

# APLIKASI PREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER (NBC) (STUDI KASUS : DI STMIK YADIKA BANGIL)

**Kurniawan Wahyu Haryanto<sup>1)</sup>, Risky Adi Saputra<sup>2)</sup>**

Teknik Informatika, STMIK Yadika Bangil,

email : kurniawan.wahyu@stmik-yadika.ac.id<sup>1)</sup>, riskydeliver16@gmail.com<sup>2)</sup>

**Abstract** :Tingkat kelulusan dianggap sebagai salah satu parameter efektivitas lembaga pendidikan. Jumlah siswa yang lulus tepat waktu menjadi indikator efektivitas perguruan tinggi baik negeri maupun swasta. Kelulusan yang tepat waktu merupakan masalah penting yang perlu ditangani secara bijak oleh lembaga pendidikan, karena penurunan angka kelulusan akan mempengaruhi akreditasi perguruan tinggi, sehingga dibutuhkan solusi untuk dapat meningkatkan masalah kelulusan tepat waktu. Penambahan data adalah proses melakukan analisis untuk mencari informasi dalam basis data. Algoritma yang digunakan untuk klasifikasi kelulusan adalah algoritma Naïve Bayes Classifier (NBC). The Naïve Bayes Classifier (NBC) adalah teknik prediksi berbasis probabilistik sederhana berdasarkan penerapan teorema bayes atau aturan dengan asumsi independensi yang kuat pada fitur, yang berarti bahwa fitur pada data Tidak terkait dengan ada atau tidak adanya fitur lain dalam data yang sama. Dalam tugas akhir ini, algoritma Naïve Bayes Classifier (NBC) digunakan untuk memprediksi masa studi siswa dengan kelulusan tepat waktu dari hasil yang diperoleh dan sistem hanya sebagai pembantu siswa untuk memecahkan masalah dalam menentukan prediksi masa studi mahasiswa Teknik Informatika STMIK Yadika Bangil. Dari algoritma yang dipilih dapat menampilkan NIM, Nama siswa, IPK, Grade, didukung oleh ilmu probabilitas dan ilmu statis dalam penggunaan data pelatihan sebagai pendukung keputusan klasifikasi.

**Kata kunci:** Tingkat Kelulusan Diprediksi, Data Mining, Naïve Bayes Classifier.

## 1. Pendahuluan

### 1.1 Latar Belakang

Jurusan teknik informatika STMIK Yadika Bangil merupakan program pendidikan yang memiliki beban sekurang-kurangnya 148 (seratus lima puluh) sks (satuan kredit semester) yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8(delapan) semester paling lama 14 (empat belas) semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) reguler di jurusan Teknik Informatika STMIK Yadika Bangil yang menempuh lama studi dari 8 semester dari yang dijadwalkan 8 semester. Sebutan mahasiswa abadi atau mahasiswa paling lama (mapala) yang kuliah S1 hingga tujuh tahun (14 semester) sudah tidak ada lagi. Pasalnya Kementerian Pendidikan dan Kebudayaan (Kemendikbud) mengeluarkan kebijakan lama kuliah sarjana S1 yaitu 4-5 tahun saja. Aturan baru ini tertuang dalam Permendikbud 49/2014 tentang Standar Nasional Pendidikan Tinggi (SNPT) . Jadi peneliti dapat menarik sebuah kesimpulan bahwa keputusan baru dengan lama studi untuk S1 yaitu maksimal 5 (lima) tahun dan S2 selama 4 (empat) tahun jika lebih dari yang sudah ditetapkan maka terancam di DO.

Oleh karena itu, untuk meningkatkan kelulusan tepat waktu disuatu perguruan tinggi, maka dibutuhkan sebuah aplikasi prediksi masa studi mahasiswa menggunakan algoritma Naïve Bayes Classifier . Aplikasi tersebut diharapkan dapat membantu dalam memprediksi masa studi mahasiswa lebih dini, sehingga pihak jurusan Teknik Informatika STMIK Yadika Bangil dapat melakukan tindakan-tindakan yang diperlukan agar mahasiswa dapat lulus tepat 8 semester dari yang sudah ditetapkan.

