

Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Dipanegara)

Implementation of Data Mining for Evaluation of Student Academic Performance using a NBC Algorithm

M. Syukri Mustafa*¹, I Wayan Simpen²

^{1,2,3}Program Studi Teknik Informatika, STMIK Dipanegara, Makassar
E-mail: *¹moh.syukri@gmail.com, ²simpen.dp@gmail.com

Abstrak

Penelitian ini difokuskan untuk mengevaluasi kinerja akademik mahasiswa STMIK Dipanegara Makassar pada dua tahun pertama dengan menggunakan teknik data mining algoritma Naive Bayes Classifier (NBC) untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kinerja akademik mahasiswa yang kelulusannya akan diklasifikasikan dan memberikan rekomendasi untuk proses kelulusan tepat waktu yang paling tepat dengan nilai optimal berdasarkan histori nilai yang telah ditempuh mahasiswa. Sampel nilai yang digunakan untuk data latih dan testing adalah nilai mahasiswa angkatan 2008-2011 yang sudah dinyatakan lulus, sedangkan mahasiswa angkatan 2013-2014 dan belum lulus akan digunakan sebagai data target. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa faktor yang paling mempengaruhi penentuan klasifikasi kinerja akademik seorang mahasiswa adalah Indeks Prestasi (IP) pada semester 1,2,3,4 dan jenis kelamin, sehingga faktor tersebut dapat menjadi bahan evaluasi terhadap pihak pengelola STMIK Dipanegara. Pengujian pada beberapa data mahasiswa angkatan 2008-2011 yang diambil secara acak, algoritma NBC menghasilkan nilai akurasi 92,3%.

Kata kunci — Algoritma Naive Bayes Classifier, Kinerja Akademik Mahasiswa.

Abstract

This study focused on evaluating the academic performance of STMIK Dipanegara Makassar students in the first two years using Naive Bayes Classifier (NBC) data mining algorithm to form a probability table as the basis of the classification process of academic performance of students whose graduation will be classified and provide recommendations for timely graduation process the most appropriate with the optimal value based on the value history that has been taken by students. Sample value used for train and testing data is the value of students of class of 2008-2011 which have been passed, while student of class of 2013-2014 and not yet pass will be used as target data. The results obtained from this study indicate that the factors that most influence the classification of academic performance of a student is the Grade Index (IP) in semester 1,2,3,4 and gender, so that these factors can be used as an evaluation for the manager STMIK Dipanegara. Tests on some data 2008-2011 generation students are captured at random, the algorithm NBC produces a value 92.3% accuracy.

Keywords— Naive Bayes classifier algorithm, Student Academic Performance.

1. Pendahuluan

Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Pemantauan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, dan rasio kelulusan terhadap jumlah total mahasiswa, dan kompetensi lulusan semestinya mendapatkan perhatian yang serius untuk memperoleh kepercayaan *stakeholder* dalam menilai dan menetapkan kelulusannya.

Berdasarkan uraian di atas, pada penelitian ini akan dibuat sebuah sistem untuk mengklasifikasikan kelulusan mahasiswa dengan cara mengevaluasi kinerja pada tahun pertama dan tahun kedua. Pada penelitian ini, digunakan teknik data mining untuk menemukan pola kelulusan mahasiswa yang sudah lulus, kemudian dijadikan dasar untuk mengklasifikasikan mahasiswa tahun kedua yang bisa lulus tepat waktu dan yang tidak.

Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan trend dengan menambang sejumlah repositori data dalam jumlah yang besar, memanfaatkan teknologi pengenalan pola seperti statistik ataupun teknik matematika [1].

Penelitian sejenis yang dilakukan oleh Green [2], menerapkan proses data mining untuk mengevaluasi performa akademik mahasiswa dengan metode klasifikasi di Universitas Klabat. Metode yang digunakan



pada penelitian ini adalah metode eksperimen dan deskriptif, Hasil yang didapat yaitu untuk mencapai semester akhir dalam 4 tahun seorang mahasiswa diawal semester berdasarkan pohon keputusan yang terbentuk yaitu mahasiswa yang mempunyai indeks prestasi 3.5 – 3.87. Dalam penelitian ini, Teknik yang digunakan adalah algoritma Naïve Bayes Classifier (NBC) yang merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan teorema bayes. Algoritma NBC dapat mengolah data kuantitatif dan data diskrit yang hanya memerlukan sejumlah kecil data pelatihan untuk perhitungan estimasi peluang yang dibutuhkan untuk klasifikasi. Perhitungan Algoritma NBC dibandingkan dengan algoritma klasifikasi yang lain lebih cepat karena hanya menguji probabilitas dengan menemukan class yang sama dari data training.

Berdasarkan uraian latar belakang di atas, maka pokok permasalahannya adalah bagaimana mengklasifikasikan mahasiswa yang telah menyelesaikan 4 semester yang dapat lulus tepat waktu dan yang tidak dapat lulus tepat waktu dengan melihat pola kelulusan mahasiswa STMIK Dipanegara Makassar beberapa periode sebelumnya menggunakan Algoritma *Naive Bayes Classifier*?

