Teknik: Jurnal Ilmu Teknik dan Informatika
Volume 6 Nomor 2 Oktober 2026
e-ISSN: 2798-2513; p-ISSN: 2808-8751, Hal *204-222*
DOI: https://doi.org/10.51903/teknik.v6i1.1171
Available online at: https://journal.stiestekom.ac.id/index.php/TEKNIK

Penerapan Data Mining Untuk Analisis Pola Kehadiran Serta Pengaruhnya Terhadap Prestasi Belajar Siswa Di SMA Negeri 7 Palembang

Resti Wulandari[1], Herri Setiawan[2], Indah Permatasari[3]

[1-3]Departemen/Informatika, Program Studi Teknik Informatika, Universitas Indo Global Mandiri Palembang

Korespondensi penulis: 202111078@students.uigm.ac.id

Abstract. *Student absenteeism is not merely an administrative issue; it also reflects discipline and motivation to learn, which influence academic achievement. Unfortunately, in many schools, absenteeism data is merely recorded without further analysis. The researchers are interested in examining how data mining techniques can be used to uncover patterns of absenteeism and their relationship to academic achievement. This study was conducted at SMA Negeri 7 Palembang using data from 318 11th-grade students for the 2024/2025 academic year. The analysis involved data preprocessing, the creation of classification labels, and the application of the K-Nearest Neighbor (KNN) and Logistic Regression algorithms. The results showed that both algorithms were able to predict achievement categories with high accuracy, namely 95.83%. KNN with a value of K=3 was more balanced in identifying high-achieving students (F1-score 0.75), while Logistic Regression excelled in precision (1.00) but had lower recall (0.56). Regression analysis indicates that each additional absence reduces a student's probability of achieving high academic performance by a factor of 0.74. These findings confirm that absences are not merely a number but a critical factor that can influence academic achievement. Therefore, schools are advised to utilize attendance and absence data not only for administrative purposes but also as an academic monitoring system so that learning strategies can be implemented more effectively.*
Keywords: *Data mining; Attendance patterns; Academic performance; SMA Negeri 7 Palembang; Student analysis*

Abstrak. Ketidakhadiran siswa bukan hanya persoalan administratif, tetapi juga mencerminkan kedisiplinan dan motivasi belajar yang berpengaruh terhadap prestasi akademik. Sayangnya, di banyak sekolah data ketidakhadiran masih sebatas dicatat tanpa analisis lebih lanjut. Peneliti tertarik untuk mengkaji bagaimana teknik *data mining* dapat digunakan untuk mengungkap pola ketidakhadiran dan hubungannya dengan prestasi belajar. Penelitian ini dilakukan di SMA Negeri 7 Palembang dengan menggunakan data 318 siswa kelas XI tahun ajaran 2024/2025. Analisis dilakukan melalui pra-pemrosesan data, pembentukan label klasifikasi, serta penerapan algoritma *K-Nearest Neighbor (KNN)* dan *Logistic Regression.* Hasil menunjukkan kedua algoritma mampu memprediksi kategori prestasi dengan akurasi tinggi, yaitu 95,83%. *KNN* dengan nilai *K*=3 lebih seimbang dalam mengenali siswa berprestasi tinggi (*F1-score* 0,75), sedangkan *Logistic Regression* unggul dalam *precision* (1,00) namun *recall* lebih rendah (0,56). Interpretasi regresi memperlihatkan bahwa setiap tambahan satu kali ketidakhadiran menurunkan peluang siswa mencapai prestasi tinggi menjadi 0,74 kali lipat. Temuan ini menegaskan bahwa ketidakhadiran bukan sekadar angka, melainkan faktor penting yang dapat memengaruhi capaian akademik. Oleh karena itu, sekolah disarankan memanfaatkan data kehadiran dan ketidakhadiran tidak hanya untuk administrasi, tetapi juga sebagai sistem pemantauan akademik agar strategi pembelajaran dapat dilakukan lebih tepat sasaran.
Kata kunci: *Data mining; Pola kehadiran; Prestasi belajar; SMA Negeri 7 Palembang; Analisis siswa*

1. LATAR BELAKANG

Di era digital, sistem pendidikan menghadapi tantangan dalam pemerataan akses, peningkatan mutu, dan adaptasi terhadap teknologi informasi. Transformasi digital mendorong sekolah untuk memanfaatkan data sebagai dasar pengambilan keputusan, termasuk dalam memantau dan meningkatkan prestasi siswa (Purnomo, 2025; Dadang, 2025).

Diterima: 24 April 2026; Direvisi: 15 Juni 2026; Disetujui: 10 Agustus 2026; Tersedia Online: 25 September 2026; Diterbitkan: 21 Oktober 2026.

Merujuk pada Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, pendidikan diartikan sebagai usaha sadar dan terencana untuk menciptakan suasana belajar serta proses pembelajaran yang memungkinkan peserta didik secara aktif mengembangkan potensi dirinya. Tujuan tersebut mencakup penguatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, dan keterampilan yang dibutuhkan dalam kehidupan bermasyarakat, berbangsa, dan bernegara (Indonesia, 2003).

Secara umum, pendidikan bertujuan untuk mengembangkan potensi siswa secara menyeluruh baik secara akademik maupun nonakademik sesuai dengan yang tercantum didalam Undang-undang. Namun, keberhasilan siswa tidak hanya ditentukan oleh kualitas pengajaran, tetapi juga oleh keaktifan siswa dalam mengikuti kegiatan pembelajaran, salah satunya adalah tingkat kehadiran di kelas. Kehadiran yang konsisten mencerminkan kedisiplinan, motivasi belajar, dan keterlibatan aktif dalam kegiatan pembelajaran (Annisa et al., 2023).

Nilai-nilai pendidikan karakter seperti disiplin, tanggung jawab, dan semangat belajar tidak hanya membentuk kepribadian siswa, tetapi juga memberikan pengaruh nyata terhadap pencapaian akademik mereka. Kehadiran yang teratur menjadi wujud kedisiplinan yang mendukung ketertiban proses belajar serta meningkatkan motivasi untuk meraih hasil optimal (Leobisa et al., 2022). Selain itu, penerapan nilai positif dalam pendidikan karakter terbukti berkaitan erat dengan motivasi belajar dan capaian siswa (Utami & Nizaar, 2023). Dengan demikian, pendidikan berperan bukan hanya sebagai sarana belajar ilmu, tetapi juga sebagai faktor penting dalam menentukan prestasi akademik, di mana kehadiran menjadi indikator penting yang mencerminkan disiplin dan motivasi belajar.

Hasil wawancara dengan berbagai pihak, baik staf kesiswaan, wakil kesiswaan, guru mata pelajaran, wakil kurikulum, maupun guru BK, menunjukkan adanya hubungan yang erat antara tingkat kehadiran siswa dengan prestasi belajar mereka. Seluruh narasumber menegaskan bahwa kehadiran merupakan faktor penting dalam menunjang keberhasilan akademik. Ketidakhadiran siswa, meskipun hanya beberapa kali, berdampak langsung pada penurunan prestasi karena siswa akan tertinggal materi, tidak mengerjakan tugas, dan berkurang pemahaman terhadap pelajaran yang diberikan.

