

## Analisis Hierarchical Clustering untuk Segmentasi Pelanggan pada Dataset Mall Customers

Maissy Angelica Pakpahan<sup>a</sup>, Sirlia Sahid<sup>b</sup>, Mika M.F Simanullang<sup>c</sup>, Rifqi Putra Winanda<sup>d</sup>

<sup>a</sup>Ilmu Komputer, FMIPA, Universitas Negeri Medan, [maissyangelica63@gmail.com](mailto:maissyangelica63@gmail.com)

<sup>b</sup>Ilmu Komputer, FMIPA, Universitas Negeri Medan, [sirliasahid@gmail.com](mailto:sirliasahid@gmail.com)

<sup>c</sup>Ilmu Komputer, FMIPA, Universitas Negeri Medan, [simanullangmika6@gmail.com](mailto:simanullangmika6@gmail.com)

<sup>d</sup>Ilmu Komputer, FMIPA, Universitas Negeri Medan, [putrarifqy416@gmail.com](mailto:putrarifqy416@gmail.com)

### Abstract

*This study aims to analyze customer segmentation using the Hierarchical Clustering method on the Mall Customers dataset. The main objective is to group customers based on similarities in Annual Income and Spending Score. The research method involves data preprocessing using Z-score normalization, calculation of Euclidean distance, and clustering using Ward linkage. The optimal number of clusters is determined using several evaluation metrics such as Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. The results show that the optimal number of clusters is five, with good clustering performance indicated by a high Silhouette Score and low Davies-Bouldin value. Each cluster represents a unique customer segment, such as high-income high-spending customers and low-income low-spending customers. This clustering can be used as a basis for targeted marketing strategies.*

**Keywords:** clustering, hierarchical clustering, customer segmentation, mall customers, data mining

### Abstrak

Penelitian ini bertujuan untuk menganalisis segmentasi pelanggan menggunakan metode Hierarchical Clustering pada dataset Mall Customers. Tujuan utama penelitian adalah mengelompokkan pelanggan berdasarkan kemiripan Annual Income dan Spending Score. Metode penelitian meliputi preprocessing data menggunakan normalisasi Z-score, perhitungan jarak Euclidean, serta proses clustering menggunakan metode Ward linkage. Penentuan jumlah cluster optimal dilakukan dengan menggunakan beberapa metrik evaluasi seperti Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah lima dengan performa clustering yang baik ditunjukkan oleh nilai Silhouette yang tinggi dan Davies-Bouldin yang rendah. Setiap cluster merepresentasikan segmen pelanggan yang berbeda seperti pelanggan dengan pendapatan tinggi dan belanja tinggi maupun rendah. Hasil ini dapat digunakan sebagai dasar strategi pemasaran yang lebih efektif.

**Kata Kunci:** clustering, hierarchical clustering, segmentasi pelanggan, mall customers, data mining

*This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license*



### PENDAHULUAN

Penelitian dalam bidang data mining terus berkembang seiring dengan meningkatnya kebutuhan akan pengolahan data yang efektif dan efisien. Berbagai metode telah dikembangkan untuk mengatasi permasalahan analisis data, khususnya dalam mengidentifikasi pola dan struktur yang tersembunyi di dalam data. Salah satu pendekatan yang banyak digunakan adalah teknik clustering, yang menjadi bagian penting dalam proses eksplorasi data karena kemampuannya dalam mengelompokkan data berdasarkan karakteristik tertentu.

Perkembangan teknologi informasi yang semakin pesat telah mendorong peningkatan jumlah data yang dihasilkan di berbagai bidang, seperti bisnis, kesehatan, pendidikan, dan industri. Data yang dihasilkan tidak hanya berukuran besar, tetapi juga memiliki kompleksitas yang tinggi, sehingga diperlukan metode analisis yang mampu mengolah data tersebut menjadi informasi yang bermakna. Salah satu pendekatan yang широко digunakan adalah data mining, yaitu proses ekstraksi pola atau pengetahuan tersembunyi dari sekumpulan data dengan memanfaatkan teknik statistik dan pembelajaran mesin. Pendekatan ini dinilai efektif dalam mendukung proses pengambilan keputusan berbasis data [1].

