



# Inventa: Journal of Science, Technology, and Innovation

Vol 1 No 3 April 2026, Hal 225-234  
ISSN: 3123-3147 (Print) ISSN: 3123-3155 (Electronic)  
Open Access: <https://scriptaintelektual.com/inventa>

## Perbandingan Kinerja K-Means dan K-Medoids dalam Klasifikasi Siswa Berprestasi di SMP Muhammadiyah 60 Medan

Farida Nafisa Siagian<sup>1\*</sup>, Halim Maulana<sup>2</sup>

<sup>1-2</sup> Universitas Muhammadiyah Sumatera Utara, Indonesia  
email: [ridasiagian31@gmail.com](mailto:ridasiagian31@gmail.com)<sup>1</sup>

### Article Info :

Received:  
20-03-2026  
Revised:  
11-04-2026  
Accepted:  
20-04-2026

### Abstract

*This study investigates the comparative performance of the K-Means and K-Medoids clustering algorithms in classifying student achievement at SMP Muhammadiyah 60 Medan. The research applies a quantitative data mining approach using academic and non-academic variables, including Mathematics scores, Science scores, Bahasa Indonesia scores, attendance records, and extracurricular participation. Data preprocessing was conducted through cleaning, normalization using Min-Max Scaling, and feature selection to ensure data consistency and analytical reliability. The findings indicate that both algorithms successfully classified students into meaningful performance groups with consistent clustering structures. K-Means demonstrated superior computational efficiency and lower SSE values, making it suitable for homogeneous datasets. In contrast, K-Medoids exhibited greater robustness against outliers and produced more stable cluster distributions. The study concludes that K-Medoids provides more representative clustering results for educational datasets characterized by heterogeneous performance patterns.*

**Keywords:** K-Means, K-Medoids, Clustering, Student Performance, Silhouette Coefficient.

### Abstrak

Penelitian ini mengkaji perbandingan kinerja algoritma pengelompokan K-Means dan K-Medoids dalam mengklasifikasikan prestasi siswa di SMP Muhammadiyah 60 Medan. Penelitian ini menerapkan pendekatan penambangan data kuantitatif dengan menggunakan variabel akademik dan non-akademik, termasuk nilai Matematika, nilai Ilmu Pengetahuan Alam, nilai Bahasa Indonesia, catatan kehadiran, dan partisipasi dalam kegiatan ekstrakurikuler. Prasunting data dilakukan melalui pembersihan, normalisasi menggunakan Min-Max Scaling, dan seleksi fitur untuk memastikan konsistensi data dan keandalan analisis. Temuan menunjukkan bahwa kedua algoritma berhasil mengklasifikasikan siswa ke dalam kelompok kinerja yang bermakna dengan struktur pengelompokan yang konsisten. K-Means menunjukkan efisiensi komputasi yang lebih baik dan nilai SSE yang lebih rendah, sehingga cocok untuk dataset homogen. Sebaliknya, K-Medoids menunjukkan ketahanan yang lebih baik terhadap outlier dan menghasilkan distribusi kluster yang lebih stabil. Studi ini menyimpulkan bahwa K-Medoids memberikan hasil pengelompokan yang lebih representatif untuk dataset pendidikan yang ditandai dengan pola kinerja heterogen.

**Kata Kunci:** K-Means, K-Medoids, Pengelompokan, Kinerja Siswa, Koefisien Silhouette.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.  
(<https://creativecommons.org/licenses/by-nc/4.0/>)

## PENDAHULUAN

Transformasi digital dalam sektor pendidikan global telah mendorong pergeseran paradigma pengambilan keputusan dari pendekatan administratif konvensional menuju pendekatan berbasis data yang menempatkan analitik komputasional sebagai instrumen strategis dalam evaluasi performa peserta didik, terutama ketika institusi pendidikan menghadapi kompleksitas data akademik dan non-akademik yang terus meningkat akibat integrasi sistem informasi sekolah, digitalisasi arsip pembelajaran, serta tuntutan akuntabilitas mutu pendidikan yang semakin tinggi dalam ekosistem Society 5.0. Dalam konteks tersebut, data mining berkembang menjadi salah satu disiplin yang memiliki posisi sentral dalam pengolahan data pendidikan karena kemampuannya mengekstraksi pola tersembunyi dari kumpulan data berskala besar untuk mendukung proses klasifikasi, segmentasi, dan prediksi perilaku akademik siswa secara lebih objektif dan terukur. Kerangka konseptual data mining yang dikemukakan oleh Jiawei Han, Micheline Kamber, dan Jian Pei menegaskan bahwa teknik clustering memiliki signifikansi tinggi dalam proses knowledge discovery karena mampu mengidentifikasi kesamaan karakteristik antarobjek tanpa memerlukan label data sebelumnya, sehingga sangat relevan digunakan dalam pemetaan performa siswa yang memiliki dimensi akademik, perilaku, dan partisipasi sosial yang

heterogen (Han et al., 2012). Perkembangan mutakhir dalam educational data mining menunjukkan bahwa algoritma clustering tidak lagi diposisikan semata sebagai alat statistik eksploratif, melainkan sebagai instrumen strategis untuk mendukung precision education, yaitu pendekatan pendidikan yang menyesuaikan intervensi pembelajaran berdasarkan karakteristik spesifik peserta didik, sehingga kebutuhan terhadap model clustering yang akurat, stabil, dan adaptif terhadap variasi data pendidikan menjadi semakin mendesak.

