

IMPLEMENTASI DIRICHLET SMOOTHING PADA NAÏVE BAYES UNTUK KLASIFIKASI KINERJA AKADEMIK MAHASISWA UNIVERSITAS MADURA

Erwin Prasetyowati

Fakultas Teknik, Universitas Madura
email: erwinprasetyowati@unira.ac.id

Abstract

This study focused on evaluating the implementation of Dirichlet Smoothing on Naïve Bayes in predicting students' graduation at Madura University. Evaluation on the performance of students conducted in the second and fourth year so that the academics can know and take action in the form of warning or warning to students to improve learning outcomes so as to pass on time. The measurement indicators used are GPA, credits that have not been taken and the length of study. Evaluations performed in the second year using established standard values, while classification with the Naïve Bayes algorithm, are used in the evaluation of the speedy year. To improve the accuracy of classification results in Naïve Bayes, Dirichlet Smoothing is used. By comparing the results of Naïve Bayes classification performance with Dirichlet Smoothing and without Smoothing, it was found that the results of classification calculation for 5 times in 725 training data and 200 testing data taken at random, then obtained the difference of accuracy value reached 5.48%, the precision value of 7.1% and the recall value is 4%. In other words, the addition of smoothing methods can improve the performance of Naive Bayes.

Keywords: Predictions of graduation, Naïve Bayes, Dirichlet Smoothing, Accuracy

1. PENDAHULUAN

Kinerja akademik mahasiswa merupakan salah satu faktor yang harus diperhatikan oleh pihak akademisi dalam meningkatkan kualitas dan kuantitas mahasiswa. Tingkat presentase kelulusan mahasiswa tepat waktu menjadi bahan evaluasi yang harus dimonitoring secara berkelanjutan sehingga perbaikan pola pembelajaran di kelas dan di luar kelas dapat dilakukan secara kontinu. Proses pembelajaran yang berkualitas tentunya akan berpengaruh pada peningkatan kompetensi lulusan, membantu mahasiswa lulus dengan tepat waktu, serta membangun kepercayaan sebagai pengguna lulusan [1].

Pada penelitian ini dilakukan proses pengklasifikasian terhadap hasil belajar mahasiswa melalui perancangan suatu sistem, sehingga kontrol tidak dilakukan secara manual. Sistem kontrol tersebut mengevaluasi pencapaian hasil belajar mahasiswa dengan memperhatikan waktu studi, jumlah SKS yang telah ditempuh dan IPK. Pencapaian hasil belajar mahasiswa tersebut selanjutnya dijadikan bahan masukan untuk mengetahui apakah mahasiswa tersebut dapat menyelesaikan waktu studinya dengan tepat waktu atau tidak.

Penelitian yang berfokus pada pengklasifikasian mahasiswa telah dilakukan oleh beberapa peneliti. Metode yang digunakan juga beragam, diantaranya

menggunakan Algoritma Naïve Bayes [1][2][3], yang digabung dengan C4.5 [4], K-Means Clustering [5], *decision tree classifier*, *neural network*, *nearest neighbour classifier* [6].

Berbeda dengan penelitian-penelitian sebelumnya, tahapan evaluasi pada hasil pencapaian belajar mahasiswa dilakukan dua kali yaitu pada masa studi 2 tahun dan masa studi empat tahun. Hasil klasifikasi tahun kedua menjadi dasar untuk pengklasifikasian pada tahun keempat. Metode yang digunakan dalam proses klasifikasi tahun keempat adalah algoritma Naïve Bayes yang kemudian dilanjutkan dengan penambahan teknik smoothing guna meningkatkan performa Naïve Bayes. Teknik smoothing yang digunakan adalah Dirichlet Smoothing. Pada dasarnya algoritma *Naïve Bayes* sangat tepat untuk menangani masalah yang sifatnya hipotesis melalui perancangan desain suatu klasifikasi dengan memisahkan beberapa obyek berdasarkan Teorema Bayes [7].

