

# IMPLEMENTASI ALGORITMA DATA MINING NAIVE BAYES UNTUK PREDIKSI KELULUSAN MAHASISWA

Diqy Fakhrun Shiddieq, S.T., M. Kom.<sup>1</sup>, Patricia<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Program Studi Teknik Informatika, STMIK LPKIA  
BANDUNG STMIK LPKIA BANDUNG, Jln. Soekarno Hatta No. 456 Bandung  
40266 <sup>1</sup>diqy@lpkia.ac.id, <sup>2</sup>patty.tricia05@gmail.com

---

## Abstrak

Tujuan dari prediksi kelulusan mahasiswa adalah untuk menemukan informasi tentang kelulusan mahasiswa berdasarkan IPK yang diperoleh, serta dapat membantu pihak manajemen atau pimpinan dalam mencari solusi dan kebijakan untuk prestasi mahasiswa agar dapat menyelesaikan studinya dengan hasil yang memuaskan. Prediksi kelulusan mahasiswa ini menggunakan Metode *Data Mining*. *Data Mining* dapat menggali dan mengekstrak data yang belum diketahui dan menghasilkan informasi baru. *Data mining* dapat mengenali pola dari pemanfaatan data historis kelulusan mahasiswa dengan menggunakan Metode Algoritma *Naïve Bayes*. Perancangan basis data untuk prediksi kelulusan mahasiswa menggunakan suatu model data berorientasi objek dan metode pemrograman berorientasi objek, sehingga produktivitas program dapat meningkat. Pemrograman yang akan dibuat, menggunakan framework *CodeIgniter* dengan database MySQL. *CodeIgniter* digunakan untuk melakukan import dan perhitungan proses mining data historis kelulusan mahasiswa. Program yang dibangun ini akan bekerja untuk memprediksi kelulusan mahasiswa Berdasarkan data historis di STMIK & Politeknik LPKIA Bandung.

**Kata kunci:** *Data mining, Naïve Bayes, probabilitas, prediksi kelulusan mahasiswa.*

---

## 1. PENDAHULUAN

Perkembangan dunia Pendidikan semakin maju dan semakin bertambah pesat jumlahnya. Berdasarkan undang - undang Republik Indonesia No. 12 Tahun 2012 Tentang Pendidikan Tinggi BAB 3 Pasal 1 Ayat 2, pendidikan merupakan sarana untuk mendapatkan suatu ilmu atau pengalaman yang berguna untuk mengembangkan otak bagi orang yang menempuhnya. Pendidikan suatu kebutuhan yang sangat penting untuk dilaksanakan bagi siapa saja untuk mencerdaskan dan mencetak generasi generasi bangsa yang berkompeten. Salah satunya perguruan tinggi merupakan tempat atau sarana untuk Pendidikan jenjang sarjana. Perguruan tinggi di Indonesia dibagi menjadi dua, yaitu perguruan tinggi negeri (PTN) dan perguruan tinggi swasta (PTS). STMIK dan Politeknik Komputer Bisnis LPKIA Bandung merupakan salah satu perguruan tinggi swasta (PTS) yang ada di Indonesia.

STMIK dan Politeknik Komputer Bisnis LPKIA memiliki moto yaitu sebuah kampus bersih lulusannya mudah bekerja dan bersertifikasi internasional. LPKIA memiliki banyak data mengenai mahasiswanya yang meliputi data induk mahasiswa dan data kelulusan mahasiswa mulai dari tahun 1984 sampai dengan sekarang. Dengan banyaknya data yang ada, maka data – data tersebut belum bisa dimanfaatkan secara optimal. Tentunya

pemanfaatan data dapat membantu dalam pengambilan keputusan dengan menganalisa data yang ada menjadi informasi – informasi yang baru. Metode yang digunakan untuk membantu dalam menentukan informasi - informasi tersebut yaitu diperlukannya teknik data mining.

Data mining adalah teknik untuk menemukan dan mendeskripsikan pola – pola yang ada dalam data sebagai sebuah alat untuk membantu menjelaskan data tersebut dan membuat prakiraan dari data itu [1]. Untuk memanfaatkan data – data yang sudah ada tersebut, metode yang digunakan adalah *Naïve Bayes*.

