Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi tingkat Drop Out Mahasiswa

Syafi'ul Hamidani*[1], Robi Yanto[2], Yayang Eluis Bali Mawartika[3]
[1,2,3] Sistem Informasi STMIK Bina Nusantara Jaya Lubuklinggau, Lubuklinggau
[1,2,3] Jl. Yos Sudarso No 97 A Kel Jawa Kanan Kota Lubuklinggau Sumatera Selatan
e-mail: *[1]**hamidanipertama@gmail.com**, [2]wrtech30@gmail.com,
[3]yayangeluisbm@gmail.com,

Abstrak

Tingkat drop out (DO) mahasiswa merupakan salah satu permasalahan penting dalam pendidikan tinggi yang berdampak pada kualitas institusi dan keberhasilan akademik mahasiswa. Oleh karena itu, diperlukan metode yang mampu mengidentifikasi mahasiswa berpotensi DO secara dini. Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (K-NN) dalam mengklasifikasikan tingkat risiko drop out mahasiswa berdasarkan data akademik. Dataset yang digunakan merupakan data historis mahasiswa yang mencakup variabel Indeks Prestasi Kumulatif (IPK), Penerima Beasiswa, Sedang Bekerja, dan Sedang Berpacaran. Tahapan penelitian meliputi pengumpulan data, preprocessing (cleaning, transformasi, dan normalisasi), pembagian data menjadi data latih dan data uji, implementasi K-NN, serta evaluasi model menggunakan confusion matrix. Hasil pengujian menunjukkan bahwa algoritma K-NN mampu menghasilkan kinerja yang baik dengan nilai akurasi sebesar 80, dan F1-score 0,80. Hasil tersebut menunjukkan bahwa K-NN efektif dalam mengklasifikasikan mahasiswa yang berpotensi mengalami DO. Dengan demikian, metode ini dapat digunakan sebagai alat bantu dalam pengambilan keputusan untuk mendukung upaya pencegahan drop out secara lebih cepat dan tepat.

***Kata kunci**: Drop Out, K-NN, Confusion Matrix*

Abstract

The student dropout rate (DO) is a significant issue in higher education, impacting the quality of institutions and student academic success. Therefore, a method capable of early identification of students at risk of dropping out is needed. This study aims to apply the K-Nearest Neighbor (K-NN) algorithm to classify the risk level of student dropout based on academic data. The dataset used is historical student data that includes variables such as Grade Point Average (GPA), Scholarship Recipient, Currently Working, and Currently Dating. The research stages include data collection, preprocessing (cleaning, transformation, and normalization), data division into training and test data, K-NN implementation, and model evaluation using a confusion matrix. The test results show that the K-NN algorithm is capable of producing good performance with an accuracy value of 80 and an F1-score of 0.80. These results indicate that K-NN is effective in classifying students at risk of dropping out. Thus, this method can be used as a decision-making tool to support dropout prevention efforts more quickly and accurately.

***Keywords:** Drop Out, K-NN, Confusion Matrix*

1. PENDAHULUAN

Tingginya angka drop out (DO) di perguruan tinggi merupakan masalah serius yang berdampak luas, baik bagi institusi pendidikan maupun mahasiswa itu sendiri. Fenomena ini tidak hanya menurunkan tingkat kelulusan dan akreditasi institusi, tetapi juga menyebabkan kerugian

Submitted: 1 Mei 2026
Published: 12 Juni 2026

finansial, sosial, dan psikologis bagi mahasiswa yang gagal menyelesaikan studi mereka [1]. Dampak DO terhadap institusi meliputi penurunan reputasi, efisiensi operasional, serta potensi kehilangan pendapatan dari biaya pendidikan yang tidak terselesaikan [2]. Sementara itu, bagi mahasiswa, DO dapat mengakibatkan hilangnya kesempatan karir, beban ekonomi akibat investasi pendidikan yang tidak tuntas, serta penurunan kepercayaan diri dan motivasi belajar.

Metode konvensional dalam mendeteksi potensi DO seringkali terbatas pada analisis manual atau berdasarkan intuisi dosen dan staf akademik. Pendekatan ini kurang efektif karena tidak mampu mengolah data dalam jumlah besar secara objektif dan sistematis [3] . Oleh karena itu, diperlukan solusi berbasis teknologi yang mampu melakukan prediksi secara lebih akurat dan efisien [4].

