

# SISTEM EVALUASI DAN KLASIFIKASI KINERJA AKADEMIK MAHASISWA UNIVERSITAS MADURA MENGGUNAKAN NAIVE BAYES DENGAN DIRICHLET SMOOTHING

Erwin Prasetyowati<sup>1)</sup> dan Nilam Ramadhani<sup>2)</sup>

<sup>1, 2)</sup>Teknik Informatika, Fakultas Teknik, Universitas Madura

Jl. Raya Panglegur Km. 3,5 Pamekasan, Jawa Timur

e-mail: [erwinprasetyowati@unira.ac.id](mailto:erwinprasetyowati@unira.ac.id)<sup>1)</sup>, [nilam\\_ramadhani@yahoo.com](mailto:nilam_ramadhani@yahoo.com)<sup>2)</sup>

## ABSTRAK

Pengawasan terhadap kinerja akademik mahasiswa sebagai bentuk peningkatan mutu harus dilakukan secara sistematis dan terintegrasi. Namun pengawasan tersebut akan lebih efektif jika dilakukan secara periodik, yaitu pada tahun kedua dan keempat dengan maksud pihak akademik dapat mengetahui perkembangan pencapaian belajar masing-masing mahasiswa dengan cepat, sehingga peringatan atau tindakan yang akan diberikan pada mahasiswa dan evaluasi terhadap seluruh aktivitas akademis dapat segera dilakukan. Penelitian ini bertujuan mengklasifikasikan kinerja mahasiswa melalui IPK dan jumlah SKS yang belum diselesaikan selama masa studinya. Klasifikasi dilakukan pada tahun kedua dan tahun keempat masa studi. Pada tahun kedua, klasifikasi dibagi menjadi tiga status yaitu Normal, Bermasalah dan Peringatan dengan ditentukan melalui standar nilai yang ditentukan. Pada tahun keempat klasifikasi dibagi menjadi dua kelompok, yaitu kelompok Lulus dan Drop Out. Proses klasifikasi yang dilakukan pada tahun keempat, menggunakan algoritma Naïve Bayes yang terbukti memiliki tingkat keakuratan yang tinggi dengan metode maximum likelihood atau berdasarkan kemiripan tertinggi dari data yang diolah. Pada proses penghitungan Naïve Bayes terdapat sedikit keraguan apabila ada peluang yang bernilai 0. Oleh karena itu untuk memaksimalkan performa dari Naïve Bayes dalam klasifikasi ini, maka digunakan Dirichlet Smoothing. Berdasarkan hasil pengujian terhadap 200 data uji maka didapatkan nilai akurasi mencapai 91.50%, nilai precision sebesar 88.78% dan nilai recall adalah 95%. Dengan demikian dapat diketahui bahwa data memiliki nilai yang konsisten.

**Kata Kunci:** Algoritma Naïve Bayes, IPK, Kinerja Mahasiswa, SKS.

## ABSTRACT

Supervision of student academic performance as a form of quality improvement, must be done systematically and integrated. However, the supervision will be more effective if done periodically, ie in the second and fourth years with the intention of the academic can know the progress of learning achievement of each student quickly, so that the warning or action to be given to students and evaluation of all academic activities can be immediately do. This study aims to classify student performance through GPA and the number of credits that have not been completed during the study period. Classification is done in the second year and the fourth year of study. In the second year, the classification is divided into three states namely Normal, Troubled and Warning by determined through the standard value specified. In the fourth year of classification is divided into two groups, namely Graduation and Drop Out group. The classification process carried out in the fourth year, using the Naïve Bayes algorithm is proven to have a high degree of accuracy with the maximum likelihood method or based on the highest similarity of the treated data. In the Naïve Bayes calculation process there is little doubt when there is a probability of 0. Therefore to maximize the performance of Naïve Bayes in this classification, Dirichlet Smoothing is used. Based on the test results of 200 test data, the obtained accuracy value reached 91.50%, precision value of 88.78% and the recall value is 95%. Thus it can be seen that the data has a consistent value.

**Keywords:** Naïve Bayes Algorithm, IPK, Performance of student, SKS.

## I. PENDAHULUAN

**K**INERJA suatu institusi pendidikan merupakan salah satu faktor penting penentu keberhasilan mencetak lulusan yang berkualitas. Oleh sebab itu penilaian kinerja harus dilakukan secara menyeluruh pada semua elemen yang berkontribusi terhadap berlangsungnya aktivitas akademik yang dilakukan. Penilaian yang dilakukan tentunya berpedoman terhadap standar nilai yang ditetapkan oleh pihak internal maupun eksternal.

Salah satu standar penilaian bagi perguruan tinggi sebagai salah satu institusi pendidikan adalah kinerja dari mahasiswa, yang meliputi input, proses dan outputnya. Hal terpenting yang perlu diperhatikan pada saat proses pembelajaran berlangsung adalah pengawasan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang dicapai mahasiswa, rasio jumlah mahasiswa yang lulus terhadap total mahasiswa dan kompetensi lulusan. Hasil capaian tersebut tentu saja berpengaruh pada ketepatan mahasiswa dalam menyelesaikan waktu studinya, dan lulusan yang dihasilkan akan memiliki kepercayaan penuh dari penggunanya [1].

Berdasarkan uraian tersebut penelitian ini difokuskan pada perancangan suatu sistem kontrol terhadap hasil belajar mahasiswa melalui lamanya waktu studi, IPK dan jumlah Satuan Kredit Semester (SKS) yang telah ditempuh menjadi input evaluasi untuk memprediksi kemampuan mahasiswa dalam menyelesaikan studinya

dengan tepat waktu. Sistem pengawasan dan evaluasi ini dilakukan di tahun kedua dan keempat selama mahasiswa menempuh studinya. Hasil pengawasan pada tahun kedua selanjutnya dapat digunakan sebagai dasar pengklasifikasian kinerja mahasiswa menjadi tiga kelompok status yaitu status Normal, Bermasalah dan Peringatan. Pengelompokan status tersebut menggunakan standar evaluasi pencapaian mahasiswa yang dinilai yang telah ditetapkan pihak Universitas Madura.

Kelompok tersebut kemudian dievaluasi dan diklasifikasikan lagi untuk mendapatkan kelompok baru pada tahun keempat, yang kemudian dianalisis dan diputuskan tindakan lanjutan terhadap mahasiswa tersebut. Dua klasifikasi tersebut adalah kelompok Lulus yaitu kelompok mahasiswa yang berkinerja normal atau diprediksi lulus tepat waktu, dan mahasiswa *Drop Out* dimana nilai yang dicapai tidak memungkinkan lulus tepat waktu. Standar masa studi mahasiswa yang ditetapkan adalah 7 tahun.

