

Penerapan Algoritma Naïve Bayes untuk Analisis Potensi Siswa Berdasarkan Nilai Rapor

Luigi Nakata¹, Danika Ambadar Alfari²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *luiginakata2003@gmail.com, danikaambadar@gmail.com

Abstrak – Pemanfaatan data akademik di sekolah masih belum optimal karena nilai rapor umumnya hanya digunakan untuk keperluan administratif. Padahal, data tersebut memiliki potensi besar untuk dianalisis lebih lanjut guna mengidentifikasi potensi akademik siswa secara objektif. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan potensi akademik siswa MTsN 7 Kediri berdasarkan data nilai rapor menggunakan algoritma Naïve Bayes. Algoritma Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang bekerja dengan menghitung peluang setiap kelas potensi akademik berdasarkan data latih, melalui perhitungan probabilitas prior dan likelihood dari setiap atribut nilai mata pelajaran, kemudian menentukan kelas dengan nilai probabilitas tertinggi. Tahapan penelitian meliputi preprocessing data nilai rapor, pembagian data latih dan data uji, pelatihan model Naïve Bayes, serta evaluasi hasil klasifikasi. Implementasi sistem dilakukan menggunakan bahasa pemrograman Python dengan bantuan library Scikit-learn. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan potensi akademik siswa secara efektif dengan tingkat akurasi sebesar 77%, sehingga model yang dihasilkan dinilai cukup baik dalam memetakan kecenderungan potensi akademik siswa. Hasil ini diharapkan dapat menjadi pendukung pengambilan keputusan akademik berbasis data di lingkungan sekolah.

Kata Kunci — Analisis Potensi Siswa, Data Mining, Naïve Bayes, Nilai Rapor

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pemanfaatan data secara luas dalam bidang pendidikan. Salah satu pendekatan yang berkembang adalah *Educational Data Mining* (EDM), yaitu penerapan teknik data mining untuk mengekstraksi pengetahuan dari data pendidikan guna mendukung pengambilan keputusan berbasis data. Berbagai penelitian menunjukkan bahwa pendekatan EDM efektif digunakan untuk menganalisis dan memprediksi performa akademik peserta didik dengan memanfaatkan algoritma klasifikasi, termasuk *Naïve Bayes* [1],[2],[3].

Nilai rapor merupakan indikator utama capaian akademik siswa karena mencerminkan hasil belajar dalam periode tertentu. Namun, pemanfaatan nilai rapor di sekolah pada umumnya masih terbatas pada keperluan administratif seperti penentuan kenaikan kelas dan kelulusan. Padahal, data nilai akademik memiliki potensi besar untuk dianalisis lebih lanjut guna mengidentifikasi prestasi dan potensi akademik peserta didik secara lebih objektif melalui pendekatan data mining [4],[5].

Identifikasi potensi akademik siswa merupakan permasalahan yang kompleks karena kemampuan siswa tidak dapat ditentukan hanya berdasarkan nilai rata-rata. Setiap peserta didik memiliki karakteristik dan kecenderungan kemampuan yang berbeda, sehingga diperlukan metode analisis yang sistematis, terukur, dan objektif untuk memetakan potensi akademik tersebut. Tanpa adanya metode analisis berbasis data, proses pemetaan potensi akademik cenderung bersifat subjektif dan kurang optimal [6],[7].

Berbagai penelitian dalam bidang data mining pendidikan menunjukkan bahwa metode klasifikasi dapat digunakan untuk memprediksi dan mengelompokkan prestasi akademik siswa maupun mahasiswa. Studi perbandingan algoritma klasifikasi menunjukkan bahwa algoritma *Naïve Bayes* memiliki kinerja yang kompetitif dibandingkan metode lain, seperti *Decision Tree*, baik dari sisi akurasi maupun efisiensi komputasi, sehingga banyak digunakan dalam analisis data pendidikan [2],[8].

Selain itu, sejumlah penelitian membuktikan bahwa algoritma Naïve Bayes mampu memberikan hasil yang baik dalam klasifikasi prestasi akademik berdasarkan data nilai rapor atau nilai akhir, baik pada tingkat sekolah maupun perguruan tinggi [4],[9],[10]. Penerapan algoritma ini juga relatif sederhana dan dapat diimplementasikan dengan bantuan perangkat lunak serta library pendukung, sehingga memudahkan proses pengolahan data dan evaluasi hasil klasifikasi [5].

