

Initial Centroid Pada Algoritma K-Means Dan K-Medoids

Devita Anggraini¹, Muhammad Siddik Hasibuan²

Jurusan Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Addresses Jl. Lap. Golf, Medan , Indonesia

Article Info

Keywords:

Initial Centroid,
K-Means,
K-Medoids,
Clustering,
Data Mining.

ABSTRACT

Pertumbuhan data yang semakin besar, kompleks, dan beragam menuntut adanya metode analisis yang mampu mengubah data mentah menjadi informasi yang bermanfaat. Salah satu teknik yang banyak digunakan adalah clustering, yaitu proses pengelompokan data berdasarkan kemiripan karakteristik. Algoritma K-Means dan K-Medoids merupakan dua metode clustering yang populer. K-Means bekerja dengan menentukan pusat cluster berdasarkan nilai rata-rata (means) dari data, sedangkan K-Medoids menggunakan salah satu objek data sebagai pusat cluster (medoid). Permasalahan utama pada kedua algoritma ini terletak pada pemilihan titik pusat awal cluster (initial centroid) yang sangat mempengaruhi hasil akhir pengelompokan. K-Means sering kali menghasilkan klaster yang berbeda karena penentuan centroid dilakukan secara acak, sedangkan K-Medoids lebih stabil terhadap pencilan, namun memerlukan waktu komputasi yang lebih lama. Penelitian ini membahas pentingnya strategi pemilihan initial centroid pada algoritma K-Means dan K-Medoids serta perbandingan kinerja keduanya. Berdasarkan hasil analisis, K-Means lebih sesuai digunakan untuk dataset berukuran besar dengan distribusi data seragam karena lebih cepat dalam proses komputasi. Sementara itu, K-Medoids lebih akurat dalam mengelompokkan dataset kecil yang memiliki variasi tinggi serta mengandung pencilan karena tidak terdistorsi oleh nilai rata-rata. Dengan demikian, pemilihan metode yang tepat harus mempertimbangkan ukuran, distribusi, dan karakteristik dataset yang dianalisis. Penelitian ini diharapkan dapat memberikan kontribusi dalam memahami peran initial centroid dalam meningkatkan stabilitas dan akurasi hasil clustering.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



Corresponding Author:

Devita Anggraini
Universitas Islam Negeri Sumatera Utara
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten
Deli Serdang, Sumatera Utara 20353
devitaa099@gmail.com

PENDAHULUAN

Semakin berkembangnya volume, kompleksitas, dan keragaman data akibat kemajuan perangkat keras dan perangkat lunak membuat proses ekstraksi data mentah menjadi informasi yang bermanfaat semakin menantang (Safar Dwi Kurniawan & Rosalina Yani Widiastuti, S.Kom., 2024). Ketika data dalam jumlah besar melampaui kemampuan manusia untuk mengenali pola dan karakteristiknya, dibutuhkan teknologi komputasi seperti data mining. Data mining merupakan teknik untuk menemukan pola dari data mentah agar menjadi informasi berguna, salah satunya melalui proses clustering, yaitu pengelompokan data berdasarkan kemiripan karakteristik (Hendrastuty, 2024).

Algoritma K-Means diperkenalkan oleh MacQueen pada tahun 1967 dengan prinsip menghitung jarak terdekat setiap data terhadap pusat cluster, lalu memperbarui pusat cluster berdasarkan rata-rata data dalam kelompok tersebut (Fusta, 2024). Proses ini berlangsung berulang hingga data tidak lagi berpindah cluster atau batas iterasi tercapai, dengan tujuan meminimalkan jarak antar data dan pusat cluster. Meskipun sederhana dan cepat, kelemahannya terletak pada pemilihan titik pusat awal yang dilakukan secara acak, sehingga hasil pengelompokan sering kali berbeda pada setiap percobaan, tidak stabil, dan belum tentu optimal terutama pada data yang kompleks atau mengandung pencilan (Hendriansyah et al., 2025). Kondisi tersebut menyebabkan data pada algoritma K-Means konvensional sering berpindah-pindah cluster hingga mencapai batas iterasi maksimum. Hal ini terjadi karena penentuan pusat awal cluster masih dilakukan secara acak. Perbedaan konfigurasi titik awal tersebut dapat menghasilkan centroid akhir yang berbeda, sehingga hasil clustering menjadi kurang stabil dan kurang akurat. Oleh karena itu, pemilihan titik pusat awal cluster atau initial centroid memiliki peran yang sangat penting dalam menentukan kualitas pengelompokan (Rahman et al., 2025).

