



Klasifikasi Performa Akademik Siswa Menggunakan KNN Berdasarkan Keterhubungan Data Siswa dan Akademik

Saffira Izzati^{1*}, Dedy Hartama² Rizki Alfadillah Nasution³

^{1,2,3} Sitem Informasi STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Informasi Artikel

Sejarah Artikel:
Submit: 04 May 2026
Revisi: 04 Juni 2026
Diterima: 14 Juni 2026
Diterbitkan: 10 Juli 2026

Kata Kunci

Data mining, Klasifikasi, K-Nearest Neighbor, Performa Akademik, Siswa

Korespondensi

E-mail: vira59645@gmail.com

A B S T R A C T

This study aims to apply the K-Nearest Neighbor (KNN) algorithm to classify student academic performance based on several variables, namely age, study hours, attendance percentage, and student academic grades. Academic performance is classified into three categories: low, medium, and high. The study was conducted using a data mining approach through data preprocessing, categorical to numeric data transformation, data normalization using StandardScaler, division of training and test data, and the classification process using the KNN algorithm. To determine the best parameters, several K values were tested and the optimal value of K = 1 was obtained based on the highest accuracy. The results showed that the KNN algorithm was able to classify student academic performance with an accuracy level of 96.4%. Evaluation results based on precision, recall, F1-score, and confusion matrix showed that the model had good classification performance with a relatively small error rate. In addition, the results of parameter testing showed that the selection of the K value affected the performance of the classification model. Based on the results of the study, the K-Nearest Neighbor algorithm has proven effective as a classification method to assist in analyzing student academic performance and can be an alternative to support decision-making in the field of education.

Abstrak

Penelitian ini bertujuan mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan performa akademik siswa berdasarkan variabel usia, jenis kelamin, jam belajar, persentase kehadiran, aktivitas tambahan, serta nilai akademik. Dataset yang digunakan terdiri atas 24.970 data siswa yang diperoleh dari [sebutkan sumber dataset], dengan pembagian data menggunakan metode 80% data latih dan 20% data uji. Tahapan penelitian meliputi preprocessing data, transformasi variabel kategorikal ke numerik menggunakan teknik encoding, normalisasi data menggunakan StandardScaler, pembagian data, serta proses klasifikasi menggunakan algoritma KNN. Pengujian dilakukan pada beberapa nilai K untuk menentukan parameter optimal. Hasil penelitian menunjukkan bahwa nilai K=1 menghasilkan akurasi tertinggi sebesar 96,4%. Evaluasi model menggunakan *confusion matrix*, *precision*, *recall*, dan *F1-score* menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik pada sebagian besar kategori performa akademik. Meskipun demikian, penelitian ini masih terbatas pada penggunaan satu algoritma klasifikasi tanpa perbandingan dengan metode lain. Hasil penelitian menunjukkan bahwa algoritma KNN berpotensi digunakan sebagai alternatif pendekatan klasifikasi performa akademik siswa dan dapat mendukung pengambilan keputusan di bidang pendidikan.

This is an open access article under the CC-BY-SA license



1. Pendahuluan

Performa akademik siswa merupakan salah satu indikator penting dalam menilai keberhasilan proses pembelajaran di dunia pendidikan[1]. Tingkat performa akademik tidak hanya mencerminkan kemampuan kognitif siswa dalam memahami materi pelajaran, tetapi juga menggambarkan efektivitas metode pembelajaran yang diterapkan oleh tenaga pendidik[2]. Oleh karena itu, evaluasi terhadap performa akademik menjadi hal yang sangat penting untuk dilakukan secara berkelanjutan. Evaluasi ini tidak hanya berfungsi untuk mengetahui tingkat pemahaman siswa, tetapi juga sebagai dasar dalam pengambilan keputusan seperti pemberian bimbingan akademik, peningkatan kualitas proses pembelajaran, serta penyusunan strategi pendidikan yang lebih efektif dan tepat sasaran[3][4].