Sejumlah penelitian terdahulu memperlihatkan bahwa algoritma berbasis partisi, khususnya K-Means dan K-Medoids, merupakan metode clustering yang paling banyak diimplementasikan dalam analisis data pendidikan karena memiliki efisiensi komputasi serta kemampuan segmentasi yang relatif baik terhadap data multidimensi. Penelitian yang dilakukan oleh Nur Rizki Wahyudi dan kolega menunjukkan bahwa implementasi K-Means pada penentuan siswa unggulan mampu menghasilkan pengelompokan performa akademik yang cukup efektif melalui identifikasi pola nilai siswa secara otomatis, terutama ketika data memiliki struktur homogen dan distribusi yang relatif stabil (Wahyudi et al., 2024). Temuan serupa juga diperlihatkan dalam penelitian Fitri Handayani yang mengaplikasikan K-Means untuk pengelompokan mahasiswa berdasarkan gaya belajar, di mana algoritma tersebut dinilai mampu menyederhanakan kompleksitas data pendidikan menjadi struktur cluster yang lebih mudah diinterpretasikan untuk kepentingan evaluasi pembelajaran (Handayani, 2022). Di sisi lain, studi komparatif yang dilakukan oleh Muhammad Daffa Salman dan tim memperlihatkan bahwa K-Medoids memiliki ketahanan yang lebih baik terhadap noise dan outlier dalam pengelompokan sekolah berdasarkan sarana-prasarana pendidikan, terutama ketika data mengandung variasi ekstrem yang dapat menggeser posisi centroid pada K-Means (Salman et al., 2025). Penelitian I Putu Agus Vidyananta dan Kadek Tri Dermawan bahkan menegaskan bahwa efektivitas kedua algoritma sangat dipengaruhi oleh karakteristik distribusi data, ukuran sampel, serta tingkat heterogenitas variabel yang digunakan dalam proses clustering, sehingga superioritas algoritma tertentu tidak dapat digeneralisasi secara universal pada seluruh domain penelitian (Vidyananta & Dermawan, 2025).

Meskipun literatur menunjukkan perkembangan signifikan dalam penerapan clustering pada sektor pendidikan, terdapat sejumlah keterbatasan konseptual dan empiris yang masih belum terselesaikan secara memadai. Sebagian besar penelitian terdahulu cenderung berfokus pada efisiensi teknis algoritma tanpa melakukan evaluasi komprehensif terhadap relevansi pedagogis dari cluster yang dihasilkan, sehingga proses pengelompokan sering berhenti pada tahap statistik tanpa mengaitkannya dengan kebutuhan intervensi pendidikan yang nyata. Penelitian sebelumnya juga memperlihatkan inkonsistensi hasil terkait superioritas K-Means maupun K-Medoids karena sebagian studi menempatkan kecepatan komputasi sebagai indikator utama performa algoritma, sementara penelitian lain lebih menekankan stabilitas cluster terhadap outlier dan noise data. Kondisi tersebut menyebabkan belum adanya konsensus metodologis mengenai algoritma clustering yang paling representatif untuk data pendidikan yang bersifat multidimensional dan dinamis. Di samping itu, penelitian berbasis clustering dalam konteks siswa berprestasi di tingkat sekolah menengah pertama masih relatif terbatas, khususnya pada institusi pendidikan swasta berbasis keagamaan yang memiliki karakteristik data akademik dan non-akademik yang berbeda dibandingkan sekolah umum. Studi Wenny Wenny mengenai normalisasi data dengan pendekatan Min-Max memperlihatkan bahwa proses preprocessing memiliki pengaruh signifikan terhadap kualitas hasil clustering, namun kajian tersebut belum mengintegrasikan analisis komparatif antaralgoritma pada konteks data pendidikan secara spesifik (Wenny, 2024). Akibatnya, hubungan antara teknik normalisasi, sensitivitas algoritma terhadap variasi data siswa, serta kualitas cluster yang dihasilkan masih menyisakan ruang kajian yang substansial.

Ketiadaan model evaluasi clustering yang mampu mempertimbangkan simultan aspek efisiensi komputasi, stabilitas cluster, dan interpretabilitas pendidikan menimbulkan implikasi praktis yang serius terhadap proses pengambilan keputusan di sekolah, terutama ketika institusi pendidikan dituntut melakukan segmentasi siswa secara cepat namun tetap akurat untuk mendukung pembinaan akademik yang tepat sasaran. Dalam praktiknya, klasifikasi siswa berprestasi masih sering dilakukan secara manual melalui observasi subjektif guru atau sekadar berdasarkan rerata nilai akademik, padahal performa siswa sesungguhnya merupakan konstruksi multidimensional yang mencakup aspek kehadiran, konsistensi belajar, serta keterlibatan dalam aktivitas ekstrakurikuler. Ketika sekolah gagal mengidentifikasi pola performa siswa secara objektif, maka risiko kesalahan dalam pemberian intervensi pendidikan, pembinaan bakat, maupun distribusi program pengembangan akademik menjadi semakin besar. Problem tersebut menjadi lebih kompleks pada sekolah dengan jumlah siswa yang besar

dan karakteristik performa yang heterogen karena analisis manual tidak lagi memadai untuk menghasilkan klasifikasi yang konsisten. Dalam kerangka tersebut, pengujian empiris terhadap performa K-Means dan K-Medoids pada data siswa menjadi penting bukan hanya untuk menentukan algoritma dengan kualitas cluster terbaik, tetapi juga untuk menguji sejauh mana pendekatan clustering mampu diterjemahkan menjadi instrumen pengambilan keputusan pendidikan yang berbasis evidensi dan adaptif terhadap dinamika karakteristik siswa.

Penelitian ini menempatkan diri pada persimpangan antara educational data mining dan evaluasi performa algoritma clustering dengan memfokuskan analisis pada perbandingan K-Means dan K-Medoids dalam klasifikasi siswa berprestasi di SMP Muhammadiyah 60 Medan. Berbeda dengan penelitian terdahulu yang cenderung hanya mengukur kualitas cluster dari perspektif matematis, penelitian ini mengintegrasikan variabel akademik dan non-akademik secara simultan untuk membangun representasi performa siswa yang lebih komprehensif, sehingga hasil clustering tidak hanya valid secara statistik tetapi juga relevan secara substantif dalam konteks pendidikan. Posisi penelitian ini juga terletak pada upaya menguji sensitivitas kedua algoritma terhadap distribusi data siswa yang memiliki variasi performa ekstrem melalui penggunaan preprocessing berbasis normalisasi Min-Max dan evaluasi cluster menggunakan Silhouette Coefficient serta Sum of Squared Errors (SSE). Pendekatan tersebut memungkinkan penelitian ini memberikan pembacaan yang lebih kritis terhadap relasi antara karakteristik data pendidikan dan performa algoritma clustering, sekaligus memperluas cakupan literatur yang selama ini masih didominasi oleh studi berbasis dataset umum atau data institusional non-pendidikan.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengklasifikasikan siswa berprestasi berdasarkan atribut akademik dan non-akademik di SMP Muhammadiyah 60 Medan, sekaligus mengevaluasi kualitas cluster yang dihasilkan melalui indikator komputasional dan interpretasi pendidikan. Kontribusi teoretis penelitian ini terletak pada penguatan pemahaman mengenai hubungan antara karakteristik data pendidikan multidimensional dan stabilitas algoritma clustering berbasis partisi, sementara kontribusi metodologisnya diwujudkan melalui integrasi preprocessing, normalisasi data, evaluasi silhouette, serta analisis komparatif algoritma dalam satu kerangka analisis yang sistematis. Penelitian ini juga diharapkan dapat memperluas implementasi educational data mining pada level sekolah menengah pertama serta menyediakan model klasifikasi siswa yang lebih objektif, adaptif, dan berbasis data untuk mendukung pengambilan kebijakan pendidikan yang presisi.