Selain itu, para narasumber juga menekankan bahwa siswa yang memiliki prestasi tinggi umumnya menunjukkan tingkat kehadiran yang baik. Mereka rajin hadir karena memiliki motivasi dan ambisi untuk tidak ketinggalan pelajaran. Sebaliknya, siswa yang sering absen cenderung memiliki prestasi yang lebih rendah, sebab peluang mereka untuk memahami materi

secara utuh menjadi terbatas. Secara keseluruhan, wawancara ini memperkuat temuan bahwa kehadiran siswa di sekolah berperan signifikan dalam menentukan prestasi belajar. Semakin tinggi tingkat kehadiran, semakin besar kemungkinan siswa meraih prestasi akademik yang baik. Sebaliknya, semakin sering siswa tidak hadir, semakin besar risiko penurunan prestasi yang dialami.

Hal tersebut menunjukkan bahwa adanya hubungan antara tingkat kehadiran terhadap prestasi akademik siswa. Siswa yang sering hadir cenderung memiliki pemahaman materi yang lebih baik dan hasil belajar yang lebih tinggi dibandingkan siswa yang sering absen (Vrasetya et al., 2025). Meski sebagian banyak sekolah telah memiliki data yang cukup kompleks baik dari segi sistem pencatatan kehadiran dan nilai akademik, pemanfaatan data tersebut masih terbatas pada kebutuhan administratif dan belum dimaksimalkan untuk analisis hubungan antara kehadiran dan prestasi siswa (Jeperson Hutahaean, 2024).

Penelitian ini dilaksanakan di SMA Negeri 7 Palembang, merupakan sekolah negeri berakreditasi A yang telah menerapkan Kurikulum Merdeka. Sekolah ini memiliki sistem dokumentasi internal yang mencatat data kehadiran dan nilai akademik secara lengkap, informasi tersebut menjadi sumber utama dalam analisis yang dilakukan. Informasi tersebut diperoleh melalui akses internal terhadap sistem dokumentasi sekolah. Untuk mengungkap pola hubungan antara tingkat kehadiran dan prestasi akademik siswa, penelitian ini menerapkan pendekatan *data mining*. *Data mining* merupakan proses untuk mengidentifikasi pola-pola yang bermakna dari kumpulan data berukuran besar (Pradita, 2022).

Dalam konteks pendidikan, data mining telah digunakan oleh peneliti sebelumnya untuk memprediksi prestasi akademik siswa berdasarkan pola kehadiran dan aktivitas belajar, dengan menggunakan algoritma *Naïve Bayes dan Logistic Regression*. Dengan hasil yang menunjukkan bahwa *Logistic Regression* memiliki performa yang lebih akurat dengan hasil 98 % dalam kasus prediksi akademik siswa (Pambudi et al., 2024). Adapun peneliti lain menerapkan data mining untuk memprediksi siswa berprestasi dengan menggunakan Algoritma *K-Nearest Neighbor*. Hasil penelitian tersebut menunjukkan bahwa penerapan algoritma KNN dengan nilai k=1 berhasil memprediksi siswa berprestasi secara akurat dan terbukti efektif dalam membantu pihak sekolah untuk menyajikan informasi sebagai acuan dalam menentukan siswa berprestasi yang benar sesuai kemampuannya, baik segi akademik maupun nonakademik (Widaningsih, 2022).

Berdasarkan penelitian sebelumnya, penelitian ini bertujuan untuk menerapkan kombinasi antara algoritma *K-Nearest Neighbor* dan *Logistic Regression* dalam menganalisis

pola kehadiran siswa serta pengaruhnya terhadap prestasi belajar di SMA Negeri 7 Palembang. Pemanfaatan kedua metode ini diharapkan mampu memperlihatkan pola hubungan antara tingkat kehadiran siswa dan prestasi akademik. Informasi tersebut dapat membantu pihak sekolah terutama guru, wali kelas, dan staf administrasi untuk memahami karakter siswa yang memiliki kecenderungan kehadiran rendah. Dengan pemahaman ini, sekolah dapat meninjau kembali strategi pembelajaran dan sistem kedisiplinan yang telah diterapkan, sebagai bahan evaluasi untuk mendukung peningkatan hasil belajar.

2. KAJIAN TEORITIS

Pendidikan

Dalam Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, pendidikan diartikan sebagai usaha sadar dan terencana untuk menciptakan suasana belajar serta proses pembelajaran yang memungkinkan peserta didik secara aktif mengembangkan potensi dirinya (Indonesia, 2003). Dalam konteks pendidikan formal keberhasilan sering diukur dengan prestasi belajar, serta dari bagaimana keterlibatan aktif siswa dalam aktivitas belajar yang menjadi dasar pencapaian prestasi tersebut (Ariani Hrp Nurlina et al., 2022).

Perkembangan dan prestasi peserta didik dipengaruhi oleh faktor internal yaitu kondisi fisiologis, psikologis, serta kemampuan kognitif dan faktor eksternal yaitu seperti halnya peran pendidik, keluarga, sekolah, teman sebaya, masyarakat, budaya, serta lingkungan fisik yang dapat membentuk perilaku, motivasi, kehadiran dan nilai belajar siswa (Ramli, M, 2025). Adapun secara umum dalam dunia pendidikan formal untuk menilai prestasi belajar siswa, maka digunakan indikator seperti ujian, keaktifan dalam kelas, kemampuan menyelesaikan tugas, serta sikap terhadap proses pembelajaran. Indikator inilah yang akan menjadi acuan dalam menentukan tingkat prestasi siswa (Hanim et al., 2022).

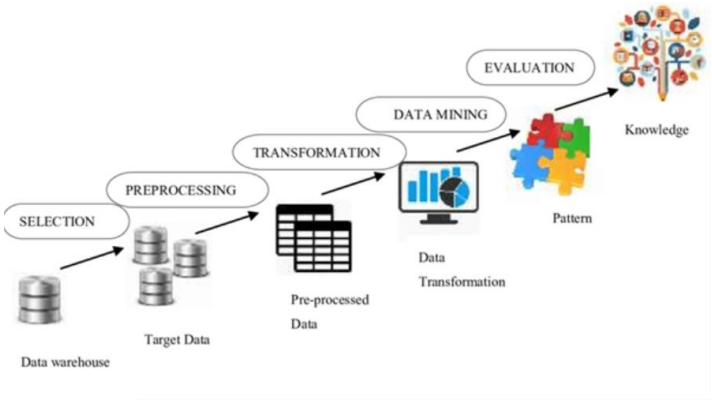
Namun, indikator tersebut tidak berdiri sendiri. Nilai karakter yang tertanam dalam diri peserta didik, seperti kedisiplinan dan tanggung jawab, turut membentuk pola kehadiran dan semangat belajar yang pada akhirnya memengaruhi pencapaian prestasi. Oleh karena itu, untuk memahami pencapaian hasil belajar secara lebih komprehensif, perlu dikaji tiga aspek utama yang saling berkaitan: nilai, kehadiran, dan prestasi.

