PRINSIP : Portal Riset & Inovasi Sistem Perangkat Lunak
Volume 4, Nomor 3, Juli 2026 e-ISSN : 2985-802X
http://doi.org/10.59696/prinsip.v4i3.243 p-ISSN : 2987-0534

Perbandingan Kinerja K-Means Dan K-Medoids Dalam Klasifikasi Siswa Berprestasi Di SMP Muhammadiyah 60 Medan

[1]Farida Nafisa Siagian, [2] Halim Maulana
Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
ridasiagian31@gmail.com

Submit : 15 Juni 2026 | **Diterima** : 25 Juni 2026 | **Terbit** : 26 Juli 2026

ABSTRAK

Pengelompokan data siswa berprestasi merupakan langkah penting dalam proses evaluasi dan pengambilan keputusan di lingkungan pendidikan. Metode clustering seperti K-Means dan K-Medoids banyak digunakan untuk mengelompokkan data berdasarkan prestasi akademik, kehadiran, dan aktivitas ekstrakurikuler. Penelitian ini bertujuan untuk membandingkan kinerja kedua algoritma tersebut dalam mengelompokkan siswa berprestasi di SMP Muhammadiyah 60 Medan. Metode yang digunakan meliputi proses clustering data menggunakan algoritma K-Means dan K-Medoids, serta evaluasi hasil menggunakan nilai silhouette untuk menilai kualitas cluster. Hasil penelitian menunjukkan bahwa kedua metode mampu mengelompokkan siswa ke dalam kategori berprestasi tinggi, sedang, dan rendah. K-Means lebih cepat dalam komputasi dan cocok untuk data yang homogen serta berukuran besar, namun sensitif terhadap outlier. Sebaliknya, K-Medoids lebih stabil dan tahan terhadap outlier meskipun membutuhkan waktu komputasi yang lebih lama. Berdasarkan hasil evaluasi, K-Medoids memberikan hasil yang lebih representatif pada kondisi data dengan variasi nilai ekstrem. Penelitian ini diharapkan dapat membantu pihak sekolah dalam melakukan segmentasi siswa secara efektif dan efisien guna mendukung pengambilan kebijakan yang lebih tepat sasaran.
Kata Kunci: K-Means; K-Medoids; Clustering; Kinerja Siswa; Nilai Silhouette.

PENDAHULUAN

SMP Muhammadiyah 60 Medan merupakan salah satu lembaga pendidikan swasta yang memiliki peran penting dalam menghasilkan siswa berprestasi. Sekolah ini menampung siswa dengan latar belakang akademik dan non-akademik yang beragam, sehingga penting bagi pihak sekolah untuk memahami pola prestasi siswa. Identifikasi pola tersebut memungkinkan sekolah untuk merancang strategi pembinaan yang tepat serta memberikan intervensi yang lebih terarah guna meningkatkan kinerja siswa.

Namun demikian, sekolah menghadapi tantangan dalam mengklasifikasikan siswa ke dalam kategori prestasi berdasarkan data yang tersedia, seperti nilai Matematika, IPA, Bahasa Indonesia, catatan kehadiran, dan aktivitas ekstrakurikuler. Volume data siswa yang cukup besar serta kompleksitas analisis secara manual menyebabkan proses klasifikasi menjadi sulit, memakan waktu, dan rentan terhadap subjektivitas. Akibatnya, pihak sekolah mengalami kesulitan dalam membedakan secara akurat antara siswa berprestasi tinggi dan siswa dengan prestasi rendah.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan analitis berbasis teknologi yang mampu mengolah data siswa dan mengidentifikasi pola yang bermakna. Data mining menjadi salah satu solusi yang potensial, khususnya melalui teknik clustering, yang dapat mengelompokkan siswa berdasarkan kemiripan data performa mereka. Pendekatan ini memungkinkan sekolah memperoleh wawasan yang lebih objektif dan berbasis data dalam mengklasifikasikan tingkat prestasi siswa.

Penelitian sebelumnya menunjukkan bahwa algoritma K-Means banyak digunakan karena kesederhanaan dan efisiensi komputasinya, terutama dalam menangani dataset berukuran besar. Namun, algoritma ini sensitif terhadap outlier yang dapat memengaruhi akurasi hasil clustering (Ahmed & Ashour, 2022). Di sisi lain, algoritma K-Medoids memiliki ketahanan yang lebih baik terhadap outlier karena menggunakan data aktual sebagai pusat cluster, sehingga menghasilkan hasil clustering yang lebih stabil, meskipun membutuhkan waktu komputasi yang lebih lama (Hardiyani et al., 2019).

Selain itu, Wahyudi (2020) melaporkan bahwa K-Means unggul dalam efisiensi waktu komputasi, sementara Ahyudi (2023) menemukan bahwa K-Medoids menghasilkan clustering yang lebih akurat pada dataset dengan variasi tinggi.