Oleh sebab itu, untuk mempermudah sistem kontrol terhadap kinerja mahasiswa, diperlukan suatu aplikasi yang dapat digunakan mengklasifikasikan mahasiswa berdasarkan tingkat pencapaian belajarnya dengan menggunakan teknik *data mining*. Teknik *data mining* yang digunakan adalah algoritma *Naive Bayes Classifier* yang berfungsi untuk menangani masalah yang sifatnya hipotesis melalui perancangan desain suatu klasifikasi dengan memisahkan beberapa obyek berdasarkan Teorema Bayes [2]. Melalui aplikasi ini pengawasan dan penelusuran data mahasiswa dapat dilakukan dengan lebih efektif dan efisien mengingat jumlah mahasiswa setiap tahunnya mengalami peningkatan. Untuk lebih memaksimalkan kinerja dari Naive Bayes dalam penelitian ini penulis menggunakan *Dirichlet Smoothing* (Dir).

Penelitian tentang klasifikasi pencapaian mahasiswa telah banyak dilakukan sebelumnya menggunakan teknik *data mining*. Adapun teknik data mining yang digunakan meliputi Naive Bayes dan C4.5 [3] yang mengevaluasi apakah mahasiswa lulus tepat waktu atau tidak, bahkan evaluasi ada yang dimulai dari tahun pertama[4] dan tahun kedua [1], serta menggunakan Algoritma *K-Means Clustering* [5] dan jaringan syaraf tiruan dan K-NN [6]. Namun pembahasan yang dilakukan pada penelitian sebelumnya belum ada yang melakukan perhitungan menggunakan Naive Bayes dengan *Dirichlet Smoothing*. Oleh sebab itu penelitian ini bermaksud untuk menjawab kebutuhan tersebut.

## II. TINJAUAN PUSTAKA

### A. Data Mining

Data mining merupakan langkah analisis terhadap sekumpulan data yang umumnya berukuran besar untuk mendapatkan hubungan antar data tersebut dan merangkumnya dalam bentuk yang mudah dipahami dan digunakan [7]. Hubungan dan rangkuman yang dihasilkan tersebut pada umumnya berupa model (*models*) atau pola (*pattern*). Dalam prosesnya data mining menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* yang berfungsi untuk melakukan ekstraksi dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database [8][10]. Sehingga tujuan utama dari *data mining* adalah menemukan dan menggali pengetahuan dari data atau informasi yang ada.

Data mining merupakan salah satu bagian atau proses utama dari *Knowledge Discovery in Database* (KDD) yang bentuk kegiatannya yaitu mengumpulkan dan, menggunakan data masa lalu untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang lebih besar [2]. Secara garis besar KDD meliputi tiga tahapan, yaitu *pre processing*, *process* (data mining) dan *post processing*. Pada tahapan data mining, informasi dari sekumpulan data dapat ditentukan dengan menggunakan kemiripan data satu dengan yang lainnya. teknik yang sering digunakan adalah asosiasi, proses klasifikasi dan pengklasteran data.

*Clustering* merupakan teknik data mining yang banyak digunakan dalam penelitian yang bertujuan untuk mengevaluasi kinerja. Salah satunya adalah *clustering* untuk penilaian kinerja keuangan koperasi menggunakan K-Means yang dilakukan penelitian [9]. Penelitian ini menghasilkan evaluasi terhadap koperasi dan mengelompokkannya dalam dua kelompok yaitu koperasi Sehat dan Tidak Sehat dalam segi keuangan menggunakan analisis rasio. Penelitian [5] menggunakan K-Means Clustering untuk klasifikasi prestasi mahasiswa ke dalam tiga kelompok yaitu Tinggi, Medium dan Rendah, sehingga aktivitas belajar mahasiswa dapat diprediksi. K-Means Clustering merupakan teknik data mining yang tanpa arahan (*unsupervised learning*) sehingga pusat cluster dapat ditentukan secara acak dari data yang digunakan.

### B. Naive Bayes Classifier

*Naive Bayes Classifier* merupakan teknik klasifikasi probabilistik berdasarkan Teori Bayes dengan asumsi independensi (ketidaktergantungan) yang kuat dalam proses pengklasifikasiannya. Teorema keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Teorema Bayes berfungsi untuk menentukan atau menghitung probabilitas dari sebuah hipotesis berdasarkan probabilitas prior. Prior probabilitas adalah nilai dari probabilitas yang sejak awal diyakini benar, sebelum melakukan pengamatan dan analisis

terhadap data yang dimiliki. Dengan demikian algoritma dasar untuk menghitung probabilitas pada setiap kemungkinan dari hipotesis yang selanjutnya menentukan kemungkinan yang paling mendekati adalah dengan menggunakan Teorema Bayes.

*Naive Bayes* adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *machine learning* dan *data mining*. Berdasarkan ciri alami dari sebuah model probabilitas, klasifikasi *Naive Bayes* bisa dibuat lebih efisien dalam bentuk pembelajaran. Dalam beberapa bentuk praktiknya, parameter untuk perhitungan model *Naive Bayes* menggunakan metode *maximum likelihood* atau kemiripan tertinggi.

Formula untuk klasifikasi berdasarkan pada Teori Bayes untuk prediksi *Naive Bayes* [11] adalah sebagai berikut:

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

Penelitian [7] melakukan klasifikasi kelulusan mahasiswa dengan menggunakan algoritma *Naive Bayes* yang menghasilkan suatu pola, informasi dan pengetahuan baru sesuai dengan tujuan data mining. Dalam hal ini klasifikasi dibantu dengan digunakannya *Rapidminer* untuk meningkatkan akurasi alat klasifikasi. Penelitian ini hanya sampai pada tahap klasifikasi sehingga tidak digunakan untuk memprediksi kelulusan mahasiswa. Pembahasan yang tidak jauh berbeda dengan penelitian [7] adalah pada penelitian [4]. Data dibagi menjadi tiga kelompok: siswa yang tidak lulus atau sampai dua mata kuliah (rendah), siswa yang lulus tiga atau empat kelas (tengah), dan siswa yang lulus semua lima mata kuliah (tinggi).

Kinerja akademik mahasiswa telah dilakukan di tahun kedua pada penelitian [1]. Pada penelitian ini dapat diketahui bahwa data training dapat mempengaruhi hasil pengujian, karena pola data training tersebut akan dijadikan sebagai rule untuk menentukan kelas pada data testing. Sehingga besar atau kecilnya prosentase tingkat *precision*, *recall*, dan *accuracy* dipengaruhi juga oleh penentuan data training.

Tingkat keakurasian pengujian *Naive Bayes* pada penelitian [3] diperbandingkan dengan algoritma *C4.5*. Algoritma *Naive Bayes* akan memiliki akurasi yang lebih tinggi apabila diimplementasikan ke data yang berbeda dari data training dan jumlah datanya lebih besar. Penelitian ini menghasilkan gambaran bahwa tingkat kesalahan prediksi dengan algoritma *C4.5* lebih rendah daripada *Naive Bayes* karena *C4.5* melakukan klasifikasi *record-record* ke dalam kelas tujuan yang ada.