## **METODE PENELITIAN**

Penelitian ini merupakan studi empiris dengan pendekatan kuantitatif berbasis data mining yang dirancang untuk membandingkan performa algoritma K-Means dan K-Medoids dalam klasifikasi siswa berprestasi di SMP Muhammadiyah 60 Medan. Proses eksperimental diawali dengan pengumpulan dataset akademik dan non-akademik siswa yang meliputi nilai Matematika, IPA, Bahasa Indonesia, tingkat kehadiran, dan aktivitas ekstrakurikuler, kemudian dilanjutkan dengan tahapan preprocessing yang mencakup data cleaning, eliminasi duplikasi, penanganan missing values, seleksi fitur, transformasi numerik, serta normalisasi menggunakan Min-Max Scaling guna mengurangi bias skala atribut dan meningkatkan stabilitas komputasi. Kerangka penelitian mengikuti tahapan Knowledge Discovery in Databases (KDD) yang menempatkan clustering sebagai mekanisme eksplorasi pola laten dalam data multidimensi sebagaimana dijelaskan oleh Pang-Ning Tan, Michael Steinbach, dan Vipin Kumar dalam pendekatan data mining modern (Tan et al., 2019). Implementasi algoritma dilakukan menggunakan lingkungan komputasi Python dengan simulasi iteratif berbasis Euclidean distance, di mana K-Means mengoptimalkan centroid melalui minimisasi jarak intra-cluster, sedangkan K-Medoids menggunakan medoid aktual untuk meningkatkan resistensi terhadap outlier dan distribusi data ekstrem. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow untuk memperoleh struktur segmentasi yang paling representatif terhadap pola performa siswa, sementara desain komparatif penelitian memungkinkan evaluasi langsung terhadap sensitivitas algoritma pada dataset pendidikan yang heterogen.

Validasi metodologis dilakukan melalui kombinasi evaluasi internal cluster dan interpretasi substantif hasil segmentasi guna memastikan bahwa model yang dihasilkan tidak hanya optimal secara matematis tetapi juga relevan dalam konteks pengambilan keputusan pendidikan. Kualitas clustering diukur menggunakan Silhouette Coefficient untuk mengevaluasi kohesi intra-cluster dan separasi inter-

cluster, sedangkan Sum of Squared Errors (SSE) digunakan untuk menilai tingkat kekompakan cluster melalui perhitungan jarak kuadrat antara titik data dan pusat cluster masing-masing. Ketahanan metodologis penelitian diperkuat melalui analisis komparatif terhadap stabilitas hasil clustering pada dua algoritma yang memiliki karakteristik optimisasi berbeda, sehingga memungkinkan identifikasi trade-off antara efisiensi komputasi dan resistensi terhadap noise data. Proses validasi juga mencakup interpretasi karakteristik cluster berdasarkan indikator performa akademik dan non-akademik untuk menguji konsistensi hasil klasifikasi terhadap kondisi riil siswa di lingkungan sekolah. Keunikan metodologi penelitian ini terletak pada integrasi preprocessing adaptif, evaluasi multi-metrik, dan pendekatan komparatif berbasis educational data mining dalam satu kerangka analisis sistematis, sehingga mampu menghasilkan model segmentasi siswa yang lebih robust, interpretatif, dan aplikatif bagi pengembangan sistem evaluasi pendidikan berbasis data.

## HASIL DAN PEMBAHASAN

### Normalisasi Data dan Karakteristik Awal Dataset

Dataset penelitian menunjukkan variasi performa akademik dan non-akademik yang cukup kontras antar siswa di SMP Muhammadiyah 60 Medan. Variabel yang digunakan mencakup nilai Matematika, IPA, Bahasa Indonesia, absensi, dan aktivitas ekstrakurikuler yang secara konseptual merepresentasikan dimensi multidimensional prestasi siswa. Pendekatan multidimensi semacam ini relevan dalam educational data mining karena mampu menggambarkan perilaku akademik secara lebih komprehensif dibandingkan penggunaan satu indikator tunggal (Tan et al., 2019). Integrasi variabel akademik dan perilaku sekolah juga memperkuat kemampuan clustering dalam mengidentifikasi pola laten antar siswa (Aulia et al., 2026).

Karakteristik data awal memperlihatkan bahwa siswa bernama Citra memiliki capaian akademik tertinggi dengan skor dominan pada seluruh indikator utama. Sebaliknya, Deiwi menunjukkan nilai akademik yang lebih rendah serta tingkat absensi tertinggi dibandingkan siswa lainnya. Pola kontras tersebut menghasilkan distribusi data heterogen yang secara statistik berpotensi memengaruhi sensitivitas algoritma berbasis centroid seperti K-Means (Kaligis & Yulianto, 2022). Variasi data semacam ini sering ditemukan pada penelitian clustering pendidikan berbasis evaluasi performa siswa (Antoni et al., 2025).

Distribusi nilai antar atribut menunjukkan adanya perbedaan rentang yang cukup besar antara variabel akademik dan non-akademik. Nilai absensi memiliki karakteristik invers karena angka yang lebih kecil merepresentasikan performa kehadiran yang lebih baik. Kondisi tersebut menuntut adanya proses standarisasi agar setiap variabel memiliki kontribusi yang seimbang terhadap pembentukan cluster (Wenny, 2024). Tahapan preprocessing menjadi penting karena kualitas clustering sangat dipengaruhi oleh konsistensi struktur data sebelum proses komputasi dilakukan (Han et al., 2012).

