PENERAPAN DATA MINING DAN ALGORITMA NAÏVE BAYES UNTUK PEMILIHAN KONSENTRASI MAHASISWA MENGGUNAKAN METODE KLASIFIKASI

Muhammad Farid Satrio Wibowo [1)] , Nila Feby Puspitasari [2)], Barka Satya [3)]

[1)] Informatika Universitas AMIKOM Yogyakarta
[2),3)] Teknik Informatika Universitas AMIKOM Yogyakarta
email : muhammad.8887@students.amikom.ac.id[1)], nilafeby@amikom.ac.id[2)] , barka.satya@amikom.ac.id[3)]

Abstraksi

Pemilihan konsentrasi atau minat studi merupakan hal yang tidak mudah dilakukan oleh seorang mahasiswa pada sebuah jurusan di Perguruan Tinggi. Mahasiswa akan berupaya memilih konsentrasi yang menurut mereka paling tepat dan sesuai dengan kompetensi dan minat studi, karena konsentrasi yang dipilih akan mempengaruhi minat belajar, prestasi, lama studi dan juga berpengaruh terhadap Indeks Prestasi Akademik (IPK) mahasiswa. Pentingnya memilih sebuah konsentrasi penjurusan bagi mahasiswa pada Institusi Perguruan Tinggi, maka perlu dibangun suatu model yang dapat membantu mahasiswa dalam memilih konsentrasi sesuai dengan kompetensi dan minat studi mahasiswa. Oleh karena itu, peneliti akan melakukan penelitian dengan membuat sistem untuk pemilihan konsentrasi mahasiswa menggunakan algoritma Naïve Bayes dengan metode klasifikasi. Untuk membantu dalam mengambil keputusan pemilihan konsentrasi, penelitian ini menggunakan teknik *data mining* sebagai proses pencarian pola yang diinginkan dalam sebuah *database* yang besar. Hasil pengujian yang telah dilakukan terhadap sample dataset sebanyak 1534 data menggunakan Algoritma Naïve Bayes, diperoleh bahwa hasil prediksi untuk menentukan konsentrasi memiliki nilai akurasi sebesar 84.27%. Variabel berpengaruh terhadap tingkat akurasi yang di hasilkan. Ukuran variabel yang sempit atau sedikit menyebabkan hasil akurasi yang kurang baik, tetapi ukuran variabel yang luas dapat menghasilkan akurasi ouput yang lebih optimal.

***Kata Kunci* :**
Data Mining, Algoritma Naive Bayes, Klasifikasi, Konsentrasi

Abstract

The choice of concentration or interest in study is not an easy thing for a student to do in a major at a university. Students will try to choose the concentration that they think is most appropriate and in accordance with their competence and interest in study, because the concentration chosen will affect interest in learning, achievement, length of study and also affect the student's Academic Achievement Index (GPA). The importance of choosing a concentration of majors for students at Higher Education Institutions, it is necessary to build a model that can assist students in choosing a concentration according to the competence and interest of student studies. Therefore, researchers will conduct research by creating a system for selecting student concentrations using the Naïve Bayes algorithm with the classification method. To assist in making concentration decisions, this study uses data mining techniques as the process of finding the desired pattern in a large database. The results of tests that have been carried out on a sample dataset of 1534 data using the Naïve Bayes Algorithm, it is found that the prediction results for determining the concentration have an accuracy value of 84.27%. Variables affect the level of accuracy that is generated. A narrow or little variable size causes poor accuracy results, but a wide variable size can produce more optimal output accuracy.

***Keyword* :**
Data Mining, Algorithm Naive Bayes, Classification, Concentration

Pendahuluan

Pendidikan tinggi merupakan salah satu institusi pendidikan yang memberikan jasa pelayanan pendidikan kepada masyarakat, sehingga peranan pendidikan sangatlah penting, dikarenakan pendidikan tinggi sangat menentukan kemampuan bangsa Indonesia untuk mencapai kemajuan dan meningkatkan kesejahteraan bangsa. Kredibilitas sebuah institusi perguruan tinggi dipengaruhi oleh beberapa faktor antara lain adalah kualitas mahasiswa. Untuk meningkatkan kualitas institusi pendidikan harus diselaraskan dengan perkembangan teknologi yang ada saat ini, sehingga perkembangan pendidikan yang ada pada institusi tersebut semakin optimal dalam mencapai tujuannya [1].

