

**IMPLEMENTASI ALGORITMA KLASIFIKASI DATA
MINING DALAM MENENTUKAN PENJURUSAN
SISWA SMA KOTA BATAM**

SKRIPSI



**Oleh:
Nofrizal
150210201**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS PUTERA BATAM
2019**

**IMPLEMENTASI ALGORITMA KLASIFIKASI DATA
MINING DALAM MENENTUKAN PENJURUSAN
SISWA SMA KOTA BATAM**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**



**Oleh:
Nofrizal
150210201**

**PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
2019**

PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, dan/atau magister), baik di Universitas Putera Batam maupun di perguruan tinggi lain.
2. Skripsi ini adalah murni gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arah pembimbing.
3. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi.

Batam, 16 Februari 2019

Yang membuat pernyataan,



**IMPLEMENTASI ALGORITMA KLASIFIKASI DATA
MINING DALAM MENENTUKAN PENJURUSAN SISWA
SMA KOTA BATAM**

**Oleh
Nofrizal
150210201**

**SKRIPSI
Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera dibawah ini**

Batam, 16 Februari 2019

A handwritten signature in black ink, appearing to read 'Sunarsan Sitohang', written in a cursive style.

Sunarsan Sitohang, S.Kom., M.TI

ABSTRAK

Dewasa ini perkembangan teknologi informasi dan komunikasi ini seolah sudah tidak bisa kita bendung lagi dalam sisi kehidupan manusia. Dimana teknologi ini sangat berperan penting dan membawa manfaat untuk kemajuan manusia, yang dimana setiap bidang teknologi sudah sangat banyak macamnya. Penjurusan adalah suatu proses penyaluran dan penempatan dalam pemilihan program pengajaran kepada siswa siswi. Kurikulum yang dipakai pada sekolah menengah atas (SMA) pada saat ini yaitu kurikulum 2013. Seiring berkembangnya teknologi maka hal seperti ini bisa di atasi dengan teknik pengelompokan data yang di dapat dari hasil *data mining*. *Data mining* adalah disiplin ilmu yang mempelajari tentang sebuah teknik untuk menggali informasi tersembunyi untuk memperoleh keuntungan dari data yang ada, sehingga data mining juga sering di sebut dengan *knowledge discovery from database*. *Data mining* memiliki bermacam-macam metode yang dimana metode ini dapat di sesuaikan dengan kebutuhan yang ada. Dalam masalah penentuan penjurusan ini, metode yang digunakan yaitu dengan penerapan algoritma *Decision Tree C4.5*. Menerapkan metode data mining dengan algoritma *Decision Tree C4.5* dalam menentukan jurusan siswa SMA. Algoritma C4.5 dapat digunakan sebagai metode klasifikasi dalam memprediksi pemilihan jurusan di sekolah MAN 1 Batam dengan memperhatikan nilai *Gain* terbesar dari tiga atribut seperti nilai rapot, nilai tes tertulis dan nilai tes praktek.

Kata kunci: *Data Mining, Decision tree, Algoritma C4.5, pemilihan jurusan*

ABSTRACT

Today the development of information and communication technology is as if we can no longer weir on the side of human life. Where this technology plays a very important role and brings benefits to human progress, which is where there are so many kinds of technology. Majoring is a process of distribution and placement in the selection of teaching programs for female students. The curriculum used in high school (SMA) is currently the 2013 curriculum. As technology develops, things like this can be overcome by data grouping techniques obtained from the results of data mining. Data mining is a discipline that learns about a technique for extracting hidden information to benefit from existing data, so data mining is also often referred to as knowledge discovery from a database. Data mining has a variety of methods where this method can be adjusted to existing needs. In the problem of determining this course, the method used is by applying the Decision Tree C4.5 algorithm. Applying data mining methods with the C4.5 Decision Tree algorithm in determining the majors of high school students. C4.5 algorithm can be used as a classification method in predicting the direction of majors at MAN 1 Batam school by paying attention to the largest Gain value of three attributes such as report scores, written test scores and practical test scores.

Keywords: Data Mining, Decision tree, C4.5 algorithm, majors selection

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini tidak akan terwujud tanpa bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Ketua Program Studi Teknik Informatika Universitas Putera Batam.
3. Bapak Sunarsan Sitohang, S.kom.,M.TI selaku pembimbing Skripsi pada Program Studi Manajemen Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam.
5. Ibu Nilawati, M.Si selaku Guru Bimbingan Konseling di sekolah MAN 1 Batam.
6. Kepada Orang Tua penulis Bapak Basrizal dan Ibu Salma yang selalu mendoakan, dan Pengorbanannya baik segi moral maupun materil kepada penulis sehingga dapat menyelesaikan penulisan skripsi ini.
7. Kepada sahabat terbaik penulis Adhitya Dewanta, Reni Afrian yang telah menemani dan selalu memberikan semangat kepada penulis dalam hal memberikan ide masukan.