### 1.2 Rumusan Masalah

Adapun perumusan masalah berdasarkan latar belakang yang telah diuraikan sebagai berikut :

- a. Bagaimana membuat aplikasi dengan menerapkan algoritma Naïve Bayes Classifier (NBC)?
- b. Bagaimana menentukan arsitektur Naïve Bayes Classifier (NBC) paling sesuai dengan persoalan yang ada ?
- c. Bagaimana menguji performance aplikasi dengan algoritma Naïve Bayes Classifier (NBC) untuk pengklasifikasian data?

**1.3 Batasan Masalah**

Didalam penulisan skripsi ini batasan – batasan masalah perlu ditetapkan agar tidak menyimpang dari maksud tujuan dari penelitian yang semula direncanakan sehingga mempermudah mendapatkan data dan informasi yang diperlukan, maka penulis menetapkan batasan-batasan sebagai berikut:

- a. Prediksi masa studi mahasiswa teknik informatika berdasarkan data nilai akademik. Data nilai akademik berupa Indeks Prestasi Kumulatif (IPK) dari semester 1 sampai semester 6.
- b. Data latih berupa data riwayat studi mahasiswa angkatan 2012 yang telah lulus dengan IPK semester 1-6 yang berjumlah 113 data mahasiswa.
- c. Data uji berupa data IPK studi mahasiswa yang telah lulus berjumlah 30 data.

**1.4 Tujuan Penelitian**

Tujuan dari penelitian yang dilakukan adalah sebagai berikut :

- a. Menerapkan dan mengimplementasi algoritma Naïve Bayes Clasification untuk memprediksi masa studi mahasiswa .
- b. Mengetahui prediksi lama masa studi dari setiap mahasiswa yang telah menempuh kuliah hingga minimal semester VI dengan menggunakan dasar data kelulusan sebenarnya.
- c. Membuat aplikasi dengan menerapkan algoritma Naive Bayes Classifier (NBC)

**2. Metode Penelitian**

Penelitian studi kasus ini menggunakan penelitian pendekatan kualitatif, Adapun langkah-langkah metode penelitian melalui beberapa tahap sebagai berikut :

- a. Tahap pendahuluan, meliputi : identifikasi masalah, tinjauan pustaka
- b. Tahap perancangan, meliputi : analisis kebutuhan sistem, desain sistem dan kontruksi sistem
- c. Tahap penyelesaian, meliputi : testing dan pelaporan

**2.1 Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan di STMIK Yadika Bangil. Waktu penelitian mulai terhitung bulan November 2016 dan Observasi pada bulan Desember 2016 – Februari 2017.

**2.2 Metode Pengumpulan Data**

Teknik pengumpulan data yang dilakukan melalui langkah-langkah seperti di bawah ini :

- a. Study kasus
- b. Observasi
- c. Studi Pustaka
- d. Studi Literatur Sejenis

**2.3 Metodologi Pengembangan Sistem**

Metode pengembangan sistem yang digunakan pada penelitian ini adalah model pengembangan sistem sekuensial linier atau disebut pula sebagai model air terjun (waterfall).

**3. Perancangan Sistem**

**3.1 Analisa Kebutuhan Sistem**

- a. Kebutuhan Hardware

No	Spesifikasi	Kebutuhan
1	CPU	AMD Quad-Core A8-7410 APU
2	RAM	4GB DDR3L-1600 SDRAM
3	HardDisk	500GB 5400 rpm SATA
4	Layar	14" HD LED
5	Graphics	Radeon R5
6	Wifi	802.11b/g/n

- b. Kebutuhan Software

No	Spesifikasi	Kebutuhan
1	OS	Windows, 7, 8, 10
2	Web Server	XAMPP (Apache Web Server)
3	Database Server	MySQL (PhpMyAdmin)
4	Web Browser	Mozilla Firefox, Google Chrome
5	Design Proses, Design Basis Data	Power Designer, Microsoft Visio 2013

