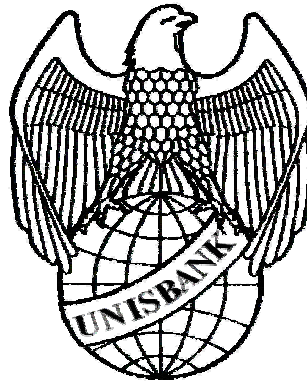


## **LAPORAN PENELITIAN**



### **PREDIKSI LAMA STUDI MAHASISWA MENGGUNAKAN TEKNIK DATAMINING (STUDI KASUS PADA FAKULTAS TEKNOLOGI INFORMASI UNISBANK)**

#### **OLEH :**

- |                                 |                                   |
|---------------------------------|-----------------------------------|
| 1. ARIEF JANANTO, S.KOM, M.CS   | / YS. 2.97.03.006 (KETUA)         |
| 2. EKO NUR WAHYUDI, S.KOM, M.Cs | / 19710912 200501 1 001 (ANGGOTA) |
| 3. ROHMAD AGUNG WIDODO          | / 11.01.55.0072 (ANGGOTA)         |
| 4. NIKKE FEBRILIANI             | / 09.01.55.0146 (ANGGOTA)         |

**FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG  
JANUARI 2013**

## HALAMAN PENGESAHAN LAPORAN PENELITIAN

1. a. Judul Penelitian : **PREDIKSI LAMA STUDI MAHASISWA MENGGUNAKAN TEKNIK DATAMINING (STUDI KASUS PADA FAKULTAS TEKNOLOGI INFORMASI UNISBANK)**
- b. Bidang Penelitian : Komputer (Sistem Informasi)
  
2. Ketua Peneliti
  - a. Nama Lengkap dan Gelar : Arief Jananto S.Kom, M.Cs
  - b. Jenis Kelamin : Laki-Laki
  - c. NIP : YS.2.97.03.006
  - e. Disiplin Ilmu : Ilmu Komputer
  - f. Pangkat / Golongan : Penata Muda / III C
  - g. Jabatan Fngsional : Lektor
  - h. Fakultas / Program Studi : Teknologi Informasi / Sistem Informasi
  - i. Alamat Kampus : Jl. Trilomba Juang No. 1 Semarang
  - j. Telp / Faks / e-mail : 8311668 / 8443240 / info@unisbank.ac.id
  - k. Alamat Rumah : Pucang Argo Tengah I/14 Pucanggading Demak
  - h. Telp / Faks / e-mail : 081325753989 / - / ajananto09@gmail.com
  
3. Jumlah Anggota Peneliti : 3 orang
  - a. Nama Anggota I : Eko Nur Wahyudi, S.Kom, M.Cs
  - b. Mahasiswa I : Rohmad Agung Widodo
  - c. Mahasiswa II : Nikke Febriliani
4. Lokasi Penelitian : Lab. Komputer UNISBANK Semarang
5. Jangka waktu Penelitian : 19 November 2012 – 31 Januari 2013
6. Jumlah biaya yang diusulkan : Rp. 3.000.000,-

Mengetahui,  
Fakultas Teknologi Informasi  
Dekan,

Semarang, 22 Januari 2013  
Ketua Pelaksana,

(Dwi Agus Diartono, S.Kom, M.Kom)

(Arief Jananto, S.Kom, M.Cs)

NIY. Y.2.90.03.054

NIY. YS.2.97.03.006

Menyetujui,  
Ketua LPPM UNISBANK

(Dr. Dra. Lie Liana, M.MSI)

NIY. Y.2.92.07.085

## ABSTRAK

### PREDIKSI LAMA STUDI MAHASISWA MENGGUNAKAN TEKNIK DATAMINING (STUDI KASUS PADA FAKULTAS TEKNOLOGI INFORMASI UNISBANK)

Arief Jananto, S.Kom, M.Cs, Eko Nur Wahyudi, S.Kom, M.Cs,  
Rohmad Agung Widodo, Nikke Febriliani

#### Abstrak

Lama studi dari mahasiswa ini sangatlah penting bagi mahasiswa, program studi serta perguruan tinggi. Permasalahan lama studi setiap mahasiswa bisa disebabkan atau dipengaruhi oleh banyak faktor. Hal tersebut telah dibuktikan dengan beberapa penelitian pada permasalahan tersebut yang mendapati sejumlah faktor yang berpengaruh terhadap lama studi mahasiswa.

Dengan menggunakan teknik data mining khususnya klasifikasi untuk prediksi dengan algoritma naive bayes dapat dilakukan prediksi terhadap ketepatan waktu studi dari mahasiswa berdasarkan data training yang ada. Data training dan testing yang digunakan diambil secara random pada tabel data master yang digunakan. Algoritma naive bayes, menghitung perbandingan peluang antara jumlah dari masing-masing kriteria nilai fields terhadap nilai hasil prediksi sesungguhnya. Fungsi untuk prediksi dibuat menggunakan Query pada MySQL dalam bentuk *function*(fbayesian).

Dari hasil uji coba diperoleh tingkat kesalahan prediksi berkisar 20% sampai dengan 50% dengan data training dan testing yang diambil secara random. Namun rata-rata tingkat kesalahan berkisar 20 % hingga 34%. Tinggi rendahnya tingkat kesalahan dapat disebabkan oleh jumlah record data dan tingkat konsistensi dari data training yang digunakan.

Sedangkan hasil prediksi dari ketepatan lama studi dari mahasiswa angkatan 2008 adalah sebesar 254 mahasiswa diprediksi "Tepat Waktu" dan sisanya yaitu 4 orang diprediksi "Tidak Tepat Waktu".

**Kata Kunci :** *Prediksi, Lama Studi, Data Mining, Naive bayes, MySQL*

## **KATA PENGANTAR**

Bismillahirrohmanirohim, puji sukur dipanjatkan ke hadirat Tuhan Yang Maha Esa hanya berkat rahmat dan hidayah NYA akhirnya telah diselesaikan laporan penelitian dengan judul “Prediksi Lama Studi Mahasiswa Menggunakan Teknik Data Mining (Studi Kasus Pada Fakultas Teknologi Informasi UNISBANK)”.

Namun demikian, selesainya penulisan laporan penelitian ini tidak terlepas dari bantuan banyak pihak yang selalu memberikan dukukung dan bimbingan serta perhatian yang cukup besar. Oleh karena itu pada kesempatan ini perkenankan penulis untuk menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Bambang Suko Priyono, M.M selaku Rektor Universitas Stikubank Semarang.
2. Ibu Dr. Lie Liana, M.MSi selaku Kepala Lembaga Penelitian dan Pengabdian Masyarakat Universitas Stikubank Semarang.
3. Bapak Dwi Agus Diartono, M.Kom selaku Dekan Fakultas Teknologi Informasi Universitas Stikubank Semarang.
4. Rekan-rekan Dosen dan Mahasiswa fakultas teknologi informasi Universitas Stikubank Semarang yang telah memberikan dukungan dan bantuannya.

Semoga laporan penelitian ini dapat bermanfaat dan menambah pengetahuan bagi siapapun yang membacanya. Tak lupa saran dan kritik dari pembaca yang budiman kami harapkan demi kebaikan bersama. Terima kasih.

