

PREDIKSI TINGKAT KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN NAÏVE BAYES: STUDI KASUS FAKULTAS EKONOMI DAN BISNIS UNIVERSITAS PENDIDIKAN NASIONAL

¹⁾I Gd I. Suwardika, ²⁾I Gst Ngr Suariana, ³⁾IB P. Bhiantara, ⁴⁾N. Y. Arso

^{1,2,3}Program Studi Ilmu Komputer Program Pascasarjana Universitas Pendidikan Ganesha

¹indrasuwardika@gmail.com, ²ngurah_suariana@kemenag.go.id, ³Gusss.yoga@gmail.com,
⁴norzayudhi@icloud.com.

Abstrak

Data merupakan kumpulan informasi yang tersimpan dari kegiatan atau transaksi. Data ini akan memberikan informasi yang lebih baik dan dapat dimanfaatkan untuk pengambilan keputusan dengan menggunakan data mining. Universitas Pendidikan Nasional merupakan Perguruan Tinggi Swasta yang memiliki keunggulan tidak hanya lingkup lokal namun mencakup Bali dan Nusa Tenggara. Fakultas Ekonomi dan Bisnis dengan Akreditasi BAN PT A pada Program Studi Manajemen dan B pada Program Studi Akuntansi pada Universitas Pendidikan Nasional merupakan salah satu Program Studi yang diminati oleh calon mahasiswanya. Tantangan Universitas dalam menyelenggarakan pendidikan adalah menghasilkan sarjana yang mempunyai kompetensi, bersaing dan beradaptasi di dunia global serta lulus sesuai waktu yang telah ditentukan. Prediksi tingkat kelulusan adalah salah satu faktor pendukung dalam memperbaiki mutu dan layanan pendidikan pada lingkup Universitas Pendidikan Nasional, oleh karena itu penelitian Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu menggunakan Naive Bayes pada Fakultas Ekonomi dan Bisnis menjadi hal yang sangat penting untuk dilakukan. Dengan menggunakan salah satu tehnik data mining yaitu Naive Bayes dapat dilakukan prediksi terhadap tingkat kelulusan tepat waktu mahasiswa fakultas tersebut terhadap data yang ada. Hasil penelitian ini dapat disimpulkan bahwa Visi dan Misi yang telah dijalankan oleh Universitas Pendidikan Nasional Fakultas Ekonomi dan Bisnis telah menghasilkan sarjana tepat waktu empat tahun dengan prosentase 98 persen.

Kata Kunci : Naïve Bayes, Data Mining, Universitas Pendidikan Nasional.

Abstract

Data is a collection of information stored from activities or transactions. This data will provide better information and can be utilized for decision making by using data mining. The University of National Education is a private university that has advantages not only in the local sphere but in Bali and Nusa Tenggara. The Faculty of Economics and Business with the BAN PT A Accreditation in the Management Study Program and B in the Accounting Study Program at the National Education University is one of the Study Programs that are of interest to prospective students. The challenge of the University in organizing education is to produce graduates who have competence, compete and adapt in the global world and graduate according to a predetermined time. Graduation prediction is one of the supporting factors in improving the quality and education services within the scope of the National Education University, therefore research on Prediction of Student Graduation on Time using Naive Bayes at the Faculty of Economics and Business is very important to do. By using one of the data mining techniques, Naive Bayes, predictions can be made on the student's on-time graduation rates for the existing data. The results of this study can be concluded that the Vision and Mission that has been carried out by the National University of

Education, Faculty of Economics and Business have produced scholars on time four years with a percentage of 98 percent.

Keywords : Naïve Bayes, Data Mining, Universitas Pendidikan Nasional

I. PENDAHULUAN

Menurut kamus besar Bahasa Indonesia klasifikasi adalah penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan [1]. Klasifikasi juga merupakan proses pengidentifikasian obyek ke dalam sebuah kelas, kelompok, atau kategori berdasarkan prosedur, karakteristik dan definisi yang telah ditentukan sebelumnya [2]. Klasifikasi merupakan salah satu komponen dalam data mining agar data bisa memberikan makna/ bisa dimanfaatkan. Klasifikasi adalah salah satu tugas yang penting dalam data mining, dalam klasifikasi, sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan sebelumnya.

