

**PENERAPAN *DATA MINING* UNTUK EVALUASI
NILAI AKADEMIK MAHASISWA KOTA BATAM
MENGUNAKAN ALGORITMA
*NATIVE BAYES CLASSIFIER***

SKRIPSI



**Oleh:
Adhitya Dewanta
150210200**

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

**PENERAPAN *DATA MINING* UNTUK EVALUASI
NILAI AKADEMIK MAHASISWA KOTA BATAM
MENGUNAKAN ALGORITMA
*NATIVE BAYES CLASSIFIER***

SKRIPSI

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



**Oleh:
Adhitya Dewanta
150210200**

**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,



Adhitya Dewanta

150210200

**PENERAPAN *DATA MINING* UNTUK EVALUASI
NILAI AKADEMIK MAHASISWA KOTA BATAM
MENGUNAKAN ALGORITMA
*NATIVE BAYES CLASSIFIER***

**Oleh
Adhitya Dewanta
150210200**

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

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

Batam, 16 Februari 2019



Yulia, S. Kom., M. Kom

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi nilai akademik dari mahasiswa Universitas Riau Kepulauan (UNRIKA) khususnya mahasiswa Fakultas Keguruan dan Ilmu Pendidikan dengan menerapkan *data mining*. Penelitian ini dilakukan pada Universitas Riau Kepulauan (UNRIKA) yang beralamat di jalan Batu Aji Baru no 99 Batam. Sampel yang terdapat di dalam penelitian ini berjumlah sebanyak 80 orang dari 80 orang populasi yang diambil dari dua kelas yang berbeda dalam satu program studi Matematika di Fakultas Keguruan dan Ilmu Pengetahuan. Dalam pengumpulan data pada penelitian ini, metode yang digunakan adalah dengan cara melakukan observasi dan wawancara terhadap narasumber dari UNRIKA. Aplikasi pembantu dalam penelitian ini yang digunakan adalah WEKA untuk menghitung hasil evaluasi nilai akademik mahasiswa. Algoritma yang digunakan dalam penelitian ini adalah algoritma *naïve bayes classifier*. Hasil pengujian ini berupa sebuah keputusan terhadap nilai akademik mahasiswa. Sebesar 97,5% mahasiswa mendapatkan kriteria lulus dari data yang diinput yang berjumlah 80 *data testing*. Pengujian ini memiliki tingkat keakurasian sebesar 96%. Penelitian ini membuktikan bahwa data yang memiliki jumlah yang banyak dapat dengan mudah untuk diproses dalam hal mencari hasil evaluasi nilai akademik mahasiswa. Penelitian ini juga membuktikan bahwa semakin banyak jumlah *data training* yang dimasukan akan menjadi lebih akurat dalam menentukan hasil pengujiannya terhadap *data testing*.

Kata kunci: Evaluasi, Nilai Akademik Mahasiswa, *Data Mining*, Algoritma *Naïve Bayes Classifier*

ABSTRACT

This study aims to evaluation of student academic score from Universitas Riau Kepulauan (UNRIKA) student especially from student of Fakultas Keguruan dan Ilmu Pendidikan by applying of data mining. This study was conducted in Universitas Riau Kepulauan (UNRIKA) in jalan Batu Aji Baru no 99 Batam. The sample in this study amounted to as many as 80 person form 80 person of population which taken from two different class in one study program in Fakultas Keguruan dan Ilmu Pengetahuan. Data collection in this study using observation and debriefing method from iuterviews of UNRIKA. This study using WEKA application to calculate the result of student academic score. This study using naïve bayes classifier algorithm. Result of this study is a decision form student academic score. 97,5% student have pass criteria from inputted data as many as 80 data testing. This study has 96% accuracy. This study prove if a lot of data can easily processed in case to find result of student academic score. This study prove more and more count of inputted data training will be more accurate to determine of student academic score.

Keywords: Evaluation, Student Academic Score, Data Mining, Naïve Bayes Classifier Algorithm

KATA PENGANTAR

Puji dan syukur kepada Allah SWT 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 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 Manajemen Universitas Putera Batam.
3. Ibu Yulia, S.Kom., M.Kom selaku pembimbing Skripsi pada Program Studi Manajemen Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam.
5. Bapak Yarsi Efendi, M.Pd., S.Si selaku wakil dekan Fakultas Keguruan dan Ilmu Pengetahuan Universitas Riau Kepulauan.
6. Kepada Orang Tua penulis Bapak Hari Handogo dan Ibu Erni Puspawati yang selalu mendoakan, dan Pengorbanannya baik segi moral maupun materil kepada penulis sehingga dapat menyelesaikan penulisan skripsi ini.
7. Kepada sahabat terbaik penulis Nofrizal, Muhammad Rizal 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 perumahan Taman Pesona Indah Batu Aji dan bapak Dedy Syaifulah Nasution selaku kepala BAK dari Fakultas Keguruan dan Ilmu Pengetahuan Universitas Riau Kepulauan.
10. Pihak-pihak yang membantu dalam penyelesaian skripsi ini yang tidak bisa penulis ucapkan satu persatu

Semoga Tuhan membalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Aamiin.

