

## Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Ketidakhadiran Mahasiswa Berdasar Presensi dan IPK

Hidayat Muhammad Nur<sup>1</sup>, Vadlya Maarif<sup>2\*</sup>, Fabriyan Fandi Dwi Imaniawan<sup>3</sup>, Sardiarinto<sup>4</sup>,  
Eko Saputro<sup>5</sup>

Universitas Bina Sarana Informatika<sup>1,2,3,4,5</sup>

[hidayat.hmm@bsi.ac.id](mailto:hidayat.hmm@bsi.ac.id)<sup>1</sup>, [vadlya.vlr@bsi.ac.id](mailto:vadlya.vlr@bsi.ac.id)<sup>2\*</sup>, [fabriyan.fbf@bsi.ac.id](mailto:fabriyan.fbf@bsi.ac.id)<sup>3</sup>, [sardiarinto.sdo@bsi.ac.id](mailto:sardiarinto.sdo@bsi.ac.id)<sup>4</sup>,  
[eko.eto@bsi.ac.id](mailto:eko.eto@bsi.ac.id)<sup>5</sup>,

---

Diterima (26-03-2025)	Direvisi (11-04-2025)	Disetujui (21-04-2025)
--------------------------	--------------------------	---------------------------

---

**Abstrak** - Tingkat ketidakhadiran mahasiswa sering dipengaruhi berbagai faktor yaitu rendahnya tingkat presensi mahasiswa, ipk yang tidak memenuhi kriteria, tugas yang tidak diselesaikan, kondisi finansial, keterlibatan dalam kegiatan organisasi, dan faktor lainnya. Maka, penting untuk memprediksi ketidakhadiran mahasiswa sejak dini. Atas pertimbangan hal tersebut, sehingga bisa dilakukan pencegahan seperti monitoring presensi secara aktif, program remedial, konseling akademik dan lainnya. Model klasifikasi Penelitian ini mengaplikasikan algoritma K-Nearest Neighbor (K-NN) untuk menganalisis data kehadiran dan IPK mahasiswa dalam memprediksi ketidakhadiran di salah satu Perguruan tinggi swasta di Kabupaten Cilacap yang terdiri dari 30 mahasiswa pada tiap kelas. Mencakup tentukan parameter  $k$ , hitung jarak antara dataset data training dan data testing (*Euclidean Distance*), urutkan data training jarak terkecil, berikut menetapkan kelas. Di antara faktor tersebut Hasil menunjukkan bahwa perkiraan mahasiswa dengan tingkat kehadiran  $\geq 75\%$  dan IPK  $\geq 3,00$  memiliki peluang kelulusan sebesar 75%, sedangkan yang tidak dapat memenuhi kriteria memiliki peluang ketidakhadiran sebesar 77,27%.

Kata Kunci : *K-Nearest Neighbor (k-NN), prediksi kelulusan, ipk, data mining, klasifikasi mahasiswa.*

**Abstract** - *The failure rate of students is often influenced by various factors, namely low student attendance, GPA that does not meet the criteria, unfinished assignments, financial conditions, involvement in organizational activities, and others. Therefore, it is important to predict student failure early on. Based on these considerations, prevention can be carried out such as active attendance monitoring, remedial programs, academic counseling and others. Classification model This study applies the K-Nearest Neighbor (K-NN) algorithm to analyze student attendance and GPA data in predicting failure at a private university in Cilacap Regency consisting of 30 students in each class. Includes determining the parameter  $k$ , calculating the distance between the training data dataset and the testing data (Euclidean Distance), sorting the training data with the smallest distance, and determining the class. Among these factors, the results show that the estimated students with an attendance rate of  $\geq 75\%$  and a GPA of  $\geq 3.00$  have a 75% chance of passing, while those who do not meet the criteria have a 77.27% chance of failing.*

