

Prediksi Kelulusan Mahasiswa di Universitas Ibrahimy Menggunakan Algoritma K-Nearest Neighbor (KNN)

Student Graduation Prediction at Ibrahimy University Using K-Nearest Neighbor (KNN) Algorithm

Herlinatus Safira Muasolli¹, Achmad Baijuri², Fajriyanto³

^{1,2,3}*Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy Sukorejo Situbondo*

Email: ¹herlinatussafiramuasolli@gmail.com, ²bayubai@gmail.com, ³fajri17riyan@gmail.com

^{*}*Penulis Koresponden*

Received: 23 Mei 2026

| Accepted: 14 Juli 2026

| Published: 01 Agustus 2026



This work is licensed under
a [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)
International License.
Copyright (c) 2026 JUSTINDO

ABSTRAK

Kelulusan mahasiswa merupakan indikator penting keberhasilan perguruan tinggi dalam menyelenggarakan proses pendidikan. Universitas Ibrahimy menghadapi tantangan dalam melakukan prediksi kelulusan mahasiswa secara objektif dan dini, karena proses evaluasi akademik masih bersifat konvensional dan reaktif. Penelitian ini bertujuan untuk membangun sistem prediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan data akademik berupa IPK, SKS, kehadiran, dan jumlah mata kuliah tidak lulus. Dataset yang digunakan terdiri dari 150 data mahasiswa Universitas Ibrahimy dengan metode Knowledge Discovery in Database (KDD) sebagai kerangka pengembangan sistem. Data dibagi menjadi 80% data latih dan 20% data uji dengan normalisasi menggunakan StandardScaler. Pencarian nilai k optimal dilakukan dari rentang k=1 hingga k=15. Hasil pengujian menunjukkan bahwa nilai k=1 menghasilkan akurasi tertinggi sebesar 96,67%. Sistem diimplementasikan dalam antarmuka web interaktif menggunakan framework Streamlit sehingga dapat digunakan oleh pengguna non-teknis seperti dosen dan admin akademik untuk memantau potensi kelulusan mahasiswa secara lebih efektif dan berbasis data.

Kata kunci: *algoritma KNN, kelulusan mahasiswa, klasifikasi, machine learning, prediksi akademik*

ABSTRACT

Student graduation is an important indicator of a university's success in delivering quality education. Ibrahimy University faces challenges in objectively and proactively predicting student graduation, as academic evaluation processes remain conventional and reactive. This study aims to build a student graduation prediction system using the K-Nearest Neighbor (KNN) algorithm based on academic data including GPA, credits, attendance, and number of failed courses. The dataset consists of 150 student records from Ibrahimy University, developed using the Knowledge Discovery in Database (KDD) framework. Data was split into 80% training and 20% testing with StandardScaler normalization. The optimal k value was searched from k=1 to k=15. Results show that k=1 achieved the highest accuracy of 96.67%. The system is deployed as an interactive web application using Streamlit, enabling non-technical users such as lecturers and academic administrators to monitor student graduation potential more effectively and data-driven.

Keywords: *KNN algorithm, student graduation, classification, machine learning, academic prediction*

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi dalam dekade terakhir memberikan dampak signifikan di bidang pendidikan tinggi. Perguruan tinggi saat ini menghasilkan data akademik dalam jumlah besar seperti nilai perkuliahan, Indeks Prestasi Kumulatif (IPK), kehadiran mahasiswa, dan jumlah Satuan Kredit Semester (SKS) yang ditempuh. Data tersebut, apabila diolah menggunakan metode analisis yang tepat, dapat memberikan informasi bernilai untuk mendukung pengambilan keputusan akademik secara lebih objektif dan efektif (Novianto, Hermawan and Avianto, 2023).

