



Penerapan K-Means Clustering untuk Klasterisasi Hasil Belajar Matematika Siswa dari Data Nilai Ujian

Divia Ilham Hasyanah*

Universitas Singaperbangsa Karawang, *Penulis Korespondensi: 2310631050131@student.unsika.ac.id

Karunia Eka Lestari

Universitas Singaperbangsa Karawang

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan hasil belajar matematika siswa dengan menerapkan algoritma *k-means clustering* serta mengidentifikasi karakteristik setiap *cluster* yang terbentuk. Populasi penelitian adalah seluruh siswa kelas XI di SMA Negeri 4 Babelan, dengan sampel sebanyak 72 siswa yang diambil secara acak menggunakan teknik random sampling dari kelas XI.C dan XI.G sebagai representasi populasi. Metode penelitian yang digunakan adalah kuantitatif dengan pendekatan deskriptif untuk menganalisis data nilai ujian matematika. Hasil klasterisasi menunjukkan terbentuk dua *cluster*, di mana *cluster* pertama terdiri dari 49 siswa dengan karakteristik nilai ujian yang lebih tinggi, sedangkan *cluster* kedua terdiri dari 23 siswa dengan nilai ujian relatif lebih rendah. Data *cluster* tersebut dapat digunakan sebagai informasi evaluasi yang bermanfaat dalam perbaikan pembelajaran matematika di sekolah tersebut.

Kata kunci: Hasil belajar, klasterisasi, k-means clustering

PENDAHULUAN

Kemampuan berpikir analitis, metodis, dan logis siswa meningkat secara signifikan melalui pendidikan matematika. Untuk membantu siswa secara aktif membangun pemahaman dan meningkatkan hasil belajar, pendidikan matematika menekankan pengembangan kemampuan berpikir kritis dan pemecahan masalah dalam dunia nyata, selain perolehan ide dan metode (Kafillah dkk., 2025). Namun demikian, hasil belajar matematika siswa di berbagai jenjang pendidikan masih menunjukkan variasi yang cukup signifikan. Sebagian siswa mampu mencapai hasil yang optimal, sementara sebagian lainnya masih mengalami kesulitan dalam memahami dan menerapkan konsep-konsep dasar (Pangesti & Yudhanegara, 2024). Selain kemampuan akademik, minat siswa terhadap pelajaran matematika juga menjadi faktor penting yang memengaruhi pencapaian belajar (Tiana dkk., 2024).

Hasil belajar matematika di SMA Negeri 4 Babelan seringkali dinilai dengan cara tradisional, dengan rata-rata kelas atau perbandingan dengan KKM sebagai dasar evaluasi. Metode ini dapat merangkum kemajuan siswa menuju tujuan pembelajaran, tetapi tidak dapat mengisolasi karakteristik individu berdasarkan kinerja mereka. Guru harus lebih memahami karakteristik kelompok siswa untuk mengembangkan praktik pembelajaran yang lebih efektif, adaptif, dan tepat sasaran. Hendrastuty, (2024) menyatakan bahwa evaluasi berbasis data dapat mengungkap pola capaian belajar yang tidak terlihat melalui penilaian konvensional, sekaligus membantu guru merancang pembelajaran berdiferensiasi sesuai kebutuhan masing-masing siswa.

Perkembangan teknologi informasi dan analisis data telah membuka peluang besar bagi dunia pendidikan dalam memahami perilaku belajar peserta didik secara lebih mendalam. Educational data mining (EDM) merupakan pendekatan yang menggabungkan metode data mining, machine learning, dan analisis statistik untuk menemukan pola-pola tersembunyi dalam data pendidikan guna meningkatkan efektivitas proses pembelajaran. Salah satu teknik yang banyak digunakan dalam EDM adalah clustering, yang berfungsi

untuk mengelompokkan data ke dalam kelompok dengan karakteristik serupa agar dapat dianalisis lebih lanjut (Andre dkk., 2023).

Salah satu algoritma clustering yang paling banyak digunakan adalah k-means clustering, yang berfungsi untuk mengelompokkan data ke dalam beberapa cluster berdasarkan karakteristik tertentu (Akbar dkk., 2025). Untuk menentukan jumlah cluster yang optimal, beberapa metode evaluasi dapat digunakan, seperti elbow method dan silhouette coefficient, yang berfungsi menilai kualitas hasil pengelompokan (Diantika dkk., 2025). Sholeh dan Aeni, (2023) menemukan bahwa kombinasi kedua metode ini meningkatkan validitas hasil klasterisasi sehingga pengelompokan lebih representatif.

