**JURNAL**

# PENERAPAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI LAMA STUDI MAHASISWA PADA JURUSAN SISTEM INFORMASI SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER (STMIK) KADIRI

***IMPLEMENTATION OF NAIVE BAYES ALGORITHM TO PREDICTING LONG STUDY OF STUDENT ON DEPARTMENT OF INFORMATION SYSTEMS AT HIGH SCHOOL INFORMATION MANAGEMENT AND COMPUTERS ( STMIK ) KADIRI***

**Oleh:**

**RENDY YUNA SIMONTIKA**

**10.1.03.02.0383**

**Dibimbing oleh :**

1. **Dr. RR. Forijati, S.E., M.M**
2. **Agustono Heriadi, S.ST., M.Kom**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS NUSANTARA PGRI KEDIRI**

**2014**

Artikel Skripsi
Universitas Nusantara PGRI Kediri

# SURAT PERNYATAAN
# ARTIKEL SKRIPSI TAHUN 2014

**Yang bertanda tangan di bawah ini:**

| | |
|---|---|
| Nama Lengkap | : Rendy Yuna Simontika |
| NPM | : 10.1.03.02.0383 |
| Telepun/HP | : 082226177746 |
| Alamat Surel (Email) | : Rysimontika@gmail.com |
| Judul Artikel | : Penerapan Algoritma Naïve Bayes untuk Prediksi Lama Studi Mahasiswa Pada Jurusan Sistem Informasi Sekolah Tinggi Manajemen Informatika dan Komputer ( STMIK ) Kadiri |
| Fakultas – Program Studi | : Fakultas Teknik – Teknik Informatika |
| Nama Perguruan Tinggi | : Universitas Nusantara PGRI Kediri |
| Alamat Perguruan Tinggi | : Jl. K.H. Ahmad Dahlan No. 76, Mojoroto, Kediri, Jawa Timur 64112 |

Dengan ini menyatakan bahwa :

a. artikel yang saya tulid merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

| Mengetahui | | Kediri, 26 Januari 2017 |
|---|---|---|
| Pembimbing I | Pembimbing II | Penulis, |
| Dr. RR. Forijati, S.E., M.M<br>NIDN : 0028016701 | Agustono Heriadi, S.ST., M.Kom<br>NIDN: 9907004548 | Rendy Yuna Simontika<br>10.1.03.02.0383 |



Rendy Yuna Simontika | 10.1.03.02.0383
Fakultas Teknik – Teknik Informatika
simki.unpkediri.ac.id
|| 1||

# ENERAPAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI LAMA STUDI MAHASISWA PADA JURUSAN SISTEM INFORMASI SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER (STMIK) KADIRI

Rendy Yuna Simontika
10.1.03.02.0383
Fakultas Teknik – Teknik Informatika
Email: Rysimontika@gmail.com
Dr. RR. Forijati, S.E., M.M dan Agustono Heriadi, S.ST., M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

## ABSTRAK

Tujuan dari penelitian ini adalah membangun sebuah aplikasi guna mengetahui prediksi lama studi dari setiap mahasiswa yang telah menempuh perkuliahan minimal hingga semester empat dengan menggunakan dasar data – data kelulusan mahasiswa sebelumnya, sehingga dapat menjadi sebuah aplikasi yang mampu memberikan solusi yaitu membantu meningkatkan kinerja Program Studi Sistem Informasi di STMIK Kadiri dalam menangani masalah ketepatan waktu kelulusan mahasiswa.
Dalam penelitian ini peneliti menggunakan salah satu algoritma dari metode *data mining* yaitu algoritma *naive bayes*. *Naive bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class, atau yang lebih mudah dipahami *naive bayes* dapat diterapkan untuk memprediksi sesuatu yang akan terjadi dimasa depan dengan mempertimbangkan kejadian yang telah terjadi dimasa lampau. Pada penelitian ini peneliti akan memprediksi ketepatan kelulusan mahasiswa dengan mempertimbangkan beberapa faktor yaitu jenis kelamin, status kerja, status pernikahan, kota asal, domisili, asal sekolah, tipe sekolah dan *Indeks Prestasi Komulatif* semester satu sampai empat. Penelitian ini dilakukan di STMIK Kadiri untuk S1 Program Studi Sistem Informasi dan menggunakan 88 data *training*.