Peran data mining khususnya algoritma K-Nearest Neighbor (KNN) semakin penting dalam upaya prediksi dan klasifikasi mahasiswa berisiko DO. Algoritma KNN dikenal karena kemudahannya dalam implementasi serta kemampuannya menangani berbagai tipe data untuk mengidentifikasi pola-pola risiko DO berdasarkan atribut akademik maupun non-akademik mahasiswa [5]. Dengan memanfaatkan teknik data mining seperti KNN, institusi dapat melakukan deteksi dini terhadap mahasiswa yang berpotensi DO sehingga intervensi preventif dapat dilakukan lebih cepat dan tepat sasaran[6] . Implementasi KNN telah terbukti memberikan tingkat akurasi yang tinggi dalam klasifikasi risiko DO di berbagai konteks pendidikan tinggi, menjadikannya salah satu metode unggulan dalam pengembangan sistem peringatan dini (*early warning system*) untuk meningkatkan retensi dan kualitas pendidikan di perguruan tinggi.

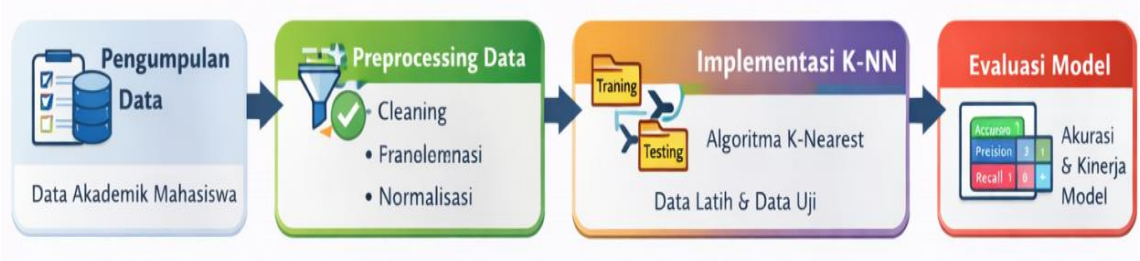
Meskipun demikian, penelitian terdahulu masih menunjukkan beberapa keterbatasan. Seperti penelitian [3] hanya memanfaatkan atribut akademik seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS, dan nilai mata kuliah sebagai variabel prediktor, sementara faktor-faktor non-akademik yang juga berpotensi memengaruhi risiko DO belum banyak dieksplorasi secara komprehensif. Selain itu penelitian [6] lebih berfokus pada pengukuran tingkat akurasi model tanpa melakukan evaluasi yang lebih mendalam menggunakan metrik seperti precision, recall, dan F1-score yang penting untuk menilai kemampuan model dalam mengidentifikasi mahasiswa berisiko DO secara tepat. Di sisi lain, implementasi algoritma KNN pada konteks perguruan tinggi yang berbeda juga dapat menghasilkan karakteristik dan pola data yang beragam sehingga diperlukan pengujian lebih lanjut untuk memastikan efektivitas model pada lingkungan institusi tertentu. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan menerapkan algoritma KNN pada data mahasiswa serta mengevaluasi kinerjanya menggunakan beberapa metrik evaluasi guna memperoleh model klasifikasi risiko DO yang lebih akurat dan representatif.

Pada STMIK Bina Nusantara Jaya Lubuklinggau tingkat drop out cukup tinggi dan menjadi salah satu permasalahan akademik yang juga mengganggu proses kegiatan di perguruan tinggi yang artinya ini harus segera diperbaiki, dengan melihat data historis drop out mahasiswa dengan mengklasifikasi tingkat drop out mahasiswa menggunakan algoritma k-nearest neightbor dengan variable IPK, penerima beasiswa, sedang bekerja, dan sedang berpacaran.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengkaji pengaruh perlakuan tertentu terhadap variabel akademik mahasiswa [7]. Pendekatan ini dipilih karena memungkinkan pengujian hipotesis secara terstruktur melalui manipulasi variabel independen dan pengukuran dampaknya pada variabel dependen[8] .