Algoritma K-Medoids, yang juga dikenal sebagai Partitioning Around Medoid (PAM), diperkenalkan oleh Leonard Kaufman dan Peter J. Rousseeuw sebagai metode clustering yang mirip dengan K-Means karena sama-sama bersifat partitional, yaitu membagi dataset ke dalam beberapa kelompok. Perbedaan utama terletak pada penentuan pusat cluster, di mana K-Means menggunakan nilai rata-rata data, sedangkan K-Medoids memilih salah satu objek data sebagai pusat (medoid). Keunggulan K-Medoids adalah kemampuannya mengurangi kelemahan K-Means yang rentan terhadap pencilan, karena data yang jauh berbeda dari mayoritas tidak akan mendistorsi hasil pengelompokan (Supriyadi et al., 2021). Penelitian yang dilakukan oleh (Arifandi et al., 2021) dengan judul "Implementasi Algoritma K-Medoids untuk Clustering Wilayah Terinfeksi Kasus Covid-19 di DKI Jakarta" menggunakan teknik data mining dalam pengelompokan kasus terinfeksi Covid-19 berdasarkan kelurahan di Provinsi DKI Jakarta. Hasil penelitian menunjukkan bahwa penerapan algoritma K-Medoids, baik melalui perhitungan manual maupun menggunakan perangkat lunak Rapid Miner, menghasilkan keluaran yang valid dan dapat digunakan sebagai dasar analisis penyebaran kasus.

Penelitian lain yang dilakukan oleh (Meiriza et al., 2023) dengan judul "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan" membahas penerapan kedua algoritma dalam mengelompokkan program bukan penerima upah berdasarkan fitur-fitur tertentu seperti umur, pekerjaan, upah, dan jenis program. Hasil penelitian menunjukkan bahwa K-Means lebih unggul dalam hal kecepatan komputasi, sedangkan K-Medoids memberikan hasil yang lebih akurat karena menggunakan medoid sebagai pusat kelompok. Oleh karena itu, pemilihan algoritma disarankan disesuaikan dengan tujuan analisis dan ketersediaan sumber daya. Penelitian ini juga merekomendasikan penambahan jumlah dataset serta variasi parameter pada penelitian selanjutnya untuk meningkatkan kinerja algoritma clustering.

Penelitian sebelumnya mengenai algoritma *clustering* seperti K-Means dan K-Medoids umumnya menitikberatkan pada penerapan keduanya dalam berbagai kasus maupun perbandingan performa berdasarkan aspek kecepatan komputasi dan akurasi, dengan temuan bahwa K-Means lebih unggul dari sisi efisiensi sedangkan K-Medoids lebih stabil terhadap keberadaan pencilan. Meskipun demikian, sebagian besar kajian tersebut belum

memberikan perhatian mendalam terhadap pemilihan titik pusat awal (*initial centroid*), padahal tahap inisialisasi ini memiliki peran krusial dalam menentukan kestabilan, kualitas, serta konsistensi hasil pengelompokan. Oleh karena itu, penelitian ini difokuskan pada analisis strategi penentuan *initial centroid* pada algoritma K-Means dan K-Medoids dengan tujuan memberikan kontribusi terhadap peningkatan akurasi dan reliabilitas hasil *clustering* yang masih menjadi celah dalam penelitian terdahulu.

Tujuan dari penelitian ini adalah untuk menganalisis metode pemilihan *initial centroid* pada algoritma K-Means dan K-Medoids guna memperoleh strategi inisialisasi yang optimal sehingga dapat meningkatkan akurasi, konsistensi, dan stabilitas hasil *clustering*. Selain itu, penelitian ini bertujuan memberikan pemahaman yang lebih mendalam mengenai pengaruh pemilihan *initial centroid* terhadap kinerja algoritma, serta menyajikan rekomendasi metode yang dapat memudahkan pengguna dalam menerapkan algoritma *clustering* di bidang ilmu komputer.

