

TESIS

**MENENTUKAN JURUSAN PERGURUAN TINGGI UNTUK PESERTA
DIDIK SMK BERDASAR NILAI RAPORT SERTA MINAT DAN BAKAT
DENGAN METODE *DECISION TREE* ALGORITMA C4.5
(STUDI KASUS SMK NEGERI 1 DONOROJO PACITAN)**



Disusun oleh:

Nama : Arif Baktiar
NIM : 20.52.1428
Konsentrasi : Digital Transformation Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

TESIS

**MENENTUKAN JURUSAN PERGURUAN TINGGI UNTUK PESERTA
DIDIK SMK BERDASAR NILAI RAPORT SERTA MINAT DAN BAKAT
DENGAN METODE *DECISION TREE* ALGORITMA C4.5
(STUDI KASUS SMK NEGERI 1 DONOROJO PACITAN)**

***DETERMINING THE DEPARTMENT OF HIGHER EDUCATION FOR
VOCATIONAL SCHOOL STUDENTS BASED ON RAPORT VALUE AND
INTEREST AND TALENT USING THE DECISION TREE ALGORITHM
C4.5 METHOD
(CASE STUDY OF SMK NEGERI 1 DONOROJO PACITAN)***

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama : Arif Baktiar
NIM : 20.52.1428
Konsentrasi : Digital Transformation Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PENGESAHAN

**MENENTUKAN JURUSAN PERGURUAN TINGGI UNTUK PESERTA
DIDIK SMK BERDASAR NILAI RAPORT SERTA MINAT DAN BAKAT
DENGAN METODE DECISION TREE ALGORITMA C4.5
(STUDI KASUS SMK NEGERI 1 DONOROJO PACITAN)**

***DETERMINING THE DEPARTMENT OF HIGHER EDUCATION FOR
VOCATIONAL SCHOOL STUDENTS BASED ON RAPORT VALUE AND
INTEREST AND TALENT USING THE DECISION TREE ALGORITHM
C4.5 METHOD
(CASE STUDY OF SMK NEGERI 1 DONOROJO PACITAN)***

Dipersiapkan dan Disusun oleh :

**Arif Baktiar
20.52.1428**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Magister Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Sabtu, 2 September 2023

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 2 September 2023
Rektor Universitas AMIKOM Yogyakarta

Prof. Dr. M. Suvanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

**MENENTUKAN JURUSAN PERGURUAN TINGGI UNTUK PESERTA DIDIK
SMK BERDASAR NILAI RAPORT SERTA MINAT DAN BAKAT DENGAN
METODE DECISSION TREE ALGORITMA C4.5
(STUDI KASUS SMK NEGERI 1 DONOROJO PACITAN)**

***DETERMINING DEPARTMENT OF HIGHER EDUCATION FOR VOCATIONAL
SCHOOL STUDENTS BASED ON RAPORT VALUE AND INTEREST AND
TALENT USING THE DECISSION TREE ALGORITHM C4.5 METHOD
(CASE STUDY SMK NEGERI 1 DONOROJO PACITAN)***

Dipersiapkan dan Disusun oleh

Arif Baktiar
20.52.1428

Telah Dijujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Sabtu, 2 September 2023

Pembimbing Utama

Prof. Dr. Bambang Soedijono W. A.
NIK. 555126

**Anggota Tim Penguji
Penguji 1**

Dhani Ariatmanto, M.Kom., Ph.D.
NIK. 190302197
Penguji 2

Pembimbing Pendamping

Alva Hendi Muhammad, ST., M.Eng., Ph.D.
NIK. 0518078401

Dr. Wing Wahyu Winarno, MAFIS., Ak.
NIK. 555195
Penguji 3

Prof. Dr. Bambang Soedijono W. A.
NIK. 555126

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 2 September 2023
Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Arif Baktiar
NIM : 20.52.1428
Konsentrasi : Digital Transformation Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Menentukan Jurusan Perguruan Tinggi Untuk Peserta Didik SMK Berdasar Nilai Raport Serta Minat dan Bakat dengan Metode Decision Tree Algoritma C4.5 (Studi Kasus SMK Negeri 1 Donorojo Pacitan)

Dosen Pembimbing Utama : Prof. Dr. Bambang Soedijono W.A.
Dosen Pembimbing Pendamping : Alva Hendi Muhammad, ST.,M.Eng.,Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 2 September 2023
Yang Menyatakan,



Arif Baktiar

HALAMAN PERSEMBAHAN

Puji Syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah, sehingga penulis masih diberikan kesempatan untuk menyelesaikan Tesis ini, sebagai salah satu syarat untuk memperoleh gelar Magister Komputer. Walaupun jauh dari kata sempurna, namun penulis bangga telah mencapai pada titik ini, yang akhirnya skripsi ini bisa selesai diwaktu yang tepat.

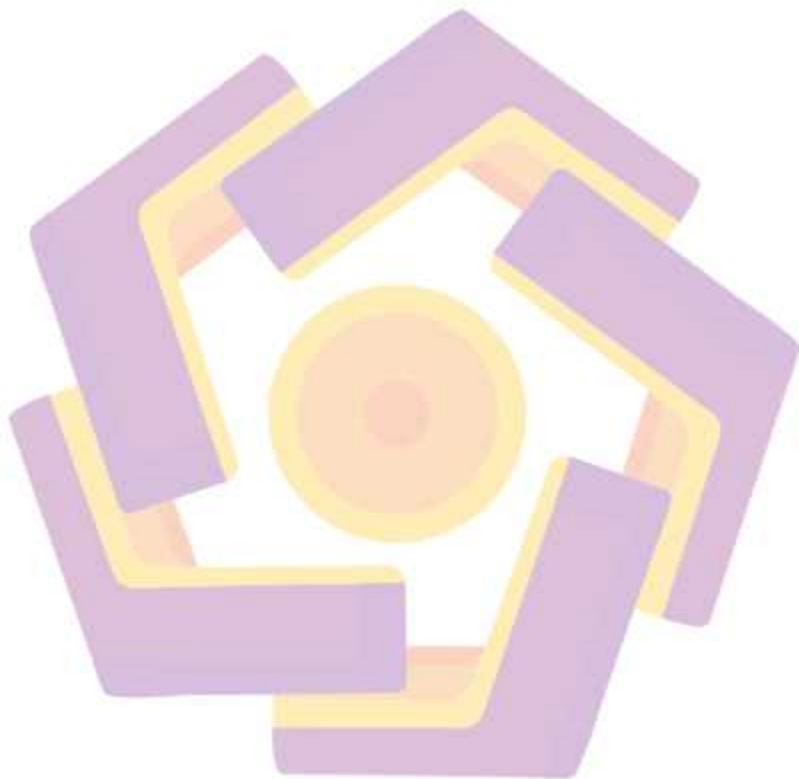
Tesis ini saya persembahkan untuk :

1. Istri dan anak-anak tercinta atas do'a, semangat dan motivasinya.
2. Orang tua dan mertua serta segenap keluarga besar atas semua dukunganya.
3. Dosen Pembimbing Bapak Prof. Dr. Bambang Soedijono W.A. dan Bapak Alva Hendi Muhammad, ST.,M.Eng.,Ph.D untuk bimbingan, masukan dan saran selama proses penyusunan Tesis.
4. Dosen Penguji Bapak Dhani Ariatmanto, M.Kom.,Ph.D. dan Bapak Dr. Wing Wahyu Winarno, MAFIS.,Ak. untuk bimbingan, masukan dan saran selama proses revisi ujian Tesis.

HALAMAN MOTTO

Pendidikan adalah senjata paling ampuh untuk mengubah dunia.

~ Nelson Mandela ~



KATA PENGANTAR

Puji syukur Alhamdulillah penulis panjatkan ke hadirat Allah SWT karena limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan hasil penelitian Tesis dengan judul “Menentukan Jurusan Perguruan Tinggi Untuk Peserta Didik SMK Berdasar Nilai Raport Serta Minat dan Bakat dengan Metode *Decision Tree* Algoritma C4.5 (Studi Kasus SMK Negeri 1 Donorojo Pacitan)”

Laporan ini disusun guna memenuhi salah satu syarat memperoleh gelar Magister Teknik Informatika Universitas AMIKOM Yogyakarta. Penulis menyadari bahwa terselesaikannya laporan ini tidak lepas dari peran serta pihak lain. Oleh karena itu pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. H. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom. selaku Direktur Pasca Sarjana Universitas AMIKOM Yogyakarta.
3. Bapak Prof. Dr. Bambang Soedijono W.A. dan Bapak Alva Hendi Muhammad, ST.,M.Eng.,Ph.D. selaku dosen pembimbing.
4. Bapak Dhani Ariatmanto, M.Kom.,Ph.D. dan Bapak Dr. Wing Wahyu Winarno, MAFIS.,Ak. selaku dosen Penguji.
5. Bapak Drs. Subagyo, M.M. selaku Kepala SMK Negeri 1 Donorojo Pacitan.
6. Semua pihak yang tidak dapat saya sebutkan satu-persatu yang telah membantu dalam penyusunan laporan Tesis ini.

Penulis menyadari bahwa laporan penelitian ini masih jauh dari sempurna, oleh karena itu saran dan kritik yang bersifat membangun sangat kami harapkan demi kesempurnaan penelitian ini.

Yogyakarta, 2 September 2023

Penulis



DAFTAR ISI

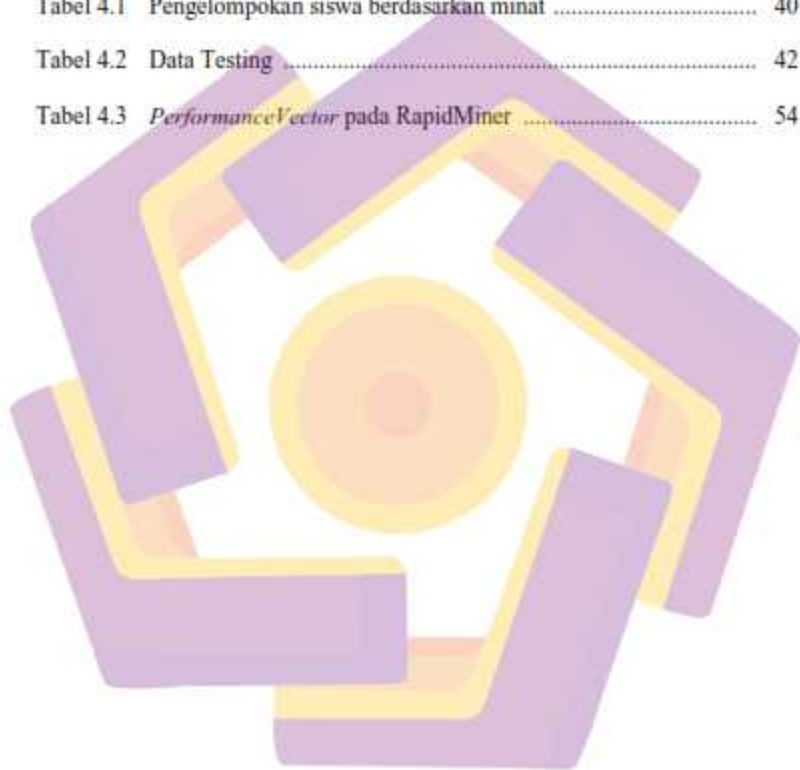
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	6
1.3. Batasan Masalah.....	7
1.4. Tujuan Penelitian.....	7
1.5. Manfaat Penelitian.....	8
BAB II TINJAUAN PUSTAKA.....	9
2.1. Tinjauan Pustaka.....	9
2.2. Keaslian Penelitian.....	14

2.3.	Landasan Teori.....	21
2.3.1	Peserta Didik	21
2.3.2	Sekolah Menengah Kejuruan (SMK).....	22
2.3.3	Penilaian Peserta Didik SMK.....	22
2.3.4	Nilai Raport Sekolah Menengah Kejuruan (SMK)	24
2.3.5	Minat Peserta Didik.....	24
2.3.6	Bakat Peserta Didik.....	25
2.3.7	<i>Decision Tree</i>	25
2.3.8	Algoritma	27
2.3.9	Algoritma C4.5.....	28
2.3.10	Data Mining.....	29
2.3.11	Pengelompokan Data Mining.....	29
2.3.12	Rapidminer.....	30
BAB III	METODE PENELITIAN	33
3.1	Jenis, Sifat Dan Pendekatan Penelitian.....	33
3.2	Metode Pengumpulan Data.....	36
3.3	Metode Analisis Data.....	37
3.4	Alur Penelitian	39
BAB IV	HASIL PENELITIAN DAN PEMBAHASAN	44
2.1.	Mengembangkan dan Memahami Domain Aplikasi.....	44
2.2.	Membuat Kumpulan Data Target.....	45

2.3. Pembersihan Dan Pemrosesan Awal Data	45
2.4. Reduksi Dan Proyeksi Data	45
2.5. Memilih <i>Software</i> Untuk Pengujian.....	48
2.6. Pengujian Data Dengan RapidMiner	48
2.6.1. Import Data	48
2.6.2. Pengujian Data	49
2.6.3. Validasi dengan <i>Cross Validation</i>	56
BAB V PENUTUP	60
5.1. Kesimpulan	60
5.2. Saran.....	61
Daftar Pustaka	62

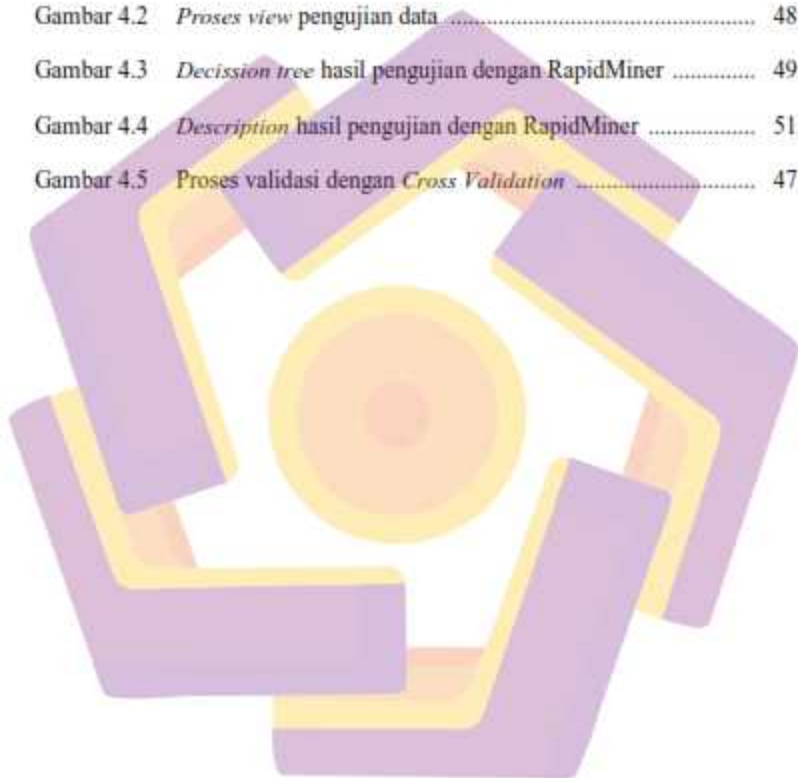
DAFTAR TABEL

Tabel 1.1	<i>Matriks Literatur Review</i> dan Posisi Penelitian	14
Tabel 3.2	Data Nilai Raport Dan Minat Siswa	38
Tabel 4.1	Pengelompokan siswa berdasarkan minat	40
Tabel 4.2	Data Testing	42
Tabel 4.3	<i>PerformanceVector</i> pada RapidMiner	54



DAFTAR GAMBAR

Gambar 3.1	Alur Penelitian	37
Gambar 4.1	Hasil Import Data Testing ke RapidMiner	47
Gambar 4.2	<i>Proses view</i> pengujian data	48
Gambar 4.3	<i>Decision tree</i> hasil pengujian dengan RapidMiner	49
Gambar 4.4	<i>Description</i> hasil pengujian dengan RapidMiner	51
Gambar 4.5	Proses validasi dengan <i>Cross Validation</i>	47



INTISARI

Penelitian ini berjudul Menentukan Jurusan Perguruan Tinggi untuk Peserta Didik SMK Berdasar Nilai Raport Serta Minat dan Bakat dengan Metode *Decision Tree* Algoritma C4.5 (Studi Kasus SMK Negeri 1 Donorojo Pacitan). Tujuan dari penelitian ini yaitu membuat metode yang baik untuk membantu dalam pengambilan keputusan yang tepat dalam pemilihan jurusan Perguruan Tinggi bagi peserta didik kelas XII SMK Negeri 1 Donorojo, dan juga membantu instansi khususnya SMK Negeri 1 Donorojo serta SMK yang lain agar dapat membuat keputusan untuk menentukan potensi yang dimiliki oleh setiap peserta didik. Batasan masalahnya yaitu untuk penentuan jurusan kuliah untuk peserta didik kelas XII di SMK Negeri 1 Donorojo dengan berdasar pada nilai rapor serta minat dan bakat, dan solusi yang ingin dihasilkan adalah adanya metode yang dapat memberikan gambaran jurusan kuliah sesuai potensi diri peserta didik sehingga dapat membantu instansi sekolah dalam membimbing peserta didik yang akan melanjutkan ke Perguruan Tinggi. Metode pengumpulan data menggunakan observasi dan kuesioner (kuesioner), sedangkan untuk analisis data menggunakan tahapan Algoritma C4.5.

