https://jurnal.pustakagalerimandiri.co.id/index.php/pustakadata
DOI : https://doi.org/10.55382/jurnalpustakadata.v5i1.1001

Jurnal Pustaka Data
Pusat Akses Kajian Database, Analisa Teknologi, dan Arsitektur Komputer
Pustaka Galeri Mandiri
Vol. 5 No. 1 (2025) 33 – 38
E ISSN: 2809-7483

Analisis Klaster Pasien Menggunakan K-Means untuk Mendukung Perawatan Medis Terpersonalisasi

Hery Oktafiandi[1], Winarnie[2], Wahid Nur Rohman[3]
[1,2]Sistem Informasi, Fakultas Teknologi Kreatif, Universitas SATU Bandung, Indonesia
[3]Teknik Rekaya Perangkat Lunak, Politeknik Sawunggalih Aji, Purworejo, Jawa Tengah
[1]winarnie@univ.satu.ac.id, [2]hery.oktafiandy@univ.satu.ac.id, [3]noerrohmanwachid@gmail.com

Abstract

This study aims to perform clustering analysis on a patient dataset using the K-Means algorithm. The dataset used contains 6,000 rows of patient data with 16 features, including age, gender, blood pressure, cholesterol, and smoking status. To find the optimal number of clusters, the Elbow method was used, which showed that the most appropriate number of clusters was 3. After that, the K-Means algorithm was applied to cluster the patient data based on the similarity of their health characteristics. The clustering results show that patients can be divided into three main groups: a group with low blood pressure and cholesterol, a group with high blood sugar levels, and a group with hypertension and obesity. These results can be used to provide further insights into grouping patients for more personalized care. This study demonstrates how clustering techniques can be used to analyze health data and aid in medical decision-making.

Keywords: clustering analysis, k-means algorithm, elbow method, health data analysis.

Abstrak

Penelitian ini bertujuan untuk melakukan analisis klasterisasi pada dataset pasien menggunakan algoritma *K-Means*. Dataset yang digunakan berisi 6.000 baris data pasien dengan 16 fitur, meliputi usia, jenis kelamin, tekanan darah, kolesterol, dan status merokok. Untuk mencari jumlah klaster yang optimal, digunakan metode Elbow yang menunjukkan jumlah klaster yang paling tepat adalah 3. Setelah itu, algoritma *K-Means* diaplikasikan untuk mengklasifikasi data pasien berdasarkan kesamaan karakteristik kesehatan mereka. Hasil klasterisasi menunjukkan bahwa pasien dapat dibagi menjadi tiga kelompok utama: kelompok dengan tekanan darah dan kolesterol rendah, kelompok dengan kadar gula darah tinggi, dan kelompok dengan hipertensi dan obesitas. Hasil ini dapat digunakan untuk memberikan wawasan lebih lanjut dalam mengelompokkan pasien untuk perawatan yang lebih personal. Penelitian ini menunjukkan bagaimana teknik klasterisasi dapat digunakan untuk menganalisis data kesehatan dan membantu dalam pengambilan keputusan medis.

Kata kunci: analisis pengelompokkan, algoritma k-means, metode elbow, analisa kesehatan

1. Pendahuluan

Di Indonesia, prevalensi penyakit jantung dan faktor risikonya terus meningkat, yang membuat upaya pencegahan dan deteksi dini semakin penting. Penting bagi tenaga medis dapat mengidentifikasi individu yang berisiko tinggi secara lebih dini, sehingga dapat diberikan perawatan yang lebih terpersonalisasi dan tepat sasaran [1]. Seiring dengan perkembangan teknologi informasi [2], jumlah dan

Submitted : 04-06-2024 | Reviewed : 08-06-2025 | Accepted : 10-06-2025

kompleksitas data kesehatan yang tersedia juga semakin besar. Data medis pasien sering kali terdiri dari banyak fitur yang saling berhubungan, yang dapat mencakup aspek demografis, riwayat medis, serta gaya hidup. Analisis data ini menggunakan metode statistik dan teknik pembelajaran mesin dapat membantu dalam menemukan pola-pola kesehatan yang tidak selalu terlihat dengan analisis konvensional. Salah satu metode analisa yang banyak dipakai dalam konteks ini yaitu *clustering*, yang bertujuan untuk mengelompokkan data yang memiliki kesamaan karakteristik[3].

