

IMPLEMENTASI DATA MINING UNTUK REKOMENDASI PENENTUAN MAHASISWA PENERIMA BEASISWA MENGGUNAKAN METODE NAIVE BAYES STUDI KASUS : UNIVERSITAS KUNINGAN

Aah Sumiah¹, Wisnu Ahmad Maulana², Deris Rinaldi³

Beasiswa adalah bantuan yang diberikan oleh lembaga Pendidikan untuk meringankan beban mahasiswa dan untuk meningkatkan prestasi bagi mahasiswa. Sehingga mahasiswa tidak perlu memikirkan beban biaya kuliah dan fokus pada prestasi akademiknya. Untuk bisa memperoleh beasiswa, mahasiswa yang mendaftar harus dinyatakan lolos seleksi administrasi dan memenuhi beberapa komponen penilaian lainnya sebagai syarat mahasiswa layak menerima beasiswa. Dalam mengambil keputusan untuk menentukan mahasiswa yang layak mendapatkan beasiswa tersebut pada prosesnya terkadang membutuhkan waktu yang cukup lama dan terkadang pula beasiswa yang diberikan dirasa kurang tepat sasaran, hal ini dikarenakan proses seleksi masih dilakukan secara manual dan pengolahan data yang belum transparan. Sehingga perlu dibuatkan sebuah sistem yang dapat membantu pihak lembaga dalam mengambil keputusan dengan cepat agar beasiswa dapat didistribusikan tepat waktu dan tepat sasaran.

Tujuan Penelitian ini adalah menghasilkan sebuah *sistem informasi* dari hasil analisis *data mining* yang dapat digunakan oleh bagian akademik sebagai rekomendasi dalam proses seleksi penerimaan beasiswa. Metode yang digunakan yaitu *naive bayes classifier* dimana metode ini akan mengklasifikasikan mahasiswa menjadi dua kelas yaitu kelas mahasiswa yang layak mendapatkan beasiswa dan kelas mahasiswa yang tidak layak mendapatkan beasiswa dan implementasi program menggunakan *visual basic.net* dan *sql server*.

Keyword : *sistem informasi, data mining, beasiswa, naive bayes classifier, visual basic.net, sql server*

1. Pendahuluan

Beasiswa adalah bantuan yang diberikan oleh lembaga Pendidikan untuk meringankan beban mahasiswa dan untuk meningkatkan prestasi bagi mahasiswa. Sehingga mahasiswa tidak perlu memikirkan beban biaya kuliah dan fokus pada prestasi akademiknya. Ada berbagai macam beasiswa yang ditawarkan oleh Lembaga Pendidikan, diantaranya adalah beasiswa Peningkatan Prestasi Akademik (PPA) yang diberikan untuk mahasiswa yang memiliki prestasi akademik yang tinggi, baik kurikuler maupun ekstrakurikuler.

Proses Penerimaan beasiswa memiliki dua tahapan yaitu pendaftaran dan seleksi. Untuk bisa memperoleh beasiswa, mahasiswa yang mendaftar harus dinyatakan lolos seleksi administrasi dan memenuhi komponen penilaian sebagai syarat mahasiswa layak menerima beasiswa. Komponen penilaian yang ditentukan oleh tiap lembaga pendidikan berbeda-beda tergantung dari kebijakan masing-masing lembaga tersebut. Tetapi pada umumnya memiliki persyaratan diantaranya mahasiswa aktif, memiliki prestasi akademik yang tinggi yang di buktikan dengan nilai IPK, memiliki prestasi di bidang tertentu dan tidak sedang menerima bantuan beasiswa dari sumber yang lain.

Dalam mengambil keputusan untuk menentukan mahasiswa yang layak mendapatkan beasiswa tersebut pada prosesnya terkadang membutuhkan waktu yang cukup lama dan terkadang pula beasiswa yang diberikan dirasa kurang tepat sasaran, hal ini dikarenakan proses seleksi masih dilakukan secara manual dan adanya pengolahan data yang belum transparan. Sehingga perlu

dibuatkan sebuah sistem yang dapat membantu pihak lembaga dalam mengambil keputusan dengan cepat agar beasiswa dapat didistribusikan tepat waktu dan tepat sasaran.

2. Tinjauan Pustaka

Tinjauan pustaka yang digunakan dalam penelitian

A. Data Mining

1. Definisi data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. 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 berbagai database besar [Turban 2005].

Menurut gartner group, data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dan sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [Larose 2005]

2. Komponen Data Mining

Secara alami, material data mining sebenarnya sudah terbentuk karena factor rutinitas dan waktu seraya perusahaan melakukan aktivitasnya. Tanpa disadari perusahaan berinvestasi dengan menggunakan budgetnya untuk penggunaan teknologi informasi atau computer. Teknologi *data mining*

mulai muncul karena akumulasi data yang besar dan pesat pertumbuhannya sehingga menimbulkan apa yang disebut *rich of data but poor information*. Tumpukan data ini tidak dapat digunakan pada aplikasi yang ada sehingga menjadi gudang data.

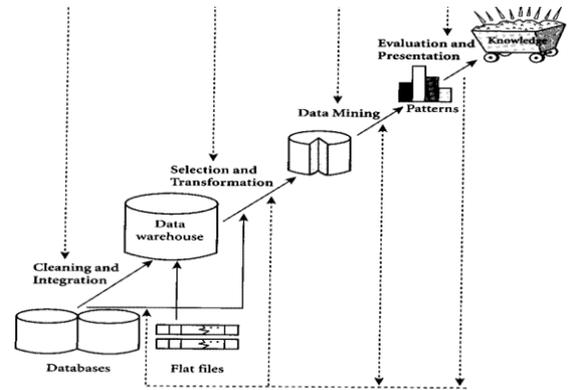
3. Pengelompokan data mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [Iarose, 2005]

- Deskripsi**
Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
- Estimasi**
Estimasi hampir sama dengan klasifikasi, kecuali *variable target*. Estimasi lebih ke arah numerik dari pada ke arah kategori.
- Prediksi**
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa datang.
- Klasifikasi**
Dalam klasifikasi terdapat *target variable* kategori, sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
- Pengklusteran**
Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan record-record dalam cluster.
- Asosiasi**
Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang biasa.

