

Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (KNN) dengan Klasifikasi Biner

Irfan Nur Sofiyanto^{1*}, Andika Bagus Saputra², Ridlo Fauzi Rakhmadianto³, Hermawan Nur Eka Febrianto⁴, Aprillio Malika Endiwan⁵, Aprilisa Arum Sari⁶

¹Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

1*220103179@mhs.udb.ac.id

²Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

220103194@mhs.udb.ac.id

³Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

3220103186@mhs.udb.ac.id

⁴Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

4220103197@mhs.udb.ac.id

⁵Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

5220103195@mhs.udb.ac.id

⁶ Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

6april_arumsari@udb.ac.id

Abstrak— Penelitian ini bertujuan mengembangkan model prediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN). Permasalahan yang diangkat adalah ketidakpastian dalam meramalkan kelulusan berdasarkan data historis akademik, khususnya Indeks Prestasi Semester (IPS). Penelitian berfokus pada pembangunan model klasifikasi biner yang dapat mengidentifikasi status kelulusan mahasiswa guna mendukung deteksi dini risiko akademik. Dataset yang digunakan terdiri dari 519 data mahasiswa Program Studi Teknologi Informasi, dengan atribut berupa nilai IPS semester 1 hingga 8, serta label target berupa status kelulusan. Karena terdapat ketidakseimbangan antar kelas, diterapkan Random Over Sampling (ROS) untuk menyeimbangkan distribusi sebelum pelatihan model. Data kemudian dinormalisasi dan dibagi menjadi data latih dan data uji. Algoritma KNN digunakan dengan parameter $k = 5$, yang mengklasifikasikan label berdasarkan mayoritas tetangga terdekat. Hasil pengujian menunjukkan bahwa model KNN mampu memberikan akurasi prediksi sebesar 90%, menandakan bahwa pendekatan berbasis data mining ini efektif untuk mendukung evaluasi dan pengelolaan akademik di lingkungan pendidikan tinggi.

Kata kunci— Prediksi Kelulusan, K-Nearest Neighbor, IPS, Data Mining, Oversampling, Pendidikan Tinggi.

Abstract— This study aims to develop a student graduation prediction model using the K-Nearest Neighbor (KNN) algorithm. The main problem addressed is the uncertainty in predicting graduation outcomes based on historical academic performance, particularly the Semester Grade Point Average (GPA). The research focuses on building a binary classification model to identify whether a student will graduate or not, serving as an early warning system for academic risk. The dataset consists of 519 student records from the Information Technology Study Program, with attributes including GPA values from semesters 1 to 8 and a target label representing graduation status. To address class imbalance, the Random Over Sampling (ROS) technique was applied. The data was normalized and split into training and testing sets. The KNN algorithm was implemented with $k = 5$, determining the predicted class based on the majority of the five nearest neighbors. Evaluation results show that the KNN model achieved a prediction accuracy of 90%, indicating that this data mining approach is effective as a decision support tool for improving academic management and the quality of learning.

Keywords— Graduation Prediction, K-Nearest Neighbor, Semester GPA, Data Mining, Oversampling, Higher Education.

I. PENDAHULUAN

Mahasiswa merupakan salah satu aset penting dalam institusi pendidikan tinggi yang perlu dikelola secara efektif dan sistematis. Tingkat kelulusan mahasiswa, terutama dalam mata kuliah yang diambil setiap semester, menjadi indikator utama dalam mengukur keberhasilan proses pembelajaran serta kualitas

penyelenggaraan program studi. Ketidakstabilan dalam capaian kelulusan dapat berdampak pada lamanya masa studi, rendahnya indeks prestasi kumulatif (IPK), hingga penurunan nilai akreditasi institusi [1].

Untuk itu, pemantauan dan evaluasi dini terhadap potensi kelulusan mahasiswa menjadi sangat penting. Salah satu pendekatan yang dapat dilakukan adalah dengan memanfaatkan data

akademik historis mahasiswa untuk memprediksi kemungkinan kelulusan [2]. Dalam konteks ini, nilai akhir mata kuliah merepresentasikan tingkat penguasaan materi yang diajarkan, yang diperoleh dari berbagai komponen penilaian seperti kuis, tugas individu, tugas kelompok, ujian tengah semester (UTS), praktik, dan ujian akhir semester (UAS) [3].

Penelitian ini memanfaatkan data indeks prestasi semester (IPS) mahasiswa dari semester 1 hingga semester 8, yang telah dirangkum dari seluruh mata kuliah yang ditempuh. Data tersebut diolah untuk membentuk pola yang dapat digunakan dalam memprediksi status kelulusan mahasiswa, baik dalam satu mata kuliah tertentu maupun sebagai prediktor terhadap kelulusan program studi secara keseluruhan [4].

