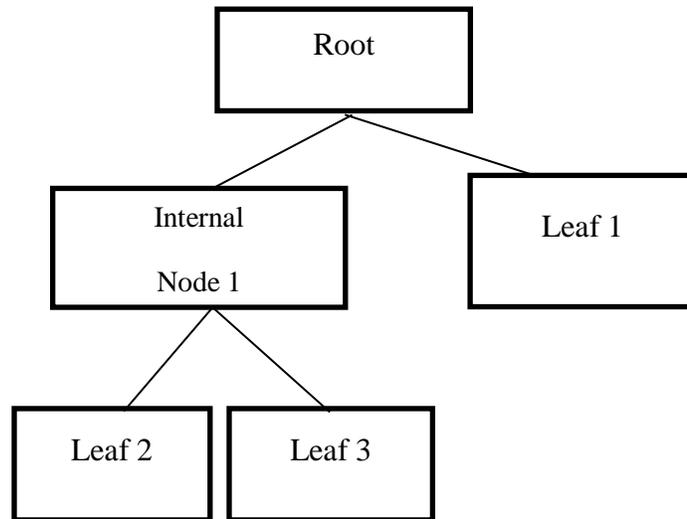
Contoh sederhana dari teknik *data mining* klasifikasi adalah pengklasifikasian hewan berdasarkan atribut jumlah kaki, habitat dan organ pernafasannya akan diklasifikasikan ke dalam dua label kelas yaitu unggas dan ikan. Label kelas unggas adalah data yang memiliki jumlah kaki dua, habitatnya di darat, dan organ pernafasannya menggunakan paru-paru, sedangkan label kelas ikan adalah data yang memiliki jumlah kaki nol (tidak memiliki kaki), habitatnya di air dan organ pernafasannya menggunakan insang. Banyak algoritma yang dapat digunakan dalam pengklasifikasian data, namun dalam penelitian ini akan menggunakan algoritma C4.5.

Menurut Mardi, Yuli (2014) Dalam klasifikasi terdapat target variable kategori. Metode-metode / model-model yang telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi antara lain :1) Pohon keputusan (2) Pengklasifikasi bayes (3)

Jaringan saraf tiruan (4) Analisis statistik (5) Algoritma genetik (6) *Rough sets* (7) Pengklasifikasi *k-nearest neighbor* (8) Metode berbasis aturan, (9) *Memory based reasoning* (10) *Support vector machine*

**b. Pohon Keputusan (*Decision Tree*)**

Pohon Keputusan (*Decision Tree*) merupakan algoritma pengklasifikasian yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan (Mantas, C. J., & Abellán, J. 2014). Menurut Sammut, G.W.C. (2011), 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. Tampilan graph pohon keputusan dapat dilihat pada Gambar 1.



**Gambar 2.2. Konsep Dasar Graph Pohon Keputusan**

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule, Basuki, A., & Syarif, I. (2003).



**Gambar 2.3. Konsep pohon keputusan secara garis besar**

Data yang akan dianalisis, dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*. Konsep data dalam pohon keputusan yaitu:

- 1) data dinyatakan dalam bentuk tabel dengan atribut dan record.
- 2) Atribut menyatakan suatu parameter yang dibuat sebagai

kriteria dalam pembentukan *tree*.

- 3) Atribut memiliki nilai-nilai yang dinamakan dengan *instance*.

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Bramer, Max (2007)).

Proses pada pohon keputusan adalah mengubah bentuk data (*table*) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Langkah pertama yang dilakukan dalam pembuatan pohon keputusan yaitu menghitung nilai *Entropy* total dari jumlah data yang dijadikan sampel, selanjutnya yaitu mengelompokkan *variable* dan menghitung nilai *gain* pada tiap atribut. Setelah dihitung menggunakan rumus algoritma C4.5 maka atribut yang mempunyai nilai *gain* tertinggi akan menjadi akar dan atribut lainnya menjadi cabang, kemudian dari cabang akan dihitung kembali atribut apalagi yang mempunyai nilai *gain* tertinggi. Tahapan perhitungan akan berulang secara *continue* sampai beberapa tingkatan hingga mencapai nilai akhir yaitu keputusan “*Yes*” dan “*No*”.

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem *break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon

keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Metode *Decision Tree* memiliki beberapa keunggulan dibandingkan metode lainnya untuk database yang besar, yaitu:

- 1) Memiliki kecepatan yang relative lebih cepat.
- 2) Dapat diubah menjadi *rule* klasifikasi dengan mudah dan sederhana.
- 3) Dapat menggunakan *query SQL* untuk mengakses *database*.
- 4) Dapat dibandingkan tingkat akurasi dengan metode lainnya.

