



Prediksi Tingkat Pemahaman Siswa Smp Swasta Tenera Berdasarkan Aktivitas Belajar Menggunakan Naive Bayes

M. Dimas Prayoga^{1*}, Mona Ayunda², Roberto Kaban³

¹⁻³Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Institut Teknologi dan Bisnis, Indonesia

Email: mdimas1836@gmail.com¹, monaayunda@icloud.com², roberto.kaban@yahoo.com³

*Penulis Korespondensi: mdimas1836@gmail.com

Abstract. *This study aims to develop a prediction model for determining the comprehension level of students at SMP Swasta Tenera using the Naive Bayes algorithm based on academic learning activities. The prediction model utilizes student learning data, including subject grades and attendance records, as the main variables to classify students' comprehension levels. The data used in this study were collected from 90 students consisting of Grade VII (28 students), Grade VIII (23 students), and Grade IX (39 students), covering performance data from 10 subjects. The research method applies a quantitative approach with data processing and classification analysis using the Naive Bayes algorithm. The evaluation results show that the developed model achieved an accuracy level of 83.33%, with a precision value of 78.12%, recall of 98.04%, and an F1-score of 86.96%. The prediction results indicate that Grade VII students have the lowest comprehension level at 42.9%, followed by Grade VIII at 60.9% and Grade IX at 64.1%. This research demonstrates that machine learning-based prediction systems can support educational decision-making by identifying students who require learning assistance and enabling schools to implement faster, more targeted, and effective academic interventions.*

Keywords: *Classification; Learning Activity; Machine Learning; Naive Bayes; Understanding Prediction.*

Abstrak. Penelitian ini bertujuan untuk memprediksi tingkat pemahaman siswa SMP Swasta Tenera menggunakan algoritma Naive Bayes berdasarkan aktivitas belajar yang mencakup nilai mata pelajaran dan kehadiran. Data penelitian diperoleh dari 90 siswa yang terdiri dari Kelas VII (28 siswa), Kelas VIII (23 siswa), dan Kelas IX (39 siswa), dengan 10 mata pelajaran yaitu PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, dan PJOK. Status pemahaman diklasifikasikan menjadi dua kategori, yaitu Paham (nilai rata-rata ≥ 75) dan Tidak Paham (nilai rata-rata < 75). Metode Naive Bayes Gaussian diterapkan untuk menghitung probabilitas posterior setiap siswa. Hasil penelitian menunjukkan bahwa dari 90 siswa, sebanyak 51 siswa (56,7%) tergolong Paham dan 39 siswa (43,3%) tergolong Tidak Paham. Model mencapai akurasi 83,33% dengan nilai presisi 78,12%, recall 98,04%, dan F1-score 86,96%. Kelas VII memiliki tingkat pemahaman terendah (42,9% Paham), diikuti Kelas VIII (60,9%) dan Kelas IX (64,1%). Penelitian ini memberikan kontribusi dalam pengembangan sistem prediksi akademik berbasis machine learning yang dapat membantu guru dan pihak sekolah dalam mengambil keputusan intervensi pembelajaran secara lebih cepat dan tepat.

Kata Kunci: Aktivitas Belajar; Klasifikasi; Naive Bayes; Pembelajaran Mesin; Prediksi Pemahaman.

1. LATAR BELAKANG

Keberhasilan pembelajaran pada tingkat sekolah menengah pertama tidak cukup dinilai hanya dari terselenggaranya kegiatan belajar mengajar, tetapi juga dari kemampuan siswa dalam memahami materi yang diberikan. Tingkat pemahaman menjadi salah satu gambaran penting untuk melihat perkembangan akademik karena menunjukkan sejauh mana siswa mampu menerima, mengolah, dan menguasai materi selama proses pembelajaran. Kondisi tersebut dapat diamati melalui berbagai rekam akademik yang tersedia di sekolah, termasuk capaian nilai mata pelajaran dan kehadiran siswa. Sholichah et al. (2022) menjelaskan bahwa perkembangan hasil belajar berkaitan dengan berbagai aktivitas yang menyertai proses pembelajaran. Permasalahan muncul ketika data akademik hanya digunakan untuk kebutuhan pencatatan dan pelaporan, sementara pola yang terkandung di dalamnya belum dimanfaatkan

untuk mengenali siswa yang berpotensi mengalami kesulitan pemahaman sejak lebih awal. Kebutuhan terhadap identifikasi dini menjadi semakin penting pada sekolah yang memiliki siswa dengan karakteristik akademik beragam. SMP Swasta Tenera memiliki siswa dari tiga jenjang, yaitu kelas VII, VIII, dan IX, dengan capaian pembelajaran yang berbeda pada setiap individu. Data penelitian mencakup 90 siswa yang terdiri atas 28 siswa kelas VII, 23 siswa kelas VIII, dan 39 siswa kelas IX. Informasi akademik yang tersedia meliputi nilai PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, PJOK, serta data kehadiran. Banyaknya atribut tersebut membuat proses identifikasi secara manual menjadi kurang praktis karena guru perlu menelaah sejumlah indikator secara bersamaan. Jika kondisi siswa baru diketahui setelah evaluasi akhir, kesempatan untuk memberikan pendampingan atau penyesuaian pembelajaran dapat menjadi lebih terbatas.

Pemanfaatan data pendidikan melalui pendekatan komputasional menawarkan cara yang lebih sistematis untuk mengenali pola akademik siswa. Salah satu bidang yang berkembang dalam konteks ini adalah *educational data mining*, yaitu pemanfaatan teknik analisis data untuk memperoleh informasi yang mendukung pemahaman terhadap proses dan hasil pendidikan (Romero & Ventura, 2020). Melalui pendekatan klasifikasi, data historis siswa dapat dipelajari untuk memperkirakan kategori tertentu berdasarkan karakteristik yang tersedia. Berbagai algoritma telah diterapkan untuk kebutuhan prediksi akademik, seperti *Decision Tree*, *K-Nearest Neighbor*, *Support Vector Machine*, dan Naive Bayes (Yagci, 2022). Di antara metode tersebut, Naive Bayes memiliki mekanisme probabilistik yang relatif sederhana dan dapat digunakan pada dataset berukuran kecil hingga menengah (Nakhipova et al., 2024).

