

Klasterisasi Kinerja Akademik Mahasiswa Berdasarkan IPK Menggunakan K-Means dan Metode Elbow

Heni Ermewaningsih^{*1}, Yulius Yus², Fransiska Elly Renni Susanti³

^{1,2,3}Rekayasa Komputer, Institut Teknologi Keling Kumang Sekadau
Email: heni.erne88@gmail.com yulius_yus@itkk.ac.id, fransiskarenni98@gmail.com

Abstrak

Data kinerja akademik mahasiswa sering kali hanya dimanfaatkan sebagai arsip administratif dan belum digunakan secara optimal untuk mengidentifikasi pola kinerja. Penelitian ini bertujuan mengelompokkan kinerja akademik mahasiswa di Institut Teknologi Keling Kumang Sekadau menggunakan algoritma K-Means. Data yang digunakan meliputi IPS semester 1-5, IPK, dan lama studi dari 118 mahasiswa yang telah melalui tahap prapemrosesan untuk memastikan kualitas data. Jumlah kluster optimal ditentukan menggunakan metode Elbow dan menghasilkan tiga kluster. Hasil penelitian menunjukkan bahwa mahasiswa terbagi ke dalam tiga kategori kinerja, yaitu tinggi, sedang, dan rendah. Temuan ini dapat dimanfaatkan sebagai dasar evaluasi akademik dan pengambilan keputusan dalam meningkatkan kualitas pendidikan.

Kata kunci: data mining, k-means, kinerja akademik, klasterisasi, mahasiswa.

Clustering of Students' Academic Performance Based on GPA Using K-Means and the Elbow Method

Abstract

Student academic performance data are often used only as administrative records and are not fully utilized to identify performance patterns. This study aims to cluster students' academic performance at Institut Teknologi Keling Kumang Sekadau using the K-Means algorithm. The data consist of semester Grade Point Averages (GPA) from semesters 1 to 5, Cumulative Grade Point Average (CGPA), and study duration from 118 students. The results are grouped into three levels of academic performance: high, medium, and low, reflecting distinct performance characteristics. These findings provide useful insights for academic evaluation and support decision-making to improve educational quality.

Keywords: clustering, data mining, k-means, academic performance, students.

1. PENDAHULUAN

Kinerja akademik mahasiswa merupakan salah satu indikator utama keberhasilan proses pembelajaran di perguruan tinggi [1]. Informasi terkait kinerja akademik, seperti Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK), umumnya digunakan sebagai tolak ukur untuk mengevaluasi hasil pembelajaran dan kualitas akademik institusi. Namun, pada banyak perguruan tinggi, data akademik masih lebih sering dimanfaatkan sebagai arsip administratif dan belum dianalisis secara optimal untuk mengidentifikasi pola kinerja mahasiswa [5]. Kondisi ini dapat membatasi institusi dalam mengambil keputusan akademik yang strategis dan berbasis data.

Dalam perkembangan analisis data, teknik data mining memungkinkan institusi mengekstraksi pola bermakna dari kumpulan data berukuran besar dan mengubah data akademik menjadi informasi strategis [9]. *Educational Data Mining* (EDM) merupakan bidang yang mengintegrasikan teknik pembelajaran mesin untuk mendukung proses pengambilan keputusan akademik [11]. Melalui pendekatan ini, institusi dapat mengklasifikasikan, memprediksi, serta menganalisis tren kinerja mahasiswa secara lebih sistematis.

Institut Teknologi Keling Kumang Sekadau merupakan perguruan tinggi yang relatif baru berdiri dengan tiga program studi, yaitu Rekayasa Komputer, Kewirausahaan, dan Agroteknologi. Sebagai institusi yang masih dalam tahap pengembangan, pemetaan kinerja akademik mahasiswa secara sistematis menjadi sangat penting untuk mendukung penjaminan mutu dan perencanaan akademik. Perbedaan karakteristik antarprogram studi berpotensi menimbulkan variasi kinerja akademik mahasiswa, sehingga diperlukan analisis yang mampu memberikan gambaran pola tersebut secara komprehensif.

