

# Perkiraan Waktu Studi Mahasiswa Menggunakan Metode Klasifikasi Dengan Algoritma Naive Bayes

Lestari Handayani<sup>1</sup>, Eka Lona Maulida<sup>2</sup>

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau  
Jl. H.R. Soebrantas No. 155 KM. 18 Simpang Baru, Pekanbaru 28293  
email : lestari.handayani@uin-suska.ac.id<sup>1</sup>, ekalona.maulida@gmail.com<sup>2</sup>

## Abstrak

Permasalahan lama studi setiap mahasiswa biasanya disebabkan atau dipengaruhi oleh banyak faktor, sehingga perlu dibuat sistem prediksi kelulusan mahasiswa untuk mengetahui faktor apa saja yang dapat mempengaruhi kelulusan mahasiswa. Sistem yang dirancang adalah sistem prediksi tingkat kelulusan mahasiswa menggunakan metode data mining dengan algoritma naive bayes. Kriteria yang digunakan dalam memprediksi tingkat kelulusan adalah IPK, nim, jenis kelamin, tempat lahir, asal sekolah, kota asal sekolah, angkatan, tanggal dan tahun lulus. Sistem ini dirancang dan dibangun menggunakan bahasa pemrograman PHP dan database MySql. Data training yang digunakan berjumlah 150 data. Setelah dilakukan analisa dan pengujian, tingkat akurasi sistem dari data dalam training berjumlah 25 data, diperoleh tingkat akurasi data sebesar 88% dan laju error data sebesar 12% sedangkan dari data luar training berjumlah 25 data, diperoleh tingkat akurasi data sebesar 92% dan laju error data sebesar 8%. Setelah dilakukan pengujian diperoleh hasil mahasiswa dengan IPK diatas 3 diprediksi lulus tepat waktu. IPK sangat berpengaruh terhadap sistem prediksi sehingga tingkat akurasi dari dalam data training lebih kecil dibandingkan akurasi diluar data training.

**Kata kunci:** data mining, Naive Bayes, waktu studi.

## Abstract

The problems of time studi are caused by many factors. So it needs to be made graduation prediction system to determine what factors can affect student graduation. The system is designed to prediction of student graduation rate using data mining with Naive Bayes algorithm. The criteria used in predicting graduation rates are GPA, nim, sex, and place of birth, home school, home town school, class, date and year of graduation. This system was designed and built using the PHP programming language and MySQL database. Training data used 150 data. After analysis and testing, the accuracy of system using 25 data testing from data training amounted 88%. The accuracy of system using 25 data testing from out of data training amounted 92%. A result of analysis is students with a GPA above 3 predicted to graduate on time. GPA is very influential on the prediction system so that the accuracy of the training data is smaller than the accuracy of the data outside the training.

**Key words:** data mining, MySql, Naive Bayes, PHP, students, time study.

## 1. Pendahuluan

Teknologi komputasi dan media penyimpanan telah memungkinkan manusia untuk mengumpulkan dan menyimpan data dari berbagai sumber dengan jangkauan yang amat luas. Meskipun teknologi basis data modern telah menghasilkan media penyimpanan yang besar, teknologi untuk membantu menganalisis, memahami, atau bahkan memvisualisasikan data belum banyak tersedia. Hal inilah yang melatarbelakangi dikembangkannya konsep data mining.

Lama studi dari mahasiswa sangatlah penting bagi mahasiswa itu sendiri dan tentunya bagi program studi dalam rangka untuk menentukan langkah-langkah yang diperlukan agar perjalanan studi mahasiswa menjadi lancar. Permasalahan lama studi setiap mahasiswa biasanya disebabkan atau dipengaruhi oleh banyak faktor, sehingga perlu dibuat sistem prediksi kelulusan mahasiswa untuk mengetahui faktor apa saja yang dapat mempengaruhi kelulusan mahasiswa.

