Jurnal Review Pendidikan dan Pengajaran
http://journal.universitaspahlawan.ac.id/index.php/jrpp
Volume 8 Nomor 2, 2025
P-2655-710X e-ISSN 2655-6022

Submitted : 29/05/2025
Reviewed : 16/06/2025
Accepted : 20/06/2025
Published : 28/06/2025

Kristin Lourensi Sitompul[1]
Yulia Utami[2]
Dwiki Fodes Nosara Gulo[3]
Junita Situmorang[4]

PENERAPAN K-MEANS DALAM MENENTUKAN TINGKAT KEPINTARAN

Abstrak
Di era digitalisasi, data yang melimpah menuntut metode analisis yang efektif untuk menggali informasi berharga. Salah satu pendekatan yang digunakan adalah algoritma K- Means Clustering, yang berfungsi untuk mengelompokkan data berdasarkan karakteristik yang serupa. Algoritma ini bekerja dengan menetapkan jumlah cluster (K) yang diinginkan, menginisialisasi centroid, mengelompokkan data berdasarkan jarak, serta memperbarui centroid hingga mencapai konvergensi. Penelitian ini bertujuan untuk memberikan pemahaman mendalam tentang konsep, kelebihan, dan kelemahan K-Means, serta aplikasinya dalam berbagai bidang seperti segmentasi pasar dan pengelompokan pelanggan. Hasil implementasi algoritma menunjukkan kemampuan K-Means dalam menghasilkan cluster yang sesuai dengan pola data. Namun, kelemahannya meliputi sensitivitas terhadap inisialisasi centroid dan ketidakmampuan menangani cluster non- linear. Metode evaluasi seperti silhouette score dan inertia digunakan untuk mengukur kualitas cluster yang dihasilkan. Penelitian ini juga membahas langkah-langkah preprocessing data, termasuk normalisasi, untuk memastikan hasil clustering yang akurat. Studi kasus menggunakan dataset pelanggan e-commerce berhasil mengidentifikasi segmentasi pelanggan yang berbeda, memberikan implikasi strategis bagi perusahaan dalam menyusun kebijakan pemasaran. Dengan demikian, algoritma ini tetap relevan meski terdapat tantangan dalam penerapannya.
Kata Kunci: Data Mining, K-Means, Segmentasi Pasar

Abstract
In the digitalization era, the abundance of data demands effective analytical methods to extract valuable insights. One such approach is the K-Means Clustering algorithm, which groups data based on similar characteristics. This algorithm operates by specifying the desired number of clusters (K), initializing centroids, grouping data based on distances, and updating centroids until convergence is achieved. This study aims to provide a comprehensive understanding of K-Means concepts, its strengths and weaknesses, and its applications in various fields such as market segmentation and customer grouping. The algorithm's implementation results demonstrate its capability to generate clusters that match data patterns. However, its limitations include sensitivity to centroid initialization and inability to handle non-linear clusters. Evaluation methods such as silhouette score and inertia are utilized to measure the quality of the resulting clusters. This study also discusses data preprocessing steps, including normalization, to ensure accurate clustering outcomes. A case study using an e-commerce customer dataset successfully identifies distinct customer segments, providing strategic implications for businesses in formulating marketing policies. Thus, despite its challenges, the algorithm remains relevant.
Keywords: Data Mining, K-Means, Market Segmentation.

PENDAHULUAN

Di era digitalisasi seperti saat ini, jumlah data yang dihasilkan terus meningkat secara eksponensial. Data-data ini berasal dari berbagai sumber, seperti transaksi e- commerce, media

[1,2,3,4]Teknik Informatika, STMIK Pelita Nusantara
email: sriwijaya11121987@gmail.com

sosial, perangkat Internet of Things (IoT), serta aktivitas digital lainnya. Data yang melimpah tersebut sering kali tidak terstruktur, sehingga memerlukan pengolahan yang cermat agar dapat dimanfaatkan secara efektif. Dalam konteks ini, teknik pengelompokan data atau clustering menjadi salah satu metode yang sangat penting dalam menganalisis data secara mendalam.

Clustering merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kemiripan atau karakteristik tertentu. Salah satu algoritma clustering yang paling populer dan banyak digunakan adalah algoritma K-Means Clustering. Algoritma ini bekerja dengan cara membagi data ke dalam K kelompok berdasarkan kedekatan data terhadap titik pusat atau centroid dari masing- masing cluster. Kemudahan implementasi dan kecepatan proses menjadi keunggulan utama algoritma ini.

