

Decision Support System Design and Development for Determine Graduate Phase of College Students with Naïve Bayes Algorithm Web-Based in Indonesia Institute of Business and Technology

(Rancang Bangun Sistem Pendukung Keputusan Menentukan Masa Kelulusan Mahasiswa dengan Algoritma *Naïve Bayes* Berbasis *Web* (Studi Kasus: INSTIKI))

Andres Adyana^{*}, Ayu Manik Dirgayusari, Nia Maharani, I Gede Andika,
I Kadek Dwi Gandika Supartha

Program Studi Sistem Komputer, Fakultas Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia,
Jl. Tukad Pakerisan No. 97 Denpasar, Bali, Indonesia

ABSTRACT

In the course of the lecture period, each student is different, various factors can affect the mental and academic achievement of students which have an impact on their graduation period. This things also impact on the campus of the Indonesian Institute of Business and Technology, until now there is no system that can determine the graduation period of students at the campus, therefore it is necessary to build a decision support system to determine the graduation period of students, especially students in the Informatics Engineering Study Program and Program Study of Computer Systems at the Indonesian Institute of Business and Technology (INSTIKI) using the Naïve Bayes algorithm that utilizes parameters, namely student academic data consisting of work status, Semester Achievement Index (IPS) scores semesters one until four, and Grade Point Average (IPK). The system designed by using the GaussianNB library in python and website based. From the results of this study, the accuracy of the model made on the system produces an accuracy value of 87.5% with a dataset rule of 800 data and the data is divided and used as test data by 20%, while the accuracy value of classifying 250 test data on the system is 80.8%.

Didalam perjalanan masa perkuliahan setiap mahasiswa berbeda-beda, berbagai faktor dapat mempengaruhi mental dan prestasi akademik mahasiswa yang berdampak pada masa kelulusannya. Hal tersebut juga dialami oleh kampus Institut Bisnis dan Teknologi Indonesia dan sampai saat ini belum ada sistem yang dapat menentukan masa kelulusan mahasiswa di kampus tersebut, maka dari itu perlu dibangunnya sistem pendukung keputusan untuk menentukan masa kelulusan mahasiswa khususnya mahasiswa pada Program Studi Teknik Informatika dan Program Studi Sistem Komputer Institut Bisnis dan Teknologi Indonesia (INSTIKI) dengan menggunakan algoritma *Naïve Bayes* yang memanfaatkan parameter yaitu data akademik mahasiswa yang terdiri dari status kerja, nilai Indeks Prestasi Semester (IPS) semester satu sampai empat, dan nilai Indeks Prestasi Kumulatif (IPK). Didalam perancangan bangun sistem menggunakan *library* GaussianNB pada *python* dan berbasis *website*. Dari hasil penelitian ini akurasi model yang dibuat pada sistem menghasilkan nilai akurasi sebesar 87,5% dengan aturan dataset berjumlah 800 data dan data tersebut dibagi dan digunakan sebagai data uji sebesar 20%, sedangkan nilai akurasi dari pengklasifikasian 250 data uji pada sistem sebesar 80,8%.

Keywords: Classification, Graduation, Naïve Bayes.

^{*}Corresponding author:
Andres Adyana
E-mail: andresadyana741@gmail.com

PENDAHULUAN

Didalam perguruan tinggi terdapat banyak permasalahan yang dapat timbul, baik dari kebutuhan sarana dan prasarana perkuliahan, kualitas tenaga

pengajar atau dosen, dan kualitas pada pelajarnya atau mahasiswa. Salah satu permasalahan yang terdapat pada mahasiswa adalah mahasiswa tidak lulus dengan tepat waktu dikarenakan berbagai faktor yang dapat mempengaruhi masa kelulusannya.

Seorang mahasiswa diharapkan mampu menyelesaikan perkuliahannya dalam kurun waktu selama 7-8 semester, salah satu permasalahan yang ada adalah adanya mahasiswa yang terlambat lulus atau tidak tepat pada waktunya sehingga menjadi pengaruh didalam lingkungan perguruan tinggi tersebut dan pada mahasiswa itu sendiri. Untuk mengatasi permasalahan tersebut perlu dilakukan identifikasi sejak dini agar pihak kampus mengetahui siapa sajakah mahasiswa yang menempuh masa studi melebihi standar kelulusan yang telah di tetapkan sehingga kecenderungan mahasiswa lulus tidak tepat waktu dapat dicegah. Salah satu penanganannya adalah dengan menentukan masa kelulusan mahasiswa melalui status masa studinya.

Institut Bisnis dan Teknologi Indonesia memiliki data yang besar, yaitu data akademik mahasiswa yang dikelola pada Sistem Akademik INSTIKI (SADS) sistem tersebut mengelola Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif Mahasiswa (IPK). IPS dan IPK berupa data informasi yang sangat bermanfaat untuk mengetahui nilai mahasiswa per-semester, tetapi sistem tersebut belum dapat memprediksi masa kelulusan mahasiswa.

Berdasarkan permasalahan diatas, dibutuhkan sistem pendukung keputusan untuk menentukan masa kelulusan mahasiswa yang tepat waktu dan tidak tepat waktu khususnya mahasiswa pada Program Studi Teknik Informatika dan Program Studi Sistem Komputer Institut Bisnis dan Teknologi Indonesia dengan algoritma *Naïve Bayes* yang memanfaatkan parameter yaitu data akademik mahasiswa yang terdiri dari status kerja, nilai Indeks Prestasi Semester (IPS) semester satu sampai empat, dan IPK.

Beberapa penelitian telah membahas mengenai pembuatan sistem pendukung keputusan memprediksi tingkat kelulusan mahasiswa diantaranya sistem pendukung keputusan untuk memprediksi kelulusan mahasiswa pada Program Studi Pendidikan Teknik Informatika dan Komputer dengan Menggunakan Algoritma *Naïve Bayes* oleh Reni, dkk [1]. Sistem pendukung keputusan tersebut dibuat untuk memprediksi tingkat kelulusan mahasiswa khususnya Program Studi Pendidikan Teknik Informatika dan Komputer Fakultas Teknik Universitas Negeri Jakarta. Metode *Naïve Bayes* digunakan untuk menganalisis data dalam pengenalan pola dan memprediksi kelulusan mahasiswa, data yang dibutuhkan adalah data mahasiswa yang sudah lulus

sebagai data training sedangkan data mahasiswa yang masih menempuh studi kuliah sebagai data uji. Dari hasil penelitian tersebut memperoleh tingkat akurasi 85%.

