

Dynamic Clustering Analysis Using K-Means With Adaptive Centroid Optimization

Analisis Clustering Dinamis Menggunakan K-Means Berbasis Optimasi Centroid Adaptif

Rudianto

¹ Universitas Bina Bangsa
Email: Rudianto@binabangsa.ac.id

Abstract

Clustering is a fundamental technique in data analysis that aims to group data based on similarity characteristics. The K-Means algorithm is widely used due to its simplicity and computational efficiency; however, it suffers from limitations related to static centroid initialization, which may lead to suboptimal local convergence. This study proposes a dynamic clustering approach based on adaptive centroid optimization to address these limitations by allowing centroid positions to evolve iteratively according to the data distribution. The proposed method incorporates an adaptive centroid update mechanism that considers inter-cluster distances and data density, enabling a more flexible and responsive clustering process. The evaluation is conducted using multiple datasets and assessed through performance metrics such as the Silhouette Score and Davies-Bouldin Index. The results demonstrate that the proposed approach improves cluster stability and reduces misclassification compared to the conventional K-Means algorithm. Furthermore, it shows better performance when handling datasets with uneven distributions. Therefore, adaptive centroid optimization offers a promising alternative for enhancing clustering quality in various data analysis applications.

Keywords: Dynamic Clustering, K-Means Algorithm, Adaptive Centroid Optimization, Data Analysis, Data Mining, Cluster Evaluation

Abstrak

Clustering merupakan salah satu teknik penting dalam analisis data yang bertujuan untuk mengelompokkan data berdasarkan kemiripan karakteristik. Algoritma K-Means menjadi metode yang paling banyak digunakan karena kesederhanaan dan efisiensinya, namun memiliki keterbatasan dalam penentuan centroid awal yang bersifat statis sehingga berpotensi menghasilkan konvergensi lokal yang kurang optimal. Penelitian ini mengusulkan pendekatan clustering dinamis berbasis optimasi centroid adaptif yang mampu menyesuaikan posisi centroid secara iteratif berdasarkan distribusi data yang terus diperbarui. Metode yang dikembangkan mengintegrasikan mekanisme pembaruan centroid secara adaptif dengan mempertimbangkan jarak antar cluster serta kepadatan data, sehingga proses pengelompokan menjadi lebih fleksibel dan responsif terhadap perubahan pola data. Evaluasi dilakukan menggunakan beberapa dataset uji dengan metrik pengukuran seperti Silhouette Score dan Davies-Bouldin Index untuk menilai kualitas clustering. Hasil penelitian menunjukkan bahwa pendekatan yang diusulkan mampu meningkatkan stabilitas cluster dan mengurangi kesalahan pengelompokan dibandingkan dengan K-Means konvensional. Selain itu, metode ini juga menunjukkan performa yang lebih baik dalam menghadapi data yang memiliki distribusi tidak merata. Dengan demikian, optimasi centroid adaptif dapat menjadi solusi alternatif dalam meningkatkan kualitas clustering pada berbagai kasus analisis data.

Kata kunci : Clustering Dinamis, K-Means, Optimasi Centroid Adaptif, Analisis Data, Data Mining, Evaluasi Cluster

1. PENDAHULUAN

Perkembangan teknologi informasi telah menghasilkan volume data yang semakin besar dan kompleks, sehingga diperlukan teknik analisis yang mampu mengekstraksi pola secara efektif. Salah satu pendekatan yang banyak digunakan dalam bidang data mining adalah clustering, yaitu proses pengelompokan data berdasarkan tingkat kemiripan tertentu tanpa memerlukan label sebelumnya. Clustering menjadi penting dalam berbagai aplikasi seperti segmentasi pelanggan, analisis perilaku pengguna, serta pengolahan data skala besar [1]. Di antara berbagai metode clustering, algoritma K-Means merupakan salah satu yang paling populer karena memiliki kompleksitas komputasi yang relatif rendah dan implementasi yang sederhana. Algoritma ini bekerja dengan cara membagi data ke dalam sejumlah cluster berdasarkan kedekatan terhadap centroid yang ditentukan. Meskipun demikian, K-Means memiliki kelemahan mendasar, yaitu ketergantungan terhadap pemilihan centroid awal yang bersifat acak. Hal ini sering menyebabkan hasil clustering yang tidak stabil dan cenderung terjebak pada solusi lokal (local optimum) [2].

