

BAB I

PENDAHULUAN

1.1 Latar Belakang

Lama masa studi yang ditempuh oleh mahasiswa merupakan salah satu standar yang termasuk ke dalam standar penilaian oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT untuk mengukur akreditasi perguruan tinggi tersebut yakni Mahasiswa dan Lulusan. (Buku III BAN-PT: Borang Akreditasi Institusi Perguruan Tinggi, 2011).

Mahasiswa sarjana yang menyelesaikan masa studi lebih dari 8 semester hingga 14 semester sudah tergolong mahasiswa yang menempuh masa studi yang lama. Paling lama 7 (tujuh) tahun akademik untuk program sarjana, program diploma empat/ sarjana terapan, dengan beban belajar mahasiswa paling sedikit 144 (seratus empat puluh empat) sks. (Buku Panduan STIKI Malang, 2015).

Pada institusi pendidikan perguruan tinggi seperti STIKI Malang, data mahasiswa dan data jumlah kelulusan mahasiswa dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan mahasiswa setiap tahunnya, profil, dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di STIKI Malang. Adanya informasi mengenai lama masa studi mahasiswa tentunya akan menjadi pendukung pengambilan keputusan yang tepat bagi manajemen STIKI Malang dalam mengambil langkah berikutnya.

Salah satu permasalahan yang sering terjadi di perguruan tinggi khususnya STIKI Malang yaitu ketidakseimbangan antara mahasiswa yang masuk dan keluar

karena telah menyelesaikan masa studi. Mahasiswa yang masuk dalam jumlah banyak, namun jumlah yang lulus tepat waktu lebih sedikit daripada mahasiswa yang masuk ke STIKI Malang. Hal ini akan mempengaruhi mutu lulusan STIKI Malang. Untuk itu diperlukan adanya prediksi. Salah satu cara untuk memprediksi masa studi mahasiswa yaitu dengan membangun model klasifikasi. Selama ini STIKI Malang belum memiliki model klasifikasi lama masa studi mahasiswa yang dapat digunakan untuk memprediksi mahasiswa yang lulus tepat waktu. Padahal data mahasiswa yang tersedia sudah sangat berlimpah, hanya saja data-data tersebut belum dimanfaatkan untuk dianalisis lebih jauh.

Algoritma *Naïve Bayes* dapat digunakan untuk mengklasifikasi dengan waktu komputasi yang cepat, menghapus fitur yang tidak relevan akan meningkatkan kinerja klasifikasi dan dapat menghasilkan akurasi yang akurat. (Ihsan A. Abu Amra, 2017, 911). Pada penelitian ini akan menggunakan algoritma *Naïve Bayes* untuk klasifikasi masa studi dan predikat kelulusan mahasiswa STIKI Malang apakah tergolong tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini adalah sebagai berikut:

1. Apakah algoritma *Naïve Bayes* dapat membantu memprediksi lama masa studi mahasiswa - predikat kelulusan di STIKI Malang?

2. Bagaimana cara implementasi sistem dengan algoritma *Naïve Bayes* untuk mengklasifikasi lama masa studi – predikat kelulusan mahasiswa di STIKI Malang?

1.3 Tujuan Penelitian

Adapun tujuan penelitian dalam tugas akhir ini adalah sebagai berikut:

1. Mengetahui apakah algoritma *Naïve Bayes* dapat membantu dalam memprediksi lama masa studi – predikat kelulusan mahasiswa di STIKI Malang.
2. Membangun sistem untuk mengklasifikasi masa studi mahasiswa di STIKI Malang menjadi 9 golongan yakni tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus dengan menggunakan algoritma *Naïve Bayes*.

1.4 Batasan Masalah

Untuk menghindari melebarnya pokok bahasan maka diperlukan batasan masalah, diantaranya :

1. Data yang digunakan adalah data mahasiswa Teknik Informatika STIKI Malang angkatan 2008, 2009, dan 2010.
2. Menggunakan klasifikasi masa studi *supervised*.
3. Data mahasiswa yang digunakan yaitu nrp, jalur penerimaan, umur mahasiswa, jenis kelamin, pekerjaan ayah, sekolah asal, jurusan asal

sekolah, nilai MK Logika & Algoritma, nilai MK Pemrograman Visual 1, nilai MK Pemrograman Visual 2, IPK Semester 2, SKS lulus sampai dengan semester 2, dan klasifikasi masa studi – predikat kelulusan.

4. Bahasa pemrograman yang digunakan adalah bahasa Java.

1.5 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memberikan manfaat-manfaat yaitu:

1. Puket dan Kaprodi Teknik Informatika STIKI Malang dapat mengetahui informasi masa studi mahasiswa-mahasiswanya sehingga dapat meningkatkan proses pembelajaran.
2. Mahasiswa-mahasiswa STIKI Malang dapat menyelesaikan masa studi tepat waktu dan mengurangi mahasiswa yang tidak lulus atau *drop out*.

1.6 Metodologi Penelitian

Berikut ini metodologi penelitian yang akan digunakan dalam penyusunan tugas akhir:

1.6.1 Tempat dan Waktu Penelitian

Tempat : STIKI Malang

Waktu : Enam bulan (Juni 2018 – November 2018)

Jadwal : Terdapat pada tabel 1.1

Tabel 1.1 Jadwal Penelitian

Tahapan	Bulan	Bulan	Bulan	Bulan	Bulan	Bulan
	Juni	Juli	Agustus	September	Oktober	November

Analisa Kebutuhan						
Perancangan Sistem						
Desain dan Implementasi						
Pengujian Sistem						
Verifikasi						
Penyusunan Laporan						

1.6.2 Alat dan Bahan

a. Perangkat keras

Penelitian ini menggunakan Laptop Asus A456URK dengan spesifikasi lengkap sebagai berikut :

- Sistem Operasi : Windows 10 Pro 64-bit
- Prosesor : Intel Core i5 – 7200U, 2.50Ghz
- Memori : 8192MB RAM
- Harddisk : 1TB
- Kartu Grafis : Intel HD Graphics 620, Nvidia Geforce 930MX

b. Perangkat lunak

Penelitian ini menggunakan perangkat lunak sebagai berikut :

- Teks Editor : Microsoft Word 2010, Microsoft Excel 2010, Netbeans
- Browser : Mozilla Firefox, Google Chrome

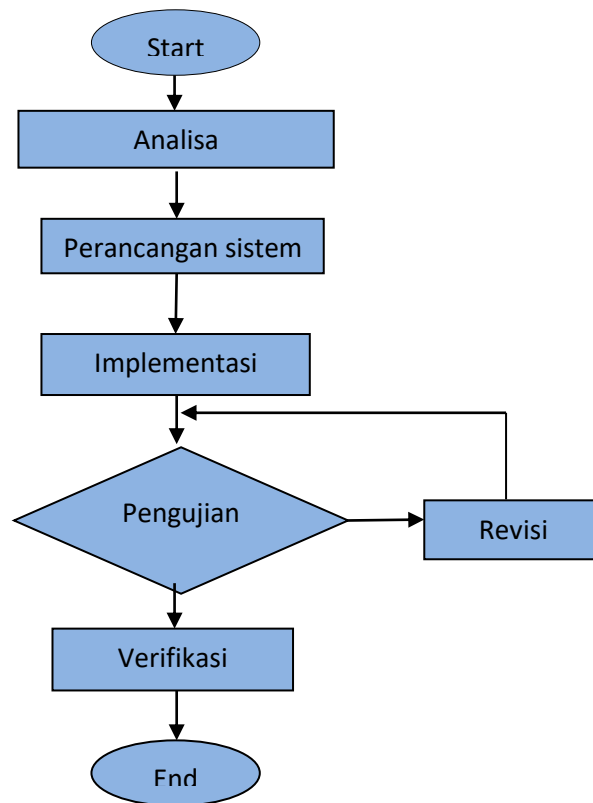
- Bahasa Pemrograman : Java
- Desain Diagram : Sybase Power Designer

1.6.3 Pengumpulan Data dan Informasi

Teknik yang digunakan dalam pengumpulan data dan informasi data mahasiswa adalah dengan melakukan pengambilan data SIAKAD kepada unit kerja BAA dan KTI untuk mendapatkan data berupa NRP, jalur penerimaan, umur mahasiswa, jenis kelamin, pekerjaan ayah, sekolah asal, jurusan asal sekolah, nilai MK Logika & Algoritma, nilai MK Pemrograman Visual 1, nilai MK Pemrograman Visual 2, IPK semester 2, total SKS lulus sampai semester 2, klasifikasi lama masa studi – predikat kelulusan.