Klasifikasi data adalah proses dua langkah. Pada langkah pertama, sebuah model dibangun menggambarkan sebuah kumpulan kelas data atau konsep dari populasi data yang telah ditentukan sebelumnya (misalkan data rekening koran/ pendapatan). Model tersebut dibangun dengan menganalisa data latih yang digambarkan oleh atribut-atribut. Tiap tuple di asumsikan untuk dimiliki oleh kelas yang telah di tentukan, seperti di tentukan oleh salah satu atribut, yang dinamakan class *label attribute*. Langkah kedua adalah menguji model yang telah dibangun kepada data uji untuk mengukur ketepatan atau performa model dalam mengklasifikasi data uji. Setelah pengukuran performa selesai dilakukan, pengambil keputusan dapat memutuskan untuk menggunakan model tersebut atau mengulang pembuatan model dengan data latih atau metode yang berbeda untuk menghasilkan model klasifikasi yang lebih baik [3].

Salah satu bentuk klasifikasi adalah naïve bayes, naïve bayes merupakan metode “supervised learning” karena menggunakan data awal sebagai data pembelajaran. Naive bayes merupakan metode penyederhanaan dari teorema bayes yang merupakan metode dengan pendekatan statistik dengan probabilitas

bersyarat pada persoalan klasifikasi. Penerapan metode Naïve Bayes ini dilakukan karena karakter metode Naïve Bayes bersifat independen pada tiap fituranya. Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang cukup efektif dan efisien untuk machine learning dan data mining. Pada makalah ini kami berusaha untuk melakukan eksplorasi terhadap algoritma naïve bayes. Dengan menggunakan contoh kasus, kami mencoba memaparkan naïve bayes dengan lebih baik sehingga hasil yang didapat melalui proses perhitungan bisa dipahami

II. TINJAUAN PUSTAKA

2.1 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 [4].

Menurut [5], 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 mning dapat digabungkan dalam kasus tertentu untuk menjawab masalah yang ada [6]. 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 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. E-Journal Teknik Informatika Vol 11, No.1 (2017) ISSN : 2301 - 8364
7. Deviation Analysis Fungsi dari deviation analysis adalah untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

2.2 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 [7]. 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.

2.3 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” [7].

$$\text{Probabilitas (H|E,e)} = p(i)$$

Dimana :

e = evidence

E = evidence observasi baru

$P(H|E,e)$ = probabilitas hipotesis H benar jika muncul evidence

baru dari E dari evidence e

$p(H|E)$ = probabilitas hipotesis 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

2.4 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.

2.5 Preprocessing

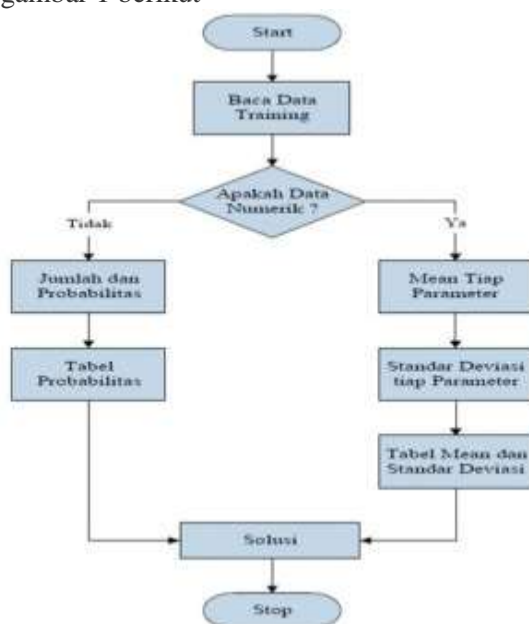
Preprocessing merupakan tahapan awal dalam pengolahan data. beberapa kasus pasti memiliki kualitas data yang kurang baik. Data yang memiliki kualitas data yang kurang baik memiliki kekurangan nilai-nilai atribut atau atribut tertentu, berisi kesalahan atau nilai-nilai outlier yang menyimpang dari yang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

III. METODE PENELITIAN

3.1 Alur dan Metode Naive Bayes

Alur dari metode Naive Bayes dapat dilihat pada gambar 1 berikut



Gambar. 1 Alur Metode Naive Bayes

Kelebihan dan Kelemahan Naive Bayes

Kelebihan metode naive bayes

- Klasifikasi yang dilakukan 40relative lebih cepat
- Tidak sensitive terhadap fitur yang tidak relevan
- Lebih mudah dipahami
- Pengkodean yang diperlukan relative sederhana

