

Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes

Mongan Winny Amelia, Arie S. M. Lumenta, Agustinus Jacobus
Teknik Informatika Universitas Sam Ratulangi

120216039@student.unsrat.ac.id, al@unsrat.ac.id, a.jacobus@unsrat.ac.id

Abstrak – *Data mining* adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang dapat dimanfaatkan untuk pengambilan keputusan penting. Data kelulusan mahasiswa Universitas Sam Ratulangi menghasilkan data akademik mahasiswa yang berpengaruh pada kualitas perguruan tinggi lewat akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional atau BAN-PT. *Naïve Bayes* merupakan algoritma yang dapat digunakan untuk prediksi masa studi mahasiswa yang bisa dikategorikan tepat waktu, terlambat dan tidak lulus. Dengan menggunakan data mining khususnya klasifikasi untuk prediksi dengan algoritma naïve bayes dapat dilakukan prediksi terhadap ketepatan waktu studi dari mahasiswa berdasarkan data *training* yang ada.. Pengujian yang dipakai yaitu *k-fold cross validation 10-fold*. Hasil pengujian didapat nilai akurasi rata-rata sebesar 85.17 % sedangkan nilai akurasi tertinggi sebesar 88.96 %.

Kata kunci : *Naïve Bayes*, *K-Fold Cross Validation*, *Data Mining*, *Prediksi*.

I. PENDAHULUAN

Banyak data yang dapat dijadikan informasi untuk memustuskan tindakan atau aktivitas masyarakat sehari – hari. Informasi dari suatu data saja tidak cukup untuk dijadikan suatu keputusan. Dalam hal ini suatu Universitas dapat menggunakan data yang telah diolah menjadi informasi, seperti data mahasiswa, data pegawai, data dosen, data prestasi mahasiswa data jumlah kelulusan mahasiswa tiap tahun serta data lainnya. Penggunaan data mining dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan khususnya faktor dalam data induk mahasiswa. Tingkat kelulusan mahasiswa berpengaruh terhadap akreditasi Universitas.

Dalam jenjang pendidikan Program Sarjana (S1) di Fakultas Teknik memiliki beban 144 sks (satuan kredit semester) yang harus diselesaikan dalam waktu delapan semester. Masing-masing mahasiswa memiliki pengontrakan jumlah dan jenis mata kuliah yang berbeda di setiap semester yang membuat juga nilai IP semester berbeda – beda. Nilai IP semester juga sangat berpengaruh terhadap IPK mahasiswa. Tahun lulus mahasiswa dipengaruhi oleh nilai IPK serta waktu pengontrakan setiap Mata Kuliah yang diambil.

Berdasarkan hal diatas, maka perlu dibuat implementasi data mining dengan suatu algoritma untuk memprediksi lama masa studi mahasiswa, yang mempunyai kemampuan melihat pola kelulusan mahasiswa, untuk selanjutnya bisa menjadi strategi dalam proses perkuliahan. Informasi lama masa studi dari

mahasiswa Universitas Sam Ratulangi sangat penting untuk meningkatkan pelayanan yang dapat membuat mahasiswa nyaman dan bisa lulus tepat waktu. Pada penelitian yang akan dilakukan ini, menjelaskan bahwa faktor yang paling berpengaruh dalam tingkat kelulusan mahasiswa adalah Indeks Prestasi Semester (IPS) dan Jumlah Sistem Kredit Semester (SKS) secara keseluruhan maupun tiap semester.

II. LANDASAN TEORI

A. Data Mining

Data mining adalah istilah yang digunakan untuk mendapatkan penemuan pengetahuan dan atau informasi di dalam *database*. Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan data yang tersimpan dalam penyimpanan dengan menggunakan teknik statistik dan matematika [1].

Menurut [2], tujuan dari adanya data mining adalah:

1. Explanatory, yaitu untuk menjelaskan beberapa kegiatan observasi atau suatu kondisi.
2. Confirmatory, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
3. Exploratory, yaitu untuk menganalisis data baru suatu relasi yang janggal

Data mining memiliki banyak fungsi yang dapat digunakan. Fungsi data mining dapat digabungkan dalam kasus tertentu untuk menjawab masalah yang ada [3]. Berikut adalah fungsi data mining secara umum :

1. Classification

Fungsi dari Classification adalah untuk mengklasifikasi suatu target class ke dalam kategori yang dipilih.

