

## IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK PREDIKSI KELULUSAN MAHASISWA

<sup>1</sup>Angga Aditya Permana, <sup>2</sup>Rohmat Taufiq, <sup>3</sup>Rachmat Destriana, <sup>4</sup>Aliya Nur'aini

<sup>1</sup>Prodi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,  
Jl. Scientia Boulevard, Gading Serpong, Kel. Curug Sangereng, Kec. Kelapa Dua, Kab.  
Tangerang, Banten

<sup>2,3,4</sup>Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang,  
Jl. Perintis Kemerdekaan 1 No 33, Kota Tangerang, Banten  
e-mail: angga.permana@umn.ac.id

Receive: 30 Januari 2024

Accepted: 01 Maret 2024

### Abstrak

Pada tingkat perguruan tinggi, pencapaian kelulusan tepat waktu adalah indikator kunci dari keberhasilan mahasiswa. Namun, mengidentifikasi faktor-faktor yang berpotensi memengaruhi kelulusan mahasiswa merupakan tantangan yang kompleks. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem prediksi tingkat kelulusan mahasiswa dengan menggunakan metode Naïve Bayes. Langkah-langkah penelitian mencakup pengumpulan data dari dataset Kaggle, pembersihan data untuk menangani nilai yang hilang atau tidak relevan, transformasi data untuk mempersiapkannya untuk analisis, dan penerapan metode Naïve Bayes sebagai model prediktif. Variabel yang digunakan dalam analisis meliputi jenis kelamin, status mahasiswa, usia, nilai Indeks Prestasi Semester (IPS), nilai Indeks Prestasi Kumulatif (IPK), dan status kelulusan. Hasil eksperimen menunjukkan bahwa model prediksi mencapai akurasi sebesar 89%, dengan presisi sekitar 88% untuk kelas 0 dan 89% untuk kelas 1. Selain itu, recall mencapai sekitar 85% untuk kelas 0 dan 91% untuk kelas 1. Diharapkan hasil penelitian ini dapat memberikan kontribusi penting dalam meningkatkan efektivitas prediksi tingkat kelulusan mahasiswa, sehingga institusi pendidikan dapat mengambil tindakan preventif yang lebih tepat untuk mendukung keberhasilan akademis mahasiswa.

**Kata Kunci:** Kelulusan Mahasiswa, Naïve Bayes, Prediksi.

### PENDAHULUAN

Tingkat keberhasilan mahasiswa di perguruan tinggi merupakan salah satu penentu utama dalam membentuk sumber daya manusia berkualitas. Di sana, pencapaian kelulusan tepat waktu memegang peran sentral sebagai tolok ukur prestasi. (Apridiansyah *et al.* 2021; Abdurrohman dan Wibowo 2022; Supangat dan Sulistyawan 2023) Namun, menemukan faktor-faktor yang memengaruhi kelulusan dengan tepat seringkali menjadi tantangan yang besar. Oleh karena itu, langkah prediksi kelulusan mahasiswa menjadi langkah krusial yang dapat membantu perguruan tinggi mengambil langkah-langkah preventif yang sesuai, membawa kemajuan bagi masa depan akademis mahasiswa.

Metode Naive Bayes, sebuah pendekatan dalam ilmu data mining dan kecerdasan buatan, telah terbukti efektif dalam berbagai bidang (Permana *et al.* 2021; PERMANA dan NOVIYANTO 2023), termasuk prediksi kelulusan mahasiswa (Mayadewi dan Rosely 2020; Abdurrohman dan Wibowo 2022) . Dengan menganalisis data historis dan memanfaatkan informasi probabilitas, metode Naive Bayes mampu memberikan prediksi

yang akurat tentang kemungkinan kelulusan seorang mahasiswa berdasarkan karakteristik dan perilaku akademik yang dimilikinya (Steven dan Wella 2020).

Pada penelitian ini, akan mengimplementasikan metode Naive Bayes untuk meramalkan kelulusan mahasiswa. Data yang akan digunakan mencakup beragam variabel seperti jenis kelamin, status mahasiswa, usia, status pernikahan, nilai rata-rata tiap semester, dan Indeks Prestasi Kumulatif (IPK) (Nafi'iyah *et al.* 2020). Dengan menerapkan analisis probabilistik menggunakan Naive Bayes, harapannya penelitian ini dapat memberikan kontribusi yang berarti dalam membantu perguruan tinggi mengidentifikasi mahasiswa yang berpotensi tidak lulus tepat waktu.

## METODE PENELITIAN

Dalam penelitian ini, tahapan penelitian yang digunakan yaitu:



Gambar 1. Tahapan Penelitian

### Pengumpulan Data

Penelitian ini menggunakan dataset Kaggle untuk memprediksi tingkat kelulusan mahasiswa. Di mana data diambil dan diolah untuk mengambil variabel-variabel yang berkorelasi dengan tingkat kelulusan mahasiswa. Output dari sistem prediksi ini adalah tingkat kelulusan mahasiswa.