Ada beberapa penelitian yang telah diteliti sebelumnya tentang prediksi kelulusan mahasiswa diantaranya menurut Huda[1] dalam penelitiannya “aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro” dengan menggunakan metode data mining dengan algoritma apriori menyebutkan bahwa data proses masuk, asal sekolah, kota asal dan program studi menjadi pertimbangan dalam menentukan tingkat kelulusan. Sedangkan menurut Meinanda [2] dalam penelitiannya tentang prediksi masa studi sarjana dengan menggunakan metode *artificial neural network* menyebutkan bahwa masa studi seorang sarjana dipengaruhi oleh IPK, jumlah mata kuliah yang diambil, jumlah mata kuliah yang mengulang, jumlah mata kuliah tertentu.

Salah satu metode yang dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa adalah metode klasifikasi Naïve Bayes. Beberapa penelitian tentang metode klasifikasi *Naive Bayes* diantaranya adalah, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classification*” yang memiliki tingkat keberhasilan 70%, sedangkan algoritma *Naive Bayes* untuk perkiraan waktu studi mahasiswa memiliki tingkat keberhasilan 80%. Berdasarkan kajian penelitian tingkat kelulusan mahasiswa menggunakan metode Naive Bayes memiliki akurasi tinggi dibandingkan metode-metode lainnya. Untuk itu peneliti menggunakan metode Naïve Bayes untuk meneliti kelulusan mahasiswa dengan kriteria yang digunakan diantaranya nama, nim, jenis kelamin, asal sekolah, kota asal sekolah, tempat lahir, IPK, angkatan, tanggal lulus, tahun lulus, kategori kelulusan (tepat waktu, tidak tepat waktu). Tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui prediksi lama masa studi dari setiap mahasiswa yang telah menempuh kuliah hingga minimal semester IV dengan menggunakan dasar data kelulusan sebelumnya.

Adapun beberapa manfaat yang diharapkan pada penelitian ini adalah dengan mengetahui prediksi dari lama masa studi, maka dapat diambil manfaat bagi mahasiswa khususnya untuk memperhatikan perjalanan studinya dan umumnya bagi program studi untuk melakukan tindakan preventif berkaitan dengan lama studi setiap mahasiswa tersebut.

## 2. Metodologi Penelitian

Menurut Gartner Group, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan[3].

Penelitian ini menggunakan teknik dalam metode mining dengan beberapa tahapan penelitian sebagai berikut:

### a. Identifikasi Masalah.

Mengidentifikasi masalah-masalah yang relevan dengan penelitian ini yaitu bagaimana cara mencari perkiraan waktu studi mahasiswa.

### b. Pengumpulan Data

Menyiapkan data-data yang dibutuhkan, yaitu data berkaitan dengan identitas pribadi dari mahasiswa dan nilai matakuliah yang telah ditempuh oleh mahasiswa hingga menyelesaikan seluruh sks yang telah ditentukan dan telah dinyatakan lulus. Data nilai matakuliah yang telah ditempuh hanya diambil sampai dengan semester IV dan kemudian dihitung indeks prestasinya (IPK) secara kumulatif sampai dengan semester IV. Data akademik mahasiswa di dapat dari bagian akademik jurusan teknik informatika UIN SUSKA. Dan data kelulusan di dapat dari Koordinator TA. Data yang di ambil merupakan data mahasiswa tahun 2006 sampai 2009 yang sudah dinyatakan lulus, yang nantinya data tersebut merupakan data training. Data mahasiswa tahun 2010 sampai 2011 yang diasumsikan belum lulus merupakan data target.

Metode yang digunakan dalam pengambilan data tersebut adalah secara random atau acak dari setiap tahun mahasiswa.

Sumber Data: data yang digunakan dalam penelitian ini terdiri dari beberapa sumber data, yaitu data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan.

Setelah seluruh data terkumpul, selanjutnya dilakukan tahapan proses data mining, yaitu (Han et al., 2012, p7):

1. Pembersihan data (*cleaning*).

Setelah data di peroleh, maka akan ditemukan data-data yang tidak lengkap. Untuk itu diperlukan tahapan pembersihan data yaitu membuang data-data yang tidak lengkap.

2. Integrasi Data ( *data integration*).  
Integrasi data yaitu penggabungan data.
3. Seleksi Data.  
Seleksi data, yaitu menyeleksi data-data yang akan digunakan.
4. Transformasi data.  
Transformasi data yaitu perubahan data atau penggabungan data ke dalam format yang sesuai untuk diproses.
4. Proses Mining.  
Proses mining yaitu merupakan proses utama saat metode diterapkan untuk menemukan hasil prediksi kelulusan mahasiswa.