Dengan banyaknya data yang tersedia dan kompleksitas hubungan antar variabel, pendekatan data mining menjadi solusi strategis untuk menggali pola tersembunyi dalam data akademik. Salah satu algoritma populer yang digunakan untuk klasifikasi adalah K-Nearest Neighbor (KNN). Algoritma ini bekerja berdasarkan kedekatan jarak antar data, bersifat non-parametrik, sederhana, mudah diimplementasikan, dan efektif untuk menangani data numerik seperti nilai akademik [5].

Penelitian ini menggunakan algoritma KNN untuk memprediksi status LULUS atau TIDAK LULUS mahasiswa dalam mata kuliah *Sistem Manajemen Basis Data* (SMBD). Pemilihan KNN didasarkan pada kemudahan interpretasi hasil, kesederhanaan metode, dan kemampuannya dalam dijadikan dasar eksplorasi model prediktif lanjutan. Selain itu, model KNN diimplementasikan dalam bentuk aplikasi web berbasis bahasa pemrograman PHP, yang dapat diakses oleh dosen atau admin untuk melakukan prediksi kelulusan berdasarkan data input mahasiswa [6].

Sistem yang dibangun memiliki antarmuka input sederhana, memungkinkan pengguna memasukkan nilai IPS mahasiswa, yang kemudian diklasifikasikan menggunakan model KNN [7]. Dataset yang digunakan dalam penelitian ini terdiri dari 519 data mahasiswa Program Studi Teknologi Informasi,

dengan fitur berupa nilai IPS semester 1 hingga 8. Untuk mengatasi masalah ketidakseimbangan kelas (label), digunakan teknik Random Over Sampling. Model KNN diuji menggunakan parameter $K = 5$, dan hasil evaluasi menunjukkan akurasi mencapai 90%.

Dengan hasil tersebut, model prediksi yang dikembangkan dapat dijadikan sebagai alat bantu dalam pengambilan keputusan akademik, terutama dalam mendeteksi dini risiko ketidakkelulusan mahasiswa. Penelitian ini memberikan kontribusi dalam pengembangan sistem informasi berbasis data mining yang dapat diterapkan secara nyata dalam dunia pendidikan tinggi.

II. METODOLOGI PENELITIAN

A. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian terapan (applied research) dengan pendekatan kuantitatif yang bertujuan untuk membangun dan mengimplementasikan sistem prediksi kelulusan mahasiswa pada mata kuliah *Sistem Manajemen Basis Data* menggunakan algoritma K-Nearest Neighbor (KNN) [8]. Sistem dikembangkan dalam bentuk aplikasi web menggunakan bahasa pemrograman PHP dan basis data MySQL [9].

B. Model Pengembangan Sistem

Model pengembangan perangkat lunak yang digunakan dalam penelitian ini adalah model Waterfall. Waterfall merupakan pendekatan sistematis dan berurutan, yang terdiri dari beberapa tahapan utama seperti analisis kebutuhan, perancangan sistem, implementasi, pengujian, dan pemeliharaan [10]. Model ini dipilih karena kebutuhan sistem telah terdefinisi dengan jelas sejak awal, serta perubahan atau iterasi sistem bersifat minimal.

Tahapan model Waterfall dalam penelitian ini meliputi:

1. Requirement Analysis (Analisis Kebutuhan) Mengidentifikasi kebutuhan fungsional dan non-fungsional sistem, termasuk input data mahasiswa (nilai IPS semester 1–8), proses klasifikasi dengan KNN, dan output berupa status kelulusan (LULUS / TIDAK LULUS).

2. **System Design (Perancangan Sistem)** Melakukan perancangan basis data, struktur tabel, algoritma klasifikasi, serta perancangan antarmuka pengguna berbasis web (form input nilai IPS dan hasil prediksi).