**KATA KUNCI :** Penerapan, Naive Bayes, Prediksi Lama Studi Mahasiswa.

## I. LATAR BELAKANG

Pendidikan sangat penting untuk setiap orang. Kemajuan jaman menuntut setiap orang untuk memperoleh pendidikan dengan sebaik-baiknya, seperti yang kita tahu dalam dunia kerja orang-orang yang memiliki pendidikan tinggi lebih diprioritaskan untuk mendapatkan pekerjaan yang lebih baik.Perguruan tinggi adalah satu wadah bagi mahasiswa untuk memperoleh pendidikan yang lebih tinggi dari sebelumnya. Perguruan tinggi diharapkan menyelenggarakan pendidikan yang berkualitas agar dapat mencetak mahasiswa yang berkualitas pula. Semakin banyak jumlah perguruan tinggi maka tidak menutup kemungkinan semakin meningkat pula jumlah sumber daya manusia yang berkualitas. Kemampuan mahasiswa untuk menyelesaikan studi tepat waktu kemungkinan juga dapat menjadi salah satu faktor yang mempengaruhi kualitas perguruan tinggi tersebut. Program studi hendaknya selalu memantau perkembangan studi dari mahasiswanya, seperti jumlah mata kuliah yang diambil, jumlah mata kuliah yang mengulang, IP semester 1-4, penyelesaian tanggungan pembiayaan, dan lain-lain. Hal ini perlu dilakukan guna mengantisipasi terjadinya mahasiswa yang “mangkrak” atau “mangkir” yang akan menyebabkan kinerja program studi menjadi kurang baik.

## II. METODE

a. Algoritma *Naive Bayes*

Menurut Kusrini (2009) *Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decesion tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Teorema *Bayes* memiliku bentuk umum sebagai berikut :

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

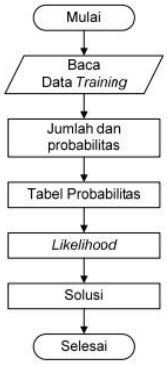
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Flowchart dari Algoritma *Naive Bayes* dapat dilihat pada gambar berikut:

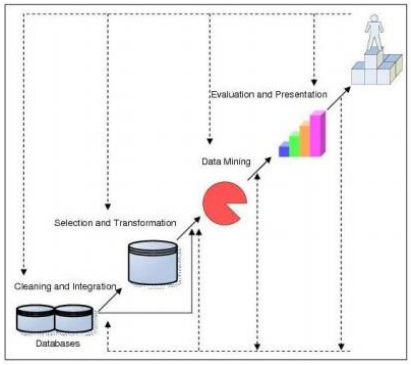



Gambar 1: Flowchart Algoritma *Naive Bayes* (Ridwan, 2013)

b. Pengertian *Data Mining*

Menurut Han dan Kamber (2001) alasan utama mengapa *data mining* diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, mulai manajemen bisnis, control produksi, kesehatan, dan lain-lain. Secara sederhana, *data mining* dapat diartikan sebagai proses mengekstrak atau "menggali" *knowledge* yang ada pada sekumpulan data. Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge Discovery in Database*, atau yang biasa disebut KDD. Dari sudut pandang yang lain, *data mining* dianggap sebagai statu langkah yang penting di dalam proses KDD. Han dan Kamber (2001) menyatakan bahwa KDD terdiri dari langkah-langkah sebagai berikut :

*Data cleaning* , *Data integration*, *Data selection*, *Data transformation*, *Data mining*, *Pattern evaluation*, dan *Knowledge presentation*.




Gambar 2: *Data Mining* adalah suatu langkah didalam proses KDD (Han dan Kamber, 2001)

c. *Database*

Menurut Kustiyaningsih (2011), "*Database* adalah Struktur penyimpanan data. Untuk menambah, mengakses dan memperoses data yang disimpan dalam sebuah database komputer,

diperlukan system manajemen database seperti *MYSQL Server*".

