



Penerapan Algoritma *Naive Bayes* untuk Evaluasi Kedisiplinan Siswa pada Sistem Kehadiran Terintegrasi Berbasis *Web*

Nina Mumtazia¹, Romi Andrianto²

^{1,2} Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang
ninamumtazia47@gmail.com, dosen02391@unpam.ac.id

Abstract

Student attendance is an important indicator of discipline in the learning process; however, at SDN Mekar Jaya attendance data has mainly been used as administrative records, so that discipline evaluation still tends to be subjective. This study aims to build a web-based integrated attendance system that applies the Naive Bayes algorithm to objectively evaluate student discipline based on attendance data. Data were collected through literature review, observation, interviews, and documentation, while the system was developed using the Agile Development Methodology. The Gaussian Naive Bayes variant was employed because it is able to process continuous numerical data, namely the percentages of attendance, sick leave, permitted absence, and unexcused absence. The dataset consisted of 57 attendance records of fourth-grade students at SDN Mekar Jaya collected over one semester, divided into 48 training data and 9 testing data. Evaluation using a confusion matrix shows that the model classifies students into the Disciplined, Moderately Disciplined, and Less Disciplined categories with an accuracy of 88.89%, correctly classifying 8 of the 9 testing data. The Disciplined and Moderately Disciplined classes achieved a precision of 100%, while the Less Disciplined class obtained 0% because one Moderately Disciplined sample was misclassified. These results indicate that the Gaussian Naive Bayes algorithm performs well in evaluating student discipline from attendance data, and that the developed system supports attendance management and discipline evaluation in a more objective, efficient, and structured manner than the manual process.

Keywords: Gaussian Naive Bayes, Student Discipline, Integrated Attendance System, Classification, Discipline Evaluation.

Abstrak

Kehadiran siswa merupakan indikator penting kedisiplinan dalam proses pembelajaran, namun di SDN Mekar Jaya data kehadiran masih dimanfaatkan sebatas catatan administratif sehingga evaluasi kedisiplinan siswa cenderung dilakukan secara subjektif dan belum didukung oleh analisis data yang sistematis. Penelitian ini bertujuan membangun sistem kehadiran terintegrasi berbasis web yang menerapkan algoritma Naive Bayes untuk mengevaluasi tingkat kedisiplinan siswa secara objektif berdasarkan data kehadiran. Data penelitian dikumpulkan melalui studi pustaka, observasi, wawancara, dan dokumentasi, sedangkan pengembangan sistem menggunakan metode Agile Development Methodology. Algoritma yang diterapkan adalah Gaussian Naive Bayes karena mampu mengolah data numerik kontinu berupa persentase hadir, sakit, izin, dan alfa. Dataset penelitian terdiri dari 57 data absensi siswa kelas 4 SDN Mekar Jaya selama satu semester yang dibagi menjadi 48 data latih dan 9 data uji. Hasil pengujian menggunakan confusion matrix menunjukkan bahwa model berhasil mengklasifikasikan siswa ke dalam kategori Disiplin, Cukup Disiplin, dan Kurang Disiplin dengan nilai akurasi sebesar 88,89%, yaitu 8 dari 9 data uji terklasifikasi dengan benar. Kelas Disiplin dan Cukup Disiplin memperoleh precision sebesar 100%, sedangkan kelas Kurang Disiplin sebesar 0% karena terdapat satu data Cukup Disiplin yang diprediksi sebagai Kurang Disiplin. Hasil tersebut menunjukkan bahwa algoritma Gaussian Naive Bayes memiliki kinerja yang baik dalam mengevaluasi kedisiplinan siswa berdasarkan data kehadiran. Dapat disimpulkan bahwa sistem yang dibangun mampu mendukung proses pengelolaan absensi dan evaluasi kedisiplinan siswa secara lebih objektif, cepat, dan terstruktur dibandingkan dengan proses manual.

Kata Kunci: Gaussian Naive Bayes, Kedisiplinan Siswa, Sistem Kehadiran Terintegrasi, Klasifikasi, Evaluasi Kedisiplinan.