Batam, Februari 2019

Penulis

DAFTAR ISI

	HALAMAN
HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	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 Rumusan Masalah.....	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian.....	6
BAB II KAJIAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i>	7
2.2 <i>Data Mining</i>	9
2.3 <i>Metode Data Mining</i>	12
2.4 <i>Software Pendukung</i>	15
2.5 Penelitian Terdahulu	16
2.6 Kerangka Pemikiran	19
BAB III METODE PENELITIAN	
3.1 Desain Penelitian	21
3.2 Pengumpulan Data.....	23
3.3 Operasional Variabel	24
3.4 Metode Perancangan Sistem.....	25
3.5 Lokasi Dan Jadwal Penelitian.....	26
BAB IV HASIL DAN PEMBAHASAN	
4.1 Analisa Data.....	28
4.2 Hasil Pengujian.....	79
BAB V SIMPULAN DAN SARAN	
5.1 Simpulan.....	86
5.2 Saran	87

**DAFTAR PUSTAKA
DAFTAR RIWAYAT HIDUP
SURAT KETERANGAN PENELITIAN
LAMPIRAN**

DAFTAR GAMBAR

	HALAMAN
Gambar 2.1 Kerangka Pemikiran	20
Gambar 3.1 Desain Penelitian	21
Gambar 3.2 Perancangan Sistem	25
Gambar 4.1 WEKA GUI Chooser	80
Gambar 4.2 Konversi CSV ke ARFF	81
Gambar 4.3 Tab Preprocess Explorer WEKA	81
Gambar 4.4 Tab Classify Explorer WEKA	82
Gambar 4.5 Hasil Uji WEKA	83
Gambar 4.6 Simpan Hasil Pengujian	83
Gambar 4.7 Hasil Pengujian	84
Gambar 4.8 Hasil Pengujian	85

DAFTAR TABEL

	HALAMAN
Tabel 3.1 Jadwal Penelitian.....	27
Tabel 4.1 Data Nilai Mahasiswa	28
Tabel 4.2 Data Training.....	32
Tabel 4.3 Data Testing	38
Tabel 4.4 Data Testing 1	39
Tabel 4.5 Hasil Data Testing 1	41
Tabel 4.6 Data Testing 2	42
Tabel 4.7 Hasil Data Testing 2	44
Tabel 4.8 Data Testing 3	44
Tabel 4.9 Hasil Data Testing 3	46
Tabel 4.10 Data Testing 4	46
Tabel 4.11 Hasil Data Testing 4	48
Tabel 4.12 Data Testing 5	49
Tabel 4.13 Hasil Data Testing 5	51
Tabel 4.14 Data Testing 6	51
Tabel 4.15 Hasil Data Testing 6	53
Tabel 4.16 Data Testing 7	53
Tabel 4.17 Hasil Data Testing 7	55
Tabel 4.18 Data Testing 8	56
Tabel 4.19 Hasil Data Testing 8	58
Tabel 4.20 Data Testing 9	58
Tabel 4.21 Hasil Data Testing 9	60
Tabel 4.22 Data Testing 10	60
Tabel 4.23 Hasil Data Testing 10	62
Tabel 4.24 Data Testing 11	63
Tabel 4.25 Hasil Data Testing 11	65
Tabel 4.26 Data Testing 12	65
Tabel 4.27 Hasil Data Testing 12	67
Tabel 4.28 Data Testing 13	67
Tabel 4.29 Hasil Data Testing 13	70
Tabel 4.30 Data Testing 14	70
Tabel 4.31 Hasil Data Testing 14	72
Tabel 4.32 Data Testing 15	72
Tabel 4.33 Hasil Data Testing 15	74
Tabel 4.34 Data Testing 16	75
Tabel 4.35 Hasil Data Testing 16	77
Tabel 4.36 Data Testing 17	77
Tabel 4.37 Hasil Data Testing 17	79

DAFTAR RUMUS

	HALAMAN
Rumus 2.1 <i>Probabilitas Bayes</i>	14
Rumus 2.2 <i>Teorema Bayes</i>	14
Rumus 4.1 <i>Teorema Bayes</i>	39

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Perkembangan teknologi informasi dan komunikasi atau sering disebut TIK pada saat ini sudah tidak dapat di hindari lagi dalam kehidupan saat ini. Semua lapisan kegiatan saat ini sudah sangat membutuhkan yang namanya teknologi informasi dan komunikasi. Perkembangan TIK ini sangat cepat, seperti kebutuhan sebuah TIK yang digunakan dalam pengolahan data dan nilai dalam instansi pendidikan. Dengan menggunakan TIK yang dapat melakukan pengolahan data dan nilai, tentunya akan mempermudah para pegawai yang terkait di dalam instansi pendidikan tersebut dalam mengolah hasil evaluasi nilai bahkan dalam menentukan metode pembelajaran selanjutnya.

Universitas Riau Kepulauan atau UNRIKA adalah salah satu dari perguruan tinggi yang terletak di kota Batam. UNRIKA sendiri merupakan pengembangan dari Sekolah Tinggi Ilmu Ekonomi (STIE) dan Sekolah Tinggi Teknik (STT) Batam yang berdiri sejak tahun 1994 oleh Yayasan Perguruan Tinggi Kota Batam. Pada tahun 2006 STIE dan STT dilebur menjadi satu karena perubahan zaman dan paradigma baru menjadi Universitas Riau Kepulauan atau UNRIKA melalui SK Dikti No.68/D/0/2006 dan 68/D/T/2006. UNRIKA memiliki 5 fakultas didalamnya, fakultas tersebut antara lain adalah Fakultas Ekonomi, Fakultas Teknik, Fakultas Keguruan dan Ilmu Pendidikan, Fakultas Hukum dan Fakultas

Ilmu Sosial dan Ilmu Politik. Universitas Riau Kepulauan sendiri merupakan lembaga pendidikan swasta yang berada di bawah naungan Kopertis Wilayah X.

Dalam perhitungan hasil evaluasi nilai akademik mahasiswa, UNRIKA menerapkan perhitungan manual. Belum adanya penerapan *data mining* dalam perhitungan evaluasi nilai akademik mahasiswa mengakibatkan permasalahan dalam hal efisiensi waktu. UNRIKA memiliki data penilaian akademik mahasiswa yang banyak, sehingga tanpa diterapkan *data mining* dalam pengolahannya menjadi lebih sulit dan memakan waktu yang lama.