Berdasarkan temuan tersebut, terdapat perbedaan kinerja antara K-Means dan K-Medoids yang mendorong perlunya penelitian lebih lanjut, khususnya dalam konteks data pendidikan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengklasifikasikan siswa berprestasi di SMP Muhammadiyah 60 Medan. Perbandingan difokuskan pada kualitas clustering dan karakteristik algoritma untuk menentukan metode yang paling sesuai dalam analisis data pendidikan.

Tujuan dari penelitian ini adalah: (1) menghasilkan klasifikasi siswa berprestasi, (2) membandingkan kinerja algoritma K-Means dan K-Medoids dalam klasifikasi siswa, dan (3) mengembangkan sistem perangkat lunak untuk menganalisis serta membandingkan kinerja kedua algoritma tersebut. Ruang lingkup penelitian ini dibatasi pada data siswa SMP Muhammadiyah 60 Medan dengan menggunakan atribut nilai akademik, kehadiran, dan aktivitas ekstrakurikuler.

Diharapkan hasil penelitian ini dapat membantu pihak sekolah dalam mengklasifikasikan siswa secara lebih objektif dan akurat, serta memberikan wawasan mengenai efektivitas algoritma clustering dalam bidang pendidikan.

TINJAUAN PUSTAKA

Data mining didefinisikan sebagai proses mengekstraksi informasi yang bermakna dan bernilai dari kumpulan data yang besar dengan menggunakan teknik statistik, matematika, dan kecerdasan buatan (Setiawan, 2021). Data mining juga sering disebut sebagai *Knowledge Discovery in Databases* (KDD), yang melibatkan beberapa tahapan, yaitu seleksi data, preprocessing, transformasi, proses data mining, dan evaluasi. Fungsi utama data mining meliputi analisis deskriptif yang bertujuan untuk memahami pola dalam data, serta analisis prediktif yang digunakan untuk memprediksi nilai yang belum diketahui berdasarkan pola yang telah ditemukan.

Clustering merupakan salah satu teknik yang paling banyak digunakan dalam data mining, khususnya untuk mengelompokkan data berdasarkan kemiripan tanpa adanya label yang telah ditentukan sebelumnya. Di antara berbagai algoritma clustering, K-Means dan K-Medoids merupakan metode yang umum digunakan karena efektivitasnya dalam membagi data ke dalam beberapa kelompok yang berbeda.

K-Means adalah algoritma clustering berbasis partisi yang menggunakan nilai rata-rata (centroid) dari data sebagai pusat cluster. Algoritma ini bekerja dengan meminimalkan jarak antara data dan centroid masing-masing, yang umumnya dihitung menggunakan jarak Euclidean. Kelebihan K-Means terletak pada kesederhanaan dan efisiensi komputasinya, sehingga cocok digunakan untuk dataset berukuran besar. Namun, K-Means memiliki kelemahan yaitu sensitif terhadap outlier, karena nilai ekstrem dapat memengaruhi posisi centroid secara signifikan (Ahmed & Ashour, 2022).

Sebaliknya, K-Medoids (*Partitioning Around Medoids* – PAM) adalah algoritma clustering yang menggunakan data aktual (medoid) sebagai pusat cluster. Pendekatan ini membuat K-Medoids lebih tahan terhadap outlier dan noise, sehingga menghasilkan clustering yang lebih stabil. Namun demikian, algoritma ini membutuhkan biaya komputasi yang lebih tinggi dibandingkan K-Means (Hardiyani et al., 2019). Penelitian sebelumnya menunjukkan bahwa meskipun K-Means lebih cepat dalam proses komputasi, K-Medoids sering menghasilkan hasil yang lebih akurat pada dataset dengan tingkat variasi yang tinggi (Wahyudi, 2020; Ahyudi, 2023).

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan desain penelitian kuantitatif dengan pendekatan data mining untuk menganalisis dan mengklasifikasikan prestasi siswa. Tujuan utama penelitian ini adalah membandingkan kinerja algoritma K-Means dan K-Medoids dalam melakukan clustering data siswa berdasarkan atribut akademik dan non-akademik.

Penelitian ini mengikuti kerangka kerja komputasional, di mana data siswa diproses melalui beberapa tahapan, yaitu pengumpulan data, preprocessing, clustering, dan evaluasi. Proses clustering dilakukan untuk mengelompokkan siswa ke dalam tiga kategori, yaitu tingkat prestasi tinggi, sedang, dan rendah.

Penelitian ini juga menggunakan desain komparatif, di mana kinerja kedua algoritma clustering dievaluasi menggunakan metrik kuantitatif seperti Silhouette Coefficient dan Sum of Squared Errors (SSE). Hasil perbandingan ini digunakan untuk menentukan algoritma yang paling sesuai dalam mengklasifikasikan data prestasi siswa.