Salah satu teknik dalam data mining yang dapat digunakan untuk tujuan tersebut adalah klusterisasi [4]. Klusterisasi merupakan metode pengelompokan data berdasarkan tingkat kemiripan tanpa memerlukan label kelas sebelumnya, sehingga mampu mengungkap pola tersembunyi dalam data akademik. Dalam konteks pendidikan, teknik ini dapat membantu mengidentifikasi kelompok mahasiswa berdasarkan tingkat kinerja akademik.

Algoritma K-Means merupakan salah satu metode klusterisasi yang banyak digunakan karena kesederhanaan dan efisiensinya dalam mengolah data numerik [4]. Algoritma ini bekerja dengan menentukan pusat klaster (centroid) dan secara iteratif meminimalkan jarak antara data dengan centroid terdekat. Selain itu, metode Elbow sering digunakan untuk menentukan jumlah klaster optimal dengan mengevaluasi nilai variasi dalam klaster.

Beberapa penelitian sebelumnya menunjukkan bahwa K-Means efektif dalam mengelompokkan mahasiswa berdasarkan indikator akademik seperti IPK [8]. Namun, sebagian besar penelitian tersebut dilakukan pada institusi yang telah mapan dengan jumlah data yang besar. Penelitian yang berfokus pada perguruan tinggi yang relatif baru dengan keterbatasan data serta keberagaman program studi masih terbatas.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis dan mengelompokkan kinerja akademik mahasiswa di Institut Teknologi Keling Kumang Sekadau menggunakan algoritma K-Means, dengan jumlah klaster optimal yang ditentukan melalui metode Elbow. Data yang digunakan meliputi IPS semester 1 hingga 5, IPK, dan lama studi mahasiswa.

Kebaruan penelitian ini terletak pada penerapan klusterisasi pada institusi yang relatif baru serta interpretasi akademik terhadap setiap klaster yang dihasilkan. Penelitian ini tidak hanya mengelompokkan data, tetapi juga memberikan analisis terhadap karakteristik masing-masing klaster sebagai dasar evaluasi akademik dan pengambilan keputusan berbasis data.

2. METODE PENELITIAN

Penelitian ini memodelkan permasalahan sebagai proses pengelompokan data mahasiswa $X = \{x_1, x_2, \dots, x_n\}$, di mana setiap data direpresentasikan dalam bentuk vektor numerik yang terdiri dari IPS semester 1 hingga semester 5, IPK, dan lama studi. Tujuan klusterisasi adalah membagi data ke dalam k klaster $C = \{C_1, C_2, \dots, C_k\}$ dengan meminimalkan variasi dalam klaster (*within-cluster variance*) serta meminimalkan perbedaan antarklaster. Optimasi dilakukan dengan meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS). Metode yang digunakan adalah algoritma K-Means dengan penentuan jumlah klaster optimal menggunakan metode Elbow.

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik data mining untuk mengelompokkan mahasiswa berdasarkan kinerja akademik. Algoritma K-Means dipilih karena efektif dalam mengelompokkan data numerik berdasarkan tingkat kemiripan [2] serta memiliki efisiensi komputasi yang baik pada data pendidikan yang terstruktur [15]. Proses penelitian meliputi tahapan pengumpulan data, prapemrosesan, proses klusterisasi, dan evaluasi hasil menggunakan metode Elbow untuk menentukan jumlah klaster optimal.

Penelitian dilakukan pada mahasiswa Institut Teknologi Keling Kumang Sekadau yang memiliki tiga program studi, yaitu Rekayasa Komputer, Kewirausahaan, dan Agroteknologi. Analisis kinerja akademik ini digunakan sebagai dasar evaluasi pembelajaran dan pendukung pengambilan keputusan akademik.

2.2. Sumber dan Jenis Data

Data yang digunakan merupakan data sekunder yang diperoleh dari bagian akademik Institut Teknologi Keling Kumang Sekadau, dengan jumlah sebanyak 118 mahasiswa. Data tersebut mencakup IPS semester 1 hingga semester 5, IPK, serta lama studi mahasiswa. Seluruh data telah dianonimkan untuk menjaga kerahasiaan identitas mahasiswa.

