

Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan

Aah Sumiah¹⁾, Nita Mirantika²⁾

*^{1,2)} Fakultas Ilmu Komputer, Universitas Kuningan
Jl. Cut Nyak Dhien, Cijoho - Kuningan*

Email : aah.sumiah2020@gmail.com¹⁾, nita.mirantika@uniku.ac.id¹⁾

Abstrak

Beberapa tahun terakhir, data semakin heterogen dan kompleks dengan volume yang meningkat cepat hingga diperkirakan akan mencapai 44 zettabyte di tahun 2020 (turner 2014). Hal ini sering disebut dengan Big data. Era Big data menghasilkan data yang menumpuk sehingga perlu dilakukan pengolahan untuk mencari knowledge dari tumpukan data tersebut menggunakan data mining. Penelitian ini merupakan penelitian lanjutan dari penelitian sebelumnya yang berjudul “ Implementasi data Mining untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa menggunakan Metode Naive Bayes Studi Kasus Universitas Kuningan”. Pada penelitian ini mencoba membandingkan dua algoritma untuk mengetahui algoritma mana yang paling cocok digunakan untuk rekomendasi penentuan mahasiswa penerima beasiswa pada universitas kuningan menggunakan algoritma K-Nearest Neighbor dan algoritma Naive Bayes. Metode ini di pilih karena kedua algoritma merupakan algoritma yang populer digunakan dalam proses pengklasifikasian data Hasil analisis data di implementasikan menjadi sebuah sistem informasi menggunakan visual basic.net dan sql server yang dapat digunakan oleh bagian akademik sebagai rekomendasi dalam proses seleksi penerimaan beasiswa.

Kata Kunci : sistem informasi; data mining; beasiswa; naive bayes classifier; K-Nearest Neighbor; visual basic.net; sql server

Abstract

In recent years, data has become increasingly heterogeneous and complex with volumes increasing rapidly until it is expected to reach 44 zettabytes by 2020 (turner 2014). This is often referred to as Big Data. The era of Big Data generates data that is piling up, so it needs to be processed to find knowledge from the pile of data using data mining. This research is a continuation of previous research entitled "Implementation of Mining Data for Recommendations for Determining Scholarship Recipients using the Naive Bayes Method of Case Study at Kuningan University". In this study, trying to compare two algorithms to find out which algorithm is the most suitable for the recommendation of determining scholarship recipients at a brass university using the K-Nearest Neighbor algorithm and the Naive Bayes algorithm. This method was chosen because both algorithms are popular algorithms used in the process of classifying data. The results of data analysis are implemented into an information system using visual basic.net and sql server that can be used by the academic department as a recommendation in the selection process for scholarship acceptance.

Keywords: Information Systems; data mining; scholarship; naive bayes classifier; K-Nearest Neighbor; visual basic.net; sql server

1. PENDAHULUAN

Beberapa tahun terakhir, data semakin heterogen dan kompleks dengan volume yang meningkat cepat hingga diperkirakan akan mencapai 44 zettabyte di tahun 2020 (turner 2014). Hal ini sering disebut dengan Big data. Era Big data menghasilkan data yang menumpuk sehingga perlu

dilakukan pengolahan untuk mencari knowledge dari tumpukan data tersebut. Hal yang dilakukan adalah dengan Data mining. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya. [Pramudiono 2006]

Kemajuan yang luar biasa dari datamining yang didorong oleh faktor pertumbuhan yang cepat dari penyimpanan data dalam database memungkinkan memberikan indikasi yang bermanfaat bagi pemilik data. Beragam riset dan pengembangan data mining telah memberikan banyak produk yang berguna bagi masyarakat luas. Contohnya di bidang perbankan dimana pihak bank sudah bisa memprediksi mana nasabah yang bermasalah dalam kredit dan mana nasabah yang tidak bermasalah. Di bidang akademik misalnya pihak manajemen kampus sudah bisa memprediksi mahasiswa yang akan melakukan registrasi dan mahasiswa yang tidak akan melakukan registrasi, sehingga pengunduran mahasiswa baru bisa diantisipasi.

Perusahaan komersial maupun non komersial berskala menengah ke atas saat ini sudah seharusnya memanfaatkan data mining untuk membantu prospek bisnis perusahaan. Jika tidak, manajerial yang dilakukan tidak ada bedanya dengan perusahaan tradisional yang beralih menggunakan komputer yang hanya menggantikan cara-cara manual.

Pada penelitian ini peneliti mencoba membandingkan dua algoritma untuk mengetahui algoritma mana yang paling cocok digunakan untuk rekomendasi penentuan mahasiswa penerima beasiswa pada universitas kuning. Metode yang di pilih adalah algoritma K-Nearest Neighbor dan algoritma Naive Bayes. Metode ini di pilih karena kedua algoritma merupakan algoritma yang populer digunakan dalam proses pengklasifikasian data. Peneliti tertarik untuk melihat tingkat akurasi data dari kedua metode terhadap data penerimaan beasiswa.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine elarning untuk mengekstrasi 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 statistic dan matematika [Larose 2005]

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

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

a. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target. Estimasi lebih kearah numeric dari pada kearah kategori.

c. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa datang.

d. Klasifikasi

Dalam kalsifikasi terdapat target variable kategori, sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. Pengklusteran