Karakteristik data pada penelitian ini didominasi oleh atribut numerik berupa nilai mata pelajaran dan persentase kehadiran. Kondisi tersebut mendukung penggunaan Gaussian Naive Bayes karena metode ini dapat memperkirakan peluang suatu data terhadap kelas tertentu melalui distribusi probabilitas pada fitur kontinu. Ariansyah et al. (2023) menunjukkan relevansi pendekatan Gaussian dalam menangani atribut numerik pada proses klasifikasi. Melalui mekanisme tersebut, penentuan tingkat pemahaman siswa tidak hanya bertumpu pada satu nilai, tetapi mempertimbangkan pola dari sejumlah atribut akademik yang dianalisis secara bersamaan. Hasil prediksi kemudian dapat digunakan untuk membedakan siswa ke dalam kategori Paham dan Tidak Paham berdasarkan karakteristik data pembelajaran yang dimiliki. Sejumlah penelitian terdahulu memperlihatkan bahwa Naive Bayes memiliki potensi dalam pengolahan data akademik. Wahyuni dan Pratiwi (2022) melaporkan penerapan metode tersebut untuk klasifikasi prestasi belajar dengan tingkat akurasi sebesar 87%. Firmansyah dan

Yulianto (2023) juga menunjukkan bahwa Naive Bayes dapat digunakan pada data berukuran terbatas, sedangkan Riyanto dan Ompusunggu (2024) menerapkannya untuk mengklasifikasikan hasil belajar siswa. Temuan tersebut memperlihatkan bahwa pendekatan probabilistik dapat dimanfaatkan untuk mendukung prediksi akademik. Meskipun demikian, konteks objek, susunan atribut, cakupan jenjang, dan sasaran klasifikasi pada setiap penelitian tidak sepenuhnya sama.

Celah penelitian ini terletak pada kebutuhan untuk memprediksi tingkat pemahaman siswa pada satu lingkungan sekolah swasta dengan menggabungkan tiga jenjang kelas dalam satu proses analisis. Penelitian tidak hanya menggunakan satu kelompok kelas, tetapi mencakup siswa kelas VII, VIII, dan IX yang memiliki variasi capaian akademik berbeda. Selain itu, prediksi dibangun dari kombinasi nilai sepuluh mata pelajaran dan kehadiran siswa, kemudian hasilnya dianalisis untuk melihat distribusi tingkat pemahaman pada masing-masing jenjang. Pendekatan tersebut memberikan ruang analisis yang lebih luas karena model tidak berhenti pada penentuan kategori individual, tetapi juga dapat menunjukkan perbedaan kecenderungan pemahaman antarkelas sebagai bahan evaluasi akademik.

Berdasarkan kondisi tersebut, penelitian ini bertujuan menerapkan algoritma Gaussian Naive Bayes untuk memprediksi tingkat pemahaman siswa SMP Swasta Tenera berdasarkan aktivitas belajar yang direpresentasikan melalui nilai mata pelajaran dan kehadiran. Penelitian juga diarahkan untuk menganalisis distribusi kategori Paham dan Tidak Paham pada kelas VII, VIII, dan IX serta mengevaluasi kinerja model menggunakan *confusion matrix*. Hasil penelitian diharapkan dapat memberikan dasar analitis bagi sekolah dalam mengenali kondisi pemahaman siswa secara lebih terstruktur dan mendukung penentuan intervensi pembelajaran yang lebih tepat sesuai karakteristik akademik siswa.

2. KAJIAN TEORITIS

Tingkat Pemahaman Siswa

Tingkat pemahaman siswa menggambarkan kemampuan peserta didik dalam menerima, menguasai, dan memaknai materi yang diperoleh selama proses pembelajaran. Pemahaman tidak hanya berkaitan dengan kemampuan mengingat informasi, tetapi juga tercermin dari kecakapan siswa dalam menghubungkan konsep serta menunjukkan penguasaan terhadap materi melalui capaian belajar. Sholichah et al. (2022) menjelaskan bahwa aktivitas belajar memiliki keterkaitan dengan pemahaman siswa karena keterlibatan selama pembelajaran dapat memengaruhi proses penerimaan dan penguasaan materi. Sejalan dengan itu, Wahyuni dan Pratiwi (2022) menunjukkan bahwa capaian belajar siswa dapat

diklasifikasikan dengan mempertimbangkan sejumlah faktor yang menyertai proses pendidikan, sehingga kondisi akademik tidak selalu tepat apabila dinilai hanya berdasarkan satu indikator.

Dalam penelitian ini, tingkat pemahaman dipandang sebagai kondisi akademik yang dapat dikenali melalui pola capaian belajar siswa secara menyeluruh. Peneliti menempatkan nilai sepuluh mata pelajaran dan data kehadiran sebagai informasi yang dapat menggambarkan kecenderungan penguasaan siswa terhadap proses pembelajaran. Berdasarkan pendekatan tersebut, siswa dikelompokkan ke dalam kategori Paham dan Tidak Paham, sehingga tingkat pemahaman tidak sekadar dibaca dari satu nilai tertentu, tetapi dianalisis melalui karakteristik data belajar yang tersedia pada siswa SMP Swasta Tenera.

Aktivitas Belajar Siswa

Aktivitas belajar merupakan serangkaian keterlibatan siswa selama mengikuti proses pembelajaran yang dapat tercermin melalui partisipasi, kedisiplinan, kehadiran, serta capaian yang diperoleh dari berbagai kegiatan akademik. Aktivitas tersebut menjadi bagian penting dalam melihat perkembangan siswa karena proses memahami materi tidak hanya berlangsung melalui evaluasi akhir, tetapi terbentuk dari keterlibatan yang dilakukan secara berkelanjutan selama pembelajaran. Sholichah et al. (2022) menjelaskan bahwa aktivitas belajar memiliki pengaruh terhadap pemahaman siswa, khususnya ketika keterlibatan dalam pembelajaran mendukung proses penerimaan dan penguasaan materi. Sinaga (2023) juga menunjukkan bahwa prediksi kondisi akademik siswa dapat mempertimbangkan variabel lain di luar capaian nilai semata, sehingga analisis terhadap perkembangan siswa dapat dilakukan melalui karakteristik belajar yang lebih beragam.

Dalam penelitian ini, peneliti memandang aktivitas belajar sebagai informasi yang dapat digunakan untuk membaca kecenderungan tingkat pemahaman siswa berdasarkan data yang tersedia di sekolah. Aktivitas tersebut direpresentasikan melalui capaian sepuluh mata pelajaran, yaitu PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, dan PJOK, serta data kehadiran siswa. Kombinasi atribut tersebut digunakan karena peneliti menilai bahwa variasi capaian antarmata pelajaran dan konsistensi kehadiran dapat membentuk pola yang membantu membedakan siswa dalam kategori Paham dan Tidak Paham.