**3.2 Perhitungan Naïve Bayes**

- a. Kriteria Penilaian

Dalam Studi Kasus ini, kriteria yang digunakan dalam sistem ini sesuai dengan data kebutuhan Mitra Statistik di BPS Kota Pasuruan yang ada, antara lain :

No	Kriteria	Jenis Data
1	Jenis Kelamin	Kategori
2	IPK	Numeric
3	Nilai Grade	Kategori
4	IP	Numeric

**b. Data Training**

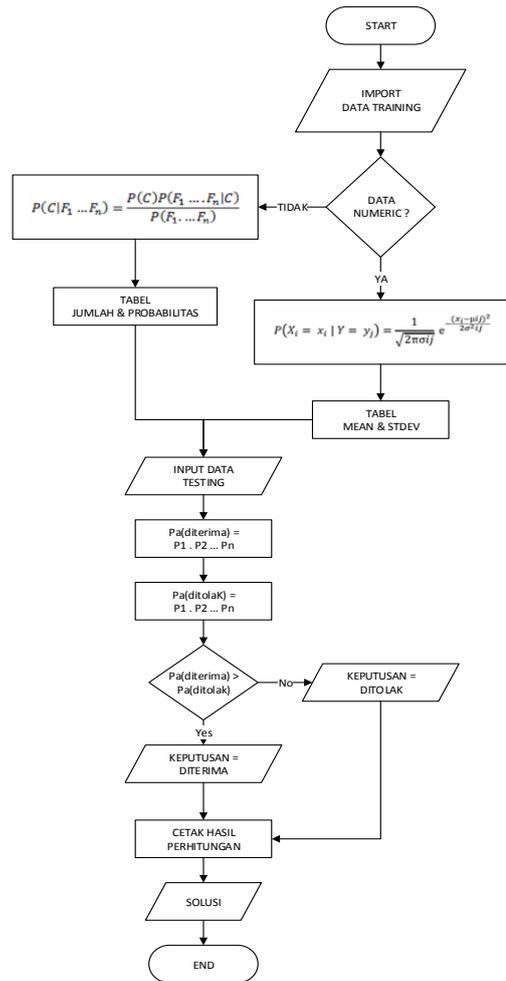
Data Training merupakan data yang penting dalam implementasi metode naïve bayes pada Proses Prediksi Masa Studi Mahasiswa di STMIK Yadika Bangil ini. Dimana probabilitas masing-masing kriteria didapat bersumber dari data training tersebut. Dalam kasus ini, peneliti mengambil data training dari Prediksi Masa Studi Mahasiswa yang telah lulus angkatan 2012 yang berjumlah 113 data, mahasiswa seleksi Mitra Statistik sebelumnya yaitu tetepat waktu berjumlah 82 mahasiswa dan terlambat berjumlah 31 mahasiswa.

**c. Data Testing**

Data Testing merupakan data uji dalam implementasi metode naïve bayes pada Proses Prediksi Masa Studi Mahasiswa di STMIK Yadika Bangil ini. Dimana data ini digunakan sebagai pengujian akurasi dalam perhitungan naïve bayes pada aplikasi prediksi masa studi mahasiswa di STMIK Yadika Bangil. Dalam kasus ini, peneliti mengambil data testing dari mahasiswa STMIK Yadika Bangil angkatan 2012 berjumlah 113 data mahasiswa.

**d. Perhitungan Naïve Bayes**

Proses Perhitungan metode Naïve Bayes ini dapat dilihat dalam FlowChart Proses berikut ini :



Dilihat dari FlowChart Proses diatas Langkah Pertama yang dilakukan adalah mengklasifikasi data training berkategori berdasarkan hasil keputusan diterima dan ditolak. Berikut Tabel Data Klasifikasi dari Data Training data berkategori Aplikasi prediksi masa studi mahasiswa di STMIK Yadika Bangil.