#### c. Konversi Data

Untuk mempermudah dalam proses mining selanjutnya maka dilakukan proses konversi kedalam bentuk yang dapat diolah dengan alat bantu data mining yang ada. Selanjutnya dengan menggunakan alat bantu MySql dibuatlah sebuah prosedur yang dapat digunakan untuk menentukan kelas(prediksi) dari sebuah record data baru dengan berdasar data training tersebut.

#### d. Analisa

1. Analisa permasalahan.  
Berkaitan dengan mengidentifikasi kebutuhan dalam suatu penelitian. Akan dijelaskan mengenai permasalahan yang terjadi dan bagaimana tahap penyelesaian masalah tersebut.
2. Analisa kebutuhan data  
Akan di jelaskan mengenai data-data apa saja yang diperlukan dalam pembuatan sistem ini.
3. Analisa metode naïve bayes

Teorema Bayes memiliki bentuk umum sebagai berikut [4]:

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

Keterangan rumus:

- X = Data dengan class yang belum diketahui
- C<sub>i</sub> = Hipotesis data X merupakan suatu class spesifik
- P(C<sub>i</sub>|X) = Probabilitas hipotesis C<sub>i</sub> berdasarkan kondisi x (posteriori prob.)
- P(C<sub>i</sub>) = Probabilitas hipotesis C<sub>i</sub> (prior prob.)
- P(X|C<sub>i</sub>) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Untuk menentukan atau memprediksi pilihan kelas, digunakan peluang maksimal dari seluruh kelas dengan fungsi:

$$\text{argmax } C_i = \frac{P(X|C_i) P(C_i)}{P(X)} \quad (2)$$

Karena nilai P(X) konstan untuk semua kelas, maka P(X) dapat diabaikan sehingga menghasilkan fungsi:

$$F_c(X) = \text{argmax } C_i = P(X|C_i) P(C_i) \quad (3)$$

Keterangan rumus :

- F<sub>c</sub>(X) = output hasil klasifikasi naive bayes argmax
- C<sub>i</sub> = nilai mak dari kelas
- P(C<sub>i</sub>) = peluang dari kategori yang diberikan
- P(X|C<sub>i</sub>) = kategori pada suatu class.

Setelah kedua peluang diatas dicari nilainya masing-masing, maka kemudian dua nilai peluang tersebut dipilih yang mempunyai nilai yang lebih besar menjadi kesimpulan kelas yang dicari dari data yang diprediksi.

Adapun tahapan algoritma Naive bayes dalam sistem adalah :

- a. Mulai

- b. Baca data training
  1. Hitung  $P(C_i)$  untuk setiap kelas
  2. Hitung  $P(X|C_i)$  untuk setiap kriteria dan setiap kelas
  3. Cari  $P(X|C_i)$  yang paling besar menjadi kesimpulan
- c. Tampilkan hasil prediksi
- d. Selesai

**e. Perancangan Perangkat Lunak**

1. Perancangan Subsistem Basis Data

Setelah menganalisa sistem yang akan dibuat, maka tahap selanjutnya adalah analisa dan perancangan basis data yang telah dilakukan untuk melengkapi komponen system menggunakan Entity Relational Diagram (ERD).

2. Perancangan Subsistem Model

Didalam perancangan subsistem model terdapat perancangan model untuk memprediksi lama studi mahasiswa yang digambarkan menggunakan *Flowchart* (Diagram alir), Diagram Flow Data (DFD). Model yang digunakan penulis adalah metode klasifikasi dengan algoritma naive bayes.

**f. Implementasi dan Pengujian**

1. Implementasi

Pada proses implementasi ini akan dilakukan pembuatan modul-modul yang telah dirancang dalam tahap perancangan ke dalam bahasa pemrograman. Implementasi sistem akan dilakukan dengan spesifikasi sebagai berikut:

Perangkat Lunak yang digunakan sebagai berikut :

- Sistem operasi : *Windows 7 Ultimate*
- Bahasa pemrograman : PHP versi 5.3.8
- *Tools* perancangan : Notepad ++ versi 5.8.2
- *Web server* : *Apache Web Server Version 2.2.21*
- *Database* : *MySQL Database Version 5.0.8*