Beberapa penelitian menunjukkan efektivitas k-means dalam konteks pendidikan. Yudhanegara dan Lestari, (2024) menemukan bahwa penerapan algoritma k-means mampu mengelompokkan siswa berdasarkan karakteristik hasil belajar mereka secara akurat, sehingga membantu guru dalam perencanaan pembelajaran diferensial, sementara Fauzie dkk., (2023) yang menunjukkan bahwa penggunaan k-means clustering dapat memberikan wawasan baru bagi pendidik dalam memahami pola capaian akademik siswa dan melakukan evaluasi pembelajaran berbasis data. Antoni dkk., (2025) menambahkan bahwa pengelompokan berbasis k-means mendukung strategi pembelajaran yang lebih adaptif, sedangkan Hidayat, (2022) menekankan relevansi algoritma ini dalam menganalisis minat siswa untuk meningkatkan motivasi belajar. Fatah dan Maghfiroh, (2025) menyoroti bahwa k-means efektif dalam mengidentifikasi pola tersembunyi dalam data pendidikan, sehingga memperkuat perannya sebagai alat perencanaan pembelajaran adaptif.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan algoritma k-means clustering dalam menganalisis data hasil belajar matematika peserta didik di SMA Negeri 4 Babelan. Melalui penerapan metode ini, penelitian diharapkan dapat memberikan manfaat praktis bagi pendidik dan pihak sekolah, antara lain membantu mengidentifikasi kelompok siswa berdasarkan tingkat pencapaian belajar, memberikan dasar dalam perancangan strategi pembelajaran berdiferensiasi, serta mendukung evaluasi efektivitas proses pembelajaran (Pratama dkk., 2025). Dengan demikian, hasil penelitian ini diharapkan dapat berkontribusi terhadap peningkatan kualitas pembelajaran matematika melalui penerapan pendekatan analisis berbasis data.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode deskriptif. Penelitian kuantitatif deskriptif merupakan jenis penelitian yang bertujuan menggambarkan keadaan atau fenomena sesuai kondisi sebenarnya dengan menggunakan data numerik seperti skor, frekuensi, atau persentase. Penelitian ini tidak bertujuan untuk menguji hipotesis atau hubungan sebab-akibat, tetapi fokus pada penggambaran data yang diperoleh sehingga memberikan gambaran faktual terkait variabel yang diteliti (Waruwu dkk., 2025). Penelitian ini dilaksanakan di SMAN 4 Babelan dengan populasi siswa kelas XI. Sampel penelitian diambil menggunakan teknik random sampling, dimana peneliti secara acak memilih beberapa kelas sebagai kelompok sampel. Dari metode ini, kelas XI.C, dan XI.G terpilih sebagai sampel yang mewakili populasi untuk keperluan penelitian. Dalam penelitian ini diambil dari data nilai ujian (PSTS 1, PSTS 2, PSAS, dan PSAT). Sugiyono, (2019) menyatakan rumus slovin, yaitu:

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n: ukuran sampel

N: ukuran populasi

e^2 : galat yang ditoleransi dalam pemilihan sampel oleh peneliti dinyatakan dalam persentase tertentu.

Dalam penelitian ini, peneliti menetapkan batas toleransi galat sebesar 10% ($e = 0,1$). Dengan populasi sebesar 252 siswa, maka perhitungan sampel menggunakan rumus Slovin menghasilkan ukuran sampel sebagai berikut:

$$n = \frac{252}{1 + 252(0,1)^2} = 71,59 \approx 72$$

Maka sampel pada penelitian ini adalah siswa kelas XI SMA Negeri 4 Babelan dengan jumlah sampel sebanyak 72 siswa.

