

PREDIKSI CALON KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

Reza Dirantara¹, Febri Sugandi²

Universitas Dharma Wacana

e-mail: ¹rezadirgantara66@gmail.com, ²fsugandi87@gmail.com

Abstract: *The Academic Information System (SIKAD) has become a crucial tool in universities for monitoring and evaluating student performance periodically. Data generated by this system, such as Grade Point Average (GPA), attendance, and student activities, can be utilized to predict timely graduation. In this study, the K-Nearest Neighbors (K-NN) algorithm is applied to predict student graduation based on academic and non-academic data. This method employs a distance-based approach to analyze historical data. The findings indicate that the K-NN algorithm provides accurate predictions, enabling educational institutions to implement early interventions for at-risk students. This study is expected to support strategic decision-making in improving student graduation rates.*

Keywords: *SIKAD, K-Nearest Neighbors, graduation prediction, distance-based algorithm, early intervention.*

Abstrak: Aplikasi Sistem Informasi Akademik (SIKAD) telah menjadi alat penting dalam perguruan tinggi untuk memonitor dan mengevaluasi hasil belajar mahasiswa secara berkala. Berbagai data yang dihasilkan oleh sistem ini, seperti Indeks Prestasi Kumulatif (IPK), kehadiran, dan aktivitas mahasiswa, dapat digunakan untuk memprediksi kelulusan tepat waktu. Dalam penelitian ini, algoritma K-Nearest Neighbors (K-NN) diterapkan untuk memprediksi kelulusan mahasiswa berdasarkan data akademik dan non-akademik. Metode ini bekerja dengan menggunakan pendekatan berbasis jarak untuk menganalisis data historis. Hasil penelitian menunjukkan bahwa algoritma K-NN memberikan prediksi yang akurat, sehingga membantu institusi pendidikan melakukan intervensi dini bagi mahasiswa yang berisiko. Penelitian ini diharapkan dapat mendukung pengambilan keputusan strategis dalam meningkatkan tingkat kelulusan mahasiswa.

Kata kunci: SIKAD, K-Nearest Neighbors, prediksi kelulusan, algoritma berbasis jarak, intervensi dini.

PENDAHULUAN

Aplikasi Sistem Informasi Akademik (SIKAD) di perguruan tinggi memiliki keunggulan dalam memonitor dan mengevaluasi hasil belajar mahasiswa. Dengan berbasis database cloud computing, sistem ini menyediakan laporan akurat mengenai kinerja mahasiswa, seperti Indeks Prestasi Kumulatif (IPK), jumlah kredit semester (SKS), dan kehadiran, yang berguna untuk pengambilan keputusan strategis. Namun, faktor-faktor seperti

prestasi akademik, kehadiran, dan partisipasi dalam kegiatan ekstrakurikuler dapat memengaruhi kelulusan tepat waktu. Oleh karena itu, diperlukan pendekatan berbasis data untuk memprediksi kelulusan. Algoritma K-Nearest Neighbors (K-NN) digunakan sebagai metode sederhana namun efektif untuk memproses data akademik dan non-akademik guna memprediksi tingkat kelulusan. Penelitian ini bertujuan mengeksplorasi penerapan K-NN untuk membantu institusi pendidikan meningkatkan angka kelulusan melalui

intervensi yang tepat sasaran.

Machine Learning

Pembelajaran mesin adalah cabang ilmu komputer, kecerdasan buatan, dan statistik yang dapat memperkirakan hasil penelitian dan analisis data (Andreas C. Muller & Sarah Guido, 2017). Penambangan data adalah teknik yang sama dengan pembelajaran mesin. Namun, dalam pembelajaran mesin, mereka mencari model dan menggunakan data untuk meningkatkan pemahaman program. Budiharto, Widodo (2016). Dengan menganalisis data yang kita ubah, pembelajaran mesin membantu kita merencanakan tindakan di masa depan (Judith Hurwitz & Daniel Kirsch, 2018).