Perangkat keras yang akan digunakan dalam pembuatan dan penerapan aplikasi ini:

- Processor : Intel Pentium
- Memori (RAM) : 2.00 GB

2. Pengujian

Pengujian merupakan tahapan dimana aplikasi akan dijalankan, tahap pengujian diperlukan untuk menjadi ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan. Tahap pengujiannya yaitu:

- i. Pengujian *black box*, pengujian ini berfokus pada tingkah laku aplikasi yang telah dirancang. Pengujian ini bertujuan untuk memastikan sistem yang telah dibangun sesuai dengan analisa dan perancangan.
- ii. Pengujian tingkat akurasi sistem yang mana terdiri dari tingkat akurasi sistem dalam data training dan tingkat akurasi sistem di luar data training. Akurasi diukur menggunakan rumus akurasi dan laju eror di bawah ini [5].

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{Jumlah data yang diprediksi}} \times 100\% \quad (4)$$

$$\text{Error} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{Jumlah data yang diprediksi}} \times 100\% \quad (5)$$

**g. Kesimpulan dan Saran**

Dalam tahapan ini dapat ditentukan kesimpulan terhadap hasil pengujian yang telah dilakukan untuk mengetahui apakah implementasi sistem yang telah dilakukan dapat beroperasi dengan baik dan sesuai dengan tujuan yang diinginkan serta memberikan saran-saran untuk penyempurnaan dan pengembangan penelitian selanjutnya.

### 3. Hasil dan Analisis

Pada bab ini akan dijelaskan hasil dari analisis masalah, analisis data, hasil analisis metode klasifikasi Naïve Bayes untuk prediksi kelulusan mahasiswa, dan perancangan sistem serta hasil implementasi, terakhir pengujian dan analisis hasil pengujian.

#### 3.1. Analisis Masalah

Setiap perguruan tinggi baik negeri maupun swasta tentunya memiliki kebijakan kurikulum yang berbeda. Di UIN Suska Riau khususnya pada program studi S1 Teknik Informatika tingkat kelulusan masih terbilang rendah hal ini disebabkan oleh beberapa faktor, yang diantaranya: kurangnya strategi mahasiswa dalam belajar, sistem pengajaran yang kurang baik dan adanya perbedaan kemampuan setiap mahasiswa. Dari beberapa faktor tersebut menyebabkan banyak mahasiswa yang menempuh masa studi lebih dari lima tahun. Untuk mengatasi permasalahan tersebut maka penulis membuat solusi mengimplementasikan *data mining* untuk memprediksi tingkat kelulusan mahasiswa. Dengan memanfaatkan sistem *data mining* khususnya algoritma klasifikasi Naïve Bayes dibuat sebuah perancangan aplikasi untuk memprediksi masa studi mahasiswa berdasarkan data pribadi mahasiswa dan data nilai (IPK) mahasiswa, dengan tiga kategori yaitu cepat, sedang dan lambat. Dengan adanya aplikasi ini diharapkan dapat dimanfaatkan dosen dan staff jurusan untuk memprediksi kelulusan mahasiswa dalam masa studinya.

#### 3.2. Analisis Data

Data yang digunakan dalam penelitian ini terdiri dari beberapa sumber data, yaitu data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan. Setelah seluruh data terkumpul, selanjutnya dilakukan proses tahapan data mining yaitu :

1. Pembersihan data (*cleaning*).

Setelah data di peroleh, maka akan ditemukan data-data yang tidak lengkap. Untuk itu diperlukan tahapan pembersihan data yaitu membuang data-data yang tidak lengkap.

2. Integrasi Data (*data integration*).

Setelah data di *Cleaning*, maka proses selanjutnya adalah integrasi data yaitu penggabungan data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan.

3. Seleksi Data.

Proses selanjutnya adalah seleksi data, yaitu menyeleksi data-data yang akan digunakan. Data yang digunakan adalah nama, nim, jenis kelamin, asal sekolah, kota asal sekolah, tempat lahir, IPK, angkatan, tanggal SKL.