Kualitas institusi dan ketersediaan infrastruktur berkaitan dengan kualitas mahasiswa. Memilih konsentrasi penjurusan yang sesuai dapat berpengaruh terhadap motivasi mahasiswa dalam belajar untuk mendapatkan nilai akademik yang optimal. Pemilihan konsentrasi jurusan di Perguruan

Tinggi merupakan hal yang tidak mudah bagi seorang mahasiswa. Mahasiswa akan melakukan pemilihan konsentrasi yang sesuai dengan minat dan bidang keilmuannya, karena ketepatan dalam pemilihan konsentrasi akan berpengaruh terhadap peluang pekerjaan yang diminati oleh mahasiswa ketika telah menyelesaikan studinya. Dalam proses pemilihan konsentrasi, mahasiswa dituntut untuk dapat memberikan penilaian dan dapat mengetahui kompetensi yang dimilikinya dalam bidang akademik. Ketepatan pemilihan konsentrasi yang diambil akan berpengaruh terhadap minat belajar sesuai dengan bidang ilmunya dan keaktifan serta ketekunan mahasiswa dalam mengerjakan tugas dan ujian selama mengikuti kegiatan perkuliahan sehingga menghasilkan Indeks Prestasi Akademik (IPK) yang optimal [2].

Pentingnya mahasiswa dalam pemilihan konsentrasi penjurusan pada sebuah perguruan tinggi, maka perlu dibangun sebuah model dapat membantu mengatasi problematika mahasiswa dalam memilih konsentrasi penjurusan yaitu berupa sistem untuk menentukan konsentrasi mahasiswa dengan studi kasus jurusan informatika Universitas Amikom Yogyakarta [2].

Peneliti akan melakukan penelitian dengan membuat sistem untuk pemilihan konsentrasi mahasiswa menggunakan *data mining* dan algoritma Naïve Bayes dengan metode klasifikasi. Adapun teknik yang digunakan untuk mendapatkan informasi aktual dengan cara mencari pola atau aturan tertentu dari sejumlah data yang sangat besar yang disebut dengan istilah *data mining*. *Data mining* merupakan serangkaian proses untuk mendapatkan nilai tambah berupa pengetahuan baru terhadap suatu kumpulan data [3], sedangkan algoritma Naïve Bayes merupakan algoritma yang cocok digunakan dalam penerapan *Data Mining*.

Tinjauan Pustaka

Salah satu hasil penelitian tentang Penerapan Data Mining untuk mengevaluasi kinerja akademik mahasiswa menggunakan algoritma Naive Bayes *Classifier* adalah pembuatan sistem menggunakan teknik data mining dengan cara mengklasifikasi mahasiswa STMIK Dipanegara Makassar berdasarkan variabel lulus dan tidak lulus tepat waktu [4]. Hasil pengujian yang telah dilakukan menunjukkan bahwa semakin banyak *data training* yang digunakan. maka tingkat *recall, presisi* dan akurasi sistem akan semakin optimal. Selain itu dilakukan pengujian terhadap akurasi. bahwa ada berbagai macam faktor yang berpengaruh terhadap tingkat kelulusan mahasiswa STMIK Dipanegara Makassar, antara lain bukan hanya dari faktor akdemik saja. tetapi faktor non-akademik juga mempengaruhi.

Hasil penelitian [5] tentang "Penerapan Data Mining Menggunakan Algoritma Naive Bayes Classifier dan C4.5 untuk memprediksi Kelulusan Mahasiswa". Hasil penelitian menunjukkan bahwa hasil prediksi terhadap tingkat kelulusan mahasiswa pada STMIK Bina Nusantara Jaya Lubuklinggau berdasarkan dataset menunjukkan nilai akurasi sebesar 78,46% dan hasil prediksi menggunakan Algoritma C4.5 diperoleh nilai akurasi yang lebih besar yaitu 79,08%. Dari hasil prediksi kedua algoritma tersebut, maka Algoritma C4.5 memiliki nilai akurasi yang lebih baik dibandingkan nilai akurasi metode Naive Bayes *Classifier*, oleh karena itu Algoritma C4.5 dapat direkomendasikan sebagai algoritma yang layak digunakan dalam menyelesaikan masalah prediksi kelulusan mahasiswa pada STMIK Bina Nusantara Jaya Lubuklinggau.

Hasil penelitian [6] tentang bagaimana memprediksi Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes (Studi Kasus 5 PTS di Banda Aceh)". Adapun atribut data yang digunakan untuk proses klasifikasi adalah Indeks Prestasi Akademik (IPK) dan masa studi. Hasil penelitian ini adalah menunjukkan bahwa prediksi mahasiswa yang kuliah di ASM Nusantara dan AMIK Indonesia memperoleh tingkat kelulusan sebesar 60%, sedangkan mahasiswa yang kuliah di STIES Banda Aceh dan Universitas Serambi Mekkah prediksi kelulusannya sebesar 52%. Mahasiswa yang kuliah di STIA Iskandar Thani, prediksi kelulusan hanya sebesar 48% dan tidak lulus tepat waktu 52%.