1.6.4 Prosedur Penelitian

Prosedur penelitian yang digunakan digambarkan dalam diagram alir Gambar 1.1.



Gambar 1.1 Prosedur Penelitian

Pada tahap analisa, dikumpulkan beberapa data yang berhubungan dengan klasifikasi masa studi mahasiswa STIKI Malang.

Pada tahap perancangan sistem, dari hasil tahap pengumpulan data yang dilakukan, dilanjutkan dengan membuat perancangan sistem.

Pada tahap implementasi, dari hasil perancangan sistem yang sudah dilakukan, dibuat program sesuai dengan rancangan yang telah dibuat. Pembuatan program dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam tahap berikutnya.

Pada tahap pengujian sistem, dari hasil tahap implementasi yang sudah dilakukan kemudian dilakukan proses pengujian program guna mencari

kekurangannya, dan dilakukan proses revisi untuk menemukan hasil yang sesuai dengan kebutuhan yang ada.

Pada tahap verifikasi, setelah melakukan pengujian sistem kemudian akan ada verifikasi sistem yang menandakan bahwa sistem telah selesai dikembangkan tanpa adanya revisi lagi.

1.7 Sistematika Penulisan

BAB 1 PENDAHULUAN

Berisi tentang Latar Belakang, Rumusan Masalah, Tujuan Penelitian, Batasan Masalah, Manfaat Penelitian, Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan mengenai penelitian pendukung yang pernah dilakukan.

BAB III ANALISA DAN PERANCANGAN

Bab ini berisi tentang uraian pemecahan masalah, algoritma, desain dan perancangan aplikasi.

BAB VI IMPLEMENTASI SISTEM

Bab ini berisi tentang implementasi dan pembahasan aplikasi

BAB V PENUTUP

Memaparkan kesimpulan yang diambil berdasarkan hasil yang telah dicapai dan saran untuk kepentingan pengembangan selanjutnya.

LAMPIRAN

Berisi dokumen-dokumen yang terkait selama pelaksanaan pengembangan sistem yang dianggap perlu untuk dilampirkan dalam laporan Tugas Akhir ini.

BAB II

LANDASAN TEORI

2.1 Kajian Penelitian Sebelumnya

Terdapat beberapa penelitian sebelumnya yang meneliti tentang kasus serupa dengan menggunakan metode algoritma *Naïve Bayes*. Adapun kajian penelitian akan diuraikan sebagai berikut.

1. *Drop out Estimation Students based on the Study Period: Comparison between Naïve Bayes and Support Vector Machines Algorithm Methods.*

(Harwati, 2016)

Pada penelitian ini dilakukan perbandingan metode klasifikasi untuk mengestimasi *drop out* mahasiswa dengan menggunakan metode *Naïve Bayes* dan Support Vector Machine. Atribut atau variabel yang digunakan adalah *Gender, High School, Majoring, NEM, Father's Education, Mother's Education, Father's Occupation, Mother's Occupation, GPA of the fourth semesters*. Hasil dari penelitian ini adalah perbandingan estimasi *drop out* mahasiswa dengan 2 algoritma. Algoritma yang digunakan yaitu algoritma *Naïve Bayes* dengan *Support Vector Machine*, dan menunjukkan bahwa *Naïve Bayes* dengan *Support Vector Machine* dapat diterapkan. Dengan akurasi masing-masing 80,67% dan 60%.

2. *Students Performance Prediction Using KNN and Naïve Bayesian.* (Ihsan A.

Abu Amra, 2017)

Pada penelitian ini dilakukan perbandingan metode klasifikasi untuk memprediksi kinerja mahasiswa baru dengan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*. Atribut atau variabel yang digunakan pada penelitian ini adalah *Gender, DOB, Specialization, City, Secondary school name, status, Father's job, student status*. Hasil penelitian ini berupa perbandingan prediksi kinerja mahasiswa baru dengan tujuan untuk membantu kementrian pendidikan dalam meningkatkan pembelajaran siswa di Gaza, dan menunjukkan *Naïve Bayes* dan *KNN* dapat diterapkan dalam kasus ini.

3. Analysis On Student Performance Using Naïve Bayes Classifier (Mokhairi Makhtar, 2017)

Pada penelitian ini dilakukan prediksi untuk klasifikasi kinerja siswa pada Mata Pelajaran tertentu seperti *English, Islamic Education subject, History, Mathematics, Additional Mathematic, Physics, Chemistry* untuk kategori atau *grade* yang telah ditentukan seperti *Excelent, Good, Average, Poor*. Data yang digunakan adalah data tahun 2011 hingga 2014 dengan total 488 siswa. Hasil penelitian ini adalah prediksi klasifikasi kinerja siswa, dan menunjukkan *Naïve Bayes Classifier* dapat diterapkan akurasi *Naïve Bayes* mencapai 73,4% pada kasus ini.

4. Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naïve Bayes (Muh Amin Nurrohmat, 2015)

Pada penelitian ini dilakukan prediksi untuk masa studi dan predikat kelulusan pada jurusan Informatika di Universitas Muhammadiyah Surakarta dengan menggunakan metode *Naïve Bayes*. Atribut yang dipakai untuk perhitungannya yaitu jurusan asal sekolah, *gender*, daerah asal, asal sekolah, dan asisten. Hasil dari penelitian ini yaitu prediksi masa studi berupa Tepat waktu, dan Terlambat serta predikat kelulusan (IPK) berupa *Cum laude*, Sangat Memuaskan, dan Memuaskan.

5. Perbedaan penelitian ini terhadap kajian penelitian sebelumnya.

Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya yaitu pada penelitian ini akan dilakukan klasifikasi dengan studi kasus data-data mahasiswa Teknik Informatika STIKI Malang ke dalam 9 kelas, yaitu tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus. Data atribut yang akan dipakai disesuaikan juga dengan jurusan Teknik Informatika STIKI sendiri yang merupakan kampus IT, dimana pada kurikulum Teknik Informatika STIKI terdapat banyak mata kuliah pemrograman yang akan berpengaruh terhadap lama masa studi mahasiswa. Oleh sebab itu, ditambah atribut atau variabel berupa nilai-nilai mata kuliah yang ada di semester 1 dan semester 2 yang berhubungan dengan pemrograman yakni Logika Algoritma, Pemrograman Visual 1, dan Pemrograman Visual 2. Data atribut yang akan digunakan berupa NRP

mahasiswa, jalur penerimaan, umur mahasiswa, jenis kelamin, pekerjaan ayah, nama sekolah asal, jurusan asal sekolah, nilai MK Logika & Algoritma, nilai MK Pemrograman Visual 1, nilai MK Pemrograman Visual 2, IPK, SKS lulus, klasifikasi masa studi – predikat kelulusan.

2.2 Konsep Dasar

2.2.1 Machine Learning

Machine Learning adalah bagian dari *Artificial Intelligence* dimana mesin dilatih untuk belajar dari pengalaman masa lalu. Pengalaman masa lalu dikembangkan melalui data yang dikumpulkan. Dalam *mesin learning* sendiri terbagi menjadi 2 jenis yaitu *supervised learning* dan *unsupervised learning*. (datascience, 2018)

2.2.1.1 Supervised Learning

Supervised Learning adalah suatu teknik pembelajaran yang digunakan dimana sudah terdapat data yang dilatih, dan terdapat variabel yang ditargetkan. *Supervised Learning* biasanya digunakan untuk tujuan klasifikasi yaitu mengelompokkan suatu data ke kelas-kelas data yang sudah ada (label). Algoritma dalam *supervised learning* antara lain *Support Vector Machine*, *Nearest Neighbor*, *Artificial Neural Network*, *Naïve Bayes*, *Decision Tree* dan *Fuzzy K-Nearest Neighbor*.

