Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

Pola Prestasi Akademik Mahasiswa Informatika Menggunakan K-Means Clustering Studi Kasus Universitas Hamzanwadi NTB

Amanah Azzahra[1*], Kusrini[2]
[1,2]Program Studi Informatika, Universitas Amikom Yogyakarta
*amanah.azzahra@students.amikom.ac.id

Abstrak

Analisis pola prestasi akademik mahasiswa penting untuk mendukung pengambilan keputusan akademik berbasis data. Penelitian ini bertujuan mengidentifikasi pola prestasi akademik mahasiswa Program Studi Informatika Universitas Hamzanwadi menggunakan pendekatan unsupervised learning berbasis algoritma K-Means. Kontribusi penelitian ini terletak pada penerapan eksperimen skenario beberapa kombinasi atribut akademik serta evaluasi multi-metrik untuk menentukan konfigurasi clustering paling optimal pada data mahasiswa semester tujuh. Dataset mencakup Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), dan total Satuan Kredit Semester (SKS). Setelah proses pembersihan dan pra-pemrosesan, diperoleh 382 data mahasiswa untuk dianalisis. Normalisasi dilakukan menggunakan Min–Max Scaling. Penentuan jumlah cluster optimal menggunakan metode Elbow menghasilkan tiga cluster. Evaluasi kualitas clustering menunjukkan Silhouette Score 0,71, Davies–Bouldin Index 0,54, dan Calinski–Harabasz Index 808,64 yang mengindikasikan kualitas cluster yang baik. Hasil menunjukkan tiga kelompok mahasiswa dengan karakteristik prestasi akademik tinggi, sedang, dan rendah. Temuan ini membuktikan K-Means efektif dalam mengelompokkan pola prestasi akademik mahasiswa dan dapat dimanfaatkan sebagai dasar evaluasi serta strategi pembinaan akademik. Penelitian ini terbatas pada data semester tujuh dan dapat dikembangkan pada dataset lain atau lintas program studi.

Kata kunci : Clustering, Data Mining, K-Means, Mahasiswa Informatika, Prestasi Akademik

Abstract

Analyzing students' academic achievement patterns is important to support data-driven academic decision-making. This study aims to identify academic achievement patterns of Informatics students at Universitas Hamzanwadi using an unsupervised learning approach based on the K-Means algorithm. The main contribution of this research lies in conducting scenario-based experiments using several combinations of academic attributes and applying multi-metric evaluation to determine the most optimal clustering configuration for seventh-semester student data. The dataset includes the Semester Grade Point Average (IPS), Cumulative Grade Point Average (IPK), and the total accumulated credit units (SKS). After data cleaning and preprocessing, 382 student records were obtained for analysis. Data normalization was performed using Min–Max Scaling. The optimal number of clusters was determined using the Elbow method, resulting in three clusters. Cluster quality evaluation produced a Silhouette Score of 0.71, a Davies–Bouldin Index of 0.54, and a Calinski–Harabasz Index of 808.64, indicating good clustering quality. The results revealed three groups of students with high, medium, and low academic achievement characteristics. These findings confirm that the K-Means algorithm is effective in clustering student academic achievement patterns and can be used as a basis for academic evaluation and student development strategies. However, this study is limited to seventh-semester data and can be extended to other datasets or across study programs.

Keywords : *Clustering, Data Mining, K-Means, Informatics Students, Academic Achievement.*

1. Pendahuluan

Prestasi akademik mahasiswa merupakan indikator penting keberhasilan pembelajaran di perguruan tinggi. Capaian akademik yang tercermin melalui Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), dan akumulasi

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

Satuan Kredit Semester (SKS) tidak hanya merepresentasikan kemampuan mahasiswa, tetapi juga menjadi dasar pengambilan keputusan akademik seperti evaluasi kurikulum, perancangan strategi pembelajaran, dan pembinaan mahasiswa. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengolah data akademik secara sistematis dan objektif. Seiring perkembangan teknologi informasi, Educational Data Mining (EDM) banyak dimanfaatkan untuk mengekstraksi pengetahuan dari data pendidikan guna mendukung peningkatan kualitas pembelajaran dan pengambilan keputusan berbasis data[1], [2]. Salah satu teknik utama dalam EDM adalah clustering, karena dapat mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label kelas[1]. Berbagai penelitian menunjukkan bahwa algoritma K-Means efektif untuk analisis data akademik. K-Means digunakan untuk mengelompokkan performa mahasiswa dan mengidentifikasi pola prestasi yang berbeda antar kelompok[3], serta mendukung evaluasi pendidikan melalui pengelompokan capaian belajar[4]. Pendekatan clustering juga dinilai mampu memberikan gambaran yang lebih komprehensif mengenai pencapaian akademik mahasiswa di perguruan tinggi[5].