Namun demikian, dalam praktiknya pengolahan data akademik di banyak institusi pendidikan masih dilakukan secara konvensional. Proses pengolahan data yang dilakukan secara manual atau menggunakan metode sederhana seringkali menyebabkan keterbatasan dalam menghasilkan informasi yang akurat dan cepat[5]. Hal ini berdampak pada kurang optimalnya pemanfaatan data akademik sebagai dasar dalam pengambilan keputusan. Selain itu, jumlah data akademik yang terus meningkat setiap tahunnya juga menjadi tantangan tersendiri dalam proses analisis data secara manual. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah data dalam jumlah besar secara efektif, efisien, dan akurat.

Seiring dengan perkembangan teknologi informasi, pemanfaatan teknik *data mining* dan *machine learning* menjadi solusi yang dapat digunakan untuk mengatasi permasalahan tersebut[6]. *Data mining* merupakan proses ekstraksi informasi atau pola penting dari sekumpulan data dalam jumlah besar, sedangkan *machine learning* memungkinkan sistem untuk belajar dari data dan menghasilkan prediksi atau klasifikasi secara otomatis. Dengan memanfaatkan teknik ini, data akademik yang sebelumnya hanya menjadi arsip dapat diolah menjadi informasi yang bernilai dan dapat digunakan sebagai dasar dalam pengambilan keputusan di bidang pendidikan.

Salah satu metode dalam *data mining* yang banyak digunakan untuk klasifikasi adalah algoritma *K-Nearest Neighbor* (KNN)[7]. Algoritma KNN merupakan metode klasifikasi berbasis jarak yang bekerja dengan cara menentukan kelas suatu data berdasarkan kedekatannya dengan data lain yang telah diketahui kelasnya. Prinsip dasar dari algoritma ini adalah bahwa data yang memiliki kemiripan atau kedekatan jarak cenderung berada dalam kelompok yang sama. KNN dikenal sebagai algoritma yang sederhana, mudah diimplementasikan, serta memiliki performa yang cukup baik dalam berbagai kasus klasifikasi. Selain itu, KNN juga tidak memerlukan proses pelatihan model yang kompleks, sehingga cocok digunakan untuk dataset dengan ukuran kecil hingga menengah[8].

Beberapa penelitian terdahulu telah menerapkan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi performa akademik siswa dengan berbagai kombinasi variabel dan ukuran dataset. Namun demikian, sebagian besar penelitian hanya berfokus pada nilai akademik atau tingkat kehadiran sebagai variabel utama, serta belum mengkaji pengaruh kombinasi variabel demografis, aktivitas belajar, dan capaian akademik secara simultan. Selain itu, penelitian sebelumnya umumnya hanya melaporkan nilai akurasi tanpa memberikan analisis pengaruh variasi parameter K terhadap performa model. Oleh karena itu, penelitian ini menawarkan kontribusi berupa penerapan KNN pada dataset performa akademik siswa yang memadukan berbagai variabel akademik dan non-akademik, disertai analisis sensitivitas parameter K untuk memperoleh konfigurasi model yang optimal.

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan terdiri atas 24.970 data siswa dengan karakteristik akademik dan perilaku belajar yang beragam. Dataset tersebut mencakup variabel usia, jenis kelamin, jam belajar, persentase kehadiran, aktivitas tambahan, serta nilai akademik yang digunakan untuk mengklasifikasikan performa akademik siswa ke dalam kategori rendah, sedang, dan tinggi [9][10]. Penggunaan dataset dengan jumlah data yang relatif besar

diharapkan dapat meningkatkan reliabilitas hasil klasifikasi dan memberikan gambaran yang lebih representatif mengenai performa model KNN pada data pendidikan.

Selain itu, dalam penerapan algoritma KNN, pemilihan nilai parameter K menjadi salah satu faktor penting yang mempengaruhi performa model. Nilai K yang terlalu kecil dapat menyebabkan model menjadi sensitif terhadap noise, sedangkan nilai K yang terlalu besar dapat menyebabkan penurunan akurasi karena terlalu banyak mempertimbangkan data yang kurang relevan. Oleh karena itu, diperlukan pengujian beberapa nilai K untuk menentukan nilai optimal yang memberikan hasil klasifikasi terbaik.