Algoritma K-Nearest Neighbor

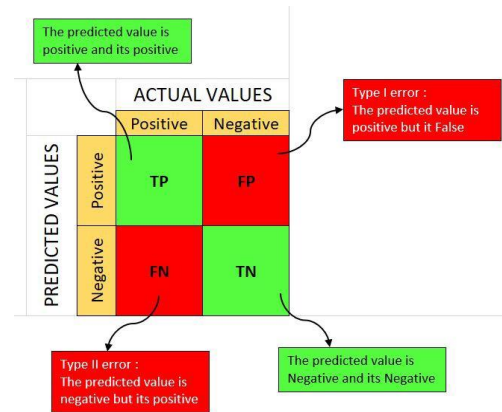
Algoritma K-Nearest Neighbor adalah algoritma pembelajaran mesin yang paling sederhana. Digunakan dengan membuat model pembelajaran mesin untuk menyimpan dataset pelatihan, membuat prediksi untuk titik data baru, kemudian menemukan titik data terdekat. Algoritma K-Nearest Neighbor adalah algoritma pembelajaran yang diawasi (Andreas C. Muller dan Sarah Guido, 2017).

Menurut Widodo Budiharto (2016), algoritma K-Nearest Neighbor menggunakan algoritma non parametrik yang biasa digunakan untuk regresi dan klasifikasi. Proses berikut digunakan untuk:

1. Mengumpulkan data dengan berbagai algoritma. Menghitung nilai jarak (*distance calculation*).
2. Menganalisis data dengan berbagai algoritma.
3. Memproses data *training*
4. Menghitung tingkat kesalahan.
5. Menalakan algoritma untuk menentukan kelas yang cocok dengan data yang dimasukkan (Widodo Budiharto, 2016).

CONFUSION MATRIK

Matriks konfusi biasanya digunakan untuk mengukur kinerja model klasifikasi, dan juga dapat digunakan untuk mengukur kinerja model melalui perhitungan metrik kinerja seperti akurasi, presisi, perolehan, dan skor F1.



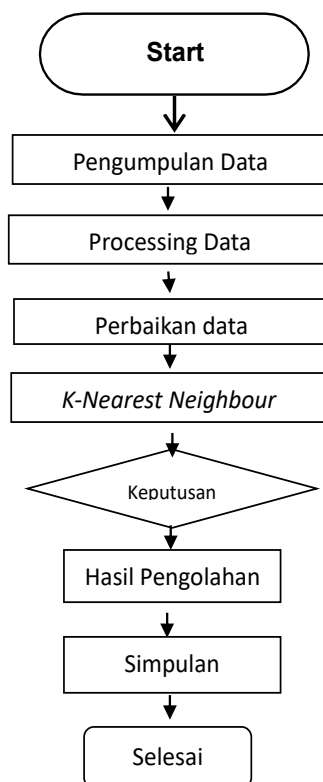
yang mana:

1. True Positive (TP) memprediksi ramalan positif dan itu benar.
 2. True Negative (TN) memprediksi ramalan negatif dan itu benar.
 3. False Positive (FP) memprediksi ramalan positif dan itu salah
 4. False Negative (FN) memprediksi ramalan negatif dan itu salah.
- Kesalahan Tipe 2 ini sangat berbahaya.

Nilai False Positive (FP), False Negative (FN), True Positive (TP), dan True Negative (TN) digunakan untuk menentukan akurasi. Nilai akurasi menunjukkan kemampuan sistem untuk mengklasifikasikan data dengan benar. Dengan kata lain, nilai presisi adalah prediksi antara data yang diklasifikasikan dengan benar dan kumpulan data. Rumus berikut dapat digunakan untuk menghitung tingkat akurasi:

$$\text{Akurasi} = \frac{\text{Total data yang diprediksi dengan benar}}{\text{Total pengujian yang dilaksanakan}} \times 100\%$$

METODE



1. **Analisis Data** Penelitian dimulai dengan analisis data melalui studi literatur, wawancara, dokumentasi, dan observasi untuk mengidentifikasi kebutuhan data.
2. **Pengumpulan Data:** Data terdiri dari dua bagian: data pelatihan untuk membentuk model dan data pengujian untuk pengujian.
3. **Preprocessing Data** Tahap ini dilakukan untuk membersihkan data mentah agar dapat digunakan dalam proses pengolahan berikutnya.
4. **K-Nearest Neighbors** Algoritma KNN digunakan untuk memprediksi kelulusan tepat waktu dan tidak tepat waktu berdasarkan data yang telah diolah.
5. **Simpulan** Penelitian membandingkan kinerja algoritma Naïve Bayes dan KNN dalam memprediksi kelulusan mahasiswa Fakultas Ilmu Social Politik Dharma Wacana Metro

