

Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa

Arief Jananto

Program Studi Sistem Informasi, Universitas Stikubank

email : ajananto09@gmail.com

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

PENDAHULUAN

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.

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, 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 paling lama 14 semester.

Untuk memprediksi dapat dilakukan dengan berbagai cara, salah satunya dapat pula dilakukan dengan menggunakan teknik datamining.

Perumusan Masalah

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.

TUJUAN DAN MANFAAT

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.

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.

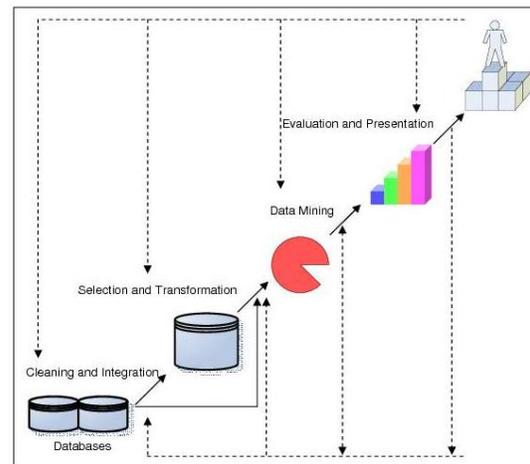
TELAAH PUSTAKA

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).

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. Langkah proses datamining dapat dilihat pada gambar 1.



Gambar 1. Data Mining adalah suatu langkah di dalam proses KDD (Han dan Kamber, 2001)

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

Penelitian Bidang Prediksi Lama Studi

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*.

METODE PENELITIAN

Obyek dan Tahapan Penelitian

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

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.

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.

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.

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 :

Tabel 1. Nilai IPK

IPK Asli	IPK Baru
IPK Asli > 3.00	3
2 <= IPK Asli <=3.00	2
IPK Asli < 2.00	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'.

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'.

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'.

Dan pada akhirnya diperoleh susunan data sebagai tampak pada tabel 1. sejumlah 266 record data yang merupakan data siap untuk di proses mining.

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

Implementasi Algoritma Naive Bayes

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

Adapun tahapan algoritma Naive bayes dalam query adalah :

a. Mulai

b. Baca data training

- 1) Hitung P(Ci) untuk setiap kelas
- 2) Hitung P(X|Ci) untuk setiap kriteria dan setiap kelas
- 3) Cari P(X|Ci) yang paling besar menjadi kesimpulan

c. Tampilkan hasil prediksi

d. Selesai

Lebih detail dalam bentuk query di MySQL :

a. Mulai

- 1) Siapkan variabel tambahan yang digunakan untuk menyimpan data hasil perhitungan.

```
BEGIN
declare total_record numeric;
declare jml_tw numeric;
declare jml_ttw numeric;
.....
declare pa_tw numeric(12,4);
declare pa_ttw numeric(20,4);
declare hasil char(40);
```

b. Baca data training

- 1) Hitung P(Ci) untuk setiap kelas

- (1) select count(*) into total_record from data_training;
- (2) select count(*) into jml_tw from data_training where lama_studi='Tepat Waktu';
- (3) select count(*) into jml_ttw from data_training where lama_studi='Tidak Tepat Waktu';
- (4) select (jml_tw/total_record) into pjml_tw;
- (5) select (jml_ttw/total_record) into pjml_ttw;

- 2) Hitung P(Ci) untuk setiap kriteria dalam kelas

- (1) select count(*) into ipk_4_tw from data_training where ipk_4=ipk_4x and lama_studi='Tepat Waktu' ;
- (2) select count(*) into ipk_4_ttw from data_training where ipk_4=ipk_4x and lama_studi='Tidak Tepat Waktu';
- (3) select count(*) into jenis_kelamin_tw from data_training where trim(jenis_kelamin)=trim(jenis_kelaminx) and lama_studi='Tepat Waktu';

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

```
(5) select count(*) into kota_lahir_tw from data_training where trim(kota_lahir)=trim(kota_lahirx) and lama_studi='Tepat Waktu';
```

```
(6) select count(*) into kota_lahir_ttw from data_training where trim(kota_lahir)=trim(kota_lahirx) and lama_studi='Tidak Tepat Waktu';
```

```
(7) select count(*) into tipe_sekolah_tw from data_training where trim(tipe_sekolah)=trim(tipe_sekolahx) and lama_studi='Tepat Waktu';
```

```
(8) select count(*) into tipe_sekolah_ttw from data_training where trim(tipe_sekolah)=trim(tipe_sekolahx) and lama_studi='Tidak Tepat Waktu';
```