8. Teman seperjuangan Program Studi Teknik Informatika Universitas Putera Batam.
9. Orang – orang di lingkungan Kav. Sagukung Baru Asri dan ibu Chuswatun Chasanah selaku staff Tata Usaha di sekolah MAN 1 Batam
10. Pihak-pihak yang membantu dalam penyelesaian skripsi ini yang tidak bisa penulis ucapkan satu persatu

Semoga Tuhan membalas kebaikan dan selalu mencurahkan hidayah dan taufik-Nya, Aamiin.

Batam , Februari 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN SAMBUNG DEPAN	
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	3
1.3 Pembatasan Masalah	4
1.4 Perumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA	
2.1 Knowledge Discovery in Database	7
2.2 Data Mining	8
2.3 Metode Data Mining	12
2.4 Software Pendukung	16
2.5 Penelitian Terdahulu	17
2.6 Kerangka Pemikiran	20
BAB III METODE PENELITIAN	
3.1 Desain Penelitian	22
3.2 Pengumpulan Data	24
3.3 Operasional Variabel.....	25
3.4 Metode Perancangan Sistem	26
3.5 Lokasi Dan Jadwal Penelitian	26
BAB IV HASIL DAN PEMBAHASAN	
4.1 Analisa Data	28
4.2 Hasil Pengujian	45
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	53
5.2 Saran	53

DAFTAR PUSTAKA
RIWAYAT HIDUP
SURAT KETERANGAN PENELITIAN
LAMPIRAN

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Kerangka Pemikiran	21
Gambar 3.1 Desain Penelitian	22
Gambar 3.2 Perancangan Sistem	26
Gambar 4.1 Pohon Node 1	39
Gambar 4.2 Pohon Node 1.2	44
Gambar 4.3 Tampilan WEKA	46
Gambar 4.4 Preproces Exploler WEKA.....	46
Gambar 4.5 Pemilihan Metode Yang Digunakan.....	47
Gambar 4.6 Tampilan Classify	47
Gambar 4.7 Hasil Perhitungan Klasifikasi	48
Gambar 4.8 Clasifier Output	49
Gambar 4.9 Cara Melihat Hasil Pada Kolom Result List	50
Gambar 4.10 Classifier Tree Visualizer	51

DAFTAR TABEL

	Halaman
Tabel 3.1 Jadwal Penelitian.....	27
Tabel 4.1 Klasifikasi Range Nilai	28
Tabel 4.2 Format Data Akhir	29
Tabel 4.3 Perhitungan Node 1	38
Tabel 4.4 Atribut Nilai Tes Praktek = > 71.30	39
Tabel 4.5 Perhitungan Node 1.2.....	44

DAFTAR RUMUS

	Halaman
Rumus 2.1 <i>Gain</i>	15
Rumus 2.2 <i>Entropy</i>	15
Rumus 2.3 <i>RasioGain</i>	16
Rumus 2.4 <i>SplitInfo</i>	16

BAB I PENDAHULUAN

1.1 Latar Belakang Penelitian

Dewasa ini perkembangan teknologi informasi dan komunikasi ini seolah sudah tidak bisa kita bendung lagi dalam sisi kehidupan manusia. Dimana teknologi ini sangat berperan penting dan membawa manfaat untuk kemajuan manusia, yang dimana setiap bidang teknologi sudah sangat banyak macamnya. Salah satu contoh dari perkembangan teknologi ini seperti dibidang pengolahan data, yang dimana dahulu untuk pengolahan data masih menggunakan kertas yang disusun di lemari arsip dan membutuhkan banyak ruangan dan bnyak tempat, sementara sekarang dalam menyimpan data dapat dilakukan di dalam *database*.

Penjurusan adalah suatu proses penyaluran dan penempatan dalam pemilihan program pengajaran kepada siswa siswi. Penjurusan ini biasa dilakukan pada saat siswa siswi menginjak bangku sekolah menengah atas (SMA). Sekolah menengah atas (SMA) yaitu salah satu lembaga pendidikan yang mulai memperkenalkan jurusan dan membaginya kedalam beberapa pilihan jurusan.

Kurikulum yang dipakai pada sekolah menengah atas (SMA) pada saat ini yaitu kurikulum 2013. Pada perkembangan kurikulum 2013 ini terdapat beberapa perbedaan dengan kurikulum sebelumnya. Salah satu perbedaannya terletak pada saat menentukan jurusan pada siswa siswi sekolah menengah atas, yang dimana pemilihan jurusan dilaksanakan pada kelas sepuluh.

Dalam penentuan jurusan ini, pihak sekolah yang terutama yaitu guru bimbingan konseling (BK) dituntut sebijaksana mungkin dalam memutuskan jurusan yang tepat bagi siswa siswinya. Penentuan jurusan ini harus memperhatikan banyak faktor yang kompleks dan dilakukan secara manual yang di mana mempunyai banyak kelemahan. Data yang banyak cukup menyita waktu dan banyak menguras tenaga, serta menuntut ketelitian ekstra.