Berdasarkan uraian tersebut, penelitian ini dilakukan dengan tujuan untuk menerapkan algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan performa akademik siswa berdasarkan beberapa variabel yang relevan[11]. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pemanfaatan teknik *data mining* di bidang pendidikan, khususnya dalam membantu proses analisis performa akademik siswa secara lebih efektif dan akurat. Selain itu, hasil penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam pengembangan metode klasifikasi pada data akademik siswa.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk mengklasifikasikan performa akademik siswa menggunakan algoritma *K-Nearest Neighbor* (KNN). Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data numerik serta analisis statistik terhadap hasil klasifikasi. Metode *data mining* digunakan untuk mengekstraksi pola dan informasi penting dari data akademik siswa sehingga dapat digunakan dalam proses pengambilan keputusan[12]. Proses penelitian dilakukan melalui beberapa tahapan, mulai dari pengumpulan data, preprocessing, hingga evaluasi model.

2.1. Rancangan Penelitian

Penelitian ini dirancang menggunakan metode klasifikasi dalam *data mining* dengan algoritma *K-Nearest Neighbor* (KNN). Metode klasifikasi digunakan untuk mengelompokkan data siswa ke dalam kategori performa akademik tertentu, yaitu rendah, sedang, dan tinggi[13]. Algoritma KNN bekerja dengan prinsip kedekatan jarak antar data, di mana suatu data akan diklasifikasikan berdasarkan mayoritas kelas dari tetangga terdekatnya[14][15]. Dalam penelitian ini, jarak antar data dihitung menggunakan metode *Euclidean Distance*. Pemilihan algoritma KNN didasarkan pada kemampuannya yang sederhana, mudah diimplementasikan, serta memiliki performa yang baik dalam berbagai kasus klasifikasi.

2.2. Data dan Sumber Data

Dataset yang digunakan dalam penelitian ini merupakan dataset publik yang diperoleh dari platform Kaggle. Dataset terdiri atas 24.970 *record* siswa dengan sembilan atribut prediktor dan satu atribut target yang merepresentasikan karakteristik demografis, aktivitas belajar, serta performa akademik siswa. Sebelum proses klasifikasi dilakukan, data diperiksa untuk memastikan tidak terdapat nilai kosong (*missing value*), duplikasi data, maupun anomali yang berpotensi memengaruhi hasil klasifikasi. Variabel-variabel dalam dataset mencakup fitur numerik dan kategorikal yang digunakan sebagai input dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Untuk memberikan gambaran yang lebih jelas mengenai jenis dan kategori fitur yang digunakan, deskripsi fitur dataset disajikan pada Tabel 1.

Tabel 1. Deskripsi Fitur Dataset

Kategori Fitur	Variabel
<i>Numerical</i>	Usia, Jenis Kelamin, Jam Belajar, Persentase Kehadiran, Aktivitas Tambahan, Nilai Matematika, Nilai Sains, Nilai, Bahasa Inggris, Nilai Keseluruhan, Kelas
<i>Categorical</i>	Jenis Kelamin, <i>Extra Activities</i>
<i>Target</i>	<i>Final Grade</i>

Tabel 1 menunjukkan kategori fitur yang digunakan dalam penelitian ini. Fitur numerik terdiri dari variabel yang memiliki nilai kuantitatif dan digunakan dalam perhitungan jarak pada algoritma KNN. Fitur kategorikal meliputi variabel yang bersifat non-numerik yang kemudian diubah menjadi bentuk numerik melalui proses *encoding*. Variabel target berupa *Final Grade* digunakan sebagai label klasifikasi yang dibagi ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi.

