

PENGELOMPOKAN CAPAIAN PEMBELAJARAN MAHASISWA PADA MATA KULIAH SISTEM MANAJEMEN BASIS DATA DENGAN K-MEDOIDS

Mohamad Ardi¹, Haryansyah², Yusni Amaliah³

^{1,2,3}STMIK PPKIA Tarakanita Rahmawati, Tarakan, Indonesia

Email: mohamadardi@ppkia.ac.id

Abstrak

Data penilaian mata kuliah menyimpan pola capaian belajar yang tidak tampak dari nilai akhir saja. Penelitian ini bertujuan mengelompokkan pola capaian pembelajaran mahasiswa pada mata kuliah Sistem Manajemen Basis Data menggunakan algoritma K-Medoids. Data berupa nilai 97 mahasiswa pada enam komponen penilaian (partisipasi, proyek, tugas, kuis, UTS, dan UAS) yang dianonimkan dan dinormalisasi dengan Min-Max. K-Medoids dengan *Partitioning Around Medoids* (PAM) diuji pada $k = 2$ sampai 6 dengan jarak Euclidean dan Manhattan, dibandingkan dengan K-Means, serta dievaluasi menggunakan *Silhouette Coefficient*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index*. Hasil menunjukkan $k = 3$ paling sesuai (*Silhouette* 0,690; DBI 0,499) dengan partisi yang identik pada kedua jarak dan stabil pada 100 pengujian subsampel ($ARI = 1,00$). Tiga klaster yang terbentuk adalah capaian konsisten tinggi (70 mahasiswa), capaian tinggi tanpa pengumpulan tugas (18 mahasiswa), dan tidak aktif berisiko gagal (9 mahasiswa). Medoid yang berupa mahasiswa nyata memudahkan penafsiran profil klaster untuk evaluasi pembelajaran dan intervensi dini.

Kata Kunci: Klasterisasi, K-Medoids, Educational Data Mining, Capaian Pembelajaran, Silhouette Coefficient.

Abstract

Course assessment data hold learning achievement patterns that final grades alone do not reveal. This study aims to cluster students' learning achievement patterns in a Database Management Systems course using the K-Medoids algorithm. The data consist of the scores of 97 anonymized students on six assessment components (participation, project, assignment, quiz, midterm, and final exam), normalized with Min-Max scaling. K-Medoids based on *Partitioning Around Medoids* (PAM) was tested for $k = 2$ to 6 with Euclidean and Manhattan distances, compared with K-Means, and evaluated using the *Silhouette Coefficient*, the *Davies-Bouldin Index* (DBI), and the *Calinski-Harabasz Index*. The results show that $k = 3$ is the most suitable solution (*Silhouette* 0.690; DBI 0.499), yielding identical partitions under both distances and remaining stable across 100 subsample tests ($ARI = 1.00$). The three clusters are consistently high achievers (70 students), high achievers without submitted assignments (18 students), and inactive students at risk of failing (9 students). Because medoids are actual students, the cluster profiles are easy to interpret for learning evaluation and early intervention.

Keywords: Clustering, K-Medoids, Educational Data Mining, Learning Achievement, Silhouette Coefficient.

A. PENDAHULUAN

Perguruan tinggi menghasilkan data penilaian dalam jumlah besar pada setiap semester, mulai dari partisipasi, tugas, kuis, proyek, hingga ujian. Data tersebut lazimnya hanya dipakai

untuk menghitung nilai akhir dan huruf mutu, sehingga pola capaian belajar yang lebih halus tidak terlihat. Dua mahasiswa dengan huruf mutu yang sama dapat memiliki profil capaian yang sangat berbeda. Educational data mining memanfaatkan teknik pembelajaran mesin untuk menggali pola semacam ini, dan klasterisasi sebagai salah satu metode unsupervised learning memungkinkan mahasiswa dikelompokkan tanpa label kelas yang ditentukan lebih dahulu (Ezugwu et al., 2022; Murphy et al., 2024).

Algoritma K-Means merupakan metode klasterisasi yang paling populer, tetapi memiliki sejumlah keterbatasan. Pemilihan pusat awal secara acak dapat menyebabkan konvergensi ke solusi lokal, dan penggunaan jarak Euclidean membatasi kemampuan algoritma ini pada bentuk klaster tertentu (Ikotun et al., 2023). Selain itu, pusat klaster K-Means berupa rata-rata yang belum tentu sesuai dengan satu mahasiswa pun. K-Medoids, khususnya Partitioning Around Medoids (PAM), memakai objek data nyata sebagai pusat klaster dan dipandang lebih tahan terhadap derau dan pencilan (Alyasyifa et al., 2025; Schubert & Rousseeuw, 2021). Dalam konteks pembelajaran, medoid adalah mahasiswa nyata yang mewakili klasternya, sehingga profil klaster lebih mudah ditafsirkan oleh dosen.