Hasil Penelitian [7] tentang "Pengujian performa Algoritma Naïve Bayes untuk memrediksi Masa studi mahasiswa. Data yang digunakan dalam penelitian ini berjumlah 300 data alumni yang diklasifikasi menggunakan Algoritma Naïve Bayes berdasarkan waktu kelulusan mahasiswa dengan model klasifikasi menghasilkan rata-rata nilai akurasi sebesar 68%, recall 65.3%, presisi 61.3%, dan f1-score sebesar 61% yang dihitung berdasarkan metode 10-*Fold Cross Validation*, dan *Confusion Matrix*. Pemilihan data training yang digunakan berpengaruh terhadap hasil pengujian, karena probabilitas yang dimiliki oleh model akan digunakan untuk menentukan kelas pada data testing, dan berpengaruh terhadap besar kecilnya nilai akurasi, *precision*, *recall*, dan *f1-score.*

Pendekatan alternatif yang dilakukan oleh peneliti adalah melakukan penelitian dengan membuat sistem untuk pemilihan konsentrasi mahasiswa menggunakan *Data Mining* dan Algoritma Naïve Bayes dengan metode *Classification*, *dimana Data Mining* merupakan suatu proses pencarian pola terhadap sejumlah data yang sangat besar sehingga menghasilkan tingkat akurasi yang sangat baik. Sedangkan algoritma Naïve Bayes merupakan algoritma yang cocok digunakan untuk *Data Mining*.

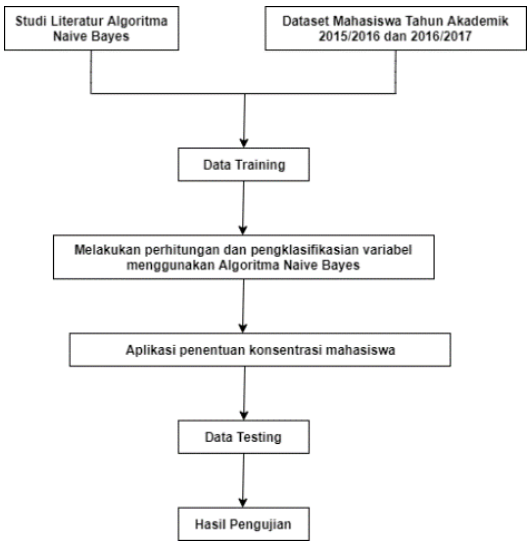
Metode Penelitian

1. Tahapan Penelitian

Ada beberapa tahapan yang digunakan dalam penelitian ini yaitu:

a. Mengumpulkan data yaitu studi literatur terkait dengan teori yang berhubungan secara mendalam tentang Algoritma *Naïve Bayes* dalam pengklasifikasian variabel yang sudah ditentukan, agar dapat membantu dalam pengembangan perangkat lunak.
b. Menganalisa sumber data. Data yang digunakan berupa *dataset* (sekumpulan data) mahasiswa Universitas Amikom Yogyakarta jurusan Informatika dengan Tahun Akademik 2015/2016 dan 2016/2017 dengan jumlah data sebanyak 1614 mahasiswa.
c. Membuat *data training / training-set* yaitu untuk meng-*training data* yang bertujuan untuk membuat model machine learning.
d. Melakukan perhitungan dengan menerapkan Algoritma *Naïve Bayes* untuk pemilihan konsentrasi mahasiswa.
e. Membuat aplikasi penentuan konsentrasi mahasiswa.
f. Membuat *data testing* yang akan digunakan untuk menguji performa dan kebenaran (terhadap korelasi) terhadap model yang bersangkutan.
g. Melakukan hasil pengujian terhadap aplikasi tersebut.

Adapun tahapan penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2. Analisa Pemodelan Data

Penelitian ini menggunakan dataset mahasiswa berupa NIM, usia, jenis kelamin, IPK selama 3 (tiga) semester dan mata kuliah yang relevan dengan konsentrasi. Data yang digunakan pada penelitian ini terdiri dari NIM, Usia, Jenis Kelamin, IPK dan nilai mata kuliah antara lain Algoritma dan Pemrograman, Struktur Data, Sistem Basis Data, Pemrograman, Pemrograman Lanjut, Komputer Grafis, Multimedia, Jaringan Komputer, Sistem Operasi, Aljabar Linear dan Matriks, Kalkulus, Komunikasi Data, Logika Informatika, Matematika Diskret, Organisasi dan Arsitektur Komputer, dan *Hardware* dan *Software*. Berikut merupakan data mentah yang belum dilakukan proses *data mining* atau biasa disebut dengan *preprocessing data.* Adapun sample data set mahasiswa dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Sample Data Set Mahasiswa

