

# Penerapan *Data Mining* Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier*

Mujib Ridwan, Hadi Suyono, dan M. Sarosa

**Abstrak**—Penelitian ini difokuskan untuk mengevaluasi kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak. Kemudian dari klasifikasi tersebut, sistem akan memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat dengan nilai optimal berdasarkan histori nilai yang telah ditempuh mahasiswa.

Input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa. Sampel mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus akan digunakan sebagai data *training* dan *testing*. Sedangkan data mahasiswa angkatan 2010-2011 dan belum lulus akan digunakan sebagai data target. Data input akan diproses menggunakan teknik *data mining* algoritma *Naive Bayes Classifier* (NBC) untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kelulusan mahasiswa. Output dari sistem ini berupa klasifikasi kinerja akademik mahasiswa yang diprediksi kelulusannya dan memberikan rekomendasi untuk proses kelulusan tepat waktu atau lulus dalam waktu yang paling tepat dengan nilai optimal.

Hasil pengujian menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa yaitu Indeks Prestasi Kumulatif (IPK), Indeks Prestasi (IP) semester 1, IP semester 4, dan jenis kelamin. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi. Pengujian pada data mahasiswa angkatan 2005-2009, algoritma NBC menghasilkan nilai *precision*, *recall*, dan *accuracy* masing-masing 83%, 50%, dan 70%.

**Kata Kunci**—Kinerja akademik mahasiswa, *data mining*, dan *Naive Bayes Classifier*.

## I. PENDAHULUAN

Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Pemantauan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang dicapai mahasiswa, rasio kelulusan terhadap jumlah total mahasiswa, dan

kompetensi lulusan seyogyanya mendapatkan perhatian yang serius untuk memperoleh kepercayaan *stakeholder* dalam menilai dan menetapkan penggunaan lulusannya.

Berdasarkan uraian di atas, pada penelitian ini akan dibuat sebuah sistem untuk mengklasifikasikan kelulusan mahasiswa dengan cara mengevaluasi kinerja pada tahun pertama dan atau tahun kedua. Pada penelitian ini, digunakan teknik *data mining* untuk menemukan pola kelulusan mahasiswa yang sudah lulus, kemudian dijadikan dasar untuk memprediksi kelulusan mahasiswa pada tahun ke-2.

*Data mining* adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan [7]. Teknik *data mining* yang akan digunakan dalam penelitian ini adalah algoritma NBC yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes. Ide dasar dari Teorema Bayes adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek [9].

Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik *data mining* untuk menggali berbagai informasi dari sebuah *database* mahasiswa, seperti untuk analisis hubungan antara jalur masuk mahasiswa baru, prestasi dan kelulusannya menggunakan algoritma K-Means [1]; *monitoring* dan evaluasi kinerja akademik mahasiswa menggunakan teknik *data mining* [8]; menerapkan *data mining* untuk membangun SPK dalam mengevaluasi dan memilih calon mahasiswa internasional yang paling memenuhi syarat [4]; prediksi kelulusan mahasiswa dengan menerapkan algoritma *Naive Bayes* dan algoritma C4.5 [11]; model prediksi kinerja akademik mahasiswa tahun pertama menggunakan NBC [2]; prediksi kinerja akademik mahasiswa dengan algoritma K-Means *clustering* yang hasilnya setelah proses *clustering* mahasiswa akan dikelompokkan ke dalam tiga kategori, yaitu kategori mahasiswa yang memiliki kinerja tinggi, sedang, dan rendah [10]; dan prediksi kinerja mahasiswa menggunakan algoritma klasifikasi *data mining* yaitu *decision tree classifier*, *neural network*, dan *nearest neighbour classifier* [5].