Tahap preprocessing dilakukan melalui data cleaning, pemeriksaan duplikasi, validasi missing values, dan normalisasi menggunakan Min-Max Scaling. Hasil observasi menunjukkan bahwa dataset tidak memiliki missing values maupun inkonsistensi data sehingga dapat langsung diproses pada tahap normalisasi. Min-Max Scaling digunakan untuk mentransformasikan seluruh atribut ke rentang 0–1 guna mengurangi dominasi variabel tertentu dalam perhitungan Euclidean distance (Wenny, 2024). Pendekatan ini lazim digunakan dalam clustering pendidikan karena meningkatkan kestabilan iterasi dan mengurangi bias jarak antar atribut (Handayani, 2022).

Hasil normalisasi menunjukkan perubahan struktur numerik dataset menjadi lebih proporsional untuk kebutuhan clustering. Citra memperoleh nilai normalisasi tertinggi pada hampir seluruh atribut akademik dan ekstrakurikuler, sedangkan Deiwi memiliki nilai minimum pada sebagian besar indikator performa. Pola tersebut memperlihatkan adanya jarak performa yang cukup signifikan antar siswa sehingga memungkinkan pembentukan cluster yang jelas. Kondisi ini mendukung asumsi dasar clustering bahwa data memiliki tingkat similaritas internal dan separasi eksternal yang dapat diidentifikasi secara matematis (Tan et al., 2019).

**Tabel 1. Hasil Normalisasi Dataset Siswa**

Nama	Mtk	IPA	Bhs	Absensi	Ekskul
Ahmad	0.7407	0.6923	0.7857	0.1428	0.6667
Buidi	0.4814	0.1923	0.4286	0.5714	0.3333

Citra	1.0000	1.0000	1.0000	0.0000	1.0000
Deiwi	0.0000	0.0000	0.0000	1.0000	0.0000
Eiko	0.8518	0.5385	0.6429	0.2857	0.8333

Sumber: Hasil preprocessing data penelitian menggunakan Min-Max Scaling berbasis Python, 2026.

Data pada Tabel 1 memperlihatkan bahwa normalisasi berhasil mempertahankan pola distribusi asli tanpa menghilangkan karakteristik ekstrem pada dataset. Nilai absensi Deiwi yang mencapai 1,0000 menunjukkan tingkat ketidakhadiran tertinggi sehingga berpotensi membentuk cluster tersendiri pada proses segmentasi. Kondisi tersebut memperkuat argumen bahwa preprocessing tidak bertujuan menghapus variasi data, melainkan menyeragamkan skala atribut agar proses clustering berjalan optimal (Han et al., 2012). Stabilitas data hasil normalisasi menjadi faktor penting dalam meningkatkan akurasi segmentasi siswa berbasis data mining (Noyari et al., 2024).

Penentuan jumlah cluster dilakukan menggunakan metode Elbow dengan membandingkan perubahan nilai SSE pada beberapa skenario jumlah cluster. Pengujian menunjukkan bahwa titik penurunan SSE mulai melandai pada nilai  $k = 2$  sehingga konfigurasi tersebut dipilih sebagai struktur cluster optimal. Pemilihan dua cluster dinilai paling representatif karena mampu memisahkan siswa dengan performa tinggi dan rendah secara jelas. Hasil ini sejalan dengan penelitian Salman et al. (2025) yang menunjukkan bahwa metode Elbow efektif dalam menentukan jumlah cluster optimal pada dataset pendidikan dengan ukuran terbatas.

Pemilihan dua cluster juga dipengaruhi oleh ukuran dataset yang relatif kecil sehingga penggunaan cluster lebih banyak dapat menghasilkan segmentasi yang terlalu sempit. Struktur cluster yang terlalu kecil berisiko menghasilkan over-segmentation dan menurunkan interpretabilitas hasil clustering. Pada konteks educational data mining, interpretabilitas cluster memiliki nilai penting karena hasil segmentasi akan digunakan sebagai dasar pengambilan keputusan pendidikan (Putri et al., 2025). Faktor interpretatif tersebut menjadi salah satu pertimbangan utama dalam desain penelitian ini.

Kondisi dataset menunjukkan bahwa karakteristik performa siswa di SMP Muhammadiyah 60 Medan memiliki pola distribusi yang cukup jelas setelah preprocessing dilakukan. Variabel akademik dan non-akademik saling membentuk struktur data yang konsisten sehingga proses clustering dapat berjalan lebih stabil. Integrasi preprocessing adaptif dan normalisasi numerik terbukti memperkuat kualitas data sebelum memasuki tahap komputasi algoritma clustering. Temuan ini memperlihatkan bahwa kualitas hasil clustering tidak hanya ditentukan oleh algoritma, tetapi juga oleh ketepatan pengolahan data pada tahap awal penelitian (Agni, 2025).

### Implementasi Clustering dan Struktur Segmentasi Siswa

Implementasi algoritma clustering pada dataset siswa SMP Muhammadiyah 60 Medan menunjukkan bahwa pola segmentasi performa siswa terbentuk secara cukup jelas setelah proses optimasi centroid dan medoid dilakukan secara iteratif. Struktur cluster yang dihasilkan memperlihatkan adanya pemisahan antara kelompok siswa dengan performa akademik tinggi dan siswa dengan performa akademik rendah berdasarkan kombinasi atribut akademik dan non-akademik. Karakteristik tersebut memperlihatkan bahwa pendekatan clustering mampu mengidentifikasi pola laten pada data pendidikan multidimensi secara efektif, khususnya pada atribut nilai akademik, absensi, dan aktivitas ekstrakurikuler (Tan et al., 2019). Temuan ini sejalan dengan penelitian Antoni et al. (2025) yang menjelaskan bahwa algoritma clustering efektif digunakan untuk memetakan variasi performa siswa dalam lingkungan pendidikan berbasis data mining.