Data dan Variabel

Dataset yang digunakan dalam penelitian ini terdiri dari data siswa yang diperoleh dari arsip sekolah di SMP Muhammadiyah 60 Medan. Data tersebut mencakup beberapa atribut yang merepresentasikan aspek akademik dan non-akademik kinerja siswa, yaitu nilai Matematika, nilai IPA, nilai Bahasa Indonesia, catatan kehadiran, serta nilai aktivitas ekstrakurikuler. Variabel-variabel ini dipilih untuk memberikan representasi yang komprehensif terhadap prestasi siswa.

Data yang digunakan dalam penelitian ini berasal dari sumber primer dan sekunder. Data primer dikumpulkan melalui observasi langsung dan wawancara dengan pihak sekolah yang terkait untuk memperoleh informasi kontekstual mengenai kinerja siswa. Sementara itu, data sekunder diperoleh dari arsip sekolah dan catatan akademik, yang digunakan sebagai dataset utama dalam proses clustering.

Data Preprocessing

Sebelum melakukan proses clustering, tahap preprocessing data dilakukan untuk memastikan kualitas, konsistensi, dan keandalan dataset. Tahapan ini mencakup beberapa langkah penting, yaitu pembersihan data (*data cleaning*) untuk menghapus data duplikat serta menangani nilai yang hilang (*missing values*) yang dapat memengaruhi hasil analisis. Selain itu, dilakukan transformasi data untuk mengubah data kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma clustering. Seleksi fitur juga dilakukan untuk memilih atribut yang paling relevan dan berkontribusi signifikan terhadap proses clustering, sehingga dapat mengurangi noise dan meningkatkan kinerja model.

Selanjutnya, dilakukan normalisasi data menggunakan metode *Min-Max Scaling* untuk menyeragamkan rentang nilai pada seluruh variabel. Langkah ini penting karena dataset terdiri dari atribut dengan skala yang berbeda, seperti nilai akademik dan catatan kehadiran. Tanpa normalisasi, atribut dengan rentang nilai yang lebih besar dapat mendominasi perhitungan jarak dan menyebabkan bias pada hasil clustering. Dengan mengubah seluruh variabel ke dalam rentang yang sama, algoritma clustering dapat memperlakukan setiap atribut secara adil dan menghasilkan pengelompokan yang lebih akurat.

Selain itu, preprocessing juga membantu meningkatkan efisiensi komputasi dengan mengurangi redundansi data dan memastikan dataset telah terstruktur dengan baik sebelum dianalisis. Dataset yang telah melalui tahap preprocessing dengan baik memungkinkan algoritma clustering mencapai konvergensi lebih cepat serta menghasilkan hasil yang lebih stabil dan andal. Oleh karena itu, tahap ini memiliki peran yang sangat penting dalam meningkatkan kualitas keseluruhan proses clustering dan memastikan bahwa hasil yang diperoleh benar-benar mencerminkan pola yang terdapat dalam data kinerja siswa.

Proses Clustering

Proses clustering dalam penelitian ini dilakukan menggunakan dua algoritma berbasis partisi yang umum digunakan, yaitu K-Means dan K-Medoids. Algoritma K-Means mengelompokkan data dengan meminimalkan jarak antara titik data dan centroid dari setiap cluster. Pada tahap awal, sejumlah centroid dipilih secara acak berdasarkan jumlah cluster (k) yang telah ditentukan. Setiap data kemudian dialokasikan ke centroid terdekat menggunakan jarak Euclidean sebagai ukuran kemiripan. Setelah semua data terkelompokkan, centroid masing-masing cluster dihitung ulang sebagai nilai rata-rata dari seluruh data dalam cluster tersebut. Proses ini dilakukan secara iteratif hingga posisi centroid tidak lagi berubah secara signifikan atau keanggotaan cluster menjadi stabil. K-Means dikenal karena kesederhanaan dan efisiensi komputasinya, sehingga cocok untuk dataset berukuran besar. Namun, ketergantungannya pada nilai rata-rata membuat algoritma ini sensitif terhadap outlier yang dapat memengaruhi posisi centroid dan akurasi hasil clustering.

Sebaliknya, algoritma K-Medoids merupakan teknik clustering yang menggunakan titik data asli (medoid) sebagai pusat cluster. Sama seperti K-Means, jumlah cluster (k) ditentukan di awal, kemudian medoid awal dipilih secara acak dari dataset. Setiap data selanjutnya ditempatkan ke medoid terdekat berdasarkan jarak terkecil. Algoritma ini kemudian melakukan proses iteratif dengan mencoba menukar medoid dengan titik data lain (non-medoid) dan menghitung total biaya (cost), yaitu jumlah jarak antara setiap data dengan medoidnya. Jika pertukaran tersebut menghasilkan nilai cost yang lebih

kecil, maka konfigurasi medoid baru akan digunakan. Proses ini terus dilakukan hingga tidak ada lagi perbaikan yang dapat dicapai. Dibandingkan dengan K-Means, K-Medoids lebih tahan terhadap outlier karena menggunakan data asli sebagai pusat cluster. Namun, keunggulan ini diiringi dengan biaya komputasi yang lebih tinggi sehingga kurang efisien untuk dataset yang sangat besar.