Kekurangan metode naive bayes

- Asumsi data yang diberikan adalah independent
- Metode bayes memerlukan pengetahuan awal dalam mengambil keputusan, tingkat keberhasilan metode ini

tergantung seberapa baik pengetahuan awal yang diberikan

3.2 Laplace Correction

Laplace Correction (Laplacian Estimator) atau additive smoothing adalah suatu cara untuk menangani nilai probabilitas 0 (nol). Dari sekian banyak data di training set, pada setiap perhitungan datanya ditambah 1 (satu) dan tidak akan membuat perbedaan yang berarti pada estimasi probabilitas sehingga bisa menghindari kasus nilai probabilitas 0 (nol). Laplace Correction ini dapat dinyatakan dalam persamaan seperti pada persamaan (2) berikut ini

$$p_i = \frac{m_i + 1}{n + k} \quad (2)$$

Keterangan

- p_i : probabilitas dari atribut m_i
- m_i : jumlah sampel dalam kelas dari atribut m_i
- k : jumlah kelas dari atribut m_i
- n : jumlah sampel

Probabilitas berbasis frekuensi mungkin menghasilkan nilai nol(0) ketika mengalikan probabilitas, yang menyebabkan kegagalan dalam menghasilkan informasi yang dikontribusikan oleh probabilitas yang tidak nol. Oleh karena itu, pendekatan *Laplace Correction*, harus diadopsi untuk mengatasi masalah ini.

Sebagai contoh, asumsikan ada class Penggunaan Listrik=rendah di suatu training set, memiliki 3 sampel, ada 0 sampel dengan Perlengkapan=sedikit, 3 sampel dengan Perlengkapan=sedang, dan 0 sampel dengan Perlengkapan=banyak. Probabilitas dari kejadian ini tanpa Laplacian Correction adalah $P(\text{Perlengkapan=sedikit}|\text{Penggunaan Listrik=rendah})=0$, $P(\text{Perlengkapan=sedang}|\text{Penggunaan Listrik=rendah})=1.000$ (dari 3/3), dan $P(\text{Perlengkapan=banyak}|\text{Penggunaan Listrik=rendah})=0$.

Menggunakan *Laplacian Correction* dari tiga kejadian diatas, diasumsikan ada 1 sampel lagi untuk masing – masing nilai Penggunaan Listrik=banyak. Dengan cara ini, didapatkanlah

probabilitas sebagai berikut (dibulatkan menjadi 3 angka dibelakang koma):

$P(\text{Perlengkapan}=\text{sedikit}|\text{Penggunaan Listrik}=\text{rendah})=0.166$ (dari $1/6$),
 $P(\text{Perlengkapan}=\text{sedang}|\text{Penggunaan Listrik}=\text{rendah})=0.666$ (dari $4/6$), dan
 $P(\text{Perlengkapan}=\text{banyak}|\text{Penggunaan Listrik}=\text{rendah})=0.166$ (dari $1/6$)

Probabilitas yang “dibenarkan” hasilnya tidak berbeda jauh dengan hasil probabilitas sebelumnya sehingga nilai probabilitas 0 (nol) dapat dihindari.

3.3 Weka Software

Waikato Environment for Knowledge Analysis atau sering disebut dengan WEKA merupakan salah satu software data mining yang dikembangkan oleh Universitas Waikato, New Zeland. Weka di implementasikan pertamakali pada tahun 1997 dan mulai menjadi open source pada tahun 1999, hingga dewasa ini Weka sudah mencapai versi 3.8.3. Weka ditulis dalam bahasa pemrograman Java, dan didukung oleh GUI yang baik serta user friendly dapat mengolah berbagai file data dan memiliki fitur utama seperti data preprocessing tools, learning algorithms dan berbagai metode evaluasi [8].



Gambar 2. Software Weka

Weka terdiri dari beberapa tools yang dapat digunakan untuk menjalankan tugas preprocessing data, classification, regression, clustering, association rules, dan visualisasi, proses klustering digunakan untuk melakukan identifikasi pengelompokan dari beberapa kejadian dalam dataset agar dapat menghasilkan

informasi yang dapat dianalisis oleh penggunanya.

Dalam proses pengelompokan di Weka, beberapa atribut juga dapat diabaikan dengan tujuan hanya menggunakan data yang memberikan hasil spesifik saja, ada 6 atribut yang digunakan terdiri dari Program Studi, Fakultas, Angkatan, Tanggal Wisuda, IPK, dan masa studi yang akan diujikan dengan menggunakan software Weka.

IV. HASIL DAN PEMBAHASAN

Proses Data Mining

I. Data Selection

Pada proses seleksi ini, dipilih dari data Universitas Pendidikan Nasional yaitu data alumni Fakultas Ekonomi dan Bisnis angkatan 2013-2017. Dari hasil seleksi akan digunakan sebanyak 227 data alumni untuk perhitungan pada proses data mining.

