



## Analisis Kinerja Algoritma *Naive Bayes* dalam Memprediksi Kelulusan Mahasiswa Menggunakan *Python*

Maulana Fansyuri<sup>1</sup>, Devi Yunita<sup>2</sup>

<sup>12</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang  
[dosen02359@unpam.ac.id](mailto:dosen02359@unpam.ac.id) <sup>2</sup>[dosen00846@unpam.ac.id](mailto:dosen00846@unpam.ac.id)

### Abstract

*The low and varied graduation rates in universities often give the challenges in maintaining academic quality, guiding students effectively and preventing delayed completion of study. Therefore, a predictive approach is required to detect potential graduation risks early. This study aims to analyze the performance of the Naive Bayes algorithm in predicting student graduation based on academic and non-academic data by using the Python programming language. The dataset includes variables such as grade point average, total earned credits, attendance rate and graduation status. The research stages involve data cleaning, categorical encoding, dataset splitting into training and testing subsets and also model training by using Naive Bayes. Model evaluation was conducted using confusion matrix, accuracy, precision, recall and F1-score. The results show that Naive Bayes achieves an accuracy of 85% with consistent value of precision, recall and F1-score indicates its reliability in distinguishing between students who graduate on time and those who do not. These findings highlight the potential of Naive Bayes as a baseline model to support academic decision-making and Early Warning systems and future research may improve its performance through feature expansion, class imbalance handling, or comparisons with other machine learning algorithms.*

**Keywords:** *Naive Bayes, Student Graduation Prediction, Machine learning, Python, Classification*

### Abstrak

Tingkat kelulusan mahasiswa sering menjadi tantangan bagi perguruan tinggi, terutama dalam upaya menjaga mutu akademik, meningkatkan efektivitas pembimbingan, serta meminimalkan risiko keterlambatan studi. Oleh karena itu, diperlukan suatu metode prediksi yang mampu mengidentifikasi potensi ketidakhadiran sejak dini. Penelitian ini bertujuan menganalisis kinerja algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa berbasis data akademik dan non-akademik dengan menggunakan bahasa pemrograman *Python*. Dataset penelitian mencakup variabel indeks prestasi, jumlah SKS, tingkat kehadiran, dan status kelulusan. Tahapan penelitian meliputi pembersihan data, penyandian variabel kategorikal, pembagian data latih serta data uji, lalu pelatihan model *Naive Bayes*. Evaluasi performa dilakukan menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model *Naive Bayes* mampu memprediksi kelulusan dengan akurasi 85%, serta nilai *precision*, *recall*, dan *F1-score* yang stabil sehingga menegaskan kemampuan model dalam membedakan mahasiswa lulus dan tidak lulus. Temuan ini berkontribusi sebagai acuan pengembangan *Early Warning system* dalam pengambilan keputusan akademik, serta dapat ditingkatkan pada penelitian berikutnya melalui penambahan fitur, penanganan ketidakseimbangan data, maupun perbandingan dengan algoritma lain

**Kata kunci:** *Naive Bayes, Prediksi Kelulusan Mahasiswa, Machine learning, Python, Klasifikasi.*



## 1. Pendahuluan

Perkembangan teknologi informasi dalam beberapa tahun terakhir telah mendorong peningkatan pemanfaatan data dalam sektor pendidikan tinggi, khususnya dalam proses pengambilan keputusan berbasis analitik untuk mendukung peningkatan mutu akademik. Perguruan tinggi kini tidak hanya menjadikan data akademik sebagai arsip administratif, tetapi juga memanfaatkannya untuk tujuan *prediktif* guna menyusun strategi pembinaan belajar yang lebih tepat sasaran. Salah satu isu penting yang menjadi perhatian adalah kemampuan untuk memprediksi kelulusan mahasiswa secara lebih dini sebagai upaya pencegahan keterlambatan studi dan penurunan performa akademik [1]. Pendekatan manual yang selama ini diterapkan terbukti memerlukan waktu panjang, tidak efisien, serta memiliki tingkat akurasi yang sulit diukur, sehingga diperlukan solusi berbasis *machine learning* yang lebih adaptif dan terukur.