Dengan permasalahan seperti itu, tentunya penerapan *data mining* akan lebih bagus dalam melakukan evaluasi nilai akademiknya. *Data mining* dianggap sesuai dalam kasus seperti ini, karena dengan menggunakan *data mining* yang dapat menampung banyak data, akan menjadi lebih mudah dalam pengolahan evaluasi nilai akademiknya. *Data mining* adalah bagian dari teknologi informasi dan komunikasi yang lazim digunakan dalam memprediksi kejadian-kejadian yang akan terjadi.

Data mining atau penambangan data sendiri memiliki arti sebagai sebuah tata cara atau proses, eksplorasi dan pemodelan dari sejumlah data yang sangat besar atau banyak untuk mencari suatu pola yang biasanya tidak dapat disadari keberadaan pola tersebut secara langsung (Pramadhani & Setiadi, 2014). Dengan metode yang sesuai *data mining* dapat di aplikasikan dalam perhitungan evaluasi nilai akademik mahasiswa. Metode *naïve bayes classifier* dapat digunakan dalam perhitungan evaluasi nilai akademik mahasiswa.

Metode *naïve bayes classifier* (NBC) merupakan sebuah pengklasifikasian atau pengelompokan dengan sifat probabilistik yang menghitung dari beberapa nilai probabilitas dengan menjumlahkan frekuensi dan kombinasi dari beberapa nilai yang diberikan dari dataset yang ada (Saleh, 2015). Algoritma yang dipakai menggunakan *naïve bayes* dengan asumsi semua atribut bersifat *independen*.

Dengan menggunakan *data mining* dan metode NBC data yang besar dapat dengan mudah diolah untuk menghitung hasil evaluasi nilai mahasiswa. Tanpa digunakan *data mining*, akan sangat susah dalam menghitung evaluasi nilai akademik mahasiswa. Kesulitan dalam perhitungan data yang besar dengan tidak menerapkan *data mining* adalah efisiensi waktu. Dalam era globalisasi ini efisiensi waktu sangat diperlukan dalam berbagai kegiatan yang ada. Dengan efisiensi waktu yang baik pekerjaan akan lebih mudah dikerjakan sehingga kuantitas dalam jumlah pekerjaan yang dilakukan akan bertambah.

Berdasarkan permasalahan di atas, peneliti mengangkat judul penelitian dengan judul “**Penerapan *Data Mining* Untuk Evaluasi Nilai Akademik Mahasiswa Kota Batam Menggunakan Algoritma *Naïve Bayes Classifier*”.**

1.2 Identifikasi Masalah

Berdasarkan latar belakang permasalahan di atas, beberapa point-point permasalahan yang akan menjadi fokus penelitian adalah:

1. Evaluasi nilai akademik pada Universitas Riau Kepulauan belum menggunakan *data mining* sehingga dalam mengevaluasi nilai akademik tersebut menjadi lebih sulit.

2. Penggunaan perhitungan manual dalam melakukan evaluasi nilai akademik mahasiswa mengurangi efisiensi waktu dalam pengolahan datanya.
3. Banyaknya jumlah data yang akan di kelola, apabila tanpa menggunakan sebuah sistem maka pengelolaan data tersebut akan menjadi lebih sulit.

1.3 Pembatasan Masalah

Beberapa hal yang dijadikan pembatasan masalah agar penelitian ini tidak menyimpang dan hasil sesuai target. Adapun batasan masalah di dalam penelitian ini adalah:

1. Penelitian ini mengambil nilai akademik dari mahasiswa program studi Matematika sebagai sumber datanya.
2. Penelitian ini menggunakan *data mining* dan metode *naïve bayes classifier* dalam menyelesaikan permasalahannya.
3. Hasil penelitian ini berupa sebuah hasil evaluasi perhitungan nilai mahasiswa yang akan datang atau yang sedang berlangsung.
4. Penelitian ini dilakukan di Fakultas Keguruan & Ilmu Kependidikan Universitas Riau Kepulauan (UNRIKA).
5. Aplikasi yang akan digunakan dalam penelitian ini adalah WEKA 3.9.3.

1.4 Perumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan di atas, dapat dirumuskan permasalahan yang akan diselesaikan di dalam penelitian ini adalah:

1. Bagaimana menerapkan *data mining* dalam mengevaluasi nilai akademik mahasiswa?
2. Bagaimana cara penerapan algoritma *naïve bayes classifier* dalam melakukan evaluasi nilai akademik mahasiswa?
3. Bagaimana cara menghitung keakuratan pengujian dalam menggunakan metode *naïve bayes classifier*?

1.5 Tujuan Penelitian

Berdasarkan perumusan masalah yang sudah dipaparkan di atas, adapun tujuan dari penelitian ini adalah:

1. Untuk menerapkan *data mining* dalam mengevaluasi nilai akademik mahasiswa.
2. Untuk menerapkan algoritma *naïve bayes classifier* dalam melakukan perhitungan evaluasi nilai akademik mahasiswa.
3. Untuk menghitung keakuratan pengujian dalam menggunakan algoritma *naïve bayes classifier*.