Berdasarkan penjabaran sebelumnya, maka penelitian ini difokuskan pada analisis terhadap penggunaan teknik smoothing pada Naïve Bayes. Melalui pengolahan data yang dilakukan akan dapat dievaluasi tingkat keakurasian klasifikasi yang dihasilkan. Dengan demikian maka akan dilakuakn perbandingan hasil klasifikasi menggunakan

Nive Bayes dengan Dirichlet Smoothing dan tanpa smoothing.

Adapun beberapa tujuan dari penelitian ini meliputi :

- Mengklasifikasikan kinerja akademik mahasiswa dalam dua kelompok yaitu Lulus dan Drop Out menggunakan Naive Bayes dan Dirichlet Smoothing
- Membandingkan dan mengevaluasi keakuratan klasifikasi melalui penerapan teknik smoothing khususnya Dirichlet Smoothing pada Naive Bayes
- Sebagai media evaluasi dan prediksi ketepatan mahasiswa menyelesaikan waktu studinya.

2. METODE PENELITIAN

Penelitian ini tentunya memiliki beberapa tahapan atau proses guna menghasilkan output yang direncanakan. Adapun beberapa proses yang harus dilakukan adalah berikut ini.

Data Input

Jumlah mahasiswa pada saat ini di Universitas Madura adalah 4719 yang tersebar pada sepuluh program studi. Jumlah tersebut tidak semua akan dimasukkan pada proses pengklasifikasian kinerja mahasiswa, karena hanya sebanyak mahasiswa yang telah dinyatakan lulus dari Universitas Madura yang dapat digunakan sebagai data uji (testing data) dan data latih (training data). Jumlah mahasiswa yang lulus tersebut adalah 925 mahasiswa dan merupakan mahasiswa angkatan 2006 - 2008. Atribut yang digunakan adalah NIM, nama mahasiswa, jurusan, nilai KHS, dan nilai pendukung lainnya yang didapat dari SIMAT Universitas Madura.

Selain pengamatan langsung, juga dilakukan interview terhadap pihak akademisi untuk menentukan beberapa hal sebagai berikut:

- Standar evaluasi nilai sks untuk masa studi dua tahun pertama yang harus ditempuh minimal 80 sks dan untuk masa studi empat tahun yang harus ditempuh minimal 140 sks.
- Status mahasiswa dalam evaluasi mahasiswa yaitu mahasiswa berstatus normal atau memenuhi, peringatan dan bermasalah.
- Standar evaluasi nilai IPK untuk mahasiswa berstatus normal adalah $\geq 2,75$ untuk IPK mahasiswa berstatus peringatan adalah $\geq 2,01$ dan $\leq 2,65$, serta IPK untuk

mahasiswa dengan status bermasalah adalah $\leq 2,00$.

- Data mahasiswa yang akan dimining adalah data mahasiswa yang berstatus peringatan dan bermasalah.
- Target yang akan dicapai adalah mahasiswa drop out dan dipertahankan.



Gambar 1. Standar Evaluasi Terhadap Mahasiswa

Perancangan Sistem

Perancangan sistem adalah penggambaran, perencanaan dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah kedalam satu kesatuan yang utuh dan berfungsi. Bentuk DFD dapat dilihat pada gambar berikut.



