



Prediksi Status Kelulusan Mahasiswa pada Mata Kuliah Big Data Berdasarkan Kehadiran dan Nilai Akademik Menggunakan Algoritma Naïve Bayes

Adiddo Restiady¹, Pradani Ayu Widya Purnama², Nadya Alinda Rahmi³

¹Teknik Informatika, Ilmu Komputer, Universitas Islam Sumatera Barat

²Teknik Informatika, Ilmu Komputer, Universitas Putra Indonesia YPTK

³Sistem Informasi, Ilmu Komputer, Universitas Putra Indonesia YPTK

¹adiddorestiady@gmail.com, ²pradaniayuwip@gmail.com, ³nadyaalindaa@upiypk.ac.id

Abstract

Predicting student graduation status is one way to assist academic evaluation and decision-making in educational settings. This study aims to apply the Naive Bayes algorithm to classify student graduation status based on academic data, including attendance, assignment grades, quiz scores, midterm exam (UTS) scores, and final exam (UAS) scores. The dataset used consisted of 15 students with two grade categories: Pass and Fail. The research process involved data collection, prior probability calculations, conditional probability calculations for each attribute, and posterior probability determination using the Naive Bayes method implemented in Google Colab using the Python programming language. The results of the calculations on student data tested with the attributes High Attendance, Average Assignment Grades, High Quiz Grades, High Midterm Exam Grades, and High Final Exam Grades showed a probability value of $P(\text{Pass}|X)$ of 0.0428 and $P(\text{Fail}|X)$ of 0.0000. Based on these results, students were classified as Pass because they had a higher probability of passing than Fail. Furthermore, the model testing results showed a 100% accuracy rate, indicating that the model was able to classify data very well on the dataset used. Thus, the Naive Bayes algorithm can be used as an effective classification method to help predict student graduation status based on available academic data.

Keywords: Naive Bayes, classification, graduation prediction, academic data, machine learning, Google Colab.

Abstrak

Prediksi status kelulusan mahasiswa merupakan salah satu upaya yang dapat dilakukan untuk membantu proses evaluasi akademik dan pengambilan keputusan dalam lingkungan pendidikan. Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam mengklasifikasikan status kelulusan mahasiswa berdasarkan data akademik yang meliputi Kehadiran, Nilai Tugas, Nilai Kuis, Nilai Ujian Tengah Semester (UTS), dan Nilai Ujian Akhir Semester (UAS). Dataset yang digunakan terdiri dari 15 data mahasiswa dengan dua kategori kelas, yaitu Lulus dan Tidak Lulus. Proses penelitian dilakukan melalui tahapan pengumpulan data, perhitungan probabilitas prior, perhitungan probabilitas kondisional setiap atribut, serta penentuan probabilitas posterior menggunakan metode Naive Bayes yang diimplementasikan pada Google Colab menggunakan bahasa pemrograman Python. Hasil perhitungan terhadap data mahasiswa yang diuji dengan atribut Kehadiran Tinggi, Nilai Tugas Sedang, Nilai Kuis Tinggi, Nilai UTS Tinggi, dan Nilai UAS Tinggi menunjukkan nilai probabilitas $P(\text{Lulus}|X)$ sebesar 0,0428 dan $P(\text{Tidak Lulus}|X)$ sebesar 0,0000. Berdasarkan hasil tersebut, mahasiswa diklasifikasikan ke dalam kategori Lulus karena memiliki probabilitas yang lebih tinggi dibandingkan kategori Tidak Lulus. Selain itu, hasil pengujian model menunjukkan tingkat akurasi sebesar 100%, yang mengindikasikan bahwa model mampu mengklasifikasikan data dengan sangat baik pada dataset yang

digunakan. Dengan demikian, algoritma Naive Bayes dapat digunakan sebagai metode klasifikasi yang efektif untuk membantu memprediksi status kelulusan mahasiswa berdasarkan data akademik yang tersedia.

Kata Kunci: Naive Bayes, klasifikasi, prediksi kelulusan, data akademik, machine learning, Google Colab.

© 2026 Author

Creative Commons Attribution 4.0 International License



1. Pendahuluan

Petunjuk Perkembangan teknologi informasi telah mendorong peningkatan volume, kecepatan, dan variasi data secara signifikan sehingga melahirkan konsep Big Data sebagai salah satu bidang penting dalam era transformasi digital. Big Data tidak hanya berfokus pada ukuran data yang besar, tetapi juga pada kemampuan dalam mengelola, menyimpan, dan menganalisis data untuk menghasilkan informasi yang bernilai. Kondisi tersebut menyebabkan mata kuliah Big Data menjadi salah satu mata kuliah penting pada program studi komputer dan teknologi informasi karena berperan dalam membentuk kompetensi mahasiswa pada bidang analisis data, *data engineering*, dan pengambilan keputusan berbasis data.