1.6 Manfaat Penelitian

Dengan tercapainya tujuan dari penelitian, maka diharapkan penelitian ini bermanfaat bagi berbagai pihak. Manfaat dari penelitian ini adalah:

1.6.1 Bagi Universitas Riau Kepulauan

Manfaat penelitian bagi objek penelitian adalah :

1. Penelitian ini dapat dijadikan bahan evaluasi dari UNRIKA untuk bahan acuan pergantian paradigma lama menjadi paradigma baru yaitu menerapkan *data mining* dalam melakukan evaluasi nilai akademik dari mahasiswanya.
2. Pihak terkait yang berhadapan langsung dengan pengolahan data menjadi terbantu dalam pengolahan nilai mahasiswa.

1.6.2 Bagi Peneliti

Manfaat penelitian bagi peneliti adalah :

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

BAB II KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Pencarian pengetahuan di dalam *database* atau lebih dikenal dengan istilah *knowledge discovery in database* (KDD) merupakan rumpun ilmu yang memiliki sebuah proses yang dikenal dengan *data mining*. *Data mining* dan KDD umumnya sering disamakan, sebenarnya kedua istilah tersebut memiliki konsep yang berbeda. Proses KDD secara garis besar akan dijelaskan sebagai berikut (Kusrini, 2009):

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum memulai tahap penggalian informasi dalam KDD. Data hasil penyeleksian akan digunakan pada proses *data mining*, data tersebut nantinya akan di simpan di dalam suatu berkas dan terpisah dari *database* operasional.

2. *Pre-processing Cleaning*

Sebelum melakukan proses *data mining*, perlu dilakukan sebuah proses *cleaning* pada data yang akan menjadi fokus KDD. Proses ini mencakup seperti membuang data yang redundan, pemeriksaan data yang inkonsisten, dan memperbaiki kesalahan yang ada pada data, seperti kesalahan penulisan (tipografi). Proses ini juga melakukan proses yang bernama *enrichment*, proses *enrichment* yaitu proses

“memperkaya” data yang sudah ada dengan data / informasi lainnya yang relevan dan diperlukan di dalam KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* di dalam KDD merupakan proses yang tergantung pada jenis atau pola informasi yang akan dicari di dalam *database*.

4. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data yang telah terpilih dari proses sebelumnya dengan menggunakan sebuah metode tertentu. Teknik, metode, atau algoritma di dalam *data mining* sangat bervariasi seperti contohnya algoritma *naïve bayes classifier*. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan proses KDD secara umum.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* harus ditampilkan dalam bentuk yang mudah dimengerti oleh *user* yang akan menggunakan hasil proses tersebut. Tahap ini merupakan salah satu bagian dari KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis 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 digunakan dalam berbagai hal yang bersifat kolektif data yang besar. *Data mining* adalah sebuah 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* menggunakan metode statistik yang digunakan dalam perhitungan secara empiris dari suatu *database* yang ada. Data yang diidentifikasi dan di ekstraksi oleh *data mining* bukanlah data yang berbentuk kecil / sedikit, melainkan *data mining* digunakan apabila data tersebut merupakan suatu data yang besar / banyak sehingga apabila tanpa menggunakan *data mining* akan menyulitkan dalam pengolahannya.

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 (Kusrini, 2009) :

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.

4. Tekanan kompetensi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat 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 suatu proses otomatisasi terhadap data yang sudah 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.

Dalam poin ketiga dari hal penting yang ada di atas, terdapat kata hubungan / pola. Maksud dari hubungan yang dihasilkan dari *data mining* adalah mencari ketergantungan dari satu objek terhadap yang lain. Contoh dari hubungan yang ada di dalam *data mining* adalah, hubungan pembelian suatu produk terhadap produk lainnya. Selain itu, hubungan juga dapat dilihat antara dua atribut atau lebih dan dua atau lebih objek.

Sementara penemuan pola merupakan keluaran lain dari *data mining*. Misalkan sebuah perusahaan yang akan meningkatkan kartu kredit dari pelanggannya, maka perusahaan tersebut akan mencari sebuah pola dari pelanggan-pelanggan yang ada.

Pola tersebut digunakan untuk mencari dan mengetahui mana pelanggan yang berpotensi dan mana pelanggan yang tidak berpotensi.

Data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan *integration and research, interpretation and visualitation* dari pola sejumlah data. Serangkaian proses tersebut memiliki tahapan antara lain (Vulandari, 2017):

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai agar dapat di *mining*).
4. Aplikasi teknik *Data Mining*, pada tahap ini data di ekstraksi berdasarkan pola yang sudah ada.
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan dalam pengambilan keputusan).
6. Presentasi pengetahuan dengan menggunakan teknik visualisasi.

Beberapa alasan digunakannya *data mining* dalam kegiatannya dapat dilihat dari beberapa sudut pandang. Pada sudut pandang komersial *data mining* dilakukan karena beberapa alasan, diantaranya adalah:

1. Besarnya *volume* data yang di simpan di dalam *data warehouse* seperti *data warehouse* dari *e-commerce*, penjualan yang ada di *supermarket*, transaksi bank.
2. Proses komputasi yang dapat diupayakan dan dioptimalkan.

3. Kerasnya persaingan dalam menyediakan layanan yang lebih baik.

Selain alasan digunakannya *data mining* dalam kegiatannya dilihat dari sudut pandang komersial terdapat alasan dari sudut pandang yang lain. Pada sudut pandang keilmuan alasan digunakannya *data mining* adalah untuk dapat meng-capture, menganalisis serta menyimpan data yang bersifat *real time* dan sangat besar, misalnya:

1. *Remote sensor* yang ditempatkan pada satelit yang mengorbit di luar angkasa.
2. *Telescope* yang digunakan untuk memindai langit oleh para astronom.
3. Simulasi saintifik yang membangkitkan data dalam ukuran sangat besar yaitu ukuran *terabytes*.