### Pembersihan Data

Pembersihan data merupakan langkah yang dilakukan sebelum masuk pada proses mining pada data. Pembersihan data berisi beberapa kegiatan yang tujuan utamanya adalah melakukan pengenalan dan perbaikan pada data yang akan diteliti. Dalam proses ini, beberapa data yang noise atau kosong dan atribut yang tidak berpengaruh dalam perhitungan akan dihapus. Untuk atribut yang dihapus yaitu nama. Sehingga atribut yang masih dipakai, antara lain: jenis kelamin, status mahasiswa, umur, status nikah, IPS, IPK, dan status kelulusan.

Tabel 1. Dataset Sebelum Proses Cleaning

NAMA	JK	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
ANIK WIDAYANTI	PR	BEKERJA	28	BM	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07	TERLAMBAT
DWI HESTYNA PRIHASTANTY	PR	MHS	32	BM	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	TERLAMBAT
MURYA ARIEF BASUKI	PR	BEKERJA	29	BM	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	TERLAMBAT

Tabel 2. Dataset Setelah Proses Cleaning

JK	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
PR	BEKERJA	28	BM	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07	TERLAMBAT
PR	MHS	32	BM	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	TERLAMBAT
PR	BEKERJA	29	BM	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	TERLAMBAT

### Transformasi Data

Data yang sudah dibersihkan selanjutnya akan dikonversikan agar sesuai untuk diproses ke dalam data mining. Data asli mempunyai beberapa atribut yang perlu diubah agar dapat diolah dengan perhitungan data mining, antara lain: jenis kelamin, status mahasiswa, status nikah, dan status kelulusan. Perubahan nilai pada atribut jenis kelamin: perempuan (0) dan laki-laki (1); pada status mahasiswa: mahasiswa (0) dan bekerja (1); pada status nikah: belum nikah (0) dan nikah (1); dan status kelulusan: terlambat (0) dan tepat (1).

Tabel 3. Dataset setelah dilakukan transformasi

JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
0	1	28	0	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07	0
0	0	32	0	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	0
0	1	29	0	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	0

## HASIL DAN PEMBAHASAN

### Memberi Nilai Pada *Missing Value*

*Missing Value* adalah informasi yang tidak tersedia untuk sebuah objek (kasus). *Missing value* terjadi karena informasi untuk sesuatu tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Menurut definisinya, *missing value* adalah hilangnya beberapa data yang telah diperoleh.

Sebelum dilakukan penelitian, perlu kiranya mengetahui dataset yang akan digunakan pada penelitian. Dimana pada dataset tersebut memiliki nilai yang tidak lengkap (*missing value*). Adapun jumlah data yang digunakan pada penelitian ini yaitu 379 data. Pengisian nilai pada *missing value* dapat dilihat pada gambar berikut ini:

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

Gambar 2. Import library pada Python

```
df.isnull().sum().sum()
10
```

Gambar 3. Total data missing value sebelum diberi nilai

### Filling Null Values

```
df2 = df.fillna(value = 0)
df2
```

Gambar 4. Pengisian data yang missing vale

```
df2.isnull().sum().sum()
0
```

Gambar 5. Total data missing value setelah diberi nilai

### Algoritma Naïve Bayes

*Naïve Bayes* adalah konsep dari probabilitas yang dipakai sebagai cara untuk menentukan kelompok kelas dokumen teks dan juga dapat mengolah data dalam jumlah besar serta menghasilkan akurasi yang cukup tinggi dan *naive bayes* memiliki tingkat akurasi yang lebih baik dibanding model *classifier* lainnya.

```
# Import Naive Bayes
from sklearn.naive_bayes import GaussianNB

# Memanggil fungsi Gaussian Naive Bayes
nbc = GaussianNB()
```

Gambar 6. Import Naïve Bayes

```
# Memasukkan data training pada fungsi kLasifikasi Naive Bayes
data_training = nbc.fit(x,y)

# Prediksi pada data training
y_predict = data_training.predict(x)
print(y_predict)

[0 1 0 1 0 0 1 1 0 0 0 0 0 0 1 1 1 0 0 1 0 0 1 0 0 0 0 0 1 0 1 0 1 1 1 0 0
 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 0 0 0 0 0 0
 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 0 1 1 1 1 0 0 1 1 0 1 0 0
 0 0 1 1 0 1 1 1 0 0 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 1 1 0
 0 1 0 1 1 1 0 1 0 0 0 0 1 1 0 0 0 1 1 0 1 1 1 1 0 0 0 1 0 1 0 1 0 1 0 1 1
 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 1 0 0 1 1 1 1 0 0 1 0 0 0 1 0 1 0 1 1
 1 1 1 1 1 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 0 1 1 1]
```

Gambar 7. Data training

```
# Prediksi data testing yang akan diprediksi
JENIS_KELAMIN = 1 #pria
STATUS_MAHASISWA = 0 #mahasiswa
UMUR = 25
STATUS_NIKAH = 1 #menikah
IPS1 = 3.12
IPS2 = 2.75
IPS3 = 2.05
IPS4 = 3.01
IPS5 = 2.56
IPS6 = 3.67
IPS7 = 2.75
IPS8 = 3.45
IPK = np.average ([IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPS8])
data_testing = [[JENIS_KELAMIN, STATUS_MAHASISWA, UMUR, STATUS_NIKAH, IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPS8, IPK]]
```