Submitted: 1 Mei 2026
Published: 12 Juni 2026



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Teknik pengumpulan data merupakan bagian yang terpenting dalam desain penelitian, karena jika judul karya ilmiah dari desain penelitian sudah disetujui untuk diteliti, maka peneliti sudah dapat mulai mengumpulkan data [9], disini peneliti mengumpulkan data akademik mahasiswa yang relevan, seperti nilai, IPK, kehadiran, serta status mahasiswa (drop out atau tidak). Data ini diperoleh dari basis data akademik institusi sebagai data historis yang akan digunakan dalam proses pelatihan dan pengujian model, kemudian penulis juga menambahkan beberapa data seperti sedang bekerja dan sedang berpacaran dimana berdasarkan hasil wawancara dan juga observasi kedua data ini sering dijadikan alasan atau kriteria mahasiswa untuk drop out dengan alasan terbanyak adalah ekonomi yang membuat mahasiswa bekerja dan tidak sanggup kuliah dan sedang berpacaran yang membuat mahasiswa tiba tiba menikah dan memutuskan tidak melanjutkan kuliah.

2. 2 Preprocessing Data

Tahapan ini adalah untuk membersihkan, menormalkan, dan mempersiapkan data sehingga dapat diolah lebih efektif oleh algoritma analisis atau model machine learning. Dataset ini terlebih dahulu dilakukan pembersihan, perbaikan, dan transformasi tanpa merubah validitas isi yang terkandung dalam dataset [10], tahapan ini bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi. Tahap ini meliputi beberapa langkah, yaitu cleaning untuk menghapus data yang tidak lengkap atau duplikat, transformasi untuk mengubah data ke dalam format yang sesuai (misalnya mengubah data kategorikal menjadi numerik), serta normalisasi untuk menyamakan skala nilai antar variabel agar tidak terjadi bias dalam perhitungan jarak pada algoritma K-NN[11]

2. 3 Pembagian Data

Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu data latih (training data) yang digunakan untuk membangun model, dan data uji (testing data) yang digunakan untuk mengukur performa model. Pada penelitian ini, dataset dibagi menjadi 80% data training dan 20% data testing. Pemilihan rasio 80:20 didasarkan pada pertimbangan bahwa model data mining memerlukan jumlah data yang cukup besar untuk proses pelatihan, sekaligus tetap menyediakan data yang memadai untuk mengevaluasi kinerja model secara objektif. Rasio 80:20 juga merupakan salah satu metode pembagian data yang paling umum digunakan dalam penelitian bidang data mining karena mampu memberikan keseimbangan antara proses pelatihan dan pengujian.

2. 4 Implementasi Algoritma K-Nearest Neighbor

Pada tahap ini, model dibangun dengan menghitung jarak antara data uji dan data latih menggunakan metode seperti Euclidean Distance.

$$d = \sqrt{\sum_{=1}^{n}(xi - ci)^2} \quad (1)$$

Kemudian, sejumlah tetangga terdekat (nilai K) dipilih untuk menentukan kelas dari data

Submitted: 1 Mei 2026
Published: 12 Juni 2026

uji berdasarkan mayoritas kelas dari tetangga tersebut, pada penelitian ini peneiliti menggunakan nilai K = 3 dikarenakan merupakan nilai ganjil yang mampu mengurangi kemungkinan terjadinya tie voting atau jumlah suara yang sama antar kelas pada proses klasifikasi. Selain itu, pemilihan K yang relatif kecil memungkinkan model mempertahankan sensitivitas terhadap pola lokal dalam data sehingga lebih efektif dalam mengidentifikasi karakteristik mahasiswa yang memiliki risiko drop out.

2. 5 Pengujian Model

Ini merupakan tahap penting untuk mengevaluasi kinerja model klasifikasi yang telah dibangun. Pengujian dilakukan menggunakan confusion matrix, Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [12]. Metode ini merupakan salah satu teknik evaluasi yang paling banyak digunakan dalam permasalahan klasifikasi karena mampu memberikan informasi yang lebih rinci dibandingkan hanya menggunakan nilai akurasi. Melalui confusion matrix, peneliti dapat mengetahui jumlah prediksi yang benar maupun prediksi yang salah untuk setiap kelas yang digunakan dalam penelitian[13].

Oleh karena itu, penelitian ini juga menggunakan metrik Precision, Recall, dan F1-Score. Precision digunakan untuk mengukur tingkat ketepatan model dalam mengidentifikasi mahasiswa yang diprediksi berisiko DO. Nilai precision yang tinggi menunjukkan bahwa sebagian besar mahasiswa yang diprediksi berisiko DO memang benar-benar termasuk dalam kategori tersebut. Sementara itu, recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh mahasiswa yang benar-benar berisiko DO. Nilai recall yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar kasus mahasiswa yang berpotensi mengalami drop out[14].