2.3 Metode *Data Mining*

Dalam pengolahan sebuah data yang kecil ataupun data yang besar, terdapat sebuah cara atau teknik penyelesaian yang digunakan. Dalam *data mining* cara atau teknik dalam pengolahan data biasa disebut dengan metode. Terdapat beberapa metode yang dapat membantu *data mining* dalam pengolahan datanya. Metode-metode tersebut antara lain:

2.3.1 Metode Klasifikasi *Decision Tree*

Decision Tree merupakan metode yang memanfaatkan ilustrasi dari sebuah pohon. Ilustrasi pohon tersebut digunakan dalam pengambilan keputusan dengan metode klasifikasi. Pohon apabila diartikan dalam pemecahan masalah

pengambilan keputusan adalah penjabaran mengenai berbagai macam alternatif pemecahan masalah yang dapat ditarik dari permasalahan tersebut (Vulandari, 2017). Pohon tersebut juga dapat menampilkan faktor-faktor probabilitas yang dapat mempengaruhi alternatif keputusan tersebut disertai dengan hipotesis dari hasil yang akan di dapat apabila mengambil alternatif keputusan tersebut.

Decision tree merupakan salah satu algoritma dari metode klasifikasi yang paling populer, dikarenakan *decision tree* merupakan algoritma yang mudah diinterpretasi oleh manusia. Konsep dari *decision tree* adalah merubah data yang ada menjadi berbentuk sebuah pohon dengan beserta aturan-aturannya. Manfaat utama dalam penggunaan *decision tree* adalah *decision tree* memiliki kelebihan dalam mem-*break down* proses pengambilan keputusan yang kompleks menjadi pengambilan keputusan yang lebih ringkas dari sebelumnya.

2.3.2 Metode Klasifikasi Teorema Bayes

Teori keputusan *Bayes* adalah suatu pendekatan yang bersifat statistika di dalam *data mining* (Vulandari, 2017). Pendekatan ini berdasarkan kuantifikasi *trade-off* antara berbagai macam keputusan klasifikasi yang ada dengan menggunakan probabilitas. *Bayesian classification* adalah pengklasifikasian statistik yang dapat diterapkan dalam memprediksi probabilitas keanggotaan dalam suatu *class* (Kusrini, 2009). Metode ini memiliki kemampuan klasifikasi yang sejenis dengan *decission tree* dan *neural network*. Metode ini sangat cocok

diterapkan apabila penelitian tersebut memiliki *database* dengan data yang besar atau banyak.

Metode *naïve bayes classifier* (NBC) merupakan sebuah pengklasifikasian atau pengelompokan dengan sifat probabilistik yang menghitung dari beberapa nilai probabilitas dengan menjumlahkan frekuensi dan kombinasi dari beberapa nilai yang diberikan dari *dataset* yang ada (Saleh, 2015). Algoritma yang dipakai menggunakan teori *Bayesian* dengan asumsi semua atribut bersifat *independen*.

Metode ini ditemukan oleh ilmuwan inggris bernama Thomas Bayes. Metode ini digunakan dalam memprediksi sebuah peluang di masa yang akan datang dengan didasari dari data-data yang berasal dari masa lampau. Metode ini didasari oleh penyederhanaan dimana atribut akan bersifat *independen* apabila atribut tersebut diberikan *output*. Metode *bayes* digunakan dalam melakukan inferensi induksi pada persoalan klasifikasi (Vulandari, 2017). Metode *bayes* menggunakan probabilitas sebagai landasarnya. Dalam ilmu probabilitas bersyarat dinyatakan sebagai berikut:

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

Rumus 2.1 *Probabilitas Bayes*

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

Teorema *Bayes* memiliki bentuk umum rumus seperti berikut:

$$P(C|X) = \frac{P(X|c)P(c)}{P(X)}$$

Rumus 2.2 *Teorema Bayes*

Dalam hal ini X merupakan data dari *class* yang belum diketahui sebelumnya. Sedangkan C merupakan hipotesis yang muncul dari variabel X dan merupakan *class* yang spesifik. $P(C|X)$ adalah probabilitas dari hipotesis C berdasarkan kondisi dari variabel X (*posteriori probability*). $P(C)$ adalah probabilitas dari hipotesis C (*prior probability*). $P(X|c)$ merupakan probabilitas dari variabel X terhadap kondisi pada hipotesis C . Sedangkan $P(X)$ merupakan probabilitas variabel X .

2.4 *Software* Pendukung

Software merupakan sebuah sistem yang dapat dijalankan di dalam komputer yang dikerjakan secara otomatis. *Software* yang digunakan di dalam penelitian ini adalah WEKA. WEKA adalah sebuah paket *tools machine learning* yang praktis (Vulandari, 2017). WEKA atau WAIKATO *Environment for Knowledge Analysis* yang dibuat di Universitas Waikato, New Zealand yang diperuntukan sebagai penelitian pendidikan dan berbagai aplikasi. WEKA dianggap mampu menyelesaikan permasalahan *data mining* yang ada di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan *machine learning*.

WEKA mudah dalam penggunaannya dan banyak diterapkan pada beberapa tingkatan yang berbeda-beda. Pada WEKA tersedia implementasi dari algoritma pembelajaran *state-of-the-art* yang dapat diimplementasikan ke dalam *dataset* dari *command line*. WEKA juga mengandung *tools* yang dapat digunakan untuk *pre-processing* data, klasifikasi, regresi, *clustering*, aturan *association*, dan visualisasi.

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.