2. Clustering

Fungsi dari clustering adalah untuk mencari pengelompokan atribut ke dalam segmentasi-segmentasi berdasarkan similaritas.

3. Association

Fungsi dari association adalah untuk mencari keterkaitan antara atribut atau item set, berdasarkan jumlah item yang muncul dan rule association yang ada.

4. Regression

Fungsi dari regression hampir mirip dengan klasifikasi. Fungsi dari regression adalah bertujuan untuk mencari prediksi dari suatu pola yang ada.

5. Forecasting

Fungsi dari forecasting adalah untuk peramalan waktu yang akan datang berdasarkan trend yang telah terjadi di waktu sebelumnya.

6. Sequence Analysis

Fungsi dari sequence analysis adalah untuk mencari pola urutan dari rangkaian kejadian.

7. *Deviation Analysis*

Fungsi dari devation analysis adalah untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

B. Klasifikasi

Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu : pertama, pembangunan model sebagai prototype untuk disimpan sebagai memori dan kedua, penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang mudah disimpan [4].

Proses untuk menemukan pola yang menjelaskan data yang penting dikenal sebagai klasifikasi. Metode klasifikasi dalam data mining ada banyak, diantaranya decision tree, k-nearest neighbor, neural network dan naïve bayes.

1. *Decision Tree*

Decision Tree merupakan metode klasifikasi dalam bentuk diagram yang direpresentasikan seperti struktur pohon.

2. *K-Nearest Neighbor*

Metode K-Nearest Neighbor merupakan metode klasifikasi pertama yang dijabarkan pada awal tahun 1950.

3. *Neural Network*

Neural Network terinspirasi oleh pengenalan sistem pembelajaran yang kompleks pada otak binatang yang terdiri atas kumpulan neuron yang saling berhubungan.

4. *Naïve Bayes*

Klasifikasi naïve bayes adalah salah satu teknik data mining yang paling populer untuk mengklasifikasikan data dalam jumlah besar dan dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class.

C. *Naïve Bayes*

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes* model yang digunakan adalah “model fitur independen” [4]

$$Probabilitas(H|E,e) = p(H|E) * \frac{p(e|E,H)}{p(e,E)}$$

Dimana :

- E = evidence
- E = evidence obeservasi baru
- P(H|E,e) = probabilitas hitosis H benar jika muncul evidence baru daru E dari evidence e
- p(H|E) = probabilitas hitosis H benar jika diberikan evidence E
- p(e|E,H) = kaitan antara e dan E jika H benar
- p(e|E) = kaitan tanpa memandang hipotesis apapun

D. Tahapan Data Mining

1. *Data Selection*

Data selection adalah pemilihan data yang relevan dengan tugas analisis diambil dari *database*.

Data selection merupakan pemilihan himpunan data atau seleksi data dari sekumpulan data. Proses seleksi perlu dilakukan sebelum tahap penggalian informasi dimulai. Data hasil seleksi akan digunakan untuk proses data mining, disimpan terpisah dengan basis data operasional.

2. *Preprocessing*

Preprocessing merupakan tahapan awal dalam pengolahan data .beberapa kasus pasti memiliki kualitas data yang kurang baik. Data yang memiliki kualitas data yang kurang baik memiliki kekurangan nilai-nilai atribut atau atribut tertentu, berisi kesalahan atau nilai-nilai outlier yang menyimpang dari yang diharapkan dan ketidakcocokan dalam penggunaan kode atau nama.

Disini kualitas data yang baik berbanding lurus dengan keputusan yang baik. Untuk memiliki keputusan yang baik, digunakan cara yaitu : Menghilangkan data yang salah, memperbaiki kekacauan data dan memeriksa data yang tidak konsisten yang disebut dengan *data Cleaning*. Menggabungkan dari dari beberapa sumber kedalam penyimpanan data yang sesuai disebut dengan *Data integration*.

3. *Naïve Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

4. *Data Interpretation and Evaluation*

Tahap *Evaluation* ini merupakan tahapan pemeriksaan apakah pola yang terbentuk atau informasi yang dihasilkan memiliki keakuratan yang baik, atau jauh bertentangan dengan data yang sebelumnya. Pada tahap evaluasi ini akan digunakan metode confusion matrix sebagai metode yang akan menguji precision, recall maupun accuracy dari pola yang terbentuk. Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

III. Metodologi Penelitian

A. Proses Data Mining

Metodologi penelitian yang dilakukan pada pembuatan tugas akhir adalah melakukan tahapan proses data mining, yaitu *data selection*, *preprocessing*, data mining dan *interpretation and evaluation*.