### c. Algoritma *Decision Tree* 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. 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.

Tahapan dari algoritma *decision tree* C4.5 adalah sebagai berikut.

Tahapan penting dalam pohon keputusan yaitu membentuk table dengan atribut dan record sesuai data yang dimiliki. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan. Berikut ini contoh penerapan algoritma C4.5 dalam menyusun pohon keputusan pada suatu data. Hal tersebut dilakukan untuk menentukan bermain tenis atau tidak, kriteria yang diperlukan meliputi cuaca, angin,

kelembapan, dan temperatur udara.

Salah satu atribut merupakan data solusi per item data yang disebut target atribut. Misalnya atribut “*play*” dengan nilai “main” atau “tidak main”. Atribut memiliki nilai-nilai yang dinamakan dengan istilah *instance*. Misalnya atribut “Cuaca” memiliki *instance* yaitu cerah, berawan, dan hujan.

**Tabel 2.2. Contoh data set**

NO	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Berdasarkan tabel di atas akan dibuat tabel keputusan untuk menentukan main tenis atau tidak dengan melihat keadaan Outlook (cuaca), *Temperature* (temperatur), *Humidity* (kelembapan) dan *Windy* (keadaan angin).

Algoritma secara umum adalah memilih atribut sebagai akar, membuat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tahap selanjutnya adalah memilih atribut berdasarkan nilai *gain* tertinggi dari atribut-atribut yang ada. Perhitungan nilai *gain* dapat dilakukan dengan menggunakan rumus:

$$Gain(S, A) = Entro(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan:

A : atribut

V : nilai untuk atribut A

Values(A) : himpunan nilai yang mungkin untuk atribut A

|S<sub>v</sub>| : jumlah sampel untuk nilai v

|S| : jumlah seluruh sampel data

Entropy (S<sub>v</sub>) : Entropy untuk sampel yang memiliki nilai v

Nilai *Gain* berfungsi untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan data. Jadi, secara heuristic akan dipilih atribut yang menghasilkan simpul yang paling bersih (*purest*). Jika dalam cabang suatu *decision tree* anggotanya berasal dari satu kelas maka cabang ini disebut *pure*. Kriteria yang digunakan adalah dengan menggunakan *information gain* paling besar.

Ukuran *information gain* digunakan untuk memilih atribut uji pada setiap *node* di dalam *tree*. Ukuran ini digunakan untuk memilih atribut atau *node* pada pohon. Atribut dengan nilai *information gain* tertinggi akan terpilih sebagai *parent* bagi *node* selanjutnya.

Berdasarkan rumus *Information Gain* di atas, berarti kita juga harus menghitung nilai *entropy*. *Entropy* adalah suatu

parameter untuk mengukur tingkat keberagaman (heterogenitas) dari kumpulan data (*dataset*). Semakin heterogen, nilai *entropy* semakin besar. Rumus untuk menghitung nilai *entropy* adalah:

$$H(X) = - \sum p(X_i) \log_b p(X_i)$$

Keterangan:

$X$  : himpunan kasus

$n$  : jumlah partisi  $X$

$p(X_i)$  : proporsi dari  $X_i$  terhadap  $X$

Berdasarkan Tabel Contoh Dataset di atas, selanjutnya adalah menghitung jumlah kasus seluruhnya, jumlah berkeputusan *Yes* maupun *No*. Menghitung *Entropy* dari semua kasus yang terbagi berdasarkan atribut *Outlook*, *Temperature*, *Humidity*, *Windy* serta melakukan perhitungan *gain* setiap atributnya.

**Tabel 2.3. Contoh Perhitungan Nilai *Gain* dan *Entropy***

Node			Jumlah Kasus(S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1	TOTAL		14	4	10	0,863120569	
	OUTLOOK						0,258521037
		CLOUDY	4	0	4	0	
		RAINY	5	1	4	0,721928095	
		SUNNY	5	3	2	0,970950594	
	TEMP						0,183850925
		COOL	4	0	4	0	
		HOT	4	2	2	1	
		MILD	6	2	4	0,918295834	
	HUMIDITY						0,005977711
		HIGH	7	4	3	0,985228136	

		NORMAL	7	0	7	0	
	WINDY						0,005977711
		FALSE	8	2	6	0,811278124	
		TRUE	6	2	4	0,918295834	
1	Total		14	4	10	0,863120569	