Data Mining dalam Dunia Pendidikan

Data Mining ialah untuk memperoleh informasi berharga dari kumpulan data yang besar dan harus diolah agar bisa menjadi informasi baru yang dapat memberikan bantuan (Liantoni, 2022). Menurut Adji et al., (2025) data mining digunakan untuk klasifikasi dan klasterisasi data

guna mendukung pengambilan keputusan berbasis data. Dalam bukunya Birant (2021) menyebutkan bahwa Data mining merupakan metode untuk mengidentifikasi pola, koneksi, dan tren signifikan dari data besar yang disimpan, menggunakan teknik seperti pendekatan statistik dan matematik. Menurut Papadogiannis & Wallace (2024) EDM bertujuan untuk merancang dan menggunakan metode analisis guna memahami data yang muncul dari aktivitas pendidikan, seperti interaksi siswa dengan platform belajar, hasil ujian, dan kegiatan belajar secara daring.

Dalam bukunya yang menjelaskankan tahapan penting, mulai dari pemilihan data yang relevan, pembersihan dan persiapan data, pengubahan data ke format yang sesuai, penerapan teknik data mining untuk menemukan pola, hingga tahap akhir berupa interpretasi dan evaluasi hasil analisis (Marisa Fitri, Anastasia L. Maukar, 2021) . Gambar 2.1 menunjukkann ilustrasi serta penjelasan menegenai proses KDD secara detail :



Gambar 2.1. *Knowledge Discovery in Database Process (KDD)*

Algoritma K-Nearest Neighbor (KNN)

KNN adalah metode klasifikasi yang bekerja dengan prinsip pengukuran jarak antar data. Algoritma ini berasumsi bahwa setiap data pelatihan memuat informasi atribut sekaligus label kelasnya, sehingga keseluruhan data pelatihan berperan langsung sebagai acuan dalam proses klasifikasi. Prinsip kerja KNN adalah menilai data baru berdasarkan kedekatannya dengan data latih yang paling mirip, kemudian menentukan kelas atau nilai berdasarkan mayoritas tetangga terdekat (Halder et al., 2024)

Algoritma *KNN* bekerja dengan mengidentifikasi sejumlah tetangga terdekat dari data latih terhadap data yang akan diklasifikasikan, berdasarkan jarak terpendek (Rahmadiah et al., 2022). Proses pelaksanaan *KNN* dimulai dengan menetapkan nilai *K* sebagai jumlah tetangga terdekat yang digunakan dalam proses klasifikasi, kemudia dilanjutkan dengan menghitung kuadrat jarak *Euclidean* antara setiap objek dalam data latih terhadap data uji.

Untuk mengidentifikasi tetangga terdekat, algoritma *KNN* menghitung jarak antara data uji dan setiap data dalam data latih. Pada umumnya, perhitungan ini menggunakan rumus jarak Euclidean, yang merupakan metode paling sering diterapkan untuk data bertipe numerik, yang dirumuskan sebagai berikut:

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

dimana:

d = *Jarak antar data*

x_i = *sampel data*

y_i = *data uji*

i = *Variabel data*

n = *Dimensi data*

Dalam dunia pendidikan, *KNN* telah diterapkan untuk memprediksi Prestasi Siswa SMA Pada Dinas Pendidikan Provinsi Jambi, dengan menggunakan parameter *K*=3, algoritma *KNN* berhasil mencapai nilai *F1-score* sebesar 61,99% dan *Mean Squared Error (MSE)* sebesar 0,38, yang mencerminkan performa yang cukup baik dalam memproyeksikan keberlanjutan pendidikan siswa (Lastari, 2023). Sementara itu menurut Lastari (2023) dalam penelitiannya dalam memprediksi kelulusan mahasiswa Pendidikan Profesi Guru (PPG) menggunakan algoritma *KNN* dengan *K* = 1 dan diuji menggunakan 4-*fold cross validation* menunjukkan bahwa tingkat kelulusan PPG mencapai 86,7%, dan *KNN* mampu melakukan prediksi dengan akurasi terbaik 91,6%.

Logistic Regression

Logistic Regression adalah metode statistik yang digunakan untuk memprediksi hasil dua kemungkinan, seperti lulus atau tidak. Metode ini cocok digunakan karena bisa membaca hubungan antar data dan hasilnya mudah dipahami (Pambudi et al., 2024). Konsep dari regresi logistik adalah fungsi logit, yaitu transformasi logaritmik dari odds (peluang kejadian terhadap tidak kejadian). Persamaan dasar regresi logistik adalah:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Selain bentuk logit, probabilitas suatu peristiwa dapat dituliskan dengan fungsi logistik:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

dimana :

$\ln\left(\frac{p}{1-p}\right)$ = Logit

ln = Logaritma natural

p = $Probabilitas$

β_0 = $Intersep\ atau\ konstanta\ model.$

β_1, β_2, = Koefisien regresi variabel 1, variabel 2

β_n = *Koefisien regresi variabel n*

X_1, X_2, = Variabel prediktor 1, Variabel prediktor 2

X_n = Variabel prediktor n

Confusion Matrix

Confusion matrix adalah metode evaluasi yang umum digunakan dalam klasifikasi untuk membandingkan label aktual dengan hasil prediksi model (Farokhah & Korespondensi, 2020). Matriks ini memungkinkan perhitungan metrik seperti akurasi, presisi, recall, dan F1-score, yang penting dalam menilai performa algoritma seperti KNN dan Logistic Regression. Dalam penelitian pendidikan, confusion matrix telah digunakan untuk mengevaluasi model prediksi capaian akademik berbasis latar belakangs siswa, seperti yang dilakukan oleh (Prasetyawan & Gatra, 2022).

Struktur dasar confusion matrix terdiri dari empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Keempat komponen ini menjadi dasar dalam menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang sangat penting dalam menilai efektivitas model klasifikasi. Tabel 2.1. menunjukkan bentuk dasar confusion matrix untuk klasifikasi biner:

Table 2.1. Confussion Matrix

Confusion Matrix	*Predicted Positive*	*Predicted Negative*
Actual Positive	*True Positive* (TP)	*False Negative* (FN)
Actual Negative	*False Positive* (FP)	*True Negative* (TN)

Dalam konteks penelitian ini, *confusion matrix* digunakan untuk mengevaluasi performa algoritma *KNN* dan *Logistic Regression* dalam mengklasifikasikan data kehadiran siswa terhadap kategori prestasi belajar di SMA Negeri 7 Palembang. Evaluasi ini penting untuk mengetahui seberapa akurat model dalam memetakan pola kehadiran siswa dan dampaknya terhadap capaian akademik, serta menjadi dasar dalam menilai efektivitas kombinasi algoritma yang diterapkan.

Google Colab

Google Colab adalah platform berbasis cloud dari Google yang memungkinkan eksekusi kode Python secara online tanpa perlu instalasi perangkat lunak tambahan di komputer.

Layanan ini terintegrasi dengan Google Drive dan dapat diakses melalui peramban, sehingga memudahkan pengembangan aplikasi pembelajaran mesin, analisis data, maupun eksperimen lain yang membutuhkan daya komputasi besar (Hadi et al., 2025). Dalam penelitian ini, Google Colab dimanfaatkan sebagai platform utama untuk menjalankan algoritma KNN dan Logistic Regression dalam menganalisis pola kehadiran siswa serta kaitannya dengan prestasi belajar. Keunggulan Colab terletak pada kemampuannya mengeksekusi komputasi tanpa instalasi lokal, sekaligus mendukung kolaborasi daring dan menyediakan akses GPU gratis.