Proses implementasi K-Means memperlihatkan bahwa algoritma mampu mencapai kondisi konvergen pada iterasi kedua setelah pembaruan centroid dilakukan secara berulang. Stabilitas iterasi tersebut menunjukkan bahwa distribusi data siswa relatif homogen pada cluster utama sehingga perpindahan anggota cluster tidak lagi terjadi secara signifikan. Fenomena ini mengindikasikan bahwa atribut akademik memiliki kontribusi dominan terhadap pembentukan pola segmentasi dibandingkan atribut lainnya, terutama pada siswa dengan performa konsisten tinggi seperti Citra dan Ahmad. Pola serupa juga ditemukan oleh Wahyudi et al. (2024) yang menunjukkan bahwa konvergensi cepat pada K-Means umumnya muncul pada dataset pendidikan dengan tingkat heterogenitas sedang.

Karakteristik cluster yang terbentuk memperlihatkan bahwa Cluster 1 diisi oleh siswa dengan kombinasi nilai akademik tinggi, tingkat absensi rendah, dan partisipasi ekstrakurikuler aktif. Citra menjadi anggota dengan posisi paling dominan dalam cluster tersebut karena memiliki nilai tertinggi

hampir pada seluruh atribut yang digunakan dalam penelitian. Sebaliknya, Deiwi membentuk posisi terpisah pada Cluster 2 akibat kombinasi nilai akademik rendah dan tingkat ketidakhadiran yang paling tinggi dibandingkan siswa lain. Pola segmentasi seperti ini menunjukkan bahwa absensi memiliki pengaruh kuat terhadap separasi cluster karena secara langsung memengaruhi kedekatan jarak antar data pada proses clustering (Asher et al., 2025).

Hubungan antara aktivitas ekstrakurikuler dan performa akademik juga terlihat cukup kuat pada hasil segmentasi yang terbentuk. Ahmad, Citra, dan Eiko menunjukkan kecenderungan berada dalam cluster performa tinggi karena memiliki keseimbangan antara pencapaian akademik dan partisipasi kegiatan sekolah. Kondisi tersebut memperlihatkan bahwa keterlibatan siswa dalam aktivitas non-akademik tidak menurunkan performa akademik, melainkan memperkuat konsistensi karakteristik cluster. Temuan ini mendukung penelitian Putriana et al. (2023) yang menyatakan bahwa atribut aktivitas tambahan mampu meningkatkan kualitas pemetaan cluster dalam analisis performa individu berbasis data mining.

**Tabel 2. Struktur Hasil Clustering K-Means dan K-Medoids**

Nama	Jarak ke C1	Jarak ke C2	Cluster
Ahmad	0.1900	1.6803	1
Buidi	0.7660	0.8645	1
Citra	0.6312	2.2361	1
Deiwi	1.6224	0.0000	2
Eiko	0.1730	1.6221	1

Sumber: Hasil olahan data penelitian menggunakan Python, 2026.

Data pada Tabel 2 memperlihatkan bahwa Deiwi memiliki jarak nol terhadap pusat Cluster 2 sehingga bertindak sebagai titik paling representatif dalam kelompok performa rendah. Kondisi ini memperlihatkan bahwa posisi Deiwi sangat berbeda dibandingkan siswa lain sehingga membentuk pemisahan cluster yang cukup tegas. Sebaliknya, Ahmad dan Eiko memiliki jarak yang sangat dekat terhadap pusat Cluster 1 sehingga keduanya menjadi representasi utama kelompok performa tinggi. Struktur tersebut memperlihatkan bahwa distribusi data dalam Cluster 1 cenderung lebih padat dibandingkan Cluster 2 yang hanya terdiri dari satu anggota dengan karakteristik ekstrem (Amaliah et al., 2023).

Keberadaan Buidi dalam Cluster 1 menunjukkan karakteristik yang berbeda dibanding anggota cluster lainnya karena nilai jaraknya terhadap kedua cluster relatif berdekatan. Posisi tersebut menunjukkan bahwa Buidi berada pada area borderline yang secara matematis memiliki potensi berpindah cluster apabila terjadi perubahan kecil pada atribut performa akademik maupun absensi. Fenomena borderline pada clustering pendidikan sering muncul ketika terdapat siswa dengan performa moderat yang tidak sepenuhnya identik dengan kelompok tertentu. Karakteristik seperti ini juga ditemukan dalam penelitian Ih'Diati et al. (2024) yang menjelaskan bahwa siswa dengan pola performa menengah cenderung memiliki sensitivitas lebih tinggi terhadap perubahan pusat cluster.

Implementasi K-Medoids menunjukkan struktur segmentasi yang konsisten dengan hasil K-Means meskipun pendekatan optimasi yang digunakan berbeda secara konseptual. Penggunaan medoid aktual membuat pusat cluster tetap berada pada titik data nyata sehingga distribusi cluster menjadi lebih stabil terhadap pengaruh nilai ekstrem. Pada dataset ini, Deiwi tetap dipertahankan sebagai representasi Cluster 2 karena karakteristik nilainya berbeda cukup jauh dibanding anggota lain. Temuan tersebut mendukung hasil penelitian Puri et al. (2024) yang menjelaskan bahwa K-Medoids memiliki ketahanan lebih baik dalam mempertahankan struktur cluster ketika data mengandung pola ekstrem atau noise.

Dari perspektif metodologis, struktur segmentasi yang konsisten antara kedua algoritma memperlihatkan bahwa pola performa siswa memang secara alami terbagi menjadi dua kelompok utama. Kondisi tersebut mengindikasikan bahwa atribut akademik, absensi, dan ekstrakurikuler memiliki korelasi yang cukup kuat dalam membentuk distribusi performa siswa. Kestabilan hasil clustering juga memperlihatkan bahwa proses preprocessing sebelumnya berhasil mengurangi bias skala atribut sehingga jarak Euclidean dapat bekerja secara optimal dalam proses segmentasi. Pola ini sejalan dengan kajian Kaligis dan Yulianto (2022) yang menyebutkan bahwa konsistensi cluster antar algoritma menjadi indikator penting dalam validasi kualitas segmentasi data.

Implementasi clustering pada penelitian ini juga menunjukkan bahwa ukuran dataset kecil tidak selalu menghasilkan segmentasi yang lemah apabila atribut yang digunakan mampu merepresentasikan karakteristik objek secara komprehensif. Pemilihan variabel akademik dan non-akademik membuat struktur cluster lebih interpretatif karena mencerminkan kondisi riil siswa di lingkungan sekolah. Segmentasi yang terbentuk memberikan gambaran bahwa siswa dengan performa akademik tinggi cenderung memiliki kedisiplinan kehadiran yang lebih baik dibandingkan kelompok performa rendah. Temuan tersebut mendukung penelitian Aulia et al. (2026) yang menjelaskan bahwa integrasi atribut akademik dan perilaku mampu meningkatkan kualitas interpretasi clustering dalam educational data mining.

