

Penerapan Algoritma k-Nearest Neighbor untuk Prediksi Kinerja Akademik Mahasiswa

Andi Syam Aswandi^{1*}, Nurjaya²

¹²Universitas Muhammadiyah Kolaka Utara, Indonesia

¹andisyamaswandi@gmail.com, ²nurjaya.9801@gmail.com

Abstract

Student academic performance is an important indicator in evaluating the quality of learning and academic success in higher education. The optimal utilization of academic data is required to support objective and data-driven academic decision making. This study aims to apply the k-Nearest Neighbor (k-NN) algorithm to predict student academic performance categories based on the similarity of historical academic data. This research employs a quantitative approach using a classification method and involves four main features, namely the latest Grade Point Average (GPA), the number of credits completed, the average score of core courses, and the number of repeated courses. The dataset consists of ten training data records that have been classified into high, moderate, and low academic performance categories, as well as one testing data record as a new case. The classification process is conducted using Euclidean distance with a k value of 3. The results indicate that the k-NN algorithm is able to effectively classify student academic performance, where the testing data are categorized as low academic performance based on the majority class of the nearest neighbors. These findings demonstrate that the k-NN algorithm has the potential to be applied as a basis for developing academic decision support systems, particularly for the early detection of students at risk of declining academic performance.

Keywords: k-Nearest Neighbor; Student Academic Performance; Academic Data; Classification; Decision Support System

Abstrak

Kinerja akademik mahasiswa merupakan indikator penting dalam evaluasi mutu pembelajaran dan keberhasilan studi di perguruan tinggi. Pemanfaatan data akademik secara optimal diperlukan untuk mendukung pengambilan keputusan akademik yang lebih objektif dan berbasis data. Penelitian ini bertujuan untuk menerapkan algoritma *k-Nearest Neighbor* (k-NN) dalam memprediksi kategori kinerja akademik mahasiswa berdasarkan kemiripan data akademik historis. Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi, melibatkan empat fitur utama, yaitu IPK terakhir, jumlah SKS yang telah ditempuh, rata-rata nilai mata kuliah inti, dan jumlah mata kuliah mengulang. Dataset yang digunakan terdiri dari sepuluh data latih mahasiswa yang telah diklasifikasikan ke dalam kategori kinerja akademik tinggi, sedang, dan rendah, serta satu data uji sebagai kasus baru. Proses klasifikasi dilakukan menggunakan jarak Euclidean dengan nilai parameter $k = 3$. Hasil penelitian menunjukkan bahwa algoritma k-NN mampu mengklasifikasikan kinerja akademik mahasiswa secara efektif, di mana data uji dikategorikan ke dalam kinerja akademik rendah berdasarkan mayoritas kelas tetangga terdekat. Temuan ini menunjukkan bahwa algoritma k-NN memiliki potensi untuk diterapkan sebagai dasar pengembangan sistem pendukung keputusan akademik, khususnya untuk deteksi dini mahasiswa yang berisiko mengalami penurunan kinerja akademik.

Kata Kunci: k-Nearest Neighbor; Kinerja Akademik Mahasiswa; Data Akademik; Klasifikasi; Sistem Pendukung Keputusan

Published Online 31-12-2025

I. PENDAHULUAN

Kinerja akademik mahasiswa merupakan indikator penting dalam menilai keberhasilan proses pembelajaran di perguruan tinggi. Informasi mengenai kinerja akademik tidak hanya digunakan untuk evaluasi, tetapi juga sebagai dasar pengambilan keputusan akademik, seperti pembinaan mahasiswa, perencanaan kurikulum, dan strategi peningkatan mutu pendidikan. Namun, proses evaluasi kinerja akademik sering kali masih bersifat deskriptif dan belum memanfaatkan potensi data akademik secara optimal.

