

# PENERAPAN DATA MINING DALAM KLASIFIKASI TINGKAT KELULUSAN MAHASISWA DI UNIVERSITAS MUHAMMADIYAH JEMBER

<sup>1</sup>Mardhia Safa (1310651125) <sup>2</sup>Dewi Lusiana, Ir. MT  
Jurusan Teknik Informatika Fakultas Teknik  
Universitas Muhammadiyah Jember  
Email : [mardhiasafa@gmail.com](mailto:mardhiasafa@gmail.com)

## ABSTRAK

Penelitian ini difokuskan untuk mengevaluasi kelulusan mahasiswa Teknik Informatika 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 2008-2012 yang sudah dinyatakan lulus akan digunakan sebagai data *training* dan *testing*. Sedangkan data mahasiswa angkatan 2014 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 kelulusan mahasiswa yang diprediksi kelulusannya dan memberikan rekomendasi untuk proses kelulusan tepat waktu atau tidak.

Hasil pengujian menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kelulusan mahasiswa yaitu Indeks Prestasi Kumulatif (IPK) dan jenis kelamin pada mahasiswa Teknik Informatika angkatan 2014 di Universitas Muhammadiyah Jember. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi. Pengujian pada data mahasiswa angkatan 2014, algoritma NBC menghasilkan nilai *precision* dan *accuracy* masing-masing 58,0%, dan 73,7%.

*Kata Kunci* — Kelulusan 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.

Pencapaian atau nilai data mining adalah istilah untuk menunjukkan suatu pencapaian tingkat keberhasilan tentang suatu tujuan karena suatu usaha belajar telah dilakukan oleh seseorang secara optimal. Usaha belajar dapat dilakukan di perguruan tinggi, karena perguruan tinggi adalah tempat untuk mendapatkan ilmu pengetahuan. Kelulusan mahasiswa dapat dilihat dari hasil proses belajar di perguruan tinggi sehingga pencapaian tingkat keberhasilan akademik mahasiswa dapat dipengaruhi mutu dari pendidikan perguruan tinggi.

Input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa. Sampel mahasiswa angkatan 2008-2012 yang sudah dinyatakan lulus akan digunakan sebagai data *training* dan *testing*. Sedangkan data mahasiswa angkatan 2014 dan belum lulus akan digunakan sebagai data target. Data input akan diproses menggunakan teknik *data mining* untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kelulusan mahasiswa. Output dari sistem ini berupa klasifikasi kelulusan mahasiswa yang diklasifikasi 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 kelulusan mahasiswa yaitu *Indeks Prestasi Kumulatif* (IPK) dan jenis kelamin. Sehingga faktor-faktor tersebut

dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi. Pengujian pada data mahasiswa angkatan 2008-2012, algoritma NBC menghasilkan nilai *precision*, dan *accuracy* masing-masing 83%, dan 70%.

## 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 *data* dalam jumlah besar. 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.

#### 1. Metode Pelatihan

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

- *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan dan tanpa ada dosen. Dosen 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 :

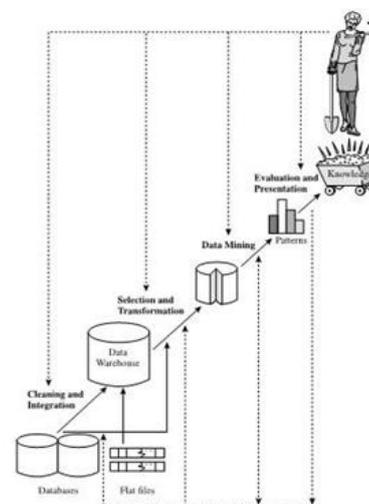
- Deskripsi  
Para peneliti biasanya mencoba

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.
- Klasifikasi  
Klasifikasi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, klasifikasi 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 2.1. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.



