

**PENERAPAN *DATA MINING* UNTUK
MENAMPILKAN INFORMASI NILAI SISWA PADA
SMK NEGERI 5 BATAM DENGAN MENGGUNAKAN
METODE *NAÏVE BAYES CLASSIFIER***

SKRIPSI



**Oleh :
Helmaida Marbun
150210114**

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

**PENERAPAN *DATA MINING* UNTUK
MENAMPILKAN INFORMASI NILAI SISWA PADA
SMK NEGERI 5 BATAM DENGAN MENGGUNAKAN
METODE *NAÏVE BAYES CLASSIFIER***

SKRIPSI

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



**Oleh :
Helmaida Marbun
150210114**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
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 arahan 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 di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku diperguruan tinggi.

Batam, 09 Agustus 2019

Yang membuat pernyataan,

Helmaida Marbun
150210114

**PENERAPAN *DATA MINING* UNTUK MENAMPILKAN
INFORMASI NILAI SISWA PADA SMK NEGERI 5 BATAM
DENGAN MENGGUNAKAN METODE *NAÏVE BAYES*
*CALSSIFIER***

Oleh
Helmaida Marbun
150210114

SKRIPSI

Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana

Telah disetujui oleh Pembimbing pada tanggal
seperti tertera dibawah ini

Batam, 2019

Koko Handoko, S.Kom., M.Kom.

ABSTRAK

Persoalan tentang nilai siswa dianggap penting sebagai tolak ukur untuk mengetahui tingkat prestasi siswa. Prestasi siswa merupakan hasil pengukuran terhadap peserta didik setelah mengikuti proses pembelajaran dalam periode tertentu. Pengelolaan nilai yang efektif dengan dengan aplikasi pendukung sangat lah membantu dalam menghasilkan perhitungan yang akurat. Dengan begitu penggunaan metode *naïve bayes* dianggap mampu menjawab menyelesaikan permasalahan dalam bidang teknologi informasi supaya setiap data terintegrasi untuk persiapan data pembelajaran dan berguna sebagai analis prediktif. Tujuan penelitian ini adalah untuk mengetahui penerapan data *mining* siswa yang berprestasi dan tidak berprestasi, dari penilaian selama per semester serta untuk mengetahui perhitungan dari data siswa dengan menggunakan metode *naïve bayes classifier* sehingga perhitungan lebih cepat dan baik. Berdasarkan pemaparan tersebut permasalahan penelitian ini adalah bagaimana penerapan *data mining* untuk mengetahui siswa teknik informatikan yang berprestasi dan tidak berprestasi menggunakan metode *naïve bayes classifier*. Metode penelitian yang digunakan adalah dengan cara melakukan observasi dan wawancara terhadap narasumber dari SMK 5 Batam. Aplikasi pembantu dalam penelitian ini yang digunakan adalah *RapidMiner* untuk menghitung hasil nilai siswa. Algoritma yang digunakan dalam penelitian ini berupa sebuah keputusan terhadap nilai siswa. Hasil penelitian ini adalah dengan menerapkan *data mining* pada nilai SMK 5 Batam untuk pengolahan dan mencari nilai siswa lebih mudah. Dalam penelitian ini telah memasukkan 43 data *testing* nilai siswa dapat dengan mudah diketahui hasil pengolahan nya. Saran dari penelitian ini adalah sebaiknya menggunakan metode *naïve bayes* diutamakan memiliki *data training* yang banyak untuk meningkatkan akurasi atau tidaknya pengujian. Hasil penelitian ini dapat digunakan sebagai bahan referensi untuk penelitian tentang *data mining* dan metode *naïve bayes* kedepannya.

Kata Kunci : *Data Mining*, Nilai siswa, Metode *Naïve Bayes*, *RapidMiner*

ABSTRACT

The issue of student grades is considered important as a benchmark to determine the level of student achievement. Student achievement is the result of measurement of students after following the learning process in a certain period. Effective value management with supporting applications is very helpful in producing accurate calculations. That way the use of the naïve bayes method is considered capable of answering problems in the field of information technology so that each data is integrated to prepare learning data and is useful as a predictive analyst. The purpose of this study is to determine the application of data mining of students who excel and not perform, from the assessment during the semester and to determine the calculation of student data using the naïve bayes classifier method so that calculations are faster and better. Based on the explanation, the problem of this research is how the application of data mining to find out information engineering students who excel and not perform using the naïve bayes classifier method. The research method used was by conducting observations and interviews with speakers from SMK 5 Batam. The helper application in this study that was used was RapidMiner to calculate the results of student grades. The algorithm used in this study is in the form of a decision on student grades. The results of this study are by applying data mining on the value of SMK 5 Batam for processing and finding students' grades easier. In this research has included 43 data testing student grades can easily know the results of its processing. The suggestion from this research is that it is better to use the naïve bayes method, preferably having a lot of training data to improve the accuracy of the test. The results of this study can be used as reference material for research on data mining and the naïve Bayes method going forward.