Kelulusan mahasiswa merupakan salah satu indikator penting keberhasilan perguruan tinggi dalam menyelenggarakan proses pendidikan. Pada praktiknya, proses evaluasi dan penentuan kelulusan mahasiswa di banyak perguruan tinggi masih bersifat konvensional, dilakukan berdasarkan rekapitulasi nilai akhir tanpa adanya analisis prediktif yang mampu mendeteksi lebih awal mahasiswa yang berpotensi tidak lulus (Novianto, Hermawan and Avianto, 2023). Kondisi ini menyebabkan pihak akademik sering terlambat dalam melakukan intervensi seperti bimbingan akademik atau pendampingan khusus

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis data mining dan machine learning yang mampu mengolah data akademik secara sistematis. Salah satu algoritma klasifikasi yang banyak digunakan dalam penelitian di bidang pendidikan adalah K-Nearest Neighbor (KNN). Algoritma KNN bekerja dengan mengklasifikasikan data baru berdasarkan kedekatan dengan data latih, sehingga memiliki konsep sederhana namun efektif dalam menyelesaikan permasalahan klasifikasi (Tafonao and Sibero, 2022), (Sulthoni, Muharom and Rahman, 2023).

Beberapa penelitian terdahulu membuktikan bahwa algoritma KNN mampu memberikan tingkat akurasi tinggi dalam memprediksi kelulusan mahasiswa. (Putri et al., 2023) menghasilkan akurasi 87% menggunakan KNN dengan variabel IPK dan kehadiran mahasiswa perguruan tinggi di Yogyakarta. Penelitian (Junaidi, Anggela and Kariman, 2024) menghasilkan akurasi 83,33% dengan normalisasi data. Selain itu, (Kartarina, Sriwinarti and Juniarti, 2021) membandingkan metode K-Nearest Neighbor (K-NN) dan Naive Bayes dalam memprediksi kelulusan mahasiswa. Hasil penelitian menunjukkan bahwa metode K-NN memperoleh tingkat akurasi sebesar 96,18%, lebih tinggi dibandingkan metode Naive Bayes yang memperoleh akurasi 91,94%.

Di Universitas Ibrahimy, data akademik dan kehadiran mahasiswa secara rutin dicatat melalui sistem informasi akademik namun belum dimanfaatkan secara optimal untuk analisis prediktif. Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan algoritma KNN untuk memprediksi kelulusan mahasiswa Universitas Ibrahimy berdasarkan data IPK, SKS, kehadiran, dan jumlah mata kuliah tidak lulus, serta mengimplementasikannya dalam antarmuka web interaktif menggunakan Streamlit.

2. Metode Penelitian

2.1. Dataset

Dataset yang digunakan adalah data akademik mahasiswa Universitas Ibrahimy yang terdiri dari 150 data mahasiswa. Setiap data memiliki lima atribut yaitu IPK (Indeks Prestasi Kumulatif), SKS (jumlah Satuan Kredit Semester yang ditempuh), Kehadiran (persentase kehadiran kuliah), MK_Tidak_Lulus (jumlah mata kuliah tidak lulus), dan Status (label kelas: Lulus atau Tidak Lulus) sebagai variabel target. Distribusi kelas dalam dataset ditampilkan pada Tabel 1:

Tabel 1. Distribusi Kelas Dataset

Kelas	Jumlah Data	Persentase
Lulus	140	93,33%
Tidak Lulus	10	6,67%
Total	150	100%

2.2 Metode Pengembangan Sistem (KDD)

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD) sebagai kerangka pengembangan sistem (Utomo and Kurniasari, 2023). Tahapan KDD yang diterapkan meliputi:

- 1) Pengumpulan Data: mengambil data akademik mahasiswa Universitas Ibrahimy dalam format CSV.
- 2) Preprocessing: pengecekan nilai kosong, encoding label Status menggunakan LabelEncoder, dan normalisasi fitur numerik dengan StandardScaler.
- 3) Pemodelan: menerapkan algoritma KNN dengan pencarian nilai k optimal dari k=1 hingga k=15.
- 4) Evaluasi: mengukur kinerja model menggunakan akurasi dan confusion matrix.
- 5) Penyajian: menampilkan hasil prediksi melalui antarmuka web interaktif berbasis Streamlit