**Keywords:** *K-Nearest Neighbor (k-NN), graduation prediction, GPA, data mining, student classification.*

### I. PENDAHULUAN

Lulus kuliah tepat waktu dalam jangka empat tahun tentu menjadi hal paling ideal bagi para mahasiswa yang berjuang meraih gelar sarjana. Namun tak semua mahasiswa mampu lulus tepat waktu karena berbagai faktor (Wijaya & Anita, 2023). Berbagai faktor yang memengaruhi ketidakhadiran ini, seperti rendahnya tingkat kehadiran di kelas, IPK yang tidak memenuhi kriteria akademik, tugas yang tidak diselesaikan, keterbatasan finansial, dan keterlibatan dalam kegiatan organisasi secara berlebihan (Yanti et al., 2019). Dari berbagai faktor tersebut, tingkat kehadiran di bawah 75% dan IPK kurang dari 3,00 terbukti menjadi penyebab utama. Data menunjukkan bahwa mahasiswa dengan tingkat

kehadiran  $\geq 75\%$  dan IPK  $\geq 3,00$  memiliki peluang kelulusan sebesar 90%, sedangkan mahasiswa yang tidak memenuhi kriteria ini memiliki peluang ketidakhadiran sebesar 75%. Fakta ini menunjukkan pentingnya memantau kedua indikator tersebut untuk mencegah ketidakhadiran mahasiswa.

Urgensi penelitian ini terletak pada perlunya prediksi ketidakhadiran mahasiswa sejak dini (Pratama et al., 2024). Dengan memprediksi potensi ketidakhadiran, institusi dapat mengambil langkah-langkah preventif seperti monitoring kehadiran secara aktif, pemberian program remedial untuk meningkatkan pemahaman akademik, serta konseling bagi mahasiswa yang menghadapi kendala non-akademik (Sibuea et

al., 2023). Prediksi yang akurat memungkinkan tindakan yang lebih cepat dan tepat sasaran, sehingga dapat meningkatkan tingkat kelulusan dan kualitas pendidikan secara keseluruhan (Fatunnisa & Marcos, 2024).

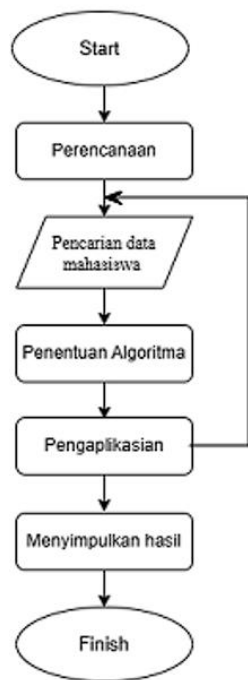
Penelitian ini bertujuan untuk mengaplikasikan algoritma *K-Nearest Neighbor (K-NN)* dalam menganalisis tingkat kehadiran dan IPK mahasiswa untuk memprediksi potensi ketidakkelulusan. *K-NN* dipilih karena kesederhanaannya dalam menangani data numerik dengan akurasi yang baik, sehingga cocok untuk kebutuhan analisis ini. Penelitian dilakukan melalui tahapan pengumpulan data, pra-pemrosesan data, implementasi algoritma *K-NN*, serta evaluasi model prediksi. Hasil dari penelitian ini diharapkan dapat

## II. METODOLOGI PENELITIAN

Berikut ini merupakan flowchart dari cara kerja penelitian ini, dalam memprediksi ketidakkelulusan mahasiswa. Peneliti membagi tahapan penelitian menjadi beberapa bagian, antara lain: dapat lihat Gambar 1.

1. Perencanaan
2. Pencarian data mahasiswa
3. Penentuan algoritma
4. Pengaplikasian algoritma
5. Menyimpulkan hasil

Jika terdapat evaluasi pada tahapan pengaplikasian, mempunyai *rules* untuk melakukan pengulangan ke tahap pencarian data.