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah implementasi dan penerapan data mining dengan teknik klasifikasi menggunakan algoritma naïve bayes classifier.

2.1. Data mining

2.1.1. Definisi Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis.

Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan lainnya [3]

2.1.2. Metode Data Mining

Data mining memiliki berbagaimetode atau fungsi yang dapat digunakan untuk menggali dan menemukan pengetahuan. Menurut Susanto & Suryadi, [4] ada enam kelompok fungsional *data mining*, yaitu :

1. Deskripsi (*description*), memberi gambaran secara ringkas terhadap sejumlah data yang berskala besar dan memiliki banyak jenis. Termasuk di dalamnya metode *Decision Tree*, *Exploratory Data Analysis* dan *Neural Network*.
2. Estimasi (*estimation*), menerka suatu nilai yang belum diketahui, misalnya menerka penghasilan seseorang ketika beberapa informasi mengenai orang tersebut sudah diketahui. Metode yang dapat digunakan adalah *Point Estimation*, *Confidence Interval Estimations*, *Simple Linear Regression*, *Correlation*, dan *Multiple Regression*.
3. Prediksi (*prediction*), memperkirakan suatu nilai di masa mendatang, misalnya memprediksi stok barang tiga tahun ke depan. Yang termasuk fungsi ini antara lain metode *Neural Network*, *Decision Tree*, dan *k-Nearest Neighbor*.
4. Klasifikasi (*Classification*), merupakan proses dalam menemukan suatu model atau fungsi yang dapat membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Yang termasuk dalam fungsi ini antara lain *Neural Network*, *Decision Tree*, *k-Nearest Neighbor*, dan *Naive Bayes*.
5. Pengelompokan (*Clustering*), pengelompokan untuk mengidentifikasi data dengan karakteristik tertentu. Yang termasuk dalam fungsi ini diantaranya model *Hierarchical Clustering*, metode *K-Means*, dan *Self Organizing Map (SOM)*
6. Asosiasi (*Association*), biasa disebut juga analisis keranjang pasar dimana fungsi ini digunakan untuk mengidentifikasi item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain. Yang termasuk dalam metode atau algoritma dalam fungsi ini antara lain *Apriori*, *Generalized Sequential Pattern (GSP)*, *FP-Growth* dan *GRI Algorithm*

2.2. Algoritma Naive Bayes Classifier

2.2.1. Definisi Algoritma

Menurut *Suarga*, [5], Algoritma adalah (1) teknik penyusunan langkah/tahap penyelesaian masalah dalam bentuk kalimat dengan jumlah kata yang terbatas, tetapi tersusun secara logis dan sistematis. (2) Suatu prosedur yang jelas untuk menyelesaikan suatu persoalan dengan menggunakan langkah-langkah tertentu dan terbatas jumlahnya. (3) Susunan langkah yang pasti, yang bila diikuti maka akan mentransformasi data input menjadi output yang berupa informasi.

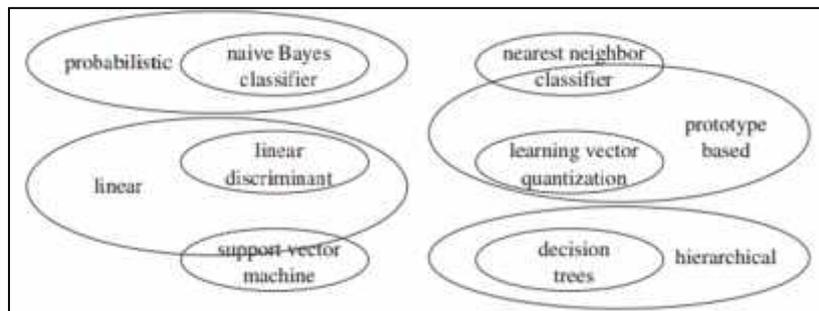
2.2.2. Pengertian Algoritma NBC.

Naive Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian.

Olson dan Delen [6] dalam bukunya, menjelaskan bahwa Naive Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar. Algoritma ini berdasar pada mengasumsikan bahwa atribut obyek adalah sesuatu yang independen. Probabilitas yang terlibat dalam membuat perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan.

Naive Bayes Classifier memiliki akurasi lebih baik dibanding dengan model *classifier* lainnya. Penelitian yang dilakukan oleh Xhemali, Hinde dan Stone dalam jurnalnya "*Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training WebPages*" mengatakan bahwa "*Naive Bayes Classifier* memiliki tingkat akurasi yang lebih baik dibanding model *classifier* lainnya".

Pada gambar 1 dapat dilihat skema yang sering digunakan dalam proses klasifikasi, yang tentunya juga menyertakan *Naive Bayes Classifier*.



