

Perbandingan Algoritma K-Nearest Neighbor dan Random Forest dalam Prediksi Kelulusan Mahasiswa Universitas Muhammadiyah Jakarta

Rizal Hadi Hidayatullah¹, Relly Sanjaya², Muhammad Jundi Alazami³, Muhamad Affifudin⁴, Riza Samsinar⁵

^{1,2,3,4,5}Teknik Elektro, Fakultas Teknik, Universitas Muhammadiyah Jakarta

Jl. Cempaka Putih Tengah 27, Cempaka Putih, Jakarta Pusat, 10510

Telp. 0214244016

E-mail: 20210410200002@student.umj.ac.id

ABSTRAK

Kelulusan mahasiswa tepat waktu menjadi salah satu indikator utama keberhasilan institusi pendidikan tinggi. Universitas Muhammadiyah Jakarta terus berupaya meningkatkan angka kelulusan tepat waktu melalui analisis dan pemantauan akademik. Meskipun demikian, berbagai tantangan seperti rendahnya Indeks Prestasi Kumulatif (IPK), jumlah SKS yang tidak mencukupi, tingkat kehadiran perkuliahan yang rendah, dan lamanya durasi penyelesaian tugas akhir menjadi penyebab utama keterlambatan kelulusan. Penelitian ini bertujuan membangun model prediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN). Data yang dianalisis meliputi IPK, jumlah SKS yang telah ditempuh, tingkat kehadiran, nilai mata kuliah inti, serta durasi penyelesaian tugas akhir. Algoritma KNN diterapkan dengan pengujian parameter K (3, 5, 7) dan metrik jarak Euclidean. Hasil penelitian menunjukkan bahwa model dengan K=5 memberikan tingkat akurasi tertinggi sebesar 86%. Faktor IPK, jumlah SKS yang telah ditempuh, dan durasi penyelesaian tugas akhir ditemukan sebagai variabel dominan yang memengaruhi kelulusan mahasiswa. Model prediksi ini memungkinkan Universitas Muhammadiyah Jakarta untuk mengidentifikasi mahasiswa yang berisiko terlambat lulus sehingga dapat merancang intervensi yang tepat. Pengembangan lebih lanjut dapat mencakup integrasi variabel non-akademik, seperti partisipasi organisasi, status pekerjaan mahasiswa, motivasi belajar, dan faktor sosial lainnya, untuk meningkatkan akurasi prediksi. Dengan demikian, model ini diharapkan menjadi alat penting dalam meningkatkan keberhasilan akademik, efisiensi pendidikan, dan reputasi institusi secara keseluruhan.

Kata Kunci: *Kelulusan mahasiswa, K-Nearest Neighbor (KNN), prediksi, Universitas Muhammadiyah Jakarta, IPK*

ABSTRACT

Timely student graduation is a key indicator of the success of higher education institutions. Universitas Muhammadiyah Jakarta strives to enhance on-time graduation rates through academic monitoring and analysis. However, challenges such as low Grade Point Average (GPA), insufficient completed credit hours, poor attendance, and extended final project duration significantly hinder timely graduation. This study aims to develop a predictive model for student graduation using the K-Nearest Neighbor (KNN) algorithm. The analyzed data include GPA, completed credit hours, class attendance, core course grades, and final project duration. The KNN algorithm was implemented with parameter testing for K values (3, 5, 7) and

the Euclidean distance metric. Results reveal that the model with $K=5$ achieved the highest accuracy of 86%. GPA, completed credit hours, and final project duration were identified as dominant factors influencing graduation outcomes. This prediction model enables Universitas Muhammadiyah Jakarta to identify at-risk students and design effective interventions. Future research may include non-academic variables, such as organizational involvement, employment status, learning motivation, and social factors, to improve predictive accuracy. Hence, this model serves as a valuable tool in enhancing institutional academic success, optimizing educational management, and strengthening the institution's overall reputation.