Sumber : Penelitian (2025)

Gambar. 1. Cara Kerja Prediksi

## III. HASIL DAN PEMBAHASAN

### 1. Hasil

Setelah melakukan analisis terhadap perancangan dengan tahapan data mining untuk menghasilkan prediksi ketidakkelulusan mahasiswa dengan menggunakan algoritma *K-Nearest Neighbor (K-NN)*. Algoritma *KNN (K-Nearest Neighbors)* adalah metode klasifikasi yang bekerja dengan cara mencari *K* data yang paling dekat (tetangga terdekat) dari data yang akan diklasifikasikan, kemudian menggunakan kelas mayoritas dari tetangga tersebut untuk menentukan kelas dari data baru (Afifah, 2020).

Data yang digunakan dalam penelitian ini yaitu data mahasiswa yang atributnya berisi nim, nama lengkap, kehadiran, ipk, dan keterangan lulus atau tidak lulus.

Peneliti menggunakan aplikasi RapidMiner untuk melakukan proses data mining, yang menghasilkan informasi prediksi ketidakkelulusan mahasiswa.

### 2. Pembahasan

Berikut adalah langkah-langkah algoritma *K-Nearest Neighbor (K-NN)* berdasarkan tahapan data mining untuk memprediksi kelulusan mahasiswa tepat waktu:

- a. Menentukan Nilai *K*  
Nilai *k* dalam algoritma *K-NN* ditentukan langsung secara manual. Dalam penelitian ini, nilai *k* yang digunakan adalah 5.
- b. Menghitung Jarak Euclidean  
Hitung jarak Euclidean adalah menghitung jarak antara data training dan data testing. Dalam penelitian ini, data training diambil 70% dari data total. Sedangkan data testing diambil 30% dari data total.
- c. Mengurutkan Jarak  
Hasil perhitungan diurutkan dari yang paling dekat hingga yang paling jauh.
- d. Menentukan Kelompok Data Berdasarkan Label Mayoritas  
Dengan nilai *k* = 5, ambil lima jarak terkecil untuk menentukan kelompok data hasil uji berdasarkan label mayoritas dari tetangga terdekat.

### 3. Implementasi Rapid Miner

RapidMiner adalah salah satu alat yang digunakan dalam data mining dan menyediakan sekitar 500 operator yang mencakup berbagai fungsi, seperti input, output, preprocessing data, dan visualisasi. RapidMiner berfungsi sebagai perangkat lunak mandiri untuk analisis data sekaligus sebagai mesin data mining yang dapat diintegrasikan dengan sistem lain (Maulid, 2023).

Data yang digunakan dalam penelitian ini yaitu data mahasiswa yang atributnya berisi nim, nama lengkap, kehadiran, ipk, dan keterangan lulus atau tidak lulus.

Nilai Kehadiran dihitung berdasarkan Jumlah pertemuan yang dihadiri mahasiswa dibandingkan dengan total jumlah pertemuan yang ada yaitu 14 pertemuan, kemudian dinyatakan dalam bentuk desimal. Rumusnya adalah:

$$\text{Kehadiran} = \frac{\text{Jumlah Kehadiran}}{14}$$

IPK (Indeks Prestasi Kumulatif) diperoleh dari rata-rata nilai mata kuliah yang diambil oleh mahasiswa selama masa studinya. Nilai akhir setiap mata kuliah biasanya dihitung berdasarkan beberapa komponen penilaian berikut:

- a. Nilai Tugas  
Nilai ini mencakup berbagai tugas individu maupun kelompok yang diberikan selama satu semester. Tugas dapat berupa laporan, esai, proyek, atau presentasi.
- b. Nilai Ujian  
Termasuk nilai ujian tengah semester (UTS) dan ujian akhir semester (UAS). Ujian ini biasanya memiliki bobot yang cukup besar dalam menentukan nilai akhir.
- c. Nilai Kehadiran dan Keaktifan  
Beberapa mata kuliah memberikan nilai untuk kehadiran di kelas dan partisipasi aktif dalam diskusi atau kegiatan perkuliahan.