### C. Metode Dirichlet Smoothing Pada Naive Bayes

Salah satu teknik *smoothing* dari fungsi klasifikasi *Naive Bayes* adalah *Dirichlet Smoothing* dengan formula seperti persamaan 2.

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

Perhitungan teknik *smoothing* ini dilakukan agar tidak ada nilai probabilitas yang bernilai 0. Pada formula tersebut  $c(w, ci)$  adalah frekuensi term  $w$  dalam kategori  $ci$ ,  $\mu$  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  $C$ .

### D. Evaluasi Kinerja Klasifikasi

Untuk mengevaluasi kinerja klasifikasi yang dilakukan dengan melakukan analisis kemampuan alat klasifikasi mengenali tuple dari kelas yang berbeda, maka digunakan *confusion matrix* [12]. Akurasi dapat dihitung dengan membagi jumlah hasil klasifikasi benar dengan jumlah seluruh data melalui persamaan 3.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

Selain menghitung nilai *accuracy*, nilai *precision* dan *recall* dapat dihitung dengan ketentuan-ketentuan sebagai berikut:

1. *Precision* (P) adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Precision* dapat dihitung dengan persamaan 4.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{4}$$

2. *Recall* (R) adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Untuk menghitung nilai *recall* digunakan persamaan 5.

$$R = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

TABEL I  
STANDAR EVALUASI PENCAPAIAN MAHASISWA YANG DINILAI

Evaluasi	Evaluasi 2 Tahun	Evaluasi 4 Tahun
Status	Aktif	Aktif
Semester	4 smt	8 smt
Min.SKs	80 sks	140 sks
IPK	$\geq 2.75$	$\geq 2.75$
	$2.00 \leq x < 2.75$	$2.00 \leq x < 2.75$
	$< 2.00$	$< 2.00$

### III. METODOLOGI PENELITIAN

#### A. Kebutuhan Data

Untuk memenuhi semua aspek yang telah dirancang sebelumnya dalam penelitian ini, maka dibutuhkan beberapa data yang menunjang tercapainya tujuan yang ditetapkan. Adapun data yang diperlukan dalam penelitian ini adalah jumlah mahasiswa Universitas Madura yang tersebar di 10 program studi, baik mahasiswa yang aktif maupun tidak aktif.

Data yang digunakan dalam proses penggalan data untuk melakukan klasifikasi mahasiswa ke dalam kelas drop out dan dipertahankan adalah berupa data mahasiswa yang terdiri dari beberapa atribut yaitu nim, nama mahasiswa, jurusan, nilai KHS, dan nilai pendukung lainnya. Sumber data didapatkan dari database sistem informasi Universitas Madura. Dalam pengujian ini, data yang dipilih sebagai sampel adalah mahasiswa angkatan tahun 2005 – 2007 yang telah dinyatakan lulus dari Universitas Madura.

#### B. Prosedur Sistem

Untuk menganalisa jalannya suatu sistem, maka alur prosedur sistem pada sistem evaluasi mahasiswa yang digunakan adalah sebagai berikut:

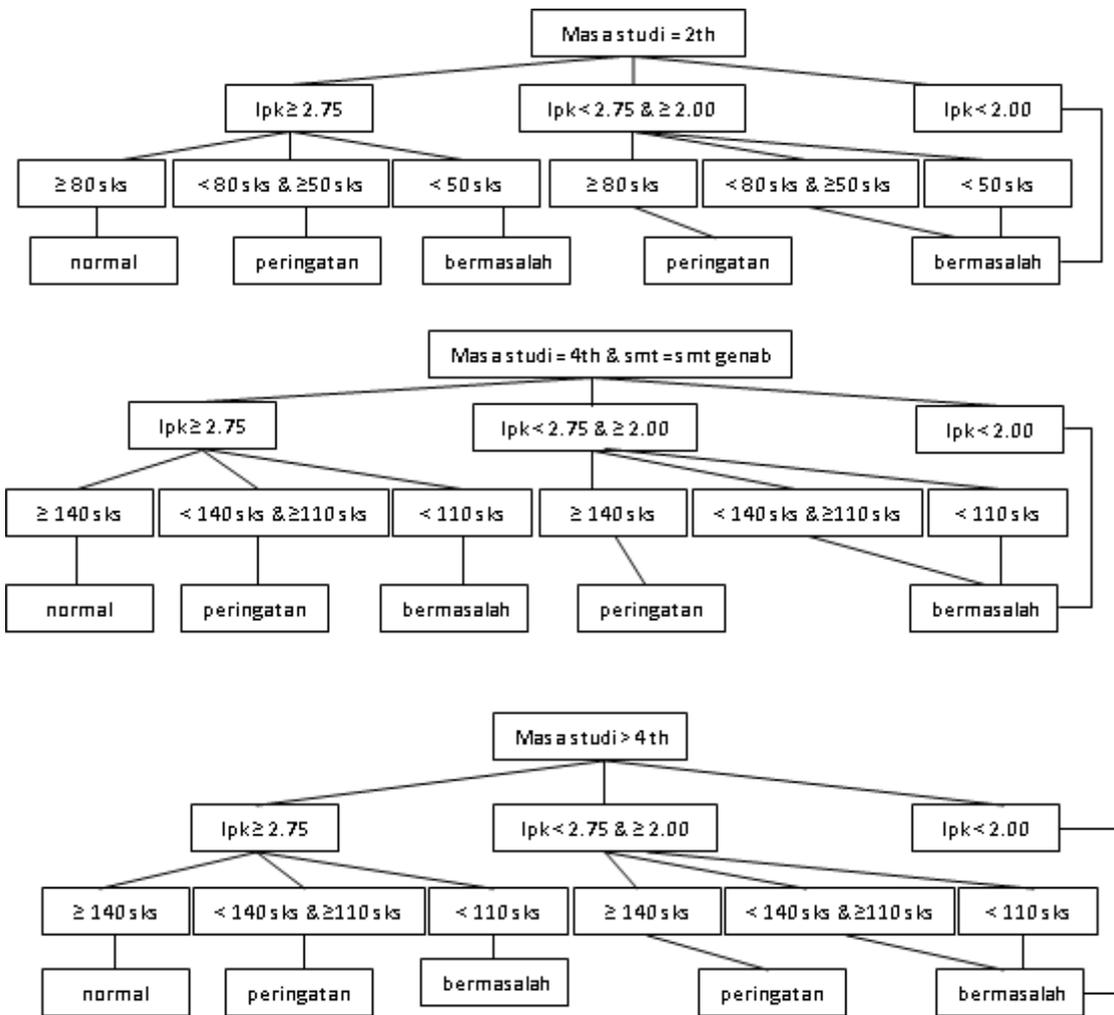
1. Admin melakukan evaluasi mahasiswa pada saat nilai akhir semester sudah dikeluarkan atau sudah tersimpan dalam database sistem informasi.
2. Sistem evaluasi mahasiswa akan mengambil data mahasiswa yang ada pada database sistem informasi untuk diproses.
3. Data mahasiswa akan dievaluasi berdasarkan standar nilai evaluasi yang sudah ditentukan. Tabel I menjelaskan standar nilai evaluasi pada sistem evaluasi.
4. Proses klasifikasi pada sistem evaluasi mahasiswa pada masa studi 2 tahun, masa studi empat tahun dan masa studi lebih dari empat t, yang alurnya dijelaskan pada Gambar 1. Pada periode tersebut, dilakukan proses pemilahan data berdasarkan IPK dan masa studi, hingga dimasukkan ke dalam tiga kelompok awal yaitu normal, peringatan dan bermasalah.
5. *Output* dari hasil evaluasi berupa laporan atau surat pemberitahuan kepada mahasiswa melalui program studinya masing-masing.