1. Pendahuluan

Pendidikan merupakan salah satu aspek fundamental dalam meningkatkan kualitas sumber daya manusia. Melalui pendidikan, siswa dituntut tidak hanya menguasai pengetahuan akademik, tetapi juga membentuk karakter, sikap, dan kedisiplinan. Kedisiplinan menjadi salah satu indikator penting keberhasilan proses pembelajaran karena berkaitan langsung dengan tanggung jawab dan keterlibatan siswa dalam kegiatan belajar, dan salah satu indikator yang dapat diamati secara langsung adalah tingkat kehadiran siswa di sekolah. Kehadiran mencerminkan komitmen siswa terhadap proses pembelajaran, sedangkan ketidakhadiran atau keterlambatan yang berulang dapat menjadi indikasi adanya permasalahan kedisiplinan.

Data kehadiran siswa pada dasarnya memiliki potensi besar untuk dianalisis sebagai dasar evaluasi perilaku siswa, namun dalam praktiknya data tersebut masih sering dimanfaatkan secara terbatas hanya sebagai pencatatan administratif. Penelitian terbaru menunjukkan bahwa pemanfaatan data pendidikan tanpa analisis yang sistematis dapat menyebabkan evaluasi siswa menjadi kurang optimal dan cenderung subjektif (Salsabila & Basri, 2026). Di

banyak sekolah, termasuk SDN Mekar Jaya, sistem kehadiran siswa telah diterapkan baik secara manual maupun berbasis web, tetapi data yang dihasilkan belum diolah lebih lanjut untuk mengevaluasi tingkat kedisiplinan siswa sehingga sekolah belum memiliki dasar yang objektif dalam menilai perilaku kedisiplinan maupun dalam melakukan intervensi yang tepat.

Seiring dengan perkembangan teknologi informasi, khususnya bidang data mining, data pendidikan dapat diolah menjadi informasi yang lebih bermakna melalui pendekatan Educational Data Mining, salah satunya dengan teknik klasifikasi yang telah banyak dimanfaatkan untuk menganalisis data siswa (*Dwi Agustina et al., 2025*). Salah satu metode klasifikasi yang dapat digunakan adalah algoritma Naive Bayes, yang bekerja berdasarkan pendekatan probabilistik dan dikenal memiliki keunggulan kesederhanaan serta efisiensi komputasi, serta mampu mengolah berbagai variabel akademik maupun non-akademik seperti kehadiran untuk menghasilkan klasifikasi sebagai dasar pengambilan keputusan dalam bidang pendidikan (*Baridah et al., 2025*).

Penelitian terdahulu menunjukkan bahwa Naive Bayes memberikan kinerja yang baik dalam berbagai konteks pendidikan: klasifikasi kedisiplinan siswa berbasis data pelanggaran yang dioptimasi dengan Genetic Algorithm mencapai akurasi 89,47% (*Utoyo & Mujiyono, 2025*); sistem penilaian akademik berbasis Naive Bayes Classifier mencapai akurasi 96,67% (*Iskandar et al., 2025*); perbandingan Naive Bayes dengan C4.5 untuk prediksi kelulusan menunjukkan Naive Bayes unggul dengan akurasi 100% berbanding 89% (*Harahap, 2024*); klasifikasi kedisiplinan siswa menggunakan Decision Tree, K-Nearest Neighbor, dan Naive Bayes menunjukkan variabel kehadiran dan prestasi akademik berpengaruh signifikan dengan akurasi Naive Bayes sebesar 77,62% (*Hutabalian et al., 2026*); serta prediksi ketuntasan belajar berbasis Naive Bayes yang mengintegrasikan data akademik dan non-akademik memperoleh akurasi 85,2% (*Ekawati et al., 2025*). Sebagian besar penelitian tersebut masih berfokus pada aspek akademik seperti performa dan kelulusan siswa, dan belum secara khusus mengevaluasi kedisiplinan siswa berbasis data kehadiran pada sistem yang terintegrasi, khususnya pada jenjang sekolah dasar.

Berdasarkan celah tersebut, penelitian ini difokuskan pada penerapan algoritma Gaussian Naive Bayes untuk mengevaluasi kedisiplinan siswa pada sistem kehadiran terintegrasi berbasis web di SDN Mekar Jaya. Varian Gaussian dipilih karena mampu mengolah atribut kehadiran yang bersifat numerik kontinu, yaitu persentase hadir, sakit, izin, dan alfa. Tujuan penelitian ini adalah: (1) memanfaatkan data kehadiran siswa pada sistem kehadiran terintegrasi untuk menghasilkan evaluasi kedisiplinan; (2) menerapkan algoritma Gaussian Naive Bayes dalam mengklasifikasikan tingkat kedisiplinan siswa ke dalam kategori Disiplin, Cukup Disiplin, dan Kurang Disiplin; dan (3) mengukur kinerja algoritma menggunakan parameter akurasi, precision, dan recall. Penelitian ini diharapkan menghasilkan klasifikasi kedisiplinan siswa yang lebih objektif dan memberikan kontribusi dalam pemanfaatan data mining pada pendidikan dasar, khususnya dalam pengolahan data kehadiran sebagai dasar evaluasi kedisiplinan.

