

Metode Naïve Bayes Classifier (NBC) untuk Memprediksi Tingkat Kelulusan Mahasiswa Tepat Waktu

Muhamad Kadafi¹

¹Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Raden Fatah Palembang
e-mail: kadafi_uin@radenfatah.ac.id

Abstrak

Untuk membantu menjaga kualitas penyelenggaraan program studi dalam hal ini adalah layanan akademik, maka penyeimbangan ratio antara penerimaan mahasiswa yang masuk dengan sarana dan prasana merupakan salah satu upaya yang perlu dilakukan oleh perguruan tinggi caranya adalah dengan memprediksi jumlah tingkat kelulusan mahasiswa yang tepat waktu, dengan mengetahui hal tersebut maka pimpinan dapat menetapkan ratio penerimaan mahasiswa baru pada tahun berikutnya berdasarkan data prediksi yang di dapat. Penelitian ini bertujuan untuk menganalisis data akademik yaitu dengan melakukan klasifikasi data kelulusan mahasiswa sebelumnya dan menggunakan pola yang terbentuk untuk memprediksi besaran persentase tingkat kelulusan mahasiswa serta mencari parameter apa saja yang mempengaruhi tingkat kelulusan. Metode yang digunakan adalah teknik klasifikasi data mining dengan Naïve Bayes Classifier (NBC). Hasil penelitian ini adalah jumlah rata – rata persentase keseluruhan kelulusan mahasiswa tepat waktu yaitu sebesar 73.44% kelulusan dan parameter yang mempengaruhi kelulusan mahasiswa tepat waktu adalah IPK, IPS4, IPS3, IPS2. Hasil penelitian dapat dijadikan landasan bagi pimpinan untuk menentukan ratio penerimaan mahasiswa baru pada tahun berikutnya, dan sebagai bahan evaluasi dalam membantu mahasiswa untuk lulus tepat waktu.

Kata kunci: Naïve Bayes Classifier (NBC); Data Mining.

Abstract

To help maintain the quality of the implementation of the study program in this case is academic services, then balancing the ratio between admission of students with facilities and infrastructures is one of the efforts that need to be done by university by means of predicting the number of student graduation rates on time, with knowing this, the leadership can determine the ratio of new student admissions in the following year based on the prediction data obtained. This study aims to analyze academic data by classifying the graduation data of previous students and using patterns that are formed to predict the percentage of student graduation rates and to find out what parameters affect graduation rates. The method used is the classification of data mining techniques with Naïve Bayes Classifier (NBC). The results of this study are the average number of percentages of overall students graduation on time that is equal to 73.44% graduation and the parameters that affect students' graduation on time are IPK, IPS4, IPS3, IPS2. The results of the study can be used as a basis for leaders to determine the ratio of new student admissions in the following year, and as an evaluation material in helping students to graduate on time.

Keywords: Naïve Bayes Classifier (NBC); Data Mining.

1. PENDAHULUAN

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Sebagai penyelenggara pendidikan, perguruan tinggi diharapkan menghasilkan sumber daya manusia yang berkualitas, berilmu, cakap dan kreatif [1].

Kualitas tersebut diukur berdasarkan 7 standar utama, salah satu pointnya adalah Mahasiswa dan Lulusan. Mahasiswa dan lulusan merupakan salah satu point 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 penggunaan kelulusannya [2].

Untuk membantu evaluasi penyelenggaraan Program Studi, Perguruan tinggi dapat memanfaatkan gudang data yang tersedia untuk melakukan analisa dalam mengambil keputusan [3]. Gudang data dalam hal ini adalah data akademik. Data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan. Serangkaian proses untuk mendapatkan pola atau pengetahuan tersebut disebut dengan *datamining* [4]. Klasifikasi merupakan salah satu masalah mendasar dalam data mining [5]. Dalam klasifikasi sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan sebelumnya performa pengklasifikasi biasanya di ukur dengan ketepatan atau tingkat galat [6].