Tabel 3.3 Data Klasifikasi Kategori

Kriteria		Tepat	Terlambat
Jenis Kelamin	L	14	3
	P	68	28

Langkah selanjutnya, yaitu menghitung probabilitas data berkategori dengan rumus

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

dimana P(C) P(F1... Fn |C) adalah probabilitas independen kelas C dari semua fitru dalam vektor

(F1...Fn) yang merupakan data klasifikasi data training dan P (F1... Fn) adalah probabilitas awal kelas X yang merupakan jumlah data atribut berdasarkan hasil keputusan.

Seperti contoh pada kategori laki-laki di atribut Jenis Kelamin yang tepat waktu sebanyak 82 dan yang terlambat sebanyak 31 mahasiswa, sedangkan jumlah dari keseluruhan kategori Jenis Kelamin = L yang Tepat Waktu sebanyak 14, dan yang Terlambat sebanyak 3. Maka dapat diambil probabilitasnya sebagai berikut :

$$P(\text{Jenis Kelamin=L} \mid \text{Tepat Waktu}) = 14 / 82 = 0,170$$

$$P(\text{Jenis Kelamin=L} \mid \text{Terlambat}) = 3 / 82 = 0,036$$

Setelah dilakukan perhitungan seluruh probabilitas data berkategori. Diperoleh Tabel Data Probabilitas berkategori sebagai berikut :

Tabel 3.4 Data Probabilitas Kategori

Kriteria	Tepat Waktu		Terlambat
	L	P	
Jenis Kelamin	L	0.170731707	0.036585366
	P	0.829268293	0.341463415

Langkah berikutnya yaitu menghitung probabilitas data numeric pada kriteria Usia,

KRITERIA	MEAN	
	TEPAT	TERLAMBAT
IPK	3.16066	1.31091
NILAI	A	12.03659
	AB	10.42683
	B	14.31707
	BC	4.79268
	C	3.03659
	D	0.15854
E	0.01220	6.19355

Berat Badan, dan Tinggi Badan. data numeric digunakan rumus Normal Distribusi Probability yaitu :

$$P(X_i = x_i \mid Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Dimana pada parameter  $\mu_{ij}$  dapat diestimasi berdasarkan sampel mean  $X_i$  (x) untuk seluruh training record yang dimiliki kelas  $Y_j$  dan  $\sigma_{ij}^2$  dapat diestimasi dari sampel varian ( $s^2$ ) training record tersebut. Dalam hal ini, perhitungan naïve bayes pada data numeric menggunakan distribusi

dikarakterisasi dengan dua parameter, yaitu mean,  $\mu$ , dan varian  $\sigma^2$  (sampel Varian / Standar Deviasi) karena data tersebut dapat diasumsikan dalam bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan dapat diperkirakan parameter distribusi dengan data training. Maka diperoleh rumus dengan parameter yang ada yaitu :

$$P(X|Y) = \frac{1}{\sqrt{2\pi S}} \cdot \exp \frac{-(x - \text{Mean})^2}{2 \cdot S^2}$$

Keterangan :

SD = (Standar Deviasi)

X = Data Testing

Mean = Rata-rata dengan Y

$\pi$  = phi (3,14)

$\exp$  = 2,718

Langkah awal untuk menentukan probabilitas data numeric yaitu melakukan perhitungan mean pada setiap atribut dalam data numeric yang ada. Sebagai contoh pada atribut IPK Semester 6 dengan keputusan (Y) tepat waktu dengan data Jumlah Data IPK Semester 6 yang Tepat Waktu yaitu, dilakukan perhitungan dengan rumus:

$$X(\text{IPK Smt 6} \mid \text{Tepat Waktu}) =$$

Penjumlahan Data dengan Y=tepat waktu

$$\frac{\text{Banyaknya Data}}{82} = \frac{259,12}{82} = 3,160$$

Berikut tabel Perhitungan Mean pada Aplikasi Prediksi Masa Studi Mahasiswa di STMIK Yadika Bangil.