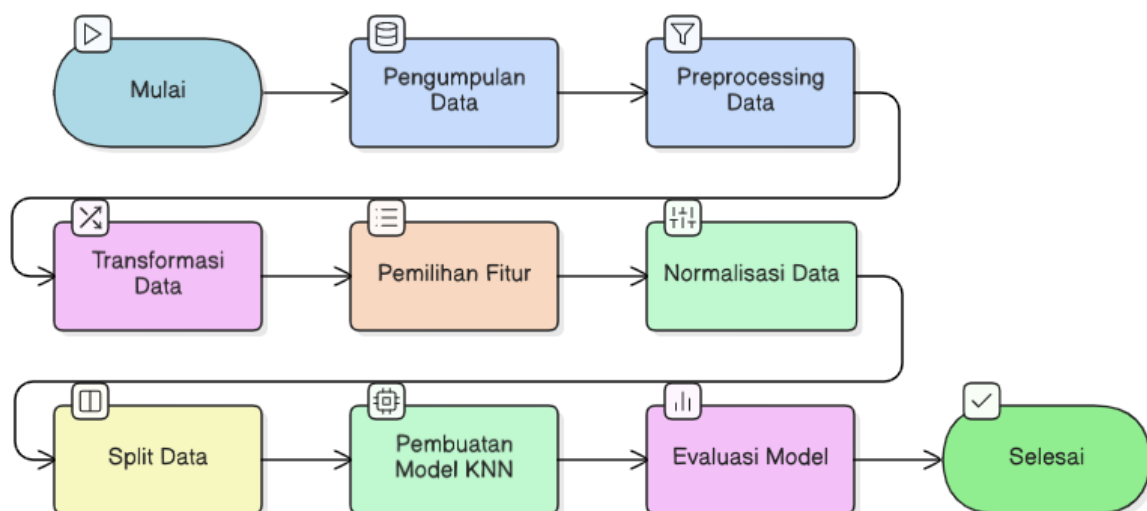
Tabel 2. Contoh Dataset Numerik Mentah

V1	V2	V3	V4	V5	V6
1	14	0	3.1	84.3	1
2	18	1	3.7	87.8	0
3	17	1	7.9	65.5	0
4	16	2	1.1	58.1	0
5	16	1	1.3	61.0	1
6	19	0	1.8	69.6	1
7	14	1	1.8	81.6	0
8	18	1	5.6	59.4	1
9	15	2	3.2	89.6	1
10	14	1	6.8	62.4	0

Tabel 2 menunjukkan contoh dataset numerik mentah yang digunakan dalam penelitian ini. Dataset tersebut telah melalui proses encoding pada variabel kategorikal seperti Jenis Kelamin dan *Extra Activities* sehingga dapat direpresentasikan dalam bentuk numerik. Setiap baris pada tabel merepresentasikan satu data siswa, sedangkan setiap kolom menunjukkan fitur yang digunakan dalam proses klasifikasi. Variabel numerik seperti usia, jam belajar, persentase kehadiran, serta nilai akademik digunakan sebagai dasar dalam perhitungan jarak pada algoritma *K-Nearest Neighbor* (KNN). Data yang ditampilkan pada tabel ini memberikan gambaran mengenai struktur dan distribusi dataset sebelum dilakukan proses normalisasi dan pembentukan model klasifikasi.

2.3. Prosedur Penelitian

Untuk menggambarkan tahapan penelitian yang dilakukan, disusun alur prosedur penelitian yang mencakup seluruh proses mulai dari pengumpulan data hingga evaluasi model. Alur tersebut disajikan dalam bentuk diagram pada Gambar 1.



Gambar 1. Alur Prosedur Penelitian

Berdasarkan Gambar 1, proses penelitian dimulai dari tahap pengumpulan data yang kemudian dilanjutkan dengan preprocessing data untuk membersihkan dan mempersiapkan dataset. Selanjutnya dilakukan transformasi data dan pemilihan fitur yang relevan sebelum dilakukan normalisasi data. Dataset kemudian dibagi menjadi data latih dan data uji untuk proses pembentukan model menggunakan algoritma *K-Nearest Neighbor* (KNN). Tahap akhir adalah evaluasi model untuk mengukur performa klasifikasi yang dihasilkan.

2.4. Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini adalah perangkat lunak berupa Python dengan library pendukung seperti *pandas* untuk pengolahan data, *numpy* untuk komputasi numerik, *scikit-learn* untuk implementasi algoritma *machine learning*, serta *matplotlib* untuk visualisasi data. Seluruh proses pengolahan data dan pembuatan model dilakukan menggunakan lingkungan pemrograman Python.

2.5. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah metode klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Model bekerja dengan menghitung jarak antara data uji dan data latih menggunakan metode *Euclidean Distance*, kemudian menentukan kelas berdasarkan mayoritas tetangga terdekat. Untuk meningkatkan performa model, dilakukan normalisasi data agar perhitungan jarak menjadi lebih optimal. Selain itu, dilakukan pengujian beberapa nilai K untuk menentukan parameter terbaik. Evaluasi model dilakukan dengan membandingkan hasil prediksi dengan data aktual menggunakan metrik akurasi, sehingga dapat diketahui tingkat keberhasilan model dalam melakukan klasifikasi performa akademik siswa.

