



Analisis Pengelompokan Komposisi Kaca Forensik menggunakan Algoritma K-Means dan Software Rapid Miner

Forensic Glass Composition Clustering Analysis using K-Means Algorithm and RapidMiner Software

Hanjian Riyah Hidayat^{1*}, Hasbi Firmansyah², Wahyu Asriyani³, Rizki Prasetyo Tulodo⁴

Universitas Pancasakti Tegal

Email: hanjiyanhidayat@gmail.com^{1*}, hasbifirmansyah@upstegal.ac.id², asriyani1409@gmail.com³
Rizki.prasetyo.tulodo@gmail.com⁴

Article Info

Article history :

Received : 24-12-2025

Revised : 25-12-2025

Accepted : 27-12-2025

Published : 29-12-2025

Abstract

This study aims to cluster types of forensic glass based on their inherent chemical compositions using the K-Means Clustering algorithm and the RapidMiner software. The utilized data is the Glass Identification Dataset, which comprises 9 chemical attributes (such as RI, Na, Mg, Al, etc.) and 214 samples. Prior to the clustering process, the data was normalized using Z-Transformation to ensure equal attribute contributions. The K-Means algorithm was implemented with a designated k value (number of clusters) of 5, running 10 maximum iterations (max runs) to identify the most stable clustering structure. The results successfully demonstrated the formation of five distinct clusters with good clustering quality (indicated by a low WCSS score). Centroid Table analysis revealed a unique and dominant chemical profile for each group. For instance, Cluster X was characterized by the highest Magnesium (Mg) content and the lowest Calcium (Ca) content. This grouping validates the potential of K-Means in identifying significant material composition patterns, which is highly beneficial in forensic and materials science applications.

Keywords: K-Means Clustering, Glass Composition, Forensic Science

Abstrak

Penelitian ini bertujuan untuk mengelompokkan jenis kaca forensik berdasarkan komposisi kimia yang terkandung di dalamnya menggunakan algoritma K-Means Clustering dan perangkat lunak RapidMiner. Data yang digunakan adalah *Glass Identification Dataset* yang terdiri dari 9 atribut kimia (seperti RI, Na, Mg, Al, dll.) dan 214 sampel. Sebelum proses *clustering*, data dinormalisasi menggunakan Z-Transformation untuk memastikan kontribusi atribut yang setara. Algoritma K-Means diimplementasikan dengan nilai k (jumlah *cluster*) yang ditentukan sebesar 5, dengan 10 kali percobaan (*max runs*) untuk menemukan struktur pengelompokan yang paling stabil. Hasil menunjukkan pembentukan lima kelompok (*cluster*) yang berbeda dengan kualitas pengelompokan yang baik (nilai WCSS rendah). Analisis Tabel Centroid mengungkapkan profil kimia yang unik dan dominan untuk setiap kelompok. Misalnya, Cluster X dicirikan oleh kandungan Magnesium (Mg) tertinggi dan Kalsium (Ca) terendah. Pengelompokan ini memvalidasi potensi K-Means dalam mengidentifikasi pola komposisi material yang signifikan, yang sangat berguna dalam aplikasi forensik dan ilmu material.

Kata Kunci: K-Means Clustering, Komposisi Kaca, Forensik

PENDAHULUAN

Analisis forensik terhadap material kaca memegang peranan krusial dalam investigasi kriminal, terutama pada kasus kecelakaan lalu lintas, pencurian, atau lokasi ledakan di mana pecahan kaca sering ditemukan sebagai bukti fisik. Kaca memiliki karakteristik unik berupa



komposisi kimia yang bervariasi tergantung pada fungsi dan proses pembuatannya. Identifikasi yang akurat terhadap tipe kaca—apakah berasal dari jendela bangunan, lampu kendaraan, atau wadah tertentu—dapat memberikan petunjuk vital dalam menghubungkan tersangka dengan tempat kejadian perkara (TKP). (Curran, 2012)