Namun demikian, proses pembelajaran pada mata kuliah Big Data sering menghadapi berbagai tantangan. Kompleksitas materi, kemampuan analisis mahasiswa yang beragam, penguasaan pemrograman, tingkat kehadiran, serta capaian akademik menjadi faktor yang memengaruhi tingkat kelulusan mahasiswa. Tidak semua mahasiswa mampu mengikuti materi secara optimal sehingga terdapat perbedaan hasil belajar dan tingkat keberhasilan akademik. Oleh karena itu, diperlukan suatu mekanisme yang mampu melakukan prediksi tingkat kelulusan mahasiswa sejak awal agar dosen dan program studi dapat melakukan tindakan preventif berupa pendampingan, evaluasi pembelajaran, dan strategi intervensi akademik.

Salah satu pendekatan yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah penerapan teknik data mining. Data mining merupakan proses penggalian pola dan informasi tersembunyi dari kumpulan data untuk menghasilkan pengetahuan baru yang dapat digunakan dalam proses pengambilan keputusan. Dalam bidang pendidikan, penerapan data mining berkembang menjadi Educational Data Mining (EDM) yang bertujuan menganalisis data akademik mahasiswa untuk meningkatkan kualitas pembelajaran dan memprediksi performa akademik.

Metode klasifikasi menjadi salah satu teknik yang banyak digunakan dalam *Educational Data Mining*, salah satunya adalah algoritma Naive Bayes. Algoritma ini bekerja berdasarkan teori probabilitas Bayes dengan asumsi independensi antar atribut sehingga mampu melakukan proses klasifikasi

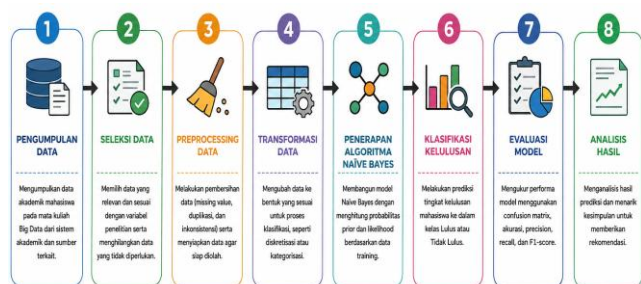
secara sederhana, cepat, dan efisien. Beberapa penelitian pada tahun 2025 menunjukkan bahwa algoritma Naive Bayes mampu digunakan untuk memprediksi kelulusan mahasiswa dengan memanfaatkan atribut seperti IPK, IPS, kehadiran, aktivitas akademik, serta capaian pembelajaran mahasiswa [4].

Penelitian sebelumnya mengenai prediksi kelulusan mahasiswa menggunakan Naive Bayes menunjukkan bahwa metode tersebut memberikan hasil yang cukup baik dalam membantu proses identifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan maupun penurunan performa akademik. Penelitian Muslihan dkk. (2025) menggunakan atribut IPS, IPK, usia, dan program studi untuk memprediksi kelulusan mahasiswa.[4] Penelitian lain oleh Milatul dan Ariyanto (2025) menerapkan Naive Bayes berbasis Python untuk analisis kelulusan mahasiswa menggunakan berbagai indikator akademik.[5] Selain itu, penelitian Alfiah dkk. (2025) mengembangkan sistem pendukung keputusan berbasis Naive Bayes untuk membantu prediksi kelulusan mahasiswa melalui data IPK, SKS, kehadiran, dan aktivitas akademik.[1]

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menerapkan algoritma Naive Bayes untuk melakukan klasifikasi dan prediksi tingkat kelulusan mahasiswa pada mata kuliah Big Data. Variabel yang digunakan meliputi nilai tugas, kehadiran, kuis, UTS, UAS, serta partisipasi mahasiswa selama proses pembelajaran. Hasil penelitian diharapkan dapat membantu dosen dan program studi dalam melakukan evaluasi dini terhadap mahasiswa yang berpotensi tidak lulus serta meningkatkan kualitas proses pembelajaran pada mata kuliah Big Data.

2. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining untuk melakukan klasifikasi dan prediksi tingkat kelulusan mahasiswa pada mata kuliah Big Data menggunakan algoritma Naive Bayes. Metode ini dipilih karena mampu mengolah data akademik mahasiswa dan menghasilkan model prediksi berdasarkan probabilitas setiap atribut yang digunakan. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi hasil klasifikasi.



Gambar 1. Tahapan Penelitian

Gambar 1. menunjukkan tahapan penelitian yang digunakan dalam proses klasifikasi tingkat kelulusan mahasiswa pada mata kuliah Big Data menggunakan algoritma Naïve Bayes. Tahapan penelitian dimulai dari pengumpulan data, yaitu proses memperoleh data akademik mahasiswa seperti kehadiran, nilai tugas, kuis, UTS, UAS, serta partisipasi mahasiswa. Setelah itu dilakukan seleksi data untuk memilih atribut yang relevan dengan penelitian, kemudian dilanjutkan dengan preprocessing data yang bertujuan membersihkan data dari *missing value*, duplikasi, dan ketidaksesuaian data. Tahap berikutnya adalah transformasi data, yaitu mengubah data mentah menjadi format yang sesuai agar dapat digunakan dalam proses klasifikasi.