Algoritma KNN bekerja dengan menghitung jarak antar data menggunakan *Euclidean Distance*. Secara matematis, perhitungan jarak pada algoritma KNN menggunakan *Euclidean Distance*:

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

$d(x, y)$ = jarak antar data

x_i = data uji

y_i = data latih

n = jumlah fitur

Persamaan (1) digunakan untuk menentukan kedekatan antar data dalam proses klasifikasi. Setelah jarak antar data dihitung, algoritma KNN akan memilih sejumlah K tetangga terdekat dari data uji. Kelas dari data uji ditentukan berdasarkan mayoritas kelas dari tetangga terdekat tersebut. Pemilihan nilai K yang tepat sangat berpengaruh terhadap performa model klasifikasi.

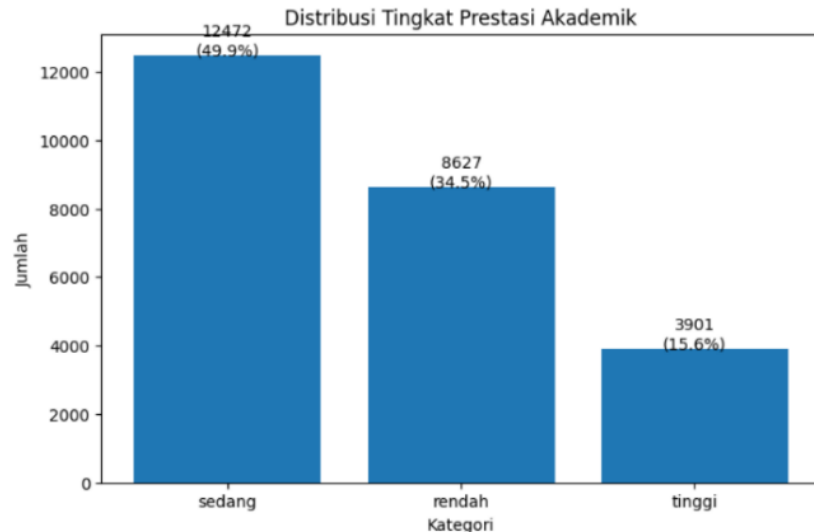
3. Hasil dan Pembahasan

Penelitian ini menerapkan algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan performa akademik siswa ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi. Variabel yang digunakan dalam penelitian ini meliputi usia, jenis kelamin, jam belajar, persentase kehadiran, aktivitas tambahan, serta nilai akademik siswa.

Tabel 3. Hasil Preprocessing Data

Usia	Jenis Kelamin	Jam Belajar	Kehadiran	Aktivitas	Nilai Matematika	Nilai Sains	Nilai Bahasa Inggris	Nilai Keseluruhan	Kelas
14	0	3.1	84.3	1	42.7	55.4	57.0	57.0	Rendah
18	1	3.7	87.8	0	57.6	68.8	64.8	61.3	Sedang
17	1	7.9	65.5	0	84.8	95.0	95.0	89.3	Tinggi
16	0	2.5	70.2	1	45.3	50.1	50.1	49.3	Rendah
19	1	6.2	90.4	1	78.5	80.3	82.7	80.5	Tinggi