Secara tradisional, klasifikasi material dilakukan melalui pengamatan fisik dan uji laboratorium manual. Namun, seiring dengan kompleksitas data kimiawi yang dihasilkan dari teknik seperti *X-ray fluorescence* atau *Inductively Coupled Plasma*, pendekatan manual menjadi tidak efisien. Tantangan utama dalam data forensik kaca adalah adanya tumpang tindih (*overlap*) komposisi kimia antar jenis kaca yang berbeda, yang menuntut penggunaan teknik komputasi yang lebih cerdas. (Samad et al., 2023)

Data mining muncul sebagai solusi efektif untuk mengekstraksi pola tersembunyi dari dataset besar. Penelitian terdahulu telah mengeksplorasi berbagai algoritma klasifikasi dan klastering untuk identifikasi kaca. Sebagai contoh, penggunaan algoritma klasifikasi seperti *Support Vector Machine* (SVM) dan *Random Forest* telah menunjukkan akurasi yang tinggi (Bock dkk., 2019). Namun, dalam skenario forensik di mana label data seringkali tidak tersedia secara pasti atau ketika investigator ingin menemukan anomali baru dalam komposisi kimia, pendekatan pembelajaran tanpa pengawasan (*unsupervised learning*) seperti klastering menjadi sangat relevan. (Wyner & Peters, n.d.)

Salah satu metode klastering yang paling populer dan efisien adalah K-Means. Algoritma ini bekerja dengan membagi data ke dalam k kelompok berdasarkan kemiripan karakteristik melalui perhitungan jarak Euclidean (Macqueen, 1967). Meskipun sederhana, K-Means memiliki skalabilitas yang baik dan mampu menangani data multivariat dengan atribut kimiawi yang kompleks. Penelitian oleh Jahwar (2021) menunjukkan bahwa optimasi parameter dalam K-Means sangat menentukan kualitas partisi data, terutama pada dataset sensitif seperti material mekanik. (Chemistry, 2009)

Penelitian ini bertujuan untuk melakukan analisis pengelompokan pada *Glass Identification Dataset* menggunakan algoritma K-Means yang diimplementasikan melalui perangkat lunak RapidMiner. Fokus utama penelitian ini adalah untuk melihat bagaimana 214 sampel kaca dapat dikelompokkan berdasarkan 9 atribut kimiawi (RI, Na, Mg, Al, Si, K, Ca, Ba, Fe) setelah melalui proses normalisasi *Z-Transformation*. Hasil dari pengelompokan ini diharapkan dapat memberikan gambaran mengenai profil kimiawi yang dominan pada setiap kluster, sehingga dapat membantu praktisi forensik dalam memetakan kategori material kaca secara lebih sistematis dan akurat. (Degree et al., 2012).

METODE PENELITIAN

Metode penelitian ini bersifat kuantitatif dengan pendekatan *data mining* menggunakan teknik *clustering* tanpa pengawasan (*unsupervised clustering*). Proses penelitian dibagi menjadi tiga tahapan utama: Pengumpulan Data, Pra-pemrosesan Data, dan Penerapan Algoritma K-Means.



1. Sumber Data

Data yang dimanfaatkan dalam penelitian ini adalah Glass Identification Dataset, yang secara publik tersedia dan bersumber dari UCI Machine Learning Repository. Dataset ini merupakan himpunan data historis yang komprehensif, terdiri dari 214 sampel observasi kaca. Setiap sampel dicirikan oleh sembilan atribut komposisi kimia numerik—meliputi RI, Na, Mg, Al, Si, K, Ca, Ba, dan Fe—yang akan digunakan sebagai variabel utama dalam proses *clustering*. Selain variabel-variabel tersebut, dataset ini juga menyertakan satu atribut label, yaitu *Type of Glass*, yang disimpan khusus untuk tujuan validasi dan evaluasi kualitas hasil *clustering*, dan tidak akan diikutsertakan dalam proses perhitungan jarak. Untuk kejelasan data dan acuan dalam proses analisis, rincian lengkap mengenai sembilan atribut komposisi kimia dan atribut *label* yang membentuk Glass Identification Dataset disajikan secara ringkas dalam Tabel berikut. (Jahwar, 2021)