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

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

- 3) Hitung Probabilitas dari P(Ci) dari setiap kelas

```
(1) 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;
```

- ```
(2) 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;

(3) select px_ttw*pjml_tw into pa_tw;
(4) select px_ttw*pjml_tw into pa_ttw;

4) Tampilkan hasil prediksi
(1) 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;
```

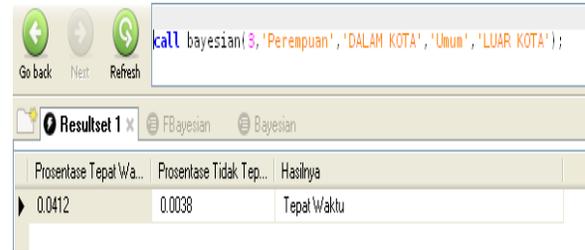
**HASIL DAN PEMBAHASAN**

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 data yang belum diketahui kelasnya sebagai berikut :

- 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 2.



Gambar 2. 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 :

**Tabel 3. Kesimpulan Lama Studi**

|                                                                     |     |             |
|---------------------------------------------------------------------|-----|-------------|
| P(Lama_Studi="Tepat Waktu")                                         | 163 | 0.815       |
| P(Lama_Studi="Tidak Tepat Waktu")                                   | 37  | 0.185       |
| P(Ipk_4=3 Tepat Waktu)                                              |     | 0.619632    |
| P(Ipk_4=3 Tidak Tepat Waktu)                                        |     | 0.513514    |
| P(Jenis_kelamin='Perempuan' Tepat Waktu)                            |     | 0.435583    |
| P(Jenis_kelamin='Perempuan' Tidak Tepat Waktu)                      |     | 0.243243    |
| P(Kota_lahir="DALAM KOTA" Tepat Waktu)                              |     | 0.460123    |
| P(Kota_lahir="DALAM KOTA" Tidak Tepat Waktu)                        |     | 0.378378    |
| P(Tipe_sekolah="Umum" Tepat Waktu)                                  |     | 0.779141    |
| P(Tipe_sekolah="Umum" Tidak Tepat Waktu)                            |     | 0.837838    |
| P(Kota_sekolah="LUAR KOTA" Tepat Waktu)                             |     | 0.521472    |
| P(Kota_sekolah="LUAR KOTA" Tidak Tepat Waktu)                       |     | 0.513514    |
| P(X Lama_Studi="TepatWaktu")                                        |     | 0.050457    |
| P(X Lama_Studi="Tidak TepatWaktu")                                  |     | 0.020334    |
| P(X Lama_Studi="TepatWaktu")P(Lama_Studi="Tepat Waktu")             |     | 0.041123    |
| P(X Lama_Studi="Tidak TepatWaktu")P(Lama_Studi="Tidak Tepat Waktu") |     | 0.003762    |
| Kesimpulan Lama_Studi                                               |     | 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 2.

**Tabel 4.** 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 2 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 3

**Tabel 5.** Daftar tingkat kesalahan dari hasil uji coba yang dilakukan

| Kegiatan   | Data testing yg digunakan | Prediksi Benar | Prediksi Salah | Tingkat Kesalahan Prediski |
|------------|---------------------------|----------------|----------------|----------------------------|
| Uji coba 1 | <code>Datatesting1</code> | 49             | 17             | 34.69388                   |
| Uji coba 2 | <code>Datatesting2</code> | 52             | 14             | 26.92308                   |
| Uji coba 3 | <code>Datatesting3</code> | 44             | 22             | 50                         |
| Uji coba 4 | <code>Datatesting4</code> | 51             | 15             | 29.41176471                |
| Uji coba 5 | <code>Datatesting5</code> | 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.

**KESIMPULAN DAN SARAN**

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.

Adapun saran yang dapat diberikan adalah :

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

- Han, J., Kamber, M. (2000). "*Data mining: Concepts and Techniques*". New York: Morgan-Kaufman.
- 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
- Kadir, K. (2008). "*Belajar Basis Data Dengan MySql*", Andi, Yogyakarta
- Kusrini, Luthfi, E.T. (2009). "*Algoritma Data Mining*", Andi Offset. Surabaya.
- Meinanda, M.H., Anisa, M., Muhandri, N., Suryadi, K. (2009). "*Prediksi masa studi sarjana dengan artificial neural network*", *Internetworking Indonesia Journal*, Vol.1 No.2, pp. 31-35.
- UNISBANK. (2010), *Buku Peraturan akademik Universitas Stikubank (UNISBANK) Semarang tahun akademik 2010/2011*.