Sebelum analisis, data melalui tahap prapemrosesan yang meliputi pembersihan, validasi, dan normalisasi untuk memastikan kelengkapan serta konsistensi data [6]. Normalisasi dilakukan untuk menghindari perbedaan skala antarvariabel yang dapat memengaruhi perhitungan jarak pada proses klusterisasi [14].

2.3. Variabel Penelitian

Variabel penelitian berupa indikator numerik yang merepresentasikan kinerja akademik mahasiswa, yaitu IPS semester 1-5, IPK, dan lama studi. Indikator ini dipilih karena mampu menggambarkan capaian akademik secara menyeluruh [5]. Rincian variabel disajikan pada Tabel 1.

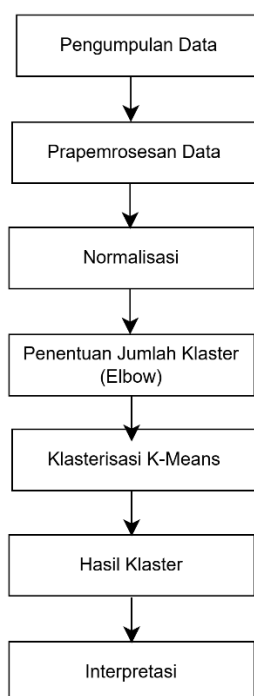
Tabel 1. Variabel Penelitian

Variabel	Tipe Data	Deskripsi
IPS Semester 1	Numerik	Nilai akademik pada semester pertama
IPS Semester 2	Numerik	Nilai akademik pada semester kedua
IPS Semester 3	Numerik	Nilai akademik pada semester ketiga
IPS Semester 4	Numerik	Nilai akademik pada semester keempat
IPS Semester 5	Numerik	Nilai akademik pada semester kelima
IPK	Numerik	Rata-rata nilai akademik kumulatif mahasiswa
Lama Studi	Numerik	Durasi masa studi mahasiswa

IPS per semester digunakan untuk menganalisis tren performa akademik dari waktu ke waktu, sedangkan IPK merepresentasikan capaian kumulatif. Lama studi digunakan untuk menggambarkan progres akademik mahasiswa. Kombinasi indikator tersebut memberikan representasi yang komprehensif dalam proses klusterisasi.

2.4. Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis yang meliputi:



Gambar 1. Diagram Alur Penelitian Metode K-Means dan Elbow

Berdasarkan diagram alur penelitian pada Gambar 1, tahapan penelitian dapat dijelaskan sebagai berikut:

1. Pengumpulan Data
Data akademik dikumpulkan dari bagian akademik Institut Teknologi Keling Kumang Sekadau, meliputi IPS semester 1-5, IPK, dan lama studi.
2. Prapemrosesan Data
Data yang diperoleh melalui tahap pembersihan, validasi, dan normalisasi untuk memastikan kualitas, kelengkapan, serta konsistensi data sebelum dianalisis.
3. Penentuan Jumlah Klaster (Metode Elbow)
Metode Elbow digunakan untuk menentukan jumlah klaster optimal dengan mengevaluasi nilai *inertia* pada berbagai nilai *k*. Titik “siku” pada grafik menunjukkan jumlah klaster yang paling representatif.
4. Proses Klusterisasi (Algoritma K-Means)
Setelah nilai *k* optimal diperoleh, algoritma K-Means diterapkan untuk mengelompokkan data mahasiswa berdasarkan kemiripan nilai akademik.

5. Analisis dan Interpretasi Hasil
Klaster yang terbentuk dianalisis untuk mengidentifikasi karakteristik masing-masing kelompok, seperti kategori kinerja tinggi, sedang, dan rendah.
6. Kesimpulan dan Rekomendasi
Hasil analisis digunakan sebagai dasar evaluasi akademik dan penyusunan strategi pengambilan Keputusan berbasis data di institusi.
Metode Elbow digunakan untuk menentukan jumlah klaster optimal dengan mengevaluasi nilai *inertia* pada berbagai nilai k , sehingga diperoleh jumlah klaster yang paling representatif [4]. Setelah nilai k optimal diperoleh, algoritma K-Means diterapkan untuk mengelompokkan mahasiswa berdasarkan kemiripan nilai IPS semester 1-5, IPK, dan lama studi.
Klaster yang dihasilkan selanjutnya dianalisis untuk menginterpretasikan karakteristik akademik masing-masing kelompok dan memberikan wawasan yang relevan bagi evaluasi institusi.