Gambar 1. DFD Level 0 Sistem Evaluasi Mahasiswa

Tabel 2. Hasil Data Integration

No	NIM	PRODI	SKS1	IPSI	SKS2	IPS2	SKS3	IPS3	SKS4	IPS4	SKS5	IPS5	SKS6	IPS6	SKS7	IPS7	SKS8	IPS8	SKSLulus	IPK	BSS	STATUS
1	060501030	ILMU ADMINISTRASINEGARA	24	3.25	24	3.63	24	3.5	9	3	21	3.57	4	4	0	0	6	4	105	2.53	0	Lulus
2	0702022001	PENDIDIKAN BAHASA DAN SASTRA I	20	3.1	24	3.17	24	3.08	22	3.45	24	3.08	20	3	12	3.17	8	3.5	154	3.17	0	Lulus
3	0702022002	PENDIDIKAN BAHASA DAN SASTRA I	20	2.7	22	2.95	22	2.91	22	3	24	3.08	22	2.82	14	2.71	8	3.5	154	2.93	0	Lulus
4	0702022004	PENDIDIKAN BAHASA DAN SASTRA I	20	2.9	22	3.05	24	2.83	22	2.82	22	3.09	22	2.91	14	2.71	8	3.5	154	2.94	0	Lulus
5	0702022008	PENDIDIKAN BAHASA DAN SASTRA I	20	3	24	3	24	2.92	22	2.91	22	3	20	3.1	14	3	14	3.29	160	3.01	0	Lulus
6	0602022505	PENDIDIKAN BAHASA DAN SASTRA I	20	2.8	22	2.68	22	1.09	0	0	0	0	0	0	0	0	0	0	64	2.17	0	Drop Out
7	0602022525	PENDIDIKAN BAHASA DAN SASTRA I	20	2.6	22	3.38	22	0.55	22	0	0	0	0	0	0	0	0	0	86	1.56	0	Drop Out
8	0602022528	PENDIDIKAN BAHASA DAN SASTRA I	10	2.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	2.8	0	Drop Out
9	0602022535	PENDIDIKAN BAHASA DAN SASTRA I	20	2.7	22	3	24	2.92	22	3.27	24	2.92	22	2.91	12	0	0	0	146	2.71	0	Drop Out
10	0602022537	PENDIDIKAN BAHASA DAN SASTRA I	20	3.1	24	3.04	24	2.67	0	0	0	0	0	0	0	0	0	0	68	2.93	0	Drop Out

d) Data Transformation

Setelah penyeleksian data, maka dilanjutkan mengubah atau menggabungkan data ke dalam format yang sesuai untuk kemudian diproses dalam data mining. Proses transformasi data mahasiswa berikut ini.

Tabel 1. Proses Transformasi Data

Standar		Hasil Transformation
IPS semt. 1-8 dan IPK	IPS 1-8 dan IPK ≥ 3	3
	2 ≤ IPS 1-8 dan IPK < 3	2
	IPS 1-8 dan IPK < 2	1
SKS semt. 1-8	SKS > 21	Baik
	18 ≤ SKS ≤ 21	Cukup
	SKS < 18	Kurang
SKS Lulus 4 tahun	SKS > 140	Baik
	SKS = 140	Cukup
	110 ≤ SKS ≤ 140	Kurang
	SKS < 110	Sangat Kurang

Hasil transformasi data dari data set adalah 925 data, yang selanjutnya dijadikan sebagai data latih sebanyak 725 data. Dengan jumlah data uji dipilih secara acak sebanyak 200 data.

Gambar 6. Penentuan Training Data dan Testing Data

Proses Mining

Setelah melewati tahapan preprocessing maka data set telah siap untuk dilanjutkan ke tahapan berikutnya yaitu menggali informasi melalui data mining. Metode yang digunakan dalam proses klasifikasi adalah algoritma Naive Bayes [9] dengan rumus berikut ini.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dengan, P(X|H) adalah peluang data sampel X, apabila diasumsikan hipotesa benar.

Beberapa metode smoothing yang dapat diimplementasikan pada Naive Bayes diantaranya adalah Jelinek-Mercer (JM), Dirichlet (Dir), Absolute discounting (AD), Modified Absolute Discounting (MAD) dan Twostage (TS). Salah satu teknik smoothing dari fungsi klasifikasi Naive Bayes yang digunakan dalam penelitian ini adalah Dirichlet Smoothing dengan formula seperti persamaan 2. Adapun fungsi perhitungan teknik smoothing ini dilakukan agar tidak ada nilai probabilitas yang bernilai 0.

$$P\mu(w|Ci) = \frac{c(w, ci) + \mu P(w|C)}{\sum_{w' \in V} c(w', ci) + \mu} \quad (2)$$

Pada formula tersebut $c(w,ci)$ adalah frekuensi term w dalam kategori ci , μ merupakan koefisien control, $\sum w' \in V^{c(w,ci)}$ adalah banyaknya term dalam kelas ci , $P(W|Ci)$ adalah peluang banyaknya term w dalam kelas C dibagi banyaknya semua term dalam kelas C .