Database juga dapat kita simpulkan yaitu sekumpulan file yang saling berhubungan dan terorganisasi atau kumpulan *record-record* yang menyimpan data dan data tersebut saling berhubungan.

d. Sistem Informasi

Pengertian Sistem

Menurut Turban, et al (2005), "Sistem adalah kumpulan objek seperti orang, sumber daya, konsep dan prosedur yang dimaksudkan untuk melakukan suatu fungsi yang dapat diidentifikasi atau untuk melayani suatu tujuan:.

Menurut Prahasta (2005), "Sistem adalah sekumpulan objek, ide,tindakan yang berikut saling keterhubungannya (*inter-relasi*) dalam mencapai tujuan atau sasaran bersama".

e. Data dan Informasi

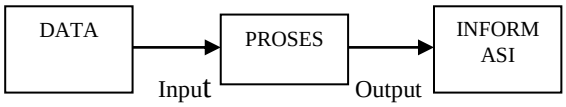
Definisi Data

Menurut Sutarman (2012), "Data adalah fakta dari sesuatu pernyataan yang berasal dari kenyataan, di mana pernyataan tersebut merupakan hasil pengukuran atau pengamatan. Data dapat berupa angka-angka, huruf-huruf, simbol-simbol khusus, atau gabungan darinya".

Definisi Informasi

Menurut Sutarman (2012), "Informasi adalah sekumpulan fakta (data) yang diorganisasikan dengan cara tertentu sehingga mereka mempunyai arti bagi si penerima".

Hubungan antara data dan informasi digambarkan sebagai berikut:




Gambar 3: Hubungan Data dan Informasi

Pengertian Sistem Informasi

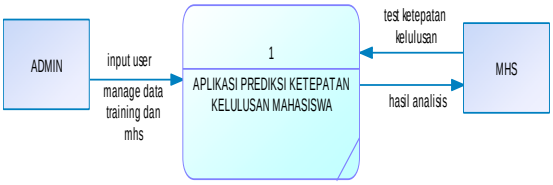
Sistem Informasi menupakan suatu sistem terpadu antar *User* dan Komputer untuk memberikan informasi untuk mendukung fungsi – fungsi operasi, manajemen, analisa dan pengambilan keputusan dalam organisasi. Sedangkan sistem informasi merupakan suatu sistem yang dibuat oleh manusia yang terdiri dari komponen – komponen dalam organisasi untuk mencapai suatu tujuan (menyajikan Informasi).

Menurut Prahasta (2005), "Sistem informasi adalah suatu sistem manusia-mesin yang terpadu untuk menyajikan informasi guna mendukung fungsi operasi, manajemen dan pengambilan keputusan organisasi".

## III. HASIL DAN KESIMPULAN

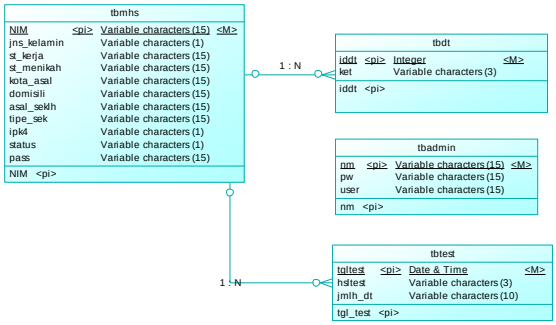
### 1. Perancangan

a. Diagram Konteks




Gambar 4: Diagram Konteks

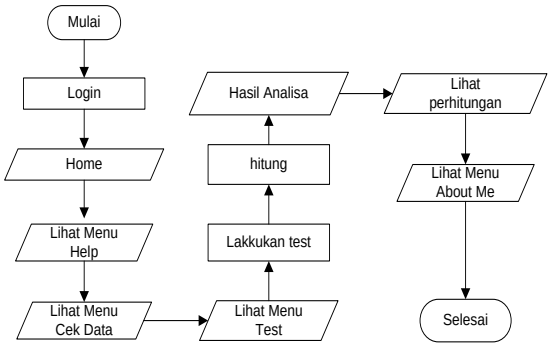
b. *Entity Relationship Diagram* Sistem




Gambar 5: *Entity Relationship Diagram*

c. *Flowchart Program*

*Flowchart system* digunakan untuk menggambarkan arus bagaimana sistem berjalan. *Flowchart* dari sistem ini adalah sebagai berikut:




Gambar 6: *Flowchart* Sistem

d. Penulisan *Indeks Prestasi Komulatif*

| IPK Asli | Penulisan |
|---|---|
| **IPK > 3.50** | A |
| **3.00 <= IPK <= 3.50** | B |
| **2.50 <= IPK < 3.00** | C |
| **IPK < 2.50** | D |

## 2. Contoh Perhitunngan Manual

Tabel 1: Data training

| JK | St. Krj | St. Pernkhn | Kota_asal | Domisili | Asal_sek | Tipe_sek | IPK semester 1-4 | Ket. Lulus |
|---|---|---|---|---|---|---|---|---|
| L | Bekerja | Single | Luar kota | Dalam kota | Luar kota | Umum | C | TTW |
| P | Tdk bekerja | Single | Dalam kota | Dalam kota | Dalam kota | Umum | B | TW |
| L | Bekerja | Single | Dalam kota | Dalam kota | Dalam kota | Kejuruan | B | TW |
| P | Bekerja | Single | Dalam kota | Dalam kota | Dalam kota | Kejuruan | B | TW |
| P | Bekerja | Menikah | Luar kota | Luar kota | Luar kota | Kejuruan | C | TTW |
| P | Tdk bekerja | Single | Luar kota | Luar kota | Luar kota | Kejuruan | B | TW |
| L | Tdk bekerja | Single | Luar kota | Dalam kota | Luar kota | Umum | D | TTW |
| L | Bekerja | Single | Dalam kota | Dalam kota | Luar kota | Kejuruan | B | TW |
| P | Bekerja | Single | Dalam kota | Dalam kota | Dalam kota | Umum | C | TTW |
| L | Bekerja | Menikah | Dalam kota | Dalam kota | Luar kota | Kejuruan | B | TW |

Ket. L : Laki-laki
P : Perempuan
TW : Tepat Waktu
Dalam Kota : Wilayah Kediri
Luar Kota : Bukan Wilayah Kediri

Tabel 2: Data Testing

| JK | St. Krj | St. Pernkhn | Kota_asal | Domisili | Asal_sek | Tipe_sek | IPK Sem1-4 | Ket. Lulus |
|---|---|---|---|---|---|---|---|---|
| L | Tdk bekerja | Single | Luar kota | Luar kota | Luar kota | Umum | B | ??? |

**Tahap 1.** Menghitung jumlah *class*

P(X= TW) = 6/10 => jumlah data "TW" pada kolom "KET. LULUS" dibagi jumlah data

P(X= TTW) = 4/10 => jumlah data "TTW" pada kolom "KET. LULUS" dibagi jumlah data

---

**Tahap 2.** Menghitung jumlah kasus yang sama dengan class yang sama

P(Jen_Kel = L | X= TW) = 3/6 => jumlah data jenis kelamin "L" dengan keterangan "TW" dibagi jumlah data TW

P(Jen_Kel = L | X= TTW) = 2/4 => jumlah data jenis kelamin "L" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(St. krj = Tdk bebekerja | X= TW) = 2/6 => jumlah data dengan status "Tdk bebekerja" dengan keterangan "TW" dibagi jumlah data TW

P(St. krj = Tdk bebekerja | X= TTW) = ¼ => jumlah data dengan status "Tdk bebekerja" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(St. Pernkhn = Single | X= TW) = 5/6 => jumlah data dengan status pernikahan "Single" dan keterangan "TW" dibagi jumlah data TW

P(St. Pernkhn = Single | X= TTW) = 2/4=> jumlah data dengan status pernikahan "Single" dan keterangan "TTW" dibagi jumlah data TTW

---

P(Kota_asal = Luar kota | X= TW) = 1/6 => jumlah data Kota _asal "Luar kota" dengan keterangan "TW" dibagi jumlah data TW