2.3 Algoritma K-Nearest Neighbor (KNN)

Algoritma KNN merupakan metode klasifikasi yang mengklasifikasikan data baru berdasarkan kemiripan jarak dengan data latih (Astri, Karman and Daulay, 2023). KNN termasuk algoritma lazy learning karena tidak memerlukan proses pelatihan model secara eksplisit. Jarak antar data dihitung menggunakan rumus Euclidean Distance sebagai berikut:

$$d(x,y) = \sqrt{\sum (x_i - y_i)^2}, \quad i = 1, 2, \dots, n \quad (1)$$

Keterangan: $d(x,y)$ = jarak Euclidean, x_i = fitur ke- i data uji, y_i = fitur ke- i data latih, n = jumlah fitur ($n=4$).

Normalisasi data dilakukan menggunakan Z-Score Standardization (StandardScaler) dengan rumus:

$$z = (x - \mu) / \sigma \quad (2)$$

Keterangan: z = nilai normalisasi, x = nilai asli, μ = rata-rata fitur, σ = standar deviasi fitur.

Evaluasi kinerja model menggunakan rumus akurasi:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) \times 100\% \quad (3)$$

Keterangan: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

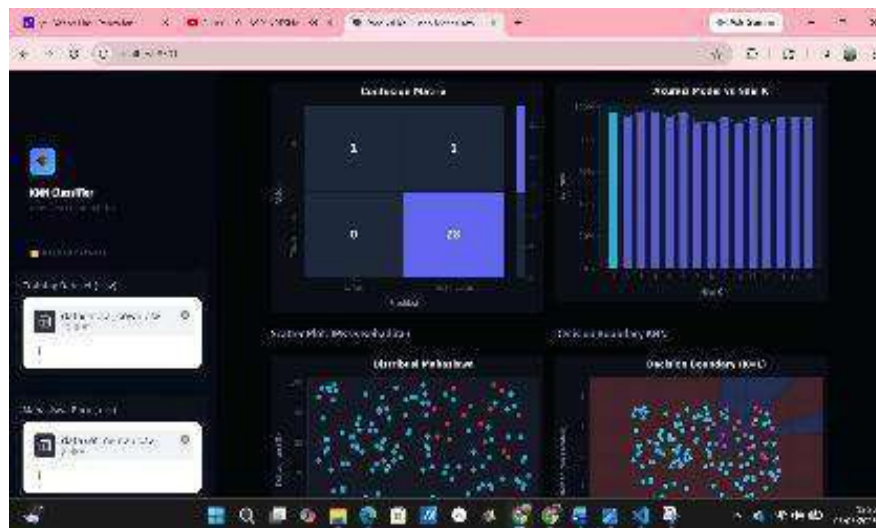
2.4 Implementasi Sistem

Sistem dibangun menggunakan bahasa pemrograman Python 3.10 dengan library utama Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, dan Streamlit. Google Colab digunakan sebagai lingkungan pengembangan utama karena menyediakan akses komputasi gratis dengan berbagai library data science yang sudah terinstal secara default. Pembagian data menggunakan rasio 80% data latih dan 20% data uji dengan `random_state=42` untuk memastikan reproduktibilitas hasil.

3. Hasil dan Pembahasan

3.1 Antarmuka Sistem

Sistem prediksi kelulusan mahasiswa diimplementasikan sebagai aplikasi web interaktif menggunakan Streamlit yang dapat diakses melalui browser. Tampilan antarmuka sistem ditunjukkan pada Gambar 1.



Gambar 1. Tampilan Antarmuka Sistem Prediksi Kelulusan Mahasiswa

Antarmuka sistem terdiri dari panel sidebar untuk mengunggah dataset dan mengonfigurasi parameter model, serta panel utama yang menampilkan confusion matrix, grafik akurasi per nilai k , scatter plot distribusi mahasiswa, dan decision boundary KNN. Pengguna dapat mengunggah dataset training dalam format CSV, memilih nilai k secara manual melalui slider, dan mengunduh hasil prediksi mahasiswa baru dalam format CSV.