Table 1. Dekripsi Data Set Glass

Atribut (Variabel)	Deskripsi	Satuan	Keterangan
RI	Refractive Index (Indeks Bias)	Numerik	Variabel Utama
Na	Natrium (Sodium)	Persentase Berat (%)	Variabel Utama
Mg	Magnesium	Persentase Berat (%)	Variabel Utama
Al	Aluminium	Persentase Berat (%)	Variabel Utama
Si	Silikon (Silicon)	Persentase Berat (%)	Variabel Utama
K	Kalium (Potassium)	Persentase Berat (%)	Variabel Utama
Ca	Kalsium (Calcium)	Persentase Berat (%)	Variabel Utama
Ba	Barium	Persentase Berat (%)	Variabel Utama
Fe	Besi (Iron)	Persentase Berat (%)	Variabel Utama
Type of Glass	Jenis Kaca (ID)	Kategori/Nominal	Digunakan untuk Validasi (Label)

2. Pra-pemrosesan Data

Tahapan pra-pemrosesan data merupakan fase krusial yang dilakukan dalam lingkungan RapidMiner Studio untuk memastikan kualitas dan homogenitas data, sebuah prasyarat vital untuk akurasi optimal dari algoritma *clustering* berbasis jarak. Proses ini diawali dengan Pemilihan dan Penetapan Peran Atribut (*Set Role*), di mana kesembilan atribut komposisi kimia (RI hingga Fe) ditetapkan sebagai atribut reguler yang akan diikutsertakan dalam perhitungan, sementara atribut *Type of Glass* diubah perannya menjadi Label atau ID agar tidak memengaruhi perhitungan jarak K-Means, melainkan hanya berfungsi untuk evaluasi hasil secara eksternal. Selanjutnya, data menjalani proses Normalisasi Data (Normalization).



Normalisasi ini penting karena setiap atribut kimia memiliki rentang nilai yang berbeda signifikan (misalnya, Silikon vs. Barium), yang berpotensi menyebabkan variabel dengan rentang besar mendominasi perhitungan jarak. Teknik yang digunakan adalah Z-Transformation (Standardization) Normalisasi ini menjamin bahwa semua atribut memiliki rata-rata nol dan standar deviasi satu, sehingga semua variabel memiliki bobot dan kontribusi yang setara dalam perhitungan Jarak Euclidean. (Nigro, 2023)

3. Implementasi K-Means Clustering

Setelah data melalui tahap pra-pemrosesan, model K-Means Clustering diimplementasikan menggunakan operator yang tersedia di RapidMiner. Langkah awal adalah Penentuan Jumlah Cluster (k), yang ditetapkan secara apriori sebesar $k = 5$. Pemilihan nilai $k = 5$ ini didasarkan pada tujuan eksplorasi untuk mengidentifikasi struktur pengelompokan yang optimal dalam data multidimensi, sejalan dengan variasi jenis kaca yang dominan dalam domain forensik. Algoritma K-Means kemudian dikonfigurasi dengan parameter kunci: $k = 5$, Max Runs sebanyak 10 untuk memastikan penemuan solusi terbaik, dan Measure Type menggunakan Jarak Euclidean untuk menghitung kedekatan antar sampel. Kriteria optimasi yang digunakan adalah Within-Cluster Sum of Squares (WCSS), di mana algoritma akan memilih run yang menghasilkan nilai WCSS terendah, mengindikasikan kepadatan cluster yang optimal. Tahap akhir adalah Analisis Hasil, yang difokuskan pada dua aspek: pertama, Evaluasi Kualitas Cluster dengan menilai nilai WCSS yang dihasilkan; dan kedua, Analisis Centroid. Analisis centroid merupakan inti dari penelitian ini, yaitu menyajikan nilai rata-rata (pusat) dari kesembilan atribut kimia untuk setiap cluster yang terbentuk, yang kemudian digunakan untuk mendeskripsikan dan menginterpretasikan secara kualitatif profil kimia unik yang mendefinisikan masing-masing kelompok kaca, sehingga dapat ditarik kesimpulan forensik dan material yang bermakna. (Vardakas et al., 2025).