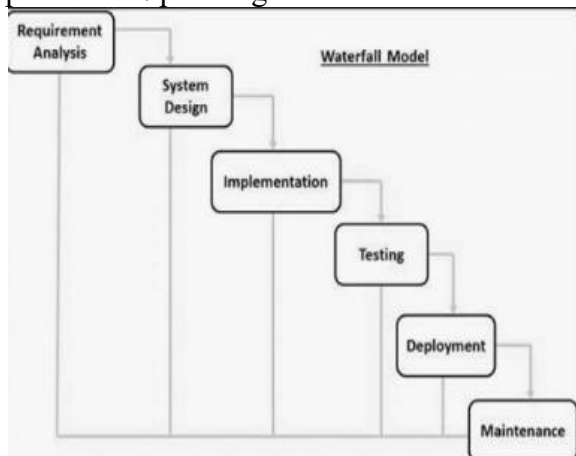
3. **Implementation (Implementasi)** Implementasi dilakukan menggunakan:

1. Bahasa pemrograman: PHP
2. Database: MySQL
3. Teknologi pendukung: HTML dan CSS

Fungsi utama sistem adalah melakukan klasifikasi status kelulusan mahasiswa berdasarkan nilai-nilai IPS.

4. **Testing (Pengujian Sistem)** Sistem diuji secara fungsional untuk memastikan proses input, perhitungan algoritma KNN, serta output berjalan dengan benar. Pengujian akurasi juga dilakukan untuk mengevaluasi performa model klasifikasi.

5. **Maintenance (Pemeliharaan)** Tahapan ini mencakup perawatan sistem, penambahan data baru, dan kemungkinan perubahan parameter seperti nilai k pada algoritma KNN.



Gambar 1. Model Waterfall untuk Prediksi Kelulusan Mahasiswa

C. Dataset dan Pra-Pemrosesan Data

Dataset yang digunakan terdiri dari 519 data mahasiswa Program Studi Teknologi Informasi.

Atribut yang digunakan meliputi:

1. Fitur Input : Nilai IPS Semester 1 hingga 8
2. Label Output : Status kelulusan (dikonversi dari A, B, C, D, S menjadi LULUS / TIDAK LULUS)

Masalah ketidakseimbangan kelas diatasi menggunakan teknik Random Over Sampling (ROS). Selanjutnya, dilakukan normalisasi

data menggunakan metode Standard Scaler agar skala antar fitur seragam sehingga perhitungan jarak tidak bias.

D. Metode Klasifikasi

Penelitian ini menggunakan pendekatan klasifikasi biner, yaitu memetakan data mahasiswa ke dalam dua kelas:

1. Lulus
2. Tidak Lulus

Penentuan label dilakukan berdasarkan IPK akhir atau nilai akhir pada mata kuliah Sistem Manajemen Basis Data.

E. Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor merupakan algoritma klasifikasi non-parametrik yang bekerja berdasarkan prinsip kedekatan jarak antar data. Proses klasifikasi dalam sistem meliputi:

1. Menghitung jarak Euclidean antara data uji dan semua data latih.
2. Mengambil k tetangga terdekat (dalam penelitian ini, digunakan $k = 5$).
3. Menentukan label hasil klasifikasi berdasarkan mayoritas kelas dari tetangga tersebut.

Rumus perhitungan jarak Euclidean adalah sebagai berikut:

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

Keterangan::

1. x_i : Nilai IPS pada data uji
2. y_i : Nilai IPS pada data latih
3. n : Jumlah fitur (8 semester)

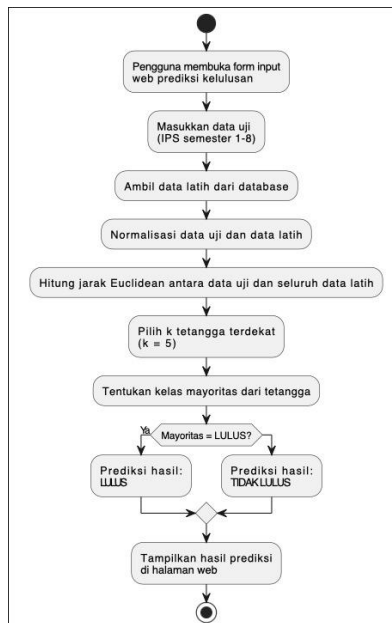
F. Alur Sistem Prediksi

Alur kerja sistem prediksi kelulusan mahasiswa adalah sebagai berikut:

1. Pengguna (admin/dosen) memasukkan data IPS semester 1–8 melalui form input berbasis web
2. Sistem mengambil data latih dari basis data
3. Sistem menghitung jarak Euclidean dari data uji ke seluruh data
4. Sistem memilih 5 tetangga terdekat

5. Sistem menentukan kelas mayoritas dari tetangga tersebut

6. Sistem menampilkan hasil prediksi: LULUS atau TIDAK LULUS



Gambar 2. Alur Sistem Prediksi

III. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Model KNN

Penelitian ini mengimplementasikan algoritma K-Nearest Neighbor (KNN) dengan nilai $k = 5$ untuk memprediksi status kelulusan mahasiswa berdasarkan data nilai Indeks Prestasi Semester (IPS) dari semester 1 hingga semester 8. Dataset yang digunakan terdiri dari 519 data mahasiswa Program Studi Teknologi Informasi.