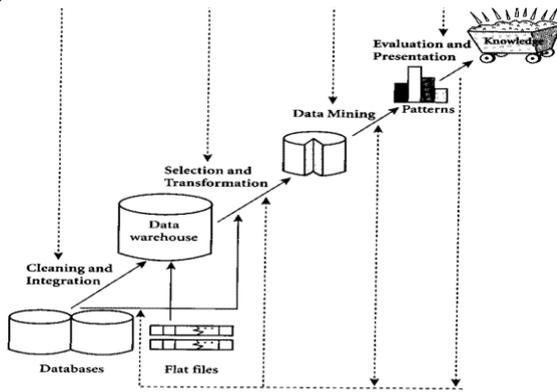
Pengkulusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan record-record dalam cluster.

f. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang biasa

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 1 Tahapan Data Mining [Pramudiono 2003]

2.2 K-NN

Naïve Bayes merupakan metode probabilistik pengklasifikasian sederhana berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien. Naïve 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

2.3 Decision Tree

Algoritma C.45 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Data didalam pohon keputusan biasanya dinyatakan dalam bentuk table dengan atribut dan record. atribut menyatakan suatu parameter yang dibuat sebagai criteria dalam pembentukan pohon. Misalnya untuk menentukan main tennis, criteria yang di perhatikan adalah cuaca, angin, temperature. Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance. Misalkan atribut cuaca mempunyai instance berupa cerah, berawan, dan hujan [Basuki & Syarif, 2003].

2.4 Confusion Matrik

Confusion matrik adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining . confusion matrik adalah media yang berguna untuk menganalisis seberapa baik classifier dapat mengenali tupel dari kelas yang berbeda.

2.5 Penelitian Sebelumnya

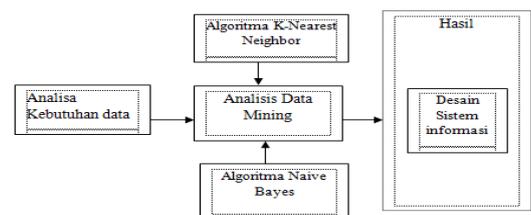
Beberapa peneliti telah melakukan proses penelitian yang sama Diantaranya:

1. Penelitian paulus dian wicaksana, Universitas Sanata Dharma dengan judul “Perbandingan K-Nearest Neighbor dan Naive Bayes untuk Studi data ”wisconsin Diagnosi Breast Cancer”. Hasil penelitian adalah: Mengukur Tingkat akurasi Data menggunakan metode naive bayes.
2. Penelitian yang dilakukan oleh Yusra, Dita Olivita, Yelfi Fitriani, UIN Sultan Syarif Kasim Riau, dengan judul “Perbandingan klasifikasi Tugas Akhir Mahasiswa jurusan Teknik Informatika Menggunakan Metode Naive Bayes Classifier dan K-Nearest Neighbor “. Hasil Penelitian adalah : Menghasilkan Tingkat Akurasi data
3. Penelitian yang dilakukan oleh Riyan Eko putri, suparti, Rita Rahmawati, Universitas Diponegoro, dengan judul “Perbandingan Metode Klasifikasi Naive Bayes dan K-nearest Neighbor Pada Analisis data status Kerja di Kabupaten Demak tahun 2012“. Hasil Penelitian adalah : Klasifikasi data pada data status Kerja Kabupaten Demak.

2.6 Kerangka Konsep

Sesuai dengan tujuan penelitian yang telah di tetapkan maka peneliti ingin melakukan perbandingan tingkat akurasi algortima K-Nearest Neighbor dan algoritma Decision tree pada data penerimaan beasiswa.

kerangka konsep di atas dapat di gambarkan seperti berikut ini :



Gambar 2 Kerangka pemikiran

Kerangka Konsep Diatas Dapat jelaskan sebagai berikut :

Desain data mining di hasilkan berdasarkan kebutuhan data untuk system rekomendasi penentuan mahasiswa penerimaan beasiswa. Data mining ini dilakukan dengan cara mengolah data operasional pada data penerimaan beasiswa dan Data SIKA (Sistem Informasi Akademik) dan melakukan Analisa data menggunakan algoritma K-Nearest Neighbor dan naive bayes. Kemudian

melakukan perbandingan tingkat akurasi data menggunakan confusion matrix dan membuat rancangan sistem informasi

2.7 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.

2. Metode Pengumpulan Data

Data untuk penelitian ini diambil dari data primer dan data sekunder. 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

Salah satu cara untuk melakukan analisis data akademik untuk rekomendasi penentuan mahasiswa penerima beasiswa adalah dengan melakukan klasifikasi data variabel dari kumpulan data penerima beasiswa yang sudah ada. analisa ini menggunakan Algoritma K-nearest Neighbor dan Naive Bayes yang kemudian di lakukan perhitungan data menggunakan excel dan tools microsoft SQL Server 2008 serta menggunakan bantuan tools weka untuk melakukan pengukuran tingkat akurasi data menggunakan confusion matrik

3. HASIL DAN PEMBAHASAN

3.1 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.