Sejumlah penelitian telah menerapkan K-Medoids pada data mahasiswa. (Bahri & Midyanti, 2023) memakai K-Medoids untuk mengelompokkan mahasiswa berpotensi drop out dan memperoleh *Silhouette* terbaik 0,394 pada dua klaster. (Qomariyah & Siregar, 2022) membandingkan K-Means dan K-Medoids pada data akademik mahasiswa dengan DBI, sedangkan (Zahrotun et al., 2023) membandingkan K-Medoids dengan hierarchical clustering berdasarkan *Silhouette*. (Alyasyifa et al., 2025) menerapkan PAM untuk membentuk kelompok belajar dan membandingkan beberapa ukuran jarak. Penelitian-penelitian tersebut umumnya memakai indikator agregat tingkat program studi, seperti IPK, asal sekolah, dan lama studi (Qomariyah & Siregar, 2022; Zahrotun et al., 2023).

Dari literatur yang ditinjau, kajian yang mengklasterkan komponen penilaian pada tingkat satu mata kuliah masih terbatas. Padahal, tingkat inilah yang paling dekat dengan keputusan pembelajaran seorang dosen. Kajian semacam itu juga perlu menguji kepekaan hasil terhadap pilihan jarak, jumlah klaster, dan algoritma pembandingan. Mata kuliah Sistem Manajemen Basis Data (SMBD) dipilih karena merupakan mata kuliah inti bidang informatika dengan komponen proyek yang berbobot besar (25%) dan jumlah peserta yang cukup untuk dianalisis.

Penelitian ini bertujuan (1) menentukan jumlah klaster dan ukuran jarak yang sesuai untuk K-Medoids pada data penilaian SMBD; (2) mengidentifikasi profil capaian pembelajaran tiap klaster; dan (3) membandingkan hasilnya dengan K-Means serta menguji stabilitas dan kepekaannya. Kontribusi penelitian ini adalah (a) penerapan K-Medoids berbasis medoid untuk evaluasi pembelajaran tingkat mata kuliah, (b) protokol evaluasi yang mencakup beberapa nilai k, dua jarak, algoritma pembandingan, dan uji stabilitas, serta (c) temuan praktis mengenai peran komponen tugas yang berbobot kecil dalam pembentukan nilai akhir.

B. TINJAUAN PUSTAKA

1. Klasterisasi Pada Data Pendidikan

Klasterisasi mengelompokkan objek sehingga objek dalam satu klaster lebih mirip satu sama lain daripada dengan objek di klaster lain. (Ezugwu et al., 2022) menyajikan tinjauan sistematis atas teknik klasterisasi tradisional maupun mutakhir beserta taksonominya. Pada data pendidikan, (Nalli et al., 2021) membandingkan beberapa algoritma klasterisasi untuk membentuk kelompok mahasiswa yang heterogen pada kuliah daring, sedangkan (Murphy et al., 2024) menyajikan tutorial komparatif analisis klaster berbasis ketakmiripan untuk data pendidikan. Studi-studi ini menunjukkan bahwa pilihan algoritma dan ukuran ketakmiripan perlu dievaluasi, bukan diasumsikan.

2. K-Means dan K-Medoids

K-Means meminimalkan jumlah kuadrat jarak objek ke rata-rata (*centroid*) klaster. (Ikotun et al., 2023) mencatat bahwa pemilihan pusat awal secara acak dapat menghasilkan konvergensi lokal dan bahwa jarak Euclidean membatasi ketahanan algoritma pada bentuk klaster tertentu. K-Medoids membatasi pusat klaster pada objek data itu sendiri, yang disebut medoid. Tujuan optimasinya adalah meminimalkan total jarak setiap objek ke medoid klasternya, seperti pada Persamaan (4) di bagian Metode pada riset ini. PAM bekerja dalam dua fase, yaitu BUILD yang memilih medoid awal secara deterministik dan SWAP yang menukar medoid dengan objek non-medoid selama total jarak masih dapat diturunkan. (Schubert & Rousseeuw, 2021) mengusulkan percepatan untuk PAM, CLARA, dan CLARANS, sehingga biaya komputasi PAM bukan lagi hambatan utama. Pada data berukuran 97 objek, biaya ini bahkan dapat diabaikan.