Gambar 8. Prediksi data testing

```
# Prediksi
y_pred = data_training.predict(data_testing)

C:\Users\ALIYA NURAINI\AppData\Local\Programs\Python\Python312\Lib\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature
ut GaussianNB was fitted with feature names
warnings.warn(

if y_pred == 0:
    hasil = 'Terlambat'
elif y_pred == 1:
    hasil = 'Tepat'
else:
    hasil = 'Error'
print('Hasil Prediksi Kelulusan Mahasiswa =', hasil)

Hasil Prediksi Kelulusan Mahasiswa = Terlambat
```

Gambar 9. Hasil dari prediksi data testing

## Evaluasi Hasil

Berikut adalah nilai akurasi model dan hasil penelitian:

```
# Akurasi model
from sklearn.metrics import accuracy_score
print('Nilai Akurasi = %0.2f' % accuracy_score(y, y_predict))

Nilai Akurasi = 0.89
```

Gambar 10. Akurasi model

```
# Hasil
from sklearn.metrics import classification_report
print(classification_report(y, y_predict))
```

	precision	recall	f1-score	support
0	0.88	0.85	0.87	163
1	0.89	0.91	0.90	216
accuracy			0.89	379
macro avg	0.89	0.88	0.88	379
weighted avg	0.89	0.89	0.89	379

Gambar 11. Hasil penelitian

## SIMPULAN

Penelitian prediksi tingkat kelulusan mahasiswa menggunakan 379 data. Atribut yang digunakan, antara lain: jenis kelamin, status mahasiswa, umur, status nikah, IPS, IPK,

dan status kelulusan. Tahapan yang dilakukan dalam penelitian, antara lain: pengumpulan data, data cleaning, data transformation, pengisian nilai pada *missing value*, klasifikasi dengan *naïve bayes*, evaluasi dan hasil. Hasil penelitian yang diperoleh yaitu akurasi = 89%, precision = 88% dari prediksi positif pada kelas 0 benar-benar positif, dan sekitar 89% dari prediksi positif pada kelas 1 benar-benar positif. dan recall = 85% dari kasus positif pada kelas 0 berhasil diidentifikasi, dan sekitar 91% dari kasus positif pada kelas 1 berhasil diidentifikasi. Hasil klasifikasi termasuk dalam kategori *good classification*.

### UCAPAN TERIMAKASIH

Terimakasih Kepada Universitas Multimedia Nusantara dan Universitas Muhammadiyah Tangerang yang telah memberikan support dalam melakukan penelitian ini, sehingga penelitian ini telah selesai dilaksanakan.

### DAFTAR PUSTAKA

- Abdurrohman S, Wibowo A. 2022. Klasifikasi Data Mining Untuk Memprediksi Status Penerimaan Di Perguruan Tinggi Negeri Bagi Lulusan Bimbel NF Dengan Algoritme Naive Bayes. *Semin. Nas. Mhs. Fak. Teknol. Inf.* 1(1):84–92.
- Apridiansyah Y, Veronika NDM, Putra ED. 2021. Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes. *JSAI (Journal Sci. Appl. Informatics)*. 4(2):236–247.doi:10.36085/jsai.v4i2.1701.
- Mayadewi P, Rosely E. 2020. Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Semin. Nas. Sist. Inf. Indones.*(November):329–334.
- Nafi'iyah N, Ahmad RA, Mujilawati S. 2020. Prediksi Nilai Calon Mahasiswa dengan Algoritma Backpropagation (Studi Kasus: Data Kaggle). *J. Nas. Komputasi dan Teknol. Inf.* 3(1):9–17.doi:10.32672/jnkti.v3i1.1945.
- Permana AA, Fahrezi MF, Kristiyanti DA, Sihotang M. 2021. Sentimen Analisis Opini Masyarakat Pada Media Sosial Twitter Terhadap Vaksin Berbayar Menggunakan Metode Naïve Bayes Classifier (Nbc). *J. Tek.* 10(2):113–121.doi:10.31000/jt.v10i2.5471.
- PERMANA AA, NOVIYANTO WA. 2023. Comparison of the Accuracy of the Lexicon-Based and Naive Bayes Classifier Methods To Public Opinions About Removing Masks on Social Media Twitter. *J. Theor. Appl. Inf. Technol.* 101(3):1174–1183.
- Steven C, Wella W. 2020. The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter. *IJNMT (International J. New Media Technol.* 7(2):102–110.doi:10.31937/ijnmt.v7i2.1732.
- Supangat, Sulistyawan MR. 2023. PEMODELAN PREDIKSI TINGKAT KELULUSAN MAHASISWA DENGAN PENDEKATAN ALGORITMA NAÏVE BAYES. *J. Inform. Polinema.* 9(4):405–414.doi:10.33795/jip.v9i4.1367.