Npm	JK	Usia	IPK	Alprog	mrogram	SBD	Pem. Lanjut	Struktur	Komgraf	Mulmed	Jarkom	SO	Aljabar	Kalkulus	Komdat	LI	Matdis	Orakom	HS 2
15.11.8482	L	22	3.28	B	B	B	A	A	A	A	B	B	A	B	B	A	D	A	B
15.11.8483	L	25	3.44	B	A	B	B	A	A	A	B	B	A	B	B	A	A	A	A
15.11.8484	L	22	2.87	C	-	C	B	A	A	A	B	C	B	C	B	A	C	A	A
15.11.8485	L	23	3.08	B	C	C	B	A	A	A	A	C	C	C	B	A	B	B	A
15.11.8486	L	21	3.46	A	B	B	A	A	A	-	B	B	B	A	B	A	A	A	A
15.11.8487	L	22	3.51	A	C	B	B	A	A	B	A	B	A	A	B	A	B	A	A
15.11.8488	L	21	3	B	B	C	B	B	A	B	B	B	B	B	B	A	-	B	A
15.11.8489	L	22	3.49	B	A	B	-	A	A	A	B	B	A	A	B	A	C	A	A
15.11.8490	L	21	3.75	A	B	A	A	A	A	A	A	A	A	A	B	A	B	A	B
15.11.8491	L	21	3.89	A	B	A	A	A	A	A	A	A	A	B	A	A	A	A	A
15.11.8492	L	24	3.47	B	C	B	B	A	A	B	A	B	B	-	B	A	A	A	A
15.11.8493	L	22	3.16	B	B	B	A	A	A	B	B	A	B	B	A	B	A	A	A
15.11.8494	L	23	3.07	B	B	B	B	A	C	C	A	B	C	B	B	A	B	A	A
15.11.8495	L	21	3.65	-	B	A	B	A	A	B	A	A	A	A	B	A	B	A	A
15.11.8496	L	22	3.28	B	B	C	B	A	A	B	B	B	A	B	B	A	A	B	A
15.11.8497	L	23	3.43	A	C	B	B	A	B	-	A	A	A	A	B	A	A	A	A
15.11.8498	P	21	3.5	B	B	B	B	A	A	B	A	B	A	A	B	A	B	A	A
15.11.8499	L	22	3.15	B	B	B	B	B	A	A	A	B	A	B	C	A	B	A	A
15.11.8500	L	24	3.75	A	C	B	A	A	A	A	A	A	A	A	B	A	A	A	A

Kemudian peneliti melakukan *filtering* terhadap banyak data yang tidak valid dan tidak memenuhi kriteria untuk masuk kedalam proses diantaranya banyak mahasiswa yang nilai mata kuliahnya nol. Untuk itu peneliti harus menghapus semua NIM yang tidak mengambil mata kuliah dan juga menghapus NIM dengan usia yang tidak logis. Pada Tabel 2 dapat dilihat hasil filtering data mahasiswa.

Tabel 2. Hasil Filtering Data Mahasiswa

Npm	JK	Usia	IPK	Alprog	mrogram	SBD	Pem. Lanjut	Struktur Data	Komgraf	Mulmed	Jarkom	SO	Aljabar	Kalkulus	Komdat	LI	Matdis	Orakom	HS 2
15.11.8482	L	22	3.28	B	B	B	A	A	A	A	B	B	A	B	B	A	D	A	B
15.11.8483	L	25	3.44	B	A	B	B	A	A	A	B	B	A	B	B	A	A	A	A
15.11.8484	L	22	2.87	C	C	C	B	A	A	A	B	C	B	C	B	A	C	A	A
15.11.8485	L	23	3.08	B	C	C	B	A	A	A	A	C	C	C	B	A	B	B	A
15.11.8486	L	21	3.46	A	B	B	A	A	A	A	B	B	B	A	B	A	A	A	A
15.11.8487	L	22	3.51	A	C	B	B	A	A	B	A	B	A	A	B	A	B	A	A
15.11.8488	L	21	3	B	B	C	B	B	A	B	B	B	B	B	B	A	B	B	A
15.11.8489	L	22	3.49	B	A	B	B	A	A	A	B	B	A	A	B	A	C	A	A
15.11.8490	L	21	3.75	A	B	A	A	A	A	A	A	A	A	A	B	A	B	A	B
15.11.8491	L	21	3.89	A	B	A	A	A	A	A	A	A	A	A	B	A	A	A	A
15.11.8492	L	24	3.47	B	C	B	B	A	A	B	A	B	B	A	B	A	A	A	A
15.11.8493	L	22	3.16	B	B	B	A	A	A	B	B	A	B	B	A	B	A	A	A
15.11.8494	L	23	3.07	B	B	B	B	A	C	C	A	B	C	B	B	A	B	A	A
15.11.8495	L	21	3.65	A	B	A	B	A	A	B	A	A	A	A	B	A	B	A	A
15.11.8496	L	22	3.28	B	B	C	B	A	A	B	B	B	A	B	B	A	A	B	A
15.11.8497	L	23	3.43	A	C	B	B	A	B	C	A	A	A	A	B	A	A	A	A
15.11.8498	P	21	3.5	B	B	B	B	A	A	B	A	B	A	A	B	A	B	A	A
15.11.8499	L	22	3.15	B	B	B	B	B	A	A	A	B	A	B	C	A	B	A	A
15.11.8500	L	24	3.75	A	C	B	A	A	A	A	A	A	A	A	B	A	A	A	A