a. Tabel Penerimaan Beasiswa Tahun 2012

Tabel 1 Tabel Penerima Beasiswa Tahun 2012

NO	NAMA	KELOMPOK	IPK	STATUS
1	ABDULLAH	1	3.50	1
2	ABDULLAH	1	3.50	1
3	ABDULLAH	1	3.50	1
4	ABDULLAH	1	3.50	1
5	ABDULLAH	1	3.50	1
6	ABDULLAH	1	3.50	1
7	ABDULLAH	1	3.50	1
8	ABDULLAH	1	3.50	1
9	ABDULLAH	1	3.50	1
10	ABDULLAH	1	3.50	1
11	ABDULLAH	1	3.50	1
12	ABDULLAH	1	3.50	1
13	ABDULLAH	1	3.50	1
14	ABDULLAH	1	3.50	1
15	ABDULLAH	1	3.50	1
16	ABDULLAH	1	3.50	1
17	ABDULLAH	1	3.50	1
18	ABDULLAH	1	3.50	1
19	ABDULLAH	1	3.50	1
20	ABDULLAH	1	3.50	1
21	ABDULLAH	1	3.50	1
22	ABDULLAH	1	3.50	1
23	ABDULLAH	1	3.50	1
24	ABDULLAH	1	3.50	1
25	ABDULLAH	1	3.50	1

b. Database Penerimaan Mahasiswa Baru tahun 2010

Tabel yang diambil dari database penerimaan mahasiswa baru meliputi tabel msclnmhs yang berisi data calon mahasiswa, pekerjaan orang tua, penghasilan dan registrasi mahasiswa.

Tabel 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

c. Database Akademik

Tabel yang diambil dari database akademik meliputi tabel tabel msmhs,tabel mspst, tabel nilai, tabel detailkrs, tabel trkrs dan tabel trakad. Penjelasan tabel dpt diuraikan sebagai berikut :

Tabel 3 Tabel Akademik

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
Tabel tbttrakad	Idmengajar	Digunakan untuk menyimpan transkrip akademik 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_tanggung, dan penghasilan.

Hasil data dilengkapi dengan menambahkan kolom baru status beasiswa untuk mendapatkan hasil layak atau tidak layak sehingga menghasilkan data sebagai berikut :

Tabel 4 Data Mentah Penerimaan Beasiswa

A	B	C	D	E	F	G	H	I	
1	NO	NIM	NAMA MAHASISWA	PROGRAM STUDI	IPK	PEKERJAAN	JML TANGGUNGAN	PENGHASILAN	STATUS BEASISWA
2	1	201001001	AAH CARSAH	Akuntansi	2	Wirawasta	2	Rp. 500.000,- s.d. Rp 2.000.000,-	Tidak Layak
3	3	201001006	ADE ROSA NURBARA PUTRI	Pendidikan Bahasa dan Sastra Indonesia	2,07	Wirawasta	1	<Rp. 500.000,-	Tidak Layak
4	4	201001004	Aldesta	Manajemen	3,33	Wirawasta	1	<Rp. 500.000,-	Tidak Layak
5	5	201001006	Amnah	Manajemen	3,32	Lainya	2	<Rp. 500.000,-	Tidak Layak
6	6	201001001	Amir Hamzah	Akuntansi	4	Wirawasta	3	<Rp. 500.000,-	Layak
7	7	201001007	Ane Nugman	Pendidikan Biologi	3,33	Wirawasta	1	<Rp. 500.000,-	Tidak Layak
8	8	201001009	ANAH NURBANAH	Pendidikan Bahasa dan Sastra Indonesia	3,19	Wirawasta	1	Rp. 500.000,- s.d. Rp 2.000.000,-	Tidak Layak
9	9	201001011	ANI NORYANI	Pendidikan Bahasa dan Sastra Indonesia	3,61	Wirawasta	1	Rp. 500.000,- s.d. Rp 2.000.000,-	Tidak Layak
10	10	201001011	Angeni Lathu Fenieta	Pendidikan Biologi	4	PNS	3	Rp. 2.000.000,- s.d. Rp 5.000.000,-	Layak
11	11	201001001	ASEP WAHIDUN	Manajemen Informatika	3	Ibu Rumah Tangga	1	<Rp. 500.000,-	Tidak Layak
12	12	201001003	Devi Sujadi	Manajemen Informatika	1	Petani	1	<Rp. 500.000,-	Tidak Layak
13	13	201001002	Dian Febriansyah	Teknik Informatika	4	Petani/Nelayan	7	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
14	14	201001002	Dhidi Walidhi	Sistem Informasi	3,49	Lainya	1	<Rp. 500.000,-	Tidak Layak
15	15	201001002	DWI SETIAWAN	Sistem Informasi	2,73	Lainya	1	<Rp. 500.000,-	Tidak Layak
16	16	201001003	EGI GUNAWAN	Sistem Informasi	3,48	Lainya	1	<Rp. 500.000,-	Tidak Layak
17	17	201001004	Egi Puspa Dewi	Manajemen Informatika	4	PNS	2	Rp. 2.000.000,- s.d. Rp 5.000.000,-	Layak
18	18	201001003	ENDANG Hidayat	Sistem Informasi	1,79	Pensunan / Purnawirawan	1	Rp. 500.000,- s.d. Rp 2.000.000,-	Tidak Layak
19	19	201001004	EMMA RAHMAMATI	Sistem Informasi	3,15	Wirawasta	1	<Rp. 500.000,-	Tidak Layak
20	20	201001006	Faisal	Pendidikan Bahasa Inggris	3	Pegawai Swasta	4	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
21	21	201001003	Gara Akda Hestu	Pendidikan Ekonomi	3	Lainya	3	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
22	22	201001006	HANIFATUL ULA	Manajemen Informatika	3,31	Wirawasta	1	<Rp. 500.000,-	Tidak Layak
23	23	201001048	Im Inayah	Pendidikan Biologi	3	Lainya	2	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
24	24	201001001	Ika Kartika	Manajemen	3	Wirawasta	3	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
25	25	201004108	Indar Sumiarti	Pendidikan Bahasa Inggris	3	Wirawasta	2	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak
26	26	201001006	Imeliesari	Manajemen	4	Lainya	4	Rp. 500.000,- s.d. Rp 2.000.000,-	Layak