Pada studi komparatif, K-Means dilaporkan unggul dari Hierarchical Clustering dan DBSCAN dalam kesederhanaan implementasi, efisiensi komputasi, dan kemudahan interpretasi, terutama pada data numerik berskala besar[6],[7]. Selain itu, clustering juga terbukti efektif untuk mengidentifikasi pola perilaku belajar mahasiswa yang kompleks dan dapat dimanfaatkan sebagai evaluasi otomatis yang adaptif[8],[9]. Pada konteks nasional, penerapan K-Means juga menunjukkan fleksibilitas untuk pengelompokan data manusia dan pendidikan dalam berbagai studi pada Jurnal Infotek[10].

Namun demikian, di Program Studi Informatika Universitas Hamzanwadi, analisis prestasi akademik mahasiswa masih didominasi evaluasi deskriptif tanpa pemodelan berbasis data mining. Pola capaian akademik mahasiswa, khususnya pada semester tingkat akhir, belum dianalisis secara sistematis menggunakan teknik clustering sehingga pemanfaatan data akademik untuk mendukung keputusan program studi belum optimal. Kontribusi utama penelitian ini adalah penerapan eksperimen skenario beberapa kombinasi atribut akademik serta evaluasi multi-metrik (Silhouette Score, Davies–Bouldin Index, dan Calinski–Harabasz Index) untuk memperoleh konfigurasi clustering yang paling optimal pada mahasiswa semester tujuh.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengidentifikasi pola prestasi akademik mahasiswa Program Studi Informatika Universitas Hamzanwadi menggunakan algoritma K-Means Clustering. Analisis dilakukan

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

menggunakan data IPS, IPK, dan total SKS pada semester tujuh, serta dievaluasi menggunakan Silhouette Score, Davies–Bouldin Index, dan Calinski–Harabasz Index untuk memastikan kualitas cluster yang dihasilkan. Hasil penelitian diharapkan dapat menjadi dasar evaluasi dan penyusunan strategi peningkatan prestasi akademik di lingkungan program studi.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Penelitian sebelumnya yang menggunakan algoritma K-Means Clustering dalam pengolahan data pendidikan dan akademik yang relevan dengan penelitian ini antara lain sebagai berikut:

1. Penelitian yang dilakukan oleh Vankayalapati et al. dengan judul *"K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques"* bertujuan mengelompokkan tingkat performa peserta didik berdasarkan data akademik menggunakan K-Means. Hasilnya menunjukkan bahwa K-Means mampu membentuk cluster performa belajar yang berbeda antar kelompok sehingga dapat mendukung pengambilan keputusan akademik[3].
2. Penelitian selanjutnya dilakukan oleh Cahapin et al. dengan judul *"Clustering of Students Admission Data Using K-Means, Hierarchical, and DBSCAN Algorithms"*. Penelitian ini membandingkan K-Means, Hierarchical Clustering, dan DBSCAN pada data mahasiswa. Hasil penelitian menyimpulkan bahwa K-Means lebih efisien, mudah diimplementasikan, dan menghasilkan cluster yang relatif mudah diinterpretasikan pada data numerik berskala besar[6].
3. Penelitian oleh Hooshyar et al. yang berjudul *"Clustering Algorithms in an Educational Context: An Automatic Comparative Approach"* mengkaji penerapan clustering untuk evaluasi pendidikan. Hasilnya menegaskan bahwa clustering efektif digunakan sebagai evaluasi otomatis untuk mengungkap pola capaian belajar yang tidak terlihat melalui analisis konvensional[9].
4. Selanjutnya adalah Liu dalam penelitiannya yang berjudul *"Data Analysis of Educational Evaluation Using K-Means Clustering Method"* menerapkan algoritma K-Means untuk mengevaluasi capaian akademik dalam lingkungan pendidikan. Penelitian ini menunjukkan bahwa pengelompokan mahasiswa berdasarkan indikator akademik menggunakan K-Means dapat memberikan gambaran yang jelas mengenai tingkat prestasi mahasiswa, serta membantu institusi pendidikan dalam melakukan evaluasi dan perencanaan akademik[4].
5. Penelitian yang dilakukan oleh Wang dengan judul *"Higher Education Management and Student Achievement Assessment Method Based on Clustering Algorithm"* berfokus pada penerapan teknik clustering dalam menilai pencapaian akademik mahasiswa di perguruan tinggi. Hasil