Sistem pendukung keputusan yang dibuat oleh Isa. Aplikasi Prediksi Kelulusan Tepat Waktu Dengan Algoritma *Naïve Bayes* (Studi Kasus Program Studiteknik Informatika, UMMI) [2]. Perancangan sistem menggunakan pemodelan UML dengan implementasi menggunakan bahasa pemrograman PHP dan basis data MySQL, dalam penelitian tersebut menerapkan tahapan-tahapan Algoritma *Naïve Bayes* pada perancangan aplikasi untuk memprediksi kelulusan tepat waktu mahasiswa berdasarkan parameter yang terdapat pada Penerimaan Mahasiswa Baru (PMB), Indeks Prestasi Kumulatif (IPK) dan Bagian Keuangan. Data yang digunakan adalah data mahasiswa Program Studi Teknik Informatika Tahun Ajaran 2016/2017. Berdasarkan hasil pengujian didapatkan tingkat error data sebanyak 1.88%.

Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah oleh Testiana [3]. Sistem tersebut didesain dengan menggunakan model pengembangan CRISP-DM (*Cross Industry Standard Process for Data Mining*), percobaan pada penelitian tersebut menggunakan Rapidminer 5.3.008. Algoritma yang digunakan adalah *Naïve Bayes*. Proses validasi menggunakan *x-validation* dan untuk *testing* menggunakan *Apply Model* untuk menjalankan algoritma atau model *Naïve Bayes* serta aksi *Performance* untuk mengukur performa dari model *Naïve Bayes* tersebut. Dari hasil penelitian nilai akurasi terhadap klasifikasi kelulusan sebesar 82.08 %.

Penelitian terdahulu selanjutnya adalah optimasi model prediksi kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* oleh Hartatik [4]. Pada sistem tersebut untuk menerapkan metode *Naïve Bayes* penulis menggunakan parameter Indeks Prestasi Kumulatif dan sosial parameter yaitu jenis kelamin dan status tinggal. Hasil dari penelitian tersebut mendapatkan nilai akurasi untuk metode *Naïve Bayes* sebesar 75% dan akurasi untuk model prediksi dengan parameter sosial sebesar 85% dengan selisih 10%.

Penelitian nurjoko dan kurniawan, aplikasi data *mining* untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma *APRIORI* di IBI Darmajaya Bandar Lampung [5]. Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, dapat diharapkan untuk

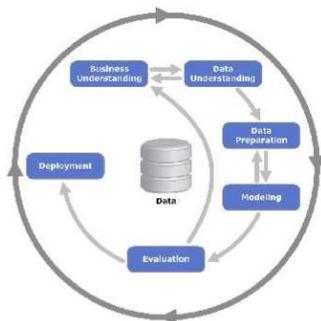
menghasilkan informasi tentang tingkat kelulusan dan data induk mahasiswa melalui teknik data *mining*. Kategori tingkat kelulusan di ukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma *apriori*, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan. Hasil prediksi dari sistem pendukung keputusan tersebut memiliki akurasi sebesar 70%.

Berdasarkan beberapa penelitian terdahulu maka, pada penelitian ini dibangun sebuah sistem pendukung keputusan menentukan masa kelulusan mahasiswa di Institut Bisnis dan Teknologi Indonesia menggunakan algoritma *Naïve Bayes* sebagai metode klasifikasi dan parameter atribut yang digunakan untuk memprediksi tingkat kelulusan mahasiswa adalah data akademik mahasiswa berupa status kerja, IPS 1, IPS 2, IPS 3, IPS 4, dan IPK. Pada penelitian ini menggunakan model metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*), dimana model metodologi tersebut merupakan standar dari proses data *mining*.

METODE PENELITIAN

Alur Penelitian

Penelitian ini didesain dengan menggunakan model CRISP-DM (*Cross Industry Standard Process for Data Mining*), dalam metode ini terdapat enam tahapan: (1) *Business/Research Understanding* (Pemahaman Penelitian). (2) *Data Understanding Phase* (Tahap Pemahaman Data). (3) *Data Preparation* (Persiapan Data). (4) *Modelling Phase* (Tahap Pemodelan). (5) *Evaluation Phase* (Tahap Evaluasi). (6) *Deployment Phase* (Tahap Penyebaran). Metodologi atau tahapan penelitian diperlukan sebagai kerangka dan panduan proses penelitian, sehingga rangkaian proses penelitian dapat dilakukan secara terarah, teratur dan sistematis [3].



Gambar 1. Model CRISP-DM (*Cross Industry Standard Process for Data Mining*)

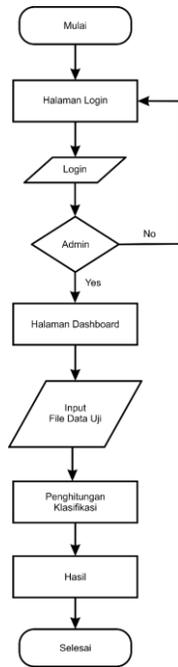
Business/Research Understanding (Pemahaman Penelitian)

Pengumpulan data pada penelitian ini dilakukan menggunakan metode observasi, wawancara, metode dokumentasi, dan metode kepustakaan. Observasi merupakan salah satu teknik pengumpulan data dengan melakukan pengamatan secara langsung terhadap objek dan peran yang terlibat [6]. Teknik pengumpulan data dengan observasi dilakukan untuk mengetahui skema dari berjalannya sistem.

Metode wawancara, wawancara dilakukan di ruangan Prodi Institut Bisnis dan Teknologi Indonesia Bersama bapak I Nyoman Buda Hartawan, S.Kom., M.Kom sebagai narasumber. Kegiatan wawancara berguna untuk mencari informasi dan menemukan permasalahan yang harus diteliti serta menghasilkan data analisis yang diperlukan guna menentukan data atribut yang di fokuskan pada perancangan sistem. Hasil dari metode wawancara adalah sistem pendukung keputusan berbasis *website* dengan peran *admin* pada sistem dapat menginputkan *file* data uji yang berisikan data atribut yang dibutuhkan, lalu sistem akan mengklasifikasikan data uji tersebut menggunakan algoritma *Naïve Bayes* berdasarkan data kelulusan mahasiswa sebelumnya atau data latih lalu sistem akan memberikan hasil *output* berupa grafik dan tabel hasil yang berisikan atribut mahasiswa serta masa kelulusannya. Gambaran umum atau *flowchart* sistem dapat dilihat pada Gambar 2.