Tabel 1. Hasil Seleksi Data

No	NIM	Nama	Jenis Kelamin	Tahun Angkatan	IPK	Kota Lahir	Sekolah	Kota Sekolah	Tanggal SKL
1	10651004284	Aidil Amal	laki-laki	2007	2	Bangkinang	SMU	Bangkinang	10-12-2014
2	10651004291	Bayu Hanif Pratama	laki-laki	2006	3	Pekanbaru	SMK	Pekanbaru	11-07-2013
3	10651004294	Eka Ramian Putra	laki-laki	2006	2	Duri	SMU	Duri	11-07-2013
4	10651004301	Anita Febriani	perempuan	2006	3	Bangkinang	SMU	Pekanbaru	15-07-2011
5	10651004302	Jhansen Risnaldo	laki-laki	2006	2	Padang	SMU	Pekanbaru	12-07-2013
6	10651004329	Bobby Handoko	laki-laki	2006	2	Pekanbaru	SMU	Pekanbaru	11-07-2013

4. Transformasi data.

Setelah data selesai diseleksi, selanjutnya dilakukan transformasi data yaitu perubahan data atau penggabungan data ke dalam format yang sesuai untuk diproses.

A. Data IPK

Transformasi nilai IPK dilakukan dengan membuat range nilai IPK yaitu :

IPK ASLI	IPK BARU
IPK > 3,00	3
2,00 ≤ IPK ≤ 3,00	2
IPK < 2,00	1

B. Data Jenis Kelamin

Data jenis kelamin, dikarenakan hanya berisi dua nilai maka tidak dilakukan

- konversi.
- C. Data Kota Lahir  
Data kota lahir dikelompokkan hanya menjadi dua nilai yaitu kota lahir yang berasal dari Pekanbaru dikonversikan menjadi 'DALAM KOTA' dan yang berasal dari luar Pekanbaru dikonversikan menjadi 'LUAR KOTA'.
- D. Data Kota Sekolah  
Data kota sekolah dikelompokkan menjadi dari dalam kota Pekanbaru atau dari luar kota Pekanbaru. Untuk yang kota sekolahnya adalah Pekanbaru maka dikonversikan datanya menjadi 'DALAM KOTA' sedangkan selain yang kota sekolahnya 'Pekanbaru' dikonversikan menjadi 'LUAR KOTA'.
- E. Data Tipe Sekolah  
Data tipe sekolah dilakukan pengelompokan yaitu dari sisi tipe sekolahnya. Untuk sekolah berkategori SMU atau SMA dikonversikan menjadi 'Umum' sedangkan selain SMU atau SMA dikonversikan menjadi 'Kejuruan'. Setelah selesai tahapan mining, diperoleh data sebanyak 150 data.
5. Proses Mining.  
Setelah keempat tahapan di atas selesai, maka tahapan terakhir adalah proses mining yaitu merupakan proses utama saat metode diterapkan untuk menemukan hasil prediksi kelulusan mahasiswa. Jumlah data yang dilakukan proses mining berjumlah 150 data.  
Tepat waktu = 68 orang,  
Tidak tepat waktu = 82 orang  
IPK 3 (tepat waktu) = 50 orang ,  
IPK 3 (tidak tepat waktu) = 27 orang  
IPK 2 (tepat waktu) = 18 orang,  
IPK 2 (tidak tepat waktu) = 53 orang  
IPK 1 (tidak tepat waktu) = 2 orang  
Perempuan (tepat waktu) = 33 orang  
Perempuan (tidak tepat waktu) = 20 orang  
Laki- laki (tepat waktu) = 35 orang  
Laki-laki (tidak tepat waktu) = 62 orang  
SMU (tepat waktu) = 53 orang  
SMU (tidak tepat waktu) = 61 orang  
SMK (tepat waktu) = 9 orang  
SMK (tidak tepat waktu) = 16 orang  
MA (tepat waktu) = 6 orang  
MA (tidak tepat waktu) = 5 orang  
Kota sekolah dalam kota (tepat waktu) = 10 orang  
Kota sekolah dalam kota (tidak tepat waktu) = 12 orang  
Kota sekolah luar kota (tepat waktu) = 58 orang  
Kota sekolah luar kota (tidak tepat waktu) = 70 orang  
Kota lahir dalam kota (tepat waktu) = 8 orang  
Kota lahir dalam kota (tidak tepat waktu) = 7 orang  
Kota lahir luar kota (tepat waktu) = 60 orang  
Kota lahir luar kota (tidak tepat waktu) = 75 orang  
Dari data IPK diatas dapat disimpulkan bahwa IPK berpengaruh besar terhadap prediksi tingkat kelulusan mahasiswa. Dimana IPK mahasiswa yang lebih dari 3 lebih banyak lulus tepat waktu di banding mahasiswa yang IPK nya 2 atau kurang dari 2.