METHODS

Data Mining

Data mining merupakan suatu pendekatan yang digunakan untuk menemukan pola, tren, serta karakteristik tertentu yang dapat memberikan gambaran mengenai kondisi di masa depan. Proses ini tidak hanya berfungsi untuk melakukan analisis prediktif, tetapi juga mampu mengungkap informasi yang bersifat tersembunyi dan sebelumnya belum pernah teridentifikasi dalam suatu basis data yang besar. Dengan demikian, data mining memiliki peran penting dalam membantu pengambilan keputusan yang lebih tepat, efektif, dan berbasis data pada berbagai bidang, mulai dari bisnis, kesehatan, pendidikan, hingga penelitian ilmiah (Sekar Setyaningtyas et al., 2022).

Data mining merupakan proses ekstraksi informasi berharga dari basis data besar melalui pembangunan model yang mampu mengenali pola tersembunyi. Fungsi utamanya terbagi dua, yaitu deskriptif untuk memahami karakteristik data dengan menemukan pola berulang, dan prediktif untuk memanfaatkan pola tersebut dalam memperkirakan variabel yang belum diketahui. Proses ini dilakukan dengan metode tertentu yang membantu dalam pengumpulan informasi, mulai dari perencanaan hingga implementasi akhir (Arifandi et al., 2021).

Centroid

Centroid merupakan titik pusat massa atau pusat geometris dari suatu objek maupun sekumpulan titik dalam suatu ruang. Konsep ini banyak dimanfaatkan dalam bidang matematika, statistik, serta disiplin ilmu lainnya. Secara umum, centroid dipahami sebagai titik pusat yang mewakili posisi keseimbangan atau representasi dari sebuah objek maupun himpunan titik tertentu (Raja et al., 2024).

Centroid adalah salah satu metode yang digunakan dalam proses pembobotan kriteria dengan memanfaatkan konsep titik pusat (centroid) sebagai acuan. Metode ini berfungsi untuk memberikan penilaian yang lebih objektif melalui perhitungan pusat data, sehingga kriteria dapat diperingkat secara sistematis. Dengan pendekatan ini, setiap kriteria tidak hanya dinilai berdasarkan bobot individualnya, tetapi juga dipertimbangkan dalam hubungan keseluruhan data, sehingga menghasilkan perankingan yang lebih akurat dan seimbang (Widodo, 2024).

K-Means

K-Means clustering adalah salah satu metode dalam analisis data dan machine learning yang digunakan untuk membagi data ke dalam beberapa kelompok (cluster) berdasarkan tingkat kesamaan fitur atau atribut yang dimilikinya. Teknik ini bekerja dengan mencari titik pusat (centroid) pada setiap cluster, lalu menempatkan data ke dalam kelompok dengan jarak terdekat ke centroid tersebut. Dengan cara ini, K-Means mampu membantu dalam menemukan pola tersembunyi, menyederhanakan kompleksitas data, serta mendukung pengambilan keputusan dalam berbagai bidang seperti bisnis, kesehatan, hingga penelitian ilmiah (Maori & Evanita, 2023). K-Means memiliki keunggulan karena algoritmanya sederhana dan mampu bekerja dengan efisien. Namun, kelemahannya terletak pada sensitivitas terhadap pemilihan centroid awal, sehingga hasil pengelompokan dapat bervariasi dan menghasilkan solusi yang berbeda pada setiap proses percobaan (Khalish et al., 2025).

K-Medoid

K-Medoid merupakan metode clustering yang berfokus pada pencarian titik paling representatif (medoid) dalam sebuah dataset dengan mempertimbangkan jarak antar data. Prosesnya dilakukan dengan mengevaluasi berbagai kemungkinan kombinasi medoid agar diperoleh konfigurasi di mana jarak antar titik dalam satu cluster relatif kecil, sedangkan jarak antar cluster menjadi besar. Berbeda dengan K-Means yang menggunakan rata-rata sebagai pusat cluster, K-Medoid lebih tahan terhadap keberadaan outlier karena pusat cluster dipilih dari data aktual, sehingga metode ini sering dianggap lebih robust dalam menangani data yang mengandung nilai ekstrem (Riskha & Farokhah, 2023). K-Medoids memanfaatkan objek representatif (medoid) sebagai pusat cluster, sehingga tidak mudah dipengaruhi oleh outlier. Hal ini membuat metode ini lebih stabil dalam melakukan pengelompokan data yang mengandung nilai ekstrem. Selain itu, karena menggunakan data asli sebagai pusat, K-Medoids mampu memberikan hasil clustering yang lebih akurat dan realistis, terutama pada dataset dengan distribusi yang tidak merata atau memiliki banyak variasi jarak antar titik (Salman et al., 2025).

