E-ISSN 3032-601X & P-ISSN 3032-7105
Vol. 3, No. 1, 2026

# MISTER

## Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research

Jurnal Penelitian Multidisiplin dalam Ilmu Pengetahuan, Teknologi dan Pendidikan

UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH
mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology and Educational Research

# Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 977-986

# Pemanfaatan Algoritma K-Means dalam Menentukan Cluster pada Glass Dataset Secara Efektif

Husni Husni, Hasbi Firmansyah, Wahyu Asriyani, Rizki Prasetyo Tulodo

Universitas Pancasakti Tegal

Article in Journal of MISTER

| | | |
|---|---|---|
| Available at | : | https://jurnal.serambimekkah.ac.id/index.php/mister/index |
| DOI | : | https://doi.org/10.32672/mister.v3i1. 4020 |

How to Cite this Article

| | | |
|---|---|---|
| APA | : | Husni, H., Firmansyah, H., Asriyani, W. ., & Prasetyo Tulodo, R. (2025). Pemanfaatan Algoritma K-Means dalam Menentukan Cluster pada Glass Dataset Secara Efektif. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 977 – 986. https://doi.org/10.32672/mister.v3i1.4020 |
| Others Visit | : | https://jurnal.serambimekkah.ac.id/index.php/mister/index |

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.

ISSN 3032-7105 ISSN 3032-601X

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research
UNIVERSITAS SERAMBI MEKKAH
https://jurnal.serambimekkah.ac.id/index.php/mister

e-ISSN3032-601X & p-ISSN3032-7105
Vol. 3 No. 1, Tahun 2026
Doi: 10.32672/mister.v3i1.4020
Hal. 977-986

# Pemanfaatan Algoritma K-Means dalam Menentukan Cluster pada Glass Dataset Secara Efektif

**Husni Husni[1], Hasbi Firmansyah[2], Wahyu Asriyani [3], Rizki Prasetyo Tulodo[4]**
Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal, Tegal, Indonesia[1,2,4]
Program Studi Pendidikan Bahasa dan Sastra Indonesia,Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Tegal, Indonesia[3]

*Email: husnioston422@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com
Rizki.prasetyo.tulodo@gmail.com

Diterima: 10-12-2025 | Disetujui: 20-12-2025 | Diterbitkan: 22-12-2025

***ABSTRACT***

*Glass is a material with complex and diverse chemical compositions, making the identification and grouping of glass types a challenging task that requires computational approaches capable of extracting latent patterns from high-dimensional numerical data. This study aims to utilize the K-Means algorithm to effectively determine clusters within the Glass Dataset based on chemical composition attributes such as refractive index, Na, Mg, Al, Si, K, Ca, Ba, and Fe. The research methodology consists of dataset collection, data preprocessing through normalization, data processing using RapidMiner, optimal cluster number determination, K-Means clustering model construction, and evaluation and interpretation of clustering results. Cluster quality was assessed using internal validation metrics, particularly the Davies–Bouldin Index and performance vector analysis. The results indicate that the K-Means algorithm successfully formed three main clusters with distinct chemical characteristics, as evidenced by a Davies–Bouldin Index value of 0.530, which reflects good cluster separation and adequate compactness. The resulting clusters represent groups of glass with standard compositions, specific compositional variations, and extreme characteristics that may correspond to rare samples or outliers. These findings demonstrate that K-Means is effective for clustering the Glass Dataset and has potential applications in material identification, forensic analysis, and chemical composition–based research.*

***Keywords**: Clustering, Glass Dataset, K-Means Algorithm, Material Identification, Unsupervised Learning*

**ABSTRAK**

Kaca merupakan material dengan komposisi kimia yang kompleks sehingga pengelompokan jenis kaca memerlukan pendekatan komputasional untuk mengekstraksi pola laten dari data berdimensi tinggi. Penelitian ini bertujuan memanfaatkan algoritma K-Means untuk menentukan cluster pada Glass Dataset berdasarkan atribut kimia seperti refractive index, Na, Mg, Al, Si, K, Ca, Ba, dan Fe. Metode penelitian meliputi pengumpulan dataset, preprocessing melalui normalisasi, pengolahan data menggunakan RapidMiner, penentuan jumlah cluster optimal, pembentukan model K-Means, serta evaluasi dan interpretasi hasil. Kualitas clustering dievaluasi menggunakan Davies–Bouldin Index dan performance vector. Hasil penelitian menunjukkan bahwa K-Means membentuk tiga cluster utama dengan karakteristik kimia berbeda, ditunjukkan oleh nilai Davies–Bouldin Index sebesar 0,530 yang menandakan pemisahan dan kekompakan cluster yang baik. Cluster yang terbentuk merepresentasikan kaca dengan komposisi umum, variasi tertentu, serta karakteristik ekstrem yang berpotensi sebagai outlier. Dengan demikian, algoritma K-Means efektif

