

# PREDICTION OF STUDENT GRADUATION USING THE K-NEAREST NEIGHBOR METHOD

Mukminina<sup>1</sup>

Program Studi Manajemen Informatika<sup>1</sup>  
STMIK IKMI Cirebon  
<https://ikmi.ac.id/page/18/?lang=de>  
[mukminina.310103@gmail.com](mailto:mukminina.310103@gmail.com)

(\*) Corresponding Author : [mukminina.310103@gmail.com](mailto:mukminina.310103@gmail.com).  
Published : 30 Desember 2025

**Abstract**—The rapid development of information technology has encouraged universities to utilize information systems to manage student academic data effectively for evaluation and graduation decisions. However, in many institutions, graduation monitoring remains manual or semi-manual, leading to time-consuming processes, data processing errors, and delayed identification of students at risk of late graduation. This study addresses the absence of an integrated graduation prediction system capable of automatically processing academic data and generating reliable predictions. The proposed system applies the K-Nearest Neighbor (KNN) method through stages including requirements analysis, academic data collection, data preprocessing, system flow design, and model implementation and testing. KNN classifies graduation outcomes based on Euclidean Distance by selecting the optimal K value and assigning the majority class among nearest neighbors. Implementation and evaluation were conducted using Google Colab. Results demonstrate that the system delivers accurate graduation predictions that support academic monitoring, targeted guidance, informed decision-making, and improved academic management.

**Keywords:** Information System; Student Graduation Prediction; K-Nearest Neighbor; Academic Data; Machine Learning

**Abstrak**—Perkembangan teknologi informasi mendorong perguruan tinggi untuk mengoptimalkan sistem informasi dalam pengelolaan data akademik, khususnya dalam evaluasi dan prediksi kelulusan mahasiswa. Namun, di beberapa institusi, proses pemantauan kelulusan masih dilakukan secara manual atau semi-manual sehingga membutuhkan waktu lama dan berpotensi menimbulkan kesalahan pengolahan data. Akibatnya, informasi mengenai potensi kelulusan mahasiswa sering tidak tersedia secara tepat waktu dan menyulitkan pihak akademik dalam memberikan bimbingan yang efektif. Penelitian ini bertujuan membangun sistem informasi prediksi kelulusan mahasiswa berbasis metode K-Nearest Neighbor (KNN). Tahapan penelitian meliputi pengumpulan dan pra-pemrosesan data akademik, perancangan sistem, serta penerapan dan pengujian metode KNN menggunakan perhitungan Euclidean Distance. Implementasi dilakukan dengan memanfaatkan Google Colab sebagai media pengolahan data. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan mampu menghasilkan prediksi kelulusan mahasiswa secara cepat dan akurat, sehingga dapat mendukung pengambilan keputusan serta meningkatkan efektivitas pemantauan dan bimbingan akademik

**Kata Kunci :** Sistem Informasi; Prediksi Kelulusan Mahasiswa; K-Nearest Neighbor; Data Akademik; Machine Learning,

## INTRODUCTION

Perkembangan teknologi informasi yang semakin pesat telah mendorong berbagai institusi, termasuk perguruan tinggi, untuk memanfaatkan teknologi dalam pengelolaan

data akademik (Laudon & Laudon, 2020). Perguruan tinggi dituntut untuk mampu mengolah data mahasiswa secara efektif dan efisien guna mendukung proses pengambilan keputusan akademik yang tepat dan berbasis data.

Perguruan tinggi dituntut untuk mampu mengolah data mahasiswa secara efektif dan efisien guna mendukung proses pengambilan keputusan akademik yang tepat dan berbasis data. [3], [4]. Salah satu data akademik yang sangat penting adalah data kelulusan mahasiswa, karena kelulusan merupakan indikator keberhasilan proses pembelajaran serta kualitas penyelenggaraan pendidikan di perguruan tinggi (Suryani et al., 2021).

Kajian sebelumnya menunjukkan bahwa landasan teoritis dan empiris untuk mendukung penelitian yang sedang dilakukan.[6], metode *K-Nearest Neighbor* (KNN) [7], Metode KNN banyak digunakan karena memiliki konsep klasifikasi [8] namun mampu memberikan hasil prediksi yang cukup akurat pada permasalahan berbasis data historis.

[9] juga menegaskan bahwa tidak ada satu metode oversampling yang unggul pada seluruh tipe data, sehingga pemilihannya harus disesuaikan dengan karakteristik dataset. Berdasarkan celah penelitian tersebut, pengembangan model analisis sentimen yang menggabungkan algoritma Naïve Bayes dengan teknik oversampling modern seperti SMOTE pada konteks ulasan e-commerce masih diperlukan. Pendekatan ini berpotensi menghasilkan model yang lebih akurat, efisien, dan mudah diimplementasikan dalam pemrosesan ulasan pelanggan, khususnya pada platform seperti Alfagift [10].

Penelitian ini bertujuan Menawarkan solusi berupa perancangan dan implementasi sistem informasi prediksi kelulusan mahasiswa menggunakan metode *K-Nearest Neighbor* sebagai upaya untuk membantu pengolahan dan pemantauan data kelulusan mahasiswa secara terkomputerisasi.

