PERANCANGAN APLIKASI PREDIKSI LAMA STUDI MAHASISWA BERDASARKAN MODEL KLASIFIKASI DATA LAMA STUDI MAHASISWA STMIK INDONESIA YANG TERBENTUK MENGGUNAKAN *DECISION TREE*

Syam Gunawan[1], Pritasari Palupiningsih[2]
[1,2] Program Studi Sistem Informasi, STMIK Indonesia
e-mail: [1]syam @stmik-indonesia.ac.id, [2]prita@stmik-indonesia.ac.id

ABSTRACT

One way to improve quality in universities is through accreditation. One of the accreditation criteria is the student. Student's performance must be monitored and evaluated. Regarding the study duration, the undergraduate bachelor's degree programs typically takes four years to complete. It is important for the university staff to quickly identify which students are less likely to finish the degree on time. Therefore, it is necessary to predict the length of study for each student. The goal of this research is to predict study duration by building Decision Tree-based classifier model using NBTree algorithm. Then, an application is built by applying the classification model. Data used in this research are the grades and academic leave. Result shows that the Naïve Bayes Decision Tree classification model could predict study duration with the accuracy of 73.45%.

Kata Kunci: Lama Studi Mahasiswa, Klasifikasi, Decision Tree, Algoritma NBTree, Prediksi

1. PENDAHULUAN

Program Studi merupakan garda terdepan dalam penyelenggaraan pendidikan dari sebuah Perguruan Tinggi, sehingga program studi senantiasa melakukan evaluasi guna meningkatkan mutu dan efisiensi perguruan tinggi termasuk peningkatan kualitas lulusan. Selain itu, salah satu kriteria penilaian pada akreditasi program studi adalah penilaian terhadap lama studi mahasiswa yang lulus tepat waktu. Masa studi mahasiswa telah diatur dalam ketetapan Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi tentang Sistem Pendidikan Tinggi yang menyebutkan bahwa untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144 - 160 satuan kredit semester (sks) dengan masa studi selama 8 - 10 semester atau 4 – 5 tahun.

Pada institusi pendidikan perguruan tinggi, terdapat data nilai mahasiswa dan data jumlah kelulusan mahasiswa berukuran besar yang dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya, profil, dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di perguruan tinggi. Adanya informasi mengenai lama studi mahasiswa tentu akan menjadi pendukung suatu pengambilan keputusan yang tepat bagi manajemen Perguruan Tinggi dalam mengambil langkah berikutnya.

Permasalahan yang sering terjadi pada beberapa Perguruan Tinggi adalah masih banyaknya jumlah mahasiswa yang lulus dengan masa studi melampaui waktu yang telah ditetapkan atau tidak tepat waktu. Hal ini dapat mempengaruhi mutu lulusan Perguruan Tinggi. Sehingga penting bagi program studi untuk mengetahui mahasiswa mana saja yang memiliki kemungkinan lulus tidak tepat waktu agar dapat diberikan penanganan lebih lanjut. Sehingga mahasiswa tersebut bisa lulus tepat waktu. Idealnya program studi mengetahui kemungkinan ini sedini mungkin. Maksimal pada saat mahasiswa memasuki awal semester 6, program studi sudah mengetahui mahasiswa mana saja yang kemungkinan akan lulus tidak tepat waktu. Akan tetapi, selama ini program studi kesulitan untuk mengetahui mahasiswa mana saja yang memiliki kemungkinan lulus tidak tepat waktu. Program studi baru akan mengetahui bahwa seorang mahasiswa akan memiliki kemungkinan lulus tidak tepat waktu saat mahasiswa tersebut sudah memasuki semester 7 atau semester 8. Hal ini akan mengakibatkan program studi kesulitan untuk mengusahakan agar mahasiswa tersebut dapat lulus tepat waktu. Untuk itu diperlukan adanya prediksi lama studi mahasiswa. Salah satu cara untuk dapat memprediksi lama studi mahasiswa adalah dengan membangun model klasifikasi. Selama ini STMIK Indonesia belum memiliki model klasifikasi lama studi mahasiswa yang dapat digunakan sebagai prediksi jumlah lulus tepat waktu. Padahal data mahasiswa sangat berlimpah, hanya saja data-data tersebut belum dimanfaatkan untuk dianalisis lebih jauh.

Data lama studi mahasiswa berukuran besar dapat dianalisis menggunakan teknik klasifikasi. Salah satu metode dalam klasifikasi adalah *decision tree* yang akan menghasilkan model klasifikasi. Salah satu algoritma yang dapat diterapkan dalam metode *decision tree* adalah NBTree. Model klasifikasi yang terbentuk akan dapat digunakan dalam melakukan prediksi. Penelitian ini bertujuan untuk menerapkan metode *decision tree* dengan algoritma NBTree untuk membentuk model klasifikasi. Model klasifikasi tersebut kemudian diterapkan pada perancangan aplikasi prediksi lama studi mahasiswa.

2. TINJAUAN PUSTAKA

2.1 Klasifikasi

Klasifikasi adalah salah satu teknik dalam menemukan model (fungsi) yang menjelaskan dan

membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk mengetahui kelas atau objek yang memiliki label kelas yang tidak diketahui. Model yang diturunkan didasarkan pada analisis dari pelatihan data.