Semarang, 22 Januari 2013

Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
KATA PENGANTAR .....	iii
ABSTRAK.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR .....	vii
DAFTAR TABEL.....	viii
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2. Perumusan Masalah .....	2
<b>BAB II TUJUAN DAN MANFAAT PENELITIAN</b>	
2.1. Tujuan Penelitian .....	3
2.2. Manfaat Penelitian .....	3
<b>BAB III TELAAH PUSTAKA</b>	
3.1. Evaluasi Belajar .....	4
3.2. Data Mining .....	5
3.3. Algoritma Naïve Bayes.....	7
3.4. Penelitian Sebelumnya.....	9
<b>BAB IV METODOLOGI PENELITIAN</b>	
4.1. Obyek Penelitian .....	10
4.2. Jenis Data .....	10
4.3. Teknik Pengumpulan Data.....	10
4.4. Tahapan Penelitian.....	11
<b>BAB V HASIL DAN PEMBAHASAN.....</b>	<b>26</b>

BAB VI KESIMPULAN DAN SARAN

6.1. Kesimpulan ..... 31  
6.2. Saran..... 31

DAFTAR PUSTAKA ..... 33

LAMPIRAN

Daftar Riwayat Peneliti(Ketua)..... 34  
Daftar Riwayat Peneliti(Anggota)..... 36  
Daftar Riwayat Peneliti(Anggota)..... 38  
Daftar Riwayat Peneliti(Anggota)..... 39  
Peta Lokasi Penelitian..... 40

## **DAFTAR GAMBAR**

Gambar 1.1. Data Mining adalah suatu langkah di dalam proses KDD ..... 7

Gambar 5.1. Hasil uji coba procedure/function bayesian ..... 27

## **DAFTAR TABEL**

Tabel 4.1. Data Asli (Master Data) .....	13
Tabel 4.2. Data kelulusan dari website EPSBED .....	13
Tabel 4.3. Data matakuliah yang telah ditempuh.....	14
Tabel 4.4. Data gabungan dari tabel data baik dan EPSBED.....	15
Tabel 4.5. Data pada tabel siap proses mining.....	19
Tabel 5.1. Tingkat error dengan data training diambil secara random. ....	28
Tabel 5.2. Daftar tingkat kesalahan dari hasil uji coba yang dilakukan .....	29



# **BAB I**

## **PENDAHULUAN**

### **1.1. LATAR BELAKANG**

Kualitas dari sebuah perguruan tinggi selain dapat dilihat dari rata-rata lama lulusannya mendapatkan pekerjaan juga dapat dilihat dari rata-rata masa atau lama studi dari mahasiswanya. Setiap perguruan tinggi akan mempunyai variasi lama studi dari mahasiswanya yang berbeda-beda.

Berhubungan dengan pengelolaan lama studi dari mahasiswa ini merupakan tugas sangat penting dari program studi. Program studi berkewajiban untuk memantau perkembangan studi dari mahasiswanya. Kemudian program studi juga mempunyai tugas untuk memprediksi lama studi dari setiap mahasiswanya guna menentukan dan mengantisipasi terjadinya mahasiswa yang 'mangkrak' atau 'mangkir' yang akan menyebabkan kinerja program studi menjadi kurang baik.

Sesuai dengan buku peraturan akademik Universitas Stikubank (UNISBANK) Semarang tahun akademik 2010/2011, pada Bab I Pengertian Umum Pasal 1 ayat 11 disebutkan bahwa : Program Sarjana(S-1) reguler adalah program pendidikan akademik setelah pendidikan menengah yang memiliki beban studi sekurang-kurangnya 144 sks dan sebanyak-banyaknya 160 sks yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan paling lama 14 semester.

Lama studi dari mahasiswa ini sangatlah penting bagi mahasiswa itu sendiri dan tentunya bagi program studi dalam rangka untuk menentukan tindakan-tindakan atau langkah-langkah yang diperlukan agar perjalanan studi mahasiswa menjadi lancar. Permasalahan lama studi setiap mahasiswa bisa disebabkan atau dipengaruhi oleh banyak faktor. Hal tersebut telah dibuktikan dengan beberapa penelitian pada permasalahan tersebut yang mendapati sejumlah faktor yang berpengaruh terhadap lama studi mahasiswa.

Menurut Nuqson Masykur Huda, dalam penelitiannya aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus

FMIPA Universitas Diponegoro menyebutkan bahwa data proses masuk, asal sekolah, kota asal dan program studi menjadi pertimbangan dalam menghitung tingkat kelulusan. Dimana masing-masing faktor tersebut dicari nilai *support* dan *confidence* nya.

Sedangkan menurut Muhamad Hanief Meinanda dkk, dalam penelitiannya tentang prediksi masa studi sarjana dengan *artificial neural network* menyebutkan bahwa masa studi seorang sarjana dipengaruhi oleh IPK, jumlah matakuliah yang diambil, jumlah matakuliah yang mengulang, jumlah matakuliah tertentu.

Untuk memprediksi dapat dilakukan dengan berbagai cara, salah satunya dapat pula dilakukan dengan menggunakan teknik datamining. Teknik data mining sebagai sebuah teknik analisis berbasis database dan statistik, sangatlah tepat untuk menggali berbagai potensi informasi yang terkandung dalam sebuah database. Pada penelitian yang akan dilakukan nantinya akan menggunakan faktor penilaian(nilai) sampai pada masa evaluasi kinerja mahasiswa tahap I, yaitu semester iv(empat). Selain itu dimungkinkan juga digunakan beberapa data terkait dengan data induk mahasiswa seperti jenis/tipe sekolah, jenis kelamin, range usia, kelompok kelas.

## 1.2. PERUMUSAN MASALAH

Berdasarkan latar belakang di atas, maka permasalahan yang dapat dirumuskan adalah bagaimana memprediksi lama masa studi dengan menggunakan teknik data mining. Dimana faktor nilai-nilai matakuliah yang telah ditempuh hingga semester iv(masa evaluasi tahap I) atau IPK sampai dengan semester IV dan faktor data induk mahasiswa yang berupa jenis kelamin, kota lahir, tipe sekolah, kota sekolah akan menjadi variabel prediktornya. Metode yang akan digunakan adalah klasifikasi dengan algoritma naive bayes.

## **BAB II**

### **TUJUAN DAN MANFAAT**

#### 2.1. TUJUAN PENELITIAN

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.

#### 2.2. MANFAAT PENELITIAN

Dengan mengetahui prediksi dari lama masa studi 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.

## **BAB III**

### **TELAAH PUSTAKA**

#### **3.1. TINJAUAN PUSTAKA**

##### **3.1.1. Evaluasi Hasil Belajar**

Evaluasi hasil belajar adalah suatu proses yang dilakukan untuk mengetahui tingkat kinerja akademik. Evaluasi hasil belajar tersebut dilakukan secara menyeluruh dan terus-menerus dengan cara yang sesuai dengan ciri-ciri pendidikan yang bersangkutan (Pedoman Akademik UNISBANK, 2010).

##### **Evaluasi Kemajuan Hasil Belajar**

(1) Evaluasi kemajuan studi mahasiswa Program Sarjana (S-1) :

Untuk mengetahui kemajuan studi mahasiswa, pada setiap empat semester dilakukan evaluasi.

(a) Kriteria evaluasi tiap tahapan sebagai berikut :

(1) Empat semester pertama.

1. Mampu mengumpulkan paling sedikit 48 sks dengan  $IPK \geq 2,75$
2. Apabila mampu mengumpulkan  $> 48$  sks, tetapi  $IPK < 2,75$  maka diambil nilai-nilai tertinggi sampai sejumlah 48 sks dengan  $IPK \geq 2,75$

(2) Empat semester kedua (semester kedelapan) .

1. Mampu mengumpulkan paling sedikit 96 sks dengan  $IPK \geq 2,75$
2. Apabila mampu mengumpulkan  $> 96$  sks, tetapi  $IPK < 2,75$  maka diambil nilai-nilai tertinggi sampai sejumlah 96 sks dengan  $IPK \geq 2,75$ .

(3) Akhir Program

Selambat-lambatnya pada akhir semester ke-14, mahasiswa harus sudah mengumpulkan (lulus) semua beban sks yang ditetapkan dengan  $IPK \geq 2,75$ .

(b) Mahasiswa akan mendapatkan peringatan akademik apabila disangsikan dapat melalui tiap tahapan evaluasi.

(c) Mahasiswa yang tidak dapat memenuhi kriteria setiap tahapan evaluasi tersebut dianggap tidak mampu mengikuti kegiatan-kegiatan akademiknya. Sehubungan dengan hal tersebut, Rektor menerbitkan surat keputusan menghentikan statusnya sebagai mahasiswa UNISBANK setelah memperoleh bahan-bahan pertimbangan seperti diatur pada Pasal 10 ayat (6) Buku Peraturan Akademik.