4. Tahapan Data Mining

Tahapan *data mining* dilakukan melalui pembentukan *data warehouse* karena hanya struktur data dari *data warehouse* yang dapat digunakan untuk *data mining*. Dengan melaksanakan ekstraksi, pembersihan dan transformasi (*Extraction, Cleansing, Transformation*) atas data transaksional yang disebut dengan data OnLine Transactional Processing (OLTP) ke OnLine Analytical Processing (OLAP) System untuk membuat sebuah *data warehouse* maka akan didapat suatu kumpulan data yang luas dan besar serta bersifat statis dan historis yang berasal dari data-data transaksi OLTP selama perusahaan menjalankan bisnisnya [Vieira 2000].



Gambar 2.1 Tahapan Data Mining [Pramudiono 2003]

5. Algoritma Naive Bayes

Naive Bayes merupakan metode probabilistik pengklasifikasian sederhana berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien. Naive Bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain. Teorema Bayes sendiri dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Dimana persamaan Teori Bayes tersebut adalah:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Dimana :

X : Data dengan kelas yang belum diketahui

C : Hipotesis data X merupakan suatu kelas spesifik

P (C|X) : Probabilitas hipotesis C berdasar kondisi X (probabilitas posterior)

P(C) : Probabilitas hipotesis C (probabilitas prior)

P(X|C) : Probabilitas X berdasarkan kondisi pada hipotesis C

P(X) : Probabilitas X

Adapun alur dari metode naive Bayes adalah :

- Menghitung nilai peluang kasus baru dari setiap hipotesis dengan kelas yang ada
 - Menghitung nilai akumulasi peluang dari setiap kelas
 - Menghitung nilai
 - Menentukan kelas dari setiap kasus baru tersebut
- Untuk menjelaskan teorema Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes pada persamaan (1) disesuaikan menjadi persamaan (2)

$$P(C|X_1 \dots X_n) = \frac{P(C)P(X_1 \dots X_n|C)}{P(X_1 \dots X_n)} \quad (2)$$

Dimana Variabel C merepresentasikan kelas, sementara variabel $X_1 \dots X_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi atau kriteria. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence).

B. SQL Server

Sql server merupakan database engine yang mendukung penggunaan arsitektur client server. Penggunaan client server bisa digambarkan bahwa aplikasi dibagi menjadi dua bagian yaitu bagian client dan server. Bagian client atau user bisa mendapatkan keuntungan bahwa sql server bisa dijalankan di system operasi personal seperti windows 95 ataupun window ME, dan juga bisa di koneksikan dengan berbagai aplikasi atau bahasa pemrograman dengan bantuan ODBC, sedangkan bagian server atau database administrator (DBA) akan mempunyai database yang handal, konsisten, metode locking dan control secara baik.

Ada dua fasilitas SQL yang digunakan dalam penelitian ini yaitu Enterprise Manager, dan Query Analyzer.

- a. Enterprise manager merupakan utility bantu utama dalam SQL Server. Enterprise manager merupakan interface terintegrasi di dalam SQL Server yang mampu melakukan hampir semua fungsi dan perintah yang kita butuhkan dalam SQL Server.
- b. Query analyzer merupakan wadah dimana data penelitian ini disimpan. Fasilitas ini juga menyediakan fungsi-fungsi untuk melakukan proses-proses yang berhubungan dengan pembuatan table, atribut, dan tipe data.

C. Tinjauan Studi

Beberapa peneliti telah melakukan proses penelitian yang sama Diantaranya:

1. Penelitian yang dilakukan oleh diasrina dahri, fahrul agusdina marisa khairina, Universitas Mulawarman dengan judul “Metode naive bayes untuk penentuan penerimaan beasiswa bidikmisi universitas mulawarman”.

Hasil penelitian adalah: Aplikasi sistem pendukung keputusan untuk seleksi beasiswa bidikmisi menggunakan metode naive bayes.

2. Penelitian yang dilakukan oleh Alfa saleh, Universitas Potensi Utama dengan judul “implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga “

Hasil Penelitian adalah : Prediksi besarnya penggunaan listrik tiap rumah tangga.

3. Penelitian yang dilakukan oleh Mulyadi, STMIK Antar Bangsa dengan judul “Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa Prestasi “

Hasil Penelitian adalah : Penerapan algoritma naive bayes dapat dijadikan rekomendasi dalam menentukan penerima beasiswa prestasi.

D. Tinjauan Objek Penelitian

Universitas Kuningan (UNIKU) didirikan sebagai perwujudan dari idealisme dan komitmen Yayasan Pendidikan Sang Adipati Kuningan untuk terus menerus berkarya khususnya dalam bidang peningkatan sumber daya manusia menuju peningkatan mutu kehidupan masyarakat pada umumnya melalui pendidikan tinggi. Gagasan tentang pendirian Universitas di Kuningan sebetulnya telah ada sejak tahun 1979 ketika Yayasan ini didirikan. Namun karena keterbatasan sumber daya, baik sumber daya manusia maupun sumber daya keuangan, maka niat itu tidak bisa langsung diwujudkan sekaligus. Mengingat keterbatasan itu yayasan menggunakan strategi bertahap melalui pendirian sekolah tinggi sebagai cikal bakal berdirinya universitas.