#### C. Diagram Alir

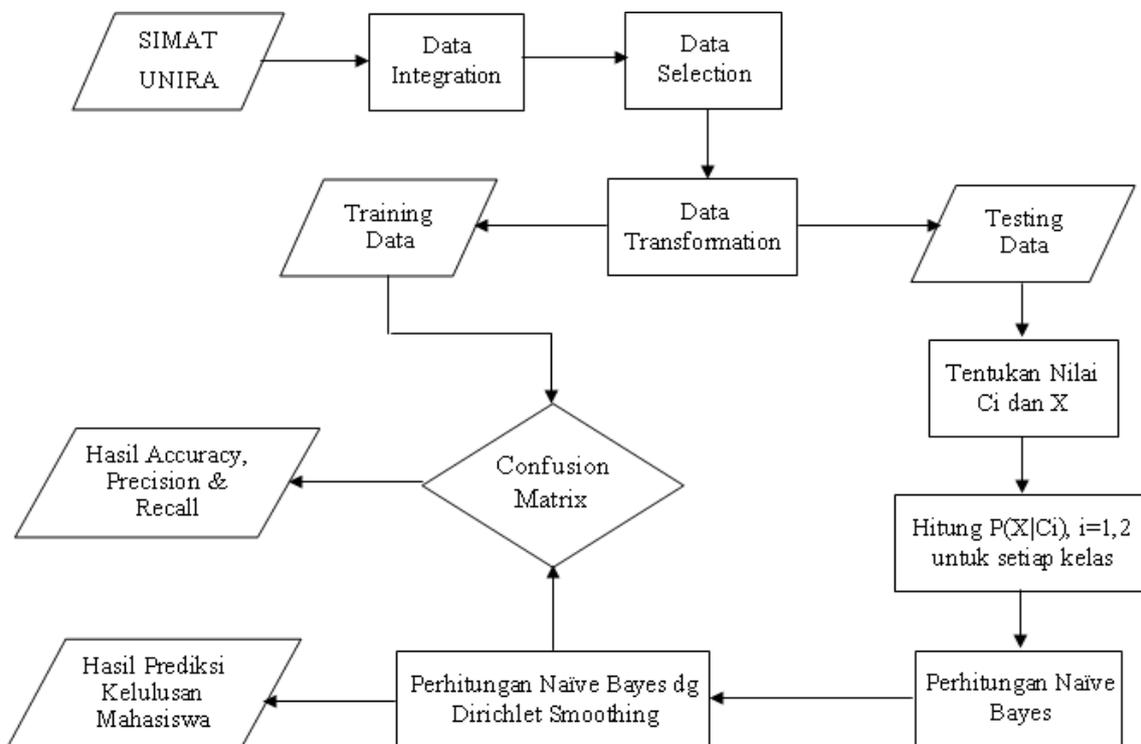
Pada bagian ini, digambarkan alur proses dari evaluasi melalui langkah-langkah klasifikasi menggunakan Algoritma Naïve Bayes dengan *Dirichlet Smoothing*. Gambar 2 menjelaskan langkah-langkah dalam evaluasi kinerja akademik mahasiswa dan kinerja klasifikator dalam melakukan prediksi. Hasil klasifikasi akan menentukan kinerja mahasiswa dan membaginya menjadi dua kelompok yaitu kelompok mahasiswa Lulus dan kelompok mahasiswa *Drop Out*.

#### D. Tahapan Data Mining

Dalam pengujian ini, data yang dipilih yaitu mahasiswa angkatan tahun 2005 – 2007 dipilah menjadi dua bagian yaitu *data training* (latih) dan *data testing* (uji). Pada algoritma Naive Bayes, *data training* digunakan untuk membentuk tabel probabilitas, sedangkan *data testing* berfungsi untuk menguji tabel probabilitas yang telah terbentuk. Sebelum dilakukan penggalan data, *data set* yang ada perlu melewati beberapa tahapan *preprocessing* data. Data *preprocessing* adalah suatu proses atau langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (*input* yang baik untuk *data mining tools*). Berikut ini adalah tahapan-tahapan yang dilakukan dalam *data mining*.



Gambar. 1. Proses klasifikasi evaluasi kinerja akademik mahasiswa selama 2 tahun, empat tahun dan lebih dari empat tahun



Gambar. 2. Diagram alir evaluasi kinerja akademik mahasiswa

### 1) Data Cleaning

Pada tahap ini dilakukan proses pembersihan noise dan data yang tidak konsisten atau data tidak relevan. Beberapa atribut yang dianggap tidak penting dan tidak berpengaruh pada proses pengklasifikasian seperti nama, jurusan, dan lainnya akan dihilangkan (*cleaning*). Pada proses pengklasifikasian mahasiswa ini, tahapan *cleaning* data tidak dilakukan karena data yang akan dipakai sudah sesuai dan relevan.

### 2) Data Integration

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Pada tahapan ini, data dari beberapa sumber dilakukan proses penggabungan. Sumber data yang digunakan berasal dari tabel-tabel pada database sistem informasi Universitas Madura, meliputi mahasiswa, mahasiswa\_bss\_mss, mahasiswa\_status1, fakultas, fakjur, krs, kurikulum dan wisuda.

### 3) Data Selection

Data-data yang sesuai dengan kebutuhan pengujian yang akan diambil dari database untuk dilakukan penyeleksian data dari data yang dibutuhkan dalam proses memining data, menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (*discovery*) akan dilakukan. Kemudian atribut yang penting dipilih dan digunakan pada proses selanjutnya, yang meliputi atribut NIM, SKS1-8, IPS1-8, IPK, SKS LULUS, BSS dan STATUS.