Tabel 3.5 Mean Data Numeric

Langkah selanjutnya yaitu melakukan perhitungan varian  $\sigma^2$  (sampel Varian / Standar Deviasi). Sebagai contoh pada atribut Usia dengan keputusan (Y) Tepat Waktu di STMIK Yadika Bangil dengan data Jumlah Data IPK Semester 6 yang tepat waktu yaitu, Digunakan rumus yang digunakan sebagai berikut :

$$S_{(IPK \text{ Smt } 6 | \text{ Tepat Waktu})} = \sqrt{\frac{(n1 - \text{mean})^2 + (n2 - \text{mean})^2 - (nx - \text{mean})^2}{\text{banyaknya data} - 1}}$$

$$= \sqrt{\frac{18.9378}{82 - 1}}$$

$$= \sqrt{\frac{18.9378}{81}} = \sqrt{0,2338}$$

$$= 0,48357$$

Dengan demikian didapat tabel Perhitungan Standar Deviasi pada Aplikasi Prediksi Masa Studi Mahasiswa di STMIK Yadika Bangil.

Tabel 3.6 Standar Deviasi Data Numeric

KRITERIA	STANDAR DEVIASI	
	TEPAT	TERLAMBAT
IPK	0.48357	1.04776
NILAI	A	6.77950
	AB	3.75164
	B	4.35147
	BC	3.27632
	C	2.36980
	D	0.92255
E	0.11043	4.43786

Perhitungan selanjutnya yaitu menentukan probabilitas numeric akhir, Seperti contoh : Pada kategori IPK Semester 6 dengan data testing bernilai 2,90 untuk Hasil Keputusan Tepat Waktu. Maka dapat diambil probabilitasnya sebagai berikut:

$$P(X|Y) = \frac{1}{\sqrt{2\pi}S} \cdot \exp\left\{-\frac{(x - \text{Mean})^2}{2 \cdot S^2}\right\}$$

$$P_{(IPK \text{ Smt } 6=2,9 | \text{ Tepat Waktu})} = \frac{1}{\sqrt{2,314 \cdot 0,48357}} \cdot 2,718 \cdot \frac{\exp\left\{-\frac{(2,90 - 3,16066)^2}{2 \cdot 0,2338}\right\}}{2 \cdot 0,2338} = 0.71188$$

Langkah terakhir dalam perhitungan naïve bayes yaitu menghitung probabilitas akhir dengan melakukan perkalian seluruh kriteria pada setiap kelas hasil keputusan. Seperti contoh menghitung Prediksi Masa Studi Mahasiswa dengan kriteria :

- Jenis Kelamin = P
- IPK Semester 6 = 2,92
- Grade Nilai A = 3
- Grade Nilai AB = 5
- Grade Nilai B = 17
- Grade Nilai BC = 6

- Grade Nilai C = 7
- Grade Nilai D = 1
- Grade Nilai E = 0

Kemudian dilakukan perhitungan setiap kriteria sesuai dengan table perhitungan tersebut di atas maka didapatkan nilai probabilitas sebagai berikut:

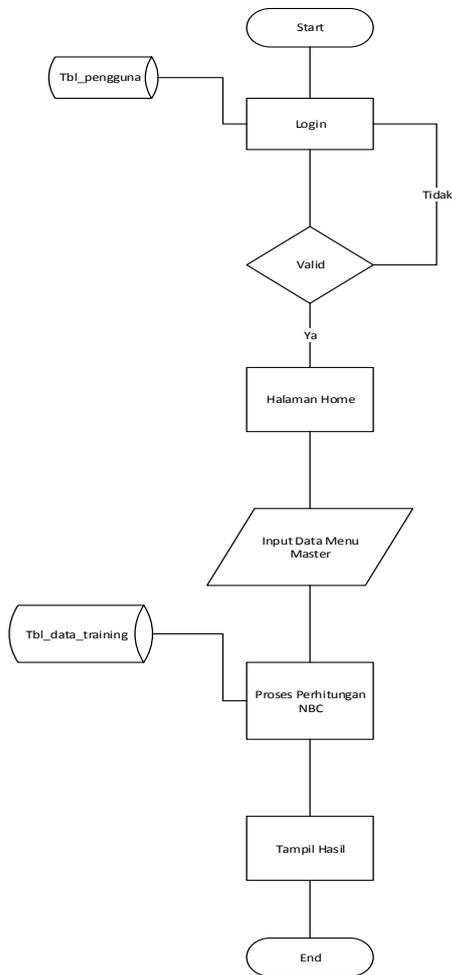
Tabel 3.7 Nilai Probabilitas Akhir

Kriteria (P)	Tepat Waktu (C)	Terlambat (C)
Jenis Kelamin = P	0.829268293	0.3414634
IPK Smt 6 = 2,92	0.72623	0.1176933
Grade Nilai A	0.02421	0.1661168
Grade Nilai AB	0.0373529	0.0908455
Grade Nilai B	0.0758102	0.0091389
Grade Nilai BC	0.1137725	0.096898
Grade Nilai C	4.247E-05	0.0493822
Grade Nilai D	0.2852773	0.1948263
Grade Nilai E	3.5906151	0.0339461
Naïve Bayes Probability	2.043E-10	1.754E-10

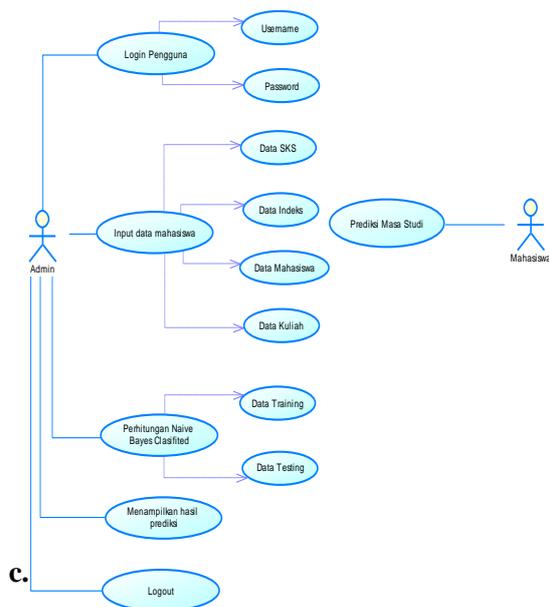
Dilihat dari hasil akhir perhitungan naïve bayes pada Contoh di atas, Nilai probabilitas Kriteria pada hasil keputusan “Diterima” sebesar 2.043E-10 sedangkan pada hasil keputusan “Ditolak” sebesar 1.754E-10. Karena nilai probabilitas tepat waktu lebih besar daripada nilai probabilitas terlambat, Maka dapat disimpulkan bahwa keputusan untuk Calon Mitra Statistik dengan kriteria tersebut di atas adalah “**Tepat Waktu**”.