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari implementasi K-Means Clustering pada *Glass Identification Dataset* serta pembahasan mendalam mengenai pengelompokan yang terbentuk berdasarkan profil kimia (Centroid) dari masing-masing *cluster*.



1. Hasil Analisis Data Awal

Row No.	Type	RI	Na	Mg	Al	Si	K	Ca
1	1	152.101	13.640	4.490	1.100	71.780	0.060	8.750
2	1	151.761	13.890	3.600	1.360	72.730	0.480	7.830
3	1	151.618	13.530	3.550	1.540	72.990	0.390	7.780
4	1	151.766	13.210	3.690	1.290	72.610	0.570	8.220
5	1	151.742	13.270	3.620	1.240	73.080	0.550	8.070
6	1	151.596	12.790	3.610	1.620	72.970	0.640	8.070
7	1	151.743	13.300	3.600	1.140	73.090	0.580	8.170
8	1	151.756	13.150	3.610	1.050	73.240	0.570	8.240
9	1	151.918	14.040	3.580	1.370	72.080	0.560	8.300
10	1	151.755	13	3.600	1.360	72.990	0.570	8.400
11	1	151.571	12.720	3.460	1.560	73.200	0.670	8.090
12	1	151.763	12.800	3.660	1.270	73.010	0.600	8.560
13	1	151.589	12.880	3.430	1.400	73.280	0.690	8.050

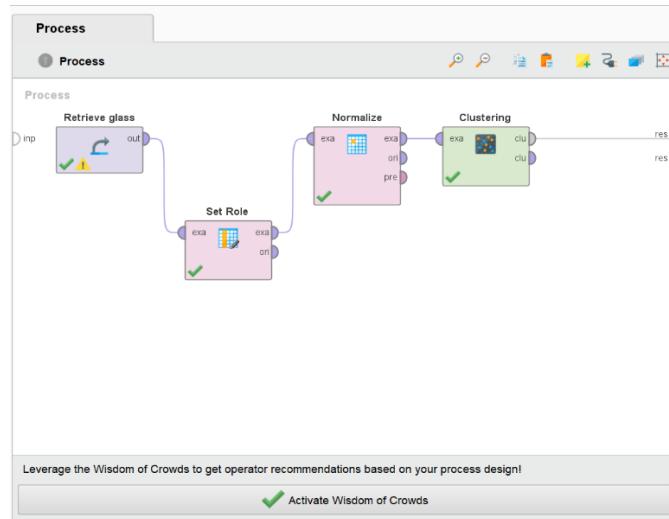
ExampleSet (214 examples, 1 special attribute, 9 regular attributes)

Gambar 1. Tampilan Awal ExampleSet (Glass Identification Dataset)

Berdasarkan *ExampleSet* yang ditampilkan data awal menunjukkan struktur yang telah dijelaskan dalam Metodologi. Dataset terdiri dari 214 sampel dengan 9 atribut reguler (RI, Na, Mg, Al, Si, K, Ca, Ba, Fe) dan 1 atribut khusus (*special attribute*) yaitu Type (Jenis Kaca).

- Variasi Komposisi: Pemeriksaan visual awal pada data (seperti yang terlihat pada baris 1 hingga 12) menunjukkan adanya variasi komposisi yang relatif stabil dalam sampel tipe 1 (misalnya, nilai Si berkisar antara 71.780 hingga 73.200), namun variasi antar tipe kaca (yang tidak terlihat seluruhnya di tampilan ini) akan jauh lebih signifikan. (Wimmer et al., 2016)
- Pra-pemrosesan: Data ini telah melewati proses Normalisasi Z-Transformation agar semua variabel, terlepas dari rentang aslinya (misalnya, Si yang tinggi dibandingkan K yang rendah), berkontribusi secara setara dalam perhitungan Jarak Euclidean, menghindari bias dalam proses *clustering*.