Educational Data Mining dan Klasifikasi

Educational Data Mining merupakan bidang analisis yang memanfaatkan data pendidikan untuk menemukan pola, hubungan, dan informasi yang dapat mendukung pemahaman terhadap proses belajar siswa. Salah satu teknik yang digunakan di dalamnya adalah klasifikasi, yaitu proses pengelompokan data ke dalam kategori tertentu berdasarkan

karakteristik yang telah dipelajari dari data sebelumnya. Romero dan Ventura (2020) menjelaskan bahwa *educational data mining* berperan dalam mengolah data yang dihasilkan dari lingkungan pendidikan untuk memperoleh informasi mengenai perilaku dan perkembangan peserta didik. Sejalan dengan itu, Yagci (2022) menunjukkan bahwa algoritma *machine learning* dapat digunakan untuk memprediksi performa akademik siswa berdasarkan pola pada data pendidikan, sedangkan Chakraborty et al. (2021) menegaskan pemanfaatan metode klasifikasi untuk mengenali kecenderungan kinerja akademik siswa.

Dalam penelitian ini, peneliti memandang data akademik siswa bukan hanya sebagai catatan administratif, tetapi sebagai sumber informasi yang dapat diolah untuk mengenali kecenderungan tingkat pemahaman. Nilai mata pelajaran dan kehadiran siswa dianalisis sebagai karakteristik yang membentuk pola tertentu, kemudian digunakan dalam proses klasifikasi untuk menentukan kategori Paham atau Tidak Paham. Pendekatan tersebut dipilih agar identifikasi kondisi siswa SMP Swasta Tenera dapat dilakukan secara lebih sistematis berdasarkan pola data belajar yang tersedia.

Algoritma Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk memperkirakan kemungkinan suatu data termasuk ke dalam kelas tertentu berdasarkan atribut yang dimilikinya. Proses klasifikasi dilakukan melalui perhitungan probabilitas awal (*prior*), peluang kemunculan atribut pada setiap kelas (*likelihood*), dan probabilitas akhir (*posterior*) sebagai dasar penentuan kelas prediksi. Nakhipova et al. (2024) menjelaskan bahwa Naive Bayes dapat dimanfaatkan dalam *machine learning* untuk memprediksi performa siswa berdasarkan karakteristik data yang dianalisis. Basri et al. (2022) juga menerapkan Naive Bayes pada klasifikasi hasil belajar, sedangkan Susanti dan Rokhim (2021) menunjukkan penggunaan algoritma tersebut untuk mengelompokkan hasil belajar siswa berdasarkan pola data yang tersedia.

Peneliti memandang Naive Bayes sebagai pendekatan yang sesuai untuk mengenali tingkat pemahaman siswa karena proses prediksi dapat dilakukan dengan mempertimbangkan sejumlah atribut belajar secara bersamaan. Nilai mata pelajaran dan kehadiran digunakan sebagai informasi masukan untuk menghitung kecenderungan probabilitas setiap siswa terhadap kategori Paham atau Tidak Paham. Melalui pendekatan tersebut, keputusan klasifikasi tidak hanya bertumpu pada satu komponen akademik, tetapi didasarkan pada pola probabilitas dari keseluruhan atribut yang digunakan dalam penelitian.

Gaussian Naive Bayes

Gaussian Naive Bayes merupakan pengembangan dari algoritma Naive Bayes yang digunakan untuk mengklasifikasikan data dengan atribut numerik kontinu melalui asumsi distribusi Gaussian atau distribusi normal. Pada pendekatan ini, karakteristik setiap atribut dalam masing-masing kelas diperkirakan berdasarkan nilai rata-rata (*mean*) dan standar deviasi, kemudian digunakan untuk menghitung peluang suatu data berada pada kelas tertentu. Ariansyah et al. (2023) menunjukkan bahwa model *Naive Bayes Gaussian* dapat digunakan dalam klasifikasi nilai siswa dengan memanfaatkan karakteristik data akademik serta pengoptimalan model. Sejalan dengan itu, Permatasari dan Nugroho (2023) menerapkan Gaussian Naive Bayes dalam sistem prediksi prestasi akademik, yang memperlihatkan relevansi pendekatan probabilistik terhadap data pendidikan berbentuk numerik.

Gaussian Naive Bayes sesuai dengan karakteristik data siswa SMP Swasta Tenera karena atribut yang dianalisis berupa nilai sepuluh mata pelajaran dan data kehadiran dalam bentuk numerik. Setiap atribut memiliki variasi nilai yang berbeda pada masing-masing siswa, sehingga distribusi data tersebut dapat digunakan untuk memperkirakan kecenderungan siswa termasuk dalam kategori Paham atau Tidak Paham. Penggunaan pendekatan Gaussian dipilih agar proses klasifikasi mempertimbangkan pola numerik dari keseluruhan atribut belajar, bukan hanya membandingkan nilai siswa secara langsung.

Penelitian Terdahulu

Penerapan metode klasifikasi dalam menganalisis capaian akademik siswa telah dilakukan pada berbagai jenjang pendidikan dengan memanfaatkan karakteristik data yang berbeda. Chakraborty et al. (2021) menggunakan sejumlah *machine learning classifiers* untuk memprediksi performa akademik siswa, sedangkan Yagci (2022) mengkaji prediksi kinerja akademik melalui pendekatan *educational data mining*.

Pada penggunaan Naive Bayes, Firmansyah dan Yulianto (2023) menerapkan metode tersebut untuk memprediksi hasil belajar pada tingkat sekolah dasar, sementara Ariansyah et al. (2023) mengembangkan klasifikasi nilai siswa melalui *Naive Bayes Gaussian* yang dipadukan dengan *tuning model* dan seleksi fitur. Penelitian lain oleh Riyanto dan Ompusunggu (2024) memanfaatkan Naive Bayes untuk klasifikasi hasil belajar siswa, sedangkan Nakhipova et al. (2024) menerapkan algoritma yang sama dalam prediksi performa siswa berbasis *machine learning*. Perbandingan penelitian yang memiliki keterkaitan dengan penelitian ini disajikan pada Tabel 1.

Tabel 1. Perbandingan Penelitian Terdahulu.