Karena itu, pada penelitian ini akan dikembangkan suatu sistem yang dapat mengklasifikasi kelulusan

Mujib Ridwan adalah Dosen Jurusan Teknik Informatika, UIN Maulana Malik Ibrahim Malang dan Mahasiswa Program Magister Teknik Elektro, Fakultas Teknik, Universitas Brawijaya-Malang (email: mu71bridwan@gmail.com)

Hadi Suyono adalah Dosen Teknik Elektro, Universitas Brawijaya Malang-Indonesia (email: hadis@ub.ac.id).

M. Sarosa adalah Dosen Teknik Elektro, Politeknik Negeri Malang-Indonesia (email: rmsarosa@gmail.com).

mahasiswa pada tahun ke-2 dengan melihat pola kelulusan mahasiswa beberapa periode sebelumnya.

## II. LANDASAN TEORI

### A. Data Mining

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [12]. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database (KDD)*. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [9].

#### 1. Metode Pelatihan

Secara garis besar metode pelatihan yang digunakan dalam teknik-teknik *data mining* dibedakan ke dalam dua pendekatan, yaitu [9]:

- *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.
- *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses *training*.

#### 2. Pengelompokan Data Mining

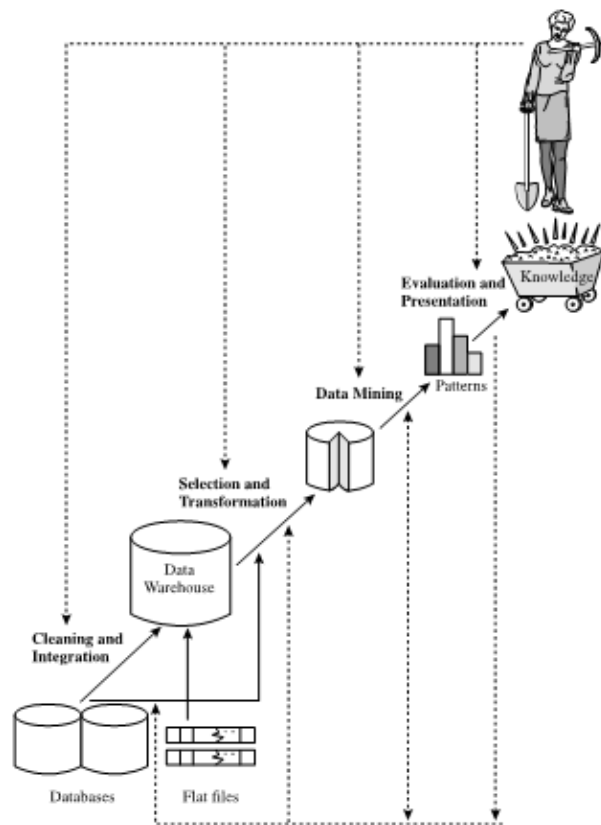
Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan, yaitu [6]:

- Deskripsi  
Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data.
- Estimasi  
Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.
- Prediksi  
Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).
- Klasifikasi  
Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

- *Clustering*  
*Clustering* lebih ke arah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.
- Asosiasi  
Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

#### 3. Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 1. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.



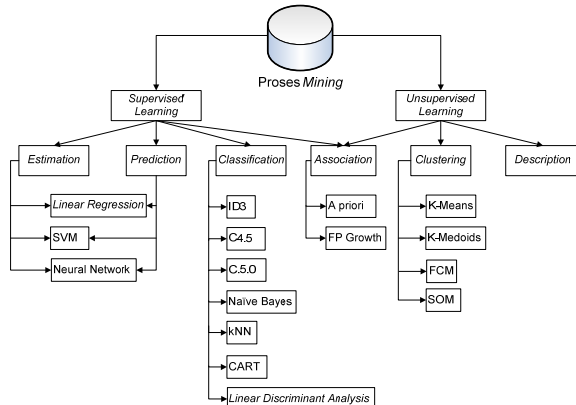
Gambar 1 Tahap-tahap Data Mining [3]

Tahap-tahap *data mining* adalah sebagai berikut:

- Pembersihan data (*data cleaning*)  
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
- Integrasi data (*data integration*)  
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- Seleksi data (*data selection*)  
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
- Transformasi data (*data transformation*)  
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

- Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *data mining* dapat dilihat pada Gambar 2.