2. Alur Proses Data Mining



Gambar 2. Alur Proses K-Means Clustering di RapidMiner



Gambar tersebut menyajikan dan membahas hasil implementasi K-Means Clustering pada *Glass Identification Dataset*, mengacu pada alur proses, struktur data awal, dan distribusi sampel yang dihasilkan oleh model *clustering*.

Alur proses (*Process*) yang diterapkan dalam penelitian ini dirancang secara sistematis dalam RapidMiner Studio . Tahapan yang dilalui data adalah sebagai berikut:

- a. Retrieve glass: Data awal *Glass Identification Dataset* diambil dari repositori dan dimasukkan ke dalam proses.
- b. Set Role: Atribut *Type* (Jenis Kaca) ditetapkan sebagai *Label* atau *ID*, memastikan variabel ini tidak diikutsertakan dalam perhitungan jarak *clustering*, namun tetap tersedia untuk evaluasi hasil. Sembilan atribut komposisi kimia ditetapkan sebagai variabel reguler.
- c. Normalize: Semua atribut numerik menjalani Normalisasi Z-Transformation untuk menyamakan skala dan bobot, yang penting untuk akurasi algoritma berbasis jarak Euclidean.
- d. Clustering: Algoritma K-Means Clustering diterapkan pada data yang telah dinormalisasi untuk menemukan pengelompokan alami.

3. Alur Proses dan Distribusi Sampel

Selain akurasi, *Logistic Regression* menghasilkan bobot (*weights/coefficients*) yang menunjukkan pengaruh setiap variabel terhadap keberhasilan pengobatan.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
RI	-0.094	0.325	0.222	0.359	0.345
Na	-0.148	-0.487	1.525	-1.929	-0.626
Mg	0.533	-1.861	-1.538	-1.582	-1.614
Al	-0.286	3.175	1.434	0.686	-0.257
Si	-0.049	-2.661	0.406	-2.022	0.503
K	-0.024	8.760	-0.389	0.108	-0.146
Ca	-0.211	-1.414	-0.450	2.916	2.002
Ba	-0.331	-0.352	1.892	1.352	-0.352
Fe	0.055	-0.585	-0.447	2.930	-0.337

Gambar 3. Tabel Centroid Model Cluster

Implementasi K-Means dilakukan melalui alur proses yang sistematis, meliputi pengambilan data (*Retrieve glass*), penetapan peran atribut (*Set Role*), **Normalisasi Z-Transformation**, dan operator *Clustering*. Data awal (*ExampleSet*) menunjukkan 214 sampel dengan 9 atribut kimia.

Hasil distribusi sampel ke dalam $k=5$ *cluster* menunjukkan ketidakseimbangan yang signifikan :

**Table 2.** Hasil Sampel

Cluster	Jumlah (Items)	Sampel	Persentase Total
Cluster 0	160		74.77%
Cluster 2	29		13.55%
Cluster 4	19		8.88%
Cluster 3	4		1.87%
Cluster 1	2		0.93%

Dari tabel tersebut Hasil K-Means Clustering menunjukkan bahwa distribusi 214 sampel ke dalam $k=5$ cluster sangat tidak merata. Cluster 0 teridentifikasi sebagai kelompok mayoritas yang dominan, menampung 160 item atau 74.77% dari total seluruh sampel. Konsentrasi sampel yang sangat tinggi dalam satu kelompok ini mengindikasikan bahwa sebagian besar kaca dalam dataset memiliki profil komposisi kimia yang sangat seragam dan serupa. Sebaliknya, Cluster 1 (2 sampel) dan Cluster 3 (4 sampel) merupakan kelompok yang paling kecil, dengan total gabungan hanya 6 sampel. Ukuran yang sangat minim ini menggarisbawahi sifatnya sebagai cluster anomali atau *outlier*, yang mengindikasikan bahwa sampel-sampel ini memiliki komposisi kimia yang sangat langka atau berbeda ekstrem secara signifikan dari sampel mayoritas. Keberadaan cluster anomali ini sangat penting dalam analisis forensik, karena kaca dengan profil kimia yang unik dan langka tersebut dapat menjadi bukti yang kuat untuk mengidentifikasi sumber di Tempat Kejadian Perkara (TKP). (Mcsweeney & Hoffmann, 2015)