### 3.3 Analisa Metode Naïve Bayes

Setelah dilakukan tahapan data mining, maka data yang digunakan dalam sistem prediksi tingkat kelulusan mahasiswa berjumlah 150 data.

Contoh kasus :

IPK:3, Jenis kelamin:perempuan, kota sekolah:dalam kota, sekolah:SMU, kota lahir:dalam kota. Berdasarkan rumus 1, dan 3 maka di dapat hasil sebagai berikut,

#### 1). Target

Ci : Tepat Waktu  $68/150 = 0.452$ .

Ci : Tidak Tepat Waktu  $82/150 = 0.55$

**2). IPK**

X || Ci : IPK = 3 || Tepat Waktu  $50/68 = 0.74$

X || Ci : IPK = 3 || Tidak Tepat Waktu  $27/82 = 0.33$

**3). Jenis Kelamin**

X || Ci : Jenis Kelamin = perempuan || Tepat Waktu  $33/68 = 0.49$

X || Ci : Jenis Kelamin = perempuan || Tidak Tepat waktu  $20/82 = 0.24$

**4). Kota Lahir**

X || Ci : Kota Lahir = dalam kota || Tepat Waktu  $8/68 = 0.12$

X || Ci : Kota Lahir = dalam kota || Tidak Tepat waktu  $7/82 = 0.09$

**5). Tipe Sekolah**

X || Ci : Tipe Sekolah = SMU || Tepat Waktu  $53/68 = 0.78$

X || Ci : Tipe Sekolah = SMU || Tidak Tepat Waktu  $61/82 = 0.74$

**6). Kota Sekolah**

X || Ci : Kota Sekolah = dalam kota || Tepat Waktu  $10/68 = 0.15$

X || Ci : Kota Sekolah = dalam kota || Tidak Tepat

Waktu  $12/82 = 0.15$

**7). Lama Studi tepat waktu**

X || Ci : Lama Studi || Tepat Waktu  $0.74 \times 0.49 \times 0.12 \times 0.78 \times 0.15 = 0.0051$

$P(X| \text{lama studi} = \text{"tepat waktu"}) P(\text{lama studi tepat waktu})$

$0.0051 \times 0.45 = 0.002295$

**8). Lama Studi tidak tepat waktu**

$P(X| \text{lama studi} = \text{"tidak tepat waktu"}) P(\text{lama studi tidak tepat waktu})$