Setelah diperoleh jumlah sampel penelitian sebanyak 72 siswa melalui perhitungan menggunakan rumus Slovin, tahap selanjutnya adalah melakukan analisis data hasil belajar matematika menggunakan pendekatan k-means clustering. Analisis ini bertujuan untuk mengelompokkan siswa berdasarkan kesamaan karakteristik capaian nilai ujian matematika, sehingga dapat diperoleh kelompok siswa dengan tingkat prestasi yang relatif homogen. Melalui proses klasterisasi ini, data hasil belajar tidak hanya ditinjau dari nilai rata-rata semata, tetapi juga dianalisis secara lebih mendalam untuk menemukan pola-pola tertentu dalam distribusi capaian akademik siswa.

Metode clustering yang digunakan dalam penelitian ini adalah algoritma k-means dengan matriks jarak Euclidean (Widiastuti & Waktu, 2022). Algoritma ini bekerja secara iteratif untuk mengelompokkan data berdasarkan tingkat kemiripan antar objek, sehingga data dalam satu cluster memiliki karakteristik yang serupa, sedangkan antar cluster berbeda secara signifikan.

Adapun langkah-langkah pelaksanaan algoritma k-means clustering adalah sebagai berikut.

1. Menentukan jumlah cluster (k)

Jumlah cluster ditentukan menggunakan metode elbow dan silhouette, yaitu dengan menganalisis nilai Within-cluster sum of squares (WCSS) dan silhouette coefficient untuk memperoleh jumlah cluster paling optimal (Widiastuti & Waktu, 2022).

2. Menyiapkan dan menormalkan data

Data nilai ujian siswa terlebih dahulu disiapkan dalam bentuk tabel terstruktur dan dinormalisasi agar setiap variabel memiliki skala yang sebanding, sehingga tidak ada atribut yang mendominasi hasil perhitungan jarak.

3. Menentukan titik pusat awal (centroid)

Sejumlah (k) data awal dipilih secara acak sebagai centroid awal dari setiap cluster.

4. Menghitung jarak setiap data ke centroid

Setiap data dihitung jaraknya terhadap setiap centroid menggunakan rumus persamaan Euclidean Distance:

$$d(x_i, y_j) = \sqrt{\sum_{i=1}^n (x_{ik} - y_{jk})^2}$$

Keterangan:

- $d(x_i, y_j)$ adalah jarak antara data ke- i dan data ke- j , yang menunjukkan tingkat kemiripan antar objek data.
- x_{ik} adalah skor ke- k dari siswa ke- i pada komponen penilaian tertentu.

- y_{jk} adalah skor ke- k dari siswa ke- j pada komponen penilaian yang sama.
- n adalah jumlah centroid.

Data kemudian dikelompokkan ke dalam cluster dengan jarak terdekat.

5. Memperbarui posisi centroid

Setelah semua data dikelompokkan, posisi centroid diperbarui dengan menghitung rata-rata nilai dari semua data yang berada dalam masing-masing cluster.

6. Iterasi hingga konvergen

Proses perhitungan jarak dan pembaruan centroid dilakukan secara berulang hingga posisi centroid tidak berubah signifikan (konvergen), menandakan bahwa pembentukan cluster telah stabil.

7. Evaluasi hasil klasterisasi

Hasil pengelompokan dievaluasi menggunakan nilai silhouette coefficient dan Between-cluster sum of squares (BCSS) untuk menilai homogenitas dan pemisahan antar cluster (Diantika dkk., 2025)

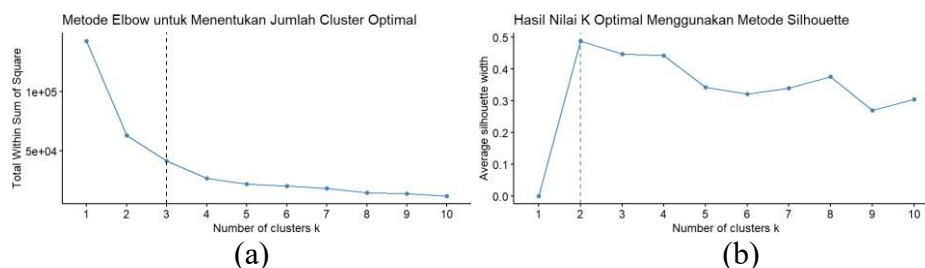
8. Interpretasi hasil cluster

Setiap cluster yang terbentuk dianalisis untuk mengidentifikasi karakteristik siswa. Misalnya, cluster dengan rata-rata nilai tinggi dikategorikan sebagai kelompok berprestasi tinggi, sedangkan cluster dengan nilai rendah dikategorikan sebagai kelompok yang memerlukan pendampingan tambahan (Widiastuti & Wakit, 2022).