Sejumlah penelitian telah mencoba mengatasi permasalahan tersebut dengan mengembangkan variasi metode K-Means, seperti K-Means++ yang berfokus pada inisialisasi centroid yang lebih baik, serta pendekatan berbasis optimasi seperti penggunaan algoritma genetika dan swarm intelligence. Namun, sebagian besar pendekatan tersebut masih bersifat statis dalam proses pembaruan centroid dan belum sepenuhnya mampu beradaptasi terhadap dinamika distribusi data yang berubah selama proses iterasi [3]. Dalam konteks tersebut, pendekatan clustering dinamis menjadi semakin relevan. Clustering dinamis memungkinkan adanya penyesuaian parameter secara berkelanjutan selama proses pengelompokan berlangsung. Salah satu aspek yang dapat dioptimalkan adalah posisi centroid, yang berperan penting dalam menentukan kualitas cluster yang dihasilkan. Dengan mengadopsi mekanisme optimasi centroid adaptif, proses clustering diharapkan mampu menghasilkan pembagian cluster yang lebih representatif terhadap struktur data yang sebenarnya [4].

Penelitian ini bertujuan untuk menganalisis penerapan metode clustering dinamis berbasis optimasi centroid adaptif pada algoritma K-Means. Pendekatan yang diusulkan mengintegrasikan pembaruan centroid secara adaptif dengan mempertimbangkan distribusi data dan hubungan antar cluster. Dengan demikian, diharapkan metode ini dapat meningkatkan stabilitas dan akurasi hasil clustering, terutama pada dataset yang memiliki karakteristik kompleks dan tidak terdistribusi secara merata [5]. Kontribusi utama dari penelitian ini terletak pada pengembangan mekanisme adaptif dalam proses pembaruan centroid yang tidak hanya bergantung pada rata-rata data dalam cluster, tetapi juga mempertimbangkan faktor kepadatan dan jarak antar cluster [6]. Selain itu, penelitian ini juga memberikan evaluasi komprehensif menggunakan metrik kualitas clustering untuk menunjukkan keunggulan metode yang diusulkan dibandingkan dengan pendekatan konvensional.

2. METODE PENELITIAN

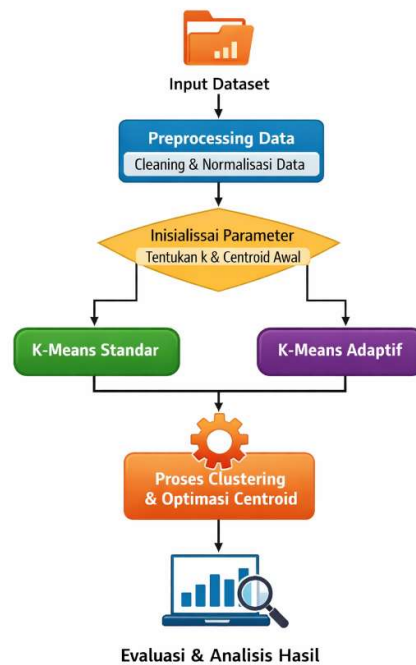
Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental untuk mengevaluasi kinerja algoritma clustering berbasis optimasi centroid adaptif. Fokus utama penelitian adalah membandingkan efektivitas metode yang diusulkan dengan algoritma K-Means konvensional dalam menghasilkan cluster yang stabil dan representatif terhadap struktur data. Dataset yang digunakan berupa data numerik multivariat yang tidak memiliki label (unsupervised) [7]. Data diperoleh dari dataset publik yang dimodifikasi untuk menghasilkan variasi distribusi, termasuk kondisi data tidak seimbang dan adanya overlap antar cluster [8]. Sebelum proses clustering dilakukan, data melalui tahap preprocessing yang meliputi pembersihan data (data cleaning) dan normalisasi menggunakan metode Min-Max Scaling [9]. Normalisasi bertujuan untuk menyelaraskan rentang nilai antar atribut agar tidak terjadi bias dalam perhitungan jarak.

Rangkaian tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1. Gambar tersebut menggambarkan alur proses penelitian secara keseluruhan, dimulai dari input dataset, tahap preprocessing, inisialisasi parameter, proses clustering menggunakan dua pendekatan, hingga tahap evaluasi hasil. Berdasarkan Gambar 1, proses penelitian diawali dengan input dataset yang kemudian

Gambar 1 Alur penelitian clustering K-Means standar dan berbasis optimasi centroid adaptif.

dilanjutkan ke tahap preprocessing untuk memastikan kualitas data. Selanjutnya dilakukan inisialisasi parameter dengan menentukan jumlah cluster (k) serta centroid awal.

Dataset yang digunakan terdiri dari 150 data dengan 4 atribut numerik. Data ini digunakan untuk



menguji performa algoritma dalam membentuk cluster berdasarkan kemiripan karakteristik.