### 4) Data Transformation

Setelah penyeleksian data, maka dilanjutkan mengubah atau menggabungkan data ke dalam format yang sesuai untuk kemudian diproses dalam data mining. Adapun proses transformasi data mahasiswa yang telah diperoleh dari tahap data cleaning sampai data selection adalah sebagai berikut :

IPS 1-8 dan IPK, transformasi nilai IPS tiap semester 1-8 dan IPK dilakukan dengan membuat range nilai sebagai berikut :

- IPS 1-8 dan IPK  $\geq 3$ , nilainya 3.
  - $2 \leq$  IPS 1-8 dan IPK  $< 3$ , nilainya 2
  - IPS 1-8 dan IPK  $< 2$ , nilainya 1
- a. SKS 1-8, transformasi jumlah SKS tiap semester 1-8 dilakukan dengan membuat range nilai antara lain:
- SKS  $> 21$ , maka nilainya Baik
  - $18 \leq$  SKS  $\leq 21$ , maka nilainya Cukup
  - SKS  $< 18$ , maka nilainya Kurang
- b. SKS Lulus 4 Tahun, transformasi jumlah SKS lulus selama 4 tahun masa studi dilakukan dengan membuat range nilai antara lain:
- SKS  $> 140$ , maka nilainya Baik
  - SKS = 140, maka nilainya Cukup
  - $110 \leq$  SKS  $\leq 140$ , maka nilainya Kurang
  - SKS  $< 110$ , maka nilainya Sangat Kurang

Hasil transformasi data dari data set sebelumnya yaitu 887 data, yang selanjutnya dijadikan sebagai data training (latih) sebanyak 687 data.

## IV. HASIL DAN PEMBAHASAN

### A. Proses Mining

Tahapan berikutnya adalah memining data sebagai proses akhir, yang digunakan untuk mengklasifikasikan mahasiswa ke dalam dua kelas atau kelompok yaitu kelompok *Drop Out* (dipindahkan atau dikeluarkan) dan Lulus (dipertahankan). Proses klasifikasi ini menggunakan 687 data training dengan menggunakan perhitungan berdasarkan pada atribut yang ada dan sesuai, dengan probabilitas yang terbesar. Proses perhitungan guna menentukan klasifikasi mahasiswa menggunakan Metode Naïve Bayes tanpa *smoothing* dan Naïve Bayes dengan *smoothing* (*Dirichlet Smoothing*) melalui beberapa langkah yang akan dijelaskan berikut ini.

#### 1) Penentuan Data Uji (*Testing Data*)

Data uji (*testing data*) ditentukan secara acak dengan mengambil beberapa data yang selanjutnya dijadikan sampel untuk penentuan prediksi kelulusan mahasiswa menggunakan metode Naïve Bayes dengan Dirichlet Smoothing. Adapun sebagian dari data uji yang digunakan dapat dilihat pada gambar 5. Pada penelitian ini data uji yang digunakan adalah 200 data yang dipilih secara acak dari data set yang berjumlah 887 data seperti pada Tabel II.

TABEL II  
DATA UJI YANG DIGUNAKAN, DIAMBIL SEBAGIAN PADA SISTEM

NIM	SKS 1	IPS 1	SKS 2	IPS 2	SKS 3	IPS 3	SKS 4	IPS 4	SKS 5	IPS 5	SKS 6	IPS 6	SKS 7	IPS 7	SKS 8	IPS 8	SKS LULUS	IPK	TOTAL BSS	STATUS
702031666	Cukup	2	Cukup	2	Baik	2	Cukup	2	Kurang	2	Kurang	1	Kurang	1	Kurang	1	Kurang	2	2	?
702031667	Cukup	3	Baik	3	Baik	3	Cukup	2	Baik	2	Cukup	2	Kurang	1	Kurang	1	Kurang	2	0	?
702031679	Cukup	2	Cukup	3	Baik	2	Cukup	2	Kurang	1	Kurang	1	Kurang	1	Kurang	1	Sangat Kurang	2	0	?
702031690	Cukup	2	Cukup	2	Cukup	1	Kurang	1	Sangat Kurang	2	0	?								
702031768	Cukup	2	Cukup	2	Baik	2	Kurang	2	Cukup	2	Kurang	2	Kurang	2	Kurang	2	Kurang	2	0	?
702031774	Cukup	2	Cukup	2	Baik	2	Kurang	2	Baik	2	Kurang	2	Cukup	2	Kurang	1	Kurang	2	0	?
702031777	Cukup	2	Cukup	3	Baik	2	Cukup	2	Kurang	2	Kurang	1	Cukup	1	Kurang	2	Kurang	2	0	?
702031787	Cukup	3	Baik	2	Baik	2	Kurang	2	Cukup	2	Kurang	2	Cukup	2	Kurang	3	Kurang	2	0	?
702031789	Cukup	2	Cukup	2	Cukup	1	Kurang	1	Baik	2	Kurang	2	Kurang	1	Baik	2	Kurang	2	0	?
702031790	Cukup	1	Cukup	1	Kurang	1	Sangat Kurang	1	0	?										
702031791	Cukup	2	Cukup	2	Baik	2	Kurang	3	Baik	2	Cukup	2	Cukup	2	Kurang	3	Baik	2	0	?
712000947	Cukup	3	Baik	3	Baik	2	Cukup	3	Baik	3	Cukup	3	Cukup	2	Kurang	2	Baik	3	0	?
712000953	Cukup	2	Baik	3	Baik	2	Cukup	3	Baik	3	Baik	2	Kurang	3	Kurang	3	Baik	3	0	?
712000957	Cukup	1	Kurang	1	Sangat Kurang	1	0	?												
712000960	Cukup	2	Baik	2	Kurang	2	Cukup	2	Cukup	2	Cukup	3	Baik	1	Kurang	1	Baik	2	1	?
712000968	Cukup	3	Baik	2	Cukup	2	Cukup	2	Cukup	3	Baik	3	Baik	3	Kurang	3	Kurang	3	0	?
7120001002	Cukup	3	Baik	3	Baik	3	Baik	1	Kurang	1	Kurang	1	Kurang	1	Kurang	1	Sangat Kurang	2	3	?
7120001013	Cukup	2	Cukup	3	Baik	2	Cukup	2	Kurang	3	Baik	3	Baik	2	Kurang	3	Baik	2	0	?
715011384	Cukup	3	Baik	3	Baik	3	Baik	3	Baik	2	Cukup	3	Kurang	3	Kurang	3	Baik	3	0	?