Data Understanding Phase (Tahap Pemahaman Data)

Hasil dari metode dokumentasi adalah hasil pengumpulan data pada SIMJAR Institut Bisnis dan Teknologi Indonesia diperoleh data mahasiswa angkatan tahun 2013 dan 2014. Total data keseluruhan berjumlah 800 data kelulusan mahasiswa yang akan digunakan sebagai dataset didalam pembuatan model algoritma dan data uji yang yang didapatkan berupa data mahasiswa angkatan tahun 2016 yang berjumlah 250 data, data uji tersebut akan digunakan sebagai data yang akan di ujikan pada sistem.

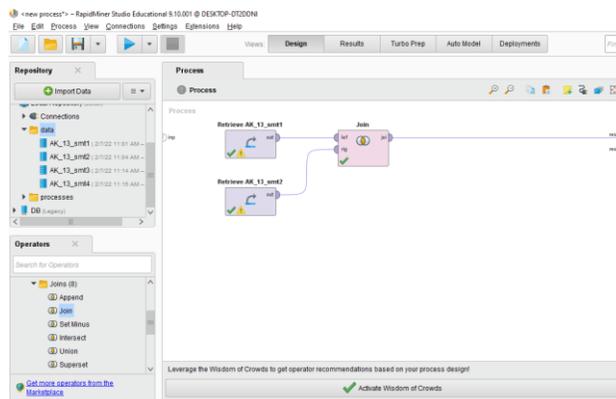


Gambar 2. Alur Flowchart Sistem

Data Preparation (Persiapan Data)

Tahap ketiga adalah tahap persiapan data yang meliputi proses penseleksian data dan integrasi data yang akan diteruskan pada tahap *extract, transform, and load*. Penseleksian data dilakukan menggunakan *Microsoft excel* lalu data dipisahkan sesuai kategori persemester dan diletakkan pada *database*[7]. Dikarenakan data yang diperoleh berupa data mentah maka perlu dilakukan proses *Extract Transform and Load* (ETL).

Proses ETL yang dilakukan antara lain melakukan proses integrasi data yang sudah dipisahkan agar menjadi suatu data baru yang terstruktur. Proses integrasi menggunakan *software rapidminer*, menggunakan fungsi *join* untuk mengintegrasikan setiap data dan setelah proses integrasi selesai data tersebut akan di *export* menjadi *file* yang berekstensi CSV. Proses integrasi dapat dilihat pada Gambar 3.



Gambar 3. Proses Integrasi Data

Tabel 1. Atribut Dataset

Atribut	Jenis Atribut	Total	Masa Kelulusan	
			Tepat Waktu (567)	Tidak Tepat Waktu (233)
Status Kerja	Bekerja	283	207	76
	Tidak Bekerja	517	360	157
IPS 1	< 3.0	204	93	111
	> 3.0	596	474	122
IPS 2	< 3.0	177	61	116
	> 3.0	623	506	117
IPS 3	< 3.0	172	55	117
	> 3.0	628	512	116
IPS 4	< 3.0	238	81	157
	> 3.0	562	486	76
IPK	< 3.0	218	65	153
	> 3.0	582	502	80

nim	no_tlp	nama	thn_aka	jurusan	sts_krj	ips1	ips2	ips3	ips4	ipk
16101001	81123123	Made Merta Yasa	2016/2017	TI-MDI	1	3.44	3.56	3.43	2.67	3.28
16101002	81123124	Ni Luh Tri Puspa Dewi	2016/2017	TI-KAB	0	3.56	3.67	3.67	3.5	3.6
16101003	81123125	I Gusti Ngurah Ambara Putra	2016/2017	TI-DM	0	3.44	3.44	3.62	3.5	3.5
16101006	81123126	Syinye Renita	2016/2017	TI-KAB	1	3.89	3.67	3.86	3.75	3.79
16101007	81123127	I Dewa Made Endra Juniarta	2016/2017	TI-KAB	1	3.33	3.44	3.22	3.5	3.37
16101009	81123128	Lisa Cahyani Puspa	2016/2017	TI-KAB	0	3.6	3.33	3.56	3.5	3.5
16101010	81123129	Kadek Riska Yuniati	2016/2017	TI-KAB	0	3.44	3.22	3.78	3	3.36
16101012	81123130	Ni Wayan Sarita Devi	2016/2017	TI-MDI	1	3.78	3.78	3.86	3.33	3.69
16101013	81123131	Ni Kadek Maryani	2016/2017	TI-KAB	0	3.78	3.56	3.78	3	3.53
16101014	81123132	I Made Surespa Yuda WG	2016/2017	TI-MDI	0	3.89	3.33	3.88	3	3.53
16101016	81123133	I Putu Karisma Putra Wibawa	2016/2017	TI-KAB	0	3.22	3	3.38	2	2.9
16101017	81123134	Niken Septiya Wulan Mawardani	2016/2017	TI-DM	1	3.56	3.44	3.62	3.67	3.57
16101018	81123135	Ni Made Sri Wahyuni	2016/2017	TI-DM	0	3.67	3.22	3.75	3.5	3.54
16101019	81123136	Ni Putu Diah Kartini	2016/2017	TI-KAB	1	3.56	4	3.89	3	3.61

Gambar 4. Format Data Uji

Terdapat dua hasil dari proses *Extract Transform and Load* (ETL) yaitu mendapatkan dataset dan data uji yang siap digunakan pada perancangan sistem. Pada dataset tersebut berupa data akademik mahasiswa yang merupakan data mentah, dimana perlu dilakukan proses integrasi dan proses transformasi terlebih dahulu. Data tersebut terdiri dari enam atribut yaitu: A1: status kerja, A2: IPS 1, A3: IPS 2, A4: IPS 3, A5: IPS 4, A6: IPK. Dari 800 data kelulusan mahasiswa, jumlah data kelulusan mahasiswa yang tepat waktu berjumlah 567 dan 233 data kelulusan mahasiswa yang tidak tepat waktu. Adapun keterangan data atribut atau fitur yang akan digunakan dapat dilihat pada Tabel 1.

Sedangkan untuk data uji diperoleh data mahasiswa angkatan tahun 2016 sebesar 250 data mahasiswa dengan rincian masa kelulusan mahasiswa yang tepat waktu berjumlah 207 mahasiswa dan mahasiswa yang kelulusannya tidak tepat waktu berjumlah 43 mahasiswa. Data tersebut yang akan digunakan sebagai data uji pada sistem. Adapun format data uji yang akan digunakan dapat dilihat pada Gambar 4.

Modelling Phase (Tahap Pemodelan)

Pada tahapan ini merupakan tahap analisis data latih yang akan diklasifikasikan oleh model algoritma dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan algoritma *Naïve Bayes* dan akan dihitung menggunakan *software text editor* Bahasa pemrograman *python* yaitu *jupyter notebook*. *Bayesian classification* merupakan pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas isi dari suatu *class*. *Bayesian classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *neural network* dan *decision tree*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan

yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [6].