Hasil implementasi kedua algoritma memperlihatkan bahwa pendekatan clustering dapat digunakan sebagai dasar segmentasi siswa untuk mendukung pengambilan keputusan pendidikan berbasis data. Struktur cluster yang terbentuk memungkinkan sekolah melakukan identifikasi dini terhadap siswa yang membutuhkan pembinaan akademik maupun pendampingan kedisiplinan. Selain menghasilkan pemetaan performa siswa secara objektif, clustering juga membantu sekolah memahami hubungan antar atribut performa yang sebelumnya sulit diamati secara manual. Pendekatan ini memiliki relevansi tinggi terhadap pengembangan sistem evaluasi pendidikan modern yang menekankan integrasi analitik data dan pengambilan keputusan adaptif (Syafrudin et al., 2025).

### Evaluasi Kinerja Algoritma dan Validasi Cluster

Evaluasi performa algoritma clustering dilakukan untuk mengidentifikasi kualitas segmentasi siswa berdasarkan kedekatan intra-cluster dan separasi antar-cluster pada data akademik dan non-akademik siswa SMP Muhammadiyah 60 Medan. Tahapan evaluasi ini difokuskan pada pengukuran efektivitas model menggunakan indikator Sum of Squared Errors (SSE), Silhouette Coefficient, stabilitas cluster, serta sensitivitas algoritma terhadap variasi data ekstrem. Pendekatan evaluatif semacam ini penting dalam penelitian educational data mining karena kualitas cluster tidak hanya diukur dari akurasi matematis, tetapi juga dari kemampuan model membentuk segmentasi yang interpretatif terhadap kondisi siswa (Tan et al., 2019). Kajian komparatif antara K-Means dan K-Medoids menunjukkan bahwa kedua algoritma menghasilkan pola pengelompokan yang relatif konsisten, meskipun memiliki karakteristik optimisasi yang berbeda pada proses iterasi dan pembentukan pusat cluster.

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow melalui pengamatan pola penurunan SSE pada beberapa skenario nilai  $k$ . Visualisasi Elbow menunjukkan bahwa penurunan SSE mengalami perlambatan signifikan pada  $k = 2$  sehingga konfigurasi tersebut dipilih sebagai struktur cluster paling representatif untuk dataset penelitian. Kondisi tersebut mengindikasikan bahwa pembentukan dua cluster telah mampu merepresentasikan distribusi performa siswa tanpa menghasilkan fragmentasi kelompok yang terlalu kecil, sebagaimana juga ditemukan dalam studi segmentasi pendidikan oleh Salman et al. (2025). Interpretasi metodologis terhadap pola elbow memperlihatkan bahwa distribusi data siswa memiliki kecenderungan polaritas performa yang cukup jelas antara kelompok dengan capaian akademik tinggi dan kelompok dengan performa rendah.

**Tabel 3. Perbandingan Kinerja Algoritma K-Means dan K-Medoids**

Algoritma	SSE	Silhouette Coefficient	Stabilitas Cluster	Sensitivitas Outlier
K-Means	1,0512	0,544	Sedang	Tinggi
K-Medoids	1,1835	0,544	Tinggi	Rendah

Sumber: Hasil Pengolahan Data Penelitian Menggunakan Python, 2026.

Berdasarkan Tabel 3, terlihat bahwa algoritma K-Means menghasilkan nilai SSE sebesar 1,0512 yang lebih rendah dibandingkan K-Medoids dengan nilai 1,1835. Nilai tersebut menunjukkan bahwa jarak rata-rata titik data terhadap centroid pada K-Means lebih kecil sehingga cluster yang terbentuk cenderung lebih kompak secara geometris. Temuan ini sejalan dengan penelitian Kaligis dan Yulianto (2022) yang menyatakan bahwa mekanisme centroid pada K-Means lebih efektif dalam meminimalkan error kuadrat pada dataset dengan distribusi relatif homogen. Karakteristik tersebut memperlihatkan

bahwa pendekatan berbasis rata-rata mampu meningkatkan efisiensi pemadatan cluster meskipun tetap memiliki kelemahan terhadap distribusi data yang mengandung nilai ekstrem.

Meskipun K-Means menghasilkan SSE lebih rendah, evaluasi terhadap stabilitas cluster memperlihatkan bahwa K-Medoids memiliki resistensi yang lebih baik terhadap noise dan outlier. Penggunaan medoid aktual sebagai pusat cluster menyebabkan perpindahan posisi pusat kelompok tidak mudah terdistorsi oleh perubahan nilai ekstrem pada atribut tertentu, terutama pada variabel kehadiran dan aktivitas ekstrakurikuler siswa. Pola serupa juga ditemukan oleh Pratama dan Purwanto (2025) serta Syafrudin et al. (2025) yang menegaskan bahwa K-Medoids lebih stabil pada dataset heterogen dengan variasi antarobjek yang cukup tinggi. Dalam konteks data pendidikan, kestabilan cluster menjadi penting karena distribusi performa siswa sering kali dipengaruhi oleh ketimpangan nilai akademik maupun perilaku belajar yang tidak seragam.

Nilai Silhouette Coefficient sebesar 0,544 pada kedua algoritma menunjukkan bahwa struktur cluster berada pada kategori baik dengan tingkat separasi antarcluster yang cukup jelas. Nilai tersebut mengindikasikan bahwa sebagian besar objek data memiliki kedekatan lebih tinggi terhadap anggota cluster-nya dibandingkan cluster lain sehingga kualitas segmentasi dapat dianggap representatif secara statistik. Interpretasi ini konsisten dengan parameter evaluasi cluster yang dijelaskan oleh Azizah et al. (2024) dan Seran et al. (2026), di mana nilai silhouette di atas 0,50 mencerminkan kualitas pengelompokan yang stabil dan tidak mengalami tumpang tindih signifikan. Kondisi tersebut memperlihatkan bahwa struktur segmentasi siswa pada penelitian ini telah memenuhi standar validasi internal clustering berbasis kohesi dan separasi.