Sekolah tinggi pertama kali lahir pada tahun 1985 yaitu Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Kuningan. Sekolah tinggi ini sebelum menjadi Universitas Kuningan mempunyai tiga program Studi yaitu Pendidikan Bahasa dan Sastra Indonesia (S1), Pendidikan Biologi (S1), dan Pendidikan Ekonomi (S1).

Sekolah tinggi kedua lahir tahun 1995 yaitu Sekolah Tinggi Ilmu Ekonomi (STIE) Kuningan, dengan 2 (dua) program studi yaitu program Studi Manajemen (S1) dan Program Studi Akuntansi (S1).

Sekolah tinggi ketiga lahir pada tahun 2001; yaitu Sekolah Tinggi Ilmu Kehutanan (STIK) Kuningan dan Sekolah Tinggi Kehutanan (STIK) Kuningan dengan 2 (dua) Program Studi yaitu Program Studi Budidaya Hutan (S1). Karena dirasakan adanya tuntutan perkembangan teknologi terutama computer maka pada waktu yang bersamaan lahir pula Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Kuningan dengan 4 program studi yaitu Program Studi Teknik Informatika (S1),

Sistem Informasi (SI), Teknik (DIII) dan Manajemen Informatika (DIII).

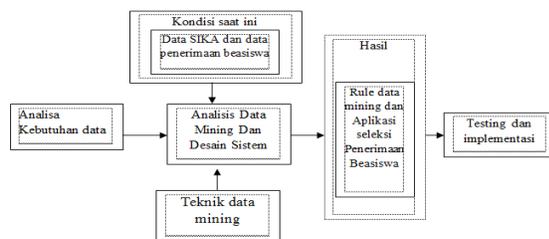
Setelah memiliki empat sekolah tinggi, yayasan merasa punya cukup pengalaman untuk mendirikan universitas. Oleh karena itu pengembangan dan penggabungan keempat sekolah tinggi lantas dilaksanakan. Berdasarkan Surat Keputusan Menteri Pendidikan Nasional Nomor : 62/D/O/2003 tanggal 6 Juni 2003, berdirilah Universitas Kuningan diharapkan yang merupakan penggabungan keempat sekolah tinggi yang ada dibawah naungan Yayasan Pendidikan Sang Adipati Kuningan. Universitas ini diresmikan oleh Menteri Pendidikan Nasional Prof. A. Malik Fajar, M.Sc. pada tanggal 17 Juni 2003.

Pada tahun 2006 Universitas Kuningan membuka dua program studi baru, yaitu Program Studi Pendidikan Bahasa Inggris Jenjang S1, dan Program Studi Pendidikan Ekonomi Jenjang S2 (Magister).

E. Kerangka Konsep

Sesuai dengan tujuan penelitian yang telah di tetapkan maka peneliti ingin melakukan analisa data mining pada sistem informasi akademik yang ada dengan konsep yang menurut penulis cocok digunakan di Universitas Kuningan. Sehingga pihak manajemen kampus dapat mengambil keputusan dengan cepat.

kerangka konsep di atas dapat di gambarkan seperti berikut ini :



Gambar 2.2 Kerangka pemikiran

Kerangka Konsep Diatas Dapat jelaskan sebagai berikut :

Desain data mining di buat berdasarkan kebutuhan data untuk system seleksi penerimaan beasiswa. Data mining ini dilakukan dengan cara melakukan analisis data operasional pada data penerimaan beasiswa dan Data SIKa (Sistem Informasi Akademik) menggunakan teknik data mining naive bayes. Hasil dari desain diharapkan akan menghasilkan sebuah sistem informasi berupa program aplikasi seleksi penerimaan beasiswa

3. Desain Penelitian

1. Metode Penelitian

Agar penelitian yang dilakukan lebih terarah maka peneliti menggunakan suatu metode deskriptif kualitatif yang bertujuan untuk mendapatkan gambaran lengkap dari proses dan subjek penelitian. Dalam penelitian ini metode yang digunakan adalah *waterfall*. Langkah yang dilakukan dimulai dengan identifikasi dan analisis kebutuhan pengguna

2. Metode Pengumpulan Data

Data untuk penelitian ini diambil dari data primer dan data skunder. Data primer di dapat dari wawancara dan observasi lapangan. Data sekunder di dapat dari data objek penelitian (database), studi literature dan tulisan ilmiah tentang data mining.

3. Teknik Analisis Data

Teknik analisa yang dilakukan yaitu Analisis Data Mining Dan hasilnya diimplementasikan menjadi sebuah Program Aplikasi penilaian Penerimaan mahasiswa baru.

4. Hasil Penelitian

4.1 Sumber Data

Data yang di ambil untuk penelitian ini mengambil sebagian dari database penerimaan mahasiswa baru, data akademik mahasiswa angkatan 2010 dan file excel penerimaan beasiswa tahun 2012. Adapun data nilai mahasiswa yang diambil dari data akademik adalah nilai IPK mahasiswa untuk semester ke 4 dengan asumsi pada semester tersebut mahasiswa sudah bisa mengajukan diri untuk mendapatkan beasiswa.