Copyright © 2025 The Author(s) This article is distributed under a
Creative Commons Attribution-ShareAlike 4.0 International License

e-ISSN3032-601X & p-ISSN3032-7105

digunakan untuk analisis pengelompokan Glass Dataset dan mendukung identifikasi material serta analisis berbasis komposisi kimia.
**Kata kunci:** Clustering, Dataset Kaca, Identifikasi Material, K-Means, Pembelajaran Tanpa Pengawasan

## PENDAHULUAN

Kaca sebagai material memiliki komposisi kimia yang kompleks dan bervariasi, sehingga identifikasi jenis kaca memerlukan pendekatan komputasional yang mampu mengungkap pola laten dalam data(Meng et al., 2023). Metode clustering unsupervised seperti K-Means sangat sesuai karena mampu mengelompokkan data berdasarkan kemiripan tanpa memerlukan label awal (Lv et al., 2023). Penelitian modern menunjukkan bahwa K-Means mampu mengklasifikasikan artefak kaca dengan konsisten pada tipe-tipe seperti high-potassium glass dan lead-barium glass (Wang & Shen, 2023). Efektivitas algoritma K-Means sangat (Kong et al., 2023). dipengaruhi oleh pemilihan parameter seperti jumlah cluster (K), inisialisasi centroid, serta teknik pra-pengolahan data (Dong, 2023). Tantangan tersebut mendorong penggunaan metode pendukung seperti K-Means++ dalam meningkatkan stabilitas dan kualitas inisialisasi centroid (Y. Zhang et al., 2023)Evaluasi jumlah cluster optimal juga umum dilakukan menggunakan metode seperti Elbow, Silhouette, maupun pendekatan validitas cluster lainnya (J. Zhang, 2022). Studi ilmiah terbaru menunjukkan bahwa penerapan pipeline K-Means yang tepat dapat menghasilkan subclassifikasi kaca dengan akurasi tinggi dan konsistensi struktural . Hal ini membuktikan bahwa K-Means efektif digunakan dalam identifikasi artefak dan konservasi material berbasis analisis komposisi kimia (Ji, 2023). Glass Dataset sebagai objek penelitian menyediakan fitur kimia yang relevan sehingga memungkinkan penerapan K-Means untuk mengkaji pola komposisi serta struktur kelompok secara sistematis (Yin, 2022). Berdasarkan kondisi tersebut, penelitian ini berfokus untuk mengevaluasi pemanfaatan K-Means pada Glass Dataset secara efektif melalui penentuan K optimal, interpretasi cluster, dan analisis hasil pengelompokannya (Yan et al., 2023). Implementasi K-Means pada konteks material gelas dan pengolahan data berbasis citra turut menunjukkan fleksibilitas algoritma ini dalam menangani data material yang kompleks dan heterogen (Chen & Ji, 2023). Selain itu, pendekatan clustering berbasis K-Means yang dikombinasikan dengan analisis statistik multivariat terbukti mampu meningkatkan interpretabilitas hasil pengelompokan pada data material kompleks (Loh et al., 2024)

## METODE PENELITIAN

Metode penelitian ini meliputi analisa permasalahan, perancangan arsitektur metode, tahapan review, serta ketentuan teknis penulisan sesuai standar publikasi ilmiah. Penelitian ini berjudul "Pemanfaatan Algoritma K-Means dalam Menentukan Cluster pada Glass Dataset Secara Efektif", sehingga seluruh metode diarahkan untuk memperoleh jumlah cluster optimal, memetakan pola komposisi kimia pada Glass Dataset, serta mengevaluasi kualitas pengelompokan.

Permasalahan utama dalam penelitian ini adalah bagaimana menentukan kelompok (cluster) yang paling representatif pada *Glass Dataset* berdasarkan fitur-fitur komposisi kimia seperti Na, Mg, Al, Si, K, Ca, Ba, dan Fe. Variasi kandungan kimia pada kaca sangat kompleks dan sering kali tidak menunjukkan pola yang dapat diamati secara langsung menggunakan analisis konvensional. Oleh karena itu, identifikasi jenis kaca memerlukan pendekatan komputasional yang mampu mengekstraksi pola laten dari data numerik berdimensi tinggi (Gao et al., 2023).

