

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Studi

Sebelum melakukan penelitian penulis terlebih dahulu melakukan tinjauan pustaka dari penelitian lain dan penelitian tentang prediksi penjurusan sekolah menengah atas ini bukanlah penelitian yang pertama kalinya. Sebelumnya sudah ada penelitian yang membahas mengenai klasifikasi penjurusan sekolah menengah atas menggunakan metode algoritma *Naive Bayes*. Berikut adalah beberapa penelitian yang terkait dengan masalah tersebut.

Penelitian yang pertama adalah penelitian yang dilakukan oleh Claudia Clarentina Ciptohartono yang berjudul “Algoritma Klasifikasi *Naive Bayes* untuk Menilai Kelayakan Kredit” [7]. Metode yang digunakan pada penelitian ini adalah Algoritma *Naive Bayes*. Data yang digunakan adalah data dari nasabah perusahaan BCA Finance Jakarta tahun 2013. Tool yang digunakan untuk implementasi sistem adalah matlab. Penelitian ini menghasilkan bukti bahwa algoritma *Naive Bayes* bisa diterapkan dalam pemberian kelayakan kredit pada BCA Finance Jakarta. Tahap pengolahan data awal dapat menghasilkan akurasi yang tinggi dan menghasilkan akurasi akhir yang *Excellent*. Dalam penelitian kelayakan kredit ini pada data awal dengan melakukan pre-processing mendapatkan akurasi sebesar 85.57%, tetapi jika data awal dilakukan pengolahan dan dengan melakukan pre-processing dapat menghasilkan akurasi sebesar 92.53%. Jadi Algoritma *Naive Bayes* akan lebih unggul dalam pemberian penilaian kelayakan kredit jika dilakukan proses pengolahan data awal, meskipun algoritma *Naive Bayes* adalah algoritma yang sanggup menangani data yang hilang.

Penelitian yang kedua adalah penelitian yang dilakukan oleh Arief Jananto. Penelitian yang berjudul “Algoritma *Naive Bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa” [8]. Metode yang digunakan pada

penelitian ini adalah Algoritma Klasifikasi *Naive Bayes*. Data yang diperoleh adalah data yang berkaitan dengan keterangan diri dari mahasiswa UNISBANK dan data nilai semua mata kuliah mahasiswa lulusan tahun 2004-2007. Kesimpulan dari penelitian tersebut adalah sebagai berikut :

- a. Ketepatan masa studi mahasiswa diprediksi berdasarkan latar belakang sekolah sebelumnya dan data akademik, serta pribadi saat berada diperguruan tinggi.
- b. Dalam memprediksi ketepatan studi mahasiswa menggunakan data *training* dan *testing* dengan memanfaatkan fungsi prediksi dari teknik data mining yaitu menggunakan algoritma klasifikasi *Naive Bayes*.
- c. Fungsi klasifikasi yang digunakan dalam prediksi ketepatan masa studi mahasiswa ini mempunyai tingkat kesalahan sebesar 20% sampai 34%, kesalahan tersebut dapat dipengaruhi oleh jumlah data *training* dan *testing* serta tingkat konsistensi data yang digunakan.

Penelitian yang ketiga adalah penelitian yang dilakukan oleh Yuda Septian Nugroho. Penelitian yang berjudul “Data Mining Menggunakan Algoritma *Naive Bayes* untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro” [9]. Teknik yang digunakan merupakan teknik klasifikasi dan menggunakan algoritma *Naive Bayes*, tujuan dari penelitian tersebut adalah mengklasifikasi kelulusan mahasiswa Udinus Fakultas Ilmu Komputer angkatan tahun 2009. Hasil dari penelitian ini dengan menggunakan metode klasifikasi *Naive Bayes* dan dataset berupa data mahasiswa Universitas Dian Nuswantoro Fakultas Ilmu Kompur angkatan tahun 2009, mendapatkan hasil akurasi kelulusan sebesar 82.08%. Data yang kurang kompleksitas menyebabkan model dapat memprediksi cukup akurat.