MTsN 7 Kediri sebagai lembaga pendidikan berbasis madrasah memiliki karakteristik kurikulum yang memadukan mata pelajaran umum dan keagamaan. Kondisi ini menuntut adanya metode analisis yang mampu mengolah data nilai rapor secara efektif.

Berdasarkan uraian tersebut, penelitian ini berfokus pada penerapan algoritma *Naïve Bayes* untuk menganalisis potensi akademik siswa berdasarkan data nilai rapor dengan pendekatan *Educational Data Mining*. Diharapkan hasil penelitian ini dapat memberikan informasi potensi akademik siswa secara Objektif.

2. METODE PENELITIAN

Penelitian ini menggunakan metode klasifikasi dengan algoritma *Naïve Bayes* untuk menganalisis potensi akademik siswa berdasarkan data nilai rapor. Algoritma *Naïve Bayes* dipilih karena memiliki kemampuan klasifikasi berbasis probabilitas yang sederhana, efisien, serta efektif dalam mengolah data numerik dengan jumlah atribut yang relatif banyak. Implementasi sistem dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *library Scikit-learn*.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa data nilai rapor siswa yang mencerminkan capaian akademik pada beberapa mata pelajaran. Mata pelajaran tersebut dikelompokkan ke dalam empat bidang utama, yaitu sains, bahasa, sosial, dan keagamaan. Pengelompokan ini bertujuan untuk mempermudah analisis serta menggambarkan kecenderungan potensi akademik siswa pada masing-masing bidang. Data nilai rapor selanjutnya digunakan sebagai data sampel dalam penerapan algoritma *Naïve Bayes* untuk mengklasifikasikan potensi akademik siswa berdasarkan pola nilai yang dimiliki.

Sebelum proses inti analisis potensi siswa dengan *naïve bayes* dilakukan, terlebih dahulu dilakukan pra-pemrosesan Data. Pra-pemrosesan data merupakan tahap awal yang bertujuan meningkatkan kualitas data sebelum dilakukan proses klasifikasi. Tahapan ini penting karena data nilai rapor yang diperoleh dapat mengandung ketidakkonsistenan format atau nilai yang tidak lengkap. Proses pra-pemrosesan yang dilakukan meliputi:

1. Pembersihan Data

Pembersihan data dilakukan dengan menghilangkan data yang tidak lengkap, duplikasi data, serta memperbaiki kesalahan penulisan atau pengisian nilai rapor. Tahap ini bertujuan memastikan data yang digunakan dalam proses klasifikasi benar-benar valid dan dapat dipertanggungjawabkan.

2. Normalisasi Data

Normalisasi data dilakukan untuk menyamakan rentang nilai antar mata pelajaran sehingga tidak terjadi dominasi atribut tertentu dalam proses perhitungan probabilitas. Normalisasi membantu meningkatkan kestabilan dan keakuratan hasil klasifikasi potensi siswa.

2.2 Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* adalah metode klasifikasi yang didasarkan pada *Teorema Bayes* dengan asumsi bahwa setiap atribut bersifat saling independen satu sama lain. Metode ini banyak digunakan dalam bidang data mining pendidikan karena kesederhanaan perhitungan, efisiensi proses, serta kemampuannya dalam menangani data dengan jumlah atribut yang relatif banyak.

Pada penelitian ini, algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan potensi akademik siswa berdasarkan data nilai rapor. Setiap siswa direpresentasikan oleh sejumlah atribut nilai mata pelajaran yang telah dikelompokkan ke dalam bidang sains, bahasa, sosial, dan keagamaan. Kelas yang dihasilkan merupakan bidang potensi akademik dominan yang dimiliki oleh siswa. Secara umum, persamaan *Naïve Bayes* dinyatakan sebagai berikut:

$$P(C | X) = \frac{P(X|C) \times P(C)}{P(X)} \dots (1)$$

Penjelasan:

C : kelas potensi siswa.

X : data nilai rapor siswa.

$P(C)$: probabilitas prior kelas.

$P(X|C)$: probabilitas likelihood.

$P(C|X)$: probabilitas posterior.

2.3 Proses Klasifikasi Potensi Siswa

Proses klasifikasi dimulai dengan melatih model *Naïve Bayes* menggunakan data latih untuk membentuk model probabilistik. Model tersebut kemudian digunakan untuk memprediksi potensi

akademik siswa pada data uji. Hasil prediksi diperoleh dengan menghitung probabilitas posterior dari setiap kelas potensi akademik.