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \tag{1}$$

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (*posteriori* probabilitas.)

P(H) = Probabilitas hipotesis H (*prior* probabilitas.)

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas dari X

Sebelum masuk langkah-langkah perhitungan, berikut adalah penjelasan mengenai *Naïve Bayes* [8]. Setiap data dipresentasikan sebagai vektor berdimensi-*n* yaitu $X=(X_1, X_2, X_3, \dots, X_n)$, *n* adalah gambaran dari ukuran yang dibuatkan *test* dari *n* atribut yaitu $A_1, A_2, A_3, \dots, A_n$. *m* adalah kumpulan kategori yaitu C_1, C_2, \dots, C_m . Diberikan data *test* X yang tidak diketahui kategorinya, maka *classifier* akan memprediksi bahwa X adalah milik kategori dengan *posterior probability* tertinggi berdasarkan kondisi X. Oleh karena itu, *Naïve Bayes* menandai bahwa *test* X yang tidak diketahui tadi ke kategori C_i jika dan hanya jika:

$P(C_i|X) > P(C_j|X)$ untuk $i \leq j \leq m, j \neq i$. Kemudian $P(C_i|X)$ perlu dimaksimalkan.

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)} \tag{2}$$

Untuk mengurangi perhitungan dalam evaluasi $P(X|C_i)$, asumsi *naïf* independensi kelas bersyarat perlu dibuat, Hal ini menganggap nilai atribut saling independen dengan yang lainnya.

$$P(X_1 | C_i) \times P(X_2 | C_i) \times \dots \times P(X_n | C_i) \quad (3)$$

Sehingga lebih mudah memperkirakan $P(X_1 | C_i)$, $P(X_2 | C_i)$, ..., $P(X_n | C_i)$ dari data training. Dimana X_k mengacu pada nilai atribut A_k untuk data X . Untuk setiap atribut yang akan dihitung harus diperhatikan terlebih dahulu, apakah termasuk atribut kategorial atau atribut numerik. Oleh karenanya, untuk menghitung $P(X | C_i)$ perlu memperhatikan hal berikut:

Untuk memprediksi kelas dari X , $P(X | C_i)P(C_i)$ dievaluasi untuk setiap kelas C_i . *Naïve Bayes* akan memprediksi bahwa kelas dari X adalah C_i jika dan hanya jika $P(X | C_i)P(C_i) > P(X | C_k)P(C_k)$, di mana $1 \leq j \leq m$, $m \neq j$. Dengan kata lain dari prediksi kelas termasuk kelas C_i jika $P(X | C_i)P(C_i)$ bernilai maksimal.

$$P(X | C_i)P(C_i) \quad (4)$$

Bayesian classification diterapkan untuk melakukan beberapa proses yang telah dijelaskan diatas, *output*

yang akan dihasilkan sistem berupa nilai peluang bersyarat, nilai peluang *prior*, dan tingkat akurasi klasifikasi. Sedangkan yang kedua adalah data hasil prediksi yaitu tingkat kelulusan mahasiswa (Tepat waktu/Tidak tepat waktu). Nilai tingkat kelulusan yang digunakan adalah yang terbesar, yaitu Tepat waktu didapatkan jika $P(\text{Tepat waktu} | X) > P(\text{Tidak tepat waktu} | X)$ dimana X adalah atribut yang sudah diketahui. Begitupun sebaliknya, Tidak tepat waktu didapatkan jika $P(\text{Tepat waktu} | X) < P(\text{Tidak tepat waktu} | X)$.

Pada penelitian ini untuk menerapkan algoritma *Naïve Bayes* pada sistem pendukung keputusan berbasis website maka penulis menggunakan *jupyter notebook* sebagai contoh pembuatan model algoritma *Naïve Bayes* dengan menggunakan *library* GaussianNB pada *python*. Adapun *script* pembuatan model algoritma *Naïve Bayes* menggunakan *library* GaussianNB, yang ditunjukkan pada Gambar 5.

```
import numpy as np
import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix
import pickle
import sklearn

dataset=pd.read_csv('dataset.csv')
x = dataset.iloc[:, [4,5,6,7,8,9]].values
y = dataset.iloc[:, 10].values

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=0)

NBmodel = GaussianNB()
NBmodel.fit(x_train, y_train)

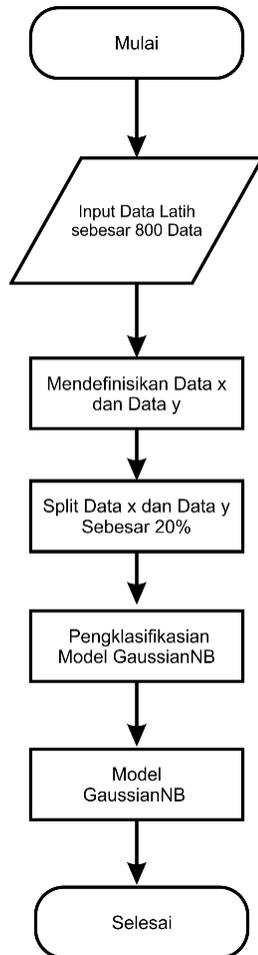
y_pred=NBmodel.predict(x_test)

accuracy_score(y_test,y_pred)*100
87.5
```

Gambar 5. Script Model Algoritma *Naïve Bayes*

Pada Gambar 5 merupakan langkah-langkah pembuatan model algoritma *Naïve Bayes* menggunakan *library* GaussianNB, dimana model tersebut akan digunakan pada proses pengklasifikasian data uji. Diawali dengan meng-*import* segala *library* dan *package* yang dibutuhkan dan dilanjutkan dengan membuat *variable* dataset lalu memanggil *file* dataset.csv yaitu dataset mahasiswa tahun 2013 dan 2014 sebesar 800 data, mengambil data atribut pada kolom 4, 5, 6, 7, 8, 9 sebagai data x dan mengambil data atribut pada kolom 10 sebagai data y atau probabilitas. Selanjutnya dataset akan di bagi sebesar 20% untuk mendapatkan

data uji pada model dengan menggunakan *library* *train test split*, dan setelah itu membuat *variable* NBmodel untuk memanggil *library* GaussianNB maka, Langkah terakhir adalah mengeksekusi data uji yang diberi nama *variable* x_test dari hasil pembagian dataset dengan perintah $y_pred = NBmodel.predict(x_test)$. Dari hasil *library* *accuracy score* model tersebut memiliki akurasi sebesar 87.5 %, gambaran umum dari pembuatan model atau *flowchart* pembuatan model GaussianNB dapat dilihat pada Gambar 6.