Berikut ini adalah Data mentah yang digunakan oleh peneliti untuk proses perancangan data

Tabel 4.1 Tabel Penerima Beasiswa Tahun 2012

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	NPM	KEMPT	JENIS BEASISWA	COUNTER	NAMA BMS	JK	KODE PRODI	ID	JENJANG	SMT	IPK	KODE PERUBAHAN	JML. TANGGUNGAN	PENGHASILAN PRESTASI
3	2010011098	041038		1	1. Suci Nurwarahmah	2	88201	51	4	3,50	6		3	900000
4	2010011138	041038		1	2. Eka Wirian	1	88201	51	4	3,04	2		5	900000
5	2010031064	041038		3	Lodan Herifah	2	84205	51	4	3,73	3		3	500000
6	2010031011	041038		4	Anjan Luthi Ferietha	1	84205	51	4	3,01	1		3	3235000
7	2009031054	041038		1	5. Ficht Fitrianti	2	84205	51	6	3,41	5		5	450000
8	2010021151	041038		6	Luthi Rahmawati	2	87203	51	4	3,76	1		5	3807000
9	2010021051	041038		7	Dini Wachulima	2	87203	51	4	3,74	2		5	1000000
10	2010021079	041038		8	Gugun Sunawan	1	87203	51	4	3,66	3		2	600000
11	2009021096	041038		9	Alex Khaerunnisa	2	87203	51	6	3,59	6		2	1442951
12	2010041175	041038		10	Nurendah	2	88203	51	4	3,87	3		4	1500000
13	2010041111	041038		11	Jaruki Andriansyah M.	1	88203	51	4	3,86	6		3	600000
14	2010041054	041038		12	Ayeng Winda Sri W.	2	88203	51	4	3,70	1		1	4130000
15	2009041016	041038		13	Ambar Pramuita S.	2	88203	51	6	3,83	1		4	5348853
16	2010031138	041038		14	Sri Wahyu Rini	2	61201	51	4	3,69	1		2	2394531
17	2009031069	041038		15	Muhammad Adam	1	61201	51	6	3,54	1		2	4774453
18	2010061019	041038		16	Amir Hamzah	1	62201	51	4	3,89	3		3	400000
19	2010061087	041038		17	Sinta Juni Nur Hibah M.	2	62201	51	4	3,72	1		2	1500000
20	2009071041	041038		18	Aggrita	2	54251	51	6	3,46	3		2	1000000
21	2010031139	041038		19	Sarif Widayat	1	55201	51	4	4,00	3		2	1000000
22	2010031037	041038		20	Dan Febriansyah	1	55201	51	4	3,96	5		7	800000
23	2009031095	041038		21	Ricky Pineddy Syam	1	57201	51	6	3,79	1		3	2754400
24	2009031044	041038		22	Sally Agustien Nurdaya	1	57201	51	6	3,70	2		3	1890000
25	2010101052	041038		23	Debra Prihatiningtyasi	2	55401	3	4	3,84	6		4	1000000

Tabel 4.2 Tabel Penerimaan Mahasiswa Baru

Nama Tabel	Primary Key	Keterangan
Tabel msclnmhs	No_daftar	Berisi data calon mahasiswa, pekerjaan orang tua, penghasilan orang tua dan registrasi mahasiswa

Tabel 4.3 Tabel Akademik

Nama Tabel	Primary Key	Keterangan
Tabel msmhs	Nimmhs	Digunakan untuk menyimpan data mahasiswa aktif
Tabel mspst	Kdpstmspst	Digunakan untuk menyimpan data program studi
Tabel nilai	Nilai	Digunakan untuk menyimpan bobot nilai mahasiswa
Tabel trdetailkrs	Iddetkrs	Digunakan untuk menyimpan detail data Krs dan nilai
Tabel trkrs	Id_krs	Digunakan untuk menyimpan data nilai IPK dan IPS mahasiswa

Data diatas kemudian di olah dengan mengambil beberapa variabel yang diperlukan untuk proses penelitian. Dan dilakukan proses integrasi dan penyamaan nama field diantaranya NPM menjadi NIM, nama_mhs menjadi nama_mahasiswa, kd_prodi menjadi program_studi, ipk, kode_pekerjaan menjadi pekerjaan, jml_tanggungan, dan penghasilan. Dari data diatas diperoleh data sebanyak 1086 data, tetapi terdapat duplikasi sehingga kemudian di lakukan penghapusan dan pembersihan data sehingga menjadi hanya 984 data yang bisa dijadikan sebagai data dasar untuk proses data mining. Hasil data dilengkapi dengan menambahkan kolom baru status beasiswa untuk mendapatkan hasil layak atau tidak layak sehingga menghasilkan data sebagai berikut :

Tabel 4.4 Data Mentah Penerimaan Beasiswa

Proses selanjutnya adalah melakukan proses penambahan data pada data beasiswa. Field data yang akan digunakan sebagai variabel penentu adalah IPK, Pekerjaan, jml_tanggungan dan penghasilan.

4.2 Melakukan Preprocessing

Pada tahap ini akan dilakukan proses perubahan data mentah menjadi data yang mudah dikelola. Berdasarkan data yang didapat, sebagian data berupa nilai dan angka sehingga untuk memudahkan proses penambahan data (data mining) maka data tersebut dikelompokkan berdasarkan jangkauan tertentu. Hal ini untuk memudahkan pengkategorian nilai. Proses transformasi data dilakukan untuk data IPK, Pekerjaan, dan penghasilan. Berdasarkan pemilihan variable-variabel tersebut, maka format data akan dibuat menjadi seperti berikut ini :

Untuk aturan penilaian IPK adalah sebagai berikut :

Table 4.5 Aturan penilaian IPK

No	Aturan	Kategori
1	IPK <= 2.75	Kurang
2	IPK < 2.76 - 3.50	Cukup
3	IPK >= 3.51	Baik

Untuk aturan penilaian nilai Pekerjaan adalah sebagai berikut :