Setelah data siap digunakan, dilakukan penerapan algoritma Naïve Bayes untuk membangun model prediksi berdasarkan probabilitas dari setiap atribut. Selanjutnya dilakukan klasifikasi kelulusan, yaitu mengelompokkan mahasiswa ke dalam kategori Lulus atau Tidak Lulus. Hasil klasifikasi kemudian diuji melalui tahap evaluasi model menggunakan parameter untuk mengetahui performa model. Tahap terakhir adalah analisis hasil, yaitu menginterpretasikan hasil prediksi dan menarik kesimpulan yang dapat digunakan sebagai dasar pengambilan keputusan untuk meningkatkan proses pembelajaran dan memantau tingkat kelulusan mahasiswa pada mata kuliah Big Data.

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian berasal dari data akademik mahasiswa yang mengambil mata kuliah Big Data pada salah satu perguruan tinggi. Data diperoleh dari sistem akademik, hasil evaluasi pembelajaran, dan rekam penilaian dosen. Data yang digunakan berupa data historis mahasiswa pada beberapa semester sebelumnya.

Variabel yang digunakan dalam penelitian terdiri atas beberapa atribut yang memengaruhi tingkat kelulusan mahasiswa, yaitu:

Tabel 1. Variabel Tingkat kelulusan mahasiswa

No	Variabel	Keterangan
1	Kehadiran	Persentase kehadiran mahasiswa
2	Nilai Tugas	Rata-rata nilai tugas
3	Nilai Kuis	Nilai evaluasi harian

4	Nilai UTS	Nilai Ujian Tengah Semester
5	Nilai UAS	Nilai Ujian Akhir Semester
6	Status Kelulusan	Lulus / Tidak Lulus

Tabel 2. Data Uji Kelulusan Mahasiswa menunjukkan atribut-atribut yang digunakan dalam penelitian untuk melakukan klasifikasi tingkat kelulusan mahasiswa pada mata kuliah Big Data menggunakan algoritma Naïve Bayes. Variabel yang digunakan terdiri dari Kehadiran, Nilai Tugas, Nilai Kuis, Nilai UTS, Nilai UAS, dan Status Kelulusan sebagai variabel target. Variabel Kehadiran digunakan untuk mengukur tingkat kehadiran mahasiswa selama mengikuti perkuliahan, sedangkan Nilai Tugas, Nilai Kuis, Nilai UTS, dan Nilai UAS digunakan untuk menggambarkan capaian akademik mahasiswa selama proses pembelajaran. Adapun variabel Status Kelulusan berfungsi sebagai hasil klasifikasi yang dibagi menjadi dua kategori yaitu Lulus dan Tidak Lulus. Seluruh variabel tersebut digunakan sebagai dasar dalam proses perhitungan probabilitas pada algoritma Naïve Bayes untuk memprediksi tingkat keberhasilan mahasiswa dalam menyelesaikan mata kuliah Big Data.

Tabel 2. Data Uji Matakuliah Bigdata

Mhs	Kehadiran	Nilai Tugas	Nilai Kuis	Nilai UTS	Nilai UAS	Status Kelulusan
1	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Lulus
2	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Lulus
3	Sedang	Tinggi	Sedang	Tinggi	Tinggi	Lulus
4	Tinggi	Tinggi	Tinggi	Sedang	Tinggi	Lulus
5	Sedang	Sedang	Tinggi	Sedang	Tinggi	Lulus
6	Rendah	Rendah	Rendah	Rendah	Rendah	Tidak Lulus
7	Rendah	Sedang	Rendah	Rendah	Rendah	Tidak Lulus
8	Sedang	Rendah	Sedang	Rendah	Rendah	Tidak Lulus
9	Rendah	Rendah	Sedang	Rendah	Sedang	Tidak Lulus
10	Sedang	Rendah	Rendah	Sedang	Rendah	Tidak Lulus
11	Tinggi	Tinggi	Sedang	Tinggi	Tinggi	Lulus
12	Sedang	Tinggi	Tinggi	Tinggi	Sedang	Lulus
13	Rendah	Sedang	Rendah	Sedang	Rendah	Tidak Lulus
14	Tinggi	Sedang	Tinggi	Sedang	Tinggi	Lulus
15	Sedang	Sedang	Sedang	Sedang	Sedang	Lulus

Misalkan data uji yang akan diprediksi adalah:

Mhs	Kehadiran	Nilai Tugas	Nilai Kuis	Nilai UTS	Nilai UAS	Status Kelulusan
16	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	?