2.2.1.2 Unsupervised Learning

Unsupervised learning berbeda dengan *supervised learning*, dimana *unsupervised learning* tidak memiliki data-data sebelumnya termasuk kelas beserta

variabel dikelompokkan data tersebut ke dalam beberapa kelompok berdasarkan kriteria-kriteria nya masing-masing yang biasa disebut juga dengan *clustering*. Algoritma dalam *unsupervised learning* antara lain *K-Means*, *Hierachical Clustering*, *DBSCAN*, *Fuzzy C-Means* dan *Self-Organizing Map*.

2.2.2 Naïve Bayes Classifier

Naïve bayes classifier adalah suatu implementasi dari algoritma *Naïve Bayes* yang dipergunakan untuk mengklasifikasi data. *Naïve bayes* sendiri berasal dari teorema *Bayes* (aturan *Bayes*). Aturan *Bayes* merupakan suatu teknik prediksi yang menggunakan probabilitas sederhana berdasarkan pengalaman di masa lalu. Disebut *Naïve Bayes* karena asumsi independensi yang kuat pada pada setiap atribut. Maksud dari independensi (ketidaktergantungan) yang kuat pada atribut adalah ada atau tidak nya suatu atribut tidak berkaitan terhadap atribut-atribut lain.

$$p(Ck|x) = \frac{p(x|Ck)p(Ck)}{p(x)}$$

Gambar 2. 1 Rumus Perhitungan *Naive Bayes*

$$Posterior = \frac{likelihood \times prior}{evidence}$$

Gambar 2. 2 Rumus Perhitungan *Naive Bayes*

Keterangan

Ck = Kategori lama studi – predikat kelulusan mahasiswa k = 1,2,3,...9. Dimana dalam penelitian ini k1 = Tepat waktu – *summa cum laude*, k2 = Tepat waktu – *cum laude*, k3 = Tepat waktu – sangat memuaskan, k4 =

Tepat waktu – memuaskan, k_5 = Terlambat – *summa cum laude*, k_6 = Terlambat – *cum laude*, k_7 = Terlambat – sangat memuaskan, k_8 = Terlambat – memuaskan, k_9 = Tidak lulus.

$P(X_i|C_k)$ = Probabilitas x_i pada kategori C_k

X_i = Atribut yang mempengaruhi lama studi mahasiswa

$P(C_k)$ = Probabilitas dari C_k

2.2.3 Masa Studi

Menurut Kamus Besar Bahasa Indonesia (KBBI), istilah Sarjana merupakan gelar strata satu yang dicapai oleh seseorang yang telah menamatkan pendidikan tingkat terakhir di perguruan tinggi. Masa studi paling lama 7 (tujuh) tahun akademik untuk program sarjana, program diploma empat/ sarjana terapan, dengan beban belajar mahasiswa paling sedikit 144 (seratus empat puluh empat) sks. (Buku Pedoman STIKI, 2015).

2.2.4 Teknik Evaluasi

Teknik evaluasi merupakan teknik pengukuran sebuah efektivitas klasifikasi yang dapat dilakukan dengan menggunakan teknik *Confusion Matrix*, dan *Precision-Recall*.

2.2.4.1 Confusion Matrix

Confusion matrix adalah salah satu cara untuk mendapatkan akurasi dengan tabel yang menyatakan jumlah data uji klasifikasi yang benar dan salah dengan cara menentukan nilai *True* dan *False*. Contoh *confusion matrix* untuk mengklasifikasi

masa studi mahasiswa – predikat kelulusan mahasiswa ke dalam 9 kategori ditunjukkan pada tabel 2.1.

Tabel 2.1 *Confusion Matriks*

<i>K</i> <i>E</i> <i>L</i> <i>A</i> <i>S</i> <i>A</i> <i>C</i> <i>T</i> <i>U</i> <i>A</i> <i>L</i>	Nilai	KELAS PREDIKSI								
		1	2	3	4	5	6	7	8	9
	1	T	F	F	F	F	F	F	F	F
	2	F	T	F	F	F	F	F	F	F
	3	F	F	T	F	F	F	F	F	F
	4	F	F	F	T	F	F	F	F	F
	5	F	F	F	F	T	F	F	F	F
	6	F	F	F	F	F	T	F	F	F
	7	F	F	F	F	F	F	T	F	F
	8	F	F	F	F	F	F	F	T	F
	9	F	F	F	F	F	F	F	F	T

Keterangan Kelas

1. Tepat waktu – *Summa Cum laude*
2. Tepat waktu – *Cum laude*
3. Tepat waktu – Sangat memuaskan
4. Tepat waktu – Memuaskan
5. Terlambat – *Summa Cum laude*
6. Terlambat – *Cum laude*

7. Terlambat – Sangat memuaskan
8. Terlambat – Memuaskan
9. Tidak Lulus

Akurasi dapat dihitung dengan cara menjumlahkan data hasil uji klasifikasi yang benar dibagi dengan total data uji klasifikasi. Perhitungan akurasi dinyatakan dalam Gambar 2.3.

$$\text{Akurasi} = \frac{T1+T2+\dots+T9}{\text{jumlah total data uji}} \times 100\%$$

Gambar 2.3 Rumus Perhitungan Akurasi

2.2.5 Java

Java adalah suatu bahasa pemrograman yang *multi platform*. *Multi platform* disini berarti program yang dibuat dengan bahasa Java dapat dijalankan di semua komputer yang mendukung bahasa Java tanpa ada perubahan dalam kodenya. Beberapa kelebihan bahasa java yaitu :

- OOP (*Object Oriented Programing*) atau pemrograman berbasis objek.
- Library yang lengkap.

2.2.6 Unified Modeling Language (UML)

UML merupakan sekumpulan alat untuk melakukan abstraksi terhadap suatu sistem yang akan dibuat. UML dapat juga digunakan untuk mendokumentasi dan menspesifikasikan sistem. UML dapat mempermudah pengembangan sistem aplikasi dengan menyediakan berbagai macam diagram dalam pemodelan, diagram yang akan dipakai untuk aplikasi ini, diantaranya:

- *Use Case Diagram*
- *Activity Diagram*

2.2.7 XAMPP

XAMPP berasal dari huruf X, A, M, P, P yang memiliki artinya masing-masing. X berarti program dapat di jalankan pada berbagai sistem operasi. A berarti apache yang merupakan aplikasi web server. M berarti Mysql yang merupakan database server. P pertama yang berarti bahasa pemrograman PHP. P kedua yang berarti bahasa pemrograman Perl. XAMPP adalah sebuah *free software* yang mendukung banyak sistem operasi yang merupakan kumpulan dari beberapa program lain. Adapun beberapa fungsi XAMPP yaitu sebagai server (localhost), yang terdiri dari Apache HTTP Server, Mysql database dan terjemahan bahasa yang ditulis dengan bahasa PHP dan Perl.

BAB III

ANALISA DAN PERANCANGAN

2.1 Analisa Masalah

Siswa-siswa sekolah menengah atas (SMA / sederajat) yang melanjutkan studi ke STIKI Malang setiap tahun terus mengalami peningkatan, namun mahasiswa yang lulus atau yang telah menyelesaikan studi STIKI Malang setiap tahun tidak sebanding dengan mahasiswa baru nya. Jika terus dibiarkan, akan mempengaruhi mutu dan kualitas dari mahasiswa itu sendiri dan akreditasi STIKI Malang. Ada berbagai penyebab mahasiswa tidak dapat menyelesaikan studi tepat waktu atau bahkan di *drop out*, diantara lain yaitu kemampuan mahasiswa itu sendiri dikarenakan tidak mengerti dan tidak memahami kuliah yang ditelaah diajarin dosen, tidak mengulanginya kembali dirumah, atau bahkan tidak mengerjakan tugas sama sekali sehingga berpengaruh terhadap nilai mata kuliah tersebut.