Hasil penelitian menunjukkan Accuracy sebesar 45.21%, *Class Precision* untuk Bahasa Inggris 33.33%, Matematika 14.29%, Bahasa Indonesia 66.67%, Pendidikan Agama Islam 36.36%, dan Pendidikan Kewarganegaraan 0.00%. sedangkan untuk *Class Recall* Bahasa Inggris 27.27%, Matematika 18.18%, Bahasa Indonesia 70.59%, Pendidikan Agama Islam 30.77%, dan Pendidikan Kewarganegaraan 0.00%.

Kesimpulan dari penelitian ini bahwa minat siswa dalam memilih jurusan di perguruan tinggi tidak selalu berdasarkan nilai mata pelajaran tetapi lebih banyak berdasarkan variabel atau atribut lain yang belum diketahui.

Kata Kunci : *Decision Tree*, Algoritma C4.5, SMKN 1 Donorojo

ABSTRACT

This research is entitled Determining Higher Education Majors for Vocational High School Students Based on Report Card Values and Interests and Talents with the C4.5 Algorithm Decision Tree Method (Case Study of SMK Negeri 1 Donorojo Pacitan). The purpose of this research is to create a good method to assist in making the right decision in selecting tertiary majors for class XII students at SMK Negeri 1 Donorojo, and also to help agencies, especially SMK Negeri 1 Donorojo and other SMK, to be able to make decisions to determine the potential of each student. The limitation of the problem is to determine college majors for class XII students at SMK Negeri 1 Donorojo based on report cards and interests and talents, and the solution to be produced is a method that can provide an overview of college majors according to students' potential so that they can help agencies schools in guiding students who will continue on to higher education. Methods of data collection using observation and questionnaires (questionnaire), while for data analysis using the stages of Algorithm C4.5.

The results showed that accuracy was 45.21%, Class Precision for English was 33.33%, Mathematics was 14.29%, Indonesian was 66.67%, Islamic Religious Education was 36.36%, and Citizenship Education was 0.00% while for Class Recall English 27.27%, Mathematics 18.18%, Indonesian 70.59%, Islamic Religious Education 30.77%, and Citizenship Education 0.00%.

The conclusion from this study is that students' interest in choosing majors in tertiary institutions is not always based on subject scores but more based on other unknown variables or attributes.

Keywords: Decision Tree, Algorithm C4.5, SMKN 1 Donorojo

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pendidikan adalah salah satu kebutuhan dasar dari setiap manusia, tanpa adanya pendidikan maka manusia tidak akan dapat mengembangkan ilmu pengetahuan. Untuk meningkatkan pendidikan maka pemerintah mendirikan sekolah. Sekolah didirikan secara berjenjang dari tingkat dasar sampai menengah. Dimana setiap masyarakat dapat memperoleh pendidikan secara merata.

Penilaian hasil belajar peserta didik SMK (Sekolah Menengah Kejuruan) dapat dikategorikan menjadi 3 (dua) bagian yaitu: penilaian sikap, penilaian pengetahuan dan penilaian keterampilan. Penilaian sikap merupakan kegiatan penilaian untuk memperoleh informasi mengenai perilaku peserta didik, penilaian sikap akan berkaitan dengan butir nilai-nilai karakter yang meliputi: (1) Integritas, (2) Religius, (3) Nasionalis, (4) Mandiri, (5) Gotong Royong. Penilaian Pengetahuan dipergunakan untuk mengukur ketercapaian aspek pengetahuan berdasarkan Taksonomi Bloom, kemampuan tersebut adalah kemampuan mengingat, memahami, menerapkan, menganalisis, mengevaluasi dan mengkreasi, sedangkan teknik penilaian yang digunakan dalam penilaian pengetahuan ini yaitu: tes lisan, tes tertulis dan penugasan. Penilaian keterampilan adalah suatu penilaian untuk mengetahui kemampuan peserta didik dalam mengaplikasikan pengetahuan, teknik penilaian keterampilan meliputi: (1)

penilaian kinerja, (2) penilaian proyek, (3) portofolio. Hasil penilaian peserta didik setiap semester akan di dokumentasikan ke dalam rapor, pelaporan penilaian ini dilakukan setiap semester atau setiap 6 (enam bulan) sekali. SMK dengan masa studi 3 tahun akan menempuh pendidikan selama 6 semester termasuk juga Praktek Kerja Lapangan selama 6 bulan.

SMK merupakan satuan Pendidikan dengan konsentrasi kejuruan yang memiliki standar atau tolok ukur keberhasilan tertentu, SMK juga merupakan satuan Pendidikan yang memiliki tujuan tidak jauh dari pendidikan vokasi. Suatu tingkat keberhasilan pendidikan di SMK diidentifikasi dari beberapa faktor salah satunya adalah tingkat kerteserapan tamatan di dunia industri maupun Perguruan Tinggi. Menurut data dari bidang Bursa Kerja Khusus (BKK) di SMKN 1 Donorojo tahun 2021 dapat dijabarkan sebagai berikut: 50% tamatan bekerja, 25% tamatan melanjutkan ke Perguruan Tinggi, 5% tamatan berwirausaha dan 20% tamatan tidak diketahui.

Menurut data survei dari bagian Kesiswaan dan Bimbingan Konseling SMK Negeri 1 Donorojo yang dilakukan pada peserta didik kelas XII yang ingin melanjutkan ke Perguruan Tinggi permasalahan yang sering terjadi adalah peserta didik belum bisa mengetahui bakat dan minatnya sehingga hal tersebut berpengaruh pada pemilihan program studi ketika akan mendaftar di Perguruan Tinggi. Meskipun pada awal masuk di SMK telah memilih program keahlian, akan tetapi menurut data Penerimaan Peserta Didik Baru (PPDB) 20% diantara calon peserta didik yang mendaftarkan di SMK Negeri 1 Donorojo diterima pada program keahlian pilihan ke-2 maupun ke-3, menurut data tersebut juga dapat

dilihat bahwa hampir 40% calon peserta didik tidak mengetahui latar belakang mereka dalam menetapkan pilihan program keahlian di SMK Negeri 1 Donorojo, sehingga hal tersebut juga akan berpengaruh kepada pemilihan program studi ketika nanti ingin melanjutkan ke Perguruan Tinggi, oleh karena hal tersebut maka diperlukan suatu metode untuk memberikan gambaran kepada peserta didik SMK Negeri 1 Donorojo tentang jurusan kuliah yang sesuai, metode tersebut mengambil data dengan berdasar pada nilai raport semester 1-5 serta uji minat dan bakat.

Menurut Kemendikbud (2016), dilakukannya tes bakat dan minat agar seseorang dapat meningkatkan kelebihan dan mengatasi kelemahan yang dimiliki, sehingga masing-masing individu dapat memperoleh manfaatnya. Tes minat dan bakat dapat digunakan dalam mengklasterisasi seseorang dalam menentukan pilihan penjurusan Perguruan Tinggi yang sesuai dengan kompetensi dan keterampilan yang dimiliki.

Dalam Kemendikbud (2016), tujuan dari dilakukannya tes minat dan bakat adalah 1) mengetahui dan memahami apa bakat dan minat Individu dan menemukan cara untuk mengembangkan kekuatan tersebut secara optimal. 2) Individu dapat melakukan pemetaan lebih jauh mengenai bakat yang dimilikinya dan minat yang diinginkannya. 3) Individu dapat mengambil keputusan untuk memilih bakat dan minatnya dengan tepat sesuai dengan kemampuan dan keinginannya. Keputusan tersebut memungkinkan memberikan hasil yang bervariasi untuk setiap individu bahkan ketika mereka memiliki kesempatan yang sama untuk mendapatkan hasil tertentu, karena setiap individu dapat

mempersepsikan nilai pentingnya suatu hasil tes secara berbeda. 4) Tujuan dalam kerangka institusional yaitu, nilai dari keputusan ditentukan oleh institusi dan berlaku sama untuk setiap keputusan tertentu. Contoh dari keputusan institusional adalah penerimaan siswa untuk masuk sekolah pada bidang keahlian/kejuruan yang diminati atau juga bias digunakan untuk seleksi perusahaan. 5) Pengelompokan bakat dan minat peserta didik dapat dilakukan untuk mempermudah pengembangan bakat dan minat masing-masing peserta didik SMK.

Proses penjurusan yang dilakukan melalui minat dan bakat bertujuan untuk menempatkan peserta didik sesuai dengan bidang atau keahlian yang dimiliki, sehingga kedepannya peserta didik dapat memiliki motivasi dan kenyamanan dalam setiap prosesnya. Permasalahan yang sering dijumpai di bidang Pendidikan setelah peserta didik menyelesaikan pendidikannya dan melanjutkan ke Perguruan Tinggi adalah seringkali mendapatkan jurusan yang tidak sesuai dengan minat dan bakatnya, atau bisa juga jurusan kuliah yang dilakukan sama sekali tidak ada ilmu pengetahuan yang dimilikinya selama bersekolah. Dan yang belum banyak disadari menekuni bidang keahlian selama bersekolah bisa menjadi modal peserta didik nantinya di dunia kerja.

Relevansi minat dan bakat dalam penjurusan diartikan sebagai kesesuaian modal yang dipelajari oleh peserta didik dengan tuntutan di perkuliahan, hal tersebut dapat ditinjau dari aspek kualitasnya. Tidak hanya itu penentuan minat dan bakat dinilai penting bagi peserta didik yang memilih untuk melanjutkan Pendidikan ke Perguruan Tinggi. Secara psikologi menurut (J.P Chapling et.al

2019) menjelaskan bahwa minat merupakan sikap yang berlangsung dan terus menerus dalam membuat pola perhatian seseorang sehingga membuat dirinya selektif terhadap minatnya. Sedangkan bakat diartikan sebagai kemampuan bawaan atau potensi yang masih perlu dikembangkan dan dilatih agar lebih maksimal (Kemendikbud, 2016).

Tes minat dan bakat tidak hanya berupa kalkulasi dari sebuah nilai, tetapi juga bentuk cerminan gambaran pola pikir dari peserta didik (Asmara & Haryanto, 2015). Sehingga dapat diartikan bahwa tes minat dan bakat dalam mengambil suatu keputusan dinilai rumit, oleh karena itu dibutuhkan metode dalam memecahkan masalahnya untuk melihat dan membedakan bentuk polanya.

Pada penelitian ini bertujuan untuk menentukan penjurusan peserta didik berdasarkan data dari hasil tes minat dan bakat yang sudah dilakukan dengan menggunakan metode yang sudah dipilih. Pelaksanaan tes minat dan bakat membentuk data personal, data dari pelaksanaan tes tersebut akan terus bertambah seiring berjalannya waktu. Dari kumpulan data tersebut dapat diolah informasi baru melalui data mining. Istilah data mining merupakan serangkaian proses untuk mendapatkan informasi berguna dari gudang basis data besar (Prasetyo, 2012).

Proses mendapatkan informasi dari kumpulan data berskala besar juga diartikan sebagai penerapan teknik data mining, singkatnya teknik data mining tersebut diartikan sebagai pencarian informasi yang belum diketahui sebelumnya dari sekumpulan data yang berskala besar (Selvia, Wendi, 2014). Pada penelitian data mining yang digunakan adalah hasil nilai raport peserta didik SMKN 1

Donorojo semester 1 - 6. Nilai raport tersebut juga dijadikan hasil nilai tes minat dan bakat peserta didik dalam menentukan penjurusannya di Perguruan Tinggi.

Metode *decision tree* dibangun menggunakan data dengan variable input, nilai raport, jenis kelamin, dan usia peserta didik dengan label penjurusan. Selanjutnya konsep algoritma C4.5 menjadi salah satu teknik dalam pembentukan *decision tree*. Adapun dalam prosedurnya pemilihan atribut sebagai cabang pada algoritma C4.5 menggunakan hasil perhitungan dari gain ratio. Berdasarkan konsep perhitungan yang dirancang, maka algoritma C4.5 digunakan dalam membangun *decision tree* dalam memberikan rekomendasi penjurusan Perguruan Tinggi bagi peserta didik di tingkat SMK. Penelitian ini akan menganalisis secara kualitatif tentang implementasi dari metode *decision tree* dalam menentukan penjurusan peserta didik dengan menggunakan algoritma C4.5 dari data mining berupa information gain nilai raport, minat dan bakat peserta didik.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka permasalahan yang akan dibahas/diteliti dalam Tesis ini dirumuskan sebagai berikut:

1. Bagaimana penerapan *Decision Tree Algoritma C4.5* dalam menentukan minat dan bakat siswa dalam memilih jurusan di perguruan tinggi untuk peserta didik kelas XII di SMK Negeri 1 Donorojo?
2. Faktor apa yang mempengaruhi peserta didik dalam memilih jurusan di perguruan tinggi?

1.3. Batasan Masalah

Batasan masalah sebagai berikut:

- a. Peneliti membatasi penelitian hanya pada penentuan jurusan kuliah untuk peserta didik kelas XII di SMK Negeri 1 Donorojo dengan berdasar pada nilai rapor serta minat dan bakat.
- b. Asumsi dasar yang digunakan untuk penelitian adalah keraguan peserta didik SMK di SMK Negeri 1 Donorojo dalam mengenali potensi diri yang sebenarnya.
- c. Solusi yang ingin dihasilkan oleh peneliti adalah adanya metode yang dapat memberikan gambaran jurusan kuliah sesuai potensi diri peserta didik sehingga dapat membantu instansi sekolah dalam membimbing peserta didik yang akan melanjutkan ke Perguruan Tinggi.