Keywords : *Data Mining, Student Scores, Naïve Bayes Method, RapidMiner*

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat 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 takkan 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. Bapak Andi Maslan, S.T., M.Si selaku ketua Program Studi Teknik Informatika Universitas Putera Batam.
3. Bapak Koko Handoko, S.Kom., M.Kom selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Dosen dan staff Universitas Putera Batam.
5. Bapak Agus Sahrir, M.Pd selaku kepala sekola SMK Negeri 5 Batam.
6. Kepada Orangtua penulis Bapak Normal Marbun dan Mama Rotua Nainggolan yang selalu mendoakan, dan pengorbanannya baik segi moral maupun materi kepada penulis sehingga dapat menyelesaikan penulisan skripsi ini.

7. Kepada Saudara kandung dan teman-teman 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 dilingkungan perumahan bida ayu terlebih jemaat dan Bapak Ibu Gembala GPDI Eben Haezer Bida Ayu Batam.
10. Pihak-pihak yang membantu dalam penyelesaian skripsi ini yang tidak bisa penulis ucapkan satu per satu.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan berkatNya kepada kita semua, Amin.

Batam, 09 Agustus 2019

Penulis

DAFTAR ISI

	Hal
PERNYATAAN	iii
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	2
1.3 Pembatasan Masalah	3
1.4 Perumusan Masalah.....	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	5
BAB II KAJIAN PUSTAKA	6
2.1 <i>Knowledge Discovery in Database (KDD)</i>	6
2.1.1 Proses <i>Knowledge Discovery in Database</i>	6
2.2 <i>Data Mining</i>	8
2.2.1 Pengelompokan <i>Data Mining</i>	9
2.3 Metode <i>Data Mining</i>	11
2.3.1 Metode Klasifikasi <i>Decision Tree</i>	11
2.3.2 Metode Klasifikasi Teorema Bayes	12
2.4 SMK Negeri 5 Batam	13
2.5 <i>Software</i> Pendukung.....	13
2.6 Penelitian Terdahulu.....	14
2.7 Kerangka Pemikiran	16
BAB III METODE PENELITIAN	17
3.1 Desain Penelitian	17
3.2 Teknik Pengumpulan Data	19
3.2.1 Observasi.....	19
3.2.2 Wawancara.....	19
3.2.3 Studi Pustaka.....	19
3.3 Operasional Variabel	19
3.4 Metode Analisis dan Rancangan Sistem	20
3.5 Lokasi dan Jadwal Penelitian	21
BAB IV HASIL DAN PEMBAHASAN	22
4.1 Arsitektur Sistem	22

4.1.1	Seleksi Data dan Eliminasi Atribut.....	27
4.1.2	Pengujian Data (Data <i>Training</i>).....	35
4.2	Hasil Pengujian.....	76
BAB V SIMPULAN DAN SARAN.....		82
5.1	Simpulan.....	82
5.2	Saran.....	82

DAFTAR PUSTAKA

RIWAYAT HIDUP

SURAT KETERANGAN PENELITIAN

SURAT BALASAN

LAMPIRAN

DAFTAR GAMBAR

	Hal
Gambar 3.1 Desain Penelitian	17
Gambar 3.2 Metode Perancangan Sistem.....	20
Gambar 4.1 Aplikasi <i>Rapidminer</i>	76
Gambar 4.2 Membuat Data <i>Training</i>	77
Gambar 4.3 Data <i>Import Wizard</i>	77
Gambar 4.4 Membuat Data <i>Testing</i>	78
Gambar 4.5 Tampilan Data <i>Testing</i> Dan Data <i>Training</i>	79
Gambar 4.6 Pembuatan <i>Naïve Bayes</i>	79
Gambar 4.7 Koneksi <i>Process</i>	80
Gambar 4.8 Hasil <i>Run</i>	81
Gambar 4.9 Hasil <i>Run</i>	81