Tabel 4.6 Aturan nilai pekerjaan

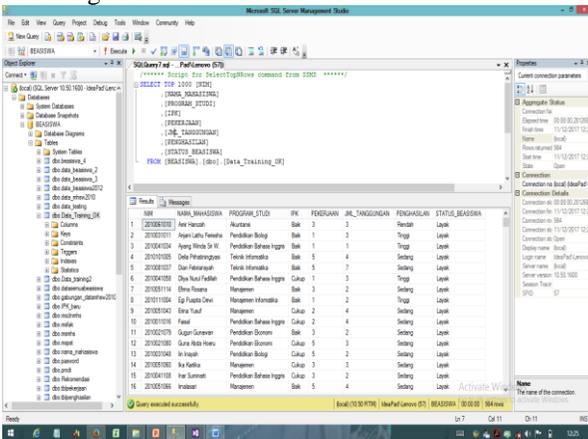
No	Aturan	Kategori
1	PNS/Pegawai negara	1
2	Pegawai swasta	2
3	Wiraswasta	3
4	Anggota TNI/Polri	4
5	Petani /Nelayan	5
6	Lainya	6

Untuk aturan penilaian nilai Penghasilan adalah sebagai berikut :

Tabel 4.7 Aturan nilai penghasilan

No	Aturan	Kategori
1	< Rp. 500.000,-	Rendah
2	Rp. 500.000 ,- s.d. Rp 2.000.000,-	Sedang
3	> Rp. 2.000.000,- s.d. Rp. 5.000.000,-	Tinggi

Sehingga pemrosesan data akhir menghasilkan data sebagai berikut :



Gambar 4.1 Informasi Data Training

Data diatas akan dijadikan sebagai data dasar untuk proses perhitungan menggunakan algoritma naive bayes classifier .

4.3 Perhitungan Nilai

Proses berikutnya adalah menghitung nilai maksimal posterior class untuk menentukan status basiswa layak atau tidak layak

Jika ada mahasiswa yang mengajukan basiswa dengan data sebagai berikut maka apakah mahasiswa layak atau tidak layak mendapatkan basiswa?

Tabel 4.8 Data Testing Mahasiswa

Nim	Nama Mahasiswa	Program Studi	IPK	Pekerjaan	Jml tanggungan	Penghasilan	Status Basiswa
20160810053	Marwan Dhiaur Rahman	Teknik Informatika	3,86	PNS/Pegawai Negara	4	RP > 2000000-5000000	?

Untuk menjawab Pertanyaan diatas terlebih dahulu dilakukan proses perubahan data menjadi seperti berikut :

Tabel 4.9 Preprocessing Data Testing Mahasiswa

Nim	Nama Mahasiswa	Program Studi	IPK	Pekerjaan	Jml tanggungan	Penghasilan	Status Basiswa
20160810053	Marwan Dhiaur Rahman	Teknik Informatika	Baik	1	4	Tinggi	?

Proses Perhitunganya adalah sebagai berikut :

1. Menghitung jumlah class atau label

Tahap 1 menghitung jumlah class/label

$P(Y= Layak) = 34/984 = \text{'jumlah data "layak" pada kolom status basiswa dibagi jumlah data} = 0,03$

$P(Y= Tidak layak) = 950/984 = \text{'jumlah data "Tidak layak" pada kolom status basiswa' dibagi jumlah data} = 0,96$

2.Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

$P(IPK = Baik | Y= Layak) = 19/34 = 0,55$

'jumlah data ipk= "Baik" dengan keterangan "layak dibagi jumlah data layak

$P(IPK = Baik | Y= Tidak Layak) = 42/950 = 0,04$

'jumlah data ipk= "Baik" dengan keterangan "tidak layak dibagi jumlah data tidak layak

$P(jml_tanggungan = 4 | Y= Layak) = 10/34 = 0,29$

'jumlah data tanggungan= '4' dengan keterangan "layak" dibagi jumlah data layak

$P(jml_tanggungan = 4 | Y= Tidak Layak) = 0/950 = 0$

'jumlah data tanggungan='4' dengan keterangan "tidak layak" dibagi jumlah data tidak layak

$P(pekerjaan = 1 | Y= Layak) = 7/34 = 0,20$

'jumlah data pekerjaan= '1' dengan keterangan "layak" dibagi jumlah data layak

$P(pekerjaan = 1 | Y= Tidak Layak) = 240/950 = 0,25$

'jumlah data pekerjaan='1' dengan keterangan "tidak layak" dibagi jumlah data tidak layak

$P(penghasilan = Tinggi| Y=Layak)= 6/34 = 0,17$

'jumlah data penghasilan= 'Tinggi' dengan keterangan "layak" dibagi jumlah data layak

$P(Penghasilan =Tinggi | Y=tidak Layak) = 19/950 = 0,02$

'jumlah data penghasilan= 'Tinggi' dengan keterangan "tidak layak" dibagi jumlah data tidak layak

- Kalikan semua hasil variabel layak dan tidak layak
 $P(IPK = Baik), (Pekerjaan = 1), (jml_tanggung\ respon = 4), (penghasilan = Tinggi)$
 $= \{P(P(IPK = Baik | Y = layak), (Pekerjaan = 1 | Y = layak), (jml_tanggung\ respon = 4 | Y = layak), (penghasilan = Tinggi | Y = layak))\}$
 $= 0,55 \times 0,20 \times 0,17 \times 0,29 \times 0,03 = 0,000561$
 $= \{P(P(IPK = Baik | Y = tidak layak), (Pekerjaan = 1 | Y = tidak layak), (jml_pekerjaan = 4 | Y = tidak layak), (penghasilan = Tinggi | Y = tidak layak))\}$
 $= 0,04 \times 0,25 \times 0,02 \times 0 \times 0,96 = 0$
- Bandingkan hasil class layak dan tidak layak
 $P(Layak = 0,000206334116058441)$
 $P(Tidak Layak = 0)$
 Karena hasil $(P | layak)$ lebih besar dari $(P | Tidak layak)$ maka keputusannya adalah “Layak”