Pada penelitian ini, algoritma *K-Means clustering* digunakan untuk mengelompokkan [4] pasien berdasarkan kesamaan pola kesehatan mereka, yang diwakili oleh berbagai fitur, seperti usia, jenis kelamin, tekanan darah, kolesterol, kadar glukosa, dan kebiasaan merokok[5]. Menggunakan metode ini, dapat dihasilkan kluster-kluster yang berisi pasien-pasien dengan kondisi kesehatan serupa, yang nantinya dapat digunakan untuk memberikan perawatan medis yang lebih efisien dan sesuai dengan kondisi masing-masing pasien[6].

2. Metode Penelitian

Percobaan memakai dataset yang terdiri dari 6.000 baris data pasien dengan 16 fitur, yang mencakup faktor demografis dan kesehatan seperti usia, jenis kelamin, tekanan darah, kolesterol, kadar glukosa plasma, indeks massa tubuh (BMI), dan status merokok. Langkah - langkah yang dipakai pada percobaan yaitu [7]:

1. Pengumpulan Data: Dataset [8] yang digunakan diperoleh dari Kaggle dan berisi data pasien dengan berbagai fitur kesehatan, seperti terlihat gambar 1

```
import pandas as pd
df = pd.read_csv('/content/sample_data/patient_dataset.csv')
df.head()
```

	age	gender	chest_pain_type	blood_pressure	cholesterol	max_heart_rate	exer
0	24	1.0	4	250	139	212	
1	29	0.0	4	132	187	147	
2	46	0.0	3	271	185	193	
3	73	NaN	2	102	200	125	
4	49	1.0	3	91	163	192	

Gambar 1. Dataset

2. Pembersihan dan Transformasi Data [9]: Tahap ini adalah menangani nilai yang hilang (*missing values*) dan mengubah kolom kategorikal menjadi format numerik menggunakan *One-Hot Encoding*[10], seperti terlihat pada gambar 2

```
# One-Hot Encoding untuk kolom kategorikal
df = pd.get_dummies(df, columns=['gender', 'chest_pain_type', 'residence_type', 'smoking_status'], dro

# Melihat hasil setelah encoding
print(df.head())
```

```
   age  blood_pressure  cholesterol  max_heart_rate  exercise_angina  \
0   24             250          139             212                0
1   29             132          187             147                0
2   46             271          185             193                0
3   73             102          200             125                0
4   49              91          163             192                0

   plasma_glucose  skin_thickness  insulin        bmi  diabetes_pedigree  \
0           108.0            33.0    109.0  37.999303           0.480277
1           202.0            42.0      NaN  25.588346           0.283986
2           149.0            43.0    102.0  37.892029           2.472309
3           105.0            77.0    165.0  18.660241           1.472052
4           162.0            31.0    170.0  12.767980           0.537627

   hypertension  heart_disease  gender_1.0  chest_pain_type_2  \
0             1              1        True              False
1             1              1       False              False
2             1              0       False              False
3             0              1       False               True
4             1              1        True              False

   chest_pain_type_3  chest_pain_type_4  residence_type_Urban  \
0              False               True                 True
1              False               True                 True
2               True              False                False
3              False              False                False
4               True              False                False

   smoking_status_Smoker  smoking_status_Unknown
0                   True                   False
1                  False                    True
2                  False                   False
3                   True                   False
4                   True                   False
```

Gambar 2. Proses Pembersihan dan Transformasi Data

3. Normalisasi Data: Semua fitur dinormalisasi [10] menggunakan *StandardScaler* untuk memastikan skala yang konsisten antar fitur, seperti terlihat pada gambar 3 [11].

```
from sklearn.preprocessing import StandardScaler

# Menstandarisasi data (normalisasi ke distribusi dengan mean 0 dan std dev 1
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df)

# Mengubah hasil scaling menjadi DataFrame agar lebih mudah dibaca
```

Gambar 3. Proses Normalisasi Data

4. *Clustering* dengan *K-Means* [12]: Algoritma *K-Means clustering* diterapkan pada dataset setelah dilakukan normalisasi, dengan jumlah kluster optimal [13] yang ditentukan menggunakan *Elbow Method*. Seperti terlihat pada gambar 4 [14].