2. Metode Penelitian

2.1. Metode Pengumpulan Data

Data penelitian dikumpulkan melalui empat teknik, yaitu studi pustaka terhadap literatur mengenai sistem kehadiran, kedisiplinan siswa, *data mining*, dan algoritma *Naive Bayes*; observasi langsung terhadap proses pencatatan kehadiran di SDN Mekar Jaya; wawancara dengan guru dan staf tata usaha mengenai kriteria penilaian kedisiplinan; serta dokumentasi data kehadiran siswa (hadir, izin, sakit, alfa) yang digunakan sebagai dataset utama penelitian.

2.2 Metode Pengembangan Sistem

Sistem kehadiran terintegrasi dikembangkan menggunakan metode Agile Development Methodology yang dipilih karena memberikan fleksibilitas terhadap perubahan kebutuhan pengguna serta mendukung pengembangan secara iteratif dan inkremental. Tahapan yang dilalui meliputi perencanaan kebutuhan sistem, perancangan antarmuka pengguna dan basis data, implementasi menggunakan bahasa pemrograman berbasis web dengan framework Laravel dan basis data MySQL (*Christanto & Somya, 2023*), pengujian fungsi sistem dan algoritma klasifikasi, penerapan sistem pada lingkungan sekolah, hingga pemeliharaan sistem secara berkelanjutan.

2.3. Dataset dan Variabel Penelitian

Dataset penelitian berupa data absensi siswa kelas 4 SDN Mekar Jaya selama satu semester yang terdiri atas 57 data, dengan variabel input dan keterangannya disajikan pada Tabel 1, sedangkan variabel output berupa kategori kedisiplinan.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Jumlah Hadir	Jumlah kehadiran siswa
Jumlah Sakit	Ketidakhadiran karena sakit
Jumlah Izin	Ketidakhadiran dengan izin
Jumlah Alfa	Ketidakhadiran tanpa keterangan
Persentase Kehadiran	Persentase kehadiran siswa

DOI: <https://doi.org/10.69693/ijmst.v4i2.12997>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Seluruh 57 data pada penelitian ini diperiksa melalui tahap pembersihan data (data cleaning) dan dinyatakan valid, tanpa data kosong maupun tidak lengkap. Selanjutnya dilakukan transformasi data, yaitu mengubah nilai atribut hadir, sakit, izin, dan alfa ke dalam bentuk desimal berdasarkan total pertemuan (115 kali), serta menghitung persentase kehadiran sebagai dasar penentuan kategori kedisiplinan, menggunakan Persamaan (1) dan (2).

$$\text{Persentase} = \frac{H_{\text{dir}}}{\text{Total Pertemuan}} \times 100\% \quad (1)$$

$$\text{Nilai Desimal} = \frac{\text{Nilai Atribut}}{\text{Total Pertemuan}} \quad (2)$$

Berdasarkan hasil transformasi tersebut, setiap data diberi label kelas kedisiplinan menggunakan kriteria persentase kehadiran, dengan ketentuan sebagaimana disajikan pada Tabel 2.

Tabel 2. Penentuan Label Kelas

Persentase Kehadiran	Kategori
$\geq 90\%$	Disiplin
$75\% - 89\%$	Cukup Disiplin
$< 75\%$	Kurang Disiplin

Dari 57 data yang tersedia, sebanyak 48 data (80%) digunakan sebagai data latih (training data) untuk membangun model klasifikasi melalui perhitungan prior probability, mean, dan standar deviasi pada setiap kategori kedisiplinan, sedangkan 9 data (20%) digunakan sebagai data uji (testing data) untuk mengukur akurasi model. Distribusi data latih pada masing-masing kelas kedisiplinan ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Data Latih