Penelitian yang terakhir adalah penelitian yang dilakukan Faid Ari Prastya yang berjudul “Penerapan Algoritma C4.5 Untuk Prediksi Jurusan Siswa SMAN 3 Rembang”[10]. Algoritma yang digunakan pada penelitian ini adalah Algoritma C4.5. Data yang digunakan adalah data siswa SMAN 3

Rembang tahun ajaran 2014/2015. Berdasarkan hasil pengujian menggunakan algoritma C4.5 didapat akurasi ketepatan hasil prediksi sebesar 74,65%. Dengan jumlah true positif (tp) sebanyak 107 record, false positif (fp) sebanyak 34 record. Jumlah true negative (tn) sebanyak 108 record dan jumlah false negative (fn) sebanyak 39 record. Jadi klasifikasi data siswa baru Sekolah Menengah Atas Negeri 3 Rembang Tahun Ajaran 2014 / 2015 dengan algoritma C4.5 bisa menjadi pendukung keputusan yang digunakan oleh pihak Sekolah Menengah Atas Negeri 3 Rembang dalam proses penentuan jurusan siswa.

Tabel 2.1 : Ringkasan Penelitian

Peneliti	Judul	Metode	Tahun	Hasil Penelitian
Claudia Clarentia Ciptohartono	Algoritma Klasifikasi <i>Naive Bayes</i> Untuk Menilai Kelayakan Kredit	Algoritma <i>Naive Bayes Classifier</i>	2014	Penelitian ini menghasilkan bukti bahwa algoritma <i>Naive Bayes</i> bisa diterapkan dalam pemberian kelayakan kredit pada BCA Finance Jakarta. Tahap pengolahan data awal dapat menghasilkan akurasi yang tinggi dan menghasilkan akurasi akhir yang <i>Excellent</i> . Dalam penelitian kelayakan kredit ini pada data awal dengan melakukan pre-processing mendapatkan akurasi sebesar 85.57%, tetapi jika data awal dilakukan pengolahan dan dengan melakukan pre-processing dapat menghasilkan

				akurasi sebesar 92.53%. Jadi Algoritma <i>Naive Bayes</i> akan lebih unggul dalam pemberian penilaian kelayakan kredit jika dilakukan proses pengolahan data awal, meskipun algoritma <i>Naive Bayes</i> adalah algoritma yang sanggup menangani data yang hilang.
Arief Jananto	Algoritma <i>Naive Bayes</i> untuk Mencari Perkiraan Waktu Studi Mahasiswa	Algoritma <i>Naive Bayes</i>	2013	Dengan menguji coba data <i>training</i> dan <i>testing</i> secara random memperoleh kesalahan prediksi mencari perkiraan waktu studi mahasiswa sebesar 20% sampai 50%. Jumlah record data dan konsistensi dari data <i>training</i> dan <i>testing</i> yang digunakan dapat menyebabkan tinggi dan rendahnya tingkat kesalahan/akurasi. Hasil dari prediksi ketepatan mahasiswa angkatan tahun 2008 dengan menggunakan algoritma <i>Naive Bayes</i> ini adalah diprediksi 254 mahasiswa “Tepat Waktu” dan diprediksi 4 mahasiswa “Tidak Tepat Waktu”.

Yuda Septian Nugroho	Data Mining Menggunakan Algoritma <i>Naive bayes</i> Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro	Algoritma <i>Naive Bayes</i>	2014	Hasil dari penelitian ini dengan menggunakan metode klasifikasi <i>Naive Bayes</i> dan dataset berupa data mahasiswa Universitas Dian Nuswantoro Fakultas Ilmu Komputer angkatan tahun 2009, mendapatkan hasil akurasi kelulusan sebesar 82.08%. Data yang kurang kompleksitas menyebabkan model dapat memprediksi cukup akurat.
Faid Ari Prastya	Penerapan Algoritma C4.5 Untuk Prediksi Jurusan Siswa SMAN 3 Rembang	C4.5	2015	Dengan menggunakan Algoritma C4.5 menghasilkan akurasi sebesar 74,65%. Dengan jumlah true positif (tp) sebanyak 107 record, false positif (fp) sebanyak 34 record. Jumlah true negative (tn) sebanyak 108 record dan jumlah false negative (fn) sebanyak 39 record. Jadi klasifikasi data siswa baru SMAN 3 Rembang Tahun Ajaran 2014/2015 dengan algoritma C4.5 bisa menjadi pendukung keputusan yang digunakan oleh pihak SMAN 3 Rembang dalam proses penentuan jurusan siswa.