3.2 Pemrosesan Data Awal

Proses selanjutnya adalah melakukan proses penambangan data pada data beasiswa. Berdasarkan hasil pengolahan data diatas, data mentah kemudian di ekstrak ke dalam database untuk proses data mining. Dari data diatas diperoleh Informasi Sebagai berikut :

Table 5 Tabel Informasi Data Mentah

No	Aturan	Keterangan
1	No	Nomor urut
2	Nim	Nomor induk mahasiswa
3	Nama_mahasiswa	Nama mahasiswa
4	Program_studi	Program studi
5	Ipk	Indeks prestasi kumulatif
6	Pekerjaan	Pekerjaan orang tua
7	Jml_tanggung	Jumlah tanggungan
8	Penghasilan	Penghasilan orang tua
9	Status_beasiswa	Status beasiswa

3.3 Melakukan Pre-processing dan tranformation

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 penambangan data (data mining) maka data tersebut dikelompokkan berdasarkan jangkauan tertentu. 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 6 Aturan penilaian IPK

No	Aturan	Klasifikasi
1	IPK <= 2.75	1
2	IPK < 2.76 - 3.50	2
3	IPK >=3.51	3

Untuk aturan penilaian nilai Pekerjaan adalah sebagai berikut :

Tabel 7 Aturan nilai pekerjaan

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

Untuk aturan penilaian nilai Penghasilan adalah sebagai berikut :

Tabel 8 Aturan nilai penghasilan

No	Aturan	Klasifikasi
1	< Rp. 500.000,-	1
2	Rp. 500.000,- s.d. Rp 2.000.000,-	2
3	> Rp. 2.000.000,- s.d. Rp. 5.000.000,-	3

Data diatas akan dijadikan sebagai data dasar untuk proses perhitungan menggunakan algoritma K-Nearest Neighbor dan naive bayes classifier.

3.4 Perhitungan Nilai

3.4.1 Perhitungan data menggunakan Naive Bayes

Proses berikutnya adalah menghitung nilai maksimal posterior class untuk menentukan status beasiswa layak atau tidak layak Untuk menghitung nilai posterior maksimal dilakukan dengan cara melakukan uji coba terhadap data testing. Sebagai contoh Jika ada mahasiswa yang mengajukan beasiswa dengan data sebagai berikut maka apakah mahasiswa tersebut layak atau tidak layak mendapatkan beasiswa?

Tabel 9 Preprocessing Data Testing Mahasiswa

IPK	Pekerjaan	Jml tanggungan	Penghasilan	Status Beasiswa
3	1	4	3	?

Proses Perhitungannya adalah sebagai berikut :

1. Menghitung jumlah class atau label

Tahap 1 menghitung jumlah class/label
 $P(Y = \text{Layak}) = 34/984 = \text{'jumlah data "layak" pada kolom status beasiswa dibagi jumlah data}$
 $= 0,03$
 $P(Y = \text{Tidak_layak}) = 950/984 = \text{'jumlah data "Tidak layak" pada kolom status beasiswa' dibagi jumlah data}$
 $= 0,96$

- Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama
 $P(\text{IPK} = 3 \mid Y = \text{Layak}) = 19/34 = 0,55$
 'jumlah data ipk= "3" dengan keterangan "layak dibagi jumlah data layak'
 $P(\text{IPK} = 3 \mid Y = \text{Tidak Layak}) = 42/950 = 0,04$
 'jumlah data ipk= "3" dengan keterangan "tidak layak dibagi jumlah data tidak layak'
 $P(\text{jml_tanggung} = 4 \mid Y = \text{Layak}) = 10/34 = 0,29$
 'jumlah data tanggungan= '4' dengan keterangan "layak" dibagi jumlah data layak
 $P(\text{jml_tanggung} = 4 \mid Y = \text{Tidak Layak}) = 0/950 = 0$
 'jumlah data tanggungan='4' dengan keterangan "tidak layak" dibagi jumlah data tidak layak
 $P(\text{pekerjaan} = 1 \mid Y = \text{Layak}) = 7/34 = 0,20$
 'jumlah data pekerjaan= '1' dengan keterangan "layak" dibagi jumlah data layak
 $P(\text{pekerjaan} = 1 \mid Y = \text{Tidak Layak}) = 240/950 = 0,25$
 'jumlah data pekerjaan='1' dengan keterangan "tidak layak" dibagi jumlah data tidak layak
 $P(\text{penghasilan} = 3 \mid Y = \text{Layak}) = 6/34 = 0,17$
 'jumlah data penghasilan= '3' dengan keterangan "layak" dibagi jumlah data layak
 $P(\text{Penghasilan} = 3 \mid Y = \text{tidak Layak}) = 19/950 = 0,02$
 'jumlah data penghasilan= '3' dengan keterangan "tidak layak" dibagi jumlah data tidak layak

