

Optimasi Prediksi Hasil Belajar Algoritma dengan Metode kNN

Syahraeni Rahmadi¹, Jimsan², Muh. Hajar Akbar³

¹²³Sistem Informasi, Universitas Sembilanbelas November Kolaka, Indonesia

Email : renhyrahmadi@gmail.com¹, jimsan@usn.ac.id², hajarakbar16@gmail.com³

Abstract

Assessing students' academic achievement is a vital component of educational evaluation, particularly in Algorithm courses, which necessitate robust logical and analytical comprehension. This study employs the k-Nearest Neighbors (kNN) algorithm to categorize students' learning outcomes based on academic metrics, including assignment scores, mid-term test scores, final exam scores, number of coding exercises, and attendance rate. The dataset comprises 53 students, with the classification procedure executed by Euclidean Distance computations to ascertain the proximity between test data and training data. The findings demonstrate that the kNN approach can accurately identify pupils, with those who participate more actively in coding exercises and maintain regular attendance often attaining superior academic success. When implemented correctly, this strategy can function as a supplementary resource for academic analysis and aid educators in recognizing kids who need additional support.

Keywords: k-Nearest Neighbors, Academic Classification, Student Assessment, Data Mining

Abstrak

Penentuan performa akademik mahasiswa merupakan aspek penting dalam evaluasi pendidikan, terutama dalam mata kuliah Algoritma yang membutuhkan pemahaman logis dan analitis yang kuat. Penelitian ini menerapkan metode k-Nearest Neighbors (kNN) untuk mengklasifikasikan hasil belajar mahasiswa berdasarkan parameter akademik, seperti nilai tugas, nilai UTS, nilai UAS, jumlah latihan koding, dan tingkat kehadiran. Dataset yang digunakan terdiri dari 53 mahasiswa, dengan proses klasifikasi dilakukan melalui perhitungan Euclidean Distance untuk menentukan jarak antara data uji dan data latih. Hasil penelitian menunjukkan bahwa metode kNN dapat mengklasifikasikan mahasiswa dengan akurasi yang baik, di mana mahasiswa dengan tingkat latihan koding yang tinggi dan kehadiran yang konsisten cenderung memiliki performa akademik lebih baik. Dengan penerapan yang tepat, metode ini dapat digunakan sebagai alat bantu dalam analisis akademik serta membantu pendidik dalam mengidentifikasi mahasiswa yang membutuhkan bimbingan lebih lanjut.

Kata Kunci : k-Nearest Neighbors, Klasifikasi Akademik, Evaluasi Mahasiswa, Data Mining

Pendahuluan

Kemajuan teknologi dalam bidang pendidikan telah mendorong pemanfaatan metode berbasis data untuk meningkatkan efektivitas pembelajaran. Salah satu tantangan utama dalam dunia akademik adalah mengukur dan menganalisis performa mahasiswa secara objektif. Dalam konteks mata kuliah Algoritma, pemahaman mahasiswa terhadap konsep dasar pemrograman sangat beragam, yang menyebabkan hasil belajar mereka bervariasi. Oleh karena itu, diperlukan pendekatan berbasis kecerdasan buatan untuk membantu dalam klasifikasi dan analisis performa akademik mahasiswa. Salah satu metode yang dapat digunakan adalah *k-Nearest Neighbors* (kNN), yang memiliki keunggulan dalam klasifikasi berbasis kedekatan data[1].

Metode kNN merupakan salah satu algoritma dalam machine learning yang digunakan untuk klasifikasi dan regresi. Prinsip utama dari kNN adalah membandingkan karakteristik suatu data dengan data lain yang sudah diketahui kategorinya, kemudian menentukan kelas berdasarkan mayoritas tetangga

terdekatnya. Algoritma ini memiliki fleksibilitas dalam menangani berbagai jenis data tanpa memerlukan asumsi distribusi tertentu, menjadikannya pilihan yang relevan dalam analisis hasil belajar mahasiswa.