Gambar 2 Beberapa Metode *Data Mining*

- Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

- Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

### B. Naive Bayes Classifier (NBC)

NBC merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi [9]. Teorema keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Naive bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu.

Dengan memasukkan Persamaan 1 ke Persamaan 2 akan diperoleh pendekatan yang digunakan dalam NBC.

$$P(a_1, a_2, a_3, \dots, a_n | V_j) = \prod_i P(a_i | v_j) \dots \dots \dots (1)$$

$$V_{NB} = \arg \max_{v_j \in V} P(V_j) \prod_i P(a_i | v_j) \dots \dots \dots (2)$$

dengan

$V_{NB}$  : nilai output hasil klasifikasi Naïve Bayes

$P(a_i | v_j)$  : rasio antara  $n_c / n$ , di mana  $n_c$  adalah jumlah data *training* untuk  $v = v_j$  dan  $a = a_i$ ; dan  $n$  adalah total kemungkinan output

## III. METODE PENELITIAN

### A. Data

Pada penelitian ini digunakan data sebagai berikut:

1. Data *training* dan data *testing*

Data ini akan digunakan sebagai proses *mining* dan pengujian, berupa sampel data induk dan data akademik mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus. Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orangtua, IP semester 1-4, IPK semester 1-4, dan keterangan lulus.

2. Data target

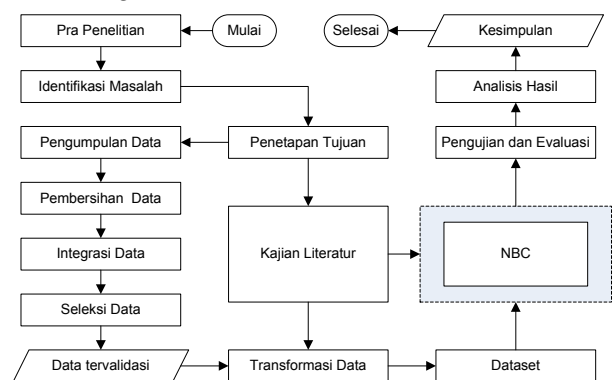
Data ini berupa sampel data induk data akademik mahasiswa angkatan 2010-2011 yang diasumsikan belum lulus. Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orangtua, IP semester 1-4, dan IPK semester 1-4. Setelah proses *mining*, data ini akan memiliki kelas berdasarkan tabel probabilitas yang diperoleh dari data *training*.

3. Data riwayat matakuliah

Data ini digunakan untuk mengevaluasi data target ketika diklasifikasikan lulus tidak tepat waktu. Data ini akan dianalisis untuk memberikan rekomendasi dalam proses perkuliahan berikutnya. Data ini memiliki atribut NIM, kode matakuliah, dan nilai.

### B. Tahapan Penelitian

Pada penelitian ini, tahapan penelitian yang dilakukan adalah seperti pada Gambar 3. Penelitian ini secara garis besar meliputi beberapa kegiatan inti yaitu pembuatan proposal, pengumpulan data, pengolahan data, implementasi NBC, pengujian, dan analisis hasil. Pada tahap pengolahan ada beberapa kegiatan sesuai dengan tahapan yang ada pada data mining, yaitu pembersihan data, integrasi data, seleksi data, transformasi data, dan pembentukan *dataset* yang dalam penelitian akan digunakan sebagai data *training* dan data *testing*.



