

**LAPORAN AKHIR**  
**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI PREDIKAT KELULUSAN**  
**MAHASISWA DENGAN MODEL NAÏVE BAYES**

Disusun untuk Memenuhi Tugas Mata Kuliah Data Mining

Oleh:

Kelompok 1

Khairul Zikria Burhan	1911522001
Salsabila Rahmah	2111521010
Annisa Hasifah Cantika	2111522012
Fahri Andika Sanjaya	2111523006



**Dosen Pengampu:**

Dwi Welly Sukma Nirad, S.Kom, M.T	199108122019032018
Aina Hubby Aziira, M.Eng	199504302022032013
Febby Apri Wenando, M.Eng	199104172022031007

**PROGRAM STUDI S1 SISTEM INFORMASI**  
**FAKULTAS TEKNOLOGI INFORMASI**  
**UNIVERSITAS ANDALAS**

**2023**

## DAFTAR ISI

<b>DAFTAR ISI.....</b>	<b>1</b>
<b>DAFTAR GAMBAR.....</b>	<b>2</b>
<b>DAFTAR TABEL.....</b>	<b>3</b>
<b>BAB 1 PENDAHULUAN.....</b>	<b>4</b>
1.1 Latar Belakang.....	4
1.2 Batasan Penelitian.....	4
1.3 Rumusan Masalah.....	5
1.4 Tujuan.....	5
<b>BAB 2 METODOLOGI.....</b>	<b>6</b>
2.1 Tahapan Penelitian.....	6
2.2 Identifikasi Masalah.....	6
2.3 Studi Pustaka.....	7
2.4 Pengumpulan Data.....	7
2.5 Preprocessing data.....	7
2.6 Algoritma Metode Classification.....	8
2.7 Algoritma Naive Bayes.....	8
<b>BAB 3 HASIL.....</b>	<b>9</b>
3.1 Proses Perhitungan Naive Bayes.....	9
3.1.1 Data Mahasiswa.....	9
3.1.2 Perhitungan.....	11
3.1.3 Pengujian Data.....	15
3.2 Hasil Evaluasi dengan Confusion Matrix menggunakan Jupyter.....	16
3.3 Analisis Hasil Evaluasi dengan Confusion Matrix menggunakan Jupyter.....	17
3.3.1 Analisis Hasil Tabel Confusion Matrix.....	17
3.3.2 Analisis Hasil Akurasi, Presisi, Recall, F1-score.....	19
<b>PENUTUP.....</b>	<b>21</b>
4.1 Kesimpulan.....	21
4.2 Saran.....	21
<b>DAFTAR PUSTAKA.....</b>	<b>22</b>

## DAFTAR GAMBAR

Gambar 2.1 Tahapan Penelitian.....	8
Gambar 3.1 Hasil Confusion Matrix.....	22
Gambar 3.2 Hasil Akurasi, Presisi, Recall, F1-score.....	23

## DAFTAR TABEL

Tabel 3.1 Data Alumni Mahasiswa.....	15
Tabel 3.2 Probabilitas IPK.....	16
Tabel 3.3 Probabilitas Durasi Kuliah.....	16
Tabel 3.4 Probabilitas Ada Tidaknya Nilai D.....	17
Tabel 3.5 Probabilitas Sikap.....	18
Tabel 3.6 Data Uji.....	18
Tabel 3.7 Matriks Probabilitas.....	19
Tabel 3.8 Matriks Hasil Probabilitas.....	19
Tabel 3.9 Confusion Matrix.....	20

## **BAB 1 PENDAHULUAN**

### **1.1 Latar Belakang**

Di era digital dan perkembangan teknologi informasi saat ini, data yang dihasilkan oleh perguruan tinggi atau lembaga pendidikan lainnya semakin melimpah. Data tersebut meliputi berbagai atribut dan variabel mengenai prestasi akademik mahasiswa, seperti nilai mata kuliah, frekuensi kehadiran, aktivitas ekstrakurikuler, dan faktor-faktor lain yang mempengaruhi kelulusan mahasiswa.

Dalam konteks tersebut, data mining menjadi alat yang penting untuk menganalisis data yang ada dan mengidentifikasi pola atau hubungan yang tersembunyi di dalamnya. Salah satu teknik data mining yang populer adalah Naïve Bayes, yang merupakan metode klasifikasi probabilitas yang didasarkan pada teorema Bayes.