3. Validasi Klaster

Silhouette Coefficient mengukur seberapa baik sebuah objek cocok dengan klasternya dibandingkan klaster terdekat lainnya, dengan rentang -1 sampai 1 dan nilai yang lebih tinggi menunjukkan struktur yang lebih baik (Shutaywi & Kachouie, 2021). *Davies-Bouldin Index* (DBI) mengukur rasio sebaran dalam klaster terhadap jarak antar klaster, dengan nilai lebih rendah menunjukkan klaster yang lebih terpisah, dan telah dipakai untuk membandingkan K-Means dan K-Medoids pada data mahasiswa (Qomariyah & Siregar, 2022). *Calinski-Harabasz Index* (CH) membandingkan sebaran antar klaster dengan sebaran dalam klaster, dengan nilai lebih tinggi lebih baik.

4. Penelitian Terkait

Tabel 1 meringkas penelitian terdahulu yang paling dekat dengan penelitian ini. Perbedaan utamanya terletak pada tingkat data (komponen penilaian satu mata kuliah), perbandingan dua jarak dan algoritma pembandingan, serta pengujian stabilitas.

Tabel 1. Ringkasan Penelitian Terdahulu dan Posisi Penelitian

Penelitian	Data	Metode dan Evaluasi	Temuan Utama
Bahri & Midyanti (2023)	Mahasiswa berpotensi <i>drop out</i>	K-Medoids; <i>Silhouette</i>	<i>Silhouette</i> terbaik 0,394 pada $k = 2$
Qomariyah & Siregar (2022)	Data akademik mahasiswa angkatan 2013–2015	K-Means dan K-Medoids; DBI	K-Means ($k = 5$; DBI 0,781) lebih optimal daripada K-Medoids ($k = 3$; DBI 0,929)
Zahrotun et al. (2023)	Asal sekolah, asal daerah, nilai matematika, TOEFL, IPK, lama studi (empat program studi)	K-Medoids dan <i>hierarchical clustering</i> ; <i>Silhouette</i> ; $k = 2-4$	<i>Silhouette</i> hierarchical clustering (0,87–0,88) lebih tinggi daripada K-Medoids (0,35–0,67)
Alyasyifa et al. (2025)	Profil 41 mahasiswa (kuis, IPK, nilai mata kuliah, MBTI, VARK, minat)	K-Medoids (PAM); jarak Euclidean, Manhattan, Gower; $k = 2-5$	Euclidean dengan PCA memberi <i>Silhouette</i> tertinggi
Penelitian ini	Enam komponen penilaian satu mata kuliah (97 mahasiswa)	K-Medoids (PAM) dan K-Means; Euclidean dan Manhattan; $k = 2-6$; <i>Silhouette</i> , DBI, CH; uji stabilitas	Lihat bagian Hasil dan Pembahasan pada riset ini

C. METODE

1. Desain Penelitian

Penelitian ini bersifat kuantitatif eksploratif dengan tahapan yang diadaptasi dari CRISP-DM (Schröer et al., 2021), yaitu pemahaman masalah, pemahaman data, persiapan data, pemodelan, dan evaluasi. Hasil evaluasi kemudian ditafsirkan sebagai profil kluster untuk keperluan pembelajaran.

2. Data

Tabel 2. Dataset Penelitian

<i>ID</i>	<i>Partisipasi</i>	<i>Proyek</i>	<i>Tugas</i>	<i>Kuis</i>	<i>UTS</i>	<i>UAS</i>	<i>Total</i>	<i>Grade</i>
M001	85.0	25.0	65.0	75.0	86.0	85.0	69.0	B
M002	100.0	100.0	97.0	80.0	86.0	98.0	96.0	A
M003	100.0	100.0	90.0	75.0	93.0	100.0	97.0	A
M004	100.0	100.0	97.0	80.0	86.0	100.0	96.0	A
M005	100.0	100.0	93.0	80.0	93.0	93.0	96.0	A
...	...	...	...	...	...	...	...	...
M047	85.0	100.0	97.0	75.0	93.0	95.0	92.0	A
M048	100.0	100.0	97.0	75.0	100.0	93.0	97.0	A
M049	85.0	100.0	87.0	80.0	93.0	98.0	93.0	A
M050	100.0	100.0	97.0	75.0	93.0	95.0	96.0	A
M051	85.0	92.0	100.0	80.0	93.0	80.0	88.0	A
...	...	...	...	...	...	...	...	...
M093	100.0	92.0	70.0	80.0	100.0	93.0	94.0	A
M094	85.0	100.0	100.0	80.0	100.0	88.0	93.0	A
M095	85.0	100.0	0.0	75.0	93.0	93.0	87.0	A
M096	85.0	100.0	0.0	75.0	100.0	88.0	88.0	A
M097	85.0	100.0	0.0	75.0	86.0	85.0	84.0	AB