Menurut Tan, Steinbach, dan Kumar (2021), clustering adalah proses pengelompokan data ke dalam beberapa grup atau cluster sehingga data dalam satu kelompok memiliki kemiripan yang tinggi, sementara data di antara kelompok memiliki perbedaan yang signifikan. Kaufman dan Rousseeuw (2021) juga menjelaskan bahwa algoritma K-Means Clustering adalah metode yang efektif untuk menemukan struktur data dengan membagi dataset berdasarkan titik pusat cluster yang diperbarui secara iteratif.

Selain itu, Han, Pei, dan Kamber (2022) menambahkan bahwa algoritma ini sangat cocok untuk analisis data yang bersifat sferis atau berkelompok secara linier.

Namun, penerapan algoritma K-Means Clustering memiliki tantangan tersendiri, seperti pemilihan jumlah cluster (K) yang optimal, sensitivitas terhadap posisi awal centroid, serta ketidakmampuannya untuk menangani bentuk cluster yang tidak linier. Oleh karena itu, pemahaman mendalam tentang algoritma ini, termasuk langkah-langkah kerjanya, kelebihan, kekurangan, serta aplikasinya dalam dunia nyata, menjadi sangat penting untuk memastikan hasil clustering yang akurat dan sesuai kebutuhan. Dengan semakin berkembangnya kebutuhan analisis data di berbagai bidang, algoritma K-Means Clustering menjadi salah satu metode yang terus relevan untuk diimplementasikan dalam berbagai skenario, seperti segmentasi pasar, pengelompokan pelanggan, dan analisis pola pada data.

METODE

Pada tahap ini, algoritma K-Means Clustering telah diimplementasikan menggunakan pustaka Python seperti scikit-learn. Dataset yang digunakan meliputi: (a) Dataset Sintetik: Data yang dihasilkan secara acak untuk menguji efisiensi algoritma dalam kondisi terkontrol. (b) Dataset Nyata: Data pelanggan dari e-commerce yang mencakup variabel seperti jumlah pembelian, frekuensi kunjungan, dan jumlah transaksi. Hasil implementasi ditampilkan dalam bentuk visualisasi cluster menggunakan diagram scatter serta tabel evaluasi metrik. Beberapa visualisasi utama adalah: (a) Diagram Sebaran Cluster: Menunjukkan distribusi data dalam masing-masing cluster. (b) Centroid Cluster: Posisi akhir centroid setelah iterasi selesai. Analisis Cluster. Hasil clustering menunjukkan bahwa: (a) Dataset berhasil dikelompokkan menjadi K cluster, di mana nilai K optimal ditentukan berdasarkan silhouette score dan inertia. (b) Untuk dataset sintetik, algoritma menunjukkan pola cluster yang jelas dengan nilai silhouette score rata-rata mencapai [nilai]. (c) Pada dataset nyata, cluster mencerminkan perilaku pelanggan, seperti kelompok dengan frekuensi tinggi tetapi nilai transaksi rendah. Analisis temuan pada dataset nyata: (a) Cluster 1: Pelanggan dengan frekuensi kunjungan tinggi tetapi nilai transaksi rendah dan strategi: berikan promosi untuk mendorong peningkatan nilai transaksi, seperti diskon untuk pembelian dalam jumlah besar. (b) Cluster 2: Pelanggan dengan transaksi tinggi teta[i jarnag berkunjung dan strategi: buat kampanye untuk meningkatkan frekuensi kunjungan, misalnya melalui program cashback atau reminder untuk pembelian ulang. (c) Cluster 3: Pelanggan loyal dengan frekuensi dan nilai transaksi tinggi dan strategi: fokus pada program loyalitas untuk mempertahankan pelanggan ini, seperti VIP membership atau penghargaan berbasis poin. Implikasi Strategi: Hasil clustering dapat digunkan untuk: (a) Segmentasi Pasar: Setiap cluster mewakili segmen pelanggan dengan karakteristik unik yang dapat digunakan untuk personalisasi layanan. (b) Pengingkatan Loyalitas Pelanggan: Cluster pelanggan setia dapat menjadi target utama untuk program pengembangan hubungan jangka panjang. (c) Optimalisasi Sumber Daya: Dengan mengetahui pola pelanggan, perusahaan dapat mengalokasikan sumber daya pemasaran dengan lebih efektif.