Madrasah Aliyah Negeri (MAN) 1 Batam merupakan salah satu sekolah unggulan yang menerapkan proses penjurusan bagi siswa siswinya untuk dibagi menjadi 3 jurusan, yaitu IPA, IPS, dan Agama. Proses penjurusan dengan menggunakan nilai rapot, nilai tes tertulis dan nilai tes praktek. Dalam proses penjurusan, siswa siswi di beri kesempatan memilih jurusan, baik itu jurusan IPA, IPS, ataupun Agama, sebelum nantinya di klasifikasi keputusan jurusannya menurut nilai dan minat masing-masing siswa siswi tersebut.

Permasalahan yang sering terjadi pada saat pemilihan jurusan ada bermacam-macam dari hal yang biasa saja sampai ke masalah yang penting yang nantinya bisa berdampak pada pemilihan minat pada saat di perguruan tinggi. Terkadang ada juga siswa siswi yang tidak berpikir panjang dalam menentukan jurusan yang mereka ambil, tidak mempunyai pilihan atau tujuan setelah mereka lulus sekolah menengah atas, dan bisa juga berakibat pada menurunnya prestasi belajar mereka. Untuk mengatasi masalah tersebut siswa hanya berkonsultasi secara langsung kepada guru wali kelas, guru BK maupun dengan orang tua masing-masing yang dimana dilakukan secara manual dan membutuhkan waktu yang sangat lama.

Seiring berkembangnya teknologi maka hal seperti ini bisa di atasi dengan teknik pengelompokan data yang di dapat dari hasil *data mining*. *Data mining* adalah disiplin ilmu yang mempelajari tentang sebuah teknik untuk menggali informasi tersembunyi untuk memperoleh keuntungan dari data yang ada, sehingga data mining juga sering di sebut dengan *knowledge discovery from database* (Derick, 2013). *Data mining* juga bisa untuk melakukan klasifikasi, mengestimasi dan memprediksi data yang ada dan mengubahnya menjadi informasi yang berguna.

Data mining memiliki bermacam-macam metode yang dimana metode ini dapat di sesuaikan dengan kebutuhan yang ada. Dalam masalah penentuan jurusan ini, metode yang digunakan yaitu dengan penerapan algoritma *Decision Tree C4.5*. Dengan demikian peluang yang tepat dalam pemilihan jurusan ini semakin besar.

Dari permasalahan yang ada diatas maka penulis tertarik untuk melakukan sebuah penelitian untuk menghasilkan sebuah informasi dari data mining yang bisa berguna untuk sekolah MAN 1 Batam yang berjudul: “ **IMPLEMENTASI ALGORITMA KLASIFIKASI DATA MINING DALAM MENENTUKAN PENJURUSAN SISWA SMA KOTA BATAM**”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, maka dapat di identifikasi masalahnya sebagai berikut:

1. Penentuan jurusan ini harus memperhatikan banyak faktor yang kompleks dan dilakukan secara manual yang di mana mempunyai banyak kelemahan. Data yang banyak cukup menyita waktu dan banyak menguras tenaga, serta menuntut ketelitian ekstra.
2. Terkadang ada juga siswa siswi yang tidak berpikir panjang dalam menentukan jurusan yang mereka ambil, tidak mempunyai pilihan atau tujuan setelah mereka lulus sekolah menengah atas, dan bisa juga berakibat pada menurunnya prestasi belajar mereka.
3. Berkonsultasi secara langsung kepada guru wali kelas, guru BK maupun dengan orang tua masing-masing yang dimana dilakukan secara manual dan membutuhkan waktu yang sangat lama.

1.3 Pembatasan Masalah

Dalam melakukan penelitian ini agar pembahasan menjadi terarah dan tidak melebar, maka penulis membatasi kajian mengenai masalah yang akan di bahas. Adapun pembahasan yang akan di angkat yaitu:

1. Data mining dengan menggunakan metode penerapan algoritma Decision Tree C4.5
2. Proses pengambilan data dilakukan di sekolah MAN 1 Batam.
3. Data yang digunakan yaitu data anak kelas sepuluh jurusan IPA tahun ajaran 2016.

1.4 Perumusan Masalah

Berdasarkan latar belakang diatas dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini adalah:

1. Bagaimana penerapan data mining dengan menggunakan metode *Decision Tree* C4.5 dalam menentukan penjurusan siswa MAN 1 Batam?
2. Bagaimana cara menerapkan algoritma C4.5 dengan metode pohon keputusan untuk menggambarkan proses penjurusan siswa MAN 1 Batam?
3. Bagaimana cara menentukan keputusan dalam penjurusan siswa MAN 1 Batam menggunakan metode *Decision Tree* C4.5?

1.5 Tujuan Penelitian

Setiap kegiatan yang dilakukan dan dilaksanakan dengan terencana pastinya memiliki tujuan, begitu juga dengan penelitian ini. Adapun tujuan dari penelitian ini adalah:

1. Menerapkan metode data mining dengan algoritma *Decision Tree* C4.5 dalam menentukan jurusan siswa MAN 1 Batam.
2. Menggambarkan proses penjurusan dengan menggunakan metode pohon keputusan.
3. Menentukan keputusan dari pemilihan jurusan Siswa MAN 1 Batam dengan metode *Decision Tree* C4.5.