3. HASIL DAN PEMBAHASAN

3. 1 Dataset

Dataset yang digunakan dalam penelitian ini merupakan data akademik mahasiswa yang diperoleh dari sistem informasi akademik perguruan tinggi. Data tersebut bersifat historis dan digunakan untuk mengidentifikasi pola mahasiswa yang berpotensi mengalami drop out. Setiap data merepresentasikan satu individu mahasiswa dengan sejumlah atribut yang relevan terhadap kondisi akademik.

Variabel yang digunakan dalam penelitian ini meliputi IPK Terakhir, Penerima Beasiswa, Sedang Bekerja, Sedang Berpacaran. Selain itu, terdapat variabel target (label) yaitu status mahasiswa yang diklasifikasikan ke dalam dua kategori, yaitu drop out (DO) dan tidak drop out (Non-DO), dan Jumlah data yang digunakan dalam penelitian ini sebanyak 25 data mahasiswa yang terdiri dari 20 data training dan 5 data uji, seperti pada tabel berikut:

Tabel 1 Data Training

Nama	IPK	Penerima Beasiswa	Sedang Bekerja	Sedang Berpacaran	Drop Out
Mhs1	0	Ya	Tidak	Tidak	Ya
Mhs2	3.6	Ya	Ya	Ya	Tidak
Mhs3	3.66	Ya	Tidak	Tidak	Tidak
Mhs4	2.06	Ya	Tidak	Ya	Ya
Mhs5	3.15	Tidak	Tidak	Tidak	Tidak
Mhs6	3.16	Tidak	Tidak	Tidak	Tidak
Mhs7	3.16	Ya	Ya	Tidak	Ya
Mhs8	3.88	Ya	Tidak	Tidak	Tidak
Mhs9	3.96	Ya	Tidak	Tidak	Tidak

Submitted: 1 Mei 2026
Published: 12 Juni 2026

Mhs10	3.64	Ya	Ya	Tidak	Ya
Mhs11	1.21	Ya	Ya	Tidak	Ya
Mhs12	3.33	Ya	Tidak	Tidak	Tidak
Mhs13	3.83	Ya	Tidak	Ya	Tidak
Mhs14	3.36	Ya	Ya	Tidak	Tidak
Mhs15	3.56	Ya	Tidak	Ya	Tidak
Mhs16	2.93	Ya	Tidak	Ya	Tidak
Mhs17	1.5	Ya	Tidak	Ya	Ya
Mhs18	1.76	Tidak	Tidak	Ya	Ya
Mhs19	0	Ya	Tidak	Tidak	Ya
Mhs20	3.55	Ya	Tidak	Tidak	Tidak

Tabel 2 Data Uji

Nama	IPK	Penerima Beasiswa	Sedang Bekerja	Sedang Berpacaran	Drop Out
Mhs21	3.87	Ya	Ya	Tidak	Tidak
Mhs22	3.36	Ya	Tidak	Tidak	Tidak
Mhs23	1.46	Ya	Ya	Ya	Ya
Mhs24	2.53	Tidak	Ya	Ya	Ya
Mhs25	3.45	Tidak	Tidak	Ya	Tidak

3. 2 Preprocessing Data

Preprocessing data adalah teknik pemrosesan awal untuk mengubah data mentah menjadi format bersih, terstruktur, dan siap digunakan, dalam penelitian ini data dirubah dari kategorikal menjadi data numerik, seperti pada tabel berikut ini.

Tabel 3 Data numerik

Nama	IPK	Penerima Beasiswa	Sedang Bekerja	Sedang Berpacaran	Drop Out
Mhs1	0	1	0	0	1
Mhs2	3.6	1	1	1	0
Mhs3	3.66	1	0	0	0
Mhs4	2.06	1	0	1	1
Mhs5	3.15	0	0	0	0
Mhs6	3.16	0	0	0	0
Mhs7	3.16	1	1	0	1
Mhs8	3.88	1	0	0	0
Mhs9	3.96	1	0	0	0
Mhs10	3.64	1	1	0	1
Mhs11	1.21	1	1	0	1
Mhs12	3.33	1	0	0	0
Mhs13	3.83	1	0	1	0
Mhs14	3.36	1	1	0	0
Mhs15	3.56	1	0	1	0
Mhs16	2.93	1	0	1	0
Mhs17	1.5	1	0	1	1
Mhs18	1.76	0	0	1	1
Mhs19	0	1	0	0	1
Mhs20	3.55	1	0	0	0