Proses klasifikasi dibagi menjadi dua fase yaitu pelatihan dan pengujian. Pada fase pelatihan, data yang telah diketahui kelas datanya (training set) digunakan untuk membentuk model. Selanjutnya pada fase testing, model yang sudah terbentuk diuji dengan sebagian data lainnya (test set) untuk mengetahui akurasi dari model tersebut. Jika akurasinya mencukupi maka model tersebut dapat dipakai untuk prediksi kelas data yang belum diketahui (Han & Kamber, 2006)

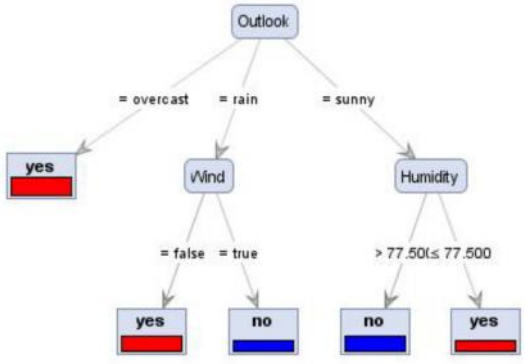
2.2 Decision Tree

Decision Tree adalah sebuah struktur pohon, dimana setiap *node* pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan *node* daun merepresentasikan kelompok kelas tertentu. Level *node* teratas dari sebuah *decision Tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *decision Tree* melakukan strategi pencarian secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari *node* akar (*root*) sampai *node* akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu. (Hastuti, 2012)

Terdapat 3 jenis node yang terdapat pada decision tree, yaitu:

a. *Root node*, merupakan *node* paling atas, pada*node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
b. *Internal Node*, merupakan *node* percabangan. Pada *node* ini terdapat percabangan. Pada *node* ini terdapat satu *input* dan memiliki *output* mininmal dua.
c. *Leaf node* atau terminal *node*, merupakan *node* akhir. Pada *node* ini terdapat satu *input* dan tidak mempunyai *output*.

Contoh decision tree dapat dilihat dari Gambar 1 berikut :



Gambar 1. Contoh *Decision Tree*

2.2 Algoritma NBTree

Salah satu algoritme pembentukan *decision tree* adalah algoritme NBTree. Algoritme NBTree merupakan algoritme hasil penggabungan teknik *decision tree classifier* dengan *naïve-bayes classifier*. Algoritme ini akan membangun *decision tree* dengan *node* yang mengandung *univariate split* seperti *decision tree* biasa, tetapi pada node *leaf* terkandung *naïve-bayes classifier.* (Kohavi, 1996)

Algoritma NBTree (Kohavi, 1996)

Input : himpunan T yang terdiri dari *instance* dengan label

Output : sebuah pohon keputusan dengan pengkategori Naïve-bayes pada daun

1. Hitung *utility* untuk setiap atribut X_i, $\mu(X_i)$ Untuk atribut kontinyu, dibuat sebuah *threshold.*
2. Misalkan $j = arg\,max_i(\mu_i)$, adalah atribut dengan nilai *utility* tertinggi.
3. Jika μ_j tidak lebih tinggi dibanding nilai *utility* yang dimiliki *node* yang sekarang, buat model Naïve-Bayes untuk *node* yang sekarang dan kembali ke langkah 1.
4. Bagi *T* menurut pengujian di X_j. Jika X_j adalah kontinyu, sebuah pembagian menggunakan *threshold* dibuat untuk semua nilai yang mungkin.
5. Untuk setiap *child*, panggil algoritme secara rekursif untuk membagi *T* yang sesuai dengan pengujian dari *child.*

Dengan memberikan sekumpulan *instance* ke suatu *node*, algoritme NBTree akan melakukan evaluasi *utility of split* untuk setiap atribut. Jika *utility* terbesar dari semua atribut lebih tinggi dibanding *utility* yang dimiliki node yang sekarang, maka akan dilakukan pembagian *instance-instance* yang ada berdasarkan atribut tersebut. (William *et all*, 2006)

Utility of node dihitung dengan melakukan diskretisasi pada data yang ada dan menghitung estimasi akurasi 5-*fold cross validation* dari penggunaan naïve-bayes di *node* tersebut. Sedangkan *utility of split* adalah jumlah bobot dari *utility of node*, dimana bobot yang diberikan ke sebuah node sebanding dengan jumlah *instance* yang diturunkan *node* tersebut. Pembagian ditetapkan signifikan jika reduksi relatif terhadap kesalahan lebih bagus dari 5% dan setidaknya terdapat 30 *instance* di *node* tersebut. Hal ini untuk menghindari terjadinya pembagian dengan nilai yang kecil. (Kohavi, 1996)

2.3 Naïve Bayes Classifier

Klasifikasi Naive Bayes dapat diuraikan sebagai berikut :