Peneliti	Fokus Penelitian	Metode	Hasil/Kontribusi Utama	Keterkaitan dengan Penelitian Ini
Chakraborty et al. (2021)	Prediksi performa akademik siswa	<i>Machine Learning Classifiers</i>	Menunjukkan pemanfaatan algoritma klasifikasi untuk mengenali performa akademik siswa	Menjadi dasar penggunaan data pendidikan dalam prediksi kondisi akademik siswa
Yagci (2022)	Prediksi kinerja akademik siswa melalui <i>educational data mining</i>	Algoritma <i>Machine Learning</i>	Membandingkan pendekatan pembelajaran mesin dalam memprediksi performa akademik	Memperkuat penggunaan pola data pendidikan sebagai dasar prediksi tingkat pemahaman
Firmansyah & Yulianto (2023)	Prediksi hasil belajar siswa sekolah dasar	Naive Bayes	Menerapkan klasifikasi probabilistik untuk memprediksi hasil belajar siswa	Memiliki kedekatan pada objek siswa, tetapi penelitian ini berfokus pada tingkat pemahaman siswa SMP
Ariansyah et al. (2023)	Klasifikasi nilai siswa	<i>Naive Bayes Gaussian, tuning model</i> , dan seleksi fitur	Mengembangkan model klasifikasi nilai siswa dengan pengoptimalan Gaussian Naive Bayes	Mendukung penggunaan Gaussian Naive Bayes pada atribut akademik berbentuk numerik
Riyanto & Ompusunggu (2024)	Klasifikasi hasil belajar siswa	Naive Bayes	Memanfaatkan data mining untuk mengelompokkan hasil belajar berdasarkan pola data siswa	Relevan dari sisi klasifikasi hasil belajar, sedangkan penelitian ini memprediksi kategori Paham dan Tidak Paham
Nakhipova et al. (2024)	Prediksi performa siswa berbasis <i>machine learning</i>	Naive Bayes Classifier	Menunjukkan pemanfaatan Naive Bayes dalam prediksi performa siswa	Memperkuat pemilihan Naive Bayes untuk mengolah karakteristik belajar siswa

Sumber: Data Penelitian, 2026

Berdasarkan perbandingan tersebut, penelitian terdahulu menunjukkan bahwa metode klasifikasi telah digunakan untuk memprediksi performa akademik, hasil belajar, dan nilai siswa pada berbagai konteks pendidikan. Meskipun demikian, penelitian ini menempatkan fokus yang berbeda pada prediksi tingkat pemahaman siswa SMP Swasta Tenera dengan menggabungkan data dari tiga jenjang kelas, yaitu VII, VIII, dan IX, dalam satu proses analisis.

Data penelitian mencakup 90 siswa dengan atribut berupa nilai sepuluh mata pelajaran dan kehadiran, kemudian diklasifikasikan ke dalam kategori Paham dan Tidak Paham menggunakan Gaussian Naive Bayes. Posisi tersebut menjadi pembeda penelitian karena analisis tidak hanya diarahkan pada prediksi individual, tetapi juga digunakan untuk melihat

distribusi tingkat pemahaman antarkelas berdasarkan pola aktivitas belajar siswa. Rancangan ini konsisten dengan karakteristik data dan tujuan yang tercantum dalam naskah sumber.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan penerapan teknik *machine learning* untuk memprediksi tingkat pemahaman siswa melalui algoritma Gaussian Naive Bayes. Penelitian dilakukan di SMP Swasta Tenera dengan menggunakan data akademik tahun ajaran 2023/2024 yang mencakup 90 siswa dari tiga jenjang kelas, yaitu 28 siswa kelas VII, 23 siswa kelas VIII, dan 39 siswa kelas IX. Seluruh data siswa digunakan sebagai sampel penelitian. Data diperoleh melalui teknik dokumentasi terhadap arsip akademik sekolah yang memuat nilai sepuluh mata pelajaran, meliputi PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, dan PJOK, serta data kehadiran siswa. Variabel target dibedakan menjadi kategori Paham untuk siswa dengan nilai rata-rata ≥ 75 dan Tidak Paham untuk siswa dengan nilai rata-rata < 75 . Susunan data, jumlah siswa, atribut, dan batas kategorisasi tersebut mengikuti data pada naskah penelitian sumber.

Tahapan penelitian dimulai dari pengumpulan data akademik, pemeriksaan kelengkapan data, penyesuaian format atribut, penentuan label kelas, perhitungan probabilitas awal, pembentukan parameter distribusi Gaussian, proses prediksi, dan evaluasi hasil klasifikasi. Tahap *preprocessing* dilakukan untuk memastikan setiap atribut dapat digunakan dalam proses analisis, termasuk pemeriksaan *missing value* dan konversi format data apabila diperlukan. Selanjutnya, setiap siswa diberi label Paham atau Tidak Paham berdasarkan batas nilai yang telah ditetapkan. Wijaya dan Kusnandar (2022) menjelaskan bahwa tahap persiapan data memiliki peran penting sebelum proses klasifikasi dilakukan karena kualitas dan kesesuaian data dapat memengaruhi hasil prediksi.

Proses klasifikasi menggunakan algoritma Naive Bayes yang menentukan probabilitas suatu kelas berdasarkan atribut yang diamati. Nakhipova et al. (2024) menjelaskan bahwa Naive Bayes dapat digunakan untuk prediksi performa siswa melalui perhitungan probabilitas terhadap karakteristik data yang tersedia. Secara matematis, hubungan probabilitas tersebut dinyatakan melalui Teorema Bayes sebagai berikut:

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

Karena atribut penelitian berbentuk numerik kontinu, perhitungan *likelihood* dilakukan menggunakan distribusi Gaussian. Pendekatan ini memanfaatkan nilai rata-rata dan standar deviasi setiap atribut pada masing-masing kelas untuk memperkirakan peluang kemunculan

suatu nilai. Ariansyah et al. (2023) menunjukkan penggunaan *Naive Bayes Gaussian* dalam klasifikasi nilai siswa pada data akademik. Fungsi densitas probabilitas Gaussian dirumuskan sebagai berikut:

$$P(x_i | C) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \dots\dots\dots (ii)$$

Untuk mengurangi risiko *underflow* numerik akibat perkalian sejumlah probabilitas bernilai kecil, proses penentuan kelas menggunakan pendekatan *log-posterior* sebagai berikut:

$$\hat{C} = \arg \max_C [\log P(C) + \sum_{i=1}^n \log P(x_i | C)] \dots\dots\dots (iii)$$