Tabel 4.10 Hasil Rekomendasi

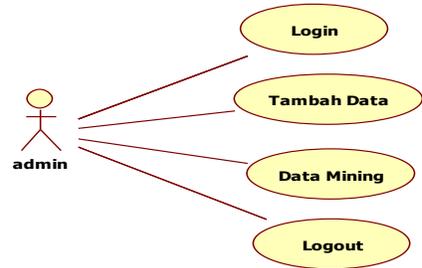
Nim	Nama Mahasiswa	Program Studi	IPK	Pekerjaan	Jml tanggungan	Penghasilan	Status Beasiswa
20160810053	Marwan Dhiaur Rahman	Teknik Informatika	3,86	PNS/Pegawai Negara	4	RP > 2000000-5000000	Layak

4.4 Perancangan Sistem

Dalam melakukan proses perancangan ini, peneliti menggunakan diagram-diagram UML untuk memberikan gambaran terintegrasi terhadap sistem yang akan dibangun. Beberapa diagram UML yang penulis gunakan yaitu use case diagram, Aktivitas Diagram, dan Sekuensial Diagram. Adapun proses Perancangannya adalah sebagai berikut ini :

4.4.1 Diagram Use Case

Use case adalah fungsionalitas atau persyaratan-persyaratan sistem yang harus di penuhi oleh sistem yang akan dikembangkan tersebut menurut pandangan pemakai sistem. Sedangkan actor bisa berupa orang, peralatan, atau sistem lain yang berinteraksi terhadap sistem yang akan dibangun. [Sholiq 2010]. Dibawah ini merupakan penggambaran umum mengenai use case yang terdapat dalam sistem informasi klasifikasi latar belakang mahasiswa berdasarkan prestasi akademiknya

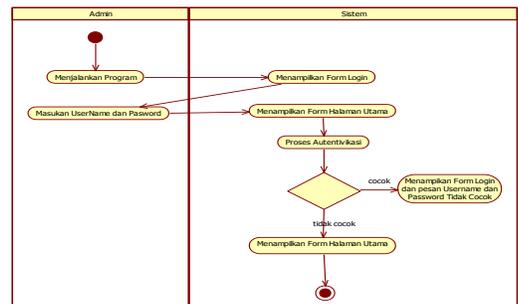


Gambar 4.2 Use case data rekomendasi penerima beasiswa

4.4.2 Diagram Aktivitas

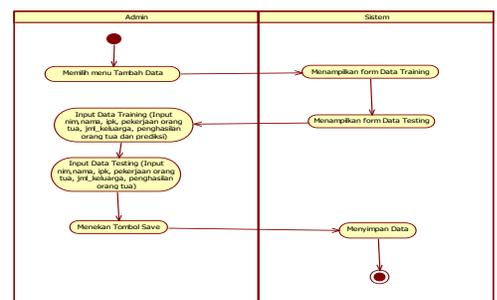
Diagram aktivitas merupakan diagram yang menjelaskan aktivitas pengguna dengan program. Berikut ini adalah diagram aktivitas pada pengolahan data akademik dan PMB

a. Diagram aktivitas login



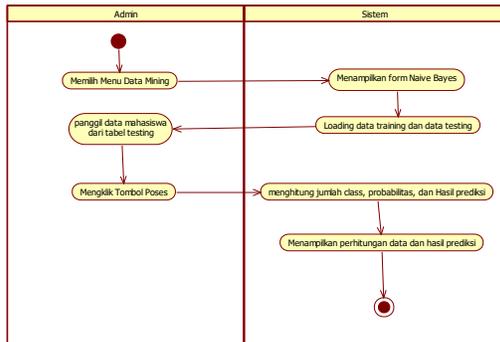
Gambar 4.3 Diagram Aktivitas login

b. Diagram aktivitas tambah data



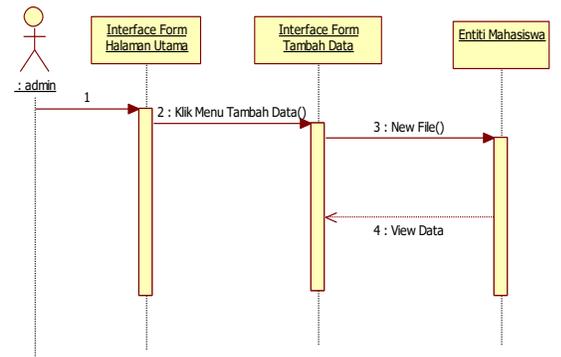
Gambar 4.4 Diagram Aktivitas Tambah Data

c. Diagram aktivitas Data mining



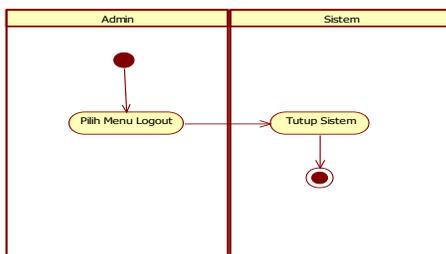
Gambar 4.5 Diagram Aktivitas Data Mining