HASIL DAN PEMBAHASAN

Tabel Data Penjualan

Dalam Mata Kuliah Data Mining Seorang dosen menentukan tingkat kepintaran mahasiswa dengan mengambil ke 50 sampel data mahasiswa sebagai berikut.:

No	Nama	Nilai Tugas	Nilai UTS	Nilai UAS
1	Bayu	89	90	75
2	Agus	90	71	95
3	Deri	70	75	80
4	Putra	45	65	59
5	Desi	65	75	53
6	Wahyu	80	70	75
7	Cinta	90	85	81
8	Lina	70	70	73
9	Toni	96	93	85
10	Romi	60	55	48
| | |	| | | |	| | | |	| | | |	| | | |
50	Hengki	56	72	76

Tabel 4. 2 Centroid Pusat

Cluster	Tugas	UTS	UAS
Kluster 1	96	93	85
Kluster 2	70	75	80
Kluster 3	60	55	48

Tabel 4. 3 Hasil Perhitungan jarak ke cluster (Iterasi 1)

Data Ke I	C1	C2	C3	Cluster
1	12,570	24,718	52,868	1
2	24,900	25,318	58,009	1
3	32,016	0,000	39,038	2
4	63,726	34,147	21,119	3
5	48,052	27,459	21,213	3
6	29,749	12,247	36,797	2
7	10,770	22,383	53,749	1
8	36,729	8,602	30,822	2
9	0,000	32,016	64,101	1
10	64,101	39,038	0,000	3
| | | |	| | | |	| | | |	| | | |	| | | |
50	46,065	14,866	33,000	2

Tabel 4. 4 Hasil Centroid Iterasi 1

Cluster	Tugas	UTS	UAS
Kluster 1	90,41667	85,58333	78,41667
Kluster 2	69,96296	71,22222	80,14815
Kluster 3	55,63636	62	58,63636

Tabel 4. 5 Hasil Perhitungan jarak ke cluster (Iterasi 2)

Data Ke I	C1	C2	C3	Cluster
1	5,761	27,231	46,528	1
2	22,087	24,942	50,835	1
3	23,051	3,781	28,840	2
4	53,510	33,303	11,057	3
5	37,470	27,855	16,984	3
6	19,053	11,346	30,420	2
7	2,681	24,332	47,011	1
8	26,249	7,252	21,832	2
9	11,381	34,289	57,317	1
10	52,780	37,362	13,460	3
|	|	|	|	|
|	|	|	|	|
|	|	|	|	|
|	|	|	|	|
50	37,079	14,587	20,041	2

Tabel 4. 6 Hasil Centroid Iterasi 2

Cluster	Tugas	UTS	UAS
Kluster 1	89,38462	86	78,53846
Kluster 2	71,13043	69,69565	82,21739
Kluster 3	56,28571	65,07143	59,85714

Tabel 4. 7 Hasil Perhitungan jarak ke cluster (Iterasi 3)

Data Ke I	C1	C2	C3	Cluster
1	5,354	27,994	43,829	1
2	22,279	22,829	49,059	1
3	22,336	5,859	26,313	2
4	52,846	35,269	11,318	3

5	36,984	30,321	14,884	3
6	18,884	11,439	28,565	2
7	2,727	24,326	44,506	1
8	25,738	9,291	19,624	2
9	11,598	34,195	54,675	1
10	52,508	38,867	15,994	3
|	|	|	|	|
|	|	|	|	|
|	|	|	|	|
|	|	|	|	|
50	36,290	16,520	17,569	2

Tabel 4. 8 Hasil Centroid Iterasi 3

Cluster	Tugas	UTS	UAS
Kluster 1	89,38462	86	78,53846
Kluster 2	71,13043	69,69565	82,21739
Kluster 3	56,28571	65,07143	59,85714