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

# Menentukan jumlah kluster optimal dengan Elbow Method
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=42)
    kmeans.fit(df_scaled)
    wcss.append(kmeans.inertia_)
```

Gambar 4. Proses Clustering dengan K-Means

Evaluasi Model [15]: Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur kualitas pemisahan antar kluster[16] .

3. Hasil dan Pembahasan

3.1. Proses Penerapan *K-Means Clustering*

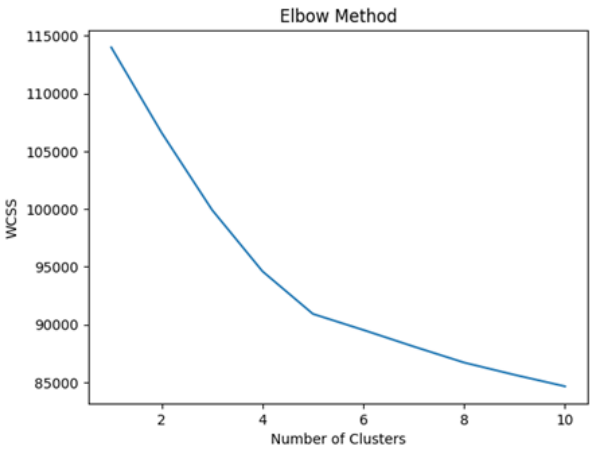
Percobaan ini, dataset yang digunakan terdiri dari 6.000 baris data pasien dengan 16 fitur yang mencakup faktor demografis dan kesehatan pasien, seperti *age* (usia), *gender* (jenis kelamin), *chest_pain_type* (jenis nyeri dada), *blood_pressure* (tekanan darah), *cholesterol* (kolesterol), *max_heart_rate* (detak jantung maksimal), *exercise_angina* (apakah mengalami angina saat berolahraga), *plasma_glucose* (kadar glukosa plasma), *skin_thickness* (ketebalan kulit), *insulin* (kadar insulin), *bmi* (indeks massa tubuh), *diabetes_pedigree* (riwayat keturunan diabetes), *hypertension* (hipertensi), *heart_disease* (penyakit jantung), *residence_type* (jenis tempat tinggal), dan *smoking_status* (status merokok).

Sebelum menerapkan algoritma *K-Means*, langkah-langkah pembersihan dan transformasi data dilakukan. Kolom-kolom kategorikal seperti *chest_pain_type* dan *smoking_status* diubah menjadi format numerik melalui *One-Hot Encoding*, sementara nilai yang hilang diisi dengan median untuk fitur numerik. Setelah itu, data dinormalisasi menggunakan *StandardScaler* untuk memastikan setiap fitur berada pada skala yang sama, yang sangat penting agar algoritma *clustering* dapat berfungsi dengan baik[2].

3.2. Menentukan Jumlah Kluster Optimal dengan Elbow Method

Untuk menentukan jumlah kluster yang optimal, *Elbow Method* digunakan dengan mengukur *Within-Cluster Sum of Squares* **(WCSS)** [16] untuk berbagai jumlah kluster. WCSS mengukur seberapa rapat data dalam setiap kluster terhadap pusat kluster (*centroid*). Semakin kecil nilai WCSS, semakin homogen data dalam kluster[3]

Berdasarkan analisis WCSS, seperti yang ditunjukkan pada tabel 1 didapatkan grafik *Elbow Method* yang menunjukkan penurunan tajam pada awalnya, tetapi setelah 4 kluster, penurunan WCSS mulai melambat. Ini menandakan bahwa penambahan kluster lebih lanjut tidak memberikan pengurangan WCSS yang signifikan. Oleh karena itu, jumlah kluster optimal yang ditemukan adalah 4 seperti yang ditunjukan pada gambar 5 dan tabel 1.



Gambar 5. *Elbow Method WCSS*

Tabel 1.*Tabel WCSS*

No	Jumlah Kluster	WCSS
1	1	114000
2	2	106604.024
3	3	99925.1576
4	4	94608.738
5	5	90922.4692
6	6	89533.4425
7	7	88097.1292