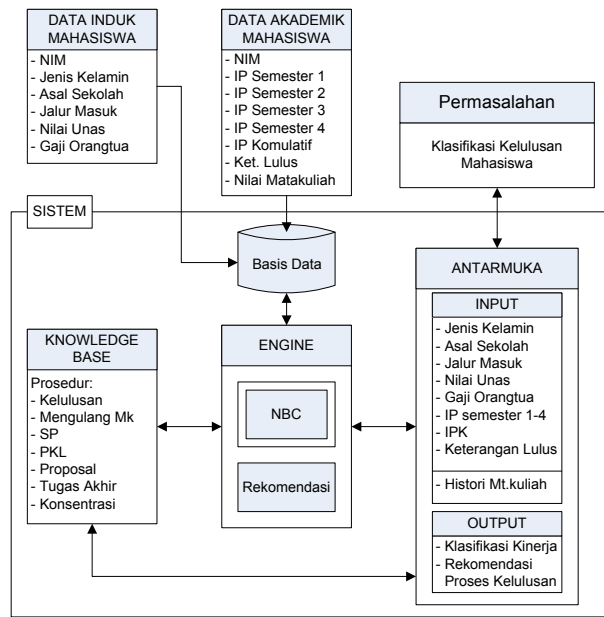
Gambar 3 Tahapan Penelitian

### C. Desain Sistem

Berikut ini adalah desain sistem yang digunakan pada penelitian ini:

1. Arsitektur Sistem

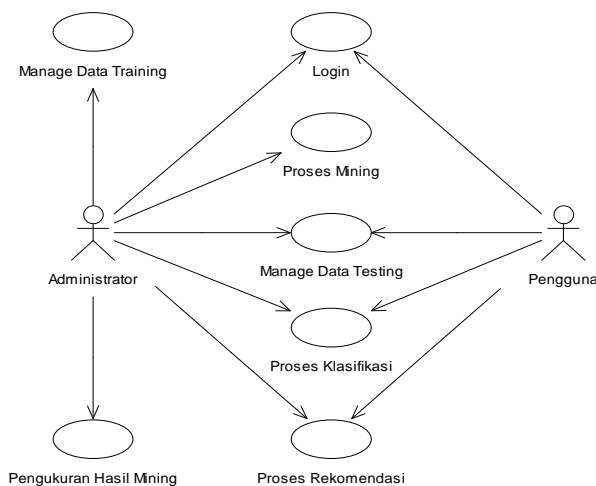
Pada bagian komponen sistem terbagi menjadi empat lingkungan, yaitu basis data, *engine*, basis pengetahuan, dan antarmuka pengguna. Arsitektur sistem dapat dilihat pada Gambar 4.



Gambar 4 Arsitektur Sistem

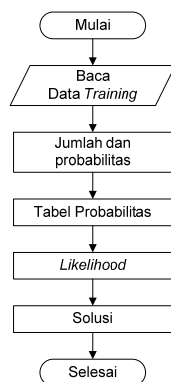
## 2. Pemodelan Sistem

Pada penelitian ini sistem dimodelkan menggunakan *use case diagram* untuk memodelkan tingkah laku (*behavior*) sistem yang akan dibuat.



Gambar 5 Use Case Diagram

## 3. Algoritma NBC



Gambar 6 Flowchart Algoritma NBC

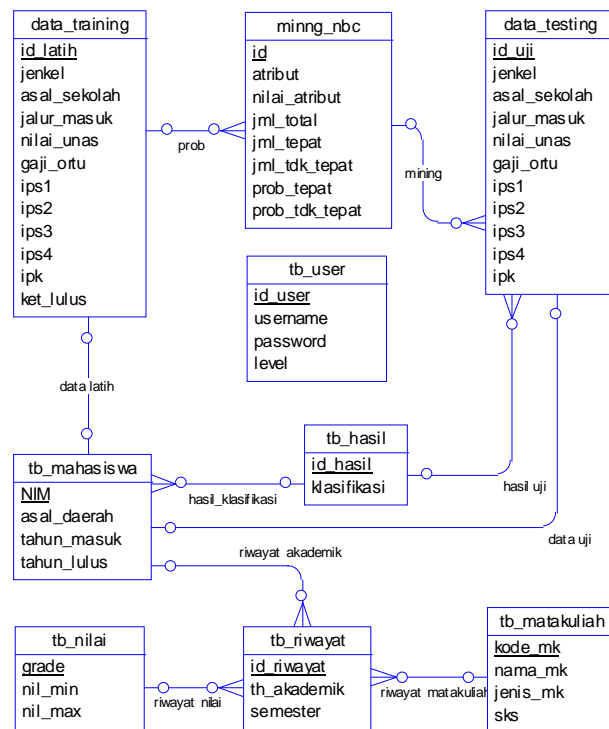
Langkah-langkah algoritma NBC pada Gambar 6 dapat diuraikan sebagai berikut:

- Baca data *training*

- Cari nilai probabilitasnya dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
- Mendapatkan nilai dalam tabel probabilitas.