Kategori	Jumlah Data
Disiplin	35
Cukup Disiplin	7
Kurang Disiplin	6
Total	48

2.4 Algoritma Gaussian Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi dalam data mining yang didasarkan pada Teorema Bayes dan digunakan untuk menghitung probabilitas suatu data terhadap kelas tertentu berdasarkan nilai atribut yang dimilikinya (Liantoni, 2022). Metode ini bekerja dengan menghitung probabilitas posterior suatu kelas berdasarkan probabilitas prior dan likelihood data yang diamati, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi. Algoritma ini mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya sehingga proses perhitungan menjadi lebih sederhana, dan meskipun asumsi tersebut tidak selalu sesuai dengan kondisi nyata, metode ini tetap memiliki performa yang baik pada berbagai kasus klasifikasi (Shi, 2022). Secara matematis, Teorema Bayes dirumuskan pada Persamaan (3).

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)} \quad (3)$$

dengan $P(H|X)$ adalah probabilitas hipotesis H (kelas tertentu) jika diberikan data X , $P(X|H)$ adalah probabilitas munculnya data X jika hipotesis H benar, $P(H)$ adalah probabilitas awal (prior) hipotesis H , dan $P(X)$ adalah probabilitas munculnya data X .

Penelitian ini menerapkan varian Gaussian Naive Bayes karena mampu mengolah atribut berupa data numerik atau kontinu dengan asumsi bahwa data mengikuti distribusi normal (Gaussian), sehingga perhitungan probabilitas dilakukan berdasarkan nilai rata-rata (mean) dan standar deviasi setiap atribut pada masing-masing kelas (Destiana et al., 2023). Pendekatan ini lebih sesuai digunakan pada data berbentuk angka dibandingkan Naive Bayes konvensional yang umumnya digunakan pada data kategorikal (Çınar, 2024). Fungsi distribusi Gaussian dirumuskan pada Persamaan (4).

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

dengan $P(x)$ adalah probabilitas kemunculan nilai x berdasarkan distribusi Gaussian, μ (mean) adalah nilai rata-rata data pada suatu kelas, σ (standar deviasi) adalah ukuran penyebaran data terhadap rata-rata, σ^2 adalah varians, e adalah bilangan eksponensial, dan π adalah konstanta matematika. Nilai mean dan standar deviasi setiap atribut dihitung menggunakan Persamaan (5) dan (6).

$$\mu = \frac{\sum x}{n} \quad (5)$$

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n-1}} \quad (6)$$

dengan x adalah nilai atribut dan n adalah jumlah data pada kelas tersebut. Nilai probabilitas awal (prior probability) setiap kelas dihitung berdasarkan proporsi jumlah data pada tiap kelas terhadap total data latih

DOI: <https://doi.org/10.69693/ijmst.v4i2.12997>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

menggunakan Persamaan (7), sedangkan penentuan kelas hasil klasifikasi dilakukan dengan memilih nilai probabilitas posterior tertinggi (argmax) menggunakan Persamaan (8).

$$P(C) = \frac{\text{Jumlah Data pada Kelas}}{\text{Total Data}} \quad (7)$$

$$\hat{C} = \arg \max P(C | X) \quad (8)$$

2.5 Evaluasi Model

Kinerja model klasifikasi dievaluasi menggunakan confusion matrix, yaitu tabel yang membandingkan hasil prediksi model dengan kelas aktual pada data uji, sehingga dapat dihitung nilai akurasi, presisi, dan recall (Çınar, 2024). Akurasi menunjukkan tingkat ketepatan model dalam mengklasifikasikan keseluruhan data uji, presisi menunjukkan ketepatan prediksi positif yang dihasilkan model, dan recall menunjukkan kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya (Tanjung et al., 2023). Ketiga metrik tersebut dihitung menggunakan Persamaan (9) hingga (11).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (10)$$