Data tersebut dibagi menjadi dua bagian:

1. Data Latih sebanyak 415 data (80%)
2. Data Uji sebanyak 104 data (20%)

Pembagian dilakukan secara stratifikasi untuk menjaga proporsi antara kelas LULUS dan TIDAK LULUS tetap seimbang pada kedua subset data.

B. Evaluasi Kinerja Model

Model KNN diuji menggunakan data uji untuk mengevaluasi performa klasifikasi. Hasil pengujian menunjukkan bahwa model mampu memprediksi kelulusan mahasiswa dengan tingkat akurasi tinggi.

Tabel 1. Evaluasi Kinerja Model

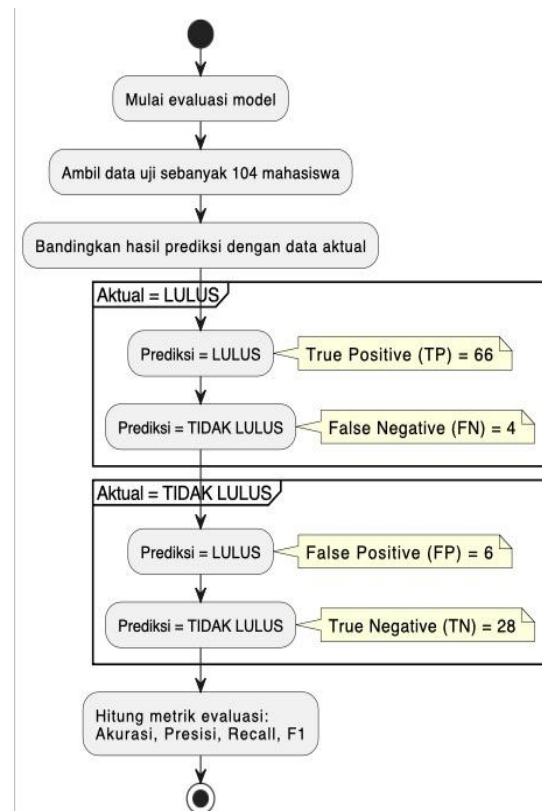
Metrik Evaluasi	Nilai
Akurasi	90,38%
Presisi (Precision)	91,67%
Recall (Sensitivity)	94,29%
F1-Score	92,96%
Jumlah prediksi benar	94 dari 104
Jumlah prediksi salah	10 dari 104

C. Confusion Matrix

Confusion matrix digunakan untuk mengidentifikasi jumlah prediksi yang benar dan salah pada masing-masing kelas. Berikut distribusi hasil prediksi terhadap kelas LULUS dan TIDAK LULUS:

Tabel 2. Confusion Matrix

	Prediksi LULUS	Prediksi TIDAK LULUS
Aktual LULUS	66 (TP)	4(FN)
Aktual TIDAK LULUS	6 (FP)	28(TN)



Gambar 3. Evaluasi Kinerja Model

D. Analisis dan Interpretasi

1. Akurasi Mode Tinggi

Model menunjukkan tingkat akurasi yang tinggi, yaitu 90,38%, yang mengindikasikan bahwa algoritma KNN cukup efektif dalam memprediksi kelulusan mahasiswa hanya berdasarkan nilai IPS semester 1 hingga 8.

2. Efektivitas Algoritma KNN

Algoritma KNN dipilih karena kemampuannya dalam menangani klasifikasi berbasis data numerik sederhana. Keunggulannya meliputi:

- a. Tidak memerlukan proses pelatihan (*lazy learning*)
 - b. Sederhana dalam perhitungan dan implementasi
 - c. Memberikan hasil klasifikasi yang dapat dijelaskan secara logis
- ##### 3. Kualitas Data dan Pra-Pemrosesan
- a. Normalisasi Min-Max dilakukan agar skala nilai IPS seragam dan tidak mendistorsi perhitungan jarak.
 - b. Random Over Sampling (ROS) digunakan untuk menyeimbangkan jumlah kelas antara LULUS dan TIDAK LULUS, sehingga model tidak bias terhadap kelas mayoritas.
 - c. Pembagian data dengan stratified sampling menjaga proporsi kelas saat evaluasi.
- ##### 4. Pola Nilai IPS