(d) Keberhasilan Menyelesaikan Studi

Mahasiswa berhasil menyelesaikan pendidikan program sarjana (lulus sarjana), yang dinyatakan dalam yudisium kelulusan apabila telah memenuhi persyaratan akademik dan kemahasiswaan sebagai berikut :

(1) Telah berhasil mengumpulkan sejumlah sks yang ditetapkan di dalam kurikulum program studi.

(2) Tidak memiliki nilai E untuk semua matakuliah, nilai D dan C-maksimal 10% dari jumlah matakuliah yang ditetapkan, dan minimal C untuk MPK pada kurikulum inti.

(3)  $IPK \geq 2,75$

(4) Telah memperoleh minimal 100 skp untuk program diploma

(e) Tanggal kelulusan adalah tanggal penetapan IPK akhir program.

(f) Syarat pengambilan ijazah harus memiliki sekurang-kurangnya 3 (tiga) macam sertifikasi dan memiliki sertifikat entrepreneurial.

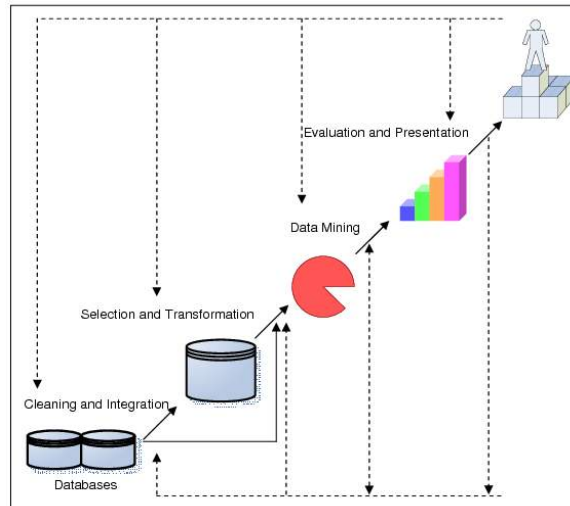
### 3.1.2. DATA MINING

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 satu langkah yang penting di dalam proses KDD. Han dan Kamber (2001) menyatakan bahwa KDD terdiri dari langkah-langkah sebagai berikut :

1. *Data cleaning* : proses menghapus data yang tidak konsisten dan noise
2. *Data integration* : penggabungan multiple data source
3. *Data selection* : pengambilan data yang relevan dari database
4. *Data transformation* : proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining
5. *Data mining* : suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data
6. *Pattern evaluation* : proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada di dalam data itu sendiri
7. *Knowledge presentation* : proses visualisasi dan teknik menyajikan knowledge untuk menampilkan knowledge hasil mining kepada user

Untuk lebih jelasnya, dapat dilihat pada Gambar 1.1



Gambar 1.1. Data Mining adalah suatu langkah di dalam proses KDD  
(Han dan Kamber, 2001)

Beberapa algoritma dan teknik yang biasa digunakan dalam data mining antara lain adalah sebagai berikut :

1. *Prediction*, yaitu bagaimana suatu atribut tertentu termasuk datanya akan bereaksi di masa yang akan datang
2. *Identification*, yaitu mengenali keberadaan suatu item, suatu kejadian atau suatu kegiatan
3. *Classification*, yaitu membagi data ke dalam beberapa kategori
4. *Clustering*, yaitu mengelompokkan objek
5. *Assosiation*, yaitu menggabungkan beberapa objek

### 3.1.3. ALGORITMA NAIVE BAYES

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

Teorema Bayes memiliki bentuk umum sebagai berikut :

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

- 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

Peluang kejadian H sebagai X ditentukan dari peluang X saat A, peluang H dan peluang X. Kemudian pada pengaplikasiannya menjadi :

$$P(D|C_i) = \frac{P(D|C_i)P(C_i)}{P(D)}$$

Dimana :

- $P(C_i|D)$  adalah peluang data pada kategori  $C_i$
- $P(D|C_i)$  adalah peluang pada kategori  $C_i$ , data D muncul pada kategori tersebut
- $P(C_i)$  adalah peluang dari kategori yang diberikan, dibandingkan dengan kategori-kategori lainnya yang dianalisa
- $P(D)$  adalah peluang data tersebut secara spesifik.

Penerapan Naive bayes

Dalam query rumus bayesian akan berubah menjadi :

$$P(\text{Tepat Waktu} | D) = (P(D | \text{Tepat Waktu}) * P(\text{Tepat Waktu}))$$

dan

$$P(\text{Tidak Tepat Waktu} | D) = (P(D | \text{Tidak Tepat Waktu}) * P(\text{Tidak Tepat Waktu}))$$

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.



#### 3.1.4. Lama Studi Jenjang Strata 1

Program Sarjana(S-1) reguler adalah program pendidikan akademik setelah pendidikan menengah yang memiliki beban studi sekurang-kurangnya 144 sks dan sebanyak-banyaknya 160 sks yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan paling lama 14 semester (Buku Peraturan akademik Universitas Stikubank (UNISBANK) Semarang tahun akademik 2010/2011).

### 3.2. PENELITIAN SEBELUMNYA

Setiap mahasiswa mempunyai lama studi yang bisa sama ataupun berbeda dengan mahasiswa lain dengan berbagai faktor penyebab. Seperti penelitian yang dilakukan oleh Nuqson Masykur Huda, dalam penelitiannya aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro menyebutkan bahwa data proses masuk, asal sekolah, kota asal dan program studi menjadi pertimbangan dalam menghitung tingkat kelulusan. Nuqson dalam penelitiannya menggunakan metoda *association rule* serta algoritma *apriori*. Dimana dalam metoda dan algoritma ini masing-masing faktor tersebut dicari nilai *support* dan *confidence* nya sehingga akan diperoleh faktor mana yang paling berperan atau paling mempunyai pengaruh yang cukup besar berkaitan dengan tingkat kelulusan mahasiswa.

Sedangkan menurut Muhamad Hanief Meinanda dkk, dalam penelitiannya tentang prediksi masa studi sarjana dengan *artificial neural network* menyebutkan bahwa masa studi seorang sarjana dipengaruhi oleh IPK, jumlah matakuliah yang diambil, jumlah matakuliah yang mengulang, jumlah matakuliah tertentu. Metoda penelitian yang digunakan merupakan adopsi dari CRISP-DM atau *Cross-Industry Standard Process for Data Mining*, dimana di dalamnya terdapat enam fase yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation* dan *Deployment*. Dalam. Setelah memahami masalahnya kemudian melakukan processing data dengan membuat *cross-tabulation*, koreksi data yang *misclassification*, *missing value* maupun *outlier*.

## **BAB IV**

### **METODE PENELITIAN**

#### **4.1. Obyek Penelitian**

Pada penelitian ini, yang menjadi obyek penelitian adalah Fakultas Teknologi Informasi UNISBANK, Kampus Mugas Jl. Tri Lomba Juang No 1 Mugas Semarang.

#### **4.2. Jenis Data**

##### **a. Data Primer**

Merupakan data-data utama yang digunakan atau diolah dalam penelitian ini. Data primer ini diambil dari bagian biro administrasi akademik. Data berupa data nilai-nilai matakuliah yang telah ditempuh hingga semester iv(masa evaluasi tahap I) dan data induk mahasiswa yang berupa jenis kelamin, kota lahir, tipe sekolah, kota sekolah.

##### **b. Data Sekunder**

Merupakan data yang diperoleh dari pihak lain yaitu dapat melalui media tercetak maupun non tercetak. Dalam hal ini data tentang tahun lulus diambil dari website EPSBED yang berkaitan dengan data tahun lulus dari program studi yang berjenjang Strata 1 di Fakultas Teknologi Informasi.