Digunakannya metode *Naïve Bayes* dalam penelitian ini yaitu *Naïve Bayes* merupakan salah satu metode klasifikasi yang memiliki tingkat akurasi yang lebih baik dibandingkan Metode classifier lainnya. Hal tersebut dibuktikan oleh *Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone* [1]. Tujuan untuk menemukan informasi tentang kelulusan mahasiswa berdasarkan IPK mahasiswa yang diperoleh, dan dapat membantu pihak manajemen atau pimpinan dalam mencari solusi dan kebijakan untuk meningkatkan prestasi mahasiswa agar dapat menyelesaikan studinya dengan hasil yang memuaskan.

Dari uraian latar belakang masalah diatas, identifikasi permasalahan yang ada yaitu data mahasiswa belum dimanfaatkan secara optimal untuk melakukan prediksi kelulusan mahasiswa sehingga manajemen atau pimpinan kesulitan dalam mengambil keputusan.

Adapun batasan masalah dalam penelitian ini adalah:

1. Data yang diolah adalah data mahasiswa dan data kelulusan mahasiswa Prodi Manajemen Informatika jenjang D3 dan S1 dari tahun 2013 – 2015.
2. Data yang digunakan dalam proses mining adalah menggunakan data asumsi yang sudah melalui proses tahap *cleaning*.
3. Dalam penelitian hanya menggunakan beberapa atribut yaitu Jurusan, asal Sekolah, keterangan nilai NEM, dan keterangan nilai IPK.
4. Metode yang digunakan adalah data mining dengan algoritma Naive Bayes.
5. Hasil output dari penelitian ini yaitu berupa prediksi kelulusan mahasiswa dengan predikat cukup, memuaskan dan pujian.

Adapun tujuan dari penelitian ini yaitu, memanfaatkan data mahasiswa dengan optimal untuk menampilkan prediksi kelulusan mahasiswa agar mempermudah manajemen atau pimpinan mengambil keputusan untuk strategi pengembangan mahasiswa.

## 2. LANDASAN TEORI

### 2.1 Data Mining

Data mining adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara interaktif pada *database* yang besar, dengan tujuan mengekstrak informasi dan *knowledge* yang akurat dan berpotensi berguna untuk *knowledge workers* yang berhubungan dengan pengambilan keputusan dan pemecahan masalah [3].

### 2.2 Naive Bayes

*Naive Bayes* merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Metode ini pertama kali dikenalkan oleh ilmuwan Inggris *Thomas Bayes*, yaitu digunakan untuk memprediksi peluang yang terjadi di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes* [3].

*Bayes rule* digunakan untuk menghitung probabilitas suatu class. Algoritma *Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi.

Metode *Bayes rule* digunakan dan diterapkan untuk melakukan perhitungan terhadap posterior dan probabilitas dari data sebelumnya. Dalam analisis bayesian, fungsi klasifikasi akhir dihasilkan dengan menggabungkan kedua sumber informasi (*prior* dan *posterior*) untuk menghasilkan probabilitas menggunakan aturan *Bayes* [3]. Adapun tahapan algoritma *Naive Bayes*, yaitu:

1. Menghitung jumlah kelas / label
2. Menghitung jumlah kasus per kelas
3. Kalikan semua variabel kelas
4. Bandingkan hasil per kelas

Metode algoritma *Naive Bayes* merupakan penyederhanaan metode *Bayes*. Untuk mempermudah pemahaman, maka *Teorema Bayes* disederhanakan menjadi

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

Keterangan:

*X*: Data dengan class yang belum diketahui

*H*: Hipotesis data *X* merupakan suatu class spesifik

*P(H|X)*: Probabilitas hipotesis *H* berdasar kondisi *X* (posteriori probability)

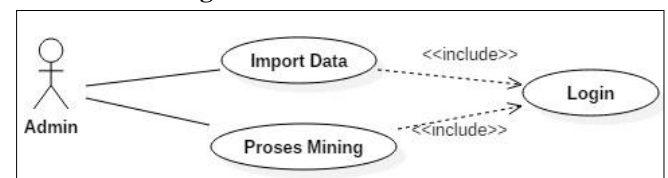
*P(H)*: Probabilitas hipotesis *H* (prior probability)