2.5. Metode Analisis Data

Proses klasterisasi menggunakan algoritma K-Means yang membagi data ke dalam k klaster berdasarkan kedekatan jarak terhadap centroid. Jarak antar data dihitung menggunakan *Euclidean distance* sebagai berikut:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

Dimana x_i merepresentasikan titik data ke- i , c_j merupakan centroid klaster ke- j , dan n adalah jumlah variabel [14].

Tujuan utama algoritma ini adalah meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS):

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2 \quad (2)$$

Algoritma bekerja secara iteratif dengan memperbarui centroid hingga mencapai kondisi konvergen, yaitu ketika tidak terjadi perubahan signifikan pada keanggotaan klaster [4].

Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow, yaitu dengan mengamati titik perubahan signifikan pada grafik *inertia* terhadap jumlah klaster.

2.6. Alat Penelitian

Proses analisis data dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka seperti NumPy, Pandas, dan *scikit-learn*. Pustaka tersebut digunakan untuk prapemrosesan data serta implementasi algoritma K-Means secara efisien [7]. Penggunaan alat ini memungkinkan perhitungan *inertia* dan penerapan algoritma K-Means dilakukan secara sistematis dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengelompokan kinerja akademik mahasiswa menggunakan algoritma K-Means berdasarkan data 118 mahasiswa yang telah melalui tahap prapemrosesan. Analisis dilakukan menggunakan variabel IPS semester 1 hingga 5, IPK, serta lama studi. Hasil klasterisasi tidak hanya menunjukkan pembagian kelompok secara numerik, tetapi juga memberikan interpretasi terhadap karakteristik akademik masing-masing klaster.

3.1. Penentuan Jumlah Klaster Menggunakan Metode Elbow

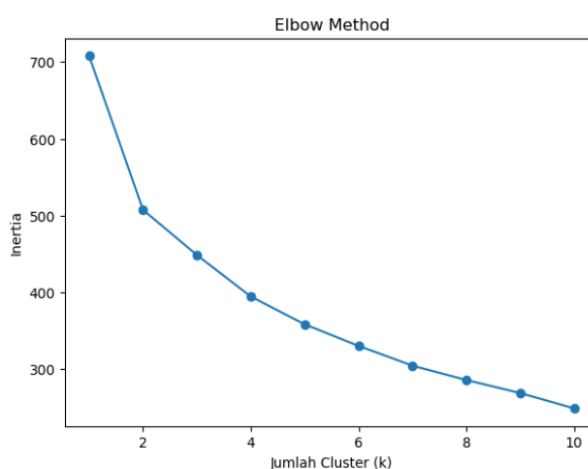
Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow dengan mengevaluasi nilai *inertia* pada berbagai nilai k . Hasil menunjukkan bahwa penurunan *inertia* signifikan terjadi hingga $k = 3$, kemudian melambat pada nilai berikutnya.

Nilai *inertia* untuk setiap k disajikan pada Tabel 2.

Tabel 2. Nilai *Inertia* Menggunakan Metode Elbow

No	Jumlah Klaster (k)	Nilai <i>Inertia</i>
1	1	705
2	2	510
3	3	450
4	4	395
5	5	360
6	6	330
7	7	305
8	8	290
9	9	270
10	10	250

Untuk memperjelas pola tersebut, visualisasi metode Elbow ditunjukkan pada Gambar 2.



Gambar 2. Grafik Metode Elbow untuk Menentukan Jumlah Klaster Optimal

Berdasarkan Gambar 2, terlihat pola “siku” (*elbow*) pada $k = 3$, sehingga jumlah klaster optimal dalam penelitian ini ditetapkan sebanyak tiga klaster.