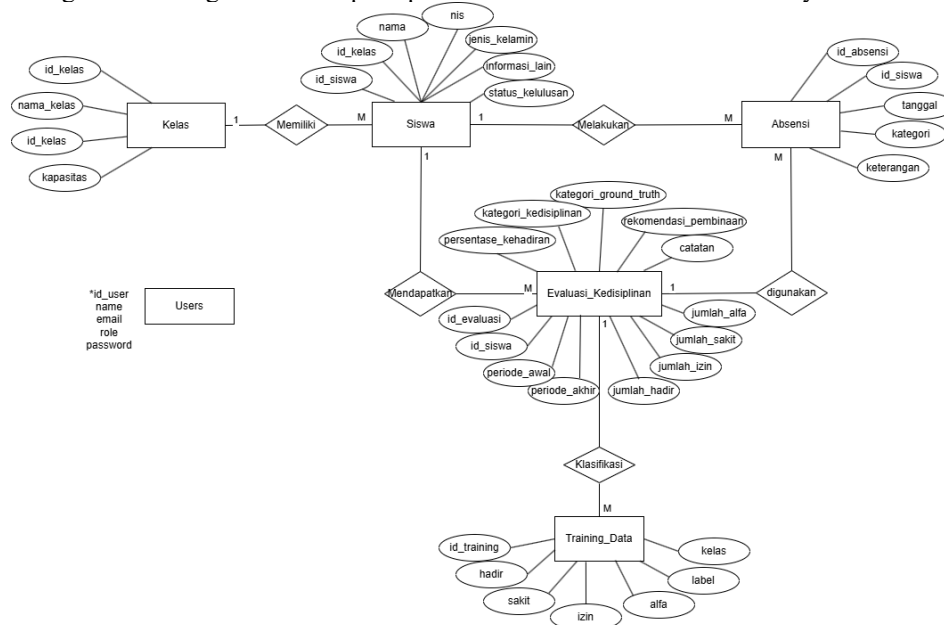
$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (11)$$

dengan TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) merupakan empat komponen utama confusion matrix yang menjadi dasar perhitungan seluruh metrik evaluasi model klasifikasi pada penelitian ini.

3. Hasil dan Pembahasan

3.1 Implementasi Sistem Kehadiran Terintegrasi

Struktur basis data Sistem Kehadiran Terintegrasi digambarkan melalui Entity Relationship Diagram (ERD) pada Gambar 1, yang terdiri dari entitas Kelas, Siswa, Absensi, Evaluasi_Kedisiplinan, dan Training_Data. Entitas Kelas berelasi one-to-many dengan Siswa, Siswa berelasi dengan Absensi dan Evaluasi_Kedisiplinan, sedangkan Training_Data digunakan sebagai data latih pada proses klasifikasi Gaussian Naive Bayes.



Gambar 1. Entity Relationship Diagram (ERD) Sistem

Gambar 2 menunjukkan halaman evaluasi kedisiplinan yang menampilkan hasil klasifikasi *Gaussian Naive Bayes* untuk setiap siswa berdasarkan data hadir, izin, sakit, alfa, dan persentase kehadiran, sedangkan Gambar 3 menunjukkan halaman monitoring yang merekap jumlah siswa pada setiap kategori kedisiplinan per kelas.

Gambar 2. Halaman Evaluasi Kedisiplinan

Gambar 2. Halaman Monitoring

3.2 Prior Probability, Mean, dan Standar Deviasi

Berdasarkan 48 data latih yang terdiri dari 35 data kelas Disiplin, 7 data kelas Cukup Disiplin, dan 6 data kelas Kurang Disiplin, diperoleh nilai prior probability berturut-turut sebesar 0,729; 0,146; dan 0,125 menggunakan Persamaan (7). Selanjutnya, nilai mean dan standar deviasi setiap atribut pada masing-masing kelas dihitung menggunakan Persamaan (5) dan (6), dengan hasil ditunjukkan pada Tabel 4 dan Tabel 5.

Tabel 4. Nilai Mean Setiap Atribut per Kelas

Kelas	Hadir	Sakit	Izin	Alfa
Disiplin	0,9675	0,0045	0,0040	0,0241
Cukup Disiplin	0,8175	0,0332	0,0200	0,1322
Kurang Disiplin	0,6383	0,0567	0,0367	0,2667

Tabel 5. Nilai Standar Deviasi Setiap Atribut per Kelas

Kelas	Hadir	Sakit	Izin	Alfa
Disiplin	0,0193	0,0083	0,0080	0,0203
Cukup Disiplin	0,0456	0,0102	0,0100	0,0380
Kurang Disiplin	0,0598	0,0082	0,0103	0,0589

3.3 Perhitungan Probabilitas Gaussian dan Hasil Klasifikasi

Perhitungan probabilitas Gaussian dicontohkan pada salah satu data uji dengan nilai atribut hadir 0,94; sakit 0,04; izin 0,00; dan alfa 0,02. Perhitungan atribut hadir terhadap kelas Disiplin menggunakan Persamaan (4) menghasilkan $P(\text{Hadir}|\text{Disiplin}) = 7,0226$. Hasil perhitungan probabilitas Gaussian untuk seluruh atribut dan kelas ditunjukkan pada Tabel 6.