DAFTAR TABEL

	Hal
Tabel 3.1 Jadwal Penelitian.....	21
Tabel 4.1 Data Nilai Siswa.....	22
Tabel 4.2 Data Nilai Siswa.....	24
Tabel 4.3 Data Nilai Siswa.....	26
Tabel 4.4 Data Training.....	28
Tabel 4.5 Data Nilai Siswa.....	30
Tabel 4.6 Data Nilai Siswa.....	32
Tabel 4.7 Data <i>Testing</i>	34
Tabel 4.8 Data <i>Testing</i> 1.....	35
Tabel 4.9 Hasil Data <i>Testing</i> 1	37
Tabel 4.10 Data <i>Testing</i> 2.....	38
Tabel 4.11 Hasil Data <i>Testing</i> 2	40
Tabel 4.12 Data <i>Testing</i> 3.....	40
Tabel 4.13 Hasil Data <i>Testing</i> 3	42
Tabel 4.14 Data <i>Testing</i> 4.....	42
Tabel 4.15 Hasil Data <i>Testing</i> 4	44
Tabel 4.16 Data <i>Testing</i> 5.....	44
Tabel 4.17 Hasil Data <i>Testing</i> 5	46
Tabel 4.18 Data <i>Testing</i> 6.....	46
Tabel 4.19 Hasil Data <i>Testing</i> 6	48
Tabel 4.20 Data <i>Testing</i> 7.....	49
Tabel 4.21 Hasil Data <i>Testing</i> 7	51
Tabel 4.22 Data <i>Testing</i> 8.....	51
Tabel 4.23 Hasil Data <i>Testing</i> 8	52
Tabel 4.24 Data <i>Testing</i> 9.....	53
Tabel 4.25 Hasil Data <i>Testing</i> 9	55
Tabel 4.26 Data <i>Testing</i> 10.....	56
Tabel 4.27 Hasil Data <i>Testing</i> 10	57
Tabel 4.28 Data <i>Testing</i> 11.....	57
Tabel 4.29 Hasil Data <i>Testing</i> 11	59
Tabel 4.30 Data <i>Testing</i> 12.....	59
Tabel 4.31 Hasil Data <i>Testing</i> 12	61
Tabel 4.32 Data <i>Testing</i> 13.....	62
Tabel 4.33 Hasil Data <i>Testing</i> 13	64
Tabel 4.34 Data <i>Testing</i> 14.....	64
Tabel 4.35 Hasil Data <i>Testing</i> 14	66
Tabel 4.36 Data <i>Testing</i> 15.....	66
Tabel 4.37 Hasil Data <i>Testing</i> 15	68
Tabel 4.38 Data <i>Testing</i> 16.....	68
Tabel 4.39 Hasil Data <i>Testing</i> 16	71

DAFTAR RUMUS

	Hal
Rumus 2.1 Probabilitas Bayes.....	12
Rumus 2.2 Teorema Bayes.....	12
Rumus 4.1 Teorema Bayes.....	35

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Seiring berkembangnya kemajuan teknologi telah membuat kita semakin mudah dalam melakukan kegiatan dalam kehidupan sehari-hari. Teknologi sekarang sangat membantu kita dalam menyelesaikan masalah. Diantara beberapa persoalan yang ada yaitu tentang nilai siswa. Informasi nilai siswa ini akan diproses dan dijadikan sebagai informasi yang efektif dan akurat dengan menggunakan *data mining* dan metode *naïve bayes*. Teknologi informasi adalah salah satu yang sangat penting untuk mengelola nilai tersebut yang dapat digunakan untuk meningkatkan keunggulan dalam bersaing dan teknologi dapat digunakan untuk mengolah dan menyebarkan informasi.

Sekolah SMK Negeri 5 Batam adalah salah satu sekolah menengah kejuruan di kota Batam. Sekolah ini mempunyai banyak jurusan salah satunya adalah teknik informatika. Salah satu yang terpenting di dalam sebuah sekolah adalah nilai. Baik tengah semester dan akhir semester. Nilai siswa yang terkumpul tersebut akan menjadi sebuah informasi nilai siswa yang berprestasi dalam sekolah SMK negeri 5 Batam. Sesuai hasil nilai yang di dapat oleh siswa akan ditentukan siapakah yang mendapat nilai berprestasi dan tidak berprestasi dalam kelas tersebut. Di lihat dari hasil penelitian sebelumnya sekolah masih kurang efektif dan aplikasi yang dipakai belum memadai dibidang teknologi

Sehingga peneliti akan menggunakan algoritma *naïve bayes* sehingga efisien dalam menghasilkan perhitungan yang dapat membantu menentukan siswa yang berprestasi dan siswa yang tidak berprestasi.