Pengembangan WEKA sejalan dengan model yang dikeluarkan oleh LINUX, digit kedua yang genap pada nama *software* WEKA menandakan *release* yang stabil dan pada digit yang lainnya menandakan *release* pengembangan. Beberapa versi dari WEKA antara lain:

1. WEKA 3.0 : “versi buku” versi ini sesuai dengan yang di deskripsikan di dalam buku *data mining*.
2. WEKA 3.2 : “versi GUI” versi ini terdapat penambahan GUI (*Graphic User Interface*) dan CLI.
3. WEKA 3.3 : “versi pengembangan” versi ini sudah ditambahkan beberapa peningkatan.

2.5 Penelitian Terdahulu

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

1. Alfa Saleh, (2015). “**Implementasi Metode Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga**”. ISSN: 2354-

5771. Penelitian ini menjelaskan permasalahan tidak dapat memprediksi penggunaan listrik rumah tangga. Berbagai alat rumah tangga yang mengandalkan listrik sebagai sumber dayanya seperti lemari es, penanak nasi, kipas angin, *air conditioner*, televisi dan lainnya. Hal ini menjadikan permintaan akan listrik semakin meningkat namun permintaan ini berbanding terbalik dengan ketersediaan pasokan listrik yang semakin menipis. Penelitian ini menerapkan metode *naïve bayes* yang diharapkan mampu memprediksi besarnya penggunaan listrik rumah tangga sehingga dalam mengatur penggunaan listrik dapat menjadi lebih mudah. Pengujian dilakukan kepada 60 pengguna listrik dengan metode *naïve bayes*, didapatkan hasil presentase sebesar 78,3333% untuk keakuratan prediksi, dimana dari 60 data pemakaian listrik rumahan yang telah diuji, sebanyak 47 data pemakai listrik rumahan yang berhasil diklasifikasikan secara tepat.

2. Aline Embun Pramadhani dan Tedy Setiadi, (2014). “**Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernafasan Akut) Dengan Algoritma Decission Tree (ID3)**”. e-ISSN: 2338-5197. Dalam penelitian ini mengangkat permasalahan banyaknya data penderita penyakit ISPA di Klinik Dharma Husana yang hanya disimpan saja tidak menjadi representasi pengetahuan dari gejala penyakit ISPA sebelumnya. Sehingga perlu adanya klasifikasi penyakit yang paling banyak di klinik ini. Klasifikasi ini bertujuan membentuk sebuah pohon keputusan dimana pohon ini akan digunakan untuk memprediksi penyakit ISPA dari variabel yang paling

mempengaruhi penyakit ISPA dengan kategorinya. Dalam membentuk sebuah pohon keputusan penelitian ini menggunakan metode *decision tree* untuk pengolahan datanya. Dari pohon keputusan yang terbentuk dari 200 data pasien maka dapat diketahui bahwa jenis kelamin tidak berpengaruh terhadap penyakit ISPA.

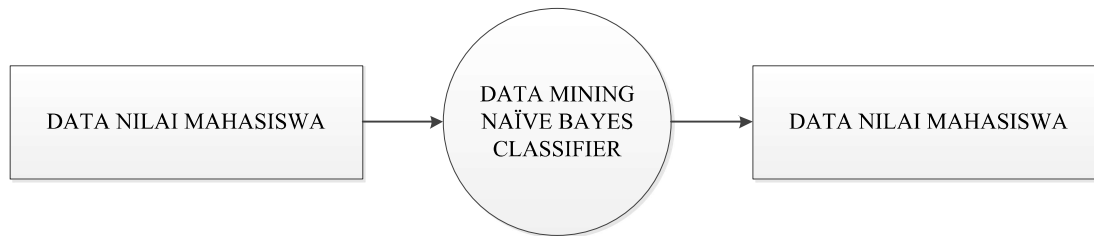
3. Andhika Novandya dan Isni Oktria, (2017). “**Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi**”. ISSN: 2089-5615. Mengangkat permasalahan perkiraan cuaca di daerah Bekasi. Data yang digunakan dalam penelitian ini mengacu pada data dari *World Wheater Online*. Dengan menggunakan algoritma C4.5 penelitian ini diharapkan dapat menghasilkan pola klasifikasi cuaca. Hasil pengujian algoritma C4.5 menggunakan *10-fold cross validation* dan dibuktikan dengan pembuatan aplikasi web untuk pengujian sehingga menghasilkan nilai akurasi sebesar 88.89%.
4. Katrina Shin dan Loganatan Muthu, (2015). “**Application of Big Data In Education Data Mining And Learning Analytics – A Literature Review**”. ISSN: 2229-6956. Mengangkat permasalahan aktivitas *online* dari mahasiswa yang menimbulkan sejumlah data yang tidak terpakai saat mahasiswa tersebut *online* akan menjadi tidak berguna. Penelitian ini menggunakan metode *big data* dan beberapa variabel di dalam pendidikan untuk memproses sejumlah data yang tidak terpakai tadi oleh kegiatan *online* mahasiswa. Hasil penelitian ini dimasukkan kedalam aplikasi *big data* untuk pendidikan sehingga dapat

menampilkan pembelajaran yang tersedia di dalam pendidikan *data mining* dan analisis.

5. P. Thangaraju, B. Deepa, dan T. Karthikeyan, (2014). “**Comparison of Data Mining Techniques for Forecasting Diabetes Mellitus**”. ISSN: 2778-1021. Melakukan penelitian berdasarkan permasalahan penyakit diabetes. Pengambilan permasalahan penyakit diabetes dilakukan karena penyakit diabetes adalah penyakit yang tergolong sebagai penyakit berbahaya dan biasanya terjadi dengan periode yang lama. Penelitian ini menggunakan metode klustering *K-Means* untuk meramal penyakit diabetes. Hasil dari penelitian ini akan terbagi menjadi tiga pengelompokan, yaitu pengelompokan hierarki, pengelompokan kepadatan dan pengelompokan *K-Means*.