Salah satu teknik dalam data mining yang banyak digunakan adalah clustering, yaitu metode pengelompokan data tanpa label (unsupervised learning). Clustering bertujuan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristiknya, sehingga objek dalam satu kelompok memiliki tingkat kesamaan yang tinggi dibandingkan dengan objek pada kelompok lainnya. Teknik ini telah banyak diterapkan dalam berbagai bidang, seperti segmentasi pelanggan, analisis kesehatan, dan pengelompokan data wilayah [2]. Selain itu, clustering juga digunakan dalam pengelompokan data kasus tertentu guna membantu proses analisis yang lebih terarah [3].

Salah satu teknik dalam data mining yang banyak digunakan adalah clustering, yaitu metode pengelompokan data tanpa label (unsupervised learning). Clustering bertujuan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristiknya, sehingga objek dalam satu kelompok memiliki tingkat

kesamaan yang tinggi dibandingkan dengan objek pada kelompok lainnya. Teknik ini telah banyak diterapkan dalam berbagai bidang, seperti segmentasi pelanggan, analisis kesehatan, dan pengelompokan data wilayah [2]. Selain itu, clustering juga digunakan dalam pengelompokan data kasus tertentu guna membantu proses analisis yang lebih terarah [3]. Salah satu metode clustering yang cukup populer adalah Hierarchical Clustering, yang mampu membentuk struktur pengelompokan data secara bertahap dalam bentuk hierarki. Metode ini memiliki keunggulan dalam memberikan representasi visual berupa dendrogram, sehingga memudahkan dalam memahami hubungan antar data serta menentukan jumlah cluster yang optimal. Pendekatan ini juga dinilai fleksibel karena tidak memerlukan penentuan jumlah cluster di awal proses [4]. Selain itu, hierarchical clustering dapat digunakan untuk menganalisis struktur data secara lebih mendalam, bahkan dalam konteks pengelompokan algoritma atau sistem yang kompleks [5].

Beberapa penelitian sebelumnya menunjukkan bahwa metode hierarchical clustering mampu menghasilkan pengelompokan data yang lebih stabil dan mudah diinterpretasikan dibandingkan metode lainnya. Penggunaan metode ini dalam berbagai studi, seperti analisis ketahanan pangan dan kesejahteraan wilayah, menunjukkan bahwa hierarchical clustering dapat memberikan hasil yang representatif dalam mengidentifikasi pola data [6]. Di sisi lain, keberhasilan proses clustering juga dipengaruhi oleh tahap preprocessing, seperti normalisasi data, yang bertujuan untuk menyamakan skala antar variabel agar tidak terjadi bias dalam perhitungan jarak .

Beberapa penelitian sebelumnya menunjukkan bahwa metode hierarchical clustering mampu menghasilkan pengelompokan data yang lebih stabil dan mudah diinterpretasikan dibandingkan metode lainnya. Penggunaan metode ini dalam berbagai studi, seperti analisis ketahanan pangan dan kesejahteraan wilayah, menunjukkan bahwa hierarchical clustering dapat memberikan hasil yang representatif dalam mengidentifikasi pola data [6]. Di sisi lain, keberhasilan proses clustering juga dipengaruhi oleh tahap preprocessing, seperti normalisasi data, yang bertujuan untuk menyamakan skala antar variabel agar tidak terjadi bias dalam perhitungan jarak.

Meskipun demikian, masih terdapat permasalahan dalam penerapan clustering, terutama dalam menentukan metode yang tepat serta memastikan kualitas hasil pengelompokan data. Oleh karena itu, diperlukan penerapan metode clustering yang sesuai dengan karakteristik data yang digunakan, serta tahapan analisis yang sistematis agar hasil yang diperoleh lebih optimal.

Berdasarkan permasalahan tersebut, penelitian ini menggunakan dataset Iris sebagai objek studi, yang merupakan salah satu dataset klasik dalam bidang machine learning. Dataset ini terdiri dari beberapa atribut numerik yang merepresentasikan karakteristik bunga Iris, serta memiliki tiga kelas utama, yaitu Setosa, Versicolor, dan Virginica. Meskipun sederhana, dataset ini memiliki tingkat kemiripan antar kelas yang cukup tinggi, sehingga menjadi tantangan dalam proses clustering.