Analisis lebih lanjut menunjukkan bahwa terdapat satu objek data yang memiliki kecenderungan borderline karena berada pada posisi yang relatif dekat dengan dua cluster sekaligus. Fenomena tersebut muncul akibat adanya kemiripan parsial antara performa akademik dan non-akademik siswa tertentu sehingga proses alokasi cluster menjadi lebih sensitif terhadap perubahan pusat kelompok. Putra et al. (2025) menjelaskan bahwa kondisi borderline sering muncul pada data multidimensi yang memiliki variabel dengan kontribusi tidak sepenuhnya dominan terhadap pembentukan cluster. Keberadaan objek borderline pada penelitian ini tidak menurunkan kualitas cluster secara keseluruhan, tetapi menunjukkan bahwa karakteristik siswa bersifat dinamis dan tidak selalu dapat dipisahkan secara absolut.

Efektivitas algoritma juga dianalisis melalui aspek efisiensi komputasi selama proses iterasi clustering berlangsung. K-Means menunjukkan konvergensi yang lebih cepat karena proses pembaruan centroid dilakukan melalui perhitungan rata-rata matematis yang relatif sederhana dibandingkan mekanisme pertukaran medoid pada K-Medoids. Temuan ini mendukung hasil penelitian Susanto dan Riska (2026) serta Vidyananta dan Dermawan (2025) yang menempatkan K-Means sebagai algoritma dengan efisiensi komputasi lebih tinggi untuk dataset berukuran besar maupun distribusi data yang stabil. Dari perspektif implementasi sistem pendidikan berbasis data, efisiensi tersebut memberikan keuntungan pada proses analisis berkala yang membutuhkan waktu pemrosesan cepat.

Di sisi lain, K-Medoids memperlihatkan performa yang lebih adaptif ketika dataset mengandung variasi nilai yang cukup tajam antarobjek. Penggunaan data aktual sebagai pusat cluster membuat algoritma mampu mempertahankan konsistensi struktur segmentasi meskipun terdapat perubahan distribusi atribut tertentu. Karakteristik tersebut sesuai dengan temuan Salman et al. (2025) dan Ih'Diati et al. (2024) yang menyatakan bahwa K-Medoids lebih representatif dalam membangun cluster pada data sosial dan pendidikan yang memiliki potensi noise tinggi. Dalam penelitian ini, resistensi terhadap outlier menjadi faktor penting karena atribut absensi dan aktivitas ekstrakurikuler memiliki rentang variasi yang cukup kontras dibandingkan nilai akademik utama.

Validasi cluster secara substantif memperlihatkan bahwa hasil segmentasi memiliki relevansi praktis terhadap konteks pengambilan keputusan pendidikan. Struktur cluster yang terbentuk mampu membedakan kelompok siswa dengan performa akademik dan partisipasi sekolah yang relatif baik dari kelompok dengan capaian lebih rendah tanpa menghasilkan tumpang tindih kategori yang dominan. Pola tersebut mendukung pendekatan educational data mining yang menempatkan clustering sebagai instrumen identifikasi karakteristik belajar siswa secara objektif dan berbasis data empiris (Han et al., 2012). Relevansi praktis ini memperkuat posisi clustering sebagai metode pendukung evaluasi pendidikan yang dapat membantu sekolah dalam menyusun strategi pembinaan yang lebih terarah.

Interpretasi metodologis terhadap hasil komparasi menunjukkan adanya trade-off antara efisiensi komputasi dan ketahanan struktur cluster. K-Means lebih unggul dalam meminimalkan error serta

mempercepat proses iterasi, sedangkan K-Medoids lebih konsisten dalam mempertahankan kestabilan segmentasi ketika dataset mengandung variasi ekstrem. Keseimbangan antara dua aspek tersebut juga ditemukan pada penelitian Putri et al. (2025), Darmawan et al. (2023), dan Aulia et al. (2026) yang menegaskan bahwa pemilihan algoritma clustering harus mempertimbangkan karakteristik distribusi data serta tujuan analisis yang ingin dicapai. Berdasarkan karakteristik dataset penelitian ini, K-Medoids menunjukkan kualitas interpretatif yang lebih baik untuk segmentasi siswa, sedangkan K-Means tetap relevan untuk kebutuhan analisis cepat dengan kompleksitas data yang lebih rendah.

## KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means dan K-Medoids mampu menghasilkan pengelompokan siswa yang efektif berdasarkan atribut akademik dan non-akademik, sehingga dapat mendukung proses klasifikasi prestasi siswa secara lebih objektif dan terstruktur. Hasil evaluasi menggunakan Silhouette Coefficient dan Sum of Squared Errors (SSE) memperlihatkan bahwa kedua algoritma menghasilkan kualitas cluster yang baik dengan pola pengelompokan yang konsisten. K-Means memiliki keunggulan pada efisiensi komputasi dan kemampuan meminimalkan error clustering, sehingga lebih sesuai untuk dataset homogen dan berukuran besar. Namun, sensitivitas terhadap outlier menyebabkan stabilitas cluster cenderung menurun ketika data memiliki variasi ekstrem. Sebaliknya, K-Medoids menunjukkan ketahanan yang lebih tinggi terhadap outlier karena menggunakan medoid sebagai pusat cluster, sehingga menghasilkan struktur cluster yang lebih stabil dan representatif. Temuan ini menegaskan bahwa pemilihan algoritma clustering harus mempertimbangkan karakteristik data yang digunakan, di mana K-Medoids lebih efektif untuk dataset pendidikan dengan heterogenitas tinggi, sedangkan K-Means lebih optimal untuk kebutuhan analisis yang menuntut efisiensi komputasi.