TABEL III  
JUMLAH SETIAP CLASS DALAM DATA TRAINING

CLASS	JUMLAH
LULUS	350
DROP OUT	337
TOTAL	687

2) Menentukan Nilai Ci dan Nilai X

Sebelum melakukan proses perhitungan lebih lanjut, harus terlebih dahulu menentukan Nilai Ci dan X terlebih dahulu. Penentuan Nilai Ci dan X adalah sebagai berikut:

- a. Nilai Ci, meliputi Kelas C1 adalah kelompok LULUS dan Kelas C2 adalah kelompok Drop Out.
- b. Nilai X, yang terdiri dari :

X <sub>1</sub> : SKS1 (SKS Semester1)	X <sub>8</sub> : IPS4 (IP Semester4)	X <sub>14</sub> : IPS7 (IP Semester7)
X <sub>2</sub> : IPS1 (IP Semester1)	X <sub>9</sub> : SKS5 (SKS Semester5)	X <sub>15</sub> : SKS8 (SKS Semester8)
X <sub>3</sub> : SKS2 (SKS Semester2)	X <sub>10</sub> : IPS5 (IP Semester5)	X <sub>16</sub> : IPS8 (IP Semester8)
X <sub>4</sub> : IPS2 (IP Semester2)	X <sub>11</sub> : SKS6 (SKS Semester6)	X <sub>17</sub> : SKS LULUS 4th
X <sub>5</sub> : SKS3 (SKS Semester3)	X <sub>12</sub> : IPS6 (IP Semester6)	X <sub>18</sub> : IPK
X <sub>6</sub> : IPS3 (IP Semester3)	X <sub>13</sub> : SKS7 (SKS Semester7)	X <sub>19</sub> : TOTAL BSS
X <sub>7</sub> : SKS4 (SKS Semester4)		

3) Menghitung P(X|Ci), i=1,2 Untuk Setiap Kelas

Tahapan berikutnya adalah menghitung jumlah setiap kelas dari data training berdasarkan klasifikasi yang terbentuk (*prior probability*). Adapun jumlah setiap kelas dapat dilihat pada Tabel III.

Maka probabilitas setiap kelas adalah sebagai berikut:

$$P(\text{Status} = \text{LULUS}) = \frac{350}{687} = 0.50946142649199$$

$$P(\text{Status} = \text{DROP OUT}) = \frac{337}{687} = 0.49053857350801$$

B. Perhitungan Naive Bayes

Setelah masing-masing kelas pada data latih (training data) diketahui nilai probabilitasnya, maka langkah selanjutnya adalah menghitung P(X|Ci) untuk setiap atribut pada setiap class status yaitu LULUS atau DROP OUT. Berikut hasil perhitungan Naive Bayes pada salah satu data uji.

1) Probabilitas setiap atribut pada kelas atau kelompok Lulus hasilnya adalah sebagai berikut:

P(X <sub>1</sub> =Cukup C=Lulus) = 0.71	P(X <sub>8</sub> =2 C=Lulus) = 0.52	P(X <sub>14</sub> =1 C=Lulus) = 0.07
P(X <sub>2</sub> =2 C=Lulus) = 0.48	P(X <sub>9</sub> =Kurang C=Lulus) = 0.24	P(X <sub>15</sub> =Kurang C=Lulus) = 0.90
P(X <sub>3</sub> =Cukup C=Lulus) = 0.35	P(X <sub>10</sub> =2 C=Lulus) = 0.58	P(X <sub>16</sub> =1 C=Lulus) = 0.07

$P(X_4=2 C=Lulus)$	$= 0.53$	$P(X_{11}=Kurang C=Lulus)$	$= 0.30$	$P(X_{17}=Kurang C=Lulus)$	$= 0.19$
$P(X_5=Baik C=Lulus)$	$= 0.82$	$P(X_{12}=1 C=Lulus)$	$= 0.08$	$P(X_{18}=2 C=Lulus)$	$= 0.61$
$P(X_6=2 C=Lulus)$	$= 0.56$	$P(X_{13}=Kurang C=Lulus)$	$= 0.61$	$P(X_{19}=2 C=Lulus)$	$= 0$
$P(X_7=Cukup C=Lulus)$	$= 0.36$				

### 2) Menghitung $P(X|Ci = Lulus)$

Adapun  $P(X|Ci=Lulus)$  dapat dihitung dengan mengalikan semua probabilitas setiap atribut pada kelas Lulus.

Dengan demikian hasil dari perhitungan  $P(X|Ci=Lulus)$  sebagai berikut :

$$P(X|C=Lulus)P(C=Lulus) = 0$$

### 3) Probabilitas setiap atribut pada kelas atau kelompok Drop Out (DO) hasilnya adalah sebagai berikut:

$P(X_1=Cukup C=DO)$	$= 0.75$	$P(X_8=2 C=DO)$	$= 0.24$	$P(X_{14}=1 C=DO)$	$= 0.88$
$P(X_2=2 C=DO)$	$= 0.42$	$P(X_9=Kurang C=DO)$	$= 0.78$	$P(X_{15}=Kurang C=DO)$	$= 0.97$
$P(X_3=Cukup C=DO)$	$= 0.16$	$P(X_{10}=2 C=DO)$	$= 0.18$	$P(X_{16}=1 C=DO)$	$= 0.89$
$P(X_4=2 C=DO)$	$= 0.35$	$P(X_{11}=Kurang C=DO)$	$= 0.83$	$P(X_{17}=Kurang C=DO)$	$= 0.09$
$P(X_5=Baik C=DO)$	$= 0.34$	$P(X_{12}=1 C=DO)$	$= 0.83$	$P(X_{18}=2 C=DO)$	$= 0.44$
$P(X_6=2 C=DO)$	$= 0.28$	$P(X_{13}=Kurang C=DO)$	$= 0.93$	$P(X_{19}=2 C=DO)$	$= 0.05$
$P(X_7=Cukup C=DO)$	$= 0.14$				

### 4) Menghitung $P(X|Ci = DO)$

Menentukan  $P(X|Ci=DO)$  dilakukan dengan mengalikan semua probabilitas setiap atribut pada kelas Dropout.

Hasil dari perkalian tersebut sebagai berikut :

$$P(X|C=DO)P(C=DO) = 3.69$$

Berdasarkan perhitungan Naive Bayes di atas dapat disimpulkan bahwa  $P(X|Ci = Lulus) < P(X|Ci = DO)$ , sehingga kesimpulannya data atau mahasiswa yang bersangkutan masuk pada kelompok atau kelas Drop Out.

## C. Perhitungan Naive Bayes dengan Dirichlet Smoothing

Untuk menghindari nilai nol pada model probabilitas yang dihasilkan Naive Bayes, maka perlu diterapkan smoothing pada saat pembangunan model probabilitas tersebut. Untuk menyempurnakan hasil perhitungan Naive Bayes, maka digunakan *Dirichlet Smoothing*. Dengan menggunakan koefisien control  $\mu = 0.1$ , proses perhitungan  $P_{\mu}(X|Ci)$  untuk setiap atribut pada setiap class status yaitu LULUS dan DROP OUT dihitung dengan persamaan 2. Hasil perhitungannya dijelaskan pada berikut ini.