Pengklasifikasian *Bayesian* merupakan pengklasifikasi statistik untuk dapat memprediksi probabilitas keanggotaan kelas tertentu [4], menghitung peluang untuk suatu hipotesis, menghitung peluang dari suatu kelas dari masing – masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal [7]. Pengklasifikasian Bayesian telah banyak digunakan penelitian – penelitian untuk memprediksi suatu kasus tertentu. Seperti pada : metode naïve bayes dalam memprediksi kelulusan mahasiswa [8]. Analisis komparasi algoritma klasifikasi data mining salah satunya adalah naïve bayes untuk prediksi mahasiswa non aktif [1]. Algoritma *naïve bayes classifier* juga digunakan untuk evaluasi kinerja *akademik* [2].

Secara kinerja metode naïve bayes classifier memiliki nilai akurasi dan AUC paling tinggi diantara metode klasifikasi yang di komparasikan, yaitu C45, neural network, dan logistic regression [9]. Dan juga dalam perbandingan kinerja algoritma klasifikasi naïve Bayesian, lazy-ibk, Zero-R, dan Decision Tree – J48. Disimpulkan bahwa naïve Bayesian memiliki kinerja yang paling baik dalam hal akurasi [10].

Peneliti melihat bahwa Fakultas Sains dan Teknologi Univesitas Islam Negeri Raden Fatah Palembang saat ini juga terus berupaya untuk meningkatkan kualitas layanan akademik dari masing masing prodi yang berada di bawahnya. Tetapi sebagai Fakultas Baru permasalahan yang dihadapi adalah masih terbatasnya jumlah sarana dan prasarana yang tersedia seperti jumlah kelas, jumlah lab, dan jumlah tenaga pendidik. Oleh karena itu untuk menjaga kualitas penyelenggaraan akademik perlu adanya penyeimbangan ratio antara jumlah penerimaan mahasiswa yang masuk dengan sarana dan prasarana yang tersedia. Salah satu caranya adalah dengan menggunakan teknik klasifikasi pada data mining yaitu memprediksi jumlah tingkat kelulusan mahasiswa tepat waktu, dengan mengetahui perkiraan jumlah mahasiswa yg lulus tepat waktu, maka pimpinan (dekanat dan ketua program studi) dapat menetapkan ratio penerimaan jumlah mahasiswa baru pada tahun berikutnya berdasarkan data prediksi.

Penelitian ini bertujuan untuk menganalisis data akademik dengan melakukan pengklasifikasian data kelulusan mahasiswa sebelumnya dan menggunakan pola yang terbentuk untuk memprediksi besaran persentase tingkat kelulusan mahasiswa serta mencari parameter apa saja yang mempengaruhi tingkat kelulusan. Serta metode klasifikasi menggunakan naïve bayes classfier (NBC). Hasil dari penelitian ini adalah jumlah rata – rata persentase keseluruhan kelulusan mahasiswa tepat waktu sebesar 73.44% kelulusan dan parameter yang mempengaruhi kelulusan mahasiswa tepat waktu adalah IPK, IPS4, IPS3, IPS2. Hasil penelitian dapat dijadikan landasan bagi pimpinan untuk menentukan ratio penerimaan mahasiswa baru pada tahun berikutnya, dan sebagai bahan evaluasi dalam membantu mahasiswa untuk lulus tepat waktu.

2. METODE PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif data digunakan untuk menghitung dan mengukur pada saat analisis dan interpretasi data. dan hasilnya berupa data prediksi [11].

A. Identifikasi Masalah

Permasalahan pada penelitian ini adalah bagaimana Memprediksi jumlah tingkat kelulusan mahasiswa tepat waktu pada Fakultas Sains dan Teknologi UIN Raden Fatah Palembang dengan menggunakan metode Naïve Bayes Classifier.