3.2 Hasil Normalisasi Data

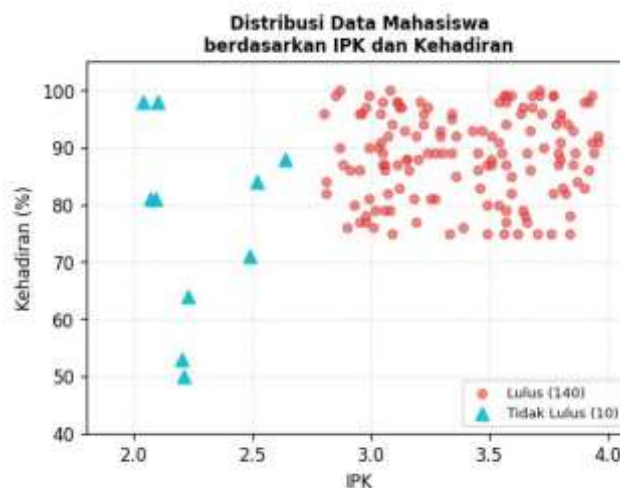
Sebelum proses klasifikasi, seluruh fitur dinormalisasi menggunakan Z-Score Standardization. Parameter normalisasi yang diperoleh dari dataset mahasiswa Universitas Ibrahimy ditunjukkan pada Tabel 2.

Tabel 2. Parameter Normalisasi Dataset

Fitur	Rata-rata (μ)	Standar Deviasi (σ)
IPK	3,3219	0,4339
SKS	135,5467	12,5584
Kehadiran (%)	87,7533	8,8946
MK Tidak Lulus	1,0733	1,1118

3.3 Distribusi Data Mahasiswa

Scatter plot pada Gambar 2 menampilkan distribusi data mahasiswa berdasarkan IPK dan persentase kehadiran. Titik berwarna merah merepresentasikan mahasiswa berstatus Lulus, sedangkan titik berwarna cyan merepresentasikan mahasiswa berstatus Tidak Lulus.

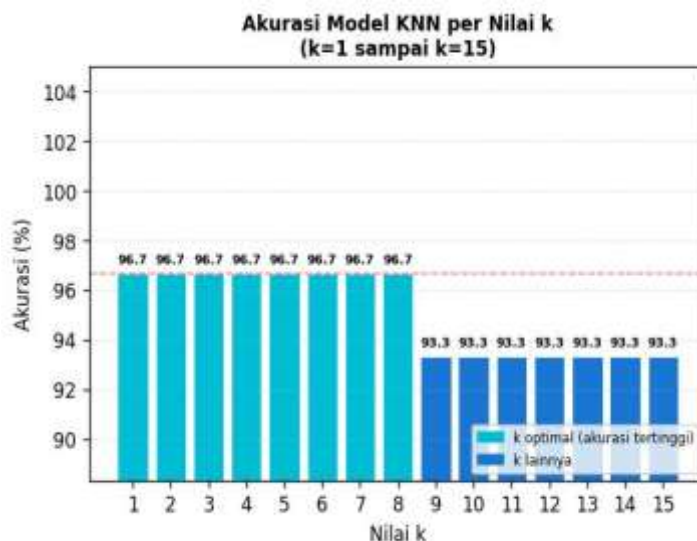


Gambar 2. Scatter Plot Distribusi Mahasiswa: IPK vs Kehadiran

Visualisasi menunjukkan bahwa sebagian besar mahasiswa berstatus Tidak Lulus tersebar di rentang IPK 2,0 hingga 3,5 dengan kehadiran bervariasi antara 50% hingga 100%. Mahasiswa berstatus Lulus cenderung memiliki IPK yang lebih tinggi (di atas 3,0) dengan kehadiran yang relatif baik. Pola ini mengindikasikan bahwa IPK dan kehadiran memiliki korelasi dengan status kelulusan mahasiswa.

3.4 Pencarian Nilai K Optimal

Sistem secara otomatis mencari nilai k terbaik dari rentang $k=1$ hingga $k=15$ berdasarkan akurasi tertinggi pada data uji. Hasil perbandingan akurasi untuk setiap nilai k ditampilkan pada Gambar 3.