## 4. Desain Basis Data

Pada desain basis data digambarkan beberapa tabel, atribut, dan relasinya yang akan digunakan sebagai penyimpanan data *training*, data *testing*, dan data target pada penelitian ini.



Gambar 7 Desain Basis Data

## IV. PENGUJIAN DAN PEMBAHASAN

### A. Persiapan Data

Pada tahap pengujian ini, data yang akan digunakan sudah dibersihkan dan ditransformasi dalam bentuk kategori. Dalam pengujian ini digunakan data sampel mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus dari perguruan tinggi. Jumlah data yang digunakan adalah 100 data dengan kelas "Tepat" dan "Tidak Tepat" masing-masing berjumlah 33 dan 67.

Dalam proses pengujian, data dibagi menjadi 2 bagian yaitu data latih dan data uji. Oleh algoritma NBC, data latih digunakan untuk membentuk tabel probabilitas, dan data uji digunakan untuk menguji tabel probabilitas yang telah terbentuk.

### B. Pengujian

Pengujian ini bertujuan untuk mengetahui unjuk kerja dari algoritma NBC dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Pada uji coba ini, diberikan data latih untuk membentuk tabel probabilitas. Langkah selanjutnya akan diberikan data uji untuk menguji tabel probabilitas yang sudah terbentuk. Unjuk kerja diperoleh dengan memberikan

nilai pada *confusion matrix* untuk menghitung nilai *precision*, *recall*, dan *accuracy* dari hasil pengujian. Berikut hasil pengujian dari beberapa percobaan:

#### 1. Percobaan ke-1

Menggunakan data latih sebanyak 20 data sampel 2005 dengan data uji sebanyak 20 set data. Setelah proses *import data training*, berikut hasil perhitungan dari 20 data *training* pada aplikasi:

NO	ALIRAN	NILAI AKHIR	JUMLAH KASUS TOTAL	JUMLAH KASUS TEPAT	JUMLAH KASUS TIDAK TEPAT	PROBABILITAS KASUS TEPAT
1	total	total	20	7	13	1
2	jawa	jawa	13	6	7	0.8571
3	jawa	jawa	7	1	6	0.1429
4	asal_jelajah	MA Negeri	7	3	4	0.4286
5	asal_jelajah	MA Swasta	4	2	2	0.5
6	asal_jelajah	MA Negeri	4	1	3	0.25
7	asal_jelajah	MA Swasta	2	0	2	0
8	asal_jelajah	MA Negeri	2	1	1	0.5
9	asal_jelajah	MA Swasta	1	0	1	0
10	asal_jelajah	MA Negeri	1	0	1	0
11	asal_jelajah	MA Swasta	1	0	1	0

Gambar 8 Perhitungan Data Training

Perhitungan pada Gambar 8 merupakan dasar pembuatan tabel probabilitas yang akan digunakan pada proses klasifikasi data *testing*.