Keywords: *Student graduation, K-Nearest Neighbor (KNN), prediction, Universitas Muhammadiyah Jakarta, GPA*

1. Pendahuluan

Kelulusan tepat waktu menjadi salah satu indikator utama keberhasilan pendidikan tinggi dalam mencetak lulusan yang berkualitas. Universitas Muhammadiyah Jakarta (UMJ), seperti halnya perguruan tinggi lainnya, menghadapi tantangan dalam memastikan kelulusan mahasiswa tepat waktu. Permasalahan ini tidak hanya berdampak pada individu mahasiswa, seperti meningkatnya beban finansial dan tekanan psikologis, tetapi juga memengaruhi reputasi institusi secara keseluruhan.

Faktor-faktor yang menyebabkan keterlambatan kelulusan dapat dibagi menjadi dua kategori utama, yaitu faktor akademik dan non-akademik. Dari sisi akademik, rendahnya Indeks Prestasi Kumulatif (IPK), jumlah SKS yang tidak mencukupi, tingkat kehadiran yang rendah, dan lamanya durasi penyelesaian tugas akhir menjadi faktor utama. Di sisi lain, faktor non-akademik seperti keikutsertaan dalam organisasi mahasiswa, status pekerjaan, dan kondisi sosial ekonomi juga turut memengaruhi.

Dalam era digital, analisis data akademik dapat menjadi solusi untuk mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu. Salah satu metode yang dapat digunakan adalah algoritma K-Nearest Neighbor (KNN). Algoritma ini bekerja dengan membandingkan jarak antara data mahasiswa baru dan data mahasiswa yang sudah ada untuk memprediksi status kelulusan. KNN merupakan metode yang sederhana namun efektif dalam menghasilkan prediksi yang akurat.

Penelitian ini bertujuan untuk membangun model prediksi kelulusan mahasiswa di UMJ menggunakan algoritma KNN. Dengan model ini, pihak universitas diharapkan dapat mengidentifikasi mahasiswa yang membutuhkan intervensi akademik lebih awal. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi variabel-variabel dominan yang memengaruhi kelulusan, sehingga dapat digunakan untuk pengambilan keputusan strategis. Dengan demikian, model prediksi ini diharapkan dapat meningkatkan efektivitas pengelolaan pendidikan di UMJ.

2. Metodologi

2.1 Pengumpulan Data

Penelitian ini dilakukan untuk membangun model prediksi kelulusan mahasiswa Universitas Muhammadiyah Jakarta menggunakan algoritma K-Nearest Neighbor (KNN). Data penelitian dikumpulkan dari catatan akademik mahasiswa Universitas Muhammadiyah Jakarta selama lima tahun terakhir. Variabel yang dianalisis meliputi Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, jumlah organisasi yang diikuti, dan durasi penyelesaian tugas akhir. Setelah data terkumpul, tahap selanjutnya adalah pra-pemrosesan data untuk memastikan kualitas dan konsistensi. Data yang

hilang dihapus atau dilengkapi, sedangkan data yang tidak dalam skala yang sama dinormalisasi agar pengukuran jarak tidak didominasi oleh satu variabel tertentu. Selain itu, data kategorikal dikonversi ke dalam bentuk numerik agar dapat diproses oleh algoritma KNN.

2.2 Penerapan K-Nearest Neighbor (KNN)

Metode ini termasuk dalam algoritma supervised learning, yang berarti memerlukan data yang telah dilabeli (dengan kelas atau nilai target) untuk melatih model. KNN bekerja berdasarkan prinsip kemiripan atau kedekatan antara data yang baru (data uji) dengan data yang sudah ada (data pelatihan). Ketika sebuah data baru harus diklasifikasikan, KNN akan menghitung jarak antara data baru tersebut dengan semua data yang ada di dataset pelatihan, lalu memilih sejumlah k data terdekat. Keputusan klasifikasi dibuat berdasarkan mayoritas kelas dari data tetangga terdekat tersebut.