8	8	86708.1303
9	9	85648.1447
10	10	84653.3034

3.3. Karakteristik Kluster yang Ditemukan

Setelah menentukan jumlah kluster optimal, *K-Means* diterapkan dengan jumlah kluster sebanyak 4. Hasil *clustering* menunjukkan bahwa pasien dapat dikelompokkan ke dalam 4 kluster berdasarkan karakteristik kesehatan mereka, dapat dilihat pada gambar 6. Berikut adalah deskripsi singkat mengenai karakteristik masing-masing kluster:

3.3.1. Kluster 0:

Karakteristik: Pasien dalam kluster ini memiliki usia rata-rata sekitar 55-60 tahun dengan tekanan darah dan kolesterol tinggi. Sebagian besar dari mereka mengalami hipertensi dan memiliki BMI yang lebih tinggi, yang mengindikasikan potensi risiko penyakit jantung dan gangguan metabolisme.

Faktor Risiko: Tekanan darah tinggi, kolesterol tinggi, obesitas (BMI tinggi), dan kecenderungan merokok.

Implicasi: Kluster ini menunjukkan pasien dengan risiko tinggi terhadap penyakit jantung, hipertensi, dan diabetes.

3.3.2. Kluster 1:
Karakteristik: Kluster ini terdiri dari pasien dengan usia lebih muda, sekitar 40-45 tahun, dengan tekanan darah normal dan kadar kolesterol yang relatif rendah. Namun, mereka cenderung memiliki kadar glukosa plasma yang lebih tinggi dan BMI yang sedikit lebih tinggi.

Faktor Risiko: Risiko diabetes (glukosa plasma tinggi) dan obesitas (BMI tinggi).

Implicasi: Kluster ini berpotensi memiliki masalah terkait dengan diabetes atau gangguan metabolisme, meskipun tidak menunjukkan hipertensi atau penyakit jantung yang jelas.

3.3.3. Kluster 2:
Karakteristik: Pasien dalam kluster ini memiliki usia yang lebih muda dengan tekanan darah dan kolesterol yang normal, serta kadar glukosa plasma yang normal. Mereka cenderung memiliki kebiasaan merokok rendah dan BMI yang lebih sehat.

Faktor Risiko: Tidak ada faktor risiko besar yang terdeteksi. Kluster ini menunjukkan pasien dengan gaya hidup sehat dan kondisi kesehatan yang relatif baik.

Implicasi: Kluster ini menunjukkan pasien dengan kondisi kesehatan optimal, dengan risiko rendah terhadap penyakit jantung, diabetes, dan hipertensi.

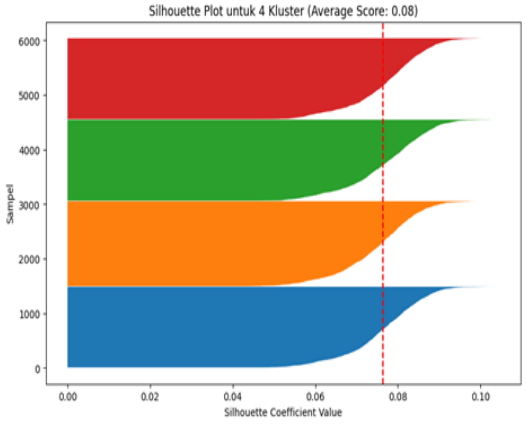
3.3.4. Kluster 3:
Karakteristik: Pasien dalam kluster ini memiliki usia yang lebih tua, sekitar 50-55 tahun, dengan tekanan darah yang cenderung tinggi dan kadar kolesterol yang relatif tinggi. Mereka juga menunjukkan tanda-tanda obesitas dan prevalensi penyakit jantung lebih tinggi dibandingkan kluster lainnya.

Faktor Risiko: Hipertensi, kolesterol tinggi, obesitas (BMI tinggi), dan kemungkinan penyakit jantung.
Implicasi: Kluster ini mencerminkan kelompok pasien yang mungkin mengalami komplikasi lebih lanjut terkait dengan penyakit jantung dan hipertensi, yang memerlukan intervensi medis lebih lanjut.