1.6 Manfaat Penelitian

Dengan adanya penelitian ini, di harapkan memberikan manfaat-manfaat yang bisa diterima sebagai berikut:

1. Bagi Tempat Yang Diteliti

Manfaat penelitian bagi objek penelitian sebagai berikut:

1. Hasil penelitian ini diharapkan bermanfaat dan memudahkan dalam pemilihan jurusan yang di inginkan dan diminati oleh siswa.
2. Hasil penelitian ini di harapkan bisa membantu guru BK dalam pengambilan keputusan untuk mentukan jurusan bagi siswa siswinya.

2. Bagi Peneliti

Manfaat penelitian bagi peneliti sebagai berikut:

1. Penelitian ini dapat dijadikan peningkatan kemampuan dari peneliti dalam menggunakan data mining di dalam dunia pendidikan khususnya dalam pengolahan dan pengambil keputusan dari pemilihan jurusan.
2. Sebagai syarat dalam tugas akhir dan syarat pengambilan gelar Sarjana.

BAB II KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database*

Istilah *knowledge discovery in database*(KDD) dan data mining sering digunakan untuk menjabarkan proses pencarian informasi yang tersembunyi dari suatu basis data yang besar. Adapun tahapan proses KDD secara garis besar dapat dijelaskan sebagai berikut (Vulandari, 2017):

1. *Data Selection*

Seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan ke dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan *noise* dilakukan. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. *Transformation*

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Proses *Data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/evaluation*

Penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2 *Data Mining*

Data mining merupakan salah satu bagian dari Teknologi Informasi dan Komunikasi dan salah satu proses yang ada di dalam *knowledge discovery in database* (KDD). *Data mining* banyak dipakai dalam berbagai hal yang bersifat kolektif data yang besar. *Data mining* adalah suatu aktivitas yang menggunakan

suatu metode statistik, matematik, kecerdasan buatan dan penggunaan bahasa mesin yang digunakan untuk ekstraksi dan identifikasi suatu informasi yang bermanfaat yang berasal dari kumpulan data yang besar (Buaton, Sundari, & Maulita, 2016). *Data mining* merupakan suatu cara penelusuran data yang ada untuk membangun sebuah model, yang kemudian menggunakan model data tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang disimpan tersebut (Siska Haryati, 2015). *Data mining* menggunakan metode statistik yang digunakan dalam perhitungan secara empiris dari suatu database yang ada. *Data mining* merupakan suatu proses yang menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan data besar yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola (Maharani, Mesran, & Suginam, 2017). Data yang diidentifikasi dan diekstraksi oleh *data mining* bukanlah data yang berbentuk sedikit, melainkan data mining digunakan apabila data tersebut merupakan suatu data yang banyak sehingga apabila tanpa menggunakan data mining akan menyulitkan dalam pengolahannya.

Data mining adalah proses interaktif dan iteratif untuk menemukan model baru yang shahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar (*massive database*) (Hermawati, 2013).

1. Sahih: sesuatu dapat digeneralisasi untuk masa yang akan datang.
2. Baru: sesuatu yang sedang tidak diketahui.
3. Bermanfaat: digunakan untuk melakukan sebuah tindakan.
4. iteratif: memerlukan beberapa proses yang harus diulang.

5. Interaktif: sesuatu hal yang memerlukan interaksi manusia dalam suatu proses.

Data mining berisi pencarian pola yang diinginkan dalam database yang besar untuk membantu dalam proses pengambil keputusan di waktu yang akan datang. Pola ini dapat dikenali oleh perangkat tertentu yang bisa memberikan suatu analisa data yang berguna, bermanfaat dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin menggunakan perangkat pendukung keputusan yang lainnya.

Data mining sendiri adalah suatu istilah yang biasa dipakai dalam penguraian dan pencarian penemuan pengetahuan yang ada di *database*. Beberapa faktor yang mendorong kemajuan di dalam bidang *data mining* antara lain:

1. Proses pertumbuhan yang cepat di dalam kumpulan data.
2. Penyimpanan data di dalam data warehouse, sehingga seluruh perusahaan bisa memiliki akses ke dalam database yang andal.
3. Adanya suatu peningkatan akses data yang melalui navigasi web dan intranet.
4. Tekanan kompetensi bisnis yang digunakan untuk meningkatkan penguasaan pasar di dalam globalisasi ekonomi.
5. Perkembangan sebuah teknologi perangkat lunak untuk data mining.
6. Perkembangan di dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Berdasarkan penjelasan-penjelasan yang sudah di ulas sebelumnya, terdapat hal-hal penting yang terkait dengan data mining. Hal penting tersebut antara lain:

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

Tahapan proses dalam penggunaan data mining yang merupakan proses *Knowledge Discovery In Database* (KDD) dapat di uraikan sebagai berikut (Hermawati, 2013):

1. Memahami domain aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna.
2. Membuat target dat-set yang meliputi pemilihan data dan fokus pada sub-set data.
3. Pembersihan dan transformasi data meliputi eliminasi derau, *outliers*, *missing value* serta pemilihan fitur dan reduksi dimensi.
4. Penggunaan algoritma data mining yang terdiri dari asosiasi, sekuensial, klasifikasi, klasterisasi, dll.
5. Interpretasi, evaluasi dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan iterasi jika diperlukan.