#### **4.3. Teknik Pengumpulan Data**

##### **a. Wawancara**

Metode wawancara merupakan suatu metode pengumpulan data dengan cara melakukan tanya jawab dengan pihak yang berwenang dalam hal ini adalah Kepala Biro Administrasi Akademik di UNISBANK Semarang.

b. Studi Pustaka

Merupakan metode pengumpulan data dengan mencari atau menggunakan sumber dari buku atau literatur, jurnal dan hasil-hasil penelitian yang berkaitan dengan penelitian ini.

#### **4.4. Tahapan Penelitian**

Langkah atau tahapan penelitian ini dilakukan sebagai berikut :

1. Melakukan studi kepustakaan terhadap berbagai referensi yang berkaitan dengan penelitian yang dilakukan. Topik-topik yang dikaji antara lain meliputi : database, data mining, klasifikasi, dan beberapa algoritma lain yang mungkin dapat digunakan, serta perangkat lunak yang digunakan untuk proses mining.
2. Menyiapkan data-data yang dibutuhkan, yaitu data berkaitan dengan identitas pribadi dari mahasiswa dan nilai matakuliah yang telah ditempuh oleh mahasiswa hingga mencapai 146 sks 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 ini nantinya merupakan data training dan testing. Mendapatkan tahun lulus dari setiap mahasiswa yang

telah lulus tadi sebagai acuan untuk menentukan criteria tepat waktu lulus atau tidak.

3. Mengolah data yang telah disiapkan menggunakan beberapa perangkat lunak bantu seperti Excel, MySql.
4. Membuat kesimpulan dari hasil pengolahan data dan mining data yang telah dilakukan untuk menghasilkan informasi mengenai prediksi lama studi dari mahasiswa maupun informasi lain yang dapat digunakan sebagai bahan analisa untuk pengambilan keputusan.

#### **4.4.1. Persiapan Data**

Tahapan awal yang dilakukan dalam penelitian ini adalah menyiapkan data, dimana data diperoleh dari bagian Biro Administrasi Akademik (BAA). Data yang diperoleh dan akan digunakan dalam penelitian ini berupa data berkaitan dengan keterangan diri dari mahasiswa yaitu kota lahir, jenis sekolah, kota sekolah, jenis kelamin, dan data akademik selama kuliah sampai dengan semester 4. Data nilai semua matakuliah diambil berdasarkan data dari semua mahasiswa yang telah lulus periode 2004 s/d 2007. Diambilnya periode angkatan mahasiswa pada periode itu adalah karena memiliki struktur kurikulum yang sama. Selanjutnya dari data itu diambil hanya nilai matakuliah sampai dengan semester 4, dengan dasar karena periode tersebut adalah merupakan tahap I penilaian kemampuan study dari mahasiswa dan hal ini mengacu pada aturan yang tertera dalam buku pedoman akademik.

Pada tahap ini data mahasiswa yang lulus dari periode akademik 2004 s/d 2007 diperoleh data sejumlah 1059 record. Adapun data waktu kelulusan didasarkan pada laporan di website EPSBED. Berikut contoh sebagian data yang berhubungan dengan data pribadi dan latar belakang sekolah menengah atas dari

mahasiswa yang diperoleh dari BAAK dengan tidak menampilkan data nim dan nama dari mahasiswa dengan alasan merupakan data pribadi.

Tabel. 4.1. Data asli (Master Data)

No	IPK-4	Jenis Kelamin	Kota Lahir	Tipe Sekolah	Kota Sekolah
1	3.10	Perempuan	TEGAL	SMK BHAKTI PRAJA TALANG	TEGAL
2	3.16	Perempuan	PATI	SMU NASIONAL	PATI
3	3.37	Perempuan	SEMARANG	SMA MASEHI 2 PSAK	SEMARANG
4	3.28	Perempuan	SEMARANG	SMAN 2 MRANGGEN	DEMAK
5	3.46	Laki-Laki	SEMARANG	SMU NEGERI 9	SEMARANG
6	2.96	Perempuan	SEMARANG	SMU NUSAPUTERA	SEMARANG
7	2.71	Laki-Laki	SEMARANG	SMU NUSAPUTERA	SEMARANG
8	2.74	Perempuan	LAHAT	SMU 3 REMBANG	REMBANG
9	2.51	Perempuan	KUDUS	SMUN 1 GEBOG	KUDUS
10	3.03	Laki-Laki	SEMARANG	SMU NUSAPUTERA	SEMARANG
11	3.08	Perempuan	SEMARANG	SMUN 11	SEMARANG
12	3.05	Perempuan	SEMARANG	SMUN 11	SEMARANG
13	2.82	Perempuan	JEPARA	SMU PGRI JEPARA	JEPARA
14	2.85	Perempuan	SEMARANG	SMU NUSAPUTERA	SEMARANG
15	3.04	Laki-Laki	SEMARANG	SMKN 3	SEMARANG
16	2.49	0	0	0	0
17	1.86	Laki-Laki	SEMARANG	SMU KEBON DALEM	SEMARANG
...	.....	.....	.....	.....	.....

Sedangkan contoh data yang berkaitan dengan tahun kelulusan yang diperoleh dari website EPSBED dapat dilihat pada tabel 4.2.

Tabel. 4.2. Data kelulusan dari website ESBED

Progdi	Nama	Angkatan	Tanggal Lulus
SI	<u>ABDUL LUQMAN SIDIK</u>	2004	20080922
SI	<u>ABI AUFA ZAIN</u>	2004	20090327
SI	<u>ACHMAD RIVAI</u>	2004	20080922
SI	<u>ADITIA DWI KURNIAWAN</u>	2004	20080922
SI	<u>AGUS ARIYANTO</u>	2004	20090929
SI	<u>AGUS BUDIYANTO</u>	2004	20090911
SI	<u>AGUS SUBEKTI CANDRA</u>	2004	20080922

SI	<u>AGUSTINA WIDONINGRUM</u>	2004	20080922
SI	<u>AJI GUNAWAN</u>	2004	20080922
SI	<u>AL- ATIKOH</u>	2004	20080922
SI	<u>ALI IMRON</u>	2004	20090327
SI	<u>ALOYSIUS DHARMA LUKITA</u>	2004	20090327
SI	<u>ANA FITRIANISA</u>	2004	20080922
....	.....	....	.....

Selain itu juga diambil data mengenai pencapaian indeks prestasi akademik hingga semester 4 dari data mahasiswa yang telah dinyatakan lulus di atas. Contoh data nilai matakuliah yang diperoleh dapat dilihat pada tabel 4.3.

Tabel 4.3. Data matakuliah yang telah ditempuh

	ALGORITMA DAN STRUKTUR DATA 1	PENDIDIKAN PANCASILA	PRAKTIKUM ALGORITMA DAN STRUKTUR	PERANCANGAN BASIS DATA 1	BAHASA INGGRIS 1	KONSEP SISTEM INFORMASI	PENGANTAR MANAJEMEN	XXXXXXXXXXXXXXXXXX	PRAKT. ANALISA DAN PERANC. SIS	PRAKTIKUM BAHASA PEMROGRAMAN	ANALISA DAN PERANCANGAN SISTEM	BAHASA PEMROGRAMAN 2	STATISTIKA	MANAJEMEN SAINS	
	2	2	1	2	2	2	2	9	1	1	2	2	4	3	75
NIM	1	2	3	4	5	6	7	9	37	38	39	40	41	42	IPK
04.01.55.0005	B+	A	A-	B	B+	C+	B	xx	B+	B-	C+	B	B	B	3.02
04.01.55.0012	B	B+	A-	B+	B+	B-	C+	xx	B	B	C+	B	B+	A-	3.11
04.01.55.0019	B+	B	A-	B	A-	A	A-	xx	A-	C+	B-	B	A-	B-	3.37
04.01.55.0021	B+	B+	A	A-	B	B-	A-	xx	B+	B	B-	B	B	B+	3.22
04.01.55.0024	A-	B	A	B	B+	A	A	xx	A-	B	B-	B	A	B+	3.48
04.01.55.0025	B+	A-	A	B+	B+	A-	C	xx	B+	B	C+	B	B	C+	2.94
04.01.55.0026	B-	B	A-	B+	C	C+	B	xx	B	C+	C+	B	C	C+	2.7
04.01.55.0027	B-	A-	A	B+	B+	C+	B	xx	C-	B-	C	C+	C+	C	2.67
04.01.55.0028	B-	B	B	B+	B	B+	B-	xx	C	B-	C+	B-	C+	D	2.55
04.01.55.0029	B	B+	B	B	C	B-	B	xx	C-	B+	B+	B-	B	C+	3.03