Beberapa penelitian sebelumnya telah menerapkan metode KNN, yaitu: [2] membahas penggunaan Quantum KNN (QKNN) sebagai pengembangan dari metode KNN konvensional. [3] mengusulkan skema voting berbasis perbedaan metrik jarak pada algoritma KNN. [4] berfokus pada perbandingan metode manual dan otomatis dalam menentukan nilai K dalam KNN. [5] KNN digunakan untuk klasifikasi uang kuliah. [6] menerapkan KNN untuk menentukan tingkat kematangan stroberi berdasarkan fitur warna menggunakan ekstraksi fitur HSV. [7] KNN digunakan untuk klasifikasi penyakit kardiovaskular berdasarkan faktor risiko seperti usia, tekanan darah, riwayat keluarga, dan pola hidup. [8] menerapkan KNN untuk mengklasifikasikan ketidaktepatan pembayaran uang SPP di sekolah swasta. [9] menggunakan KNN untuk membangun sistem rekomendasi pembelian smartphone berbasis rating pengguna. [10] KNN digunakan untuk memprediksi gizi balita. [11] mengusulkan pendekatan baru dalam KNN, yaitu K Conditional Nearest Neighbor (kCNN), yang bertujuan untuk memperbaiki estimasi probabilitas kelas pada klasifikasi multi-kelas. [12] menerapkan KNN dan Principal Component Analysis (PCA) untuk identifikasi wajah sebagai metode biometrik.

Penelitian ini bertujuan untuk menerapkan metode kNN dalam klasifikasi performa akademik mahasiswa pada mata kuliah Algoritma. Dengan menggunakan parameter seperti nilai tugas, ujian tengah semester, ujian akhir semester, jumlah latihan koding, dan kehadiran, penelitian ini akan mengevaluasi efektivitas kNN dalam mengelompokkan mahasiswa ke dalam kategori performa akademik tertentu. Hasil dari penelitian ini diharapkan dapat memberikan wawasan dalam pemanfaatan teknik data mining untuk mendukung proses evaluasi akademik secara lebih objektif dan berbasis data.

Metodologi

Penelitian ini dilakukan melalui beberapa tahapan utama, yaitu pengumpulan data, pra-pemrosesan data, implementasi metode kNN, dan analisis hasil.

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari hasil akademik mahasiswa pada mata kuliah Algoritma. Dataset mencakup variabel seperti nilai tugas, nilai UTS, nilai UAS, jumlah latihan koding, dan kehadiran mahasiswa. Data dikumpulkan dari rekam akademik mahasiswa selama beberapa periode perkuliahan.

2. Pra-pemrosesan

Data Langkah ini mencakup pembersihan data, normalisasi, dan penanganan data yang hilang. Data yang tidak lengkap akan diolah menggunakan teknik imputasi, sedangkan normalisasi dilakukan untuk memastikan bahwa setiap fitur memiliki skala yang seragam agar tidak memengaruhi kinerja kNN.

3. Implementasi Metode kNN

Algoritma kNN diterapkan dengan menentukan nilai optimal untuk parameter k . Kriteria jarak yang digunakan adalah Euclidean Distance untuk mengukur kedekatan antar data. Euclidean Distance dihitung dengan rumus berikut[13]:

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

di mana x dan y adalah dua vektor data, dan n adalah jumlah fitur dalam dataset.

Klasifikasi dilakukan dengan membandingkan data baru dengan k tetangga terdekat yang telah diketahui kelasnya. Mayoritas kelas dari k tetangga terdekat akan menentukan kelas data baru tersebut.

4. Analisis Hasil

Setelah implementasi metode kNN, hasil klasifikasi dianalisis untuk melihat pola performa akademik mahasiswa. Faktor-faktor yang berkontribusi terhadap prediksi yang akurat akan dikaji untuk memberikan rekomendasi dalam meningkatkan hasil belajar mahasiswa.

Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini terdiri dari 53 mahasiswa dengan variabel utama berupa nilai tugas, nilai UTS, nilai UAS, jumlah latihan koding, dan kehadiran mahasiswa. Data ini mencerminkan variasi performa akademik mahasiswa dalam mata kuliah Algoritma. Dataset seperti ditunjukkan Tabel 1 berikut.

Tabel 1. Dataset Akademik

ID Mahasiswa	Nilai Tugas	Nilai UTS	Nilai UAS	Jumlah Latihan Koding	Kehadiran (%)	Kelas Hasil
MH-001	78	60	55	6	60	E
MH-002	91	48	84	11	77	B
MH-003	68	78	57	3	74	A
MH-004	54	57	86	1	88	D
MH-005	82	43	92	3	82	CD
MH-006	47	64	63	5	50	E
MH-007	60	99	65	14	76	C
MH-008	78	53	64	3	62	BC
MH-009	97	89	99	1	90	AB
MH-010	58	97	99	5	52	CD
MH-011	62	48	99	10	88	CD
MH-012	50	65	84	7	55	B
MH-013	50	92	80	14	57	D
MH-014	63	41	68	7	76	E
MH-015	92	59	54	11	58	E
MH-016	75	67	84	9	86	A
MH-017	79	86	40	10	82	A
MH-018	63	99	64	10	91	BC
MH-019	42	46	46	12	93	B
MH-020	61	83	48	13	73	CD
MH-021	92	47	63	3	64	E
MH-022	41	86	40	7	81	C
MH-023	63	74	83	1	81	B
MH-024	83	53	47	4	73	BC
MH-025	69	56	63	13	90	BC
MH-026	77	75	50	4	98	B
MH-027	41	89	90	14	98	BC
MH-028	99	79	56	5	61	B
MH-029	60	43	47	7	88	AB
MH-030	72	41	74	7	51	B
MH-031	51	45	74	13	52	D
MH-032	97	93	72	11	98	B
...
MH-050	42	99	76	12	55	D
MH-051	76	96	74	12	81	D
MH-052	90	79	83	4	53	CD
MH-053	46	60	79	14	60	AB

Data yang terkumpul mengalami tahap pembersihan, termasuk penghapusan data duplikat dan penanganan data yang hilang menggunakan teknik imputasi. Selanjutnya dilakukan normalisasi untuk menyamakan skala nilai sehingga tidak ada fitur yang mendominasi proses klasifikasi. Hasil normalisasi seperti Tabel 2 berikut.

Tabel 1. Normalisasi Dataset

ID Mahasiswa	Nilai Tugas	Nilai UTS	Nilai UAS	Jumlah Latihan Koding	Kehadiran (%)	Kelas Hasil
MH-001	0.788	0.606	0.556	0.429	0.612	E
MH-002	0.919	0.485	0.848	0.786	0.786	B
MH-003	0.687	0.788	0.576	0.214	0.755	A
MH-004	0.545	0.576	0.869	0.071	0.898	D
MH-005	0.828	0.434	0.929	0.214	0.837	CD
MH-006	0.475	0.646	0.636	0.357	0.510	E
MH-007	0.606	1.000	0.657	1.000	0.776	C
MH-008	0.788	0.535	0.646	0.214	0.633	BC
MH-009	0.980	0.899	1.000	0.071	0.918	AB
MH-010	0.586	0.980	1.000	0.357	0.531	CD
MH-011	0.626	0.485	1.000	0.714	0.898	CD
MH-012	0.505	0.657	0.848	0.500	0.561	B
MH-013	0.505	0.929	0.808	1.000	0.582	D
MH-014	0.636	0.414	0.687	0.500	0.776	E
MH-015	0.929	0.596	0.545	0.786	0.592	E
MH-016	0.758	0.677	0.848	0.643	0.878	A
MH-017	0.798	0.869	0.404	0.714	0.837	A
MH-018	0.636	1.000	0.646	0.714	0.929	BC
MH-019	0.424	0.465	0.465	0.857	0.949	B
MH-020	0.616	0.838	0.485	0.929	0.745	CD
MH-021	0.929	0.475	0.636	0.214	0.653	E
MH-022	0.414	0.869	0.404	0.500	0.827	C
MH-023	0.636	0.747	0.838	0.071	0.827	B
MH-024	0.838	0.535	0.475	0.286	0.745	BC
MH-025	0.697	0.566	0.636	0.929	0.918	BC
MH-026	0.778	0.758	0.505	0.286	1.000	B
MH-027	0.414	0.899	0.909	1.000	1.000	BC
MH-028	1.000	0.798	0.566	0.357	0.622	B
MH-029	0.606	0.434	0.475	0.500	0.898	AB
MH-030	0.727	0.414	0.747	0.500	0.520	B
MH-031	0.515	0.455	0.747	0.929	0.531	D
MH-032	0.980	0.939	0.727	0.786	1.000	B
MH-033	0.616	0.818	0.990	0.929	0.878	BC
...
MH-050	0.424	1.000	0.768	0.857	0.561	D
MH-051	0.768	0.970	0.747	0.857	0.827	D
MH-052	0.909	0.798	0.838	0.286	0.541	CD
MH-053	0.465	0.606	0.798	1.000	0.612	AB