NO	NIM	JENIS KELAMIN	ASAL SEKOLAH	JALUR MASUK	NILAI UJIAN	GAKSI ORANGTUA	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	KELULUSAN
1	09001	Laki-laki	MA Swasta	Undangan	Rendah	Rendah	A	A	C	B	B	Tidak Tepat
2	09003	Perempuan	MA Negeri	SNMPN	Tinggi	Sedang	B	B	B	B	B	Tidak Tepat
3	09007	Perempuan	MA Negeri	SNMPN	Rendah	Sedang	B	C	B	B	B	Tidak Tepat
4	09008	Laki-laki	MA Negeri	SNMPN	Rendah	rendah	B	B	B	C	B	Tidak Tepat
5	09010	Laki-laki	MA Negeri	SNMPN	Rendah	rendah	B	B	B	B	B	Tidak Tepat
6	09026	Perempuan	MA Swasta	Mandiri	Tinggi	Sedang	B	C	B	B	B	Tidak Tepat
7	09030	Laki-laki	MA Swasta	Mandiri	Tinggi	Sedang	B	B	B	B	B	Tidak Tepat
8	09042	Perempuan	MA Swasta	Mandiri	Tinggi	Sedang	B	C	B	B	B	Tidak Tepat
9	09045	Laki-laki	MA Negeri	SNMPN	Sedang	Sedang	B	B	B	B	B	Tidak Tepat
10	09059	Laki-laki	MA Negeri	Mandiri	Sedang	Sedang	A	A	B	B	A	Tepat
11	09061	Perempuan	MA Negeri	Mandiri	Sedang	Sedang	B	A	B	B	B	Tidak Tepat

Gambar 9 View Hasil Testing

Tabel 1 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix* untuk percobaan ke-5.

TABEL 1  
CONFUSION MATRIX PERCOBAAN KE-1

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	2	8
	Tidak Tepat	1	9

$$\text{Precision: } P = 2 / (2 + 1) \times 100\% = 66.66\%$$

$$\text{Recall: } R = 2 / (2 + 8) \times 100\% = 20.00\%$$

$$\text{Accuracy: } A = (2 + 9) / 20 \times 100\% = 55.00\%$$

#### 2. Percobaan ke-2

Menggunakan data latih sebanyak 40 data sampel 2005-2006 dengan data uji sebanyak 20 set data. Tabel 2 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix* untuk percobaan ke-5.

TABEL 2  
CONFUSION MATRIX PERCOBAAN KE-2

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	5	5
	Tidak Tepat	4	6

$$\text{Precision: } P = 5 / (5 + 4) \times 100\% = 55.55\%$$

$$\text{Recall: } R = 5 / (5 + 5) \times 100\% = 50.00\%$$

$$\text{Accuracy: } A = (5 + 6) / 20 \times 100\% = 55.00\%$$

#### 3. Percobaan ke-3

Menggunakan data latih sebanyak 60 data sampel 2005-2007 dengan data uji sebanyak 20 set data. Tabel 3 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix* untuk percobaan ke-5.

TABEL 3  
CONFUSION MATRIX PERCOBAAN KE-3

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	5	5
	Tidak Tepat	4	6

$$\text{Precision: } P = 5 / (5 + 4) \times 100\% = 55.55\%$$

$$\text{Recall: } R = 5 / (5 + 5) \times 100\% = 50.00\%$$

$$\text{Accuracy: } A = (5 + 6) / 20 \times 100\% = 55.00\%$$

#### 4. Percobaan ke-4

Menggunakan data latih sebanyak 80 data sampel 2005-2007 dengan data uji sebanyak 20 set data. Tabel 4 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix* untuk percobaan ke-5.

TABEL 4  
CONFUSION MATRIX PERCOBAAN KE-4

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	5	5
	Tidak Tepat	2	8

$$\text{Precision: } P = 5 / (5 + 2) \times 100\% = 71.42\%$$

$$\text{Recall: } R = 5 / (5 + 5) \times 100\% = 50.00\%$$

$$\text{Accuracy: } A = (5 + 8) / 20 \times 100\% = 65.00\%$$

#### 5. Percobaan ke-5

Menggunakan data latih sebanyak 100 data sampel 2005-2009 dengan data uji sebanyak 20 set data. Tabel 5 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix* untuk percobaan ke-5.