B. Review Literatur

Literatur – literatur yang di pakai sebagai bahan refensi dalam penelitian ini adalah dari jurnal – jurnal ilmiah, modul pembelajaran, dan buku tentang data mining. literatur – literatur ini akan menjadi pedoman untuk melakukan penelitian agar memudahkan proses penelitian.

C. Menetapkan Tujuan Penelitian

Tujuan pada penelitian ini adalah memprediksi jumlah tingkat kelulusan mahasiswa tepat waktu pada Fakultas Sains dan Teknologi UIN Raden Fatah Palembang.

D. Mengumpulkan Data

Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung di Fakultas Sains dan Teknologi UIN Raden Fatah Palembang. Teknik yang digunakan adalah teknik analisis dokumen karena data yang menjadi sumber utama pada penelitian ini di ambil dari PUSTIPD (Pusat Teknologi Informasi dan Pangkalan Data) UIN Raden Fatah Palembang. Data pada penelitian ini adalah data Nilai KHS Mahasiswa dalam hal ini indeks prestasi semester dan indeks prestasi kumulatif, juga data biodata mahasiswa, dan data Yudisium Mahasiswa Yudisium ke 1 – Yudisium ke 10.

Data yang akan digunakan pada penelitian ini dikelompokan menjadi dari 3 data yaitu :

1) Data Latih

Data training adalah data yang pada penelitian ini digunakan untuk proses pembuatan pemodelan prediksi kelulusan mahasiswa di Fakultas Sains dan Teknologi. Data Training pada penelitian ini di ambil dari 2 sumber yaitu:

- a) Data yudisium mahasiswa ke 1 – 9 Fakultas Sains dan Teknologi. Data yudisium mahasiswa terdiri dari 563 dataset. Dari data ini penulis menggunakan 5 atribut yaitu: NIM, Nama Mahasiswa, Tempat Tanggal Lahir, Masa Studi.
- b) Data Nilai Mahasiswa persemester (Ganjil dan Genap) dari tahun 2013 – 2018. Dari data ini penulis menggunakan 6 atribut : Jenis Kelamin, IPS1, IPS2, IPS3, IPS4, IPK.

2) Data Uji

Data Testing adalah data yang digunakan untuk melakukan uji pada pemodelan data yang terbentuk pada proses pemodelan. Data testing di dapat dari pembagian data training yang di bagi menjadi 2 bagian : 1 bagian untuk data latih, dan sebagian

lagi untuk data uji. Teknik pembagian menggunakan metode k-fold cross – validation.

3) Data Prediksi

Data prediksi adalah data yang digunakan untuk melakukan prediksi tingkat kelulusan mahasiswa pada Fakultas Sains dan Teknologi. Mahasiswa yang di prediksi tingkat kelulusannya adalah angkatan 2014,2015,2016.

E. Menganalisis dan Interpretasi Data

Dalam pengolahan *Data Mining* penulis menggunakan langkah sebagai berikut yaitu :

- 1) Melakukan uji kedekatan objek pada setiap atribut yang ada pada data training dengan atribut tepat waktu. Yang bertujuan untuk mengetahui atribut mana saja yang paling mempengaruhi kelulusan mahasiswa tepat waktu.
- 2) Membuat model klasifikasi tingkat kelulusan tepat waktu dengan Naïve Bayes Clasfier (NBC), yang bertujuan untuk mencari model yang tepat dengan melihat nilai akurasi tertinggi.
- 3) Pattern Evaluation yang bertujuan untuk menseleksi lebih rinci model yang dapat di gunakan untuk prediksi.
- 4) Knowledge Presentation yang bertujuan untuk mengimplementasikan model untuk memprediksi jumlah tingkat kelulusan mahasiswa.

F. Melaporkan dan Evaluasi Hasil Penelitian

Membuat kesimpulan dari hasil penelitian dan memberikan saran untuk pihak Fakultas Sains dan Teknologi sebagai bahan evaluasi akademik untuk dapat menjadi lebih baik lagi.