Secara sistematis, ada tiga langkah utama dalam data mining (Prasetyo, 2012):

1. Eksplorasi / pemrosesan awal data
Eksplorasi awal data terdiri dari pembersihan data, normalisasi data, transformasi data, penanganan data yang salah, reduksi dimensi, pemilihan subset fitur, dan sebagainya.

2. Membangun model dan melakukan validasi terhadapnya

Membangun model dan melakukan validasi terhadapnya berarti melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang terbaik. Dalam langkah ini digunakan metode-metode seperti klasifikasi, regresi, analisis cluster, deteksi anomali, juga masuk dalam langkah eksplorasi. Akan tetapi, deteksi anomaly juga dapat digunakan sebagai algoritma utama, terutama untuk mencari data-data yang special.

3. Penerapan

Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan perkiraan/perdiksi masalah yang diinvestasikan.

2.3 Metode Data Mining

Terdapat beberapa metode yang digunakan pada *data mining* dalam mengelolah data. Metode-metode tersebut antara lain:

1. Metode *Klasifikasi Decision Tree*

Decision tree atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan (Prasetyo, 2014). Pohon keputusan merupakan sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun merepresentasikan kelompok kelas tersebut (Julianto, Yunitarini, & Sophan, 2014). Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. Karena pohon

keputusan memadukan antara ekspolrasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagi, anggota himpunan hasil menjadi mirip satu dengan yang lain (Kusrini, 2009). Pohon keputusan banyak digunakan untuk menyelesaikan kasus penentuan keputusan seperti di bidang psikologi (teori pengambilan keputusan), ilmu komputer (struktur data), bidang kedokteran (diagnosis penyakit pasien), dan sebagainya. Banyak algoritma yang bisa digunakan dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5.

2. Metode Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3. Dalam ID3, induksi decision tree hanya bisa dilakukan pada fitur bertipe katogerikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan (Prasetyo, 2014). Perbaikan yang membedakan algoritma C4.5 dengan ID3 adalah dapat menangani fitur dengan tipe numeric, melakukan pemotongan (pruning) pohon keputusan, dan penurunan (deriving) rule set. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang di induksi.

Yang menjadi hal terpenting dalam induksi pohon keputusan adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node:

1. Fitur biner

Fitur yang hanya memiliki dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang.

2. Fitur bertipe katogerikal

Untuk fitur yang bertipe katogerikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda.

3. Fitur bertipe numeric

Untuk fitur bertipe numeric, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan pengujian perbandingan ($A < v$) atau ($A > v$) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk $v_i \leq A < v_{i+1}$, untuk $i = 1, 2, \dots, k$.

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

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 1 berikut:

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

Rumus 2.1 *Gain*

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S_i| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut:

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

Rumus 2.2 *Entropy*

Keterangan:

S : himpunan kasus

A : fitur

n : jumlah partisi S

p_i : property dari S_i terhadap S

Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C4.5 adalah rasio *gain*, yang diformulasikan oleh persamaan berikut:

$$RasioGain(S,j) = \frac{Gain(s,j)}{SplitInfo(s,j)}$$

Rumus 2.3 *RasioGain*

Keterangan:

S : himpunan kasus

J : fitur ke- j

Sedangkan untuk split info dapat diperoleh dari persamaan berikut:

$$\boxed{SplitInfo(s_j) = - \sum_{i=1}^k p(V_i | s) \log_2 p(V_i | s)}$$

Rumus 2.4 *SplitInfo*

Keterangan:

K : jumlah pemecahan

2.4 *Software* Pendukung

Software yang digunakan pada penelitian ini yaitu *software* WEKA. Weka merupakan sebuah paket *tools machine learning* yang praktis (Vulandari, 2017). WEKA merupakan kepanjangan dari *Waikato Environment for Knowledge Analysis* yang dibuat di Universitas Waikato, New Zealand yang digunakan sebagai pendidikan, penelitian, dan berbagai aplikasi. WEKA dianggap bisa menyelesaikan masalah-masalah tentang data mining di dunia, khususnya klasifikasi yang mendasari pendekatan *machine learning*.