### 3.3 Desain Proses

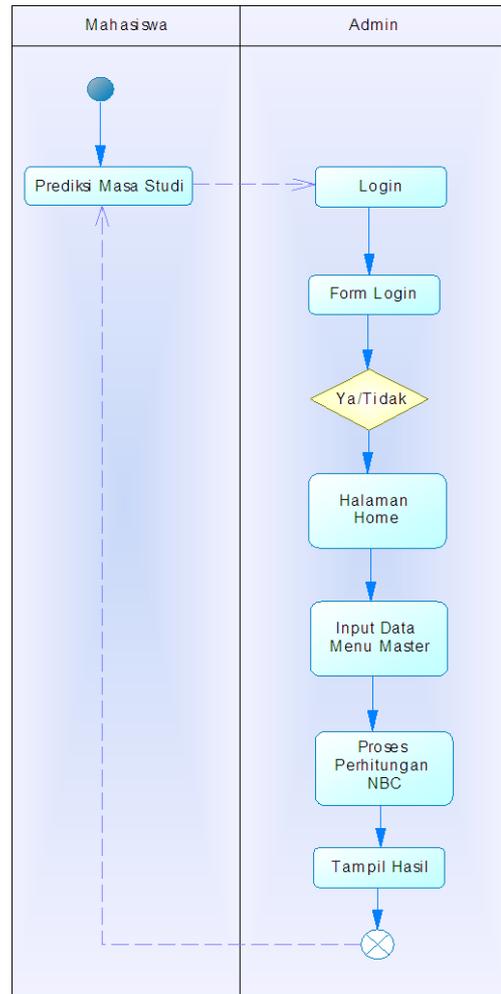
#### a. Flowchart



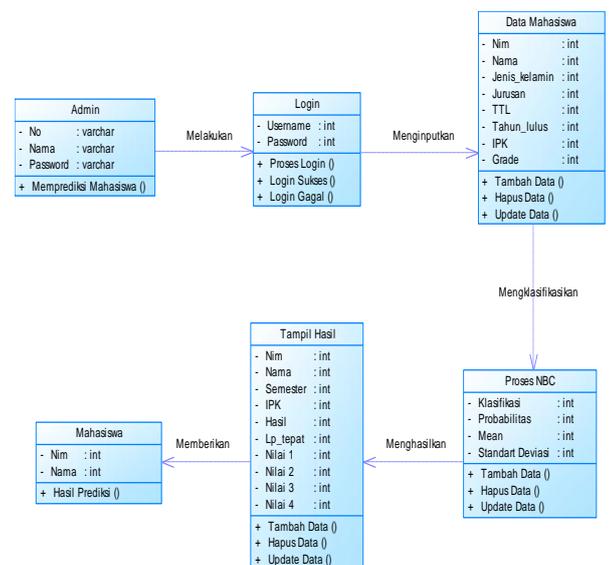
#### b. Use Case Diagram



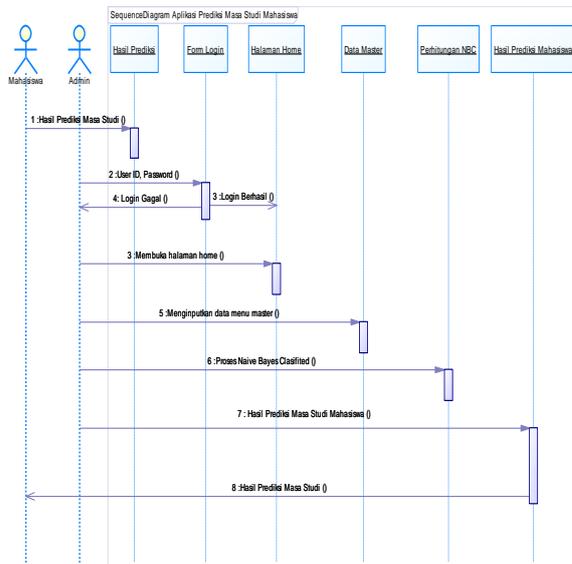
#### d. Activity Diagram



#### d. Class Diagram



**e. Squence Diagram**



**Hasil dan Pembahasan**

**4.1 Implementasi Sistem**

Teknologi yang digunakan dalam pengembangan sistem ini adalah teknologi aplikasi berbasis vb.net, yang membentuk sebuah program yang dapat berdiri sendiri dan dioperasikan dengan aplikasi prediksi masa studi mahasiswa. Pada aplikasi prediksi masa studi mahasiswa ini terdapat beberapa primary interfaces atau tampilan utama yang meliputi: halaman login, home, data master, metode, data hasil.

**4.2 Penjelasan Program**

**a. Halaman Login**

Tampilan awal adalah halaman login yang digunakan untuk validasi setiap pengguna yang akan masuk ke sistem. User pengguna harus Memasukkan username dan password kemudian pilih login. Berikut tampilan halaman login pada aplikasi prediksi masa studi mahasiswa:.