Penentuan Ci dan X

Pada tahapan ini ditentukan bahwa Ci ditentukan berdasarkan dua kelas atau kelompok yaitu $C1$ adalah kelas “Lulus” dan

$C2$ adalah Kelas “Dropout”. Sedangkan nilai X terdapat 19 macam yang mewakili IPS dan SKS masing-masing semester, IPK, jumlah SKS yang ditempuh dan total BSS.

Jumlah Kelas Pada Training Data

Melalui perhitungan jumlah setiap kelas training data berdasarkan probabilitas yang terbentuk maka jumlah masing-masing kelas diketahui bahwa $C1$: Kelas “Lulus” sebanyak 472 sedangkan $C2$: Kelas “Dropout” 253 dari 725 training data.

Tabel 3. Training Data

NIM	SKS1	IPS1	SKS2	IPS2	SKS3	IPS3	SKS4	IPS4	SKS5	IPS5	SKS6	IPS6	SKS7	IPS7	SKS8	IPS8	SKS LULUS	IPK	TOTAL BSS	STATUS
602022461	Cukup	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Kurang	2 Kurang	2 Kurang	3 Baik	3	0	LULUS	
602022464	Cukup	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022465	Cukup	3 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Cukup	3 Kurang	2 Kurang	3 Baik	3 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022466	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022469	Cukup	3 Baik	2 Baik	3 Baik	3 Baik	2 Baik	2 Baik	2 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022470	Cukup	3 Baik	3 Baik	3 Baik	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		
602022471	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022472	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022473	Kurang	3 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	3	0	DROP OUT		
602022474	Cukup	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Baik	2 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022480	Cukup	3 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022484	Cukup	2 Cukup	2 Baik	3 Baik	3 Baik	2 Baik	2 Baik	3 Cukup	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022485	Cukup	2 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Cukup	2 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022491	Cukup	2 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Kurang	2 Kurang	2 Kurang	3 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022493	Kurang	2 Cukup	2 Baik	2 Kurang	2 Kurang	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Kurang	3 Baik	2	0	LULUS	
602022495	Cukup	2 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Cukup	2 Kurang	2 Kurang	2 Kurang	3 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022498	Cukup	3 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022499	Cukup	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022502	Cukup	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Cukup	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022503	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	2 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022505	Cukup	3 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022506	Cukup	3 Baik	2 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		
602022509	Cukup	2 Baik	2 Cukup	3 Baik	2 Baik	2 Baik	2 Baik	2 Cukup	2 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022510	Cukup	2 Kurang	2 Baik	2 Baik	2 Baik	2 Cukup	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022511	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		
602022513	Cukup	3 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	2 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022514	Cukup	2 Baik	3 Baik	3 Baik	3 Baik	2 Baik	2 Baik	3 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022515	Cukup	2 Baik	2 Baik	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		
602022524	Cukup	2 Baik	3 Baik	2 Baik	2 Baik	2 Baik	2 Baik	2 Kurang	2 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022525	Cukup	2 Baik	3 Baik	1 Baik	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	1	0	DROP OUT		
602022527	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022528	Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		
602022529	Cukup	3 Baik	2 Baik	2 Baik	3 Baik	3 Baik	2 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022532	Cukup	2 Baik	2 Baik	2 Baik	2 Baik	2 Baik	1 Kurang	1 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022533	Cukup	2 Baik	2 Baik	3 Baik	3 Baik	3 Baik	3 Baik	2 Cukup	2 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022534	Cukup	2 Baik	3 Baik	3 Baik	3 Baik	2 Baik	2 Cukup	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	2 Kurang	3 Baik	3 Baik	2	0	LULUS	
602022535	Cukup	2 Baik	3 Baik	2 Baik	3 Baik	3 Baik	2 Baik	2 Baik	2 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Baik	2	0	DROP OUT		
602022536	Cukup	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Baik	3 Kurang	2 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Kurang	3 Baik	3 Baik	3	0	LULUS	
602022537	Cukup	3 Baik	3 Baik	2 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Kurang	1 Sangat Kurang	2	0	DROP OUT		