3. HASIL DAN PEMBAHASAN

Dalam melakukan prediksi tingkat kelulusan mahasiswa adapun tahapan yang digunakan penulis dalam pengolahan data mining pada penelitian ini yaitu :

3.1. Melakukan uji Kedekatan objek pada setiap atribut yang ada pada data training dengan atribut tepat waktu.

Korelasi parameter mempengaruhi hasil akurasi Naïve Bayes Clasfier, oleh karena itu sebelum melakukan klasifikasi perlu dilakukan perhitungan uji korelasi antar parameter terlebih dahulu [12]. Pada proses ini peneliti menggunakan ukuran kemiripan (similarity) dan metode yang digunakan adalah Rank Spearman Correlation. Alat bantu yang digunakan adalah software SPSS. Berikut adalah hasil uji.

			Correlations								
			JK	IPS1	IPS2	IPS3	IPS4	IPK	Umur	KotaKelahiran	TepatWaktu
Spearman's rho	JK	Correlation Coefficient	1.000	-.112*	-.146**	-.127**	-.172**	-.153**	-.020	-.030	-.178**
		Sig. (2-tailed)	.	.010	.001	.003	.000	.000	.639	.491	.000
		N	530	530	530	530	530	530	530	530	530
	IPS1	Correlation Coefficient	-.112*	1.000	.348**	.517**	.571**	.600**	.052	.060	.406**
		Sig. (2-tailed)	.010	.	.000	.000	.000	.000	.228	.168	.000
		N	530	530	530	530	530	530	530	530	530
	IPS2	Correlation Coefficient	-.146**	.348**	1.000	.451**	.489**	.641**	.064	.005	.517**
		Sig. (2-tailed)	.001	.000	.	.000	.000	.000	.142	.910	.000
		N	530	530	530	530	530	530	530	530	530
	IPS3	Correlation Coefficient	-.127**	.517**	.451**	1.000	.535**	.651**	.039	.036	.492**
		Sig. (2-tailed)	.003	.000	.000	.	.000	.000	.370	.414	.000
		N	530	530	530	530	530	530	530	530	530
	IPS4	Correlation Coefficient	-.172**	.571**	.489**	.535**	1.000	.772**	.012	-.014	.670**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.791	.753	.000
		N	530	530	530	530	530	530	530	530	530
	IPK	Correlation Coefficient	-.153**	.600**	.641**	.651**	.772**	1.000	.061	-.006	.777**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.159	.890	.000
		N	530	530	530	530	530	530	530	530	530
	Umur	Correlation Coefficient	-.020	.052	.064	.039	.012	.061	1.000	.044	.093*
		Sig. (2-tailed)	.639	.228	.142	.370	.791	.159	.	.307	.032
		N	530	530	530	530	530	530	530	530	530
	KotaKelahiran	Correlation Coefficient	-.030	.060	.005	.036	-.014	-.006	.044	1.000	.028
		Sig. (2-tailed)	.491	.168	.910	.414	.753	.890	.307	.	.515
		N	530	530	530	530	530	530	530	530	530
	TepatWaktu	Correlation Coefficient	-.178**	.406**	.517**	.492**	.670**	.777**	.093*	.028	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.032	.515	.
		N	530	530	530	530	530	530	530	530	530

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

Gambar 1. matrik korelasi antar atribut dengan rank spearman

Pada Gambar 1 dapat dilihat nilai korelasi parameter JK, IPS1, IPS2, IPS3, IPS4, IPK, Umur, Kota Kelahiran terhadap parameter tepat waktu adalah sebagai berikut :

parameter JK = - 0.178, IPS1 = 0.406, IPS2 = 0.517, IPS3 = 0.492, IPS4 = 0.670, IPK = 0,777, umur = 0.093, kota kelahiran = 0.028. Parameter yang mempunyai nilai korelasi signifikan pada level 0.01 (ditandai dengan ‘*’) adalah IPK, IPS4, IPS3, IPS2, IPS1, dan JK. Karena nilai korelasi yang di dihasilkan > 0.01, sedangkan parameter Umur bernilai signifikan pada level 0.05 (ditandai dengan ‘**’) karena bernilai > 0.05 dan < 0.01. dan Kota Kelahiran tidak signifikan karena bernilai kurang dari 0.05.