Gambar 6. Flowchart Model GaussianNB

Pada tahap pemodelan ini juga merancang bagaimana alur data dari sistem pendukung keputusan. Pada penelitian ini menggunakan diagram alur atau *Data Flow Diagram* (DFD), DFD pada sistem ini berupa konteks diagram dan DFD *level 1* Konteks diagram pada sistem menggambarkan aliran data dari *admin prodi* menuju sistem, dimana *admin* memasukkan data uji berupa status kerja, IPS 1, IPS 2, IPS 3, IPS 4, dan IPK. Setelah itu *output* pada sistem akan memberikan hasil klasifikasi sistem kepada *admin*. Konteks diagram dapat dilihat pada Gambar 7.



Gambar 7. Konteks Diagram

Agar lebih memahami seluruh proses dan cara kerja pada sistem dapat menyimak Pada *flowchart* pengklasifikasian data uji merupakan gambaran umum bagaimana alur sistem pada pengklasifikasian data uji dengan model GaussianNB yang telah dibuat sebelumnya. Pada *flowchart* pengklasifikasian data uji dimulai dengan memasukkan data uji sebesar 250 data mahasiswa tahun akademik 2016 yang diperoleh dari proses pengumpulan data penelitian. Selanjutnya mendefinisikan data atribut yang akan digunakan sebagai data x, setelah itu langkah selanjutnya adalah mengeksekusi data tersebut menggunakan model GaussianNB. *flowchart* tersebut dapat dilihat pada Gambar 8.

Untuk menyimpan semua data pada sistem diperlukan sebuah *database* agar data yang tersimpan dapat dikelola sehingga sistem dapat berjalan dengan baik. *Database* pada sistem terdiri dari *tb_user* sebagai tabel *user/admin*, jurusan sebagai tabel jurusan mahasiswa, kerja sebagai tabel status pekerjaan mahasiswa, *prob_data* sebagai tabel masa kelulusan mahasiswa, *data_uji* sebagai tabel data uji dan *hasil_spk* sebagai tabel penyimpanan dari hasil penghitungan dan pengklasifikasian sistem. Rancangan *database* pada sistem dapat dilihat pada Gambar 10.

Evaluation Phase (Tahap Evaluasi)

Pada tahapan ini dilakukan proses pengklasifikasian sesuai dengan yang telah dirancang, digambarkan dengan *flowchart* dan menggunakan *package* GaussianNB untuk mendapatkan hasil perhitungan yang akurat maka, selanjutnya adalah mengimplementasikannya kedalam sistem yang berbasis *web* dan melalui dua tahap yaitu tahap implementasi pada sistem dan tahap implementasi *user interface*.

Pada tahap ini akan dilakukan tiga metode pengujian pada sistem mulai dari uji sistem, uji *black box*, dan uji akurasi. Sistem akan di uji dengan memberikan sebuah data *input* yang berupa 250 data mahasiswa pada tahun akademik 2016. *Output* yang dihasilkan sistem akan menjadi bahan untuk dibandingkan dengan masa kelulusan pada data yang asli dan pada tahap ini akan dilakukan uji akurasi untuk mengetahui seberapa baik akurasi yang dihasilkan dari algoritma atau model yang digunakan pada sistem didalam menentukan masa kelulusan mahasiswa berdasarkan hasil perhitungan *confusion matrix* pada tahap pemodelan sebelumnya. Dimana 800 data yang terdapat pada dataset di bagi menggunakan *library train test split* dengan ukuran sebesar 0.20 atau 20% sebagai data uji dan menghasilkan akurasi sebesar 87,5% pada model.

Deployment Phase (Tahap Penyebaran)

Pada tahap ini, menjelaskan kesimpulan dari hasil perancangan sistem dan saran penulis terhadap sistem yang dirancang serta saran untuk penelitian selanjutnya apabila ingin mengembangkan sistem pendukung keputusan menentukan masa kelulusan mahasiswa.

HASIL DAN PEMBAHASAN

Implementasi Sistem

Implementasi sistem pada sistem pendukung keputusan ini membutuhkan *tools* pendukung dengan rincian seperti: *apache web server*, *python programming*, *flask web framework*, *visual studio code*, dan *MySQL*. *Script* program pada sistem dibuat menggunakan *text editor visual studio code*, hal pertama yang wajib dilakukan adalah memanggil *library*, *package* dan membuat *route database* yang digunakan pada sistem. *Script* tersebut dapat dilihat pada Gambar 11.

```
import importlib
from logging import exception
import mimetypes
from urllib import response
from flask import Flask, render_template,
redirect, request, url_for, session, flash,
Response
from flask_mysql import MySQL, MySQLdb
from fpdf import FPDF
from requests import Response
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
from sklearn.model_selection import
train_test_split
from sklearn.metrics import accuracy_score,
confusion_matrix
import mysql.connector as sql_db
import json
import pandas as pd
import numpy as np
import sklearn
import pickle
import os
import csv
import bcrypt
import werkzeug
import mimetypes
import io
import xlwt

app = Flask(__name__)
app.config['MYSQL_HOST'] = 'localhost'
app.config['MYSQL_USER'] = 'root'
app.config['MYSQL_PASSWORD'] = ''
app.config['MYSQL_DB'] = 'spk'
app.config["SECRET_KEY"] = "mysecretkey"
mysql = MySQL()
mysql.init_app(app)
```

Gambar 11. *Script Library* Program

Berdasarkan *script* program pada gambar 11 terdapat *library flask* yang digunakan sebagai *framework website* pada sistem dan memanggil *library* GaussianNB yang digunakan sebagai algoritma *Naive Bayes*. Pada Langkah selanjutnya adalah membuat *variable* yang digunakan serta memanggil *library* yang digunakan didalam proses klasifikasi pada sistem. *Script* tersebut dapat dilihat pada Gambar 12.

```
@app.route("/prosedataujifinal", methods=["POST","GET"])
def prosedataujifinal():
    cur =mysql.connect
    dataset=pd.read_csv('dataset.csv')
    x = dataset.iloc[:, [4,5,6,7,8,9]].values
    y = dataset.iloc[:, 10].values
    x_train, x_test, y_train,
    y_test = train_test_split(x, y, test_size=0.20,
    random_state=0)
    NBmodel = GaussianNB()
    NBmodel.fit(x_train, y_train)
    y_pred=NBmodel.predict(x_test)
    print(y_pred)
```

Gambar 12. Script Program GaussianNB

Berdasarkan *script* program pada gambar 10 merupakan langkah menginisialisasi data yang digunakan sebagai data x dan y pada dataset, dan menggunakan algoritma *Naive Bayes* yaitu *library* GaussianNB yang merupakan *package* didalam Bahasa pemrograman pyhton. Data latih yang digunakan adalah data latih yang sudah dibahas pada sub-bab tahap persiapan data, sedangkan untuk data uji merupakan data yang diambil dari tabel data_uji3 pada *database* yang telah melewati proses *submit* data uji. selanjutnya hasil akan disimpan kedalam tabel *prob_data*. *Script* program dapat disimak pada Gambar 13.