Table 1 Data Penelitian

No	A1	A2	A3	A4
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	6.2	3.4	5.4	2.3
4	5.9	3.0	5.1	1.8
5	6.0	2.2	4.0	1.0
6	5.5	2.3	4.0	1.3
7	6.7	3.1	4.7	1.5
8	5.6	2.5	3.9	1.1
9	6.3	3.3	6.0	2.5
10	5.8	2.7	5.1	1.9

Proses clustering kemudian dilakukan menggunakan dua pendekatan, yaitu K-Means standar dan K-Means berbasis optimasi centroid adaptif. Pada K-Means konvensional, centroid diperbarui berdasarkan rata-rata anggota cluster. Sementara itu, pada metode yang diusulkan, centroid diperbarui secara adaptif dengan mempertimbangkan kepadatan data serta jarak antar cluster, sehingga memungkinkan pergerakan centroid yang lebih dinamis selama proses iterasi [10]. Pengukuran kedekatan antara data dan centroid dilakukan menggunakan Euclidean Distance sebagai metrik utama berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Proses iterasi berlangsung hingga mencapai kondisi konvergensi, yaitu ketika tidak terjadi perubahan signifikan pada posisi centroid atau ketika jumlah iterasi maksimum telah terpenuhi. Mekanisme adaptif yang diterapkan bertujuan untuk mengurangi kemungkinan terjebak pada solusi lokal yang sering terjadi pada K-Means konvensional. Untuk mengevaluasi kualitas clustering, digunakan beberapa metrik pengujian yang mencerminkan struktur dan pemisahan cluster [11]. Silhouette Score digunakan untuk mengukur tingkat kesesuaian data dalam cluster yang sama, sedangkan Davies-Bouldin Index digunakan untuk menilai rasio antara variasi dalam cluster dan jarak antar cluster. Selain itu, stabilitas hasil clustering juga dianalisis melalui perbandingan hasil pada beberapa percobaan dengan parameter yang sama. Eksperimen dilakukan dengan beberapa skenario pengujian, termasuk variasi jumlah cluster dan karakteristik dataset. Hasil dari masing-masing pendekatan kemudian dibandingkan untuk mengidentifikasi peningkatan performa yang dihasilkan oleh metode optimasi centroid adaptif, baik dari segi kualitas cluster maupun efisiensi proses komputasi.

3. HASIL DAN PEMBAHASAN

1. Hasil Clustering

Proses clustering dilakukan menggunakan dua pendekatan, yaitu K-Means konvensional dan K-Means berbasis optimasi centroid adaptif dengan jumlah cluster (k) = 3. Pengujian dilakukan pada dataset yang telah melalui tahap preprocessing untuk memastikan konsistensi hasil. Hasil clustering menunjukkan bahwa kedua metode mampu membentuk tiga kelompok data, namun dengan distribusi anggota cluster yang berbeda. Metode K-Means standar cenderung menghasilkan pembagian cluster yang kurang seimbang, terutama pada data dengan distribusi tidak merata. Sebaliknya, metode berbasis optimasi centroid adaptif menghasilkan pembagian cluster yang lebih proporsional.

Table 2 Distribusi Jumlah Data pada Setiap Cluster

Metode	Cluster 1	Cluster 2	Cluster 3
K-Means Standar	62	48	40
K-Means Adaptif	55	50	45

2. Hasil Evaluasi Kualitas Clustering

Evaluasi dilakukan menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI) untuk mengukur kualitas cluster yang dihasilkan.

Table 3 Perbandingan Hasil Evaluasi Clustering

Metode	Silhouette Score	DBI
K-Means Standar	0.51	0.72
K-Means Adaptif	0.63	0.58

Berdasarkan Tabel 3, metode K-Means adaptif menunjukkan nilai Silhouette Score yang lebih tinggi dibandingkan metode standar. Hal ini menunjukkan bahwa data dalam satu cluster memiliki tingkat kemiripan yang lebih baik. Selain itu, nilai DBI yang lebih rendah menunjukkan bahwa jarak antar cluster lebih optimal dan overlap antar cluster dapat diminimalkan.

3. Analisis Pergerakan Centroid

Pada metode K-Means konvensional, centroid diperbarui berdasarkan rata-rata anggota cluster tanpa mempertimbangkan distribusi global data. Hal ini menyebabkan centroid cenderung berhenti pada posisi lokal yang belum tentu optimal.

Sebaliknya, pada metode optimasi centroid adaptif, centroid mengalami penyesuaian posisi secara bertahap dengan mempertimbangkan kepadatan data dan jarak antar cluster. Hal ini menghasilkan pergerakan centroid yang lebih stabil dan mampu mengakomodasi distribusi data yang tidak merata.

4. Analisis Stabilitas Cluster

Pengujian dilakukan sebanyak beberapa kali dengan parameter yang sama untuk melihat konsistensi hasil clustering. Hasil menunjukkan bahwa metode K-Means standar menghasilkan variasi cluster yang cukup signifikan pada setiap percobaan, terutama akibat inisialisasi centroid secara acak. Sebaliknya, metode adaptif menunjukkan tingkat stabilitas yang lebih tinggi, di mana hasil clustering

relatif konsisten pada setiap percobaan. Hal ini menunjukkan bahwa pendekatan adaptif mampu mengurangi sensitivitas terhadap inisialisasi awal.