*P(X|H)*: Probabilitas *X* berdasarkan kondisi pada hipotesis *H*

*P(X)*: Probabilitas *X*

## 3. GAMBARAN SISTEM

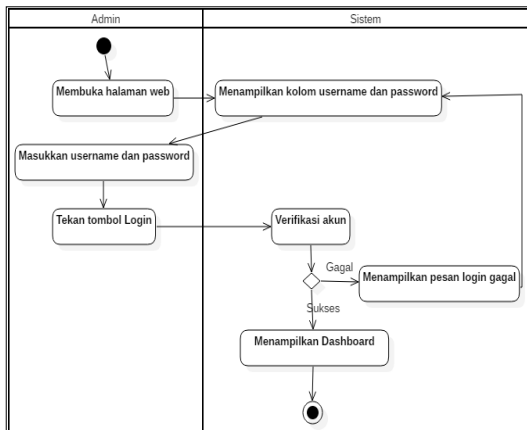
### 3.1 Use Case Diagram



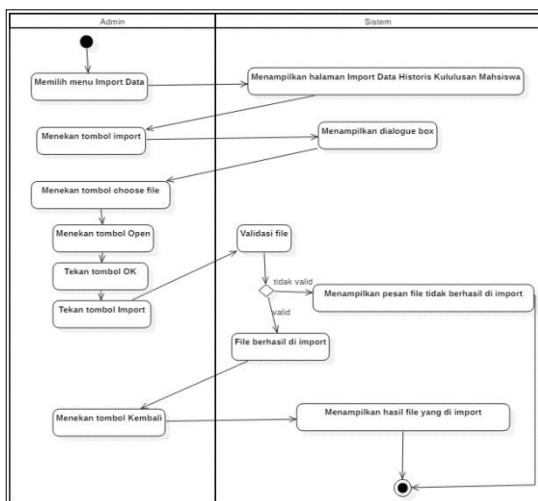
Gambar 1. Use Case Diagram

### 3.2 Activity Diagram

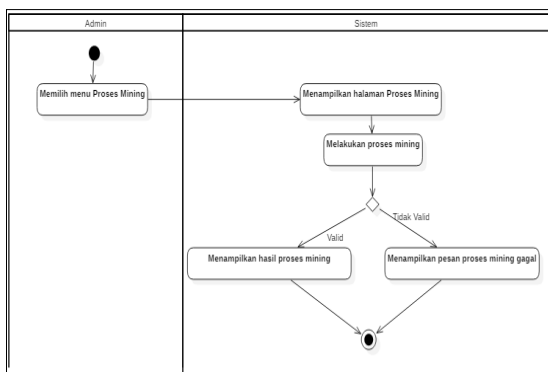
Pada sub bab berikut, akan dimodelkan aliran kegiatan yang terjadi dalam penelitian ini yang digambarkan pada Activity Diagram secara garis besar untuk memodelkan aliran kerja atau aktivitas dan operasi.



Gambar 2. Activity Diagram Login



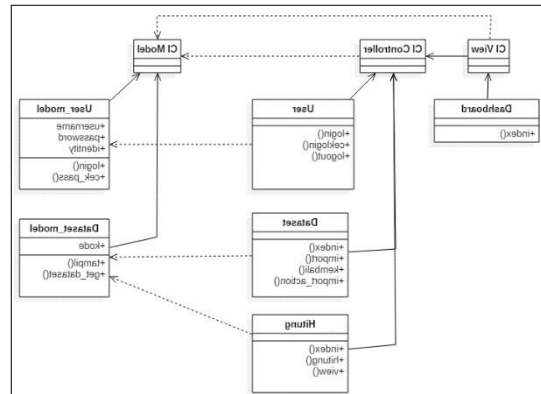
Gambar 3. Activity Diagram Import Data



Gambar 4. Activity Diagram Proses Mining

### 3.3 Class Diagram

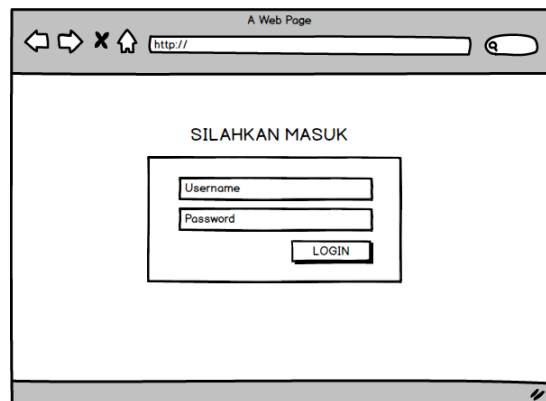
Class diagram adalah suatu diagram yang menyediakan sekumpulan class objek antar muka interface dan relasinya, dan juga untuk memodelkan database logic.