Gambar 1 Skema Klasifikasi Algoritma NBC

Formula perhitungan *Naive Bayes Classifier* berdasarkan probabilitas ditunjukkan sebagai berikut :

$$p(A|B), p(B) = p(B|A) \cdot p(A) \tag{1}$$

$$p(A_i|B) = \frac{p(A_i) \cdot p(B|A_i)}{\sum_{j=1}^r p(A_j) \cdot p(B|A_j)} \tag{2}$$

Dengan mengubah nilai A_i dan A_j kedalam vector "x" maka didapatkan bentuk formula sebagai berikut :

$$p(x|i) = \frac{p(i|x) \cdot p(x)}{\sum_{j=1}^r p(j) \cdot p(x|j)} \tag{3}$$

Adapun perhitungan *Naive Bayes Classifier* untuk data kontinu menggunakan *Distribusi Gauss* sebagai berikut:

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

Keterangan :

$p(x|i)$ = Probabilitas hipotesis x jika diberikan fakta atau record i (*Posterior probability*).

$p(i|x)$ = Mencari nilai parameter yang memberi kemungkinan yang paling besar (*Likelihood*)

$p(x)$ = *Prior probability* dari I (*Prior Probability*)

$p(i)$ = Jumlah *probability tuple* yang muncul

g = *Distribusi Gauss*

- \bar{x} = Rata-rata
- σ = Standar Deviasi

Jikap(x/i) dapat diketahui berdasarkan perhitungan diatas, maka kelas/label dari data sampel X adalah kelas (label) yang memiliki $p(x/i) * p(i)$ maksimum.

3. Hasil dan Pembahasan

Pada penelitian ini digunakan berbagai macam data, diantaranya data training, data testing, dan juga data *history* matakuliah.

3.1 Data Training

Data Training digunakan untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Data ini akan digunakan sebagai proses *mining* berupa 1115 sampel data induk akademik mahasiswa STMIK Dipanegara angkatan 2008-2011 yang sudah dinyatakan lulus, dengan pengklasifikasian 781 Mahasiswa Lulus Tepat Waktu, dan 334 Mahasiswa Tidak Lulus Tepat Waktu. Data ini memiliki atribut NIM, nama, jenis kelamin, IPS 1, IPS 2, IPS 3, IPS 4, IPK semester 4, dan keterangan lulus. Beberapa sampel data training dapat dilihat pada tabel 1.

3.2 Data Testing

Data Testing digunakan untuk mengukur sejauh mana algoritma NBC berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada pada testing set seharusnya tidak boleh ada pada training set sehingga dapat diketahui apakah model *classifier* sudah “pintar” dalam melakukan klasifikasi. Data ini akan digunakan untuk pengujian, berupa data akademik mahasiswa angkatan 2013-2014. Data ini memiliki atribut NIM, nama, jenis kelamin, IPS 1, IPS 2, IPS 3, IPS 4, dan IPK semester 4. Setelah proses *mining*, data ini akan memiliki kelas berdasarkan probabilitas yang diperoleh dari data *training*. Beberapa sampel data testing dan juga hasil klasifikasinya dapat dilihat pada tabel 2 dan tabel 3.

Tabel 1. *Data Training*

No	NIM	Nama	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	IPS5	Kelulusan
1	082008	HARTVIQ BATURNTE	Laki-Laki	3,285714	0,714286	2,571429	2,25	2,166667	Terlambat
2	082010	ILHAM	Laki-Laki	3,666667	3,904762	3,181818	3,1	3,464286	Tepat Waktu
3	082011	SYAEPUL SYAH ALAM	Laki-Laki	0	0	1,153846	1,833333	0,627119	Terlambat
4	082012	MUH. IKBAL S.	Laki-Laki	2,666667	3,428571	2,444444	2,705882	2,831169	Tepat Waktu
5	082018	MUIS NANJA	Laki-Laki	3,476191	3,476191	3,4	3,05	3,353658	Tepat Waktu
6	092052	THEO XAVER YANSEN	Laki-Laki	2,714286	2,476191	1,666667	1,888889	2,376344	Terlambat
7	092053	FERDY BUDYANTO HALIK.	Laki-Laki	3,476191	3,476191	3,045455	2,454545	3,240385	Tepat Waktu
8	092054	MUHAMMAD IMRAN L.	Laki-Laki	2,619048	2,238095	2,058824	2,5	2,329897	Tepat Waktu
9	092057	SISWANA ASIS	Perempuan	3,571429	3,904762	3,095238	3,388889	3,568627	Tepat Waktu
10	092058	ARDY MAULANA WAHAB	Laki-Laki	1,952381	2,761905	0,571429	0,6875	1,477273	Terlambat
11	092059	ANDY PUTRA ANUGRAH.	Laki-Laki	2,428571	2,476191	2,882353	2,3	2,494737	Tepat Waktu
12	102119	YUNI MARSELINA PANGESSO	Perempuan	1,857143	2,071429	2,611111	3	2,315217	Terpambat
13	102121	NITSAR PALECHE	Laki-Laki	0,571429	0,125	2,136364	1,529412	1,173333	Terlambat
14	112052	VERONICA WIJAYA	Perempuan	3,428571	3,571429	3,318182	2,952381	3,317647	Tepat Waktu
15	112059	ELSYE KIONG	Laki-Laki	3,285714	3,285714	1,117647	1,25	2,548387	Terlambat
16
:

Tabel 2. Data Testing

No	NIM	Nama	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	IPK
1	132011	HASNIATI PAMULA	Perempuan	2,857143	2,190476	2,380952	2	2,37037
2	132012	YOSIA TANDIALO	Laki-Laki	3,095238	3,333333	2,952381	3	3,095238
3	132013	SIGIT ADRIAN RH.	Laki-Laki	3,428571	3,047619	3,523809	3,454545	3,364706
4	142026	MIA AUDINA M	Perempuan	3,5	3,454545	3,333333	3,142857	3,365079
5	142028	FALMASARI SABBAN	Perempuan	3,1	2,863636	3,190476	1,809524	2,738095
6	142029	RESKI AMALIA	Laki-Laki	3,636364	3,9	3,333333	3,523809	3,68254
7	142037	CRISTIAN OGESTA M	Laki-Laki	3,00	2,0	1,047619	2,174286	2,21428
8	142038	NURAEANI	Perempuan	2,272727	2,65	2,285714	2,857143	2,587301

Tabel 3. Hasil Klasifikasi Data Testing

No	NIM	Nama	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	IPK	Hasil Klasifikasi	Rekomendasi
1	132011	HASNIATI PAMULA	Perempuan	2,85714	2,19048	2,38095	2	2,37037	Tepat Waktu	Tidak
2	132012	YOSIA TANDIALO	Laki-Laki	3,09524	3,33333	2,95238	3	3,09524	Tepat Waktu	Tidak
3	132013	SIGIT ADRIAN RH.	Laki-Laki	3,42857	3,04762	3,52381	3,45455	3,36471	Tepat Waktu	Tidak
4	142026	MIA AUDINA M	Perempuan	3,5	3,45455	3,33333	3,14286	3,36508	Tepat Waktu	Tidak
5	142028	FALMASARI SABBAN	Perempuan	3,1	2,86364	3,19048	1,80952	2,7381	Tepat Waktu	Tidak
6	142029	RESKI AMALIA	Laki-Laki	3,63636	3,9	3,33333	3,52381	3,68254	Tepat Waktu	Tidak

3.3 Data Riwayat Matakuliah

Data ini digunakan untuk mengevaluasi *data testing* ketika diklasifikasikan terlambat. Data ini akan dianalisis untuk memberikan rekomendasi dalam perkuliahan semester berikutnya.

Data Riwayat matakuliah dapat dilihat pada tabel 4.

Tabel 4. Data Riwayat Matakuliah

No	NIM	Nama	Mata Kuliah	Nilai
1	142026	MIA AUDINA M	ALGORITMA DAN PEMROGRAMAN	A
2	142026	MIA AUDINA M	PRAK. ALGORITMA DAN PEMROGRAMAN	A
3	142026	MIA AUDINA M	PENGEMBANGAN DIRI	A
4	142026	MIA AUDINA M	KALULUS I	B
5	142026	MIA AUDINA M	PRAK. PENGENALAN TEKNOLOGI INFORMASI	A
6	142026	MIA AUDINA M	PRAK. ELEKTRONIKA ANALOG	A
7	142026	MIA AUDINA M	BAHASA INGGRIS I	B
8	142026	MIA AUDINA M	PENGENALAN TEKNOLOGI INFORMASI	A
9	142026	MIA AUDINA M	PENDIDIKAN AGAMA ISLAM	B
10	142026	MIA AUDINA M	ELEKTRONIKA ANALOG	C
11	142026	MIA AUDINA M	BAHASA INGGRIS II	C
12	142026	MIA AUDINA M	ELEKTRONIKA ADIGITAL	A
13	142026	MIA AUDINA M	HUKUM DAN KOMPUTER	A
14	142026	MIA AUDINA M	INTERAKSI MANUSIA DAN KOMPUTER	A
15	142026	MIA AUDINA M	KALULUS II	C
16	142026	MIA AUDINA M	KECAKAPAN ANTAR PERSONAL	A
17	142026	MIA AUDINA M	KEWIRUSAHAAN	A
18	142026	MIA AUDINA M	SISTEM BASIS DATA	B
19	142026	MIA AUDINA M	PANCASILA DAN KEWARGANEGARAAN	A
20	142026	MIA AUDINA M	PENG. FORENSIK TEKNOLOGI INFORMASI	A
21	142026	MIA AUDINA M	PRAKTIKUM ELEKTRONIKA DIGITAL	A
22	142026	MIA AUDINA M	PRAK. SISTEM BASIS DATA	C
23	142026	MIA AUDINA M	ANALISIS DAN DESAIN SISTEM	A
24	142026	MIA AUDINA M	MIKROKONTROLER	A
25	142026	MIA AUDINA M	MULTIMEDIA DAN KONVERSI	A
26	142026	MIA AUDINA M	PEMROGRAMAN VISUAL .NET	B
27	142026	MIA AUDINA M	PENGANTAR INTELIGENSI BUATAN	C
28	142026	MIA AUDINA M	PRAK. MIKRO KONTROLLER	A
29	142026	MIA AUDINA M	PRAK. VISUAL .NET	B
30	142026	MIA AUDINA M	SISTEM OPERASI KOMPUTER	B
31	142026	MIA AUDINA M	STATISTIKA	B
32	142026	MIA AUDINA M	DATA WAREHOUSE DAN DATA MINING	B
33	142026	MIA AUDINA M	JARINGAN KOMPUTER	B
34	142026	MIA AUDINA M	METODE NUMERIK	C
35	142026	MIA AUDINA M	ORGANISASI DAN ARSITEKTUR KOMPUTER	A
36	142026	MIA AUDINA M	PEMROGRAMAN WEB	B
37	142026	MIA AUDINA M	PRAK. JARINGAN KOMPUTER	A
38	142026	MIA AUDINA M	PRAK. PEMROGRAMAN WEB	C
39	142026	MIA AUDINA M	REKAYASA PERANGKAT LUNAK	A
40	142026	MIA AUDINA M	STRUKTUR DATA	B