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk menerapkan algoritma k-means clustering dalam mengelompokkan hasil belajar siswa pada mata pelajaran Matematika di SMA Negeri 4 Babelan. Data yang digunakan berupa nilai ujian siswa yang mencerminkan capaian belajar individu. Proses penelitian ini dilakukan melalui beberapa tahap, mulai dari penentuan jumlah cluster optimal, validasi hasil klasterisasi, analisis statistik deskriptif, hingga visualisasi hasil pengelompokan data. Setiap tahap bertujuan untuk menjawab rumusan masalah penelitian, yaitu bagaimana penerapan algoritma k-means dapat mengelompokkan hasil belajar siswa, mengidentifikasi karakteristik tiap cluster yang terbentuk, serta memanfaatkan hasil klasterisasi sebagai dasar evaluasi pembelajaran matematika di sekolah.

Kemampuan komunikasi matematis siswa yang mendapatkan pembelajaran berbasis masalah berbantuan *software cabri* lebih baik dibandingkan kemampuan komunikasi matematis siswa yang mendapatkan pembelajaran ekspositori. Simpulan yang diperoleh dari penelitian ini adalah pembelajaran berbasis masalah berbantuan *software cabri* lebih efektif dibandingkan dengan pembelajaran ekspositori di kelas XI SMA Negeri 1 Karawang. Jadi model pembelajaran berbasis masalah berbantuan *software cabri* efektif untuk meningkatkan kemampuan komunikasi matematis siswa.

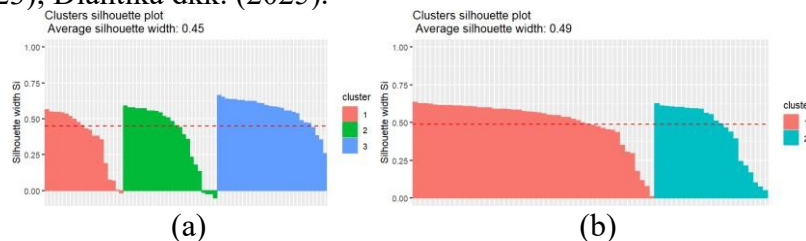


Gambar 1. Penentuan Nilai K Optimal Menggunakan Metode *Elbow* (a) dan *Silhouette* (b)

Berdasarkan metode *elbow*, terlihat bahwa titik tekukan (*elbow point*) berada pada $k = 3$. Hal ini menunjukkan bahwa tiga *cluster* dapat dianggap sebagai jumlah yang cukup representatif untuk menggambarkan variasi dalam data nilai siswa. Semakin besar nilai k , penurunan *Within-cluster sum of squares (WCSS)* menjadi semakin kecil, dan titik tekukan pada $k = 3$ menandakan bahwa setelah tiga *cluster*, peningkatan hasil pemisahan tidak lagi signifikan.

Sementara itu, metode *silhouette* menunjukkan hasil berbeda, di mana nilai rata-rata koefisien *silhouette* tertinggi diperoleh pada $k = 2$ dengan nilai sebesar 0,49. Nilai tersebut menandakan bahwa pembagian data menjadi dua kelompok memberikan hasil pengelompokan yang paling stabil, karena jarak antar *cluster* cukup jauh dan data dalam satu *cluster* memiliki kesamaan tinggi.

Dengan demikian, meskipun metode *elbow* menyarankan tiga *cluster*, hasil dari metode *silhouette* menunjukkan bahwa dua *cluster* menghasilkan pemisahan yang lebih baik secara kualitas. Oleh karena itu, analisis selanjutnya menggunakan dua *cluster* ($k = 2$) sebagai jumlah *klaster* optimal yang paling representatif untuk analisis selanjutnya Sholeh dan Aeni (2023); Diantika dkk. (2025).