2.6 Kerangka Pemikiran

Di dalam melakukan penelitian, peneliti melakukan argumentasi sementara terhadap permasalahan yang diangkat. Kerangka berpikir digunakan untuk menjelaskan secara sementara terhadap suatu gejala yang menjadi objek permasalahan. Kerangka pikir penelitian ini disajikan dalam bentuk diagram sebagai berikut :



Gambar 2.1 Kerangka Pemikiran
Sumber Data Penelitian (2018)

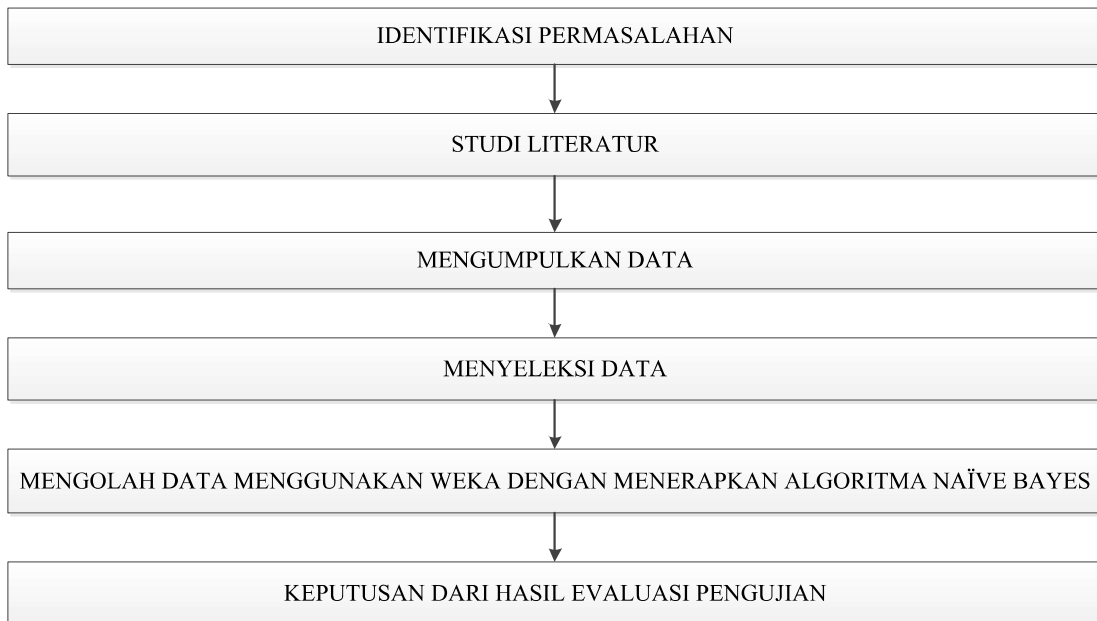
Penjelasan dari kerangka pemikiran yang di sajikan di dalam diagram tersebut akan dijelaskan di sebagai berikut :

1. Tahapan *input* meliputi memasukan variabel penelitian sebagai sumber acuan ke dalam aplikasi WEKA untuk diproses. Variabel yang dipakai adalah nilai akademik mahasiswa.
2. Tahapan proses adalah tahap melakukan olah data menggunakan aplikasi WEKA dengan metode klasifikasi teorema *naïve bayes*.
3. Tahapan *output* adalah tahap dimana keluaran dari hasil olah data dengan menggunakan aplikasi WEKA. Hasil penelitian diharapkan sesuai dengan tujuan dari penelitian ini.

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 di jelaskan melalui gambar dibawah ini.



Gambar 3.1 Desain Penelitian
Sumber : Data Penelitian (2018)

Penjelasan dari desain penelitian di atas adalah:

1. Identifikasi Permasalahan

Melakukan identifikasi permasalahan yang ada di lokasi penelitian. Peneliti melihat adanya permasalahan dimana jumlah mahasiswa UNRIKA yang besar akan mempersulit dalam menilai hasil evaluasi dari metode pembelajaran yang diterapkan. 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 didapatkan dari beberapa sumber seperti buku, jurnal dan internet.

3. Mengumpulkan Data

Peneliti melakukan pengumpulan data menggunakan metode wawancara dan observasi untuk mendapatkan nilai akademik mahasiswa UNRIKA yang akan dijadikan sumber data penelitian.

4. Menyeleksi Data

Peneliti melakukan penyeleksian data sehingga data yang akan dijadikan sampel adalah data yang benar-benar sesuai dengan tujuan penelitian ini. Data hasil seleksi ini nantinya akan digunakan sebagai sampel penelitian.

5. Mengolah Data Menggunakan WEKA Dengan Menerapkan Algoritma *Naïve Bayes*

Data yang sudah dijadikan sampel penelitian dimasukkan kedalam aplikasi WEKA dan menerapkan algoritma *naïve bayes* dalam mengolahnya.

6. Keputusan Dari Hasil Evaluasi Pengujian

Peneliti menampilkan hasil penelitian yang berisi keputusan lulus atau tidak lulusnya mahasiswa berdasarkan hasil pengujian.

3.2 Pengumpulan Data

Pengumpulan data adalah sebuah teknik yang dilakukan untuk mendapatkan sebuah data atau informasi dari narasumber atau sumber data. Teknik yang digunakan peneliti dalam mengumpulkan data adalah sebagai berikut :

3.2.1 Observasi

Observasi adalah sebuah metode pengumpulan data yang melibatkan peneliti berbaur pada objek penelitian itu sendiri. Pada observasi yang dilakukan peneliti, peneliti datang langsung ke Fakultas Keguruan dan Ilmu Pendidikan di Universitas Riau Kepulauan (UNRIKA) untuk pengambilan data.