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

penelitian menunjukkan bahwa clustering mampu memberikan pemetaan prestasi mahasiswa secara komprehensif, sehingga dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan manajemen pendidikan[5].

6. Berikut adalah Li et al. melakukan penelitian berjudul *"An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns"*. Penelitian ini menerapkan pendekatan clustering untuk mengidentifikasi pola perilaku mahasiswa. Hasil penelitian menunjukkan bahwa teknik clustering mampu mengungkap karakteristik belajar mahasiswa yang beragam, yang tidak dapat diidentifikasi melalui metode evaluasi tradisional[8].

7. Penelitian yang dilakukan oleh Suhartini dan Ria Yuliani yang berjudul "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K Means di Dusun Bagik Endep Sukamulia Timur". Penelitian ini bertujuan untuk mengelompokkan data penduduk di wilayah Sukamulia Timur yang memang dikatakan tergolong penduduk miskin. Berdasarkan hasil pengujian yang dilakukan dengan menerapkan algoritma K-Means didapatkan hasil dengan Cluster 1 berjumlah 18 penduduk dengan kriteria Penduduk ekonomi tinggi, Cluster 2 berjumlah 72 Penduduk dengan kriteria Penduduk ekonomi sedang, dan Cluster 3 berjumlah 110 penduduk dengan kriteria Penduduk ekonomi rendah[11].

2.2. Landasan Teori

1. Data Mining

Data mining merupakan proses penggalian informasi dan pola tersembunyi dari kumpulan data berukuran besar menggunakan teknik statistik, matematika, dan pembelajaran mesin. Dalam bidang pendidikan, data mining digunakan untuk menganalisis data akademik mahasiswa untuk memahami pola capaian, performa, dan karakteristik kelompok tertentu. Salah satu pendekatan penting dalam data mining adalah unsupervised learning, yang berfokus pada pembentukan pola kelompok tanpa label kelas[1].

2. Educational Data Mining

Educational Data Mining (EDM) merupakan cabang data mining yang fokus pada analisis data dalam lingkungan pendidikan untuk memahami perilaku belajar, mengevaluasi performa akademik, serta mendukung pengambilan keputusan berbasis data di institusi pendidikan[1]. Data akademik seperti nilai, indeks prestasi, dan beban studi dapat dimanfaatkan untuk menemukan pola capaian belajar yang relevan bagi perencanaan akademik dan peningkatan mutu pendidikan[9], [5].

3. Clustering

Clustering adalah teknik unsupervised learning yang mengelompokkan objek data ke dalam beberapa cluster berdasarkan kemiripan, di mana objek dalam satu cluster memiliki karakteristik yang lebih dekat dibandingkan dengan objek di

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

cluster lain[9]. Dalam pendidikan, clustering banyak digunakan untuk memetakan mahasiswa berdasarkan performa akademik dan perilaku belajar sehingga membantu institusi memahami variasi capaian secara lebih terstruktur[3], [4].

4. Algoritma K-Means

K-Means merupakan salah algoritma clustering yang populer karena sederhana dan efisien. Algoritma ini membagi data ke dalam sejumlah K cluster berdasarkan kedekatan terhadap pusat cluster (centroid). Tahapannya mencakup penentuan jumlah cluster, inisialisasi centroid, perhitungan jarak data ke centroid, dan pembaruan centroid secara iteratif hingga konvergen[7]. K-Means banyak digunakan dalam pengelompokan mahasiswa berdasarkan capaian akademik[3], [5]. dan terbukti fleksibel untuk berbagai domain lain seperti sistem rekomendasi dan analisis aktivitas manusia[12],[13].