```
   age  blood_pressure  cholesterol  max_heart_rate  exercise_angina  \
0   24             250          139             212                0
1   29             132          187             147                0
2   46             271          185             193                0
3   73             102          200             125                0
4   49              91          163             192                0

   plasma_glucose  skin_thickness  insulin        bmi  diabetes_pedigree  \
0           108.0            33.0    109.0  37.999303           0.480277
1           202.0            42.0    129.0  25.588346           0.283986
2           149.0            43.0    102.0  37.892029           2.472309
3           105.0            77.0    165.0  18.660241           1.472052
4           162.0            31.0    170.0  12.767980           0.537627

   hypertension  heart_disease  gender_1.0  chest_pain_type_2  \
0             1              1        True              False
1             1              1       False              False
2             1              0       False              False
3             0              1       False               True
4             1              1        True              False

   chest_pain_type_3  chest_pain_type_4  residence_type_Urban  \
0              False               True                 True
1              False               True                 True
2               True              False                False
3              False              False                False
4               True              False                False

   smoking_status_Smoker  smoking_status_Unknown  Cluster
0                   True                   False        0
1                  False                    True        2
2                  False                   False        1
3                   True                   False        3
4                   True                   False        1
```

Gambar 6. *Clustering*

3.4. Evaluasi Model dengan *Silhouette Score*
Untuk mengevaluasi kualitas clustering, *Silhouette Score* dihitung[4]. Nilai *Silhouette Score* yang diperoleh adalah sekitar 0.07638, yang menunjukkan bahwa pemisahan antara kluster-kluster masih dapat dikategorikan sebagai agak buruk, karena nilai yang mendekati 0 mengindikasikan bahwa beberapa data mungkin berada di dekat batas kluster dan tidak terpisah dengan jelas. [1]Hal ini dapat disebabkan oleh adanya tumpang tindih karakteristik antara beberapa kluster, yang mungkin memerlukan pemrosesan lebih lanjut atau pemilihan metode *clustering* yang lebih sesuai. Visualisasi dapat dilihat pada gambar 7.

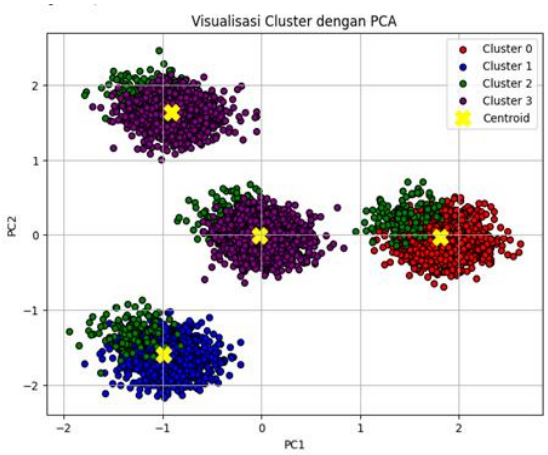


Gambar 7. *Grafik Silhouette Score*

Submitted : 04-06-2024 | Reviewed : 08-06-2025 | Accepted : 10-06-2025

3.5. Visualisasi Hasil *Clustering*

Tampilan clustering menggunakan *Principal Component Analysis* (PCA) [15] menunjukkan pemisahan yang relatif jelas antara kluster-kluster, meskipun ada beberapa data yang berada di area perbatasan antar kluster. [13] *Scatter plot* hasil PCA memperlihatkan tiga dimensi utama yang mengelompokkan pasien, berdasarkan faktor-faktor seperti tekanan darah, kolesterol, BMI, dan kadar glukosa plasma. Visualisasi ini membantu memperjelas bahwa kluster 1 dan 3 memiliki karakteristik yang lebih mirip dibandingkan dengan kluster 0 dan 2, yang dapat membantu dalam perancangan intervensi medis yang lebih terpersonalisasi. Visualisasi menggunakan PCA dapat dilihat pada gambar 8.