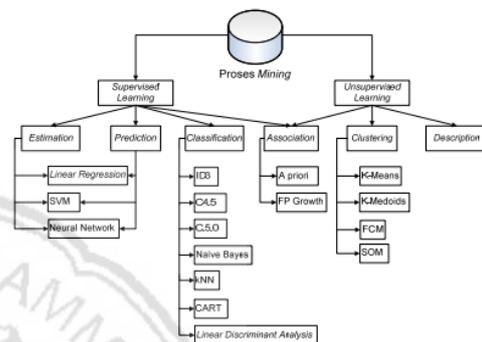
Gambar 1 Tahap-tahap *Data Mining*

Tahap-tahap *data mining* adalah sebagai berikut :

- **Pembersihan data (*data cleaning*)**  
Pembersihan data merupakan langkah menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari data suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah kompleksitasnya.
- **Integrasi data (*data integration*)**  
Integrasi data merupakan penggabungan data dari berbagai *data* ke dalam satu *data* baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu *data* tetapi juga berasal dari beberapa *data* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.
- **Seleksi data (*data selection*)**  
Data yang ada pada data sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *data*.
- **Transformasi data (*data transformation*)**  
Data di ubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering*

hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

- **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 2.2.



Gambar 2 Beberapa Metode Data Mining

- **Evaluasi pola (*pattern evaluation*)**  
Data di ubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
- **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. 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.

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)}$$

- Dengan
- $x$  : Data dengan *class* yang belum diketahui
  - $c$  : Hipotesis data merupakan suatu *class* spesifik
  - $P(c|x)$  : Probabilitas hipotesis berdasar kondisi (posteriori probabillity)
  - $P(c)$  : Probabilitas hipotesis (prior probability)
  - $P(x|c)$  : Probabilitas berdasarkan kondisi pada hipotesis
  - $P(x)$  : Probabilitas  $c$

### III. METODE PENELITIAN

#### A. Data

Pada penelitian ini digunakan data sebagai berikut :

##### 1. Data Sampel atau Acuan

Data ini akan digunakan sebagai proses *mining* dan pengujian, berupa sampel data induk dan data akademik mahasiswa angkatan 2008-2012 yang sudah dinyatakan lulus, data ini memiliki atribut jenis kelamin dan IPK.

##### 2. Data Target

Data ini berupa sampel data induk data akademik mahasiswa angkatan 2014 yang diasumsikan belum lulus. Data ini memiliki atribut jenis kelamin, dan IPK. Setelah proses *mining*, data ini akan memiliki kelas berdasarkan tabel probabilitas yang diperoleh dari data *training*.

##### 3. Data Riwayat Mata Kuliah

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

### IV. HASIL DAN PEMBAHASAN

#### A. Persiapan Data

Pada tahap pengujian ini, data yang akan digunakan sudah dibersihkan dan ditransformasikan dalam bentuk kategori. Dalam pengujian ini digunakan data sampel mahasiswa angkatan 2014 yang sudah dinyatakan lulus dari perguruan tinggi. Jumlah data yang digunakan adalah 219 data dengan kelas “Tepat” dan “Tidak Tepat”.

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*, dan *accuracy* dari hasil pengujian. Berikut hasil pengujian dari beberapa percobaan :

### Pengujian ke – 1

Menggunakan data latih sebanyak 15 data sampel dengan data uji sebanyak 25 set data. Setelah proses *import data training*, berikut hasil perhitungan dari 25 data *training* pada aplikasi :

TABEL 1  
CONFUSION MATRIX PENGUJIAN KE-1

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	11	1
	Tidak tepat	4	9

$$Precision: P = 11 / (11 + 4) \times 100\% = 73,30\%$$

$$Accuracy: A = (11 + 9) / 25 \times 100\% = 80,00\%$$

### Pengujian ke – 2

Menggunakan data latih sebanyak 25 data sampel dengan data uji sebanyak 194 set data. Tabel 2 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix*.

TABEL 2  
CONFUSION MATRIX PENGUJIAN KE-2

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	66	5
	Tidak tepat	49	74

$$Precision: P = 66 / (66 + 49) \times 100\% = 57,40\%$$

$$Accuracy: A = (66 + 74) / 194 \times 100\% = 72,20\%$$

### Pengujian ke – 3

Menggunakan data latih sebanyak 50 data sampel dengan data uji sebanyak 169 set data. Tabel 3 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix*.

TABEL 3  
CONFUSION MATRIX PENGUJIAN KE-3

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	62	4
	Tidak tepat	39	64

$$Precision: P = 62 / (62 + 39) \times 100\% = 61,40\%$$

$$Accuracy: A = (62 + 64) / 169 \times 100\% = 74,60\%$$

### Pengujian ke – 4

Menggunakan data latih sebanyak 75 data sampel dengan data uji sebanyak 144 set data. Tabel 4 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix*.