Pada Langkah selanjutnya untuk mendapatkan nilai akurasi dari model yang dibuat dan digunakan untuk mengklasifikasi data uji. Jumlah data yang di uji tingkat akurasinya pada sistem ini berjumlah 20% dari dataset. Proses pengujian tingkat akurasi menggunakan *package accuracy score* dan *metrics confusion matrix*. *Script* uji akurasi dapat dilihat pada Gambar 14.

```
akurasi=pd.read_sql_query('SELECT*FROM data_uji3',cur )
x1 = akurasi.iloc[:, [5,6,7,8,9,10]].values
y_pred1 = NBmodel.predict(x1)
pt = y_pred1.reshape(250,1)
if request.method == "POST":
    curl =mysql.connection.cursor()
for i in pt:
    curl.execute("INSERT INTO prob_data(kd_prob) VALUES(%s)",i)
mysql.connection.commit()
```

Gambar 13. Script Dan Query Simpan Hasil Program

```
n_errors1 = (y_test != y_pred).sum()
print('XG boost no of Errors:{}'.format(n_errors1))
er = n_errors1
av = accuracy_score(y_test, y_pred)*100
cv = metrics.confusion_matrix(y_test, y_pred)
pt = y_pred1
print('matrix :{}'.format(cv))
curl.execute("SELECT COUNT(*) FROM data_uji3")
cn = curl.fetchone()[0]
```

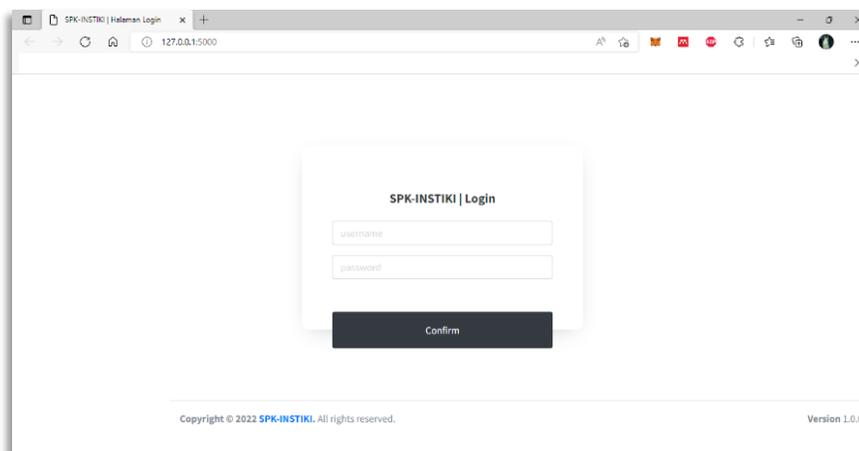
Gambar 14. Script Uji Akurasi

Implementasi User Interface

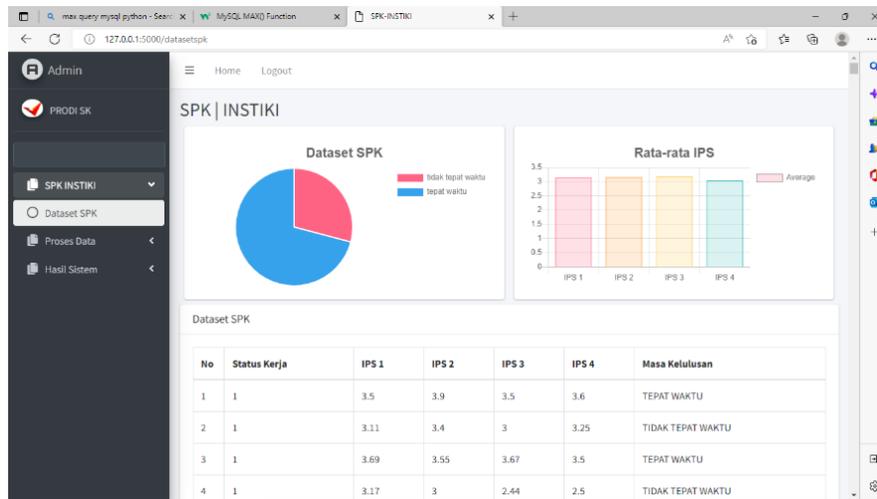
Implementasi user interface pada sistem pendukung keputusan ini membutuhkan template yang tersedia dari berbagai sumber seperti, html, css, javascript yang terdapat pada bootstrap. Tampilan user interface dan aplikasi sebagai berikut.

Halaman Login

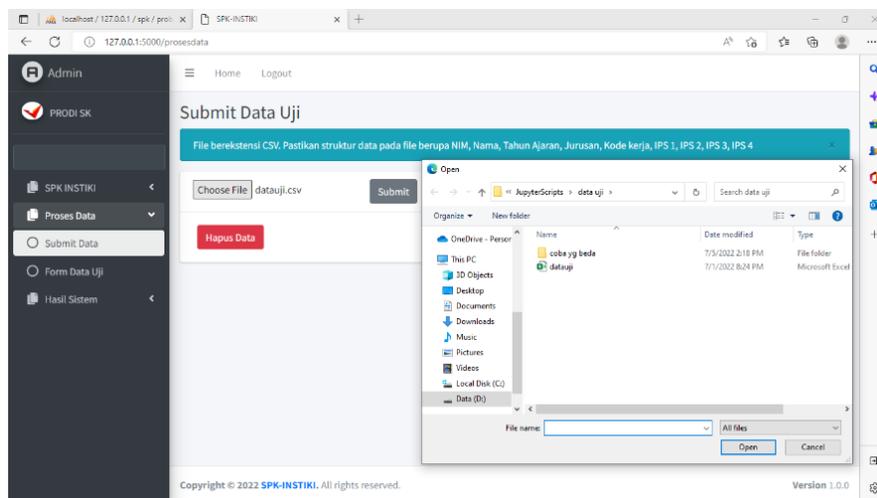
Tampilan halaman login pada website dapat dilihat pada Gambar 15.