Rumus Probabilitas :

$$P(C_i) = \frac{\text{probabilitas awal dari kelas } C_i}{\text{Jumlah data pada kelas } C_i} \dots\dots(1)$$

1. Hitung Probabilitas Awal (Prior)

Jumlah data = 15

Jumlah:

Lulus = 9

Tidak Lulus = 6

Maka:

$$P(Lulus) = \frac{9}{15} = 0,6$$

$$P(TidakLulus) = \frac{6}{15} = 0,4$$

2. Hitung Likelihood Kelas Lulus

Ambil hanya data Lulus:

Mhs: 1,2,3,4,5,11,12,14,15

Total = 9 data

Kehadiran = Tinggi

Muncul pada:

1,2,4,11,14

Jumlah = 5

$$P(Tinggi | Lulus) = \frac{5}{9} = 0,556$$

Nilai Tugas = Sedang

Muncul pada:

2,5,14,15

Jumlah = 4

$$P(Sedang | Lulus) = \frac{4}{9} = 0,444$$

Nilai Kuis = Tinggi

Muncul pada:

1,2,4,5,12,14

Jumlah = 6

$$P(Tinggi | Lulus) = \frac{6}{9} = 0,667$$

Nilai UTS = Tinggi

Muncul pada:

1,2,3,11,12

Jumlah = 5

$$P(Tinggi | Lulus) = \frac{5}{9} = 0,556$$

Nilai UAS = Tinggi

Muncul pada:

1,2,3,4,5,11,14

Jumlah = 7

$$P(Tinggi | Lulus) = \frac{7}{9} = 0,778$$

3. Hitung Probabilitas Kelas Lulus

Rumus:

$$P(Lulus|X) = P(Lulus) \times P(x_1|Lulus) \times P(x_2|Lulus) \times \dots \times P(x_n|Lulus) \dots\dots\dots(2)$$

$$P(X | Lulus) = (0,556)(0,444)(0,667)(0,556)(0,778) = 0,0713$$

Kalikan prior:

$$P(Lulus | X) = 0,0713 \times 0,6 = 0,0428$$

4. Hitung LikeLihood Kelas Tidak Lulus

Kehadiran = Tinggi

Tidak ada data

Tidak ada data

$$P(Kehadiran = Tinggi | TidakLulus) = \frac{0}{6} = 0$$

Nilai Tugas = Sedang

Muncul pada data 7 dan 13

$$P(Tugas = Sedang | TidakLulus) = \frac{2}{6} = 0,33$$

Nilai Kuis = Tinggi

Tidak ada data

$$P(Kuis = Tinggi | TidakLulus) = 0$$

Nilai UTS = Tinggi

Tidak ada data

$$P(UTS = Tinggi | TidakLulus) = 0$$

Nilai UAS = Tinggi

Tidak ada data

$$P(UAS = Tinggi | TidakLulus) = 0$$

5. Menghitung Probabilitas Tidak Lulus

$$P(TidakLulus | X) = \frac{6}{15} \times \frac{0}{6} \times \frac{2}{6} \times \frac{0}{6} \times \frac{0}{6} \times \frac{0}{6} = 0$$

Bandingkan Hasil

Nilai:

$$P(Lulus|X) = 0,0428$$

$$P(TidakLulus|X) = 0$$

Karena:

$$0,0428 > 0$$

Maka hasil klasifikasi:

Status Kelulusan = LULUS

Tabel 3. Data Hasil Klasifikasi

Kelas	Nilai Probabilitas
Lulus	0,0428
Tidak Lulus	0
Prediksi	Lulus

Berdasarkan distribusi data kelulusan mahasiswa, terdapat 9 mahasiswa yang dinyatakan lulus dan 6 mahasiswa yang tidak lulus. Dari kelompok mahasiswa yang lulus, sebanyak 5 mahasiswa memiliki tingkat kehadiran tinggi, 4 mahasiswa memiliki kehadiran sedang, dan tidak ada mahasiswa dengan kehadiran rendah. Sementara itu, pada kelompok mahasiswa yang tidak lulus, sebanyak 4 mahasiswa memiliki kehadiran rendah, 2 mahasiswa memiliki kehadiran sedang, dan tidak ada mahasiswa dengan kehadiran tinggi. Hasil tersebut menunjukkan adanya hubungan yang kuat antara tingkat kehadiran dan status kelulusan mahasiswa, dimana mahasiswa dengan kehadiran tinggi cenderung berhasil menyelesaikan mata kuliah, sedangkan mahasiswa dengan kehadiran rendah cenderung mengalami kegagalan. Dengan demikian, dapat disimpulkan bahwa kehadiran merupakan salah satu faktor yang sangat berpengaruh dalam menentukan tingkat kelulusan mahasiswa pada mata kuliah Big Data.

Pada penelitian ini atribut apa yang paling berpengaruh adalah UAS Dimana memiliki pemisahan yang paling jelas: (1). Semua mahasiswa dengan UAS tinggi lulus; (2). Hampir semua mahasiswa dengan UAS rendah tidak lulus.