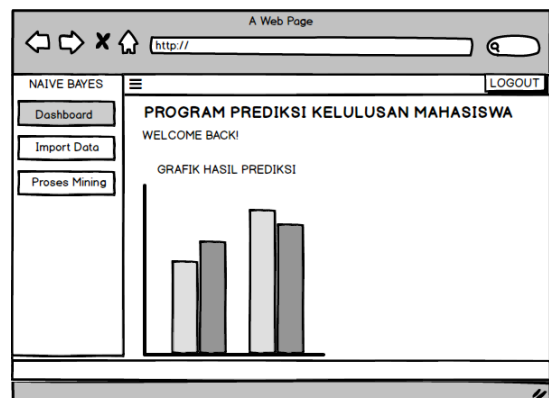
Gambar 5. Class Diagram

### 3.4 Perancangan Antarmuka

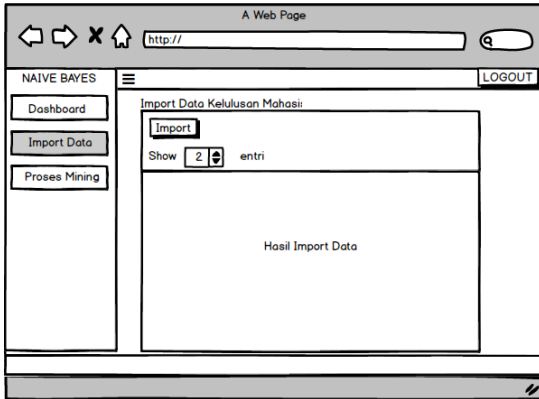
Perancangan antarmuka bertujuan untuk memberikan gambaran mengenai bentuk antarmuka dari perangkat lunak yang akan digunakan oleh user untuk berinteraksi dengan perangkat lunak. Perancangan antarmuka ini mempertimbangkan berbagai kemudahan dan fungsionalitas dari perangkat lunak itu sendiri.



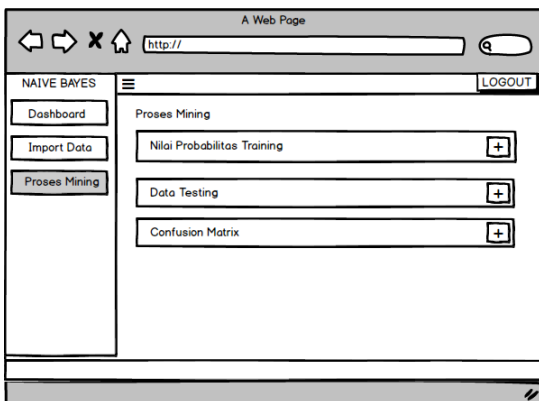
Gambar 6. Perancangan Antarmuka Login



Gambar 7. Perancangan Antarmuka Dashboard



Gambar 8. Perancangan Antarmuka Import Data



Gambar 9. Perancangan Antarmuka Proses Mining

## 4. IMPLEMENTASI

### 4.1 Implementasi Naïve Bayes

1. Persiapkan data yang akan dijadikan sebagai data training. Data yang digunakan merupakan data kelulusan mahasiswa D3 dan S1 tahun 2014 – 2018. Berikut adalah data yang akan digunakan:

Tabel 1. Data Training

NO	JURUSAN	ASAL SEKOLAH	NEM	KET. NEM	IPK	KET. IPK
1	IF	SMK	24	KURANG	3.05	MEMUASKAN
2	IF	SMA	30	CUKUP	2.50	MEMUASKAN
3	SI	SMK	34	CUKUP	2.93	MEMUASKAN
4	IF	SMA	33	CUKUP	3.36	MEMUASKAN
5	IF	SMK	35	CUKUP	2.84	MEMUASKAN
6	IF	SMK	24	KURANG	2.93	MEMUASKAN
7	SI	SMK	34	CUKUP	2.93	MEMUASKAN
8	IF	SMK	30	CUKUP	2.93	MEMUASKAN
9	IF	SMA	37	CUKUP	2.93	MEMUASKAN
10	IF	SMA	35	CUKUP	2.69	MEMUASKAN
.....	.....	.....	.....	.....	.....	.....
973	IF	SMK	27	KURANG	3.00	MEMUASKAN

2. Berikut ini adalah atribut yang akan digunakan dalam perhitungan

Tabel 2. Atribut Perhitungan

Atribut	
1	Jurusan
2	Asal Sekolah
3	Ket. NEM
4	Ket. IPK

3. Setelah menentukan atribut yang akan digunakan dalam perhitungan, kemudian data cleaning dilakukan agar sesuai dengan atribut yang akan digunakan.