Berikut adalah perbedaan penelitian yang penulis teliti dengan penelitian sebelumnya adalah :

- a. Data yang digunakan oleh penulis dalam penelitian ini adalah data siswa SMA 1 Kajen tahun ajaran 2015/2016 yang menggunakan kurikulum 2013.
- b. Data yang dianalisa dan diproses oleh penulis akan dijadikan untuk klasifikasi penjurusan siswa di SMA 1 Kajen dengan parameter yang digunakan berupa data nilai Ujian Nasional IPA dan Matematika Sekolah Menengah Pertama (SMP), nilai rapot IPA dan Matematika Sekolah Menengah Pertama (SMP) selama 5 semester, nilai kualitas, nilai IQ, minat jurusan dan jurusan.
- c. Metode yang digunakan penulis dalam penelitian ini adalah algoritma klasifikasi *Naive Bayes*, yang digunakan untuk mengklasifikasi penjurusan di SMA 1 Kajen. Untuk melihat tingkat keakurasian data yang diteliti penulis menggunakan *tools* Rapidminer dan Matlab untuk mengolah data dalam klasifikasi penjurusan siswa SMA 1 Kajen.

2.2 Tinjauan Pustaka

2.2.1 Penjurusan Siswa

Peminatan siswa adalah proses dalam pengambilan keputusan dan pilihan oleh siswa dalam bidang keahlian yang didasarkan atas pemahaman potensi diri dan peluang yang ada. Dalam konteks ini, bimbingan dan konseling membantu siswa untuk memahami diri, menerima diri, mengarahkan diri, mengambil keputusan diri, merealisasikan keputusannya secara bertanggung jawab. Implementasi kurikulum 2013 akan dapat menimbulkan masalah bagi siswa SMA/MA dan SMK yang tidak mampu dalam menetapkan pilihan peminatan, baik permintaan kelompok mata pelajaran, peminatan lintas mata pelajaran maupun pendalaman mata

pelajaran secara tepat, sehingga akan menimbulkan kesulitan dan kecenderungan gagal dalam belajar [11].

Penetapan pilihan peminatan kelompok mata pelajaran, permintaan lintas mata pelajaran, dan pemintan pedalaman materi mata pelajaran harus sesuai dengan kecerdasan, bakat, minat dan kecenderungan pilihan masing-masing siswa agar proses belajar berjalan dengan baik dan berhasil dalam belajar. Oleh karena itu peminatan sangat diperlukan bagi siswa agar dapat menetapkan pilihan peminatan sesuai kemampuan potensi yang dimilikinya dan kemungkinan berhasil dalam belajar.

2.2.2 Data Mining

Data mining merupakan proses penggunaan teknik statistik, matematika, *artificial intelligence* (kecerdasan buatan) dan *machine learning* yang digunakan untuk mengekstrak serta mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [12].

Berikut merupakan karakteristik dari data mining :

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang belum diketahui sebelumnya.
- b. Data mining dapat menggunakan data yang sangat besar. Data yang sangat besar biasanya digunakan untuk membuat hasil data mining yang lebih terpercaya.
- c. Data mining berguna untuk membuat suatu keputusan yang kritis, terutam dalam hal strategi.