Gambar 15. Halaman Login



Gambar 16. Halaman Dataset SPK



Gambar 17. Halaman Submit Data Uji

Halaman Dataset SPK

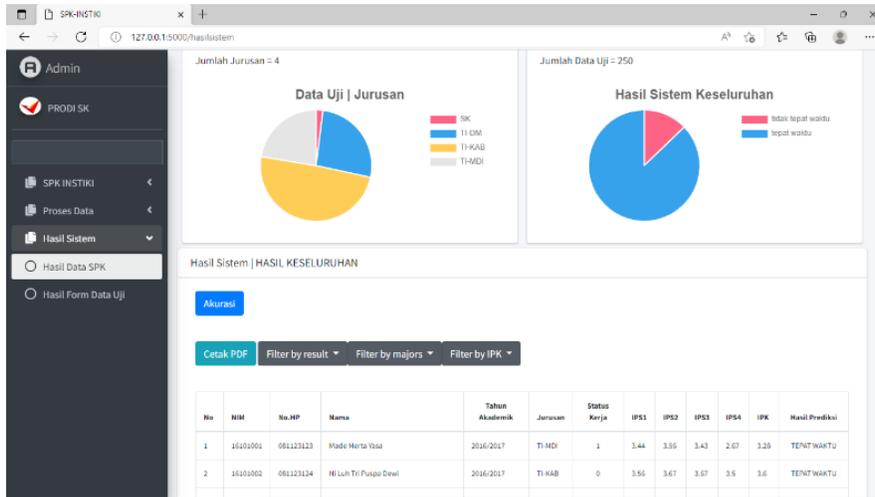
Pada halaman ini menampilkan dataset yang digunakan pada sistem. Halaman dataset SPK dapat dilihat pada Gambar 16. Pada halaman dataset SPK menampilkan grafik *pie chart* dan *bar chart* untuk memudahkan pengguna melihat dataset yang digunakan pada sistem.

Halaman *Submit Data Uji*

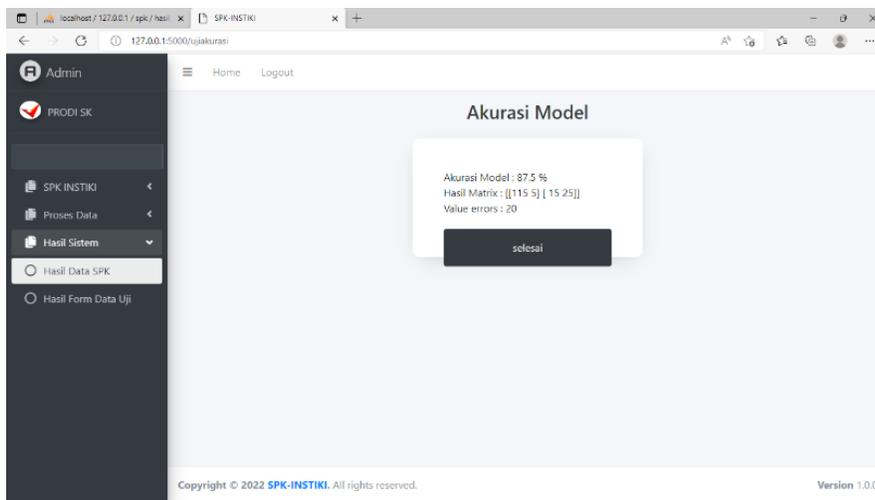
Pada halaman ini merupakan opsi untuk mengunggah *file* yang akan diuji pada sistem. Halaman *submit* data uji dapat dilihat pada Gambar 17. Pada

halaman submit data uji pengguna dapat memilih file berekstensi Comma Separated Values (CSV) sebagai data uji dengan format data uji yang telah dijelaskan pada sub-bab persiapan data. Setelah pengguna memilih file data uji dan menekan tombol upload file, maka akan dialihkan ke halaman proses data pada Gambar 18 untuk mengeksekusi file yang telah diunggah sebelumnya.

Pada halaman proses data uji terdapat tombol proses data untuk memulai pengklasifikasian data uji, apabila proses klasifikasi selesai maka pengguna akan diarahkan ke halama hasil data SPK.



Gambar 19. Halaman Hasil Data SPK



Gambar 20. Output Tombol Akurasi

Halaman Hasil Data SPK

Pada halaman ini menampilkan hasil dari pengklasifikasian pada halaman proses data uji sebelumnya. Halaman hasil data SPK dapat dilihat pada Gambar 19. Pada halaman hasil data SPK menampilkan visualisasi berupa grafik berjenis *pie chart* agar memudahkan pengguna untuk melihat hasil dari sistem dan menampilkan tabel hasil dari pengklasifikasian pada halaman proses data uji. Pada halaman hasil data SPK terdapat lima tombol yaitu tombol akurasi untuk melihat nilai akurasi yang dihasilkan oleh model pada sistem, tombol cetak PDF

untuk mencetak hasil dari pengklasifikasian data uji menjadi *file* berekstensi PDF, tombol *filter by result* yaitu menyaring hasil berdasarkan masa kelulusan, tombol *filter by majors* yaitu menyaring hasil berdasarkan jurusan, dan tombol *filter by IPK* yaitu menyaring hasil berdasarkan nilai IPK. *Output* dari tombol akurasi dan cetak PDF, seperti Gambar 20.

Pada halaman hasil akurasi model pada gambar 20 diatas menampilkan akurasi pada model algoritma Naïve Bayes sebesar 87,5 %. Hasil tersebut sesuai dengan pembahasan pada sub-bab tahap pemodelan.

No	NIM	Nama	No.HP	Jurusan	Status	IPS1	IPS2	IPS3	IPS4	IPK	Masa Kelulusan
1	16101001	Made Merta Yasa	081123123	Ti-MDI	1	3,44	3,56	3,43	2,67	3,28	Tepat Waktu
2	16101002	Mu Luth Fa Puqqa Dewi	081123124	Ti-KAB	0	3,36	3,67	3,67	3,5	3,6	Tepat Waktu
3	16101003	Fikri Nugrah Ambartha Putra	081123125	Ti-DMI	0	3,44	3,44	3,52	3,5	3,5	Tepat Waktu
4	16101004	Syafa Rizka	081123126	Ti-KAB	1	3,89	3,67	3,86	3,75	3,79	Tepat Waktu
5	16101007	Denia Made Endra Junarta	081123127	Ti-KAB	1	3,39	3,44	3,22	3,5	3,37	Tepat Waktu
6	16101008	Lisa Cahyani Putri	081123128	Ti-KAB	0	3,6	3,39	3,56	3,5	3,5	Tepat Waktu
7	16101010	Kadek Rizka Permadi	081123129	Ti-KAB	0	3,44	3,22	3,76	3,8	3,36	Tepat Waktu
8	16101012	Mu Wajidi Karla Dini	081123130	Ti-MDI	1	3,38	3,78	3,86	3,33	3,49	Tepat Waktu

Gambar 21. Output Tombol Cetak PDF

Output yang dihasilkan pada tombol cetak PDF pada gambar 21 diatas berupa data atribut mahasiswa yang terdiri dari NIM, nama lengkap mahasiswa, no HP mahasiswa, jurusan, nilai IPS 1 sampai IPS 4, nilai IPK, dan hasil klasifikasi pada sistem berupa masa kelulusan mahasiswa tepat waktu dan tidak tepat waktu.