Data mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen *database*. *Data mining* digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya *data mining* maka akan di dapatkan suatu permata berupa pengetahuan di dalam kumpulan data-data yang banyak jumlahnya (Yanto & Khoiriah, 2018). Penelitian ini menggunakan metode *naive bayes* yang merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas *Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai teorema bayes. Teorema tersebut dikombinasikan dengan *naïve* dimana nilai siswa yang akan diasumsikan kondisi antar atribut saling bebas. Nilai siswa tersebut yang akan diklasifikasi menjadi sebuah data besar (Dicky Nofriansyah, 2016).

Memandang permasalahan diatas mendorong peneliti untuk membuat sebuah aplikasi *data mining* dengan metode *naïve bayes classifier* yang dapat membantu sekolah SMK Negeri 5 menyelesaikan permasalahan dalam bidang teknologi informasi khususnya tingkat prestasi siswa. Hal ini bertujuan untuk memajukan sekolah dan semakin tahunnya akan meningkat.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah diatas peneliti dapat mengidentifikasi masalah-masalah berdasarkan latar belakang masalah yang telah dijelaskan sebelumnya yaitu :

1. Informasi nilai siswa pada SMK Negeri 5 belum menggunakan *data mining* sehingga mendapatkan hasil nilai siswa tersebut menjadi lebih sulit.
2. Penggunaan perhitungan manual dalam melakukan informasi nilai siswa mengurangi efesien waktu dalam pengolahan nilainya.
3. Kesulitan dalam mengolah informasi nilai siswa di SMK negeri 5, sehingga pengelolaan nilai siswa kurang efektif.

1.3 Pembatasan Masalah

Dari rumusan masalah di atas, perlu adanya suatu batasan masalah. Maka peneliti membatasi masalah yang ada pada penelitian ini adalah :

1. Penelitian ini menggunakan data nilai siswa SMK Negeri 5 Teknik Komputer dan Jaringan dalam kurun waktu 5 tahun terakhir.
2. Algoritma *naïve bayes classifier* sebagai metode yang digunakan dalam penelitian ini.
3. Aplikasi *data mining* yang digunakan untuk pengujian data nilai siswa yang berprestasi dan tidak berprestasi dengan menggunakan aplikasi *RapidMiner*.

1.4 Perumusan Masalah

Sebagaimana telah dibahas latar belakang diatas serta mengetahui batasan masalah dari judul yang diangkat, maka rumusan masalah yang ditemui antara lain :

1. Bagaimana penerapan *Data Mining* untuk mengetahui siswa Teknik Komputer dan Jaringan yang berprestasi dan tidak berprestasi menggunakan metode *naïve bayes classifier*?
2. Bagaimana cara perhitungan secara manual dari data nilai siswa dengan *software RapidMiner*?
3. Bagaimana manfaat dan pengaruh *data mining* dalam sekolah SMK Negeri 5 Batam?

1.5 Tujuan Penelitian

Sebagaimana rumusan permasalahan yang terjadi, maka tujuan yang akan di capai antara lain :

1. Untuk mengetahui penerapan *data mining* siswa yang berprestasi dan tidak berprestasi, dari penilaian selama per semester.
2. Untuk mengetahui perhitungan dari data siswa dengan menggunakan metode *naïve bayes classifier* sehingga perhitungan lebih cepat dan baik.
3. Untuk mengetahui manfaat dan pengaruh nilai siswa teknik informatika di sekolah SMK Negeri 5, yang bermanfaat bagi sekolah tersebut.

1.6 Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan pada penelitian ini adalah :

1. Menambah dan menguatkan teori yang telah ada tentang *data mining*.
2. Sebagai tambahan referensi untuk membantu sekolah SMK Negeri 5 Batam yang membutuhkan informasi tentang siswa yang berprestasi setiap tahunnya.
3. Diharapkan hasil penelitian ini berguna bagi Universitas Putera Batam sebagai bahan bacaan dan referensi bagi mahasiswa lainnya

BAB II

KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database merupakan metode untuk memperoleh pengetahuan dari *database* yang ada. Dalam *database* terdapat tabel-tabel yang saling berhubungan/berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan (Yuli, 2017).

Istilah *data mining* dan *Knowledge Discovery in Database (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses *Knowledge Discovery in Database (KDD)* adalah *data mining* (Dicky Nofriansyah, 2016). KDD sering disebut juga sebagai kegiatan yang meliputi pengumpulan, pemakaian data historis untuk mendapatkan keteraturan, pola hubungan dalam set data berukuran besar.