2.3 Penerapan Random Forest

Metode ini digunakan untuk tugas klasifikasi, regresi, dan berbagai tugas lainnya yang melibatkan analisis data. Random Forest bekerja dengan membangun sejumlah besar decision tree selama pelatihan dan menggabungkan hasilnya (melalui voting atau rata-rata) untuk menghasilkan prediksi yang lebih akurat dan stabil.

2.4 Google Colab

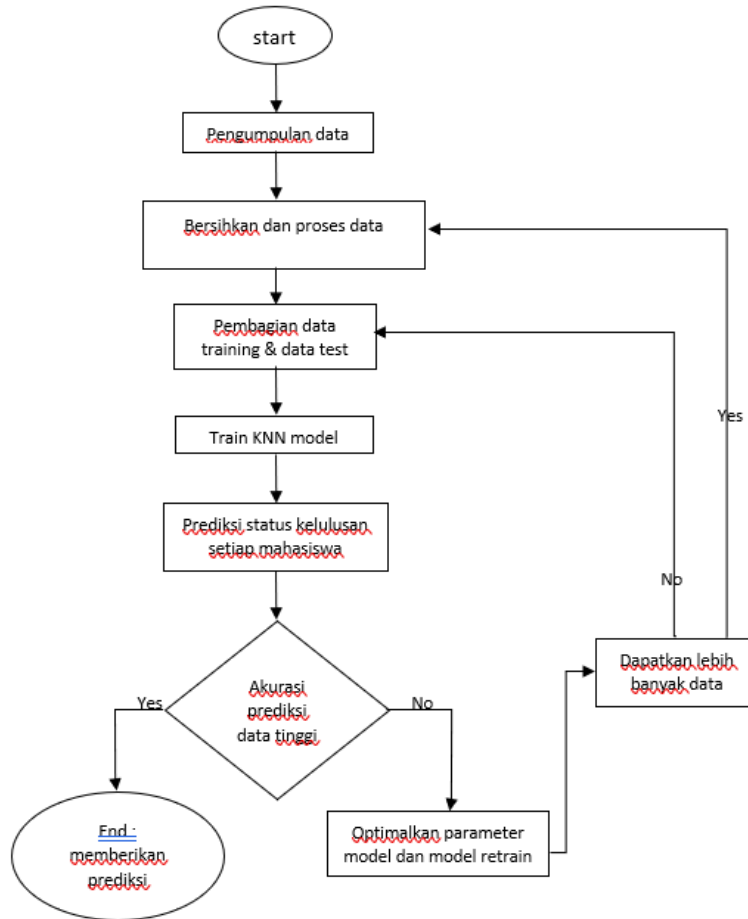
Google Colab (atau Google Colaboratory) adalah sebuah platform berbasis web yang disediakan oleh Google untuk menjalankan kode Python secara interaktif di lingkungan Jupyter Notebook. Google Colab sangat populer di kalangan pengembang, peneliti, dan pelajar karena menyediakan sumber daya komputasi seperti GPU dan TPU secara gratis, yang sangat berguna untuk pembelajaran mesin, analisis data, dan pengembangan model AI.



Gambar 2.4 Software Google Colab

2.5 Flowchart

Model KNN ini dievaluasi menggunakan teknik cross-validation dengan metrik seperti akurasi, presisi, recall, dan F1-score, di mana model dengan $K=5$ menunjukkan akurasi terbaik sebesar 86%. Dari hasil analisis, variabel seperti IPK, jumlah SKS yang ditempuh, dan durasi penyelesaian tugas akhir ditemukan sebagai faktor dominan yang memengaruhi kelulusan mahasiswa. Prediksi ini diharapkan dapat membantu Universitas Muhammadiyah Jakarta dalam mengidentifikasi mahasiswa yang berisiko terlambat lulus, sehingga dapat dilakukan intervensi dini untuk meningkatkan keberhasilan akademik mereka.