TABEL 4  
CONFUSION MATRIX PENGUJIAN KE-4

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	49	3
	Tidak tepat	33	59

$$Precision: P = 49 / (49 + 33) \times 100\% = 59,80\%$$

$$Accuracy: A = (49 + 59) / 144 \times 100\% = 75,00\%$$

### Pengujian ke – 5

Menggunakan data latih sebanyak 100 data sampel dengan data uji sebanyak 119 set data. Tabel 5 adalah perhitungan nilai *precision*, *recall*, dan *accuracy* dengan *confusion matrix*.

TABEL 5  
CONFUSION MATRIX PENGUJIAN KE-5

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	37	3
	Tidak	28	51

	tepat		
--	-------	--	--

$$\text{Precision: } P = 37 / (37 + 28) \times 100\% = 56,90\%$$

$$\text{Accuracy: } A = (37 + 51) / 119 \times 100\% = 73,90\%$$

### Pengujian ke – 6

Menggunakan data latih sebanyak 125 data sampel dengan data uji sebanyak 94 set data. Tabel 6 adalah perhitungan *nilai precision, recall, dan accuracy* dengan *confusion matrix*

TABEL 6  
CONFUSION MATRIX PENGUJIAN KE-6

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	24	3
	Tidak tepat	23	44

$$\text{Precision: } P = 24 / (24 + 23) \times 100\% = 51,10\%$$

$$\text{Accuracy: } A = (24 + 44) / 94 \times 100\% = 72,30\%$$

### Pengujian ke – 7

Menggunakan data latih sebanyak 150 data sampel dengan data uji sebanyak 69 set data. Tabel 7 adalah perhitungan *nilai precision, recall, dan accuracy* dengan *confusion matrix*

TABEL 7  
CONFUSION MATRIX PENGUJIAN KE-7

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	21	1
	Tidak tepat	15	32

$$\text{Precision: } P = 21 / (21 + 15) \times 100\% = 58,30\%$$

$$\text{Accuracy: } A = (21 + 32) / 69 \times 100\% = 76,80\%$$

### Pengujian ke – 8

Menggunakan data latih sebanyak 175 data sampel dengan data uji sebanyak 44 set data. Tabel 8 adalah perhitungan *nilai precision, recall, dan accuracy* dengan

*confusion matrix.*

TABEL 8  
CONFUSION MATRIX PENGUJIAN KE-8

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	13	2
	Tidak tepat	11	18

$$\text{Precision: } P = 13 / (13 + 11) \times 100\% = 54,20\%$$

$$\text{Accuracy: } A = (13 + 18) / 44 \times 100\% = 70,50\%$$

### Pengujian ke – 9

Menggunakan data latih sebanyak 200 data sampel dengan data uji sebanyak 19 set data. Tabel 9 adalah perhitungan *nilai precision, recall, dan accuracy* dengan *confusion matrix.*

TABEL 9  
CONFUSION MATRIX PENGUJIAN KE-9

		Predicted Class	
		Tepat	Tidak Tepat
Actual Class	Tepat	5	1
	Tidak tepat	5	8

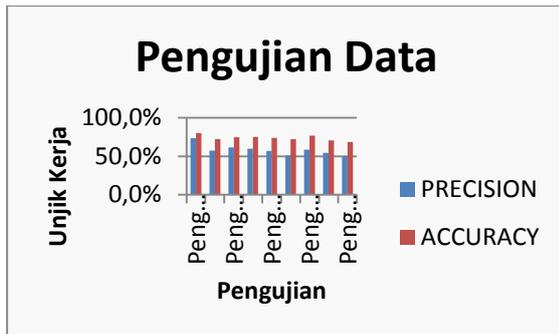
$$\text{Precision: } P = 5 / (5 + 5) \times 100\% = 50,00\%$$

$$\text{Accuracy: } A = (5 + 8) / 19 \times 100\% = 68,40\%$$

## C. Pembahasan

Dari hasil pengujian dapat diketahui nilai *precision*, dan *accuracy* untuk setiap percobaan. Perhitungan rata-rata dari semua percobaan menghasilkan nilai *precision*, dan *accuracy* masing-masing 91,1% dan 73,7%. Nilai *precision* tertinggi dari semua percobaan yaitu 95,5%, sedangkan nilai *precision* terendah dari semua percobaan yaitu 83,3%. Nilai *accuracy* tertinggi dari semua percobaan yaitu 80,0%, sedangkan nilai *accuracy* terendah dari semua percobaan yaitu 68,4%.