Submitted: 1 Mei 2026
Published: 12 Juni 2026

3. 3 Euclidean Distance

Tahapan selanjutnya adalah menghitung jarak ke semua data, dengan menggunakan rumus (1)

$$D11 = \sqrt{(0 - 3.87)^2 + (1 - 1)^2 + (0 - 1)^2 + (0 - 0)^2}$$
$$D11 = \sqrt{14{,}97 + 0 + 1 + 0}$$
$$D11 = 3.977$$

$$D12 = \sqrt{(3.6 - 3.87)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 0)^2}$$
$$D12 = \sqrt{0{,}072 + 0 + 0 + 1}$$
$$D12 = 1.036$$

….

$$D520 = \sqrt{(3.55 - 3.45)^2 + (1 - 0)^2 + (0 - 0)^2 + (1 - 1)^2}$$
$$D520 = \sqrt{0{,}01 + 1 + 0 + 1}$$
$$D520 = 1.418$$

Preprocessing data adalah teknik pemrosesan awal untuk mengubah data mentah menjadi format yang bersih, terstruktur, dan siap digunakan. Dalam penelitian ini, data diubah dari bentuk awal yang masih belum terstruktur, tidak konsisten, dan berpotensi mengandung kesalahan menjadi data yang lebih terorganisasi, valid, dan siap digunakan pada tahap analisis atau pemodelan selanjutnya.

3. 4 Hasil Jarak

Apabila sudah diketahui jarak masing-masing dataset dengan data uji maka selanjutkan dilakukan proses pengurutan dari nilai jarak terkecil ke nilai jarak yang terbesar lalu ambil data sebanyak K yang di tentukandari nilai jarak terkecil, dalam penelitian ini penulis memilih nilai K=3, seperti pada tabel salah satu data uji berikut.

Tabel 4 Jarak Mhs21

Drop out	Hasil	Peringkat
Ya	3.997	19
Tidak	1.036	7
Tidak	1.022	6
Ya	2.297	15
Tidak	1.587	13
Tidak	1.582	12
Ya	0.710	3
Tidak	1.000	4
Tidak	1.004	5
Ya	0.230	1
Ya	2.660	16
Tidak	1.136	9
Tidak	1.415	10
Tidak	0.510	2
Tidak	1.448	11
Tidak	1.698	14
Ya	2.760	18
Ya	2.730	17
Ya	3.997	19

Submitted: 1 Mei 2026
Published: 12 Juni 2026

Tidak	1.050	8

Dari hasil jarak diambil nilai dengan peringkat 3 teratas dari masing masing data uji dan diambil klasifikasi dari masing masing data uji, seperti pada tabel berikut.

Tabel 5 Hasil Klasifikasi

	Peringakat 1/ DO	Peringkat 2/ DO	Peringkat 3/ DO	Hasil Klasifikasi
Mhs21	0.230/ Ya	0.510/ Tidak	0.710/ Ya	Drop Out
Mhs22	0.030/ Tidak	0.190/ Tidak	0.300/ Tidak	Tidak DO
Mhs23	1.001/ Ya	1.031/ Ya	1.166/ Ya	Drop Out
Mhs24	0.770/ Ya	1.105/ Ya	1.105/ Ya	Drop Out
Mhs25	1.006/ Tidak	1.041/ Tidak	1.044/ Ya	Tidak DO

3. 5 Evaluasi Model

Dari hasil tersebut didapatlah perbandingan hasil klasifikasi dengan nilai real yang tertera seperti pada tabel berikut.