1. *Data Selection*

Pada proses seleksi ini, dipilih dari keseluruhan data Universitas Sam Ratulangi yaitu data alumni Fakultas Teknik angkatan 2008-2012. Dari hasil seleksi akan digunakan sebanyak 1625 data alumni untuk perhitungan pada proses data mining.

2. *Preprocessing*

Preprocessing menggambarkan kualitas data yang baik berbanding lurus dengan keputusan yang baik. Untuk memiliki keputusan yang baik, dilakukan proses Integrasi skema sistem, *data cleaning*, *data transformation* dan *data reduction*. Proses tersebut dapat dijelaskan seperti dibawah ini :

a. Integrasi Skema Aplikasi Sistem

Proses ini dilakukan karena perlu adanya perubahan skema struktur data dari data mentah ke data yang siap akan dilakukan proses preprocessing lebih lanjut dengan tahap *cleaning*.

Tabel 1. Data Sebelum Proses Integrasi

Identifier	Prodi	Semester	IP Semester	IP K	SKS	Lulus	Jumlah Semester
"ff69.....fad000"	"INFORMATIKA"	2011 1	3.05	3.05	19	1	11
"ff69.....fad000"	"INFORMATIKA"	2011 2	3.33	3.21	24	1	11
"ff69.....fad000"	"INFORMATIKA"	2012 1	3.05	3.16	19	1	11
"ff69.....fad000"	"INFORMATIKA"	2012 2	3.22	3.18	23	1	11

Data akan berubah skema atau struktur menjadi seperti pada tabel 2 yang mudah untuk diproses. Data tabel didefinisikan memanjang ke samping kanan dengan penggunaan kolom yang lebih banyak, dikarenakan hasil *pivoting* dari tiap data identifier.

Tabel 2. Data Sesudah Proses Integrasi

Identifier	Prodi	Lulus	Angkatan	IP Semester				SKS			
				2011 1	2011 2	2012 1	2012 2	2011 1	2011 2	2012 1	2012 2
"ff6"	IT	1	2011	3.05	3.33	3.05	3.22	19	24	19	23

b. *Data Cleaning*

Pada proses ini data yang tidak perlu dipakai pada penggunaan pengolahan data dan data yang akan menyebabkan kekacauan data sehingga data menyimpang atau tidak konsisten.

Tabel 3. Data Sebelum Proses Cleaning

Identifier	Prodi	Semester	IPS	IPK	SKS	Lulus	Jumlah
11021 5	"TEKNIK ELEKTRO"	2011 1	3.2	3.2	19	0	21
11021 5	"TEKNIK ELEKTRO"	2011 2	3.1	3.1	20	0	21
11021 5	"TEKNIK ELEKTRO"	2011 3	0	3.1	0	0	21
11021 5	"TEKNIK ELEKTRO"	2011 4	0	3.1	0	0	21

Data yang akan dihapus adalah data Index Prestasi (ip) semester mahasiswa yang memiliki atau mengikuti semester perbaikan. Angka terakhir pada kolom semester menjelaskan posisi semester mahasiswa. Angka paling akhir yaitu 1

atau 2, menjelaskan ganji atau genap data semester mahasiswa. Sedangkan angka 3 atau empat adalah baris data ip semester mahasiswa yang mengikuti semester perbaikan.

c. *Data Transformation*

Proses *Transformation* melakukan pencarian fitur yang dapat mempresentasikan data dan pencariannya sangat bergantung pada jenis atau pola informasi yang akan dicari dalam database.

Tabel 4. Transformasi Status Kelulusan

Jumlah Semester	Status Kelulusan
<= 10	Tepat
> 10	Terlambat
Tidak lulus	Tidak Lulus

Transformasi data yang terjadi pada data jumlah semester untuk menentukan status kelulusan dengan cara, Kategori status kelulusan yang akan dibagi ditentukan berdasarkan data mahasiswa yang lulus dan tidak lulus.

d. *Naïve Bayes*

Simulasi perhitungan ini diambil 20 sampel dari data training untuk dilakukan proses mining dengan menggunakan Microsoft excel. Untuk perhitungan secara lengkap akan menggunakan bahasa pemrograman PHP.