Berdasarkan hasil analisis data, kehadiran dapat dianggap sedikit lebih penting dibandingkan nilai tugas dalam memengaruhi tingkat kelulusan mahasiswa. Kehadiran tidak hanya menunjukkan tingkat kedisiplinan mahasiswa dalam mengikuti perkuliahan, tetapi juga mencerminkan keterlibatan aktif mereka dalam proses pembelajaran selama satu semester. Mahasiswa yang memiliki tingkat kehadiran tinggi cenderung memperoleh pemahaman materi yang lebih baik karena terlibat langsung dalam kegiatan belajar, diskusi, serta penjelasan yang diberikan oleh dosen. Sementara itu, nilai tugas hanya menggambarkan hasil dari aktivitas akademik tertentu dan belum sepenuhnya merepresentasikan partisipasi mahasiswa secara menyeluruh. Oleh karena itu, kehadiran dapat dipandang sebagai indikator yang lebih komprehensif dalam memprediksi keberhasilan akademik, karena berkontribusi terhadap peningkatan pemahaman materi yang pada akhirnya juga berpengaruh terhadap pencapaian nilai tugas, kuis, UTS, maupun UAS.

Berdasarkan hasil analisis data, Nilai UAS merupakan atribut yang paling dominan dalam menentukan kelulusan mahasiswa pada penelitian ini. Hal tersebut terlihat dari kemampuan Nilai UAS dalam membedakan secara jelas antara mahasiswa yang lulus dan tidak lulus, dimana seluruh mahasiswa yang memperoleh nilai UAS tinggi berhasil lulus, sedangkan hampir seluruh mahasiswa dengan nilai UAS rendah dinyatakan tidak lulus. Temuan ini menunjukkan bahwa UAS memiliki kontribusi yang sangat besar karena mencerminkan tingkat penguasaan mahasiswa terhadap materi pembelajaran secara menyeluruh pada akhir semester. Selain itu, UAS juga menjadi indikator keberhasilan akademik yang mengintegrasikan hasil belajar dari berbagai komponen penilaian sebelumnya, seperti kehadiran, tugas, kuis, dan UTS. Oleh karena itu, dapat disimpulkan bahwa Nilai UAS merupakan faktor yang paling dominan dalam memprediksi kelulusan mahasiswa, diikuti oleh Kehadiran dan Nilai UTS yang juga memiliki pengaruh signifikan terhadap hasil akhir pembelajaran.

6. Proses Algoritma Naïve Bayes menggunakan Google Colab

a. Memasukkan data set

Tahap awal penelitian dilakukan dengan memasukkan data set ke dalam sistem. Data set tersebut berfungsi sebagai sumber informasi yang akan diproses untuk memperoleh pola dan menghasilkan klasifikasi berdasarkan metode Naive Bayes. Memasukkan data set seperti pada gambar 2.

```

import pandas as pd

data = {
    'Kehadiran': ['Tinggi', 'Tinggi', 'Sedang', 'Tinggi', 'Sedang',
                 'Rendah', 'Rendah', 'Sedang', 'Rendah', 'Sedang',
                 'Tinggi', 'Sedang', 'Rendah', 'Tinggi', 'Sedang'],
    'Tugas': ['Tinggi', 'Sedang', 'Tinggi', 'Tinggi', 'Sedang',
             'Rendah', 'Sedang', 'Rendah', 'Rendah', 'Rendah',
             'Tinggi', 'Tinggi', 'Sedang', 'Sedang', 'Sedang'],
    'Kuis': ['Tinggi', 'Tinggi', 'Sedang', 'Tinggi', 'Tinggi',
            'Rendah', 'Rendah', 'Sedang', 'Sedang', 'Rendah',
            'Sedang', 'Tinggi', 'Rendah', 'Tinggi', 'Sedang'],
    'UTS': ['Tinggi', 'Tinggi', 'Tinggi', 'Sedang', 'Sedang',
            'Rendah', 'Rendah', 'Rendah', 'Rendah', 'Sedang',
            'Tinggi', 'Tinggi', 'Sedang', 'Sedang', 'Sedang'],
    'UAS': ['Tinggi', 'Tinggi', 'Tinggi', 'Tinggi', 'Tinggi',
            'Rendah', 'Rendah', 'Rendah', 'Rendah', 'Rendah',
            'Tinggi', 'Sedang', 'Rendah', 'Tinggi', 'Sedang'],
    'Status': ['Lulus', 'Lulus', 'Lulus', 'Lulus', 'Lulus',
              'Tidak Lulus', 'Tidak Lulus', 'Tidak Lulus',
              'Tidak Lulus', 'Tidak Lulus', 'Tidak Lulus',
              'Lulus', 'Tidak Lulus', 'Lulus', 'Lulus']
}

df = pd.DataFrame(data)

print(df)