[6], Mencapai hasil berupa sistem informasi yang mampu mengolah data akademik mahasiswa dan menghasilkan informasi prediksi kelulusan secara cepat, tepat, dan mudah dipahami oleh pihak akademik. [8], Mengaitkan penerapan metode *K-Nearest Neighbor* dalam sistem informasi dengan permasalahan kelulusan mahasiswa, sehingga sistem yang dibangun sesuai dengan judul tugas akhir dan latar belakang masalah yang telah diuraikan. [9] Secara teoritis, penelitian ini diharapkan dapat menambah wawasan dan pengetahuan dalam bidang sistem informasi. [11]. Selain itu, penelitian ini dapat mendukung

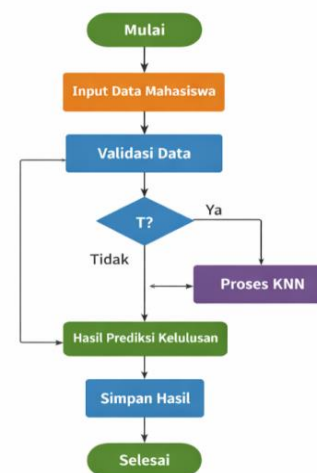
pengembangan teori terkait pemanfaatan data akademik dalam pengambilan keputusan.

## MATERIALS AND METHODS

Tahapan perancangan sistem dilakukan secara bertahap dan terstruktur untuk memastikan sistem informasi prediksi kelulusan mahasiswa dapat dibangun sesuai dengan tujuan penelitian. Tahapan perancangan sistem ini mencakup seluruh proses penelitian, mulai dari pengumpulan data hingga pengujian sistem.

tahap pertama adalah analisis kebutuhan sistem, yaitu mengidentifikasi kebutuhan data dan fungsi sistem yang diperlukan untuk melakukan prediksi kelulusan mahasiswa. Tahap kedua adalah pengumpulan dan persiapan data, di mana data dari Kaggle diperiksa dan disesuaikan agar siap digunakan dalam proses prediksi.

[12], [13].



Gambar 1 tahapan perancangan

Tahap ketiga adalah perancangan sistem, yang meliputi perancangan alur proses sistem, perancangan basis data, dan perancangan antarmuka pengguna. Tahap keempat adalah implementasi metode *K-Nearest Neighbor*, yaitu penerapan metode KNN dalam sistem untuk menghasilkan prediksi kelulusan mahasiswa.[9], [14]. Tahap terakhir adalah pengujian sistem, yang dilakukan untuk memastikan sistem dapat berjalan dengan baik dan menghasilkan output prediksi sesuai dengan tujuan penelitian.

Data yang digunakan termasuk dalam kategori data sekunder, karena data telah tersedia dan dipublikasikan oleh pihak lain. Dataset dari Kaggle digunakan sebagai data simulasi untuk

mendukung perancangan dan implementasi sistem informasi prediksi kelulusan mahasiswa menggunakan metode K-Nearest Neighbor. [15], [16].

Penggunaan dataset publik dipilih karena data mudah diakses, tidak melanggar etika penelitian, serta sesuai dengan kebutuhan penelitian.

[17]. Data yang digunakan dalam penelitian ini adalah dataset mahasiswa yang mencakup atribut yang berkaitan dengan prestasi akademik mahasiswa, seperti nilai mata kuliah, Indeks Prestasi Kumulatif Semester (IPK), Indeks Prestasi Kumulatif (IPK), dan status kelulusan mahasiswa.

[18], [19]. Dataset ini digunakan sebagai data pelatihan dan pengujian ketika menerapkan metode K - Nearest Neighbor ( KNN ) untuk memprediksi kelulusan mahasiswa

Penggunaan data Kaggle dalam penelitian ini berfungsi sebagai data simulasi untuk mendukung desain dan pengujian sistem informasi prediksi kelulusan mahasiswa. Data yang digunakan tidak mencakup identitas pribadi siswa, sehingga aman untuk tujuan akademis dan penelitian tanpa melanggar etika penggunaan data.

## RESULTS AND DISCUSSION

Hasil perancangan sistem yang telah disusun menunjukkan bahwa alur sistem prediksi kelulusan mahasiswa bersifat sederhana dan mudah dipahami, sesuai dengan karakteristik Tugas Akhir Diploma 3. Perancangan ini telah mencakup tahapan utama yang diperlukan dalam pengembangan sistem prediksi, mulai dari pengelolaan data hingga proses klasifikasi.

### Hasil Diagram Proses Klasifikasi Menggunakan Metode KNN

Diagram proses KNN menggambarkan tahapan konseptual dari metode K-Nearest Neighbor yang digunakan dalam penelitian ini. Diagram ini berfokus pada proses klasifikasi yang dilakukan oleh metode KNN terhadap data akademik mahasiswa. Proses klasifikasi dimulai dari penggunaan data latih yang telah tersedia. Selanjutnya ditentukan nilai K sebagai jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Setelah nilai K ditentukan, dilakukan perhitungan jarak antar data menggunakan metode Euclidean Distance untuk menentukan tetangga terdekat. Kelas mayoritas dari tetangga

terdekat tersebut digunakan sebagai dasar dalam menentukan hasil prediksi kelulusan mahasiswa, yaitu lulus atau tidak lulus.