3.1 Pemecahan Masalah

Berkaitan dengan analisa masalah diatas, maka didapat pemecahan masalahnya yaitu dengan sebuah sistem prediksi untuk mengklasifikasikan masa studi mahasiswa STIKI Malang. Untuk mengklasifikasi lama masa studi mahasiswa digunakan *Naïve Bayes Classifier* pada data mahasiswa. *Naïve Bayes Classification* merupakan suatu teknik prediksi dengan menggunakan probabilitas sederhana berdasarkan pengalaman masa lalu dengan asumsi independensi (ketidaktergantungan) yang kuat pada setiap atribut.

Dengan adanya data atau informasi mengenai lama masa studi mahasiswa seorang mahasiswa, tentunya akan menjadi pendukung pengambilan keputusan yang tepat bagi manajemen STIKI Malang dalam mengambil langkah berikutnya.

4.1 Perancangan Algoritma

Tahapan algoritma yang akan dilakukan untuk klasifikasi lama masa studi mahasiswa adalah sebagai berikut.

3.3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan data-data mahasiswa Teknik Informatika STIKI Malang angkatan 2008,2009,dan 2010 pada Badan Administrasi dan Akademik STIKI Malang. Pengumpulan data mahasiswa tersebut dilakukan dengan cara *query* pada *database* mahasiswa STIKI Malang oleh pihak Badan Administrasi dan Akademik, dan diketik kembali ke *Microsoft Excel*. Data mahasiswa yang telah terkumpul, dilabeli sesuai dengan kelasnya, yaitu tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus.

3.3.2 Perancangan Aplikasi

Pada perancangan aplikasi, akan dijelaskan mengenai tahapan yang akan dikerjakan. Terdapat beberapa tahap yang harus dikerjakan untuk melakukan klasifikasi. Dalam penelitian ini data yang akan dipakai yaitu data mahasiswa Teknik Informatika STIKI Malang angkatan 2008, 2009, dan 2010. Dimana data latih maupun data tes berasal dari data mahasiswa.

3.3.3 Data Preprocessing

Dalam tahap ini, data akan menjadi sangat penting untuk proses pembangunan pengetahuan dan proses klasifikasi. *Preprocessing* ini dibagi menjadi 2 tahapan, tahap-tahap nya adalah sebagai berikut:

- *Data cleaning*

Pembersihan data dengan cara menghilangkan *noise* dan data yang tidak konsisten ataupun kosong dan tidak relevan. Data-data tersebut nantinya akan dihapus.

- *Data transformation.*

Keseluruhan isi data mahasiswa berupa jalur penerimaan, umur, jenis kelamin, pekerjaan ayah, sekolah asal, jurusan asal, nilai MK Logika Algoritma, nilai MK Pemrograman Visual 1, nilai MK Pemrograman Visual 2, dan SKS lulus akan diubah ke dalam bentuk yang telah ditentukan. Untuk IPK yang bersifat nilai kontinu, tidak dilakukan *data transformation*.

Tabel 3.1 *Data Transformation Jalur Penerimaan*

Atribut	Data	Data Transformation
Jalur_penerimaan	Seleksi Raport	1
	Prestasi	2
	Tes Masuk	3

Tabel 3.2 *Data Transformation Umur*

Atribut	Data	Data Transformation
Umur	<= 18 tahun	1
	19	2
	20	3
	>=21	4

Tabel 3.3 Data Transformation Jenis Kelamin

Atribut	Data	Data Transformation
Jenis Kelamin	Laki laki	1
	Perempuan	2

Tabel 3.4 Data Transformation Pekerjaan Ayah

Atribut	Data	Data Transformation
Pekerjaan	Pegawai Negeri	1
	Pegawai Swasta	2
	Wira Swasta	3
	Pensiunan	4
	ABRI	5
	Dosen	6
	Petani	7

	Lainnya	8
--	---------	---

Tabel 3.5 *Data Transformation* Sekolah Asal

Atribut	Data	<i>Data Transformation</i>
Sekolah Asal	Tulungagung	1
	Banyuwangi	2
	Jember	3
	Kediri	4
	Situbondo	5
	Probolinggo	6
	Malang	7
	Medan	8
	Toba Samosir	9
	Tanjungpinang	10
	Muarateweh	11
	Serui	12
	Pamekasan	13
	Kotawaringin	14
	Surabaya	15
	Praya	16
	Palu	17

	Palangkaraya	18
	Flores	19
	Sorong	20
	Soe	21
	Bima	22
	Banjarmasin	23
	Negara	24
	Tabanan	25
	Tabalong	26
	Ambon	27
	Jakarta Utara	28
	Tanjung Selor	29
	Makasar	30
	Sidoarjo	31
	Manokwari	32
	Lumajang	33
	Mojokerto	34
	Rame	35
	Sampang	36
	Mataram	37
	Tuban	38

	Muna Sultra	39
	Tarakan	40
	Jogjakarta	41
	Jayapura	42
	Pulang Pisau	43
	Manggarai Barat	44
	Lamongan	45
	Tanah Grogot	46
	Wonorejo	47
	Bintuni	48
	Atambua	49
	Sumenep	50
	Tumpang	51
	Bontang	52
	Bojonegoro	53
	Kepanjen	54
	Madiun	55
	Bondowoso	56
	Blitar	57
	Nganjuk	58
	Trenggalek	59

	Jombang	60
	Batu	61
	Denpasar	62
	Kupang	63
	Ternate	64
	Rante Pao	65
	Bogor	66
	Sumbawa Besar	67
	Banyuwangi	68
	Pasuruan	69
	Waikabubak	70
	Banjarbaru	71
	Batam	72

Tabel 3.6 *Data Transformation* Jurusan Asal

Atribut	Data	<i>Data Transformation</i>
Jurusan Asal	IPA	1
	IPS	2
	Bahasa	3
	Grafika	4

	Elektro	5
	Komunikasi	6
	Listrik	7
	Mesin	8
	Akuntansi	9
	Manajemen	10
	Lainnya	11

Tabel 3.7 Data *Transformation* nilai Mata Kuliah

Atribut	Data	Data <i>Transformation</i>
Nilai MK	A	1
	B+	2
	B	3
	C+	4
	C	5
	D	6
	E	7

Tabel 3.8 Data *Transformation* SKS Lulus

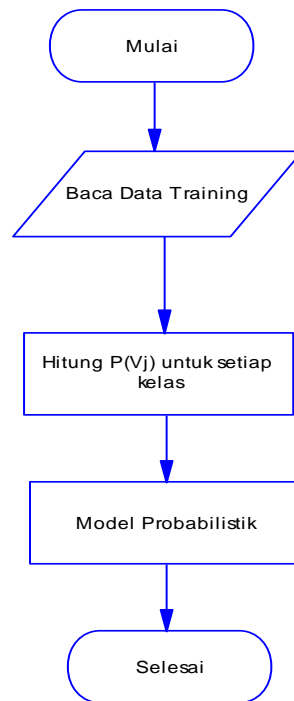
Atribut	Data	Data <i>Transformation</i>
---------	------	----------------------------

SKS Lulus	0 - 6	1
	7 - 12	2
	13 - 18	3
	19 - 24	4
	25 - 30	5
	31 - 36	6
	37 - 42	7
	43 - 48	8

3.3.4 Pelatihan Algoritma Naïve Bayes

Secara umum proses pelatihan *Naïve Bayes* dibagi menjadi beberapa tahap.