5. Penentuan Jumlah Cluster

Jumlah cluster merupakan parameter penting karena memengaruhi kualitas hasil clustering. Salah satu metode yang umum digunakan adalah Elbow Method, yaitu memilih titik siku berdasarkan perubahan nilai inertia ketika penambahan jumlah cluster tidak lagi menurunkan inertia secara signifikan[14].

Beberapa penelitian menunjukkan bahwa kombinasi K-Means dan Elbow Method meningkatkan interpretabilitas hasil clustering prestasi akademik dan menghasilkan pemetaan kelompok yang lebih optimal[15].

6. Validasi Cluster

Karena clustering tidak memiliki label kebenaran, kualitas cluster dievaluasi menggunakan indeks internal seperti Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI)[6]. Silhouette Score menilai kohesi dan separasi cluster (nilai lebih tinggi lebih baik), DBI menilai kedekatan antar cluster (nilai lebih kecil lebih baik), sedangkan CHI menilai pemisahan cluster berdasarkan rasio variasi antar cluster dan dalam cluster (nilai lebih besar lebih baik)[4], [5].

2.1. Tahapan Penelitian

Tahapan-tahapan penelitian dapat dilihat dari alur proses pada gambar 1 dibawah ini:

Gambar 1. Alur Penelitian

Tahapan dari alur proses penelitian diatas yaitu:

1. Penelitian diawali dengan pengumpulan dataset akademik mahasiswa Program Studi Informatika yang diperoleh dari sistem akademik perguruan tinggi dalam format *.xls*.
2. Tahap persiapan data (Data Preparation) yaitu dengan merubah dataset yang masih mentah

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

menjadi sebuah dataset yang siap untuk dianalisis.

3. Setelah data siap dianalisis, dilakukan penentuan jumlah cluster optimal menggunakan metode Elbow.
4. Data yang sudah siap dapat diproses menggunakan bahas pemrograman python pada Google Colab dengan algoritma K-Means.
5. Hasil clustering dievaluasi menggunakan Silhouette Score, Davies–Bouldin Index, dan Calinski–Harabasz Index, lalu dilakukan analisis untuk menginterpretasikan pola cluster yang terbentuk

3. Metode Penelitian

Penelitian ini menggunakan metode eksperimental dengan pendekatan kuantitatif untuk mengidentifikasi pola prestasi akademik mahasiswa melalui proses clustering. Pendekatan ini dipilih karena mampu mengeksplorasi struktur pola dalam data akademik tanpa memerlukan label kelas, sehingga sesuai untuk mengelompokkan mahasiswa berdasarkan kemiripan capaian akademik.

3.1. Teknik Pengumpulan Data

Data yang digunakan berupa data akademik mahasiswa Program Studi Informatika yang bersifat pribadi. Teknik pengumpulan data dilakukan melalui:

1. Teknik Observasi, yaitu ekstraksi data akademik dari sistem informasi akademik perguruan tinggi.
2. Teknik Dokumentasi, yaitu penggunaan data dalam bentuk dokumen digital *.xls*. sebagai dasar analisis clustering.

Penggunaan data dilakukan dengan izin institusi/program studi untuk kepentingan penelitian akademik. Untuk menjaga kerahasiaan, data diproses secara anonim tanpa mencantumkan identitas personal mahasiswa, dan hanya atribut akademik yang relevan yang digunakan. Dataset awal berjumlah 415 mahasiswa, dan setelah pembersihan data digunakan 382 mahasiswa. Ringkasan jumlah data ditunjukkan pada Tabel 1.