berdasarkan hasil uji korelasi di atas dapat di simpulkan bahwa atribut yang paling berpengaruh positif dalam kelulusan mahasiswa adalah IPK, IPS4, IPS2. Artinya semakin tinggi nilai atribut tersebut maka akan semakin cepat mahasiswa untuk lulus tepat waktu.

selanjutnya penulis menggunakan semua parameter yang telah di uji selama parameter tersebut masih memiliki hubungan dengan parameter tepat waktu dan melihat semua kemungkinan besaran akurasi yang dapat terjadi pada saat pembentukan model klasifikasi.

3.2. Membuat model klasifikasi tingkat kelulusan dengan tepat waktu dengan Naive Bayes Classifier (NBC).

Pada Langkah ini prosedur penentuan model menggunakan Backward Feature Selection yaitu dengan menghilangkan satu persatu fitur, dimulai dari fitur yang memiliki nilai korelasi yang terkecil. Dan metode evaluasi menggunakan k-fold cross validation. Pada proses ini data training dibagi terlebih dahulu menjadi partisi seperti pada Tabel 1.

Tabel 1. SIMULASI PENGUJIAN DATA MENGGUNAKAN K=5

Partisi ke - n Fold ke - n	Partisi 1 Baris 1 – 106	Partisi 2 Baris 107 - 212	Partisi 3 Baris 213 - 318	Partisi 4 Baris 319 - 424	Partisi 5 Baris 425 - 530
Fold 1	Data Testing				
Fold 2		Data Testing			
Fold 3			Data Testing		
Fold 4				Data Testing	
Fold 5					Data Testing

Pada Tabel 1 data terdiri dari 530 recordset dibagi menjadi 5 partisi sama besar yaitu 106 recordset yang nanti akan di jadikan sebagai data testing (data uji). Dan sisanya yaitu data yang lain akan menjadi data training (data latihan) .

Selanjutnya adalah prosedur penentuan model dengan Backward Feature Selection. berikut simulasi pemodelan yang terbentuk.

Tabel 2. Simulasi Pemodelan Dengan Backward Feature Selection

Model	Parameter yang digunakan
1	IPK,IPS4,IPS3,IPS2,IPS1,KotaKelahiran,Umur,JK
2	IPK,IPS4,IPS3,IPS2,IPS1,KotaKelahiran,Umur
3	IPK,IPS4,IPS3,IPS2,IPS1,KotaKelahiran,JK
4	IPK,IPS4,IPS3,IPS2,IPS1,Umur,JK
5	IPK,IPS4,IPS3,IPS2,KotaKelahiran,Umur,JK
6	IPK,IPS4,IPS3,IPS1,KotaKelahiran,Umur,JK
7	IPK,IPS4,IPS2,IPS1,KotaKelahiran,Umur,JK
8	IPK,IPS3,IPS2,IPS1,KotaKelahiran,Umur,JK
9	IPS4,IPS3,IPS2,IPS1,KotaKelahiran,Umur,JK
10	IPK,IPS4,IPS3,IPS2,IPS1,KotaKelahiran
11	IPK,IPS4,IPS3,IPS2,IPS1,Umur
12	IPK,IPS4,IPS3,IPS2,KotaKelahiran,Umur
13	IPK,IPS4,IPS3,IPS1,KotaKelahiran,Umur
14	IPK,IPS4,IPS2,IPS1,KotaKelahiran,Umur
15	IPK,IPS3,IPS2,IPS1,KotaKelahiran,Umur
16	IPS4,IPS3,IPS2,IPS1,KotaKelahiran,Umur
17	IPK,IPS4,IPS3,IPS2,IPS1
18	IPK,IPS4,IPS3,IPS2,KotaKelahiran
19	IPK,IPS4,IPS3,IPS1,KotaKelahiran
20	IPK,IPS4,IPS2,IPS1,KotaKelahiran
21	IPK,IPS3,IPS2,IPS1,KotaKelahiran