Gambar 2.4 Flowchart analisis dan predeksi prediksi kelulusan

2.6 Bahan Penelitian

a. Perangkat Lunak:

Google Colab digunakan sebagai platform berbasis cloud yang memungkinkan pengguna untuk menulis, menjalankan, dan berbagai kode Python.

b. Perangkat Keras:

Tabel 2.6 Spesifikasi laptop

| | |
|--------------------|----------------------------------|
| Model Laptop | HP 14-em0167AU em0168AU |
| Prosesor | AMD Ryzen 7-7730U |
| GPU | AMD Radeon™ Graphics |
| Memory and storage | 16 GB memory; 512 GB SSD storage |
| DirectX Version | DirectX 12 |
| OS | Windows 11 Home Single Language |

3. Hasil dan Pembahasan

3.1 Preproses data

Dataset ini adalah untuk secara memprediksi predikat kelulusan berdasarkan beberapa pengukuran yang tercantum dalam dataset. Beberapa batasan diterapkan pada pemilihan ini secara khusus.

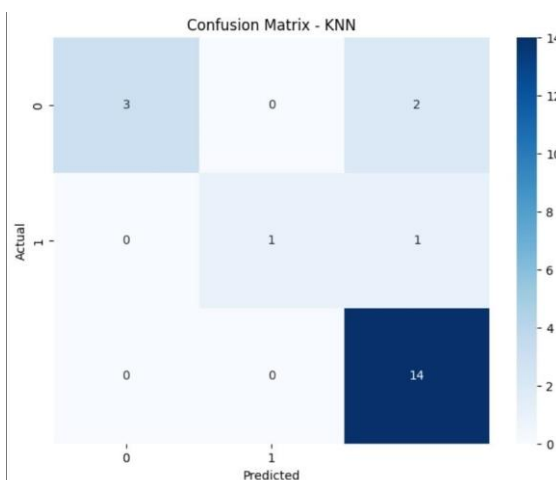
Tabel 3.1 Dataset lulusan elektro UMJ

| NIM | NAMA | TANGGAL_LULUS | TOTAL | IPK | PREDIKAT_KELULUSAN | NOMOR_SERI_TRANSKIP | PENOMORAN_UJAZAH_NASIONAL | LAMA_STUDI | STATUS_BEKERJA | JUMLAH_ORGANISASI |
|------------|----------------------|---------------|-------|------|--------------------|---------------------|---------------------------|------------|----------------|-------------------|
| 2015420047 | Govinda Gatot Aditya | 2/11/2022 | 144 | 3,15 | Sangat Memuaskan | 47099 | 202012022000014 | 4 | 0 | 1 |
| 2015420052 | Imam Abdulloh Hasan | 2/11/2022 | 144 | 3,24 | Sangat Memuaskan | 47100 | 202012022000021 | 4 | 0 | 2 |
| 2015420069 | Teguh Santoso | 2/11/2022 | 144 | 3,01 | Sangat Memuaskan | 47101 | 202012022000038 | 4 | 0 | 1 |
| 2016420048 | Ahmad Hairul Hamimi | 2/11/2022 | 144 | 3,09 | Sangat Memuaskan | 47102 | 202012022000045 | 4,5 | 1 | 3 |
| 2016420056 | Edy Purwanto | 2/11/2022 | 144 | 3,05 | Sangat Memuaskan | 47103 | 202012022000052 | 4,5 | 1 | 1 |
| 2016420057 | Fadli Aristiyan | 2/11/2022 | 144 | 3,02 | Sangat Memuaskan | 47104 | 202012022000069 | 4 | 1 | 0 |
| 2017420001 | Alfi Firman Nafis | 2/11/2022 | 144 | 3,09 | Sangat Memuaskan | 47105 | 202012022000076 | 4 | 0 | 3 |
| 2017420004 | Dandy Akhdan Akbar | 2/11/2022 | 144 | 3,05 | Sangat Memuaskan | 47106 | 202012022000083 | 4,5 | 0 | 2 |
| 2017420007 | Farhan Rizqulloh | 2/11/2022 | 144 | 3,06 | Sangat Memuaskan | 47107 | 202012022000090 | 4,5 | 0 | 2 |
| 2017420014 | Muhammad Farhan | 2/11/2022 | 144 | 2,89 | Memuaskan | 47108 | 202012022000106 | 4,5 | 0 | 1 |