Python

Python adalah bahasa pemrograman yang berjalan menggunakan interpreter, sehingga kode dapat dijalankan secara langsung tanpa proses kompilasi. Karena sifatnya yang lintas platform, Python bisa digunakan di berbagai sistem operasi seperti Windows, Linux, dan lainnya. Bahasa ini dipilih karena memiliki banyak keunggulan, antara lain mudah dipahami, efisien, fleksibel, mampu berinteraksi dengan sistem lain, serta didukung oleh komunitas pengguna yang besar dan aktif (Retnoningsih & Pramudita, 2020). Dalam penelitian ini, Python dijalankan melalui platform *Google Colab* yang mendukung pemrograman interaktif berbasis *cloud*, sehingga memudahkan pengolahan data secara efisien tanpa memerlukan instalasi lokal.

Scikit-learn

Scikit-learn merupakan pustaka *open-source* dalam bahasa pemrograman Python yang menyediakan berbagai algoritma dan fungsi untuk pembelajaran mesin (*machine learning*). Pustaka ini dibangun di atas NumPy, SciPy, dan Matplotlib, serta mendukung berbagai metode seperti klasifikasi, regresi, klastering, reduksi dimensi, dan *data preprocessing*. Dengan antarmuka yang sederhana dan efisien, *Scikit-learn* banyak digunakan dalam penelitian maupun aplikasi industri untuk analisis data prediktif (Abdullah & Kusumastuti, 2025).

Standar Scaler

Standard Scaler adalah teknik *preprocessing* yang digunakan untuk menstandarkan fitur dengan cara mengurangi nilai rata-rata dan menyamakannya ke dalam skala varian satuan. Proses ini diterapkan pada setiap fitur dalam dataset agar tidak ada perbedaan skala yang terlalu besar antar data, sehingga model dapat dilatih dengan lebih seimbang dan hasilnya sesuai dengan yang diharapkan (Prasetyo, Vincentius et al., 2024).

Stratified Splitting

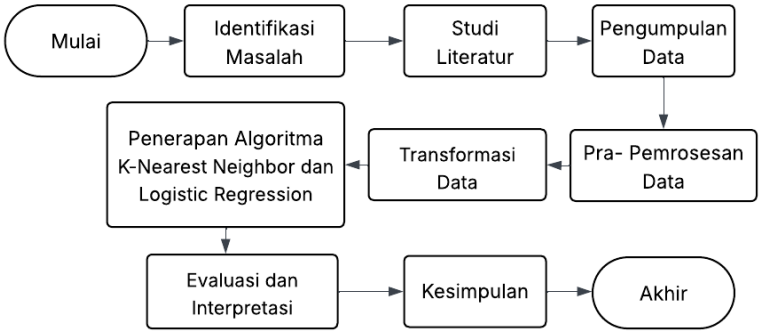
Stratified splitting merupakan teknik pembagian dataset ke dalam data latih dan data uji dengan cara mempertahankan proporsi kelas yang sama seperti pada dataset asli. Dengan kata lain, distribusi label tetap konsisten pada kedua subset, sehingga model dapat dilatih dan diuji

secara representatif. Teknik ini sangat berguna pada kasus *imbalanced dataset*, di mana jumlah data antar kelas tidak seimbang (Farias et al., 2020).

3. METODE PENELITIAN

Pengumpulan data dilakukan sebagai tahap awal dalam penelitian ini. Data yang digunakan berupa data kehadiran siswa (sakit, izin, alpa) dan data prestasi siswa (nilai rapor semester ganjil dan genap) kelas XI tahun ajaran 2024/2025 dari SMA Negeri 7 Palembang, yang diperoleh melalui metode dokumentasi dari arsip internal sekolah. Total data yang terkumpul sebanyak 636 catatan dari 318 siswa. Adapun alur penelitian ditampilkan pada Gambar 3.1

Gambar 3.1 Alur Penelitian



Setelah data terkumpul, dilakukan pra-pemrosesan data yang meliputi perhitungan rata-rata nilai rapor (dengan menjumlahkan seluruh nilai mata pelajaran lalu dibagi jumlah mata pelajaran), perhitungan total ketidakhadiran (penjumlahan sakit, izin, dan alpa), serta pembentukan label klasifikasi. Berdasarkan kriteria sekolah, siswa dengan rata-rata nilai ≥88 dikategorikan "Tinggi", nilai >75 dikategorikan "Sedang", dan nilai <75 "Rendah". Selanjutnya dilakukan transformasi data berupa normalisasi skala variabel ke rentang 0–1 agar perhitungan jarak Euclidean tidak bias, serta encoding label "Tinggi" menjadi 1 dan "Sedang" menjadi 0 untuk keperluan klasifikasi biner.

Dua algoritma klasifikasi diterapkan dalam penelitian ini, yaitu K-Nearest Neighbor (KNN) dan Logistic Regression. Pada KNN, setiap siswa direpresentasikan sebagai titik koordinat (rata-rata nilai, total absensi), kemudian dihitung jarak Euclidean ke siswa lain. Nilai K ditetapkan sebesar 3 (berdasarkan $\sqrt{N}$), dan klasifikasi ditentukan berdasarkan mayoritas label dari tiga tetangga terdekat. Pada Logistic Regression, model bekerja dengan fungsi logit untuk menghitung probabilitas seorang siswa masuk ke kategori tertentu, dengan ambang batas 0,5. Parameter yang diperoleh dari uji coba manual adalah $\beta_0 = -12$, $\beta_1 = 0{,}14$ (pengaruh nilai rata-rata), dan $\beta_2 = -0{,}25$ (pengaruh ketidakhadiran). Odds ratio menunjukkan bahwa setiap

kenaikan 1 poin nilai rata-rata meningkatkan peluang prestasi tinggi 1,15 kali, sedangkan setiap tambahan 1 kali absen menurunkan peluang menjadi 0,78 kali.

Evaluasi performa model dilakukan dengan confusion matrix dan metrik seperti akurasi, precision, recall, serta F1-score. Pada uji coba pendahuluan dengan sampel 10 siswa, KNN mencapai akurasi 90%, dengan precision kelas Tinggi 1,00 dan recall 0,50. Logistic Regression mencapai akurasi 70%, dengan recall kelas Tinggi 1,00 namun precision 0,40. Hasil evaluasi dianalisis untuk mengetahui kemampuan masing-masing algoritma dalam mengklasifikasikan prestasi siswa berdasarkan pola kehadiran. Proses penelitian diakhiri dengan interpretasi hasil, penarikan kesimpulan, serta pemberian saran untuk pengembangan penelitian selanjutnya, baik dari segi pemilihan metode klasifikasi, teknik pengumpulan data, maupun peningkatan kualitas model agar lebih akurat dalam menangani data yang tidak seimbang.