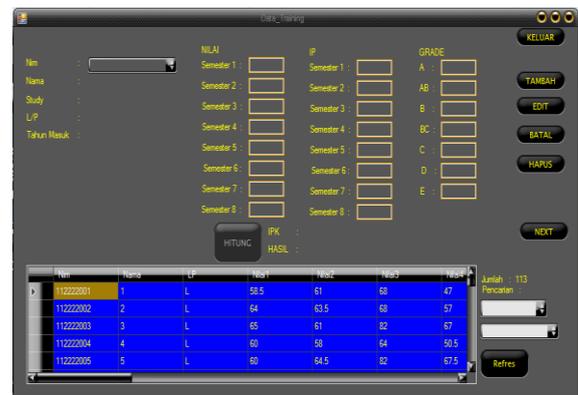
**b. Halaman Home**



Setelah melakukan Proses Login, kita akan masuk pada sebuah halaman Home Sistem dengan berbagai menu di dalamnya meliputi : data master, penghitungan metode, dan tampil hasil prediksi. Berikut tampilan Home pada aplikasi prediksi masa studi mahasiswa:

**c. Data Training**

Training merupakan sumber data dari nilai probabilitas, dimana nilai tersebut nantinya di gunakan untuk memprediksi dari data hasil sebenarnya dengan data hasil prehitungan dari algoritma Naive Bayes Classifier (NBC).



**d. Data Testing**

Testing adalah sub menu untuk memprediksi mahasiswa dimana dengan menginputkan data mahasiswa yang ingin di prediksi dan memasukan grade, dimana akan keluar hasil prediksi lulus tepat waktu atau lulus terlambat.

Kode	Nama	Semester	SKS	Grade
MKS-2401	Sistem Operasi	1	2	
MKS-2201	Aljabar Linear	1	2	
MKS-2211	Kalkulus	1	4	
MKS-2403	Algoritma dan Pe...	1	3	
MKS-2203	Bahasa Inggris K...	1	3	
MKS-2201	pendidikan agama	1	4	
MKS-2302	Pendidikan Pans...	1	2	
MKS-2402	Jaringan Komputer	2	4	
MKS-2105	Pakikan Statist...	2	1	
MKS-2204	Herika Komput...	2	2	
MKS-2205	Matematika Diskr...	2	2	

**. Simpulan**

**5.1 Kesimpulan**

Berdasarkan hasil dari aplikasi prediksi masa studi mahasiswa menggunakan algoritma Naive Bayes Classifier (NBC), maka dapat disimpulkan bahwa :

- Sistem ini digunakan untuk memprediksi masa studi mahasiswa dengan menggunakan algoritma Naive Bayes Classifier (NBC yang diimplementasikan menggunakan Bahasa Pemrograman Visual Basic.Net,dan Syntax MySQL
- Dari algoritma terpilih dapat menampilkan NIM, Nama Mahasiswa, IPK, Grade, Prediksi kelulusan tepat waktu yang merupakan hasil klasifikasi datamining dengan menggunakan Naive Bayes Classifier (NBC)
- Algoritma Naive Bayes didukung oleh ilmu Probabilistik dan ilmu Statistika khususnya dalam penggunaan data training untuk mendukung keputusan pengklasifikasian. Pada algoritma Naive Bayes, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting.
- Hasil yang diperoleh dari sistem ini hanya sebagai pendukung bagi pengguna untuk menyelesaikan permasalahan dalam menentukan prediksi masa studi mahasiswa.

**5.2 Saran**

Adapun saran yang dapat diberikan untuk mengembangkan penelitian ini adalah sebagai berikut :

- baik Metode Naive Bayes Classifier (NBC) digunakan untuk menghitung

probabilitas dengan kemungkinan tepat waktu atau terlambat dalam menentukan prediksi kelulusan mahasiswa.

- Untuk pengembangan selanjutnya, hendaknya pengujian sistem dilakukan dengan melibatkan lebih banyak data. Semakin banyak data yang terlibat dalam pengujian sistem maka kinerja dari sistem yang dibuat akan diketahui dengan.

**Daftar Pustaka**

Arief Jananto, *Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa, Jurnal Teknologi Informasi DINAMIK, vol 18, no.1, Januari 2013.*

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.

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.*

Prabowo. 2012. *Aneka Teknik, Piranti dan Penerapan Data Mining: Studi Kasus Peramalan Harga Saham Industri Telekomunikasi Berbasis Jaringan Saraf Tiruan*. Modul Perkuliahan Universitas Budi Luhur.