2.1.1 Proses *Knowledge Discovery in Database*

Proses *Knowledge Discovery in Database (KDD)* secara garis besar dapat dijelaskan sebagai berikut (Dicky Nofriansyah, 2016) :

1. Pembersihan Data (*Data Cleaning*)

Pembersih data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik

dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi Data

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file text*. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya.

3. Seleksi Data (*Data Selection*)

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

4. Transformasi Data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

5. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola

Untuk mengidentifikasi pola-pola *knowledge based* yang ditemukan. Dalam tahap ini hasil dalam teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila hasil yang dicapai tidak sesuai hipotesa, ada beberapa alternative yang dapat diambil seperti menjadikan untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

2.2 *Data Mining*

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis

data yang besar sehingga sering disebut *Knowledge Discovery Database (KDD)* (Vulandari, 2017).

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. Suatu rangkaian proses untuk menggali nilai tambah dari sekumpulan data yang berupa pengetahuan yang selama ini tidak secara manual. *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan, *machine learning*, *statistic* dan basis data (Handoko, 2016).

2.2.1 Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Yuli, 2017) :

1. Deskripsi (*Descriptioni*)

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun

menggunakan record lengkap yang menyediakan nilai dari variable target sebagai nilai prediksi. Pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

3. *Prediction* (Prediksi)

Prediksi hampir sama dengan klasifikasi, dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah :

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

4. *Classification* (Klasifikasi)

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

5. *Clustering* (Pengklusteran)

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

6. *Association* (Asosiasi)

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah :

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.3 *Metode Data Mining*

Dalam pengolahan sebuah data yang kecil atau pun data yang besar, terdapat sebuah cara atau teknik penyelesaian yang akan digunakan. Dalam *data mining* yang sering disebut sebagai metode. Terdapat beberapa metode yang membantu *data mining* dalam pengolahan data tersebut. Metode-metode tersebut adalah :

2.3.1 *Metode Klasifikasi Decision Tree*

Decision Tree merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (Vulandari, 2017).

Decision tree merupakan suatu metode yang sangat praktis dalam melakukan sebuah penelitian, metode ini sangat membantu untuk menyelesaikan sebuah permasalahan yang baik, karena manfaat dari *decision tree* ini kemampuan

untuk mem-*break down* proses sebuah keputusan agar menjadi lebih efektif solusi dari permasalahan.

2.3.2 Metode Klasifikasi Teorema Bayes

Teori keputusan *Bayes* adalah pendekatan statistika yang fundamental dalam *data mining* (Vulandari, 2017). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas. *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dan sejumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Alfa Saleh, 2015). Bayes yang memiliki akurasi dan kecepatan ke dalam *database* dengan data yang desar. Dalam ilmu probabilitas bersyarat dinyatakan sebagai :

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$

Rumus 2.1 Probabilitas Bayes

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y, atau dengan bahasa lain $P(X|Y)$ adalah presentase banyaknya X di dalam Y.

Rumus dari *teorema bayes* adalah

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

Rumus 2.2 Teorema Bayes

Keterangan diatas X adalah bukti, H adalah hipotesis, $P(H/X)$ adalah probabilitas bahwa hipotesis H benar untuk bukti X atau dengan kata lain $P(H/X)$ merupakan probabilitas posterior H dengan syarat X, $P(X/H)$ adalah probabilitas

bahwa bukti X benar untuk hipotesis H atau probabilitas posterior X dengan syarat H , $P(H)$ adalah probabilitas prior hipotesis H , dan $P(X)$ adalah probabilitas *prior* bukti X .

2.4 SMK Negeri 5 Batam

Sekolah SMK negeri 5 merupakan sekolah SMK yang ada di kota Batam yang berdiri pada tanggal 12 juni 2009 sesuai dengan Kep.Ka.Dinas pendidikan kota Batam No 345/422.4/Dikmen/VI/2009 tentang : pembentukan panitia penerimaan siswa baru SMK negeri 5 Batam Tahun pelajaran 2009/2010 yang menggunakan fasilitas SMPN 27 Batam di : kavling seroja, kel.sungai pelunggut, kec.sagulung Batam selama satu semester. Kemudian berdasarkan keputusan walikota Batam, Drs.Ahmad Dahlan, MH, No.KPTS.279/HK/X/2009 tanggal 21 oktober 2009 tentang penetapan nama sekolah SD, SMP, SMA, dan SMK negeri dilingkungan pemerintahan kota Batam, ditetapkan berdirinya sebuah SMKN 5 Batam, di kav.bukit kamboja, kel.sungai pelunggut, kec.sagulung.