22	IPS4,IPS3,IPS2,IPS1,KotaKelahiran
23	IPK,IPS4,IPS3,IPS2
24	IPK,IPS4,IPS3,IPS1
25	IPK,IPS4,IPS2,IPS1
26	IPK,IPS3,IPS2,IPS1
27	IPS4,IPS3,IPS2,IPS1
28	IPK,IPS4,IPS3
29	IPK,IPS4,IPS2
30	IPK,IPS3,IPS2
31	IPS4,IPS3,IPS2
32	IPK,IPS4
33	IPS4,IPS3
34	IPK
35	IPS4

Pada Tabel 2 Proses pemodelan dimulai dengan fitur lengkap, dan iterasi berikutnya dengan menghilangkan satu persatu fitur. Terdapat 8 kali iterasi pada pemodelan di atas, iterasi pertama menggunakan 8 fitur, iterasi ke-2 menggunakan 7 fitur, dan seterusnya hingga tinggal 1 fitur yang tersisa.

Pada tahapan ini penulis mengkombinasikan 2 proses klasifikasi pada tabel II, dan tabel III, untuk mendapatkan nilai akurasi terbaik. Berikut adalah hasil yang di dapat :

- 1) Pada fold ke – 1 nilai akurasi tertinggi terdapat pada model ke 1, 5, 7, 12, 14 yaitu sebesar 95,28%, dan model 34 sebesar 96,22%
- 2) Pada fold ke – 2 nilai akurasi tertinggi terdapat pada model ke 5, 7, 11, 12, 14, 15 sebesar 89,62%, model ke 29 sebesar 90,56%, dan model ke 34 sebesar 92,45%
- 3) Pada fold ke – 3 nilai akurasi tertinggi terdapat pada model ke 3, 12, 13, 19, 20, 24, 25, 28 sebesar 87,73%, dan model ke 17 sebesar 88,67%.
- 4) Pada fold ke – 4 nilai akurasi tertinggi terdapat pada model ke 1, 3, 4, 7, 10, 12, 15, 17, 18, 20, 23, 25, 28, 29, 30, 31, 34 sebesar 89,62 %, dan model ke 35 sebesar 91,5%.
- 5) Pada fold ke – 5 nilai akurasi tertinggi terdapat pada model ke 5, 7, 9, 10, 12, 14, 17, 18, 20, 23, 25 sebesar 96,22%

Berdasarkan pengujian model di atas dapat dilihat bahwa akurasi tertinggi pada setiap fold terdapat pada fold ke 5 yaitu sebesar 96,22%.

3.3. Evaluasi Model

Pada proses data mining sebelumnya telah di dapat bahwa nilai akurasi yang paling besar berada pada fold ke-5. pada proses ini penulis akan mengujikan kembali model tersebut dengan data uji yang berbeda yaitu peserta yudisium ke-10. Peserta yudisium 10 berjumlah 20 orang peserta terdiri dari mahasiswa angkatan 2 orang angkatan 2012, 3 orang angkatan 2013, 15 orang angkatan 2014. dan melihat apakah ada perbedaan akurasi pada keduanya. Berikut adalah hasilnya.