3.4 Implementasi Algoritma Naïve Bayes Classifier

Implementasi algoritma *Naïve Bayes Classifier* dalam penelitian ini menggunakan Data Testing yang diberikan dapat dilihat pada tabel 5.

Tabel 5. Data Testing

No	NIM	JKL	IPS 1	IPS 2	IPS 3	IPS 4	IPK
1	142037	Laki – Laki	3	2	1,047619	2,173913	2,214286
2	142038	Perempuan	2,272727	2,65	2,285714	2,857143	2,587301

Langkah-langkah Algoritma NBC dan data pendukungnya dapat dilihat pada tabel 6, tabel 7 dan tabel 8 sebagai berikut :

1. Menentukan *Prior Probability* (P)

Tabel 6. Data Kelulusan

KELULUSAN	JUMLAH KEJADIAN
Tepat Waktu	781
Terlambat	334
JUMLAH	1115

$$P(\text{Tepat Waktu}) = \frac{781}{1115} = 0,70044 \quad P(\text{Terlambat}) = \frac{334}{1115} = 0,29955$$

2. Menentukan probabilitas kemunculan setiap nilai untuk atribut Jenis Kelamin (X1).

Tabel 7. Data Jenis Kelamin

JENIS KELAMIN (X1)	JUMLAH KEJADIAN	
	Tepat Waktu	Terlambat
Laki-Laki	576	294
Perempuan	205	40
JUMLAH	781	334

3. Menentukan *Mean* dan Standar Deviasi untuk atribut IPS1 (X2), IPS2 (X3), IPS3 (X4), IPS4 (X5), IPK (X6) (Tabel 7), karena nilai untuk atribut berupa data diskrit, yang nantinya akan dimasukkan kedalam rumus Distribusi Gauss.

Tabel 8. Data Mean dan Stdv. IP Semester 1 sampai 6

IPS1 (X2)		IPS2 (X3)			
Tepat Waktu	Terlambat	Tepat Waktu	Terlambat		
3,666667	3,285714	3,904762	0,7142857		
2,666667	0	3,428571	0		
.....		
2,904762	3	2,285714	3,142857		
Mean	3,069460907	2,233810369	Mean	3,14080341	2,358325649
Stdv	0,522886664	0,831114459	Stdv	0,477485491	0,766993131

IPS3 (X4)		IPS4 (X5)	
Tepat Waktu	Terlambat	Tepat Waktu	Terlambat
3,181818	2,571429	3,1	2,25
2,444444	1,153846	2,705882	1,833333
.....

	2,888889	2,45
Mean	2,994895206	1,959281693
Stdv	0,534886209	0,813286468

	3,315789	2,555556
Mean	2,977060345	1,698225088
Stdv	0,568959766	0,883503517

IPK (X6)	
Tepat Waktu	Terlambat
3,464286	2,166667
2,831169	0,6271186
.....
2,835443	2,85
Mean	3,103901497
Stdv	0,384092914

4. Menghitung Probabilitas dari setiap atribut

a. Pengujian data testing mahasiswa ke-1 :

X1 = "Laki - Laki", X3 = 2, X5 = 2,173913,
 X2 = 3, X4 = 1,047619, X6 = 2,214286

$$P(X1 = \text{"Laki - Laki"} | \text{Tepat Waktu}) = 576/781 = 0,737516005$$

$$P(X1 = \text{"Laki - Laki"} | \text{Terlambat}) = 294/334 = 0,880239520$$

$P(X2 = 3 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,522886664)} e^{-\frac{(3-2,994895206)^2}{2(0,522886664)^2}} = 0,7562589665186071$$