Tabel 3. Data Cleaning

NO	JURUSAN	ASAL SEKOLAH	KET. NEM	KET. IPK
1	IF	SMK	KURANG	MEMUASKAN
2	IF	SMA	CUKUP	MEMUASKAN
3	SI	SMK	CUKUP	MEMUASKAN
4	IF	SMA	CUKUP	MEMUASKAN
5	IF	SMK	CUKUP	MEMUASKAN
6	IF	SMK	KURANG	MEMUASKAN
7	SI	SMK	CUKUP	MEMUASKAN
8	IF	SMK	CUKUP	MEMUASKAN
9	IF	SMA	CUKUP	MEMUASKAN
10	IF	SMA	CUKUP	MEMUASKAN
.....	.....	.....	.....	.....
973	IF	SMK	KURANG	MEMUASKAN

4. Menghitung jumlah kelas/label yaitu Ket. IPK pada data cleaning. Pada Ket. IPK ada 3 kategori yaitu cukup, memuaskan, dan pujian. Kategori tersebut dihitung masing – masing jumlahnya. Setelah menemukan jumlah setiap kategori, hitung ketiga kategori tersebut dengan jumlah data label. Berikut adalah cara perhitungannya:  
 Cukup = 101 / 973  
 Memuaskan = 752 / 973  
 Pujian = 120 / 973

Tabel 4. Jumlah Kelas/Label

Jumlah label (Ket. IPK)	Cukup	Memuaskan	Pujian
Cukup	101	0.10	0.77
Memuaskan	752		
Pujian	120		
	973		

5. Kemudian lakukan perhitungan jumlah kasus per kelas. Berikut ini adalah contoh cara perhitungannya:

Jurusan IF : Jumlah data IF / jumlah data Ket.

$$\text{IPK Cukup} = 507 / 59 = 8.59$$

Jurusan SI : Jumlah data SI / jumlah data Ket.

$$\text{IPK Cukup} = 466 / 42 = 11.10$$

Lakukan perhitungan diatas untuk semua kelas dan kategori yang ada pada data cleaning.

Tabel 5. Perhitungan Jumlah Kasus Per Kelas

PARAMETER	JUMLAH DATA	CUKUP	MEMUASKAN	PUJIAN
JURUSAN				
	IF	507	8.59	1.29
	SI	466	11.10	1.30
ASAL SEKOLAH	973			
	SMA	369	10.54	1.25
	SMK	604	9.15	1.32
KET. NEM	973			
	KURANG	217	14.47	1.25
	CUKUP	554	7.91	1.33
	MEMUASKAN	202	12.63	1.23
	973			

6. Setelah itu kalikan semua variable kelas dengan hasil perhitungan data diatas dan data cleaning yang sebelumnya. Berikut adalah contoh cara perhitungannya:
- Data untuk kategori Cukup  
Jurusan IF (8.59) \* Asal Sekolah (9.15) \* Ket. Nem(14.47) \* jumlah label ipk (0.10) = 118.09
  - Data untuk kategori Memuaskan  
Jurusan IF (1.29) \* Asal Sekolah (1.32) \* Ket. Nem(1.25) \* jumlah label ipk (0.77)= 1.65
  - Data untuk kategori Pujian  
Jurusan IF (9.39) \* Asal Sekolah (7.37) \* Ket. Nem(7.48) \* jumlah label ipk (0.12)= 63.82

Lakukan perhitungan diatas dengan 973 data yang ada.