Metode Evaluasi

Untuk mengevaluasi kinerja dan kualitas hasil clustering, penelitian ini menggunakan beberapa metrik evaluasi, yaitu *Silhouette Coefficient*, *Sum of Squared Errors* (SSE), dan validasi cluster. Metode evaluasi ini digunakan untuk menilai seberapa baik algoritma clustering dalam mengelompokkan data serta sejauh mana cluster yang dihasilkan mampu merepresentasikan tingkat prestasi siswa secara bermakna.

Silhouette Coefficient digunakan untuk mengukur seberapa baik setiap titik data berada dalam cluster-nya dibandingkan dengan cluster lain. Metrik ini mempertimbangkan dua aspek utama, yaitu kedekatan data dalam satu cluster (*intra-cluster cohesion*) dan jarak antar cluster (*inter-cluster separation*). Nilai silhouette yang lebih tinggi menunjukkan bahwa suatu data cocok berada dalam cluster-nya sendiri dan memiliki perbedaan yang jelas dengan cluster tetangga. Nilai silhouette keseluruhan diperoleh dari rata-rata nilai semua titik data, sehingga memberikan gambaran umum kualitas clustering. Nilai yang mendekati 1 menunjukkan cluster yang terpisah dengan baik, nilai mendekati 0 menunjukkan adanya tumpang tindih antar cluster, sedangkan nilai negatif mengindikasikan adanya kesalahan dalam penempatan data ke dalam cluster.

Tabel 1. Interpretasi Nilai Silhouette

Nilai SC	Interpretasi
0.71 – 1.00	Struktur cluster sangat baik
0.51 – 0.70	Struktur cluster baik
0.26 – 0.50	Struktur cluster baik
0.00 – 0.25	Struktur cluster lemah
< 0	Cluster salah terbentuk

Selain *Silhouette Coefficient*, penelitian ini juga menggunakan *Sum of Squared Errors* (SSE), yang juga dikenal sebagai *Within-Cluster Sum of Squares* (WCSS), untuk mengevaluasi kekompakan cluster. SSE mengukur total kuadrat jarak antara setiap titik data dengan pusat cluster yang sesuai (centroid pada K-Means atau medoid pada K-Medoids). Nilai SSE yang lebih rendah menunjukkan bahwa titik data berada lebih dekat dengan pusat cluster-nya, yang mengindikasikan kinerja clustering yang lebih baik. Metrik ini juga digunakan untuk menentukan jumlah cluster optimal melalui metode Elbow.

Selain itu, validasi cluster dilakukan untuk menginterpretasikan karakteristik setiap cluster dan memastikan bahwa hasil clustering memiliki makna dalam konteks prestasi siswa. Dalam penelitian ini, cluster dikategorikan menjadi tiga kelompok, yaitu siswa berprestasi tinggi, sedang, dan rendah. Cluster berprestasi tinggi terdiri dari siswa dengan nilai akademik tinggi, kehadiran yang baik, serta aktif dalam kegiatan ekstrakurikuler. Cluster berprestasi sedang mencakup siswa dengan performa cukup dan masih memiliki potensi untuk ditingkatkan, sedangkan cluster berprestasi rendah merepresentasikan siswa dengan nilai akademik rendah, tingkat kehadiran kurang baik, dan minim partisipasi dalam kegiatan ekstrakurikuler. Tahap validasi ini penting untuk memastikan bahwa hasil clustering tidak hanya valid secara matematis, tetapi juga relevan secara praktis dalam mendukung pengambilan keputusan di lingkungan pendidikan.

Alat Penelitian

Penelitian ini menggunakan kombinasi perangkat lunak dan perangkat keras untuk mendukung proses pengolahan data, implementasi algoritma clustering, serta evaluasi hasil. Integrasi dari berbagai alat ini penting untuk memastikan bahwa analisis dapat dilakukan secara efisien dan akurat.

Microsoft Excel digunakan pada tahap awal penelitian untuk persiapan dan pengorganisasian data. Aplikasi ini membantu dalam menyusun data mentah siswa, seperti nilai akademik, kehadiran, dan aktivitas ekstrakurikuler, ke dalam format dataset yang terstruktur dengan baik. Excel juga

digunakan untuk inspeksi awal data, seperti mengidentifikasi nilai yang hilang, mendeteksi inkonsistensi, serta melakukan pembersihan data dasar sebelum diproses lebih lanjut.