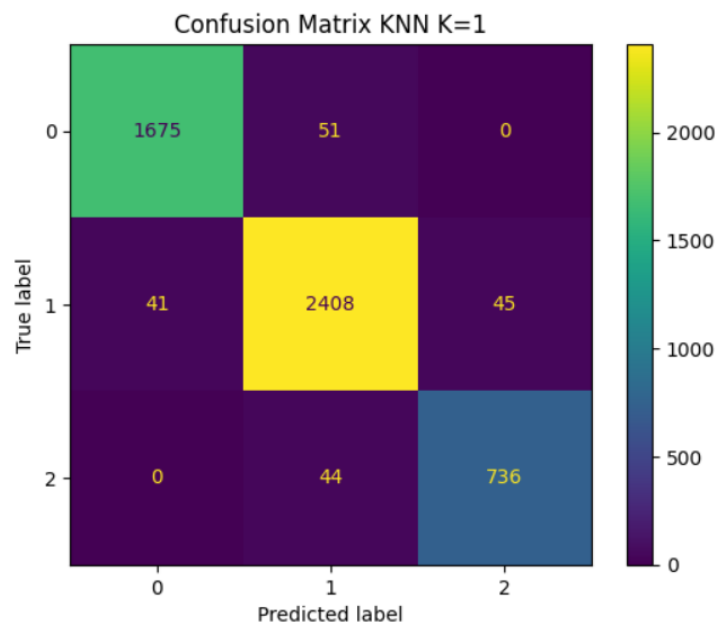
Berdasarkan Tabel 3, data telah berhasil melalui proses preprocessing, di mana variabel kategorikal seperti Jenis Kelamin dan aktivitas tambahan telah diubah ke dalam bentuk numerik. Selain itu, variabel target berupa *Final Grade* telah diklasifikasikan ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi. Hasil ini menunjukkan bahwa data telah siap digunakan dalam proses klasifikasi menggunakan algoritma KNN.



Gambar 2. Distribusi Tingkat Performa Akademik Siswa

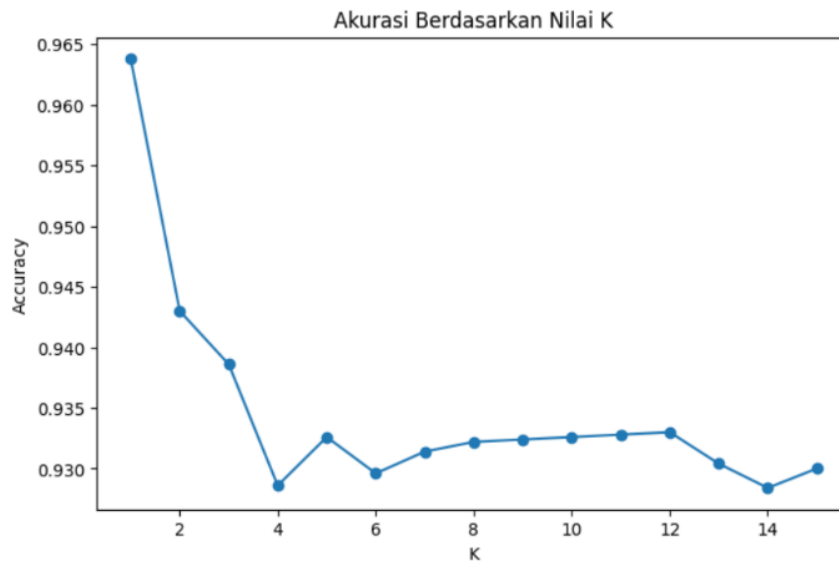
Berdasarkan distribusi data pada Gambar 2, terlihat bahwa jumlah data pada setiap kelas tidak seimbang, di mana kategori sedang memiliki jumlah data lebih banyak dibandingkan kategori rendah dan tinggi. Ketidakseimbangan ini dapat mempengaruhi performa model, karena algoritma cenderung lebih mudah mengenali kelas dengan jumlah data yang lebih dominan.

Ketidakseimbangan kelas (*class imbalance*) tersebut dapat menyebabkan model lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas. Untuk meminimalkan dampak tersebut, penelitian ini menggunakan pembagian data secara *stratified train-test split* sehingga proporsi masing-masing kelas tetap terjaga pada data latih dan data uji. Namun demikian, penelitian selanjutnya dapat mempertimbangkan penggunaan teknik penyeimbangan data seperti *SMOTE*, *oversampling*, atau *undersampling* untuk meningkatkan kemampuan model dalam mengenali seluruh kategori secara lebih merata.



Gambar 3. *Confusion matrix* KNN

Confusion matrix pada Gambar 3 menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Kesalahan klasifikasi yang terjadi relatif kecil dan umumnya terjadi pada kelas yang memiliki karakteristik yang mirip, seperti kategori sedang dan tinggi.