Tabel 6. Probabilitas Gaussian Data Uji

Kelas	Hadir	Sakit	Izin	Alfa
Disiplin	7,0226	18,5691	0,0007	44,1294
Cukup Disiplin	0,2496	0,1081	23,4642	5,3991
Kurang Disiplin	$2,14 \times 10^{-5}$	$8,68 \times 10^{-4}$	13,30	0,0708

Probabilitas akhir setiap kelas kemudian dihitung dengan mengalikan nilai prior probability dengan seluruh nilai probabilitas Gaussian pada Persamaan (8), sehingga diperoleh $P(\text{Disiplin}|X) = 2,978754$; $P(\text{Cukup Disiplin}|X) = 0,498499879$; dan $P(\text{Kurang Disiplin}|X) = 2,18255 \times 10^{-9}$. Karena nilai probabilitas terbesar diperoleh pada kelas Disiplin, maka hasil klasifikasi data uji tersebut adalah Disiplin, yang menunjukkan bahwa pola kehadiran pada data uji tersebut memiliki karakteristik yang lebih dekat dengan distribusi data kelas Disiplin dibandingkan kelas lainnya. Seluruh 9 data uji yang digunakan pada penelitian ini ditunjukkan pada Tabel 7.

DOI: <https://doi.org/10.69693/ijmst.v4i2.12997>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 7. Data Uji (Hadir, Sakit, Izin, Alfa, dan Persentase Kehadiran)

Hadir	Sakit	Izin	Alfa	Persentase
0,94	0,00	0,00	0,01	94%
0,97	0,00	0,00	0,03	97%
1,00	0,00	0,00	0,00	100%
0,97	0,03	0,00	0,00	97%
0,85	0,02	0,00	0,13	85%
1,00	0,00	0,00	0,00	100%
0,96	0,01	0,00	0,03	96%
0,70	0,03	0,00	0,27	70%
0,94	0,04	0,00	0,02	94%

3.4 Evaluasi Kinerja Model

Evaluasi model dilakukan dengan membandingkan hasil prediksi sistem terhadap kelas aktual pada 9 data uji menggunakan confusion matrix, sebagaimana ditunjukkan pada Tabel 8.

Tabel 8. Confusion Matrix

Aktual / Prediksi	Disiplin	Cukup Disiplin	Kurang Disiplin
Disiplin	7	0	0
Cukup Disiplin	0	1	1
Kurang Disiplin	0	0	0

Berdasarkan Tabel 8, algoritma Gaussian Naive Bayes berhasil mengklasifikasikan 8 dari 9 data uji dengan benar, sehingga menggunakan Persamaan (9) diperoleh nilai accuracy sebesar $8/9 \times 100\% = 88,89\%$. Nilai precision dan recall untuk setiap kelas, dihitung menggunakan Persamaan (10) dan (11).

Hasil perhitungan menunjukkan bahwa kelas Disiplin dan Cukup Disiplin memiliki nilai precision sebesar 100%, sedangkan kelas Kurang Disiplin memiliki nilai precision sebesar 0% karena terdapat satu data aktual Cukup Disiplin yang diprediksi sebagai Kurang Disiplin. Nilai recall kelas Disiplin sebesar 100% dan kelas Cukup Disiplin sebesar 50%, sedangkan recall kelas Kurang Disiplin tidak dapat dihitung karena tidak terdapat data aktual pada kelas tersebut di dalam data uji. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma Gaussian Naive Bayes memiliki kinerja yang baik dalam mengklasifikasikan tingkat kedisiplinan siswa berdasarkan data kehadiran, dengan tingkat akurasi yang sebanding bahkan lebih baik dibandingkan beberapa penelitian sejenis yang dirangkum pada Tabel 1, khususnya penelitian yang berfokus pada aspek non-akademik seperti kedisiplinan.