3.2. Hasil Klasterisasi Mahasiswa Menggunakan Algoritma K-Means

Berdasarkan hasil klasterisasi dengan $k = 3$, diperoleh tiga kelompok mahasiswa. Distribusi jumlah mahasiswa pada setiap klaster disajikan pada Tabel 3.

Tabel 3. Distribusi Jumlah Mahasiswa pada Setiap Klaster.

Klaster	Jumlah Mahasiswa
Klaster 0	54
Klaster 1	37
Klaster 2	27
Total	118

Berdasarkan Tabel 3, klaster 0 memiliki jumlah mahasiswa terbanyak, yaitu 54 mahasiswa, diikuti oleh klaster 1 sebanyak 37 mahasiswa dan klaster 2 sebanyak 27 mahasiswa. Distribusi ini menunjukkan bahwa kinerja akademik mahasiswa tidak homogen, melainkan terbagi ke dalam beberapa kelompok dengan karakteristik yang berbeda.

3.3. Karakteristik Klaster Berdasarkan IPK dan Lama Studi

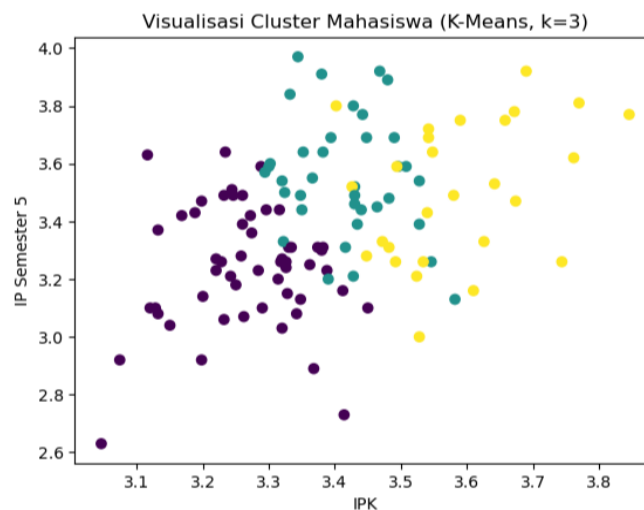
Analisis karakteristik klaster dilakukan berdasarkan rata-rata IPK dan lama studi. Nilai rata-rata setiap klaster disajikan pada Tabel 4.

Tabel 4. Rata-rata IPK dan Lama Studi pada Setiap Klaster

Klaster	Rata-rata IPK	Lama Studi (Semester)	Lama Studi (Tahun)
Klaster 0	3.266	5.0	2.5
Klaster 1	3.416	5.0	2.5
Klaster 2	3.587	5.0	2.5

Berdasarkan Tabel 4, klaster 2 memiliki rata-rata IPK tertinggi, yaitu 3,587, diikuti klaster 1 sebesar 3,416, dan klaster 0 sebesar 3,266. Meskipun rata-rata lama studi relatif serupa, perbedaan nilai IPK menunjukkan adanya variasi tingkat capaian akademik antarklaster. Dengan demikian, mahasiswa dapat dikelompokkan ke dalam kategori kinerja tinggi, sedang, dan rendah.

Untuk memperjelas perbedaan antarklaster, dilakukan visualisasi data yang ditunjukkan pada Gambar 3.



Gambar 3. Visualisasi Klaster Mahasiswa Berdasarkan IPK dan Lama Studi

Berdasarkan Gambar 3, terlihat pemisahan klaster yang cukup jelas, terutama berdasarkan variasi nilai IPK, sehingga mendukung hasil analisis numerik.

3.4. Implikasi Analitis dari Hasil Klasterisasi

Hasil penelitian menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan mahasiswa ke dalam tiga kategori kinerja akademik, yaitu tinggi, sedang, dan rendah. Temuan ini sejalan dengan penelitian sebelumnya dalam bidang *Educational Data Mining* yang menunjukkan bahwa teknik klasterisasi mampu mengungkap pola akademik secara eksploratif.