Data berupa nilai mata kuliah Sistem Manajemen Basis Data pada semester Ganjil 2025/2026 dari 97 mahasiswa di STMIK PPKIA Tarakanita Rahmawati. Data merupakan data sekunder hasil penilaian yang dikelola pengampu mata kuliah. Seluruh NIM dan nama diganti dengan kode M001 sampai M097, sebagian *dataset* penelitian ini ditampilkan pada Tabel 2. Atribut yang dipakai sebagai masukan klasterisasi adalah enam komponen penilaian pada Tabel 3. Kolom Total dan Grade tidak dipakai sebagai masukan agar kluster tidak ditentukan oleh nilai akhir, tetapi dipakai untuk menafsirkan kluster.

Tabel 3. Komponen Penilaian dan Statistik Deskriptif

<i>Komponen</i>	<i>Bobot</i>	<i>Rerata</i>	<i>Simpangan Baku</i>	<i>Minimum</i>	<i>Maksimum</i>
Aktivitas Partisipatif	25%	82,94	27,72	0	100
Hasil Proyek	25%	88,19	29,80	0	100
Tugas	5%	70,41	39,52	0	100
Kuis	5%	74,85	20,34	0	100
UTS	20%	89,45	17,99	0	100
UAS	20%	82,82	27,63	0	100

Komponen	Bobot	Rerata	Simpangan Baku	Minimum	Maksimum
Total (Tidak Dipakai Sebagai Masukan)	100%	84,51	23,13	0	99

3. Praproses Data

Data tidak memiliki nilai hilang dan seluruh nilai berada pada rentang 0 sampai 100. Karena komponen berbeda dalam sebaran, setiap atribut dinormalisasi dengan Min-Max agar tidak ada atribut yang mendominasi jarak (Ahsan et al., 2021), seperti pada Persamaan (1).

$$x' = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \quad (1)$$

Dua mahasiswa memiliki nilai nol pada seluruh komponen dan sembilan mahasiswa memiliki nilai total 30 atau kurang. Mereka tidak dibuang karena mewakili kelompok mahasiswa yang nyata dan justru merupakan sasaran evaluasi pembelajaran. Pengaruh keberadaan mereka terhadap hasil diuji pada analisis kepekaan.

4. Algoritma K-Medoids (PAM)

PAM dijalankan dengan fase BUILD, yang memilih medoid pertama sebagai objek dengan jumlah jarak terkecil ke semua objek lain, lalu menambah medoid satu per satu sampai berjumlah k. Fase SWAP kemudian menukar sebuah medoid dengan objek non-medoid apabila penukaran menurunkan total jarak, dan diulang sampai tidak ada penukaran yang memperbaiki hasil. Setiap objek ditempatkan pada kluster dengan medoid terdekat. Dua ukuran jarak dibandingkan, yaitu Euclidean (Persamaan 2) dan Manhattan (Persamaan 3), sedangkan fungsi tujuannya ditunjukkan Persamaan (4).

$$d_E(x, y) = \sqrt{\sum_j (x_j - y_j)^2} \quad (2)$$

$$d_M(x, y) = \sum_j |x_j - y_j| \quad (3)$$

$$J = \sum_k = 1^K \sum_{i \in k} d(x_i, m_k) \quad (4)$$

dengan x dan y adalah vektor atribut dua mahasiswa, $i \in k$ berarti mahasiswa i berada pada kluster k, dan m dengan indeks k adalah medoid kluster tersebut yang dipilih dari objek data. Jumlah kluster diuji pada k = 2 sampai 6.

5. Evaluasi, Pembanding, Stabilitas, dan Kepekaan

Kualitas kluster dinilai dengan *Silhouette Coefficient* (Persamaan 5), DBI, dan CH. *Silhouette* dihitung memakai jarak yang sama dengan jarak klusterisasi, sedangkan DBI dan CH dihitung pada ruang Euclidean.

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (5)$$

dengan a(i) rerata jarak objek i ke objek lain dalam klasternya dan b(i) rerata jarak terkecil objek i ke kluster lain (Shutaywi & Kachouie, 2021). Sebagai pembanding, K-Means dijalankan dengan jarak Euclidean, 50 inisialisasi acak, dan *random seed* 42. Kemiripan partisi antar algoritma diukur dengan *Adjusted Rand Index* (ARI).