Status kelulusan (0 = tidak lulus, 1 = lulus) ditentukan berdasarkan kriteria tertentu yang sudah ditetapkan. Dalam penelitian ini, status dihitung dari Kriteria berikut :

- a. Kehadiran Minimum  
Mahasiswa harus memenuhi tingkat kehadiran minimum adalah 75% (dalam bentuk desimal = 0.75).
- b. IPK Minimum  
Mahasiswa harus mencapai nilai IPK minimum adalah 3.0.
- c. Kombinasi Kehadiran dan IPK  
Status kelulusan diberikan kepada mahasiswa yang memenuhi kedua kriteria sekaligus:
  - 1) Kehadiran  $\geq 0.75$
  - 2) IPK  $\geq 3.0$
  - 3) Jika salah satu atau kedua kriteria tidak terpenuhi, status mahasiswa adalah 0 (tidak lulus). Jika keduanya terpenuhi, status adalah 1 (lulus).

Tabel 1. Datasheet

No	Nama	NIM	Kehadiran	IPK	Status
1	Lestari Widyastuti	2020012	0,625	2,6	0
2	Nadia Fadilah	2020014	0,75	3	0

3	Putri Andini	2020016	0,6875	3,1	0
4	Siti Nurhayati Ulfa	2020018	0,5625	2,7	0
5	Novitasari	2020020	0,8125	3,2	0
6	Vina Wahyuni	2020021	0,75	3,1	0
7	Xaverius Dwi Siti	2020023	0,625	2,9	0
8	Siti Nurhayati	2020018	0,5625	2,7	0
9	Amanda Putri	2020026	0,6875	3	0
10	Cinta Dewi	2020028	0,75	3,2	0
11	Dedi Mulyadi	2020029	0,875	3,5	1
12	Eka Nurhayati	2020030	0,625	2,8	0
13	Galuh Widodo	2020032	0,6875	3,1	0
14	Hana Suryani	2020033	0,875	3,6	1
15	Joko Susilo	2020035	0,75	3,2	0
16	Kurnia Dewi	2020036	0,625	2,9	0
17	Lestari Kusuma	2020037	0,9375	3,7	1
18	Mawar Sari	2020038	0,6875	3	0
19	Nurhayati Wijaya	2020039	0,875	3,5	1
20	Cinta Dewi Ulfa	2020028	0,75	3,2	0
21	Novitasari Lestari	2020020	0,8125	3,2	0
22	Widyastuti	2020012	0,625	2,6	0
23	Wahyu Nugroho	2020022	1	3,8	1
24	Dedi Mulyadi	2020029	0,875	3,5	1
25	Siti Nurhayati	2020018	0,5625	2,7	0
26	Zainal Abidin	2020025	0,9375	3,7	1
27	Hana Suryani	2020033	0,875	3,6	1
28	Oktaviani Dwi	2020015	1	3,8	1
29	Rahmat Budi	2020017	0,9375	3,9	1
30	Zainal Abidin	2020025	0,9375	3,7	1

Sumber : Penelitian (2025)

Berikut adalah proses implementasi berdasarkan data di atas :

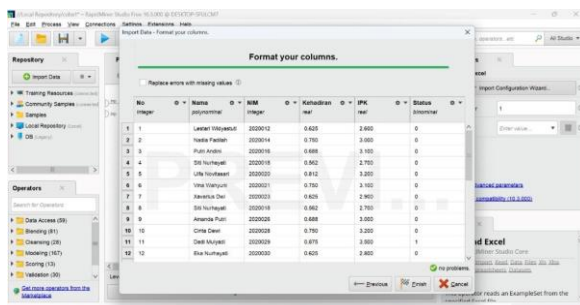


Sumber : Penelitian (2025)

Gambar 2. Subproses Data Implementasi

Pada proses tersebut terdapat beberapa operator yaitu :

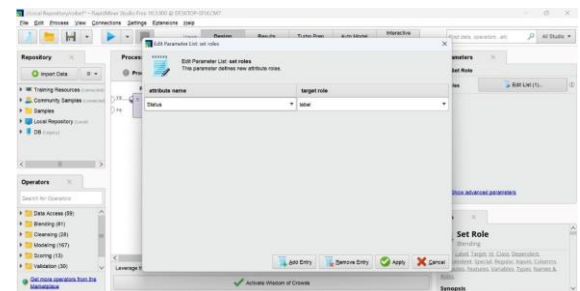
- a. Read Excel, sebagai pembaca data eksternal dari Microsoft Excel.