1.4. Tujuan Penelitian

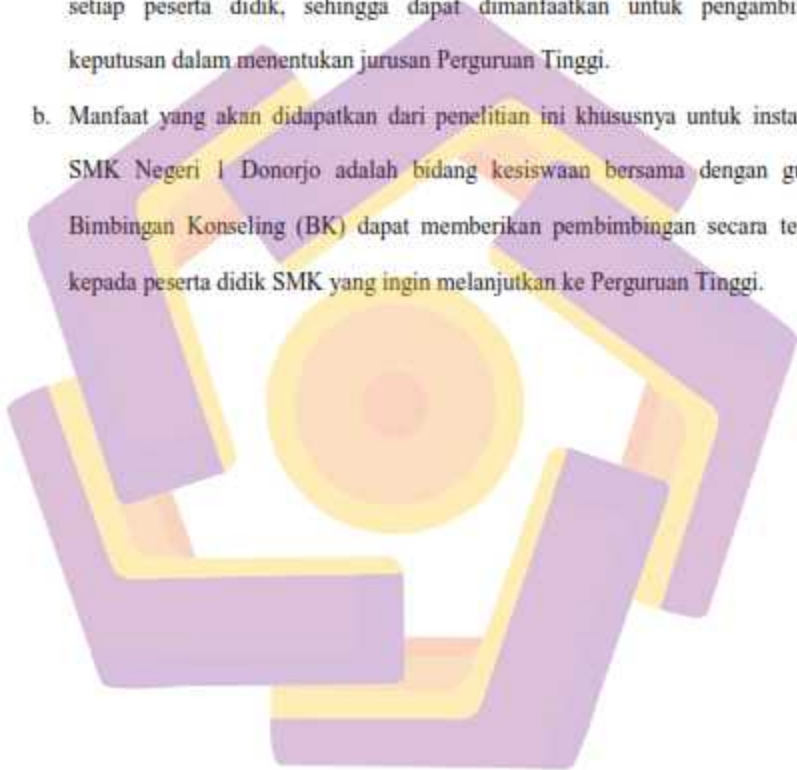
Tujuan penelitian adalah sebagai berikut :

- a. Membuat metode yang baik untuk membantu dalam pengambilan keputusan yang tepat dalam pemilihan jurusan Perguruan Tinggi bagi peserta didik kelas XII SMK Negeri 1 Donorojo.
- b. Membantu instansi khususnya SMK Negeri 1 Donorojo serta SMK yang lain pada umumnya agar dapat membuat keputusan untuk menentukan potensi yang dimiliki oleh setiap peserta didik.

1.5. Manfaat Penelitian

Manfaat penelitian adalah sebagai berikut ini :

- a. Dengan menggunakan metode *Decision Tree Algoritma C4.5* pada data rapor serta uji minat dan bakat dapat memberikan suatu gambaran tentang potensi setiap peserta didik, sehingga dapat dimanfaatkan untuk pengambilan keputusan dalam menentukan jurusan Perguruan Tinggi.
- b. Manfaat yang akan didapatkan dari penelitian ini khususnya untuk instansi SMK Negeri 1 Donorjo adalah bidang kesiswaan bersama dengan guru Bimbingan Konseling (BK) dapat memberikan pembimbingan secara tepat kepada peserta didik SMK yang ingin melanjutkan ke Perguruan Tinggi.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian yang dilakukan oleh Ramadhanu et al. (2021), dijelaskan bahwa tujuan yang ingin dicapai dalam penelitian tersebut adalah bagaimana mengetahui factor dan kondisi mahasiswa bisa memperoleh prestasi atau berprestasi. Selain itu, tujuan dari penelitian ini juga mengungkapkan bagaimana algoritma K-Means melakukan pengelompokan data mahasiswa yang akan berprestasi atau tidak dan bagaimana algoritma C4.5 memprediksi mahasiswa yang telah dikelompokkan. Pengolahan data dalam penelitian ini menggunakan RapidMiner versi 9.7.002. Hasil dari penelitian ini adalah lebih mudah untuk mengelompokkan data dalam bentuk numerik daripada data dalam bentuk polinomial. Hasil lain dalam penelitian ini adalah dari 100 siswa, 27 siswa (27%) diprediksi mahasiswa dan 73 (73%) tidak berprestasi.

Penelitian yang dilakukan oleh Dharshinni et al. (2021), dijelaskan bahwa permasalahan yang dihadapi banyak siswa yang mengambil jurusan adalah karena mengikuti teman atau orang tuanya sehingga menyulitkan siswa untuk mengikuti mata pelajaran yang tersedia sesuai dengan jurusan yang dipilih dan berdampak pada prestasi belajar siswa. Dalam menganalisis penentuan jurusan yang tepat berdasarkan keahlian siswa digunakan algoritma C4.5. Klasifikasi algoritma C45 akan menghasilkan pohon keputusan yang dapat digunakan dalam menentukan arah yang benar. Hasil *Confusion matrix*, Data klasifikasi nilai siswa dalam

menentukan jurusan menghasilkan nilai akurasi sebesar 95,92%, presisi kelas/recall kelas di IPA 97,56%, presisi kelas/recall kelas di IPS 87,50% dan kesalahan klasifikasi adalah 4,08%. Hasil pohon keputusan menunjukkan variabel mata pelajaran yang mempengaruhi pemilihan jurusan siswa adalah matematika, IPA, TIK, keterampilan, dan pariwisata, Nilai tertinggi lagi terletak pada mata pelajaran Pariwisata yang merupakan akar dari pohon keputusan yang terbentuk. Aturan yang dihasilkan adalah nilai matematika di atas 82, nilai IPA minimal 84,5, nilai TIK minimal 85,5, sehingga siswa lebih cocok untuk jurusan IPA. Sedangkan jika nilai matematika kurang dari 82, pariwisata kurang dari 90,5 dan keterampilan kurang dari 84,5 maka mahasiswa tersebut lebih cocok untuk jurusan IPS.

Penelitian yang dilakukan oleh Ramadanty et al. (2021), dijelaskan bahwa tujuan dari penelitian ini adalah mengimplementasikan algoritma C4.5 dan K-Nearest Neighbor (K-NN) dalam klasifikasi peminatan program studi. Algoritma yang digunakan pada penelitian ini yaitu C4.5 dan K-NN. Berdasarkan hasil perhitungan dan pengujian menggunakan algoritma C4.5 dan K-Nearest Neighbor, dapat disimpulkan bahwa algoritma C4.5 dan K-Nearest Neighbor mampu melakukan klasifikasi peminatan program studi dengan sangat baik. Kedua algoritma tersebut menghasilkan nilai akurasi yang tinggi dengan selisih hasil yang tipis, yaitu 100% dari algoritma C4.5 dan 98% dari algoritma K-Nearest Neighbor.

Penelitian yang dilakukan oleh Narulita (2021), dijelaskan bahwa penelitian dilakukan untuk menguji akurasi model prediksi menggunakan metode

data mining classification decision tree dalam menentukan minat peserta didik. Sedangkan untuk hasil penelitian adalah 1) tingkat akurasi dari model prediksi yang dihasilkan sebesar 86,84% dan nilai AUC sebesar 0,752 termasuk dalam kategori *fair classification* (Gorunescu, 2011). 2) Penentuan minat dari peserta didik dengan model prediksi menggunakan data mining metode *classification decision tree* dengan algoritma C4.5 menghasilkan tingkat atau nilai akurasi yang lebih tinggi atau lebih baik, sehingga dapat dijadikan sebagai acuan dalam prediksi penentuan peminatan peserta didik.

Penelitian yang dilakukan oleh Sulihati I (2022), dijelaskan bahwa hasil penelitian pada dataset *campus recruitment* dengan menggunakan metode Decision Tree didapatkan akurasi 82,34%, dengan menggunakan metode Naive Bayes didapatkan akurasi 91,08%. Dari hasil tersebut dapat disimpulkan bahwa kinerja metode algoritma Naive Bayes lebih baik dibanding metode Decision Tree.

Penelitian yang dilakukan oleh Manullang et al. (2021), dijelaskan bahwa peminatan siswa dalam pemilihan jurusan sangat penting dalam menunjang ilmu ke perguruan tinggi dan lebih semangatnya siswa tersebut dalam mengembangkan ilmu di bidang jurusan yang dipilihnya. Metode yang digunakan dalam menentukan jurusan yang akan diambil oleh siswa sesuai dengan latar belakang, minat dan kemampuannya sendiri adalah Algoritma C4.5. Variabel yang digunakan adalah jurusan siswa, hasil tes minat dan bakat siswa. Dengan pengolahan data menggunakan aplikasi RapidMiner didapat bahwa nilai *precision* sebesar 100% dan nilai *recall* sebesar 100%. Dari hasil penelitian yang dilakukan

dalam proses pengujian sebanyak 100 record data testing yang diuji menyatakan bahwa algoritma C4.5 dapat menghasilkan tingkat akurasi sebesar 100,00%. Penerapan data mining dengan menggunakan metode Algoritma C4.5 ini dapat mempercepat pengambilan keputusan dalam memprediksi peminatan jurusan siswa saat proses masuk.

Penelitian yang dilakukan oleh Bahri S (2022), dijelaskan bahwa setiap tahun banyak siswa SMK yang ingin melanjutkan ke perguruan tinggi tetapi tidak mengetahui jurusan apa yang diinginkan dan diterapkan di dunia kerja sesuai dengan bakat dan minatnya, sehingga masih banyak siswa yang mengambil keputusan yang tidak tepat sesuai dengan minat dan bakatnya. Untuk itu diperlukan suatu model untuk mengklasifikasikan permasalahan tersebut. Dalam penelitian ini digunakan tiga algoritma klasifikasi yaitu pohon keputusan, nave Bayes, dan k-nearest neighbor dengan teknik data mining untuk mencari pola dari model yang digunakan, hasil penelitian ini diharapkan dapat membantu mahasiswa dalam menentukan jurusan yang akan diambil. Dari hasil pengujian penelitian ini, faktor yang paling mempengaruhi kesalahan dalam mengambil jurusan di perguruan tinggi adalah variabel berdasarkan jurusan (diri/teman/orang tua), dan dari ketiga algoritma yang digunakan, algoritma pohon keputusan merupakan algoritma terbaik dengan tingkat akurasi tinggi 75,38%.

Penelitian yang dilakukan oleh Bahri S (2022), algoritma C4.5 digunakan untuk menentukan jurusan yang akan diambil oleh peserta didik sesuai dengan latar belakang minat dan kemampuannya sendiri. Parameter pemilihan jurusan yang ditentukan oleh tes buta warna, kesehatan, dan wawancara. Hasil pengujian

dan evaluasi menunjukkan bahwa algoritma decision tree C4.5 akurat ditetapkan untuk penentuan kesesuaian jurusan peserta didik SMK.

Sedangkan dalam penelitian tesis ini akan menggunakan metode *Decision Tree Algoritma C4.5* untuk menentukan jurusan perguruan tinggi untuk peserta didik SMK dimana variabel yang digunakan lebih banyak dari pada penelitian-penelitian sebelumnya yaitu berdasarkan nilai raport, minat, dan bakat siswa dengan studi kasus di SMK Negeri 1 Donorojo Pacitan. Implementasi menggunakan algoritma C4.5 diharapkan dapat mengatasi cabang-cabang pohon keputusan yang tidak relevan menjadi relevan, sehingga dalam mengimplementasikan Bahasa pemrograman dapat lebih mudah diterjemahkan. Algoritma C4.5 sendiri menggunakan pendekatan induksi dimana dalam pendekatan ini, algoritma C4.5. Membagi data berdasarkan kriteria yang dipilih untuk membuat sebuah pohon keputusan. Analisis Kinerja Algoritma C4.5 Pada Sistem Pendukung Keputusan Penentuan Jenis Pelatihan (Oktavia, et al. 2015).

2.2. Keaslian Penelitian

Tabel 1.1. Matriks *literatur review* dan posisi penelitian
Menentukan Jurusan Perguruan Tinggi untuk Peserta Didik SMK Berdasar Nilai Raport Serta Minat dan Bakat dengan Metode Decision Tree Algoritma C4.5
(Studi Kasus SMK Negeri 1 Donorojo Pacitan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1.	<i>Hybrid Data Mining with the Combination of K-Means Algorithm and C4.5 to Predict Student Achievement</i>	Agung Ramadhani dkk, <i>International Journal Of Artificial Intelligence Research</i> , 2021	Untuk mengetahui faktor dan kondisi mahasiswa bisa memperoleh prestasi atau berprestasi. Selain itu, tujuan dari penelitian ini juga mengungkapkan bagaimana algoritma K-Means melakukan pengelompokan data mahasiswa yang akan berprestasi atau tidak dan bagaimana algoritma C4.5 memprediksi mahasiswa yang telah dikelompokkan	Lebih mudah untuk mengelompokkan data dalam bentuk numerik daripada data dalam bentuk polinomial. Hasil lain dalam penelitian ini adalah dari 100 siswa, 27 siswa (27%) diprediksi mahasiswa dan 73 (73%) tidak berprestasi.	<ul style="list-style-type: none"> a. Sebaiknya menggunakan atribut yang lebih banyak agar hasil clustering dan predicting lebih baik. b. Sebaiknya atribut data untuk clustering dengan metode K-Means berupa data numerik kemudian hasilnya dijadikan data polinomial sehingga dapat digunakan untuk prediksi metode C4.5. c. Disarankan jumlah record data latih lebih banyak daripada jumlah record data testing agar prediksi yang dibuat lebih tepat dan akurat. 	Pada jurnal penelitian ini salah satu tujuannya adalah untuk mengetahui faktor dan kondisi mahasiswa bisa memperoleh prestasi, sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	<i>Classification of Major Selection Based on Students Expertise Using C4.5 Algorithm</i>	Dharshinni NP, dkk, Jurnal Infokum (2021)	Untuk mengklasifikasi pemilihan jurusan siswa, karena adanya permasalahan yang dihadapi banyak siswa yang mengambil jurusan karena mengikuti teman atau orang tuanya sehingga menyulitkan siswa untuk mengikuti mata pelajaran yang tersedia sesuai dengan jurusan yang dipilih dan berdampak pada prestasi belajar siswa	Variabel nilai mata kuliah yang berpengaruh dalam menentukan pemilihan jurusan mahasiswa adalah matematika, IPA, TIK, Keterampilan, dan Pariwisata	Data yang digunakan hanya sebatas nilai dari rapor dan digunakan untuk penentuan masuk jurusan ilmu pengetahuan alam atau ilmu pengetahuan sosial.	Dalam jurnal penelitian mengklasifikasi pemilihan jurusan siswa, karena adanya permasalahan yang dihadapi banyak siswa yang mengambil jurusan karena mengikuti teman atau orang tuanya sehingga menyulitkan siswa untuk mengikuti mata pelajaran, sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat
3	Penerapan Algoritma C4.5 dan K-Nearest Neighbor untuk Klasifikasi Peminatan Program Studi di Perguruan Tinggi Berdasarkan Nilai Rapor	Ramadanty MC, dkk, <i>Scientific Student Journal for Information, Technology and Science</i> , (2021)	Mengimplementasikan algoritma C4.5 dan K-Nearest Neighbor (K-NN) dalam klasifikasi peminatan program studi	Algoritma C4.5 dan K-Nearest Neighbor mampu melakukan klasifikasi peminatan program studi dengan sangat baik. Kedua algoritma tersebut menghasilkan nilai	Data yang digunakan hanya sebatas nilai rapor Matematika dan mata pelajaran produktif siswa kelas XII jurusan Teknik Komputer Jaringan (TKJ), Teknik Elektronika Industri (TEI), dan	Dalam jurnal penelitian ini dijelaskan bagaimana mengimplementasikan algoritma C4.5 dan K-Nearest Neighbor (K-NN) dalam klasifikasi peminatan program studi, sedangkan