TABEL 5  
CONFUSION MATRIX PERCOBAAN KE-5

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	5	5
	Tidak Tepat	1	9

$$\text{Precision: } P = 5 / (5 + 1) \times 100\% = 83.33\%$$

$$\text{Recall: } R = 5 / (5 + 5) \times 100\% = 50.00\%$$

$$\text{Accuracy: } A = (5 + 9) / 20 \times 100\% = 70.00\%$$

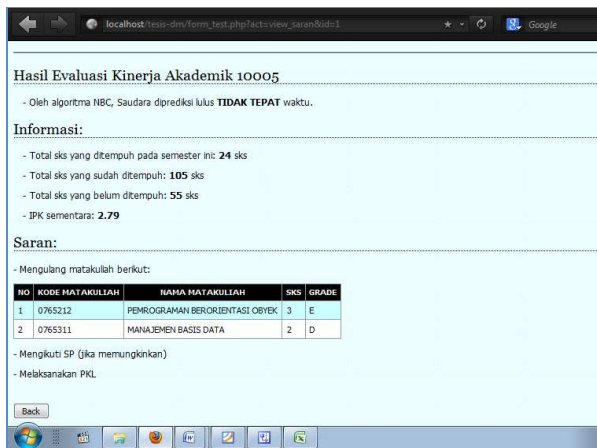
Selain pengujian pada data *testing*, juga akan dilakukan klasifikasi kelulusan pada data target. Data target berupa sampel mahasiswa angkatan 2010-2011 dan belum lulus. Data ini belum memiliki kelas seperti pada data latih dan data uji. Berikut adalah tampilan hasil klasifikasi pada data target:

NO	NIM	JENIS KELAMIN	ASAL SEKOLAH	JALUR MASUK	NILAI UJIAN	GAKSI ORANGTUA	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	KELULUSAN
1	09001	Laki-laki	MA Negeri	SNMPN	Tinggi	Sedang	B	B	C	B	B	Tidak Tepat / N/C

Gambar 10 Klasifikasi Data Target

Jika hasil klasifikasi “Tidak Tepat”, maka sistem akan memberikan rekomendasi dengan melakukan analisis pada data riwayat matakuliah yang telah ditempuh. Hasil evaluasi kinerja akademik mahasiswa meliputi informasi sks dan saran untuk proses perkuliahan pada semester berikutnya agar dapat lulus dalam waktu yang tepat.



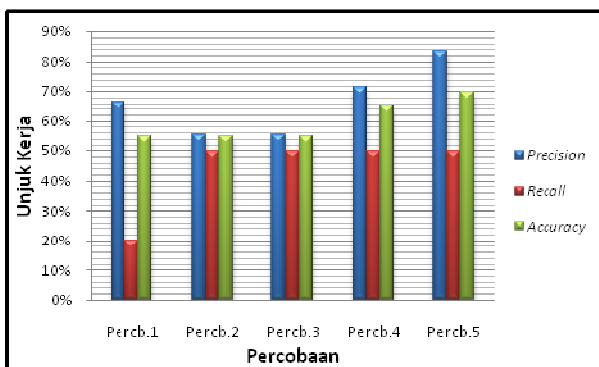


Gambar 11 Evaluasi Data Target

### C. Pembahasan

Dari hasil pengujian dapat diketahui nilai *precision*, *recall*, dan *accuracy* untuk setiap percobaan. Perhitungan rata-rata dari semua percobaan menghasilkan nilai *precision*, *recall*, dan *accuracy* masing-masing 67%, 44%, dan 60%. Nilai *precision* tertinggi dari semua percobaan yaitu 83%, sedangkan nilai *precision* terendah dari semua percobaan yaitu 56%. Untuk Nilai *recall* tertinggi dari semua percobaan yaitu 50%, sedangkan nilai *recall* terendah dari semua percobaan yaitu 20%. Nilai *accuracy* tertinggi dari semua percobaan yaitu 70%, sedangkan nilai *accuracy* terendah dari semua percobaan yaitu 55%.