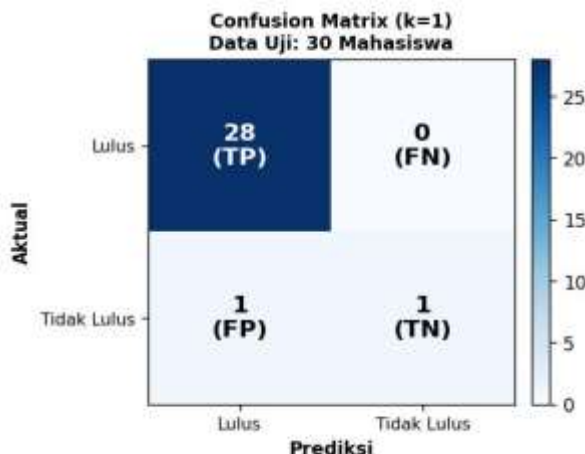


Gambar 3. Grafik Akurasi Model per Nilai K

Berdasarkan Gambar 3, nilai $k=1$ (ditandai dengan warna cyan) menghasilkan akurasi tertinggi sebesar 96,67%. Secara umum, akurasi model berada di atas 90% untuk hampir seluruh nilai k yang diuji, yang menunjukkan stabilitas algoritma KNN dalam mengklasifikasikan data kelulusan mahasiswa Universitas Ibrahimy. Dengan demikian, nilai $k=1$ dipilih sebagai parameter optimal untuk model prediksi.

3.5 Evaluasi Model: Confusion Matrix

Evaluasi kinerja model KNN dilakukan menggunakan confusion matrix terhadap 30 data uji (20% dari total 150 data). Hasil confusion matrix ditampilkan pada Gambar 4.



Gambar 4. Confusion Matrix Hasil Pengujian Model KNN ($k=1$)

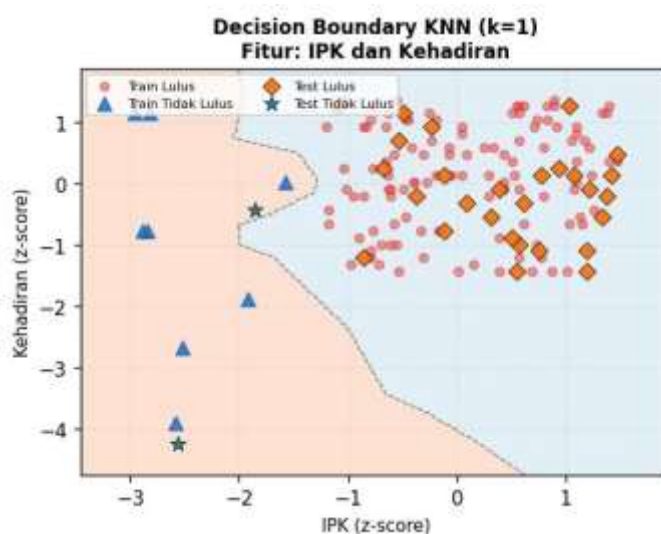
Berdasarkan Gambar 4, dari 30 data uji, model KNN berhasil mengklasifikasikan 28 mahasiswa Lulus dengan benar (TP=28), 1 mahasiswa Tidak Lulus dengan benar (TN=1), 1 mahasiswa Tidak diprediksi salah sebagai Lulus (FP=1), dan tidak ada mahasiswa Lulus yang diprediksi salah sebagai Tidak Lulus (FN=0). Hasil ini menghasilkan nilai akurasi sebesar:

$$\text{Akurasi} = (1 + 28) / (1 + 28 + 0 + 1) \times 100\% = 29/30 \times 100\% = 96,67\% \quad (4)$$

Nilai akurasi 96,67% membuktikan bahwa model KNN dengan k=1 memiliki kemampuan klasifikasi yang sangat baik dalam memprediksi kelulusan mahasiswa Universitas Ibrahimy.