Gambar 4. Akurasi Berdasarkan Nilai K

Berdasarkan Gambar 4, terlihat bahwa nilai $K=1$ menghasilkan akurasi tertinggi sebesar 96,4%. Meskipun nilai $K=1$ menghasilkan akurasi tertinggi, penggunaan nilai K yang terlalu kecil berpotensi meningkatkan sensitivitas model terhadap noise dan menyebabkan overfitting. Oleh karena itu, hasil ini perlu diinterpretasikan secara hati-hati dan sebaiknya divalidasi menggunakan teknik *cross-validation* pada penelitian lanjutan untuk memastikan stabilitas model. Temuan ini menunjukkan bahwa pemilihan parameter K memiliki pengaruh yang signifikan terhadap performa klasifikasi yang dihasilkan oleh algoritma K -Nearest Neighbor.

Evaluasi model dilakukan untuk mengukur kinerja algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan data performa akademik siswa. Pengukuran dilakukan menggunakan metrik *precision*, *recall*, dan *F1-score* untuk setiap kategori kelas.

Tabel 4. Hasil Evaluasi KNN

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Rendah	0.95	0.97	0.96
Sedang	0.82	0.79	0.80
Tinggi	0.90	0.85	0.87

Berdasarkan hasil evaluasi pada Tabel 4, model menunjukkan performa yang berbeda pada masing-masing kelas. Kelas rendah memiliki nilai *F1-score* tertinggi sebesar 0,96, yang menunjukkan kemampuan model yang sangat baik dalam mengenali kategori tersebut. Sebaliknya, kelas sedang memiliki nilai *F1-score* sebesar 0,80 yang relatif lebih rendah dibandingkan kelas lainnya. Kondisi ini mengindikasikan bahwa meskipun akurasi keseluruhan model mencapai 96,4%, kemampuan klasifikasi tidak terdistribusi secara merata pada seluruh kelas. Perbedaan performa ini diduga dipengaruhi oleh ketidakseimbangan distribusi data antar kelas serta adanya kemiripan karakteristik antara kelas sedang dan tinggi sebagaimana terlihat pada *confusion matrix*. Oleh karena itu, akurasi tidak dapat dijadikan satu-satunya indikator performa model dan perlu dipertimbangkan bersama metrik *precision*, *recall*, dan *F1-score*.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma KNN mampu menghasilkan performa klasifikasi yang baik pada dataset yang digunakan. Meskipun demikian, terdapat perbedaan

tingkat keberhasilan pada masing-masing kategori kelas sehingga evaluasi model perlu mempertimbangkan tidak hanya akurasi, tetapi juga *precision*, *recall*, dan *F1-score*. Hasil ini menunjukkan bahwa algoritma KNN berpotensi digunakan sebagai pendukung analisis performa akademik siswa dengan tetap memperhatikan karakteristik distribusi data yang digunakan.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan performa akademik siswa berdasarkan variabel usia, jenis kelamin, jam belajar, persentase kehadiran, aktivitas tambahan, dan nilai akademik. Hasil pengujian menunjukkan bahwa nilai $K=1$ menghasilkan akurasi tertinggi sebesar 96,4%. Evaluasi menggunakan *precision*, *recall*, *F1-score*, dan *confusion matrix* menunjukkan bahwa model mampu melakukan klasifikasi dengan performa yang baik, meskipun masih terdapat perbedaan performa pada masing-masing kategori kelas.

Penelitian ini menunjukkan bahwa pemilihan parameter K berpengaruh terhadap hasil klasifikasi. Namun demikian, penelitian masih memiliki beberapa keterbatasan, antara lain penggunaan satu algoritma klasifikasi tanpa metode pembandingan, belum diterapkannya validasi silang (*cross-validation*), serta adanya ketidakseimbangan distribusi kelas pada dataset. Oleh karena itu, penelitian selanjutnya disarankan untuk membandingkan KNN dengan algoritma klasifikasi lainnya dan menerapkan teknik validasi serta penyeimbangan data yang lebih komprehensif untuk memperoleh hasil yang lebih robust.