## DAFTAR PUSTAKA

- Agni, M. A. M. (2025). Optimalisasi Pengelompokan Jenis Produk dalam Penentuan Strategi Penjualan Menggunakan Algoritma K-Means di Toko Cirebon Vape Store. *MULTINETICS*, 11(1), 1-10. <https://doi.org/10.32722/multinetics.v11i1.7315>
- Amaliah, R. A., Tohidi, E., Wahyudin, E., Rinaldi, A. R., & Iin, I. (2023). Pengelompokan Data Bencana Alam Berdasarkan Wilayah Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3572-3579. <https://doi.org/10.36040/jati.v7i6.8253>
- 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), 299-306. <https://doi.org/10.37676/jmi.v21i1.8101>
- Asher, C., Fredricka, J., & Alinse, R. T. (2025). Penerapan Metode K-Means Dalam Pengelompokan Data Siswa Berdasarkan Masalah Akademik Di SMA Negeri Selangit. *JURNAL MEDIA INFOTAMA*, 21(2), 615-623. <https://doi.org/10.37676/jmi.v21i2.9370>
- Aulia, H. S., Arifin, M., & Fithri, D. L. (2026). Penerapan Algoritma Machine Learning Untuk Pengelompokan Siswa Berdasarkan Aspek Akademik Dan Non-Akademik. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 11(1), 1200-1210. <https://doi.org/10.36341/rabit.v11i1.7249>
- Azizah, M., Rahaningsih, N., & Dana, R. D. (2024). Analisis Klasterisasi Wilayah Penyandang Disabilitas di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 638-644. <https://doi.org/10.36040/jati.v8i1.8398>
- Darmawan, A., Yudhisari, I., Anwari, A., & Makruf, M. (2023). Pola prediksi kelulusan siswa madrasah aliyah swasta dengan support vector machine dan random forest. *Jurnal Minfo Polgan*, 12(1), 387-400. <https://doi.org/10.33395/jmp.v12i1.12388>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal teknologi dan Informasi*, 12(1), 46-63. <https://doi.org/10.34010/jati.v12i1.6733>
- Ih'Diati, N. O., Anto, A., & Rosmini, R. (2024). Implementasi Algoritma K-Medoids Dalam Mengelompokkan Siswa Berdasarkan Keaktifan Dalam Proses Pembelajaran. *Journal of Big Data Analytic and Artificial Intelligence*, 7(1), 1-7. <https://doi.org/10.71302/jbidai.v7i1.38>

- Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 1(3), 179-193. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Noyari, J. A., Aprillia, A., Munthe, R. G., Sutarman, A., & Kallas, E. (2024). Optimasi Kinerja Sistem Informasi Manajemen Kampus Menggunakan Teknik Data Mining: Optimizing the Performance of Campus Management Information Systems Using Data Mining Techniques. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, 3(1), 52-63. <https://doi.org/10.33050/mentari.v3i1.613>
- Pratama, B. S., & Purwanto, G. (2025). Perbandingan K-Means dan K-Medoids dalam Pengelompokan Tingkat Kejahatan pada Provinsi Jawa Tengah. *IDEALIS: InDonEsiA journal Information System*, 8(2), 295-303. <https://doi.org/10.36080/idealis.v8i2.3562>
- Puri, A., Solihudin, D., Anwar, S., Pratama, D., & Wahyudin, E. (2024). Analisis Kluster K-Medoid Untuk Pengelompokan Dan Pemetaan Provinsi Di Indonesia Berdasarkan Nilai Ujian Nasional. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 918-924. <https://doi.org/10.36040/jati.v8i1.8653>
- Putra, J. L., Kanedi, I., & Al Akbar, A. (2025). Klasterisasi Data Karyawan Berdasarkan Penilaian Kinerja Menggunakan Metode K-Medoid. *JURNAL MEDIA INFOTAMA*, 21(1), 143-151. <https://doi.org/10.37676/jmi.v21i1.8178>
- Putri, A., Jasmir, J., & Purnama, B. (2025). Klasifikasi Kelayakan Penerima Beasiswa Menggunakan Naive Bayes dengan Optimasi Atribut Berbasis K-Means Clustering: Classification of Scholarship Eligibility Using Naïve Bayes with Attribute Optimization Based on K-Means Clustering. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(4), 1450-1462. <https://doi.org/10.57152/malcom.v5i4.2312>
- Putriana, P., Suarna, N., & Prihartono, W. (2023). Analisis Clustering Prestasi Atlet Pada Berbagai Cabang Olahraga Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3435-3442. <https://doi.org/10.36040/jati.v7i6.8211>
- Salman, M. D., Rahmaddeni, R., Pratama, N. R., Setiawan, A. A., Zalianti, F., & Huda, I. B. (2025). Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids dalam Pengelompokan Sekolah di Provinsi Riau Berdasarkan Ketersediaan Sarana dan Prasarana: Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Province Based on Availability of Facilities and Infrastructure. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 797-806. <https://doi.org/10.57152/malcom.v5i3.1950>
- Seran, K. J. T., Soares, J. T., Bobu, F. R., & Chrisinta, D. (2026). Analisis Perbandingan Algoritma K-Means dan K-Medoids dalam Penentuan Status Gizi Balita. *JIKTEKS: Jurnal Ilmu Komputer dan Teknologi Informasi*, 4(02), 42-50. <https://doi.org/10.70404/jikteks.v4i02.648>
- Susanto, S. P. A., & Riska, S. Y. (2026). Perbandingan K-Means dan K-Medoids Untuk Clustering Lagu Setipe di Spotify Berdasarkan Karakteristik Audio. *INTEGER: Journal of Information Technology*, 11(1). <https://doi.org/10.31284/j.integer.2026.v11i1.8394>
- Syafrudin, T., Hermawan, A., Avianto, D., & Maulana, I. (2025). Analisis Komparasi Algoritma K-Means Dan K-Medoids Dalam Segmentasi Data Untuk Strategi Promosi Mahasiswa Baru Di Universitas X. *Komputika: Jurnal Sistem Komputer*, 14(2), 203-211. <https://doi.org/10.34010/komputika.v14i2.16698>
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.
- Vidyananta, I. P. A., & Dermawan, K. T. (2025). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Keadaan Sekolah Dasar. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.8145>
- Wahyudi, N. R., Rahmawati, Y., & Supriyanto, S. (2024). Implementasi Rapidminer Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means. *SKANIKA: Sistem Komputer dan Teknik Informatika*, 7(2), 109-118. <https://doi.org/10.36080/skanika.v7i2.3173>
- Wenny, W. (2024). Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokan Tingkat Ekonomi Masyarakat. *Bulletin of Information System Research*, 2(2), 53-63. <https://doi.org/10.62866/bios.v2i2.141>