Tabel 3. Pengujian Dengan Data Yudisum Ke-10

Model	Akurasi fold ke - 5	Akurasi dengan data yudisium ke - 10
5	96,22%	65%
7	96,22%	60%
9	96,22%	55%
10	96,22%	60%
12	96,22%	60%
14	96,22%	60%
17	96,22%	60%
18	96,22%	65%
20	96,22%	60%
23	96,22%	65%
25	96,22%	60%

Pada tabel 3 Terlihat bahwa model yang diujikan terjadi penurunan akurasi, hal ini disebabkan karena terdapat beberapa recordset yang ada pada data yudisium ke - 10 memiliki nilai atribut yang bertentangan dengan analisis uji korelasi yang sudah dilakukan pada data training sebelumnya (Gambar 1). Misalnya jika nilai IPK, IPS4, IPS3, IPS2 berpengaruh positif terhadap kelulusan mahasiswa, maka sebaliknya record tersebut berpengaruh negatif.

3.4. Representasi Knowledge

Selanjutnya Pengujian dengan data yudisium 10 digunakan sebagai dasar untuk menseleksi lebih rinci lagi pemodelan yang telah di dapat pada fold ke – 5. Berdasarkan tabel 3 nilai akurasi terbaik pada pengujian data yudisium ke - 10 di dapat pada model ke 5, 18, dan 23. Selanjutnya model tersebut dijadikan acuan untuk memprediksi tingkat kelulusan mahasiswa di Fakultas Sains dan Teknologi UIN Raden Fatah. Data yang digunakan untuk prediksi adalah data mahasiswa angkatan 2014, 2015, dan 2016 yang aktif pada semester 1, 2, 3, dan 4. Berikut ini adalah hasil prediksi dengan menggunakan pemodelan 5, 18, 23 :

1) Program Studi Sistem Informasi

Jumlah total mahasiswa yang di prediksi untuk program studi Sistem Informasi berjumlah 340 orang mahasiswa terdiri dari 92 mahasiswa angkatan 2014, 139 mahasiswa angkatan 2015, dan 109 mahasiswa angkatan 2016. Jumlah mahasiswa yang lulus tepat waktu angkatan 2014 adalah 55 orang mahasiswa atau 16.18% dari jumlah total mahasiswa yang di prediksi, untuk angkatan 2015 jumlah mahasiswa yang lulus tepat waktu ada 109 orang mahasiswa atau 32.06%, sedangkan angkatan 2016 mahasiswa yang lulus tepat waktu sebanyak 82 mahasiswa atau 24.12%. Total keseluruhan mahasiswa yang lulus tepat waktu pada program studi Sistem informasi adalah 246 mahasiswa atau sekitar **72.35%** kelulusan untuk angkatan 2014, 2015, 2016.

2) Program Studi Biologi

Jumlah total mahasiswa yang di prediksi untuk program studi Biologi berjumlah 65 orang mahasiswa terdiri dari 34 mahasiswa angkatan 2015, dan 31 mahasiswa

angkatan 2016. Jumlah mahasiswa yang lulus tepat waktu angkatan 2015 berjumlah 24 orang mahasiswa atau sebanyak 36.92% dari jumlah total mahasiswa yang di prediksi. Dan untuk angkatan 2016 jumlah mahasiswa yang lulus tepat waktu adalah sebanyak 24 orang mahasiswa atau sebanyak 36.92% mahasiswa. Total keseluruhan mahasiswa yang lulus tepat waktu pada prodi biologi sebanyak 48 orang mahasiswa atau **73.84%** kelulusan untuk angkatan 2015, dan 2016.

3) Program Studi Kimia

Jumlah total mahasiswa yang di prediksi pada Program Studi Kimia berjumlah 58 orang mahasiswa terdiri dari 31 orang mahasiswa angkatan 2015, dan 27 orang mahasiswa angkatan 2016. Jumlah mahasiswa yang lulus tepat waktu angkatan 2015 berjumlah 22 orang atau 37.93% dari jumlah total mahasiswa yang di prediksi. Sedangkan untuk angkatan 2016 berjumlah 21 orang mahasiswa atau 36.20% kelulusan. Jumlah total mahasiswa yang lulus tepat waktu pada program studi kimia berjumlah 43 orang atau **74.13%** kelulusan.