Hasil prediksi selanjutnya dievaluasi menggunakan *confusion matrix* untuk membandingkan kelas aktual dengan kelas hasil prediksi. Komponen evaluasi terdiri atas *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan komponen tersebut, performa model diukur menggunakan akurasi, presisi, *recall*, dan *F1 score*. Akurasi menunjukkan proporsi keseluruhan prediksi yang benar, presisi mengukur ketepatan prediksi kelas positif, *recall* menunjukkan kemampuan model mengenali data positif yang sebenarnya, sedangkan *F1-score* menggambarkan keseimbangan antara presisi dan *recall*. Rumus evaluasi model dinyatakan sebagai berikut:

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

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (v)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (vi)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots\dots (vii)$$

4. HASIL DAN PEMBAHASAN

Statistik Deskriptif Data

Data penelitian terdiri atas 90 siswa SMP Swasta Tenera dari tiga jenjang kelas, yaitu 28 siswa kelas VII, 23 siswa kelas VIII, dan 39 siswa kelas IX. Setiap data siswa memuat nilai sepuluh mata pelajaran yang terdiri atas PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, dan PJOK, serta data kehadiran. Berdasarkan pengolahan data, rata-rata nilai keseluruhan siswa mencapai 74,10 dengan rata-rata kehadiran sebesar 75,78%. Status tingkat pemahaman ditentukan berdasarkan nilai rata-rata siswa, yaitu kategori Paham untuk nilai rata-rata ≥ 75 dan Tidak Paham untuk nilai rata-rata < 75 . Hasil pengelompokan menunjukkan bahwa 51 siswa atau 56,7% berada pada kategori Paham,

sedangkan 39 siswa atau 43,3% termasuk kategori Tidak Paham. Distribusi tingkat pemahaman pada setiap jenjang kelas disajikan pada Tabel 2.

Tabel 2. Distribusi Status Pemahaman Siswa per Kelas.

Kelas	Jumlah Siswa	Paham	Tidak Paham	Rata-rata Nilai	Rata-rata Kehadiran
VII	28	12 (42,9%)	16 (57,1%)	72,76	76,29%
VIII	23	14 (60,9%)	9 (39,1%)	74,00	75,57%
IX	39	25 (64,1%)	14 (35,9%)	75,12	75,54%
Total	90	51 (56,7%)	39 (43,3%)	74,10	75,78%

Sumber: Data Penelitian, 2026

Berdasarkan Tabel 2, distribusi tingkat pemahaman menunjukkan perbedaan pada setiap jenjang kelas. Kelas VII memiliki proporsi siswa Paham paling rendah, yaitu 12 dari 28 siswa atau 42,9%, sedangkan 16 siswa atau 57,1% berada pada kategori Tidak Paham. Kondisi berbeda terlihat pada kelas VIII, dengan 14 dari 23 siswa atau 60,9% termasuk kategori Paham dan 9 siswa atau 39,1% tergolong Tidak Paham. Proporsi pemahaman tertinggi ditemukan pada kelas IX, yaitu 25 dari 39 siswa atau 64,1%, sementara 14 siswa atau 35,9% berada pada kategori Tidak Paham. Pola tersebut memperlihatkan adanya peningkatan proporsi siswa Paham dari kelas VII menuju kelas IX, meskipun rata-rata kehadiran ketiga jenjang relatif berdekatan.

Temuan tersebut menunjukkan bahwa tingkat kehadiran tidak dapat secara langsung diperlakukan sebagai satu-satunya penentu pemahaman siswa. Kelas VII memiliki rata-rata kehadiran tertinggi sebesar 76,29%, tetapi justru mencatat proporsi siswa Paham paling rendah. Sebaliknya, kelas IX memiliki rata-rata kehadiran 75,54% dengan proporsi siswa Paham tertinggi sebesar 64,1%. Dari sudut pandang peneliti, kondisi ini mengindikasikan bahwa prediksi tingkat pemahaman perlu mempertimbangkan pola sejumlah atribut secara bersamaan, terutama variasi capaian pada sepuluh mata pelajaran dan kehadiran. Pandangan tersebut selaras dengan Sholichah et al. (2022) yang menempatkan aktivitas belajar sebagai bagian yang berkaitan dengan pemahaman siswa, serta Yagci (2022) yang menunjukkan bahwa pola data pendidikan dapat dimanfaatkan dalam prediksi performa akademik melalui pendekatan *machine learning*.

Perhitungan Probabilitas Prior

Probabilitas *prior* menunjukkan peluang awal setiap kelas sebelum atribut nilai mata pelajaran dan kehadiran siswa diperhitungkan dalam proses klasifikasi. Berdasarkan 90 data siswa SMP Swasta Tenera, sebanyak 51 siswa berada pada kategori Paham, sedangkan 39 siswa termasuk kategori Tidak Paham. Distribusi tersebut menghasilkan probabilitas awal

sebesar 0,5667 untuk kelas Paham dan 0,4333 untuk kelas Tidak Paham. Perhitungan probabilitas *prior* dilakukan berdasarkan proporsi jumlah anggota setiap kelas terhadap keseluruhan data penelitian. Data jumlah kelas tersebut konsisten dengan naskah sumber.

$$P(\text{Paham}) = \frac{51}{90} = 0,5667 \dots\dots\dots(\text{viii})$$

$$P(\text{Tidak Paham}) = \frac{39}{90} = 0,4333 \dots\dots\dots(\text{ix})$$

Hasil perhitungan probabilitas *prior* disajikan pada Tabel 3.

Tabel 3. Probabilitas Prior Kelas Pemahaman Siswa.

Kelas	Jumlah Siswa	Probabilitas Prior	Persentase
Paham	51	0,5667	56,67%
Tidak Paham	39	0,4333	43,33%
Total	90	1,0000	100,00%

Sumber: Data Penelitian, 2026

Berdasarkan Tabel 3, kelas Paham memiliki probabilitas *prior* lebih tinggi sebesar 0,5667 dibandingkan kelas Tidak Paham sebesar 0,4333. Selisih probabilitas kedua kelas mencapai 0,1334 atau sekitar 13,34 poin persentase. Kondisi ini menunjukkan bahwa sebelum nilai setiap atribut dianalisis, data siswa memiliki kecenderungan awal yang lebih besar untuk berada pada kategori Paham. Meskipun demikian, distribusi tersebut tidak memperlihatkan dominasi kelas yang terlalu ekstrem karena kedua kategori masih memiliki proporsi yang relatif berdekatan. Temuan ini penting dalam proses klasifikasi karena probabilitas *prior* menjadi salah satu komponen yang dikombinasikan dengan nilai *likelihood* setiap atribut untuk memperoleh probabilitas akhir kelas.