4. HASIL DAN PEMBAHASAN

Data yang Dikumpulkan

Penelitian ini menggunakan data sekunder dari SMA Negeri 7 Palembang berupa catatan kehadiran dan nilai rapor siswa kelas XI tahun ajaran 2024/2025. Total siswa yang digunakan sebagai sampel adalah 318 siswa, dengan total catatan sebanyak 636 (gabungan semester ganjil dan genap). Rata-rata total ketidakhadiran siswa dalam satu tahun ajaran adalah 2,92 hari, dengan median 2 hari dan modus 0 hari. Rata-rata nilai rapor tahunan siswa adalah 85,97 dengan standar deviasi 1,625. Dari 318 siswa, sebanyak 287 siswa (90%) berada pada kategori prestasi Sedang dan 31 siswa (10%) pada kategori Tinggi, sementara tidak ditemukan siswa dalam kategori Rendah. Korelasi Pearson antara total ketidakhadiran dan nilai rata-rata tahunan adalah -0,35 (p-value < 0,001), yang menunjukkan hubungan linier negatif bersifat sedang. Artinya, semakin tinggi ketidakhadiran, cenderung semakin rendah prestasi akademik siswa.

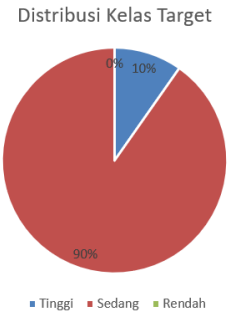
Pengolahan Data

Tahap pra-pemrosesan data meliputi *data cleaning* untuk penyeragaman *header* dan format, seleksi kolom relevan (nilai mata pelajaran, kehadiran, identitas siswa), serta penggabungan data semester menjadi data tahunan berdasarkan NISN atau nama. Setiap siswa akhirnya memiliki satu baris data unik yang mencakup total ketidakhadiran tahunan dan rata-rata nilai tahunan. Pembentukan label klasifikasi awalnya menggunakan tiga kategori (Tinggi, Sedang, Rendah) berdasarkan ketentuan sekolah (nilai ≥88 = Tinggi, >75 = Sedang, <75 = Rendah), namun karena tidak ada siswa dengan kategori Rendah, label disederhanakan menjadi dua kelas biner: Tinggi (1) dan Sedang (0). Distribusi label pada dataset akhir dapat dilihat pada

Tabel 4.1 dan Gambar 4.1, yang menunjukkan ketidakseimbangan kelas (imbalanced) dengan dominasi kelas Sedang.

Tabel 4.1. Kategori belajar siswa

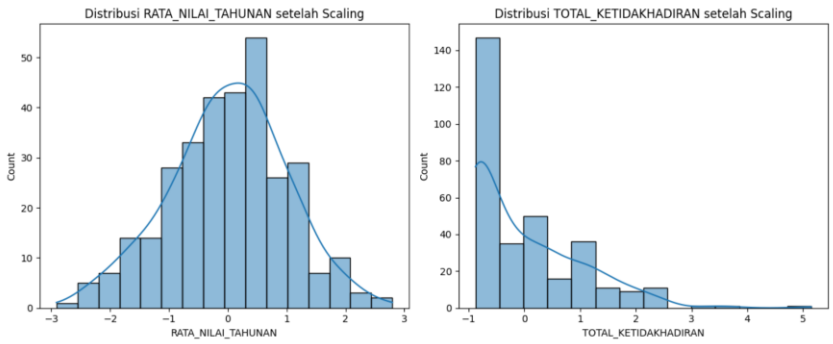
Kategori	Jumlah Siswa	Persentase (%)
Tinggi	31	10%
Sedang	287	90%
Rendah	0	0



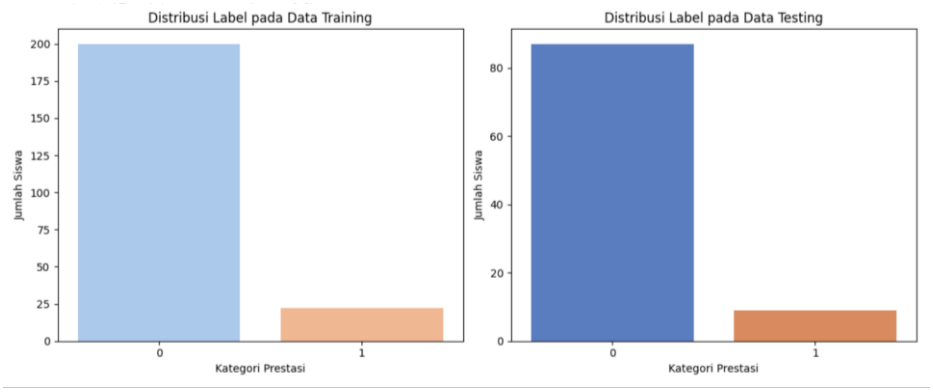
Gambar 4.1. Diagram Distribusi Label Kelas Target

Normalisasi Data

Agar semua variabel berada dalam skala yang sama dan tidak menimbulkan bias pada algoritma yang sensitif terhadap jarak (KNN) maupun regresi logistik, dilakukan standarisasi menggunakan *StandardScaler* dari *scikit-learn*. Teknik ini mengubah distribusi data menjadi memiliki rata-rata (mean) 0 dan standar deviasi 1. Hasil standarisasi pada variabel total ketidakhadiran dan rata-rata nilai tahunan ditunjukkan pada Gambar 4.2. Setelah distandarisasi, data dibagi menjadi data latih (*training*) sebesar 70% (222 siswa) dan data uji (*testing*) sebesar 30% (96 siswa) menggunakan *stratified split* agar proporsi kelas tetap terjaga. Visualisasi distribusi label pada kedua subset dapat dilihat pada Gambar 4.3



Gambar 4.2. Distribusi Total Ketidakhadiran setelah terstandarisasi.



Gambar 4.3. Hasil Visualisasi Distribusi Label Train-Test Split.

Klasifikasi dengan KNN dan Logistic Regression

Penelitian ini menerapkan dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Logistic Regression. Untuk algoritma KNN, dilakukan pencarian nilai K terbaik menggunakan *cross-validation* berbasis F1-score pada rentang K ganjil 1 hingga 19. Hasil uji coba Tabel 4.2 menunjukkan bahwa nilai K=3 memberikan rata-rata F1-score tertinggi yaitu 0,9201, sehingga dipilih sebagai parameter optimal. Kurva performa F1-score terhadap nilai K ditampilkan pada Gambar 4.6. Model KNN dengan K=3 kemudian dilatih dengan data latih dan diuji pada data uji. *Source code* penerapan KNN ditunjukkan pada Gambar 4.4 dan hasil eksekusi pada Gambar 4.5