Method

Kerangka penelitian merupakan serangkaian tahapan yang disusun secara sistematis untuk mencapai tujuan penelitian secara optimal dan menghasilkan data yang akurat. Tahapan ini berfungsi sebagai pedoman dalam proses penelitian agar lebih terarah, efisien, serta meminimalisir kesalahan. Dengan adanya kerangka penelitian, setiap langkah yang dilakukan dapat dipertanggungjawabkan secara ilmiah dan mendukung tercapainya hasil penelitian yang valid. Langkah-langkah dalam kerangka penelitian adalah sebagai berikut:



Gambar 1 Kerangka Berpikir

Analisa Masalah

Tahap ini bertujuan untuk mengidentifikasi permasalahan utama dalam penelitian, menentukan data yang diperlukan untuk membangun sistem, serta menetapkan kriteria masalah yang akan dijadikan acuan.

Rumusan Masalah

Pada tahap ini penulis menganalisa masalah yang terjadi. Pada Penelitian ini yang mengacu pada permasalahan yang timbul adalah bagaimana pengaruh penggunaan initial centroid terpilih pada performa algoritma K-Means dalam mengelompokkan data masyarakat kurang mampu pada Kabupaten Deli Serdang.

Pengumpulan Data

Tahap ini dilakukan peneliti untuk mengumpulkan informasi yang mendukung tujuan penelitian. Data diperoleh melalui dua cara, yaitu studi literatur dengan mengumpulkan referensi dari buku, jurnal, dan sumber internet yang relevan dengan topik penelitian, serta pengambilan langsung berupa data akademik 250 siswa yang mencakup atribut seperti identitas unik siswa dan jumlah jam belajar.

RESULTS AND DISCUSSION

Analisa Data

Data yang digunakan dalam penelitian ini berasal dari analisis akademik siswa yang diambil dari situs www.kaggle.com, dengan jumlah sekitar 250 siswa. Atribut yang tercakup meliputi identitas, ID siswa, jumlah jam belajar, partisipasi dalam kegiatan ekstrakurikuler, nilai akademik, serta tingkat pendidikan orang tua. Contoh sampel data dapat dilihat pada tabel berikut:

Tabel 1 Data Siswa

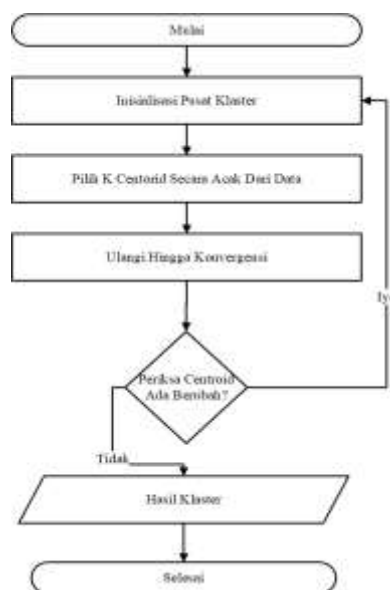
ID	Study Hours/ Week	Attendance Rate	Previous Grades	Participation in Extracurricular Activities	Parent Education Level	Passed
S00001	12.5	95.8	75.0	Yes	Master	Yes

S00002	9.3	95.3	60.6	No	High School	No
S00003	13.2	95.8	64.0	No	Associate	No
S00004	17.6	76.8	62.4	Yes	Bachelor	No
S00005	8.8	89.3	72.7	No	Master	No
S00006	8.8	73.8	69.3	Yes	High School	Yes
...
S00246	8.3	42.3	65.5	No	Associate	Yes
S00247	7.6	78.8	86.2	nan	Doctorate	No
S00248	6.7	85.3	67.3	No	Doctorate	Yes
S00249	18.8	93.8	43.0	No	Bachelor	Yes
S00250	12.0	45.6	34.7	No	Master	No

Penerapan Metode K-Means dan K-Medoids

1. K-Means

Gambar di bawah ini menunjukkan diagram alir (*flowchart*) dari metode K-Means, yang menggambarkan tahapan-tahapan proses *clustering* secara sistematis. Flowchart ini memvisualisasikan langkah-langkah mulai dari inisialisasi centroid awal, perhitungan jarak setiap data terhadap centroid, penentuan keanggotaan data dalam cluster, pembaruan posisi centroid, hingga proses iterasi yang berlanjut sampai mencapai konvergensi. Dengan menggunakan flowchart, alur kerja metode K-Means menjadi lebih jelas dan mudah dipahami.