Asumsi bahwa setiap *instance* direpresentasikan dengan sebuah vektor $\boldsymbol{X}=(x_1,x_2,\ldots,x_n)$, dimana $x_1,x_2,\ldots,xn$ adalah ukuran dari atribut $A_1,A_2,\ldots,A_n$. Andaikan terdapat kelas sejumlah *m* yaitu $C_1,C_2,\ldots,C_m$. Diberikan suatu *instance* $\boldsymbol{X}$ yang belum diketahui kelasnya, dengan menggunakan teorema Bayesian, *posterior probability* dari $\boldsymbol{X}$ terhadap C_i ditunjukkan pada persamaan (1).

$$P(C_i|\boldsymbol{X}) = [P(\boldsymbol{X}|C_i)P(C_i)]/P(\boldsymbol{X}) = [P(\boldsymbol{X}|C_i)P(C_i)]/[\sum_{k=1}^{m} P(Ck)P(\boldsymbol{X}|Ck)] \quad (1)$$

Class prior probability dapat diduga dengan $P(C_l)=S_l/S$, dimana s_l adalah jumlah dari data pelatihan dengan kelas C_l dan *s* adalah jumlah total data pelatihan.

Naive Bayes menduga *conditionally independent* antara satu atribut dan atribut lainnya dengan menggunakan persamaan (2).

$$P(X|Cl)= \prod_{k=1}^{n} P(X_k|C_l) \quad (2)$$

$P(x_k|C_l)$ dapat diduga dari data. Sehingga dengan menggunakan persamaan (3) dapat diperoleh nilai peluang $P(Cl|\boldsymbol{X})$.

$$P(Cl|\boldsymbol{X})= P(C_l) \prod_{k=1}^{n} P(X_k|C_l)/[\sum_{h=1}^{m} P(Ch)P(X_k|C_h)] \quad (3)$$

Untuk menggolongkan sebuah data ***X*** yang belum diketahui kelasnya, $P(C_l|\boldsymbol{X})$ dievaluasi untuk setiap kelas C_l. Data ***X*** akan dimasukkan dalam kelas C_l jika dan hanya jika $P(C_l|\boldsymbol{X}) > P(Cj|\boldsymbol{X})$, $1 \le j \le m$, $j \ne l$. (Deng & Peng 2006)

2.4 *Confusion Matrix*

Confusion matrix merupakan sebuah tabel yang berisi jumlah banyaknya *test record* yang diprediksi secara benar dan tidak benar oleh model klasifikasi. Bentuk dari *confusion matrix* terlihat pada Tabel 1. Setiap entri pada f_{ij} pada tabel ini menyatakan banyaknya *record* dari kelas i yang diprediksi ke dalam kelas j.

Tabel 1. *Confussion Matrix*

		Kelas yang diprediksi	
		Kelas = 1	Kelas = 0
Kelas aktual	Kelas = 1	f_{11}	f_{11}
	Kelas = 2	f_{01}	f_{00}

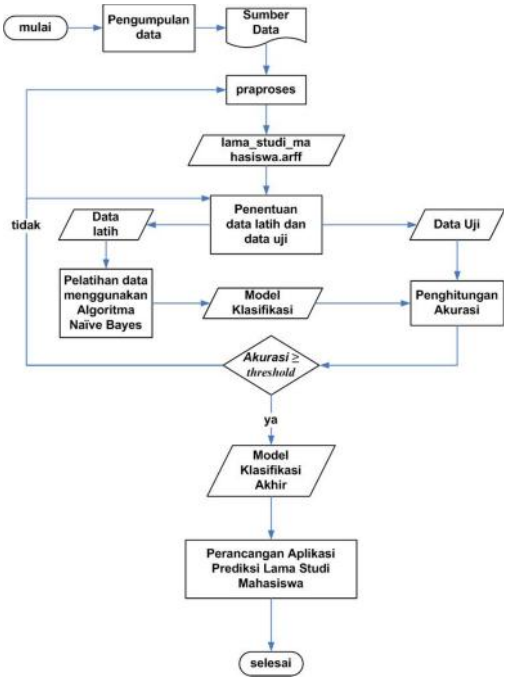
Informasi dari *confusion matrix* diperlukan untuk menentukan kinerja suatu model klasifikasi. Informasi ini dapat diringkas ke dalam suatu nilai seperti akurasi. (Tan et al, 2005)

$$\text{akurasi}= \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}} = \frac{f_{11}+ f_{00}}{f_{11}+ f_{10}+ f_{01}+ f_{00}}$$

3. METODE PENELITIAN

3.1 Metode Penelitian Yang Digunakan

Data yang digunakan dalam penelitian ini merupakan data lama studi mahasiswa yang terdiri dari nilai mahasiswa per mata kuliah mulai dari semester 1 sampai semester 5, data cuti kuliah, dan data ketepatan waktu lulus. Data mahasiswa yang digunakan adalah data mahasiswa kelulusan tahun 2013 sampai dengan tahun 2015 di program studi sistem informasi STMIK Indonesia. Penelitian ini dilakukan secara bertahap sesuai tahapan yang telah disusun pada Gambar 2. Nilai *threshold* yang digunakan pada penelitian ini adalah 70%.



Gambar 2. Metode Penelitian

3.2 Praproses Data

Data dari sumber dikumpulkan dan dilakukan tahapan praproses data sebagai berikut :

1 Seleksi data
Pada tahap ini, dilakukan pengelompokkan mahasiswa berdasarkan lama studinya dan memilih atribut-atribut yang sesuai dengan kategori permasalahan.