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				akurasi yang tinggi dengan selisih hasil yang tipis, yaitu 100% dari algoritma C4.5 dan 98% dari algoritma K-Nearest Neighbor.	Rekayasa Perangkat Lunak (RPL) Sekolah Menengah Kejuruan Negeri (SMKN) 1 Karawang	dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat
4	Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4.5 untuk Penentuan Peminatan Peserta Didik	Narulita S. dkk. Jurnal Media Aplikom (2021)	Untuk menguji akurasi model prediksi menggunakan metode <i>data mining classification decision tree</i> dalam menentukan minat peserta didik	Penentuan minat dari peserta didik dengan model prediksi menggunakan data mining metode <i>classification decision tree</i> dengan algoritma C4.5 menghasilkan tingkat atau nilai akurasi yang lebih tinggi atau lebih baik, sehingga dapat dijadikan sebagai acuan dalam prediksi penentuan peminatan peserta didik.	Tidak ada penjelasan lebih lanjut mengenai angkset yang digunakan dalam penelitian ini.	Dalam jurnal penelitian ini dilakukan pengujian akurasi model prediksi menggunakan metode <i>data mining classification decision tree</i> dalam menentukan minat peserta didik, sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi	Sulihati E, <i>Teenoscientia</i> (2022)	penelitian ini dilakukan untuk membantu menyelesaikan permasalahan dikarenakan dalam pendataan telah menciptakan situasi dimana data berlimpah tetapi informasi langka, seperti informasi kepada orang tua tentang hasil penerimaan anak mereka ke perguruan tinggi dengan menggunakan teknik data mining	Kinerja metode algoritma Naïve Bayes lebih baik dibanding metode <i>Decision Tree</i> .	Analisa perbandingan algoritma klasifikasi diatas perlu dilakukan lagi penelitian lanjutan agar klasifikasi data seleksi perguruan tinggi dapat dilakukan secara lebih valid agar akurasi yang dihasilkan lebih tinggi.	Dalam jurnal penelitian Algoritma C4.5 digunakan untuk memberikan informasi kepada orang tua tentang hasil penerimaan anak mereka ke perguruan tinggi, sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat
6	Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5	Manullang N, dkk, <i>Jurnal Ilmu Komputer dan Teknologi</i> (2021)	Untuk mengimplementasikan teknik data mining untuk prediksi peminatan siswa dalam pemilihan	Penerapan data mining dengan menggunakan metode Algoritma C4.5 ini dapat mempercepat	Untuk memaksimalkan atau menambah atribut yang lebih spesifik dan lebih banyak dalam menentukan peminatan	Dalam jurnal penelitian ini teknik data mining untuk prediksi peminatan siswa dalam pemilihan jurusan, sedangkan dalam penelitian

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			jurusan sangat penting dalam menunjang ilmu ke perguruan tinggi dan lebih semangatnya siswa tersebut dalam mengembangkan ilmu di bidang jurusan yang dipilihnya.	pengambilan keputusan dalam memprediksi peminatan jurusan siswa saat proses masuk.	jurusan siswa seperti data siswa baru dan penambahan dataset dalam data training dan data testing serta perlu adanya penelitian lebih lanjut dengan melakukan pengujian dengan metode lain seperti naive bayes, ID3 dan lain sebagainya	adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat
7	Implementasi Data Mining untuk Menentukan Minat Siswa dalam Menentukan Jurusan pada Perguruan Tinggi	Bahe S, Jurnal Sistem Informasi (JUSIN) (2022)	Mengimplementasikan data mining untuk membantu mahasiswa dalam menentukan jurusan yang akan diambil.	Dari hasil pengujian penelitian ini, faktor yang paling mempengaruhi kesalahan dalam mengambil jurusan di perguruan tinggi adalah variabel berdasarkan jurusan (diri/teman/orang tua), dan dari ketiga algoritma yang digunakan, algoritma pohon keputusan	Data yang digunakan dalam penelitian ini adalah kuesioner yang disebar ke mahasiswa yang sudah masuk kuliah bukan kepada calon mahasiswa baru yang akan masuk kuliah.	Dalam jurnal penelitian ini data mining untuk membantu mahasiswa dalam menentukan jurusan yang akan diambil, sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				merupakan algoritma terbaik dengan tingkat akurasi tinggi 75,38%.		
8	Rules Generation untuk Klasifikasi Data Bakat dan Minat Berdasarkan Rumpun Ilmu dengan Decision Tree	Suryadi Syamsu dkk, Inspiration : Jurnal Teknologi Informasi dan Komunikasi, 2019	merancang aplikasi dan mengimplementasikan <i>Rules Generation</i> untuk klasifikasi data bakat dan minat berdasarkan rumpun ilmu dengan decision tree	Implementasi <i>Rules Generation</i> terhadap <i>Decision Tree</i> (RGFDT) penelitian tentang penentuan jurusan seseorang yang menggunakan algoritma C4.5 dapat mengatasi cabang-cabang pohon keputusan yang tidak relevan menjadi relevan, sehingga dalam mengimplementasi ke dalam bahasa pemrograman dapat lebih mudah diterjemahkan, dan aplikasi tes bakat dan minat dapat berjalan dengan baik dan memudahkan	<ul style="list-style-type: none"> a. Pengembangan interface sesuai kebutuhan pengembangan sistem selanjutnya b. Perbaikan pada class diagram dan database sesuai dengan kebutuhan sistem c. Pengembangan pada soal-soal Tes Minat dan Bakat yang sesuai dengan saran dari ahli psikologi d. Perbaikan pada metode penelitian RGFDT yang belum dapat diaplikasi dan diterapkan oleh peneliti 	Dalam jurnal penelitian ini merancang aplikasi dan mengimplementasikan <i>Rules Generation</i> untuk klasifikasi data bakat dan minat berdasarkan rumpun ilmu dengan <i>decision tree</i> , sedangkan dalam penelitian adalah menentukan jurusan perguruan tinggi untuk peserta didik SMK berdasar nilai raport serta minat dan bakat

Tabel 1.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				responden dalam melakukan tes bakat dan minat dengan mudah	c. Pengembangan untuk dibuat dalam sistem Android lebih baik lagi	

2.3. Landasan Teori

2.3.1 Peserta Didik

Pengertian siswa atau peserta didik menurut ketentuan umum undang-undang RI No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional adalah anggota masyarakat yang berusaha mengembangkan potensi diri melalui proses pembelajaran yang tersedia pada jalur, jenjang, dan jenis pendidikan tertentu. Dengan demikian peserta didik adalah orang yang mempunyai pilihan untuk menempuh ilmu sesuai dengan cita-cita dan harapan masa depan. (Septiana & Sholeh, 2021).

Menurut (Oemar Hamalik et.al, 2011) peserta didik sebagai suatu komponen masukan dalam sistem pendidikan, yang selanjutnya diproses dalam proses pendidikan, sehingga menjadi manusia yang berkualitas sesuai dengan tujuan pendidikan Nasional. Menurut (Abu Ahmadi et.al, 2021) peserta didik adalah sosok manusia sebagai individu/pribadi (manusia seutuhnya). Individu diartikan "orang seorang tidak tergantung dari orang lain, dalam arti benar-benar seorang pribadi yang menentukan diri sendiri dan tidak dipaksa dari luar, mempunyai sifat-sifat dan keinginan sendiri".

Sedangkan Hasbullah berpendapat bahwa siswa sebagai peserta didik merupakan salah satu input yang ikut menentukan keberhasilan proses pendidikan. Tanpa adanya peserta didik, sesungguhnya tidak akan terjadi proses pengajaran. Sebabnya ialah karena peserta didiklah yang membutuhkan pengajaran dan bukan guru, guru hanya berusaha memenuhi kebutuhan yang ada pada peserta didik. (Rajawali Pers, 2010).

Berdasarkan pengertian-pengertian di atas, dapat dikatakan bahwa peserta didik adalah orang/individu yang mendapatkan pelayanan pendidikan sesuai dengan bakat, minat, dan kemampuannya agar tumbuh dan berkembang dengan baik serta mempunyai kepuasan dalam menerima pelajaran yang diberikan oleh pendidikannya.

2.3.2 Sekolah Menengah Kejuruan (SMK)

Pendidikan menengah kejuruan adalah pendidikan pada jenjang pendidikan menengah yang mengutamakan pengembangan kemampuan siswa untuk melaksanakan jenis pekerjaan tertentu. Pendidikan menengah kejuruan mengutamakan penyiapan siswa untuk memasuki lapangan kerja serta mengembangkan sikap profesional. Sesuai dengan bentuknya, sekolah menengah kejuruan menyelenggarakan program-program pendidikan yang disesuaikan dengan jenis-jenis lapangan kerja (Peraturan Pemerintah No. 29, 1990).

Sekolah Menengah Kejuruan (SMK) merupakan salah satu bentuk satuan pendidikan formal yang menyelenggarakan pendidikan kejuruan pada jenjang pendidikan menengah sebagai lanjutan dari SMP, MTs, atau bentuk lain yang sederajat. Sekolah di jenjang pendidikan dan jenis kejuruan dapat bernama Sekolah Menengah Kejuruan (SMK) atau Madrasah Aliyah Kejuruan (MAK), atau bentuk lain yang sederajat (Undang-undang RI No. 20, 2003).

2.3.3 Penilaian Peserta Didik Sekolah Menengah Kejuruan (SMK)

Penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik. Proses tersebut dilakukan melalui berbagai teknik penilaian, menggunakan berbagai instrumen, dan berasal

dari berbagai sumber agar lebih komprehensif. Penilaian harus dilakukan secara efektif. Oleh sebab itu, pengumpulan informasi yang akan digunakan untuk mengukur pencapaian hasil belajar peserta didik harus lengkap dan akurat agar dihasilkan keputusan yang tepat.

Pengumpulan informasi pencapaian hasil belajar peserta didik membutuhkan teknik dan instrumen penilaian, serta prosedur analisis sesuai dengan karakteristik penilaian masing-masing. Kurikulum 2013 merupakan kurikulum berbasis kompetensi dengan KD sebagai kompetensi minimal yang harus dicapai oleh peserta didik.

Untuk mengetahui ketercapaian KD, pendidik harus merumuskan sejumlah indikator pencapaian kompetensi (IPK). IPK digunakan sebagai acuan penilaian. Pendidik atau satuan pendidikan (sekolah) juga harus menentukan pencapaian kriteria ketuntasan minimal (KKM).

Penilaian tidak hanya difokuskan pada hasil belajar, tetapi juga pada proses belajar. Peserta didik dilibatkan dalam proses penilaian terhadap dirinya sendiri dan penilaian antar peserta didik (penilaian antar teman) sebagai sarana untuk berlatih melakukan penilaian.

Pembelajaran konsep penilaian ini meningkatkan kompetensi guru dalam melaksanakan pengembangan penilaian berorientasi pada Keterampilan Berpikir Tingkat Tinggi. (Kemendikbud, 2016).

2.3.4 Nilai Raport Sekolah Menengah Kejuruan (SMK)

SMK merupakan sekolah yang sudah memberikan penjurusan langsung untuk peserta didiknya untuk siap di dunia kerja tetapi tidak menutup kemungkinan bagi peserta didik untuk melanjutkan ke perguruan tinggi, dan bagi peserta didik yang akan melanjutkan ke perguruan tinggi mungkin dalam mengambil jurusan tidak atau belum sesuai dengan minat dan bakat, maka dalam penelitian ini digunakan mata pelajaran selain kejuruan yang dimaksudkan untuk menjangking bakat minat peserta didik.

Raport adalah buku yang berisi nilai kepandaian dan prestasi belajar murid di sekolah, berfungsi sebagai laporan resmi guru kepada orangtua wali murid yang wajib menerimanya. Raport itu sendiri merupakan salah satu pertanggung jawaban sekolah terhadap masyarakat tentang kemampuan yang dimiliki siswa yang berupa sekumpulan hasil penilaian. (Pangastuti PAD, 2019)

2.3.5 Minat Peserta Didik

Terdapat banyak pendapat tentang istilah minat dan bakat, Minat adalah suatu pemusatan perhatian yang mengandung unsur-unsur perasaan, kesenangan, kecenderungan hati, keinginan yang tidak disengaja yang sifatnya aktif untuk menerima sesuatu dari luar (lingkungan). (Andi AP, 2019)

. Minat berasal dari dalam diri individu yang memiliki potensi dan perlu dikembangkan atau dilatih untuk mencapai suatu prestasi. Hal tersebut sejalan dengan pendapat Monk dalam (Terttiaavini, et al, 2016), menambahkan bahwa keberbakatan (*Gifedness*) adalah suatu potensi bawaan yang memerlukan binaan guna mencapai prestasi sesuai dengan prestasinya.

2.3.6 Bakat Peserta Didik

Bakat adalah potensi yang sudah ada dalam diri seseorang yang diperlukan pengembangan diri. Bakat masih bersifat potensial, bakat adalah kemampuan seseorang yang sangat membutuhkan usaha dalam mengembangkannya dan latihan dengan sistematis dan berkesinambungan supaya bisa menjadi sebuah prestasi. Oleh karena itu, dapat diartikan bahwa bakat dapat terasah dengan baik dengan latihan yang rutin dan mengeksplorasi bakat sehingga menjadi sebuah prestasi dan dapat mencapai prestasi tersebut dalam bidang tertentu. Pengalaman, pengetahuan, dan latihan yang rutin sangat diperlukan untuk mewujudkan bakat dalam suatu prestasi. Sebagai contoh, seseorang yang mempunyai bakat bernyanyi namun tidak memiliki kesempatan untuk mengembangkannya, berbeda dengan peserta didik yang bakatnya sangat diperhatikan oleh orang tua nya sehingga orang tua sangat mendukung potensi yang ada pada diri sang anak. Orang tua akan berusaha mengembangkan potensi anaknya dengan latihan sehingga anak dapat mengembangkan potensi yang ada dengan optimal. Dukungan orang tua juga dapat memupuk rasa percaya diri pada seorang anak, sehingga anak mampu untuk mengeksplorasi kemampuannya didepan semua orang sampai menghasilkan prestasi yang gemilang. (Basri et al., 2021)

2.3.7 Decision Tree

Decision Tree atau Pohon keputusan adalah algoritma pembelajaran terawasi yang membutuhkan data dengan atribut kelas. Jika data hilang atau proses klasifikasi tidak dapat dilakukan, data harus beragam. Salah satu aspek yang paling menarik dari pohon keputusan adalah bagaimana aturan dibentuk.

Pohon keputusan adalah aturan keputusan khusus yang diatur dalam struktur pohon. Aturan keputusan dapat dibangun dari pohon keputusan hanya dengan melintasi jalur apa pun dari simpul akar ke daun apa pun. Satu set lengkap aturan keputusan dihasilkan. Pohon keputusan membagi ruang dokumen di daun menjadi area yang tidak tumpang tindih, dan prediksi dibuat pada setiap daun.

Algoritma pohon keputusan dibangun di atas dataset. *Divide-and-conquer* digunakan untuk membangun model pohon keputusan dengan menggunakan IG untuk memilih atribut dari kumpulan data gaya pohon. Pada setiap langkah dalam membangun pohon keputusan, satu dari setiap atribut dipilih untuk mengisolasi data.

Nilai atribut digunakan untuk menentukan nilai pembatas berdasarkan atribut yang dipilih. GI dan entropi biasanya digunakan untuk pohon klasifikasi. Untuk menghitung entropi didefinisikan (Shannon, 1948) sebagai berikut:

$$H_e(S) = 1 - \sum_{y \in C} p(y)^2$$

Di mana S adalah dataset, C adalah kelas dan $p(y)$ adalah perbandingan jumlah data terhadap kelas C . entropi akan bernilai 0 apabila hanya ada 1 kelas dan mencapai nilai maksimum ketika semua kelas memiliki kemungkinan yang sama. (Sulihati I, 2022)

Decision tree atau pohon keputusan adalah salah satu metode klasifikasi yang populer dan banyak digunakan secara praktis. Metode ini berusaha menemukan model klasifikasi yang tahan terhadap derau (Sari & Harman, 2020). Metode *decision tree* juga diartikan sebagai sebuah struktur *flowchart* yang mirip dengan struktur pohon, setiap titik pohon adalah atribut yang telah diuji, setiap

cabang adalah hasil uji, dan titik akhir adalah pembagian kelas yang dihasilkan (Sucipto et al., 2017). Pendapat lain juga dikemukakan oleh (Yudha 2015) menjelaskan bahwa *decision tree* merupakan salah satu metode klasifikasi pada *text mining*.