Tabel 6 Hasil Perbandingan Klasifikasi

	Hasil Klasifikasi	Hasil Real
Mhs21	Drop Out	Tidak DO
Mhs22	Tidak DO	Tidak DO
Mhs23	Drop Out	Drop Out
Mhs24	Drop Out	Drop Out
Mhs25	Tidak DO	Tidak DO

Kemudian dilakukan dengan uji model confusion matrix dan didapatlah hasil seperti berikut :

Tabel 7 Hasil Confusion Matrix

	Prediksi Tidak DO	Prediksi DO
Aktual Tidak DO	2 (TP)	1 (FN)
Aktual DO	0 (FP)	2 (TN)

Selanjutnya adalah menghitung nilai accuracy, precision, recall dan F-1 score nya:

Accuracy:

$$\frac{TP+TN}{Total} = \frac{2+2}{5} = 0{,}8 = 80\ \%$$

Precision:

$$\frac{TP}{TP+FP} = \frac{2}{2+0} = 1.0$$

Recall:

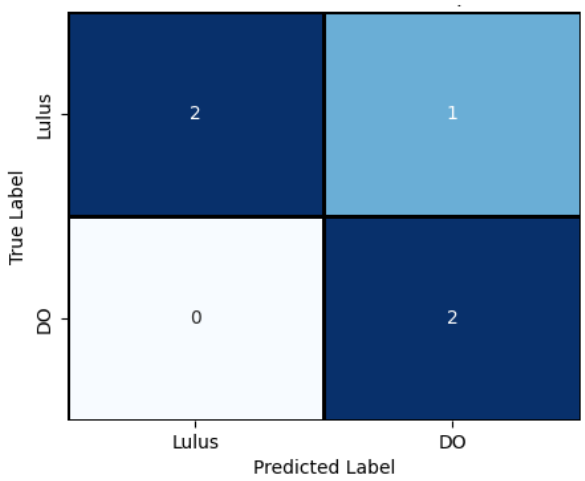
$$\frac{TP}{TP+FN} = \frac{2}{2+1} = 0{,}66$$

F1- Score:

Submitted: 1 Mei 2026
Published: 12 Juni 2026

$$2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall} = 2x\ \frac{1x0,66}{1+0,66} = 0,80$$

Sehingga didapatlah model confusion matrix dalam bentuk heatmap seperti gambar berikut:



Gambar 2. Heatmap Hasil Confusion Matrix

Berdasarkan hasil evaluasi menggunakan Confusion Matrix, model klasifikasi menunjukkan kinerja yang cukup baik dengan nilai F1-score sebesar 80%. Hal ini mengindikasikan bahwa model memiliki keseimbangan yang baik antara precision dan recall dalam mengklasifikasikan data, selain itu, tidak ditemukan kesalahan dalam memprediksi kelas Drop Out (DO) sebagai Lulus (L), yang ditunjukkan dengan nilai False Positive (FP) sebesar 0. Kondisi ini menunjukkan bahwa model memiliki tingkat ketepatan yang tinggi dalam menghindari kesalahan klasifikasi terhadap kelas positif.

Namun demikian, masih terdapat kesalahan dalam memprediksi data yang sebenarnya termasuk dalam kelas Lulus (L) menjadi Drop Out (DO), yang tercermin dari nilai False Negative (FN) yang tidak nol. Hal ini menunjukkan bahwa model masih memiliki keterbatasan dalam mengenali seluruh data pada kelas Lulus secara optimal.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma K-Nearest Neighbor (K-NN) mampu digunakan secara efektif dalam mengklasifikasikan tingkat drop out (DO) mahasiswa. Model yang dibangun menunjukkan tingkat keberhasilan yang cukup baik dalam membedakan antara mahasiswa yang berpotensi DO dan yang tidak, yang ditunjukkan melalui nilai evaluasi seperti akurasi yang cukup besar yaitu 80%, dan F1-score senilai 0,80. Hal ini menunjukkan bahwa K-NN memiliki kemampuan dalam mengenali pola dari data historis mahasiswa berdasarkan kedekatan karakteristik antar data.

5. SARAN

Berdasarkan hasil penelitian, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya. Pertama, disarankan untuk menambahkan variabel lain yang lebih beragam, seperti faktor ekonomi, kondisi psikologis, dan latar belakang sosial mahasiswa, sehingga model yang dihasilkan dapat memberikan hasil klasifikasi yang lebih akurat dan komprehensif. Selain itu, penelitian selanjutnya perlu diarahkan pada pengembangan

Submitted: 1 Mei 2026
Published: 12 Juni 2026

metodologi yang lebih beragam, penambahan jumlah dan cakupan dataset agar model dapat merepresentasikan kondisi yang lebih luas, serta melakukan perbandingan dengan berbagai algoritma klasifikasi lainnya, seperti Decision Tree, Naïve Bayes, Random Forest, maupun Support Vector Machine (SVM). Langkah tersebut diharapkan dapat meningkatkan validitas, reliabilitas, dan kemampuan generalisasi hasil penelitian sehingga diperoleh model prediksi risiko drop out yang lebih optimal dan dapat diterapkan secara efektif pada berbagai lingkungan perguruan tinggi.