2 Pembersihan data
Pada data dilakukan pembersihan data untuk memperbaiki data yang hilang atau kosong, data yang mengandung *noise*, dan data yang tidak konsisten.

3 Integrasi data
Pada tahap ini dilakukan penggabungan data dari berbagai sumber ke suatu basis data. Kemudian dilakukan proses reduksi data, dimana data yang tidak relevan dan data yang redudansi dibuang.

4 Transformasi data
Proses perubahan bentuk ke dalam bentuk data yang tepat agar dapat digunakan untuk proses selanjutnya. Proses ini meliputi penyeragaman nama atribut.

3.3 Pembagian Data Latih dan Data Uji

Proses pembagian data menjadi data latih dan data uji dilakukan dengan menggunakan 10 *fold cross validation*. Data latih akan digunakan untuk membentuk model klasifikasi. Sedangkan data uji akan digunakan untuk menghitung akurasi yang diperoleh dari model klasifikasi.

3.4 Klasifikasi

Pada proses klasifikasi dilakukan pembentukan model klasifikasi menggunakan metode *decision tree*. Kemudian dilakukan penghitungan akurasi dari model klasifikasi yang terbentuk. Dari proses klasifikasi ini akan diperoleh

model klasifikasi yang dapat digunakan untuk mengisi label kelas dari data baru yang belum diketahui label kelasnya.

3.5 Model Klasifikasi

Proses pembentukan model klasifikasi dari data mahasiswa menggunakan data tabel gabungan dan kelas targetnya. Pembentukan model klasifikasi ini menggunakan salah satu algoritme pembentukan *decision tree* yaitu algoritme NBTree.

3.6 Penghitungan Akurasi

Tahap ini adalah tahap untuk menghitung akurasi dari model klasifikasi yang diperoleh dari proses klasifikasi. Metode yang digunakan dalam proses penghitungan akurasi ini adalah dengan menggunakan *confussion matrix*.

Jika hasil akurasi yang diperoleh sudah memenuhi nilai *threshold*, maka model klasifikasi itu akan digunakan untuk menentukan label kelas dari data baru. Akan tetapi jika akurasi yang diperoleh belum memenuhi nilai *threshold*, maka proses klasifikasi akan diulang dengan menggunakan proporsi data latih dan data uji yang berbeda atau mengulang tahap praproses dengan objek yang berbeda.

3.7 Perancangan Aplikasi Prediksi Lama Studi Mahasiswa

Pada tahap ini akan dilakukan perancangan aplikasi untuk melakukan prediksi lama studi mahasiswa. Model klasifikasi yang telah diperoleh dari tahap sebelumnya akan diimplementasikan pada aplikasi ini. Aplikasi ini akan dibangun dengan menggunakan bahasa pemrograman Java dan basis data MySQL.

4. HASIL DAN PEMBAHASAN

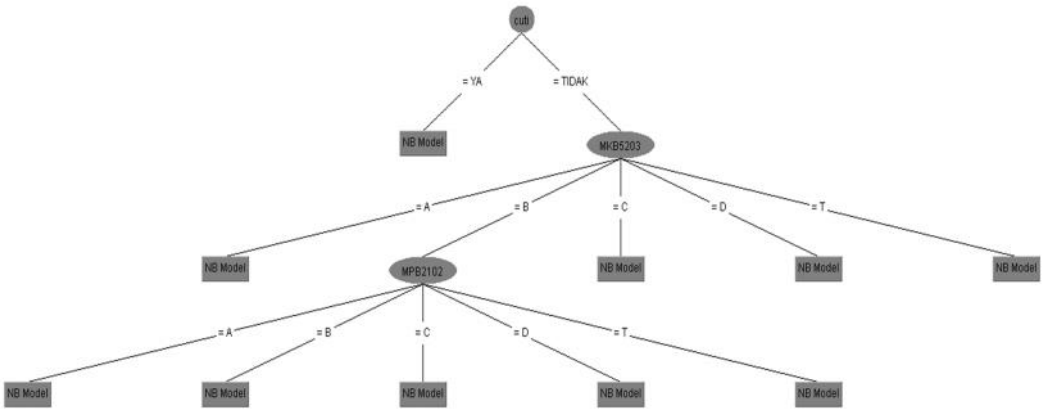
4.1 Praproses Data

Data yang digunakan pada penelitian ini adalah data nilai akademik mahasiswa, data cuti akademik, dan data ketepatan waktu lulus mahasiswa program studi sistem informasi tahun 2013 – 2015. Tidak semua data yang terdapat pada data nilai akademik mahasiswa digunakan pada penelitian ini. Setelah melalui tahap pemilihan atribut, terdapat 36 atribut yang digunakan pada penelitian ini. Atribut itu terdiri dari 34 mata kuliah semester 1 sampai semester 5, cuti kuliah dan ketepatan lulus studi. Atribut ketepatan lulus studi menjadi kelas dari data yang digunakan pada penelitian ini.