Implementasi data mining dengan model Naïve Bayes dalam prediksi kelulusan mahasiswa memiliki tujuan untuk membantu pengambilan keputusan terkait dengan kelulusan mahasiswa berdasarkan data yang telah dikumpulkan. Dengan menerapkan teknik ini, lembaga pendidikan dapat mengidentifikasi faktor-faktor yang mempengaruhi kelulusan mahasiswa dan memberikan rekomendasi atau intervensi yang sesuai untuk meningkatkan tingkat kelulusan.

Melalui penggunaan model Naïve Bayes, diharapkan dapat dibangun suatu sistem klasifikasi yang mampu memprediksi predikat kelulusan mahasiswa dengan akurasi yang tinggi. Dengan demikian, lembaga pendidikan dapat mengoptimalkan penggunaan sumber daya dan menyediakan bantuan yang tepat kepada mahasiswa yang membutuhkannya, serta meningkatkan efisiensi proses akademik.

### **1.2 Batasan Penelitian**

Adapun batasan masalah dalam pembuatan laporan ini yaitu:

- a. Laporan ini menggunakan data mahasiswa yang sudah lulus atau alumni dari jurusan Sistem Informasi fakultas Teknologi Informasi.
- b. Metode yang digunakan pada analisis ini yaitu metode Naïve Bayes Classifier.
- c. Bahasa pemrograman yang digunakan pada proses analisis yaitu bahasa pemrograman Python.

### 1.3 Rumusan Masalah

Berdasarkan latar belakang penelitian terhadap data alumni semasa studi di Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Andalas, maka rumusan masalah pada penelitian ini adalah bagaimana metode klasifikasi dengan algoritma naive bayes digunakan dalam implementasi data mining untuk klasifikasi kelulusan mahasiswa dan bagaimana akurasi dalam implementasinya.

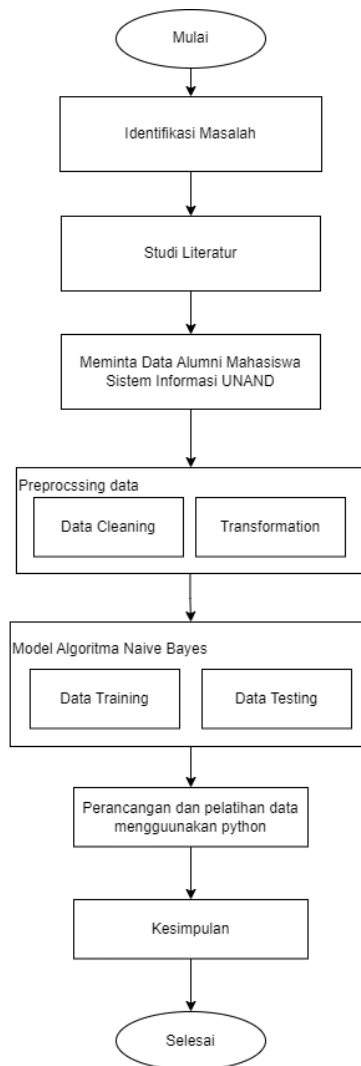
### 1.4 Tujuan

Berdasarkan rumusan masalah yang diuraikan, tujuan dari laporan ini adalah mengetahui metode klasifikasi dengan algoritma naive bayes serta dapat diketahui akurasi. Dapat mengetahui faktor-faktor yang mempengaruhi predikat kelulusan mahasiswa menggunakan metode naive bayes, dapat memberikan wawasan yang berguna bagi stakeholder terkait sehingga dapat meningkatkan predikat kelulusan mahasiswa, dan meningkatkan akreditasi program studi dengan melakukan identifikasi dan tindakan preventif terhadap mahasiswa berisiko memiliki predikat kurang baik.

## BAB 2 METODOLOGI

### 2.1 Tahapan Penelitian

Metodologi penelitian digunakan adalah sebagai acuan dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan. Gambar 2.1 akan menggambarkan bagaimana alur dari tahap penelitian.



**Gambar 2.1** Tahapan Penelitian

### 2.2 Identifikasi Masalah

Pada tahapan ini dilakukanya analisis dan mencari berbagai masalah yang akan dijadikan objek penelitian. Kemudian membuat rumusan masalah, mencari

teori-teori pendukung terkait dengan penelitian, selanjutnya menentukan tujuan penelitian agar penelitian lebih terarah.