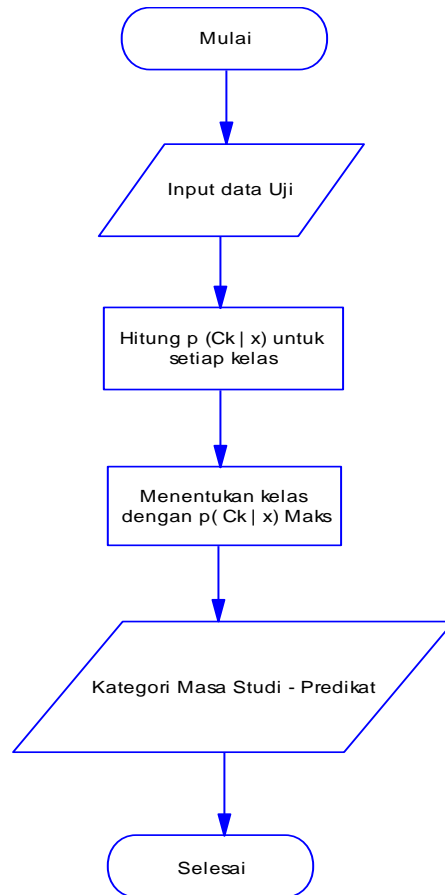
Tahap-tahap tersebut dapat dilihat pada gambar berikut:



Gambar 3.1 Tahap Pelatihan Algoritma *Naïve Bayes*

3.3.5 Pengujian Algoritma Naïve Bayes

Pada proses ini, data uji akan melewati proses klasifikasi berdasarkan data latih. *Flowchart* untuk tahap klasifikasi dapat dilihat pada gambar



Gambar 3.2 Tahap Pengujian Algoritma *Naïve Bayes*

Pengujian akurasi dapat dilakukan dengan *confusion matriks*, yaitu dengan data hasil uji klasifikasi yang benar dibagi dengan total data uji klasifikasi.

$$\text{Akurasi} = \frac{T1+T2+\dots+T9}{\text{jumlah total data uji}} \times 100\%$$

Gambar 3.3 Rumus Perhitungan Akurasi

5.1 Database

a. Tabel Data *Training*

Tabel data *training* digunakan untuk data pelatihan sebagian data set.

Tabel 3.9 Tabel *Training* pada *database*

Nama Field	Tipe	Keterangan
nrp	Int	Nomor unik mahasiswa (PK)
jalur_penerimaan	Varchar(10)	Jalur penerimaan mahasiswa
umur_mhs	Int	Umur mahasiswa angkatan 2008, 2009, dan 2010
Jenis_kelamin	Char(1)	Jenis kelamin mahasiswa
Pekerjaan_ayah	Varchar(15)	Pekerjaan ayah/ wali
Asal_sekolah	Varchar(20)	Kode id dari kategori
Jurusan_asal	Varchar(10)	Jurusan asal sekolah
Nilai_MK_Logika_Algoritma	Varchar(2)	Nilai MK Logika Algoritma
Nilai_MK_pem_visual_1	Varchar(2)	Nilai MK Pemrograman Visual semester 1
Nilai_MK_pem_visual_2	Varchar(2)	Nilai MK Pemrograman Visual semester 2
IPK	Int	Nilai IP kumulatif semester 1 dan 2
SKS_lulus	Int	Jumlah SKS lulus
kategori	Varchar(30)	Kategori lama masa studi - predikat mahasiswa.

b. Tabel Data *Testing*

Tabel data *testing* digunakan untuk data uji coba.

Tabel 3.10 Tabel *Testing* pada *database*

Nama Field	Tipe	Keterangan
nrp	Int	Nomor unik mahasiswa (PK)
jalur_penerimaan	Varchar(10)	Jalur penerimaan mahasiswa
umur_mhs	Int	Umur mahasiswa angkatan 2010
Jenis_kelamin	Char(1)	Jenis kelamin mahasiswa
Pekerjaan_ayah	Varchar(15)	Pekerjaan ayah/ wali
Asal_sekolah	Varchar(20)	Kode id dari kategori
Jurusan_asal	Varchar(10)	Jurusan asal sekolah
Nilai_MK_Logika_Algoritma	Varchar(2)	Nilai MK Logika Algoritma
Nilai_MK_pem_visual_1	Varchar(2)	Nilai MK Pemrograman Visual semester 1
Nilai_MK_pem_visual_2	Varchar(2)	Nilai MK Pemrograman Visual semester 2
IPK	Int	Nilai IP kumulatif semester 1 dan 2
SKS_lulus	Int	Jumlah SKS lulus
kategori	Varchar(12)	Kategori lama masa studi - predikat mahasiswa.

c. Tabel *Record*

Tabel *Record* digunakan untuk menyimpan informasi training yang sudah dilakukan.

Tabel 3.11 Tabel *Record* pada *database*

Nama Field	Tipe	Keterangan
No_training	Int	Nomor unik (PK)
Data_Training	Int	Jumlah data training
Data_Testing	Int	Jumlah data testing
Tanggal	String	Tanggal testing
Akurasi	Int	Akurasi dari training

d. Tabel *Detail Testing*

Tabel *Detail Testing* digunakan untuk menyimpan informasi data *testing* yang telah di tes pada uji coba.

Tabel 3.12 Tabel *Detail Testing* pada database

Nama Field	Tipe	Keterangan
No_training	Int	Nomor unik no training (FK)
nrp	Int	Nrp(FK)
Probabilitas tepat waktu – <i>summa cum laude</i>	Int	Hasil uji coba
Probabilitas tepat waktu – <i>cum laude</i>	Int	Hasil uji coba
Probabilitas tepat waktu – sangat memuaskan	Int	Hasil uji coba
Probabilitas tepat waktu – memuaskan	Int	Hasil uji coba
Probabilitas terlambat – <i>summa cum laude</i>	Int	Hasil uji coba

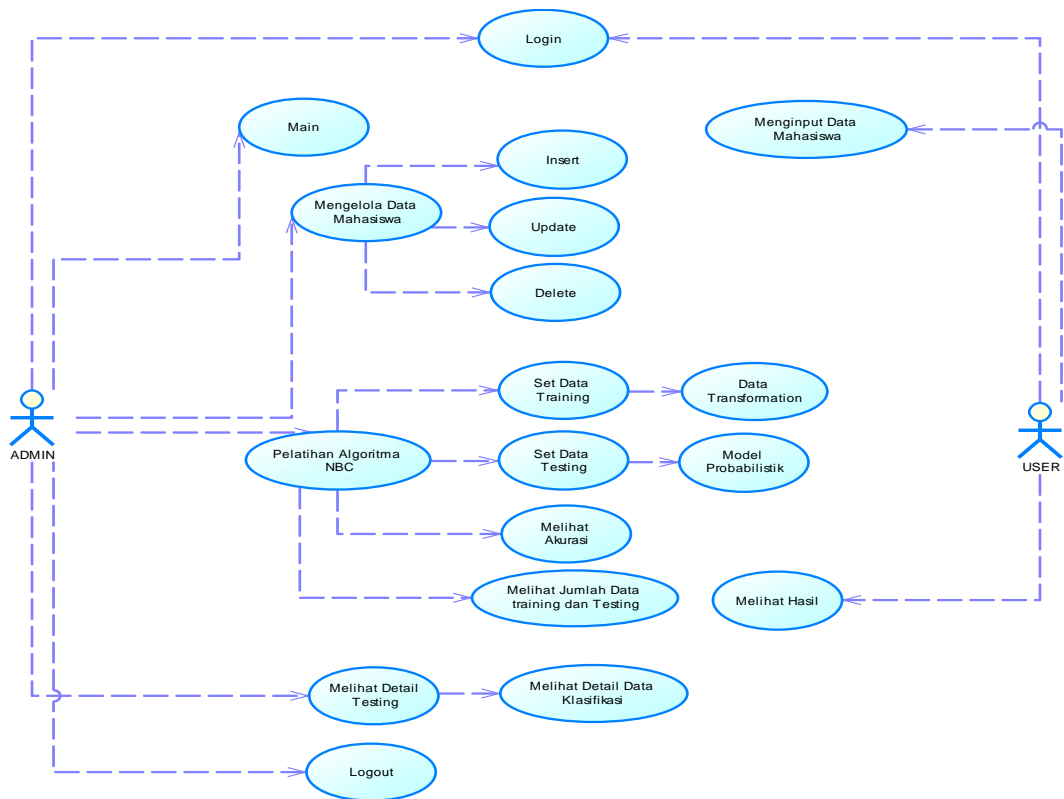
Probabilitas terlambat – <i>cum laude</i>	Int	Hasil uji coba
Probabilitas terlambat – sangat memuaskan	Int	Hasil uji coba
Probabilitas terlambat - memuaskan	Int	Hasil uji coba
Probabilitas tidak lulus	Int	Hasil uji coba