Data mining mempunyai beberapa teknik berdasarkan tugas yang dilakukan dan setiap teknik mempunyai algoritma masing-

masing. Berikut adalah teknik dalam data mining yang terbagi menjadi enam kategori [9] :

a. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan *trend* yang tersembunyi dalam data.

b. Estimasi

Teknik estimasi ini mirip dengan teknik kasifikasi, kecuali variabel tujuan lebih kearah numerik dari pada kategori.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Namun prediksi hasilnya menunjukkan sesuatu yang belum pernah terjadi atau mungkin terjadi dimasa depan.

d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Contoh, kita akan mengklasifikasi penghasilan dalam tiga kelas, yaitu penghasilan tinggi, penghasilan sedang dan penghasilan rendah.

e. Klastering

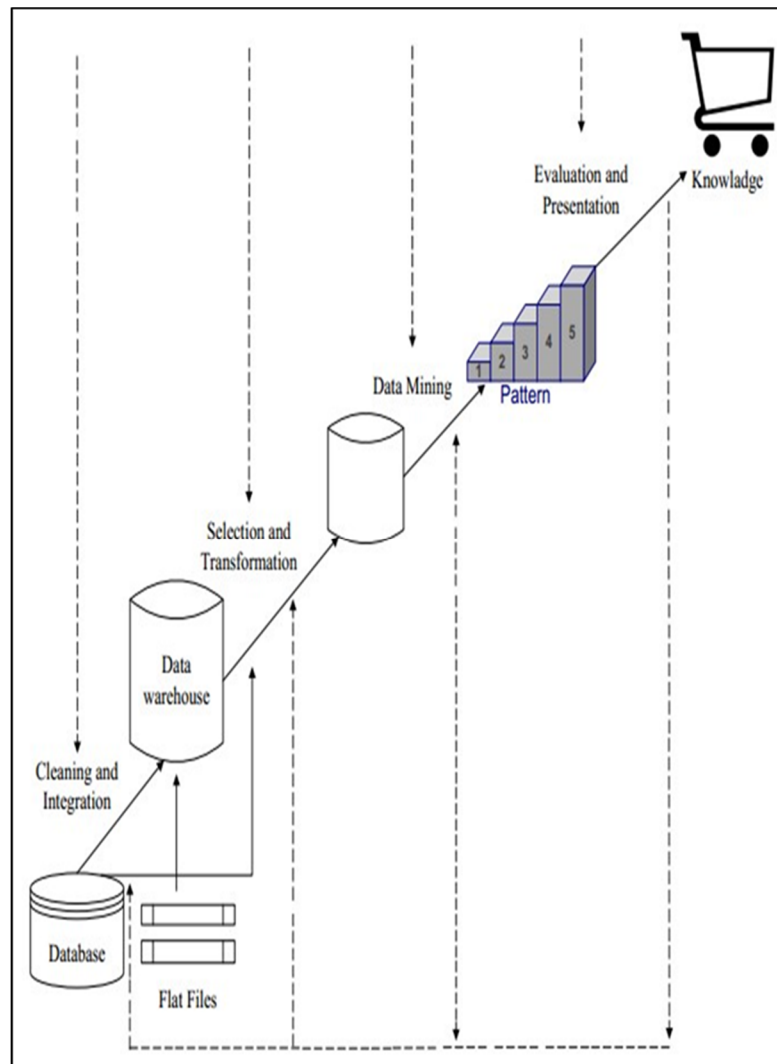
Klastering lebih ke arah pengelompokan *record*, pengamatan dan kasus dalam kelas yang memiliki kemiripan.

f. Asosiasi

Asosiasi mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

2.2.2.1 Tahap-tahap Data Mining

Data mining dapat dibagi menjadi beberapa tahap, Tahap data mining dilakukan sebagai suatu rangkaian proses. Tahap-tahap tersebut bersifat interaktif dimana pemakai terlibat langsung atau dengan perantaraan knowledge base [13]. Berikut adalah tahap-tahap dalam data mining :



Gambar 2.1 : Tahapan Data Mining

Keterangan:

1. Pembersihan Data

Pembersihan data dilakukan untuk menghilangkan noise dan data yang tidak konsisten atau tidak relevan. Sering kali data yang diperoleh dari *database* suatu perusahaan maupun diperoleh melalui hasil eksperimen, memiliki isian-isian data yang tidak lengkap seperti data yang hilang, tidak valid dan atau hanya salah ketik. Selain itu terdapat atribut-atribut data yang tidak relevan dengan

hipotesa data mining yang dimiliki. Data-data yang tidak relevan lebih baik dibuang dan tidak digunakan dalam proses. Pembersihan data akan mempengaruhi performansi dari teknik data mining. Karena data yang diproses akan berkurang jumlah dan kompleksitasnya.