b. Diagram sekuensial tambah data



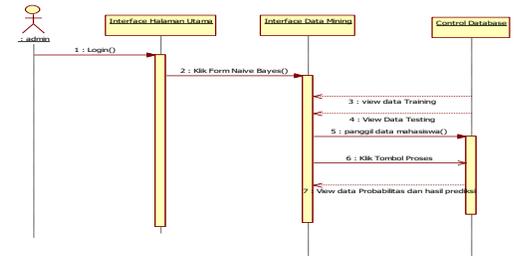
Gambar 4.8 Diagram Sekuensial Tambah Data

d. Diagram aktivitas logout



Gambar 4.6 Diagram Aktivitas Logout

c. Diagram sekuensial Data mining



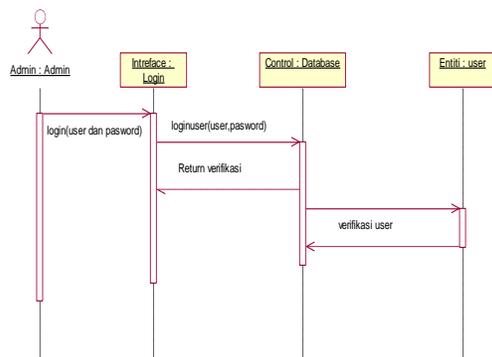
Gambar 4.9 Diagram Sekuensial Data Mining

4.4.3 Diagram Sekuensial

Diagram sekuensial adalah diagram interaksi yang disusun berdasarkan urutan waktu.

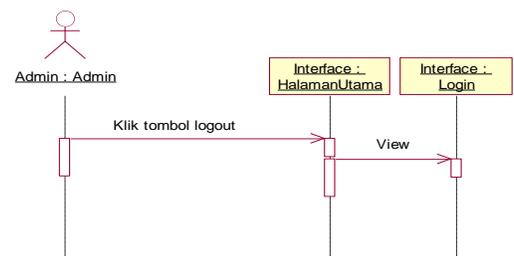
Berikut ini adalah diagram sekuensial dari sistem yang akan dibangun :

a. Diagram sekuensial login



Gambar 4.7 Diagram Sekuensial Login

d. Diagram sekuensial logout



Gambar 4.10 Diagram Sekuensial Logout

4.4 Implementasi Program

Sistem Informasi Rekomendasi Penerimaan Beasiswa Mahasiswa sudah berhasil di bangun. Berikut ini merupakan tampilan antar muka dari sistem tersebut

a. Form Login



Gambar 4.11 Form Login

Form ini merupakan form yang digunakan untuk dapat memasuki program Sistem informasi rekomendasi. Form ini terdiri atas input user name dan password. Hanya administrator yang telah diberi hak akses yang dapat memasuki program ini

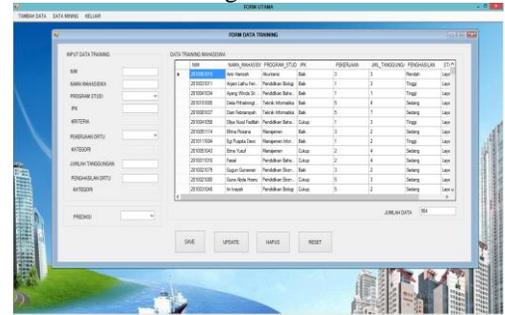
b. Form Menu Utama



Gambar 4.12 Form Menu Utama

Untuk mengendalikan jalanya program peneliti membuat sebuah form induk yang dibuat dengan menggunakan MDIform dengan nama Form_utama. Form ini terdiri dari 3 buah menu utama yang terdiri dari Tambah Data , Data mining, dan Exit. Menu Tambah Data terdiri dari sub menu Data Training dan Data Testing sedangkan Data Mining terdiri dari sub menu Proses Perhitungan serta Keluar terdiri dari sub menu logout. Tugas utama dari form ini adalah mengendalikan seluruh jalanya program dan melakukan pemanggilan pada form-form anak.

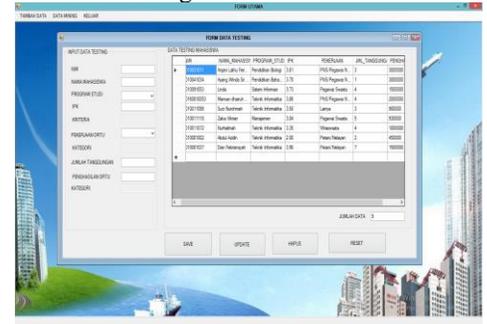
c. Form Data Training



Gambar 4.13 Form Data Training

Form ini digunakan untuk melakukan proses perhitungan naive bayes yang nanti akan membentuk rule dan diterapkan pada proses berikutnya. Form ini terdiri dari tombol save, update , hapus dan reset.

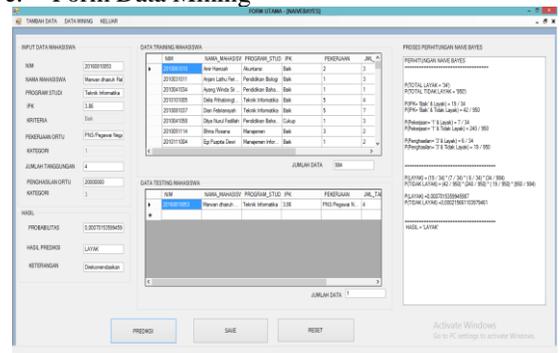
d. Form Data Testing



Gambar 4.14 Form Data Testing

Form ini digunakan untuk melakukan input data mahasiswa yang dipakai untuk keperluan data testing yang nantinya akan diproses dan menghasilkan hasil rekomendasi beasiswa dari masing-masing mahasiswa. Form ini terdiri dari tombol save, update , hapus dan reset.