Stabilitas diuji dengan 100 kali pengambilan subsampel 80% tanpa pengembalian. Pada setiap ulangan, PAM dijalankan ulang pada subsampel dan ARI dihitung antara partisi subsampel dan partisi data penuh pada mahasiswa yang sama. Uji kepekaan dilakukan dengan mengulang analisis tanpa sembilan mahasiswa bernilai total 30 atau kurang ($n = 88$). Seluruh analisis dijalankan dengan Python (pandas dan scikit-learn); PAM diimplementasikan sendiri sesuai langkah di atas.

D. HASIL DAN PEMBAHASAN

1. Karakteristik Data

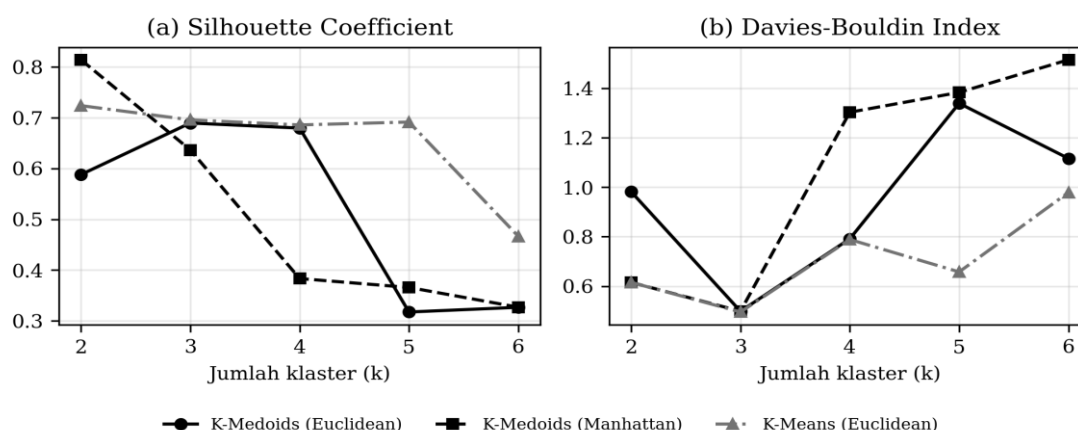
Rerata nilai total adalah 84,51 dengan 79 mahasiswa (81,4%) bergrade A, 8 AB, 1 B, dan 9 E. Distribusi nilai sangat condong ke atas, tetapi sebaran komponen Tugas (median 90; rerata 70,41; simpangan baku 39,52) dan Partisipasi menunjukkan adanya kelompok dengan nilai nol (Tabel 2). Struktur seperti ini sulit dilihat dari huruf mutu saja dan menjadi alasan klasterisasi dilakukan pada tingkat komponen.

2. Penentuan Jumlah Klaster dan Jarak

Tabel 4 dan Gambar 1 memuat hasil evaluasi untuk $k = 2$ sampai 6. Pada K-Medoids dengan jarak Euclidean, $k = 3$ memberi *Silhouette* tertinggi (0,690), DBI terendah (0,499), dan CH tertinggi (202,53) sekaligus. Pada jarak Manhattan, *Silhouette* tertinggi memang dicapai pada $k = 2$ (0,814), tetapi partisi tersebut hanya memisahkan 9 mahasiswa tidak aktif dari 88 mahasiswa lainnya sehingga tidak membedakan profil di antara mahasiswa aktif. Pada $k = 3$, jarak Manhattan menghasilkan DBI dan CH yang sama dengan Euclidean, dan partisi kedua jarak identik (ARI = 1,00). Untuk $k \geq 5$, *Silhouette* PAM Euclidean turun hingga sekitar 0,32 karena kelompok utama terbelah menjadi dua bagian yang hampir sama besar (36 dan 35 mahasiswa) tanpa makna pembelajaran yang jelas. Berdasarkan ketiga indeks dan keterbacaan hasil, $k = 3$ dipilih.