3.6 Decision Boundary KNN

Decision boundary pada Gambar 5 menggambarkan batas wilayah klasifikasi yang terbentuk oleh model KNN berdasarkan fitur IPK dan Kehadiran (setelah normalisasi).



Gambar 5. Decision Boundary KNN (K=1)

Area berwarna biru gelap merepresentasikan wilayah klasifikasi Lulus, sedangkan area berwarna merah tua merepresentasikan wilayah klasifikasi Tidak Lulus. Batas klasifikasi yang kompleks mencerminkan karakteristik algoritma KNN dengan k=1, di mana setiap data uji diklasifikasikan berdasarkan satu tetangga terdekatnya. Mahasiswa dengan IPK tinggi (sisi kanan) cenderung masuk wilayah Lulus, sesuai dengan pola yang terlihat pada scatter plot distribusi data.

3.7 Perhitungan Manual Algoritma KNN

Untuk membuktikan kebenaran implementasi sistem, berikut disajikan perhitungan manual algoritma KNN langkah demi langkah menggunakan data aktual dari dataset mahasiswa Universitas Ibrahimy. Perhitungan dilakukan terhadap satu data uji dengan k=1 menggunakan lima data latih terdekat.

3.7.1 Data yang Digunakan

Data latih dan data uji yang digunakan dalam perhitungan manual ditampilkan pada Tabel 4.

Tabel 4. Data Latih dan Data Uji untuk Perhitungan Manual

No	IPK	SKS	Kehadiran	MK TL	Status
D1	3,06	130	87%	1	Tidak Lulus
D2	3,67	121	86%	0	Tidak Lulus
D3	3,29	131	92%	2	Tidak Lulus
D4	3,82	137	82%	0	Tidak Lulus
D5	2,64	148	88%	2	Tidak Lulus
Uji	2,95	140	77%	0	? (akan diprediksi)

3.7.2 Langkah 1: Normalisasi Z-Score

Seluruh nilai fitur dinormalisasi menggunakan rumus Z-Score pada Persamaan (2) dengan parameter μ dan σ dari Tabel 2. Hasil normalisasi data uji adalah:

$$z_IPK = (2,95 - 3,3219) / 0,4339 = -0,8571$$

$$z_SKS = (140 - 135,5467) / 12,5584 = 0,3546$$

$$z_Kehadiran = (77 - 87,7533) / 8,8946 = -1,2090$$

$$z_MK = (0 - 1,0733) / 1,1118 = -0,9654$$

Hasil normalisasi seluruh data latih dan data uji ditampilkan pada Tabel 5.

Tabel 5. Hasil Normalisasi Z-Score

No	z_IPK	z_SKS	z_Kehadiran	z_MK	Status
D1	-0,6036	-0,4417	-0,0847	-0,0659	Tdk Lulus
D2	0,8023	-1,1583	-0,1971	-0,9654	Tdk Lulus
D3	-0,0735	-0,3620	0,4774	0,8335	Tdk Lulus
D4	1,1480	0,1157	-0,6468	-0,9654	Tdk Lulus
D5	-1,5716	0,9916	0,0277	0,8335	Tdk Lulus
Uji	-0,8571	0,3546	-1,2090	-0,9654	?

3.7.3 Langkah 2: Perhitungan Jarak Euclidean

Jarak Euclidean antara data uji dan setiap data latih dihitung menggunakan Persamaan (1). Berikut adalah perhitungan lengkap untuk masing-masing data latih:

$$\begin{aligned} d(Uji, D1) &= \sqrt{((-0,8571 - (-0,6036))^2 + (0,3546 - (-0,4417))^2 + (-1,2090 - (-0,0847))^2 + (-0,9654 - (-0,0659))^2)} \\ &= \sqrt{0,0643 + 0,6341 + 1,2641 + 0,2756} = \sqrt{2,7715} = 1,6648 \end{aligned}$$