Untuk menguji model kNN, sebuah data uji coba dengan karakteristik nilai tugas sebesar 75, nilai UTS sebesar 80, nilai UAS sebesar 85, jumlah latihan koding sebanyak 10 sesi, dan persentase kehadiran sebesar 90%. Data ini digunakan untuk menguji performa metode kNN dalam mengklasifikasikan hasil belajar mahasiswa berdasarkan parameter akademik yang telah ditentukan.

Tahapan perhitungan k-Nearest Neighbors (kNN) dalam penelitian ini dilakukan melalui beberapa langkah, yaitu:

a. Menghitung Jarak

Jarak antara data uji dan setiap data dalam dataset dihitung menggunakan Euclidean Distance. Hasil perhitungan jarak seperti Tabel 3 berikut.

Tabel 3. Hasil Perhitungan Jarak

ID Mahasiswa	Nilai Tugas	Nilai UTS	Nilai UAS	Jumlah Latihan Koding	Kehadiran (%)	Jumlah	Kelas Hasil
MH-001	0.030	0.202	0.303	0.286	0.306	1.127	E
MH-002	0.162	0.323	0.010	0.071	0.133	0.699	B
MH-003	0.071	0.020	0.283	0.500	0.163	1.037	A
MH-004	0.212	0.232	0.010	0.643	0.020	1.118	D
MH-005	0.071	0.374	0.071	0.500	0.082	1.097	CD
MH-006	0.283	0.162	0.222	0.357	0.408	1.432	E
MH-007	0.152	0.192	0.202	0.286	0.143	0.974	C
MH-008	0.030	0.273	0.212	0.500	0.286	1.301	BC
MH-009	0.222	0.091	0.141	0.643	0.000	1.097	AB
MH-010	0.172	0.172	0.141	0.357	0.388	1.230	CD
MH-011	0.131	0.323	0.141	0.000	0.020	0.616	CD
MH-012	0.253	0.152	0.010	0.214	0.357	0.986	B
MH-013	0.253	0.121	0.051	0.286	0.337	1.047	D
MH-014	0.121	0.394	0.172	0.214	0.143	1.044	E
MH-015	0.172	0.212	0.313	0.071	0.327	1.095	E
MH-016	0.000	0.131	0.010	0.071	0.041	0.254	A
MH-017	0.040	0.061	0.455	0.000	0.082	0.637	A
MH-018	0.121	0.192	0.212	0.000	0.010	0.535	BC
MH-019	0.333	0.343	0.394	0.143	0.031	1.244	B
MH-020	0.141	0.030	0.374	0.214	0.173	0.933	CD
MH-021	0.172	0.333	0.222	0.500	0.265	1.493	E
MH-022	0.343	0.061	0.455	0.214	0.092	1.165	C
MH-023	0.121	0.061	0.020	0.643	0.092	0.937	B
MH-024	0.081	0.273	0.384	0.429	0.173	1.339	BC
MH-025	0.061	0.242	0.222	0.214	0.000	0.740	BC
MH-026	0.020	0.051	0.354	0.429	0.082	0.934	B
MH-027	0.343	0.091	0.051	0.286	0.082	0.852	BC
MH-028	0.242	0.010	0.293	0.357	0.296	1.199	B
MH-029	0.152	0.374	0.384	0.214	0.020	1.144	AB
MH-030	0.030	0.394	0.111	0.214	0.398	1.148	B
...
MH-051	0.010	0.162	0.111	0.143	0.092	0.518	D
MH-052	0.152	0.010	0.020	0.429	0.378	0.988	CD
MH-053	0.293	0.202	0.061	0.286	0.306	1.147	AB