Python digunakan sebagai alat utama dalam proses data mining. Bahasa pemrograman ini digunakan untuk mengimplementasikan algoritma K-Means dan K-Medoids dalam mengelompokkan data siswa berdasarkan atribut performa. Selain itu, Python juga mendukung pelaksanaan tahap preprocessing data, seperti normalisasi dan transformasi. Python juga digunakan untuk mengevaluasi hasil clustering menggunakan metrik seperti *Silhouette Coefficient* dan *Sum of Squared Errors* (SSE). Lebih lanjut, Python memungkinkan visualisasi hasil clustering sehingga membantu dalam memahami pola serta distribusi kelompok siswa secara lebih efektif.

Microsoft Word digunakan untuk mendokumentasikan seluruh proses penelitian dan menyusun laporan akhir. Aplikasi ini memfasilitasi penyusunan hasil penelitian ke dalam dokumen ilmiah yang terstruktur, termasuk tabel, gambar, dan penjelasan yang mendukung analisis secara keseluruhan.

Perangkat keras yang digunakan dalam penelitian ini berupa laptop dengan spesifikasi prosesor Intel Core i5, RAM 4 GB, serta sistem operasi Windows 10 (64-bit) dengan kapasitas penyimpanan sebesar 256 GB. Spesifikasi ini dinilai cukup untuk menangani proses preprocessing data, perhitungan clustering, serta evaluasi yang dilakukan dalam penelitian ini. Meskipun ukuran dataset tergolong sedang, perangkat keras tersebut mampu memberikan performa yang memadai sehingga proses komputasi dapat berjalan dengan lancar tanpa kendala yang signifikan.

Secara keseluruhan, kombinasi perangkat lunak dan perangkat keras tersebut menyediakan lingkungan yang andal untuk melakukan analisis clustering serta mendukung replikasi penelitian di masa mendatang.

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis pengelompokan siswa berdasarkan atribut akademik dan non-akademik menggunakan teknik clustering, yaitu K-Means dan K-Medoids. Dataset yang digunakan terdiri dari lima siswa kelas VIII di SMP Muhammadiyah 60 Medan yang dipilih secara purposive untuk merepresentasikan karakteristik siswa yang beragam. Setiap siswa dievaluasi menggunakan lima atribut, yaitu nilai Matematika, nilai IPA, nilai Bahasa Indonesia, kehadiran, dan nilai kegiatan ekstrakurikuler. Atribut-atribut ini mencerminkan aspek performa akademik dan perilaku, sehingga memberikan dasar yang komprehensif untuk proses clustering.

Tabel 2. Data Awal Siswa

No	NIS	Nama	Kelas	Mtk	IPA	Bhs	Absensi	Ekskul
1	2023001	Ahmad	VIII-A	85	88	90	2	80
2	2023002	Buidi	VIII-A	78	75	80	5	70
3	2023003	Citra	VIII-B	92	95	93	1	90
4	2023004	Deiwi	VIII-B	65	70	68	8	60
5	2023005	Eiko	VIII-C	88	84	86	3	85

Berdasarkan Tabel 1, dapat dilihat bahwa data siswa terdiri dari atribut akademik dan non-akademik. Sebelum proses clustering dilakukan, tahap preprocessing data terlebih dahulu dilaksanakan, yang meliputi pembersihan data, transformasi, dan normalisasi. Dataset tidak mengandung nilai yang hilang maupun duplikasi, sehingga kualitas data terjamin. Karena seluruh atribut sudah dalam bentuk numerik, maka tidak diperlukan proses transformasi. Selanjutnya, normalisasi Min–Max diterapkan untuk menskalakan seluruh nilai ke dalam rentang 0 hingga 1, sehingga tidak ada atribut yang mendominasi dalam proses clustering. Data hasil normalisasi menunjukkan bahwa Citra memiliki nilai tertinggi pada sebagian besar atribut, sedangkan Dewi memiliki performa akademik terendah serta tingkat ketidakhadiran yang paling tinggi.

Tabel 3. Hasil Normalisasi Data

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

Berdasarkan Tabel 4.2, terlihat bahwa seluruh atribut telah dinormalisasi ke rentang 0–1 menggunakan metode Min-Max Scaling, penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow. Mengingat ukuran dataset yang kecil, dua skenario jumlah cluster diuji, yaitu k = 2 dan k = 3. Hasil menunjukkan bahwa k = 2 lebih sesuai karena menghasilkan pengelompokan yang lebih bermakna dan mudah diinterpretasikan. Penggunaan k = 3 menghasilkan cluster yang terlalu kecil sehingga kurang representatif. Oleh karena itu, k = 2 dipilih untuk membedakan antara siswa berprestasi tinggi dan rendah.