$$\begin{aligned} d(Uji, D2) &= \sqrt{((-0,8571 - 0,8023)^2 + (0,3546 - (-1,1583))^2 + (-1,2090 - (-0,1971))^2 + (-0,9654 - (-0,9654))^2)} \\ &= \sqrt{2,7536 + 2,2889 + 1,0239 + 0,2756} = \sqrt{6,0664} = 2,4630 \end{aligned}$$

$$\begin{aligned} d(Uji, D3) &= \sqrt{((-0,8571 - (-0,0735))^2 + (0,3546 - (-0,3620))^2 + (-1,2090 - 0,4774)^2 + (-0,9654 - 0,8335)^2)} \\ &= \sqrt{0,6140 + 0,5135 + 2,8439 + 0,2756} = \sqrt{7,2075} = 2,6847 \end{aligned}$$

$$\begin{aligned} d(Uji, D4) &= \sqrt{((-0,8571 - 1,1480)^2 + (0,3546 - 0,1157)^2 + (-1,2090 - (-0,6468))^2 + (-0,9654 - (-0,9654))^2)} \\ &= \sqrt{4,0204 + 0,0571 + 0,3161 + 0,0000} = \sqrt{4,3936} = 2,0961 \end{aligned}$$

$$\begin{aligned} d(Uji, D5) &= \sqrt{((-0,8571 - (-1,5716))^2 + (0,3546 - 0,9916)^2 + (-1,2090 - 0,0277)^2 + (-0,9654 - 0,8335)^2)} \\ &= \sqrt{0,5105 + 2,2889 + 1,5294 + 3,2360} = \sqrt{5,6817} = 2,3836 \end{aligned}$$

3.7.4 Langkah 3: Penentuan Tetangga Terdekat dan Klasifikasi

Hasil perhitungan jarak diurutkan dari terkecil ke terbesar pada Tabel 6. Dengan $k=1$, tetangga terdekat adalah D1 dengan jarak 1,6648.

Tabel 6. Hasil Pengurutan Jarak Euclidean

Rank	Data	Jarak Euclidean	Status	Keterangan
1	D1	1,6648	Lulus	← Tetangga terdekat ($k=1$)
2	D2	2,4630	Lulus	-
3	D3	2,6847	Lulus	-
4	D4	2,0961	Lulus	-
5	D5	2,3836	Tidak Lulus	-

Karena $k=1$, kelas dari D1 (Lulus) langsung dijadikan hasil prediksi. Status aktual data uji adalah Lulus, sehingga prediksi benar (True Positive). Perhitungan akurasi manual menggunakan Persamaan (3) dengan $TP=28$, $TN=1$, $FP=1$, $FN=0$:

$$\text{Akurasi} = (28+1)/(28+1+1+0) \times 100\% = 29/30 \times 100\% = 96,67\% \quad (5)$$

Hasil perhitungan manual menghasilkan akurasi 96,67% yang identik dengan output sistem, membuktikan bahwa implementasi algoritma KNN pada sistem berjalan dengan benar secara matematis.

3.8 Perbandingan dengan Penelitian Terdahulu

Perbandingan hasil penelitian ini dengan penelitian terdahulu yang menggunakan algoritma KNN untuk prediksi kelulusan mahasiswa ditampilkan pada Tabel 7.

Tabel 7. Perbandingan dengan Penelitian Terdahulu

Penelitian	Dataset	Nilai k	Akurasi
Putri & et al. (2023)	200 data	$k=5$	87,00%
Astri, Karman & Daulay (2023)	104 data	$k=5$	83,00%
Junaidi et al. (2024)	200 data	$k=3$	83,33%
Penelitian Ini (2026)	150 data	$k=1$ (optimal)	96,67%

Berdasarkan Tabel 7, penelitian ini menghasilkan akurasi tertinggi sebesar 96,67% dibandingkan dengan penelitian terdahulu yang menggunakan dataset dan konteks serupa. Keunggulan ini dimungkinkan oleh pemilihan nilai k yang optimal secara otomatis, penggunaan StandardScaler yang lebih tepat untuk data dengan skala berbeda, serta penggunaan empat fitur akademik yang lebih komprehensif sebagai variabel prediksi.