Gambar 2 Flowchart K-Means

a. Sampel Data

Tabel 2 Sampel Data K-Means

ID	Study Hours/Week	Attendance Rate	Previous Grades	Participation in Extracurricular Activities	Parent Education Level	Passed
S00001	12.5	95.8	75.0	Yes	Master	Yes
S00002	9.3	95.3	60.6	No	High School	No
S00003	13.2	95.8	64.0	No	Associate	No
S00004	17.6	76.8	62.4	Yes	Bachelor	No
S00005	8.8	89.3	72.7	No	Master	No

b. Menentukan Klaster

Tabel 3 Klaster

Kode Klaster	Jenis Klaster
K1	Passed
K2	Not Passed

c. Centroid Awal

dua data pertama (S00001 dan S00002) sebagai centroid awal.

Tabel 4 Centroid Awal

Feature	Centroid 1 (S00001)	Centroid 2 (S00002)
Study Hours/Week	12.5	9.3
Attendance Rate	95.8	95.3
Previous Grades	75.0	60.6
Participation in Extracurricular Activities	1	0
Parent Education Level	4	1

d. Hitung Jarak Setiap Data ke Centroid

Tabel 5 Jarak Data Ke Centroid

ID	Study Hours/Week	Attendance Rate	Previous Grades	Participation	Parent Education Level
S00001	12.5	95.8	75.0	1	4
S00002	9.3	95.3	60.6	0	1
S00003	13.2	95.8	64.0	0	2
S00004	17.6	76.8	62.4	1	3
S00005	8.8	89.3	72.7	0	4

e. Perhitungan Jarak untuk Setiap Data

Berikut adalah contoh perhitungan jarak menggunakan rumus:

Untuk S00001 ke C1:

$$d = \sqrt{(12.5 - 12.5)^2 + (98.5 - 98.5)^2 + (75.0 - 75.0)^2 + (1 - 1)^2 + (4 - 4)^2} = 0$$

Untuk S00002 ke C2:

$$d = \sqrt{(12.5 - 9.3)^2 + (95.8 - 95.3)^2 + (75.0 - 60.6)^2 + (1 - 0)^2 + (4 - 1)^2} = 0$$

$$d = \sqrt{(3.2)^2 + (0.5)^2 + (14.4)^2 + (1)^2 + (3)^2} = \sqrt{10.24 + 0.25 + 207.36 + 1 + 9} = \sqrt{227.85} \approx 15.10$$

Lakukan perhitungan yang sama kepada data seluruhnya sampai ke data 250.

f. Hasil Jarak dan Klaster

Tabel 6 Klaster iterasi 1

ID	CLUSTER 1	CLUSTER 2
S00001	0	23,383114
S00002	15,094701	20,478769
S00003	11,246777	19,619378
S00004	23,383114	0
S00005	7,8885994	18,487293

Berikut adalah centroid baru untuk masing-masing klaster setelah iterasi pertama:

Centroid C1:

Study Hours/Week: 10.65

Attendance Rate: 92.55

Previous Grades: 73.85

Participation: 0.50

Parent Education Level: 4.00

Centroid C2:
 Study Hours/Week: 13.37
 Attendance Rate: 89.30
 Previous Grades: 62.33
 Participation: 0.33
 Parent Education Level: 2.00

Tabel 7 Hasil Klaster iterasi 2

ID	Jarak ke C1	Jarak ke C2	Klaster
S00001	3.94	14.42	C1
S00002	13.94	7.53	C2
S00003	10.88	6.72	C2
S00004	20.71	13.25	C2
S00005	3.94	11.51	C1

Pembagian Klaster:

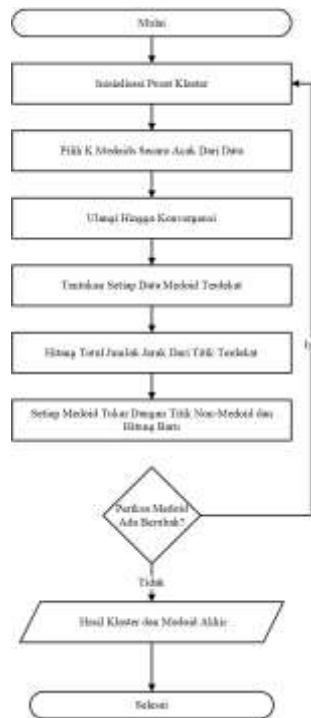
C1: S00001, S00005

C2: S00002, S00003, S00004

Karena klaster tidak berubah dari iterasi sebelumnya (C1 tetap berisi S00001 dan S00005, C2 berisi S00002, S00003, dan S00004), proses ini dapat dihentikan. Klasterisasi telah konvergen.

2. K-Medoid

Gambar di bawah ini memperlihatkan diagram alir (*flowchart*) dari metode K-Medoid, yang menggambarkan secara sistematis tahapan proses *clustering* menggunakan medoid sebagai pusat cluster. Flowchart ini mencakup langkah-langkah mulai dari pemilihan medoid awal, perhitungan jarak setiap data terhadap medoid, penentuan anggota cluster berdasarkan jarak terdekat, hingga pembaruan medoid dan iterasi yang berlanjut sampai kondisi optimal tercapai. Visualisasi ini membantu memahami alur kerja K-Medoid dan memudahkan implementasi algoritma secara terstruktur.



Gambar 3 Flowchart K-Medoid

a. Data Awal

Tabel 8 Data Awal

ID	Study Hours/Week	Attendance Rate	Previous Grades	Participation in Extracurricular Activities	Parent Education Level	Passed
S00001	12.5	95.8	75.0	Yes	Master (4)	Yes
S00002	9.3	95.3	60.6	No	High School (1)	No
S00003	13.2	95.8	64.0	No	Associate (2)	No
S00004	17.6	76.8	62.4	Yes	Bachelor (3)	No

b. Normalisasi Data

Tabel 9 Normalisasi

ID	Study Hours/Week	Attendance Rate	Previous Grades	Participation in Extracurricular Activities	Parent Education Level	Passed
S00001	$\frac{12.5 - 9.3}{8.3} = 0.38$	$\frac{95.8 - 76.8}{19.0} = 1.0$	$\frac{75.0 - 60.6}{14.4} = 1.0$	Yes = 1	Master = 4	Yes

S00002	$\frac{9.3 - 9.3}{8.3} = 0$	$\frac{95.3 - 76.8}{19.0} = 0.97$	$\frac{60.6 - 60.6}{14.4} = 0.0$	No = 0	High School = 1	No
S00003	$\frac{13.2 - 9.3}{8.3} = 0.47$	$\frac{95.8 - 76.8}{19.0} = 1.0$	$\frac{64.0 - 60.6}{14.4} = 0.24$	No = 0	Associate = 2	No
S00004	$\frac{17.6 - 9.3}{8.3} = 1.0$	$\frac{76.8 - 76.8}{19.0} = 0$	$\frac{62.4 - 60.6}{14.4} = 0.13$	Yes = 1	Bachelor = 3	No

c. Menghitung Jarak (Euclidean Distance)

Jarak antara S00001 dan S00002:

$$d(S00001, S00002) = \sqrt{(0.38 - 0.0)^2 + (1.0 - 0.97)^2 + (1.0 - 0.0)^2 + (1 - 0)^2 + (4 - 1)^2}$$

$$d(S00001, S00002) = 3.34$$

$$d(S00001, S00003) = \sqrt{(0.38 - 0.47)^2 + (1.0 - 1.0)^2 + (1.0 - 0.24)^2 + (1 - 0)^2 + (4 - 2)^2}$$

$$d(S00001, S00003) = 2.36$$

$$d(S00001, S00004) = \sqrt{(0.38 - 1.0)^2 + (1.0 - 0.0)^2 + (1.0 - 0.13)^2 + (1 - 1)^2 + (4 - 3)^2}$$

$$d(S00001, S00004) = 1.77$$

d. Penugasan Data ke Medoid Terdekat

S00001 lebih dekat ke medoid S00002 dengan jarak 3.34.

S00003 lebih dekat ke medoid S00001 dengan jarak 2.36.

S00004 lebih dekat ke medoid S00001 dengan jarak 1.77.