- Kalikan semua hasil variabel layak dan tidak layak
 $P(\text{IPK} = 3), (P(\text{pekerjaan} = 1), (\text{jml_tanggung} = 4), (\text{penghasilan} = 3))$
 $= \{P(P(\text{IPK} = 3 \mid Y = \text{layak}), (P(\text{pekerjaan} = 1 \mid Y = \text{layak}), (\text{jml_tanggung} = 4 \mid Y = \text{layak}), (\text{penghasilan} = 3 \mid Y = \text{layak}))\}$
 $= 0,55 \times 0,20 \times 0,17 \times 0,29 \times 0,03 = 0,000561$
 $= \{P(P(\text{IPK} = 3 \mid Y = \text{tidak layak}), (P(\text{pekerjaan} = 1 \mid Y = \text{tidak layak}), (\text{jml_pekerjaan} = 4 \mid Y = \text{tidak layak}), (\text{penghasilan} = 3 \mid Y = \text{tidak layak}))\}$
 $= 0,04 \times 0,25 \times 0,02 \times 0 \times 0,96 = 0$

- Bandingkan hasil class layak dan tidak layak
 $P(\text{Layak} = 0,000206334116058441)$
 $P(\text{Tidak Layak} = 0)$
 Karena hasil (P| layak) lebih besar dari (P| Tidak layak) maka keputusannya adalah " Layak"

Tabel 10 Hasil Rekomendasi

IPK	Pekerjaan	Jml tanggungan	Penghasilan	Status Beasiswa
3,86	PNS/Pegawai Negara	4	RP > 2000000-5000000	Layak

3.4.2 Perhitungan data menggunakan K-Nearest Neighbor

Metode yang digunakan untuk menentukan klasifikasi data testing adalah dengan menghitung jarak euclidean distance masing-masing objek terhadap data training. Contoh Langkah awal adalah dengan menentukan nilai parameter K= 7.

Knn tahap 1

Menghitung jarak menggunakan rumus euclidean distance :

$$D1(x,y) = \sqrt{(3-3)^2 + (1-3)^2 + (4-3)^2 + (3-1)^2} = 9$$

$$D2(x,y) = \sqrt{(3-1)^2 + (1-3)^2 + (4-4)^2 + (3-3)^2} = 1$$

$$D3(x,y) = \sqrt{(3-3)^2 + (1-3)^2 + (4-1)^2 + (3-1)^2} = 9$$

$$D4(x,y) = \sqrt{(3-3)^2 + (1-5)^2 + (4-4)^2 + (3-2)^2} = 17$$

$$D5(x,y) = \sqrt{(3-3)^2 + (1-5)^2 + (4-7)^2 + (3-2)^2} = 26$$

$$D6(x,y) = \sqrt{(3-2)^2 + (1-1)^2 + (4-3)^2 + (3-3)^2} = 2$$

$$D7(x,y) = \sqrt{(3-3)^2 + (1-3)^2 + (4-2)^2 + (3-2)^2} = 9$$

$$D8(x,y) = \sqrt{(3-3)^2 + (1-1)^2 + (4-2)^2 + (3-1)^2} = 4$$

$$D9(x,y) = \sqrt{(3-2)^2 + (1-2)^2 + (4-4)^2 + (3-2)^2} = 3$$

$$D10(x,y) = \sqrt{(3-2)^2 + (1-2)^2 + (4-4)^2 + (3-2)^2} = 3$$

$$D11(x,y) = \sqrt{(3-3)^2 + (1-3)^2 + (4-2)^2 + (3-2)^2} = 9$$

$$D12(x,y) = \sqrt{(3-2)^2 + (1-5)^2 + (4-3)^2 + (3-2)^2} = 19$$

$$D13(x,y) = \sqrt{(3-2)^2 + (1-5)^2 + (4-2)^2 + (3-2)^2} = 22$$

$$D14(x,y) = \sqrt{(3-2)^2 + (1-3)^2 + (4-3)^2 + (3-2)^2} = 7$$

$$D15(x,y) = \sqrt{(3-2)^2 + (1-3)^2 + (4-2)^2 + (3-2)^2} = 10$$

$$D16(x,y) = \sqrt{(3-3)^2 + (1-5)^2 + (4-4)^2 + (3-2)^2} = 17$$

$$D17(x,y) = \sqrt{(3-3)^2 + (1-5)^2 + (4-3)^2 + (3-2)^2} = 18$$

$$D18(x,y) = \sqrt{(3-3)^2 + (1-5)^2 + (4-4)^2 + (3-2)^2} = 2$$

$$D19(x,y) = \sqrt{(3-3)^2 + (1-3)^2 + (4-3)^2 + (3-2)^2} = 6$$

$$D20(x,y) = \sqrt{(3-3)^2 + (1-1)^2 + (4-5)^2 + (3-3)^2} = 1$$

$$\text{Dst..}$$

$$D984(x,y) = \sqrt{(3-2)^2 + (1-1)^2 + (4-1)^2 + (3-2)^2} = 11$$

KNN tahap 2

Tahap selanjutnya adalah memilih data terdekat. Dari perhitungan data diatas akan ditentukan data terdekat dengan cara mengurutkan (ascending) data hasil pengukuran jarak diatas. Berdasarkan Hasil yang diperoleh dengan memilih K-7 adalah sebagai berikut :