Algoritma K-Means dipilih karena telah banyak digunakan dan terbukti efektif dalam mengelompokkan sampel kaca berdasarkan komposisi kimia pada berbagai penelitian material dan forensik. Metode ini mampu mendeteksi kelompok seperti *high-potassium glass*, *lead-barium glass*, maupun *soda-lime glass* tanpa memerlukan label awal (*unsupervised learning*) ("The Many Faces of Ceramics: New Developments in the Study of Archaeological Ceramics from the 15th European Meeting on Ancient Ceramics (Barcelona, 16th–18th September 2019)," 2022). Namun, penggunaan K-Means bukan tanpa kelemahan. Algoritma ini sangat sensitif terhadap pemilihan nilai K, pemilihan centroid awal, serta perbedaan skala antar-fitur yang dapat menghasilkan cluster bias atau tidak stabil (Yang et al., 2021).

Dengan mempertimbangkan permasalahan tersebut, penelitian ini memfokuskan pada dua aspek penting seperti menentukan nilai K yang optimal melalui pendekatan sistematis seperti *Elbow Method* dan metrik validitas internal, serta mengevaluasi kualitas cluster berdasarkan *Silhouette Score*, *Davies–Bouldin Index*, dan analisis karakteristik kimia dari setiap cluster. Pendekatan ini diharapkan mampu menghasilkan pemetaan cluster yang tidak hanya akurat tetapi juga representatif terhadap variasi komposisi kimia pada Glass Dataset.

**Pengumpulan Dataset**

| id number | Ri | Na | Mg | Ai | Si | K | Ca | Ba | Fe |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 152.101 | 13.64 | 4.49 | 1.10 | 71.78 | 0.06 | 8.75 | 0.00 | 0.00 |
| 2 | 151.761 | 13.89 | 3.60 | 1.36 | 72.73 | 0.48 | 7.83 | 0.00 | 0.00 |
| 3 | 151.618 | 13.53 | 3.55 | 1.54 | 72.99 | 0.39 | 7.78 | 0.00 | 0.00 |
| 4 | 151.766 | 13.21 | 3.69 | 1.29 | 72.61 | 0.57 | 8.22 | 0.00 | 0.00 |
| 5 | 151.742 | 13.27 | 3.62 | 1.24 | 73.08 | 0.55 | 8.07 | 0.00 | 0.00 |
| 6 | 151.596 | 12.79 | 3.61 | 1.62 | 72.97 | 0.64 | 8.07 | 0.00 | 0.26 |
| 7 | 151.743 | 13.30 | 3.60 | 1.14 | 73.09 | 0.58 | 8.17 | 0.00 | 0.00 |
| … | … | … | … | … | … | … | … | … | … |
| 214 | 151.711 | 14.23 | 0.00 | 2.08 | 73.36 | 0.00 | 8.62 | 1.67 | 0.00 |

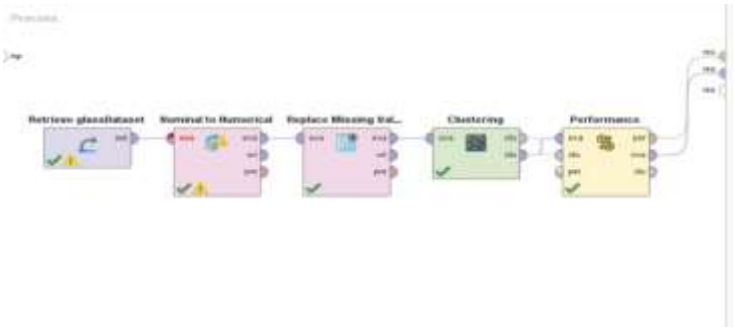
Pada tahap pertama, proses penelitian dimulai dengan pengumpulan dataset yang digunakan sebagai dasar analisis. Dalam studi ini, peneliti memanfaatkan *Glass Dataset*, yaitu kumpulan data yang berisi berbagai karakteristik kimia dan fisik dari sampel kaca yang berasal dari beberapa jenis kategori. Dataset ini diperoleh dari repositori data publik, sehingga setiap atribut dan label di dalamnya sudah terstandarisasi dan siap digunakan dalam proses *data mining*. Tahap pengumpulan dataset ini sangat penting karena memastikan bahwa data yang digunakan akurat, memiliki struktur yang jelas, serta dapat mendukung proses analisis clustering menggunakan algoritma K-Means. Data kemudian ditinjau secara menyeluruh untuk memahami jumlah atribut, jumlah kelas, serta variasi nilai yang terdapat di dalamnya.