Berikut ini merupakan penjelasan dari kelebihan dan kekurangan dari *decision tree* sebagai berikut (Syamsu dkk, 2019):

1. Kelebihan dari metode *decision tree*:
 - a. Pengambilan keputusan lebih simple dan spesifik.
 - b. Sampel diuji berdasarkan kriteria atau kelas tertentu.
 - c. Keflesibeln dari fiturnya meningkatkan hasil kualitas keputusan.
2. Kekurangan dari metode *decision tree*:
 - a. *Overlap* jika kriterianya terlalu banyak.
 - b. Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan.
 - c. Kesulitan dalam mendesain keputusan yang optimal.
 - d. Hasil kualitas berdasarkan desain pohon dibuat.

Penerapam metode tersebut dapat terhindari dari permasalahan kualitas pengambilan keputusan.

2.3.8 Algoritma

Menurut (Maulana et.al, 2017), Algoritma adalah sistem kerja komputer memiliki *brainware*, *hardware*, dan *software*. Tanpa salah satu dari ketiga sistim tersebut, komputer tidak akan berguna. Kita akan lebih fokus pada *software* komputer. *Software* terbangun atas susunan program) dan *syntax* (cara

penulisan/pembuatan program). Untuk menyusun program atau *syntax*, diperlukannya langkahlangkah yang sistematis dan logis untuk dapat menyelesaikan masalah atau tujuan dalam proses pembuatan suatu *software*.

Maka, algoritma berperan penting dalam penyusunan program atau *syntax* tersebut. Pengertian algoritma adalah susunan yang logis dan sistematis untuk memecahkan suatu masalah atau untuk mencapai tujuan tertentu. Dalam dunia komputer, algoritma sangat berperan penting dalam pembangunan suatu *software*. Dalam dunia sehari-hari, mungkin tanpa kita sadari algoritma telah masuk dalam kehidupan kita.

2.3.9 Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3. Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Algoritma C4.5 juga menggunakan kriteria *gain* dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi (Suherman et al., 2021).

Menurut (Sukma et.al, 2019), Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross quinlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja, yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*).

Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

2.3.10 Data Mining

Data mining adalah proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan didalam suatu basis data, data *warehouse*, atau tempat penyimpanan data lainnya. Kunci dari data mining meliputi data, informasi dan keputusan bisnis. Tujuan akhir dari data mining yaitu menggunakan informasi untuk membuat lebih efisien keputusan bisnis dan mengambil keputusan secara tepat (Hardiani, dkk, 2014). Data mining disebut juga sebagai *Knowledge Discovery In Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso et.al, 2016).

Menurut (Kusrini dan Luthfi et.al 2009) hal-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 untuk mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.3.11 Pengelompokan Data Mining

Menurut (Han dan Kamber et.al, 2006) data mining dikelompokkan berdasarkan fungsinya, yaitu:

1. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai variable target sebagai nilai prediksi

3. Prediksi

Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

4. Klasifikasi

Dalam klasifikasi terdapat target variable kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.

5. Klustering

Merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.3.12 Rapidminer

Menurut (CTI et.al, 2017), RapidMiner merupakan *software*/perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data

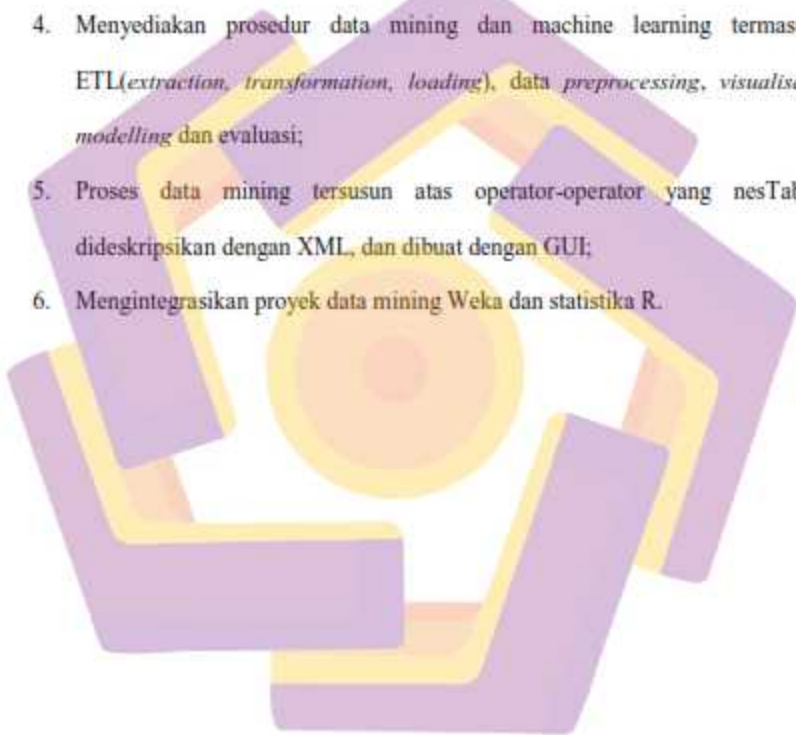
mining, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. RapidMiner memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan RapidMiner pun dapat ditampilkan secara visual dengan grafik. Menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining.

Menurut Aprilla et.al, 2013), RapidMiner memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan diberbagai sistem operasi;
2. Proses penemuan pengetahuan dimodelkan sebagai operator trees;
3. Representasi XML internal untuk memastikan format standar pertukaran data;
4. Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen;
5. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data;
6. Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Menurut (Aprilla et.al, 2013), beberapa fitur RapidMiner antara lain:

1. Banyaknya algoritma data mining, seperti *decision tree* dan *selforganization map*;
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots;
3. Banyaknya variasi plugin, seperti *text plugin* untuk melakukan analisis teks;
4. Menyediakan prosedur data mining dan machine learning termasuk: ETL(*extraction, transformation, loading*), data *preprocessing*, *visualisasi*, *modelling* dan evaluasi;
5. Proses data mining tersusun atas operator-operator yang nesTabel, dideskripsikan dengan XML, dan dibuat dengan GUI;
6. Mengintegrasikan proyek data mining Weka dan statistika R.



BAB III

METODE PENELITIAN

3.1 Jenis, Sifat Dan Pendekatan Penelitian

Penelitian ini menggunakan teknik klasifikasi yang terbagi menjadi pohon keputusan (*Decision Tree*). Metode klasifikasi yang digunakan untuk memproses permodelan dalam menggambarkan dan membedakan kelas data atau konsep label kelas yang tidak diketahui. *Decision Tree* merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Tujuan klasifikasi ini juga untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai. *Decision Tree* digunakan untuk mengubah fakta yang sangat besar untuk mempresentasikan aturan (Wulandari et al., 2021).

Pada penelitian ini dilakukan analisis data menggunakan metode *Decision Tree Algoritma C4.5*. Data yang akan digunakan dan dianalisa, yaitu data peserta didik kelas XII SMK Negeri 1 Donorojo berdasar nilai rapor serta minat dan bakat.

Algoritma C4.5 diperkenalkan oleh Quinlan sebagai versi perbaikan dari ID3. Dalam ID3, induksi *decision tree* hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Algoritma C4.5 juga

menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi. (Ramadanty et al., 2021)

Adapun tahapan algoritma C4.5 sebagai berikut (Suherman et al., 2021) :

1. *Developing and understanding the application domain*

Tujuan yang ingin dicapai dari proses ini yaitu pengelompokan siswa berdasarkan minat, dimana untuk mengetahui minat peserta didik dengan cara memberikan kuesioner yang disebar dengan *Google Form*.

2. *Creating a target data set*

Variabel (atribut) yang digunakan yaitu nama, nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris, dan *data point* (*sample*) yang akan digunakan yaitu minat.

Walaupun SMK merupakan sekolah yang sudah memberikan penjurusan langsung untuk peserta didiknya untuk siap di dunia kerja tetapi tidak menutup kemungkinan bagi peserta didik untuk melanjutkan ke perguruan tinggi, dan bagi peserta didik yang akan melanjutkan ke perguruan tinggi mungkin dalam mengambil jurusan tidak atau belum sesuai dengan minat dan bakat, maka dalam penelitian ini digunakan mata pelajaran selain kejuruan yang dimaksudkan untuk menjangkau bakat minat peserta didik.

3. *Data cleaning and preprocessing*

Pada tahap ini dilakukan proses penghapusan data apabila ada data yang tidak digunakan. Dimana jika ada data yang tidak digunakan dan akan dihapus

maka bisa dihapus sebelum data diolah di *RapidMiner*, atau dihapus di dalam tabel *Microsoft Excel*.

4. *Data reduction and projection*

Karena tidak ada data yang dihapus pada proses *Data cleaning and preprocessing* karena atribut yang ada pada tabel sudah sesuai dengan data yang akan diolah, maka atribut yang digunakan yaitu nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris. Kelima atribut tersebut merupakan kriteria numerik/kontinyu.

5. *Choosing the data mining task*

Metode data mining yang digunakan yaitu pohon klasifikasi, karena sesuai dengan tujuan yang ingin dicapai.

6. *Choosing the data mining algorithm*

Algoritma yang digunakan yaitu algoritma C4.5. Kelebihan utama dari Algoritma C4.5 yaitu dapat membuat pohon keputusan (*Decision Tree*) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima (Muhamad et al., 2019).

7. *Data mining*

Pengujian menggunakan *RapidMiner*, dimana *RapidMiner* adalah software yang menggunakan prinsip dan algoritme data mining, *RapidMiner* mengekstrak pola dari big data, menggabungkan metode statistik, kecerdasan buatan, dan database. *RapidMiner* memudahkan penggunaanya untuk

menghitung data dalam jumlah besar menggunakan operator. Operator ini digunakan untuk mengubah data. Data tersebut digabungkan dengan node di operator, dan kemudian menghubungkannya ke node hasil untuk melihat hasilnya (Ainurrohmah, 2021).

Data yang diolah dalam RapidMiner yaitu data nilai rapor peserta didik kelas XII nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris serta minat dan bakat yang diperoleh dari kuesioner yang disebar lewat *Google Form*.

Hasil dari pengolahan datanya yaitu pohon keputusan yang menggambarkan pengaruh nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris terhadap pemilihan jurusan kuliah di perguruan tinggi.

3.2 Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini adalah :

1. Observasi (Pengumpulan Data)

Observasi merupakan metode pengumpulan data yang kompleks dikarenakan melibatkan beberapa faktor dalam pelaksanaannya. Dalam penelitian ini data yang dikumpulkan adalah nilai raport siswa SMKN 1 Donorojo Pacitan yang meliputi nilai PAI, nilai PKn, nilai Bahasa Indonesia, nilai Matematika, nilai Bahasa Inggris, dengan populasi kelas XII sebanyak 234 siswa, dan sampel sebanyak 73 siswa sesuai dengan jumlah peserta didik yang mengisi kuesioner yang telah disebar melalui *Google Form*.

2. Angket (*Kuesioner*)

Kuesioner merupakan metode pengumpulan data yang dilakukan dengan cara memberi serangkaian pertanyaan-pertanyaan yang ditujukan kepada responden untuk dijawab. Metode ini merupakan salah satu metode yang efisien karena peneliti telah mengetahui variabel-variabel apa yang akan diukur dan tahu apa yang diharapkan dari responden. Dalam penelitian kuesioner dibuat atau disebarakan menggunakan *Google Form*. (Blanko kuesioner terlampir)

3.3 Metode Analisis Data

1. *Developing and understanding the application domain*

Tujuan yang ingin dicapai dari proses ini yaitu pengelompokan siswa berdasarkan minat.

2. *Creating a target data set*

Tahap ini adalah menentukan atribut-atribut yang akan digunakan, dimana atribut yang digunakan yaitu nama, nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris, dan data point (*sample*) yang akan digunakan yaitu minat.

3. *Data cleaning and preprocessing*

Pada tahap ini dilakukan proses penghapusan data apabila ada data yang tidak digunakan. Data yang digunakan adalah data yang akan diolah atau data yang mempengaruhi pemilihan jurusan di perguruan tinggi sedangkan data yang tidak diolah akan dihapus.

4. *Data reduction and projection*

Karena tidak ada data yang dihapus pada proses *Data cleaning and preprocessing* maka atribut yang digunakan nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris.

5. *Choosing the data mining task*

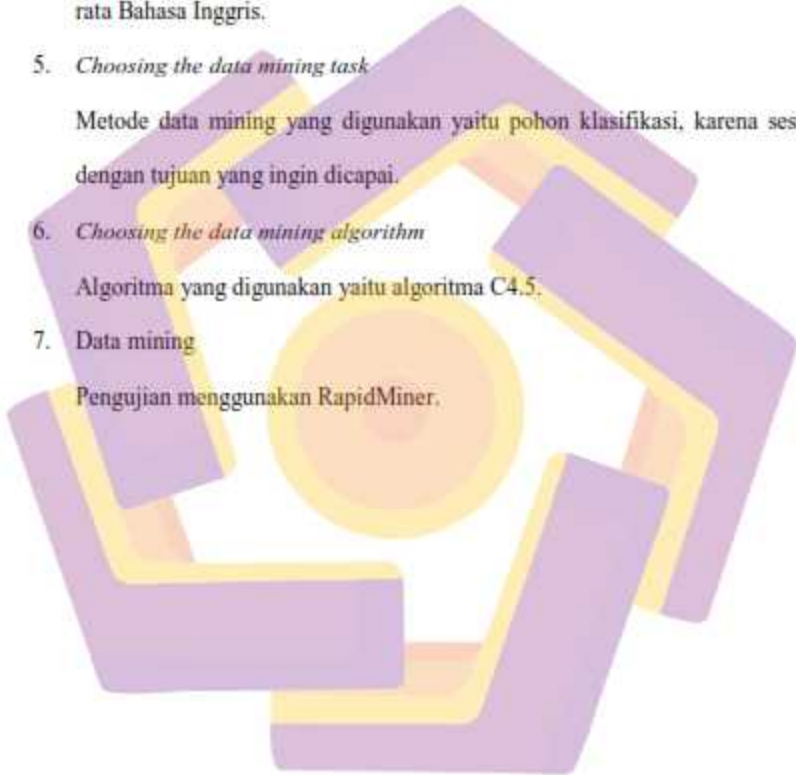
Metode data mining yang digunakan yaitu pohon klasifikasi, karena sesuai dengan tujuan yang ingin dicapai.

6. *Choosing the data mining algorithm*

Algoritma yang digunakan yaitu algoritma C4.5.