Tabel 11. Hasil Scoring KNN tahap 2

A	B	C	D	E	F	G	H	I	J
No	NIM	NAMA MAHASISWA	PROGRAM STUDI	IPK	PEKERJAAN	JML TANGGUNGAN	PENGHASILAN	STATUS BEASISWA	KNN_2
1	2020031011	Anjani Lathu Fenteha	Pendidikan Biologi	3	1	3	3	Layak	1
20	2020031131	Lefti Rahmawati	Pendidikan Ekonomi	3	1	5	3	Layak	1
4	2020041058	Oliva Nurul Fadhilah	Pendidikan Bahasa Inggris	2	1	3	3	Layak	2
5	2020031053	Linda	Sistem Informasi	3	2	4	2	Layak	2
6	2020031043	Erna Yusuf	Manajemen	2	2	4	2	Layak	3
20	2020031016	Faisal	Pendidikan Bahasa Inggris	2	2	4	2	Layak	3
8	2020113004	Egi Puspa Dewi	Manajemen Informatika	3	1	2	3	Layak	4
3	2020031118	Sri Mithya Rini	Manajemen	3	1	2	3	Layak	4
34	2020031118	Leka Viletan	Pendidikan Bahasa Dan Sastra Indonesia	2	2	5	2	Layak	4
11	2020081068	Nia Irma Dewi	Akuntansi	3	3	4	2	Layak	5
12	2020041175	Nurendah	Pendidikan Bahasa Inggris	3	3	4	2	Layak	5
13	2020081067	Sinta Juni Nur Hithah M.	Akuntansi	3	1	2	2	Layak	5
14	2020031064	Lisiani Herfah	Pendidikan Biologi	3	3	3	2	Layak	6
15	2020041170	Nunung Nurhasanah	Pendidikan Bahasa Inggris	3	3	3	2	Layak	6
16	2020011072	Nurhalimah	Pendidikan Bahasa Dan Sastra Indonesia	2	3	4	2	Layak	6
17	202011186	Nurlieli	Pendidikan Ekonomi	2	3	4	2	Layak	6
18	2020031069	Nia Kartika	Manajemen	2	3	3	2	Layak	7
19	2020031115	Purni Wulansari	Pendidikan Biologi	2	3	3	2	Layak	7
40	2020081093	ABDUL AZIS	Teknik Informatika	1	1	2	2	Tidak Layak	7
21	2020031093	ABUS HRAN	Pendidikan Bahasa Inggris	1	1	2	2	Tidak Layak	7
127	2020031093	ARIF HADIRAN	Teknik Informatika	1	1	2	2	Tidak Layak	7
127	2020031093	Asep Saiful Hilmah	Teknik Informatika	1	1	2	2	Tidak Layak	7
241	2020031112	Budi Heryono	Akuntansi	1	1	2	2	Tidak Layak	7
251	2020031094	Denang Dewaningsih	Akuntansi	1	1	2	2	Tidak Layak	7
251	2020031094	Denang Dewaningsih	Akuntansi	1	1	2	2	Tidak Layak	7

Dari hasil tersebut diatas dapat diketahui bahwa 7 jarak terdekat adalah 1, 1, 2,2,3,3,4 Dan semua data memiliki status layak sehingga dapat diambil kesimpulan bahwa dengan K-7 Hasilnya adalah layak

3.5 Pengukuran Akurasi Klasifikasi

Untuk menghitung tingkat akurasi dengan confuse matrik. peneliti menggunakan bantuan tools weka. berikut ini adalah hasil yang diperoleh menggunakan weka :

Dari data di atas diperoleh
TP = 33, FN = 1, FP = 0 , TN=950
Akurasi dari data diatas dapat dihitung $\frac{TP+TN}{P+N}$

$$\begin{aligned} P+N &= 33+1+0+950 \\ \text{Accurasi} &= \frac{33+950}{33+1+0+950} \\ &= 0.998984 \\ &= 0.998984 \times 100 \% \\ &= 99,8984 \% \end{aligned}$$

Jadi akurasi data nya adalah 99,89 %
a. Confuse Matrik K Nearest neighbor

Dari data di atas diperoleh
TP = 34, FN = 0, FP = 0, TN=980
Akurasi dari data diatas dapat dihitung $\frac{TP+TN}{P+N}$

$$\begin{aligned} P+N &= 34+0+0+980 \\ \text{Accurasi} &= \frac{34+980}{34+0+0+980} \\ &= 1 \\ &= 1 \times 100 \% \\ &= 100 \end{aligned}$$

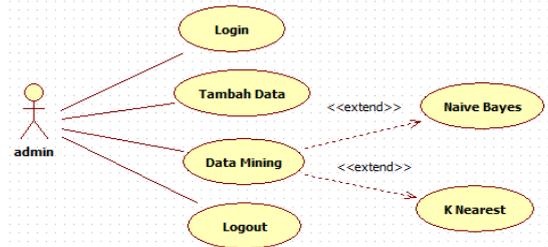
Jadi akurasi data nya adalah 100 %

3.6 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 :

3.6.1. 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 prediksi prestasi mahasiswa.