Tabel 5. 20 Sampel Data yang akan dikelola dengan perhitungan manual

I d	An gkt	Prodi	Lulus	ips _1	ips _2	sk s_ 4	sk s_ 5	sk s_ 6
1 2	200 9	TEKNIK MESIN	Cepat	0	2.1 3	30	0	\N
1 3	200 9	ARSITEKTUR	Cepat	4	0	8	\N	\N
1 4	200 9	TEKNIK SIPIL	Cepat	0	2.9 7	0	\N	\N
...								
1 5	201 0	TEKNIK SIPIL	Cepat	4	2.7 9	10	34	24
2 9	201 0	INFORMATIKA	Tidak Lulus	2.5 6	2.8 5	18	17	17
3 0	201 0	ARSITEKTUR	Tidak Lulus	1.8	0.5	17	16	15
3 1	201 0	TEKNIK SIPIL	Tidak Lulus	0.7 3	2.3 8	15	20	17

1. Menghitung *Standar Deviasi* dan *Mean*

Nilai Standar Deviasi dan Mean diambil dari masing – masing variable yang bernilai kontinu antara lain IPS_1, IPS_2, IPS_3, IPS_4, IPS_5, IPS_6, SKS_1, SKS_2, SKS_3, SKS_4, SKS_5, dan SKS_6 pada setiap kategori. Berikut proses perhitungan yang menunjukkan nilai *Mean* dan *Standar Deviasi*.

Tabel 7.Perhitungan Mean dan Standar Deviasi pada IP dan SKS dalam semester 1 s/d 6

kelas	Mean/stde	ips_1	ips_2	.	ipk_4	ipk_5	ipk_6
cepat	Mean	2.186	2.122	.	15.6	6.8	4.8
	Standar Deviasi	2.042787	1.227383	.	13.66748	15.20526	10.73313
tepat	Mean	3.39	3.032	.	15.8	23.8	22.2
	Standar Deviasi	0.601457	0.693376	.	9.038805	2.48998	4.969909
terlambat	Mean	1.51	2.198	.	15.4	19.8	21
	Standar Deviasi	1.46506	1.691544	.	8.619745	0.447214	4.301163
tidak lulus	Mean	1.246	0.908	.	14.2	15.2	16
	Standar Deviasi	1.154807	1.243411	.	8.043631	8.58487	9.192388

Dengan aturan tersebut jika diberikan data baru yang terdapat pada tabel di bawah ini maka prediksi masa studi mahasiswa dapat dikategorikan dengan menggunakan langkah-langkah sebagai berikut :

Tabel 8.Data hitung manual

NIM	Prodi	IPS1	IPS2	..	SKS3	SKS4	SKS5
12345	Sipil	2.52	2.98	..	19	9	14

2. Menghitung Probabilitas untuk setiap kelas berdasarkan fitur

Dari kasus diatas maka akan dilakukan perhitungan nilai probabilitas dari variabel yang bersifat kontinu yaitu IPS1 s/d IPS6. Berikut proses perhitungan masing-masing IPS :

a. Probabilitas IPS dan SKS untuk kelas Tepat dari IPS1 s/d IPS6

1) Probabilitas IPS1 untuk kelas Tepat

$$\begin{aligned} \text{Standar Deviasi } \sigma &= 0.601456565347825 \\ \text{Mean } \mu &= 3.39 \\ f(\text{ips1} = 2.52 | \text{Tepat}) &= \frac{1}{\sqrt{2\pi}(0.601456565347825)} e^{-\frac{(2.52-3.39)^2}{2(0.601456565347825)^2}} \\ &= 0.216730648 \times 2,7183^{-0.20084401} \\ &= 0.177294107 \end{aligned}$$

2) Probabilitas IPS2 untuk kelas Tepat

$$\begin{aligned} \text{Standar Deviasi } \sigma &= 0.693375799981511 \\ \text{Mean } \mu &= 3.032 \\ f(\text{ips2} = 2.98 | \text{Tepat}) &= \frac{1}{\sqrt{2\pi}(0.693375799981511)} e^{-\frac{(2.98-3.032)^2}{2(0.693375799981511)^2}} \\ &= 0.22917 \times 2,7183^{-0.5947} \\ &= 0.12644 \end{aligned}$$

b. Probabilitas IPS dan SKS untuk kelas Terlambat dari IPS1 s/d IPS6

1) Probabilitas IPS1 untuk kelas Terlambat

$$\begin{aligned} \text{Standar Deviasi } \sigma &= 1.46506 \\ \text{Mean } \mu &= 1.51 \\ f(\text{ips1} = 2.52 | \text{Tepat}) &= \frac{1}{\sqrt{2\pi}(1.46506)} e^{-\frac{(2.52-1.51)^2}{2(1.46506)^2}} \\ &= 0.329679933 \times 2,7183^{-0.009160196} \end{aligned}$$