Proses selanjutnya adalah proses pembersihan data. Salah satu tujuan proses pembersihan data adalah untuk mengganti data yang kosong. Jika terdapat niali atribut yang kosong untuk suatu record, akan diganti dengan nilai T. Dimana nilai T ini berarti mahasiswa tersebut tidak mengambil mata kuliah tersebut. Hal ini terjadi karena terdapat perbedaan kurikulum antara mahasiswa yang lulus tahun 2013 dengan mahasiswa yang lulus tahun 2014 – 2015. Selain itu nilai masing-masing atribut mata kuliah terdiri dari A, B, C, D, dan T. Tidak terdapat nilai E, dikarenakan salah satu persyaratan sidang skripsi adalah tidak terdapat nilai E. Sehingga mahasiswa yang sudah lulus, otomatis tidak memiliki nilai E. Kemudian masing-masing data tersebut digabungkan menjadi satu.

4.2 Klasifikasi

Proses klasifikasi dilakukan dalam dua tahap, yaitu pembentukan model klasifikasi berupa *decision tree* dan penghitungan akurasi dari *decision tree* yang terbentuk. Pembentukan *decision tree* dilakukan dengan menggunakan algoritme NBTree. Proses pembentukannya dilakukan dengan perangkat lunak Weka. Gambar 3 merupakan *decision tree* yang dibentuk. Pada *decision tree* yang dihasilkan, diketahui bahwa tidak semua atribut yang digunakan muncul sebagai *node* pada *decision tree*.



Gambar 3. Model Klasifikasi Data Lama Studi Mahasiswa Program Studi Sistem Informasi STMIK Indonesia

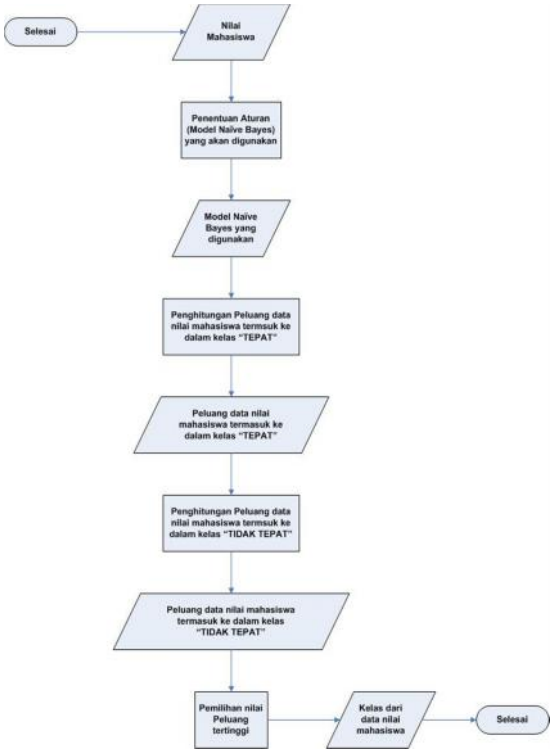
Dari gambar 3 di atas, dapat dilihat bahwa dari 35 atribut yang digunakan, hanya 3 atribut yang muncul pada *decision tree* tersebut, yaitu cuti kuliah, MKB5203 (Mata kuliah Sistem Operasi) dan MPB2102 (Mata kuliah Komputer dan Masyarakat). *Leaf node* yang dihasilkan dari *decision tree* yang dibentuk dari algoritme NBTree merupakan sebuah model naïve bayes, dimana model ini berisi peluang

untuk masing-masing kelas, dan peluang setiap atribut terhadap masing-masing kelas.
Berdasarkan *decision tree* yang terbentuk, dapat dibuat 10 aturan. Sebagai contoh, aturan yang terbentuk dari *decision tree* pada Gambar 3, adalah sebagai berikut :

- Aturan 1:
  Jika nilai atribut cuti adalah YA, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 1.
- Aturan 2:
  Jika nilai atribut cuti adalah TIDAK dan nilai atribut MKB5203 adalah A maka penentuan kelas dapat dihitung menggunakan model naïve bayes 2.
- Aturan 3:
  Jika nilai atribut cuti adalah TIDAK dan nilai atribut MKB5203 adalah C maka penentuan kelas dapat dihitung menggunakan model naïve bayes 3.
- Aturan 4:
  Jika nilai atribut cuti adalah TIDAK dan nilai atribut MKB5203 adalah D maka penentuan kelas dapat dihitung menggunakan model naïve bayes 4.
- Aturan 5:
  Jika nilai atribut cuti adalah TIDAK dan nilai atribut MKB5203 adalah T maka penentuan kelas dapat dihitung menggunakan model naïve bayes 5.
- Aturan 6:
  Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2012 adalah A, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 6.
- Aturan 7:
  Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2012 adalahB, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 7.
- Aturan 8:
  Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2012 adalah C, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 8.
- Aturan 9:
  Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2012 adalah D, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 9.
- Aturan 10:
  Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2012 adalah E, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 10.

Masing-masing aturan tersebut akan menggunakan model Naïve Bayes yang berbeda. Model Naïve Bayes adalah rangkaian nilai peluang sebuah atribut terhadap sebuah kelas serta nilai peluang untuk masing-masing kelas. Rangkaian nilai peluang ini akan berbeda untuk setiap model Naïve Bayes. Model Naïve Bayes ini yang akan digunakan dalam penentuan kelas dari sebuah data baru.