6.1 Perancangan Sistem

Dalam pembahasan ini akan dibahas mengenai perancangan sistem yang dikembangkan pada tugas akhir. Tujuan perancangan sistem ini adalah untuk menerapkan algoritma *Naïve Bayes Classifier* pada sistem pengklasifikasian lama masa studi – predikat mahasiswa. Sistem yang dikembangkan diharapkan dapat mengetahui atribut-atribut pada data mahasiswa dan secara otomatis mengklasifikasi sesuai dengan kategori yang ada, yaitu tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus.

3.5.1 Use Case Diagram

Diagram *use case* dari sistem aplikasi pengklasifikasian lama masa studi mahasiswa yang akan dikembangkan dapat dilihat pada gambar 3.4.



Gambar 3.4 Diagram *Use Case* Aplikasi

Keterangan gambar :

1. Admin

- a. Login : Admin harus melakukan login terlebih dahulu untuk mengakses halaman admin.
- b. Main : Admin dapat melihat tampilan awal sistem yang berisi tentang informasi jumlah data mahasiswa yang termasuk kategori lulus tepat waktu, terlambat dan tidak lulus.
- c. Mengelola Data Mahasiswa : Admin data memasukan data mahasiswa baru, mengedit data mahasiswa yang sudah ada, dan menghapus berita yang sudah ada.

- d. Pelatihan Algoritma NBC : Admin melatih algoritma.
- e. Detail Testing : Admin dapat melihat data yang sudah dilatih.
- f. Logout : Admin keluar dari sistem aplikasi.

2. User

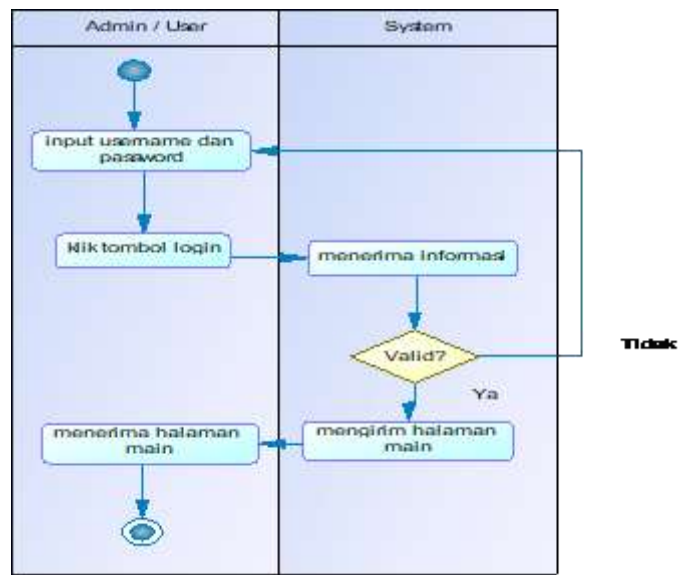
- a. Login : User PUKET / KAPRODI / BAA harus melakukan login terlebih dahulu untuk mengakses halaman user.
- b. Menginput Data Mahasiswa : User PUKET / KAPRODI / BAA dapat memasukan data mahasiswa yang akan diklasifikasi.
- c. Melihat Hasil : User PUKET / KAPRODI / BAA dapat melihat hasil klasifikasi dari data mahasiswa yang dimasukkan.
- d. Logout : User keluar dari sistem aplikasi.

3.5.2 Activity Diagram

Activity diagram atau diagram aktifitas digunakan untuk menggambarkan aktifitas yang dapat dilakukan oleh sistem. Pada tahapan ini dibuat *activity diagram* pada masing-masing fitur berdasarkan *use case diagram* yaitu login, main, data mahasiswa, pelatihan algoritma, pengujian algoritma, dan pengecekan kategori. Berikut adalah *activity diagram* dari masing-masing fitur.

1. Activity Diagram Login

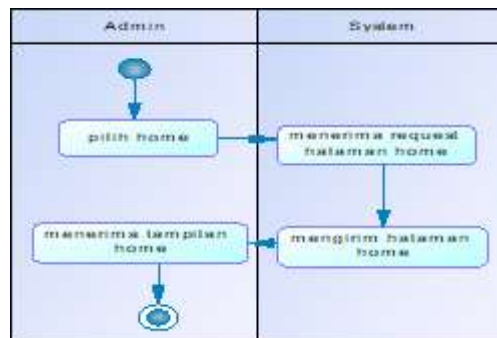
Activity ini menjelaskan bahwa admin atau user harus masuk terlebih dahulu ke dalam sistem untuk dapat melakukan tahapan proses klasifikasi.



Gambar 3.5 Activity Diagram Login

2. Activity Diagram Main

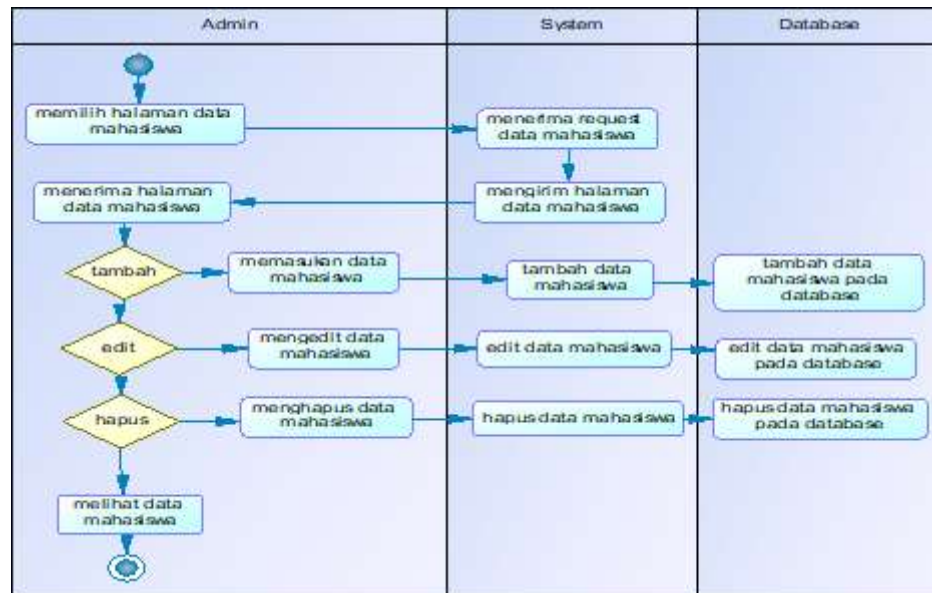
Activity ini menjelaskan tampilan awal dari sistem, yang berisi tentang informasi jumlah data mahasiswa yang masuk dalam kategori lulus tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus.