**Preprocessing Data**

Tahap *preprocessing data* dilakukan untuk meningkatkan kualitas dataset sebelum masuk ke tahap analisis. Pada Glass Dataset, preprocessing dilakukan melalui beberapa langkah, seperti pembersihan data, normalisasi, dan penanganan nilai yang tidak konsisten. Pembersihan data bertujuan untuk memastikan tidak terdapat nilai kosong atau data yang tidak relevan. Selanjutnya, proses normalisasi diterapkan karena algoritma K-Means sangat sensitif terhadap perbedaan skala antar atribut. Dengan normalisasi, setiap fitur seperti Fe, Na, Mg, Si, dan Ca dipastikan berada pada rentang nilai yang seragam sehingga tidak ada atribut yang mendominasi proses perhitungan jarak. Tahap ini juga mencakup pengecekan outlier untuk mencegah terbentuknya cluster yang tidak representatif. Semua langkah preprocessing ini bertujuan menghasilkan dataset yang bersih, konsisten, dan siap untuk dianalisis lebih jauh.

**Pengolahan Data di RapidMiner**

e-ISSN3032-601X & p-ISSN3032-7105



Setelah dataset siap digunakan, tahap berikutnya adalah pengolahan data menggunakan perangkat lunak RapidMiner. Pada tahap ini, dataset diimpor ke dalam RapidMiner dan dilakukan konfigurasi awal seperti penentuan atribut, pemeriksaan tipe data, serta pemilihan operator yang diperlukan. RapidMiner dipilih karena memiliki antarmuka visual yang memudahkan proses *data mining* tanpa memerlukan coding kompleks, serta menyediakan operator K-Means yang lengkap. Data yang telah melalui preprocessing kemudian dimasukkan ke dalam *workflow* yang berisi rangkaian operator seperti *Normalize*, *K-Means*, dan *Cluster Model Reader*. Tahap ini memastikan proses pengolahan data berjalan sistematis dan sesuai dengan rancangan penelitian.

**Pembentukan Model Clustering K-Means**

Tahap pembentukan model dilakukan dengan menggunakan algoritma K-Means untuk mengelompokkan Glass Dataset menjadi beberapa cluster berdasarkan kemiripan atribut. Proses ini dimulai dari penentuan nilai K (jumlah cluster), yang dapat ditentukan menggunakan metode seperti *Elbow Method* untuk mendapatkan jumlah cluster yang optimal. Setelah nilai K ditentukan, RapidMiner akan menjalankan algoritma K-Means dengan melakukan inisialisasi centroid awal, menghitung jarak Euclidean antar data, serta memperbarui posisi centroid hingga proses konvergensi tercapai. Hasil dari tahap ini adalah sebuah model cluster yang menunjukkan pengelompokan data berdasarkan karakteristik kimia yang paling dominan.

**Penerapan Model dan Evaluasi**

Model clustering yang telah terbentuk kemudian diterapkan pada keseluruhan dataset untuk menghasilkan pengelompokan akhir. Pada tahap penerapan ini, setiap data sampel kaca akan masuk ke cluster tertentu berdasarkan jarak terdekat dengan centroid. Setelah penerapan model, dilakukan evaluasi kualitas clustering menggunakan beberapa metrik, seperti *Davies–Bouldin Index*, *Silhouette Coefficient*, atau visualisasi hasil cluster melalui grafik scatter. Evaluasi ini dilakukan untuk memastikan bahwa cluster yang terbentuk benar-benar menggambarkan pola alami dalam Glass Dataset dan memiliki tingkat pemisahan antar cluster yang baik.

**Interpretasi Hasil dan Analisis Prediksi**

Tahap terakhir adalah interpretasi hasil dan analisis prediksi, yang dilakukan untuk memahami makna dan implikasi dari pembentukan cluster. Pada tahap ini, setiap cluster dianalisis berdasarkan nilai rata-rata pada atribut-atribut tertentu, sehingga dapat diketahui karakteristik khas dari masing-masing cluster. Misalnya, adanya cluster yang mewakili kaca jenis bangunan, kaca botol, atau kaca kendaraan berdasarkan komposisi kimianya. Selain itu, analisis prediksi dilakukan dengan melihat bagaimana pola cluster yang terbentuk dapat digunakan pada kasus nyata, seperti identifikasi jenis kaca pada laboratorium

e-ISSN3032-601X & p-ISSN3032-7105

forensik. Tahap ini juga mencakup pembahasan mengenai keterbatasan penelitian dan rekomendasi untuk pengembangan metode clustering yang lebih efektif di masa mendatang.

**Rumus / Persamaan Matematika**

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

Jika diperlukan (misalnya untuk menghitung jarak Euclidean atau metrik evaluasi), persamaan matematika akan ditulis menggunakan Microsoft Equation Editor atau MathType, ditempatkan di tengah halaman dan diberi nomor urut (1), (2), …, kemudian dirujuk dalam teks dengan format "(Persamaan 1)". Contoh persamaan yang umum digunakan:

**Pengacuan Pustaka**

Semua pustaka yang dirujuk dalam naskah mengikuti format sitasi IEEE. Referensi diurutkan berdasarkan kemunculan pertama kali dalam teks, dan daftar pustaka hanya berisi publikasi yang benar-benar disitasi. Disarankan penggunaan manajer referensi seperti Mendeley atau Zotero agar format tetap konsisten.