Penelitian ini bertujuan untuk menerapkan metode Hierarchical Clustering dalam mengelompokkan data Iris berdasarkan tingkat kemiripan fitur. Proses analisis dilakukan melalui beberapa tahapan, yaitu pengolahan data dari file Excel, normalisasi data menggunakan Z-score, perhitungan jarak menggunakan Euclidean distance, serta pembentukan cluster menggunakan metode Ward linkage. Hasil dari proses ini divisualisasikan dalam bentuk dendrogram untuk mengetahui struktur pengelompokan data secara hierarkis.

Diharapkan hasil penelitian ini dapat memberikan pemahaman yang lebih jelas mengenai penerapan hierarchical clustering dalam analisis data, serta menjadi referensi dalam pengembangan metode clustering pada berbagai permasalahan yang serupa.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining dengan menerapkan metode *Hierarchical Clustering* pada dataset Mall Customers. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, normalisasi, pembangunan model clustering, dan evaluasi performa cluster. Seluruh proses komputasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka *scikit-learn*, *scipy*, dan *matplotlib*.

### 2.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah Mall Customers Dataset yang terdiri dari 200 pelanggan sebuah pusat perbelanjaan. Dataset ini memiliki 5 atribut yang ditunjukkan pada Tabel 1. Tidak ditemukan nilai yang hilang (*missing values*) pada seluruh atribut sehingga data dapat langsung digunakan tanpa proses imputasi.

Tabel 1. Deskripsi Atribut Dataset Mall Customers

Atribut	Tipe Data	Deskripsi	Min	Max
CustomerID	Integer	Identitas unik pelanggan	1	200
Genre	Kategorikal	Jenis kelamin (Male/Female)	-	-

Age	Integer	Usia pelanggan (tahun)	18	70
Annual Income	Integer (k)	Pendapatan tahunan dalam ribuan dolar	15	137
Spending Score	Integer (1–100)	Skor belanja yang ditetapkan mall	1	99

## 2.2 Pra – Pemrosesan Data

Pra-pemrosesan data dilakukan dengan normalisasi Z-score pada dua fitur utama yang digunakan dalam proses clustering, yaitu *Annual Income* dan *Spending Score*. Normalisasi dilakukan untuk menyetarakan skala kedua fitur sehingga tidak ada dominasi satu fitur terhadap perhitungan jarak. Formula normalisasi Z-score yang digunakan adalah:

$$1. Z = (X - \mu) / \sigma \quad (1)$$

2.

Keterangan: X adalah nilai asli,  $\mu$  adalah rata-rata, dan  $\sigma$  adalah standar deviasi. Tabel 2 menyajikan statistik deskriptif sebelum dan setelah normalisasi.

Tabel 2. Statistik Deskriptif Data Sebelum dan Sesudah Normalisasi

Statistik	Age	Annual Income (k)	Spending Score	Z_Income ( $\mu/\sigma$ )	Z_Spending ( $\mu/\sigma$ )
Mean	38.85	60.56	50.20	0.00 / 1.00	0.00 / 1.00
Std Dev	13.97	26.26	25.82	-	-
Min	18	15	1	-1.739	-1.906
Max	70	137	99	2.925	1.892

## 2.3 Perhitungan Matriks Jarak Euclidean

Setelah normalisasi, jarak antar titik data dihitung menggunakan metrik jarak *Euclidean*. Formula jarak Euclidean untuk dua titik data yang telah dinormalisasi adalah:

$$d(i, j) = \sqrt{[(z^1_i - z^1_j)^2 + (z^2_i - z^2_j)^2]} \quad (2)$$

Matriks jarak dihitung untuk seluruh 200 data pelanggan. Sebagai ilustrasi, Tabel 3 menyajikan sebagian matriks jarak dari 6 sampel pertama (C001–C006) setelah normalisasi Z-score.