Sumber : Penelitian (2025)

Gambar 3. Proses Input Data

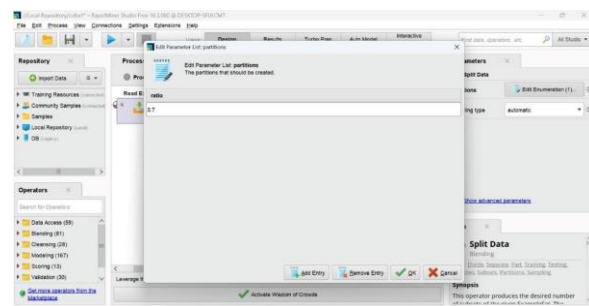
- b. Set Role, untuk memastikan target yang akan diprediksi. Dalam penelitian ini atribut yang diambil adalah status dan target role nya berupa label.



Sumber : Penelitian (2025)

Gambar 4. Edit Set Role

- c. Split Data, untuk membagi antara data training dan data testing. Dalam penelitian ini diambil ratio sebesar 0,7 data testing, dan sisanya data training.



Sumber : Penelitian (2025)

Gambar 5. Edit Ratio

- d. k-NN, untuk membangun model. Dalam penelitian ini nilai  $k$  yang digunakan adalah 5.

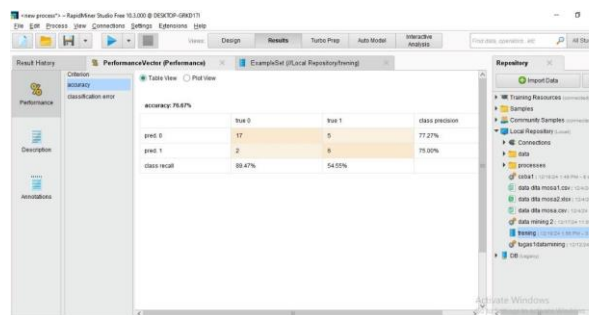


Sumber : Penelitian (2025)

Gambar 6. Penelitian K

- e. Apply Model, untuk menguji model yang telah dilatih.
- f. Performance (Classification) untuk melihat hasil prediksi. Classification dipilih karena K-NN adalah algoritma klasifikasi yang digunakan untuk memprediksi berdasarkan data training.

Tingkat hasil akurasi dari Performance (Classification) sebesar 77,27% untuk prediksi 0 (tidak lulus), dan 75% untuk prediksi 1 (lulus) dari model diatas, dapat dilihat pada gambar dibawah:



Sumber : Penelitian (2025)

Gambar 7. Hasil Accuracy Algoritma K-NN

Dari hasil diatas dapat diketahui bahwa :

- 1) Terdapat 17 data dengan status tidak lulus (0) yang tepat prediksi
- 2) Terdapat 6 data dengan status lulus (1) yang tepat prediksi
- 3) Terdapat 2 data dengan status tidak lulus (0) yang salah prediksi
- 4) Terdapat 5 data dengan status lulus (1) yang salah prediksi.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian yang menggunakan algoritma K-Nearest Neighbor (K-NN) untuk memprediksi ketidakhadiran mahasiswa, beberapa kesimpulan dapat diambil sebagai berikut:

Penerapan Algoritma K-NN: Algoritma K-NN terbukti efektif dalam memprediksi potensi ketidakhadiran mahasiswa berdasarkan dua indikator utama, yaitu tingkat kehadiran dan IPK. Nilai K yang dipilih adalah 5, yang berhasil memberikan model dengan akurasi yang cukup baik.