Keterangan:

C = Cukup

M = Memuaskan

P = Pujian

Table 6. Perkalian semua variable kasus per kelas

No	CUKUP			MEMUASKAN			PUJIAN		
	J	A	N	J	A	N	J	A	N
1	8.59	9.15	14.47	1.29	1.32	1.25	9.39	7.37	7.48
2	8.59	10.54	7.91	1.29	1.25	1.33	9.39	9.71	8.03
3	11.10	9.15	7.91	1.30	1.32	1.33	7.06	7.37	8.03
4	8.59	10.54	7.91	1.29	1.25	1.33	9.39	9.71	8.03
5	8.59	9.15	7.91	1.29	1.32	1.33	9.39	7.37	8.03
6	8.59	9.15	14.47	1.29	1.32	1.25	9.39	7.37	7.48
7	11.10	9.15	7.91	1.30	1.32	1.33	7.06	7.37	8.03
8	8.59	9.15	7.91	1.29	1.32	1.33	9.39	7.37	8.03
9	8.59	10.54	7.91	1.29	1.25	1.33	9.39	9.71	8.03
10	8.59	10.54	7.91	1.29	1.25	1.33	9.39	9.71	8.03
973	8.59	9.15	14.47	1.29	1.32	1.25	9.39	7.37	7.48

7. Kemudian, langkah terakhir adalah bandingkan hasil per kelas. Caranya yaitu lihat dari perhitungan data sebelumnya dan cari nilai yang paling tinggi untuk menjadi hasil kelas prediksi.

Table 7. Perbandingan kasus per kelas

No	CLASS PREDICTION	C	M	P
1	CUKUP	118.09	1.65	63.82
2	PUJIAN	74.43	1.66	90.28
3	CUKUP	83.42	1.78	51.50
4	PUJIAN	74.43	1.66	90.28
5	PUJIAN	64.61	1.76	68.48
6	CUKUP	118.09	1.65	63.82
7	CUKUP	83.42	1.78	51.50
8	PUJIAN	64.61	1.76	68.48
9	PUJIAN	74.43	1.66	90.28
10	PUJIAN	74.43	1.66	90.28
973	CUKUP	118.09	1.65	63.82

8. Mencoba perhitungan dengan data testing dengan cara memasukkan nama, jurusan, asal sekolah, dan nem. Kemudian lihat hasilnya apakah data yang dimasukkan tersebut akan menghasilkan memuaskan, cukup atau pujian.

Berikut ini adalah contoh menggunakan data testing.

Nama : Adam

Jurusan : SI

Asal Sekolah : SMK

Ket. NEM : CUKUP

Table 8. Contoh perhitungan Data Testing

	Cukup	Memuaskan	Pujian	Class Prediction
SI	11.10	1.30	7.06	
SMK	9.15	1.32	7.37	
CUKUP	7.91	1.33	8.03	
Jumlah label IPK	0.10	0.77	0.12	
Jumlah	83.42	1.78	51.50	CUKUP

Hasil dari perhitungan di atas yaitu mahasiswa bernama Adam, jurusan SI, asal Sekolah SMK dan nilai NEM yang CUKUP menghasilkan nilai IPK yang CUKUP.

#### 4.2 Implementasi Perangkat Lunak

Dalam subbab ini akan dijelaskan langkah-langkah pembuatan, penggunaan dan jadwal pelaksanaan dalam implementasi rancangan perangkat lunak. Jadwal pelaksanaan digambarkan menggunakan *ganttchart*.

Table 9. Gantt Chart

Kegiatan dan Waktu Pelaksanaan	Maret				April				Mei				Juni				Juli				Agustus			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Analisis Kebutuhan	█	█	█	█																				
Analisis Perancangan					█	█	█	█	█	█	█	█												
Pemrograman													█	█	█	█	█	█	█	█				
Pengujian																					█	█	█	█
Dokumentasi																								█

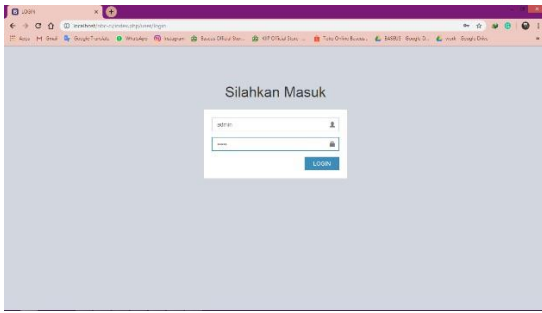
#### 4.3 Lingkup dan Batasan Implementasi

Penjelasan mengenai lingkup dan batasan implementasi termasuk modul program yang akan diimplementasikan, fungsional *system* dan lain sebagainya termasuk lingkungan *system*.