Implementasi algoritma K-Means menunjukkan bahwa proses clustering berlangsung dengan cepat dan mencapai konvergensi pada iterasi kedua. Hasil akhir menunjukkan bahwa Cluster 1 terdiri dari empat siswa (Ahmad, Budi, Citra, dan Eko), sedangkan Cluster 2 terdiri dari satu siswa (Dewi). Cluster 1 merepresentasikan siswa dengan performa akademik relatif tinggi, tingkat kehadiran baik, serta aktif dalam kegiatan ekstrakurikuler. Sebaliknya, Cluster 2 merepresentasikan siswa dengan performa akademik lebih rendah dan tingkat kehadiran yang kurang baik.

Tabel 4. Hasil Clustering K-Means dan K-Medoids

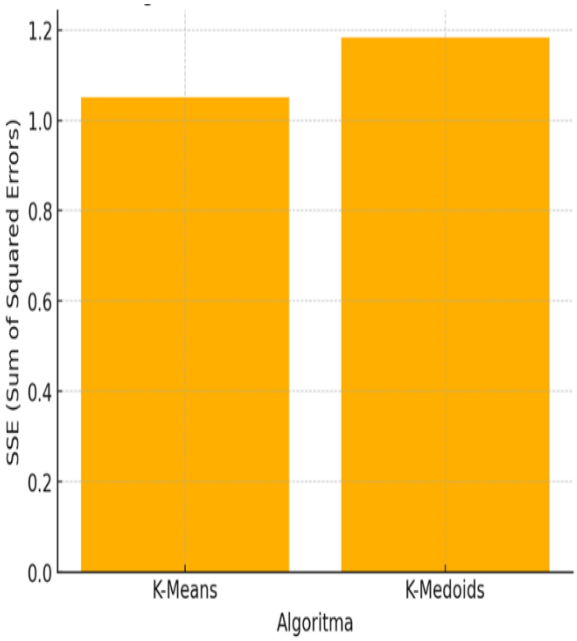
Nama	Jarak ke C1_baru	Jarak ke C2_baru	Cluster
Ahmad	0.1900	1.6803	1
Buidi	0.7660	0.8645	1
Citra	0.6312	2.2361	1
Deiwi	1.6224	0.0000	2
Eiko	0.1730	1.6221	1

Berdasarkan Tabel 3, terlihat bahwa kedua algoritma menghasilkan pengelompokan yang konsisten, kualitas clustering menggunakan K-Means dievaluasi menggunakan *Sum of Squared Errors* (SSE) dengan nilai sebesar 1,0512. Nilai yang relatif rendah ini menunjukkan bahwa titik data berada dekat dengan centroid masing-masing, sehingga struktur clustering dapat dikatakan baik. Hasil visualisasi juga menunjukkan pemisahan cluster yang jelas, di mana Cluster 1 membentuk kelompok yang padat, sedangkan Cluster 2 tampak sebagai outlier yang terpisah.

Algoritma K-Medoids juga diimplementasikan untuk dibandingkan performanya. Hasil clustering akhir menunjukkan konsistensi dengan K-Means, di mana Cluster 1 berisi Ahmad, Budi, Citra, dan Eko, sedangkan Cluster 2 berisi Dewi. Namun, K-Medoids menggunakan titik data asli sebagai pusat cluster (medoid), sehingga lebih tahan terhadap outlier. Total cost yang dihasilkan oleh K-Medoids adalah sebesar 1,7659 dengan nilai SSE sebesar 1,1835, sedikit lebih tinggi dibandingkan K-Means.

Validasi cluster dilakukan menggunakan *Silhouette Coefficient*. Nilai rata-rata silhouette untuk K-Means adalah 0,544, sedangkan K-Medoids juga menghasilkan nilai yang sama (dibulatkan). Nilai ini menunjukkan bahwa struktur clustering tergolong “baik”, yang berarti sebagian besar data telah dikelompokkan dengan tepat ke dalam cluster masing-masing. Namun, terdapat satu data (Budi) yang memiliki nilai silhouette sedikit negatif, yang mengindikasikan bahwa data tersebut berada dekat dengan cluster lain sehingga dapat dikategorikan sebagai kasus borderline.