04.01.55.0030	B+	B	C+	B	C	B	B-	xx	B-	A	B-	B	B	C+	3.04
04.01.55.0031	B+	B	C+	B	B+	B-	B-	xx	B-	A-	B-	B	B-	C+	2.84
04.01.55.0033	B	A-	C+	C+	B+	B-	B+	xx	B+	B	C	B	C	C+	2.76
04.01.55.0035	B	A-	A-	B-	B	B	B-	xx	B	B+	C	B	B-	C+	2.83
04.01.55.0038	B-	B	B	B-	B	C	B+	xx	C+	A	C+	B	B	C+	3.07
04.01.55.0039	B+	B+	D	B-	B-	C+	B-	xx	B	C+	B-	B-	C	B	2.53
04.01.55.0042	B+	B+	B-	B+	B+	B	B+	xx	A-	C+	C	B	B-	B+	2.84

Selanjutnya data-data tersebut di atas digabungkan menjadi satu tabel dengan mengacu nim sebagai key primernya sehingga diperoleh data dengan struktur sesperti tampak pada tabel 4.4.

Tabel 4.4. Data gabungan dari tabel data baak dan EPSBED

IPK-4	Jenis Kelamin	Kota Lahir	Tipe Sekolah	Kota Sekolah	Tahun Lulus
3.10	Perempuan	TEGAL	Kejuruan	TEGAL	2008
3.16	Perempuan	PATI	Umum	PATI	2011
3.37	Perempuan	SEMARANG	Umum	SEMARANG	2008
3.28	Perempuan	SEMARANG	Umum	DEMAK	2008
3.46	Laki-Laki	SEMARANG	Umum	SEMARANG	2008
2.96	Perempuan	SEMARANG	Umum	SEMARANG	2008
2.71	Laki-Laki	SEMARANG	Umum	SEMARANG	2008
2.74	Perempuan	LAHAT	Umum	REMBANG	2011
2.51	Perempuan	KUDUS	Umum	KUDUS	2008
3.03	Laki-Laki	SEMARANG	Umum	SEMARANG	2011
3.08	Perempuan	SEMARANG	Umum	SEMARANG	2011
3.05	Perempuan	SEMARANG	Umum	SEMARANG	2008
2.82	Perempuan	JEPARA	Umum	JEPARA	2008
2.85	Perempuan	SEMARANG	Umum	SEMARANG	2008
3.04	Laki-Laki	SEMARANG	Kejuruan	SEMARANG	2008
2.49	Laki-Laki	Tegal	Kejuruan	Tegal	2008
1.86	Laki-Laki	SEMARANG	Umum	SEMARANG	2008
2.88	Laki-Laki	SEMARANG	Umum	SEMARANG	2012
3.21	Perempuan	SEMARANG	Kejuruan	SEMARANG	2008
3.17	Laki-Laki	KAB. SEMARANG	Umum	UNGARAN	2008
3.18	Perempuan	SEMARANG	Umum	SEMARANG	2008

Dari hasil penggabungan data tersebut diperoleh sejumlah permasalahan kondisi data yaitu terutama data yang tidak lengkap. Untuk itu maka dilakukan proses CLEANING data. Maka untuk data yang bermasalah tersebut di hilangkan atau tidak ikut digunakan dalam proses mining nantinya, hingga akhirnya sampai proses semua data lengkap diperoleh record data sebanyak 266 record data.

#### 4.4.2. Konversi Data

Untuk mempermudah dalam proses mining selanjutnya maka data yang ada dilakukan proses konversi kedalam bentuk yang dapat diolah dengan alat bantu data mining yang ada. Adapun proses konversi data data yang telah diperoleh dari tahap persiapan data adalah sebagai berikut :

a. Untuk IPK

Konversi nilai IPK dilakukan dengan membuat range nilai IPK yaitu :

IPK Asli	IPK Baru
IPK Asli > 3.00	3
2 <= IPK Asli <=3.00	2
IPK Asli < 2.00	1

Contoh hasil konversi :

IPK-4	IPK Baru
3.10	3
3.16	3
3.37	3
3.28	3
3.46	3
2.96	2
2.71	2
2.74	2



2.51	2
3.03	3
3.08	3
3.05	3
2.82	2
2.85	2
3.04	3
2.49	2
1.86	1

b. Untuk Jenis Kelamin

Untuk jenis kelamin , dikarenakan hanya berisi dua nilai maka tidak dilakukan konversi.

c. Untuk Kota Lahir

Untuk kota lahir dikelompokkan hanya menjadi dua nilai yaitu kota lahir yang berasal dari Semarang dikonversikan menjadi 'DALAM KOTA' dan yang berasal dari luar Semarang dikonversikan menjadi 'LUAR KOTA'.

Berikut contoh datanya :

Kota Lahir (Asli)	Kota Lahir (Baru)
TEGAL	LUAR KOTA
PATI	LUAR KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
LAHAT	LUAR KOTA
KUDUS	LUAR KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
JEPARA	LUAR KOTA

d. Untuk Tipe Sekolah

Untuk tipe sekolah dilakukan pengelompokan yaitu dari sisi tipe sekolahnya. Untuk sekolah berkategori SMU atau SMA di konversikan menjadi 'Umum' sedangkan selain SMU atau SMA dikonversikan menjadi 'Kejuruan'. Berikut contoh datanya :

<b>Tipe Sekolah</b>		
SMK BHAKTI PRAJA TALANG	SMK	Kejuruan
SMU NASIONAL	SMU	Umum
SMA MASEHI 2 PSAK	SMA	Umum
SMAN 2 MRANGGEN	SMA	Umum
SMU NEGERI 9	SMU	Umum
SMU NUSAPUTERA	SMU	Umum
SMU NUSAPUTERA	SMU	Umum
SMU 3 REMBANG	SMU	Umum
SMUN 1 GEBOG	SMU	Umum
SMU NUSAPUTERA	SMU	Umum
SMUN 11	SMU	Umum
SMUN 11	SMU	Umum
SMU PGRI JEPARA	SMU	Umum
SMU NUSAPUTERA	SMU	Umum
SMKN 3	SMK	Kejuruan

e. Untuk Kota Sekolah

Untuk kota sekolah dikelompokkan menjadi dari dalam kota Semarang atau dari luar kota Semarang. Untuk yang kota sekolahnya adalah Semarang maka dikonversikan datanya menjadi 'DALAM KOTA' sedangkan selain yang kota sekolahnya 'SEMARANG' dikonversikan menjadi 'LUAR KOTA'. Berikut contoh datanya :

Kota Sekolah	
TEGAL	LUAR KOTA
PATI	LUAR KOTA
SEMARANG	DALAM KOTA

DEMAK	LUAR KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
REMBANG	LUAR KOTA
KUDUS	LUAR KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
JEPARA	LUAR KOTA
SEMARANG	DALAM KOTA
SEMARANG	DALAM KOTA
Tegal	LUAR KOTA

Dan pada akhirnya diperoleh susunan data sebagai tampak pada tabel 4.5. sejumlah 266 record data yang merupakan data siap untuk di proses mining.