Penelitian terdahulu menunjukkan bahwa algoritma *supervised learning* banyak digunakan dalam analisis data akademik karena kemampuannya dalam mengolah data historis menjadi pola *prediktif*. Di antara berbagai algoritma klasifikasi, *Naive Bayes* menjadi salah satu metode yang populer karena sederhana, efisien, serta mampu menangani data kategorikal maupun numerik dalam jumlah besar [2]. Studi oleh Imanuel dkk [3] melaporkan bahwa *Naive Bayes* dapat memprediksi keberhasilan akademik mahasiswa dengan akurasi di atas 80%. Sementara itu, penelitian lain dalam sistem *Early Warning* menggunakan *Naive Bayes* untuk mendeteksi potensi keterlambatan studi dan menghasilkan performa prediksi yang stabil pada berbagai variasi atribut akademik [4]. Temuan-temuan tersebut memperkuat potensi *Naive Bayes* sebagai alat prediksi akademik yang efektif dalam mendukung pengambilan keputusan institusi pendidikan.

Namun demikian, beberapa penelitian sebelumnya masih memiliki keterbatasan yang menjadi ruang kontribusi penelitian ini. Sebagian besar studi lebih berfokus pada prediksi IPK atau performa belajar secara umum, bukan pada prediksi kelulusan mahasiswa sebagai indikator akhir capaian studi [5][6]. *Naive Bayes* pada beberapa penelitian ditempatkan sebagai model pembandingan tanpa analisis mendalam terhadap kinerjanya menggunakan metrik evaluasi lengkap seperti *accuracy*, *precision*, *recall*, dan *f1-score* berbasis *confusion matrix*. Belum banyak studi yang mengkaji performa *Naive Bayes* secara spesifik dalam konteks prediksi kelulusan dengan mempertimbangkan karakteristik dataset akademik multidimensi terkini. Gap penelitian ini menunjukkan adanya kebutuhan untuk melakukan evaluasi komprehensif terhadap kinerja *Naive Bayes* pada prediksi kelulusan mahasiswa menggunakan data akademik aktual agar kebaruan dapat ditunjukkan secara jelas.

Berdasarkan adanya celah penelitian tersebut, studi ini berfokus pada evaluasi kinerja algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa dengan memanfaatkan variabel-variabel akademik seperti kehadiran, nilai mata kuliah, indeks prestasi, serta indikator evaluasi pembelajaran lainnya. Analisis dilakukan menggunakan pendekatan komputasi berbasis *Python* untuk memperoleh performa model melalui metrik evaluasi prediksi secara kuantitatif. Dari latar belakang tersebut, penulis membuat rumusan yang diangkat dalam penelitian ini yaitu, Seberapa baik performa algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa berdasarkan data akademik?

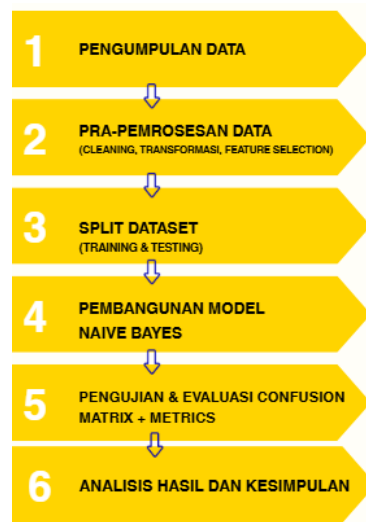
Adapun tujuan penelitian ini adalah untuk menganalisis performa algoritma *Naive Bayes* dalam prediksi kelulusan mahasiswa, mengevaluasi hasil klasifikasi menggunakan *confusion matrix* dan metrik turunannya (*accuracy*, *precision*, *recall*, *f1-score*), serta memberikan insight bagi institusi pendidikan mengenai potensi penggunaan model *prediktif* sebagai pendukung kebijakan pembinaan akademik.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional untuk menganalisis kinerja algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa berbasis data akademik.