### 1) Menghitung $P(X|Ci = Lulus)$

$P_{\mu}(X_1=Cukup C=Lulus)$	$= 0.71$	$P_{\mu}(X_8=2 C=Lulus)$	$= 0.52$	$P_{\mu}(X_{14}=1 C=Lulus)$	$= 0.07$
$P_{\mu}(X_2=2 C=Lulus)$	$= 0.48$	$P_{\mu}(X_9=Kurang C=Lulus)$	$= 0.24$	$P_{\mu}(X_{15}=Kurang C=Lulus)$	$= 0.90$
$P_{\mu}(X_3=Cukup C=Lulus)$	$= 0.35$	$P_{\mu}(X_{10}=2 C=Lulus)$	$= 0.58$	$P_{\mu}(X_{16}=1 C=Lulus)$	$= 0.07$
$P_{\mu}(X_4=2 C=Lulus)$	$= 0.53$	$P_{\mu}(X_{11}=Kurang C=Lulus)$	$= 0.30$	$P_{\mu}(X_{17}=Kurang C=Lulus)$	$= 0.19$
$P_{\mu}(X_5=Baik C=Lulus)$	$= 0.82$	$P_{\mu}(X_{12}=1 C=Lulus)$	$= 0.08$	$P_{\mu}(X_{18}=2 C=Lulus)$	$= 0.61$
$P_{\mu}(X_6=2 C=Lulus)$	$= 0.56$	$P_{\mu}(X_{13}=Kurang C=Lulus)$	$= 0.61$	$P_{\mu}(X_{19}=2 C=Lulus)$	$= 6.65$
$P_{\mu}(X_7=Cukup C=Lulus)$	$= 0.36$				

Hasil perhitungan probabilitas setiap atribut pada kelas LULUS menggunakan Naive Bayes tanpa smoothing menunjukkan bahwa nilai  $P(X_{19}=2|C=Lulus) = 0$  sehingga menyebabkan  $P(X|Ci = Lulus)$  adalah 0. Sedangkan perhitungan Naive Bayes dengan Dirichlet Smoothing menghasilkan nilai  $P_{\mu}(X_{19}=2|C=Lulus) = 6.65$ , sehingga nilai  $P_{\mu}(X|Ci = Lulus)$  adalah 1.91.

### 2) Menghitung $P(X|Ci = Drop Out)$

$P_{\mu}(X_1=Cukup C=DO)$	$= 0.75$	$P_{\mu}(X_8=2 C=DO)$	$= 0.24$	$P_{\mu}(X_{14}=1 C=DO)$	$= 0.88$
$P_{\mu}(X_2=2 C=DO)$	$= 0.42$	$P_{\mu}(X_9=Kurang C=DO)$	$= 0.78$	$P_{\mu}(X_{15}=Kurang C=DO)$	$= 0.97$
$P_{\mu}(X_3=Cukup C=DO)$	$= 0.16$	$P_{\mu}(X_{10}=2 C=DO)$	$= 0.18$	$P_{\mu}(X_{16}=1 C=DO)$	$= 0.89$
$P_{\mu}(X_4=2 C=DO)$	$= 0.35$	$P_{\mu}(X_{11}=Kurang C=DO)$	$= 0.83$	$P_{\mu}(X_{17}=Kurang C=DO)$	$= 0.09$
$P_{\mu}(X_5=Baik C=DO)$	$= 0.34$	$P_{\mu}(X_{12}=1 C=DO)$	$= 0.83$	$P_{\mu}(X_{18}=2 C=DO)$	$= 0.44$
$P_{\mu}(X_6=2 C=DO)$	$= 0.28$	$P_{\mu}(X_{13}=Kurang C=DO)$	$= 0.93$	$P_{\mu}(X_{19}=2 C=DO)$	$= 0.05$
$P_{\mu}(X_7=Cukup C=DO)$	$= 0.14$				

Hasil perhitungan probabilitas setiap atribut pada kelas DROP OUT menggunakan Naive Bayes tanpa smoothing dan dengan Dirichlet Smoothing memiliki nilai yang sama (apabila dilakukan pembulatan angka). Dengan demikian dapat diketahui bahwa nilai  $P(X|Ci=DO)$  adalah 3.69.

TABEL IV  
CONFUSION MATRIX

	LULUS	DROP OUT
LULUS	95	5
DROP OUT	12	88

Berdasarkan hasil evaluasi dua dan empat tahun masa studi, maka didapatkan hasil prediksi ketepatan mahasiswa lulus tepat waktu atau tidak melalui status dan rekomendasi yang dihasilkan sistem. Status didapatkan dari evaluasi 2 tahun sedangkan rekomendasi didapatkan melalui proses perhitungan Naive Bayes dengan Dirichlet Smoothing. Perhitungan Naive Bayes dengan Dirichlet Smoothing tersebut, menunjukkan bahwa  $P_{\mu}(X | C_i=Lulus) < P_{\mu}(X | C_i=Drop\ out)$ . Dengan demikian data mahasiswa yang diuji masuk pada kelompok atau kelas DROP OUT, sesuai dengan perhitungan sebelumnya yaitu Naive Bayes tanpa *smoothing*.

*D. Evaluasi Terhadap Kinerja Naive Bayes*

Pengujian terhadap nilai akurasi klasifikasi ini bertujuan untuk mengetahui kinerja dari algoritma Naive Bayes dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Pada uji coba ini, diberikan data latih untuk membentuk tabel probabilitas yang kemudian diuji dengan data uji. Kinerja klasifikasi diperoleh dengan memberikan nilai pada *confusion matrix* untuk menghitung nilai *precision*, *recall*, dan *accuracy* dari hasil pengujian berdasarkan Tabel IV.

Melalui training data sebanyak 687 dengan 200 testing data yang diambil secara acak, maka berdasarkan hasil dari klasifikasi Naive Bayes dengan Dirichlet Smoothing, 183 data dinyatakan benar dan 17 data dinyatakan salah. Berikut hasil perhitungan *precision*, *recall* dan *accuracy*.

$$\text{Precision (P)} = \frac{95}{95+12} \times 100\% = 88,78 \%$$

$$\text{Recall (R)} = \frac{95}{95+5} \times 100\% = 95 \%$$

$$\text{Accuracy (A)} = \frac{95+88}{95+88+12+5} \times 100\% = 91.50\%$$

*E. Analisis Pengujian Data*

Proses klasifikasi data mahasiswa dengan Naive Bayes tanpa *smoothing* pada perhitungan probabilitas atribut setiap kelas ( $P(X|C_i)$ ) di beberapa data, memberikan gambaran bahwa beberapa diantaranya menghasilkan nilai 0. Sesuai dengan tujuan di awal penelitian, penggunaan metode *smoothing* dalam Naive Bayes dilakukan agar tidak ada nilai probabilitas dari hasil pendugaan yang bernilai 0. Misalnya ada data yang masuk ke dalam training data, maka nilai akan menjadi 0, dan kita hanya bisa melakukan prediksi pada P(X) saja. Penelitian ini memperkuat kesimpulan bahwa tujuan dari metode *smoothing* adalah mengurangi probabilitas dari output yang terobservasi dan menambahnya dengan output yang belum terobservasi [13]. Dengan demikian penelitian ini membuktikan bahwa teknik *smoothing* mampu meningkatkan kinerja Naive Bayes dengan lebih baik dari sebelumnya.