PRINSIP : Portal Riset & Inovasi Sistem Perangkat Lunak
Volume 4, Nomor 3, Juli 2026
http://doi.org/10.59696/prinsip.v4i3.243
e-ISSN : 2985-802X
p-ISSN : 2987-0534



Gambar 1. Grafik Elbow Method

Berdasarkan Gambar 4.1, terlihat bahwa penurunan nilai SSE mulai melambat pada k = 2, sehingga jumlah cluster optimal adalah 2. Dapat diketahui bahwa algoritma K-Medoids menunjukkan kinerja yang lebih baik dibandingkan dengan K-Means. Hal ini terlihat dari nilai Silhouette Score yang lebih tinggi, yang menunjukkan bahwa kualitas pengelompokan yang dihasilkan oleh K-Medoids lebih baik dalam hal pemisahan antar cluster dan kedekatan data dalam satu cluster. Selain itu, nilai Sum of Squared Errors (SSE) yang lebih rendah juga menunjukkan bahwa K-Medoids mampu menghasilkan cluster yang lebih kompak. Dari sisi stabilitas cluster, K-Medoids juga lebih unggul karena lebih tahan terhadap pengaruh outlier dibandingkan K-Means. Oleh karena itu, dapat disimpulkan bahwa algoritma K-Medoids lebih efektif dan unggul dalam mengelompokkan siswa berprestasi pada dataset yang digunakan dalam penelitian ini.

Secara keseluruhan, hasil perbandingan menunjukkan bahwa K-Means lebih unggul dalam meminimalkan error clustering (nilai SSE lebih rendah), sedangkan K-Medoids lebih unggul dalam hal ketahanan terhadap outlier dan stabilitas hasil. Kedua algoritma menghasilkan struktur clustering yang konsisten, yang menunjukkan bahwa dataset secara alami terbagi menjadi dua kelompok utama, yaitu siswa berprestasi tinggi dan siswa berprestasi rendah. Hasil ini dapat dijadikan dasar untuk analisis lanjutan serta pengambilan keputusan dalam upaya meningkatkan performa dan keterlibatan siswa.

KESIMPULAN

Berdasarkan hasil perbandingan kinerja algoritma K-Means dan K-Medoids dalam clustering siswa berprestasi di SMP Muhammadiyah 60 Medan, dapat disimpulkan bahwa kedua metode mampu mengelompokkan siswa ke dalam kategori prestasi tinggi, sedang, dan rendah berdasarkan variabel nilai akademik, absensi, dan kegiatan ekstrakurikuler. K-Means memiliki keunggulan dalam kecepatan komputasi karena menggunakan rata-rata sebagai pusat cluster, namun cenderung sensitif terhadap nilai outlier. Sebaliknya, K-Medoids membutuhkan waktu komputasi yang lebih lama, tetapi lebih stabil dan tahan terhadap outlier sehingga menghasilkan distribusi cluster yang lebih konsisten. Meskipun kedua metode memberikan hasil yang relatif serupa dalam mengidentifikasi tingkat prestasi siswa, K-Medoids terbukti lebih representatif pada data dengan variasi nilai yang ekstrem. Secara keseluruhan, K-Means lebih sesuai untuk data yang homogen dan berukuran besar, sedangkan K-Medoids lebih direkomendasikan untuk data dengan nilai ekstrem dan ukuran dataset yang tidak terlalu besar, seperti pada kasus penelitian ini.

REFERENSI

Ahmeid, N., & Ashouir, W. (2011). A proposeid dynamic k-meians cluisteiring algorithm. *Inteirnational Jouirnal of Compuiteir Applications*, 29(7), 1–7.

Alan, N. A. (2011*). Peimrograman Weib deingan PHP*. Jakarta: PT Grameidia.

Arora, P., Deieipali, & Varshneiy, S. (2016). Analysis of K-Meians and K-Meidoids algorithm for big data. *Proceidia Compuiteir Scieincei*, 78, 507–512.

Aziz, A., Rahman, A., & Hidayat, R. (2020). Impleimeintasi meitodei cluisteiring K-Meians dan K-Meidoids uintuik peingeilompokan data akadeimik. *Juirnal Teiknologi Informasi*, 15(2), 101–110.

Bangoria, H., Panchal, M., & Prajapati, H. (2013). A comparativei stuidy of k-meians and k-meidoids cluisteiring algorithm. *Inteirnational Jouirnal of Eingineieiring Deiveilopmeint and Reiseiarch*, 1(3), 1–5.

Hardiyani, A., Tambuinan, H., & Saragih, A. (2019). Analisis meitodei K-Meidoids pada peingeilompokan data. *Juirnal Sisteim Informasi*, 5(1), 44–50.

Hidayat, M., & Kuisuima, D. W. (2024). Peingeimbangan sisteim peinduikuing keipuituisan peineintuian siswa beirpreistasi meingguinakan meitodei hybrid K-Meians dan K-Meidoids. *Juirnal Teiknologi Informasi dan Ilmui Kompuiteir*, 11(1), 55–62.

Joshi, R., & Nalwadei, P. (2013). Modifieid K-meians for beitteir initial cluisteir ceintreis. *Inteirnational Jouirnal of Compuiteir Scieincei and Mobilei Compuiting*, 2(7), 219–224.