Seluruh proses eksperimen dilakukan menggunakan bahasa pemrograman *Python* dengan dukungan *library* *scikit-learn*, *pandas*, dan *numpy* sebagai alat pemrosesan data serta pembangunan model klasifikasi.

Penelitian ini dilaksanakan melalui beberapa tahapan yang dirancang secara sistematis sebagaimana ditunjukkan pada gambar 1 dibawah ini.



Gambar 1. Desain Alur Penelitian

Tahapan tersebut meliputi:

1. **Pengumpulan Data.** Data akademik mahasiswa diperoleh dari basis data institusi pendidikan yang memuat variabel seperti kehadiran, nilai mata kuliah, IP, serta indikator evaluasi akademik lainnya.
2. **Pra-pemrosesan Data (Preprocessing).** Tahapan ini meliputi pembersihan data (*handling missing values*), normalisasi/standarisasi atribut bila diperlukan, transformasi tipe data, dan pemilihan fitur (*feature selection*).
3. **Pembagian Dataset.** Data dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) dengan perbandingan tertentu (misal 80:20) untuk memastikan evaluasi model lebih objektif.
4. **Pembangunan Model dengan Algoritma Naive Bayes.** Model *Naive Bayes* dibangun menggunakan *training data* untuk menghasilkan model klasifikasi kelulusan mahasiswa.
5. **Pengujian dan Evaluasi Model.** Model diuji menggunakan *testing data* dan kinerja dievaluasi berdasarkan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *f1-score*.
6. **Analisis Hasil.** Hasil prediksi dianalisis untuk melihat performa model serta potensinya sebagai pendukung keputusan akademik.

## 2.1 Dataset dan Sumber Data

Dataset yang digunakan merupakan data akademik mahasiswa pada salah satu program studi dengan total 1600 baris data dan 4 variabel prediktor, yang mencakup indikator seperti IPK setiap semesternya. Variabel target berupa status kelulusan mahasiswa (Lulus / Tidak Lulus). Data yang tidak lengkap ditangani melalui *handling missing values* menggunakan metode *mean imputation* untuk data numerik dan *mode imputation* untuk data kategorikal.

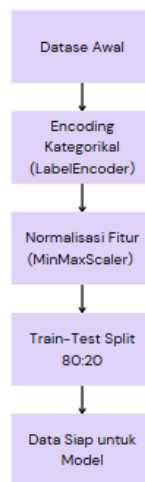
Tabel 1. Dataset Penelitian

No	IPK1	IPK2	IPK3	IPK4	Lulus Tepat Waktu
1	2.3	1.97	1.8	1.56	Tidak
2	1.81	1.68	1.57	1.86	Tidak
3	3.07	3	2.75	3.21	Tidak
4	2.71	2.33	2.61	1.98	Tidak
5	3.17	3.02	3.28	2.96	Tidak

6	3.16	3.45	3.02	3.06	Tidak
7	2.72	2.5	2.92	3	Tidak
8	2.97	3.27	2.9	2.83	Tidak
9	2.72	2.61	2.64	2.46	Tidak
10	2.78	2.85	3.08	3.35	Tidak
11	2.5	1.56	3.17	3.33	Tidak
12	2.19	2.92	2.54	2.67	Tidak
13	3.43	3.18	3.41	3.45	Ya
14	3.14	2.75	3.01	3.38	Tidak
15	3.04	2.95	3.14	2.98	Ya
16	2.83	2.56	3	2.79	Ya
17	2.6	2.72	3.08	2.54	Ya
18	2.98	3.02	2.98	2.9	Ya
19	3.1	3.34	3.38	3.42	Ya
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
1598	3.16	3.21	2.92	2.98	Tidak
1599	3.11	3.46	3.25	3.16	Ya
1600	3.28	3.48	3.42	3.39	Ya

## 2.2 Pra-Pemrosesan Data

Sebelum proses pelatihan model, data dipersiapkan melalui beberapa seperti pada gambar dibawah ini:



Gambar 2. Langkah-langkah pra-proses data

Langkah tersebut meliputi :

1. **Encoding variabel kategorikal**, Seluruh fitur *non-numerik* dikonversi menjadi nilai numerik menggunakan *LabelEncoder* agar dapat dikenali oleh model dalam proses klasifikasi.