DAFTAR PUSTAKA

[1] M. Mardolkar and N. Kumaran, "Forecasting and Avoiding Student Dropout Using the K-Nearest Neighbor Approach," *SN COMPUT. SCI.*, vol. 1, no. 2, p. 96, Mar. 2020, doi: 10.1007/s42979-020-0102-0.

[2] T. A. Marzuqi, E. Kristiani, and Marcel, "Prediksi Mahasiswa Drop-Out Di Universitas XYZ," *JTIIK*, vol. 11, no. 6, pp. 1345–1350, Dec. 2024, doi: 10.25126/jtiik.1168689.

[3] P. S. Saputra, "ANALISIS PREDIKTIF DROPOUT MAHASISWA BERDASARKAN KINERJA AKADEMIK SEMESTER AWAL MENGGUNAKAN," *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, vol. 07, no. 01, 2026.

[4] S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *JTIIK*, vol. 10, no. 1, pp. 165–172, Feb. 2023, doi: 10.25126/jtiik.20231016643.

[5] M. A. Hassan, A. H. Muse, and S. Nadarajah, "Predicting Student Dropout Rates Using Supervised Machine Learning: Insights from the 2022 National Education Accessibility Survey in Somaliland," *Applied Sciences*, vol. 14, no. 17, p. 7593, Aug. 2024, doi: 10.3390/app14177593.

[6] V. Tania, I. N. Khasanah, S. Albani, N. Azizah, and N. Jihan, "PENERAPAN K-NEAREST NEIGHBORS (K-NN) DALAM MENILAI POTENSI DROP OUT MAHASISWA: STUDI PADA ASPEK AKADEMIK, SOSIAL, DAN EKONOMI," *MATHunesa Jurnal Ilmiah Matematika*, vol. 13, no. 2, 2025.

[7] N. S. Aprilia and N. H. Yoenanto, "Pengaruh Regulasi Emosi dan Persepsi Dukungan Sosial terhadap Stres Akademik Mahasiswa yang Menyusun Skripsi," *BRPKM*, vol. 2, no. 1, pp. 19–30, Jan. 2022, doi: 10.20473/brpkm.v2i1.31924.

[8] R. D. Putra, M. A. Murtadho, M. Isnaini, and M. W. Afgani, "Design Penelitian Kuantitatif: Pengertian dan Macam-macam Jenisnya," vol. 5, no. 3, 2025.

[9] S. Hamidani and R. Yanto, "Sistem Cerdas Pemilihan Makanan Sehat Berbasis Case-Based Reasoning dan SMART untuk Edukasi Pemenuhan Gizi Masyarakat," *Jurnal Pustaka AI*, vol. 5, no. 2, Aug. 2025, doi: 10.55382/jurnalpustakaai.v5i2.1209.

[10] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia: Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information," *MALCOM*, vol. 4, no. 1, pp. 273–281, Jan. 2024, doi: 10.57152/malcom.v4i1.1085.

[11] P. Meriyana, A. Rizky Pratama, E. Nurlaelasari, and A. Ratna Juwita, "Penerapan Algoritma KNN dan Naive Bayes untuk Klasifikasi Stunting pada Balita di Desa Pasirjengkol," *jurnalpetisi*, vol. 6, no. 2, pp. 82–94, Jul. 2025, doi: 10.36232/jurnalpetisi.v6i2.2020.

[12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.

[13] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "PENGUKURAN KINERJA SISTEM KUALITAS UDARA DENGAN TEKNOLOGI WSN MENGGUNAKAN CONFUSION MATRIX," *JIU*, vol. 6, no. 2, Jan. 2021, doi: 10.26877/jiu.v6i2.6552.

Submitted: 1 Mei 2026
Published: 12 Juni 2026

[14] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak: Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children's Growth," *MALCOM*, vol. 4, no. 2, pp. 413–422, Feb. 2024, doi: 10.57152/malcom.v4i2.1210.

Submitted: 1 Mei 2026
Published: 12 Juni 2026