Setelah peneliti melakukan *filtering* data, untuk variabel mata kuliah Algoritma dan Pemrograman, Struktur Data, Sistem Basis Data, Pemrograman, Pemrograman Lanjut, Komputer Grafis, Multimedia, Jaringan Komputer, Sistem Operasi, Aljabar Linear dan Matriks, Kalkulus, Komunikasi Data, Logika Informatika, Matematika Diskret, Organisasi dan Arsitektur Komputer, dan Hardware dan Software II. Selanjutnya peneliti melakukan klasifikasi terhadap *variable* usia dan IPK. Berikut *variabel* atau *class* yang digunakan untuk melakukan prediksi mahasiswa diantaranya NIM, Usia, Jenis Kelamin, dan IPK semua mata kuliah tersebut. Setiap *class* memiliki variabel penentu yang digunakan dalam pengklasifikasian data mahasiswa, diantaranya :

1. Usia : 20-21 tahun, dan lebih dari sama dengan 22 tahun
2. Jenis Kelamin : Laki-laki dan Perempuan

3. IPK : lebih dari 3.25 dan kurang dari sama dengan 3.25
4. Klasifikasi : Pemrograman, Multimedia, dan Jaringan Komputer.

Adapun hasil klasifikasi data disajikan dalam bentuk data training yang dapat dilihat pada Tabel 3.

Tabel 3. Data Training

NIM	JK	USIA	IPK	KONSENTRASI
15.11.8482	Laki-laki	22+	+3.25	Pemrograman
15.11.8483	Laki-laki	22+	+3.25	Pemrograman
15.11.8484	Laki-laki	22+	3.25-	Multimedia
15.11.8485	Laki-laki	22+	3.25-	Multimedia
15.11.8486	Laki-laki	20-21 Tahun	+3.25	Pemrograman
15.11.8487	Laki-laki	22+	+3.25	Multimedia
15.11.8488	Laki-laki	20-21 Tahun	3.25-	Pemrograman
15.11.8489	Laki-laki	22+	+3.25	Multimedia
15.11.8490	Laki-laki	20-21 Tahun	+3.25	Jaringan Komputer
15.11.8491	Laki-laki	20-21 Tahun	+3.25	Multimedia
15.11.8492	Laki-laki	22+	+3.25	Pemrograman
15.11.8493	Laki-laki	22+	+3.25	Pemrograman
15.11.8494	Laki-laki	22+	3.25-	Jaringan Komputer
15.11.8495	Laki-laki	20-21 Tahun	+3.25	Pemrograman
15.11.8496	Laki-laki	22+	3.25-	Pemrograman
15.11.8497	Laki-laki	22+	+3.25	Pemrograman
15.11.8498	Perempuan	20-21 Tahun	+3.25	Pemrograman
15.11.8499	Laki-laki	22+	3.25-	Pemrograman
15.11.8500	Laki-laki	22+	+3.25	Jaringan Komputer
15.11.8501	Laki-laki	22+	+3.25	Multimedia

a. Analisa Pemodelan menggunakan Algoritma Naïve Bayes

Algoritma Naïve Bayes yaitu pengklasifikasian dengan menggunakan metode probabilitas dan statistik. Naive Bayes dapat dilatih dengan efisien dalam pembelajaran terawasi *(supervised learning)*. Karena variable independen diasumsikan, hanya variasi dari variabel untuk masing – masing kelas yang harus ditentukan. Naïve Bayes merupakan metode klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naive Bayes telah terbukti memiliki akurasi dan tingkat kecepatan yang tinggi saat diimplementasikan ke dalam database menggunakan data yang besar [4].

Probabilitas bayes adalah salah satu cara untuk mengatasi ketidakpastian dengan menggunakan formula bayes yang dapat dilihat pada Persamaan 1.