## HASIL DAN PEMBAHASAN

### Hasil Pengujian

Penelitian ini bertujuan untuk melakukan proses klasterisasi terhadap dataset Glass Identification yang terdiri dari 214 data dengan sejumlah atribut kimia seperti *Refractive Index (Ri), Sodium (Na), Magnesium (Mg), Aluminium (Al), dan Silikon (Si)*. Pengujian dilakukan menggunakan algoritma klasterisasi sehingga menghasilkan pembagian data ke dalam tiga klaster utama. Untuk memastikan kualitas model, dilakukan pengukuran menggunakan *Davies Bouldin Index* serta analisis *Performance Vector*.

### Evaluasi Davies Bouldin Index


Hasil perhitungan Davies Bouldin Index (DBI) ditampilkan pada Gambar *Davies Bouldin*, di mana nilai DBI yang diperoleh adalah **0,530**. Davies Bouldin Index merupakan salah satu ukuran kualitas klaster yang digunakan untuk menilai tingkat pemisahan dan kekompakan klaster. Nilai DBI berada pada rentang 0 sampai dengan ∞, dan semakin rendah nilai DBI, semakin baik struktur klaster yang terbentuk.

Nilai DBI sebesar 0,530 menunjukkan bahwa klaster-klaster yang terbentuk memiliki pemisahan antarklaster yang baik, dengan tumpang tindih (overlapping) antar klaster yang relatif kecil. Selain itu, nilai

e-ISSN3032-601X & p-ISSN3032-7105

tersebut juga menggambarkan bahwa rata-rata jarak antar data dengan centroid masing-masing relatif stabil dan tidak terlalu besar. Dengan demikian, struktur klaster dapat dianggap cukup kuat dan representatif untuk digunakan dalam analisis lanjutan.

**Hasil Performance Vector**

**PerformanceVector**

```
PerformanceVector:
Avg. within centroid distance: 24332.934
Avg. within centroid distance_cluster_0: 19683.481
Avg. within centroid distance_cluster_1: 23498.532
Avg. within centroid distance_cluster_2: 66324.682
Davies Bouldin: 0.530
```

Selain DBI, evaluasi kualitas klaster juga ditinjau melalui *Performance Vector* yang ditampilkan pada gambar kedua. Komponen yang dianalisis antara lain **average within centroid distance** secara keseluruhan serta jarak rata-rata tiap klaster. Nilai-nilai yang diperoleh adalah sebagai berikut:

- Avg. within centroid distance: **24332.934**
- Avg. within centroid distance cluster_0: **19683.481**
- Avg. within centroid distance cluster_1: **23498.532**
- Avg. within centroid distance cluster_2: **66324.682**

Nilai-nilai tersebut menunjukkan seberapa jauh rata-rata data di dalam masing-masing klaster dari pusat klaster (centroid). Semakin kecil nilai jaraknya, semakin tinggi tingkat kekompakan klaster tersebut.
Dari hasil tersebut terlihat bahwa:

- **cluster_0 dan cluster_1** memiliki jarak rata-rata yang rendah (19 ribu–23 ribu), menandakan tingkat kekompakan klaster yang cukup baik.
- **cluster_2** memiliki jarak rata-rata paling tinggi yaitu **66324.682**, jauh lebih besar dibandingkan dua klaster lainnya. Hal ini menunjukkan bahwa cluster_2 memiliki sebaran data yang lebih luas serta keragaman nilai atribut yang lebih tinggi.

**Distribusi Klaster dalam Dataset**

Open in Turbo Prep Auto Model Filter (214 / 214 examples): all

| Row No. | id | cluster | id number | Ri | Na | Mg | Al | Si |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | cluster_0 | 1 | 152101 | 13.640 | 4.490 | 1.100 | 71.780 |
| 2 | 2 | cluster_1 | 2 | 151761 | 13.890 | 3.600 | 1.360 | 72.730 |
| 3 | 3 | cluster_1 | 3 | 151618 | 13.530 | 3.550 | 1.540 | 72.990 |
| 4 | 4 | cluster_1 | 4 | 151766 | 13.210 | 3.690 | 1.290 | 72.610 |
| 5 | 5 | cluster_1 | 5 | 151742 | 13.270 | 3.620 | 1.240 | 73.080 |
| 6 | 6 | cluster_1 | 6 | 151596 | 12.790 | 3.610 | 1.620 | 72.970 |
| 7 | 7 | cluster_1 | 7 | 151743 | 13.300 | 3.600 | 1.140 | 73.090 |
| 8 | 8 | cluster_1 | 8 | 151756 | 13.150 | 3.610 | 1.050 | 73.240 |
| 9 | 9 | cluster_1 | 9 | 151918 | 14.040 | 3.580 | 1.370 | 72.080 |
| 10 | 10 | cluster_1 | 10 | 151755 | 13 | 3.600 | 1.360 | 72.990 |
| 11 | 11 | cluster_1 | 11 | 151571 | 12.720 | 3.460 | 1.560 | 73.200 |