Evaluasi Performa Klasifikator

Kinerja klasifikator dapat dievaluasi dan dianalisis berdasarkan confusion matrix yang terbentuk [10], dengan tiga metode yaitu :

a. Accuracy, yaitu menghitung tingkat keakurasian metode dalam memberikan informasi.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

b. Precision (P) adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$P = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

c. Recall (R) adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

$$R = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

3. HASIL DAN PEMBAHASAN

Proses penyiapan data dalam preprocessing secara otomatis dilakukan oleh sistem. Selanjutnya dilakukan pengujian untuk mengetahui sistem kerja pada Naive Bayes tanpa smoothing dan Naive Bayes dengan Dirichlet Smoothing. Koefisien control (μ) ditentukan sebesar 0,1.

Sesuai dengan langkah yang dibahas sebelumnya, training data ditentukan dari data set yang ada, dan testing data dipilih

secara acak. Hasil pengujian atau pengolahan data ini berupa confusion matrix yang dapat digunakan untuk menghitung accuracy, precision dan recall. Melalui 5 kali ujicoba, maka perbandingan accuracy, precision

dan recall hasil perhitungan Naive Bayes tanpa smoothing dan Naive Bayes dengan Dirichlet Smoothing ditunjukkan pada tabel berikut

Tabel 4. Nilai Hasil Evaluasi Klasifikator

Evaluasi	Metode	Pengujian ke-					Mean
		1	2	3	4	5	
Accuracy (%)	Naive Bayes tanpa smoothing	83.1	85.7	86.3	82.6	85.9	84.72
	Naive Bayes dengan Dirichlet Smoothing	88.5	91.5	91.5	88	91.5	90.2
Selisih Accuracy (%)							5.48
Precision (%)	Naive Bayes tanpa smoothing	83.6	86.2	84.4	80.2	82.6	83.4
	Naive Bayes dengan Dirichlet Smoothing	91.4	94.6	91.1	86.5	88.78	90.5
Selisih Precision (%)							7.1
Recall (%)	Naive Bayes tanpa smoothing	81	84	88	86	91	86
	Naive Bayes dengan Dirichlet Smoothing	85	88	92	90	95	90
Selisih Recall (%)							4

Berdasarkan hasil pengujian pada data set berjumlah 925 data yang terbagi menjadi 725 *training data* dan 200 *testing data* menggunakan Naive Bayes tanpa smoothing dan Naive Bayes dengan Dirichlet Smoothing, dapat diketahui bahwa nilai precision, recall, dan accuracy. Perbandingan nilai yang dihasilkan kedua metode adalah memiliki selisih nilai masing-masing, meliputi nilai accuracy sebesar 5.48 %, nilai precision sebesar 7.1%, sedangkan nilai recall sebesar 4%. Secara umum, hasil dari pengujian menunjukkan nilai akurasi yang cukup tinggi.

Dalam hal ini tingkat konsistensi data ikut membawa pengaruh terhadap hasil prediksi kelulusan mahasiswa. Selain itu hasil pengujian data menunjukkan peran metode smoothing pada Naive Bayes. Penelitian ini berkontribusi pada pihak akademisi dan mahasiswa dalam mengevaluasi hasil belajarnya. Selain itu penelitian ini diharapkan menjadi evaluasi atau contoh penerapan metode smoothing pada Naive Bayes.

Kesimpulan

Berdasarkan hasil pengujian dan pembahasan yang dilakukan pada bagian sebelumnya, maka diperoleh kesimpulan berikut ini:

- Sistem ini dapat memberikan prediksi pada pihak akademisi hasil pencapaian mahasiswa pada tahun kedua dan tahun keempat masa studi.
- Dengan sistem ini mahasiswa mampu mengevaluasi, memperbaiki dan meningkatkan kinerja belajarnya hingga penyelesaian studinya tidak melampaui batas masa studi yang ditentukan yaitu tujuh tahun.
- Dengan menerapkan data mining pada Sistem Evaluasi Mahasiswa menggunakan Naive Bayes tanpa smoothing dan Naive Bayes dengan Dirichlet Smoothing, sistem dapat mengevaluasi perbedaan kedua metode, sehingga menjadikan gambaran tentang kontribusi metode smoothing pada Naive Bayes.
- Dari hasil perhitungan klasifikasi selama 5 kali dengan menggunakan algoritma Naive Bayes tanpa smoothing dan Naive Bayes dengan Dirichlet Smoothing pada 725 *training data* dan 200 *testing data* yang diambil secara acak, maka didapatkan

4. KESIMPULAN DAN SARAN

selisih nilai akurasi mencapai 5.48%, nilai *precision* sebesar 7.1% dan nilai *recall* adalah 4%.