Seiring dengan berkembangnya teknologi informasi, data akademik mahasiswa tersimpan secara terstruktur dalam sistem informasi akademik. Data tersebut dapat dianalisis lebih lanjut menggunakan pendekatan *data mining* untuk menghasilkan informasi prediktif. Salah satu permasalahan yang relevan adalah bagaimana memprediksi kinerja akademik mahasiswa secara akurat berdasarkan data historis yang tersedia.

Algoritma *k-Nearest Neighbor* (k-NN) merupakan salah satu metode klasifikasi yang sederhana namun efektif. Algoritma ini bekerja berdasarkan kedekatan jarak antar data dan sangat sesuai untuk data numerik. Dalam konteks akademik, k-NN dapat digunakan untuk memprediksi kategori kinerja mahasiswa dengan membandingkan kemiripan karakteristik akademik mahasiswa satu dengan lainnya.

Penelitian terdahulu menunjukkan bahwa algoritma *k-Nearest Neighbor* (k-NN) banyak digunakan dan efektif dalam berbagai domain prediksi dan klasifikasi. Penelitian [1] membuktikan k-NN mampu mengklasifikasikan data pendidikan secara efektif berbasis kemiripan data historis mahasiswa. Selanjutnya, [2] menunjukkan bahwa k-NN memiliki akurasi kompetitif dibanding algoritma lain setelah dilakukan normalisasi fitur numerik. Studi [3] menegaskan bahwa k-NN dapat memprediksi prestasi mahasiswa menggunakan fitur IPK dan SKS, dengan akurasi yang dipengaruhi oleh pemilihan nilai k . Penelitian [4] menyoroti kesederhanaan dan efektivitas k-NN dalam klasifikasi kinerja akademik mahasiswa di pendidikan tinggi. Pendekatan peningkatan akurasi melalui seleksi fitur pada k-NN ditunjukkan dalam [5], khususnya pada dataset akademik dengan atribut numerik dominan. Penelitian [6] mengungkapkan bahwa k-NN sesuai digunakan untuk prediksi kelulusan mahasiswa pada dataset kecil hingga menengah. Sementara itu, [7] dan [8] menekankan pentingnya normalisasi data serta kemudahan implementasi k-NN dalam sistem pendukung keputusan pendidikan. Analisis pada [9] menunjukkan bahwa k-NN sensitif terhadap distribusi fitur dan pemilihan jarak Euclidean, sedangkan [10] menegaskan efektivitas k-NN dalam mengklasifikasikan tingkat kinerja mahasiswa berbasis nilai akademik dan riwayat studi. Selain domain pendidikan, penelitian [11] membuktikan bahwa k-NN efektif digunakan pada bidang kesehatan untuk memprediksi status gizi balita berbasis variabel numerik, memperkuat generalisasi k-NN sebagai metode klasifikasi berbasis jarak yang andal.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma k-NN dalam memprediksi kinerja akademik mahasiswa menggunakan empat fitur utama, yaitu IPK terakhir, jumlah SKS yang telah ditempuh, rata-rata nilai mata kuliah inti, dan jumlah mata kuliah mengulang. Hasil penelitian diharapkan dapat menjadi dasar pengembangan sistem pendukung keputusan akademik.

II. METODE PENELITIAN

A. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen. Fokus penelitian adalah penerapan algoritma klasifikasi *k-Nearest Neighbor* untuk memprediksi kinerja akademik mahasiswa berdasarkan data akademik historis.

B. Dataset dan Fitur

Dataset yang digunakan berasal dari data akademik mahasiswa dan direpresentasikan dalam empat fitur utama.

Tabel 1. Fitur Akademik

Kode	Fitur
F1	IPK terakhir
F2	Jumlah SKS yang telah ditempuh
F3	Rata-rata nilai matakuliah
F4	Jumlah matakuliah mengulang

Label kelas yang digunakan adalah kinerja akademik mahasiswa, yang dikategorikan menjadi Kinerja Tinggi, Kinerja Sedang, dan Kinerja Rendah.

C. Pra-pemrosesan Data

Tahapan pra-pemrosesan meliputi:

- Pembersihan data, untuk menghilangkan data tidak lengkap.