Tabel 4.5. Data pada tabel siap proses mining

IPK_4	Jenis_Kelamin	Kota_Lahir	Tipe_Sekolah	Kota_Sekolah	Lama_Studi
3	Perempuan	LUAR KOTA	Kejuruan	LUAR KOTA	Tepat Waktu
3	Perempuan	LUAR KOTA	Umum	LUAR KOTA	Tepat Waktu
3	Perempuan	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
3	Perempuan	DALAM KOTA	Umum	LUAR KOTA	Tepat Waktu
3	Laki-Laki	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
2	Perempuan	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
2	Laki-Laki	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
2	Perempuan	LUAR KOTA	Umum	LUAR KOTA	Tepat Waktu
2	Perempuan	LUAR KOTA	Umum	LUAR KOTA	Tepat Waktu
3	Laki-Laki	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
3	Perempuan	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
3	Perempuan	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
2	Perempuan	LUAR KOTA	Umum	LUAR KOTA	Tepat Waktu
2	Perempuan	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu
3	Laki-Laki	DALAM KOTA	Kejuruan	DALAM KOTA	Tepat Waktu
2	Laki-Laki	LUAR KOTA	Kejuruan	LUAR KOTA	Tepat Waktu
1	Laki-Laki	DALAM KOTA	Umum	DALAM KOTA	Tepat Waktu

2	Laki-Laki	DALAM KOTA	Umum	DALAM KOTA	Tidak Tepat Waktu
3	Perempuan	DALAM KOTA	Kejuruan	DALAM KOTA	Tepat Waktu

#### 4.4.3. Algoritma Naive Bayes

Data yang telah diolah dengan microsoft excel tersebut di atas, kemudian dimasukkan ke dalam MySql sebagai basis data.

Proses *import* data dari excel ke dalam MySql melalui MySqlQuery Browser dilakukan dengan memanfaatkan fungsi LOAD DATA INFILE. Sintak dari proses *import* data adalah :

```
LOAD DATA INFILE 'lokasi:\nama_file'
INTO TABLE 'nama_tabel'
FIELDS TERMINATED BY ','
IGNORE 1 LINES;
```

Selanjutnya dengan menggunakan alat bantu MySql dibuatlah sebuah prosedur/function yang dapat digunakan untuk menentukan kelas(prediksi) dari sebuah record data baru dengan berdasar data training tersebut. Adapun tahapan algoritma Naive bayes dalam query adalah :

- Mulai
- Baca data training
  - o Hitung  $P(C_i)$  untuk setiap kelas
  - o Hitung  $P(X|C_i)$  untuk setiap kriteria dan setiap kelas
  - o Cari  $P(X|C_i)$  yang paling besar menjadi kesimpulan
- Tampilkan hasil prediksi
- Selesai

Lebih detail dalam bentuk query di MySql :

- Mulai
  - o Siapkan variabel tambahan yang digunakan untuk menyimpan data hasil perhitungan.

```
BEGIN
declare total_record numeric;

declare jml_tw numeric;

declare jml_ttw numeric;

declare pjml_tw numeric(8,4);
declare pjml_ttw numeric(8,4);
declare ipk_4_tw numeric(8,4);
declare ipk_4_ttw numeric(8,4);
declare jenis_kelamin_tw numeric(8,4);
declare jenis_kelamin_ttw numeric(8,4);
declare kota_lahir_tw numeric(8,4);
declare kota_lahir_ttw numeric(8,4);
declare tipe_sekolah_tw numeric(8,4);
declare tipe_sekolah_ttw numeric(8,4);
declare kota_sekolah_tw numeric(8,4);
declare kota_sekolah_ttw numeric(8,4);
declare px_tw numeric(12,4);
declare px_ttw numeric(12,4);
declare p_tw numeric(12,4);
declare p_ttw numeric(12,4);
```

```
declare pa_tw numeric(12,4);
```

```
declare pa_ttw numeric(20,4);
```

```
declare hasil char(40);
```

- Baca data training

o Hitung  $P(C_i)$  untuk setiap kelas

✓ `select count(*) into total_record from data_training;`

✓ `select count(*) into jml_tw from data_training where lama_studi='Tepat Waktu';`

✓ `select count(*) into jml_ttw from data_training where lama_studi='Tidak Tepat Waktu';`

✓ `select (jml_tw/total_record) into pjml_tw;`

✓ `select (jml_ttw/total_record) into pjml_ttw;`

o Hitung  $P(C_i)$  untuk setiap kriteria dalam kelas

✓ `select count(*) into ipk_4_tw from data_training where ipk_4=ipk_4x and lama_studi='Tepat Waktu' ;`

✓ `select count(*) into ipk_4_ttw from data_training where ipk_4=ipk_4x and lama_studi='Tidak Tepat Waktu';`

✓ `select count(*) into jenis_kelamin_tw from data_training where trim(jenis_kelamin)=trim(jenis_kelaminx) and lama_studi='Tepat Waktu';`

- ✓ `select count(*) into jenis_kelamin_ttw from data_training  
where trim(jenis_kelamin)=trim(jenis_kelaminx) and  
lama_studi='Tidak Tepat Waktu';`
  
- ✓ `select count(*) into kota_lahir_tw from data_training where  
trim(kota_lahir)=trim(kota_lahirx) and lama_studi='Tepat  
Waktu';`
  
- ✓ `select count(*) into kota_lahir_ttw from data_training  
where trim(kota_lahir)=trim(kota_lahirx) and  
lama_studi='Tidak Tepat Waktu';`
  
- ✓ `select count(*) into tipe_sekolah_tw from data_training  
where trim(tipe_sekolah)=trim(tipe_sekolahx) and  
lama_studi='Tepat Waktu';`
  
- ✓ `select count(*) into tipe_sekolah_ttw from data_training  
where trim(tipe_sekolah)=trim(tipe_sekolahx) and  
lama_studi='Tidak Tepat Waktu';`
  
- ✓ `select count(*) into kota_sekolah_tw from data_training  
where trim(kota_sekolah)=trim(kota_sekolahx) and  
lama_studi='Tepat Waktu';`

```
✓ select count(*) into kota_sekolah_ttw from data_training
where trim(kota_sekolah)=trim(kota_sekolahx) and
lama_studi='Tidak Tepat Waktu';
```

- o Hitung Probabilitas dari  $P(C_i)$  dari setiap kelas

```
select truncate(
(
(ipk_4_tw/jml_tw)*
(jenis_kelamin_tw/jml_tw)*
(kota_lahir_tw/jml_tw)*
(tipe_sekolah_tw/jml_tw)*
(kota_sekolah_tw/jml_tw)
),
14) into px_tw;
```

```
select truncate(
(
(ipk_4_ttw/jml_ttw)*
(jenis_kelamin_ttw/jml_ttw)*
(kota_lahir_ttw/jml_ttw)*
```



```
(tipe_sekolah_ttw/jml_ttw)*  
(kota_sekolah_ttw/jml_ttw)  
)  
14) into px_ttw;
```

```
select px_tw*pjml_tw into pa_tw;  
select px_ttw*pjml_ttw into pa_ttw;
```

- o Tampilkan hasil prediksi

```
select  if(pa_tw > pa_ttw,concat('Tepat Waktu'," ==>  
",truncate(pa_tw,4)," > ",truncate(pa_ttw,4)),  
        concat('Tidak Tepat Waktu'," ==>  
",truncate(pa_tw,4)," < ",truncate(pa_ttw,4)))  
        into hasil;  
  
return hasil;
```

## **BAB V**

### **HASIL DAN PEMBAHASAN**

Berdasarkan data yang ada, maka setelah dilakukan pengujian terhadap fungsi yang dibuat untuk menentukan kelas pada data testing diperoleh hasil dengan tingkat kesalahan klasifikasi sebanyak 20 record dari total data testing sebanyak 66 record. Jadi tingkat kesalahan prediksi sebesar  $20/66$  % atau 34%. Selanjutnya diberikan contoh penggunaan fungsi baik dengan menggunakan fungsi MySQL secara langsung maupun dihitung menggunakan microsoft excel terhadap data uji coba yang sama.

**Berikut diberikan contoh prediksi sebuah record data dengan data :**

IPK = 3

Jenis Kelamin = Perempuan

Kota Lahir = DALAM KOTA

Tipe Sekolah = Umum

Kota Sekolah = LUAR KOTA

Dengan menggunakan mysql query browser dan function bayesian yang telah dibuat maka diperoleh hasil prediksi seperti tampak pada gambar 5.1.