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

Keterangan :

X : Data berupa kelas yang belum diketahui
H : Hipotesa data X merupakan suatu kelas yang spesifik
P(H|X) : Probabilitas hipotesa H berdasar kondisi X
P(H) : Probabilitas hipotesa H
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesa H
P(X) : Probabilitas X

b. Perhitungan menggunakan ALgoritma Naïve Bayes

Berdasarkan Tabel 3, dapat dihitung menggunakan metode klasifikasi data mahasiswa apabila diberikan *input* jenis kelamin, usia, dan IPK menggunakan Algoritma *Naïve Bayes.* Adapun contoh sample ata sebagai berikut:

Langkah 1 :
Menentukan atribut X dan nilai kelas Y
Himpunan atribut X dan nilai kelas Y
Nilai kelas Y = Prediksi

Langkah 2 :
Menghitung probabilitas P1, P2, P3
P (Y | X) = P (Y | Usia >= 22 tahun , IPK <= 3,25 , JK = x)

Keterangan :
P1, P2, P3 = Probabilitas 1, 2, dan 3
P = Pemrograman
M = Multimedia
J = Jaringan Komputer
JK = Jenis Kelamin (y = perempuan | x = Laki-laki)

P1 = P (P) * P (Usia >= 22 tahun | P) * P1 (IPK <= 3,25 | P) * P (JK= x | P)
P2 = P (M) * P (Usia >= 22 tahun | M) * P (IPK <= 3,25 | M) * P (JK = x | M)
P3 = P (J) * P (Usia >= 22 tahun | J) * P (IPK <= 3,25+ | J) * P (JK = x | J)

Keterangan :
ΣP = jumlah dari seluruh data Pemrograman.
ΣJK = jumlah dari seluruh data Jenis Kelamin.
ΣM = jumlah dari seluruh data Multimedia.
ΣJ = jumlah dari seluruh data Jaringan Komputer.

Perhitungan manual yang tercantum dibawah ini merupakan perhitungan probabilitas terhadap contoh sample data.

1. P1 = Probabilitas Pemrograman

a. P1 (Pemrograman) = $\frac{\sum P}{\sum Seluruh\ Data} = \frac{11}{20}$

b. P1 (Usia >= 22 | P) = $\frac{\sum Usia \geq 22\ tahun \mid P}{\sum Data\ P} = \frac{7}{11}$

c. P1 (IPK <= 3,25 | P) = $\frac{\sum IPK \leq 3.25 \mid P}{\sum Data\ P} = \frac{3}{11}$

d. P1 (JK = L | P) = $\frac{\sum JK\ L \mid P}{\sum Data\ P} = \frac{10}{11}$

2. P2 = Probabilitas Multimedia

a. P2 (Multimedia) = $\frac{\sum M}{\sum \text{Seluruh Data}} = \frac{6}{20}$

b. P2 (Usia >= 22 | M) = $\frac{\sum \text{Usia} \geq 22 \text{ tahun} \mid M}{\sum \text{Data M}} = \frac{5}{6}$

c. P2 (IPK <= 3,25 | M) = $\frac{\sum \text{IPK} \leq 3.25 \mid M}{\sum \text{Data M}} = \frac{2}{6}$

d. P2 (JK = L | M) = $\frac{\sum \text{JK L} \mid M}{\sum \text{Data M}} = \frac{6}{6}$

3. P3= Probabilitas Jaringan

a. P3 (Jaringan Komputer) = $\frac{\sum J}{\sum \text{Seluruh Data}} = \frac{3}{20}$

b. P3 (Usia >= 22 | J) = $\frac{\sum \text{Usia} \geq 22 \text{ tahun} \mid J}{\sum \text{Data J}} = \frac{2}{3}$

c. P3 (IPK <= 3,25 | J) = $\frac{\sum \text{IPK} \leq 3.25 \mid J}{\sum \text{Data J}} = \frac{1}{3}$

d. P3 (JK = L | J) = $\frac{\sum \text{JK L} \mid J}{\sum \text{Data J}} = \frac{3}{3}$

Langkah 3 :
Bandingkan Probabilitas 1, 2 dan 3 dimana,

a. P1 = $\frac{11}{20} \times \frac{7}{11} \times \frac{3}{11} \times \frac{10}{11}$ = 0.550 x 0.636 x 0.272 x 0.909 = 0.086

Jadi nilai probabilitas pemrograman = 0.086

b. P2 = $\frac{6}{20} \times \frac{5}{6} \times \frac{2}{6} \times \frac{6}{6}$ = 0.3 x 0.844 x 0.333 x 1 = 0.084

Jadi nilai probabilitas multimedia = 0.084

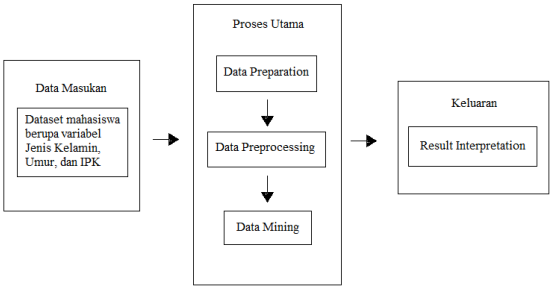
c. P3 = $\frac{3}{20} \times \frac{2}{3} \times \frac{1}{3} \times \frac{3}{3}$ = 0.15 x 0.666 x 0.333 x 1 = 0.033

Jadi nilai probabilitas jaringan = 0.033

Hasil perhitungan diatas menunjukkan bahwa nilai probabilitas tertinggi ada pada kelas P1 yaitu Pemrograman. Sehingga dapat disimpulkan bahwa prediksi mahasiswa A masuk dalam klarifikasi Pemrograman.