Berdasarkan hasil penugasan, data sementara dikelompokkan sebagai berikut:

Cluster 1 (Medoid S00001): S00001, S00003, S00004

Cluster 2 (Medoid S00002): S00002

Menghitung Total Jarak dalam Setiap Cluster: Hitung jarak total antar anggota cluster untuk menentukan medoid baru.

e. Jarak antara S00001 dan S00003:

$$d(S00001, S00003) = \sqrt{(0.38 - 0.47)^2 + (1.0 - 1.0)^2 + (1.0 - 0.24)^2 + (1 - 0)^2 + (4 - 2)^2}$$

$$d(S00001, S00003) = 2.36$$

$$d(S00001, S00004) = \sqrt{(0.38 - 1.0)^2 + (1.0 - 0.0)^2 + (1.0 - 0.13)^2 + (1 - 1)^2 + (4 - 3)^2}$$

$$d(S00001, S00004) = 1.17$$

$$d(S00003, S00004) = \sqrt{(0.47 - 1.0)^2 + (1.0 - 0.0)^2 + (0.14 - 0.13)^2 + (0 - 1)^2 + (2 - 3)^2}$$

$$d(S00003, S00004) = 1.81$$

Total Jarak untuk S00001 sebagai *Medoid*:

$$d_{total} = 2.36 + 1.77 = 4.13$$

Total Jarak untuk S00003 sebagai *Medoid*:

Jarak antara S00003 dan S00001:

Sudah dihitung sebelumnya, $d(S00003, S00001) = 2.36$

Jarak antara S00003 dan S00004:

Sudah dihitung sebelumnya, $d(S00003, S00004) = 1.81$

$$d_{total} S00003 = 2.36 + 1.81 = 4.17$$

Total Jarak untuk S00004 sebagai *Medoid*:

Jarak antara S00004 dan S00001:

Sudah dihitung sebelumnya, $d(S00004, S00001) = 1.77$

Jarak antara S00004 dan S00003:

Sudah dihitung sebelumnya, $d(S00004, S00003) = 1.81$

$$d_{total} S00004 = 1.77 + 1.81 = 3.58$$

Medoid Baru untuk *Cluster* 1:

S00004 (karena memiliki total jarak terendah, yaitu 3.58).

Cluster 2 (*Medoid* S00002): S00002X

Karena hanya memiliki satu data, *medoid* tetap S00002S00002.

Setelah *update medoid*, *medoid* baru adalah:

Cluster 1: *Medoid* = S00004

Cluster 2: *Medoid* = S00002

Kita ulangi perhitungan jarak untuk setiap data ke *medoid* baru.

Jarak ke *Medoid* S00004 (*Cluster* 1):

S00001 ke S00004:

Sudah dihitung sebelumnya, $d(S00001, S00004) = 1.77$

S00003 ke S00004:

Sudah dihitung sebelumnya, $d(S00003, S00004) = 1.81$

S00004 ke S00004: $d(S00004, S00004) = 0.0$

Jarak ke *Medoid* S00002 (*Cluster* 2):

S00001 ke S00002:

Sudah dihitung sebelumnya, $d(S00001, S00002) = 3.34$ $d(S00001, S00002) = 3.34$.

S00003 ke S00002:

$$d(S00001, S00003) = \sqrt{(0.38 - 0.47)^2 + (1.0 - 1.0)^2 + (1.0 - 0.24)^2 + (1 - 0)^2 + (4 - 2)^2}$$

$$d(S00001, S00003) = 1.13$$

$$d(S00001, S00004) = \sqrt{(1.0 - 0.0)^2 + (0.0 - 0.97)^2 + (0.13 - 0.0)^2 + (1 - 0)^2 + (3 - 1)^2}$$

$$d(S00001, S00004) = 2.64$$

Penugasan Akhir:

Cluster 1 (Medoid S00004): S00001, S00003, S00004

Cluster 2 (Medoid S00002): S00002

Karena hasil penugasan tidak berubah, proses selesai.