2. Integrasi Data

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database yang baru. Data yang digunakan dalam data mining tidak hanya didapatkan dari satu database namun juga didapatkan dari beberapa database atau *file teks*. Integrasi data dapat dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomer pelanggan dan lainnya. Dalam melakukan integrasi data harus dilakukan dengan cermat agar hasil tidak menyimpang dan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk namun menggabungkan produk dari kategori yang berbeda, maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam melakukan integrasi data diperlukan transformasi dan pembersihan data dikarenakan sering kali data dari dua database berbeda cara penulisannya dan bahkan data yang ada di satu database tidak ada di database lainnya.

3. Seleksi Data

Tidak semua data yang ada dalam database dipakai, oleh sebab itu hanya data yang sesuai akan diambil untuk dianalisa. Sebagai contoh sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, tetapi cukup dengan id pelanggan.

4. Transformasi Data

Beberapa teknik data mining memerlukan format data yang khusus sebelum dapat diaplikasikan. Sebelum diproses dalam data mining data akan diubah dan digabungkan ke dalam format yang sesuai. Beberapa metode data mining memerlukan format data khusus agar dapat diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Dalam proses ini sering disebut transformasi data. Transformasi dan pemilihan data ini menentukan kualitas dari hasil data mining nantinya, karena ada beberapa karakteristik teknik data mining tertentu yang tergantung pada tahap ini.

5. Proses Mining

Tahap ini merupakan proses utama saat metode diterapkan untuk menemukan informasi atau pengetahuan yang berharga dan tersembunyi dari data.

6. Evaluasi Pola

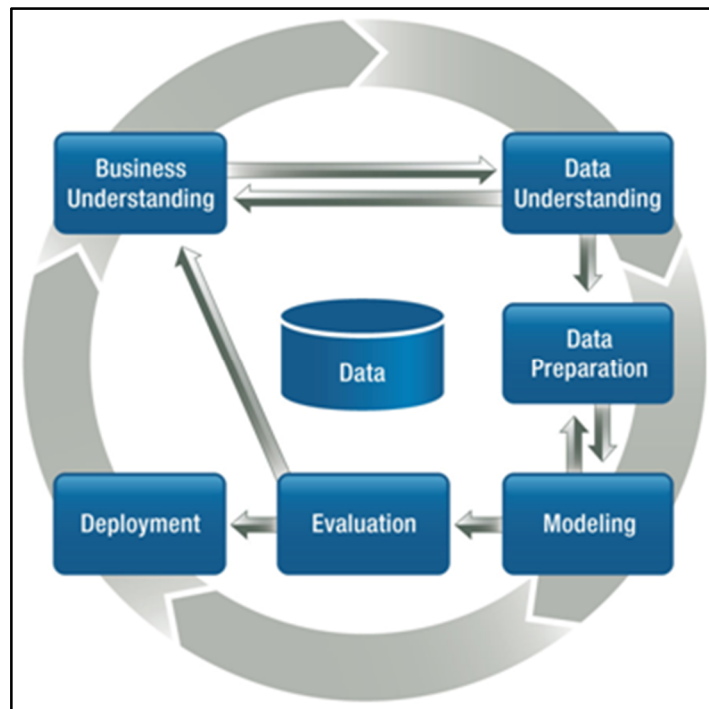
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas ataupun model prediksi akan dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Namun bila hasil yang didapatkan tidak sesuai hipotesa maka akan dilakukan beberapa alternatif, seperti menjadikan umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain dan menerima hasil ini sebagai hasil yang diluar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan

Tahap yang terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Adakalanya hal ini harus melibatkan orang-orang yang tidak memahami tentang data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang mampu dipahami semua orang dalam satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga dapat membantu mengkomunikasikan hasil dari data mining

2.2.3 CRISP-DM (*Cross Industry Standard Process for Data Mining*)

CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan sudah ditetapkan sebagai proses standar dalam data mining yang bisa diaplikasikan diberbagai sektor industri. Berikut merupakan gambar dari proses siklus hidup pengembangan dari CRISP-DM [12] :