Ketika sebuah data nilai mahasiswa baru yang belum memiliki kelas ingin diketahui kelasnya, ada beberapa perhitungan peluang yang dilakukan. Gambar 4 menunjukkan tahapan penghitungan peluang yang digunakan dalam penentuan kelas (lama studi mahasiswa) dari sebuah data baru.



Gambar 4. Tahap Penghitungan Peluang Dalam Penentuan Kelas dari Data Nilai Mahasiswa Baru

Dari Gambar 4 dapat dilihat ada beberapa tahap penghitungan peluang. Penghitungan peluang itu dilakukan dengan menggunakan nilai Naïve Bayes. Nilai yang dihitung dengan menggunakan model naïve bayes adalah *conditionally independent* antara satu atribut dan atribut lainnya dengan menggunakan persamaan (2). Kemudian dilakukan penghitungan peluang suatu *record* termasuk ke dalam setiap kelas yang dihitung dengan Persamaan (3). Pengitungan ini dilakukan untuk semua kelas yang ada. Untuk penentuan kelas, diambil berdasarkan nilai peluang suatu record termasuk ke dalam suatu kelas yang paling tinggi.

Tahapan penentuan kelas dari data nilai akademik mahasiswa baru adalah :

1. Tahap awal dari penentuan kelas ini adalah mengambil data nilai akademik dari mahasiswa yang telah dipilih.
2. Kemudian dilakukan penentuan aturan yang sesuai dengan kondisi data yang ada. Aturan ini diambil dari model klasifikasi yang digunakan. Setelah didapatkan aturan yang sesuai, maka akan diperoleh model Naïve Bayes yang akan digunakan dalam penghitungan peluang.
3. Kemudian dilakukan penghitungan *conditionally independent* antara satu atribut

dan atribut lainnya terhadap kelas “TEPAT” dengan menggunakan persamaan (2). Setelah itu dilakukan penghitungan peluang suatu *record* termasuk ke dalam kelas “TEPAT” yang dihitung dengan Persamaan (3).

4. Tahap selanjutnya adalah penghitungan *conditionally independent* antara satu atribut dan atribut lainnya terhadap kelas “TIDAK TEPAT” dengan menggunakan persamaan yang sama dengan saat melakukan penghitungan terhadap kelas “TEPAT”. Kemudian dilakukan juga penghitungan penghitungan peluang suatu *record* termasuk ke dalam kelas “TIDAK TEPAT” seperti pada kelas “TEPAT”.
5. Setelah itu ditentukan pemilihan nilai peluang yang tertinggi dari nilai peluang data nilai mahasiswa terhadap masing-masing kelas. Kelas dengan nilai peluang tertinggi yang akan dipilih sebagai kelas dari data nilai mahasiswa baru tersebut.

4.3 Penghitungan Akurasi

Untuk menghitung akurasi dari model klasifikasi yang terbentuk digunakan *confusion matrix*. *Confusion matrix* yang diperoleh dari model klasifikasi tersebut terdapat pada Tabel 2.

Tabel 2. *Confusion matrix* dari model klasifikasi

		Kelas yang diprediksi	
		Kelas = Tepat	Kelas = Tidak Tepat
Kelas Aktual	Kelas = Tepat	270	35
	Kelas = Tidak Tepat	115	145

Penghitungan akurasi dengan menggunakan *confusion matrix* adalah sebagai berikut:

$$\text{akurasi} = \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}}$$

Dengan menggunakan data pada tabel *confusion matrix*, dapat dihitung akurasi dari model klasifikasi.

$$\text{akurasi} = \frac{270 + 145}{565}$$

$$= \frac{415}{565}$$

$$= 0{,}7345$$

Hasil akurasi yang diperoleh adalah 73,45%. Sedangkan nilai *threshold* yang digunakan adalah 70%. Sehingga nilai akurasi yang diperoleh telah memenuhi *threshold* yang diberikan.

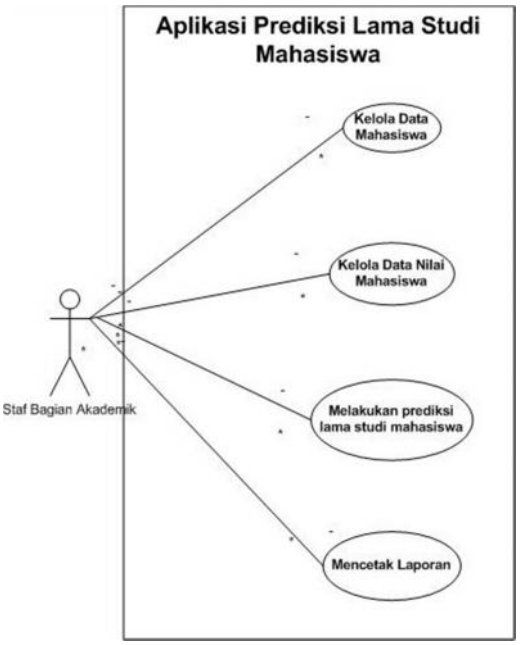
4.4 Perancangan dam Pembuatan Aplikasi

Dalam pengembangan rancangan aplikasi desktop yang mengimplementasi model klasifikasi digunakan kerangka spesifik UML. Diagram UML yang digunakan adalah *Use Case* diagram. Daftar *Use Case* pada aplikasi prediksi lama studi mahasiswa dapat dilihat pada Tabel 2.