WEKA mudah dalam penggunaannya dan banyak di terapkan pada beberapa tingkat yang berbeda. Pada WEKA tersedia implementasi dari algoritma-algoritma pembelajaran *state-of-the-art* yang dapat digunakan pada dataset dari *command line*. WEKA juga mengandung *tools* yang digunakan untuk *pre-processing* data, regresi, klasifikasi, *clustering*, visualisasi dan aturan asosiasi. Pengguna bisa melakukan preprocess pada data, memasukkannya ke dalam proses skema pembelajaran, dan menganalisa classifier yang dihasilkan dan performansinya tanpa harus mengetikkan kode programnya sama sekali.

Contoh dalam penggunaan WEKA adalah penerapan sebuah metode pembelajaran yang diimplementasikan ke dalam sebuah dataset dan di analisa sehingga output dari pemrosesan tersebut digunakan untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan dibandingkan performanya untuk nantinya akan dipilih. WEKA berkembang mengikuti model *releases* Linux. Beberapa versi awal dari WEKA adalah sebagai berikut:

1. WEKA 3.0 : “versi buku” versi ini sesuai dengan deskripsi yang ada di dalam buku data mining.
2. WEKA 3.2 : “versi GUI” versi ini yang menambahkan GUI dari CLI awal.
3. WEKA 3.3 : “versi pengembangan” versi ini bersudah ditambah peningkatan.

2.5 Penelitian Terdahulu

Berikut beberapa penelitian terdahulu yang menjadi acuan penulisan dalam melakukan penelitian sehingga peneliti dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan.

1. Chandra Purnmaningsih, Ristu Saptono, dan Abdul Aziz (vol 3. No 1. Juni 2014) ISSN : 2301-7201 yang berjudul **Pemanfaatan Metode K-Means Clustering dalam Penentuan Jurusan Siswa SMA**. Penelitian ini menjelaskan dalam penentuan penjurusan siswa SMA dilakukan berdasarkan nilai akademik yang menjadi ciri dari masing-masing jurusan IPA/IPS, dengan demikian kemungkinan bagi siswa untuk memenuhi kriteria di terima di

jurusan IPA/IPS atau di tolak di jurusan tersebut. Cara untuk mempermudah penentuan jurusan ini adalah dengan cara pengelompokan (*clustering*) data siswa tersebut. Untuk mengelompokkan data siswa tersebut, metode yang digunakan adalah *K-Means Clustering*. Hasil penelitian pengujian terbaik pada preprocessing clustering K-Means IPA dengan hasil akurasi 0.905882, tingkat kesesuaian hasil prediksi dengan data sebenarnya (recall) 1, ketepatan hasil pengujian dalam memprediksi clustering (sensitivity) 0.876923, kesesuaian prediksi negatif terhadap aktual negatif (specificity) 0.714285. Sedangkan pengujian terbaik juga pada preprocessing clustering K-Means IPS didapatkan akurasi 0.905882, recall 0.714285, sensitivity 1, dan specificity 1. Hasil perbandingan clustering terbaik pada preprocessing clustering K-Means IPA dengan preprocessing clustering K-Means IPS menunjukkan bahwa tidak ada siswa yang diterima di dua jurusan IPA/IPS atau siswa ditolak di keduanya.

2. Adhika Novandya dan Isni Oktria (vol. 6 no 2 tahun 2017) ISSN : 2089-5615 dengan judul **Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi**. Dalam penelitian ini menjelaskan permasalahan dalam perkiraan cuaca wilayah Bekasi dengan penggunaan ini pengetahuan dan teknologi dalam memperkirakannya. Data yang dipakai pada penelitian ini mengacu pada World Weather Online. Penelitian ini menggunakan algoritma C4.5 menggunakan 10-fold cross validation dan dibuktikan dengan pembuatan aplikasi web untuk pengujian sehingga menghasilkan nilai akurasi sebesar 88.89%.

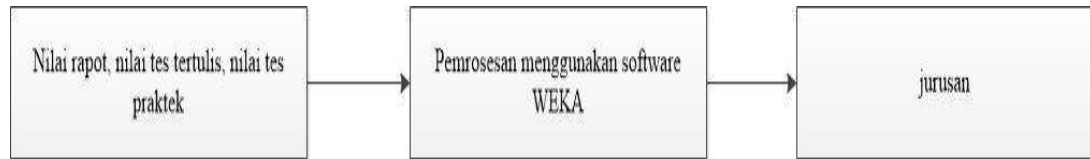
3. Kennedy Tampubolon, Hoga Saragih, dan Bobby Reza (vol 1. No 1. Oktober 2013) ISSN : 2339-210X dengan judul **Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan**. Penelitian ini menjelaskan Data Mining telah diimplementasikan ke berbagai bidang, diantaranya bidang bisnis atau perdagangan, bidang pendidikan, dan telekomunikasi. Dibidang bisnis misalnya hasil implementasi data mining menggunakan algoritma Apriori dapat membantu para pebisnis dalam kebijakan pengambilan keputusan terhadap apa yang berhubungan dengan persediaan barang. Misalnya pentingnya sistem persediaan barang di suatu Apotek dan jenis barang apa yang menjadi prioritas utama yang harus di stok untuk mengantisipasi kekosongan barang. Karena minimnya stok barang dapat berpengaruh pada pelayanan konsumen dan pendapatan Apotek. Oleh sebab itu ketersediaan berbagai jenis alat-alat kesehatan di Apotek sebagai salah satu supplier alat-alat kesehatan, mutlak untuk mendukung kelancaran penyalurannya kepada konsumen, sehingga aktivitas pelayanan konsumen berjalan dengan baik. Dalam penelitian ini membahas tentang implementasi *data mining* menggunakan algoritma apriori.
4. Katrina Sin dan Loganathan Muthu (vol.5 Juli 2015) ISSN: 2229-6956 dengan judul ***Application Of Big Data In Education Data Mining And Learning Analytics – A Literature Review***. Penelitian ini menjelaskan penggunaan sistem manajemen pembelajaran dalam pendidikan yang telah meningkat saat ini. Saat ini banyak siswa yang menggunakan ponsel, baik untuk kegiatan sehari-hari maupun mengakses konten online. Aktifitas online siswa tersebut