Tabel 4. Hasil Evaluasi Klaster

Algoritma	Jarak	k	<i>Silhouette</i>	DBI	CH	Ukuran Klaster
K-Medoids	Euclidean	2	0,588	0,982	50,32	74, 23
K-Medoids	Euclidean	3	0,690	0,499	202,53	70, 18, 9
K-Medoids	Euclidean	4	0,680	0,791	183,59	70, 18, 5, 4
K-Medoids	Euclidean	5	0,317	1,339	149,84	36, 35, 17, 5, 4
K-Medoids	Euclidean	6	0,327	1,116	148,88	36, 35, 17, 4, 3, 2
K-Medoids	Manhattan	2	0,814	0,615	121,39	88, 9
K-Medoids	Manhattan	3	0,637	0,499	202,53	70, 18, 9
K-Medoids	Manhattan	4	0,383	1,303	143,58	41, 30, 17, 9
K-Medoids	Manhattan	5	0,366	1,383	149,06	41, 30, 17, 5, 4
K-Medoids	Manhattan	6	0,327	1,514	123,60	37, 19, 17, 15, 5, 4
K-Means	Euclidean	2	0,724	0,615	121,39	88, 9
K-Means	Euclidean	3	0,696	0,496	202,68	71, 17, 9
K-Means	Euclidean	4	0,686	0,788	183,77	71, 17, 5, 4
K-Means	Euclidean	5	0,692	0,658	170,17	71, 17, 3, 3, 3
K-Means	Euclidean	6	0,467	0,979	161,66	59, 17, 12, 3, 3, 3



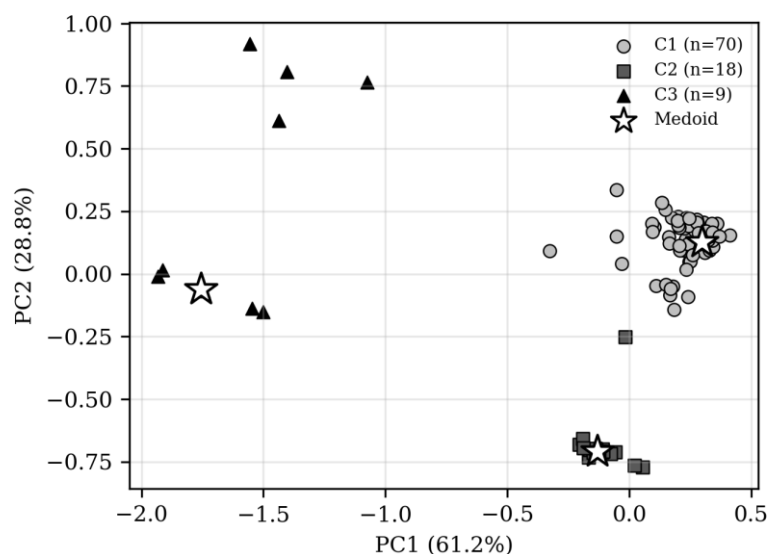
Gambar 1. Silhouette Coefficient dan Davies-Bouldin Index

3. Profil Kluster

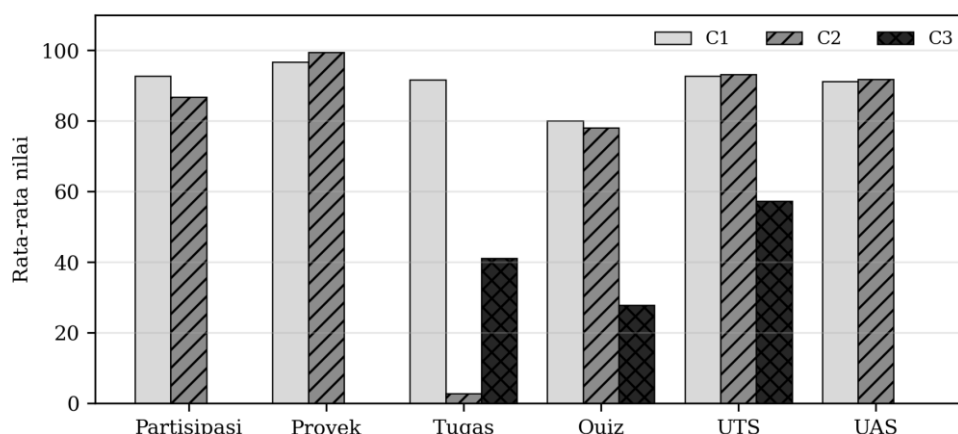
Tabel 5 merangkum profil ketiga kluster pada $k = 3$, sedangkan Gambar 2 menampilkan sebaran mahasiswa pada dua komponen utama (PCA) beserta posisi medoid. Gambar 3 membandingkan rerata nilai per komponen.