2. **Normalisasi fitur numerik**, Normalisasi dilakukan menggunakan *MinMaxScaler* untuk menyamakan rentang nilai antar variabel, sehingga tidak terjadi dominasi pada fitur tertentu dalam proses perhitungan.
3. **Pembagian dataset (*train-test split*)**, Dataset dipisahkan menjadi **80% data pelatihan (*training*)** dan **20% data pengujian (*testing*)**. Pembagian ini bertujuan memastikan evaluasi model dilakukan secara objektif menggunakan data yang belum pernah dilatih sebelumnya.

Langkah-langkah tersebut merupakan tahapan standar dalam *data preprocessing* pada penelitian berbasis *machine learning*, tanpa dilakukan modifikasi konsep dasar.

### 2.3 Pembangunan Model

Algoritma yang digunakan adalah *Naive Bayes*, dipilih sebagai representasi metode probabilistik untuk klasifikasi dengan asumsi independensi antar fitur [7]. Pelatihan model dilakukan pada data pelatihan dengan ukuran 600 sampel, sementara 1000 sampel lainnya digunakan sebagai data pengujian. Proses pembangunan model dilakukan melalui tiga tahapan berulang (replikasi trial) untuk memastikan kestabilan hasil prediksi dan meminimalkan variabilitas performa [8].

### 2.4 Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan metrik umum pada sistem klasifikasi, yaitu *confusion matrix*, *accuracy*, *precision*, *recall*, dan *f1-score* [9]. *Confusion matrix* digunakan untuk menunjukkan distribusi prediksi benar dan salah pada setiap kelas sehingga dapat menggambarkan kesalahan klasifikasi secara lebih detail [10]. Berdasarkan *confusion matrix*, beberapa metrik evaluasi diturunkan sebagai berikut:

- **Accuracy** mengukur persentase prediksi benar dibandingkan total data uji.
- **Precision** menilai proporsi prediksi positif yang benar terhadap seluruh data yang diprediksi positif.
- **Recall** mengukur tingkat keberhasilan model dalam menemukan seluruh data positif sebenarnya.
- **F1-score** merupakan harmonisasi antara *precision* dan *recall* sehingga dapat menggambarkan performa model secara seimbang [11] [12].

Proses evaluasi dilakukan pada hasil pengujian untuk tiap replikasi pelatihan–pengujian model, kemudian nilai metrik dirata-ratakan guna memperoleh hasil evaluasi yang stabil dan *reliabel*. Metrik-metrik ini dipilih karena secara luas digunakan dalam penelitian berbasis *machine learning classification* dan mampu memberikan gambaran objektif terkait kualitas model dalam memprediksi kelulusan mahasiswa [13].

### 2.5 Implementasi

Seluruh eksperimen dijalankan pada lingkungan *Google Colab* dengan spesifikasi *runtime Python 3.10*, RAM 12 GB. Kode program sepenuhnya dituliskan secara manual tanpa penggunaan *auto-ML* untuk menjaga keterulangan percobaan. Semua parameter algoritma menggunakan konfigurasi standar dari *scikit-learn* tanpa *hyperparameter tuning* agar fokus penelitian tetap pada analisis performa algoritma *Naive Bayes*.

## 3. Hasil dan Pembahasan

### 3.1 Hasil Penelitian

Bagian ini menyajikan temuan utama dari proses pelatihan dan pengujian model yang telah dibangun sebelumnya. Fokus evaluasi diarahkan pada performa algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa berdasarkan rekam jejak akademik berupa nilai IPK setiap semester [14]. Penyajian hasil dilakukan dalam bentuk tabel dan visualisasi untuk memperjelas persebaran prediksi model pada masing-masing kelas.