### 2.3 Studi Pustaka

Studi Pustaka merupakan metode pengumpulan data dengan menganalisa dan mengutip informasi terkait melalui referensi tertentu. Beberapa sumber yang dijadikan referensi pada penelitian ini diantaranya adalah jurnal internasional maupun jurnal nasional, buku, e-book, artikel dan Skripsi.

### 2.4 Pengumpulan Data

Di tahapan pengumpulan data pada penelitian ini kami menggunakan data mahasiswa alumni Sistem Informasi Universitas Andalas. Data tersebut merupakan hasil wawancara dengan pihak jurusan Sistem Informasi Universitas Andalas dan juga berdasarkan referensi jurnal atau studi literatur yang data tersebut akan disimpan dalam bentuk CSV.

### 2.5 Preprocessing data

Dalam data mining, yaitu proses ekstraksi pengetahuan yang berharga, pola, atau informasi yang tersembunyi dalam data. Memiliki tujuan yaitu untuk mengidentifikasi pola-pola yang tersembunyi, pengetahuan yang berharga, dan informasi yang relevan dari sebuah dataset. Dalam pengolahan data atau preprocessing data di dalam data mining juga terdapat beberapa teknik yang dapat digunakan agar data bisa digunakan, maka data perlu diolah terlebih dahulu. Data akan diolah melalui beberapa tahapan, yaitu:

#### 1. Data Cleaning

Tahap ini melibatkan pembersihan data dari kesalahan, nilai yang hilang, atau outlier yang tidak relevan. Data yang tidak valid atau tidak berkualitas dibuang sehingga tersisa data yang bersih dan siap untuk diolah.

#### 2. Transformation

Pada tahap ini, data yang telah diintegrasikan diubah formatnya agar sesuai dengan teknik atau metode data mining yang akan digunakan. Transformasi data juga melibatkan pengkodean variabel kategori, normalisasi data, atau reduksi dimensi jika diperlukan.



## 2.6 Algoritma Metode Classification

Metode ini merupakan metode yang digunakan untuk mengidentifikasi kategori atau label pada data berdasarkan atribut-atribut yang ada. Algoritma klasifikasi mempelajari pola dan hubungan antara atribut untuk membangun model yang dapat digunakan untuk mengklasifikasikan data baru. Metode klasifikasi sendiri memiliki beberapa algoritma, diantaranya algoritma Naive Bayes, Decision Tree, KNN, SVM, Neural Network, dan beberapa algoritma lainnya. Dalam penelitian ini kami menggunakan algoritma Naive Bayes dan kemudian data tersebut akan dibagi menjadi 2 data yaitu data *training* dan data *testing*.

## 2.7 Algoritma Naive Bayes

Metode yang dikemukakan oleh ilmuwan Inggris Thomas Bayes ini berakar pada teorema Bayes yang memiliki asumsi bahwa setiap atribut dalam data adalah independen satu sama lain. Algoritma ini sangat populer dan sering digunakan dalam berbagai aplikasi, terutama dalam pengolahan teks, klasifikasi dokumen, dan analisis sentimen.

Teorema Bayes sendiri dapat ditulis sebagai:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Keterangan:

$P(A|B)$ : Probabilitas A terjadi dengan bukti bahwa B telah terjadi (probabilitas superior)

$P(B|A)$ : Probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$  : Peluang terjadinya A

$P(B)$  : Peluang terjadinya B

## BAB 3 HASIL

### 3.1 Proses Perhitungan Naive Bayes

#### 3.1.1 Data Mahasiswa

Kumpulan data alumni mahasiswa Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Andalas berupa data primer diperoleh melalui wawancara dengan bagian Akademik Jurusan Sistem Informasi Universitas Andalas.

Nama	NIM	IPK	Angkatan	Tahun Lulus	Durasi Kuliah	Ada tidaknya D	Sikap	Status Kelulusan
Putri Citra Denezi	1010963009	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Lola Ardi Lovina	1010961012	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Meriza Putri	1010962025	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Rona Mutia Sari	1010961008	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Muslim	1010962022	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Reni Fitria	1010962015	Diatas 3	2010	2015	5	Ada	Baik	Biasa
Refki Indra Hefiandes	1010961007	Diatas 3	2010	2015	5	Ada	Baik	Biasa
Gita Febriani	1010962001	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Nova Yeni	1010962011	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Rahmi Prisalia	1010962008	Diatas 3	2010	2015	5	Ada	Baik	Biasa
Anoki Antonia Caesar	1010961010	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa

Novya Ardiesa	1010963010	Di Bawah 3	2010	2015	5	Ada	Kurang Baik	Biasa
Ari Bespriadi	1010962029	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Devi Syani	1010962026	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Rian Ade Hidayat	1010963006	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Pramita Nanda Putri	1010961014	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Afriniati	1010961016	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Rahmi Nur Fitria	1010962018	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Fitri Muliany	1010962021	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Muhammad Junaidi	1010963015	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Karlina Fitriana	1010963005	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Fajar Lazuari	1010961003	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Zulhendri	1010961013	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Nindy Irzavika	1110961028	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Rahmatika Pratama Santi	1110962001	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Andre Nofriandi	1010961017	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa
Tuti Rahmawati	1010963001	Dibawah3	2010	2015	5	Ada	Kurang Baik	Biasa
Wike Ulfiani Aresa	1110962007	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Juliatriza	1010962024	Diatas 3	2010	2015	5	Tidak Ada	Baik	Biasa

Riza Perdamaian	1010961011	Diatas 3	2010	2015	5	Ada	Kurang Baik	Biasa
Mardaleni	1010963003	Dibawah3	2010	2015	5	Ada	Kurang Baik	Biasa
Sri Rahmina	1110963003	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Ovilia Mega Sari	1110962015	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Winni Rahdini	1110961001	Diatas 3	2011	2015	4	Ada	Baik	Terpuji
Rezie Novina	1110963015	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Fandi Ihsan	1110961011	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Yudhi Hartadi	1110961013	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Alberta Rahmat Ramadhan	1110962030	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Adjie Cahyana Putra	1110961010	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji
Hafiz Fajrin	1110962018	Diatas 3	2011	2015	4	Ada	Baik	Terpuji
Ilham Aminuddin	1110962002	Diatas 3	2011	2015	4	Ada	Kurang Baik	Terpuji
Septrian Yuwilda Yunahar	1110963004	Diatas 3	2011	2015	4	Tidak Ada	Baik	Terpuji

**Tabel 3.1** Data Alumni Mahasiswa

### 3.1.2 Perhitungan

Proses perhitungan menggunakan algoritma Naive Bayes metode klasifikasi dengan menghitung probabilitas masing-masing atribut berdasarkan data yang sudah didapatkan. Langkah proses perhitungan Naive Bayes adalah sebagai berikut :

a. Jumlah Atribut Target

Atribut target pada data diatas adalah ‘Status Kelulusan’. Terdapat dua elemen didalamnya, yaitu ‘Terpuji’ dan ‘Biasa’. Jumlah dari masing-masing elemen dalam atribut target adalah sebagai berikut :

- Terpuji : 14
- Biasa : 28

b. Probabilitas IPK

Perhitungan probabilitas dari atribut ‘IPK’ dengan rumus sebagai berikut :

$$P(T) = \frac{14}{14} = 1,00$$

$$P(T) = \frac{0}{14} = 0,00$$

$$P(B) = \frac{25}{28} = 0,89$$

$$P(B) = \frac{3}{28} = 0,11$$

IPK	Terpuji	Biasa	P(T)	P(B)
Diatas 3	14	25	1,00	0,89
Dibawah 3	0	3	0,00	0,11

**Tabel 3.2** Probabilitas IPK

Pada tabel diatas menunjukkan bahwa probabilitas terpuji dengan IPK diatas 3 adalah 1,00 dan probabilitas terpuji dengan IPK dibawah 3 adalah 0,00. Sedangkan probabilitas biasa dengan IPK diatas 3 adalah 0,89 dan probabilitas biasa dengan IPK dibawah 3 adalah 0,11.

c. Probabilitas Durasi Kuliah

Perhitungan probabilitas dari atribut ‘Durasi Kuliah’ dengan rumus sebagai berikut :

$$P(T) = \frac{14}{14} = 1,00$$

$$P(T) = \frac{0}{14} = 0,00$$

$$P(B) = \frac{0}{28} = 0,00$$

$$P(B) = \frac{28}{28} = 1,00$$

Durasi Kuliah (tahun)	Terpuji	Biasa	P(T)	P(B)
4	14	0	1,00	0,00
5	0	28	0,00	1,00

**Tabel 3.3** Probabilitas Durasi Kuliah

Pada tabel diatas menunjukkan bahwa probabilitas terpuji dengan durasi kuliah 4 tahun adalah 1,00 dan probabilitas terpuji dengan durasi kuliah 5 tahun adalah 0,00. Sedangkan probabilitas biasa dengan durasi kuliah 4 tahun adalah 0,00 dan probabilitas biasa dengan durasi kuliah 5 tahun adalah 1,00.