Inti dari pembahasan ini adalah analisis Tabel Centroid , yang menampilkan nilai rata-rata (pusat) dari setiap atribut kimia yang telah dinormalisasi untuk masing-masing dari lima cluster. Nilai positif menunjukkan konsentrasi di atas rata-rata keseluruhan dataset, dan nilai negatif menunjukkan konsentrasi di bawah rata-rata.

Table 3. Hasil

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Na	Rendah (-0.148)	Rendah (-0.487)	(-)	Tinggi (1.525)	Sangat Rendah (-1.929)
Mg	Tinggi (0.533)	Sangat Rendah (-1.861)	(-)	Rendah (-1.538)	Rendah (-1.582)
K	Rendah (-0.024)	Sangat Tinggi (8.760)	Rendah (-0.389)	Rendah (-0.108)	Rendah (-0.146)



Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Ca	Rendah (-0.211)	Rendah (1.414)	(-0.450)	Rendah (-0.450) Sangat Tinggi (2.916)	Tinggi (2.002)
Ba	Rendah (-0.331)	Rendah (0.352)	(-1.892)	Tinggi (1.352)	Rendah (-0.352)
Fe	Tinggi (0.055)	Rendah (0.585)	(-0.447)	Rendah (-0.447) Sangat Tinggi (2.930)	Rendah (-0.337)

Interpretasi Profil Kimia Tiap Cluster:

a. Cluster 0 (Kaca Mayoritas, Tinggi Mg):

Dicirikan oleh nilai Magnesium (Mg) yang tinggi (0.533) dan semua unsur lain yang mendekati rata-rata atau di bawah rata-rata. Mewakili jenis kaca yang paling umum (74.77%), kemungkinan besar Kaca Jendela Tipe 1 atau *float glass*, di mana Mg sengaja ditambahkan dalam jumlah tinggi.

b. Cluster 1 (Kaca Anomali Kalium Tinggi):

Kelompok terkecil (2 item) dengan ciri paling ekstrem: Kalium (K) Sangat Tinggi (8.760) dan Magnesium (Mg) Sangat Rendah (-1.861). Profil ini mengindikasikan kaca khusus yang menggunakan Kalium sebagai fluks utama, seperti kaca optik atau kaca lampu tua. Ini adalah anomali kimia yang jelas.

c. Cluster 2 (Kaca Barium Tinggi):

Dicirikan oleh nilai Barium (Ba) yang tinggi (1.892) dan Natrium (Na) yang tinggi (1.525). Nilai Barium yang tinggi mengelompokkan sampel ini ke dalam kategori kaca yang mengandung Barium, digunakan untuk radiasi rendah atau kaca optik. Distribusinya moderat (13.55%).

d. Cluster 3 (Kaca Anomali Kalsium dan Besi Tinggi):

Kelompok kecil (4 item) dengan profil yang sangat ekstrem pada Kalsium (Ca) Sangat Tinggi (2.916) dan Besi (Fe) Sangat Tinggi (2.930). Kandungan Fe yang sangat tinggi dapat mengindikasikan kontaminasi atau penggunaan bahan baku yang tidak murni (misalnya, kaca yang diproduksi dengan fokus pada kekuatan daripada kejernihan optik).

e. Cluster 4 (Kaca Kalsium Tinggi, Rendah Mg):