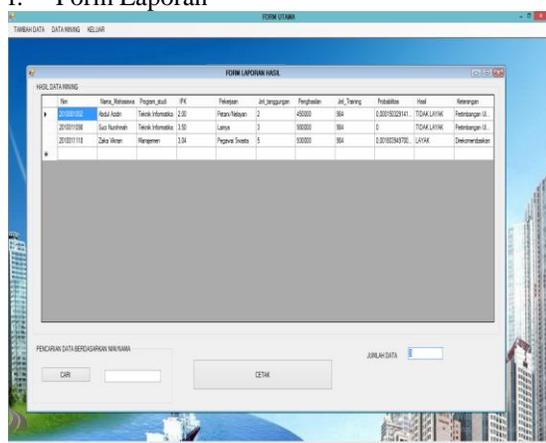
e. Form Data Mining



Gambar 4.15 Form Data Mining

Form ini digunakan untuk melakukan perhitungan data mining. Dimana form ini akan menampilkan data training dan data testing. Data testing yang akan di proses dipanggil dan ditampilkan pada kolom input data mahasiswa, kemudian data tersebut dilakukan proses perhitungan sesuai aturan algoritma naive bayes classifier dengan menekan tombol prediksi untuk menghasilkan keputusan Layak atau tidak Layak. Data yang layak akan diberi keterangan “Direkomendasikan” dan data yang tidak layak akan diberikan keterangan “ Pertimbangan Ulang” dan Proses perhitungan data akan ditampilkan pada kolom proses perhitungan naive bayes.

f. Form Laporan



Gambar 4.16 Form Laporan

Form ini menampilkan laporan dari hasil data mining dan dapat digunakan untuk melihat hasil rekomendasi sistem apakah calon peserta tersebut layak/tidak layak mendapatkan beasiswa.

5. Kesimpulan Dan Saran

Kesimpulan dan saran hasil penelitian

5.1 Kesimpulan.

1. Metodologi yang digunakan penulis dalam melakukan perhitungan dan pengukuran akurasi terhadap model data mining cukup mudah untuk diimplementasikan.
2. Pendekatan naive bayes dapat menjawab apa yang menjadi kebutuhan manajemen yaitu memprediksi kelayakan mahasiswa menerima beasiswa dan dapat dijadikan dasar untuk rekomendasi.
3. Jumlah data training dapat mempengaruhi tingkat akurasi.

5.2 Saran

5. Data yang digunakan memiliki variabel yang kurang lengkap diantaranya tidak ada kelas pembagian untuk mahasiswa yang bekerja dan tidak bekerja, kelas karyawan dan kelas reguler. Hal ini berimbas pada perhitungan akurasi data training sehingga untuk penelitian lebih lanjut agar menambahkan variabel yang lebih lengkap.
6. Dalam penelitian ini menggunakan 4 kategori untuk proses penyeleksian mahasiswa penerima beasiswa. Untuk penelitian selanjutnya dapat ditambah lagi lebih banyak kategori untuk hasil yang lebih optimal.
7. Untuk mengetahui hasil akurasi perhitungan naive bayes, pada penelitian selanjutnya dapat dibandingkan menggunakan metode yang lain.

DAFTAR PUSTAKA

- [Dahri 2016] Diasrina Dahri, Fahrul Agus, Dyna Marisa Khairina, “Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman” , Jurnal Informatika Mulawarman Vol. 11, No. 2, September 2016 29 ISSN 1858-4853
- [Jogiyanto 2008] Prof. Jogiyanto HM, Akt, MBA, Ph.D, “Metodologi Penelitian Sistem Informasi”, Penerbit Andi, Yogyakarta 2008.
- [Kusrini 2009] Kusrini, dan Luthfi Emha Taufiq, “Algoritma Data Mining”, Penerbit Andi, Yogyakarta 2009
- [Kuswanto 2016] Kuswanto, Ayu Milati Nur Azizah, “Sistem Pendukung Keputusan Penyeleksian Mahasiswa Penerima Beasiswa Menggunakan Naive Bayes Berbasis Web”, Jurnal Antivirus, Vol 10 No. 1 Mei 2016 p-ISSN: 1978-5232 e-ISSN: 2527-337X
- [Larose 2005] Larose, Daniel T. “Discovering Knowledge ind Data :An Introduction to Data Mining. John Willey & Sons.inc, 2005
- [Mulyadi 2016] Mulyadi, “ Penerapan Algoritma Naive Bayes untuk Klasifikasi Penerima Beasiswa Prestasi”. Jurnal Sistem Informasi Vol V. No 2 - Agustus 2016.
- [Rahman 2016] Fatur Rahman, Muhammad iqbal Firdaus, “Penerapan Data Mining Metode Naive Bayes Untuk Prediksi Hasil Belajar Siswa Sekolah Menengah Pertama (SMP)”, *Al Ulum Sains dan Teknologi Vol.1 No.2 Mei 2016*
- [Rizky 2004] Rizky, Soetam, “Panduan Belajar SQL Server”, Prestasi Pustaka Publisher , Jakarta 2004
- [Saleh 2015] Alfa Saleh, “Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga” , Citec Journal,

Vol. 2, No. 3, Mei 2015 – Juli 2015 ISSN: 2354-5771

- [Sulianta 2010] Sulianta, Feri dan Juj, Dominikus, “Data Mining, Meramalkan Bisnis Perusahaan ” , Elex Media Komputindo, 2010
- [Suyanto 2017] Dr Suyanto, S.T., M.Sc., “Data Mining untuk Klasifikasi dan Klasterisasi Data”, Informatika, 2017
- [Turban 2005] Turban, E.,dkk , “Decision Support System And Intelegent System”, Andi , Yogyakarta,2005