Tidak hanya berpengaruh pada nilai probabilitas atribut setiap kelas, penggunaan metode *smoothing* juga berpengaruh pada nilai *accuracy*. Metode *smoothing* terbukti mampu meningkatkan akurasi klasifikasi setelah melalui beberapa kali uji coba sesuai dengan penelitian sebelumnya. Peningkatan nilai juga terjadi pada nilai *precision* dan *recall*, sehingga menambah daftar kontribusi metode *smoothing* terhadap kinerja Naive Bayes.

Berdasarkan hasil pengujian pada 687 *training data* dan 200 *testing data* menggunakan Naive Bayes dengan Dirichlet Smoothing, dapat diketahui nilai *precision*, *recall*, dan *accuracy*. Masing-masing nilai yang dihasilkan adalah untuk nilai *precision* sebesar 88.78%, nilai *recall* sebesar 95%, sedangkan nilai *accuracy* sebesar 91.50%. Secara umum, hasil dari pengujian menunjukkan nilai akurasi yang cukup tinggi. Hal ini dapat disebabkan karena faktor penentu prediksi evaluasi mahasiswa Lulus atau *Drop Out* memiliki nilai yang konsisten. Meskipun di beberapa penelitian salah satunya menyebutkan bahwa besar kecilnya jumlah *training data* (data latih) tidak selalu berbanding lurus dengan nilai *precision*, *recall*, dan *accuracy* dari *testing data* [1].

Melalui sistem evaluasi kinerja mahasiswa ini, pihak akademik terutama masing-masing program studi dapat mengevaluasi capaian belajar mahasiswanya, sehingga menjadi rujukan bagi program studi untuk melaksanakan perbaikan dan peningkatan mutu pembelajaran serta memberikan saran-saran yang bermanfaat bagi mahasiswa dalam meningkatkan hasil belajarnya melalui bimbingan belajar atau dosen wali masing-masing. Sistem ini juga memberikan rekomendasi terhadap mahasiswa berupa hasil prediksi kelulusannya yang dapat digunakan

mahasiswa dalam meningkatkan kualitas dan kuantitas belajarnya, serta memberikan pedoman bagi mahasiswa untuk mengulang mata kuliah yang memiliki nilai rendah, meningkatkan nilai mata kuliah yang ditempuh sehingga jumlah sks terpenuhi dan mahasiswa dapat menyelesaikan studinya tepat waktu sesuai dengan batas masa studi yang ditetapkan yaitu selama tujuh tahun.

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan pembahasan yang dilakukan sebelumnya, dapat ditarik kesimpulan, yakni sebagai berikut :

1. Dengan adanya Sistem Evaluasi Mahasiswa dapat membantu pihak Universitas Madura untuk melakukan evaluasi hasil studi mahasiswa pada dua tahun dan empat tahun masa studi.
2. Dengan menerapkan data mining pada Sistem Evaluasi Mahasiswa menggunakan Naive Bayes dengan Dirichlet Smoothing, sistem dapat memberikan rekomendasi mengenai prediksi kelulusan mahasiswa pada empat tahun masa studi.
3. Dengan sistem ini mahasiswa mampu mengevaluasi, memperbaiki dan meningkatkan kinerja belajarnya hingga penyelesaian studinya tidak melampaui batas masa studi yang ditentukan yaitu tujuh tahun.
4. Dari hasil perhitungan klasifikasi dengan menggunakan algoritma Naive Bayes dengan Dirichlet Smoothing pada 687 *training data* dan 200 *testing data* yang diambil secara acak, maka didapatkan nilai akurasi mencapai 91.50%, nilai *precision* sebesar 88.78% dan nilai *recall* adalah 95%. Dengan demikian dapat diketahui bahwa data memiliki nilai yang konsisten.

### B. Saran

Penelitian ini memiliki banyak keterbatasan, oleh sebab itu penelitian selanjutnya dapat mengembangkan evaluasi kinerja mahasiswa menggunakan Algoritma Naive Bayes dengan *Dirichlet Smoothing*. Metode pengukuran yang digunakan dapat ditambah atau diganti dengan menggunakan *smoothing* lainnya sebagai bentuk analisis perbandingan hasil pengujian data sehingga optimalisasi tingkat akurasi klasifikasi yang didapatkan 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 berupa bantuan dana 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.

## DAFTAR PUSTAKA

- [1] M. Ridwan, H. Suyono dan M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier", *Jurnal EECCIS*, Vol. 7, No. 1, Hal. 59-64, Juni 2013.
- [2] B.Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu: Yogyakarta, 2007.
- [3] M.S. Suhatinah, dan Ernastuti, "Graduation Prediction of Gunadarma University Students Using Naive Bayes Algorithm and C4.5 Algorithm", 2010. Tersedia: <http://papers.gunadarma.ac.id/files/journals/3/articles/816/public/816-2182-1-PB.pdf/>.
- [4] E.P.I. Garcia dan P.M. Mora, "Model Prediction of Academic Performance for First Year Students", dalam *IEEE*, Mexico, 26 November – 4 Oktober 2011.
- [5] Md.H.I. Shovon dan M. Haque, "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm", *IJARCSSE*, Vol. 2, Issue 7, Hal. 353-355, Juli 2012.
- [6] D. Kabakchieva, "Student Performance Prediction by Using Data Mining Classification Algorithms", *IJCSMR*, Vol. 1, Issue 4, Hal. 686-690, November 2012.
- [7] J. Han, dan M.Kamber, *Data Mining: Concepts and Techniques, Second Edition*. The Morgan Kaufmann: San Fransisco, 2006.
- [8] E. Turban, J.E. Aronson dan T.P. Liang, *Decision Support System and Intelligent Systems - 7th ed*. Pearson Education, Inc. 2005.

- [9] E. Prsetyowati dan A.A. Rofiq, “Penilaian Kinerja Keuangan Koperasi pada Dinas Koperasi dan UMKM Pamekasan dengan K-Means”, *Jurnal Simantec*, Vol. 5, No. 2, Hal. 67-74, Juni 2016.
- [10] E. Prasetyo, *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*, Andi: Yogyakarta, 2012.
- [11] Q. Yuan, G. Cong dan N. M. Thalmann. “Enhancing Naive Bayes With Various Smoothing Methods For Short Text Classification”, dalam *International Conference on World Wide Web*, France, hal. 645-646, 16-20 April 2012.
- [12] B. Said dan Y.M. Pranoto, “Klasifikasi Data Sms Center Bupati Pamekasan Menggunakan Naïve Bayes Dengan Mad Smoothing”, dalam *IdeaTech*, STTS Surabaya, 2015, hal. 92-99.
- [13] E. Prsetyowati, “Implementasi Dirichlet Smoothing Pada Naïve Bayes Untuk Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura”, dalam *SENSEI*, Universitas Muhammadiyah Jember, hal. 171-178, 4 Oktober 2017.