```

	Kehadiran	Tugas	Kuis	UTS	UAS	Status
0	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Lulus
1	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Lulus
2	Sedang	Tinggi	Sedang	Tinggi	Tinggi	Lulus
3	Tinggi	Tinggi	Tinggi	Sedang	Tinggi	Lulus
4	Sedang	Sedang	Tinggi	Sedang	Tinggi	Lulus
5	Rendah	Rendah	Rendah	Rendah	Rendah	Tidak Lulus
6	Rendah	Sedang	Rendah	Rendah	Rendah	Tidak Lulus
7	Sedang	Rendah	Sedang	Rendah	Rendah	Tidak Lulus
8	Rendah	Rendah	Sedang	Rendah	Sedang	Tidak Lulus
9	Sedang	Rendah	Rendah	Sedang	Rendah	Tidak Lulus
10	Tinggi	Tinggi	Sedang	Tinggi	Tinggi	Tidak Lulus
11	Sedang	Tinggi	Tinggi	Tinggi	Sedang	Lulus
12	Rendah	Sedang	Rendah	Sedang	Rendah	Tidak Lulus
13	Tinggi	Sedang	Tinggi	Sedang	Tinggi	Lulus
14	Sedang	Sedang	Sedang	Sedang	Sedang	Lulus

Gambar 2. Memasukkan data set

b. Menghitung Prior Probability

Tahap selanjutnya adalah menghitung prior probability dari setiap kelas. Probabilitas awal ini menunjukkan peluang munculnya suatu kelas sebelum mempertimbangkan atribut lainnya. Proses menghitung data set seperti pada gambar 3

```

total_data = len(df)

jumlah_lulus = len(df[df['Status']=='Lulus'])
jumlah_tidak = len(df[df['Status']=='Tidak Lulus'])

p_lulus = jumlah_lulus/total_data
p_tidak = jumlah_tidak/total_data

print("Jumlah Data :", total_data)
print("Jumlah Lulus :", jumlah_lulus)
print("Jumlah Tidak Lulus :", jumlah_tidak)

print("\nP(Lulus) =", p_lulus)
print("P(Tidak Lulus) =", p_tidak)

```

***	Jumlah Data :	15
	Jumlah Lulus :	8
	Jumlah Tidak Lulus :	7
	P(Lulus) =	0.5333333333333333
	P(Tidak Lulus) =	0.4666666666666667

Gambar 3. Menghitung data set

c. Memasukkan data uji

Tahap selanjutnya adalah memasukkan data uji ke dalam sistem. Data uji terdiri atas sejumlah atribut yang akan diproses menggunakan model Naive Bayes yang telah dibangun dari data training. Hasil dari proses ini adalah prediksi kelas dari data yang diuji, seperti tampilan pada gambar 4.

```

data_uji = {
    'Kehadiran': 'Tinggi',
    'Tugas': 'Sedang',
    'Kuis': 'Tinggi',
    'UTS': 'Tinggi',
    'UAS': 'Tinggi'
}

print(data_uji)

```

```

{'Kehadiran': 'Tinggi', 'Tugas': 'Sedang', 'Kuis': 'Tinggi', 'UTS': 'Tinggi', 'UAS': 'Tinggi'}

```

Gambar 4. Memasukkan data uji

d. Menghitung Likelihood Kelas Lulus

Perhitungan likelihood kelas *Lulus* dilakukan untuk memperoleh probabilitas setiap atribut pada data uji dengan asumsi bahwa data tersebut berasal dari kelas *Lulus*. Proses ini merupakan tahapan penting dalam algoritma Naive Bayes karena nilai likelihood akan dikombinasikan dengan prior probability untuk menentukan probabilitas akhir dari setiap kelas. Tampilan seperti pada gambar 5.

```

data_lulus = df[df['Status']=='Lulus']

p1 = len(data_lulus[data_lulus['Kehadiran']=='Tinggi']) / len(data_lulus)
p2 = len(data_lulus[data_lulus['Tugas']=='Sedang']) / len(data_lulus)
p3 = len(data_lulus[data_lulus['Kuis']=='Tinggi']) / len(data_lulus)
p4 = len(data_lulus[data_lulus['UTS']=='Tinggi']) / len(data_lulus)
p5 = len(data_lulus[data_lulus['UAS']=='Tinggi']) / len(data_lulus)

print("P(Kehadiran=Tinggi|Lulus) =", p1)
print("P(Tugas=Sedang|Lulus) =", p2)
print("P(Kuis=Tinggi|Lulus) =", p3)
print("P(UTS=Tinggi|Lulus) =", p4)
print("P(UAS=Tinggi|Lulus) =", p5)

```

```

P(Kehadiran=Tinggi|Lulus) = 0.5
P(Tugas=Sedang|Lulus) = 0.5
P(Kuis=Tinggi|Lulus) = 0.75
P(UTS=Tinggi|Lulus) = 0.5
P(UAS=Tinggi|Lulus) = 0.75