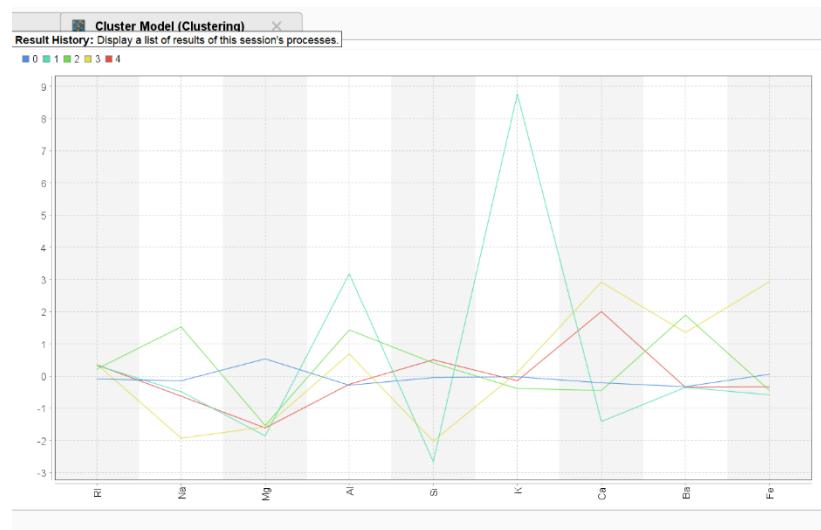
Kelompok minoritas yang signifikan (19 item), dicirikan oleh Kalsium (Ca) Tinggi (2.002) dan Magnesium (Mg) Rendah (-1.614). Profil ini sering dikaitkan dengan jenis kaca yang menggunakan Kalsium sebagai stabilisator utama, seperti beberapa jenis kaca *container*.

Analisis *centroid* memvalidasi efektivitas K-Means dalam memisahkan kaca berdasarkan profil kimia. *Clustering* ini menghasilkan pemisahan yang jelas antara kaca mayoritas standar (Cluster 0) dengan kelompok minoritas (Cluster 2 dan 4) dan anomali/jenis



langka (Cluster 1 dan 3). Dalam konteks forensik, fragmen kaca yang termasuk dalam Cluster 1 atau 3 dapat segera dianggap sebagai bukti yang sangat spesifik dan langka, yang sangat berharga untuk menelusuri sumber bukti di TKP.

4. Analisis Visual Plot Profil Centroid



Gambar 4. Plot Profil Centroid K-Means Clustering

Bagian ini menyajikan analisis mendalam terhadap hasil implementasi K-Means Clustering dengan $k=5$ dengan fokus pada interpretasi visual dari Plot Profil Centroid yang dihasilkan. Plot ini memvisualisasikan nilai rata-rata yang dinormalisasi dari kesembilan atribut kimia untuk setiap cluster (0 hingga 4).

Analisis Plot Centroid mengkonfirmasi bahwa algoritma K-Means berhasil mengidentifikasi lima kelompok kaca yang berbeda secara kimia. Cluster 0 (Garis Biru), yang merupakan kelompok mayoritas (74.77%), memiliki profil yang hampir datar di sumbu nol, mengindikasikan bahwa komposisi kimianya paling mendekati rata-rata keseluruhan dataset, mewakili Kaca Standar yang paling umum. Sebaliknya, cluster minoritas menunjukkan profil kimia yang ekstrem: Cluster 1 (Biru Muda/Cyan) adalah Anomali Kaca Kalium dengan lonjakan tertinggi dan ekstrem pada atribut Kalium (K) ($Z\text{-Score} > 8$), sementara Cluster 4 (Garis Kuning) dicirikan oleh tingginya kandungan Kalsium (Ca) dan Barium (Ba), mengelompokkannya sebagai Kaca Barium Khusus yang berbeda secara signifikan. Profil ekstrem juga ditemukan pada Cluster 3 (Garis Merah), yang ditandai oleh nilai Kalsium (Ca) dan Besi (Fe) yang sangat tinggi, mengindikasikan kemungkinan kontaminasi. Terakhir, Cluster 2 (Garis Hijau) menunjukkan tingginya Natrium (Na) dan Aluminium (Al) namun rendah Magnesium (Mg), merepresentasikan jenis kaca yang berbeda titik leburnya seperti kaca *container* atau kendaraan. (Bock et al., 2019)