Gambar 3.6 Activity Diagram Main Admin

3. Activity Diagram Data Mahasiswa

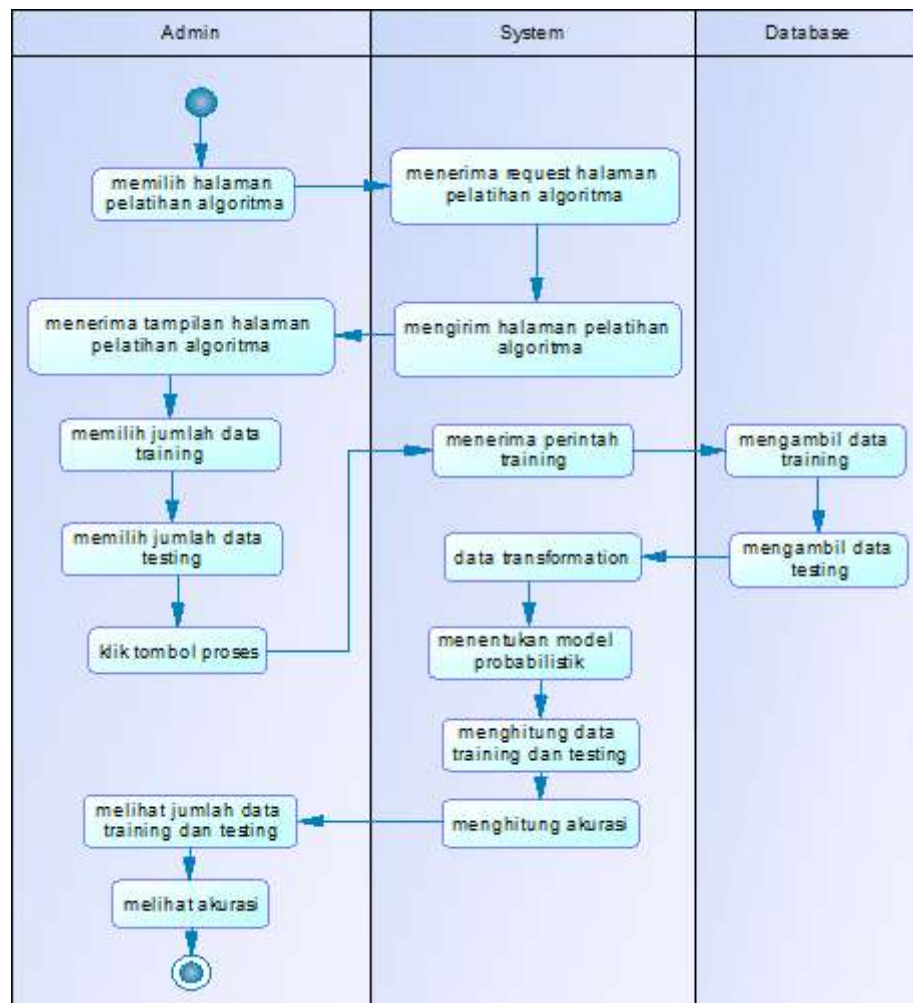
Activity ini menjelaskan tentang isi data mahasiswa, dimana admin juga dapat menambah data mahasiswa baru, melakukan edit pada data mahasiswa yang sudah ada dan dapat menghapus data mahasiswa.



Gambar 3.7 Activity Diagram data mahasiswa

4. Activity Diagram Pelatihan Algoritma

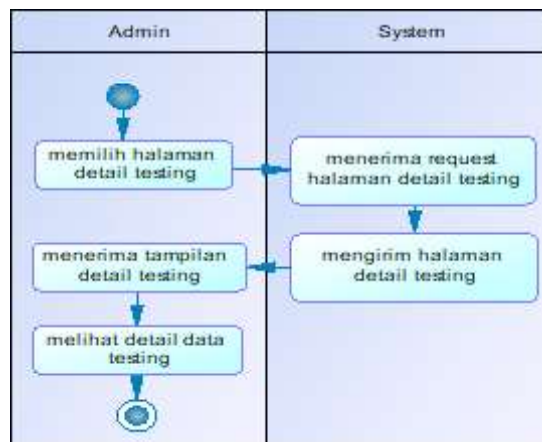
Activity ini menjelaskan tentang proses pelatihan algoritma dimana admin dapat menginputkan jumlah data training dan data testing yang diambil dari database untuk dilakukan data transformation, perhitungan model probabilistik, mengetahui jumlah data keseluruhan, dan keakuratan data dalam bentuk persen.



Gambar 3.8 Activity Diagram Pelatihan Algoritma

5. Activity Diagram Detail Testing

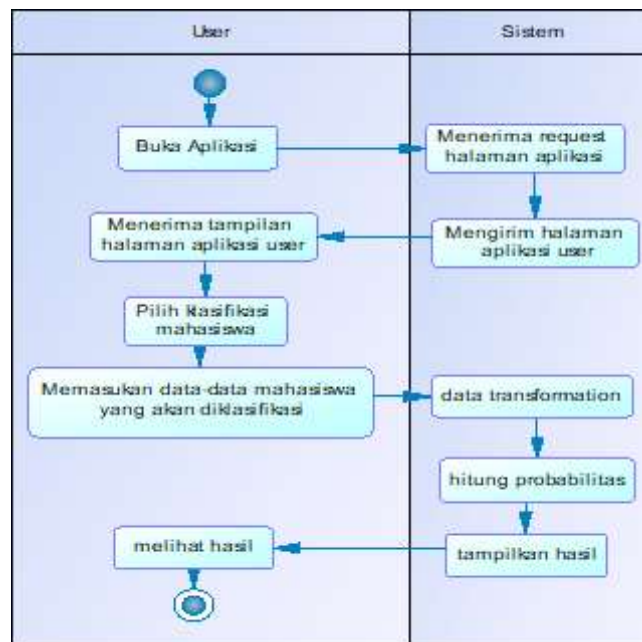
Activity ini menjelaskan tentang detail testing yang sudah diproses pada pelatihan algoritma agar dapat melihat detail perhitungan serta proses data *transformation* yang hanya dapat dilihat oleh admin.



Gambar 3.9 Activity Diagram Detail Testing

6. Activity Diagram User

Activity ini menjelaskan user atau pihak user PUKET / KAPRODI / BAA yang dapat melakukan pengecekan terhadap data mahasiswa untuk mengetahui apakah data mahasiswa yang diinputkan termasuk ke dalam kategori lulus tepat waktu – *summa cum laude*, tepat waktu – *cum laude*, tepat waktu – sangat memuaskan, tepat waktu – memuaskan, terlambat – *summa cum laude*, terlambat – *cum laude*, terlambat sangat memuaskan, terlambat – memuaskan dan tidak lulus. Dan sistem menentukan nilai probabilitasnya.



Gambar 3.10 Activity Diagram User

7.1 Perancangan Desain Interface

Perancangan desain *interface* (antarmuka pengguna) bertujuan untuk memberikan gambaran awal dan sederhana mengenai tampilan yang akan dikembangkan pada sistem pengklasifikasian lama masa studi mahasiswa. Tampilan-tampilan tersebut antara lain adalah tampilan halaman login, halaman *main* untuk admin dan user, halaman data mahasiswa berupa data *training* dan data *testing*, halaman pelatihan algoritma, halaman pengujian algoritma berupa *detail testing*, dan halaman pengecekan kategori.

1. Halaman Login

The screenshot shows a window titled "Login" with a light blue background. At the top, it says "Naive Bayes Classifier" in bold, followed by "Klasifikasi Masa Studi Mahasiswa" and "Teknik Informatika STIKI Malang". Below this, there are two input fields: "Username :" and "Password :". To the right of the password field is a checkbox labeled "Show Password". At the bottom, there are two buttons: "Login" and "Batal".

Gambar 3.11 Halaman *Login*

2. Halaman Main Admin

The screenshot shows a window titled "Naive Bayes Classifier | STIKI Malang". Below the title bar, there are three tabs: "Master Data", "Naive Bayes Classifier", and "Admin". The "Admin" tab is selected. The main content area has a light blue background and displays the following information: "Naive Bayes Classifier", "Klasifikasi Masa Studi Mahasiswa", and "Teknik Informatika STIKI Malang". Below this, there are three lines of text: "Data Latih : 150", "Data Testing : 30", and "Akurasi : 70%".