II. Preprocessing

a. Data Konversion

KONVERSI NILAI (NILAI PADA LCK MENGACU PERMENDIKBUD 81A)		
INTERVAL SKOR	PREDIKAT	NILAI KOMPETENSI
$3.66 < x \leq 4.00$	A	4.00
$3.33 < x \leq 3.66$	A-	3.66
$3.00 < x \leq 3.33$	B+	3.33
$2.66 < x \leq 3.00$	B	3.00
$2.33 < x \leq 2.66$	B-	2.66
$2.00 < x \leq 2.33$	C+	2.33
$1.66 < x \leq 2.00$	C	2.00
$1.33 < x \leq 1.66$	C-	1.66
$1.00 < x \leq 1.33$	D+	1.33
$= 1.00$	D	1.00

Tabel 1. Tabel Interval Konversi Nilai

Proses dilakukan dengan mengkonversi data angka IPK ke dalam predikat kelulusan mahasiswa alumni undiknas pada setiap prodi, seperti pada tabel 1. Konversi juga dilakukan untuk mengubah data angkatan dan tahun lulus untuk mengetahui apakah mahasiswa tersebut lulus tepat waktu atau tidak seperti tabel 2. Data mahasiswa sebanyak 226 data training tersebut sebelum dihitung menggunakan software weka akan dihilangkan data yang tidak perlu, dan mengubah data agar memberikan informasi yang lebih jelas. Data yang dihilangkan adalah data id, nim, dan nama, serta mengubah IPK menjadi nilai sesuai tabel 2 dan mengubah informasi angkatan dan

date_wisuda menjadi lulus tepat waktu atau tidak.

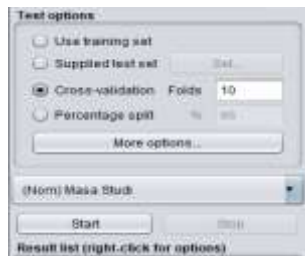
id	nim	nama	ipk	nama_prodi	nama_fakultas	angkatan	date_wisuda
6228	1.13.2.9528	Putu Jaka Widiartha	3.5	Akuntansi	Ekonomi dan Bisnis	2013	2017-02-17
6336	1.13.2.9637	Ni Made Dewi Parantika	3.4	Akuntansi	Ekonomi dan Bisnis	2013	2017-02-17
6199	1.13.2.9499	Ni Putu Naritha Kusumasari Putri	3.7	Akuntansi	Ekonomi dan Bisnis	2013	2017-02-17
6243	1.13.2.9543	Ni Made Winda Kusumadewi	3.6	Akuntansi	Ekonomi dan Bisnis	2013	2017-02-17

Tabel 2. Data sebelum dan setelah Preprocessing

Konversi data alumni yang dilakukan setiap prodi untuk menentukan status predikat kelulusan prodi alumni dengan cara mengkategorikan nilai ipk setiap prodi alumni yang akan dibagi berdasarkan data alumni yang lulus dari tahun 2013-2017.

b. Naïve Bayes

Simulasi perhitungan ini diambil dengan menggunakan 226 data training yang tersedia dan mengujinya dengan software weka.



Gambar 3. Proses penghitungan Naïve Bayes dengan Cross-Validation

Menggunakan software weka secara langsung didapatkan hasil berdasarkan test option yang telah dipilih yaitu cross-validation fold 10 dengan hasil log yang berada pada gambar 4 dibawah.

==== Run information ====
 Scheme: weka.classifiers.bayes.NaiveBayes

Relation: tes
 Instances: 226
 Attributes: 6
 nama_prodi
 nama_fakultas
 angkatan
 date_wisuda
 IPK
 Masa Studi

Test mode: 10-fold cross-validation
 Classifier model (full training set) ==
 Naive Bayes Classifier

Attribute	Class	Tepat Waktu	Tidak
		(0.98)	(0.02)
=====			
nama_prodi			
Akuntansi		106.0	1.0
Manajemen		100.0	1.0
Teknik Elektro		1.0	5.0
Ilmu Hukum		11.0	1.0
Ilmu Administrasi Negara			4.0
1.0			
Ilmu Komunikasi		6.0	1.0
[total]			228.0
10.0			
nama_fakultas			
Ekonomi dan Bisnis			205.0
1.0			
Teknik dan Informatika			1.0
5.0			
Hukum dan Ilmu Sosial			19.0
1.0			
[total]			225.0
7.0			
angkatan			
mean	2013		2012
std. dev.			0.1667
0.1667			
weight sum	222		4
precision	1		1
date_wisuda			
2017-02-17		222.0	5.0
2017-01-16		2.0	1.0
[total]		224.0	6.0