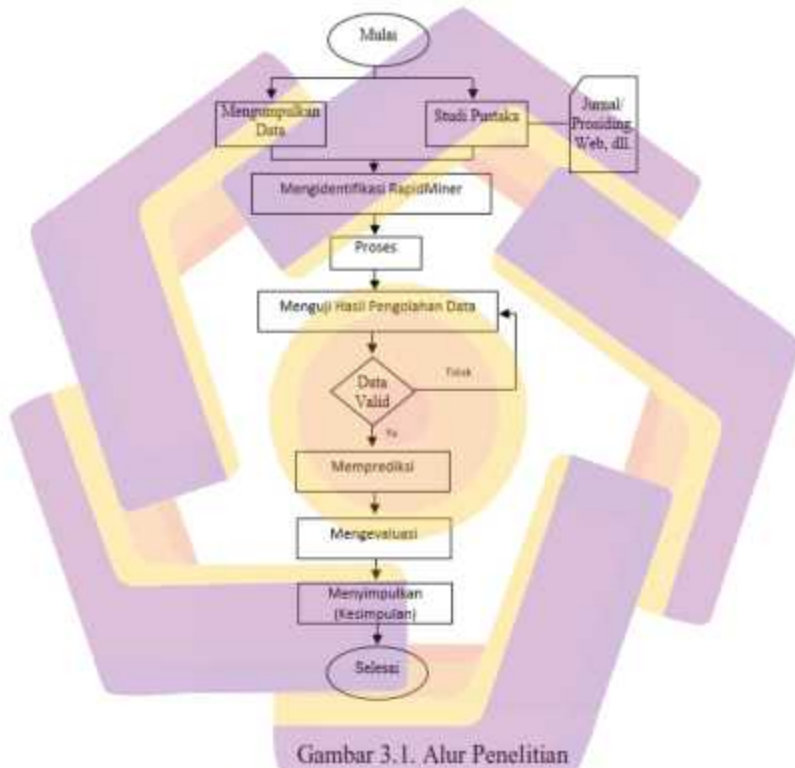
7. *Data mining*

Pengujian menggunakan RapidMiner.



3.4. Alur Penelitian

Alur penelitian menjelaskan tahap demi tahap dalam proses untuk mendapatkan keputusan terbaik dalam penelitian. Berikut alur penelitian yang digunakan :



Dari gambar 3.1. dapat dijelaskan :

- Data yang digunakan dalam penelitian ini adalah nilai raport siswa kelas XII SMKN 1 Donorojo Pacitan dan juga kuesioner yang dibagikan ke siswa
- Pustaka yang digunakan dalam penelitian ini adalah jurnal/prosiding, buku, web, dan lainnya yang relevan dengan tema penelitian.

- c. Mengidentifikasi dengan RapidMiner yaitu proses import data ke RapidMiner yang didalamnya terdapat proses seleksi data.
- d. Menguji hasil penelitian adalah proses pengolahan data di dalam RapidMiner untuk mengetahui *information gain* yang digunakan untuk menentukan pohon keputusan.
- e. Memprediksi adalah hasil dari pohon keputusan yang dihasilkan dari olah data dengan RapidMiner.
- f. Mengevaluasi adalah proses evaluasi dari hasil atau pohon keputusan yang dihasilkan dari olah data dengan RapidMiner untuk mengetahui minat peserta didik bila peserta didik akan melanjutkan pendidikan ke perguruan tinggi.

Tabel 3.1. Data Nilai Raport dan Minat Siswa

No	Nama	Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
1	DC	86	84	82	78	80	Bahasa Inggris
2	LDES	87	85	84	84	81	Matematika
3	RO	81	83	84	77	72	Matematika
4	BTK	86	84	84	83	75	Bahasa Indonesia
5	YNW	82	84	84	81	80	Bahasa Inggris
6	SK	83	84	85	85	84	Bahasa Inggris
7	AAM	84	84	82	78	76	Bahasa Indonesia
8	ADP	77	78	82	71	67	Agama Islam
9	ET	87	84	82	77	76	Bahasa Indonesia
10	AN	79	81	80	77	68	Bahasa Indonesia
11	NBL	87	85	82	88	76	Matematika
12	DACM	84	84	81	70	73	Agama Islam
13	RDA	83	84	82	80	73	Bahasa Indonesia
14	SDT	90	84	85	83	80	Bahasa Indonesia

Tabel 3.1. (lanjutan)

No	Nama	Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
15	FAAR	84	85	84	84	79	Matematika
16	RSP	76	81	81	72	70	Bahasa Indonesia
17	PRS	81	83	82	73	67	Agama Islam
18	JAS	81	84	82	75	67	Matematika
19	RCF	87	85	88	81	81	Matematika
20	RKS	83	83	83	72	67	Agama Islam
21	AAN	89	85	83	75	77	PPKn
22	SRJE	78	83	82	76	66	Agama Islam
23	DPT	85	84	82	74	79	Bahasa Indonesia
24	EH	83	84	84	75	81	Bahasa Inggris
25	HTPA	82	84	81	73	71	PPKn
26	SNF	82	84	83	76	76	Agama Islam
27	YSN	89	84	83	84	78	Bahasa Inggris
28	ADW	86	85	82	79	78	PPKn
29	ATY	84	83	82	79	73	Bahasa Indonesia
30	APT	85	84	84	78	76	Bahasa Indonesia
31	DTW	90	84	85	83	81	Agama Islam
32	YAK	76	83	80	73	70	Bahasa Indonesia
33	RK	84	84	83	77	90	Bahasa Inggris
34	NW	85	84	84	77	76	Matematika
35	KS	87	85	84	77	81	Bahasa Inggris
36	DS	80	83	81	73	71	Bahasa Indonesia
37	NTF	88	85	84	80	73	Bahasa Indonesia
38	MRS	83	84	82	80	75	Bahasa Indonesia
39	MAE	82	83	84	73	75	Agama Islam
40	ND	85	84	84	70	79	Bahasa Indonesia
41	LS	76	83	76	69	74	Bahasa Indonesia
42	Er	85	84	85	78	77	Bahasa Indonesia
43	BFR	85	85	87	83	82	Bahasa Inggris

Tabel 3.1. (lanjutan)

No	Nama	Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
44	APD	76	82	79	76	78	Bahasa Indonesia
45	DSR	86	84	79	70	77	Bahasa Indonesia
46	MFK	80	83	81	73	70	Bahasa Indonesia
47	ES	89	85	88	87	82	Matematika
48	ATN	79	84	79	74	79	Agama Islam
49	SRA	87	84	83	76	77	Bahasa Indonesia
50	AA	79	83	81	76	70	Bahasa Indonesia
51	DAPA	84	85	85	81	75	Agama Islam
52	DAP	84	84	87	80	77	Bahasa Indonesia
53	ESa	81	84	82	81	77	Bahasa Indonesia
54	AVD	85	84	84	76	79	Bahasa Inggris
55	WDA	79	83	84	75	68	Agama Islam
56	MLW	85	85	81	76	79	Bahasa Indonesia
57	HDP	88	85	85	80	78	Bahasa Inggris
58	ATK	84	84	82	74	77	Agama Islam
59	JS	84	85	84	75	77	Bahasa Indonesia
60	ANA	84	84	85	77	72	Bahasa Indonesia
61	SM	86	83	86	73	81	Bahasa Indonesia
62	MY	86	83	81	78	71	Matematika
63	HS	86	84	83	76	68	Bahasa Indonesia
64	UAL	89	84	85	80	76	Bahasa Indonesia
65	GES	86	84	79	72	77	Bahasa Indonesia
66	CF	86	83	82	76	78	PPKn
67	ADC	88	84	86	87	82	Matematika
68	TSH	78	83	78	70	79	Bahasa Indonesia
69	DA	86	84	81	75	77	Bahasa Inggris

Tabel 3.1. (lanjutan)

No	Nama	Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
70	YM	82	83	82	85	76	Bahasa Indonesia
71	MRN	82	85	83	82	74	Matematika
72	ZAF	85	83	83	72	72	Agama Islam
73	RW	87	83	82	76	75	Bahasa Indonesia

Tabel 3.1. adalah data alumni SMKN 1 Donorojo yang telah mengisi kuesioner yang disebar melalui Google Form. Data tersebut yang selanjutnya akan diolah dalam penelitian ini dengan menggunakan RapidMiner untuk menghasilkan pohon keputusan dalam menentukan jurusan peserta didik saat melanjutkan ke perguruan tinggi.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Menurut Suherman et al. (2021), langkah-langkah penyelesaian *Decision Tree* menggunakan algoritma C4.5 sebagai berikut :

2.1. Mengembangkan dan Memahami Domaln Aplikasi (*Developing and Understanding The Application Domain*)

Dari Tabel Data Nilai Raport Dan Minat Siswa pada BAB III didapatkan pengelompokan siswa berdasarkan minat dengan hasil sebagai berikut :

Tabel 4.1 Pengelompokan siswa berdasarkan minat

Minat	Jumlah Siswa
PAI (Agama Islam)	13
PKn (Pendidikan Kewarganegaraan)	4
Bahasa Indonesia	34
Matematika	11
Bahasa Inggris	11
Jumlah Total	73

Tabel 4.1. merupakan hasil pengelompokan siswa berdasarkan minat pada mata pelajaran PAI, PKn, Bahasa Indonesia, Matematika, dan Bahasa Inggris. Dari data di Tabel 4.1. yang akan diproses untuk menghitung information gain yang selanjutnya akan digunakan untuk menentukan pengaruh dari atribut dalam pemilihan jurusan di perguruan tinggi.

2.2. Membuat Kumpulan Data Target (*Creating A Target Data Set*)

Variabel (atribut) yang digunakan yaitu nama, nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris, dan minat.

2.3. Pembersihan Dan Pemrosesan Awal Data (*Data Cleaning And Preprocessing*)

Pada tahap ini untuk memastikan kebenaran, konsistensi, dan kegunaan suatu data yang ada dalam dataset maka dilakukan pembersihan dan pemrosesan awal data (*data cleaning and preprocessing*), yaitu dengan mendeteksi adanya error atau corrupt pada data, kemudian memperbaiki atau menghapus data jika memang diperlukan.

Pada proses penggabungan beberapa data, tidak terdeteksi adanya data terduplikasi atau salah label, maka pada tahap ini tidak ada data yang dihapus.

2.4. Reduksi Dan Proyeksi Data (*Data Reduction And Projection*)

Karena tidak ada data yang dihapus pada proses *Data cleaning and preprocessing* maka atribut yang digunakan nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris.

Tabel 4.2. Data Testing

Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
86	84	82	78	80	Bahasa Inggris
87	85	84	84	81	Matematika

Tabel 4.2. (Lanjutan)

Rata-rata PAI	Rata-rata PKn	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
81	83	84	77	72	Matematika
86	84	84	83	75	Bahasa Indonesia
82	84	84	81	80	Bahasa Inggris
83	84	85	85	84	Bahasa Inggris
84	84	82	78	76	Bahasa Indonesia
77	78	82	71	67	Agama Islam
87	84	82	77	76	Bahasa Indonesia
79	81	80	77	68	Bahasa Indonesia
87	85	82	88	76	Matematika
84	84	81	70	73	Agama Islam
83	84	82	80	73	Bahasa Indonesia
90	84	85	83	80	Bahasa Indonesia
84	85	84	84	79	Matematika
76	81	81	72	70	Bahasa Indonesia
81	83	82	73	67	Agama Islam
81	84	82	75	67	Matematika
87	85	88	81	81	Matematika
83	83	83	72	67	Agama Islam
89	85	83	75	77	PPKn
78	83	82	76	66	Agama Islam
85	84	82	74	79	Bahasa Indonesia
83	84	84	75	81	Bahasa Inggris
82	84	81	73	71	PPKn
82	84	83	76	76	Agama Islam
89	84	83	84	78	Bahasa Inggris
86	85	82	79	78	PPKn
84	83	82	79	73	Bahasa Indonesia
85	84	84	78	76	Bahasa Indonesia
90	84	85	83	81	Agama Islam
76	83	80	73	70	Bahasa Indonesia
84	84	83	77	90	Bahasa Inggris
85	84	84	77	76	Matematika
87	85	84	77	81	Bahasa Inggris
80	83	81	73	71	Bahasa Indonesia
88	85	84	80	73	Bahasa Indonesia
83	84	82	80	75	Bahasa Indonesia
82	83	84	73	75	Agama Islam

Tabel 4.2. (Lanjutan)

Rata-rata PAI	Rata-rata PKN	Rata-rata Bahasa Indonesia	Rata-rata Matematika	Rata-rata Bahasa Inggris	Minat
85	84	84	70	79	Bahasa Indonesia
76	83	76	69	74	Bahasa Indonesia
85	84	85	78	77	Bahasa Indonesia
85	85	87	83	82	Bahasa Inggris
76	82	79	76	78	Bahasa Indonesia
86	84	79	70	77	Bahasa Indonesia
80	83	81	73	70	Bahasa Indonesia
89	85	88	87	82	Matematika
79	84	79	74	79	Agama Islam
87	84	83	76	77	Bahasa Indonesia
79	83	81	76	70	Bahasa Indonesia
84	85	85	81	75	Agama Islam
84	84	87	80	77	Bahasa Indonesia
81	84	82	81	77	Bahasa Indonesia
85	84	84	76	79	Bahasa Inggris
79	83	84	75	68	Agama Islam
85	85	81	76	79	Bahasa Indonesia
88	85	85	80	78	Bahasa Inggris
84	84	82	74	77	Agama Islam
84	85	84	75	77	Bahasa Indonesia
84	84	85	77	72	Bahasa Indonesia
86	83	86	73	81	Bahasa Indonesia
86	83	81	78	71	Matematika
86	84	83	76	68	Bahasa Indonesia
89	84	85	80	76	Bahasa Indonesia
86	84	79	72	77	Bahasa Indonesia
86	83	82	76	78	PPKn
88	84	86	87	82	Matematika
78	83	78	70	79	Bahasa Indonesia
86	84	81	75	77	Bahasa Inggris
82	83	82	85	76	Bahasa Indonesia
82	85	83	82	74	Matematika
85	83	83	72	72	Agama Islam
87	83	82	76	75	Bahasa Indonesia

Tabel 4.2. merupakan data yang akan diuji pada RapidMiner untuk mengetahui pengaruhnya terhadap pemilihan jurusan saat peserta didik melanjutkan ke perguruan tinggi.

2.5. Memilih Software Untuk Pengujian (*Choosing The Data Mining Task*)

Dalam penelitian ini pengujian dengan menggunakan software Data Mining, yaitu RapidMiner. Untuk pengujian dengan menggunakan aplikasi RapidMiner ini digunakan data testing sebanyak 73 data.

2.6. Pengujian Data Dengan RapidMiner

Berikut langkah-langkah pengujian dengan RapidMiner (Nas, 2021):

2.6.1. Import Data

Pengujian dilakukan dengan meng-*import* data pengujian dalam file Excel ke RapidMiner. Proses ini merupakan proses untuk seleksi data dan juga memilih format kolom dari file atau data yang di-*import* dari Excel.

Beberapa pilihan (tipe) format data, yaitu :

- a. Polynominal = untuk atribut yang memiliki lebih dari 2 kategori
- b. Binominal = untuk atribut yang memiliki 2 kategori saja
- c. Real = untuk tipe data yang memiliki nilai koma, atau decimal
- d. Integer = untuk file bernilai integer atau bilangan bulat tanpa koma.

Row No.	Minat	Rata-rata PAI	Rata-rata PKs	Rata-rata B...	Rata-rata M...	Rata-rata B...
1	E.ing	85	84	82	78	80
2	MTK	87	85	84	84	81
3	MTK	81	83	84	77	72
4	E.ind	86	84	84	83	75
5	E.ing	82	84	84	81	80
6	E.ing	83	84	85	85	84
7	E.ind	84	84	82	78	76
8	PAI	77	76	82	71	87
9	E.ind	87	84	82	77	76
10	E.ind	79	81	80	77	68
11	MTK	87	85	82	88	76
12	PAI	84	84	81	70	73
13	E.ind	83	84	82	80	73
14	E.ind	80	84	85	82	80
15	MTK	84	85	84	84	79

ExampleSet (73 examples, 1 special attribute, 5 regular attributes)

Gambar 4.1. Hasil *Import Data Testing* di RapidMiner

Pada gambar 4.1, merupakan hasil import data yang telah melalui proses seleksi data dengan menghilangkan kolom yang berisi nama siswa karena tidak digunakan.