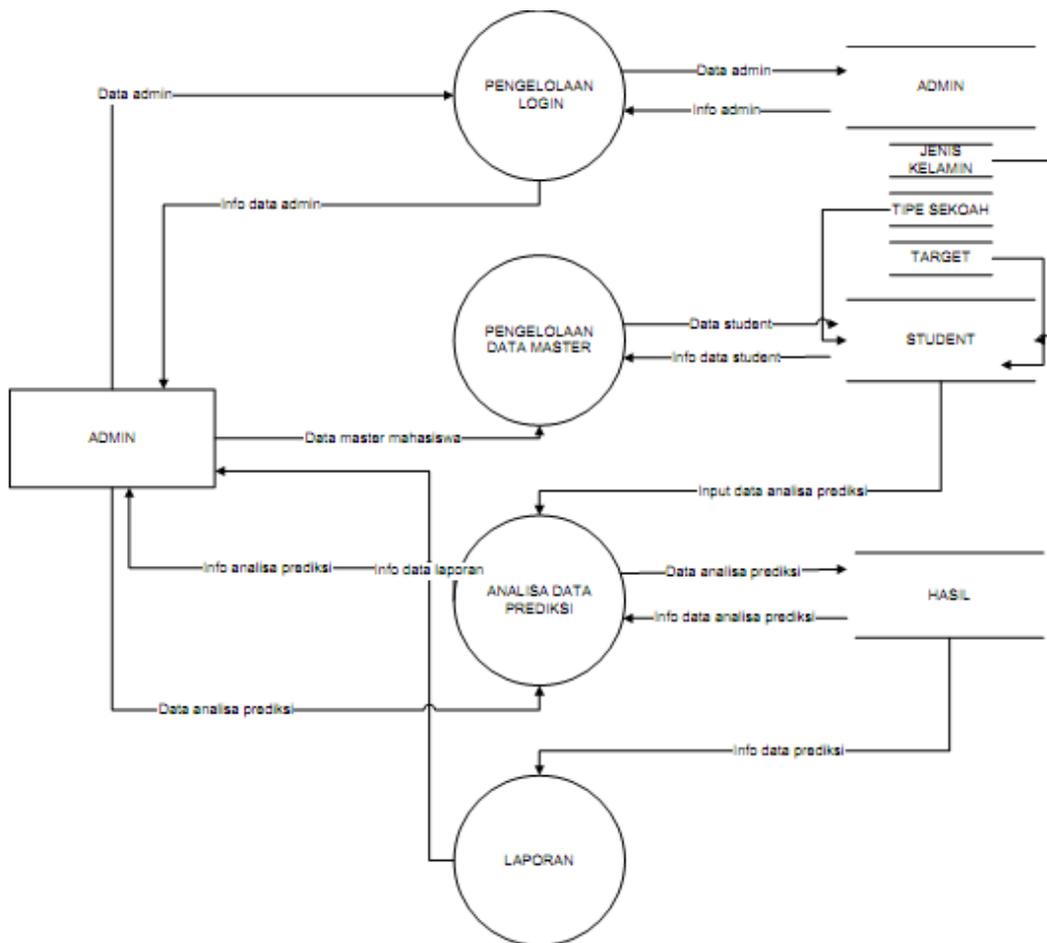
X || Ci : lama Studi || Tidak Tepat Waktu  $0.33 \times 0.24 \times 0.09 \times 0.74 \times 0.15 = 0.0008$

$0.0008 \times 0.55 = 0.00044$

" $P(X| \text{lama studi} = \text{"tepat waktu"}) P(\text{lama studi tepat waktu})$  lebih besar dari pada  
 $P(X| \text{lama studi} = \text{"tidak tepat waktu"}) P(\text{lama studi tidak tepat waktu})$   
maka mahasiswa "lulus tepat waktu"

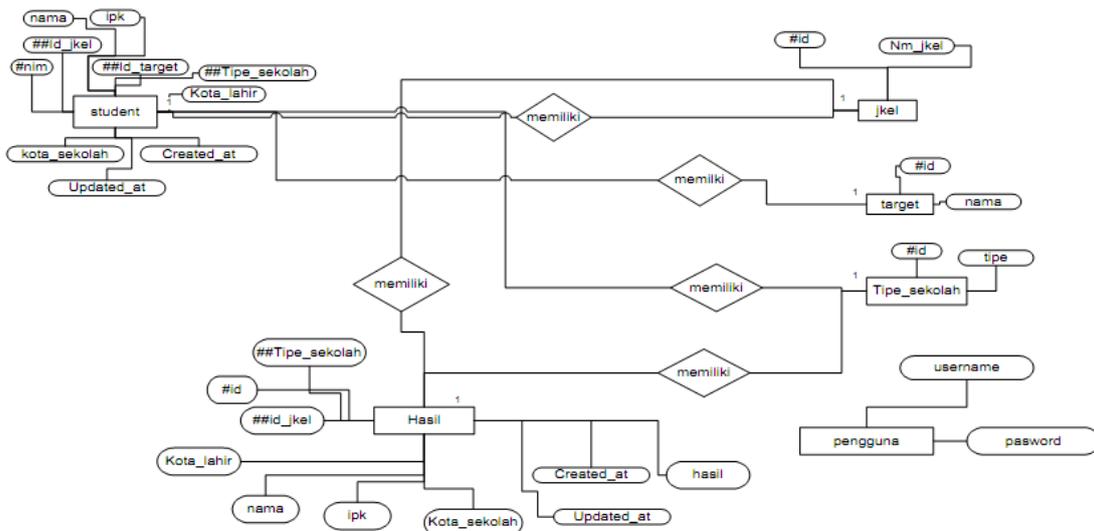
**3.4. Perancangan Perangkat Lunak**

Model perancangan perangkat lunak yang digunakan dalam sistem ini adalah model fungsional yang dipilih karena kerja sistem lebih ditekankan pada transformasi data masukan. Identifikasi data masukan hingga menghasilkan data keluaran akan digambarkan melalui *Data Flow Diagram* (DFD). Gambar 1 merupakan gambar DFD Level 1 Sistem.