IPK		
A-	173.0	1.0
A	39.0	1.0
B+	13.0	3.0
B	1.0	3.0
[total]	226.0	8.0

Time taken to build model: 0.01 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	226
100 %	
Incorrectly Classified Instances	0
0 %	
Kappa statistic	1
Mean absolute error	0
Root mean squared error	0.0001
Relative absolute error	0.0359 %
Root relative squared error	0.0654 %
Total Number of Instances	226

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall
F-Measure	MCC	ROC	Area	PRC
Class				Area
1.000	1.000	0.000	1.000	1.000
1.000	1.000	1.000	1.000	Tepat Waktu
1.000	1.000	0.000	1.000	1.000
1.000	1.000	1.000	1.000	Tidak Tepat Waktu
Weighted Avg.	1.000	0.000	1.000	1.000
1.000	1.000	1.000	1.000	

==== Confusion Matrix ====

a	b	<-- classified as
222	0	a = Tepat Waktu
0	4	b = Tidak Tepat Waktu

Dari Hasil diatas dapat disimpulkan banyaknya persentase mahasiswa yang **lulus tepat waktu** adalah 98% dan yang **tidak lulus tepat waktu** adalah 2% dengan menggunakan algoritma Naïve Bayes. Kemudian kesimpulan dengan confusion matrix [9] maka didapat *True Positive*(TP) = 222, *True Negative*(TN) = 0, *False Positive*(FP) = 0 dan *False Negative*(FN) = 4. Dengan hasil ini dapat direkomendasikan kepada Universitas

Pendidikan Nasional khususnya Fakultas Ekonomi dan Bisnis untuk tetap menjaga kualitas Program Studi tersebut dan terus meningkatkan program yang sudah tertuang dalam Visi Misi Universitas.

V. PENUTUP

5.1 Kesimpulan

Masa studi siswa dapat dilihat dengan nilai IPK yang akan diprediksi apakah mahasiswa tersebut akan lulus tepat waktu atau tidak. Dengan menggunakan algoritma Naïve Bayes dan menggunakan 226 data training mahasiswa Universitas Pendidikan Nasional Fakultas Ekonomi dan Bisnis didapatkan kesimpulan bahwa mahasiswa yang lulus tepat waktu adalah **98%** dan yang tidak lulus tepat waktu adalah **2%** sehingga arah kebijakan Visi dan Misi yang dilakukan sudah tepat hanya perlu sedikit peningkatan untuk hasil yang lebih sempurna.

5.2 Saran

Perlunya penambahan fitur lain seperti kegiatan ekstrakurikuler atau organisasi mahasiswa untuk dapat menjadi salah satu faktor apakah kelulusan masa studi tepat waktu atau tidak. Dapat juga dilakukan algoritma lain untuk menentukan apakah keakuratan algoritma ini pada kasus mahasiswa Universitas Pendidikan Nasional fakultas Ekonomi dan Bisnis sudah tepat.

DAFTAR PUSTAKA

- E. Setiawan, "Klasifikasi," 2018. [Online]. Available:<https://kbbi.web.id/klasifikasi>. [Accessed: 31-Oct-2018].
- I. G. A. Socrates, A. L. Akbar, and M. S. Akbar, "Optimasi Naïve Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio," vol. 7, no. 1, pp. 22–30, 2016.
- M. A. SHADIQ, "KEOPTIMALAN NAÏVE BAYES DALAM KLASIFIKASI," Universitas Pendidikan Indonesia, 2009.
- Larose, D. (2005). *Discovering Knowledge In Data*. Canada: Willey Interscience. metodealgoritma.com. (2013, Maret 27).

- Hoffer, Ramesh and Topi. (2012). Modern Database Management. Harlow United Kingdom : Pearson Education Limited.
- MacLennan, J., Zhao Hui Tang, Bog, Crivat, "Data Mining with Microsoft® SQL Server® 2008", Wiley Publishing, Inc., Indianapolis, Indiana. 2009.
- Prasetyo, Eko. 2012. Data Mining Konsep dan Aplikasi Menggunakan MATLAB. Yogyakarta: Andi
- F. Eibe, "Machine Learning with WEKA," Departement of Computer Science, University of Waikato, New Zeland. 2011.
- A. Indriani and D. Nbc, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier," pp. 5–10, 2014