Gambar 2. Visualisasi Plot Nilai Silhouette dengan (a) Tiga Cluster ($k = 3$) dan (b) Dua Cluster ($k = 2$)

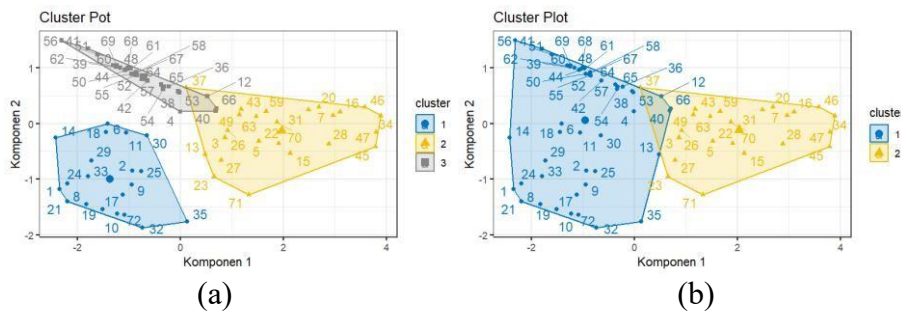
Pada *silhouette* plot dua *cluster* ($k = 2$), sebagian besar data memiliki nilai *silhouette* width positif dengan distribusi yang cukup merata. Hal ini menunjukkan bahwa data sudah terkelompok dengan baik sesuai kesamaannya. Sebaliknya, pada *silhouette* plot tiga *cluster* ($k = 3$), tampak adanya beberapa bagian dengan nilai *silhouette* negatif dan area yang tidak seimbang, yang menandakan bahwa sebagian data mungkin kurang tepat berada dalam *cluster* yang ditetapkan. Oleh karena itu, walaupun metode *elbow* menunjukkan bahwa tiga *cluster* cukup baik secara statistik, *silhouette* plot memperlihatkan bahwa dua *cluster* menghasilkan pengelompokan yang lebih stabil dan homogen, sehingga dipilih sebagai model *klasterisasi* yang paling optimal. Oleh karena itu, dua *cluster* ($k = 2$) dipilih sebagai model *klasterisasi* yang paling optimal untuk analisis selanjutnya (Sholeh & Aeni, 2023; Widiastuti & Wakit, 2022).

Tabel. 1 Statistika Deskriptif

	Cluster 2		Cluster 3			
	1	2	1	2	3	
<i>Within Cluster Sum of Square</i>	41106.93	21847.65	9835.737	23689.448	8183.857	
<i>Cluster Center</i>	psts1	21.04082	67.95652	22.00000	66.37500	20.03571
	psts2	29.46939	79.30435	28.60000	78.87500	28.67857
	psas	57.03061	76.28261	32.07500	75.47917	74.85714
	psat	80.65306	85.78261	80.60000	85.70833	80.57143
<i>Res total</i>	142495.1		142495.1			
<i>BetweenCSS / TSS</i>	55.8%		70.7%			

Dari tabel tersebut terlihat bahwa untuk dua *cluster*, nilai *Within-cluster sum of squares (WCSS)* adalah 41.106,93 dan 21.847,65, dengan rasio *Between-cluster Sum of squares*

(BCSS) terhadap Total sum of squares (TSS) sebesar 55,8%. Sedangkan pada tiga cluster, nilai WCSS masing-masing adalah 9.835,737, 23.689,448, dan 8.183,857, dengan rasio BCSS/TSS sebesar 70,7%. Nilai BCSS/TSS yang lebih tinggi pada tiga cluster menunjukkan bahwa variasi antar cluster sedikit lebih besar. Namun, hasil tersebut perlu dilihat bersama dengan homogenitas dalam cluster. Berdasarkan visualisasi dan validasi sebelumnya, pembagian menjadi dua cluster memberikan kelompok yang lebih konsisten dan mudah diinterpretasikan. Dengan demikian, dua cluster dinilai paling seimbang antara tingkat pemisahan dan keseragaman antar data, sehingga dipilih sebagai model klasterisasi yang paling optimal (Sholeh & Aeni, 2023; Diantika dkk., 2025; Widiastuti & Waktu, 2022).