2.5 Software Pendukung

Software merupakan sebuah sistem yang dapat dijalankan di dalam komputer yang dikerjakan secara otomatis. *Software* yang digunakan didalam penelitian ini adalah *RapidMiner*. *RapidMiner* adalah perangkat lunak yang dibuat oleh Dr.Markus Hofmann dari *Institute Of Technologi Blanchardstown* dan Ralf Klinkenberg dari *Rapid-i.com* dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program java dibawah lisensi GNU *Public Licence* dan *RapidMiner* dapat

dijalankan di sistem operasi manapun. Dengan menggunakan *RapidMiner*, tidak dibutuhkan kemampuan koding khusus, semua fasilitas sudah disediakan (Haryati, SudarsonHaryati, S., Sudarsono, A., &Suryana, E. (2015).

2.6 Penelitian Terdahulu

Dalam melakukan penelitian tentang penerapan *data mining* dalam nilai siswa, peneliti menggunakan beberapa referensi yang diambil dari jurnal-jurnal terdahulu yang menggunakan teknik *data mining*.

1. Dicky Nofriansyah, (2016). “**Penerapan *Data Mining* Dengan Algoritma *Naïve Bayes Classifier* Untuk Mengetahui Minat Beli Pelanggan Terhadap Kartu *Internet XL* (Studi Kasus di CV. Sumber Utama Telekomunikasi)**”. ISSN: 1978-6603. Penelitian ini menjelaskan salah satu perusahaan yang bergerak di bidang bisnis telekomunikasi, perusahaan ini menarik minat pelanggan dengan cara pemasaran sehingga tidak kalah saing dengan perusahaan lainnya. Metode *naïve bayes* yang digunakan dapat memprediksi peluang di masa depan berdasarkan pemasaran sebelumnya. Penelitian ini memprediksi terjual atau tidak kartu yang baru, sehingga perusahaan lebih cepat mengambil sebuah keputusan untuk meningkatkan pemasaran tersebut.
2. Alfa Saleh, (2015). “**Implementasi Metode Klasifikasi *Naïve Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga**” ISSN: 2354-5771. Pada penelitian ini menjelaskan permasalahan dalam memprediksi besarnya penggunaan listrik rumah tangga. Listrik adalah peranan penting bagi masyarakat, tetapi persediaan listrik yang belum

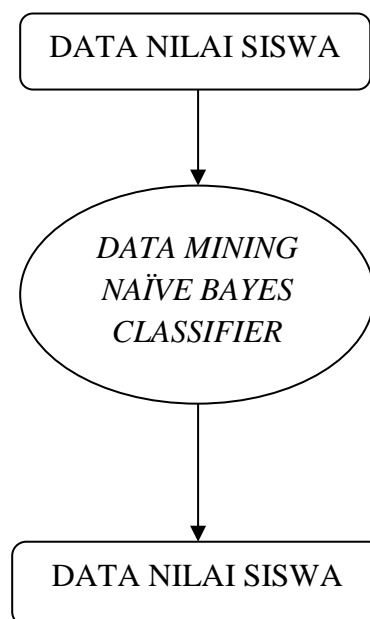
mampu memenuhi permintaan listrik yang besar. Mengatasi hal seperti ini harus adanya campur tangan pemerintah dan masyarakat dalam menggunakan listrik dengan bijak sehingga penggunaan listrik lebih efektif. 60 data penggunaan listrik di uji menggunakan metode *naïve bayes classifier* dengan hasil persentase 78,3333% untuk keakuratan prediksi. Dari 60 data tersebut hanya 47 data penggunaan listrik rumah tangga yang berhasil diklasifikasi secara tepat.