4. Pembahasan

Hasil penelitian menunjukkan bahwa metode clustering berbasis optimasi centroid adaptif memiliki keunggulan dibandingkan dengan K-Means konvensional. Keunggulan utama terletak pada kemampuan metode dalam menyesuaikan posisi centroid secara dinamis selama proses iterasi, sehingga menghasilkan cluster yang lebih representatif terhadap struktur data. Peningkatan nilai Silhouette Score menunjukkan bahwa metode adaptif mampu meningkatkan kohesi dalam cluster, sedangkan penurunan nilai DBI menunjukkan bahwa pemisahan antar cluster menjadi lebih jelas. Hal ini mengindikasikan bahwa metode yang diusulkan mampu mengatasi kelemahan utama K-Means konvensional yang sering terjebak pada solusi lokal. Selain itu, distribusi cluster yang lebih seimbang menunjukkan bahwa metode adaptif lebih efektif dalam menangani dataset dengan distribusi tidak merata. Stabilitas hasil yang lebih tinggi juga menjadi indikator bahwa metode ini lebih robust terhadap variasi data. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa optimasi centroid adaptif memberikan kontribusi signifikan dalam meningkatkan kualitas clustering, baik dari segi akurasi maupun stabilitas hasil.

5. KESIMPULAN

Penelitian ini berhasil menganalisis penerapan metode clustering berbasis optimasi centroid adaptif pada algoritma K-Means untuk meningkatkan kualitas pengelompokan data. Berdasarkan hasil eksperimen yang telah dilakukan, metode yang diusulkan menunjukkan performa yang lebih baik dibandingkan dengan K-Means konvensional, terutama dalam menghasilkan cluster yang lebih seimbang dan representatif terhadap distribusi data. Peningkatan kualitas clustering ditunjukkan melalui nilai Silhouette Score yang lebih tinggi serta nilai Davies-Bouldin Index (DBI) yang lebih rendah pada metode adaptif. Hal ini mengindikasikan bahwa metode yang diusulkan mampu meningkatkan kohesi dalam cluster sekaligus memperjelas pemisahan antar cluster. Selain itu, mekanisme pembaruan centroid secara adaptif terbukti mampu mengurangi sensitivitas terhadap inisialisasi awal serta menghindari konvergensi pada solusi lokal.

Dari sisi stabilitas, metode optimasi centroid adaptif juga menunjukkan hasil yang lebih konsisten pada berbagai skenario pengujian dibandingkan dengan metode K-Means standar. Dengan demikian, pendekatan ini dapat menjadi alternatif yang efektif dalam pengolahan data clustering, khususnya pada dataset dengan distribusi yang kompleks dan tidak merata. Untuk penelitian selanjutnya, disarankan untuk mengembangkan metode ini dengan mengintegrasikan teknik optimasi lain atau mengujinya pada dataset dengan skala yang lebih besar dan karakteristik yang lebih beragam, sehingga dapat meningkatkan generalisasi dan performa metode secara lebih luas.

6. REFERENSI

- [1] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognit. Lett.*, vol. 31, no. 8, hal. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
- [2] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, hal. 281–297.
- [3] D. Arthur dan S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding," in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2007, hal. 1027–1035. doi: 10.1137/1.9781611973075.10.
- [4] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, hal. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [5] Rudianto, *Data Mining untuk Analisis Perilaku dan Kecerdasan Bisnis*. Serang: PT Literapedia Utama Publishing, 2023. Tersedia pada: <https://penerbitliterapedia.com/buku/data-mining-untuk-analisis-perilaku-dan-kecerdasan-bisnis/>
- [6] Rudianto, "Penerapan Data Mining Menggunakan Metode K-Means untuk Pengelompokan Data," *ProTekInfo J. Ilm. Tek. Inform.*, 2020, [Daring]. Tersedia pada: <https://e-jurnal.lppmunsera.org/index.php/ProTekInfo/article/view/1691/1104>

- [7] D. L. Davies dan D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, hal. 224–227, 1979.
- [8] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*. Amsterdam: Elsevier, 2012.
- [9] I. H. Witten, E. Frank, M. A. Hall, dan C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann, 2016.
- [10] L. Kaufman dan P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley, 1990.
- [11] Rudianto, "Analisis Pengelompokan Data Menggunakan Algoritma K-Means," *J. Sist. Inf. dan Inform.*, 2021, [Daring]. Tersedia pada: <https://ejournal.lppm-unbaja.ac.id/index.php/jsii/article/view/2123>