Gambar 3. Visualisasi Hasil Clusterisasi (a) 3 Cluster dan (b) 2 Cluster

Pada cluster plot dua cluster ($k = 2$), data terlihat terbagi jelas ke dalam dua kelompok warna yang terpisah, menunjukkan bahwa terdapat perbedaan yang signifikan antara dua kelompok siswa berdasarkan nilai ujian matematikanya. Sementara itu, pada cluster plot tiga cluster ($k = 3$), meskipun terdapat tiga kelompok warna, beberapa titik data tampak saling tumpang tindih di area perbatasan, sehingga batas antar cluster menjadikurang tegas. Secara visual, pembentukan dua cluster memberikan hasil yang lebih mudah diinterpretasikan karena menunjukkan pemisahan yang jelas dan struktur data yang stabil. Oleh karena itu, hasil analisis menginformasikan bahwa dua cluster merupakan model klasterisasi paling optimal untuk menggambarkan perbedaan tingkat capaian hasil belajar siswa

Berdasarkan hasil klasterisasi yang diperoleh dari penerapan algoritma k-means clustering memberikan gambaran penting mengenai pola dan karakteristik hasil belajar siswa. Dua cluster yang terbentuk dapat diinterpretasikan sebagai kelompok siswa berprestasi tinggi dan kelompok siswa yang masih memerlukan pendampingan tambahan. Sholeh dan Aeni, (2023) menunjukkan bahwa penggunaan kombinasi elbow dan silhouette dapat menentukan jumlah cluster optimal secara akurat, sehingga distribusi data siswa tercermin dengan realistis. Informasi ini dapat dimanfaatkan guru sebagai dasar dalam evaluasi dan perencanaan pembelajaran diferensial. Misalnya, siswa pada cluster berprestasi tinggi dapat diberikan tantangan berupa soal Higher order thinking skills (HOTS) atau kegiatan pengayaan, sedangkan siswa pada cluster rendah dapat difokuskan pada penguatan konsep dasar dan latihan rutin untuk meningkatkan pemahaman. Hendrastuty, (2024) menegaskan bahwa pengelompokan k-means efektif membedakan kelompok siswa berprestasi tinggi dan rendah, sehingga hasilnya memiliki relevansi empiris yang kuat dalam konteks pembelajaran nyata.

(Antoni dkk., 2025) menunjukkan bahwa pengelompokan k-means dapat menjadi dasar penerapan strategi pembelajaran adaptif sesuai tingkat kemampuan akademik siswa. Hidayat, (2022) menambahkan bahwa k-means juga bermanfaat untuk menganalisis minat siswa, yang dapat meningkatkan motivasi dan keterlibatan belajar. Fatah dan Maghfiroh, (2025) menegaskan bahwa algoritma ini efektif dalam mengidentifikasi pola tersembunyi dalam data pendidikan, sehingga memperkuat perannya sebagai instrumen perencanaan pembelajaran adaptif dan pengambilan keputusan berbasis bukti.

Dengan demikian, hasil analisis klasterisasi ini tidak hanya berfungsi sebagai alat statistik, tetapi juga sebagai bahan pertimbangan pedagogis dalam meningkatkan efektivitas dan pemerataan kualitas pembelajaran matematika di SMA Negeri 4 Babelan. Temuan ini sekaligus menjawab tujuan penelitian, yaitu menerapkan algoritma k-means untuk mengelompokkan hasil belajar siswa serta menyajikan informasi yang bermanfaat bagi proses evaluasi dan pengambilan keputusan dalam pembelajaran.

SIMPULAN

Penelitian ini membuktikan bahwa algoritma k-means clustering efektif dalam mengelompokkan hasil belajar matematika siswa SMA Negeri 4 Babelan berdasarkan kesamaan capaian akademik. Melalui analisis metode elbow dan silhouette, diperoleh jumlah cluster optimal $k = 2$, yang menunjukkan tingkat homogenitas tinggi dan pemisahan kelompok yang jelas. Dua cluster tersebut merepresentasikan kelompok siswa berprestasi tinggi dan rendah, sehingga dapat menjadi dasar penerapan pembelajaran berdiferensiasi siswa berprestasi tinggi difokuskan pada pengayaan dan pengembangan HOTS, sedangkan siswa berprestasi rendah diberikan penguatan konsep dasar serta bimbingan intensif. Secara keseluruhan, penerapan metode EDM berbasis k-means tidak hanya berperan sebagai alat analisis data, tetapi juga mendukung evaluasi pembelajaran berbasis data (data-driven learning) dan menjadi landasan bagi penelitian lanjutan dalam pengembangan model analisis hasil belajar berbasis machine learning di bidang pendidikan.