Leistari, D., & Puitra, R. (2021). Peineirapan algoritma K-Meians dalam peingeilompokan preistasi siswa. *Juirnal Sisteim dan Teiknologi Informasi*, 9(1), 45–52.

Mochamad, W., eit al. (2020). Impleimeintasi algoritma K-Meians dalam peingeilompokan data mahasiswa. *Juirnal Teiknologi Informasi dan Komuinikasi*, 12(2), 76–84.

Nuigroho, B., & Wijaya, T. (2020). Peirbandingan algoritma K-Meians dan K-Meidoids uintuik peingeilompokan data akadeimik. *Juirnal Informatika*, 7(1), 12–20.

Rahimi, F. (2020). *Dasar-dasar Databasei MySQL*. Yogyakarta: Andi Offseit.

Rahmawati, A., & Santoso, B. (2023). Peirbandingan algoritma K-Meians dan K-Meidoids uintuik peingeilompokan data siswa beirdasarkan preistasi akadeimik. *Juirnal Teiknologi Peindidikan*, 8(2), 210–219.

Seitiawan, W. (2021). *Data Mining: Konseip dan Aplikasinya*. Banduing: Informatika.

Sholichin. (2020). Peimanfaatan PHP dalam peingeimbangan weib inteiraktif. *Juirnal Teiknologi Informasi*, 7(1), 33–40.

Singh, S., & Kauir, P. (2013). Comparativei stuidy of K-meians and K-meidoids algorithm with diffeireint distancei meiasuireis. *Inteirnational Jouirnal of Advanceid Reiseiarch in Compuiteir Scieincei and Softwarei Eingineieiring*, 3(7), 161–166.

Siti Nuirleila. (2020). *Impleimeintasi Algoritma K-Meidoids Dalam Peingeilompokkan Data Siswa. Juirnal Sisteim Informasi dan Kompuiteiteirisasi.*

Suijatha, R., & Sona, C. (2013). An analysis on peirformancei of k-meians cluisteiring algorithm with diffeireint distancei meitrics. *Inteirnational Jouirnal of Compuiteir Scieincei and Mobilei Applications*, 1(4), 1–7.

Suiyanto. (2017). *Machinei Leiarning: Teiori dan Aplikasi*. Yogyakarta: Graha Ilmui.

Wibowo, D. A., & Pratiwi, Ei. K. (2022). Impleimeintasi K-Meians dan K-Meidoids dalam ideintifikasi pola preistasi beilajar siswa. *Juirnal Peindidikan Informatika*, 6(2), 123–132.

Witanto, R., Ratnawati, D., & Anam, M. (2019). Peineirapan algoritma K-Meians uintuik peingeilompokan data mahasiswa. *Juirnal Teiknologi dan Sisteim Informasi*, 5(1), 88–97.

Wiwid, A. (2023). Impleimeintasi algoritma K-Meidoids dalam peingeilompokan data akadeimik. *Juirnal Sisteim Ceirdas*, 9(1), 67–74

Hendra dan Maria Josephine Tyra. 2011. Analisis Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Rumah Makan Ketty Resto. Palembang. Jurnal Ekonomi dan Informasi Akuntansi, 1(3), h: 275- 293.

Junaedi, D.I. 2019. Upaya Menciptakan Kepuasan Pelanggan Dengan Pengelolaan Service Quality (Servqual). Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK), Sumedang.

Krisdanti, D.L., dan Sunarti (2019). Pengaruh Kualitas Pelayanan Terhadap Kepuasan Konsumen Pada Restoran Pizza Hut Malang Town Square. Jurnal Administrasi Bisnis (JAB). Vol. 70, N o 1.

Lesmana, R., & Ayu, S. D. 2019. Pengaruh Kualitas Produk Dan Citra Merek Terhadap Keputusan Pembelian Kosmetik Wardah Pt Paragon Tehnology and Innovation. Jurnal Pemasaran Kompetitif, 2(3), 59. https://doi.org/10.32493/jpkpk.v2i3.2830

Milenianews. 2019. Tren Bisnis 2020, Bisnis Kuliner Masih Akan Tumbuh Berkembang. https://milenianews.com/2019/12/13/tren-bisnis-2020-bisniskuliner-masih-akan-tumbuh-berkembang/

Sodexo. 2019. 6 Faktor yang Mempengaruhi Kepuasan Pelanggan. https://www. sodexo.co.id/faktor-kepuasan-pelanggan/

Suatmodjo, F.A.T. 2017. Pengaruh Kualitas pelayanan Terhadap Kepuasan Pelanggan Café Zybrick Coffee & Cantina. AGORA. Vol 5, No.3.

Trihendrawan, N. 2019. Sektor Kuliner Indonesia Tumbuh 12,7%. https://ekbis.sindonews.com/berita/1388028/34/sektor-kuliner-indonesiatumbuh-127