Gambar 5.1. Hasil uji coba procedure bayesian

Selain itu dilakukan perhitungan juga dengan menggunakan bantuan microsoft Excel dengan menggunakan data training dan data uji coba yg sama maka diperoleh hasil sebagai berikut :

$P(\text{Lama\_Studi}=\text{"Tepat Waktu"})$	163	0.815
$P(\text{Lama\_Studi}=\text{"Tidak Tepat Waktu"})$	37	0.185
$P(\text{Ipk\_4}=3 \text{Tepat Waktu})$		0.619632
$P(\text{Ipk\_4}=3 \text{Tidak Tepat Waktu})$		0.513514
$P(\text{Jenis\_kelamin}=\text{'Perempuan'} \text{Tepat Waktu})$		0.435583
$P(\text{Jenis\_kelamin}=\text{'Perempuan'} \text{Tidak Tepat Waktu})$		0.243243
$P(\text{Kota\_lahir}=\text{"DALAM KOTA"} \text{Tepat Waktu})$		0.460123
$P(\text{Kota\_lahir}=\text{"DALAM KOTA"} \text{Tidak Tepat Waktu})$		0.378378
$P(\text{Tipe\_sekolah}=\text{"Umum"} \text{Tepat Waktu})$		0.779141
$P(\text{Tipe\_sekolah}=\text{"Umum"} \text{Tidak Tepat Waktu})$		0.837838
$P(\text{Kota\_sekolah}=\text{"LUAR KOTA"} \text{Tepat Waktu})$		0.521472
$P(\text{Kota\_sekolah}=\text{"LUAR KOTA"} \text{Tidak Tepat Waktu})$		0.513514
$P(X \text{Lama\_Studi}=\text{"Tepat Waktu"})$		0.050457
$P(X \text{Lama\_Studi}=\text{"Tidak Tepat Waktu"})$		0.020334
$P(X \text{Lama\_Studi}=\text{"Tepat Waktu"})P(\text{Lama\_Studi}=\text{"Tepat Waktu"})$		0.041123

Waktu")  
 $P(X|Lama\_Studi="Tidak$   
 $TepatWaktu")P(Lama\_Studi="Tidak Tepat Waktu")$  0.003762

Kesimpulan Lama\_Studi Tepat Waktu

Berdasarkan uji coba dari prediksi terhadap data dengan spesifikasi nilai tiap faktor seperti sudah disebutkan di atas, maka diperoleh hasil yang sama baik dengan menggunakan formulasi microsoft excel maupun MySQL. Dimana hasil lama studinya adalah "Tepat Waktu".

Selain itu telah juga dilakukan uji coba lain yaitu dengan mengambil data training dan testing secara random dari master data yang digunakan dalam penelitian ini. Pengambilan secara random dilakukan dengan cara menggunakan fungsi random ( rand() ) di MySQL untuk merandom urutan record data. Selanjutnya diambil 75% record pertama atau 200 record dari keseluruhan data yang diperoleh yaitu 266 record sebagai data training dan 25% sisanya sebagai data testing. Kemudian dilakukan uji coba sebanyak 5 kali dan diperoleh hasil selisih kesalahan seperti tampak pada tabel 5.1.

Tabel 5.1. Tingkat error dengan data training diambil secara random

Kegiatan	Tingkat Error	
Uji Coba 1	0.242424	.24
Uji Coba 2	0.212121	.21
Uji Coba 3	0.287879	.29
Uji Coba 4	0.196970	

		.20
Uji Coba 5	0.212121	.21

Dari tabel 5.1. terlihat bahwa tingkat error atau kesalahan prediksi adalah sebesar antara 20% hingga 30%. Namun demikian meski error rate nya bisa dikatakan rendah, hal ini dapat sangat dimungkinkan terjadi karena kondisi data yang masih bias. Dimana ada beberapa record data yang memiliki nilai variabel/field prediktor yang sama dan mempunyai nilai class yang sama ataupun berbeda.

Selanjutnya dilakukan uji coba terhadap dengan menggunakan data training yang diambil pada uji coba 1 dengan membuang record-record yang mempunyai nilai field/faktor prediktor yang sama dari ke lima field prediktor. Dari hal itu diperoleh jumlah record sebanyak 42 record dari 200 record sebelumnya yang selanjutnya diberi nama tabel datatraining1\_x. Dengan menggunakan datatraining1\_x dilakukan prediksi terhadap data testing yang digunakan pada ujicoba 1 hingga ujicoba 5. Dari hari ujicoba tersebut diperoleh hasil seperti tampak pada tabel 5.2.

Tabel 5.2. Daftar tingkat kesalahan dari hasil uji coba yang dilakukan

Kegiatan	Data testing yg digunakan	Prediksi Benar	Prediksi Salah	Tingkat Kesalahan Prediski
Uji coba 1	Datatesting1	49	17	34.69388
Uji coba 2	Datatesting2	52	14	26.92308
Uji coba 3	Datatesting3	44	22	50
Uji coba 4	Datatesting4	51	15	29.41176471
Uji coba 5	Datatesting5	52	14	26.92307692
		Rata-rata		33.59036033

Dengan melihat hasil pada tabel 5.2 Maka terlihat bahwa tingkat kesalahan prediksi lebih tinggi dibanding pada saat menggunakan data yang mempunyai tingkat inkonsistensi yang cukup tinggi. Namun jika dilihat rata-rata kesalahan prediksi masih berkisar sama yaitu 34% (33.59036033).

Selain itu telah dilakukan uji prediksi terhadap data mahasiswa angkatan 2008/2009 untuk program studi S1 Sistem Informasi dan S1 Teknik Informatika dengan jumlah record data sebanyak 258 record. Dari hasil uji coba dengan data training yg mempunyai inkonsistensi data yang tinggi tersebut diperoleh hasil hanya 1 record yang diprediksi lama studinya "Tidak Tepat Waktu" sedangkan sisanya sebanyak 257 diprediksi "Tepat Waktu". Sedangkan saat digunakan datatraining1-x maka diperoleh hasil prediksi 4 record "Tidak Tepat Waktu" dan selebihnya atau 254 dipredikis tepat waktu.

## **BAB VI**

### **KESIMPULAN DAN SARAN**

#### **6.1. KESIMPULAN**

Berdasarkan dari percobaan yang telah dilakukan maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Lama masa studi atau dalam hal ini ketepatan masa studi setiap mahasiswa dapat diprediksi berdasarkan faktor-faktor yang berkaitan dengan latar belakang sekolah sebelumnya dan data akademik serta pribadi saat berada di perguruan tinggi.
2. Fungsi prediksi dengan memanfaatkan teknik data mining menggunakan algoritma naive bayes telah dapat dibuat dan digunakan untuk memprediksi (menentukan kelas) dari masa studi atau ketepatan masa studi dari mahasiswa dengan data training dan data testing yang telah diperoleh.
3. Tingkat kesalahan dari fungsi klasifikasi yang digunakan untuk prediksi masih berkisar pada 20% hingga 34% yang hal ini dimungkinkan dapat dipengaruhi oleh jumlah data training maupun testing yang digunakan serta tingkat konsisten data yang digunakan.

#### **6.2. SARAN**

1. Sebaiknya jumlah data yang digunakan training maupun testing ditambah hingga dapat diperoleh hasil akurasi fungsi algoritma yang lebih baik.

2. Untuk pengembangan ke depannya dimungkinkan dilakukan uji coba menggunakan algoritma lain yang hasilnya bisa dilakukan perbandingan dan analisa.
3. Cakupan faktor atau variabel prediktor dapat ditambah lebih banyak dan variasi nilai data yang juga bisa lebih banyak serta konsistensi data juga diperhatikan.