= 0.32667377

3. Membandingkan hasil perkalian ketiap kelas

- Likelihood Cepat : $P(\text{Sipil}|\text{Cepat}) * P(\text{IPS1}|\text{Cepat}) * P(\text{IPS2}|\text{Cepat}) * P(\text{IPS3}|\text{Cepat}) * P(\text{IPS4}|\text{Cepat}) * P(\text{IPS5}|\text{Cepat}) * P(\text{SKS1}|\text{Cepat}) * P(\text{SKS2}|\text{Cepat}) * P(\text{SKS3}|\text{Cepat}) * P(\text{SKS4}|\text{Cepat}) * P(\text{SKS5}|\text{Cepat})$
 $= 0.4 * 0.27549 * 0.28211 * 0.22815 * 0.30901 * 0.31946 * 0.08218 * 0.06121 * 0.06275 * 0.09606 * 0.09148$
 $= 0.954276465$
- Likelihood Tepat : $= 2.81285E-05$
- Likelihood Terlambat : $= 0$
- Likelihood Tidak Lulus : $= 0.045695407$

- Probabilitas of Cepat = 82 %
- Probabilitas of Tepat = 13 %
- Probabilitas of Terlambat = 0 %
- Probabilitas of Tidak Lulus = 5 %

Dari contoh kasus di atas bisa diprediksi bahwa mahasiswa dengan Program Studi Teknik Sipil, dengan data IP Semester 1 s/d 5 memiliki Hasil Prediksi = Cepat

B. Pemodelan Proses Sistem

Pemodelan proses pada penelitian ini dilakukan dengan menggunakan Data Flow Diagram (DFD). DFD adalah diagram yang menggambarkan aliran data melalui sistem dan kerja atau pengolahan yang dilakukan oleh sistem tersebut.

1. Diagram Konteks

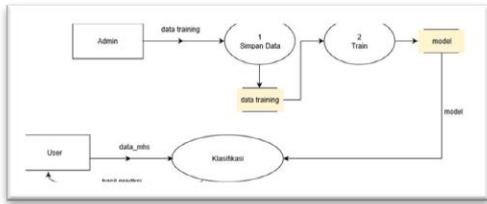
Diagram konteks adalah diagram yang menampilkan gambaran keseluruhan sistem. Gambar 1 dapat dilihat gambaran sistem yang akan dibuat memiliki dua aktor yaitu Admin (Penguji/Pengolah data) dan User (Dosen).



Gambar 1. Diagram Konteks

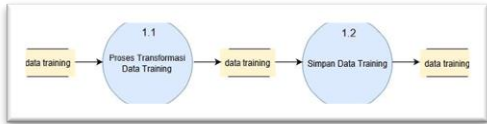
2. DFD Level 0

Diagram Nol bertujuan untuk menggambarkan proses aliran data secara terperinci yang ada pada suatu sistem. Diagram Nol merupakan tingkat lanjutan dari diagram konteks. Gambar.2 menjelaskan secara rinci bagaimana proses prediksi masa studi mahasiswa bekerja.



Gambar 2. DFD Level 0

3. DFD Level 1 *Preprocessing Data Training*
 Pada proses *preprocessing data training* terdapat proses dimana data training yang dimasukan oleh aktor admin sebelum disimpan pada *storage* harus melalui proses transformasi terlebih dahulu. Dimana data mentah dari user diubah menjadi data yang dapat dikelola oleh sistem. Diagram dijelaskan pada Gambar 3.