$P(X2 = 3 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,813286468)} e^{-\frac{(3-1,959281693)^2}{2(0,813286468)^2}} = 0,31388617123865404$$

$P(X3 = 2 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,477485491)} e^{-\frac{(2-2,831169)^2}{2(0,477485491)^2}} = 0,048130909257542674$$

$P(X3 = 2 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,766953131)} e^{-\frac{(2-2,835443)^2}{2(0,766953131)^2}} = 0,4663630964137604$$

$P(X4 = 1,047619 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,534886209)} e^{-\frac{(1,047619-2,994895206)^2}{2(0,534886209)^2}} = 9,87816404341 \times 10^{-7}$$

$P(X4 = 1,047619 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,813286468)} e^{-\frac{(1,047619-1,959281693)^2}{2(0,813286468)^2}} = 0,261703308008$$

$P(X5 = 2,173913 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,568959766)} e^{-\frac{(2,173913-2,977060345)^2}{2(0,568959766)^2}} = 0,25890099899$$

$P(X5 = 2,173913 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,883503517)} e^{-\frac{(2,173913-1,698225088)^2}{2(0,883503517)^2}} = 0,390619371335$$

$P(X6 = 2,214286 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,384092914)} e^{-\frac{(2,214286-3,103901497)^2}{2(0,384092914)^2}} = 0,0710525685$$

$P(X6 = 2,214286 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,531462498)} e^{\frac{-(2,214286-2,154584084)^2}{2(0,531462498)^2}} = 0,7459206464469$$

Sehingga:

Likelihood Tepat Waktu

$$= 0,737516005 \times 0,7562589865186071 \times 0,048130909257542674 \times 9,87816404341 \times 10^{-7} \times 0,25890099899 \times 0,0710525685 \times 0,70044 = 3,4168976744141945 \times 10^{-7}$$

Likelihood Terlambat

$$= 0,880239520 \times 0,31383617123865404 \times 0,4663630964137604 \times 0,261703308008 \times 0,390619371335 \times 0,7459206464469 \times 0,29955 = 0,002942762778336564$$

Kesimpulan kelulusan = **Terlambat**,
karena nilai *Likelihood* Terlambat Lebih besar dari *Likelihood* Tepat Waktu



Gambar 2. Hasil Pengujian Data Testing NIM 142037

b. Pengujian data testing mahasiswa ke-1

$$\begin{aligned} X1 &= \text{"Perempuan"}, & X4 &= 2,285714, \\ X2 &= 2,272727, & X5 &= 2,857143 \\ X3 &= 2,65, & X6 &= 2,587301 \\ P(X1 = \text{"Perempuan"} | \text{Tepat Waktu}) &= 205/781 = 0,262483994 \\ P(X1 = \text{"Perempuan"} | \text{Terlambat}) &= 40/334 = 0,119760479 \end{aligned}$$

$P(X2 = 2,272727 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,522886664)} e^{\frac{-(2,272727-2,069460907)^2}{2(0,522886664)^2}} = 0,2389711749794$$

$P(X2 = 2,272727 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,831114459)} e^{\frac{-(2,272727-2,255310869)^2}{2(0,831114459)^2}} = 0,4794828990530$$

$P(X3 = 2,65 | \text{Tepat Waktu})$

$$= \frac{1}{\sqrt{2\pi} (0,477485491)} e^{\frac{-(2,65-2,14020841)^2}{2(0,477485491)^2}} = 0,49262956228067$$

$P(X3 = 2,65 | \text{Terlambat})$

$$= \frac{1}{\sqrt{2\pi} (0,766993131)} e^{\frac{-(2,65-2,552225649)^2}{2(0,766993131)^2}} = 0,4838556637577135$$

$$\begin{aligned}
 &P(X4 = 2,285714 | \text{Tepat Waktu}) \\
 &= \frac{1}{\sqrt{2\pi} (0,534886209)} e^{-\frac{(2,285714 - 2,994895206)^2}{2(0,534886209)^2}} = 0,309690408669 \\
 &P(X4 = 2,285714 | \text{Terlambat}) \\
 &= \frac{1}{\sqrt{2\pi} (0,813286468)} e^{-\frac{(2,285714 - 1,959281698)^2}{2(0,813286468)^2}} = 0,452567961823 \\
 &P(X5 = 2,857143, | \text{Tepat Waktu}) \\
 &= \frac{1}{\sqrt{2\pi} (0,568959766)} e^{-\frac{(2,857143 - 2,977060845)^2}{2(0,568959766)^2}} = 0,68577607974897 \\
 &P(X5 = 2,857143, | \text{Terlambat}) \\
 &= \frac{1}{\sqrt{2\pi} (0,883503517)} e^{-\frac{(2,857143 - 1,698225088)^2}{2(0,883503517)^2}} = 0,19101638092868 \\
 &P(X6 = 2,587301 | \text{Tepat Waktu}) \\
 &= \frac{1}{\sqrt{2\pi} (0,384092914)} e^{-\frac{(2,587301 - 2,108901497)^2}{2(0,384092914)^2}} = 0,420393299 \\
 &P(X6 = 2,587301 | \text{Terlambat}) \\
 &= \frac{1}{\sqrt{2\pi} (0,531462498)} e^{-\frac{(2,587301 - 2,154584084)^2}{2(0,531462498)^2}} = 0,538830867
 \end{aligned}$$