3. Yuli, (2017). “**Data Mining : Klasifikasi Menggunakan Algoritma C4.5**”. ISSN: 2407-0491. Pada penelitian ini menjelaskan pengklasifikasi dari *data mining* sehingga mendapatkan informasi yang tepat sehingga bermanfaat bagi kumpulan data yang besar. Pengklasifikasi yang digunakan dengan algoritma C4.5. Dengan metode ini keputusan akan lebih efektif dan mudah dipahami.
4. Handoko, (2016). “**Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi TKJ Akademi Komunitas Solok Selatan)**”. ISSN: 2476-8812. Penelitian ini membahas tentang penerapan *data mining* untuk meningkatkan mutu pembelajaran pada instansi perguruan tinggi di program studi TKJ. *K-Means Clustering* adalah metode yang digunakan dalam penelitian ini dengan pengujian *RapidMiner5.3* sehingga akan meningkatkan pembelajaran.
5. Widiastuti, Santosa, & Supriyanto, (2014). “**Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Optimization Untuk**

Deteksi Penyakit Jantung". ISSN: 2355-5920. Pada penelitian ini membahas tentang tingginya faktor kematian akibat penyakit jantung. Salah satu yang menyebabkan penyakit jantung adalah kurangnya aliran darah dan oksigen ke jantung. Dengan menggunakan metode *naïve bayes* menghasilkan akurasi 82,14%, sementara dengan metode *naïve bayes* berbasis *particle swarm optimization* akurasi menjadi meningkat 92,86%.

2.7 Kerangka Pemikiran

Kerangka pemikiran yang dapat penulis gambarkan untuk penelitian ini bisa kita lihat pada gambar berikut :



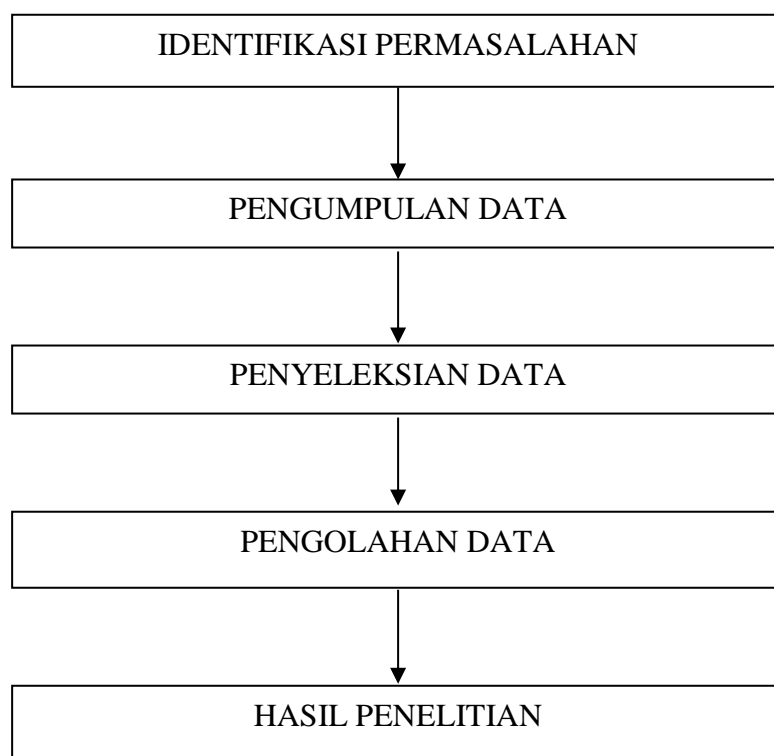
Gambar 2.1 Kerangka Pemikiran
Sumber Data Penelitian (2018)

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Dalam melakukan penelitian, peneliti membuat sebuah desain penelitian yang merepresentasikan seluruh kegiatan yang dilakukan peneliti dalam melakukan penelitian tentang *data mining* dari awal penelitian sampai akhir dari penelitian. Desain penelitian *data mining* akan dijelaskan melalui gambar dibawah ini.



Gambar 3.1 Desain Penelitian
Sumber : Data Penelitian (2019)

Sesuai dengan gambar diatas, dapat dijelaskan sebagai berikut :

1. Identifikasi Masalah

Menjelaskan tentang apa saja yang menjadi permasalahan dalam penelitian yang dilakukan oleh peneliti. Peneliti melihat adanya permasalahan dalam sekolah SMK negeri 5 sehingga peneliti akan menggunakan *data mining* dalam menyelesaikan permasalahan tersebut.

2. Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan untuk mencapai tujuan penelitian. Data yang diambil oleh peneliti adalah menggunakan metode wawancara dan observasi untuk mendapatkan nilai siswa SMK negeri 5 yang akan dijadikan sumber data penelitian.

3. Penyeleksian Data

Peneliti melakukan penyeleksian data sehingga data yang diperoleh akan dijadikan sampel yang benar-benar sesuai dengan tujuan penelitian ini.