Gambar 3. Diagram Level 1 *Preprocessing Data Training*

IV. HASIL DAN PEMBAHASAN

A. Interface Sistem

1. Halaman Utama Sistem



Gambar 4. Tampilan Halaman Utama Sistem

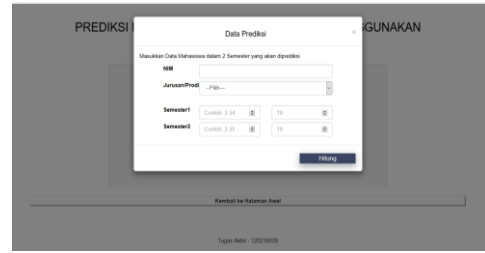
Gambar 4 adalah Tampilan pada halaman awal yang menampilkan tiga menu utama pada bagian tengah tampilan halaman, serta ada menu tambahan pada bagian kanan atas halaman web

2. Halaman Utama Admin



Gambar 5. Tampilan Halaman Utama Admin
 Halaman ini berisi beberapa menu yang digunakan untuk mengakses menu – menu seperti pada gambar 5 diatas

3. Halaman Prediksi di Semester



Gambar 6. Tampilan Halaman Prediksi di Semester

Gambar 6 ini menampilkan format isian data prediksi mahasiswa di semester dimana mahasiswa sedang menjalankan atau baru selesai perkuliahan. Isian yang ada berupa NIM, Jurusan atau Program Studi dan IP Semester serta SKS yang disesuaikan dengan semester mahasiswa.

4. Halaman Hasil Prediksi



Gambar 7. Tampilan Halaman Hasil Prediksi

Gambar 7 adalah halaman hasil prediksi yang akan langsung keluar setelah data mahasiswa dimasukan untuk dilakukan prediksi. Hasil prediksi ini berupa kesimpulan status kelulusan mahasiswa yang diuput untuk dilakukan prediksi.

B. Pengujian Model

Pengujian Model ini merupakan pengujian data terhadap algoritma. Metode pengujian yang dipakai adalah Pengujian *Cross Validation 10-Fold* yang diambil dan diuji pada bagian jumlah semester yang dilakukan prediksi.

Total Perhitungan pengujian ini yaitu dengan membagi total data umenjadi 10 bagian sama banyak. Proses pertama yang dilakukan yaitu mengambil bagian pertama menjadi data uji (*data testing*) dan dilakukan perhitungan akurasi berdasarkan data pada kelompok bagian tersebut. Sementara data yang lain (selain bagian pertama) menjadi data latih (*data training*).

1. Model untuk *Confusion Matrix* Dua Semester

Tabel 9. *Confusion Matrix* Dua Semester

		Data Alumni		
		Tepat	Terlambat	Tidak Lulus
Data Prediksi	Tepat	63	250	42
	Terlambat	15	413	133
	Tidak Lulus	4	155	550
Total Data : 1625				

$$Akurasi = \frac{65+413+550}{1625} * 100 = 63.14 \%$$

Untuk pengujian dua semester, jumlah data prediksi dengan data alumni di tampilkan dalam bentuk *Confusion Matrix*. Dalam proses pengujian ini didapatkan akurasi data sebesar 63.14 %

2. Model untuk *Confusion Matrix* Tiga Semester

Tabel 10. *Confusion Matrix* Tiga Semester

		Data Alumni		
		Tepat	Terlambat	Tidak Lulus
Data Prediksi	Tepat	55	191	16
	Terlambat	25	563	65
	Tidak Lulus	2	64	644
Total Data : 1625				

$$\text{Akurasi} = \frac{55+563+644}{1625} * 100 = 77.66 \%$$

Untuk pengujian empat semester, jumlah data prediksi dengan data alumni di tampilkan dalam bentuk *Confusion Matrix*. Dalam proses pengujian ini didapatkan akurasi data sebesar 77.66 %

3. Model untuk *Confusion Matrix* Empat Semester

Tabel 11. *Confusion Matrix* Empat Semester

		Data Alumni		
		Tepat	Terlambat	Tidak Lulus
Data Prediksi	Tepat	66	205	0
	Terlambat	15	580	34
	Tidak Lulus	1	33	691
Total Data : 1625				

$$\text{Akurasi} = \frac{66+580+691}{1625} * 100 = 82.27 \%$$

Untuk pengujian tiga semester, jumlah data prediksi dengan data alumni di tampilkan dalam bentuk *Confusion Matrix*. Dalam proses pengujian ini didapatkan akurasi data sebesar 82.27 %