3. Analisa Pemodelan Sistem

Pemodelan sistem yang akan dibangun pada penelitian ini yaitu menggunakan Bahasa pemrograman *php*. Sedangkan Algoritma yang digunakan yaitu Algoritma *Naïve Bayes*. Berikut gambaran umum sistem aplikasi dapat dilihat pada Gambar 2.



Gambar 2. Gambaran Umum Sistem

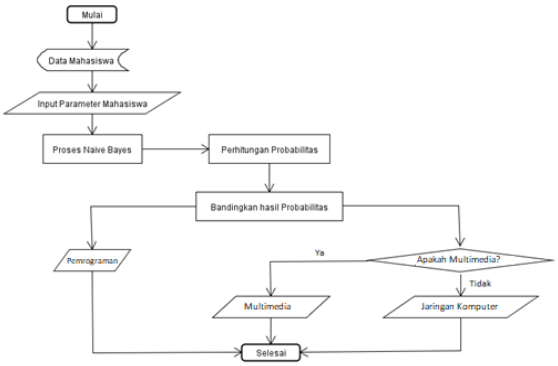
Berdasarkan pada Gambar 2 dapat diketahui bahwa program *data mining* ini memiliki beberapa proses yaitu berupa data masukkan, proses utama, dan keluaran yang dapat dijelaskan sebagai berikut :

a. Data masukan yang dibutuhkan oleh sistem untuk proses Algoritma *Naïve Bayes* berupa *dataset* mahasiswa yang dijadikan sebagai objek. Untuk variabel yang dipakai berupa jenis kelamin, umur dan IPK mahasiswa selama 3 (tiga) semester.
b. Proses utama menyatakan proses-proses berupa :
   1) *Data Preparation*, menyiapkan *dataset* mahasiswa Tahun Akademik 2015/2016 dan 2016/2017.
   2) *Data Preprocessing*, sebelum masuk ke bagian inti harus dilakukan pembersihan data *(data cleaning)* atau bisa disebut juga *tuning database*, dimana data harus dibersihkan terlebih dahulu yang sudah dijelaskan penulis pada sub bab Teknik Mengolah Data. Hal ini bertujuan untuk memperbaiki atau meningkatkan kinerja dan *performance* sebuah *database*.
   3) *Data Mining*, merupakan bagian inti dari sistem yaitu sebagai proses yang menggunakan teknik statistika, matematika dan kecerdasan buatan *(machine learning)* untuk mengekstrak dan melakukan identifikasi informasi aktual dan pengetahuan yang terkait dalam *database* berupa *dataset* mahasiswa.
c. Keluaran merupakan hasil dari proses utama yang terjadi pada sistem. *Output* dari sistem berupa pemilihan konsentrasi mahasiswa, diantaranya konsentrasi Pemrograman, Multimedia dan Jaringan Komputer.

4. Alur Algoritma Naïve bayes dalam Sistem

Berikut adalah alur dari Algoritma Naïve Bayes:

a. Baca Dataset Mahasiswa, kemudian lakukan filtering agar tidak terjadi redundancy data.
b. Cari nilai rata-rata (mean) dari setiap parameter.
c. Cari nilai probabilistik dari parameter atau class dengan cara menghitung jumlah data yang sesuai dari kelas yang sama dibagi dengan jumlah data pada kategori tersebut.
d. Lakukan pengklasifikasian Pemrograman, Jaringan dan Multimedia. alur proses naïve bayes disajikan pada Gambar 3.



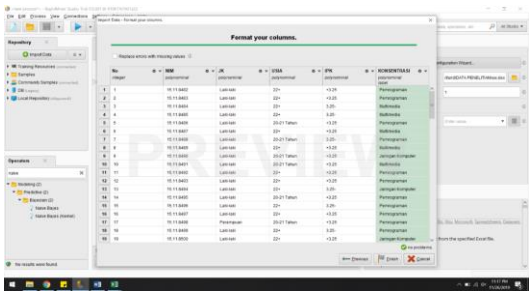
Gambar 3. Alur Proses *Naïve Bayes*

Hasil dan Pembahasan

Pada bagian ini akan dibahas tentang pengujian sistem yang bertujuan untuk mengetahui tingkat kaekurasian data serta akan dilakukan analisa hasil pengujian.