K-Nearest Neighbor yang termasuk ke dalam supervised learning, adalah salah satu metode klasifikasi data mining. Data instruksional dan atribut telah diklasifikasikan, sehingga data baru diklasifikasikan berdasarkan perbandingan data instruksional pada kesamaan mayoritas. Nilai jarak dihitung dengan melakukan pengujian data terhadap data pelatihan dengan menggunakan nilai terkecil dari K-Nearest Neighbor (Krisandi et al., 2015). Untuk menghitung jarak, matematika geometri biasanya menggunakan jarak sebagai berikut:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Di mana :

- D = Jarak Kedekatan
- x = Data Training
- y = Data Testing
- i = Fitur ke –
- n = Jumlah Fitur

Karena kesamaan data dapat lebih dari satu, K-Nearest Neighbor dapat mengklasifikasikan sejumlah k data yang serupa sebagai data yang memiliki banyak kesamaan. Dalam K-Nearest Neighbour, jarak geometris digunakan. Persyaratan untuk menghitung metode K-Nearest Neighbour adalah sebagai berikut:

1. Menemukan parameter awal k
2. Menemukan jarak antara data yang akan dievaluasi
3. Urutkan jarak yang didapat
4. Temukan jarak yang paling dekat dengan orde nilai k
5. Menggabungkan dengan kelas yang sesuai
6. Menemukan jumlah kelas di mana kategori obyek yang paling umum dan dapat diprediksi adalah

Selain kemudahan untuk menemukan jarak terpendek antar data, algoritma ini juga memiliki kemampuan untuk menggeneralisasi jumlah data pelatihan yang relatif kecil (Satriya & Santoso, 2018):

HASIL DAN PEMBAHASAN

Analisis Kebutuhan Data

Sebelum memulai perhitungan algoritma, proses pengumpulan data dilakukan. Penelitian ini menggunakan data mahasiswa Fakultas Ilmu Sosial Politik Dharma Wacana Metro dari angkatan 2019–2023, yang terdiri dari 543 mahasiswa yang akan lulus dari angkatan 2019–2023. Nomor Induk Mahasiswa (NIM), Nama, Jenis Kelamin (JK), Tanggal Lahir, Alamat, Indeks Prestasi Kumulatif (IPK), Penghasilan Orang Tua, UKOM, Tes Potensi Akademik, dan Keterangan Lulus Tepat Waktu dan Tidak Tepat Waktu Mahasiswa digunakan dalam penelitian ini. Tabel 1 menunjukkan data penelitian.

| ID | NAMA | JK | UMUR | ALAMAT | STATUS LULUS |
|----|----------------|---------|---------|---------|--------------|
| 11 | Fachrudin | Sejenis | Sejenis | Tinggi | Selesai |
| 14 | Muhammad Rizki | Sejenis | Tinggi | Tinggi | Selesai |
| 15 | Alvin | Sejenis | Sejenis | Sejenis | Selesai |
| 16 | Alvin | Sejenis | Tinggi | Sejenis | Tinggi |
| 17 | Alvin | Sejenis | Sejenis | Sejenis | Selesai |
| 18 | Alvin | Sejenis | Sejenis | Tinggi | Selesai |
| 19 | Laili | Sejenis | Sejenis | Sejenis | Selesai |
| 20 | Rizki | Sejenis | Tinggi | Sejenis | Tinggi |
| 21 | Alvin | Sejenis | Sejenis | Sejenis | Selesai |
| 22 | Alvin | Sejenis | Tinggi | Sejenis | Selesai |
| 23 | Alvin | Sejenis | Sejenis | Sejenis | Selesai |
| 24 | Alvin | Sejenis | Tinggi | Sejenis | Selesai |
| 25 | Alvin | Sejenis | Sejenis | Sejenis | Selesai |