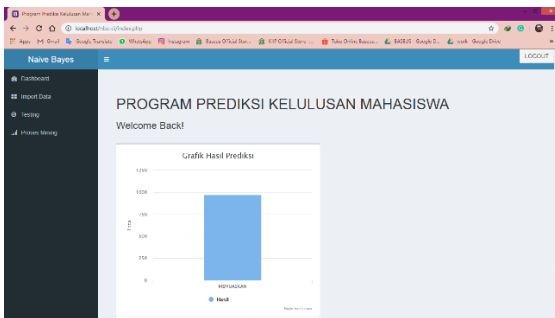
Adapun lingkup dan batasan implementasi agar pada saat digunakan *system* dapat berjalan sesuai dengan modul, fungsional *system* maka lingkup dan batasan implementasi sebagai berikut :

1. Fungsi login : Jalan masuk kedalam *system*.
2. Fungsi import data: Mengimport data kelulusan mahasiswa
3. Fungsi proses mining: hasil dari proses perhitungan

## 4.4 Implementasi Antarmuka



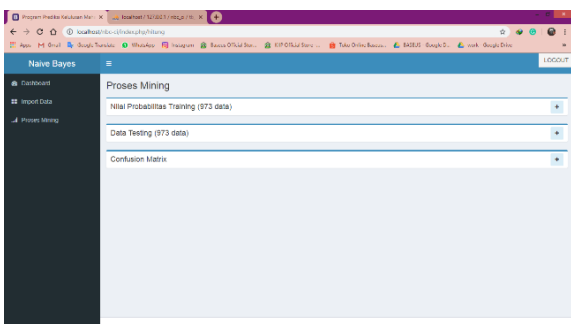
Gambar 10. Antarmuka Login



Gambar 11. Antarmuka Dashboard

Nomer	Jurusan	Ases Beban	NDM	IPK
1	IP	BAK	KUTANG	MEMULAKAN
2	IP	BAK	CORUP	MEMULAKAN
3	SI	BAK	CORUP	MEMULAKAN
4	IP	BAK	CORUP	MEMULAKAN
5	IP	BAK	CORUP	MEMULAKAN
6	IP	BAK	KONTING	MEMULAKAN
7	SI	BAK	CORUP	MEMULAKAN
8	IP	BAK	CORUP	MEMULAKAN
9	IP	BAK	CORUP	MEMULAKAN
10	IP	BAK	CORUP	MEMULAKAN
11	IP	BAK	KUTANG	MEMULAKAN

Gambar 12. Antarmuka Import Data



Gambar 13. Antarmuka Proses Mining

menjadi lebih optimal untuk melakukan prediksi kelulusan mahasiswa. Sehingga jajaran manajemen STMIK & Politeknik LPKIA mudah mendapatkan informasi kelulusan mahasiswa yang mendukung untuk pengambilan keputusan dalam solusi atau kebijakan untuk meningkatkan prestasi mahasiswa agar dapat menyelesaikan studinya dengan hasil yang memuaskan.

Proses testing digunakan untuk memprediksi data mahasiswa tentang predikat kelulusan yang diperoleh. Pada proses ini atribut yang digunakan adalah IPK. Atribut tersebut dipilih karena memiliki nilai perkalian support dan confidence yang tinggi dibandingkan atribut yang lain.

Pada analisa data yang dilakukan di proses testing, didapatkan keakuratan sistem sekitar 77.20% predikat memuaskan, 20.35% predikat cukup dan 2.45% predikat pujian. Pengujian tersebut berdasarkan dari 973 data induk mahasiswa dan data kelulusan mahasiswa tahun 2013 – 2015.

## 5.2 Saran

1. Sistem dikembangkan agar dapat melakukan *data cleaning* dalam sistem.
2. Sistem dikembangkan agar dapat menerima input ekstensi file selain *excel*.

## DAFTAR PUSTAKA

- [1] R. M. S. A.S, Modul Pembelajaran Rekayasa Perangkat Lunak, Bandung: Informatika, 2014.
- [2] D. D. David L. Olson, Advanced Data Mining Techniques, USA: Springer, 2018.
- [3] Informatikalogi, "Algoritma Naive Bayes," 18 April 2017. [Online]. Available: <https://informatikalogi.com/algoritma-naive-bayes/>. [Accessed 24 April 2019].
- [4] C. B. Thomas Connolly, Database Systems : A Practical Approach to Design, Implementation, and Management, England: Pearson Education Limited, 2015.

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Dengan adanya sistem prediksi kelulusan mahasiswa menggunakan *data mining Naive Bayes*, pemanfaatan data historis kelulusan mahasiswa