b. Menentukan Tetangga Terdekat

Dari seluruh dataset, dipilih k data dengan jarak terdekat sebagai tetangga terdekat. Dalam penelitian ini, digunakan $k = 5$, sehingga diambil tiga data dengan jarak terkecil.

Tabel 4. Tetangga Terdekat

ID Mahasiswa	Jumlah	Kelas Hasil
MH-016	0.254	A
MH-051	0.518	D
MH-018	0.535	BC
MH-033	0.538	BC
MH-011	0.616	CD
MH-036	0.619	BC
MH-017	0.637	A
MH-032	0.638	B
MH-045	0.649	B
MH-039	0.691	E
MH-002	0.699	B
MH-042	0.728	AB
MH-025	0.740	BC
MH-035	0.742	D
MH-049	0.851	AB
MH-027	0.852	BC
MH-040	0.872	E
MH-037	0.914	A
MH-047	0.932	A
MH-020	0.933	CD
MH-026	0.934	B
MH-023	0.937	B
MH-048	0.963	A
MH-007	0.974	C
MH-012	0.986	B
MH-052	0.988	CD
MH-044	1.027	A
MH-003	1.037	A
MH-014	1.044	E
MH-013	1.047	D
MH-015	1.095	E
MH-005	1.097	CD
...
MH-006	1.432	E
MH-021	1.493	E
MH-041	1.542	D

c. Menentukan Kelas Data Uji

Mayoritas kelas dari k tetangga terdekat digunakan sebagai hasil klasifikasi untuk data uji. Dari hasil perhitungan, ditemukan bahwa tiga tetangga terdekat memiliki kelas [A, D, BC, BC, CD]. Dengan

menggunakan metode voting mayoritas, hasil klasifikasi menunjukkan bahwa data uji dikategorikan sebagai BC.

Hasil klasifikasi menunjukkan bahwa mahasiswa dengan nilai tugas tinggi, partisipasi aktif dalam latihan koding, serta tingkat kehadiran yang baik cenderung memiliki performa akademik yang lebih baik. Dari penelitian ini, dapat disimpulkan bahwa metode kNN dapat digunakan secara efektif dalam mengklasifikasikan performa akademik mahasiswa berdasarkan parameter akademik tertentu. Dengan penerapan yang tepat, metode ini dapat menjadi alat bantu dalam mendukung evaluasi akademik serta membantu pendidik dalam mengidentifikasi mahasiswa yang memerlukan perhatian lebih dalam pembelajaran Algoritma.