2.6.2. Pengujian Data

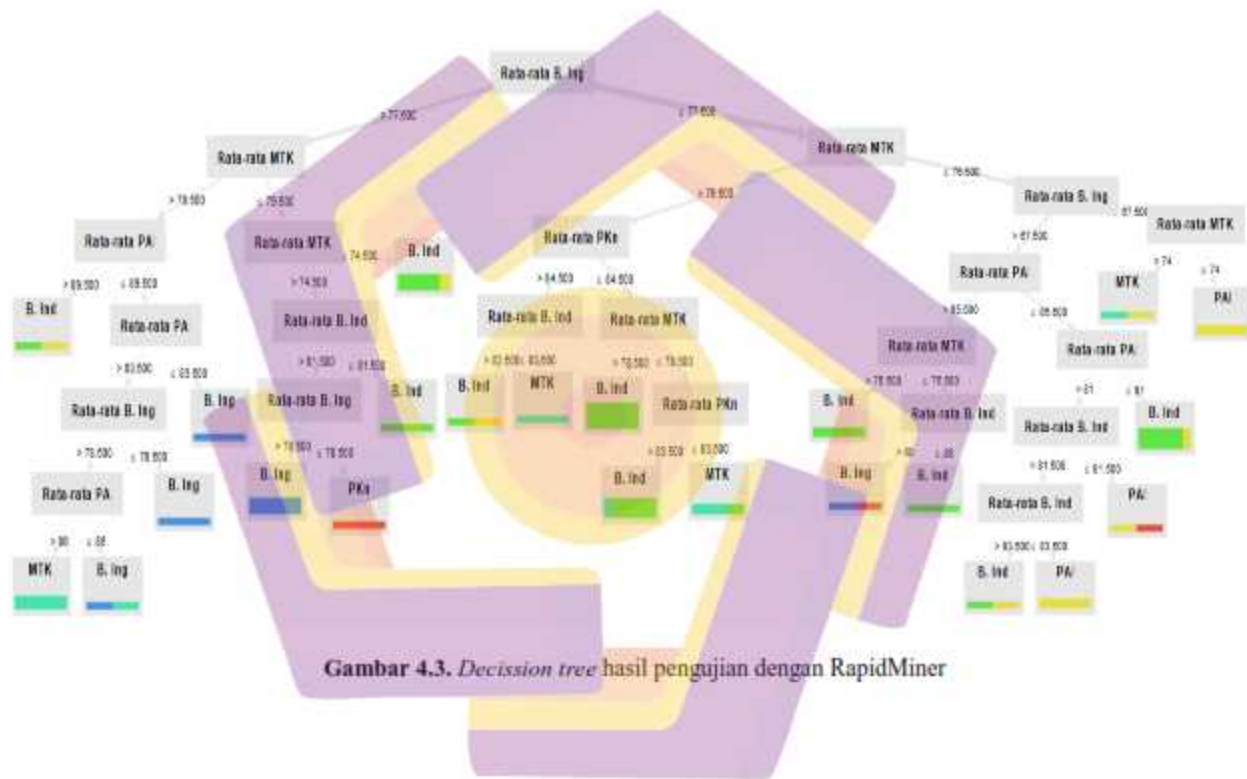
Tahap ini merupakan proses untuk menghitung *information gain* yang digunakan untuk mengetahui seberapa tingkat pengaruh masing-masing atribut terhadap minat peserta didik yang berkaitan dengan pemilihan jurusan di perguruan tinggi.

Pengujian data dilakukan dengan melakukan *drag* data yang telah di-*import* ke lembar kerja RapidMiner pada bagian *Proses* dan juga pemilihan *Decision Tree* pada kolom *Operators*, seperti terlihat pada gambar 4.2.



Gambar 4.2. Proses view pengujian data

Gambar 4.2 menunjukkan *process view* untuk menentukan atau menghitung *information gain*. Selanjutnya dilakukan *running* dan menghasilkan *decision tree* seperti pada Gambar 4.3.



Gambar 4.3. Decision tree hasil pengujian dengan RapidMiner

Gambar 4.3. dapat dijelaskan bahwa nilai rata-rata Bahasa Inggris mempunyai pengaruh paling besar terhadap minat peserta didik, dimana :

- a. Jika nilai rata-rata Bahasa Inggris >77.500 maka selanjutnya ditentukan oleh nilai rata-rata Matematika, jika nilai rata-rata Matematika >79.500 maka selanjutnya ditentukan oleh nilai rata-rata IPA, tetapi jika nilai rata-rata Matematika ≤ 79.500 maka selanjutnya ditentukan oleh rata-rata nilai Matematika lagi dimana jika nilai rata-rata Matematika ≤ 74.500 maka minat peserta didik adalah Bahasa Indonesia sedangkan jika nilai rata-rata >74.500 selanjutnya minat peserta didik ditentukan oleh rata-rata nilai Bahasa Indonesia, dan seterusnya.
- b. Jika nilai rata-rata Bahasa Inggris ≤ 77.500 maka selanjutnya ditentukan oleh nilai rata-rata Matematika, dimana jika rata-rata nilai Matematika >76.500 maka selanjutnya ditentukan nilai Pendidikan Kewarganegaraan, jika nilai rata-rata Pendidikan Kewarganegaraan >84.500 maka selanjutnya ditentukan oleh nilai rata-rata Bahasa Indonesia, sedangkan jika nilai rata-rata kewarganegaraan ≤ 84.500 selanjutnya ditentukan oleh nilai rata-rata Matematika. Sedangkan jika nilai rata-rata Matematika ≤ 76.500 selanjutnya ditentukan oleh nilai rata-rata Bahasa Inggris, dimana jika rata-rata Bahasa Inggris >67.500 selanjutnya ditentukan oleh nilai rata-rata Pendidikan Agama Islam, sedangkan jika nilai rata-rata Bahasa Inggris ≤ 67.500 selanjutnya ditentukan oleh nilai rata-rata Matematika, dan seterusnya.

Dari Gambar 4.3. juga diketahui bahwa dalam pohon keputusan atribut nilai rata-rata Bahasa Inggris berada pada *node* paling atas yang menggambarkan

atau menunjukkan bahwa nilai rata-rata Bahasa Inggris mempunyai pengaruh paling besar (mempunyai nilai *information gain* paling tinggi) terhadap minat peserta didik dalam memilih jurusan di perguruan tinggi. Maka akar dari pohon keputusan adalah atribut rata-rata nilai Bahasa Inggris.

```

Rata-rata B. Ing > 77.500
|
| Rata-rata MTK > 79.500
| |
| | Rata-rata PAI > 88.500: B. Ind (B. Ing=0, MTK=0, B. Ind=1, PAI=1, PKN=0)
| | Rata-rata PAI ≤ 87.800
| | |
| | | Rata-rata PAI > 83.800
| | | |
| | | | Rata-rata B. Ing > 78.500
| | | | |
| | | | | Rata-rata PAI > 86: MTK (B. Ing=0, MTK=1, B. Ind=0, PAI=0, PKN=0)
| | | | | Rata-rata PAI ≤ 86: B. Ing (B. Ing=1, MTK=1, B. Ind=0, PAI=0, PKN=0)
| | | | | Rata-rata B. Ing > 70.500: B. Ing (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
| | | | | Rata-rata PAI < 81.000: B. Ing (B. Ing=1, MTK=0, B. Ind=0, PAI=0, PKN=0)
| | | Rata-rata MTK < 78.500
| | Rata-rata MTK < 74.500
| | |
| | | Rata-rata B. Ind > 81.500
| | | |
| | | | Rata-rata B. Ing > 78.500: B. Ing (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
| | | | Rata-rata B. Ing > 79.500: PKN (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=1)
| | | | Rata-rata B. Ind < 81.500: B. Ind (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
| | | Rata-rata MTK < 74.500: B. Ind (B. Ing=0, MTK=0, B. Ind=0, PAI=1, PKN=0)
Rata-rata B. Ing < 77.500
Rata-rata MTK > 76.500
|
| Rata-rata PKN > 84.800
| |
| | Rata-rata B. Ind > 83.500: B. Ind (B. Ing=0, MTK=0, B. Ind=1, PAI=1, PKN=0)
| | Rata-rata B. Ind < 83.500: MTK (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
Rata-rata PKN < 84.800
|
| Rata-rata MTK > 76.500: B. Ind (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
| Rata-rata MTK < 74.500
| |
| | Rata-rata PKN > 83.500: B. Ind (B. Ing=0, MTK=1, B. Ind=0, PAI=0, PKN=0)
| | Rata-rata PKN < 83.500: MTK (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)
Rata-rata MTK < 74.500
|
| Rata-rata B. Ing > 67.500
| Rata-rata PAI > 85.500
| |
| | Rata-rata MTK > 71.500: B. Ind (B. Ing=0, MTK=0, B. Ind=1, PAI=0, PKN=0)
| | Rata-rata MTK < 71.500
| | |
| | | Rata-rata B. Ind > 80: B. Ing (B. Ing=1, MTK=0, B. Ind=0, PAI=0, PKN=1)
| | | Rata-rata B. Ind < 80: B. Ind (B. Ing=0, MTK=0, B. Ind=1, PAI=0, PKN=0)
| Rata-rata PAI < 81.500
| |
| | Rata-rata B. Ind > 81.500
| | |
| | | Rata-rata B. Ind > 83.500: B. Ind (B. Ing=0, MTK=0, B. Ind=1, PAI=1, PKN=0)
| | | Rata-rata B. Ind < 83.500: PAI (B. Ing=0, MTK=0, B. Ind=0, PAI=1, PKN=0)
| | | Rata-rata B. Ind < 81.500: PAI (B. Ing=0, MTK=0, B. Ind=0, PAI=1, PKN=1)
| | Rata-rata PAI < 81: B. Ind (B. Ing=0, MTK=0, B. Ind=0, PAI=1, PKN=0)
Rata-rata B. Ing < 67.500
|
| Rata-rata MTK > 74: MTK (B. Ing=0, MTK=1, B. Ind=0, PAI=1, PKN=0)
| Rata-rata MTK < 74: PAI (B. Ing=0, MTK=0, B. Ind=0, PAI=0, PKN=0)

```

Gambar 4.4. Description hasil pengujian dengan RapidMiner

Dari pengujian yang telah dilakukan dapat diketahui bahwa atribut-atribut yang mempengaruhi minat siswa dalam memilih jurusan di perguruan tinggi menggunakan algoritma C4.5 yaitu :

- a. Jika nilai rata-rata Bahasa Inggris >77.500 maka selanjutnya ditentukan oleh nilai rata-rata Matematika, jika nilai rata-rata Matematika >79.500 maka selanjutnya ditentukan oleh nilai rata-rata IPA, tetapi jika nilai rata-rata Matematika ≤ 79.500 maka selanjutnya ditentukan oleh rata-rata nilai Matematika lagi dimana jika nilai rata-rata Matematika ≤ 74.500 maka minat peserta didik adalah Bahasa Indonesia sedangkan jika nilai rata-rata >74.500 selanjutnya minat peserta didik ditentukan oleh rata-rata nilai Bahasa Indonesia, dan seterusnya.
- b. Jika nilai rata-rata Bahasa Inggris ≤ 77.500 maka selanjutnya ditentukan oleh nilai rata-rata Matematika, dimana jika rata-rata nilai Matematika >76.500 maka selanjutnya ditentukan nilai Pendidikan Kewarganegaraan, jika nilai rata-rata Pendidikan Kewarganegaraan >84.500 maka selanjutnya ditentukan oleh nilai rata-rata Bahasa Indonesia, sedangkan jika nilai rata-rata kewarganegaraan ≤ 84.500 selanjutnya ditentukan oleh nilai rata-rata Matematika.
- Sedangkan jika nilai rata-rata Matematika ≤ 76.500 selanjutnya ditentukan oleh nilai rata-rata Bahasa Inggris, dimana jika rata-rata Bahasa Inggris >67.500 selanjutnya ditentukan oleh nilai rata-rata Pendidikan Agama Islam, sedangkan jika nilai rata-rata Bahasa Inggris ≤ 67.500 selanjutnya ditentukan oleh nilai rata-rata Matematika, dan seterusnya.

Dari gambar 4.3, dapat dicontohkan :

- a. Seorang peserta didik dengan nilai rata-rata Bahasa Inggris 80, nilai rata-rata Matematika 81, nilai rata-rata Bahasa Indonesia 93, dan nilai rata-rata PAI 92.

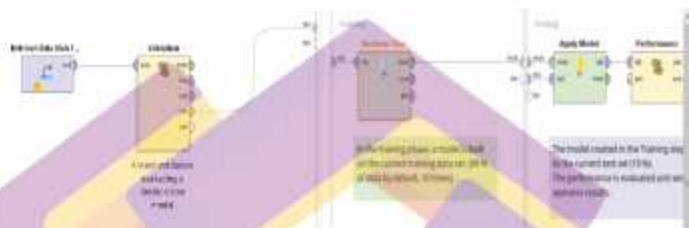
Maka dalam pohon keputusan yang pertama mempengaruhi adalah nilai rata-rata Bahasa Inggris, kemudian level di bawahnya dipengaruhi oleh nilai rata-rata Matematika, dan kemudian dipengaruhi oleh nilai Bahasa Indonesia, kemudian dipengaruhi nilai rata-rata PAI, dan jika nilai rata-rata PAI ≤ 89.500 akan dipengaruhi lagi oleh nilai rata-rata yang lain, akan tetapi jika nilai rata-rata PAI >89.500 maka hasil pohon keputusannya adalah jurusan yang tepat untuk peserta didik adalah prodi Bahasa Indonesia.

- b. Seorang peserta didik dengan nilai rata-rata Bahasa Inggris 75, nilai rata-rata Matematika 78, nilai rata-rata Bahasa Indonesia 82, nilai rata-rata PKn 86.

Maka dalam pohon keputusan yang pertama mempengaruhi adalah nilai rata-rata Bahasa Inggris, kemudian level di bawahnya dipengaruhi oleh nilai rata-rata Matematika, dan kemudian dipengaruhi oleh nilai rata-rata PKn, dan terakhir dipengaruhi oleh nilai rata-rata Bahasa Indonesia, dimana jika nilai rata-rata Bahasa Indonesia >83.500 maka jurusan yang tepat untuk peserta didik adalah prodi Bahasa Indonesia dan jika nilai rata-rata Bahasa Indonesia ≤ 83.500 maka prodi yang tepat adalah prodi Matematika.

2.6.3. Validasi dengan *Cross Validation*

Cross Validation adalah suatu dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi (tingkat kedekatan klasifikasi dengan hasil fakta) yang maksimal. Proses *Cross Validation* seperti ditunjukkan pada gambar 4.5.



Gambar.4.5. Proses validasi dengan *Cross Validation*

Proses *Cross Validation* akan menghasilkan *PerformanceVector* yang di dalamnya diperoleh hasil *accuracy*, *recall* dan *precision*. Berikut hasil atau tingkat akurasi yang diperoleh proses *data mining* dengan metode *decision tree* dengan algoritma C4.5.

Tabel 4.3 *PerformanceVector* pada RapidMiner

accuracy: 45.19% (+/- 16.79%) (macro average: 45.21%)

	True Neg	True Pos	False Neg	False Pos	True FPR	Class precision
pred. B. Ing	3	3	2	0	1	33.33%
pred. BTK	3	2	4	4	1	14.29%
pred. B. Iku	1	5	34	5	1	96.67%
pred. PM	3	1	2	4	1	30.30%
pred. Pcs	1	0	2	0	0	0.00%
macro recall	37.27%	16.13%	70.59%	36.77%	1.03%	

Dari Tabel *Performance Vector* di atas diketahui, yaitu :

a. *Accuracy*

Accuracy merupakan tingkat kedekatan klasifikasi dengan hasil fakta. Dari Tabel 4.3 diketahui bahwa tingkat akurasi yang dihasilkan sebesar 45,21%.