Analisis Mean dan Standar Deviasi Setiap Atribut

Kedua parameter tersebut digunakan dalam Gaussian Naive Bayes untuk membentuk distribusi probabilitas masing-masing atribut berdasarkan karakteristik kelasnya. Nilai *mean* menunjukkan kecenderungan pusat data, sedangkan standar deviasi menggambarkan tingkat penyebaran nilai terhadap rata-rata. Perhitungan dilakukan secara terpisah terhadap sebelas atribut yang terdiri atas nilai PPKN, Bahasa Indonesia, Matematika, IPA, IPS, Agama, TIK, Bahasa Inggris, Prakarya, PJOK, dan kehadiran. Hasil perhitungan parameter Gaussian berdasarkan data penelitian disajikan pada Tabel 4.

Tabel 4. Mean dan Standar Deviasi Setiap Atribut Berdasarkan Kelas Pemahaman.

Atribut	μ Paham	σ Paham	μ Tidak Paham	σ Tidak Paham
PPKN	78,4902	4,8554	62,5897	15,7693
Bahasa Indonesia	79,3333	2,7905	71,1282	14,6596
Matematika	78,2941	5,5256	61,7179	15,6608
IPA	79,3333	2,7905	71,1282	14,6596
IPS	78,0392	6,0892	73,8718	9,1943
Agama	76,3529	9,1276	75,1282	10,2884
TIK	75,1765	7,2214	75,6667	7,6031

Bahasa Inggris	77,6471	3,8721	75,1282	10,5660
Prakarya	74,8039	8,3571	73,3077	7,4061
PJOK	76,8431	8,9093	77,4359	2,6036
Kehadiran	79,3333	2,7905	71,1282	14,6596

Sumber: Data Penelitian, 2026

Dari Tabel 4, perbedaan *mean* paling besar terlihat pada atribut Matematika dan PPKN. Nilai rata-rata Matematika pada kelas Paham mencapai 78,2941, sedangkan kelas Tidak Paham sebesar 61,7179, dengan selisih sekitar 16,58 poin. Pada atribut PPKN, kelas Paham memiliki rata-rata 78,4902 dan kelas Tidak Paham 62,5897, sehingga terdapat perbedaan sekitar 15,90 poin. Pola tersebut menunjukkan bahwa kedua atribut mempunyai pemisahan nilai yang relatif kuat antara dua kategori pemahaman. Bahasa Indonesia dan IPA juga memperlihatkan perbedaan rata-rata sekitar 8,21 poin, sedangkan IPS memiliki selisih sekitar 4,17 poin. Berdasarkan pola ini, atribut dengan perbedaan *mean* yang lebih besar berpotensi memberikan informasi yang lebih kuat dalam membedakan karakteristik kelas Paham dan Tidak Paham.

Dari sisi penyebaran data, kelas Tidak Paham menunjukkan variasi yang lebih besar pada beberapa atribut utama. Standar deviasi PPKN mencapai 15,7693 dan Matematika sebesar 15,6608, jauh lebih tinggi dibandingkan kelas Paham yang masing-masing bernilai 4,8554 dan 5,5256. Perbedaan tersebut mengindikasikan bahwa kelompok Tidak Paham memiliki capaian yang lebih heterogen, sedangkan nilai siswa pada kelas Paham cenderung lebih terkonsentrasi di sekitar rata-ratanya. Dari sudut pandang peneliti, pola ini memperkuat alasan penggunaan Gaussian Naive Bayes karena proses klasifikasi tidak hanya mempertimbangkan besar kecilnya nilai rata-rata, tetapi juga memperhitungkan variasi distribusi setiap atribut pada masing-masing kelas. Pendekatan tersebut selaras dengan Ariansyah et al. (2023) yang menggunakan Gaussian Naive Bayes pada klasifikasi nilai siswa dan menunjukkan relevansi parameter numerik dalam pembentukan model klasifikasi akademik.

Perhitungan Gaussian Naive Bayes

Setelah nilai *prior*, *mean*, dan standar deviasi diperoleh, tahap berikutnya adalah menghitung probabilitas setiap atribut menggunakan fungsi densitas Gaussian. Sebagai ilustrasi, perhitungan dilakukan pada data siswa Abdillah Iqbal Sinurat kelas VII dengan nilai PPKN 63, Bahasa Indonesia 80, Matematika 63, IPA 80, IPS 80, Agama 80, TIK 80, Bahasa Inggris 70, Prakarya 77, PJOK 80, dan kehadiran 80%. Siswa tersebut memiliki rata-rata aktual 73,2, sehingga berdasarkan batas klasifikasi nilai rata-rata < 75 termasuk kategori Tidak Paham. Data contoh dan parameter perhitungannya mengikuti naskah sumber. Tahap awal dilakukan dengan menghitung nilai logaritma probabilitas *prior* pada kedua kelas. Berdasarkan

hasil pada Subbab 4.2, probabilitas awal kelas Paham sebesar 0,5667 dan kelas Tidak Paham sebesar 0,4333. Nilai tersebut ditransformasikan ke bentuk logaritmik sebagai berikut:

$$\log P(\text{Paham}) = \log(0,5667) = -0,5686 \dots \dots \dots (x)$$

$$\log P(\text{Tidak Paham}) = \log(0,4333) = -0,8362 \dots \dots \dots (xi)$$

Selanjutnya, nilai setiap atribut siswa dihitung terhadap distribusi kelas Paham dan Tidak Paham menggunakan parameter *mean* dan standar deviasi pada Tabel 4. Hasil perhitungan *likelihood* Gaussian untuk data siswa tersebut disajikan pada Tabel 5.

Tabel 5. Hasil Perhitungan Gaussian Likelihood pada Data Siswa Uji.