Perbandingan nilai *precision*, *recall*, dan *accuracy* untuk setiap percobaan dapat divisualisasikan dalam bentuk grafik seperti pada Gambar 11. Dari grafik tersebut dapat dilihat bahwa besar kecilnya jumlah data latih tidak selalu berbanding lurus dengan nilai *precision*, *recall*, dan *accuracy* dari data uji.



Gambar 11 Grafik Hasil Pengujian

Secara umum, hasil dari pengujian menunjukkan nilai akurasi yang sedang. Hal ini disebabkan karena faktor penentu kelulusan mahasiswa tepat atau tidak tepat waktu pada kenyataannya memiliki nilai yang tidak konsisten.

Setelah proses pengujian terhadap data uji, hasil *mining* yang terbaik akan digunakan untuk mengklasifikasikan data target ke dalam kelas “tepat” atau “tidak tepat” waktu kelulusan. Mahasiswa yang menjadi data target akan dievaluasi data riwayat akademik yang telah ditempuh dan akan diberikan saran-saran meliputi saran untuk mengulang matakuliah

pada semester berikutnya atau mengikuti SP, saran untuk melaksanakan PKL, saran untuk ujian seminar, saran untuk ujian komprehensif, dan saran untuk ujian skripsi. Saran-saran ini akan diberikan jika kondisi yang disyaratkan pada proses rekomendasi terpenuhi. Misalkan untuk saran mengulang matakuliah, akan diberikan jika pada riwayat akademik ditemukan nilai matakuliah yang lebih kecil dari C, yaitu nilai D atau E.

### V. KESIMPULAN

Dari hasil penelitian dapat disimpulkan bahwa:

1. Pengujian pada data mahasiswa angkatan 2005-2009 *mining* NBC menghasilkan nilai *precision*, *recall*, dan *accuracy* masing-masing 83%, 50%, dan 70%.
2. Penentuan 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*.
3. Hasil *mining* NBC dapat digunakan untuk mengklasifikasikan kinerja akademik mahasiswa tahun ke-2 yang dalam penelitian ini dijadikan data target.

### REFERENSI

- [1] Erdogan, S.Z. dan M. Timor. 2005. A Data Mining Application In A Student Database. *Journal Of Aeronautics And Space Technologies*. Volume 2 Number 2: 53-57.
- [2] Garcia, E.P.I. dan P.M. Mora. 2011. Model Prediction of Academic Performance for First Year Students. *IEEE Computer Society*.
- [3] Han, J. dan M. Kamber. 2006. *Data Mining: Concepts and Techniques, Second Edition*. Morgan Kaufmann Publishers. San Francisco.
- [4] Hien, N.T.N. dan P. Haddawy. 2007. A Decision Support System for Evaluating International Student Applications. *IEEE Frontiers in Education Conference*.
- [5] Kabakchieva, D. 2012. Student Performance Prediction by Using Data Mining Classification Algorithms. *IJCSMR*. Vol 1 Issue 4: 686-690.
- [6] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.
- [7] McLeod, Jr.R. dan G.P. Schell. 2007. Management Information System. 10<sup>th</sup> ed. Pearson Education, Inc. Ali Akbar Yulianto dan Afia R. Fitriati (penterjemah). 2008. Sistem Informasi Manajemen. Edisi 10. Nina Setyaningsih (editor). Salemba Empat. Jakarta.
- [8] Ogor, E.N. 2007. Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. *IEEE Computer Society*.
- [9] Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [10] Shovon, Md.H.I. dan M. Haque. 2012. Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm. *IJARCSSE*. Volume 2 Issue 7: 353-355.
- [11] 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>. 13 Maret 2013.
- [12] Turban, E., J.E. Aronson dan T.P. Liang. 2005. Decision Support System and Intelligent Systems - 7<sup>th</sup> ed. Pearson Education, Inc. Pearson Education, Inc. Dwi Prabantini (penterjemah). 2005. Sistem Pendukung Keputusan dan Sistem Cerdas. Penerbit ANDI. Yogyakarta.