4. Model untuk *Confusion Matrix* Lima Semester

Tabel 12. *Confusion Matrix* Lima Semester

		Data Alumni		
		Tepat	Terlambat	Tidak Lulus
Data Prediksi	Tepat	74	147	3
	Terlambat	8	631	43
	Tidak Lulus	0	40	679
Total Data : 1625				

$$\text{Akurasi} = \frac{74+631+679}{1625} * 100 = 85.17 \%$$

Untuk pengujian lima semester, jumlah data prediksi dengan data alumni di tampilkan dalam bentuk *Confusion Matrix*. Dalam proses pengujian ini didapatkan akurasi data sebesar 85.17 %

Tabel 13. Perbandingan Akurasi Semester

Pengujian ke -	Dua Semester	Tiga Semester	Empat Semester	Lima Semester
1	60.74 %	79.75 %	83.43 %	87.12 %
2	66.87 %	79.14 %	84.04 %	88.96 %
3	61.96 %	74.23 %	81.59 %	85.27 %
4	63.80 %	79.75 %	81.59 %	82.21 %
5	62.58 %	76.07 %	79.14 %	80.98 %
6	62.34 %	74.69 %	82.09 %	85.18 %
7	58.02 %	75.92 %	79.62 %	82.72 %
8	67.90 %	74.69 %	77.77 %	84.58 %
9	63.58 %	83.33 %	85.80 %	87.65 %
10	63.58 %	79.01 %	87.65 %	87.04 %
	63.14 %	77.66 %	82.27 %	85.17 %

Tabel perbandingan akurasi di tiap semester merupakan data yang membandingkan antara data dua semester hingga lima semester. Pengujian ini dilakukan berulang kali berdasarkan 10 kelompok pembagian yang digunakan dalam proses pengujian. Dari tabel hasil perbandingan, perhitungan akurasi dengan nilai rata-rata di tiap kelompok pada lima semester yaitu 85.17 % memiliki nilai tertinggi.

V. PENUTUP

A. Kesimpulan

Masa studi mahasiswa dapat diprediksi berdasarkan faktor yang berkaitan dengan akademik mahasiswa, seperti program studi, nilai ip semester dan jumlah sks saat di perguruan tinggi. Algoritma *Naive Bayes* yang dipakai dapat menentukan prediksi masa studi mahasiswa dengan tingkat Akurasi pada pengujian algoritma bernilai 85.17 % pada nilai rata-rata pengujian di lima semester.

B. Saran

Pada prediksi masa studi mahasiswa ini dapat ditambahkan fitur – fitur yang bernilai dan faktor diluar akademik mahasiswa (*eksternal*) untuk dapat menjadi salah satu faktor penentu lama masa studi.

Dapat dilakukan uji coba menggunakan algoritma lain dan hasilnya dapat dilakukan perbandingan dan analisa untuk pengembangan penelitian kedepan. Pada Penelitian ini dapat memperbaiki kualitas data lebih baik agar mendapatkan hasil klasifikasi yang baik. Dalam proses data mining seharusnya dilakukan proses *cleaning data*, integrasi data, transformasi data, aplikasi data mining dan evaluasi pola yang ditemukan dalam data mining agar menghasilkan nilai keakuratan yang baik.

DAFTAR REFERENSI

[1] Larose, D. (2005). *Discovering Knowledge In Data*. Canada: Willey Interscience. metode-algoritma.com. (2013, Maret 27).

[2] Hoffer, Ramesh and Topi. (2012). *Modern Database Management*. Harlow United Kingdom : Pearson Education Limited.

[3] MacLennan, J., Zhao Hui Tang, Bog, Crivat, "Data Mining with Microsoft® SQL Server® 2008", Wiley Publishing, Inc., Indianapolis, Indiana. 2009.

[4] Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi

SEKILAS TENTANG PENULIS



Saya bernama Mongan Winny Amelia dan merupakan anak kedua dari pasangan Satyano W. Mongan dan Jenny N. Sege, lahir di Manado pada tanggal 15 Juli 1994. Asal daerah Manado.

Saya mulai menempuh pendidikan di sekolah dasar SDN Malalayang (2000 - 2006). Kemudian melanjutkan studi tingkat pertama di SMPN 8 Manado (2006 - 2009) dan selanjutnya saya menempuh pendidikan tingkat atas di SMAN 1 Manado (2009- 2012).

Setelah itu, di tahun 2012 saya melanjutkan pendidikan ke salah satu perguruan tinggi yang berada di Manado yaitu Universitas Sam Ratulangi Manado, dengan mengambil Program Studi S-1 Teknik Informatika di Jurusan Elektro Fakultas Teknik.