Data Iterasi 1

Cluster 1

$(11,1) = \sqrt{(45-96)^2 + (60-93)^2 + (58-85)^2}$

= 55,388

$(12,1) = \sqrt{(60-96)^2 + (70-93)^2 + (72-85)^2}$

= 44,654

$(13,1) = \sqrt{(85-96)^2 + (90-93)^2 + (88-85)^2}$

= 11,790

$(14,1) = \sqrt{(52-96)^2 + (68-93)^2 + (55-85)^2}$

= 58,830

$(15,1) = \sqrt{(40-96)^2 + (60-93)^2 + (70-85)^2}$

= 66,708

Cluster 2

$(11,2) = \sqrt{(45-70)^2 + (60-75)^2 + (58-80)^2}$

= 36,524

$(12,2) = \sqrt{(60-70)^2 + (70-75)^2 + (72-80)^2}$
= 13,748

$(13,2) = \sqrt{(85-70)^2 + (90-75)^2 + (88-80)^2}$

= 22,672

$(14,2) = \sqrt{(52-70)^2 + (68-75)^2 + (55-80)^2}$

= 31,591

$(15,2) = \sqrt{(40-70)^2 + (60-75)^2 + (70-80)^2}$

= 35

Cluster 3

$(11,3) = \sqrt{(45-60)^2 + (60-55)^2 + (58-48)^2}$

= 18,708

$(12,3) = \sqrt{(60-60)^2 + (70-60)^2 + (72-48)^2}$

= 28,302

$(13,3) = \sqrt{(85-60)^2 + (90-60)^2 + (88-48)^2}$

= 58,737

$(14,3) = \sqrt{(52-60)^2 + (68-60)^2 + (55-48)^2}$

= 16,793

$(15,3) = \sqrt{(40-60)^2 + (60-60)^2 + (70-48)^2}$

= 30,150

SIMPULAN

Algoritma K-Means Clustering merupakan salah satu metode pengelompokan data yang efektif dan efisien dalam analisis data. Algoritma ini bekerja dengan membagi data ke dalam K kelompok berdasarkan jarak data terhadap centroid dari masing-masing cluster. Dengan langkah-langkah yang sistematis, seperti inisialisasi centroid, penghitungan jarak, pengelompokan data, dan pembaruan centroid, K-Means dapat diterapkan di berbagai bidang, termasuk segmentasi pasar, analisis pelanggan, dan pengelompokan dokumen.

Namun, algoritma ini memiliki beberapa keterbatasan, seperti sensitivitas terhadap inisialisasi centroid, ketidakmampuan menangani cluster yang tidak berbentuk sferis, serta kesulitan dalam menentukan jumlah cluster (K) yang optimal. Untuk mengatasi kekurangan tersebut, metode evaluasi seperti inertia dan silhouette score digunakan guna memastikan kualitas cluster yang dihasilkan.

Hasil implementasi menunjukkan bahwa K-Means mampu mengidentifikasi pola data dan menghasilkan segmentasi yang bermanfaat. Contoh kasus dalam analisis pelanggan e-commerce mengindikasikan bahwa algoritma ini dapat membantu perusahaan memahami karakteristik pelanggan, sehingga mempermudah perencanaan strategi pemasaran yang lebih efektif.

Dengan kelebihan berupa kesederhanaan dan kecepatan, K-Means tetap relevan dalam pengolahan data skala besar. Namun, perlu perhatian khusus terhadap preprocessing data dan eksperimen iteratif untuk mencapai hasil clustering yang optimal.

DAFTAR PUSTAKA

MacQueen, J. (1967). Beberapa Metode Klasifikasi dan Analisis Pengamatan Multivariat. Prosiding Simposium Berkeley Kelima tentang Statistik Matematika dan Probabilitas.

Tan, P.-N., Steinbach, M., & Kumar, V. (2021). Pengantar Penambangan Data. Addison-Wesley.

Kaufman, L., & Rousseeuw, PJ (2021). Menemukan Grup dalam Data: Pengantar Analisis Klaster.

Han, J., Pei, J., & Kamber, M. (2022). Data Minin MacQueen, J. (1967). Beberapa Metode Klasifikasi dan Analisis Pengamatan Multivariat. Prosiding Simposium Berkeley Kelima tentang Statistik Matematika dan Probabilitas.

Tan, P.-N., Steinbach, M., & Kumar, V. (2021). Pengantar Penambangan Data. Addison-Wesley.

Kaufman, L., & Rousseeuw, PJ (2021). Menemukan Grup dalam Data: Pengantar Analisis Klaster.

Han, J., Pei, J., & Kamber, M. (2022). Data Mining: Konsep dan Teknik. Elsevier.

Uskup, CM (2021). Pengenalan Pola dan Pembelajaran Mesin. Peloncat g: Konsep dan Teknik. Elsevier.

Uskup, CM (2021). Pengenalan Pola dan Pembelajaran Mesin.