- b. Normalisasi data, menggunakan metode *Min–Max normalization* agar setiap fitur memiliki skala yang sebanding.

D. Algoritma *k*-Nearest Neighbor

Algoritma *k*-NN melakukan klasifikasi berdasarkan mayoritas kelas dari *k* data terdekat. Jarak antar data dihitung menggunakan jarak Euclidean sebagai berikut [12].

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

Nilai *k* yang digunakan dalam penelitian ini diuji pada beberapa nilai ganjil (misalnya *k* = 3, 5, dan 7) untuk memperoleh hasil klasifikasi terbaik.

III. HASIL DAN PEMBAHASAN

1. Hasil Prediksi

Penelitian ini menggunakan 10 data latih mahasiswa yang merepresentasikan kondisi akademik berdasarkan empat fitur utama, yaitu IPK terakhir (F1), jumlah SKS yang telah ditempuh (F2), rata-rata nilai mata kuliah inti (F3), dan jumlah mata kuliah mengulang (F4). Keempat fitur tersebut dipilih karena bersifat numerik, mudah diperoleh dari sistem informasi akademik, serta mampu merepresentasikan capaian dan risiko akademik mahasiswa secara komprehensif.

Data latih selanjutnya diklasifikasikan ke dalam tiga kategori kinerja akademik, yaitu Tinggi, Sedang, dan Rendah, berdasarkan karakteristik nilai akademik yang dimiliki mahasiswa. Pengelompokan ini bertujuan untuk memudahkan proses klasifikasi menggunakan algoritma *k*-Nearest Neighbor, sekaligus memberikan interpretasi yang jelas terhadap hasil prediksi kinerja akademik mahasiswa.

Tabel 1. Data Latih Kinerja Akademik Mahasiswa

No	IPK Terakhir (F1)	Jumlah SKS (F2)	Rata-rata Nilai MK Inti (F3)	Jumlah MK Mengulang (F4)	Kinerja Akademik
1	3.85	120	88	0	Tinggi
2	3.60	110	85	1	Tinggi
3	3.25	105	80	1	Sedang
4	3.10	100	78	2	Sedang
5	2.95	98	75	2	Sedang
6	2.70	90	72	3	Rendah
7	2.60	85	70	4	Rendah
8	2.45	80	68	4	Rendah
9	3.40	108	82	1	Tinggi
10	2.80	95	74	3	Sedang

Data uji dalam penelitian ini merupakan kasus baru mahasiswa yang belum memiliki label kinerja akademik dan digunakan untuk menguji kemampuan algoritma *k*-Nearest Neighbor dalam melakukan prediksi. Data uji direpresentasikan menggunakan empat fitur akademik, yaitu IPK terakhir (F1), jumlah SKS yang telah ditempuh (F2), rata-rata nilai mata kuliah inti (F3), dan jumlah mata kuliah mengulang (F4), dengan nilai masing-masing 3.35, 85, 72, dan 2, yang mencerminkan kondisi akademik mahasiswa secara aktual.

Data uji tersebut digunakan untuk menguji kemampuan algoritma *k*-Nearest Neighbor dalam memprediksi kategori kinerja akademik mahasiswa dengan mengukur tingkat kemiripan antara data uji dan data latih. Proses prediksi dilakukan melalui perhitungan jarak antar data berdasarkan keempat fitur akademik yang digunakan, sehingga kelas kinerja akademik ditentukan oleh mayoritas kategori dari data latih yang memiliki jarak terdekat.

Pengujian dalam penelitian ini dilakukan menggunakan algoritma *k*-Nearest Neighbor (*k*-NN) dengan nilai parameter *k* = 3 serta perhitungan jarak menggunakan *Euclidean Distance*. Pemilihan nilai

k yang ganjil bertujuan untuk menghindari hasil klasifikasi yang ambigu. Berdasarkan hasil perhitungan jarak antara data uji dan seluruh data latih, diperoleh tiga data latih dengan jarak terdekat yang seluruhnya berasal dari kategori kinerja akademik rendah, sehingga proses *majority voting* menghasilkan prediksi kelas yang konsisten.