4. Kesimpulan

Penelitian ini berhasil membangun sistem prediksi kelulusan mahasiswa Universitas Ibrahimy menggunakan algoritma K-Nearest Neighbor (KNN) berbasis data akademik berupa IPK, SKS, kehadiran, dan jumlah mata kuliah tidak lulus. Hasil penelitian menunjukkan tiga kesimpulan utama.

Pertama, data akademik mahasiswa Universitas Ibrahimy yang mencakup empat fitur (IPK, SKS, kehadiran, dan MK tidak lulus) terbukti relevan dan memadai sebagai dasar prediksi status kelulusan mahasiswa.

Kedua, model KNN dengan nilai $k=1$ yang diperoleh melalui pencarian optimal dari $k=1$ hingga $k=15$ menghasilkan akurasi tertinggi sebesar 96,67% pada 30 data uji, dengan nilai confusion matrix $TP=28$, $TN=1$, $FP=1$, $FN=0$. Hasil ini melampaui akurasi penelitian terdahulu dengan konteks serupa.

Ketiga, implementasi sistem dalam antarmuka web interaktif menggunakan Streamlit berhasil menyajikan hasil prediksi, evaluasi model, dan visualisasi data secara komprehensif sehingga dapat dimanfaatkan oleh pengguna non-teknis seperti dosen dan admin akademik Universitas Ibrahimy dalam pengambilan keputusan akademik berbasis data.

Untuk penelitian selanjutnya, disarankan untuk menggunakan dataset yang lebih besar, menambahkan variabel akademik baru, membandingkan KNN dengan algoritma lain seperti Random Forest dan SVM, serta mengintegrasikan sistem ke dalam SIAKAD resmi Universitas Ibrahimy untuk pemantauan kelulusan mahasiswa secara real-time.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Bapak Achmad Baijuri, M.Kom. dan Bapak Fajriyanto, M.Kom. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan dalam penyusunan penelitian ini, serta kepada pihak Universitas Ibrahimy yang telah memberikan akses data akademik yang digunakan dalam penelitian ini.

Daftar Pustaka

- Astri, J., Karman, J. and Daulay, N.K. (2023) 'Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbor (KNN) pada Fakultas Ilmu Teknik Universitas Bina Insan', *Jurnal Riset Sistem Informasi dan Teknik Informatika*, 8(1), pp. 169–173.
- Junaidi, S., Anggela, R. V and Kariman, D. (2024) 'Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naive Bayes, Random Forest, SVM dan ANN', *Journal of Applied Computer Science and Technology*, 5(1), pp. 109–119. doi:10.52158/jacost.v5i1.489.
- Kartarina, K., Sriwinarti, N.K. and Juniarti, N.L.P. (2021) 'Analisis metode K-Nearest Neighbors (K-NN) dan Naive Bayes dalam memprediksi kelulusan mahasiswa', *Jurnal Teknologi Informasi dan Multimedia (JTIM)*, 3(2), pp. 106–112.
- Novianto, E., Hermawan, A. and Avianto, D. (2023) 'Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1', *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 8(2), pp. 146–154. doi:10.36341/rabit.v8i2.3434.
- Putri, A. et al. (2023) 'Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir', *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), pp. 20–26. doi:10.57152/malcom.v3i1.610.
- Sulthoni, R.I., Muharom, L.A. and Rahman, M. (2023) 'Analisis Tingkat Kepuasan Siswa Dalam Pembelajaran Hybyrid Menggunakan Algoritma K-Nearest Neighbor (KNN)', 4(4), pp. 406–411.
- Tafonao, I.P. and Sibero, A.F.K. (2022) 'Teknik Klasifikasi Prediksi Kelulusan Mahasiswa Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan K-Nearest Neighbors', *Jurnal Mahajana Informasi*, 7(1), pp. 83–90. doi:10.51544/jurnalmi.v7i1.2956.
- Utomo, I.Y.Y.B. and Kurniasari, I. (2023) 'Penerapan Knowledge Discovery in Database', *Jurnal Teknik Informatika Kaputama* [Preprint].