Perbandingan nilai *precision*, dan *accuracy* untuk setiap percobaan dapat divisualisasikan dalam bentuk grafik seperti pada Gambar 4.1. Dari grafik tersebut dapat dilihat bahwa besar kecilnya jumlah data latih tidak selalu berbanding lurus dengan nilai *precision*, dan *accuracy* dari data uji.



**Gambar 8 Grafik Hasil Penguujian**

Secara umum, hasil dari penguujian 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 penguujian 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. PENUTUP

### A. Kesimpulan

Dalam penelitian ini, dapat disimpulkan bahwa metode *naïve Bayes* dapat digunakan dalam klasifikasi data yang berupa tekstual. Hal ini berdasarkan uji coba peneliti dalam mengambil contoh perhitungan yang telah dilakukan sebelumnya. Untuk membandingkan hasil akurasi klasifikasi data laporan menggunakan algoritma *naïve bayes*.

Penguujian pada data mahasiswa angkatan 2014 *data mining* NBC menghasilkan nilai *precision* dan *accuracy* masing-masing 58,0%, dan 73,7%.

Penentuan data *training* dapat mempengaruhi hasil penguujian, karena pola data *training* tersebut akan dijadikan sebagai *rule* untuk menentukan kelas pada data *testing*. Sehingga besar atau kecilnya persentase tingkat *precision*, dan *accuracy* dipengaruhi juga oleh penentuan data *training*

### B. Saran

Penelitian yang dilakukan tentunya tak lepas dari sebuah kekurangan. Oleh karena itu, ada beberapa saran yang penulis berikan sebagai acuan dalam penelitian selanjutnya mengenai klasifikasi *text mining* yaitu :

1. Dalam tahapan pembobotan fitur, dapat dilakukan ekstraksi fitur setelah proses TF-IDF yaitu dapat menggunakan ekstraksi *Principal Component Analysis* (PCA), *Singular Value Decomposition* (SVD) atau metode lainnya yang dapat digunakan sehingga proses perhitungan fitur nilai terdapat lebih akurat.
2. Diharapkan penelitian selanjutnya dapat menggunakan algoritma *stemming* Bahasa Indonesia lain seperti algoritma *Stemming Porter* Bahasa Indonesia sehingga dapat diketahui algoritma *stemming* yang lebih optimal dengan metode *naïve bayes classifier*.
3. Pada penelitian selanjutnya, dapat menggunakan metode data mining yang lain dalam mengukur pengaruh tahap *stemming* pada dokumen Bahasa Indonesia seperti metode SVM (*Support Vector Machine*), *Decision treedan* dan sebagainya.

### DAFTAR PUSTAKA

- 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.
- Garcia, E.P.I. dan P.M. Mora. 2011. Model Prediction of Academic Performance for First Year Students. *IEEE Computer Society*.
- Han, J. dan M. Kamber. 2006. *Data Mining: Concepts and Techniques, Second Edition*. Morgan Kaufmann Publishers. San Francisco.
- Hien, N.T.N. dan P. Haddawy. 2007. A

Decision Support System for Evaluating International Student Applications. *IEEE Frontiers in Education Conference*.

- Kabakchieva, D. 2012. Student Performance Prediction by Using Data Mining Classification Algorithms. *IJCSMR*. Vol 1 Issue 4: 686-690.
- Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- 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.
- Ogor, E.N. 2007. Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. *IEEE Computer Society*.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- 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.
- 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/8162182-1-PB.pdf>. 13 Maret 2013.
- Turban, E., J.E. Aronson dan T.P. Liang. 2005. *Decision Support System and Intelligent Systems* - 7<sup>th</sup> ed. Pearson Education, Inc. Dwi Prabantini (penterjemah). 2005. *Sistem Pendukung Keputusan dan Sistem Cerdas*. Penerbit ANDI. Yogyakarta.
- Penerapan data mining untuk memprediksi klasifikasi jumlah pembaca sebuah artikel pada situs beranda.co.id menggunakan algoritma bayesian classification