Gambar 8. *Grafik PCA*

3.6. *Implicasi* Hasil *Clustering*

Hasil *clustering* ini memberikan wawasan yang sangat berguna bagi tenaga medis dalam memahami pola kesehatan pasien berdasarkan karakteristik yang lebih spesifik. Kluster-kluster yang terbentuk dapat digunakan untuk merancang program pencegahan dan perawatan yang lebih terfokus. Sebagai contoh:

1. Pasien dalam Kluster 0 dan Kluster 3 yang memiliki faktor risiko penyakit jantung, hipertensi, dan obesitas dapat diberikan perhatian khusus untuk pengelolaan penyakit jantung dan hipertensi.
2. Pasien dalam Kluster 1 dapat diberi perhatian lebih terhadap pencegahan diabetes dan pengelolaan berat badan.
3. Pasien dalam Kluster 2 dapat diberi informasi mengenai gaya hidup sehat untuk mempertahankan kondisi kesehatan mereka yang sudah baik.

4. Kesimpulan

Penelitian ini telah berhasil mengelompokkan pasien ke dalam empat kluster berdasarkan karakteristik kesehatan mereka menggunakan metode *K-Means Clustering*. Melalui langkah-langkah pembersihan dan transformasi data yang melibatkan *One-Hot Encoding* untuk kolom kategorikal dan normalisasi data menggunakan *StandardScaler*, data siap dianalisis lebih lanjut.

Hasil analisis dengan *Elbow Method* menunjukkan bahwa jumlah kluster optimal adalah 4, yang kemudian digunakan untuk melakukan *clustering* pada data pasien. Setiap kluster memiliki karakteristik kesehatan yang berbeda, yang dapat digunakan untuk memahami pola-pola risiko penyakit yang ada pada pasien. Kluster-kluster yang ditemukan adalah:

Kluster 0: Pasien dengan usia lebih tua, tekanan darah dan kolesterol tinggi, serta BMI tinggi, menunjukkan risiko tinggi terhadap penyakit jantung dan hipertensi.

Kluster 1: Pasien yang lebih muda dengan tekanan darah normal, kadar glukosa plasma lebih tinggi, dan BMI sedikit lebih tinggi, menunjukkan potensi risiko diabetes dan gangguan metabolisme.

Kluster 2: Pasien dengan kondisi kesehatan relatif baik, memiliki gaya hidup sehat, dan risiko rendah terhadap penyakit jantung, diabetes, dan hipertensi.

Kluster 3: Pasien lebih tua dengan tekanan darah dan kolesterol tinggi, serta obesitas, menunjukkan kemungkinan risiko penyakit jantung dan hipertensi yang lebih tinggi.

Meskipun *Silhouette Score* yang diperoleh menunjukkan hasil yang relatif rendah (0.07638), yang menandakan ada beberapa tumpang tindih antara kluster, visualisasi menggunakan *Principal Component Analysis* (PCA) menunjukkan pemisahan yang cukup jelas antar kluster. Hal ini memberikan dasar untuk merancang program-program perawatan yang lebih terpersonalisasi, dengan mengutamakan pasien berdasarkan kelompok kesehatan mereka.

Hasil clustering ini memberikan kontribusi yang penting bagi praktik medis, memungkinkan tenaga medis untuk merancang program pencegahan yang lebih terfokus pada kelompok pasien dengan risiko kesehatan yang serupa, seperti pencegahan penyakit jantung, pengelolaan diabetes, dan intervensi untuk mengurangi obesitas.

Secara keseluruhan, teknik *K-Means Clustering* dapat digunakan sebagai alat bantu untuk analisis lebih lanjut dalam perawatan kesehatan, meskipun evaluasi lebih lanjut menggunakan metode *clustering* lain dan pemilihan fitur yang lebih tepat mungkin dapat meningkatkan hasil pemisahan kluster.

Daftar Rujukan

[1] N. Darapaneni *et al.*, “Machine learning approach for clustering of countries to identify the best strategies to combat Covid-19,” *2021 IEEE Int. IOT, Electron.*