Gambar 3 Use case data rekomendasi penerima beasiswa

Dibawah ini merupakan penggambaran umum mengenai use case yang terdapat dalam sistem informasi rekomendasi penerimaan beasiswa

Nomor	Nama use case	Deskripsi	Actor
1	Login	Use ini menggambarkan proses login sebelum masuk ke sistem. Dimana pengguna harus memasukan username dan passwords untuk dapat masuk ke dalam sistem	Admin
2	Tambah data	Use case ini menggambarkan proses input data untuk proses data training ataupun data pengujian	Admin
5	Data Mining	Use case ini menggambarkan proses penambangan data menggunakan metode Naive bayes atau K Nearest Neighbor terhadap data beasiswa mahasiswa dan menghasilkan perhitungan hasil layak dan tidak layak	Admin
4	Logout	Use case ini menggambarkan proses dimana pengguna ingin keluar dari sistem	Admin

3.6.2 Narasi use case

Berikut ini adalah narasi use case yang merupakan deskripsi tekstual dari kejadian bisnis dan bagaimana pengguna berinteraksi dengan sistem

a. Narasi Tambah Data

Nama use case	TambahData
Kebutuhan terkait	Use case ini menggambarkan proses input data pelatihan dan pengujian
Tujuan	Use case ini digunakan untuk melakukan input data training dan data testing berdasarkan aturan yang dibuat.
Prasyarat	aturan sudah dibentuk
Kondisi Akhir sukses	Jika semua data sudah dipilih
Kondisi Akhir Gagal	Jika ada data yang masih kosong atau belum dipilih
Aktor utama	Admin

Pemicu	Actor mengklik tombol save
---------------	----------------------------

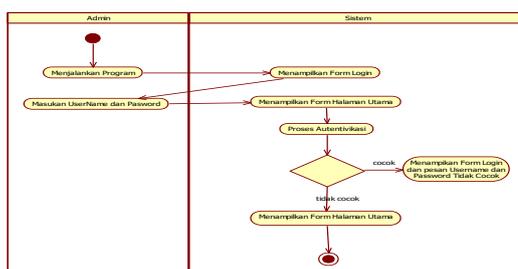
b. Narasi Data Mining

Nama use case	DataMining
Kebutuhan terkait	Usecase ini menggambarkan proses perhitungan data menggunakan algoritma naive bayes atau k-nearest neighbor
Tujuan	Use case ini digunakan untuk menghasilkan nilai layak atau tidak layak
Prasyarat	Jika semua data sudah diisi
Kondisi Akhir sukses	Menghasilkan hasil perhitungan layak dan tidak layak
Kondisi Akhir Gagal	Tidak menghasilkan keputusan karena data kosong
Aktor utama	Admin
Pemicu	Actor memilih memilih menu naive bayes atau k-nearest dan memilih tombol proses prediksi

3.6.3 Diagram Aktivitas

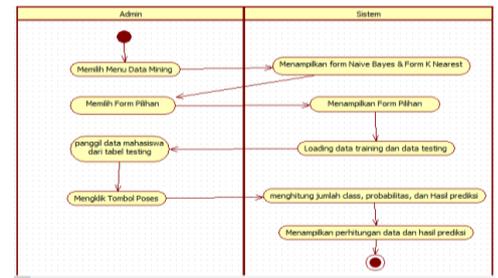
Diagram aktivitas merupakan diagram yang menjelaskan aktivitas pengguna dengan program. Berikut ini adalah diagram aktivitas pada pengolahan sistem informasi rekomendasi penerimaan beasiswa

a. Diagram aktivitas login



Gambar 4 Diagram Aktivitas login

b. Diagram aktivitas Data mining

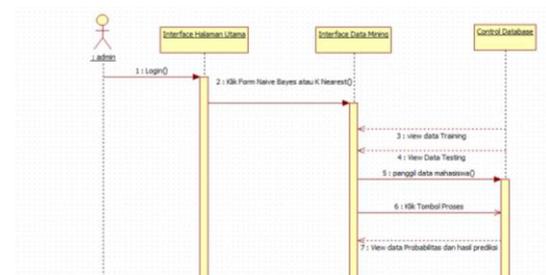


Gambar 5 Diagram Aktivitas Data Mining

3.6.4 Diagram Sekuensial

Diagram sekuensial adalah diagram interaksi yang disusun berdasarkan urutan waktu. Berikut ini adalah diagram sekuensial dari sistem yang akan dibangun :

Diagram sekuensial Data mining



Gambar 6 Diagram Sekuensial Data Mining

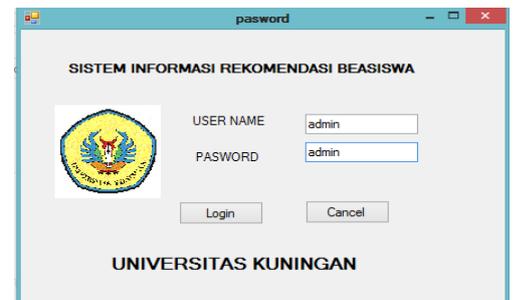
3.7 Perancangan Antar Muka

Tahapan selanjutnya adalah membuat perancangan antar muka berupa desain menu utama dan desain form-form masukan.