Tabel 2. Daftar *Use Case*

Nama *Use Case*	Deskripsi *Use Case*
Kelola Data Mahasiswa	Menambahkan data pribadi mahasiswa baru, mengubah data pribadi mahasiswa, dan menghapus data pribadi mahasiswa
Kelola Data Nilai Akademik Mahasiswa	Menambahkan data nilai mahasiswa dan mengubah data nilai mahasiswa.
Melakukan Prediksi Lama Studi Mahasiswa	Melakukan prediksi lama studi mahasiswa dengan memilih terlebih dahulu mahasiswa mana yang akan diprediksi.
Mencetak Laporan	Melihat dan mencetak laporan hasil prediksi lama studi mahasiswa

Dari daftar use case di atas, dapat dibuat diagram use case. Diagram *Use Case* dapat dilihat pada Gambar 5.

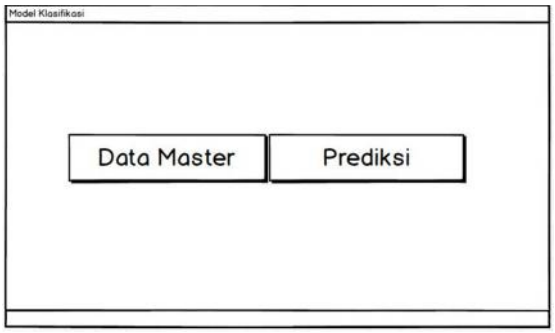


Gambar 5. *Use Case* Diagram Aplikasi Prediksi Lama Studi Mahasiswa

Dari Gambar 5 di atas, dapat dilihat bahwa pengguna dari aplikasi ini adalah Staf Bagian Akademik. Pengguna inilah yang akan melakukan kelola data mahasiswa, kelola data nilai mahasiswa, melakukan prediksi lama studi mahasiswa, dan mencetak laporan. Fungsi kelola data mahasiswa merupakan fungsi pengelolaan data pribadi mahasiswa bukan data nilai akademik mahasiswa. Fungsi ini terdiri dari menambahkan data mahasiswa baru, mengubah data mahasiswa, dan menghapus data mahasiswa. Fungsi kedua dari aplikasi ini adalah kelola data nilai mahasiswa. Fungsi ini memungkinkan pengguna untuk

melakukan pengelolaan data nilai akademik setiap mahasiswa. Dengan adanya fungsi ini, pengguna dapat melakukan penambahan data nilai akademik baru untuk setiap mahasiswa dan pengubahan data nilai akademik setiap mahasiswa. Fungsi ketiga adalah melakukan prediksi lama studi mahasiswa. Dengan fungsi ini, pengguna dapat melakukan prediksi lama studi mahasiswa dan melihat hasil prediksinya. Fungsi terakhir dari aplikasi ini adalah fungsi mencetak laporan. Pengguna dapat melihat laporan selama periode tertentu yang dipilih. Laporan ini berisi hasil prediksi yang dilakukan selama periode tertentu. Selain itu, laporan prediksi dapat dilihat untuk sekelompok mahasiswa yang dipilih. Laporan ini kemudian dapat dicetak untuk diolah lebih lanjut.

Rancangan tampilan aplikasi prediksi lama studi mahasiswa dapat dilihat di Gambar 6, Gambar 7, Gambar 8, Gambar 9, dan Gambar 10 di bawah ini. Pada Gambar 5 dapat dilihat rancangan antar muka dari menu utama. Pada halaman menu utama ini terdapat 2 sub menu yaitu Menu Data Master dan Menu Prediksi.

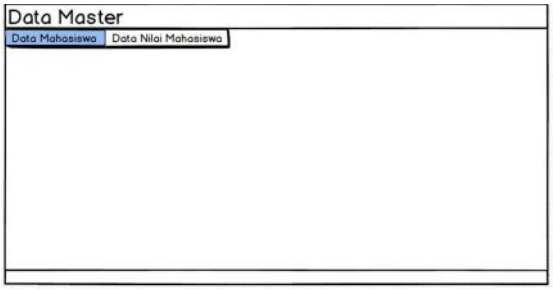


Gambar 6. Rancangan Tampilan Menu Utama dari Aplikasi Prediksi Lama Studi Mahasiswa

Menu Data Master adalah menu yang dapat digunakan pengguna untuk melakukan 2 fungsi dari aplikasi prediksi lama studi mahasiswa yaitu kelola data mahasiswa dan kelola data nilai mahasiswa. Rancangan tampilan dari Menu Data Master ini dapat dilihat pada Gambar 6. Terdapat 2 sub menu yaitu Menu Data Mahasiswa dan Menu Data Nilai Mahasiswa.