Gambar 2.2 : Siklus Hidup CRISP-DM

Berikut merupakan tahap dari siklus pengembangan data mining CRISP-DM :

1. *Business Understanding*

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis. Kemudian pengetahuan ini diterjemahkan kedalam pendefinisian masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding*

Pada tahap ini dimulai dengan pengumpulan data kemudian dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, dan untuk mendeteksi bagian menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

3. *Data Preparation*

Tahap ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diperoleh pada *modeling*) dari data mentah. *Data preparation* ini dapat diulang beberapa kali. Dalam tahap ini juga mencakup pemilihan tabel, *record*, dan atribut-atribut data, termasuk juga proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap *modeling*.

4. *Modeling*

Pada tahap ini dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Dipihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga memungkinkan pada tahap ini dapat kembali ke tahap sebelumnya.

5. *Evaluation*

Pada tahap *evaluation* ini model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Dalam tahap ini sebelum model digunakan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal yaitu *Business Understanding* akan dilakukan evaluasi terhadap keefektifan dan kualitas model. Kunci pada tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan.

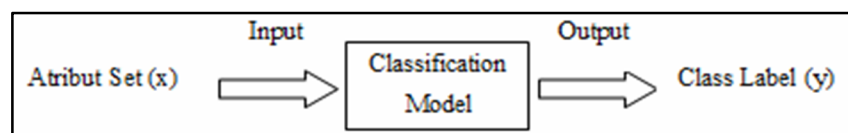
6. *Deployment*

Pada tahap ini pengetahuan dan informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus, sehingga dapat digunakan oleh pengguna. Tahap *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses

data mining yang berulang dalam perusahaan. Dalam banyak kasus, tahap *deployment* melibatkan konsumen, disamping analisis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.

2.2.4 Klasifikasi

Klasifikasi merupakan salah satu tugas yang penting dalam data mining. mengorganisasikan serta mengelompokkan data ke dalam kelas-kelas yang berbeda merupakan tujuan utama dari klasifikasi. Definisi pengklasifikasian adalah sebuah fungsi yang bersifat prediksi dan menggolongkan data item tertentu ke dalam sebuah kelas. Sebuah pengklasifikasian dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan dan dikenal ciri-cirinya sebelumnya. Performa pengklasifikasian biasanya diukur dengan ketepatan [14].



Gambar 2.3 : Blok Diagram Model Klasifikasi

2.2.5 Algoritma Naive Bayes

Algoritma *Naive Bayes* adalah salah satu algoritma yang terdapat pada teknik data mining klasifikasi. *Naive bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris yaitu Thomas bayes, *Naive Bayes* memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal dengan Teorema Bayes. Teorema

tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya [15].

Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

$P(H)$: Probabilitas hipotesis H (prior probability)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Adapun alur dari metode *Naive Bayes* adalah sebagai berikut :

1. Baca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai *mean* dan standar deviasi dari masing – masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas.

2.2.6 Pengujian *Cross Validation*

Validation merupakan proses untuk mengevaluasi keakurasian prediksi dari model data mining. Validasi digunakan untuk mendapatkan prediksi menggunakan model yang sudah ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model [16].

Cross Validation merupakan teknik validasi dengan membagi data secara acak ke dalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dalam *Cross Validation*, jumlah tetap lipatan atau partisi dari data ditentukan sendiri. Cara standar untuk memprediksi *error rate* dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan *tenfold cross validation*.

2.2.7 *Confusion Matrix*

Confusion Matrix matrix memberikan keputusan yang diperoleh dalam *training* dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi [7].

