



Studi Kasus

## IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI PREDIKAT KELULUSAN MAHASISWA DENGAN MODEL NAÏVE BAYES

Salsabila Rahmah<sup>a</sup>, Annisa Hasifah Cantika<sup>b</sup>, Fahri Andika Sanjaya<sup>c</sup>, Khairul Zikra Burhan<sup>d</sup>

<sup>a,b,c,d</sup>Departmen Sistem Informasi, Fakultas Teknologi Informasi, Universitas Andalas, Limau Manih, Kota Padang 25163, Indonesia

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 00 Februari 00

Revisi Akhir: 00 Maret 00

Diterbitkan Online: 00 April 00

### KATA KUNCI

*Naïve Bayes*

*Classification*

Prediksi Kelulusan Mahasiswa

*Accuracy*

### KORESPONDENSI

E-mail: 2111521010\_salsabila@gmail.com

### ABSTRACT

Predikat kelulusan mahasiswa adalah penilaian atau klasifikasi yang diberikan kepada mahasiswa berdasarkan pencapaian akademik mereka selama masa studi. Predikat kelulusan dalam konteks jurusan atau program studi tertentu dapat memiliki tujuan yang lebih spesifik. Tujuan predikat kelulusan dalam Jurusan adalah untuk memberikan penilaian yang relevan terhadap kemampuan dan prestasi akademik mahasiswa dalam bidang studi khusus tersebut. 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. PENDAHULUAN

Di era digital dan perkembangan teknologi informasi saat ini, perguruan tinggi dan lembaga pendidikan lainnya menghasilkan jumlah data yang semakin melimpah. Data ini mencakup berbagai atribut dan variabel yang berkaitan dengan prestasi akademik mahasiswa, seperti nilai mata kuliah, frekuensi kehadiran, aktivitas ekstrakurikuler, dan faktor-faktor lain yang memengaruhi kelulusan mahasiswa.

Dalam konteks ini, data mining menjadi alat 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[1]. Model *Naïve Bayes* telah digunakan secara luas dalam berbagai aplikasi, termasuk prediksi kelulusan mahasiswa [2].

Implementasi data mining dengan model *Naïve Bayes* dalam prediksi kelulusan mahasiswa bertujuan untuk membantu pengambilan keputusan terkait dengan kelulusan mahasiswa berdasarkan data yang telah dikumpulkan. Dengan menerapkan teknik ini, lembaga pendidikan dapat mengidentifikasi faktor-

faktor yang memengaruhi kelulusan mahasiswa dan memberikan rekomendasi atau intervensi yang sesuai untuk meningkatkan tingkat kelulusan.

Penggunaan model *Naïve Bayes* diharapkan dapat membangun suatu sistem klasifikasi yang mampu memprediksi predikat kelulusan mahasiswa dengan akurasi yang tinggi [3]. Dengan demikian, lembaga pendidikan dapat mengoptimalkan penggunaan sumber daya dan menyediakan bantuan yang tepat kepada mahasiswa yang membutuhkannya, serta meningkatkan efisiensi proses akademik.

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

Data mining adalah suatu pencarian dan analisa pada suatu koleksi data (*database*) sehingga ditemukan suatu pola yang menarik dengan tujuan mengekstrak informasi dan pengetahuan yang akurat dan potensial, serta dapat dipahami dan berguna bagi pengambilan keputusan[4].

Menurut MacLennan, Tang, & Crivat (2011), data mining memiliki beberapa fungsi yaitu (Saleh & Nasari, 2018) :

1. *Classification* adalah proses untuk mencari model atau fungsi yang menggambarkan dan membedakan kelas – kelas atau konsep data.
2. *Clustering* berfungsi untuk mencari pengelompokan atribut kedalam segmentasi – segmentasi berdasarkan similaritas.
3. *Association* berfungsi untuk mencari keterkaitan antara atribut atau item set, berdasarkan item yang muncul dan rule *association* yang ada.
4. *Regression* bertujuan untuk mencari prediksi dari suatu pola yang ada.
5. *Forecasting* berfungsi untuk meramalkan waktu yang akan datang berdasarkan trend yang telah terjadi di waktu sebelumnya.
6. *Sequence Analysis* berfungsi untuk mencari pola urutan dari rangkaian kejadian.
7. *Deviation Analysis* berfungsi untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

## 2.2. Algoritma Naïve Bayes

*Naïve Bayes* merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. *Bayesian classification* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class[5]. Cara kerja *Naïve Bayes* yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya[6].