Tabel 3. Matriks Jarak Euclidean (6 sample pertama)

ID \ ID	C001	C002	C003	C004	C005
C001	0	1.2619	1.0848	1.2236	0.881
C002	1.2619	0	2.2959	0.4563	1.5141
C003	1.0848	2.2959	0	2.1331	1.1123
C004	1.2236	0.4563	2.1331	0	1.1956
C005	0.881	1.5141	1.1123	1.1956	0
C006	1.4181	0.8932	2.1487	0.4412	1.0816

## 2.4 Metode Hierarchical Clustering – Ward Linkage

Proses clustering dilakukan menggunakan metode *Hierarchical Agglomerative Clustering* dengan *Ward Linkage*. Metode Ward dipilih karena terbukti menghasilkan cluster yang paling seimbang dan kompak dibandingkan metode linkage lainnya. Prinsip kerja Ward Linkage adalah meminimalkan peningkatan total *Sum of Squared Error* (SSE) dalam satu cluster ketika dua cluster digabungkan. Formula Ward adalah:

$$\Delta(A, B) = (n_A \cdot n_B) / (n_A + n_B) \cdot \|\mu_A - \mu_B\|^2 \quad (3)$$

Keterangan:  $n_A$  dan  $n_B$  adalah jumlah anggota cluster A dan B, sedangkan  $\mu_A$  dan  $\mu_B$  adalah centroid masing-masing cluster. Proses penggabungan dilakukan secara iteratif mulai dari setiap titik data sebagai cluster tunggal hingga seluruh data tergabung dalam satu cluster (dendrogram).

## 2.5 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal ( $k$ ) dilakukan dengan membandingkan tiga metrik evaluasi, yaitu: (1) *Silhouette Score* yang mengukur kekompakan dan separasi cluster (rentang 0–1, semakin tinggi semakin baik); (2) *Calinski-Harabasz Index* (CH Score) yang mengukur rasio dispersi antar-cluster terhadap dalam-cluster; dan (3) *Davies-Bouldin Index* (DB Score) yang mengukur kemiripan rata-rata cluster dengan cluster tetangganya (semakin rendah semakin baik). Selain itu, *Cophenetic Correlation Coefficient* juga dihitung untuk menilai seberapa baik dendrogram merepresentasikan jarak antar data asli.

## HASIL DAN PEMBAHASAN

### 3.1. Hasil Normalisasi Data

Normalisasi Z-score berhasil diterapkan pada dua fitur utama. Setelah normalisasi, nilai rata-rata kedua fitur menjadi 0 dengan standar deviasi 1, memastikan kesetaraan skala dalam proses perhitungan jarak. Sebagai contoh, pelanggan dengan *Annual Income* = 15 k menghasilkan Z-score sebesar  $-1.739$ , sedangkan pelanggan dengan *Spending Score* = 39 menghasilkan Z-score  $-0.435$ . Data setelah normalisasi siap digunakan sebagai masukan pada proses clustering.

### 3.2. Matriks Jarak dan Proses Linkage

Matriks jarak Euclidean berhasil dihitung untuk seluruh 200 data pelanggan (matriks  $200 \times 200$ ). Analisis terhadap 20 sampel pertama menunjukkan jarak terdekat antara C013 dan C015 sebesar 0.0601, yang kemudian menjadi pasangan pertama yang digabungkan pada langkah pertama proses linkage Ward. Secara keseluruhan, terdapat 19 tahapan penggabungan cluster pada 20 sampel manual. Jarak Ward mulai menunjukkan lonjakan signifikan pada langkah 17 (jarak 2.8226) dan langkah 18 (4.7341), yang mengindikasikan adanya batas pemisahan cluster yang optimal.

### 3.3. Penentuan Jumlah Cluster Optimal

Evaluasi dilakukan terhadap empat metode linkage untuk menentukan jumlah cluster optimal. Hasil perbandingan disajikan pada Tabel 4. *Cophenetic Correlation Coefficient* pada metode Ward mencapai nilai 0.7209, yang melebihi ambang batas 0.7, artinya dendrogram merepresentasikan struktur jarak data asli dengan baik.

Tabel 4. Perbandingan Metrik Evaluasi Clustering

Metode Linkage	Silhouette Score	k Optimal	CH Score	DB Score
<b>Ward</b>	<b>0.5538</b>	<b>5</b>	<b>244.41</b>	<b>0.5779</b>
Complete	0.5531	5	-	-
Average	0.5393	7	-	-
Single	0.4339	2	-	-

Berdasarkan Tabel 4, metode *Ward Linkage* menghasilkan *Silhouette Score* tertinggi sebesar 0.5538 dengan jumlah cluster optimal  $k = 5$ . Nilai CH Score yang tinggi (244.41) menunjukkan separasi antar-cluster yang baik, sementara nilai DB Score yang rendah (0.5779) mengkonfirmasi kekompakan cluster. Oleh karena itu, analisis selanjutnya menggunakan  $k = 5$  cluster dengan metode Ward.