1. Uji Validasi

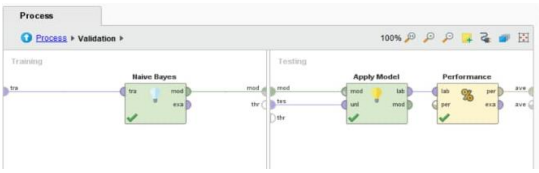
Data yang akan diuji yaitu berupa dataset sebanyak 1534 data dan menggunakan variabel Jenis Kelamin, Usia, dan IPK. Pada variabel Konsentrasi merupakan variabel penentu atau *ouput* dari uji validasi tersebut. Berikut beberapa *sample* data yang ditampilkan dan selanjutnya akan diuji menggunakan Aplikasi Rapid Miner. Pada Gambar 4 menunjukan uji validasi dataset mahasiswa.



Gambar 4. Uji Validasi dataset mahasiswa

2. Metode Uji Validasi

Pada pengujian tingkat akurasi menggunakan metode *Split Validation.* Operator ini melakukan validasi sederhana. Secara acak membagi SampleSet menjadi set pelatihan dan set tes dan mengevaluasi model. Operator ini melakukan validasi split untuk memperkirakan kinerja data *training*. Pada Gambar 5 menunjukkan Proses Training dan Testing dengan Algoritma Bayes.



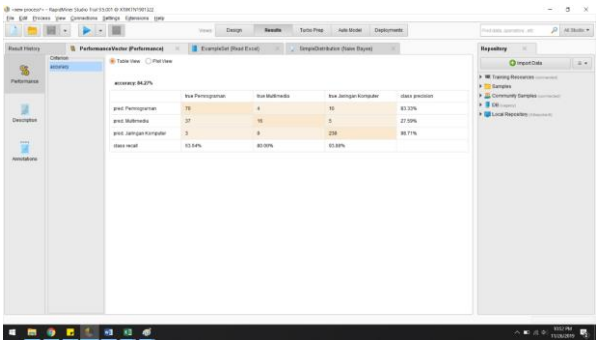
Gambar 5. Proses Training dan Testing dengan Algoritma Bayes.

Hal ini digunakan untuk memperkirakan seberapa akurat suatu model. Kemudian pada Gambar 6 dapat dilihat proses input dan output pada proses uji validasi menggunakan split validation.

Gambar 6. Uji Validasi *Split Validation*

3. Hasil Uji Validasi

Pengujian ini menggunakan nilai parameter yang diinputkan oleh pengguna (*user)* sistem, tetapi masih tetap mempertahankan variabel NIM, jenis kelamin,usia,IPK mata kuliah terkait dengan konsentrasi. Berikut hasil uji validasi dari dataset mahasiswa untuk 3 (tiga) klasifikasi Pemrograman, Multimedia, dan Jaringan Komputer. Gambar 7 menunjukan hasil uji validasi.



Gambar 7. Hasil uji validasi

Berdasarkan Gambar 7 dapat dijelaskan bahwa hasil uji validasi menghasilkan nilai akurasi sebesar 84.27%.

Kesimpulan dan Saran

Hasil pengujian yang telah dilakukan terhadap sample dataset sebanyak 1534 data menggunakan Algoritm Naïve Bayes, maka diperoleh hasil bahwa :

1. Hasil prediksi untuk pemilihan matakuliah konsentrasi memiliki nilai akurasi sebesar 84.27%.

2. Variabel berpengaruh terhadap hasil akurasi yang di hasilkan. Ukuran variabel yang sempit atau sedikit menyebabkan hasil akurasi yang kurang baik,Tetapi ukuran variabel yang luas bisa menghasilkan akurasi *ouput* yang lebih baik lagi.

Daftar Pustaka

[1] A. P. Fadillah and B. Hardiyana, "PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK PEMILIHAN KONSENTRASI

MATA KULIAH," *J. Teknol. dan Inf.*, vol. 8, no. 2, Nov. 2018, doi: 10.34010/jati.v8i2.1039.

[2] I. Verawati, "SISTEM PAKAR PENENTUAN KONSENTRASI PENJURUSAN MAHASISWA MENGGUNAKAN ALGORITMA BAYES," *J. Ilm. DASI*, vol. 16, no. 4, pp. 31–36, 2015.

[3] B. Bustami, "Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI-Jurnal Tek. Inform.*, vol. 5, no. 2, 2013, doi: https://doi.org/10.29103/techsi.v5i2.154.

[4] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, Jan. 2018, doi: 10.24076/citec.2017v4i2.106.

[5] E. Etriyanti, D. Syamsuar, and N. Kunang, "Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56–67, 2020, doi: 10.35671/telematika.v13i1.881.

[6] M. Munawir and T. Iqbal, "Prediksi Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes (Studi Kasus 5 PTS di Banda Aceh)," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 3, no. 2, p. 59, Sep. 2019, doi: 10.35870/jtik.v3i2.77.

[7] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, Apr. 2020, doi: 10.24076/citec.2019v6i1.178.