id (id) *Integer* Missing: 0 Min: 1 Max: 214 Average: 107.500

Visualisasi dataset yang ditampilkan pada gambar tabel menunjukkan pembagian klaster untuk setiap baris data. Distribusi klaster yang dihasilkan adalah:

- **cluster_1** menjadi klaster dengan jumlah anggota terbanyak, yaitu **163 data**.
- **cluster_0** memiliki jumlah anggota menengah.
- **cluster_2** hanya berisi **8 data**, menjadikannya klaster dengan jumlah anggota paling sedikit.

Ketidakseimbangan distribusi klaster ini menunjukkan bahwa sebagian besar data memiliki karakteristik kimia yang serupa dan cenderung mengumpul pada cluster_1. Sementara itu, cluster_2 berisi data yang memiliki karakteristik sangat berbeda dari mayoritas data lainnya sehingga menyebabkan jarak centroid yang besar dan jumlah anggota yang sedikit.

## Pembahasan

### Interpretasi Nilai Davies–Bouldin Index

Interpretasi nilai Davies–Bouldin Index (DBI) sebesar 0,530 menunjukkan bahwa kualitas struktur klaster yang terbentuk berada pada kategori baik, terutama dalam konteks data dengan keragaman komposisi seperti Glass Dataset. Nilai DBI yang rendah mengindikasikan keseimbangan optimal antara pemisahan antar-klaster dan kekompakan dalam-klaster. Semakin rendah DBI, semakin baik kualitas klaster karena jarak antar-klaster semakin besar dan penyebaran data dalam setiap klaster semakin kecil. Pada penelitian ini, nilai 0,530 mencerminkan bahwa klaster yang terbentuk memiliki struktur yang terdefinisi dengan baik, dengan setiap kelompok menunjukkan identitas kimia yang cukup berbeda namun tetap stabil dalam ruang multidimensi. Secara teoritis, DBI mengukur rasio antara jarak pemisahan antar-klaster dengan tingkat dispersi internal klaster; sehingga apabila nilainya tinggi, klaster dianggap terlalu berdekatan atau memiliki penyebaran yang terlalu luas. Dalam konteks penelitian ini, nilai DBI yang rendah mempertegas bahwa pendekatan K-Means telah menghasilkan klaster dengan batasan yang jelas dan homogenitas internal yang kuat, sehingga dapat dikatakan bahwa model bekerja secara efektif dalam mengelompokkan sampel kaca berdasarkan atribut kimia.

### Analisis Kekompakan dan Penyebaran Klaster

Hasil Performance Vector memberikan gambaran bahwa terdapat perbedaan mencolok dalam tingkat kekompakan dan penyebaran antar-klaster. Klaster 0 dan klaster 1 menunjukkan nilai *within centroid distance* yang relatif rendah, mengindikasikan bahwa anggota dalam kedua klaster tersebut memiliki kemiripan atribut yang kuat dan berada dalam jarak Euclidean yang dekat satu sama lain. Hal ini mencerminkan tingkat homogenitas tinggi dalam klaster tersebut. Sebaliknya, klaster 2 menunjukkan nilai jarak intrasentroid yang sangat besar, yaitu 66.324, yang menandakan bahwa penyebaran data dalam klaster ini sangat luas dan tidak terfokus. Kondisi ini memperlihatkan bahwa klaster 2 kemungkinan berisi data-data ekstrem atau sampel yang memiliki sifat kimia jauh berbeda dari mayoritas data. Variasi besar tersebut juga menunjukkan bahwa klaster ini memiliki karakteristik unik yang mungkin berasal dari jenis kaca khusus atau sampel dengan karakteristik manufaktur yang berbeda. Dalam konteks data berbasis komposisi kimia, perbedaan kecil pada atribut seperti Na, Mg, atau Al dapat menyebabkan pembentukan kelompok yang sangat berbeda, sehingga penyebaran besar pada klaster tertentu merupakan fenomena yang wajar.