3.2.2 Wawancara

Wawancara adalah sebuah metode pengumpulan data yang melibatkan narasumber dan peneliti. Peneliti memberikan butir-butir pertanyaan seputar penelitiannya kepada narasumbernya. Peneliti mewawancarai Dekan Fakultas Keguruan dan Ilmu Pengetahuan Bapak Yarsi Efendi selaku wakil dekan FKIP dan

Bapak Dedy Syaifulah Nasution selaku Kepala BAK. Hasil wawancara tersebut atau jawaban yang didapat dari narasumber akan digunakan sebagai sumber data.

3.2.3 Kajian Pustaka

Untuk memperdalam pemahaman dalam penelitian ini, peneliti melakukan kajian pustaka yang berhubungan dengan penelitian kali ini. Kajian pustaka mengambil sumber-sumber ilmiah seperti jurnal terdahulu, buku referensi yang berkaitan dengan *data mining*.

3.3 Operasional Variabel

Dalam penelitian ini operasional variabel yang dipakai adalah nilai akademik mahasiswa pada program studi Matematika di Fakultas Keguruan dan Ilmu Pengetahuan Universitas Riau Kepulauan (UNRIKA). Poin-poin yang terkandung didalam nilai akademik yang diangkat sebagai variabel sebagai berikut:

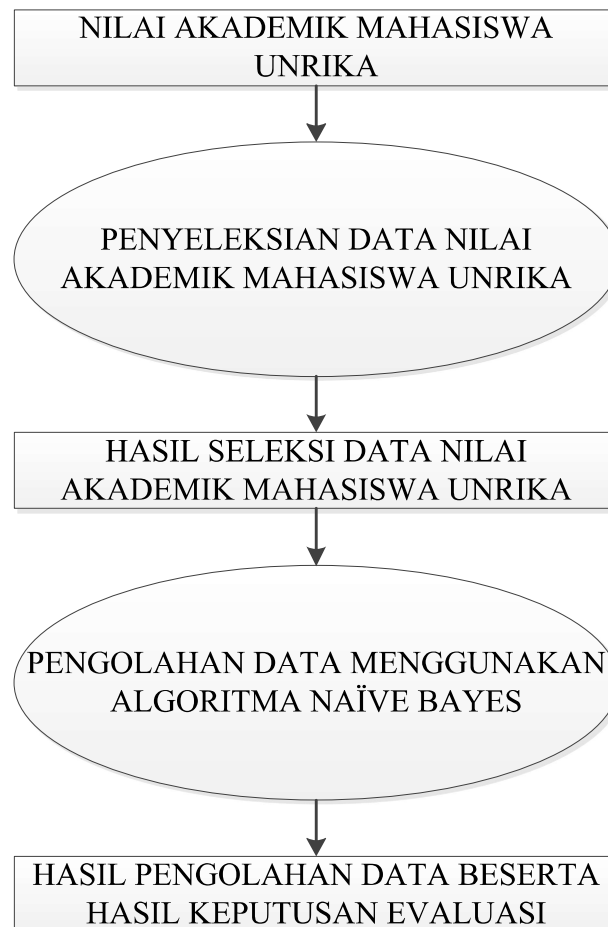
3.3.1 Nilai Tugas Mahasiswa

Untuk mendukung penelitian ini, peneliti mengambil nilai tugas mahasiswa yang diberikan oleh Dosen. Nilai tugas mahasiswa ini digunakan sebagai patokan pemahaman mahasiswa terhadap materi dari cara pembelajaran yang dilakukan tiap minggunya atau dari beberapa pertemuannya. Nilai tugas mahasiswa ini termasuk dengan nilai harian dan nilai tugas dari dosen selama satu semester.

3.3.2 Nilai Ujian Tengah Semester dan Ujian Akhir Semester

Nilai UTS dan UAS juga digunakan didalam penelitian ini, penggunaan nilai UTS dan UAS dalam penelitian ini ditujukan untuk menilai evaluasi dari mahasiswa selama setengah semester berjalan dan selama satu semester berjalan.

3.4 Metode Perancangan Sistem



Gambar 3.2 Perancangan Sistem
Sumber : Data Penelitian (2018)

Peneliti mengumpulkan data berupa nilai akademik mahasiswa UNRIKA dengan menggunakan metode wawancara dan observasi. 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 algoritma *naïve bayes* dengan menginput semua sampel penelitian ke dalam aplikasi WEKA untuk diproses. Hasil dari pemrosesan data dari aplikasi WEKA akan dipergunakan untuk menilai hasil evaluasi nilai akademik mahasiswa untuk menentukan keputusan kelulusan dalam mata kuliah yang di jalani

3.5 Lokasi dan Jadwal Penelitian

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

3.5.1 Lokasi Penelitian

Lokasi penelitian ini dilakukan di Universitas Riau Kepulauan jalan Batu Aji Baru no 99 Batam, di Fakultas Keguruan dan Ilmu Pendidikan (FKIP).

3.5.2 Jadwal Penelitian

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

Tabel 3.1 Jadwal Penelitian

No	Kegiatan	SEP 2018				OKT 2018				NOV 2018				DES 2018				JAN 2019				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
1	Input Judul	■																				
2	Pengumpulan Data		■	■																		
3	Pembuatan BAB 1				■	■																
4	Pembuatan BAB 2						■	■														
5	Pembuatan BAB 3								■	■	■	■										
6	Pembuatan BAB 4												■	■	■	■	■	■				
7	Pembuatan BAB 5																		■	■		
8	Pengumpulan Skripsi																					■

Sumber : Data Penelitian (2018)