Tabel 3. Tetangga Terdekat Data Uji

No Data	Jarak	Kelas
7	3.33	Rendah
6	5.13	Rendah
8	7.45	Rendah

Berdasarkan mekanisme *majority voting* pada algoritma *k-Nearest Neighbor*, data uji diklasifikasikan ke dalam kategori kinerja akademik rendah karena mayoritas tetangga terdekat yang memiliki jarak paling kecil berasal dari kelas tersebut. Hasil ini menunjukkan bahwa secara karakteristik akademik, data uji memiliki tingkat kemiripan yang lebih tinggi dengan mahasiswa berkinerja rendah dibandingkan dengan kategori lainnya.

Untuk memvalidasi hasil perhitungan secara implementatif, algoritma k-NN diimplementasikan dalam sebuah sistem sederhana yang menampilkan hasil prediksi kinerja akademik mahasiswa. Sistem menerima input nilai fitur akademik mahasiswa dan menampilkan kategori kinerja akademik berdasarkan hasil klasifikasi.

Prediksi Kinerja Akademik Mahasiswa (k-Nearest Neighbor)
 Fitur: F1=IPK, F2=SKS, F3=Nilai MK Inti, F4=MK Mengulang. Jarak: Euclidean. k default = 3.

Form Prediksi

IPK Terakhir (F1)	Jumlah SKS (F2)	Rata-rata Nilai MK Inti (F3)	Jumlah MK Mengulang (F4)	Nilai k (ganjil)
3,35	85	72	2	3

Hitung Prediksi

Hasil Prediksi **Rendah**
 Voting: Tinggi=0, Sedang=0, Rendah=3

Gambar 1. Halaman Hasil Prediksi Kinerja Akademik Mahasiswa

Gambar 1 menunjukkan tampilan halaman hasil prediksi kinerja akademik mahasiswa, yang memuat nilai input fitur akademik, proses klasifikasi menggunakan k-Nearest Neighbor, serta hasil akhir kategori kinerja akademik. Keberadaan tampilan ini membuktikan bahwa metode yang diusulkan tidak hanya diuji secara numerik, tetapi juga telah diimplementasikan dalam bentuk sistem.

2. Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *k-Nearest Neighbor* mampu mengklasifikasikan kinerja akademik mahasiswa secara efektif dengan memanfaatkan empat fitur utama, yaitu IPK terakhir, jumlah SKS yang telah ditempuh, rata-rata nilai mata kuliah inti, dan jumlah mata kuliah mengulang. Penggunaan fitur-fitur tersebut memberikan gambaran yang cukup komprehensif terhadap kondisi akademik mahasiswa tanpa menimbulkan kompleksitas model yang berlebihan.

Meskipun data uji memiliki nilai IPK yang relatif cukup baik, hasil klasifikasi menunjukkan bahwa data tersebut lebih dekat secara jarak dengan mahasiswa berkinerja rendah. Hal ini disebabkan oleh kombinasi jumlah SKS yang belum tinggi, nilai mata kuliah inti yang berada pada tingkat sedang, serta adanya mata kuliah mengulang, yang secara keseluruhan mencerminkan potensi risiko akademik. Temuan ini menegaskan bahwa IPK bukan satu-satunya indikator dalam menilai kinerja akademik mahasiswa, melainkan perlu dipertimbangkan bersama faktor akademik lainnya.