Tabel 1. Jumlah Data

No	Tahap	Jumlah Data
1	Data awal	415 mahasiswa
2	Data setelah dibersihan	382 mahasiswa

3.2. Teknik Analisis Data

Analisis data dilakukan secara kuantitatif menggunakan pendekatan data mining, khususnya unsupervised learning pada tahapan clustering. Analisis bertujuan mengelompokkan mahasiswa berdasarkan kemiripan nilai akademik untuk menemukan struktur kelompok tanpa variabel target.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

3.3. Teknik Pengolahan Data

Pengolahan data dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab dengan algoritma K-Means. Tahapan pengolahan data meliputi:

1. Pembersihan Data (Data Cleaning), menghapus data yang tidak valid atau tidak lengkap, misalnya nilai IPK atau Total SKS bernilai nol, serta nilai IPS kosong.
2. Seleksi Atribut, yaitu IPS, IPK, dan Total SKS sebagai representasi capaian akademik. Deskripsi atribut ditunjukkan pada Tabel 2.

Tabel 2. Atribut Data

No	Atribut	Keterangan
1	IPS	Indeks Prestasi Semester
2	IPK	Indeks Prestasi Kumulatif
3	Total SKS	Jumlah SKS kumulatif

3. Normalisasi Data, menggunakan Min–Max Scaling untuk menyetarakan rentang nilai antar atribut karena K-Means sensitif terhadap skala.

3.4. Model Yang Diusulkan

Model yang diusulkan adalah K-Means Clustering. Penentuan jumlah cluster dilakukan menggunakan Elbow Method dengan menguji rentang k = 2 hingga k = 10 berdasarkan perubahan nilai inertia, lalu dipilih pada titik siku ketika penurunan inertia mulai melambat signifikan. Proses clustering dijalankan menggunakan parameter k-means++ sebagai inisialisasi centroid, random_state = 42, n_init = 10, serta max_iter = 300 untuk memastikan hasil yang stabil.

Evaluasi kualitas cluster dilakukan menggunakan Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Silhouette Score menilai kohesi dan separasi cluster, DBI menilai kedekatan antar cluster (lebih kecil lebih baik), sedangkan CHI menilai pemisahan cluster berdasarkan rasio variasi antar cluster dan dalam cluster (lebih besar lebih baik).

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Bagian ini menyajikan hasil penerapan algoritma K-Means Clustering untuk mengidentifikasi pola prestasi akademik mahasiswa Program Studi Informatika Universitas Hamzanwadi Nusa Tenggara Barat. Analisis dilakukan menggunakan data akademik mahasiswa semester tujuh yang telah melalui tahap pembersihan dan normalisasi data.

1. Deskripsi Data dan Pra-pemrosesan

Penelitian menggunakan data akademik mahasiswa semester tujuh Program Studi Informatika Universitas Hamzanwadi. Proses clustering dilakukan berdasarkan tiga atribut utama, yaitu IPS, IPK, dan Total SKS, karena mewakili capaian akademik semesteran dan kumulatif. Pra-pemrosesan melalui konversi tipe numerik, penghapusan data tidak valid, serta

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

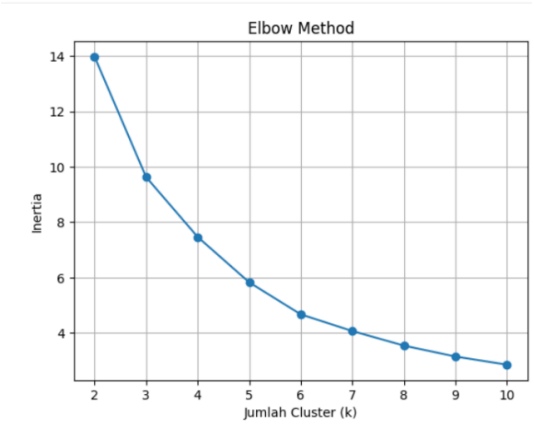
e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

normalisasi menggunakan Min–Max Scaling agar rentang nilai antar atribut seragam, mengingat K-Means berbasis perhitungan jarak dan sensitif terhadap perbedaan skala.

2. Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan mengamati nilai *inertia* pada rentang k = 2 hingga k = 10.



Gambar 2. Visualisasi Elbow Method

Gambar 2 menampilkan hasil visualisasi Elbow Method. Berdasarkan grafik tersebut terlihat titik siku (elbow) pada k = 3, sehingga jumlah cluster optimal yang digunakan dalam penelitian ini ditetapkan sebanyak tiga cluster.