Gambar Tabel data mahasiswa

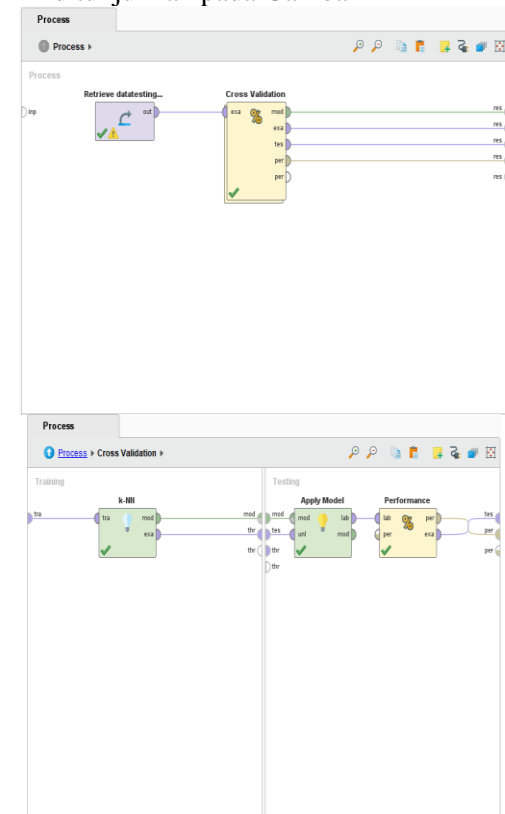
Transformasi Data

Setelah proses analisis kebutuhan data, langkah selanjutnya adalah melakukan transformasi data. Berdasarkan data mahasiswa dari angkatan 2019 sampai 2023 di Fakultas Ilmu Sosial Politik Dharma Wacana Metro, data yang ditransformasikan adalah Lulus Tepat Waktu dan Tidak Tepat Waktu Mahasiswa. Data ini digunakan sebagai atribut input untuk dianalisis menggunakan algoritma K-Nearest Neighbor.

Hasil Pengujian

Berdasarkan 543 data mahasiswa angkatan 2019 hingga 2023 yang telah diuji, diperoleh hasil Akurasi mencapai 98,19%. Proses pengambilan data menggunakan format CSV di RapidMiner untuk algoritma K-Nearest Neighbor.

Pengolahan data menggunakan algoritma ini ditunjukkan pada Gambar



Gambar Proses Pengolahan Data dengan Algoritma K-Nearest Neighbor

SIMPULAN

Pada tahap pengujian, analisis, dan desain sistem prediksi tingkat kelulusan mahasiswa menggunakan data dari mahasiswa fakultas ilmu Sosial Politik dharma wacana metro angkatan 2019 hingga 2023 dengan algoritma k-Nearest Neighbor, dapat disimpulkan Hasil perbandingan akurasi dari evaluasi algoritma menggunakan metrik akurasi dan validasi silang 2-fold menunjukkan akurasi 96,86%

DAFTAR PUSTAKA

Sugandi, F. (2023). PREDIKSI KELULUSAN MAHASISWA STMIK DHARMAWACANA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS:

- PREDIKSI KELULUSAN MAHASISWA STMIK DHARMAWACANA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS. *Sienna*, 4(1), 20-26.
- Astri, J., Karman, J., & Daulay, N. K. (2023). Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbor (KNN) pada Fakultas Ilmu Teknik, Universitas Bina Insan. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 8(1), 169-173.
- Susanto, E. S., Kusriani, K., & Al Fatta, H. (2018). Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Respati*, 13(2).
- Banjarsari, M. A., Budiman, I., & Farmadi, A. (2016). Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4. *Klik-Kumpulan Jurnal Ilmu Komputer*, 2(2), 159-173.
- Zulfallah, F. H. (2022). Implementasi algoritma KNN dalam mengukur ketepatan kelulusan mahasiswa UIN Syarif Hidayatullah Jakarta (Bachelor's thesis, Perpustakaan Fakultas Sains dan Teknologi UIN Jakarta).
- Saputra, A. Y., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour. *Techno. Com*, 17(4).
- Nasution, T. (2020). Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu. *Jurnal Perangkat Lunak*, 2(1), 1-14.
- Sumarlin, S., & Anggraini, D. (2019). Implementasi K-nearest neighbor pada rapidminer untuk prediksi kelulusan mahasiswa. *HOAQ (High Education of Organization Archive Quality): Jurnal Teknologi Informasi*, 10(1), 35-41.