Model *Naive Bayes* dilatih menggunakan 80% data pelatihan dan 20% data pengujian. Contoh representasi data uji ditunjukkan pada Tabel 2.

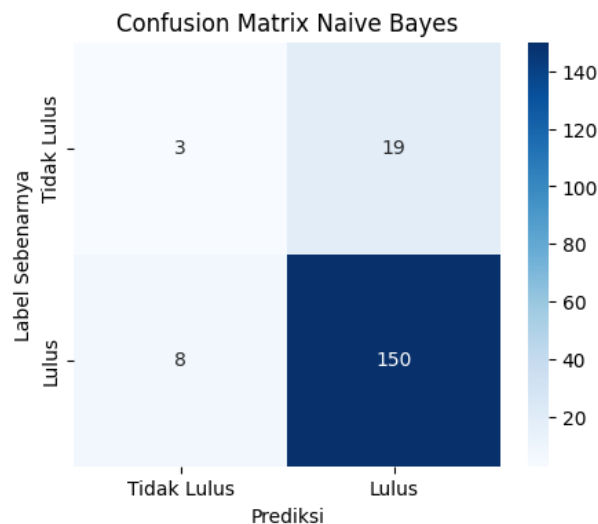
Table 2. Hasil akurasi data menggunakan metode evaluasi *confusion matrix python*

No	IPK1	IPK2	IPK3	IPK4	Lulus_Tepat_Waktu
0	2.3	1.97	1.8	1.56	Tidak
1	1.81	1.68	1.57	1.86	Tidak

2	3.07	3	2.75	3.21	Tidak
3	2.71	2.33	2.61	1.98	Tidak
4	3.17	3.02	3.28	2.96	Tidak

Evaluasi performa dilakukan menggunakan lima metrik, yaitu *Accuracy*, *precision*, *recall*, *F1-score*, dan *Confusion matrix*[15]. Ringkasan hasil evaluasi menunjukkan bahwa algoritma *Naive Bayes* menghasilkan nilai akurasi sebesar 0.85 (85%), yang mengindikasikan model mampu melakukan klasifikasi dengan tingkat ketepatan yang baik pada data pengujian. Nilai metrik diperoleh dengan menghitung rata-rata dari beberapa replikasi proses pelatihan-pengujian .

Visualisasi *Confusion matrix* pada Gambar 4 menunjukkan distribusi prediksi model terhadap label aktual. Nilai *True Positive* (TP) dan *True Negative* (TN) tercatat lebih dominan dibanding *False Positive* (FP) dan *False Negative* (FN) [16]. Kondisi ini memperlihatkan bahwa sebagian besar prediksi berada pada diagonal utama matriks, menandakan bahwa model cukup konsisten dalam membedakan mahasiswa yang lulus tepat waktu dan yang tidak.



Gambar 4. Visualisasi *Confusion matrix*

### 3.2 Pembahasan

Hasil eksperimen menunjukkan bahwa algoritma *Naive Bayes* mampu memberikan performa yang baik dalam memprediksi kelulusan tepat waktu mahasiswa berdasarkan fitur akademik berupa nilai IPK tiap semester. Model dilatih dengan menggunakan data training dan menghasilkan prediksi pada data testing. Nilai akurasi serta metrik evaluasi lainnya diperoleh melalui proses komputasi *Python* menggunakan *classification\_report* dan *confusion\_matrix* pada *library scikit-learn*, di mana nilai prediksi model dibandingkan dengan label aktual. Akurasi dihitung menggunakan persamaan:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots (1)$$

Nilai *precision*, *recall*, dan *F1-score* diperoleh melalui perhitungan otomatis pada *scikit-learn* berdasarkan distribusi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang dihasilkan model setelah proses prediksi. Dengan demikian, seluruh metrik evaluasi bersumber langsung dari *confusion matrix* yang kemudian dirangkum dalam laporan klasifikasi.