Pengujian Sistem

Data uji yang digunakan pada penelitian ini sebanyak 250 data, data tersebut merupakan data mahasiswa tahun akademik 2016. Pengklasifikasian data uji menggunakan model *Naive Bayes* berupa model GaussianNB yang telah dibahas pada sub-bab tahap pemodelan. Pengklasifikasian data uji pada sistem dilakukan dengan memasukkan file data uji berekstensi CSV kedalam sistem pada halaman *submit* data uji. Adapun hasil dari pengklasifikasian data uji pada sistem, dapat dilihat pada Tabel 2.

Tabel 2. Hasil Klasifikasi Data Uji

Jumlah Data Uji	Hasil Sistem	
	Tepat Waktu	Tidak Tepat Waktu
250	218	32

Berdasarkan proses klasifikasi 250 data uji pada sistem menghasilkan masa kelulusan mahasiswa yang tepat waktu berjumlah 218 mahasiswa dan kelulusan mahasiswa yang tidak tepat waktu berjumlah 32 mahasiswa. Pengujian sistem dilakukan tanpa ada kendala ataupun terdapat *error*, maka dapat disimpulkan pada metode uji sistem pada penelitian ini berhasil.

Hasil Uji Akurasi

Pada tahap ini dilakukan proses komparasi atau perbandingan antara hasil klasifikasi data uji pada sistem dengan data asli yang telah dibahas pada sub-bab tahap persiapan data untuk mendapatkan nilai akurasi pada sistem terkait pengklasifikasian data uji. Adapun perbandingan hasil yang diperoleh pada sistem dapat dilihat pada tabel 3 berikut.

Tabel 3. Perbandingan Hasil Sistem dengan Data Asli

Jumlah Data Uji	Kesesuaian Hasil	
	Sesuai	Tidak Sesuai
250	202	48

Perbandingan hasil klasifikasi data uji oleh sistem dengan data asli yang tertera pada tabel 3 didapatkan 48 data yang diklasifikasikan oleh sistem dengan hasil yang tidak sesuai dengan data asli, sedangkan data yang sesuai diklasifikasikan oleh sistem berjumlah 202 data. Dengan demikian dapat dihitung secara manual untuk mendapatkan nilai akurasi dari sistem dengan perhitungan hasil akurasi sebagai berikut.

$$\begin{aligned}
 \text{Akurasi Sistem} &= \frac{\text{Jumlah data sesuai}}{\text{Jumlah data uji}} \times 100 \% \quad (5) \\
 &= \frac{202}{250} \times 100 \% \\
 &= 80,8 \%.
 \end{aligned}$$

KESIMPULAN

Berdasarkan hasil evaluasi sistem yang sudah dilakukan pada penelitian ini, dapat diambil kesimpulan yaitu, pada penelitian ini, sistem berhasil menentukan masa kelulusan mahasiswa yang tepat

waktu dan yang tidak tepat waktu. Nilai akurasi yang dihasilkan oleh model GaussianNB sebesar 87,5% dan nilai akurasi yang dihasilkan pada pengklasifikasian data uji dengan menggunakan model pada sistem sebesar 80,8%. Sistem pendukung keputusan pada penelitian ini berbasis website yang menggunakan *python flask* sebagai *web framework* dan dipadukan dengan Bahasa pemrograman *web* lainnya seperti *javascript*, HTML, dan CSS. Untuk merancang bangun sistem pendukung keputusan menentukan masa kelulusan menggunakan algoritma *Naïve Bayes* diperlukan model pengklasifikasian dengan tahap pendefinisian variabel *x* dan *y* pada dataset, dan pada sistem ini menggunakan *library* GaussianNB pada *python* sebagai algoritma *Naïve Bayes*.

Adapun saran didalam penelitian ini untuk penelitian selanjutnya adalah pada klasifikasi kelulusan mahasiswa ini diharapkan data atau fitur yang terdapat pada sistem dapat ditambah, memperkaya variabel dan memiliki subyektifitas yang rendah sehingga hasil pengklasifikasian memiliki nilai akurasi yang lebih tinggi.

DAFTAR PUSTAKA

- [1] Reni, Widodo, and B. P. Adhi, "Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Pada Program Studi Pendidikan Teknik Informatika Dan Komputer Dengan Menggunakan Algoritma Naïve Bayes," *PINTER: Jurnal Pendidikan Teknik Informatika dan Komputer*, vol. 4, no. 1, pp. 26-29, 2020.
- [2] I. G. T. Isa, "Aplikasi Prediksi Kelulusan Tepat Waktu dengan Algoritma Naive Bayes (Studi Kasus Program Studi Teknik Informatika UMMI)," *Jurnal Santika*, vol. 7, no. 2, pp. 591-599, 2017.
- [3] G. Testiana, "Perancangan model prediksi kelulusan mahasiswa tepat waktu pada UIN Raden Fatah," *JUSIFO (Jurnal Sistem Informasi)*, vol. 4, no. 1, pp. 49-62, 2018.
- [4] H. Hartatik, "Optimasi Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes," *Indonesian Journal of Applied Informatics*, vol. 5, no. 1, p. 32, Apr. 2021.
- [5] H. Kurniawan and J. Sistem Informasi, "Aplikasi Datamining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori Di Ibi Darmajaya Bandar Lampung," vol. 02, 2016.
- [6] P. S. Dewi, C. K. Sastradiprja, and D. Gustian, "Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier," *Jurnal Teknologi dan Informasi*, vol. 11, no. 1, pp. 66-80, 2021.
- [7] "Basis Data - Google Books." https://www.google.co.id/books/edition/Basis_Data/nDRrEAAAQBAJ?hl=en&gbpv=0 (accessed Aug. 05, 2022).
- [8] P. U. Airlangga, S. Rancang, and B. Sistem, "Adln - Perpustakaan Universitas Airlangga," 2016.