Hasil Prediksi: Dari model yang diimplementasikan menggunakan RapidMiner, didapatkan akurasi sebesar 77,27% untuk prediksi mahasiswa yang tidak lulus (0) dan 75% untuk prediksi mahasiswa yang lulus (1). Hal ini menunjukkan bahwa algoritma K-NN mampu memberikan prediksi yang cukup akurat meskipun terdapat beberapa kesalahan prediksi. langkah preventif : Dengan adanya prediksi ini, institusi dari universitas dapat mengambil tindakan pencegahan, seperti melakukan pemantauan kehadiran secara aktif, menyediakan program remedial untuk membantu mahasiswa yang mengalami kesulitan, serta Dengan akurasi prediksi yang cukup baik, institusi dapat dengan cepat menangani masalah akademik maupun non-akademik yang mungkin dihadapi mahasiswa.

Secara keseluruhan, penerapan algoritma K-NN sangat mendukung kebijakan akademik yang lebih efektif dalam upaya meningkatkan tingkat kelulusan mahasiswa. Dengan kemampuan untuk memprediksi potensi ketidakhadiran berdasarkan data tingkat kehadiran dan IPK, algoritma ini memungkinkan institusi pendidikan mengambil tindakan preventif yang lebih tepat sasaran, seperti program remedial, pemantauan intensif, dan konseling bagi mahasiswa yang berisiko. Penerapan ini tidak hanya membantu meningkatkan peluang kelulusan tepat waktu, tetapi juga mendorong peningkatan kualitas pendidikan secara keseluruhan.

#### V. REFERENSI

- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127. <https://doi.org/10.31294/ijcit.v6i2.10438>
- Fatunnisa, A., & Marcos, H. (2024). Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest. *Jurnal Manajemen Informatika (JAMIKA)*, 14(1), 101–111. <https://doi.org/10.34010/jamika.v14i1.12114>
- Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 3(2), 59–66. <https://doi.org/10.36040/jati.v3i2.860>
- Muningsih, E., Nur, H. M., Dwi Imaniawan, F. F., Saifudin, Handayani, V. R., & Endiarto, F. (2020). Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012101>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Pratama, R., Herdiana, R., Hamonangan, R., & Anwar, S. (2024). Analisis Prediksi Kelulusan Mahasiswa Menggunakan Metode Artificial Neural Network. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 687–693. <https://doi.org/10.36040/jati.v8i1.8762>
- Rohiyah, S., Handayani, F., & Widiarti, A. (2023). Pengaruh Status Sosial Ekonomi Dan Lingkungan Kampus Terhadap Ipk Mahasiswa. *Journal of Business Education and Social*, 3(1), 17–24. <https://doi.org/10.33592/jbes.v3i1.3381>
- Sibuea, P., Lusianti, E. F., Aprilia, S. P., Ilmanun, L., Dalimunthe, W. V. P., & Adelia, T. (2023). Konsep Program Remedial dan Pengayaan Sebagai Upaya Tindak Lanjut Evaluasi Pembelajaran. *Jurnal Pendidikan Tambusai*, 7(3), 31993–32000.
- Wijaya, A. S., & Anita, G. F. (2023). Kebiasaan Belajar Mahasiswa Jalur Prestasi Berdasarkan Pilihan Kelulusan dan Jenis Kelamin. *Jurnal Wahana Ilmiah Pendidikan*,

9(April), 1–8.  
<https://doi.org/https://doi.org/10.5281/zenodo.7865525>  
Yanti, R., Windarto, A. P., & Suhada, S. (2019). Analisis Faktor Yang Mempengaruhi Ipk Mahasiswa Terhadap Penentuan Kualitas

Proses Pembelajaran Dengan Metode Electre Ii. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 36–41.  
<https://doi.org/10.30865/komik.v3i1.1696>