Tabel 2.2 : Confusion Matrix untuk 2 Kelas

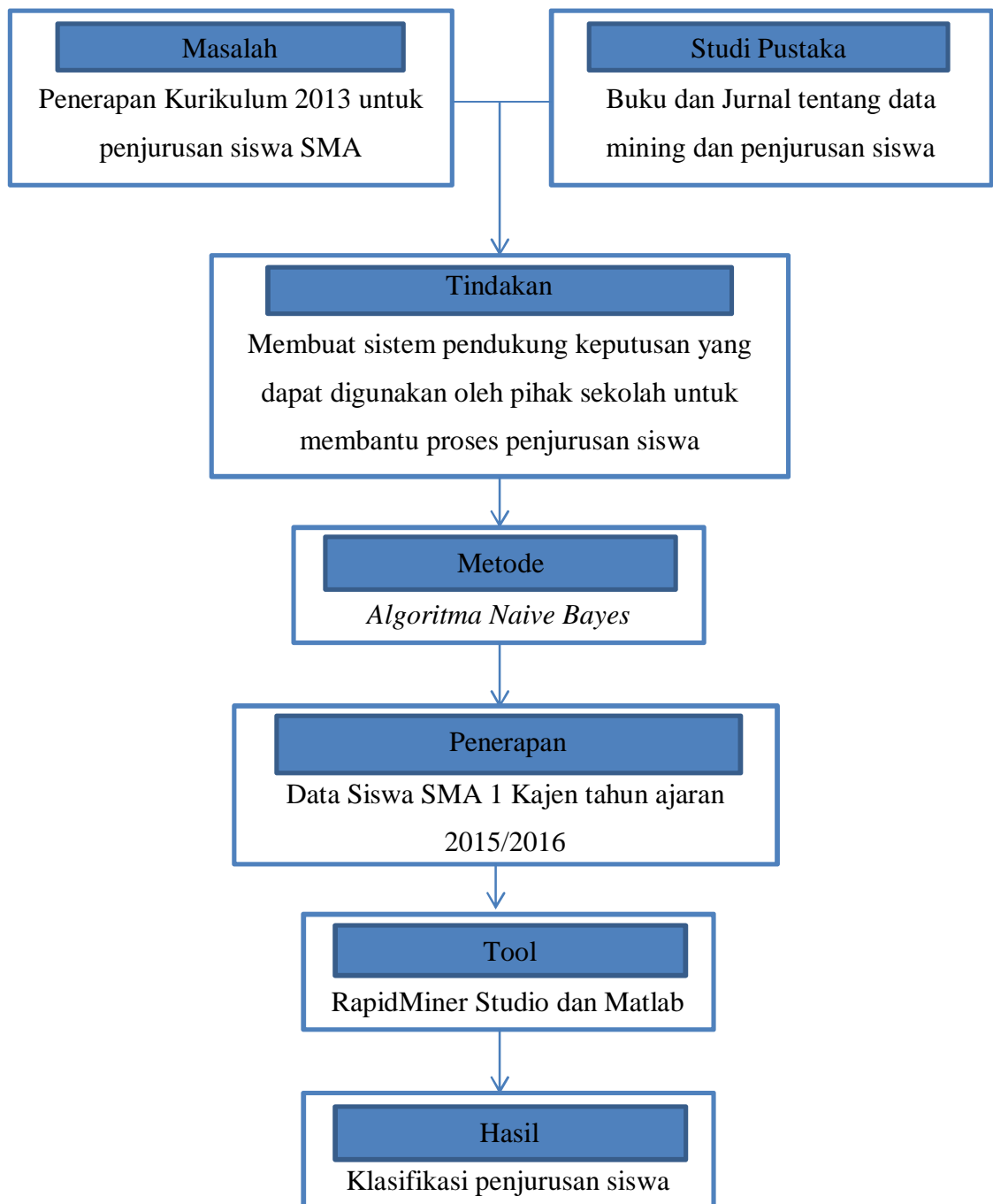
Classification	Predicted class	
	Class = Yes	Class = No
Class=Yes	a (true positive-TP)	b (false negative-FN)
Class=No	c (false positive-FP)	d (true negative-TN)

Berikut merupakan persamaan model *confusion matrix* untuk menghitung akurasi.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

2.3 Kerangka Pemikiran

Penulis perlu membuat gambaran singkat sebagai alur penyusunan laporan ini dengan kerangka pemikiran sebagai berikut:



Gambar 2.4 : Kerangka Pemikiran.