3.5 Pengujian Sistem

Pengujian sistem dilakukan menggunakan metode Black Box Testing dan White Box Testing. Pengujian Black Box dilakukan terhadap fitur utama sistem, meliputi login, pengelolaan data pengguna, kelas, dan siswa, pencatatan absensi, evaluasi kedisiplinan, monitoring, rekomendasi, dan analisis kedisiplinan, dan seluruh skenario pengujian dinyatakan diterima sesuai dengan hasil yang diharapkan. Pengujian White Box difokuskan pada alur logika fungsi prediksi (predict) yang menjalankan proses pelatihan model, pengambilan nilai prior, perhitungan likelihood Gaussian untuk setiap atribut, penjumlahan log probabilitas, normalisasi nilai posterior, hingga penentuan kategori kedisiplinan dengan nilai posterior tertinggi sebagai hasil prediksi akhir, dan seluruh jalur pengujian dinyatakan valid serta berjalan sesuai dengan rancangan algoritma Gaussian Naive Bayes.

4. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan mengenai penerapan algoritma Naive Bayes untuk evaluasi kedisiplinan siswa pada sistem kehadiran terintegrasi berbasis web di SDN Mekar Jaya, dapat disimpulkan bahwa data kehadiran siswa yang sebelumnya hanya digunakan sebagai catatan administrasi dapat dimanfaatkan sebagai dasar evaluasi kedisiplinan yang lebih objektif melalui pengolahan variabel hadir, sakit, izin, alfa, dan persentase kehadiran. Penerapan algoritma Gaussian Naive Bayes pada sistem yang dibangun mampu mengklasifikasikan tingkat kedisiplinan siswa ke dalam tiga kategori, yaitu Disiplin, Cukup Disiplin, dan Kurang Disiplin, melalui perhitungan probabilitas setiap variabel kehadiran dan penentuan kategori berdasarkan nilai probabilitas tertinggi. Sistem yang dibangun juga terbukti mampu membantu admin, guru, dan kepala sekolah dalam mengelola data absensi, merekap kehadiran, mengevaluasi kedisiplinan, memantau kondisi siswa, memberikan rekomendasi pembinaan, serta menganalisis kedisiplinan siswa secara lebih cepat dan terstruktur dibandingkan proses manual yang masih bergantung pada penilaian subjektif guru.

Hasil pengujian model menggunakan data uji menunjukkan bahwa algoritma Gaussian Naive Bayes memperoleh nilai accuracy sebesar 88,89%, yaitu 8 dari 9 data uji terklasifikasi dengan benar, dengan nilai precision kelas Disiplin dan Cukup Disiplin masing-masing sebesar 100% serta kelas Kurang Disiplin sebesar 0% karena terdapat satu data aktual Cukup Disiplin yang diprediksi sebagai Kurang Disiplin, sedangkan nilai recall kelas Disiplin sebesar 100% dan kelas Cukup Disiplin sebesar 50%, sementara recall kelas Kurang Disiplin tidak dapat dihitung karena tidak terdapat data aktual pada kelas tersebut di dalam data uji. Secara keseluruhan, hasil tersebut menunjukkan bahwa algoritma Gaussian Naive Bayes memiliki kinerja yang baik dalam mengevaluasi

kedisiplinan siswa berdasarkan data kehadiran, yang juga didukung oleh hasil pengujian Black Box dan White Box yang menunjukkan bahwa seluruh fungsi sistem serta alur logika proses prediksi telah berjalan sesuai dengan rancangan.

Penelitian selanjutnya disarankan untuk membandingkan kinerja algoritma Gaussian Naive Bayes dengan metode klasifikasi lain, seperti Decision Tree atau K-Nearest Neighbor, serta membiasakan pengelolaan data kehadiran secara rutin dan konsisten agar hasil evaluasi kedisiplinan yang dihasilkan sistem tetap relevan dengan kondisi siswa di sekolah.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, kepada Bapak Romi Andrianto, S.Kom., M.Kom. selaku dosen pembimbing atas arahan dan bimbingan yang diberikan, serta kepada kepala sekolah, guru, dan staf SDN Mekar Jaya atas izin dan data yang diberikan untuk mendukung terlaksananya penelitian ini.