3.2 Implementasi algoritma

a. *K-Nearest Neighbor*

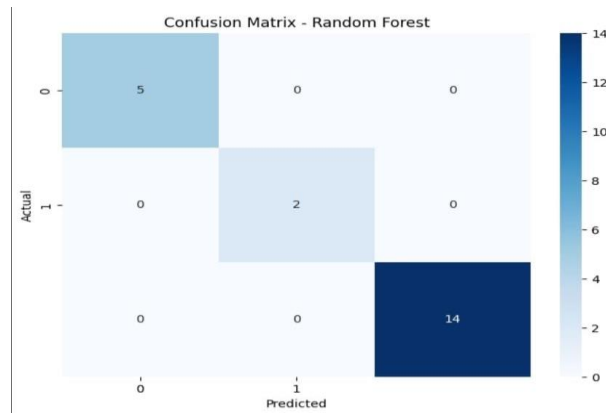
Metode ini termasuk dalam kelompok lazy learning atau instance-based learning, karena tidak mempelajari model eksplisit selama proses pelatihan. Sebaliknya, KNN menyimpan semua data pelatihan dan membuat prediksi berdasarkan kesamaan dengan data baru saat melakukan klasifikasi atau regresi.



Gambar 3.2 a. Confusion matrix K-NN

b. Random Forest

Random Forest adalah algoritma yang terdiri dari sejumlah Decision Tree yang tumbuh hingga selesai tanpa perlu dilakukan pemangkasan. Semakin banyak pohon yang dibangun, semakin akurat hasil prediksinya dan cenderung mengurangi risiko *overfitting*. Algoritma ini menghasilkan perkiraan keseluruhan dari semua pohon dan memiliki keunggulan dalam pemilihan fitur secara otomatis (Lin et al., 2017)



Gambar 3.1 b. Confusion matrix Random Forest

3.3 Evaluasi Hasil Penggabungan K-NN dan Random Forest

Tabel 3.3 Tabel evaluasi

| Model | Class | Precision | Recall | F1-Score | Support |
|---------------|--------------|-----------|--------|----------|---------|
| K-NN | 0 | 1.00 | 0.60 | 0.75 | 5 |
| | 1 | 1.00 | 0.50 | 0.67 | 2 |
| | 2 | 0.82 | 1.00 | 0.90 | 14 |
| | Accuracy | | | 0.86 | 21 |
| | Macro Avg | 0.94 | 0.70 | 0.77 | 21 |
| | Weighted Avg | 0.88 | 0.86 | 0.84 | 21 |
| Random Forest | 0 | 1.00 | 1.00 | 1.00 | 5 |
| | 1 | 1.00 | 1.00 | 1.00 | 2 |
| | 2 | 1.00 | 1.00 | 1.00 | 14 |
| | Accuracy | | | 1.00 | 21 |
| | Macro Avg | 1.00 | 1.00 | 1.00 | 21 |
| | Weighted Avg | 1.00 | 1.00 | 1.00 | 21 |

4. Kesimpulan dan Saran

4.1 Kesimpulan

(KNN). Dengan data akademik yang mencakup Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, tingkat kehadiran, nilai mata kuliah inti, dan durasi penyelesaian tugas akhir, model menunjukkan akurasi tertinggi sebesar 86% saat menggunakan parameter K=5 dengan metrik jarak Euclidean. Hasil penelitian mengungkapkan bahwa variabel IPK, jumlah SKS, dan durasi penyelesaian tugas akhir adalah faktor dominan yang memengaruhi kelulusan mahasiswa. Model prediksi ini memungkinkan identifikasi mahasiswa yang berisiko terlambat lulus sehingga pihak universitas dapat memberikan intervensi akademik yang lebih dini dan terarah.