d. Probabilitas Adanya Nilai D

Perhitungan probabilitas dari atribut ‘Adanya Nilai D’ dengan rumus sebagai berikut :

$$P(T) = \frac{3}{14} = 0,21$$

$$P(T) = \frac{11}{14} = 0,79$$

$$P(B) = \frac{18}{28} = 0,64$$

$$P(B) = \frac{10}{28} = 0,36$$

Adanya D	Terpuji	Biasa	P(T)	P(B)
Ada	3	18	0,21	0,64
Tidak Ada	11	10	0,79	0,36

**Tabel 3.4** Probabilitas Ada Tidaknya Nilai D

Pada tabel diatas menunjukkan bahwa probabilitas terpuji dengan adanya nilai D adalah 0,21 dan probabilitas terpuji dengan tidak ada nilai D adalah 0,79. Sedangkan probabilitas biasa dengan adanya nilai D adalah 0,64 dan probabilitas biasa dengan tidak ada nilai D adalah 0,36.

e. Sikap

Perhitungan probabilitas dari atribut ‘Sikap’ dengan rumus sebagai berikut :

$$P(T) = \frac{13}{14} = 0,93$$

$$P(T) = \frac{1}{14} = 0,07$$

$$P(B) = \frac{13}{28} = 0,46$$

$$P(B) = \frac{15}{28} = 0,54$$

Sikap	Terpuji	Biasa	P(T)	P(B)
Baik	13	13	0,93	0,46
Kurang Baik	1	15	0,07	0,54

**Tabel 3.5** Probabilitas Sikap

Pada tabel diatas menunjukkan bahwa probabilitas terpuji dengan Sikap Baik adalah 0,93 dan probabilitas terpuji dengan sikap kurang baik adalah 0,07. Sedangkan probabilitas biasa dengan sikap Baik adalah 0,46 dan probabilitas biasa dengan sikap Kurang Baik adalah 0,54.

### 3.1.3 Pengujian Data

Pengujian data dilakukan dengan menguji data setelah ditemukannya pola dari dataset alumni mahasiswa Jurusan Sistem Informasi Fakultas teknologi Informasi angkatan 2010 Universitas Andalas.

Misalkan ada seorang mahasiswa bernama Alya yang akan menentukan status kelulusannya menggunakan pola yang telah ditemukan dari data yang ada. Berikut data uji yang akan ditentukan status kelulusannya :

Nama	NIM	IPK	Angkatan	Tahun Lulus	Durasi Kuliah	Ada tidaknya D	Sikap	Status Kelulusan
Alya	1110190031	Diatas 3	2010	2015	5	Tidak Ada	Baik	????

**Tabel 3.6** Data Uji

Berdasarkan data Alya pada tabel, maka dibuatkan matriks dengan menyesuaikan dengan matriks tabel probabilitas atribut penentu. Dapat dilihat pada matriks tabel berikut :

Alya	Diatas 3	5	Tidak ada	Baik
Terpuji	1	0	0,79	0,93
Biasa	0,89	1	0,36	0,46

**Tabel 3.7** Matriks Probabilitas

Alya	Diatas 3	5	Tidak ada	Baik	Hasil
Terpuji	1	0	0,79	0,93	0
Biasa	0,89	1	0,36	0,46	0,147384

**Tabel 3.8** Matriks Hasil Probabilitas



Tabel diatas menunjukkan hasil dari perkalian masing-masing probabilitas atribut penentu secara horizontal. Hasil yang dipilih adalah hasil paling tinggi diantara keduanya. Pada Tabel, Biasa bernilai 0,147384 yang merupakan probabilitas paling tinggi dibandingkan probabilitas Terpuji. Jadi, status kelulusan Alya diklasifikasikan kepada Biasa setelah dilakukannya perhitungan dengan algoritma Naive Bayes.