Hasil Akhir:

Cluster 1: S00001, S00003, S00004 (Medoid = S00004)

Cluster 2: S00002 (Medoid = S00002)

CONCLUSION

Berdasarkan penerapan metode K-Means dan K-Medoids untuk pengelompokan data, dapat disimpulkan bahwa kedua algoritma memiliki karakteristik dan keunggulan masing-masing. Metode K-Means berhasil membagi data menjadi kluster berdasarkan jarak terpendek ke centroid awal, dengan proses iterasi yang cepat hingga mencapai konvergensi. Kelebihan K-Means terletak pada kecepatannya dalam menghitung dan kemudahan interpretasi hasil, terutama untuk dataset besar, namun algoritma ini sensitif terhadap outlier dan hasil klusterisasi sangat tergantung pada pemilihan centroid awal. Sementara itu, K-Medoids menggunakan medoid sebagai pusat kluster yang berasal dari data aktual, dengan proses normalisasi dan perhitungan jarak yang iteratif hingga pembagian kluster stabil. Keunggulan K-Medoids adalah ketahanannya terhadap outlier dan kecocokan untuk dataset berukuran kecil hingga sedang, namun algoritma ini membutuhkan waktu komputasi lebih lama dan hasil klusterisasi dapat dipengaruhi oleh jumlah kluster yang ditentukan sebelumnya. Dengan demikian, pemilihan metode sebaiknya disesuaikan dengan karakteristik dataset, tujuan analisis, serta prioritas antara kecepatan komputasi dan akurasi pengelompokan.

REFERENCE

- Arifandi, M., Hermawan, A., Hermawan, A., Avianto, D., & Avianto, D. (2021). Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid-19 Di DKI Jakarta. *JTT (Jurnal Teknologi Terapan)*, 7(2), 120. <https://doi.org/10.31884/jtt.v7i2.353>
- Fusta, D. (2024). *PENENTUAN ALGORITMA TERBAIK DALAM ANALISIS KLASSTER TINDAK PIDANA DI INDONESIA MENGGUNAKAN K-*. 19(2).
- 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>
- Hendriansyah, B. A., Harjanta, A. T. J., & Latifah, K. (2025). Implementasi Algoritma K-Means

- Clustering Pada Sistem Informasi Geografis Fasilitas Kesehatan Bpjs Kesehatan Kota Semarang. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 7(1), 438–448. <https://doi.org/10.51401/jinteks.v7i1.5618>
- Khalish, F., Piranti, N. M., & Martadireja, O. (2025). *Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means*. 8, 5392–5397.
- Maori, N. A., & Evanita. (2023). Metode Elbow dalam Optimasi pada K-Means Clustering. *Jurnal SIMETRIS*, 14(2), 277–287.
- Meiriza, A., Ali, E., Rahmiati, & Agustin. (2023). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan. *Indonesian Journal of Computer Science*, 12(1), 656–664.
- Rahman, F. K., Sanjaya, J. S., Handayani, L., & Insani, F. (2025). Penerapan Algoritma K-Means Clustering pada Kinerja Mesin Screw press. *Bulletin of Information Technology (BIT)*, 6(2), 59–70. <https://doi.org/10.47065/bit.v5i2.1783>
- Raja, N. A., Tinungki, G. M., & Sirajang, N. (2024). Implementasi Algoritma Centroid Linkage dan K-Medoids dalam Mengelompokkan Kabupaten/Kota di Sulawesi Selatan Berdasarkan Indikator Pendidikan. *ESTIMASI: Journal of Statistics and Its Application*, 5(1), 61–74. <https://doi.org/10.20956/ejsa.v5i1.13605>
- Riska, S. Y., & Farokhah, L. (2023). Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia. *INTEGER: Journal of Information Technology*, 8(1), 1–8. <https://doi.org/10.31284/j.integer.2023.v8i1.3659>
- Safar Dwi Kurniawan, M. K., & Rosalina Yani Widiastuti, S.Kom., M. (2024). *Big Data: Mengenal Big Data dan Implementasinya di Berbagai Bidang*. https://www.researchgate.net/deref/http%3A%2F%2Fwww.buku.sonpedia.com%2F?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19
- Salman, M. D., Pratama, N. R., & A, M. N. F. (2025). *Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Province Based on Availability of Facilities and Infrastructure* Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids dalam Pengelompokan Sekolah di. 5(July), 797–806.
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 229–240. <https://doi.org/10.29100/jupi.v6i2.2008>
- Widodo, T. (2024). Kombinasi Simple Additive Weighted dan Rank Order Centroid Dalam Pemilihan Vendor Catering. *CHAIN: Journal of Computer Technology, Computer Engineering, and Informatics*, 2(1), 11–18. <https://doi.org/10.58602/chain.v2i1.90>