4. Pengolahan Data

Pengolahan data dalam penelitian ini menggunakan aplikasi *RapidMiner* dengan menerapkan Algoritma *Naïve Bayes Classifier*. Data yang telah dipilih dan diseleksi akan diproses dengan cara algoritma *naïve bayes* yaitu secara manual sehingga diuji kembali dengan software *RapidMiner*.

5. Hasil Penelitian

Peneliti menampilkan hasil penelitian yang telah diuji, yang berisi keputusan dari hasil pengujian data.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang peneliti lakukan pada penelitian ini adalah sebagai berikut:

3.2.1 Observasi

Melakukan pengamatan langsung ke sekolah SMK negeri 5 Batam untuk mendapatkan data yang dibutuhkan.

3.2.2 Wawancara

Sebuah metode yang melibatkan peneliti dan narasumbernya. Peneliti mewawancarai Kepala Sekolah SMK Negeri 5 Bapak Agus Sahrir, M.Pd. Hasil wawancara yang dilakukan peneliti akan digunakan sebagai sumber data sehingga permasalahan yang sedang dibahas ditugas akhir ini memperoleh gambaran dan penjelasan secara efektif.

3.2.3 Studi Pustaka

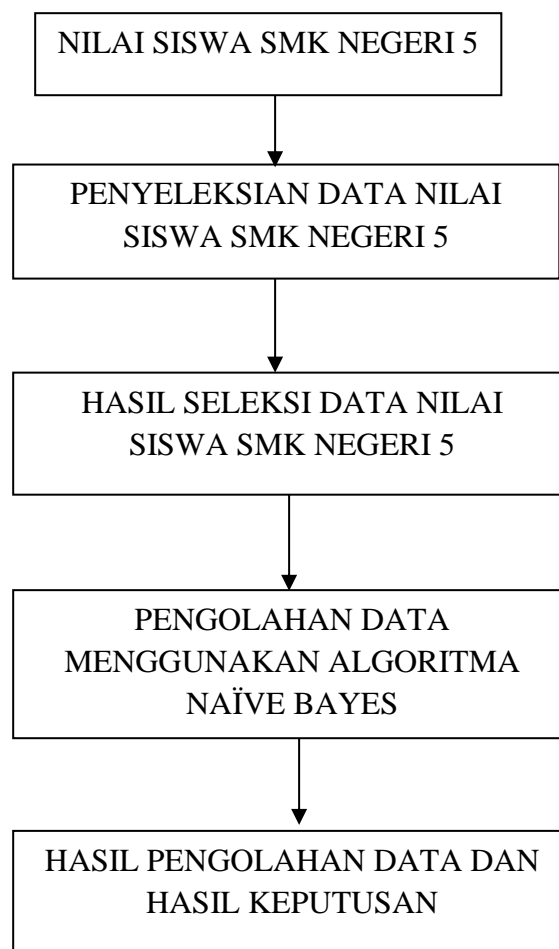
Peneliti mengumpulkan berbagai referensi, peneliti melakukan kajian pustaka yang membahas tentang penelitian kali ini. Kajian pustaka mengambil sumber-sumber yang berkaitan dengan *data mining* yaitu buku dan jurnal.

3.3 Operasional Variabel

Dalam penelitian ini operasional variabel yang dipakai adalah nilai siswa pada jurusan teknik komputer dan jaringan di sekolah SMK Negeri 5 Batam. Untuk mendukung penelitian ini, peneliti mengambil hasil nilai siswa yaitu UTS dan UAS setiap tahunnya.

3.4 Metode Analisis dan Rancangan Sistem

Metode analisis dan rancangan sistem pada penelitian bertujuan untuk memilih data yang akan dijadikan sampel penelitian. Dapat dilihat pada gambar dibawah ini :



Gambar 3.2 Metode Perancangan Sistem

Sumber : Data Penelitian (2019)

3.5 Lokasi dan Jadwal Penelitian

Penelitian ini dilakukan dari bulan maret 2019 sampai dengan juli 2019.

Jadwal penelitian ini dapat dilihat pada tabel dibawah ini:

Tabel 3.1 Jadwal Penelitian

No	Kegiatan	MARET 2019				APRIL 2019				MEI 2019				JUNI 2019				JULI 2019			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Judul			■	■																
2	Penyusunan Bab I					■	■														
3	Penyusunan Bab II									■	■	■	■								
4	Penyusunan Bab III													■	■	■	■				
5	Penyusunan Bab IV																	■	■	■	■
6	Penyusunan Bab V, daftar pustaka, lampiran																				
7	Pengumpulan Skripsi																				

Sumber : Data Penelitian (2019)