Tabel 5. Profil Kluster K-Medoids ($k = 3$, Jarak Euclidean)

Karakteristik	C1	C2	C3
Label	Capaian Konsisten Tinggi	Capaian Tinggi Tanpa Tugas	Tidak Aktif, Berisiko Gagal
Jumlah (persen)	70 (72,2%)	18 (18,6%)	9 (9,3%)
Partisipasi	92,64	86,67	0,00
Proyek	96,66	99,33	0,00
Tugas	91,57	2,78	41,11
Kuis	80,07	78,06	27,78
UTS	92,64	93,11	57,33
UAS	91,19	91,72	0,00
Rerata Nilai Total	92,69	87,50	14,89
Distribusi Grade	A 65; AB 4; B 1	A 14; AB 4	E 9
Medoid (Nilai Total, Grade)	M005 (96; A)	M095 (87; A)	M029 (16; E)



Gambar 2. Sebaran Kluster pada Dua Komponen Utama (PC1 dan PC2) dan Posisi Medoid



Gambar 3. Rerata Nilai Komponen Pada Setiap Kluster

Kluster C1 (70 mahasiswa) terdiri atas mahasiswa dengan nilai tinggi dan merata pada hampir semua komponen. Medoidnya, M005, memperoleh nilai total 96. Kluster C2 (18 mahasiswa) hampir identik dengan C1 pada semua komponen kecuali Tugas, yang rata-ratanya hanya 2,78 karena sebagian besar anggotanya tidak memperoleh nilai tugas. Kluster C3 (9 mahasiswa) memiliki nilai nol pada partisipasi, proyek, dan UAS, dan seluruhnya bergrade E. Simpangan baku C3 yang besar pada Tugas, Kuis, dan UTS (42 sampai 49) menunjukkan bahwa sebagian dari mereka sempat mengerjakan komponen awal, sedangkan tiga di antaranya tidak memperoleh nilai UTS. Medoid C3, M029, hanya memiliki nilai UTS (79).

4. Perbandingan Dengan K-Means, Stabilitas, dan Kepekaan

Pada $k = 3$, K-Means menghasilkan *Silhouette* 0,696, DBI 0,496, dan CH 202,68, yang praktis sama dengan K-Medoids (0,690; 0,499; 202,53). Kesesuaian partisi sangat tinggi (ARI = 0,962); satu-satunya perbedaan adalah M071, mahasiswa dengan nilai Tugas 47 yang ditempatkan K-Medoids pada C2 dan K-Means pada C1. Pada 100 subsampel 80%, PAM selalu menghasilkan partisi yang sama dengan partisi penuh (ARI rerata = 1,00; simpangan baku 0,00), dengan *Silhouette* rerata 0,689. Setelah sembilan mahasiswa tidak aktif dikeluarkan ($n = 88$), nilai k terbaik adalah 2 (*Silhouette* 0,717; DBI 0,355) dan keanggotaannya sama persis dengan C1 dan C2 (70 dan 18 mahasiswa), sementara $k = 3$ menurun ke *Silhouette* 0,318. Hasil

ini menunjukkan bahwa pemisahan C1 dan C2 tidak bergantung pada keberadaan mahasiswa tidak aktif.

Struktur tiga klaster ini kuat. *Silhouette* 0,690 lebih tinggi daripada nilai terbaik (Bahri & Midyanti, 2023) sebesar 0,394 dan sedikit di atas kisaran K-Medoids pada (Zahrotun et al., 2023) yang berkisar 0,35 sampai 0,67, tetapi perbandingan ini hanya indikatif karena data dan atributnya berbeda. Kesamaan hasil kedua jarak sejalan dengan (Alyasyifa et al., 2025), yang juga menemukan bahwa Euclidean dan Manhattan memberi hasil yang bersaing pada data mahasiswa.

K-Medoids tidak menghasilkan partisi yang lebih baik daripada K-Means pada data ini, sejalan dengan (Qomariyah & Siregar, 2022) yang menemukan K-Means unggul berdasarkan DBI. Alasannya, pencilaan pada data ini (mahasiswa tidak aktif) membentuk klaster tersendiri, sehingga tidak menggeser pusat klaster lain. Keunggulan K-Medoids di sini lebih bersifat interpretatif, yaitu setiap klaster diwakili mahasiswa nyata (M005, M095, M029) yang dapat ditelaah dosen secara langsung, tidak sekadar vektor rata-rata.