Sehingga:

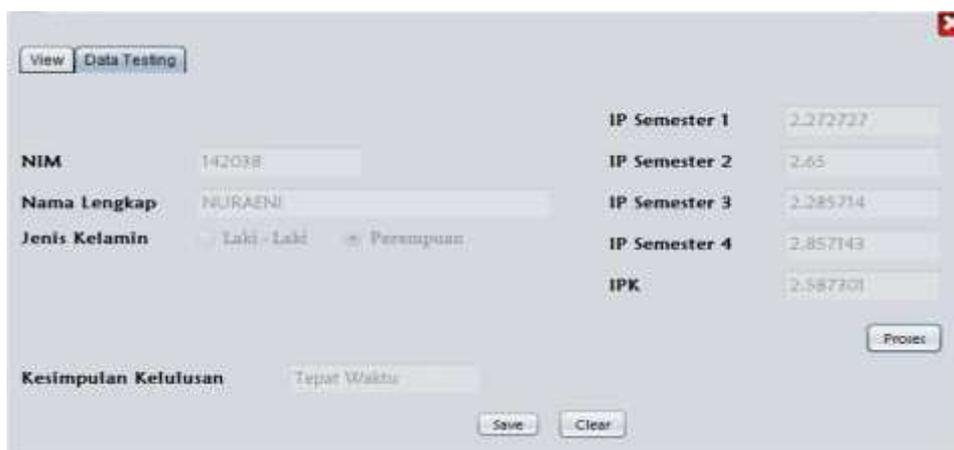
Likelihood Tepat Waktu

$$\begin{aligned}
 &= 0,262483994 \times 0,2389711749794 \times 0,49262956228067 \times \\
 &= 0,309690408669 \times 0,68577607974897 \times 0,420393299 \times 0,70044 \\
 &= \mathbf{0,0019324615231346462}
 \end{aligned}$$

Likelihood Terlambat

$$\begin{aligned}
 &= 0,119760479 \times 0,4794828990530 \times 0,4838556637577135 \times \\
 &= 0,452567961823 \times 0,19101638092868 \times 0,538830867 \times 0,29955 \\
 &= \mathbf{3,8768675256774254 \times 10^{-4}}
 \end{aligned}$$

Kesimpulan kelulusan = Tepat Waktu, karena nilai Likelihood Tepat Waktu lebih besar dari Likelihood Tepat Waktu, hasil tersebut dapat dilihat pada gambar 3.



Gambar 3. Hasil Pengujian Data Testing Nim 142038

3.5 Pengujian dan Akurasi Data

Dalam menguji akurasi dan ketepatan hasil pengklasifikasian pada penelitian ini, digunakan 26 data alumni yang diambil secara acak. 26 data tersebut tidak terdapat di dalam data training, hal ini

dimaksudkan agar hasil pengklasifikasian kelulusan dari sistem yang dirancang dapat dibandingkan dengan hasil kelulusan yang sesuai dengan data alumni mahasiswa STMIK Dipanegara Makassar.

Metode pengujian yang digunakan dalam penelitian ini adalah *Confusion Matrix*, yaitu suatu metode yang biasanya digunakan dalam melakukan perhitungan akurasi pada suatu data mining.

Perhitungan yang dihasilkan dari rumus ini terdiri dari 4 jenis keluaran, antara lain : *recall*, *accuracy*, *precision*, dan *error rate*.

1. *Recall* adalah proporsi kasus positif yang diidentifikasi dengan benar.
Rumus dari *recall* = $D/(C+D)$
2. *Accuracy* adalah perbandingan kasus yang teridentifikasi benar dengan jumlah semua kasus.
Rumus dari *accuracy* = $(A+D)/(A+B+C+D)$
3. *Precision* adalah proporsi kasus dengan hasil positif yang benar.
Rumus dari *Precision* = $D/(B+D)$
4. *Error Rate* adalah kasus yang teridentifikasi salah yang dibandingkan dengan jumlah semua kasus.
Rumus dari *Error Rate* = $(B+C)/(A+B+C+D)$

Keterangan:

- A = jika hasil prediksi Terlambat dan data sebenarnya Terlambat.
- B = jika hasil prediksi Tepat Waktu sedangkan nilai sebenarnya Terlambat.
- C = jika hasil prediksi Terlambat sedangkan nilai sebenarnya Tepat Waktu.
- D = jika hasil prediksi Tepat Waktu dan nilai sebenarnya Tepat Waktu.

$$A = 9, B = 1, C = 1, D = 15$$

$$Recall = \frac{15}{(1 + 15)} = 0,9375$$

$$Precision = \frac{15}{(1 + 15)} = 0,9375$$

$$Accuracy = \frac{(9 + 15)}{(9 + 1 + 1 + 15)} = 0,923076923$$

$$Error Rate = \frac{(1 + 1)}{(9 + 1 + 1 + 15)} = 0,076923077$$

Hasil pengujian menunjukkan *accuracy* sebesar 92.30%. Detail perbandingannya dapat dilihat pada gambar 4.