Mechatronics Conf. IEMTRONICS 2021 - Proc., no. January 2020, 2021, doi: 10.1109/IEMTRONICS52119.2021.9422621.

[2] X. Zhao, G. Xie, Y. Luo, F. Liu, and H. Bai, "Enhancing Web Text Clustering Accuracy and Efficiency With a Maximum Entropy Function Model: Overcoming High-Dimensional and Directional Challenges," *IEEE Access*, vol. 12, no. March, pp. 42961–42973, 2024, doi: 10.1109/ACCESS.2024.3374770.

[3] J. Kauffmann, M. Esders, L. Ruff, G. Montavon, W. Samek, and K. R. Muller, "From Clustering to Cluster Explanations via Neural Networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 35, no. 2, pp. 1926–1940, 2024, doi: 10.1109/TNNLS.2022.3185901.

[4] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, no. October, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.

[5] X. Sun and P. Sajda, "Circular Clustering With Polar Coordinate Reconstruction," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 21, no. 5, pp. 1591–1600, 2024, doi: 10.1109/TCBB.2024.3406341.

[6] J. Oyelade *et al.*, "Data Clustering: Algorithms and Its Applications," *Proc. - 2019 19th Int. Conf. Comput. Sci. Its Appl. ICCSA 2019*, no. ii, pp. 71–81, 2019, doi: 10.1109/ICCSA.2019.000-1.

[7] T. K. Titus and M. Jajuli, "Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means," *Gener. J.*, vol. 6, no. 1, pp. 1–12, 2022, doi: 10.29407/gj.v6i1.16103.

[8] D. Deng, "DBSCAN Clustering Algorithm Based on Density," *Proc. - 2020 7th Int. Forum Electr. Eng. Autom. IFEEA 2020*, pp. 949–953, 2020, doi: 10.1109/IFEEA51475.2020.00199.

[9] J. Jokinen, T. Raty, and T. Lintonen, "Clustering structure analysis in time-series data with density-based clusterability measure," *IEEE/CAA J. Autom. Sin.*, vol. 6, no. 6, pp. 1332–1343, 2019, doi: 10.1109/JAS.2019.1911744.

[10] A. Damayanti, W. D. Utami, D. C. R. Novitasari, P. K. Intan, and M. L. Kurniawan, "Cluster Analysis of Environmental Pollution in Indonesia Using Complete Linkage Method with Elbow Optimization," *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 7, no. 2, p. 399, 2023, doi: 10.31764/jtam.v7i2.12961.

[11] J. Wen *et al.*, "Adaptive Graph Completion Based Incomplete Multi-View Clustering," *IEEE Trans. Multimed.*, vol. 23, no. c, pp. 2493–2504, 2021, doi: 10.1109/TMM.2020.3013408.

[12] H. Shao, P. Zhang, X. Chen, F. Li, and G. Du, "A Hybrid and Parameter-Free Clustering Algorithm for Large Data Sets," *IEEE Access*, vol. 7, pp. 24806–24818, 2019, doi: 10.1109/ACCESS.2019.2900260.

[13] S. M. Al-Ghuribi, S. A. M. Noah, M. A. Mohammed, S. N. Qasem, and B. A. H. Murshed, "To Cluster or Not to Cluster: The Impact of Clustering on the Performance of Aspect-Based Collaborative Filtering," *IEEE Access*, vol. 11, no. May, pp. 41979–41994, 2023, doi: 10.1109/ACCESS.2023.3270260.

[14] D. H. Duc, C. Dang, T. T. Thao, V. Van Yem, and H. M. Son, "A Novel Clustering Algorithm based on Cooperative and Iterative Evaluation Exchange for University Classification," *ICCE 2020 - 2020 IEEE 8th Int. Conf. Commun. Electron.*, pp. 307–312, 2021, doi: 10.1109/ICCE48956.2021.9352112.

[15] R. Tariq, K. Lavangnananda, P. Bouvry, and P. Mongkolnam, "Partitioning Graph Clustering With User-Specified Density," *IEEE Access*, vol. 11, no. October, pp. 122273–122294, 2023, doi: 10.1109/ACCESS.2023.3329429.

[16] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

Submitted : 04-06-2024 | Reviewed : 08-06-2025 | Accepted : 10-06-2025

Submitted : 04-06-2024 | Reviewed : 08-06-2025 | Accepted : 10-06-2025