**Gambar 2 Diagram Proses Klasifikasi Metode K-Nearest Neighbor**

### Hasil Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data sekunder berupa dataset simulatif yang diperoleh dari platform Kaggle. Dataset ini digunakan untuk merepresentasikan karakteristik umum aktivitas dan capaian belajar mahasiswa dalam konteks akademik, serta berfungsi sebagai media pembelajaran dan pengujian penerapan metode K-Nearest Neighbor, bukan sebagai data akademik resmi dari suatu perguruan tinggi tertentu. metode K-Nearest Neighbor. Selanjutnya, dataset digunakan sebagai data latih dan data uji dalam proses pengujian sistem prediksi kelulusan mahasiswa.

**Tabel 1 Data**

No	Nama Atribut	Deskripsi	Tipe Data
1	hours_studied	Jumlah jam belajar mahasiswa	Numerik
2	sleep_hours	Jumlah jam tidur mahasiswa per hari	Numerik
3	attendance_percent	Persentase kehadiran mahasiswa	Numerik
4	previous_scores	Nilai akademik sebelumnya	Numerik
5	exam_score	Nilai ujian akhir	Numerik

### Evaluasi Informasi Struktur Dataset dan Tipe Data

tahap selanjutnya adalah melakukan pemeriksaan struktur dataset dan tipe data dari masing-masing atribut. Pemeriksaan ini bertujuan untuk memastikan bahwa data tidak memiliki kesalahan format dan seluruh atribut berada dalam bentuk numerik sehingga dapat diproses oleh algoritma KNN.

Informasi struktur dataset ditunjukkan, yang memperlihatkan jumlah data, jumlah atribut, serta tipe data dari setiap kolom. Berdasarkan hasil pemeriksaan tersebut, dapat disimpulkan bahwa dataset telah memenuhi syarat untuk digunakan dalam proses klasifikasi tanpa memerlukan konversi tipe data yang kompleks.

pra-pemrosesan data dilakukan untuk menyiapkan data sebelum diterapkan ke dalam algoritma KNN. Pada tahap ini dilakukan penentuan label kelas kelulusan mahasiswa. Nilai *exam\_score* digunakan sebagai acuan untuk menentukan status kelulusan mahasiswa.

Mahasiswa dojjjytkategorikan lulus apabila nilai ujian akhir memenuhi atau melebihi batas kelulusan yang telah ditentukan, dan tidak lulus apabila berada di bawah batas tersebut. Proses ini bertujuan untuk mengubah data numerik menjadi label kelas yang dapat digunakan dalam proses klasifikasi.

### CONCLUSION

Berdasarkan hasil perancangan, penerapan, dan pengujian yang telah dibahas pada bab-bab sebelumnya, dapat ditarik beberapa kesimpulan sebagai berikut. Penelitian ini berhasil menjawab permasalahan yang dirumuskan pada latar belakang, yaitu kebutuhan akan sistem informasi yang mampu membantu pihak akademik dalam memantau dan memprediksi kelulusan mahasiswa secara lebih efektif dan akurat.

1). Sistem informasi prediksi kelulusan mahasiswa berbasis web berhasil dirancang sesuai dengan kebutuhan pengguna dan alur sistem yang telah ditetapkan. Perancangan sistem mencakup pengelolaan data akademik, proses pra-pemrosesan data, serta penyajian hasil prediksi kelulusan mahasiswa. Perancangan ini menjadi dasar yang kuat dalam proses implementasi dan pengujian sistem.

2). Metode *K-Nearest Neighbor* (KNN) berhasil diterapkan untuk melakukan prediksi kelulusan

mahasiswa berdasarkan data akademik. Proses klasifikasi dilakukan melalui tahapan penentuan nilai  $K$ , perhitungan jarak menggunakan *Euclidean Distance*, pemilihan tetangga terdekat, dan penentuan kelas mayoritas

### REFERENCE

- [1] Han, J., Kamber, M., & Pei, J. (2018). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- [2] Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics*. John Wiley & Sons.
- [3] Laudon, K. C., & Laudon, J. P. (2020). *Management Information Systems*. Pearson.
- [4] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- [5] Putra, A., Rahmawati, D., & Hidayat, R. (2021). Penerapan Metode K-Nearest Neighbor untuk Prediksi Kelulusan Mahasiswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(3), 455-462.
- [6] Putra, R. A., & Nugroho, Y. (2019). Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 6(2), 123-130.
- [7] Sari, N., & Wijaya, A. (2022). Sistem Informasi Prediksi Kelulusan Mahasiswa Berbasis Data Mining. *Jurnal Sistem Informasi*, 18(2), 120-129.
- [8] Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Informatika.
- [9] Turban, E., Pollard, C., & Wood, G. (2019). *Information Technology for Management*. Wiley.