Saran

Penelitian ini memiliki banyak keterbatasan, oleh sebab itu penelitian selanjutnya dapat dilakukan dengan penambahan atau penggantian metode Dirichlet Smoothing dengan metode *smoothing* lainnya sebagai bentuk analisis perbandingan hasil pengujian data sehingga optimalisasi tingkat akurasi klasifikasi berdasarkan penerapan metode *smoothing* pada Naive Bayes dapat diketahui.

Selain menambah atau mengganti metode pengukuran, indikator yang menentukan kinerja mahasiswa dapat pula ditambah, karena pada prinsipnya, semakin banyak indikator evaluasi yang digunakan, maka tingkat keakurasian penelitian akan lebih baik. Selain itu dengan menambah metode *smoothing*, peneliti dapat membandingkan keakurasian perhitungan Naive Bayes dengan masing-masing *smoothing*, sehingga diketahui metode *smoothing* yang paling tepat untuk diterapkan pada evaluasi kinerja mahasiswa.

Penelitian selanjutnya juga dapat menambahkan konten saran pada mahasiswa mengenai perbaikan pada ipk misalnya jumlah mata kuliah yang harus diulang atau jumlah dan batas nilai mata kuliah yang belum ditempuh. Dengan demikian hasil evaluasi kinerja mahasiswa dalam bentuk laporan dapat diberikan bukan hanya untuk pihak akademik, namun juga dapat diakses langsung oleh mahasiswa.

UCAPAN TERIMAKASIH

Pada kesempatan ini, kami mengucapkan terimakasih dan penghargaan yang sebesar-besarnya pada Direktorat Riset dan Pengabdian kepada Masyarakat khususnya Direktorat Jenderal Penguatan Riset dan Pengembangan, yang telah memberikan kontribusi kepada kami sehingga kami dapat melaksanakan penelitian ini dengan baik. Terimakasih juga kami sampaikan kepada semua pihak yang secara tidak langsung membantu pelaksanaan penelitian ini.

5. REFERENSI

[1] Ridwan, M., Suyono, H. dan Sarosa, M. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier”, *Jurnal EECCIS*. Vol. 7, No. 1. Hal. 59-64.

- [2] Suhatinah, M.S. dan Ernastuti. 2010. *Graduation Prediction of Gunadarma University Students Using Naive Bayes Algorithm and C4.5 Algorithm*. http://papers.gunadarma.ac.id/files/journals/3/articles/816/public_816-2182-1-PB.pdf/. Diakses 23 April 2016.
- [3] Garcia, E.P.I. dan Mora, P.M. 2011. Model Prediction of Academic Performance for First Year Students”, *IEEE Computer Society*.
- [4] Shovon, Md.H.I. dan Haque, M. 2012. Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm. *IJARCSSE*. Vol. 2, Issue 7. Hal. 353-355.
- [5] Kabakchieva, D. 2012. Student Performance Prediction by Using Data Mining Classification Algorithms. *IJCSMR*. Vol. 1, Issue 4. Hal. 686-690.
- [6] Nugroho, Y. S. 2014. *Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro*. <http://eprints.dinus.ac.id/5542/1/13789.pdf>. Diakses 26 Maret 2016.
- [7] Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu, Yogyakarta
- [8] Han, J. dan Kamber, M. 2006. *Data Mining: Concepts and Techniques*. Second Edition. The Morgan Kaufmann. San Fransisco.
- [9] Prasetyo, E. 2012. *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*. Andi, Yogyakarta.
- [10] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.