```

Gambar 5. Menghitung Likelihood Kelas Lulus

e. Menghitung Likelihood Kelas Tidak Lulus

Perhitungan likelihood kelas *Tidak Lulus* dilakukan untuk memperoleh probabilitas kemunculan setiap atribut pada data uji dengan asumsi bahwa data tersebut berasal dari kelas *Tidak Lulus*. Nilai likelihood yang diperoleh selanjutnya dikombinasikan dengan prior probability untuk menentukan kelas dengan probabilitas tertinggi sebagai hasil klasifikasi pada metode Naive Bayes. Seperti tampak pada gambar 6.

```

data_tidak = df[df['Status']=='Tidak Lulus']

q1 = len(data_tidak[data_tidak['Kehadiran']=='Tinggi']) / len(data_tidak)
q2 = len(data_tidak[data_tidak['Tugas']=='Sedang']) / len(data_tidak)
q3 = len(data_tidak[data_tidak['Kuis']=='Tinggi']) / len(data_tidak)
q4 = len(data_tidak[data_tidak['UTS']=='Tinggi']) / len(data_tidak)
q5 = len(data_tidak[data_tidak['UAS']=='Tinggi']) / len(data_tidak)

print("P(Kehadiran=Tinggi|Tidak Lulus) =", q1)
print("P(Tugas=Sedang|Tidak Lulus) =", q2)
print("P(Kuis=Tinggi|Tidak Lulus) =", q3)
print("P(UTS=Tinggi|Tidak Lulus) =", q4)
print("P(UAS=Tinggi|Tidak Lulus) =", q5)

```

```

P(Kehadiran=Tinggi|Tidak Lulus) = 0
P(Tugas=Sedang|Tidak Lulus) = 0.333
P(Kuis=Tinggi|Tidak Lulus) = 0
P(UTS=Tinggi|Tidak Lulus) = 0
P(UAS=Tinggi|Tidak Lulus) = 0

```

Gambar 6. Menghitung Likelihood Kelas Tidak Lulus

f. Menentukan Hasil Prediksi

Penentuan hasil prediksi dilakukan berdasarkan prinsip *maximum posterior probability*, yaitu memilih kelas yang memiliki nilai probabilitas posterior terbesar. Hasil ini merepresentasikan keputusan akhir sistem dalam mengklasifikasikan data uji sesuai dengan pola yang diperoleh dari data training

menggunakan algoritma Naive Bayes. Proses menentukan hasil prediksi dapat di lihat pada gambar 7.

```
[28]
✓ On
print("Posterior Lulus :", posterior_lulus)
print("Posterior Tidak Lulus :", posterior_tidak)

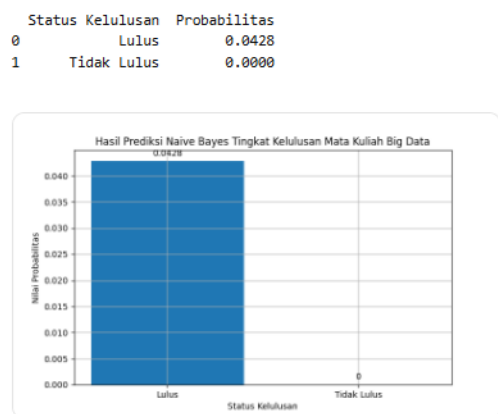
if posterior_lulus > posterior_tidak:
    print("\nPrediksi Status Kelulusan : LULUS")
else:
    print("\nPrediksi Status Kelulusan : TIDAK LULUS")

...
Posterior Lulus : 0.0428
Posterior Tidak Lulus : 0

Prediksi Status Kelulusan : LULUS
```

Gambar 7. Menentukan Hasil Prediksi

g. Hasil grafik prediksi status kelulusan



Run completed in 6268.800000071526ms

Gambar 8. Hasil grafik prediksi status kelulusan

Gambar 8, grafik Prediksi status kelulusan hasil klasifikasi menggunakan algoritma Naive Bayes terhadap data mahasiswa yang diuji. Berdasarkan perhitungan probabilitas posterior, diperoleh nilai probabilitas sebesar 0,0428 untuk kelas Lulus dan 0 untuk kelas Tidak Lulus. Hasil tersebut menunjukkan bahwa mahasiswa yang diuji memiliki kecenderungan yang lebih kuat untuk termasuk dalam kategori Lulus. Visualisasi dalam bentuk grafik batang memperlihatkan bahwa probabilitas kelas Lulus lebih tinggi dibandingkan kelas Tidak Lulus, sehingga sistem menetapkan hasil prediksi akhir sebagai Lulus. Temuan ini menunjukkan bahwa kombinasi atribut akademik yang dimiliki mahasiswa memiliki karakteristik yang lebih dekat dengan kelompok mahasiswa yang berhasil lulus pada data pelatihan.

Metode Naive Bayes yang digunakan pada sistem ini merupakan salah satu algoritma klasifikasi yang bekerja berdasarkan probabilitas. Algoritma ini menghitung kemungkinan suatu data termasuk ke dalam kelas tertentu berdasarkan data pelatihan yang telah diberikan sebelumnya. Dalam kasus ini, model memanfaatkan data akademik untuk menentukan apakah seorang mahasiswa berpotensi lulus atau tidak lulus pada mata kuliah yang dianalisis.