Secara praktis, hasil ini dapat dimanfaatkan oleh institusi untuk merancang strategi akademik yang lebih terarah, seperti program pendampingan bagi mahasiswa dengan kinerja rendah serta pengembangan bagi mahasiswa berprestasi tinggi. Pendekatan ini mendukung pengambilan keputusan berbasis data dalam pengelolaan pendidikan tinggi.

Namun, penelitian ini memiliki keterbatasan pada jumlah data yang relatif terbatas dan hanya mencakup satu institusi, sehingga hasilnya belum dapat digeneralisasikan secara luas. Selain itu, keseragaman lama studi antarklaster menunjukkan bahwa variabel ini kurang berkontribusi secara signifikan dalam membedakan kelompok.

Penelitian selanjutnya disarankan untuk menggunakan data yang lebih besar serta mempertimbangkan variabel tambahan atau metode klasterisasi lain guna meningkatkan akurasi segmentasi.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan kinerja akademik mahasiswa ke dalam tiga klaster dengan karakteristik yang berbeda. Hasil klasterisasi menunjukkan adanya perbedaan rata-rata IPK, di mana klaster 2 memiliki kinerja akademik tertinggi, diikuti oleh klaster 1 dan klaster 0. Selain itu, distribusi mahasiswa yang tidak merata menunjukkan bahwa kinerja akademik tidak bersifat

homogen. Temuan ini dapat dimanfaatkan sebagai dasar evaluasi akademik serta perumusan strategi pendampingan dan kebijakan berbasis data di perguruan tinggi.

DAFTAR PUSTAKA

- [1] M. A. Rahman, N. Islam, and A. Rahman, "Academic Performance Analysis Using Clustering Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 421–428, 2020.
- [2] A. A. Nugroho, R. R. Isnanto, and K. T. Martono, "Clustering Mahasiswa Berdasarkan Prestasi Akademik Menggunakan Metode K-Means," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 2, pp. 85–92, 2021.
- [3] R. A. Pratama, D. Puspitasari, and H. Tolle, "Analisis Prestasi Akademik Mahasiswa Menggunakan Teknik Clustering," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1451–1458, 2022.
- [4] A. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [5] R. K. Jain and V. Dubes, "Applications of Clustering Methods in Educational Data Mining," *Journal of Educational Data Mining*, vol. 12, pp. 1–18, 2020.
- [6] S. García, J. Luengo, and F. Herrera, "A Survey on Data Preprocessing in Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 294–313, 2015.
- [7] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [8] N. A. Rahman, S. S. Salleh, and R. Mohamed, "Student Academic Performance Clustering Using K-Means Algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 2, pp. 1051–1059, 2021.
- [9] C. Azzahra and S. Sriani, "Clustering of High School Students' Academic Scores Using K-Means Algorithm," *Journal of Information System and Informatics*, vol. 7, no. 1, pp. 572–586, 2025.
- [10] S. Khodijah, A. R. Hariyanto, B. Salsabiilah, W. S. Aulia, and M. Fanyusri, "Student Clustering Based on Academic Performance Using the K-Means Algorithm in RapidMiner," *Journal of Information Technology and Informatics Engineering*, vol. 1, no. 1, pp. 46–50, 2025.
- [11] S. Durachman and A. W. Bin Abdul Rahman, "Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement," *Artificial Intelligence in Learning*, vol. 1, no. 1, pp. 35–53, 2025.
- [12] D. Kurniadi and A. Sugiyono, "Pengelompokan Data Akademik Menggunakan Algoritma K-Means pada Data Akademik Unissula," *Jurnal Transformatika*, vol. 18, no. 1, 2018.
- [13] M. A. Al Fauzie, Yuliadi, and J. A. Putra, "Clustering Data Menggunakan Metode K-Means untuk Rekomendasi Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, 2023.
- [14] N. Rahayu, A. H. Anshor, and L. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *JUKI: Jurnal Komputer dan Informatika*, vol. 6, no. 1, 2024.
- [15] H. A. Almohamad, M. M. Alshammari, and A. A. Alshammari, "Student Performance Analysis Using K-Means Clustering Algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 350–357, 2023.