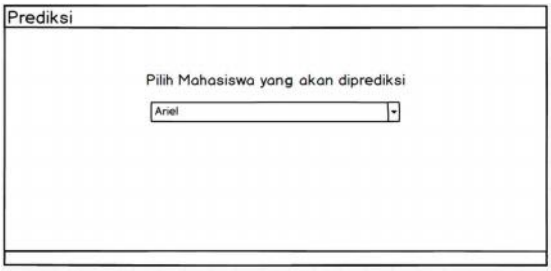
Menu Data Mahasiswa merupakan tampilan untuk fungsi kelola data mahasiswa. Sesuai dengan deskripsi yang terdapat pada Tabel 2, bahwa kelola data mahasiswa adalah fungsi untuk mengelola data pribadi mahasiswa seperti NIM, Nama, Tempat Tanggal Lahir, Alamat, Sekolah Asal, dan beberapa atribut lainnya. Fungsi kelola data mahasiswa ini terdiri dari menambahkan data mahasiswa baru, mengubah data mahasiswa, dan menghapus data mahasiswa.

Menu Data Nilai Mahasiswa merupakan tampilan untuk fungsi kelola data nilai mahasiswa. Pada menu ini, pengguna dapat melakukan kelola data nilai mahasiswa per mata kuliah, dimana dapat menambahkan data nilai yang baru dan mengubah data nilai mahasiswa. Pada menu ini tidak terdapat fungsi menghapus data nilai mahasiswa. Karena data nilai mahasiswa merupakan salah satu data penting yang sebaiknya tidak hilang.

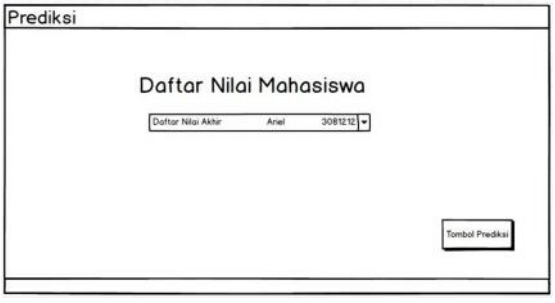


Gambar 7. Rancangan Tampilan Menu Data Master dari Aplikasi Prediksi Lama Studi Mahasiswa

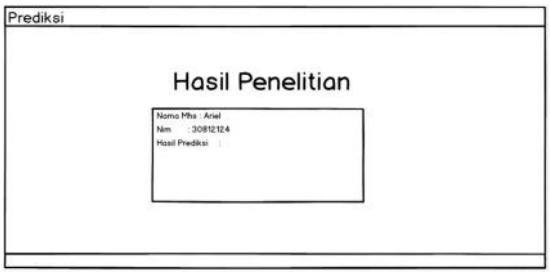
Rancangan Menu Prediksi dapat dilihat pada Gambar 7. Menu Prediksi ini digunakan pengguna untuk melakukan prediksi lama studi mahasiswa. Langkah pertama yang dilakukan pengguna adalah memilih mahasiswa yang akan dilakukan prediksi lama studinya. Kemudian aplikasi akan menampilkan daftar nilai akademik per mata kuliah dari mahasiswa tersebut. Nilai yang akan muncul adalah nilai mata kuliah dari semester 1 sampai semester 2. Rancangan tampilannya dapat dilihat pada Gambar 8. Kemudian setelah tombol prediksi di klik, maka aplikasi akan menampilkan hasil prediksi lama studi mahasiswa tersebut. Rancangan tampilannya dapat dilihat pada Gambar 9.



Gambar 8. Rancangan Tampilan Menu Prediksi (Pemilihan mahasiswa yang akan diprediksi) dari Aplikasi Prediksi Lama Studi Mahasiswa



Gambar 9. Rancangan Tampilan Menu Prediksi (Daftar Nilai Akademik Mahasiswa) dari Aplikasi Prediksi Lama Studi Mahasiswa



Gambar 10. Rancangan Tampilan Menu Prediksi (Hasil Prediksi) dari Aplikasi Prediksi Lama Studi Mahasiswa

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan dalam membentuk model klasifikasi untuk data lama studi mahasiswa, dapat diambil kesimpulan sebagai berikut :

1 Telah terbentuk model klasifikasi untuk data lama studi mahasiswa yang memiliki 10 aturan klasifikasi dengan akurasi 73,45%.

2 Lama studi mahasiswa dapat dideskripsikan oleh nilai akademik dan data cuti akademik mahasiswa.

5.2 Saran

1. Data yang digunakan bisa ditambah lebih dari 3 tahun,agar penelitiannya, model data klasifikasinya lebih variatif.
2. Gunakan metode penelitian yang lain, agar bisa membandingkan hasil penelitian NBTree

DAFTAR PUSTAKA

Han J, Kamber M. (2006). Data Mining : Concepts and Techniques. San Francisco : Morgan Kaufman Publisher.

Hastuti K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. Semarang : Seminar Nasional Teknologi Informasi & Komunikasi Terapan.

Kohavi R. (1996). Scaling Up the Accuracy of Naïve-Bayes Classifiers : a Decision-Tree Hybrid. AAAI.

Tan P, Michael S, dan Vipin K. (2005). Introduction to Data mining. Boston : Pearson Education, Inc.

Williams N, Zander S, & Armitage G. (2006). Evaluating Machine Learning Algorithms for Automated Network Application Identification. CAIA.