Perhitungan entropy total secara manual adalah:

$$\begin{aligned} \text{Entropy (Total)} &= (-4/14 * \log_2 (4/14)) + (-10/14 * \log_2 (10/14)) \\ &= 0,863120569 \end{aligned}$$

Menghitung gain pada baris Outlook:

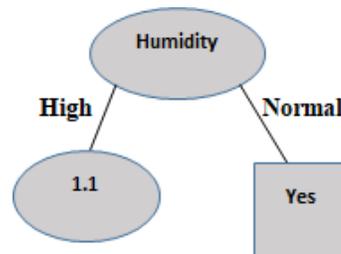
$$\begin{aligned} \text{Gain (Total, Outlook)} &= \\ &= 0,863120569 - ((4/14*0)+(5/14*0,72)+(5/14*0,97)) \\ &= 0,258521037 \end{aligned}$$

Selanjutnya menghitung nilai gain untuk *Temperature*, *Humidity*, dan *Windy*. Seperti yang terlihat pada tabel, diperoleh hasil bahwa atribut dengan gain tertinggi adalah *Humidity* 0,37, maka *Humidity* menjadi *node* akar. *Humidity* memiliki dua nilai yaitu *High* dan *Normal*. *Humidity Normal* sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya *Yes*. Untuk *Humidity High* masih perlu dilakukan perhitungan lagi (Karena masih terdapat *Yes* dan *No*).

Berdasarkan hasil pada Tabel 2.2 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah *Humidity* yaitu sebesar 0,37. Dengan demikian *Humidity* dapat menjadi *node* akar. Terdapat 2 nilai atribut dari *Humidity* yaitu *High* dan *Normal*.

Berdasarkan kedua nilai atribut tersebut, nilai atribut *Normal* sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya *Yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut *High* masih perlu dilakukan perhitungan lagi.

Berdasarkan hasil tersebut dapat di gambarkan pohon keputusan sementara sebagai berikut:



**Gambar 2.4. Pohon Keputusan Hasil Perhitungan Node 1.**

Tahap selanjutnya adalah menghitung jumlah kasus yang memiliki keputusan *Yes* dan jumlah kasus yang memiliki keputusan *No*. *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut *Outlook*, *Temperature*, dan *Windy* yang dapat dilakukan menjadi *node* akar dari nilai atribut *High*. Setelah itu lakukan perhitungan *Gain* untuk masing-masing atribut.

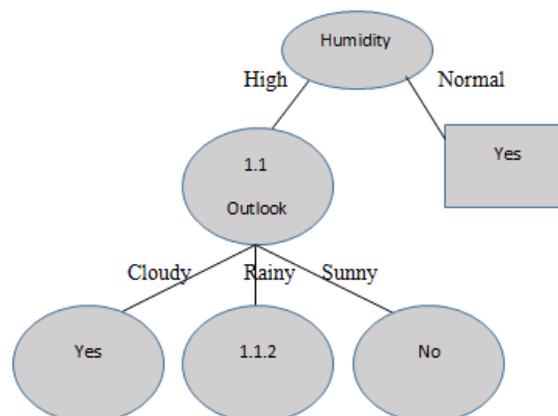
**Tabel 2.4. Contoh Perhitungan Nilai *Gain* dan *Entropy* Node 1.1**

Node			Jumlah Kasus(S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1.1	HUMIDITY-HIGH		7	4	3	0,965228136	

	OUTLOOK						0,69951385
		CLOUDY	2	0	2	0	
		RAINY	2	1	1	1	
		SUNNY	3	3	0	0	
	TEMP						0,020244207
		COOL	0	0	0	0	
		HOT	3	2	1	0,918295834	
		MILD	4	2	2	1	
	WINDY						0,020244207
		FALSE	4	2	2	1	
		TRUE	3	2	1	0,918295834	

Berdasarkan hasil pada Tabel 2.3 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah *Outlook* yaitu sebesar 0,67. Oleh karena itu *Outlook* dapat menjadi *node* cabang dari nilai atribut *High*. Terdapat 3 nilai atribut dari *Outlook* yaitu *Cloudy*, *Rainy* dan *Sunny*. Berdasarkan ketiga nilai atribut tersebut, nilai atribut *Cloudy* sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya *Yes* dan nilai atribut *Sunny* sudah mengklasifikasikan kasus menjadi satu dengan keputusan *No*, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut *Rainy* masih perlu dilakukan perhitungan lagi

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan padagambar 2.5.