Reference

- Baridah, L., Putri, R. A., Info, A., Mining, E. D., Outcomes, S. L., & Education, V. (2025). Predicting Student Learning Outcomes in Vocational Computer and Network Engineering Using Naïve Bayes. 8(3), 667–676.
- Christanto, H., & Somya, R. (2023). Sistem Informasi Berbasis Web Menggunakan Framework Laravel pada Rancangan Sistem Inventory Gudang. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 12(3), 1204. <https://doi.org/10.35889/jutisi.v12i3.1588>
- Çınar, A. (2024). Multi-Class Classification with the Gaussian Naive Bayes Algorithm. *Journal of Data Applications*, 0(2), 1–13. <https://doi.org/10.26650/joda.1389471>
- Destiana, T., Umaidah, Y., & Enri, U. (2023). Penerapan Algoritma Gaussian Naive Bayes dalam Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis. *INFOTECH Journal*, 9(2), 507–513. <https://doi.org/10.31949/infotech.v9i2.6501>
- Dwi Agustina, F., Arif, M., Ahmad, S., Studi Sistem dan Teknologi Informasi, P., & Sains dan Teknologi, F. (2025). Systematic Literature Review atas Kinerja Algoritma KNN, Naive Bayes, dan Decision Tree pada Berbagai Studi Prediksi dan Klasifikasi. *Jurnal Jawara Sistem Informasi*, 3(1).
- Ekawati, R. R., Atina, V., & Maulidar, J. (2025). Prediksi Ketuntasan Belajar Siswa Menggunakan Naive Bayes dengan Integrasi Data (Student Completion Prediction System Using the Naive Bayes Method at SMA Negeri 1 Plupuh). 5(4), 1161–1173.
- Felicia, N., Agustini, P., Hikmah, N. N., & Sari, N. E. (2024). Kehadiran dan Ketidakhadiran Peserta Didik di Sekolah.
- Harahap, S. Z. (2024). Implementation of the C4.5 and Naive Bayes Algorithms to Predict Student Graduation. 8(3), 1741–1757.
- Hidayat, A. N., & Yuningsih, D. (2026). Etika Administrasi Sekolah untuk Meningkatkan Kedisiplinan Siswa di SD Negeri Jati 2. 5(1), 3104–3111.
- Hutabalian, D. M., Hutabarat, P., Prasetyo, M., & Irnanda, M. A. (2026). Klasifikasi Tingkat Kedisiplinan Siswa Menggunakan Algoritma Machine Learning: Decision Tree, KNN, dan Naive Bayes. 4(3), 1987–1992.
- Iskandar, A., Retnawati, H., Hadi, S., Utami, L., & Amiruddin, E. G. (2025). Naïve Bayes Classifier-Based Intelligent System for Student Academic Performance Assessment. 30(6), 1609–1620.
- Liantoni, F. (2022). Data Mining dan Penerapan Metode.
- Roudhotum Mawardania, & Henny Dwi Bhakti. (2024). Perancangan Sistem Presensi Sekolah Berbasis Web di SD Muhammadiyah Gresik. *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, 4(1), 78–90. <https://doi.org/10.55606/juisik.v4i1.745>
- Pardede, N. (2025). Pengaruh Manajemen Kesiswaan terhadap Tingkat Kedisiplinan Siswa Kelas XI TKJ di SMK Negeri 1 Sidikalang Tahun Pembelajaran 2024/2025. *Jurnal Pendidikan Sosial dan Humaniora*, 4(4), 7358–7371.
- Salsabila, D., & Basri, M. (2026). Analisis Penilaian Kinerja Siswa untuk Menentukan Siswa Terbaik di Sekolah Melalui Metode Naive Bayes di Songserm Sasana School, Hat Yai. *Jurnal Informatika dan Teknologi Komputer*, 6.
- Shi, Y. (2022). The Iris Classification Based on Gaussian Naive Bayes Algorithm. In 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE 2022). Association for Computing Machinery. <https://doi.org/10.1145/3573428.3573559>
- Tanjung, J. P., Tampubolon, F. C., Panggabean, A. W., & Nandrawan, M. A. A. (2023). Customer Classification Using Naive Bayes Classifier With Genetic Algorithm Feature Selection. *Sinkron*, 8(1), 584–589. <https://doi.org/10.33395/sinkron.v8i1.12182>
- Utoyo, B. D., & Mujiyono, S. (2025). Penerapan Algoritma Naive Bayes dengan Optimasi Genetic Algorithm untuk Memprediksi Kedisiplinan Siswa. 2169–2180. <https://doi.org/10.33364/algoritma/v.22-2.3084>