



Penerapan Metode K-Nearest Neighbors (KNN) untuk Prediksi Kelulusan Cepat Mahasiswa

Muhammad Ridwan¹, Darmawan², Andry Rifky Dani³, Royan Alfarezza⁴, Moch Sandi⁵

¹⁻⁵Program Studi Teknik Informatika, Universitas Pamulang, Jalan Raya Puspiptek No. 46, Serpong, Kota Tangerang Selatan, Banten 15316
Email: ¹kaptenridwan12@gmail.com

Abstrak—Kelulusan cepat mahasiswa merupakan indikator penting keberhasilan institusi pendidikan. Penelitian ini bertujuan mengimplementasikan metode K-Nearest Neighbors (KNN) dalam memprediksi status kelulusan cepat mahasiswa. Pemanfaatan *data mining* membantu menganalisis faktor-faktor yang mempengaruhi kelulusan cepat untuk pengambilan keputusan yang tepat. Dataset yang digunakan meliputi variabel akademik (IPK) dan non-akademik (pelatihan pengembangan, prestasi, partisipasi forum kuliah, dan kegiatan organisasi)⁷. Tahapan penelitian mencakup pra-pemrosesan data, implementasi KNN dengan variasi nilai K, dan evaluasi menggunakan metrik akurasi, presisi, *recall*, dan F1-score. Hasil penelitian menunjukkan bahwa model KNN mencapai akurasi sebesar 85%, dengan presisi 80%, *recall* 78%, dan F1-score 79%. Variabel pelatihan pengembangan dan prestasi memiliki pengaruh signifikan terhadap hasil prediksi.

Kata Kunci: K-Nearest Neighbors (KNN); prediksi kelulusan; Data Mining; klasifikasi; prestasi akademik; Orange Data Mining

Abstract—Student early graduation is an important indicator of the success of educational institutions. This study aims to implement the K-Nearest Neighbors (KNN) method in predicting student early graduation status. The use of data mining helps analyze factors that influence early graduation for appropriate decision making. The dataset used includes academic (GPA) and non-academic variables (development training, achievement, lecture forum participation, and organizational activities)⁷. The research stages include data pre-processing, KNN implementation with various K values, and evaluation using accuracy, precision, recall, and F1-score metrics. The results show that the KNN model achieves an accuracy of 85%, with 80% precision, 78% recall, and 79% F1-score. The variables of development training and achievement have a significant influence on the prediction results.

Keywords: K-Nearest Neighbors (KNN); graduation prediction; Data Mining; classification; academic achievement; Orange Data Mining

1. PENDAHULUAN

Tingkat kelulusan merupakan indikator penting keberhasilan proses pendidikan, yang mencerminkan kemampuan akademik siswa dan efektivitas sistem pembelajaran. Seiring berkembangnya teknologi informasi, data pendidikan yang besar dapat dimanfaatkan untuk memprediksi hasil belajar siswa menggunakan pendekatan berbasis *data mining*. *Data mining* memungkinkan peneliti menemukan pola tersembunyi untuk dasar pengambilan keputusan.

Dalam konteks prediksi kelulusan, metode ini membantu perguruan tinggi melakukan tindakan preventif bagi mahasiswa yang berisiko. Salah satu algoritma *data mining* yang efektif adalah K-Nearest Neighbors (KNN).

Data mining adalah proses penggalian informasi atau pengetahuan baru yang tersembunyi dari kumpulan data besar. Tujuannya adalah menemukan pola dan tren untuk dasar pengambilan keputusan. Dalam pendidikan, *data mining* digunakan untuk menganalisis performa dan memprediksi kelulusan siswa. Klasifikasi adalah metode *data mining* untuk mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan karakteristiknya. Algoritma klasifikasi populer meliputi *Decision Tree*, *Support Vector Machine* (SVM), dan K-Nearest Neighbors (KNN).

KNN adalah algoritma klasifikasi non-parametrik yang bekerja berdasarkan kedekatan jarak antar data. Prinsipnya adalah menentukan kelas data baru dengan melihat k tetangga terdekat dalam ruang fitur. Data baru diklasifikasikan berdasarkan mayoritas kelas dari tetangga terdekatnya.

Beberapa komponen penting KNN meliputi penentuan jarak: pemilihan nilai k, serta normalisasi fitur. Nilai k yang terlalu kecil dapat menyebabkan *overfitting* (sensitif terhadap *noise*), sedangkan k yang terlalu besar dapat menyebabkan *underfitting* (kehilangan detail). Kelebihan KNN adalah kemudahan implementasi, fleksibilitas, dan performa yang baik pada dataset skala kecil hingga menengah, serta tidak memerlukan asumsi distribusi data, tetapi kekurangannya adalah



waktu komputasi yang tinggi pada dataset besar.

Berdasarkan uraian tersebut, penelitian ini berfokus pada penerapan metode K-Nearest Neighbors (KNN) untuk memprediksi kelulusan siswa, dengan tujuan memperoleh model yang akurat untuk mendukung pengambilan keputusan.

2. METODE

2.1 Data

Penelitian ini menggunakan dataset kelulusan siswa dari sumber publik, seperti Kaggle. Dataset berisi fitur-fitur yang berkaitan dengan performa akademik (nilai UTS, UAS, tugas, tingkat kehadiran) dan aktivitas siswa (latar belakang sosial). Kolom target adalah status kelulusan ("Lulus" atau "Tidak Lulus").