Temuan yang paling relevan bagi pembelajaran adalah klaster C2. Selisih rerata nilai total C1 dan C2 sebesar 5,19 poin, dan 4,44 poin (sekitar 86%) di antaranya berasal dari komponen Tugas yang bobotnya hanya 5%. Akibatnya, 14 dari 18 mahasiswa C2 tetap memperoleh grade A tanpa mengumpulkan tugas. Bagi dosen, ini menunjukkan bahwa komponen Tugas pada bobot sekarang nyaris tidak berkonsekuensi pada nilai akhir. Bobot, rubrik, atau syarat minimal pengumpulan tugas dapat ditinjau kembali, dengan catatan bahwa penelitian ini tidak membuktikan hubungan sebab-akibat dan tidak dapat membedakan mahasiswa yang tidak mengumpulkan tugas dari tugas yang tidak tercatat. Pada data ini nilai partisipasi nol hanya ditemukan pada klaster C3, dan seluruh anggota C3 juga bernilai nol pada proyek dan UAS. Partisipasi dan proyek yang masih kosong menjelang akhir semester dapat dijadikan penanda dini mahasiswa berisiko, sehingga dosen berpeluang menindaklanjuti sebelum UAS. Penanda ini perlu divalidasi pada semester lain sebelum dijadikan aturan.

E. KESIMPULAN

Penelitian ini menunjukkan bahwa K-Medoids (PAM) pada enam komponen penilaian mata kuliah Sistem Manajemen Basis Data menghasilkan struktur terbaik pada $k = 3$ (*Silhouette* 0,690; DBI 0,499; CH 202,53), dengan partisi identik pada jarak Euclidean dan Manhattan serta stabil pada 100 subsampel. Tiga klaster yang terbentuk adalah capaian konsisten tinggi (70 mahasiswa), capaian tinggi tanpa tugas (18 mahasiswa), dan tidak aktif berisiko gagal (9 mahasiswa), masing-masing diwakili oleh medoid berupa mahasiswa nyata. Hasil K-Medoids hampir sama dengan K-Means ($ARI = 0,962$), sehingga keunggulan K-Medoids pada data ini terletak pada keterbacaan profil, bukan pada kualitas partisi. Temuan praktisnya adalah bahwa komponen Tugas berbobot 5% hampir tidak memengaruhi nilai akhir, dan bahwa partisipasi serta proyek yang kosong dapat menjadi penanda dini mahasiswa berisiko.

Penelitian ini memiliki beberapa keterbatasan. Data berasal dari satu mata kuliah pada satu semester dengan 97 mahasiswa, sehingga generalisasi terbatas. Atribut hanya mencakup komponen penilaian, tanpa kehadiran, aktivitas pada sistem pembelajaran, atau latar belakang mahasiswa. Nilai nol tidak dapat dibedakan antara tidak mengerjakan dan tidak tercatat. Penelitian selanjutnya dapat menambah mata kuliah dan semester, memasukkan data kehadiran dan log pembelajaran, membandingkan metode lain seperti *hierarchical clustering* atau DBSCAN, serta menguji apakah intervensi dini pada mahasiswa yang ditandai benar-benar meningkatkan capaian.

DAFTAR PUSTAKA

- Ahsan, M., Mahmud, M., Saha, P., Gupta, K., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*, 9(3), 52. <https://doi.org/10.3390/technologies9030052>
- Alyasyifa, S., Pratiwi, O. N., & Darmawan, I. (2025). Data-Driven Student Group Formation for Group Investigation: A K-Medoids Clustering Approach in Cooperative Learning. *Jurnal Teknik Informatika (Jutif)*, 6(5), 3886–3898. <https://doi.org/10.52436/1.jutif.2025.6.5.4765>
- Bahri, S., & Midyanti, D. M. (2023). Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(1), 165–172. <https://doi.org/10.25126/jtiik.2023106643>
- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743. <https://doi.org/10.1016/j.engappai.2022.104743>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Murphy, K., López-Pernas, S., & Saqr, M. (2024). Dissimilarity-Based Cluster Analysis of Educational Data: A Comparative Tutorial Using R. In *Learning Analytics Methods and Tutorials* (pp. 231–283). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54464-4_8
- Nalli, G., Amendola, D., Perali, A., & Mostarda, L. (2021). Comparative Analysis of Clustering Algorithms and Moodle Plugin for Creation of Student Heterogeneous Groups in Online University Courses. *Applied Sciences*, 11(13), 5800. <https://doi.org/10.3390/app11135800>
- Qomariyah, & Siregar, M. U. (2022). Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(2), 91–99. <https://doi.org/10.14421/jiska.2022.7.2.91-99>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 101, 101804. <https://doi.org/10.1016/j.is.2021.101804>
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. *Entropy*, 23(6), 759. <https://doi.org/10.3390/e23060759>
- Zahrotun, L., Linarti, U., Suandi As, B. H. T., Kurnia, H., & Sabila, L. Y. (2023). Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students' Data Grouping. *JOIV: International Journal on Informatics Visualization*, 7(2), 446. <https://doi.org/10.30630/joiv.7.2.1204>