Table 4.2. Nilai F1-Score Pada Beberapa Nilai K

Nilai K	*F1-Score*
1	0.8526
3	0.9201
5	0.8499
7	0.8368
9	0.7971
11	0.7712
13	0.7712
15	0.7853
17	0.7853
19	0.7853

```python
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

k_values = range(1, 21, 2)  # nilai K ganjil 1-19
mean_f1_scores = []

for k in k_values:
    knn = KNeighborsClassifier(n_neighbors=k)
    scores = cross_val_score(knn, X, y, cv=5, scoring='f1_macro')
    mean_f1_scores.append(scores.mean())
    print(f"K={k}, Rata-rata F1={scores.mean():.4f}")

best_index = mean_f1_scores.index(max(mean_f1_scores))
best_k = k_values[best_index]
print(f"\nK terbaik: {best_k}")

# 2. Latih model dengan train-test split menggunakan K terbaik
knn_best = KNeighborsClassifier(n_neighbors=best_k)
knn_best.fit(X_train, y_train)
y_pred_knn = knn_best.predict(X_test)


# 3. Evaluasi akhir
print("\n=== Classification Report KNN ===")
print(classification_report(y_test, y_pred_knn))

plt.figure(figsize=(6, 5))
sns.heatmap(confusion_matrix(y_test, y_pred_knn), annot=True, fmt='d', cmap='Blues',
            xticklabels=le.classes_, yticklabels=le.classes_)
plt.title(f'Confusion Matrix KNN (K={best_k})')
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.show()

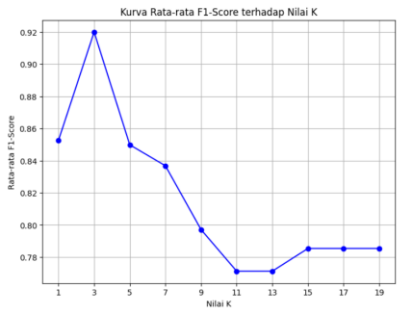
# Plot kurva F1-score terhadap nilai K
plt.figure(figsize=(8, 6))
plt.plot(k_values, mean_f1_scores, marker='o', linestyle='-', color='b')
plt.title('Kurva Rata-rata F1-Score terhadap Nilai K')
plt.xlabel('Nilai K')
plt.ylabel('Rata-rata F1-Score')
plt.xticks(k_values)
plt.grid(True)
plt.show()
```

Gambar 4.4. Source Code Penerapan KNN

```
K=1, Rata-rata F1=0.8526
K=3, Rata-rata F1=0.9201
K=5, Rata-rata F1=0.8499
K=7, Rata-rata F1=0.8368
K=9, Rata-rata F1=0.7971
K=11, Rata-rata F1=0.7712
K=13, Rata-rata F1=0.7712
K=15, Rata-rata F1=0.7853
K=17, Rata-rata F1=0.7853
K=19, Rata-rata F1=0.7853

K terbaik: 3
```

Gambar 4.5 Hasil Penerapan KNN



Gambar 4.6. Kurva Akurasi dan F1-Score

Untuk algoritma Logistic Regression, model dilatih pada data latih dengan dua variabel prediktor (total ketidakhadiran dan rata-rata nilai tahunan). Hasil perhitungan koefisien (β) dan *Odds Ratio* (OR) diperoleh melalui kode program pada Gambar 4.7 dan dirangkum dalam Tabel 4.3. Variabel rata-rata nilai tahunan memiliki koefisien sebesar 3,8028 dengan OR = 44,8282, artinya setiap kenaikan 1 poin nilai rata-rata tahunan meningkatkan peluang siswa mencapai prestasi tinggi sebesar 44,83 kali lipat. Sebaliknya, variabel total ketidakhadiran

memiliki koefisien negatif -0,3049 dengan OR = 0,7372, yang berarti setiap tambahan 1 kali ketidakhadiran menurunkan peluang prestasi tinggi menjadi 0,74 kali lipat.

```
# 8. Logistic Regression
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report

# Training & prediksi
lr = LogisticRegression()
lr.fit(X_train, y_train)
y_pred_lr = lr.predict(X_test)

print("\n=== Evaluasi Logistic Regression ===")
print("Akurasi :", accuracy_score(y_test, y_pred_lr))
print("Precision:", precision_score(y_test, y_pred_lr, average='binary'))
print("Recall   :", recall_score(y_test, y_pred_lr, average='binary'))
print("F1-Score :", f1_score(y_test, y_pred_lr, average='binary'))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_lr))


# Tambahan: Macro Avg, Weighted Avg, Support
print("\n=== Classification Report ===")
print(classification_report(y_test, y_pred_lr))

# Visualisasi confusion matrix Logistic Regression
plt.figure(figsize=(6,5))
sns.heatmap(confusion_matrix(y_test, y_pred_lr), annot=True, fmt='d', cmap='Greens',
            xticklabels=le.classes_, yticklabels=le.classes_)
plt.title('Confusion Matrix Logistic Regression')
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.show()

import numpy as np

# Interpretasi koefisien Logistic Regression + Odds Ratio
coef = lr.coef_[0]
for i, f in enumerate(features):
    odds_ratio = np.exp(coef[i])
    print(f"Pengaruh {f} terhadap prestasi (koefisien): {coef[i]:.4f}, Odds Ratio: {odds_ratio:.4f}")
```

Gambar 5.7. Source code penerapan dan evaluasi Logistic Regression.

Parameter	Nilai Variabel Ketidakhadiran	Variabel Rata Nilai Tahunan
Koefisien (β)	-0.3049	3.8028
Odds Ratio (e^β)	0.7372	44.8282

Table 4.3. Nilai koefisien (β) dan Odds Ratio

Evaluasi Model

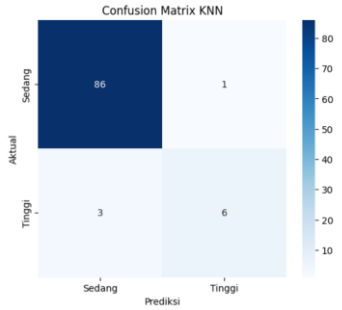
Evaluasi kedua model dilakukan pada data uji (96 siswa) dengan metrik akurasi, *precision*, *recall*, F1-score, serta *confusion matrix*. Model KNN (K=3) menghasilkan akurasi sebesar 95,83% (92 dari 96 data benar). *Classification report* KNN (Gambar 4.8) menunjukkan untuk kelas Sedang: *precision* 0,97, *recall* 0,99, F1-score 0,98; untuk kelas Tinggi: *precision* 0,86, *recall* 0,67, F1-score 0,75. *Confusion matrix* KNN (Gambar 4.9) memperlihatkan bahwa 86 siswa kelas Sedang diklasifikasikan dengan benar, 1 siswa Sedang salah diprediksi sebagai Tinggi, 6 siswa Tinggi benar, dan 3 siswa Tinggi salah diprediksi sebagai Sedang. Model Logistic Regression menghasilkan akurasi sebesar 95,83% (sama dengan KNN). *Classification report* Logistic Regression (Gambar 4.10) menunjukkan untuk kelas Sedang: *precision* 0,95, *recall* 1,00, F1-score 0,97; untuk kelas Tinggi: *precision* 1,00, *recall* 0,56, F1-score 0,714. *Confusion matrix* Logistic Regression (Gambar 4.11) menunjukkan semua 87 siswa Sedang benar, 4 siswa Tinggi salah diprediksi sebagai Sedang, dan 5 siswa Tinggi benar.

```
=== Classification Report KNN (K=3) ===
              precision    recall  f1-score   support

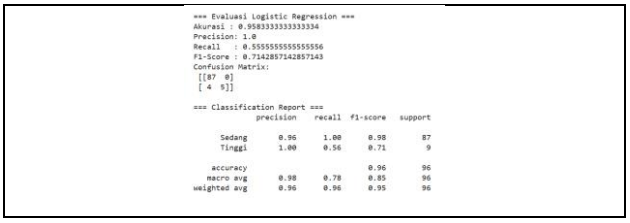
      Sedang       0.97      0.99      0.98        87
      Tinggi       0.86      0.67      0.75         9

    accuracy                           0.96        96
   macro avg       0.91      0.83      0.86        96
weighted avg       0.96      0.96      0.96        96
```