Gambar 3.12 Halaman *Main Admin*

3. Halaman Master Data, yaitu Data *Training* dan Data *Testing* Mahasiswa.

The screenshot shows a window titled "Master Data Mahasiswa". At the top, there are four buttons: "Tambah Data", "Edit Data", "Hapus Data", and "Close". Below these buttons, there is a search bar labeled "Search". The main content area is titled "Data Training Mahasiswa" and contains a table with the following columns: "NRP", "Jalan Per...", "Umur", "JK", "Pekerjaan", "Alamat", "Jumlah...", "MK Logik...", "MK Prog...", "MK Prog...", "IPK", "SKS Lulus", and "Masa Stu...". The table is currently empty. At the bottom right, there is a label "Jumlah Data : 0".

Gambar 3.13 Halaman *Master Data Training*

The screenshot shows a window titled "Master Data Testing". At the top, there are buttons: "Tambah Data", "Edit Data", "Hapus Data", and "Close". Below these is a search bar labeled "Search". The main area contains a table with the following columns: NRP, Jalur Pen..., Umur, JK, Pekerjaan..., Asal Seko..., Jurusan A..., MK Logik..., MK Prog..., MK Prog..., IPK, SKS Lulus, and Masa Stu... The table is currently empty. At the bottom right, it says "Jumlah Data : 0".

Gambar 3.14 Halaman *Master Data Testing*

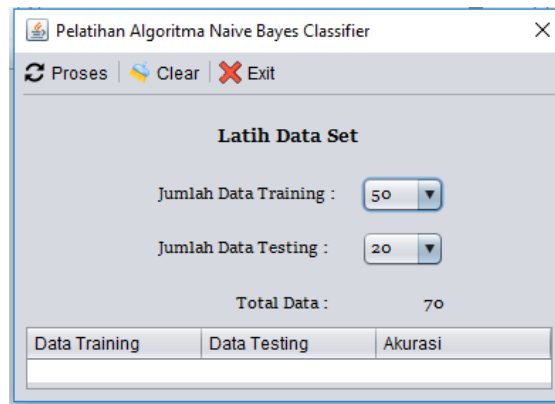
4. Halaman Tambah atau Edit data Training dan Testing Mahasiswa.

The screenshot shows a window titled "Master Data Mahasiswa" with a close button (X). Below the title bar are buttons: "Save", "Clear", and "Exit". The main section is titled "Tambah / Edit Data Training". It contains a form with the following fields and options:

- NRP: Text input field with "NRP" entered.
- Jalur Penerimaan: Dropdown menu with "Seleksi Raport" selected.
- Umur: Dropdown menu with "<= 18 tahun" selected.
- Jenis Kelamin: Radio buttons for "Laki-laki" (selected) and "Perempuan".
- Pekerjaan Ortu: Dropdown menu with "Pegawai Negeri" selected.
- Asal Sekolah: Dropdown menu with "Tulungagung" selected.
- Jurusan Asal: Dropdown menu with "IPA" selected.
- Nilai MK Logika Algoritma: Dropdown menu with "A" selected.
- Nilai MK Pem. Visual 1: Dropdown menu with "A" selected.
- Nilai MK Pem. Visual 2: Dropdown menu with "A" selected.
- Indeks Prestasi Kumulatif: Text input field with "IPK Semester 2" entered.
- SKS Lulus: Dropdown menu with "0 - 6" selected.
- Klasifikasi Masa Studi: Dropdown menu with "Tepat Waktu" selected.

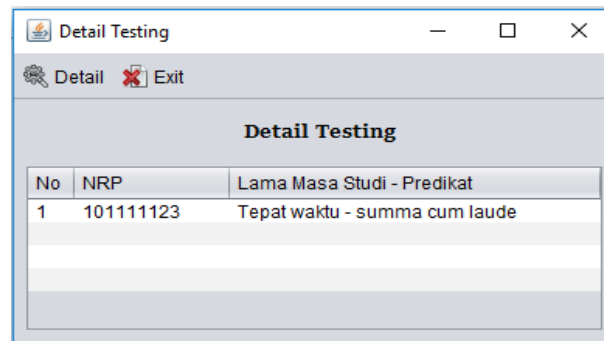
Gambar 3.15 Halaman Tambah atau *Edit Data Training* dan *Testing*

5. Halaman Pelatihan Algoritma



Gambar 3.16 Pelatihan Algoritma

6. Halaman Detail Testing

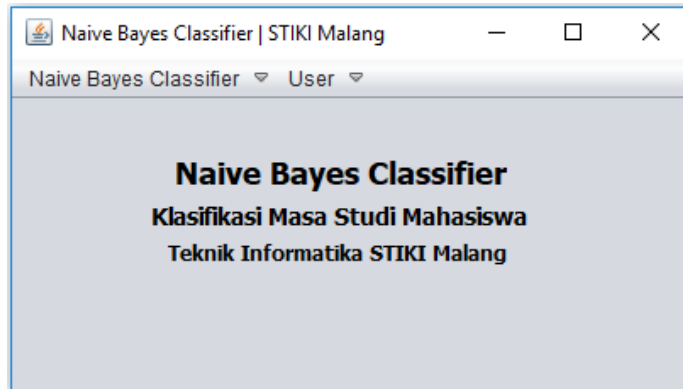


Gambar 3.17 *Detail Testing*



Gambar 3.18 *Detail Probabilitas Testing*

7. Halaman Main User



Gambar 3.19 Halaman *Main User*

8. Halaman User Untuk Klasifikasi.

The screenshot shows a window titled "Klasifikasi Mahasiswa". It has a menu bar with "Klasifikasi", "Clear", and "Exit". The main area contains a list of input fields for student data: "Jalur Penerimaan" (dropdown: Seleksi Raport), "Umur" (dropdown: <= 18 tahun), "Jenis Kelamin" (radio buttons: Laki-laki, Perempuan), "Pekerjaan Ortu" (dropdown: Pegawai Negeri), "Asal Sekolah" (dropdown: Tulungagung), "Jurusan Asal" (dropdown: IPA), "MK. Logika Algoritma" (dropdown: A), "MK. Prog Visual 1" (dropdown: A), "MK. Prog Visual 2" (dropdown: A), "Indeks Prestasi Kumulatif" (text input: 3,5), and "SKS Lulus" (dropdown: 0 - 6). At the bottom, it displays the classification result: "Klasifikasi : Tepat Waktu - cum laude".

Gambar 3.20 Halaman *User Untuk Klasifikasi*

DAFTAR PUSTAKA

- Amelia, M. W., Lumenta, A. S., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. *E-Journal Teknik Informatika*, 1-6.
- Amra, I. A., & Maghari, A. Y. (2017). Students Performance Prediction Using KNN and Naive Bayesian. *International Conference on Information Technology (ICIT)*, 909-913.
- BAN-PT. (2011). *Buku III BAN-PT: Borang Akreditasi Institusi Perguruan Tinggi*.
- Chandra, A. (2017, 03 12). *PERBEDAAN SUPERVISED AND UNSUPERVISED LEARNING*. Retrieved 05 30, 2018, from datascience: https://www.datascience.or.id/detail_artikel/52/supervised-and-unsupervised-learning
- Harwati. (2016). Drop out Estimation Students based on the Study Period: Comparison between Naïve Bayes and Support Vector Machines Algorithm Methods. *IOP Conference Series: Materials Science and Engineering*, 1-8.
- Kemdikbud. (2016). *KBBI Daring*. Retrieved 05 30, 2018, from KBBI: <https://kbbi.kemdikbud.go.id/entri/sarjana>
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: ANDI OFFSET.

Makhtar, M., Nawang, H., & Syamsuddin, S. N. (2017). ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE. *Journal of Theoretical and Applied Information Technology*, 3993-4000.

Nurrohmat, M. A., & Nugroho, Y. S. (2015). Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naive Bayes. *KHAZANAH INFORMATIKA*, 29-34.

Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 1-5.