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

Keterangan :

$P(C_i|X)$  = Probabilitas hipotesis  $C_i$  jika diberikan fakta atau *record*  $X$  (*Posterior probability*)

$P(X|C_i)$  = mencari nilai parameter yang memberi kemungkinan yang paling besar (*likelihood*)

$P(C_i)$  = Prior probability dari  $X$  (*Prior probability*)

$P(X)$  = Jumlah probability *tuple* yg muncul.

Pada perkembangannya,  $P(X)$  dapat dihilangkan karena nilai tetap, sehingga saat dibandingkan dengan tiap kategori, nilai ini dapat dihapuskan karena asumsi atribut tidak selalu terkait (*conditionally indenpenden*) maka:[5]

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^n P(x_k|C_i) \\ &= P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i). \end{aligned}$$

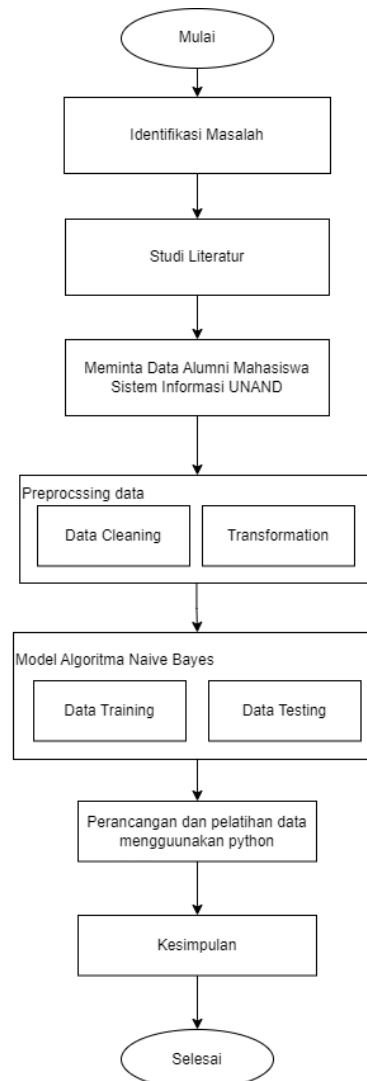
Bila  $P(X)$  dapat diketahui melalui perhitungan diatas, maka klas (label) dari data sampel  $X$  adalah klas (label) yang memiliki  $P(X|C_i) * P(C_i)$  maksimum.

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior probability}}{\text{Evidence}}$$

## 3. METODOLOGI

### 3.1. Tahapan Penelitian

Tahapan penelitian digunakan sebagai acuan dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan[1].



Gambar 1. Tahapan Penelitian

### 3.2. 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[7].

### 3.3. Studi Pustaka

Studi Pustaka merupakan metode pengumpulan data dengan menganalisa dan mengutip informasi terkait melalui referensi tertentu[8]. Beberapa sumber yang dijadikan referensi pada penelitian ini diantaranya adalah jurnal internasional maupun jurnal nasional, buku, *e-book*, artikel dan Skripsi.

### 3.4. 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[9].

### 3.5. 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[10]. 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[11]. Data akan diolah melalui beberapa tahapan, yaitu:

#### 3.5.1. Data Cleaning

Tahap ini melibatkan pembersihan data dari kesalahan, nilai yang hilang, atau outlier yang tidak relevan[12]. Data yang tidak valid atau tidak berkualitas dibuang sehingga tersisa data yang bersih dan siap untuk diolah.

#### 3.5.2. Transformation

Pada tahap ini, data yang telah diintegrasikan diubah formatnya agar sesuai dengan teknik atau metode data mining yang akan digunakan[13]. Transformasi data juga melibatkan pengkodean variabel kategori, normalisasi data, atau reduksi dimensi jika diperlukan.