### 3.4. Hasil Clustering Final

Proses Hierarchical Clustering Ward dengan  $k = 5$  menghasilkan lima segmen pelanggan yang terdefinisi dengan baik. Tabel 5 menyajikan ringkasan karakteristik setiap cluster yang terbentuk dari seluruh 200 data pelanggan.

Tabel 5. Karakteristik Lima Cluster Pelanggan Mall

Cluster	N (%)	Avg Income (k)	Avg Spending	Avg Age
Cluster 1	39 (20%)	86.5	82.1	32.7
Cluster 2	32 (16%)	89.4	15.6	41.0
Cluster 3	21 (10%)	25.1	80.0	25.3
Cluster 4	23 (12%)	26.3	20.9	45.2
Cluster 5	85 (42%)	55.8	49.1	42.5

### 3.5 Interpretasi dan Pembahasan Tiap Cluster

Cluster 1 (N=39, 20%) – Pelanggan VIP dengan Pendapatan Tinggi dan Belanja Tinggi. Cluster ini merupakan segmen paling berharga bagi pihak mall. Pelanggan memiliki rata-rata pendapatan tahunan 86.5 k dan skor belanja 82.1 dengan usia rata-rata 32.7 tahun, menunjukkan segmen muda-dewasa yang memiliki daya beli tinggi dan motivasi berbelanja yang kuat. Strategi yang direkomendasikan adalah program loyalitas premium dan penawaran eksklusif.

Cluster 2 (N=32, 16%) – Pelanggan Pendapatan Tinggi, Belanja Rendah. Segmen ini memiliki rata-rata pendapatan tertinggi (89.4 k) namun skor belanja sangat rendah (15.6) dengan usia rata-rata 41 tahun. Pola ini mengindikasikan pelanggan yang mampu secara finansial namun belum termotivasi untuk berbelanja di mall. Intervensi berupa penawaran produk premium, program diskon bertarget, dan peningkatan pengalaman berbelanja dapat meningkatkan keterlibatan segmen ini.

Cluster 3 (N=21, 10%) – Pelanggan Muda dengan Pendapatan Rendah dan Belanja Tinggi. Dengan rata-rata usia 25.3 tahun, pendapatan 25.1 k, dan skor belanja 80.0, segmen ini menggambarkan pelanggan muda yang konsumtif meskipun memiliki keterbatasan pendapatan. Program cicilan, diskon produk tertentu, dan promosi berbasis media sosial relevan untuk segmen ini.

Cluster 4 (N=23, 12%) – Pelanggan Pendapatan Rendah, Belanja Rendah. Merupakan segmen dengan daya beli dan motivasi belanja paling rendah (pendapatan 26.3 k, skor belanja 20.9). Usia rata-rata 45.2 tahun mengindikasikan segmen paruh baya dengan anggaran terbatas. Program promosi berbasis harga, produk kebutuhan sehari-hari, dan event diskon besar dapat menjadi daya tarik segmen ini.

Cluster 5 (N=85, 42%) – Segmen Terbesar: Pendapatan Menengah, Belanja Standar. Dengan proporsi terbesar (42% dari total data), segmen ini memiliki rata-rata pendapatan 55.8 k dan skor belanja 49.1. Kelompok ini merepresentasikan pelanggan tipikal mall yang menjadi tulang punggung pendapatan. Strategi retensi pelanggan, program poin reward, dan penawaran lintas kategori produk sangat relevan untuk mempertahankan dan meningkatkan frekuensi kunjungan segmen ini.

### 3.6. Diskusi Temuan

Secara keseluruhan, hasil penelitian mengkonfirmasi bahwa Hierarchical Clustering dengan metode Ward Linkage mampu mengidentifikasi struktur segmentasi pelanggan yang bermakna pada dataset Mall Customers. Nilai Cophenetic Correlation sebesar 0.7209 menunjukkan bahwa dendrogram yang dihasilkan merepresentasikan hubungan jarak antar data dengan akurasi yang dapat diterima. Pemilihan  $k = 5$  secara konsisten dikonfirmasi oleh dua metode linkage berbeda (Ward dan Complete), yang memperkuat validitas jumlah cluster tersebut.