Potensi akademik siswa ditentukan berdasarkan kelas dengan nilai probabilitas tertinggi, yaitu *sains*, bahasa, sosial, atau keagamaan. Hasil klasifikasi ini digunakan sebagai dasar analisis kecenderungan potensi akademik siswa. Hasil proses klasifikasi kemudian disajikan dalam bentuk informasi potensi siswa yang mencerminkan kecenderungan bidang akademik dominan, yaitu *sains*, bahasa, sosial, atau keagamaan. Selain itu, hasil klasifikasi divisualisasikan dalam bentuk grafik pada dashboard sistem untuk memudahkan pengguna dalam memahami dan menganalisis distribusi potensi akademik siswa secara keseluruhan.

2.4 Evaluasi Sistem

Evaluasi sistem dilakukan untuk menilai kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan potensi akademik siswa berdasarkan nilai rapor. Proses evaluasi bertujuan untuk mengetahui tingkat keakuratan hasil klasifikasi yang dihasilkan oleh sistem dibandingkan dengan data aktual atau penilaian yang telah ditentukan.

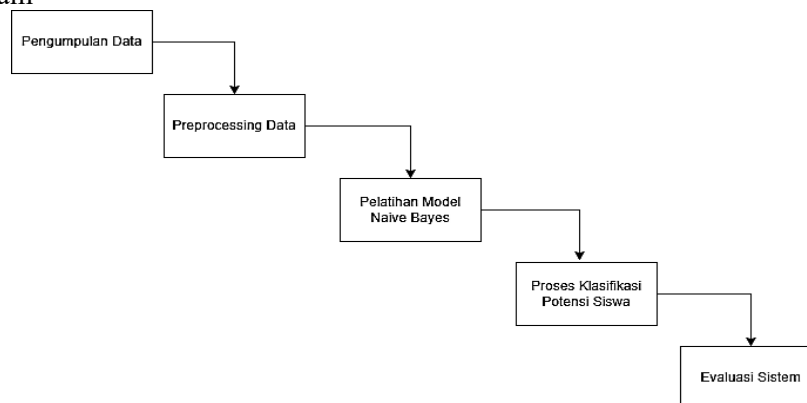
Pengujian dilakukan dengan membandingkan hasil klasifikasi sistem terhadap data uji dengan kelas potensi yang menjadi acuan. Metrik evaluasi yang digunakan dalam penelitian ini adalah tingkat akurasi, yang menunjukkan persentase data yang berhasil diklasifikasikan dengan benar oleh sistem.

Nilai akurasi yang diperoleh digunakan sebagai indikator efektivitas penerapan algoritma *Naïve Bayes* dalam analisis potensi siswa. Semakin tinggi nilai akurasi yang dihasilkan, maka semakin baik kinerja sistem dalam mengklasifikasikan potensi akademik siswa secara tepat.

3. HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.

3.1 Alur Diagram



Gambar 3.1 Alur Diagram

1. Pengumpulan Data

Tahap ini merupakan proses pengambilan data nilai rapor siswa yang mencerminkan capaian akademik pada bidang *sains*, bahasa, sosial, dan keagamaan.

2. Pra-pemrosesan Data

Data nilai rapor yang telah dikumpulkan dibersihkan dan dinormalisasi untuk memastikan kualitas data serta menghindari ketidakonsistenan sebelum dilakukan klasifikasi.

3. Pelatihan Model *Naïve Bayes*

Data yang telah melalui pra-pemrosesan digunakan sebagai data latih untuk membangun model klasifikasi menggunakan algoritma *Gaussian Naïve Bayes*.

4. Proses Klasifikasi Potensi Siswa

Model yang telah dilatih digunakan untuk menentukan potensi akademik siswa berdasarkan nilai probabilitas tertinggi pada masing-masing bidang.

5. Evaluasi Sistem

Tahap akhir dilakukan untuk menilai kinerja sistem dengan mengukur tingkat akurasi hasil klasifikasi terhadap data uji.

3.2 Hasil Pengolahan Data

Pada tahap pengumpulan data, Data yang digunakan dalam penelitian ini diambil dari MTsN 7 Kediri dan berupa data nilai rapor siswa. Data tersebut mencerminkan capaian akademik siswa pada beberapa mata pelajaran yang dikelompokkan ke dalam bidang sains, bahasa, sosial, dan keagamaan. Data nilai rapor ini digunakan sebagai dasar dalam proses analisis dan klasifikasi potensi akademik siswa dengan menerapkan algoritma *Naïve Bayes*.