Pendekatan berbasis kemiripan data yang digunakan oleh algoritma k-NN memungkinkan sistem mengenali pola kinerja akademik mahasiswa berdasarkan pengalaman kasus sebelumnya. Dengan demikian, algoritma ini berpotensi diterapkan sebagai alat bantu dalam sistem pendukung keputusan

akademik, khususnya untuk mendukung deteksi dini mahasiswa yang berisiko mengalami penurunan kinerja, sehingga pihak akademik dapat melakukan intervensi yang lebih tepat dan terarah.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *k-Nearest Neighbor* untuk memprediksi kinerja akademik mahasiswa berdasarkan empat fitur utama, yaitu IPK terakhir, jumlah SKS yang telah ditempuh, rata-rata nilai mata kuliah inti, dan jumlah mata kuliah mengulang. Hasil pengujian menunjukkan bahwa algoritma k-NN mampu mengklasifikasikan kinerja akademik mahasiswa secara efektif dengan pendekatan berbasis kemiripan data. Temuan penelitian ini menegaskan bahwa kombinasi beberapa indikator akademik memberikan hasil prediksi yang lebih representatif dibandingkan penggunaan satu indikator saja, serta menunjukkan potensi penerapan k-NN sebagai dasar pengembangan sistem pendukung keputusan akademik untuk deteksi dini mahasiswa berisiko.

V. DAFTAR PUSTAKA

- [1] J. J. Valero-Mas, C. Penarrubia, F. J. Castellanos, A. J. Gallego, and J. Calvo-Zaragoza, "Insights into imbalance-aware Multilabel Prototype Generation mechanisms for k-Nearest Neighbor classification in noisy scenarios," *Pattern Recognit.*, vol. 169, no. May 2025, p. 111884, 2026, doi: 10.1016/j.patcog.2025.111884.
- [2] R. Karmakar, A. K. Das, D. Sarkar, S. K. Biswas, A. Mandal, and A. Bhattacharya, "Breast Cancer Prediction using Stacking Models & Hyperparameter Tuning," *Intel. Artif.*, vol. 28, no. 76, pp. 149–165, 2025, doi: 10.4114/intartif.vol28iss76pp149-165.
- [3] M. Faisal, W. S. Utami, and S. Parmica, "Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan," *J. Sensi*, vol. 9, no. 1, pp. 11–23, 2023, doi: 10.33050/sensi.v9i1.2616.
- [4] C. A. Wint's II, A. I. Purnamasari, and T. Suprpti, "Prediksi Hunian Hotel Menggunakan Algoritma K-Nearest Neighbors," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1897–1903, 2024, doi: 10.36040/jati.v8i2.8342.
- [5] A. Alfani W.P.R., F. Rozi, and F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 1, pp. 155–160, 2021, doi: 10.29100/jupi.v6i1.1910.
- [6] N. Nurwati, N. Azizah, and Y. Santoso, "Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Mahasiswa Berpotensi Dropout," *J. Sensi*, vol. 9, no. 1, pp. 74–83, 2023, doi: 10.33050/sensi.v9i1.2624.
- [7] L. Alfath, H. Hermawan, A. Rustandiputri, R. Inzhagi, and R. Tandjilal, "Prediksi Saham PT. Aneka Tambang Tbk. dengan K-Nearest Neighbors," *JSai (Journal Sci. Appl. Informatics)*, vol. 5, no. 3, pp. 236–243, 2022, doi: 10.36085/jsai.v5i3.3975.
- [8] N. K. B. Dewi, N. K. A. Wirdiani, and D. M. S. Arsa, "Klasifikasi Kecanduan Smartphone pada Pelajar Sekolah Menengah Atas menggunakan Metode Machine Learning Berbasis Feature Weighting," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 95–103, 2022.
- [9] A. A. Amer, S. D. Ravana, and R. A. A. Habeeb, "Effective k-nearest neighbor models for data classification enhancement," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01137-2.
- [10] N. Ibrahim, N. R. Rajalakshmi, V. Sivakumar, and L. Sharmila, "An optimized hybrid ensemble machine learning model combining multiple classifiers for detecting advanced persistent threats in networks," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01272-w.
- [11] M. N. Sutoyo, "Rancang Bangun Aplikasi Untuk Memprediksi Status Gizi Balita," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, p. 136, 2018, doi: 10.20527/klik.v5i2.140.

- [12] A. Paliling and M. Sutoyo, "Combination of The MADM Model Yager and k-NN to Group Single Tuition Payments," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 326–334, 2023.