3. Hasil Clustering Algoritma K-Means

Proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster k = 3sehingga setiap mahasiswa dikelompokkan berdasarkan kemiripan nilai IPS, IPK, dan Total SKS. Distribusi jumlah anggota tiap cluster ditunjukkan pada Tabel 3.

Tabel 3. Hasil Clustering

No	Cluster	Jumlah Mahasiswa
1	Cluster 0	299
2	Cluster 1	71
3	Cluster 2	12
4	Total	382

Untuk memperoleh gambaran karakteristik tiap cluster, dilakukan perhitungan nilai rata-rata IPS, IPK, dan Total SKS pada masing-masing cluster. Hasil ringkasan statistik ditunjukkan pada Tabel 4.

Tabel 4. Rata-Rata Atribut Prestasi Akademik per Cluster

No	Cluster	IPS	IPK	Total SKS
1	Cluster 0	3.346	3.589	141.231
2	Cluster 1	2.768	3.553	25.831
3	Cluster 2	2.730	1.479	123.000

4. Evaluasi

Evaluasi kualitas clustering dilakukan menggunakan metrik evaluasi internal, yaitu Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Hasil evaluasi ditunjukkan pada tabel 5.

Tabel 5. Hasil Evaluasi Kualitas Cluster

No	Metode Evaluasi	Nilai
1	Silhouette Score	0.71
2	Davies-Bouldin Index	0.54
3	Calinski-Harabasz Index	808.64

Nilai Silhouette Score sebesar 0.71 menunjukkan struktur cluster yang terbentuk memiliki tingkat kohesi dan separasi yang baik. Nilai DBI sebesar 0.54 mengindikasikan jarak antar cluster relatif jelas dengan sebaran intra-cluster yang cukup

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

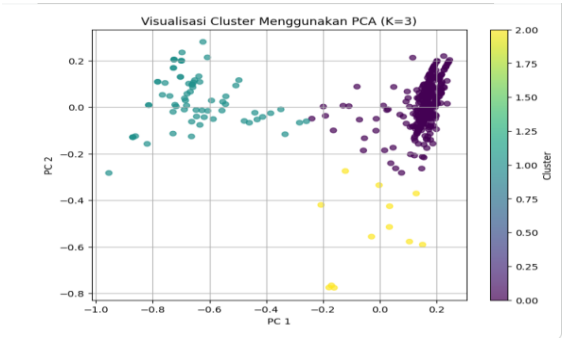
e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

kecil. Selain itu, nilai CHI sebesar 808.64 juga menunjukkan pemisahan cluster yang cukup signifikan untuk mendukung interpretasi akademik.

5. Visualisasi Cluster Menggunakan PCA

Untuk membantu interpretasi hasil clustering, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA) menjadi dua komponen utama. Hasil PCA menunjukkan bahwa dua komponen utama mampu menjelaskan sekitar 93,63% variasi data, sehingga visualisasi dua dimensi dianggap representatif ditunjukkan pada Gambar 3.



Gambar 3. Visualisasi Hasil Clustering Menggunakan PCA

Berdasarkan visualisasi tersebut, tiga cluster terbentuk secara relatif terpisah, sehingga mendukung hasil evaluasi kualitas cluster.

6. Skenario Pemilihan Model Clustering Terbaik

Untuk memperoleh konfigurasi clustering dengan performa terbaik, dilakukan eksperimen tambahan menggunakan beberapa kombinasi atribut akademik dan variasi jumlah cluster. Kombinasi yang diuji meliputi IPS dan IPK, IPK dan Total SKS, serta IPS, IPK, dan Total SKS, dengan variasi jumlah cluster k = 2 hingga k = 5. Evaluasi dilakukan menggunakan Silhouette Score, DBI, dan CHI. Ringkasan hasil ditunjukkan pada Tabel 6.