Pengembangan model dengan memasukkan variabel non-akademik, seperti keikutsertaan dalam organisasi, status pekerjaan, dan motivasi belajar, diharapkan dapat meningkatkan akurasi dan fleksibilitas model dalam memprediksi kelulusan mahasiswa. Dengan demikian, hasil penelitian ini memberikan

kontribusi nyata dalam mendukung pengelolaan pendidikan yang lebih baik di Universitas Muhammadiyah Jakarta.

4.2 Saran

1. **Peningkatan Kualitas Data:** Universitas Muhammadiyah Jakarta perlu meningkatkan kualitas dan kelengkapan data akademik untuk mendukung prediksi yang lebih akurat. Pengumpulan data non-akademik, seperti partisipasi mahasiswa dalam organisasi atau status pekerjaan, juga perlu diperhatikan.
2. **Penggunaan Algoritma Tambahan:** Selain KNN, disarankan untuk menguji algoritma lain seperti Random Forest atau Gradient Boosting untuk membandingkan hasil dan mendapatkan model terbaik yang sesuai dengan data mahasiswa.
3. **Penerapan Sistem Prediksi di Universitas:** Model prediksi ini dapat diimplementasikan dalam bentuk sistem berbasis web yang dapat diakses oleh staf akademik untuk membantu memantau mahasiswa secara real-time.
4. **Pengembangan Penelitian Lanjutan:** Penelitian lebih lanjut dapat dilakukan dengan memperluas cakupan variabel yang digunakan, seperti data psikologis atau sosial ekonomi mahasiswa, untuk memberikan prediksi yang lebih holistik.

Daftar Pustaka

- Putri, O. S. D., & Putri, A. Y. P. (2021). *Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ*.
- Wati, N. (2022). *Prediksi Kelulusan Mahasiswa Menggunakan K-Nearest Neighbor Berbasis Particle Swarm Optimization*. *Jurnal Teknologi Informasi Indonesia (JTII)*, 6(2), 118-127.
- Hakim, L. R., Rizal, A., & Ratnasari, D. (2019). *Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)*. *Jurnal Teknologi Informasi dan Multimedia (JTIM)*, 1(1), 10-20.
- Putri, A. Y. P., & Putri, O. S. D. (2019). *Model Algoritma K-Nearest Neighbor (K-NN) untuk Prediksi Kelulusan Mahasiswa*. *Jurnal Nasional Teknologi dan Sistem Informasi*, 5(1), 50-60.
- Manullang, R. A., & Sianturi, F. A. (2021). *Penerapan Algoritma K-Nearest Neighbour untuk Memprediksi Kelulusan Mahasiswa*. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 4(2), 42-50.
- Prasetyawan, D., & Gatra, R. (2022). *Algoritma KNN untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan*. *Jurnal Informatika Sunan Kalijaga (JISKA)*, 7(1), 56-67.
- Hakim, L. R., & Rizal, A. (2019). *Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)*. *Jurnal Teknologi Informasi dan Multimedia (JTIM)*, 1(1), 10-20.
- Putri, O. S. D., & Putri, A. Y. P. (2021). *Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ*. *Seminar Nasional Energi dan Teknologi Informasi Komunikasi (SNETIK)*, 4(1), 100-110.
- Wati, N. (2022). *Prediksi Kelulusan Mahasiswa Menggunakan K-Nearest Neighbor Berbasis Particle Swarm Optimization*. *Jurnal Teknologi Informasi Indonesia (JTII)*, 6(2), 118-127.
- Putri, A. Y. P., & Putri, O. S. D. (2019). *Model Algoritma K-Nearest Neighbor (K-NN) untuk Prediksi Kelulusan Mahasiswa*. *Jurnal Nasional Teknologi dan Sistem Informasi*, 5(1), 50-60.