Model *Naive Bayes* menunjukkan performa yang cukup unggul, dengan nilai *Accuracy* sebesar 0.85, yang berarti 85% hasil prediksi konsisten dengan label aktual pada data uji. Nilai *Precision* sebesar 0.8876 mengindikasikan bahwa prediksi mahasiswa yang dinyatakan lulus tepat waktu sebagian besar benar, sedangkan *Recall* 0.9494 menunjukkan kemampuan model dalam menangkap seluruh kelas positif meskipun masih terdapat kasus *False*

*Negative*. *F1-score* 0.9174 memperlihatkan keseimbangan performa antara *precision* dan *recall*, menandakan model bekerja stabil meskipun terdapat variasi nilai antar fitur.

Tabel 3. Ringkasan metrik evaluasi model

<b>Metrik</b>	<b>Nilai</b>
<i>Accuracy</i>	0.85
<i>Precision</i>	0.8876
<i>Recall</i>	0.9494
<i>F1-score</i>	0.9174

Kinerja ini sejalan dengan karakteristik *Naive Bayes* yang efektif pada dataset tabular dan fitur yang relatif *independen*. Tidak ditemukan indikasi *overfitting*, mengingat performa pada data pelatihan dan data pengujian tidak berbeda signifikan. Visualisasi laporan klasifikasi dapat dilihat pada Gambar 3 sebagai bentuk pendukung hasil *kuantitatif*.

Tabel 4. *Classification Report* hasil evaluasi model

<b>Kategori</b>	<b><i>Precision</i></b>	<b><i>Recall</i></b>	<b><i>F1-score</i></b>	<b><i>Support</i></b>
0	0.27	0.14	0.18	22
1	0.89	0.95	0.92	158
<b><i>Accuracy</i></b>			0.85	180
<b><i>Macro Avg</i></b>	0.58	0.54	0.55	180
<b><i>Weighted Avg</i></b>	0.81	0.85	0.83	180

*Naive Bayes* dapat dijadikan *baseline* model dalam sistem *Early Warning* kelulusan mahasiswa, khususnya untuk prediksi berbasis performa akademik. Meski demikian, ruang pengembangan masih terbuka untuk penelitian lanjutan, seperti optimasi *hyperparameter*, penanganan ketidakseimbangan kelas, atau perbandingan dengan model *tree-based* maupun *ensemble* untuk meningkatkan akurasi secara lebih signifikan.

#### 4. Kesimpulan

Penelitian ini dilakukan untuk menjawab rumusan masalah mengenai seberapa baik performa algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa berdasarkan data akademik. Berdasarkan hasil eksperimen yang telah dilakukan, model *Naive Bayes* menunjukkan performa klasifikasi yang baik dengan capaian akurasi sebesar 85%, *precision* 0,8876, *recall* 0,9494, dan *F1-score* 0,9174. Nilai evaluasi tersebut menunjukkan bahwa algoritma ini mampu memprediksi status kelulusan mahasiswa dengan tingkat ketepatan yang tinggi, serta dapat membedakan mahasiswa yang berpotensi lulus dan tidak lulus secara efektif.

Visualisasi *confusion matrix* menampilkan dominasi nilai pada diagonal utama, yang berarti prediksi model cenderung konsisten serta memiliki kemampuan generalisasi yang stabil. Dengan demikian dapat disimpulkan bahwa *Naive Bayes* merupakan metode yang layak diterapkan sebagai *baseline* model dalam sistem pendukung keputusan akademik, khususnya sebagai alat *Early Warning* untuk mendeteksi potensi keterlambatan kelulusan mahasiswa.

Adapun penelitian selanjutnya dapat dikembangkan melalui *hyperparameter tuning*, penanganan *imbalanced dataset*, penambahan fitur akademik maupun non-akademik, serta melakukan perbandingan dengan algoritma klasifikasi lain seperti *tree-based* atau *ensemble* guna meningkatkan akurasi dan kemampuan generalisasi model.