Tabel 6. Perbandingan Performa Model Clustering Berdasarkan Skenario

No	Kombinasi Atribut	K	Silhouette	DBI	CHI
1	IPK, Total SKS	2	0.81	0.26	1525.15
2	IPK, Total SKS	3	0.79	0.44	1525.19
3	IPS, IPK, Total SKS	2	0.74	0.42	999.15
4	IPS, IPK, Total SKS	3	0.71	0.54	808.64

Berdasarkan Tabel 6, kombinasi IPK dan Total SKS dengan k = 2 memberikan performa terbaik secara numerik karena menghasilkan Silhouette Score tertinggi (0,81) dan DBI terendah (0,26). Namun demikian, model utama penelitian ini tetap menggunakan kombinasi atribut IPS, IPK, dan Total SKS dengan k = 3 karena memberikan representasi capaian akademik yang lebih komprehensif serta menghasilkan pembagian kelompok yang lebih informatif untuk kebutuhan pembinaan akademik

4.2 Pembahasan

Pembahasan ini menguraikan interpretasi karakteristik setiap cluster, relevansi hasil evaluasi kualitas cluster, serta perbandingan temuan penelitian dengan studi terdahulu.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

1. Interpretasi Karakteristik Cluster

Hasil clustering menunjukkan terbentuknya tiga kelompok mahasiswa dengan karakteristik capaian akademik yang berbeda. Cluster 0 merupakan kelompok terbesar yang merepresentasikan capaian akademik tinggi karena memiliki rata-rata IPS dan IPK yang relatif tinggi serta Total SKS mendekati beban studi akhir. Cluster 1 menunjukkan capaian akademik sedang, ditandai oleh IPK yang masih cukup baik namun Total SKS jauh lebih rendah dibanding cluster lainnya, sehingga mengindikasikan progres studi yang belum optimal atau belum menuntaskan beban studi. Sementara itu, Cluster 2 merepresentasikan capaian akademik rendah dengan nilai IPK paling rendah dibandingkan cluster lain, meskipun Total SKS relatif tinggi, yang dapat mengarah pada kondisi mahasiswa dengan progres studi berjalan tetapi performa akademik kumulatif belum baik.

2. Pembahasan Kualitas Cluster Berdasarkan Metrik Evaluasi

Nilai Silhouette Score sebesar 0.71 menunjukkan kualitas clustering berada pada kategori baik karena mencerminkan pemisahan antar cluster yang jelas dan tingkat kemiripan tinggi di dalam cluster. Nilai DBI 0,54 juga menunjukkan cluster relatif optimal karena jarak antar cluster cukup baik (nilai lebih kecil lebih baik). Selain itu, nilai CHI 808,64 mengindikasikan variasi antar cluster yang lebih besar dibanding variasi di dalam cluster, sehingga struktur pengelompokan dinilai cukup stabil dan layak diinterpretasikan untuk konteks evaluasi akademik.

3. Perbandingan dengan Penelitian Terdahulu

Temuan penelitian ini sejalan dengan penelitian Vankayalapati et al. yang menunjukkan bahwa K-Means mampu memetakan kelompok performa akademik berdasarkan kemiripan capaian belajar sehingga menghasilkan pola prestasi yang berbeda antar kelompok mahasiswa[3]. Hasil ini juga konsisten dengan Liu yang menegaskan bahwa penerapan K-Means dalam evaluasi pendidikan dapat memberikan gambaran variasi capaian akademik secara lebih terstruktur[4]. Selanjutnya, Wang menyatakan bahwa clustering dapat mendukung pemetaan prestasi mahasiswa yang lebih komprehensif untuk pengambilan keputusan manajemen pendidikan[5]. Penggunaan evaluasi multi-metrik pada penelitian ini juga mendukung pendekatan Cahapin et al. yang menekankan keunggulan K-Means dalam kemudahan interpretasi pada data akademik numerik[6], sekaligus memperkuat kontribusi penelitian melalui eksperimen skenario untuk memilih konfigurasi clustering yang paling optimal dan relevan secara akademik

5. Kesimpulan

Penelitian ini menerapkan K-Means Clustering untuk mengelompokkan pola prestasi akademik

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

mahasiswa Informatika Universitas Hamzanwadi berdasarkan IPS, IPK, dan Total SKS. Penentuan jumlah cluster menggunakan Elbow Method menghasilkan tiga cluster dengan karakteristik capaian akademik yang berbeda. Kualitas clustering menunjukkan hasil yang baik dengan Silhouette Score 0,71, Davies–Bouldin Index 0,54, dan Calinski–Harabasz Index 808,64. Kontribusi penelitian ini terletak pada eksperimen skenario kombinasi atribut serta evaluasi multi-metrik untuk memilih konfigurasi clustering yang optimal dan relevan secara interpretatif. Hasilnya dapat dimanfaatkan sebagai dasar pemetaan capaian mahasiswa dan penyusunan strategi pembinaan akademik. Penelitian selanjutnya disarankan membandingkan metode lain (misalnya Hierarchical Clustering atau DBSCAN) serta menambahkan variabel non-akademik agar profil cluster lebih komprehensif.