Tabel 3.1 Contoh Hasil olah Data Rapor

Nama	Sains	Sosial	Bahasa	Keagamaan
ACHMAD MUHYIDIN SHOBRI	80	85	94	80
ACHMAD RIKO ADITIYA	88	84	86	82
AHMAD KHOIRUL FIKRI	95	82	90	82
AJENG ROSSITA ANGGRAINI	92	80	82	94
ALFINA FEBRIANA M	82	90	95	90
ANA FENORIKHA MUNAWAROH	90	88	84	88
DAVID SAMUDRA	84	82	95	96

3.3 Klasifikasi Potensi Akademik

Berdasarkan hasil penerapan algoritma *Gaussian Naïve Bayes*, sistem mampu mengklasifikasikan potensi akademik siswa ke dalam empat kelas potensi, yaitu *sains*, bahasa, sosial, dan keagamaan. Proses klasifikasi dilakukan dengan menghitung probabilitas posterior dari setiap kelas berdasarkan nilai rapor yang dimiliki siswa.

Hasil klasifikasi menunjukkan bahwa algoritma *Naïve Bayes* dapat mengidentifikasi kecenderungan potensi akademik siswa berdasarkan pola nilai rapor yang dimiliki. Siswa dengan nilai dominan pada mata pelajaran sains cenderung diklasifikasikan ke dalam potensi *sains*, demikian pula untuk bidang bahasa, sosial, dan keagamaan.

Tabel 3.2 Klasifikasi Potensi

Nama	Potensi Aktual	Hasil Prediksi
ACHMAD MUHYIDIN SHOBRI	Bahasa	Bahasa
ACHMAD RIKO ADITIYA	Sains	Sains
AHMAD KHOIRUL FIKRI	Sains	Sains
AJENG ROSSITA ANGGRAINI	Keagamaan	Keagamaan
ALFINA FEBRIANA M	Bahasa	Bahasa
ANA FENORIKHA MUNAWAROH	Sains	Sains
DAVID SAMUDRA	Keagamaan	Keagamaan

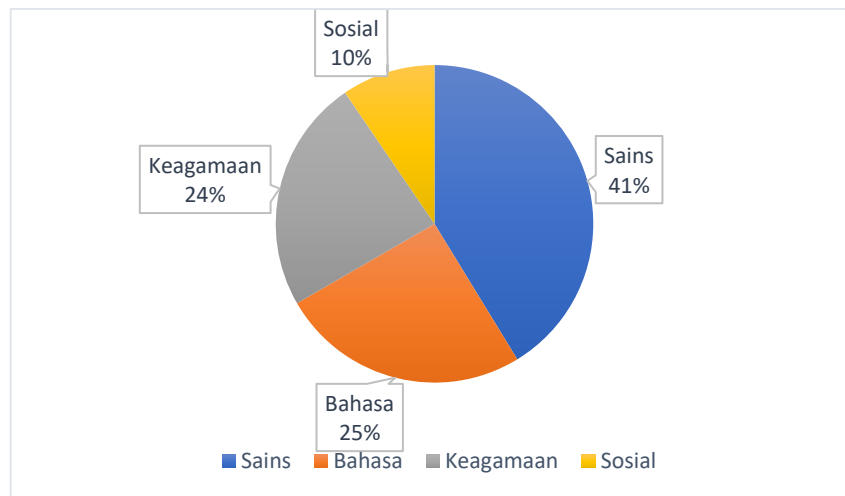
3.4 Evaluasi Kinerja

Evaluasi kinerja model dilakukan menggunakan metrik akurasi, yaitu perbandingan antara jumlah data yang berhasil diklasifikasikan dengan benar terhadap total data uji. Berdasarkan hasil pengujian, model *Naïve Bayes* menghasilkan nilai akurasi yang menunjukkan bahwa sistem mampu melakukan klasifikasi potensi akademik siswa dengan baik.

Nilai akurasi yang diperoleh menunjukkan bahwa algoritma *Naïve Bayes* efektif dalam mengolah data nilai rapor yang bersifat numerik dan kontinu. Hal ini disebabkan oleh kesederhanaan perhitungan probabilitas serta kemampuan algoritma dalam menangani data dengan jumlah atribut yang relatif sedikit namun informatif.