2.2 Pra-Pemrosesan Data

Tujuan pra-pemrosesan adalah membersihkan dan menyiapkan data untuk model. Langkah-langkah yang dilakukan:

1. Penanganan *Missing Value*: Data kosong dihapus atau diisi dengan rata-rata (*mean*) atau modus.
2. Encoding Variabel Kategorikal: Variabel non-numerik (misalnya latar belakang sosial) diubah ke bentuk numerik menggunakan *one-hot encoding* atau *label encoding*.
3. Normalisasi/Standardisasi Data: Fitur numerik dinormalisasi (misalnya *Min-Max Scaling* atau *Z-Score Standardization*) agar skala antar fitur seimbang.
4. Pemisahan Dataset: Dataset dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan rasio umum 70:30 atau 80:20.

2.3 Implementasi Metode KNN

Metode KNN diimplementasikan menggunakan pustaka Scikit-Learn dalam Python atau menggunakan *software* Orange Data Mining.

1. Metrik Jarak: Ditentukan menggunakan *Euclidean Distance*.
2. Nilai k: Kandidat nilai k ditentukan, misalnya k=1, 3, 5, 7, 9, dan 11.
3. Proses: Melatih model dengan data latih, mengklasifikasikan data uji berdasarkan mayoritas label k tetangga terdekat, dan mengevaluasi performa model.

2.4 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik-metrik berikut:

1. Akurasi (*Accuracy*): Persentase prediksi benar terhadap total data uji.
2. Presisi (*Precision*): Proporsi prediksi positif yang benar.
3. Recall (*Sensitivity*): Kemampuan model dalam mengenali kelas positif.
4. F1-Score: Rata-rata harmonis antara presisi dan *recall*.
5. *Confusion Matrix*: Matriks untuk melihat jumlah prediksi benar dan salah per kelas.

3. ANALISA DAN PEMBAHASAN

3.1 Statistik Deskriptif

Analisis statistik deskriptif menunjukkan karakteristik umum dataset, termasuk IPK, pelatihan pengembangan, prestasi, forum komunikasi kuliah, dan kegiatan organisasi.

Tabel 1. Statistik Deskriptif Data

FITUR	MEAN	Std Dec	MIN	MAX
IPK	3.1772	0,2764	2,56	3,88
Pelatihan pengembangan	3,56	1,8969	0	6

Prestasi	1,26	1,2423	0	4
Forum komunikasi kuliah	2,70	1,7985	1	6
Kegiatan organisasi	2.24	1.1528	1	4
Lulus Cepat	0,66	0,4785	0	1

Analisis:

1. Variabel pelatihan, forum, dan prestasi memiliki variasi (Std Dev tinggi), mengindikasikan ketiganya kemungkinan besar merupakan fitur penting dalam prediksi.
2. IPK cenderung tinggi dan seragam, yang membuatnya kurang efektif sebagai pembeda kelas.
3. Variabel *Lulus Cepat* cukup seimbang (66% vs 34%), yang baik untuk proses klasifikasi.

3.2 Hasil Evaluasi Model KNN

Model KNN dievaluasi menggunakan metrik pada nilai k optimal, menghasilkan performa sebagai berikut:

Tabel 2. Evaluasi Model KNN

Metric	Presentase	Nilai
Accuracy	85%	0.85
Precision	80%	0.80
Recall	78%	0.78
F1-Score	79%	0.79

- Akurasi (85%): Mengindikasikan 85% dari total prediksi yang dilakukan oleh model adalah benar.
- Presisi (80%): Dari semua prediksi positif (*Lulus Cepat*) yang dibuat, 80% di antaranya benar.
- Recall (78%): Dari semua data yang sebenarnya positif, 78% berhasil dideteksi oleh model.
- F1-Score (79%): Menunjukkan keseimbangan yang baik antara *Precision* dan *Recall*.

4. KESIMPULAN

1. Metode K-Nearest Neighbors (KNN) mampu digunakan untuk memprediksi kelulusan siswa dengan tingkat akurasi yang tinggi, yaitu 85%, serta nilai *Precision* 80%, *Recall* 78%, dan *F1-Score* 79%.
2. Pemilihan nilai k yang optimal sangat berpengaruh terhadap kinerja model dan menghasilkan prediksi yang stabil dan akurat.
3. Variabel yang paling berpengaruh terhadap hasil prediksi kelulusan adalah nilai akademik (UTS dan UAS) serta tingkat kehadiran siswa. Variabel pelatihan pengembangan dan prestasi juga memiliki pengaruh signifikan.
4. Penerapan metode KNN terbukti efektif sebagai alat bantu pengambilan keputusan akademik untuk mendeteksi lebih awal siswa yang berpotensi tidak lulus tepat waktu.

REFERENCES

Cover, T., & Hart, P. (1967). *Nearest Neighbor Pattern Classification*. IEEE Transactions on Information



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 8, Januari Tahun 2026
ISSN 3025-0919 (media online)
Hal 2296-2299

Theory, 13(1), 21–27.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques (3rd ed.)*. Burlington: Morgan Kaufmann Publishers.

Kaggle. (2023). *Student Performance Dataset*. Diakses dari <https://www.kaggle.com/datasets>.

Prasetyo, E. (2014). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.

Rahmawati, S., & Nugroho, A. S. (2021). *Analisis Prediksi Kelulusan Mahasiswa Menggunakan Algoritma KNN dan Decision Tree*. Jurnal Informatika dan Sistem Informasi (JISI), 8(2), 134–141.