6. Daftar Pustaka

[1] C. Baek and T. Doleck, “Educational Data Mining: A Bibliometric Analysis of an Emerging Field,” *IEEE Access*, vol. 10, pp. 31289–31296, 2022, doi: 10.1109/ACCESS.2022.3160457.

[2] G. Czibula, G. Ciubotariu, M. I. Maier, and H. Lisei, “IntelliDaM: A Machine Learning-Based Framework for Enhancing the Performance of Decision-Making Processes. A Case Study for Educational Data Mining,” *IEEE Access*, vol. 10, no. August, pp. 80651–80666, 2022, doi: 10.1109/ACCESS.2022.3195531.

[3] R. Vankayalapati, K. B. Ghutugade, R. Vannapuram, and B. P. S. Prasanna, “K-means algorithm for clustering of learners performance levels using machine learning techniques,” *Rev. d’Intelligence Artif.*, vol. 35, no. 1, pp. 99–104, 2021, doi: 10.18280/ria.350112.

[4] R. Liu, “Data Analysis of Educational Evaluation Using K-Means Clustering Method,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3762431.

[5] Z. Wang, “Higher Education Management and Student Achievement Assessment Method Based on Clustering Algorithm,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/4703975.

[6] E. L. Cahapin, B. A. Malabag, C. S. Santiago, J. L. Reyes, G. S. Legaspi, and K. L. Adrales, “Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 6, pp. 3647–3656, 2023, doi: 10.11591/eei.v12i6.4849.

[7] B. et al. Karthikeyan, “A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering,” *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 1600–1604, 2020, doi: 10.30534/ijeter/2020/20852020.

[8] X. Li, Y. Zhang, H. Cheng, F. Zhou, and B. Yin, “An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns,” *IEEE Access*, vol. 9, pp. 7076–7091, 2021, doi: 10.1109/ACCESS.2021.3049157.

[9] D. Hooshyar, Y. Yang, M. Pedaste, and Y. M. Huang, “Clustering Algorithms in an Educational Context: An Automatic Comparative Approach,” *IEEE Access*, vol. 8, pp. 146994–147014, 2020, doi: 10.1109/ACCESS.2020.3014948.

[10] V. No, A. R. Nabella, H. Z. Zahro, and Y. A. Pranoto, “Rancang Bangun Sistem TOEFL Untuk Analisis Kelemahan Peserta Dengan Penerapan Algoritma K-Means Clustering,” vol. 8, no. 1, 2025.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.351- 362

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.33944 Link : https://dx.doi.org/10.29408/jit.v9i2.33944

[11] Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K- Means di Dusun Bagik Endep Sukamulia Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021.

[12] A. muliawan Nur, M. Saiful2, H. Bahtiar, and Muhammad Taufik Hidayat, "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 478–488, 2024, doi: 10.29408/jit.v7i2.26283.

[13] N. D. Salsabila, K. Aulisari, and H. Z. Zahro, "Penerapan Algoritma K-Means Untuk Klasterisasi Produktivitas Tanaman Jahe," vol. 8, no. 1, pp. 228–238, 2025.

[14] M. Qusyairi, Z. Hidayatullah, and A. Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 500–510, 2024.

[15] M. M. K-means and R. Ishak, "Clustering Prestasi Akademik Lulusan," *Jambura J. Electr. …*, vol. 6, no. 1, pp. 76–81, 2024, [Online]. Available:https://ejurnal.ung.ac.id/index.php/jjeee/article/view/23967%0Ahttps://ejurnal.ung.ac.id/index.php/jjeee/article/download/23967/8053.