Dari data testing sebanyak 73 data (siswa) diketahui atau bahwa siswa yang nilai rata-rata paling tinggi Bahasa Inggris dan yang minat masuk program studi Bahasa Inggris sebanyak 3 siswa, siswa yang nilai rata-rata paling tinggi Matematika dan yang minat masuk program studi Matematika sebanyak 2 siswa, siswa yang nilai rata-rata paling tinggi Bahasa Indonesia dan yang minat masuk program studi Bahasa Indonesia sebanyak 24 siswa, siswa yang nilai rata-rata paling tinggi Pendidikan Agama Islam dan yang minat masuk program studi Pendidikan Agama Islam sebanyak 4 siswa, siswa yang nilai rata-rata paling tinggi Pendidikan Kewarganegaraan dan yang minat masuk program studi Kewarganegaraan sebanyak 0 siswa. Sedangkan sisanya minat masuk program studi di perguruan tinggi tidak sesuai atau berdasarkan nilai rata-rata paling tinggi dari mata pelajaran Bahasa Inggris, Matematika, Bahasa Indonesia, Pendidikan Agama Islam, dan Pendidikan Kewarganegaraan.

Dalam penelitian ini atribut yang digunakan bukan mata pelajaran kejuruan di SMK dikarenakan tujuannya adalah untuk mengetahui bakat peserta didik dari sisi mata pelajaran umum, dikarenakan adanya fakta-fakta dilapangan yang telah dijelaskan pada awal penelitian ini.

b. Precision

Precision merupakan tingkat akurasi antara informasi yang diminta dan jawaban yang diberikan oleh sistem. Dari Tabel 4.3 diketahui bahwa *class precision* Bahasa Inggris 33.33% dimana dari 9 peserta didik yang berdasarkan hasil olah RapidMiner jurusan yang tepat Bahasa Inggris dan peserta didik juga mengambil jurusan Bahasa Inggris sebanyak 3 peserta didik, Matematika

14.29% dimana dari 14 peserta didik yang berdasarkan hasil olah RapidMiner jurusan yang tepat Matematika dan peserta didik juga mengambil jurusan Matematika sebanyak 2 peserta didik, Bahasa Indonesia 66.67% dimana dari 36 peserta didik yang berdasarkan hasil olah RapidMiner jurusan yang tepat Bahasa Indonesia dan peserta didik juga mengambil jurusan Bahasa Indonesia sebanyak 24 peserta didik, Pendidikan Agama Islam 36.36% dimana dari 11 peserta didik yang berdasarkan hasil olah RapidMiner jurusan yang tepat Pendidikan Agama Islam dan peserta didik juga mengambil jurusan Pendidikan Agama Islam sebanyak 4 peserta didik, dan Pendidikan Kewarganegaraan 0.00% dimana dari 3 peserta didik yang berdasarkan hasil olah RapidMiner jurusan yang tepat Pendidikan Kewarganegaraan dan peserta didik juga mengambil jurusan Pendidikan Kewarganegaraan sebanyak 0 peserta didik.

c. Recall

Recall merupakan tingkat keberhasilan dengan menemukan sebuah informasi. Dari Tabel 4.3 diketahui bahwa *class recall* Bahasa Inggris 27.27% yang artinya dari 11 peserta didik yang mengambil jurusan Bahasa Inggris yang sesuai dengan hasil RapidMiner sebanyak 3 peserta didik, Matematika 18.18% yang artinya dari 13 peserta didik yang mengambil jurusan Matematika yang sesuai dengan hasil RapidMiner sebanyak 2 peserta didik, Bahasa Indonesia 70.59% yang artinya dari 34 peserta didik yang mengambil jurusan Bahasa Indonesia yang sesuai dengan hasil RapidMiner sebanyak 24 peserta didik, Pendidikan Agama Islam 30.77% yang artinya dari 13 peserta didik yang mengambil jurusan Pendidikan Agama Islam yang sesuai dengan hasil

RapidMiner sebanyak 4 peserta didik, dan Pendidikan Kewarganegaraan 0.00% yang artinya dari 4 peserta didik yang mengambil jurusan Pendidikan Kewarganegaraan yang sesuai dengan hasil RapidMiner sebanyak 0 peserta didik.

Dari hasil tersebut dapat diketahui bahwa ketepatan atau kesesuaian siswa dalam memilih jurusan di perguruan tinggi tidak selalu berdasarkan nilai mata pelajaran dan juga minat dari tetapi lebih banyak berdasarkan variabel atau atribut lain yang belum diketahui.



BAB V

PENUTUP

5.1. Kesimpulan

Kesimpulan dari penelitian ini sebagai berikut :

- a. Penerapan *Decision Tree Algoritma C4.5* dengan menggunakan atribut nilai rata-rata PAI, nilai rata-rata PKn, nilai rata-rata Bahasa Indonesia, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris serta minat dan bakat siswa dapat membantu SMK Negeri 1 Donorojo dalam memberikan informasi mengenai jurusan yang akan diambil dalam melanjutkan ke perguruan tinggi kepada peserta didik.
- b. Dari data testing sebanyak 73 data (siswa) diketahui atau bahwa siswa yang minat masuk program studi di perguruan tinggi sesuai dengan rata-rata paling tinggi baik Bahasa Inggris, Matematika, Bahasa Indonesia, Pendidikan Agama Islam, dan Pendidikan Kewarganegaraan sebanyak 33 siswa.
- c. Dari data testing sebanyak 73 data (siswa) diketahui atau bahwa siswa yang minat masuk program studi di perguruan tinggi tidak sesuai dengan rata-rata paling tinggi baik Bahasa Inggris, Matematika, Bahasa Indonesia, Pendidikan Agama Islam, dan Pendidikan Kewarganegaraan sebanyak 40 siswa.
- d. Dari pengolahan data dengan RapidMiner diketahui bahwa tingkat akurasi yang dihasilkan sebesar 45,21%, yang dapat disimpulkan bahwa minat siswa dalam memilih jurusan di perguruan tinggi tidak selalu berdasarkan nilai mata

pelajaran tetapi lebih banyak berdasarkan variabel atau atribut lain yang belum diketahui.

5.2. Saran

Saran untuk perbaikan penelitian selanjutnya yaitu :

- a. Atribut dalam penelitian ini tidak berdasarkan mata pelajaran kejuruan atau jurusan yang diambil peserta didik saat di SMK tetapi berdasarkan mata pelajaran non kejuruan dikarenakan tujuannya adalah untuk mengetahui bakat peserta didik dari sisi mata pelajaran umum, saran untuk penelitian selanjutnya adalah mengupayakan untuk lebih banyak lagi keberagaman data peserta didik SMK yang diproses sehingga akan menghasilkan akurasi data yang lebih baik.
- b. Tingkat akurasi yang dihasilkan sebesar 45,21%, untuk atribut penelitian dapat ditambahkan dengan atribut yang lain misalkan data lampau dari alumni yang telah lulus kemudian melanjutkan ke perguruan tinggi.

DAFTAR PUSTAKA

- Ainurrohmah. (2021). Akurasi Algoritma Klasifikasi pada software Rapidminer dan Weka. *Prosiding Seminar Nasional Matematika*, 4, 493-499. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Andi AP. (2019). Pengembangan minat belajar dalam pembelajaran. *Jurnal IDAARAH*, III(36), 205-215.
- Asmara, A., & Haryanto, H. (2015). Pengembangan Tes Minat Dan Bakat Dengan Metode Jaringan Syaraf Tiruan (Jst) Untuk Memprediksi Potensi Siswa Bidang Robotika. *Jurnal Pendidikan Yokasi*, 5(3), 273. <https://doi.org/10.21831/jpv.v5i3.6483>
- Astuti. (2011). Manajemen Peserta Didik Astuti. *Astuti*, 11(2), 134.
- Bahri S. (2022). Implementasi Data Mining Untuk Menentukan Minat Siswa Dalam Menentukan Jurusan Pada Perguruan Tinggi. 3(1), 23-33. <http://ars.ilkom.unsri.ac.id/472>
- Basri, H., Afdal, A., & Yusuf, A. M. (2021). Kesesuaian Antara Bakat dan Minat dalam Menentukan Jurusan Pendidikan Tinggi Melalui Bimbingan Karir di Sekolah Menengah Atas. ... *Journal of School Counseling*, 6, 157-163. <https://jurnal.iicet.org/index.php/schoulid/article/view/885>
- Dharshinni, N. P., Putri, R., Dameria, M., Christian, E., Friando, D., & Very, Y. (2021). Classification Of Major Selection Based On Students Expertise Using C4. 5 Algorithm. 9(2), 412-418.
- Kemendikbud No.23. (2016). Pengembangan Bakat dan Minat Siswa SMK. *Kemendikbud No.23. (2016). Pengembangan Bakat dan Minat Siswa SMK. Kemendikbud No.23. (2016). Pengembangan Bakat dan Minat Siswa SMK. Kemendikbud No.23. (2016). Pengembangan Bakat dan Minat Siswa SMK.* Direktorat Jenderal Pendidikan Dasar Dan Menengah Direktorat Pembinaan Sekolah Menengah Kejuruan, 77.
- Manullang, N., Sembiring, R. W., Gunawan, I., Parlina, I., & Irawan, I. (2021). Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 1-5. <https://doi.org/10.35960/ikomti.v2i2.700>
- Merawati, D., & Rino. (2019). Penerapan data mining penentu minat Dan bakat siswa Smk dengan metode C4. 5. *Jurnal Algor*, 1(1), 28-37.
- Muhamad, M., Windarto, A. P., & Suhada, S. (2019). Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 1-8. <https://doi.org/10.30865/komik.v3i1.1688>
- Narulita, S. O. A. T. S. I. (2021). Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4.5 Untuk Penentuan Peminatan Peserta Didik. *Jurnal Media Aplikom*, 13, 68-82.
- Nas, C. (2021). Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131-145. <https://doi.org/10.34010/jamika.v11i2.5506>
- Pangastuti PAD, P. D. (2019). Penerapan Sistem Informasi Raport Online.

- Ekuitas: Jurnal Pendidikan Ekonomi*, 7(1), 1-4.
<https://doi.org/10.23887/ekuitas.v7i1.16488>
- Peraturan Pemerintah No. 29. (1990). *Peraturan Pemerintah Republik Indonesia Nomor 29 Tahun 1990 Tentang Pendidikan Menengah Presiden Republik Indonesia*, 9-25.
- Prasetyo, E. (2012). *Data mining : konsep dan aplikasi menggunakan MATLAB*. CV Andi Offset.
- Ramadanty, M. C., Siregar, A. M., & Kusumaningrum, D. S. (2021). Penerapan Algoritma C4 . 5 dan K-Nearest Neighbor untuk Klasifikasi Peminatan Program Studi di Perguruan Tinggi Berdasarkan Nilai Rapor. *Scientific Student Journal for Information, Technology and Science*, 11, 76-84.
- Ramadhanu, A., Defit, S., & Kareem, S. W. (2021). Hybrid Data Mining with the Combination of K-Means Algorithm and C4.5 to Predict Student Achievement. *International Journal of Artificial Intelligence Research*, 5(2), 180-189. <https://doi.org/10.29099/ijair.v6i1.225>
- Ristianah, N., & Ma'sum, T. (2021). Peserta Didik Ideal Dalam Perspektif Pendidikan Islam. *Aulada*, 11(2), 233-249.
- Sari, I. P., & Harman, R. (2020). Decision Tree Technique Dalam Menentukan Penjurusan Siswa Menengah Kejuruan. *Journal of Information System Research (JOSH)*, 1(4), 296-304. <http://repository.upbatam.ac.id/314/>
- Selvia, Wendi, H. (2014). Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, November, 6.
- Septiana, W. O. A., & Sholeh, M. M. A. (2021). Pengaruh Motivasi Belajar terhadap Kemandirian Belajar Siswa Kelas Xi SMA Islam Sultan Agung 1 Semarang pada Mata Pelajaran Fiqih selama Masa Pandemi Covid-19. *Sultan Agung Fundamental Research Journal*, 2(1), 33-40. <http://jurnal.unissula.ac.id/index.php/safjr>
- Suherman, Purnamasari, M., & Fitriani Dwi Hastuti. (2021). Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas Minat Menggunakan Metode Decision Tree C4.5. *JSiI (Jurnal Sistem Informasi)*, 8(2), 141-149. <https://doi.org/10.30656/jsii.v8i2.3508>
- Sulihati I. (2022). Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi. *Jurnal Tecnoscienza*, 6(2), 311-320. <https://doi.org/10.51158/tecnoscienza.v6i2.694>
- Undang-undang RI No. 20. (2003). Undang Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional. *Demographic Research*, 49(0), 1-33 : 29 pag texts + end notes, appendix, referen.
- Wulandari, W., Rosnelly, R., & Wanayumini, W. (2021). Analisis Metode Decision Tree Dalam Memprediksi Kelulusan Mahasiswa. *CSRID (Computer Science Research and Its Development Journal)*, 13(3), 131. <https://doi.org/10.22303/csrid.13.3.2021.131-140>

Lampiran 1: Kuesioner

**ANGKET MINAT DAN BAKAT SISWA
SMKN 1 DONOROJO
PACITAN**

Nama Siswa :

N I S N :

Jurusan :

No	Pernyataan/Pertanyaan	SS	S	KS	TS
1	Saya merasa senang ketika membicarakan tentang perguruan tinggi.				
2	Bila dilihat dari prestasi saya selama ini, saya kemungkinan diterima di perguruan tinggi.				
3	Lulusan perguruan tinggi akan memperoleh kesempatan untuk mendapatkan pekerjaan yang lebih baik.				
4	Saya yakin kemampuan/keterampilan saya akan bertambah setelah lulus dari perguruan tinggi.				
5	Saya akan dapat memperoleh peluang pekerjaan yang lebih baik di bursa tenaga kerja bila lulus PT				
6	Masuk perguruan tinggi favorit merupakan cita-cita saya setelah lulus SMK.				
7	Saya ingin menjadi seorang yang ahli di bidang saya.				
8	Saya bercita-cita mendapatkan kehidupan yang lebih layak dan lebih baik dari sekarang.				
9	Dengan melanjutkan kuliah, akan melatih kemandirian dalam berfikir dan bersikap.				
10	Ketika membaca surat kabar, saya cenderung mencari informasi tentang perguruan tinggi, seperti tawaran beasiswa yang mungkin bisa saya dapatkan.				
11	Saya berusaha mencari informasi tentang perguruan tinggi melalui internet, koran, maupun teman-teman.				
12	Dengan melanjutkan ke perguruan tinggi akan menjanjikan masa depan saya lebih baik.				
13	Dengan melanjutkan pendidikan di perguruan tinggi akan membuat saya lebih profesional.				
14	Lulusan perguruan tinggi akan mendapatkan posisi pekerjaan yang lebih baik dari lulusan SMK.				
15	Saya mendapatkan dukungan dari orang tua untuk melanjutkan ke perguruan tinggi.				

16	Dalam memilih jurusan di perguruan tinggi orang tua memberikan saya kebebasan.				
17	Orang tua banyak memberikan masukan tentang perguruan tinggi yang baik untuk saya.				
18	Biaya kuliah beberapa perguruan tinggi yang berkualitas masih bisa dijangkau keluarga saya.				
19	Dengan masuk perguruan tinggi, saya ingin meningkatkan ekonomi keluarga, karena lulusan perguruan tinggi mendapat gaji lebih baik dari lulusan SMK.				
20	Bila ada kesempatan, saya akan melanjutkan studi ke perguruan tinggi dengan jurusan sesuai dengan nilai saya.				