Temuan ini sejalan dengan karakteristik umum segmentasi pelanggan ritel yang biasanya terbagi dalam kelompok berdasarkan kombinasi daya beli dan perilaku konsumsi. Implikasi manajerial dari penelitian ini adalah bahwa strategi pemasaran mall tidak dapat bersifat seragam, melainkan perlu disesuaikan dengan profil unik masing-masing segmen untuk memaksimalkan efektivitas program promosi dan pengelolaan hubungan pelanggan.

## SIMPULAN

Penerapan Hierarchical Clustering dengan Ward Linkage pada dataset Mall Customers menunjukkan hasil yang cukup memuaskan dalam membentuk segmentasi pelanggan yang terstruktur. Dari serangkaian pengujian menggunakan tiga metrik evaluasi Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index diperoleh bahwa pembagian ke dalam 5 cluster memberikan performa terbaik, dengan nilai Silhouette 0.5538, CH Score 244.41, dan DB Score 0.5779. Selain itu, kualitas dendrogram yang dibuat dievaluasi sebagai representatif, seperti yang ditunjukkan oleh nilai Cophenetic Correlation sebesar 0.7209, yang melampaui ambang batas 0.7.

Kelima cluster yang terbentuk menunjukkan betapa beragamnya profil pelanggan; mereka terdiri dari pelanggan muda yang sangat berdaya beli hingga pelanggan paruh baya yang lebih konservatif. Hasil ini menunjukkan bahwa metode pemasaran yang seragam tidak akan efektif jika diterapkan pada semua pelanggan. Dengan memahami karakteristik tiap segmen secara lebih mendalam, pihak manajemen mall dapat merancang strategi promosi dan program loyalitas yang lebih tepat sasaran.

## UCAPAN TERIMAKASIH

Kami mengucapkan terima kasih kepada dosen pembimbing atas arahan, bimbingan, dan masukan yang diberikan selama proses penelitian ini. Kami juga menyampaikan apresiasi kepada seluruh pihak yang telah membantu dalam pelaksanaan penelitian, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat diselesaikan dengan baik. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu pengetahuan.

#### DAFTAR PUSTAKA

- [1] B. Zhang, A. Panagiotelis, and H. Li, "Constructing hierarchical time series through clustering: Is there an optimal way for forecasting?," *Int. J. Forecast.*, vol. 41, no. 3, 2025, doi: <https://doi.org/10.1016/j.ijforecast.2024.10.002>.
- [2] A. Ramadhan, F. Achmad, I. Zulkarnain, and M. Aritsugi, "Evaluation of K-Means, DBSCAN, and Hierarchical Clustering for Strategic Segmentation of Tourism SMEs in Rembang, Indonesia," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 3, pp. 1605–1630, Jul. 2025, doi: [10.52436/1.jutif.2025.6.3.4602](https://doi.org/10.52436/1.jutif.2025.6.3.4602).
- [3] A. Rahmah, N. Faoziatun Khusna, S. A. Sanmas, S. Aulia, S. Amaria, and F. Fauzi, "Comparison Analysis of Hierarchical," 2025.
- [4] H. A. Amhimid, F. A. Alzahra Aljehany, M. A. Mohamed, and K. A. Farag, "Comparative Study of Four Methods in Hierarchical Cluster Analysis," 2025.
- [5] M. S. Calvo and H. S. Lee, "Systematic taxonomic framework of metaheuristic algorithms using hierarchical clustering and structural criteria: how novel is the novelty?," *Artif. Intell. Rev.*, vol. 59, no. 2, Dec. 2025, doi: [10.1007/s10462-025-11456-8](https://doi.org/10.1007/s10462-025-11456-8).
- [6] R. F. Sinaga, M. A. Prabukusumo, and J. Manurung, "Journal of Intelligent Decision Support System (IDSS) Comparison of k-means clustering with hierarchical agglomerative clustering for the analysis of food security of rice sector in Indonesia," 2025.