menghasilkan sejumlah data yang besar yang tidak digunakan terkadang terbuang sia-sia karena tidak adanya kemampuan untuk memprosesnya. Penelitian ini mengacu kepada aplikasi data yang besar dalam pendidikan dan menampilkan literatur pendidikan data mining dan analisis pembelajaran.

5. P. Thangaraju, B. Deepa dan T. Karthikeyan (vol.3 Agustus 2014) ISSN: 2278-1021 dengan judul *Comparison Of Data Mining Techniques For Forecasting Diabetes Mellitus*. Penelitian ini menjelaskan *data mining* sangat memainkan peran penting dalam industri perawatan kesehatan. *Data mining* paling sering digunakan dalam industri perawatan kesehatan untuk proses peramalan penyakit. Penelitian ini melakukan peramalan penyakit diabetes menggunakan teknik *clustering* dan menggunakan *software* WEKA.

2.6 Kerangka Pemikiran

Kerangka berfikir merupakan suatu model konseptual tentang bagai mana teori berhubungan dengan berbagai factor yang telah didefinisikan sebagai masalah dalam penelitian. Berdasarkan teori-teori yang telah di deskripsikan pada bagian sebelumnya, selanjutnya dianalisis secara kritis dan sistematis sehingga menghasilkan sistema atau kesimpulan tentang hubungan antara variable yang diteliti. Kerangka pemikiran penelitian ini di sajikan dalam bentuk diagram di bawah ini:



Gambar 2.1 Kerangka Pemikiran
Sumber: Data Peneliti(2018)

Penjelasan dari kerangka berfikir yang telah disajikan pada diagram diatas, sebagai berikut:

1. Tahap pertama meliputi memasukkan variabel penelitian sebagai sumber acuan ke dalam aplikasi WEKA untuk dilakukan pemrosesan. Variabel tersebut terdiri dari nilai rapor, nilai tes tertulis, dan nilai tes praktek.
2. Tahap selanjutnya yaitu melakukan pengolahan data menggunakan aplikasi WEKA dengan menggunakan algoritma C4.5.
3. Tahap terakhir adalah tahap keluaran dari hasil olah data dengan menggunakan aplikasi WEKA. Hasil penelitian diharapkan sesuai dengan tujuan dari penelitian.

BAB III METODE PENELITIAN

3.1 Desain Penelitian

Dalam melakukan penelitian peneliti membuat sebuah desain penelitian yang mempresentasikan seluruh kegiatan yang dilakukan peneliti dalam melakukan penelitian tentang data mining. Desain penelitian data mining akan dijelaskan melalui gambar di bawah ini.



Gambar 3.1 Desain Penelitian

Sumber: Data Peneliti (2018)

Penjelasan dari desain penelitian diatas sebagai berikut:

1. Identifikasi Masalah

Pada tahap ini, peneliti melakukan identifikasi permasalahan yang ada di lokasi penelitian. Peneliti melihat adanya permasalahan yang terjadi pada saat pemilihan jurusan yang terkadang tidak sesuai dengan minat dan bakat dari siswa-siswi tersebut. Peneliti menggunakan *data mining* dalam menyelesaikan permasalahan tersebut.

2. Studi Literatur

Setelah mengidentifikasi permasalahan dan menemukan permasalahan beserta cara menangani permasalahan tersebut, peneliti melakukan studi literatur yang berguna untuk memperdalam pemahaman peneliti tentang penelitian yang diangkat oleh peneliti. Studi literatur menggunakan sumber-sumber resmi seperti buku dan jurnal dari penelitian terdahulu yang membahas data mining.

3. Mengumpulkan Data

Peneliti melakukan pengumpulan data menggunakan metode dokumentasi dan wawancara untuk mendapatkan data nilai dari siswa-siswi MAN yang dijadikan sumber penelitian

4. Menyeleksi Data

Peneliti melakukan penyeleksian data sehingga data yang dijadikan sampel merupakan data yang sesuai dengan tujuan penelitian ini. Dan data hasil seleksi ini akan di jadikan sampel di dalam penelitian.