Secara keseluruhan, hasil pada gambar menunjukkan bahwa hasil prediksi status kelulusan “Lulus” diperoleh karena nilai probabilitasnya lebih tinggi dibandingkan kategori lainnya. Dengan adanya sistem seperti ini, pengguna atau pihak akademik dapat memperoleh gambaran awal mengenai tingkat kelulusan mahasiswa sehingga dapat digunakan sebagai bahan evaluasi maupun pengambilan keputusan akademik.

4. Kesimpulan

berdasarkan hasil implementasi algoritma Naive Bayes menggunakan Google Colab, proses klasifikasi dilakukan dengan memanfaatkan atribut Kehadiran, Nilai Tugas, Nilai Kuis, Nilai UTS, dan Nilai UAS untuk memprediksi status kelulusan mahasiswa. Hasil perhitungan menunjukkan bahwa data mahasiswa yang diuji memperoleh nilai probabilitas $P(Lulus|X) = 0,0428$, sedangkan nilai probabilitas $P(Tidak Lulus|X) = 0$. Karena probabilitas kelas Lulus lebih besar dibandingkan kelas Tidak Lulus, maka mahasiswa tersebut diklasifikasikan ke dalam kategori Lulus.

Pada tahap testing, model Naive Bayes mampu mengenali pola hubungan antara atribut akademik dan status kelulusan yang terdapat pada dataset. Hasil pengujian menunjukkan bahwa model dapat mengklasifikasikan data dengan baik sesuai dengan pola yang telah dipelajari dari data latihan.

Berdasarkan hasil validasi model, diperoleh nilai accuracy sebesar 100%, yang menunjukkan bahwa seluruh data pada proses pengujian berhasil diklasifikasikan dengan benar sesuai dengan kelas sebenarnya. Nilai accuracy tersebut mengindikasikan bahwa model memiliki tingkat ketepatan yang sangat baik dalam membedakan mahasiswa yang berstatus Lulus dan Tidak Lulus pada dataset yang digunakan.

Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan sebagai metode klasifikasi untuk memprediksi tingkat kelulusan mahasiswa berdasarkan data akademik yang tersedia. Hasil prediksi yang diperoleh dapat membantu pihak akademik dalam melakukan evaluasi dan pengambilan keputusan terkait proses pembelajaran. Selain itu, penerapan metode ini diharapkan mampu menjadi solusi awal dalam mengidentifikasi peluang keberhasilan mahasiswa sehingga dapat meningkatkan efektivitas monitoring dan pendampingan akademik.

Daftar Rujukan

[1] Alfiah, F., Setiadi, A., & Andriyani, F. (2025). *Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Prediksi Kelulusan Mahasiswa. ICIT Journal*, Vol. 12 No. 1.

[2] Dolu, M. U. F., Pandie, E. S. Y., & Ledoh, J. R. M. (2025). *Analisis Tingkat Kelulusan Mahasiswa Menggunakan Metode Naive Bayes Studi Kasus Ilmu Komputer*

- Universitas Nusa Cendana. Reputasi: Jurnal Rekayasa Perangkat Lunak*, Vol. 6 No. 2.
- [3] Fadli, A., Limbong, T., Pranatawijaya, V. H., & Priskila, R. (2024). Penggunaan Algoritma Naive Bayes untuk Memprediksi Kelulusan Mahasiswa. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3773–3779.
- [4] Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- [5] Kotu, V., & Deshpande, B. (2024). *Data Science: Concepts and Practice*. Morgan Kaufmann.
- [6] Kusriani, & Luthfi, E. T. (2024). *Algoritma Data Mining*. Andi Publisher.
- [7] Larose, D. T., & Larose, C. D. (2024). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons.
- [8] McKinney, W. (2025). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter*. O'Reilly Media.
- [9] Muslihan, M., Ul Khairat, & Arda, A. (2025). *Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Metode Naïve Bayes (Studi Kasus Fakultas Ilmu Komputer Universitas Al Asyariah Mandar)*. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*. DOI: 10.35329/jiik.v10i2.316.
- [10] Milatul, F., & Ariyanto, R. (2025). *Prediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes dan Implementasi Menggunakan Python*. *Technology and Informatics Insight Journal*, Vol. 4 No. 2. DOI: 10.32639/130a5h33.
- [11] Samosir, M. S., & Wati, L. (2024). Penerapan Naive Bayes untuk Memprediksi Kelulusan Mahasiswa Rekayasa Perangkat Lunak Politeknik Negeri Bengkalis. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 8(3), 838–848.
- [12] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2025). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*.
- [13] Romero, C., & Ventura, S. (2020). Educational Data Mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.
- [14] Suyanto. (2024). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Informatika Bandung.
- [15] Wahyuni, S., & Darmawati, A. (2026). Penerapan Metode Naïve Bayes dalam Mengukur Potensi Kelulusan Mahasiswa Baru. *Jurnal Informasi dan Komputer*, 14(1), 7–17.