Data menunjukkan bahwa mahasiswa dengan tren IPS yang menurun atau konsisten rendah dari semester awal cenderung masuk dalam kelas *TIDAK LULUS*, sedangkan mahasiswa dengan nilai yang stabil atau meningkat lebih banyak diprediksi *LULUS*.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membangun sistem prediksi kelulusan mahasiswa pada mata kuliah tertentu menggunakan algoritma K-Nearest Neighbor (KNN) dengan pendekatan klasifikasi biner. Dataset yang digunakan berasal dari Program Studi Teknologi Informasi dengan total 519 data mahasiswa, di mana setiap entri berisi nilai Indeks Prestasi Semester (IPS) dari semester 1 hingga 8. Proses pengembangan sistem dilakukan melalui pendekatan model Waterfall, sedangkan tahapan pemrosesan data

mencakup normalisasi, penyeimbangan kelas menggunakan Random Over Sampling, serta pembagian data secara stratifikasi.

Berdasarkan hasil evaluasi, model KNN dengan parameter $k = 5$ menunjukkan performa yang sangat baik, dengan akurasi sebesar 90,38%, precision sebesar 91,67%, recall sebesar 94,29%, dan F1-score sebesar 92,96%. Hal ini mengindikasikan bahwa algoritma KNN mampu memprediksi status kelulusan mahasiswa dengan tingkat ketepatan yang tinggi hanya berdasarkan nilai IPS. Selain itu, implementasi sistem berbasis web dengan bahasa pemrograman PHP dan MySQL memungkinkan pengguna (dosen/operator akademik) untuk melakukan evaluasi kelulusan secara mudah, cepat, dan efisien.

Dengan demikian, penelitian ini memberikan kontribusi nyata dalam pemanfaatan teknik data mining untuk mendukung proses pengambilan keputusan di bidang akademik. Sistem prediksi yang dibangun dapat dijadikan sebagai alat bantu evaluatif untuk mendeteksi dini potensi ketidakhadiran mahasiswa dan mendorong intervensi lebih awal oleh pihak institusi pendidikan.

REFERENSI

- [1] D. P. Sari, S. S. Hilabi, and A. Hananto, "Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama," *SMARTICS J.*, vol. 9, no. 1, pp. 14–19, 2023, doi: 10.21067/smartics.v9i1.8088.
- [2] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [3] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) Algorithm with Euclidean and Manhattan in Classification of Student Graduation," *J. Eng. Appl. Technol.*, vol. 2, no. 2, pp. 86–91, 2021, doi: 10.21831/jeatech.v2i2.42777.
- [4] A. Sanjaya and T. Wahyana, "Penerapan Metode K-Nearest Neighbour Untuk Sistem Prediksi Kelulusan Siswa MTs Nurul Muslimin Berbasis Website," *J. Transform. Mandalika*, vol. 3, no. 1, pp. 31–47, 2022, [Online]. Available: <https://www.ojs.cahayamandalika.com/index.php/jtm/article/view/866%0Ahttps://www.ojs.cahayamandalika.com/index.php/jtm/article/download/866/863>
- [5] R. Situmorang *et al.*, "Model Algoritma K-Nearest Neighbor (K-Nn) Dan Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 250–254, 2023, [Online]. Available: <https://pdfs.semantic scholar.org/dfa/77cfd511957000cafb74eb4e24e7775f05c2.pdf>
- [6] N. Ahmad, S. Hafizh, and R. Sulthanah, "Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma K-Nearest Neighbor," *J. Manaj. Inform.*, vol. 14, no. 2, pp. 135–149, 2024, doi: 10.34010/jamika.v14i2.12454.
- [7] I. P. . Tafonao and A. F. K. Sibero, "Teknik Klasifikasi Prediksi Kelulusan Mahasiswa Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan K- Nearest Neighbors," *J. Mahajana Inf.*

- vol. 7, no. 1, pp. 83–90, 2022, doi: 10.51544/jurnalmi.v7i1.2956.
- [8] E. Novianto, A. Hermawan, and D. Avianto, “Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, 2023, doi: 10.36341/rabit.v8i2.3434.
- [9] M. Wati, W. H. Rahmah, N. Novirasari, Haviluddin, E. Budiman, and Islamiyah, “Analysis K-Means Clustering to Predicting Student Graduation,” *J. Phys. Conf. Ser.*, vol. 1844, no. 1, pp. 1–8, 2021, doi: 10.1088/1742-6596/1844/1/012028.
- [10] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.