### Gambar 2.5. Pohon Keputusan Hasil Perhitungan Node 1.1

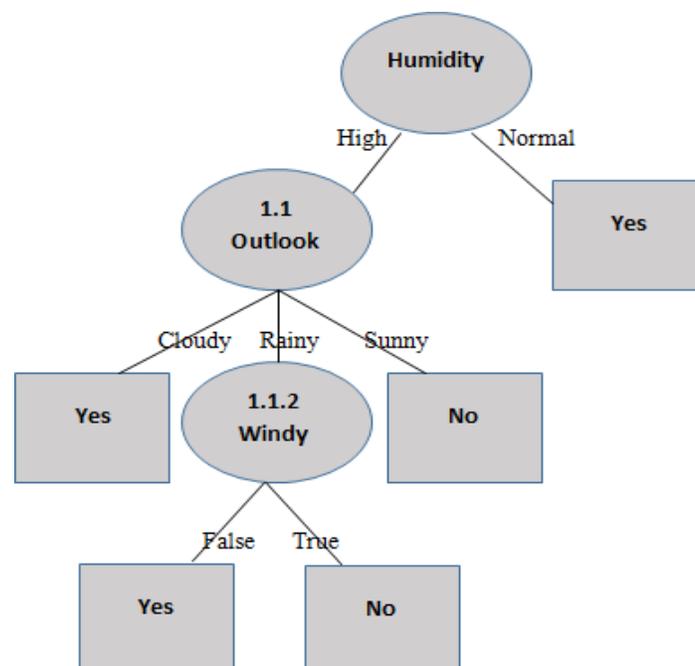
Tahap selanjutnya adalah menghitung jumlah kasus, jumlah kasus yang memiliki keputusan *Yes*, jumlah kasus yang memiliki keputusan *No*, dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut *Temperature* dan *Windy* yang dapat menjadi *node* cabang dari nilai atribut *Rainy*. Setelah itu lakukan perhitungan *Gain* untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel berikut ini.

**Tabel 2.5. Contoh Perhitungan *Gain* dan *Entropy* Node 1.1.2**

Node			Jumlah Kasus(S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1.1.2	Humidity High dan Outlook, Rainy		2	1	1	1	
	Temp						0
		COOL	0	0	0	0	
		HOT	0	0	0	0	
		MILD	2	1	1	1	

Node			Jumlah Kasus(S)	Tidak (S1)	Ya (S2)	Entropy	Gain
	Windy						1
		FALSE	1	0	1	0	
		TRUE	1	1	0	0	

Berdasarkan hasil pada Tabel 2.4 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah *Windy* yaitu sebesar 1. Oleh karena itu *Windy* dapat menjadi *node* cabang dari nilai atribut *Rainy*. Terdapat 2 nilai atribut dari *Windy* yaitu *False* dan *True*. Berdasarkan kedua nilai atribut tersebut, nilai atribut *False* sudah mengklasifikasikan kasus menjadi 1 yaitu yang berkeputusan *Yes* dan nilai atribut *True* sudah mengklasifikasikan kasus menjadi satu dengan keputusan *No*, sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini.



### **Gambar 2.6. Pohon Keputusan Hasil Perhitungan Node 1.1.**

Berdasarkan pohon keputusan pada Gambar 2.4, diketahui bahwa semua kasus sudah masuk dalam kelasnya masing-masing. Oleh karena itu, pohon keputusan pada Gambar 2.4 merupakan pohon keputusan terakhir yang terbentuk.

#### **5. *Rapid Miner***

*Rapid Miner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapid Miner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining*, dan analisis prediksi (Dennis Aprilla, 2013). *Rapid Miner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapid Miner* memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. *Rapid Miner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *Rapid Miner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

#### **C. Penelitian Relevan**

Untuk memperkuat teori-teori yang sudah di jelaskan di dalam kajian teori, kini penulis mengambil beberapa penelitian yang relevan dengan penelitian ini:

1. Husni Naprin (2014) yang berjudul “KLASIFIKASI PEMINATAN SISWA SMA MENGGUNAKAN METODE *NAÏVE BAYES*” kesimpulan pada penelitian ini yaitu dari hasil penelitian dengan menggunakan algoritma *naive bayes* sendiri sudah memiliki akurasi yang sangat baik yaitu sebesar 99,47 % dalam mengklasifikasikan status peminatan siswa SMA.
2. Liliana Swatina (2013) yang berjudul “PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN JURUSAN MAHASISWA” kesimpulan pada penelitian ini yaitu algoritma C4.5 akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat keakuratan 93,31 % dan akurasi rekomendasi jurusan sebesar 82,64 %.
3. Eka Budi Rahayu yang berjudul “ALGORITMA C4.5 UNTUK PENJURUSAN SISWA SMA NEGERI 3 PATI” kesimpulan pada penelitian ini adalah klasifikasi penjurusan siswa dengan algoritma C4.5 dapat mengklasifikasi siswa untuk penjurusan kelas IPA dan IPS
4. Beti Novianti, Tedy Rismawan, Syamsul Bahri (2016) yang berjudul “IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK PENJURUSAN SISWA (STUDI KASUS SMAN 1 PONTIANAK)” kesimpulan pada penelitian ini adalah hasil penilaian tingkat kinerja hasil klasifikasi penjurusan siswa dengan

menggunakan tabel penilaian *confussion matrix* menghasilkan nilai *precision* sebesar 89,47, nilai *precision* sebesar 96,55%, dan nilai *recall* sebesar 90,32%.

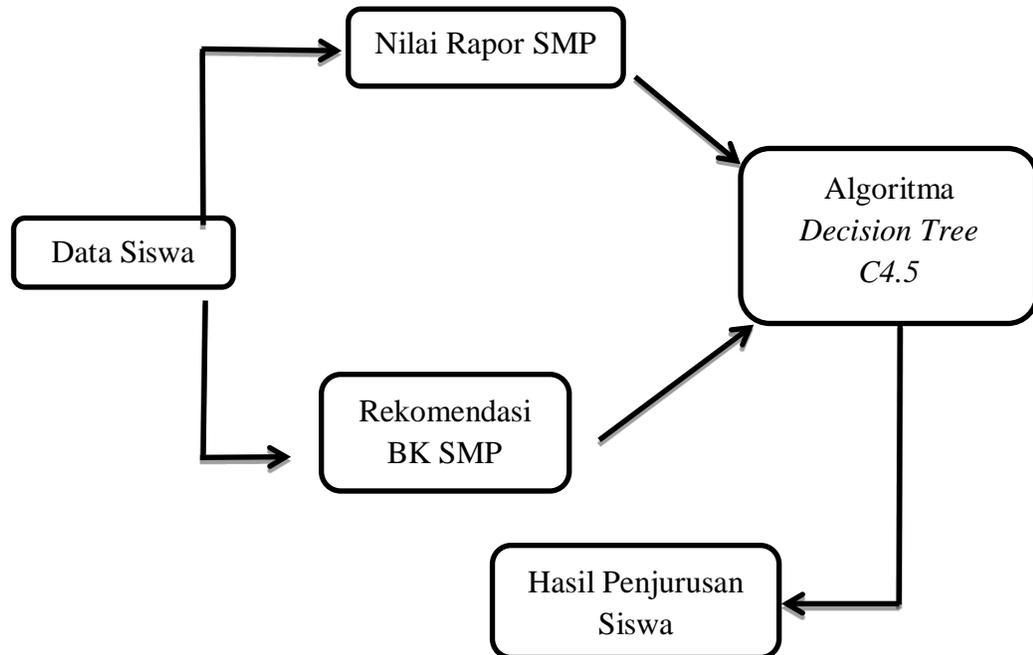
5. Alfa Saleh (2015) yang berjudul “PENERAPAN DATA MINING DALAM MENENTUKAN PENJURUSAN SISWA” kesimpulan pada penelitian ini adalah berdasarkan data akademik siswa yang diperoleh, metode *naive bayes* berhasil mengklasifikasi 90 data siswa dari 100 data yang diuji, sehingga metode *naive bayes* ini berhasil memprediksi jurusan siswa dengan persentase keakuratan sebesar 90%.

#### **D. Kerangka Berfikir**

Penjelasan kerangka pikir yang dikutip oleh Sugiyono (2007:91) bahwa kerangka pikir merupakan model konseptual tentang bagaimana teori berhubungan dengan berbagai faktor yang telah diidentifikasi sebagai masalah yang penting. Kerangka pikir dalam penelitian ini adalah penerapan metode data mining untuk memprediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi dan kedisiplinan.

Kerangka berfikir konseptual merupakan gambaran dari pengukuran yang dilakukan untuk menentukan penjurusan siswa dengan menggunakan algoritma C4.5 yang dimana studinya di SMA

Negeri 2 Padang. Pada penelitian ini ada beberapa variabel dan parameternya yaitu.