DAFTAR PUSTAKA

- Akbar, I., Samad, I. S., & Rosmiana, S. (2025). Data mining analysis of k-means algorithm and decision tree for early detection of students at risk of dropping out. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 7(2), 148–162.
- Andre, S., Suciati, N., Fabroyir, H., & Pardede, E. (2023). Educational data mining clustering approach: Case study of undergraduate student thesis topic. *IEEE Access*, 11, 130072–130088. <https://doi.org/10.1109/ACCESS.2023.3332818>
- Antoni, A. R., Sartika, D., & Trianggana, D. A. (2025). Penerapan data mining dalam pengelompokan data nilai siswa SMA Negeri 7 Kota Bengkulu menggunakan metode k-means clustering. *Jurnal Media Infotama*, 21(1), 341–139.
- Diantika, M. W., Rusgiyono, A., & Saputra, B. A. (2025). Perbandingan metode optimasi silhouette, elbow, dan gap statistics dalam menentukan nilai k terbaik pada analisis k-means clustering. *Jurnal Gaussian*, 14(2), 335–344. <https://doi.org/10.14710/j.gauss.14.2.335-344>
- Fatah, Z., & Maghfiroh, S. (2025). Analisis data mining dengan algoritma k-means clustering untuk menentukan siswa berprestasi di MTs Miftahul Ulum Bengkang. *JAMASTIKA: Jurnal Mahasiswa Teknik Informatika*, 4(2).
- Fauzie, M. A. Al, Yuliadi, & Putra, J. A. (2023). Clustering data menggunakan metode k-means untuk merekomendasikan pembelajaran akademik bagi siswa aktif dalam ekstrakurikuler. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 642–648. <https://doi.org/10.30865/klik.v4i1.1116>
- Hendrastuty, N. (2024). Penerapan data mining menggunakan algoritma k-means clustering dalam evaluasi hasil pembelajaran siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Hidayat, R. (2022). Pemanfaatan data mining untuk melihat minat siswa setelah menyelesaikan pendidikan sekolah menengah atas (SMA) dengan algoritma k-means clustering. *Technology and Informatics Insight Journal*, 1(2), 85–97. <https://doi.org/10.32639/tij.v1i2.220>

- Kafillah, M., Firmansyah, D., & Lestari, K. E. (2025). Analisis sistematis pengaruh model pembelajaran berbasis masalah terhadap peningkatan hasil belajar matematika siswa SMP. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 9(2), 946–956. <https://doi.org/10.31004/cendekia.v9i2.4110>
- Pangesti, S., & Yudhanegara, M. R. (2024). Analisis hubungan antara hasil belajar matematika dengan bahasa Indonesia menggunakan k-means clustering. *Didactical Mathematics*, 6(1), 1–12. <https://doi.org/10.31949/dm.v6i1.8030>
- Pratama, A., Amalia, S. N., & Fatih, M. (2025). Pengembangan kartu flashcard materi pengurangan melalui model KAMP untuk meningkatkan self-efficacy. *Jurnal Teknologi Pendidikan (JTP)*, 10, 211–224.
- Sholeh, M., & Aeni, K. (2023). Perbandingan evaluasi metode Davies–Bouldin, elbow, dan silhouette pada model clustering menggunakan algoritma k-means. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 8(1), 56. <https://doi.org/10.30998/string.v8i1.16388>
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Tiana, M., Siswanto, S., & Kanedi, I. (2024). Penerapan metode k-means clustering dalam mengetahui minat siswa terhadap mata pelajaran matematika di SMP Negeri 19 Kota Bengkulu. *Jurnal Media Infotama*, 20(2), 617–623. <https://doi.org/10.37676/jmi.v20i2.6567>
- Widiastuti, N. A., & Wakit, A. (2022). Evaluasi kinerja algoritma k-means dengan matriks jarak Euclidean pada penentuan siswa bermasalah. *Jurnal SIMETRIS*, 13(1), 1–8.
- Yudhanegara, M. R., & Lestari, K. E. (2024). Pelatihan penggunaan algoritma k-means clustering untuk mengidentifikasi karakteristik siswa. *JALIYE: Jurnal Abdimas, Loyalitas, dan Edukasi*, 3(1), 19–25. <https://doi.org/10.47662/jaliye.v3i1.710>