5. Mengolah Data Menggunakan WEKA Dengan Algoritma C4.5

Data yang telah dijadikansampel penelitian di olah kedalam aplikasi WEKA dengan menggunakan algoritma C4.5

6. Pengambilan Keputusan Pemilihan Penjurusan

Peneliti mengambil keputusan dari hasil pengolahan data sebelumnya untuk menentukan pemilihan jurusan yang sangat cocok untuk siswa-siswi tersebut.

3.2 Pengumpulan Data

Pengumpulan data merupakan sebuah teknik yang digunakan untuk mendapatkan sebuah data atau informasi dari sumber data atau narasumber yang berguna dalam penelitian. Teknik yang digunakan peneliti dalam pengumpulan data penelitian adalah melalui wawancara dan dokumentasi.

1. Wawancara

Wawancara merupakan alat pengumpulan informasi yang dilakukan dengan cara mengajukan sejumlah pertanyaan secara verbal kepada orang-orang yang dianggap bisa memberikan informasi mengenai hal- yang dianggap perlu (Simbolon, 2014). Peneliti memberikan pertanyaan seputar penelitian kepada narasumber. Di dalam penelitian ini wawancara di gunakan oleh peneliti untuk menentukan apakah di lokasi tersebut bisa dilakukan penelitian.

2 Dokumentasi

Dokumentasi merupakan pencarian, pengumpulan, penyelidikan, penyusunan, pemakaian, penyedia dokumen untuk mendapatkan berbagai keterangan serta

penerapan dan bukti (Sutarman & Noviana, 2018). Hasil dari dokumentasi adalah sebuah berkas atau file yang berisi data yang didapat dari sumber data atau tempat penelitian. Dokumentasi yang dilakukan peneliti dalam melakukan penelitian adalah dengan cara meminta file yang berisi nilai-nilai siswa-siswi MAN 1 Batam.

3.3 Operasional Variabel

Dalam penelitian ini peneliti menggunakan operasional variabel adalah nilai dari siswa-siswi MAN 1 Batam angkatan 2016. Hal-hal yang terkandung didalam nilai siswa-siswi tersebut sebagai berikut:

1. Nilai Rapot Siswa

Untuk mendukung penelitian ini, peneliti mengambil nilai rapot siswa yang di berikan oleh guru BK. Nilai siswa ini akan digunakan sebagai acuan dalam penentuan jurusan yang akan di pilih oleh siswa tersebut.

2. Nilai Tes Tertulis

Nilai tes tertulis digunakan dalam penelitian ini, penggunaan nilai tes tertulis dalam penelitian ini yaitu untuk menentukan kelayakan apakah siswa itu memiliki intelektual sesuai dengan jurusan yang telah di pilihnya.

3. Nilai Tes Praktek

Nilai tes praktek digunakan dalam penelitian ini, penggunaan nilai tes praktek ini digunakan untuk menentukan diterimanya atau tidaak di terimanya siswa.

3.4 Metode Perancangan Sistem



Gambar 3.2 Perancangan Sistem
Sumber: Data Peneliti(2018)

Peneliti mengumpulkan data berupa nilai siswa MAN dengan menggunakan metode dokumentasi dan wawancara. Setelah data terkumpul peneliti melakukan seleksi data yang bertujuan untuk memilih data yang akan dijadikan sampel penelitian dan menyeleksi beberapa data yang tidak digunakan dalam penelitian ini. Data yang sudah dijadikan sampel penelitian diolah menggunakan aplikasi WEKA dan menggunakan algoritma C4.5 untuk melakukan pengolahan data. Hasil dari proses tersebut akan digunakan untuk menentukan penjurusan yang tepat untuk siswa-siswi.

3.5 Lokasi dan jadwal penelitian

Berikut lokasi dan jadwal penelitian yang dilakukan peneliti. Lokasi dan jadwal penelitiannya sebagai berikut:

1. Lokasi Penelitian

Lokasi penelitian dilakukan di Madrasah Aliyah Negeri 1 Batam yang beralamat di Jl. Brigjen Katamso No.10, Tj. Uncang, Sagulung, Kota Batam, Kepulauan Riau

2. Jadwal Penelitian

Penelitian ini dilakukan dari bulan September 2018 sampai dengan bulan Januari 2019. Jadwal penelitian dapat dilihat pada tabel dibawah ini.

Tabel 3.1 Jadwal Penelitian

No	Kegiatan	September 2018				Oktober 2018				November 2018				Desember 2019				Januari 2019				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
1	Tahap Persiapan																					
	Pengajuan Judul	■																				
	Pengumpulan Data		■	■																		
2	Tahap Pembuatan Skripsi																					
	Pendahuluan				■	■																
	Kajian Pustaka						■	■														
	Metode penelitian								■	■	■	■										
	Hasil Penelitian												■	■	■	■	■	■	■			
	Kesimpulan dan Saran																			■	■	
3	Pengumpulan Skripsi																					■

Sumber: Data Peneliti (2018)