## BAB V

### PENUTUP

#### A. Kesimpulan

Dari hasil penelitian dan pembahasan yang dilakukan, maka penulis dapat mengambil kesimpulan sebagai berikut:

1. Dari hasil penelitian yang dilakukan dalam proses pengujian sebanyak 304 *rec ord* data *testing* yang diuji menyatakan bahwa algoritma C4.5 dapat menghasi lkan tingkat akurasi sebesar 68,42%.
2. Pembangunan model dengan menggunakan Algoritma C4.5 dapat diimplementasikan pada penelitian ini dengan menggunakan data siswa tahun masuk 2020/2021 dan menggunakan 5 atribut yang berupa nilai rata-rata akhir rapor SMP, dan rekomendasi guru BK. Namun model yang dihasilkan memiliki kekurangan yakni atribut data yang kurang komplek sehingga nilai akurasi yang didapatkan belum optimal.
3. Penerapan *data mining* dengan menggunakan metode Algoritma C4.5 ini dapat mempercepat pengambilan keputusan dalam pengklasifikasian jurusan siswa saat proses masuk.
4. Berdasarkan pola yang dihasilkan terdapat beberapa faktor yang mempengaruhi pemrosesan data pada diantaranya: faktor dalam pemilihan atribut dan banyaknya jumlah atribut yang akan digunakan sebagai *data training*, faktor banyaknya data yang diujikan terkait *data training* dan *data testing*, faktor spesifikasi *tools* data mining yang digunakan sebagai pengujian data dan faktor kompleksitas data

## **B. Saran**

1. Dalam penelitian selanjutnya diharapkan peneliti mencoba lebih banyak *item set data*, parameter dan atribut yang akan digunakan dalam pemrosesan data sehingga dapat menemukan pola yang lebih baik dan lebih akurat.
2. Diperlukan tingkat ketelitian yang tinggi dan pembersihan data yang sempurna sehingga tidak terjadi *Error* dalam pemrosesan data.
3. Mencoba menggunakan aplikasi selain RapidMiner dalam analisis data dan mencoba menggunakan metode lain selain C4.5

## DAFTAR PUSTAKA

- Obbie Kristanto. 2016. "Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang." *American journal of public health and the nation's health* 44(6): 760–65.
- Allolayuk, Yuriani Rinni. 2008. "Faktor Yang Mempengaruhi Siswa Memilih Jurusan IPA Pada Siswa Kelas XI SMA Negeri 72 Jakarta." : 166.
- Altanova, Reza, Abdul Syukur, dan M. Arief Soeleman. 2017. "Penentuan Jurusan Siswa Sekolah Menengah Atas Disesuaikan Dengan Minat Siswa Menggunakan Algoritma Fuzzy C-Means." *Jurnal Teknologi Informasi* 13(1): 57–68.
- Informasi, Jurnal Teknologi, dan Komunikasi Stmik Subang. 2016. "PENCARIAM JURUSAN SUBANG DENGAN ALGORITMA C 4.5 DAN DATA MINING STMIK SUBANG Timbo Faritcan Parlaungan Siallagan." 3: 1–10.
- Lisdawati, Henny Indrawati, dan Hendripides. 2016. "the Influence of Internal and External Factors on Learning Difficulties in Understanding the Material on the Economic Subjects of Class X Students in Sman 2 Siak Kabupaten Siak." : 1–12.
- Mardi, Yuli. 2017. "Data Mining : Klasifikasi Menggunakan Algoritma C4.5." *Jurnal Edik Informatika* 2(2): 213–19.
- Novianti, Beti, Tedy Rismawan, dan Syamsul Bahri. 2016. "Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)." *Jurnal Coding, Sistem Komputer Untan* 04(3): 75–84.
- Saleh, Alfa. 2015. "Penerapan Data Mining Dalam Menentukan Jurusan Siswa." *Seminar Nasional Informatika*: 351–55.
- Sari, Ika Purnama, dan Rika Harman. 2020. "Decission Tree Technique Dalam Menentukan Penjurusan Siswa Menengah Kejuruan." *Journal of Information System Research (JOSH)* 1(4): 296–304.
- Swastina, Liliana. 2013. "Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa." *Gema Aktualita* 2(1): 93–98.
- Terrae, Scientia Sinica. 2017. "郭正堂 1,2,3\* 1." 10: 421–37.