3.8 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 7 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

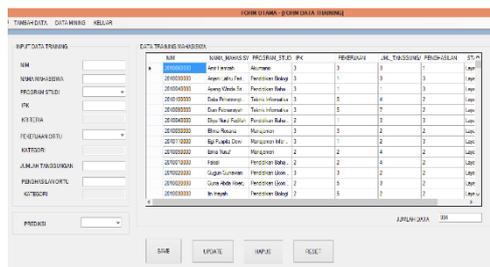
b. Form Menu Utama



Gambar 8 Form Menu Utama

Untuk mengendalikan jalannya 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.

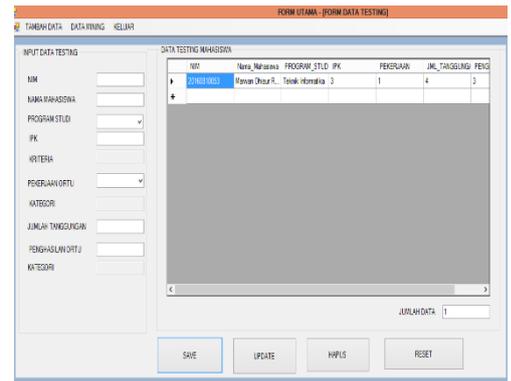
c. Form Data Training



Gambar 9 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.

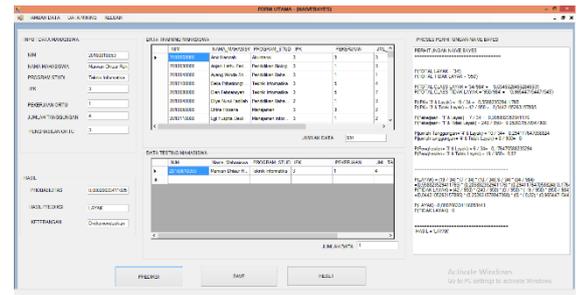
d. Form Data Testing



Gambar 10 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.

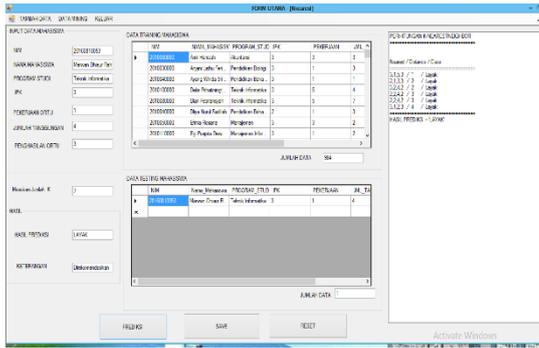
e. Form Naive bayes



Gambar 11 Form Naive Bayes

Form ini digunakan untuk melakukan perhitungan Menggunakan Naive Bayes. 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 K-Nearest Neighbor



Gambar 12 Form K-Nearest Neighbor

Form ini digunakan untuk melakukan perhitungan Menggunakan K-Nearest Neighbor Dimana form ini akan menampilkan data training dan data testing. Data testing yang akan di proses dipanggil dan ditampilkan pada kolom input data mahasiswa, User memasukan jumlah K. Kemudian data tersebut dilakukan proses perhitungan sesuai aturan algoritma K-Nearest Neighbor 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 Hasil Data yang sudah dihitung Sejumlah K yang dimasukan akan ditampilkan pada kolom proses perhitungan K-Nearest Neighbor

4. KESIMPULAN

1. Metodologi yang digunakan penulis dalam melakukan perhitungan dan pengukuran akurasi terhadap model data mining cukup mudah untuk diimplementasikan.
2. Hasil pengukuran tingkat akurasi data algoritma K-Nearest Neighbor Diperoleh akurasi data sebesar 100% dan Algoritma Naive Bayes diperoleh akurasi sebesar 99,89 % . Hal ini menunjukan bahwa pada data Penerimaan Beasiswa ini tingkat akurasi K-nearest Neighbor lebih besar dibandingkan dengan Naive Bayes sehingga untuk rekomendasi penentuan mahasiswa penerima beasiswa lebih cocok menggunakan algoritma K-Nearest Neighbor.
3. Jumlah data training dapat mempengaruhi tingkat akurasi data.

5. SARAN

1. 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.
2. Pada pembuatan program selanjutnya agar diperbaiki untuk penghitungan data mining K-nearest neighbor dan naive bayes dilakukan dalam satu form.
3. Untuk mengetahui hasil akurasi perhitungan naive bayes dan k-nearest neighbor, 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

[Larose 2005] Larose, Daniel T. “Discovering Knowledge ind Data :An Introduction to Data Mining. John Willey & Sons.inc, 2005

[Riyan 2014] Riyan Eko putri, suparti, Rita Rahmawati, “Perbandingan Metode Klasifikasi Naive Bayes dan K-nearest Neighbor Pada Analisis data status Kerja di Kabupaten Demak tahun 2012“ , Jurnal Gaussian Vol 3, nomor 4, Tahun 2014 ISSN:2339-2541.

[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 Juju, Dominikus, “Data Mining, Meramalkan Bisnis Perusahaan” , Elex Media Komputindo, 2010

[Suyanto 2017] Dr Suyanto, S.T., M.Sc., “Data Mining untuk Klasisfikasi dan Klasterisasi Data”, Informatika, 2017

[Turban 2005] Turban, E.,dkk , “Decision Support System And Intelegent System”, Andi , Yogyakarta,2005