3.5 Visualisasi Hasil

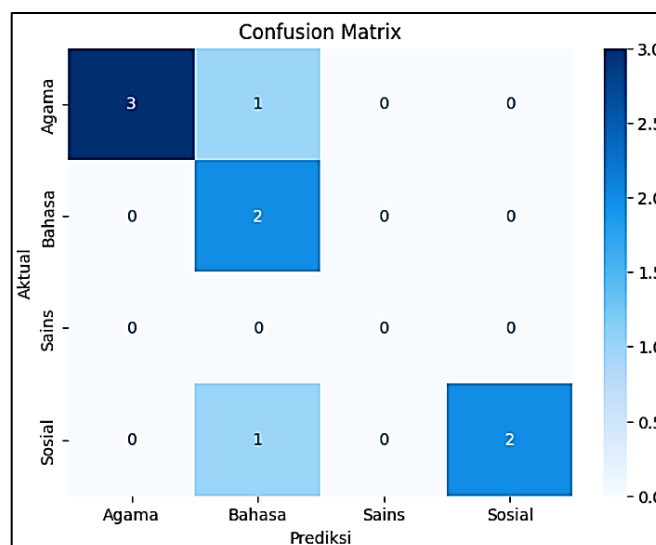
Untuk mempermudah pemahaman terhadap hasil klasifikasi, sistem menampilkan visualisasi dalam bentuk grafik. Grafik tersebut menunjukkan distribusi potensi akademik siswa berdasarkan hasil klasifikasi algoritma *Naïve Bayes*.



Gambar 3.2 Grafik Potensi

Pada gambar 3.1 Grafik potensi akademik siswa berdasarkan hasil klasifikasi menggunakan algoritma *Naïve Bayes*. Grafik ini merepresentasikan jumlah siswa yang diklasifikasikan ke dalam empat kategori potensi akademik, yaitu sains, bahasa, sosial, dan keagamaan. Grafik potensi siswa dapat dilihat bahwa setiap kategori potensi memiliki jumlah siswa yang berbeda-beda. Hal ini menunjukkan adanya variasi kecenderungan potensi akademik siswa yang dipengaruhi oleh pola nilai rapor pada masing-masing bidang mata pelajaran.

Kategori dengan jumlah siswa tertinggi mencerminkan bidang akademik yang paling dominan dimiliki oleh siswa secara keseluruhan, sedangkan kategori dengan jumlah terendah menunjukkan bidang potensi yang relatif lebih sedikit. Visualisasi grafik potensi ini membantu pengguna, seperti guru atau pihak sekolah, dalam memahami persebaran potensi akademik siswa secara cepat. Dengan adanya grafik ini, proses pemetaan potensi siswa menjadi lebih mudah, objektif, dan dapat digunakan sebagai dasar dalam pengambilan keputusan, seperti penentuan program pembinaan, pengembangan minat bakat, maupun strategi pembelajaran yang lebih tepat sasaran.



Gambar 3.3 Confusion Matrix

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* dapat digunakan sebagai metode klasifikasi yang efektif dalam analisis potensi akademik siswa berdasarkan nilai rapor. Pendekatan berbasis probabilitas memungkinkan sistem melakukan klasifikasi secara objektif tanpa dipengaruhi penilaian subjektif.

Hasil yang diperoleh menunjukkan kinerja yang baik, penelitian ini masih memiliki keterbatasan, antara lain penggunaan jumlah data yang terbatas serta hanya mengandalkan nilai rapor sebagai atribut utama. Oleh karena itu, pengembangan lebih lanjut dapat dilakukan dengan menambahkan atribut lain seperti kehadiran siswa, nilai ekstrakurikuler, atau hasil tes minat dan bakat untuk meningkatkan akurasi klasifikasi.

4. SIMPULAN

Hasil penelitian yang telah dilakukan, algoritma *Naïve Bayes* terbukti dapat digunakan secara efektif untuk menganalisis dan mengklasifikasikan potensi akademik siswa MTsN 7 Kediri berdasarkan data nilai rapor. Hasil pengujian menunjukkan bahwa model yang dibangun menghasilkan tingkat akurasi sebesar 77%, yang menunjukkan bahwa algoritma *Naïve Bayes* memiliki kemampuan klasifikasi yang baik dalam memetakan kecenderungan potensi akademik siswa ke dalam kategori yang telah ditentukan.