### Analisis Karakteristik Atribut Tiap Klaster

Berdasarkan evaluasi nilai atribut pada setiap klaster, terlihat bahwa variabel seperti Ri, Na, Mg, Al, dan Si memiliki pengaruh signifikan dalam proses pembentukan klaster pada Glass Dataset. Klaster 1 memiliki nilai atribut yang cenderung mendekati nilai rata-rata dataset secara keseluruhan, sehingga klaster ini dapat dianggap sebagai representasi umum dari sampel kaca dengan komposisi standar. Klaster 0 menunjukkan sejumlah penyimpangan ringan di mana beberapa atribut berada sedikit di atas atau di bawah rata-rata, sehingga menciptakan kelompok dengan karakteristik kimia yang berbeda namun masih berada dalam kisaran wajar. Klaster 2 menjadi kelompok yang paling mencolok karena memiliki nilai atribut kimia yang sangat ekstrem, baik jauh lebih tinggi maupun jauh lebih rendah dibanding klaster lain. Kondisi tersebut menyebabkan centroid klaster 2 berada pada jarak yang sangat berbeda dan menjadi alasan utama

tingginya nilai *within centroid distance*. Keunikan klaster 2 mengindikasikan bahwa kelompok ini mungkin menggambarkan sampel kaca langka atau sampel yang diproduksi dengan metode tertentu yang tidak umum, sehingga memiliki komposisi kimia yang lebih radikal dibanding kaca generik.

**Hubungan Antar Klaster dan Implikasi Penelitian**

Struktur klaster yang terbentuk dari hasil pengolahan Glass Dataset memiliki beberapa implikasi penting bagi pengembangan penelitian selanjutnya. Pertama, perbedaan karakteristik antar-klaster memungkinkan proses identifikasi jenis kaca berdasarkan komposisi kimia menjadi lebih sistematis. Dengan mengetahui klaster mana yang paling mendekati nilai atribut sampel baru, identifikasi awal dapat dilakukan dengan lebih akurat. Kedua, keberadaan klaster dengan jumlah anggota yang sangat sedikit dan memiliki penyebaran nilai atribut yang ekstrem, seperti klaster 2, mengindikasikan adanya kelompok data yang langka atau outlier. Hal ini penting bagi bidang kriminalistik, forensik material, dan ilmu material (material science) dalam mendeteksi jenis kaca yang tidak umum. Ketiga, nilai DBI yang rendah memberikan bukti bahwa model klasterisasi yang digunakan tidak hanya stabil tetapi juga valid untuk digunakan sebagai dasar pengembangan metode klasifikasi atau pengelompokan lanjutan. Oleh karena itu, model K-Means yang diterapkan dalam penelitian ini memiliki potensi kuat untuk digunakan dalam studi komposisi kaca, analisis manufaktur, hingga sistem prediksi berbasis pola kimia.

**Kesimpulan Pembahasan**

Berdasarkan hasil analisis yang diperoleh, dapat disimpulkan bahwa model klasterisasi yang dibangun menggunakan algoritma K-Means mampu menghasilkan struktur klaster yang berkualitas baik, ditandai dengan nilai Davies–Bouldin Index sebesar 0,530 yang menunjukkan pemisahan dan kekompakan klaster yang optimal. Klaster yang terbentuk menunjukkan distribusi data yang tidak merata dengan klaster 1 sebagai kelompok terbesar yang mewakili komposisi kimia standar, klaster 0 sebagai kelompok dengan penyimpangan ringan, dan klaster 2 sebagai kelompok dengan karakteristik ekstrem yang mengindikasikan keberadaan data unik atau langka. Variasi nilai atribut kimia menjadi faktor utama dalam pembentukan struktur klaster ini. Secara keseluruhan, model klasterisasi yang dihasilkan memberikan gambaran mendalam mengenai pola komposisi kimia kaca serta dapat digunakan secara efektif untuk analisis identifikasi, eksplorasi pola, maupun pengembangan model klasifikasi lanjutan berdasarkan karakteristik kimia.

## KESIMPULAN

Berdasarkan rangkaian proses penelitian yang melibatkan analisis permasalahan, perancangan metodologi, penerapan algoritma K-Means, evaluasi metrik validitas internal, serta interpretasi hasil pengelompokan pada Glass Dataset, dapat disimpulkan bahwa algoritma K-Means mampu memberikan struktur klaster yang efektif dan representatif dalam memetakan variasi komposisi kimia sampel kaca. Penggunaan preprocessing seperti normalisasi serta pemilihan nilai *K* optimal melalui indeks Elbow, Silhouette, dan Davies–Bouldin terbukti berkontribusi signifikan dalam meningkatkan kualitas klaster yang dihasilkan. Nilai Davies–Bouldin Index sebesar 0,530 mengindikasikan bahwa klaster yang terbentuk memiliki tingkat pemisahan yang jelas dan konsistensi internal yang memadai, sehingga struktur klaster dapat dinyatakan stabil dan dapat dipertanggungjawabkan secara metodologis.