### 3.6. 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 *Naïve Bayes*, *Decision Tree*, KNN, SVM, *Neural Network*, dan beberapa algoritma lainnya[14]. Dalam penelitian ini kami menggunakan algoritma *Naïve Bayes* dan kemudian data tersebut akan dibagi menjadi 2 data yaitu data *training* dan data *testing*.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Proses Perhitungan Naive Bayes

#### 4.1.1. 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 Priselila        | 1010962008 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Baik        | Biasa            |
| Anoki Antonia Caesar   | 1010961010 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
|                        |            |            |          |             |               |                |             |                  |
| Novya Ardiansa         | 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 Ada Hidayat       | 1010963006 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
| Pranita Nanda Putri    | 1010961014 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
| Afrinarti              | 1010961016 | Diatas 3   | 2010     | 2015        | 5             | Tidak Ada      | Baik        | Biasa            |
| Rahmi Nur Fitria       | 1010962018 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
| Fitri Muliary          | 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 Lazuardi         | 1010961003 | Diatas 3   | 2010     | 2015        | 5             | Tidak Ada      | Baik        | Biasa            |
| Zulhendri              | 1010961013 | Diatas 3   | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
| Nindy Irvavika         | 1110961028 | Diatas 3   | 2011     | 2015        | 4             | Tidak Ada      | Baik        | Terpuji          |
| Rahmatika Pratama Sami | 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 | Dibawah 3  | 2010     | 2015        | 5             | Ada            | Kurang Baik | Biasa            |
| Wike Ulfiani Aresa     | 1110962007 | Diatas 3   | 2011     | 2015        | 4             | Tidak Ada      | Baik        | Terpuji          |
| Julianrice             | 1010962024 | Diatas 3   | 2010     | 2015        | 5             | Tidak Ada      | Baik        | Biasa            |

|                          |            |          |      |      |   |           |             |         |
|--------------------------|------------|----------|------|------|---|-----------|-------------|---------|
| Riza Perdanians          | 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 |
| Renzie 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 |
| Iham Aminuddin           | 1110962002 | Diatas 3 | 2011 | 2015 | 4 | Ada       | Kurang Baik | Terpuji |
| Septrian Yuvilda Yunahar | 1110963004 | Diatas 3 | 2011 | 2015 | 4 | Tidak Ada | Baik        | Terpuji |

Gambar 2. Data Mahasiswa

#### 4.1.2. Perhitungan

Proses perhitungan menggunakan algoritma *Naïve Bayes* metode klasifikasi dengan menghitung probabilitas masing-masing atribut berdasarkan data yang sudah didapatkan. Langkah proses perhitungan *Naïve 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 :

Tabel 1. Atribut dan Value

| Atribut | Value |
|---------|-------|
| Terpuji | 14    |
| Biasa   | 28    |

##### b. Probabilitas IPK

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

$$P(T) = 14/14 = 1,00$$

$$P(T) = 0/14 = 0,00$$

$$P(B) = 25/28 = 0,89$$

$$P(B) = 3/28 = 0,11$$

Tabel 2. IPK

| IPK       | Terpuji | Biasa | P(T) |
|-----------|---------|-------|------|
| Diatas 3  | 14      | 25    | 1,00 |
| Dibawah 3 | 0       | 3     | 0,00 |

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 :

$$P(T) = 14/14 = 1,00$$

$$P(T) = 0/14 = 0,00$$

$$P(B) = 0/28 = 0,00$$

$$P(B) = 28/28 = 1,00$$

Tabel 3. Durasi Kuliah

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

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 1,00 dan probabilitas biasa dengan durasi kuliah 5 tahun adalah 0,00.