## Daftar Rujukan

- [1] a. Tholib, m. N. F. Hidayat, r. Wulanningrum, and e. Daniati, “comparison of c4 . 5 and naive bayes for predicting student graduation using machine learning algorithms,” *int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, pp. 71–78, 2023, doi: 10.30812/ijecsa.v2i2.3364.
- [2] v. M. Putri, b. A. Wisesa, i. A. Edyyul, and s. A. Darma, “sistem pakar deteksi keterlambatan bicara anak menggunakan forward chaining dan naive bayes,” *j. Pustaka ai*, vol. 5, no. 3, pp. 7–12, 2025.
- [3] a. D. Imanuel, n. N. Pusparini, and a. Sani, “klasifikasi untuk memprediksi tingkat kelulusan mahasiswa stmik widuri menggunakan algoritma naive bayes,” *j. Ilm. Inform.*, vol. 12, no. 1, pp. 1–7, 2024.
- [4] a. S. Learning, “g-tech : jurnal teknologi terapan,” *g-tech j. Teknol. Terap.*, vol. 7, no. 2, pp. 386–395, 2023.
- [5] z. Fatah et al., “klasifikasi kelulusan mahasiswa menggunakan metode decision tree,” *j. Ris. Tek. Komput.*, vol. 1, no. 4, pp. 58–64, 2024.
- [6] u. N. Makassar and p. Kelulusan, “development of graduation prediction application for faculty of engineering students at makassar state university based on website,” *j. Media elektr.*, vol. 20, no. 2, pp. 109–113, 2023.
- [7] e. J. Sudarman and s. Budi, “pengembangan model kecerdasan mesin extreme gradient boosting untuk prediksi keberhasilan studi mahasiswa,” *j. Strateg.*, vol. 5, no. 2, pp. 297–314, 2023.
- [8] i. Oktadiani, h. Fitriawan, and m. Nurwahidin, “penerapan machine learning untuk prediksi masa studi mahasiswa di perguruan tinggi x,” *j. Rekayasa dan teknol. Elektro*, vol. 17, no. 3, 2023.
- [9] u. Kristen, k. Wacana, and j. Barat, “prediksi mahasiswa drop-out di universitas xyz prediction of student drop-out at xyz university,” *j. Teknol. Inf. Dan ilmu komput.*, vol. 11, no. 6, pp. 1345–1350, 2024, doi: 10.25126/jtiik.2024118689.
- [10] t. Gori et al., “preprocessing data dan klasifikasi untuk prediksi kinerja data preprocessing and classification for predicting student,” *j. Teknol. Inf. Dan ilmu komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.2024118074.
- [11] f. Rahman, “waktu menggunakan binning dan synthetic minority,” *j. Artif. Intelligence data sci.*, vol. 4, no. 1, pp. 30–36, 2024.
- [12] t. Menggunakan and c. Algoritma, “penerapan model klasifikasi pada kelulusan mahasiswa perjuangan,” *jitet (jurnal inform. Dan tek. Elektro ter.*, vol. 12, no. 3, pp. 3402–3411, 2024.
- [13] d. A. Punkastyo, f. Septian, and a. Syaripudin, “implementasi data mining menggunakan algoritma naive bayes untuk prediksi kelulusan siswa,” *j. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 24–35, 2024.
- [14] r. S. Asrul and a. L. Pramana, “e issn : 2809-4069 sistem prediksi penerimaan karyawan di pt . Surya prima semesta berbasis web menggunakan metode naive bayes,” *j. Pustaka ai*, vol. 5, no. 1, pp. 80–88, 2025.
- [15] v. A. Renedominick and s. B. Pranata, “e issn : 2809-4069 analisis sentimen pada trailer deadpool vs wolverine menggunakan model machine learning,” *j. Pustaka ai*, vol. 5, no. 1, pp. 1–6, 2025.
- [16] z. Fatah et al., “klasifikasi data mining untuk memprediksi kelulusan mahasiswa,” *J. Ilmu Multidisiplin Ilmu*, vol. 2, no. 1, pp. 29–37, 2025.