Gambar 1. DFD Level 1

Perancangan data digambarkan menggunakan *Entity Relationship Diagram* (ERD) yang tergambar pada Gambar 2.



Gambar 2. ER-D

Perancangan antar muka sistem (*Interface*) salah satunya tergambar pada gambar 3. Dan Hasil tampilan data tergambar pada gambar 4.

Gambar 3. Perancangan Antar Muka Proses Perhitungan Prediksi Naïve Bayes

Gambar 4. Tampilan Menu Data Mahasiswa

### 3.5 Pengujian

Pengujian sistem dilakukan untuk melihat kesesuaian sistem yang telah dibangun dengan analisa dan perancangan sistem sehingga menghasilkan hasil akhir, pengujian ini menggunakan metode *black box*. Pengujian dilakukan dengan menjalankan semua proses pada *form* dengan memasukkan data dan melihat hasil keluaran dari proses yang ada. Hasil yang diperoleh adalah sistem yang dibangun berjalan baik, dan benar.

Pengujian juga dilakukan untuk mengukur tingkat akurasi sistem, pengukuran menggunakan rumus 4 dan 5.

Pengujian tingkat akurasi sistem terbagi 2, yaitu :

Uji coba 1. Dari dalam data training diambil 25 sampel.

Uji coba 2. Dari luar data training diambil 25 sampel.

Pengujian tingkat akurasi sistem dari kedua uji coba diatas terlihat pada tabel 2.

Tabel 2. Hasil Pengujian Tingkat Akurasi

Kegiatan Pengujian	Akurasi	Eror
Uji Coba 1	88 %	12 %
Uji Coba 2	92 %	8 %
<b>Rata-rata</b>	<b>90 %</b>	<b>10 %</b>

### 4. Kesimpulan

Berdasarkan hasil dan analisis sistem prediksi tingkat kelulusan mahasiswa menggunakan metode data mining dengan algoritma naïve bayes di Jurusan Teknik Informatika Fakultas Sains dan Teknologi ini, dapat diambil kesimpulan sebagai berikut:

1. Sistem prediksi tingkat kelulusan mahasiswa menggunakan metode data mining dengan algoritma naïve bayes ini berhasil dirancang dan diimplementasikan dapat

- digunakan untuk memprediksi dari masa studi atau ketepatan masa studi dari mahasiswa dengan data training dan data testing yang telah diperoleh.
2. Tingkat kesalahan dari fungsi klasifikasi yang digunakan untuk prediksi masih berkisar pada 8% hingga 12% yang hal ini dimungkinkan dapat dipengaruhi oleh jumlah data training maupun testing yang digunakan.
  3. Berdasarkan pengujian terhadap 50 data, diperoleh hasil mahasiswa dengan IPK 3 memperoleh tingkat prediksi lulus tepat waktu. Sedangkan untuk mahasiswa dengan IPK dibawah 3 akan memperoleh tingkat prediksi lulus tidak tepat waktu. Hal ini membuktikan bahwa IPK sangat berpengaruh terhadap tingkat prediksi kelulusan mahasiswa.

### Referensi

- [1] Huda, N.M. 2010. "*Aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro*", Skripsi, Program Studi Teknik Informatika Jurusan MIPA UNDIP, Semarang
- [2] Meinanda, M.H., Anisa M .Muhandri, N., Suryadi, K. 2009. "Prediksi masa studi sarjana dengan *artificial neural network*", Internet working Indonesia Journal, Vol.1No.2, pp. 31-35.
- [3] Santosa, Budi, 2007, "*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*", Graha Ilmu, Yogyakarta
- [4] Kusriani, Luthfi, E.T. 2009. "*Algoritma Data Mining*", Andi Offset. Surabaya.
- [5] Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.