Atribut	(x _i)	μ Paham	σ Paham	μ Tidak Paham	σ Tidak Paham	P((x _i) Paham)	P((x _i) Tidak Paham)
PPKN	63	78,4902	4,8554	62,5897	15,7693	0,0000	0,0253
Bahasa Indonesia	80	79,3333	2,7905	71,1282	14,6596	0,1427	0,0254
Matematika	63	78,2941	5,5256	61,7179	15,6608	0,0000	0,0253
IPA	80	79,3333	2,7905	71,1282	14,6596	0,1427	0,0254
IPS	80	78,0392	6,0892	73,8718	9,1943	0,0603	0,0392
Agama	80	76,3529	9,1276	75,1282	10,2884	0,0415	0,0381
TIK	80	75,1765	7,2214	75,6667	7,6031	0,0521	0,0524
Bahasa Inggris	70	77,6471	3,8721	75,1282	10,5660	0,0003	0,0358
Prakarya	77	74,8039	8,3571	73,3077	7,4061	0,0473	0,0521
PJOK	80	76,8431	8,9093	77,4359	2,6036	0,0417	0,1068
Kehadiran	80	79,3333	2,7905	71,1282	14,6596	0,1427	0,0254

Sumber: Data Penelitian, 2026

Berdasarkan Tabel 5, setiap atribut menghasilkan nilai *likelihood* yang berbeda terhadap kedua kelas. Nilai Bahasa Indonesia 80, IPA 80, dan kehadiran 80 memiliki probabilitas sebesar 0,1427 pada kelas Paham, sedangkan pada kelas Tidak Paham masing-masing bernilai sekitar 0,0254. Sebaliknya, nilai PPKN 63 dan Matematika 63 memiliki peluang yang sangat kecil terhadap distribusi kelas Paham karena posisinya relatif jauh dari *mean* kelas tersebut. Pada kelas Tidak Paham, kedua atribut menghasilkan *likelihood* sekitar 0,0253. Pola ini menunjukkan bahwa satu siswa dapat memiliki atribut yang mendukung kelas

berbeda, sehingga keputusan akhir tidak ditentukan hanya dari satu mata pelajaran. Seluruh nilai *likelihood* kemudian digabungkan dengan probabilitas *prior* menggunakan pendekatan *log-posterior*. Berdasarkan hasil perhitungan pada data siswa uji, diperoleh:

$$\log P (\text{Paham} | X) = -43,82 \dots \dots \dots (xii)$$

$$\log P (\text{Tidak Paham} | X) = -28,47 \dots \dots \dots (xiii)$$

Karena nilai *log-posterior* kelas Tidak Paham sebesar -28,47 lebih tinggi daripada kelas Paham sebesar -43,82, model menghasilkan prediksi Tidak Paham. Hasil tersebut sesuai dengan status aktual siswa yang memiliki rata-rata nilai 73,2. Dengan demikian, contoh perhitungan ini menunjukkan bahwa Gaussian Naive Bayes mampu menggabungkan kontribusi seluruh atribut dalam menentukan kelas akhir. Dari sudut pandang peneliti, mekanisme tersebut relevan untuk data akademik karena seorang siswa dapat menunjukkan capaian tinggi pada beberapa mata pelajaran tetapi rendah pada atribut lainnya. Temuan ini sejalan dengan Ariansyah et al. (2023) yang menempatkan Gaussian Naive Bayes sebagai pendekatan klasifikasi nilai siswa berbasis karakteristik numerik, serta Nakhipova et al. (2024) yang menunjukkan pemanfaatan Naive Bayes dalam prediksi performa siswa.

Hasil Prediksi Tingkat Pemahaman Siswa

Setelah parameter probabilitas *prior*, *mean*, dan standar deviasi diperoleh, model Gaussian Naive Bayes diterapkan pada seluruh data siswa untuk menghasilkan kategori prediksi tingkat pemahaman. Proses klasifikasi dilakukan dengan membandingkan nilai probabilitas akhir antara kelas Paham dan Tidak Paham berdasarkan sebelas atribut yang digunakan, yaitu nilai sepuluh mata pelajaran dan kehadiran. Kelas dengan nilai probabilitas akhir lebih tinggi ditetapkan sebagai hasil prediksi. Berdasarkan pengolahan terhadap 90 data siswa, model menghasilkan 75 prediksi yang sesuai dengan status aktual dan 15 prediksi yang berbeda. Ringkasan hasil prediksi siswa disajikan pada Tabel 6 sesuai data hasil pada naskah sumber.

Tabel 6. Ringkasan Hasil Prediksi Gaussian Naive Bayes.

No.	Nama Siswa	Kelas	Kehadiran	Rata-rata	Status Aktual	Hasil Prediksi	Keterangan
1	Abdillah Iqbal Sinurat	VII	80	73,2	Tidak Paham	Tidak Paham	Benar
2	Ade Mulya Akbar	VII	80	74,6	Tidak Paham	Paham	Salah
3	Adinda Ardilabina Br. Munthe	VII	85	80,8	Paham	Paham	Benar
...	...	...	...	...	...	...	...
90	Zhaira Anggraini	IX	78	76,2	Paham	Paham	Benar

Sumber: Data Penelitian, 2026

Berdasarkan Tabel 6, hasil prediksi menunjukkan bahwa model tidak selalu menghasilkan kelas yang sama dengan status aktual siswa. Pada data Abdillah Iqbal Sinurat, nilai rata-rata sebesar 73,2 menempatkan siswa pada kategori aktual Tidak Paham dan model juga menghasilkan prediksi Tidak Paham. Kondisi berbeda ditemukan pada Ade Mulya Akbar yang memiliki rata-rata 74,6 dengan status aktual Tidak Paham, tetapi diprediksi sebagai Paham. Sementara itu, Adinda Ardilabina Br. Munthe dengan rata-rata 80,8 berhasil dikenali secara tepat sebagai Paham. Perbedaan tersebut memperlihatkan bahwa keputusan Gaussian Naive Bayes tidak hanya mengikuti nilai rata-rata akhir, tetapi dibentuk dari pola distribusi setiap atribut terhadap karakteristik kedua kelas.