Kesimpulan

Berdasarkan hasil penelitian, metode k-Nearest Neighbors (kNN) terbukti dapat digunakan untuk mengklasifikasikan performa akademik mahasiswa berdasarkan parameter akademik yang telah ditentukan, seperti nilai tugas, nilai UTS, nilai UAS, jumlah latihan koding, dan tingkat kehadiran. Dengan menggunakan perhitungan Euclidean Distance dan pemilihan k tetangga terdekat, hasil klasifikasi menunjukkan bahwa mahasiswa dengan keterlibatan aktif dalam latihan koding serta kehadiran yang tinggi cenderung memiliki performa akademik yang lebih baik. Penerapan metode kNN dalam evaluasi akademik dapat membantu pendidik dalam mengidentifikasi kelompok mahasiswa yang memerlukan perhatian lebih serta memberikan wawasan berbasis data dalam meningkatkan kualitas pembelajaran. Oleh karena itu, metode ini dapat menjadi alat yang efektif dalam mendukung pengambilan keputusan akademik di lingkungan pendidikan tinggi.

Daftar Pustaka

- [1] P. Rahayu *et al.*, *Buku Ajar Data Mining*. SONPEDIA Publishing Indonesia, 2024.
- [2] J. Li, J. Zhang, J. Zhang, and S. Zhang, “Quantum KNN Classification With K Value Selection and Neighbor Selection,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 43, no. 5, pp. 1332–1345, 2024, doi: 10.1109/TCAD.2023.3345251.
- [3] G. A. Pradipta, M. Liandana, P. D. W. Ayu, D. P. Hostiadi, and P. S. E. Putra, “Voting Scheme Nearest Neighbors by Difference Distance Metrics Measurement,” *JUITA J. Inform.*, vol. 11, no. 2, p. 165, 2023, doi: 10.30595/juita.v11i2.19298.
- [4] T. Mladenova and I. Valova, “Classification with K-Nearest Neighbors Algorithm: Comparative Analysis between the Manual and Automatic Methods for K-Selection,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 396–404, 2023, doi: 10.14569/IJACSA.2023.0140444.
- [5] 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.
- [6] A. Widodo, F. D. Prasetya, and H. Nugroho, “Implementasi Metode K-Nearest Neighbors (KNN) Guna Mengetahui Klasifikasi Kematangan Stroberi,” *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 3, no. 2, pp. 31–36, 2023, doi: 10.31284/j.kernel.2022.v3i2.4185.
- [7] V. Artanti, M. Faisal, and F. Kurniawan, “Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN),” *Techno.COM*, vol. 23, no. 2, pp. 467–479, 2024.
- [8] M. D. Anggraeni, K. Kusrini, and M. R. Arief, “K-Nearest Neighbor Algorithm for Classification Inaccuracy in Payment of SPP Money,” *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. dan Komput.*, vol. 4, no. 1, p. 67, 2024, doi: 10.32503/jtecs.v4i1.4891.
- [9] B. R. Setiaji, D. Q. Utama, and A. Adiwijaya, “Smartphone Purchase Recommendation System Using the K-Nearest Neighbor (KNN) Algorithm,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2180, 2022, doi: 10.30865/mib.v6i4.4753.

- [10] M. Sutoyo, “Rancang Bangun Aplikasi Untuk Memprediksi Status Gizi Balita,” *Klik-Kumpul*, vol. 5, no. 2, pp. 136–142, 2018.
- [11] H. Gweon, M. Schonlau, and S. H. Steiner, “The k conditional nearest neighbor algorithm for classification and class probability estimation,” *PeerJ Comput. Sci.*, vol. 2019, no. 5, pp. 1–21, 2019, doi: 10.7717/peerj-cs.194.
- [12] N. K. A. Wirdiani, P. Hridayami, N. P. A. Widiari, K. D. Rismawan, P. B. Candradinata, and I. P. D. Jayantha, “Face Identification Based on K-Nearest Neighbor,” *Sci. J. Informatics*, vol. 6, no. 2, pp. 150–159, 2019, doi: 10.15294/sji.v6i2.19503.
- [13] M. Sutoyo and A. Mangkona, “Schedule Implementing the Modified Euclidean Distance Method in the Course Planning of the USN Kolaka Information Systems Curriculum,” *IT J. Res. Dev.*, vol. 7, no. 1, pp. 12–22, 2022, doi: 10.25299/itjrd.2022.7976.