Gambar 6.8. Nilai Evaluasi, Precision, Recall, F1 Score Data Testing



Gambar 7.9. Confusion Matrix KNN Pada Data Testing



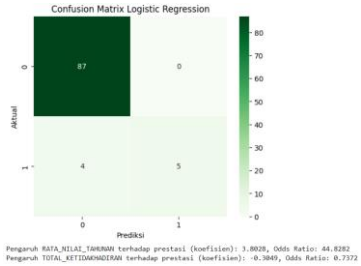
```
=== Evaluasi Logistic Regression ===
Akurasi : 0.9583333333333334
Precision: 1.0
Recall   : 0.5555555555555556
F1-Score : 0.7142857142857143
Confusion Matrix:
 [[87  0]
 [ 4  5]]

=== Classification Report ===
              precision    recall  f1-score   support

      Sedang       0.96      1.00      0.98        87
      Tinggi       1.00      0.56      0.71         9

    accuracy                           0.96        96
   macro avg       0.98      0.78      0.85        96
weighted avg       0.96      0.96      0.95        96
```

Gambar 8.10. Nilai Classificatin Report Logistic Regression Data Testing

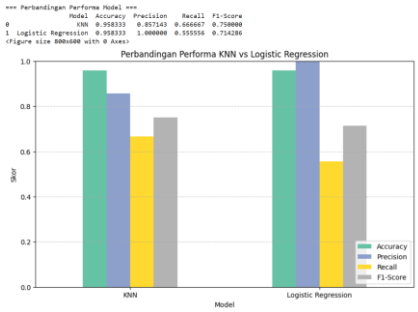


Gambar 9.11. Confusion Matrix Logistic Regression Data Testing

Perbandingan performa

Perbandingan performa kedua algoritma disajikan pada Gambar 4.24. Kedua model memiliki akurasi yang sama (95,83%). Namun, KNN unggul dalam *recall* kelas Tinggi (0,667 vs 0,556) dan F1-score kelas Tinggi (0,75 vs 0,714), sehingga lebih seimbang dalam mengenali kelas minoritas. Logistic Regression unggul dalam *precision* kelas Tinggi (1,00) serta memberikan interpretasi pengaruh variabel melalui *odds ratio* yang jelas. Dengan demikian, KNN lebih cocok untuk aplikasi yang membutuhkan deteksi kasus positif (siswa berprestasi tinggi) secara lebih sensitif, sementara Logistic Regression lebih tepat jika prioritas utama adalah meminimalkan kesalahan prediksi positif palsu. Kedua algoritma terbukti mampu

menganalisis hubungan antara kehadiran dan prestasi belajar dengan sangat baik, meskipun masih terdapat kesalahan klasifikasi pada kelas minoritas yang jumlah datanya hanya 10%. Hal ini mengindikasikan bahwa variabel kehadiran dan nilai tahunan cukup kuat sebagai prediktor, namun faktor lain di luar kedua variabel tersebut juga turut mempengaruhi prestasi akademik siswa.



Gambar 10.12. Perbandingan Performa KNN dan Logistic Regression data testing

5. KESIMPULAN DAN SARAN

Berdasarkan analisis pola kehadiran siswa dan kaitannya dengan prestasi belajar menggunakan algoritma KNN dan Logistic Regression, diperoleh beberapa temuan penting. Pertama, dari segi akurasi kedua algoritma menunjukkan performa yang sama, yaitu 95,83%. Namun demikian, algoritma KNN terbukti lebih seimbang dalam mengenali kelas minoritas, dengan nilai recall 0,667 dan F1-Score 0,750, sementara Logistic Regression memiliki recall 0,556 dan F1-Score 0,714. Meskipun demikian, Logistic Regression unggul dalam precision dengan nilai 1,000, yang berarti semua prediksi positif yang dihasilkannya benar tanpa adanya kesalahan. Kedua, hasil interpretasi dari Logistic Regression menunjukkan koefisien negatif sebesar -0,3049 dan odds ratio 0,7372. Hal ini mengindikasikan bahwa setiap tambahan satu kali ketidakhadiran siswa menurunkan peluang siswa untuk berprestasi tinggi menjadi 0,74 kali lipat. Dengan kata lain, semakin tinggi tingkat kehadiran siswa, semakin besar peluang mereka mencapai kategori prestasi “Tinggi”, dan sebaliknya, semakin banyak ketidakhadiran maka semakin kecil peluang untuk masuk ke dalam kategori berprestasi. Ketiga, baik algoritma KNN maupun Logistic Regression berhasil diterapkan untuk menganalisis pola kehadiran terhadap prestasi belajar siswa di SMA Negeri 7 Palembang, dengan jumlah data yang digunakan sebanyak 318 siswa yang kemudian dibagi menjadi data latih (training) dan data uji (testing).

DAFTAR REFERENSI

Abdullah, R. W., & Kusumastuti, R. (2025). *Analisis Pengolahan Ekstraksi Fitur Citra Untuk Klasifikasi Jenis Apel Menggunakan Scikit-Learn Dengan Algoritma K- Image Feature Extraction Processing Analysis For Apple Type Classification Using Scikit-Learn With*

The K-Nearest NEIGHBOR. *12*(1), 165–174. Https://Doi.Org/10.25126/Jtiik.2025129149

Adji, D. R., Lutfina, E., Ferdianto, B. E., Prashanti, E., Amelia, K., Anwarri, P., Informasi, S., Komputer, F. I., Nuswantoro, U. D., No, J. P., Lor, B., & Kota, K. (2025). *Metode Dan Algoritma Dalam Data Clustering : Systematic Literature Review*. *5*(1), 1–4.

Amal, B., Damayanti, S., Khonsa, A. N., Zahra, M. H., Rahmadhani, V. A., Anggraeni, W., Davina, K., Zendrato, P., Sipil, S. T., Teknik, F., Singaperbangsa, U., Pendidikan, I., & Karawang, U. S. (2025). *Penilaian Mahasiswa PBSI Unsika Terhadap Keefektifan Penggunaan Google Colab Dalam Pembelajaran Coding*. *6*(1), 66–76.

Amer, A. A., Ravana, S. D., Ahamed, R., & Habeeb, A. (2025). Effective K - Nearest Neighbor Models For Data Classification Enhancement. *Journal Of Big Data*. Https://Doi.Org/10.1186/S40537-025-01137-2

Annisa, F., Sidqie, R., Fisika, P. P., Biologi, P. P., Ar-Raniry, U., Info, A., Information, A., & Didik, P. (2023). *Educator Development Journal*. *1*, 1–8.

Ariani Hrp Nurlina, Zulaini Masruro, Siti Zahara Saragih, R. H., & Siti Suharni Simamora, T. (2022). *Buku Ajar Belajar Dan Pembelajaran* (N. Rismawati (Ed.); I). Widina Bhakti Persada Bandung.