P(Kota_asal = Luar kota | X= TTW) = 3/4=> jumlah data Kota _asal "Luar kota" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(Domisili = Luar kota | X= TW) = 1/6 => jumlah data Domisili "Luar kota" dengan keterangan "TW" dibagi jumlah data TW
P(Kota_asal = Luar kota | X= TTW) = ¼ => jumlah data Domisili "Luar kota" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(Asal_sek= Luar kota | X= TW) = 3/6 => jumlah data Asal_sek "Luar kota" dengan keterangan "TW" dibagi jumlah data TW
P(Asal_sek = Luar kota | X= TTW) = 3/4=> jumlah data Asal_sek "Luar kota" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(Tipe_sek= Umum | X= TW) = 1/6 => jumlah data Tipe_sek "Umum" dengan keterangan "TW" dibagi jumlah data TW
P(Tipe_sek= Umum | X= TTW) = 3/4 => jumlah data Tipe_sek "Umum" dengan keterangan "TTW" dibagi jumlah data TTW

---

P(IPK semester 1-4 = B | X= TW) = 6/6 => jumlah data IPK semester 1-4 "B" dengan keterangan "TW" dibagi jumlah data TW

P(IPK semester 1-4 = B | X= TTW) = 0/4 => jumlah data IPK semester 1-4 "B" dengan keterangan "TTW" dibagi jumlah data TTW

**Tahap 3.** Kalikan semua hasil variable TW dan TTW

P ((Jen_Kel = L), (St. Krj = Tdk bebekerja), (St. Pernkhn = Single), (Kota_asal = Luar kota ), (domisili = Luar kota), (Asal_sek = Luar kota), (Tipe_sek = Umum), (IPK semester 1-4 = B) | TW)
= {P(P(Jen_Kel=L|X =TW) * P(St. Krj=Tdk bebekerja|X= TW) * P St. Pernkhn=Single|X=TW) * P(Kota_asal = Luar kota | X= TW) * P(domisili = Luar kota | X= TW) * P(Asal_sek = Luar kota | X= TW) * P(Tipe_sek = Umum | X= TW) * P(IPK semester 1-4 = B | X= TW)}
= 6/10 * 3/6 * 2/6 * 5/6 * 1/6 * 1/6 * 3/6 * 1/6 * 6/6 = 3240/16796160
= 1,929012345679012e-4

---

P ((Jen_Kel = L), (St. Krj = Tdk bebekerja), (St. Pernkhn = Single), (Kota_asal = Luar kota ), (domisili = Luar kota), (Asal_sek = Luar kota), (Tipe_sek = Umum), (IPK semester 1-4 = B) | TTW)
= {P(P(Jen_Kel=L|X =TTW) * P(St. Krj=Tdk bebekerja|X= TTW) * P St. Pernkhn=Single|X=TTW) * P(Kota_asal = Luar kota | X= TTW) * P(domisili = Luar kota | X= TW) * P(Asal_sek = Luar kota | X= TTW) * P(Tipe_sek = Umum | X=

TTW) * P(IPK semester 1-4 = B | X= TTW)}

= 4/10 * 2/4 * 1/4 * 2/4 * 3/4 * 1/4 * 3/4 * 3/4 * 0/4

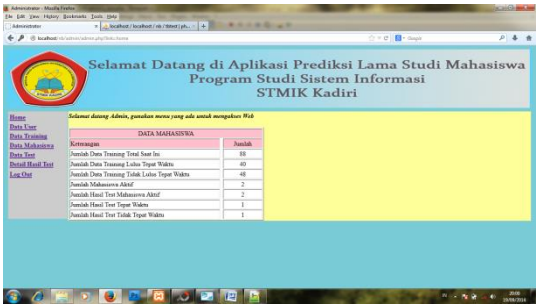
= 0

**Tahap 4**. Bandingkan hasil class TW dan TTW

Karena hasil (P|TW) lebih besar dari (P|TTW) maka keputusannya adalah "TW atau TEPAT WAKTU"