Secara keseluruhan, sebanyak 75 dari 90 data siswa berhasil diklasifikasikan secara tepat, sedangkan 15 data lainnya mengalami kesalahan prediksi. Dengan demikian, proporsi prediksi yang sesuai mencapai 83,33%, sementara ketidaksesuaian prediksi sebesar 16,67%. Hasil ini menunjukkan bahwa sebagian besar pola akademik siswa dapat dikenali oleh model, meskipun masih terdapat sejumlah data dengan karakteristik yang saling beririsan antara kelas Paham dan Tidak Paham. Kondisi tersebut dapat terjadi ketika seorang siswa memiliki nilai tinggi pada beberapa mata pelajaran, tetapi rendah pada atribut lainnya, sehingga kombinasi probabilitas Gaussian menghasilkan kelas yang berbeda dari status aktual yang ditentukan melalui batas rata-rata nilai.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma Gaussian Naive Bayes dapat diterapkan untuk memprediksi tingkat pemahaman siswa SMP Swasta Tenera berdasarkan pola nilai sepuluh mata pelajaran dan kehadiran. Dari 90 siswa yang dianalisis, sebanyak 51 siswa atau 56,7% berada pada kategori Paham dan 39 siswa atau 43,3% termasuk kategori Tidak Paham. Distribusi antarkelas menunjukkan bahwa proporsi siswa Paham meningkat dari kelas VII sebesar 42,9%, kelas VIII sebesar 60,9%, hingga kelas IX sebesar 64,1%. Model menghasilkan 75 prediksi benar dari 90 data atau setara dengan akurasi 83,33%, yang menunjukkan bahwa sebagian besar karakteristik tingkat pemahaman siswa dapat dikenali melalui kombinasi atribut akademik yang digunakan. Perbedaan *mean* yang cukup besar pada atribut Matematika dan PPKN juga memperlihatkan adanya pola numerik yang membantu membedakan karakteristik kedua kelas. Secara keseluruhan, hasil penelitian menunjukkan bahwa Gaussian Naive Bayes dapat menjadi pendekatan analitis untuk mendukung identifikasi awal kondisi pemahaman siswa, meskipun hasil prediksi tetap perlu diposisikan sebagai informasi pendukung bagi guru dan pihak sekolah, bukan sebagai satu-satunya dasar penilaian akademik. Dari temuan tersebut,

sekolah disarankan memanfaatkan hasil prediksi sebagai bahan pendukung untuk mengenali siswa yang membutuhkan perhatian pembelajaran lebih awal, terutama pada jenjang dengan proporsi Tidak Paham yang lebih tinggi. Penelitian berikutnya dapat memperluas jumlah data dan melibatkan atribut lain yang relevan, seperti keaktifan di kelas, penyelesaian tugas, hasil asesmen formatif, atau faktor pembelajaran lainnya agar karakteristik siswa dapat direpresentasikan secara lebih beragam. Pengembangan selanjutnya juga dapat membandingkan Gaussian Naive Bayes dengan algoritma klasifikasi lain, seperti *Decision Tree*, *K-Nearest Neighbor*, atau *Support Vector Machine*, serta menggunakan skema pengujian yang lebih kuat untuk menilai kestabilan model pada data yang belum pernah dianalisis sebelumnya.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada pihak SMP Swasta Tenera yang telah memberikan dukungan dan kesempatan dalam pelaksanaan penelitian, khususnya dalam penyediaan data akademik yang digunakan sebagai bahan analisis. Apresiasi juga disampaikan kepada seluruh pihak yang telah memberikan arahan, masukan, dan dukungan selama proses penelitian hingga penyusunan artikel ini. Dukungan tersebut berperan penting dalam penyelesaian penelitian mengenai prediksi tingkat pemahaman siswa berdasarkan aktivitas belajar menggunakan algoritma Gaussian Naive Bayes.

DAFTAR REFERENSI

- Ariansyah, M. H., Fitri, E. N., & Winarno, S. (2023). Peningkatan kinerja model klasifikasi kelas siswa menggunakan model tuning Naive Bayes Gaussian dan seleksi fitur. *Jurnal Teknik Informatika (JUTIF)*, 4(3), 493–501. <https://doi.org/10.52436/1.jutif.2023.4.3.737>
- Basri, H., Azis, M. S., Malau, Y., Fridayanthie, E. W., Rizal, K., & Rianto, H. (2022). Penerapan optimasi gerombolan partikel pada algoritma Naive Bayes untuk klasifikasi hasil belajar. *Sistem Informasi Pendidik dan Profesi (ISBI)*, 6(2), 97–106. <https://doi.org/10.51211/isbi.v6i2.1752>
- Chakraborty, P., Bhattacharyya, T., Bag, R., & Bhowmick, A. (2021). Memprediksi kinerja akademik siswa menggunakan pengklasifikasi pembelajaran mesin. *International Journal of Engineering and Technology*, 7(4), 34–39.
- Firmansyah, & Yulianto, A. (2023). Prediksi hasil belajar menggunakan Naive Bayes classifier pada tingkat sekolah dasar. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 7(2), 1174–1182. <https://doi.org/10.33395/remik.v7i2.12375>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

- Nakhipova, V., Kerimbekov, Y., Umarova, Z., Suleimenova, L., Botayeva, S., Ibashova, A., & Zhumatayev, N. (2024). Use of the Naive Bayes classifier algorithm in machine learning for student performance prediction. *International Journal of Information and Education Technology*, 14(1), 92–98. <https://doi.org/10.18178/ijiet.2024.14.1.2028>
- Permatasari, E., & Nugroho, A. (2023). Sistem prediksi prestasi akademik mahasiswa menggunakan Gaussian Naive Bayes. *Jurnal Teknologi Informasi*, 9(2), 88–97.
- Riyanto, A., & Ompusunggu, E. S. (2024). Implementasi data mining untuk mengklasifikasi hasil belajar siswa dengan metode Naive Bayes. *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, 7(2), 129–141.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), Article e1355. <https://doi.org/10.1002/widm.1355>
- Sholichah, I. S., Hartono, B., & Putri, A. R. (2022). Pengaruh aktivitas belajar terhadap pemahaman siswa dalam pembelajaran daring. *Jurnal Pendidikan dan Kebudayaan*, 7(2), 112–125.
- Sinaga, B. (2023). Variabel nonakademik untuk memprediksi prestasi siswa dengan data mining menggunakan metode Naive Bayes. *Nuansa Informatika*, 17(2), 35–41. <https://doi.org/10.25134/ilkom.v17i2.26>
- Susanti, M., & Rokhim, F. (2021). Penerapan algoritma Naive Bayes untuk klasifikasi hasil belajar siswa. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(3), 512–521.
- Wahyuni, D., & Pratiwi, L. (2022). Klasifikasi prestasi belajar siswa menggunakan Naive Bayes berdasarkan faktor internal dan eksternal. *Jurnal Ilmu Komputer dan Informasi*, 15(1), 28–36.
- Wijaya, H., & Kusnandar, D. (2022). Pemrosesan data dan klasifikasi untuk prediksi kinerja akademik siswa. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 9(4), 731–740.
- Yagci, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9, Article 11. <https://doi.org/10.1186/s40561-022-00192-z>