Performa algoritma *Naïve Bayes* bersifat kompetitif dan telah memenuhi target akurasi yang diharapkan. Tingkat akurasi yang dihasilkan dipengaruhi oleh beberapa faktor utama, antara lain proses *preprocessing data* yang baik, pemilihan atribut nilai rapor yang relevan, serta pembagian data latih dan data uji yang proporsional. Namun demikian, penelitian ini masih memiliki keterbatasan karena hanya menggunakan data nilai rapor sebagai atribut utama, tanpa mempertimbangkan faktor non-akademik seperti minat, kehadiran, dan aktivitas siswa. Oleh karena itu, penelitian selanjutnya disarankan untuk menambahkan variabel tersebut serta melakukan perbandingan langsung dengan algoritma klasifikasi lain guna meningkatkan akurasi dan memperkaya hasil analisis.

5. SARAN

Berdasarkan keterbatasan yang terdapat dalam penelitian ini, beberapa saran untuk penelitian selanjutnya dapat disampaikan sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar dan beragam agar hasil klasifikasi potensi akademik siswa menjadi lebih representatif dan akurat.
2. Penambahan atribut pendukung selain nilai rapor, seperti kehadiran siswa, nilai ekstrakurikuler, serta hasil tes minat dan bakat, dapat dipertimbangkan untuk meningkatkan kualitas dan ketepatan hasil klasifikasi.
3. Penelitian lanjutan dapat melakukan perbandingan dengan algoritma klasifikasi lain, guna mengetahui metode yang paling optimal dalam analisis potensi akademik siswa.
4. Pengembangan sistem lebih lanjut dapat diarahkan pada implementasi berbasis aplikasi dengan antarmuka yang lebih interaktif, sehingga hasil analisis dapat dimanfaatkan secara lebih luas dalam mendukung pengambilan keputusan di bidang pendidikan.

DAFTAR PUSTAKA

- [1] M. Yağcı, "Educational data mining : prediction of students ' academic performance using machine learning algorithms," 2022, doi: 10.1186/s40561-022-00192-z.
- [2] B. Ningsi and M. Munarsih, "Performance Comparison of Data Mining Classification Algorithms on Student Academic Achievement Prediction," vol. 6, no. 1, pp. 29–39, 2023.
- [3] D. Ayu, N. Wulandari, R. Annisa, L. Yusuf, and T. Prihatin, "AN EDUCATIONAL DATA MINING FOR STUDENT ACADEMIC PREDICTION USING K-MEANS CLUSTERING AND NAÏVE BAYES," *J. pilar mandiri*, vol. 16, no. 2, pp. 155–160, 2020.
- [4] H. Amalia, A. Puspita, and A. F. Lestari, "APPLICATION OF DECISION TREE AND NAIVE

- BAYES ON STUDENT,” vol. 18, no. 1, 2022, doi: 10.33480/pilar.v18i1.2714.
- [5] A. Rahman, “Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes,” vol. 9, pp. 22–31, 2023.
 - [6] S. Laila and Z. Fatah, “Implementasi Metode Naive Bayes Pada Pola Prestasi Siswa Dengan Menggunakan Rapidminer,” vol. 4, no. April, pp. 69–74, 2025.
 - [7] R. Ekawati, V. Atina, and J. Maulidar, “Prediksi Ketuntasan Belajar Siswa Menggunakan Naive Bayes dengan Integrasi Data Student Completion Prediction System Using the Naive Bayes Method at SMA Negeri 1 Plupuh,” vol. 5, no. 4, pp. 1161–1173, 2025.
 - [8] A. Mueen, B. Zavar, and U. Manzoor, “Modeling and Predicting Students’ Academic Performance Using Data Mining Techniques,” 2016.
 - [9] A. Yulianto, U. Bina, S. Informatika, U. N. Mandiri, and P. Korespondensi, “Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar,” vol. 7, no. April, pp. 1174–1182, 2023.
 - [10] B. Rizky, S. Dwinanda, A. N. Ilhami, H. Choerunisya, and I. J. Thira, “Jurnal Sistem dan Teknologi Informasi Indonesia Prediksi Nilai Akhir Mahasiswa Dengan Metode Naïve Bayes Prediction of Final Student Grades Using Naive Bayes Method,” vol. 9, no. 1, pp. 27–35, 2024.