## DAFTAR PUSTAKA

- Abdul Kadir, "*Belajar Basis Data Dengan MySql*", Andi, Yogyakarta, 2008
- Han, J., Kamber, M. (2000). "*Data mining: Concepts and Techniques*". New York: Morgan-Kaufman.
- Kusrini, Emha Taufi Luthfi (2009), "*Algoritma Data Mining*", Andi Offset. Surabaya.
- Muhamad Hanief Meinanda, Metri Anisa, Narendi Muhandri, Kadarsyah Suryadi (2009), "Prediksi masa studi sarjana dengan *artificial neural network*", *Internetworking Indonesia Journal*, Vol.1 No.2, pp. 31-35.
- Nuqson Masykur Huda, 2010, "*Aplikasi data mining untk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro*", *Skripsi*, Program Studi Teknik Informatika Jurusan MIPA UNDIP, Semarang
- UNISBANK(2010), *Buku Peraturan akademik Universitas Stikubank (UNISBANK) Semarang tahun akademik 2010/2011*.

## DAFTAR RIWAYAT PENELITIAN (KETUA)

### I. DATA DIRI

1. Nama Lengkap : Arief Jananto, S.Kom, M.Cs
2. NIY : YS.2.97.03.006
3. Pangkat / Golongan : Penata Muda / III C
4. Jabatan Fungsional : Lektor
5. Tempat, Tgl. Lahir : Pemalang, 6 Januari 1974
6. Jenis Kelamin : Laki-laki
7. Alamat Rumah : Jl. Pucang Argo Tengah I/14 Pucanggading  
Demak
8. Telp / Faks / e-mail : 081325753989 / ..... /  
ajananto09@gmail.com
9. Alamat Kampus : Jl. Trilomba Juang 1 Semarang
10. Telp / Faks / e-mail : 8311668 / 8443240 / info@unisbank.ac.id

### II. RIWAYAT PENELITIAN

No	Judul	Tahun	Keterangan
1	Sistem Pendukung Keputusan Kontribusi Optimal Pada Produk Pompa Air dan Mesin Molen dengan Pendekatan Program Linear Metode Simplex (Studi Kasus CV. Suratman Surakarta)	2002	Anggota
2	Rancang Bangun Sistem Informasi Penelitian dan Pengabdian Masyarakat Pada LPPM Unisbank	2005	Anggota
3	Evaluasi Kinerja Sistem Informasi Penerimaan Siswa Kantor Dinas P dan K Semarang Tahun Ajaran 2005/2006	2005	Anggota
4	Pemanfaatan Macro Excel pada Sistem Informasi Akutansi Perusahaan Jasa	2007	Anggota
5	Pemanfaatan Object (Class) pada Pengembangan Model Menu Aplikasi Sistem Informasi	2007	Ketua
6	Rancang Bangun Program Aplikasi Pengelolaan Data Alumni dan Mencetak Kartu Alumni	2007	Anggota

7	Sistem Informasi Geografi Transportasi dan Jalur Angkutan Kota untuk Penataan Ruang Wilayah Kota Semarang guna Membantu Pengambilan Keputusan (Studi Kasus: Bagian Wilayah Kota III dan IV Kotamadya Daerah Tingkat II Semarang )	2008	Anggota
8	Perbandingan Performasi Algoritma Nearest Neighbor dan SLIQ Untuk Prediksi Kinerja Akademik Mahasiswa baru ( Studi kasus : Data akademik Mahasiswa Fakultas Teknologi Informasi Unisbank)	2010	Ketua

Demikian daftar riwayat penelitian ini dibuat dengan sebenarnya.

Semarang, 22 Januari 2013

Arief Jananto, S.Kom, M.Cs

## DAFTAR RIWAYAT PENELITI (ANGGOTA)

### I. DATA DIRI

1. Nama Lengkap : Eko Nur Wahyudi, S.Kom, M.Cs
2. NIP : 19710912 200501 1 001
3. Pangkat / Golongan : Penata Muda / III B
4. Jabatan Fungsional : Lektor
5. Tempat, Tgl. Lahir : Semarang, 12 September 1971
6. Jenis Kelamin : Laki-laki
7. Alamat Rumah : Jl. Mahesa Barat I/26, Kekancan Mukti  
Pedurungan Semarang
8. Telp / Faks / e-mail : 081 724 9891 / - / ewahyudi@ymail.com
9. Alamat Kampus : Jl. Trilomba Juang 1 Semarang
10. Telp / Faks / e-mail : 8311668 / 8443240 / info@unisbank.ac.id

### II. RIWAYAT PENELITIAN

No	Judul	Tahun	Keterangan
1	Analisa Profil Data Mahasiswa Baru Unisbank Tahun 2005 – 2010 Dengan Teknik Data Mining.	2010	Ketua
2	Membangun Corpus Secara Otomatis Dengan Berbagai Format Data Dari Hasil Crawling	2010	Anggota
3	Perancangan Sistem Pakar Untuk Mendeteksi Kerusakan Pada Perangkat Keras Komputer	2007	Anggota
4	Rancang Bangun Aplikasi Bank Data Soal Ujian Menggunakan XML	2006	Anggota
5	Survay Lulusan Program Studi Teknik Informatika Universitas Stikubank : Sebuah Analisa Prospek	2005	Anggota
6	Aplikasi Belajar Mandiri (Self Learning Application) Berbasis Multi Media Untuk Mata Kuliah Statistika	2004	Anggota

7	Pembuatan Software Database Perpustakaan Penelitian Dengan Menggunakan Delphi : Studi Kasus Kantor Badan Perencanaan Daerah (BAPEDA) Kabupaten Kendal	2003	Anggota
8	Faktorisasi Variabel Penerimaan Mahasiswa Baru Unisbank Semarang	2003	Anggota
9	Analisis Resiko Dalam Manajemen Proyek Pengembangan Sistem	2002	Anggota

Demikian daftar riwayat penelitian ini dibuat dengan sebenarnya.

Semarang, 22 Januari 2013

Eko Nur Wahyudi, S.Kom, M.Cs

## DAFTAR RIWAYAT PENELITI (ANGGOTA)

### DAFTAR RIWAYAT ANGGOTA PENELITI (MAHASISWA)

#### I. DATA DIRI

1. Nama Lengkap : Rohmad Agung Widodo
2. NIM : 11.01.55.0072
3. Tempat, Tgl. Lahir : Semarang, 18/08/1975
4. Jenis Kelamin : Laki-Laki
5. Alamat Rumah : JL SENDANG UTARA III NO. 39  
RT.06/08 KEL. GEMAH KEC.  
PEDURUNGAN
6. Telp / Faks / e-mail : 085269308024/.../  
[www.agoeng\\_75@yahoo.com](mailto:www.agoeng_75@yahoo.com)
7. Alamat Kampus : Jl. Trilomba Juang 1 Semarang
8. Telp / Faks / e-mail : 8311668 / 8443240 / info@unisbank.ac.id

#### II. RIWAYAT PENELITIAN

No.	Judul	Tahun	Keterangan

Demikian daftar riwayat penelitian ini dibuat dengan sebenarnya.

Semarang, 22 Januari 2013

Rohmad Agung Widodo

## DAFTAR RIWAYAT ANGGOTA PENELITI (MAHASISWA)

### I. DATA DIRI

1. Nama Lengkap : Nikke Febriliani
2. NIM : 09.01.55.0146
3. Tempat, Tgl. Lahir : Pekalongan, 02/02/1991
4. Jenis Kelamin : Perempuan
5. Alamat Rumah : JL MERPATI GG02 NO 317  
PEKALONGAN 51152
6. Telp / Faks / e-mail : 0285 4417464 / .. /
7. Alamat Kampus : Jl. Trilomba Juang 1 Semarang
8. Telp / Faks / e-mail : 8311668 / 8443240 / info@unisbank.ac.id

### II. RIWAYAT PENELITIAN

No.	Judul	Tahun	Keterangan

Demikian daftar riwayat penelitian ini dibuat dengan sebenarnya.

Semarang, 22 Januari 2013

Nikke Febriliani

## LOKASI PENELITIAN