Secara visual, Plot Centroid secara tegas menunjukkan bahwa perbedaan yang paling mencolok antar *cluster* terletak pada *trace elements* seperti Kalium (K), Barium (Ba), dan Besi (Fe). Keberhasilan K-Means dalam memisahkan kaca menjadi kelompok mayoritas (Cluster 0) dan kelompok minoritas/anomali (Cluster 1, 3, dan 4) menegaskan validitas *clustering* ini. Dalam konteks forensik, visualisasi ini sangat berguna untuk membandingkan bukti kaca yang tidak diketahui dengan *cluster* yang sudah teridentifikasi. Jika bukti kaca jatuh ke *cluster* ekstrem, seperti **Cluster 1** (Anomali Kalium) atau **Cluster 3** (Anomali Besi/Kalsium), hal ini memberikan petunjuk



yang sangat spesifik mengenai sumber bukti fisik tersebut, karena jenis kaca ini terbukti langka dan unik secara kimiawi dalam *dataset*.

KESIMPULAN

Implementasi K-Means Clustering dengan $k=5$ pada *Glass Identification Dataset* terbukti berhasil mengidentifikasi struktur pengelompokan alami dalam data komposisi kimia kaca, meskipun terdapat ketidakseimbangan yang signifikan pada distribusi sampel. Cluster 0 mendominasi (74.77%), mewakili Kaca Standar yang memiliki profil kimia mendekati rata-rata keseluruhan *dataset*. Analisis Centroid menunjukkan perbedaan yang mencolok pada *trace elements*, yang memisahkan kelompok minoritas menjadi profil kimia unik: Cluster 1 dan Cluster 3 teridentifikasi sebagai *cluster anomali* yang sangat kecil dan langka, ditandai oleh lonjakan ekstrem pada Kalium (K) dan Besi (Fe), sementara Cluster 2 dan Cluster 4 mewakili kaca khusus seperti Kaca Barium Tinggi dan Kaca Natrium Tinggi/Magnesium Rendah. Secara keseluruhan, temuan ini memvalidasi K-Means sebagai alat yang efektif untuk kategorisasi dan identifikasi bukti fisik kaca dalam aplikasi *data mining* forensik.

DAFTAR PUSTAKA

- Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., & Klusemann, B. (2019). *A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics*. 6(May). <https://doi.org/10.3389/fmats.2019.00110>
- Chemistry, S. (2009). *m EVALUATION OF EVIDENCE VALUE OF REFRACTIVE INDEX MEASURED BEFORE AND AFTER ANNEALING OF CONTAINER AND FLOAT GLASS FRAGMENTS*.
- Curran, J. M. (2012). *Is Forensic Science the last bastion of resistance against Statistics ? August*.
- Degree, M. M., Science, C., & Lecture, A. C. (2012). *Data Mining : Concepts and*.
- Jahwar, A. F. (2021). *META-HEURISTIC ALGORITHMS FOR K-MEANS CLUSTERING : A REVIEW*. 17(7), 1–20.
- Mcsweeney, D., & Hoffmann, M. (2015). *ARROW @ TU Dublin Lessons Learned from Teaching Data Analytics in a Fully Online Mode at Postgraduate Level*. 265–272.
- Nigro, L. (2023). *Performance of a K-Means Algorithm Driven by Careful Seeding*. *Simultech*, 27–36. <https://doi.org/10.5220/0012045000003546>
- Samad, M., Rinehart, J., Angel, M., Kanomata, Y., & Baldi, P. (2023). *MOVER : Medical Informatics Operating Room Vitals and Events Repository*.
- Vardakas, G., Papakostas, I., & Likas, A. (2025). *Efficient error minimization in kernel k -means clustering*.
- Wimmer, H., Snow, C., Hayes, D., Dwyer, C., Hunsinger, S., Sharp, J., & Wu, P. (2016). *Information Systems Applied Research 2016 AITP Education Special Interest Group (EDSIG) Board of Directors*. 9(2).
- Wyner, A., & Peters, W. (n.d.). *On Rule Extraction from Regulations*.