##### d. Probabilitas Adanya Nilai D

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

$$P(T) = 3/14 = 0,21$$

$$P(T) = 11/14 = 0,79$$

$$P(B) = 18/28 = 0,64$$

$$P(B) = 10/28 = 0,36$$

Tabel 4. Adanya Nilai D

| Adanya D  | Terpuji | Biasa | P(T) |
|-----------|---------|-------|------|
| Ada       | 3       | 18    | 0,21 |
| Tidak Ada | 11      | 10    | 0,79 |

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 :

$$P(T) = 13/14 = 0,93$$

$$P(T) = 1/14 = 0,07$$

$$P(B) = 13/28 = 0,46$$

$$P(B) = 15/28 = 0,54$$

Tabel 5. Sikap

| Sikap       | Terpuji | Biasa | P(T) |
|-------------|---------|-------|------|
| Baik        | 13      | 13    | 0,93 |
| Kurang Baik | 1       | 15    | 0,07 |

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.

#### 4.1.3. 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         | Angkat<br>an | Tahu<br>n<br>Lulus | Durasi<br>Kuliah | Ada<br>tidaknya<br>D | Sikap | Status<br>Kelulus<br>an |
|------|----------------|-------------|--------------|--------------------|------------------|----------------------|-------|-------------------------|
| Alya | 11101900<br>31 | Diatas<br>3 | 2010         | 2015               | 5                | Tidak<br>Ada         | Baik  | ???                     |

Gambar 3. Data Uji

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

Tabel 6. Data Uji

| Alya    | Diatas<br>3 | 5 | Tidak<br>ada | Baik |
|---------|-------------|---|--------------|------|
| Terpuji | 1           | 0 | 0,79         | 0,93 |
| Biasa   | 0,89        | 1 | 0,36         | 0,46 |

Tabel 7. Hasil Uji

| Alya    | Diatas<br>3 | 5 | Tidak<br>ada | Baik | Hasil    |
|---------|-------------|---|--------------|------|----------|
| Terpuji | 1           | 0 | 0,79         | 0,93 | 0        |
| Biasa   | 0,89        | 1 | 0,36         | 0,46 | 0,147384 |

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 *Naïve Bayes*.

#### 4.2. 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[15]. Dengan menggunakan jupyter didapatkan hasil evaluasi sebagai berikut :

Tabel 8. Hasil Evaluasi *Confusion Matix*

|         | BIASA | TERPUJI |
|---------|-------|---------|
| BIASA   | 8     | 0       |
| TERPUJI | 0     | 5       |

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

1. Nilai *True* Positif (TP) adalah 26
2. Nilai *True* Negatif (TN) 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 *Naïve Bayes* sebagai berikut :

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

#### 4.3. Analisis Hasil Evaluasi dengan Confusion Matrix menggunakan Jupyter

##### 4.3.1. 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)
```

Gambar 4. Kodingan

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

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

#### 4.3.2. 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 5. Akurasi

## 5. KESIMPULAN DAN SARAN

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%.

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

- [1] "Data Mining Third Edition."
- [2] M. J. Berry, G. S. Linoff, and F. Marketing, "Customer Relationship Management Second Edition Data Mining Techniques."
- [3] T. M. (Tom M. Mitchell, *Machine Learning*.
- [4] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Faktor Exacta*, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [5] M. Ridho Handoko, "SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [6] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1.
- [7] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [8] S. Janu *et al.*, "ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING UNTUK MEMREDIKSI KELULUSAN," 2021.
- [9] R. Amalia, "Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode Naïve Bayes," *JUISI*, vol. 06, no. 01, 2020.
- [10] M. Kamil, W. Cholil, D. Palembang Jl Jendral Ahmad Yani No, and S. Selatan, "Perbandingan Algoritma C4.5 dan Naive Bayes Pada Lulusan Tepat Waktu Mahasiswa," *JURNAL INFORMATIKA*, vol. 7, no. 2, pp. 97–106, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [11] S. Alim, "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."
- [12] S. Widaningsih, "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*, vol. 13, no. 1, pp. 16–25, Apr. 2019, doi: 10.36787/jti.v13i1.78.
- [13] I. Widhi Saputro and B. Wulan Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction," *Citec Journal*, vol. 6, no. 1, 2019.
- [14] N. Khasanah and A. Salim, "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," 2022.
- [15] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, E. Sutoyo, and A. Almaarif, "Terakreditasi SINTA Peringkat 2 Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *masa berlaku mulai*, vol. 1, no. 3, pp. 95–101, 2017.