Hasil klasterisasi menunjukkan adanya tiga kelompok utama dengan karakteristik kimia yang berbeda secara substansial, di mana satu klaster merepresentasikan komposisi kaca umum, klaster lain menunjukkan variasi komposisi yang lebih spesifik, dan satu klaster lagi menampilkan nilai atribut ekstrem yang memungkinkan keberadaan sampel langka atau outlier. Perbedaan distribusi dan karakteristik ini memberikan wawasan mendalam terkait pola komposisi kimia kaca dan menegaskan bahwa pengelompokan berbasis K-Means dapat digunakan sebagai pendekatan valid untuk mengidentifikasi

kategori material, mengamati fenomena kimia, atau mendukung proses analisis dalam konteks forensik dan penelitian bahan.

Secara keseluruhan, penelitian ini membuktikan bahwa pemanfaatan algoritma K-Means dalam mengelompokkan Glass Dataset secara efektif tidak hanya memberikan struktur pengelompokan yang jelas, tetapi juga dapat memperkaya proses analisis material berbasis data. Model yang dibangun bersifat adaptif dan dapat dikembangkan lebih lanjut untuk aplikasi prediksi, deteksi outlier, sistem identifikasi otomatis, maupun integrasi dalam sistem pakar atau machine learning yang memerlukan klasifikasi berbasis komposisi kimia. Dengan demikian, penelitian ini berkontribusi pada pemahaman yang lebih komprehensif mengenai karakteristik kaca dan sekaligus membuka peluang untuk peningkatan performa metode pengelompokan pada studi-studi berbasis data kimia di masa mendatang.

## DAFTAR PUSTAKA

Chen, Y., & Ji, K. (2023). *Glass Classification Method Based on K-Means ++ Algorithm*. *41*, 331–339.

Dong, K. (2023). *Composition analysis and identification of ancient glass based on K-Means clustering*. *42*, 346–355. https://doi.org/10.54097/hset.v42i.7114

Gao, M., Tang, Y., Liu, H., & Ma, R. (2023). Statistics of fingerprint minutiae frequency and distribution based on automatic minutiae detection method. *Forensic Science International*, *344*, 111572. https://doi.org/https://doi.org/10.1016/j.forsciint.2023.111572

Ji, Z. (2023). *regression model and K-means clustering analysis*. *40*. https://doi.org/10.54097/hset.v40i.6526

Kong, Y., Liu, Z., Tan, W., & Wu, W. (2023). *A study of glass classification and identification based on K- means clustering and Euclidean distance*. *58*. https://doi.org/10.54097/hset.v58i.10031

Loh, C., Chen, Y., Su, C., & Lin, S. (2024). *applied sciences Multi-Objective Decision Support for Irrigation Systems Based on Skyline Query*.

Lv, Y., Tan, H., Meng, F., Shao, G., & Yang, S. (2023). *Glass Classification and Identification Based on K-Means Clustering Model*. *42*, 28–37. https://doi.org/10.54097/hset.v42i.7057

Meng, J., Yu, Z., & Cai, Y. (2023). *applied sciences K-Means ++ Clustering Algorithm in Categorization of Glass Cultural Relics*. https://doi.org/10.3390/app13084736

The many faces of ceramics: New developments in the study of archaeological ceramics from the 15th European meeting on ancient ceramics (Barcelona, 16th–18th september 2019). (2022). *Journal of Archaeological Science: Reports*, *44*, 103550. https://doi.org/https://doi.org/10.1016/j.jasrep.2022.103550

Wang, Z., & Shen, W. (2023). *Analysis and identification of the compositional correlations of glasses and the variability of different classes of compositions*. *4*(5), 25–30. https://doi.org/10.25236/AJMC.2023.040504

Yan, S., Zhang, K., & Sun, J. (2023). *Classification of Glass Products Based on Clustering Algorithm*. *43*. https://doi.org/10.54097/hset.v43i.7445

Yang, J., Wang, Y., Yao, X., & Lin, C. (2021). *Adaptive Initialization Method for K-Means Algorithm*. *4*(November), 1–13. https://doi.org/10.3389/frai.2021.740817

Yin, H. (2022). *Identification model of surface composition of glass relics based on K-means clustering and boosting learning strategy*. *22*, 340–346. https://doi.org/10.54097/hset.v22i.3401

Zhang, J. (2022). *Identification and Analysis of Glass Components by Fusing K-Means Clustering and Ridge Regression*. *5*(13), 30–37. https://doi.org/10.25236/AJCIS.2022.051305

Zhang, Y., Guo, J., & Deng, Q. (2023). *Identification of ancient glass products based on K- means composition analysis*. *4*(3), 39–46. https://doi.org/10.25236/AJMC.2023.040306