Berdasarkan data prediksi di atas rata – rata jumlah kelulusan mahasiswa angkatan 2014, 2015, dan 2016 pada Fakultas Sains dan Teknologi adalah **73.44%** kelulusan. Dan jika dilihat dari frekuensi jumlah kemunculan parameter yang digunakan yaitu model 5, model 18, dan model 23. Parameter yang mempengaruhi jumlah kelulusan adalah pada penelitian ini adalah **IPK, IPS4, IPS3, IPS2** dan sesuai dengan Analisis Uji Korelasi sebelumnya.

4. KESIMPULAN

Metode Naïve Bayes Classifier dengan penambahan penggunaan Backward Feature Selection dalam K-Fold Cross Validation merupakan Evaluasi pemodelan bertingkat pada setiap pengujian data uji serta menghasilkan nilai akurasi yang baik pada setiap model klasifikasi. Pemodelan dengan nilai akurasi tertinggi yaitu model 5, 8, 23 digunakan untuk memprediksi tingkat kelulusan mahasiswa dan menghasilkan rata rata total jumlah kelulusan sebesar 73.44%. nilai tersebut dapat digunakan oleh pimpinan sebagai landasan dalam penentuan ratio jumlah penerimaan mahasiswa pada tahun berikutnya. Parameter yang mempengaruhi tingkat kelulusan adalah IPK, IPS4, IPS3, IPS2, dapat di jadikan bahan evaluasi bagi pimpinan, ketua program studi, dan dosen dalam membatu mahasiswa untuk lulus tepat waktu.

DAFTAR PUSTAKA

- [1] Hastuti, K., 2012. Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012, Semarang, 23 Juni 2012, INFRM 241 – 249.
- [2] Ridwan, M., Suyono, H., Sarosa, M., 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier, Jurnal EECCIS Vol.7 No.1 Juni 2013, 59–64.
- [3] Kadafi, M., 2018. Penerapan Algoritma FP-GROWTH untuk Menemukan Pola Peminjaman Buku Perpustakaan UIN Raden Fatah Palembang, Jurnal MATICS Vol.10 No.2 (2018), 52 – 58.
- [4] Han, J., Kamber, M., Pei, J., 2012. Data Mining Concept and Techniques. Morgan Kaufmann Publishers is an imprint of Elsevier, 225 Wyman Street, Waltham, MA

02451, USA.

- [5] Zhang, Harry, 2004, The Optimality of Naïve Bayes, FLAIRS Conference.
- [6] Walpole, E. R., Myers, R. H., Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan, Edisi ke-4 Bandung, ITB.
- [7] Hadjaratie, L., 2011. Jaringan Saraf Tiruan Untuk Prediksi Tingkat Kelulusan Mahasiswa Diploma Program Studi Manajemen Informatika Universitas Negeri Gorontalo, Thesis, Postgraduated IPB, Bogor.
- [8] Syarli, Muin, A. A., 2016. Metode Naïve Bayes untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi), Jurnal Ilmiah Ilmu Komputer, Vol.2, No. 1, April 2016.
- [9] Saputra, R. A., Widodo, P. P., 2014. Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (TB), Seminar Nasional Inovasi dan Tren (SNIT) 2014, A120 – A126.
- [10] Fitri, S., 2014. Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian, Lazy-Ibk, Zero-R, dan Decision Tree-J48, Jurnal DASIS, Vol.15, No.1, Maret 2014, 33 – 37.
- [11] Ghozali, 2016. Desain Penelitian Kuantitatif Kualitatif untuk Akuntansi, Bisnis, dan Ilmu Sosial Lainnya. Yoga Pratama. Semarang 50143
- [12] Cahyanti, A. F., Saptono, R., Sihwi, S. W., 2015. Penentuan Model Terbaik pada Metode Naïve Bayes Classifier dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter, Jurnal ITSMART Vol. 4 No.1 Juni 2015, 28 - 35.