Nama	Jenis Kel...	IPS 1	IPS 2	IPS 3	IPS 4	IPK	P.Sistem	Kelulusan	Kecocokan
SANTI AB...	Perempu...	3.523809	3.714286	2.909091	2.888889	3.268293	Tepat Wa...	Tepat Wa...	Ya
RIVALDI ...	Laki - Laki	1.285714	2.952381	0.8571429	1.142857	1.671429	Terlambat	Terlambat	Ya
SITI NUR...	Perempu...	3.333333	3.428571	3.1	3.1	3.243902	Tepat Wa...	Tepat Wa...	Ya
MARKUS...	Laki - Laki	3.761905	3.523809	3	3.05	3.414634	Tepat Wa...	Tepat Wa...	Ya
FAJAR	Laki - Laki	2	2.428571	1.5825	2.4	2.109589	Terlambat	Terlambat	Ya
ASRUL P...	Laki - Laki	0.952381	1.952381	2.9375	2.7	2.131579	Terlambat	Tepat Wa...	Tidak
I KETUT ...	Laki - Laki	2.857143	3.476191	2.291667	2.888889	2.926829	Tepat Wa...	Tepat Wa...	Ya
ANDI AS...	Laki - Laki	1.619048	1.333333	0.6470588	1.882353	1.662921	Terlambat	Terlambat	Ya
BAHRUL	Laki - Laki	2.714286	2.809524	3.142857	3	3.131313	Tepat Wa...	Tepat Wa...	Ya
HARDIYA...	Perempu...	2.571429	2.523809	2.217391	2.695652	2.666667	Tepat Wa...	Tepat Wa...	Ya
SAHALTI...	Perempu...	3.428571	3.238095	3.428571	3.6	3.534653	Tepat Wa...	Tepat Wa...	Ya
SARTIKA...	Perempu...	2.523809	2.428571	2.565217	2.826087	2.754902	Tepat Wa...	Tepat Wa...	Ya
TEANGU...	Perempu...	3.692308	3.428571	3.454545	3.789474	3.402429	Tepat Wa...	Tepat Wa...	Ya

Recall : 93.75%
Precision : 93.75%
Accuracy : 92.3076923076923%
Error Rate : 7.6923076923076925%

Gambar4. Hasil Pengujian Akurasi

4. Kesimpulan

Berdasarkan hasil yang diperoleh pada penelitian ini, maka dapat disimpulkan bahwa :

1. Aplikasi data *mining* ini dapat mengklasifikasikan mahasiswa STMIK Dipanegara Makassar yang dapat lulus tepat waktu dan yang tidak dapat lulus tepat waktu dengan menggunakan Algoritma *Naive Bayes Classifier*.
2. Semakin banyak data *training* yang digunakan, maka tingkat Recall, Precision, Accuracion sistem akan semakin baik.
3. Berdasarkan hasil pengujian akurasi, bahwa ada faktor-faktor yang mempengaruhi kelulusan mahasiswa STMIK Dipanegara Makassar, bukan hanya dari faktor akademik saja, tetapi faktor non-akademik juga mempengaruhi.

5. Saran

Aplikasi ini hanya menggunakan satu metode klasifikasi data mining saja, yaitu menggunakan Algoritma Naive Bayes Classifier, jadi disarankan untuk pengembangan yang lebih lanjut bisa dibuat beberapa metode klasifikasi data mining yang lain, agar hasil dari beberapa metode tersebut dapat dibandingkan keakuratannya.

Daftar Pustaka

- [1] Fina Nasari, 2016, "*Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare di Kabupaten Langkat*", Cogito Smart Journal, Vol. 2, No. 2, Desember 2016.
- [2] Fajar Astuti Hermawati, 2013, "*Data Mining*", Penerbit Andi, Yogyakarta.
- [3] Green Ferry Mandias, 2015, "*Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi*", Konferensi Nasional Sistem & Informatika 2015, STIKOM Bali, 9-10 Oktober 2015.
- [4] Susanto, Sani dan Suryadi, Dedy, 2010, "*Pengantar Data Mining Menggali dari Bongkahan Data*", Penerbit Andi, Yogyakarta.
- [5] Drs. Suarga, M.Sc, M.Math, Ph.D, 2012, "*Algoritma dan Pemrograman*", Penerbit Andi, Yogyakarta.
- [6] David L. Olsen dan Dursun Delen, 2008, "*Advanced Data Mining Techniques*", Penerbit pringer, USA.