Daftar Pustaka

- [1] H. S. M. Zaehol Fatah, "Penerapan Algoritma *K-Nearest Neighbor* (K-NN) untuk Memprediksi Kelulusan Mahasiswa Menggunakan RapidMiner," vol. 4, 2025.
- [2] U. I. Wahyuni and R. Kurniawan, "Algoritma *K-Nearest Neighbor* Classification Sebagai Sistem Pengelompokan Kemampuan Akademik Siswa Berbasis Web," vol. 5, no. 2, pp. 589–600, 2025.
- [3] R. A. Syaputra, M. Dewantara, D. Arya, and D. Putra, "Analisis Komparatif Algoritma KNN dan SVM dalam Klasifikasi Tingkat Keparahannya COVID-19 Global," vol. 12, no. 4, 2025.
- [4] J. Suprianto and R. A. Riesha, "Klasifikasi Kemampuan Berpikir Kreatif Peserta Didik Madrasah Ibtidaiyah Menggunakan Algoritma *K-Nearest Neighbor* Classification of Creative Thinking Abilities in Madrasah Ibtidaiyah Students Using *K-Nearest Neighbor* Algorithm," vol. 5, no. 2, pp. 158–168, 2025.
- [5] F. Tatimu, V. R. Palilingan, and I. Rianto, "Analisis Klasifikasi Kelulusan Mahasiswa menggunakan Algoritma *K-Nearest Neighbor* di Fakultas Teknik Universitas Negeri Manado," vol. 3, no. 1, pp. 117–123, 2025.
- [6] J. J. J. Manajemen, I. Komputer, and R. N. Utari, "Identifikasi Tingkat Keaktifan Media Sosial Terhadap Pola Belajar Mahasiswa Dengan *K-Nearest Neighbor* Dan Naïve Bayes," vol. 2, no. 1, 2025.
- [7] E. Asfira, A. Yanuha, I. Wisma, D. Prastya, and I. Aristia, "Classification of Post-School Tendencies of Madrasah Aliyah Students Using the *K-Nearest Neighbor* Algorithm," vol. 10, no. 2, 2026.
- [8] K. N. Knn, S. Vector, and M. Svm, "Perbandingan Algoritma *K-Nearest Neighbor* (KNN) dan Support Vector Machine (SVM) Dalam Memprediksi Kepuasan Pelanggan," vol. 04, no. 3, pp. 135–146, 2025.
- [9] N. Suarna, N. Rahaningsih, and A. A. Suarna, "Optimalisasi Prestasi Akademik Siswa Melalui Pengelompokan Indeks Prestasi Dengan K-Means Clustering Optimization Of Students ' Academic Performance Through Grade Point Index Clustering Using K-Means Clustering," vol. 4, no. 2, pp. 198–207, 2025.
- [10] P. Akademik, S. Smp, and W. Mejobo, "Perbandingan Kinerja Knn Dan Decision Tree Dalam Klasifikasi Potensi Akademik Siswa Smp Wilayah Mejobo," vol. 10, no. 2, pp. 887–897, 2025.
- [11] T. Hasil, B. Siswa, and S. Dasar, "Pengaruh Variasi Jarak Tempuh Dan Tingkat Kehadiran Terhadap Hasil Belajar Siswa Sekolah Dasar," vol. 10, pp. 263–272, 2025.
- [12] D. M. Hutabalian, P. Hutabarat, M. Prasetyo, and M. A. Irnanda, "Klasifikasi Tingkat Kedisiplinan Siswa Menggunakan Algoritma *Machine learning*: Decision Tree, KNN, dan Naive Bayes," vol. 4, no. 3, pp. 1987–1992, 2026.

- [13] F. D. Agustina, M. Arif, and S. Ahmad, "Systematic Literature Review atas Kinerja Algoritma KNN , Naïve Bayes , dan Decision Tree pada Berbagai Studi Prediksi dan Klasifikasi," vol. 3, no. 1, 2025.
- [14] J. M. Polgan, T. Hidrasi, and A. Air, "Analisis Klasifikasi Tingkat Hidrasi Manusia Menggunakan Algoritma *Machine learning* Berbasis Dataset Asupan Air Harian," vol. 14, pp. 3533–3541, 2026.
- [15] V. No, D. Hal, and A. G. Mannang, "Implementasi Sistem Informasi Berbasis Web untuk Pengelolaan Data Akademik," vol. 1, no. 1, pp. 16–22, 2025.