Birant, D. (2021). *Data Mining*. Intechopen. Https://Doi.Org/10.5772/Intechopen.87784

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. *Jurnal Pendidikan Dan Teknologi Indonesia (JPTI)*, *5*(4), 1161–1173.

Farias, F. C., Ludermir, T. B., & Bastos-Filho, C. J. A. (2020). *Similarity Based Stratified Splitting : An Approach To Train Better Classifiers*.

Farokhah, L., & Korespondensi, P. (2020). *Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Implementation Of K-Nearest Neighbor For Flower Classification With Extraction Of Rgb Color Features*. *7*(6). Https://Doi.Org/10.25126/Jtiik.202072608

Hadi, A., Ryansyah, D., & Brotosaputro, G. (2025). *Penerapan Algoritma K-Nearest Neighbors (KNN) Dalam Menentukan Jenis KB Menggunakan Google Colab*. *18*(2), 120–128.

Halder, R. K., Uddin, M. N., Uddin, A., & Aryal, S. (2024). Enhancing K - Nearest Neighbor Algorithm : A Comprehensive Review And Performance Analysis Of Modifications. *Journal Of Big Data*. Https://Doi.Org/10.1186/S40537-024-00973-Y

Hanim, I., Rachmi, T., Utami, D. T., Hapsari, W., Umaroh, S. K., & Mardiana, D. (2022). *Psikologi Belajar*.

Indonesia, R. (2003). *Uu Ri Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional*. Https://Peraturan.Bpk.Go.Id/Details/43920/Uu-No-20-Tahun-2003

Jeperson Hutahaean, S. A. (2024). Penggunaan Teknologi Big Data Untuk Menganalisis Perilaku Belajar Siswa. *Jurnal Teknologi Pendidikan, Vol. 3*. Https://Doi.Org/Https://Doi.Org/10.56854/Tp.V3i1.232

Lastari, W. (2023). *Penerapan Data Mining Untuk Memprediksi Prestasi Siswa SMA Pada Dinas Pendidikan Provinsi Jambi*. *8*(2), 310–321.

Leobisa, J., Namah, M. G., Agama, I., & Negeri, K. (2022). Pengaruh Disiplin Dan Motivasi Belajar Terhadap Prestasi Belajar. *Edukatif : Jurnal Ilmu Pendidikan*, *4*(3), 3301–3309.

Liantoni, F. (2022). *Data Mining Dan Penerapan Metode* (E. Setiawan (Ed.); I). Eureka Media Aksara.

Luthfia, H. U., & Mustofa, T. A. (2024). *Upaya Meningkatkan Prestasi Akademik Dan Non-Akademik Peserta Didik Melalui Manajemen Kesiswaan Di SMA Al-Islam 1 Surakarta*. *13*(2), 1609–1616.

Marisa Fitri, Anastasia L. Maukar, T. M. A. (2021). *Data Mining Konsep Dan Penerapan*.

Deepublish.

Novita Triyasria, Egi Safitrib, M. Hardi Saputrac, Amran Rahman Syidadad, R. S., & Kurniawanf, H. (2025). *Keberhasilan, Prediksi Menggunakan, Akademik Regressi, Metode Dan, Logistik Machine, Support Vector*. *13*(4), 417–422. Https://Doi.Org/10.26418/Justin.V13i4.

Nuraeni, S., Putri, S., Syam, A., Wajdi, M. F., & Malkan, M. (2023). Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa Di SMAN 02 Manokwari. *G-Tech : Jurnal Teknologi Terapan*, *7*(1), 89–91.

Oktavianus Yopi Wardana, Mewati Ayub, A. W. (2023). *Perbandingan Akurasi Model Pembelajaran Mesin Untuk Prediksi Seleksi Masuk Perguruan Tinggi Negeri*. *9*(April), 141–153.

Pambudi, R. E., Purnomo, H., & Irawan, R. (2024). Pemanfaatan Data Mining Untuk Prediksi Prestasi Akademik Siswa Berdasarkan Pola Kehadiran, Aktivitas Belajar Mengguakan Naive Bayes Logistic Regression. *Jurnal Teknologi Informasi Mura*, *16*(2), 132–141.

Papadogiannis, I., & Wallace, M. (2024). *Educational Data Mining : A Foundational Overview*. 1644–1664.

Pradita, N. (2022). *Penerapan Data Mining Sebagai Cara Untuk Memprediksi Prestasi Siswa Berdasarkan Status Ekonomi Dan Kedisiplinan Menggunakan Metode Regresi Linier Berganda*. *4*(5), 683–696.

Prasetyawan, D., & Gatra, R. (2022). *Algoritma K-Nearest Neighbor Untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan Dan Ekonomi*. *7*(1), 56–67.

Prasetyo, Vincentius, R., Mercifia, M., Averina, A., Sunyoto, L., & Budiarjo. (2024). Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network. *Jurnal Ilmiah NERO*, *7*(1).

Purnomo, H. (2025). *Jurnal Teknologi Informasi Mura Pemanfaatan Data Mining Untuk Prediksi Prestasi Akademik Siswa Berdasarkan Pola Kehadiran ,. January 2024*.

Rahmadiah, M., Suparman, P., Panca, U., Bekasi, S., & Bks, K. (2022). *Penerapan Metode K-Nearest Neighbour Untuk Sistem Penentuan Peminjaman Modal Nasabah Bank Syariah*. *2*, 189–197.

Rahman, S. (2021). *Pentingnya Motivasi Belajar Dalam Meningkatkan Hasil Belajar*. *November*, 289–302.

Ramli, M, L. (2025). *Kajian Teoretis Tentang Faktor Perkembangan Peserta Didik Internal Dan Eksternal Dalam*. *4*(2), 13894–13904.

Retnoningsih, E., & Pramudita, R. (2020). *Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python*. *7*(2), 156–165.

Rido, M. (2022). *Peningkatan Prestasi Akademik Dan Non Akademik Peserta Didik : Implementasi Strategi Kepala Madrasah MAN 1 Bandar Lampung Abstract : 2*(2), 1–14.

Sari, D. Y., & Wang, J. (2024). *K-Nearest Neighbors (K-NN) Algorithm Model In Predicting The Graduation Rate Of Teacher Professional Education Students In Indonesia*. *4*(August), 291–310.

Sutanto, F. A., Yulianton, H., & Hartono, B. (2021). *Application Of Logistic Regression To Analyze Student Performance In Elective Courses*. *7*(12), 20–25.

Uswatun Hasanah, Isti Fatonah, Haiatin Chasanatin, M. D. (2019). *PSIKOLOGI PENDIDIKAN*.

Utami, M. D., & Nizaar, M. (2023). *Pengaruh Pendidikan Karakter Terhadap Motivasi Belajar Siswa Sekolah Dasar*. *3*, 121–128.

Vrasetya, A., Yanty, E., & Nasution, P. (2025). *Korelasi Antara Kehadiran Siswa Dan Hasil*. *8*, 561–568.

Widaningsih, S. (2022). Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(3), 2598–2611. Https://Doi.Org/10.35957/Jatisi.V9i3.859