### 3. Desain *Interface*

Gambar 7: Tampilan Login



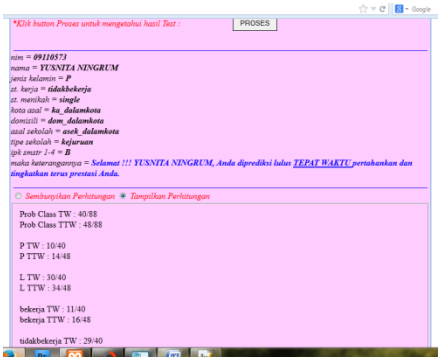

Gambar 8 :Tampilan Halaman Utama Admin

Gambar 9: Tampilan Menu Utama Web

Gambar 10: Tampilan Menu Cek Data

Gambar 11: Tampilan Menu Test

Gambar 12 : Tampilan hasil prediksi ketepatan kelulusan




Gambar 13: Tampilan perhitungan

## 4. Kesimpulan

Berdasarkan uraian bab-bab sebelumnya, maka dapat diambil kesimpulan dari penelitian ini adalah sebagai berikut :

a. Prediksi lama studi mahasiswa dapat dilakukan menggunakan teknik data mining melalui metode *naive bayes,* dimana dalam metode ini mengacu pada kejadian lampau untuk memprediksi kejadian yang akan datang.
b. Aplikasi prediksi lama studi mahasiswa menggunakan metode *naive bayes* diterapkan guna membantu meningkatkan kinerja Program Studi Sistem Informasi di STMIK Kadiri dalam menangani masalah ketepatan waktu kelulusan mahasiswa. Metode ini menetapkan beberapa kriteria sebagai acuan perhitungan dan pada penelitian ini peneliti menggunakan beberapa kriteria, yaitu jenis kelamin, status kerja, status menikah, kota asal, domisili, asal sekolah, tipe sekolah serta IPK semester satu sampai empat sebagai kriterianya.

## 5. Saran

Aplikasi prediksi lama studi mahasiswa yang dibangun masih memiliki keterbatasan, beberapa hal yang dapat di kembangkan adalah :

a. Diperlukan kriteria-kriteria lain sebagai acuan perhitungan, agar perhitungan yang dihasilkan akan semakin baik.
b. Penggunaan data training yang lebih banyak dan bervariasi juga dapat mempengaruhi hasil perhitungan menjadi lebih baik lagi.
**c.** Pengembangan aplikasi prediksi lama studi mahasiswa ini dapat dikembangkan untuk program studi yang lain, sehingga dapat memberikan manfaat yang lebih luas.

## IV. DAFTAR PUSTAKA

[1] Efraim Turban, Jay E. Aronson, Ting-Peng Liang. 2005, *Decision Support Systems and Intelligent Systems Edisi 7*. Yogyakarta : Andi.

[2] Han, J., Kamber, M. 2001. *Data Mining: Concept and Techniques.* New York: Morgan Kaufmann.

[3] Hastuti, K. 2012. *Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Preediksi Mahasiswa Non Aktif.* Makalah disajikan dalam Seminar Nasional Teknologi Informasi & Komunikasi Terapan, Universitas Dian Nuswantoro, Semarang, 23 Juni 2012. (Online), tersedia: http://eprints.dinus.ac.id/98/, diunduh 6 Nopember 2013.

[4] Jananto, A. 2013. Algoritma Naive Bayes untuk Mencari perkiraan Waktu Studi Mahasiswa. *Jurnal Teknologi Informasi*, (Online) 18 (1): 09-16, tersedia: http://www.unisbank.ac.id/ojs/index.php/fti/article/download/1669/603, diunduh 6 Nopember 2013.

[5] Kusrini, Luthfi, E. T. 2009. *Algoritma Data Mining.* Yogyakarta: Andi Offset.

[6] Kustiyaningsih, Y. 2011. *Pemrograman Basis Data Berbasis Web Menggunakan PHP & MySQL.* Yogyakarta: Graha Ilmu.

[7] Oktavian, D. P. 2010. *Menjadi Programer Jempolan Menggunakan PHP.* Yogyakarta: MediaKom.

[8] Prahasta, E. 2005. *Konsep-konsep Dasar Sistem Informasi Geografis*. Bandung : Informatika

[9] Raharjo, B. 2011. *Belajar Otodidak Membuat Database Menggunakan MySql.* Bandung: Informatika.

[10] Ridwan, M., Suyono, H. & Sarosa, M. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, (Online) 7 (1): 59-64, tersedia: http://jurnaleeccis.ub.ac.id/index.php/eeccis/article/view/204, diunduh 6 Nopember 2013.

[11] Sutarman. 2012. *Pengantar Teknologi Informasi*. Jakarta: Bumi Aksara.