### 3.2 Hasil Evaluasi dengan *Confusion Matrix* menggunakan Jupyter

Confusion Matrix merupakan pengukuran performa buat permasalahan klasifikasi machine learning dimana keluaran bisa berbentuk 2 kelas ataupun lebih. Confusion Matrix merupakan tabel dengan 4 campuran berbeda dari nilai prediksi serta nilai aktual. Dengan menggunakan jupyter didapatkan hasil evaluasi sebagai berikut :

	BIASA	TERPUJI
BIASA	8	0
TERPUJI	0	5

**Tabel 3.9** Confusion Matrix

Pada gambar diatas menunjukan nilai-nilai dari *Confusion Matrix* seperti berikut :

1. Nilai True Positif (TP) adalah 8
2. Nilai True Negatif (TP) adalah 5
3. Nilai False Positif (FP) adalah 0
4. Nilai False Negatif(FN adalah 0

Berdasarkan nilai-nilai diatas didapatkan nilai Accuracy, Precision dan Recall dengan menggunakan metode Naive Bayes sebagai berikut :

1. Akurasi: 100%
2. Presisi: 100%
3. Recall: 100%
4. F1-score: 100%

### 3.3 Analisis Hasil Evaluasi dengan *Confusion Matrix* menggunakan Jupyter

#### 3.3.1 Analisis Hasil Tabel Confusion Matrix

```
import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report

# Read the dataset
dataset = pd.read_csv('tebe.csv', delimiter=';', decimal=',')

# Replace values in 'IPK' column
dataset['IPK'] = dataset['IPK'].replace('Diatas 3', 1)
dataset['IPK'] = dataset['IPK'].replace('Dibawah3', 0)

# Replace values in 'Ada tidaknya D' column
dataset['Ada tidaknya D'] = dataset['Ada tidaknya D'].replace('Tidak Ada',
1)
dataset['Ada tidaknya D'] = dataset['Ada tidaknya D'].replace('Ada', 0)

# Replace values in 'Sikap' column
dataset['Sikap'] = dataset['Sikap'].replace('Baik', 1)
dataset['Sikap'] = dataset['Sikap'].replace('Kurang Baik', 0)

# Replace values in 'Durasi Kuliah' column
dataset['Durasi Kuliah'] = dataset['Durasi Kuliah'].apply(lambda x: 1 if x <=
4 else 0)
print(dataset)

# memilih fitur yang akan digunakan sebagai kolom klasifikasi
X = dataset[['Sikap', 'IPK', 'Durasi Kuliah', 'Ada tidaknya D']]
```

```
# memilih target sebagai kolom Biasa/Terpuji
y = dataset['Status Kelulusan']

# membagi dataset menjadi data training dan data testing
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=0)

# membuat model Naive Bayes
model = GaussianNB()

# melakukan training pada model
model.fit(X_train, y_train)

# melakukan prediksi pada data testing
y_pred = model.predict(X_test)

# menghitung nilai Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

Jika kodingan diatas dirun, maka akan didapatkan hasil confusion matrix sebagai berikut :

```
[[8 0]
 [0 5]]
```

**Gambar 3.1** Hasil Confusion Matrix

### 3.3.2 Analisis Hasil Akurasi, Presisi, Recall, F1-score

```
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score

# menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print('Akurasi:', accuracy)

# menghitung presisi
precision = precision_score(y_test, y_pred, average='weighted')
print('Presisi:', precision)

# menghitung recall
recall = recall_score(y_test, y_pred, average='weighted')
print('Recall:', recall)

# menghitung F1-score
f1 = f1_score(y_test, y_pred, average='weighted')
print('F1-score:', f1)

print(classification_report(y_test, y_pred))
```

Selanjutnya jika kodingan diatas dilanjutkan pada kodingan sebelumnya maka akan didapatkan hasil sebagai berikut :

Akurasi: 1.0  
Presisi: 1.0  
Recall: 1.0  
F1-score: 1.0

	precision	recall	f1-score	support
Biasa	1.00	1.00	1.00	8
Terpuji	1.00	1.00	1.00	5
accuracy			1.00	13
macro avg	1.00	1.00	1.00	13
weighted avg	1.00	1.00	1.00	13

**Gambar 3.2** Hasil Akurasi, Presisi, Recall, F1-score

## PENUTUP

### 4.1 Kesimpulan

Laporan ini membahas tentang penerapan metode data mining untuk memprediksi tingkat predikat kelulusan mahasiswa terpuji atau biasa dengan menggunakan algoritma Naïve Bayes. Algoritma ini mengklasifikasikan data berdasarkan probabilitas sederhana dari setiap atribut. Penelitian ini menggunakan data mahasiswa dari Jurusan Sistem Informasi angkatan 2010 sebagai studi kasus. Hasil penelitian menunjukkan bahwa metode Naïve Bayes dapat menghasilkan akurasi klasifikasi sebesar 100%.

### 4.2 Saran

Melakukan evaluasi dan perbandingan dengan metode klasifikasi lainnya: Meskipun model Naïve Bayes memberikan hasil yang baik, disarankan untuk melakukan evaluasi dan perbandingan dengan metode klasifikasi lainnya seperti Decision Tree, Random Forest, atau Support Vector Machines. Dengan membandingkan beberapa metode klasifikasi, peneliti dapat memperoleh pemahaman yang lebih komprehensif tentang kinerja setiap metode dalam hal akurasi, presisi, recall, atau matrik evaluasi lainnya yang relevan. Hal ini dapat membantu dalam memilih metode yang paling sesuai dengan data yang ada dan memperoleh hasil yang lebih andal dan dapat diandalkan.

Selain itu, membandingkan beberapa metode klasifikasi juga dapat mengidentifikasi kekuatan dan kelemahan masing-masing metode, sehingga peneliti dapat mempertimbangkan aspek-aspek tersebut dalam konteks implementasi dan kebutuhan penelitian. Dengan demikian, penelitian akan memiliki landasan yang kuat dan argumentasi yang mendukung dalam pemilihan metode klasifikasi yang optimal untuk klasifikasi predikat kelulusan mahasiswa.

## DAFTAR PUSTAKA

- Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). Introduction to Data Mining. Pearson.
- Berry, M. J., & Linoff, G. (2011). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.
- Mitchell, T. M. (1997). Machine learning. McGraw Hill.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Zhang, H. H. (2004). The optimality of Naive Bayes. Proceedings of the 17th International FLAIRS Conference.
- Ardhana, F. (2021). Sentiment Analysis Tweet Vaksinasi Covid-19 Menggunakan Metode Naive Bayes Classifier. Tugas Akhir, Universitas Andalas
- ‘4 - Data Mining Techniques For Marketing, Sales, and Customer Relationship Management’ (no date).
- Friedman, J., Hastie, T. and Tibshirani, R. (2008) *The Elements of Statistical Learning Preface to the Second Edition*.
- Han, J., Kamber, M. and Pei, J. (2011) *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Manning, C.D., Raghavan, Prabhakar. and Schütze, Hinrich. (2008) *Introduction to information retrieval*. Cambridge University Press.
- Alim, S. (no date) *IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR, DECISION TREE AND NAIVE BAYES MODELS, Jurnal Ilmiah NERO*.
- Amalia, R. (2020) ‘Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode Naïve Bayes’, *JUISI*, 06(01).
- Annur, H. (2018) *KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES, Agustus*.
- Damuri, A. *et al.* (2021) ‘Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako’, *JURIKOM (Jurnal Riset Komputer)*, 8(6), p. 219. Available at: <https://doi.org/10.30865/jurikom.v8i6.3655>.

Dirjen, S.K. *et al.* (2017) ‘Terakreditasi SINTA Peringkat 2 Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier’, *masa berlaku mulai*, 1(3), pp. 95–101.

Janu, S. *et al.* (2021) *ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN*.

Kamil, M. *et al.* (2020) ‘Perbandingan Algoritma C4.5 dan Naive Bayes Pada Lulusan Tepat Waktu Mahasiswa’, *JURNAL INFORMATIKA*, 7(2), pp. 97–106. Available at: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>.

Khasanah, N. and Salim, A. (2022) *Rachman Komarudin 4) , Yana Iqbal Maulana 5) 1) Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 2,3) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika 4) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 5) Teknik Informatika, Fakultas Teknologi Informasi*.

Setiyani, L. *et al.* (2020) ‘Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review’, *Faktor Exacta*, 13(1), p. 35. Available at: <https://doi.org/10.30998/faktorexacta.v13i1.5548>.

Widaningsih, S. (2019) ‘PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4,5, NAÏVE BAYES, KNN DAN SVM’, *Jurnal Tekno Insentif*, 13(1), pp. 16–25. Available at: <https://doi.org/10.36787/jti.v13i1.78>.

Widhi Saputro, I. and Wulan Sari, B. (2019) ‘Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction’, *Citec Journal*, 6(1).