

# ANALISIS KLASTER PADA TINGKAT AKSES LISTRIK DI NEGARA-NEGARA DUNIA MELALUI PENDEKATAN K-MEANS DAN HIERARCHICAL CLUSTERING

Bayu Eko Saputro<sup>1</sup>

<sup>1</sup>) Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: [bayu.535210109@stu.untar.ac.id](mailto:bayu.535210109@stu.untar.ac.id)

## Abstrak

*Penelitian ini menganalisis tingkat akses listrik di berbagai negara menggunakan pendekatan analisis kluster K-Means dan Hierarchical Clustering. Dengan memanfaatkan data tingkat akses listrik yang bersumber dari lembaga terpercaya seperti Bank Dunia dan IEA, penelitian ini mengelompokkan negara-negara berdasarkan kesamaan karakteristik akses listrik, termasuk variabel seperti persentase populasi dengan akses listrik, Indeks Pembangunan Manusia (HDI), dan Produk Domestik Bruto (PDB) per kapita. Hasil analisis menunjukkan bahwa metode K-Means unggul dalam efisiensi dan pemisahan kluster, sementara Hierarchical Clustering memberikan wawasan hierarkis yang lebih mendalam. Dengan metrik evaluasi seperti Silhouette Score dan Davies-Bouldin Index, klusterisasi memberikan pemahaman yang signifikan terkait pola distribusi listrik global, mendukung kebijakan untuk meningkatkan akses listrik terutama di negara-negara berkembang.*

**Kata kunci**— Akses Listrik, Analisis Kluster, Distribusi Energi, Hierarchical Clustering, K-Means.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Dengan berkembangnya teknologi dan industri, kebutuhan akan energi khususnya energi listrik semakin meningkat. Akses terhadap listrik merupakan salah satu faktor kunci dalam pembangunan ekonomi dan sosial suatu negara. Seperti yang dikemukakan Dutta dan Khan, peningkatan akses terhadap listrik dapat meningkatkan produktivitas dan kualitas hidup masyarakat serta mendorong kemajuan ekonomi [1]. Secara umum, negara-negara dengan akses listrik yang lebih baik cenderung lebih maju dalam hal infrastruktur, pendidikan, dan layanan

kesehatan. Sebaliknya, negara-negara dengan akses terbatas terhadap listrik seringkali menghadapi tantangan yang lebih besar dalam mengembangkan perekonomian dan meningkatkan kualitas hidup masyarakatnya.

Dalam skala yang lebih kecil, menunjukkan bahwa terbatasnya akses listrik di daerah pedesaan dapat mempengaruhi kesejahteraan sosial dan menurunkan kualitas hidup masyarakat [2]. Peneliti lain juga menekankan pentingnya solusi teknologi untuk meningkatkan akses listrik di negara-negara berkembang [3]. Pembangunan infrastruktur energi yang tidak merata berkontribusi terhadap kesenjangan yang signifikan dalam akses terhadap listrik di Afrika sub-Sahara [4]. Salah satu permasalahan yang muncul di negara-negara berkembang adalah ketimpangan akses terhadap listrik. Ketimpangan ini juga diamati oleh peneliti sebelumnya, studi ini menunjukkan bahwa distribusi energi yang tidak merata menyebabkan kesenjangan yang besar antara negara maju dan negara berkembang dalam hal akses terhadap listrik [5].

Oleh karena itu penting untuk melakukan analisis tingkat akses listrik di berbagai negara untuk memahami pola distribusi dan perbedaan akses yang ada. Dalam konteks ini, analisis kluster dapat memberikan wawasan yang lebih luas mengenai karakteristik akses listrik berdasarkan kelompok negara yang memiliki kesamaan. Analisis kluster, dengan menggunakan metode seperti K-Means dan Hierarchical Clustering, memungkinkan negara-negara dikelompokkan berdasarkan tingkat akses listrik yang serupa, sehingga memberikan gambaran yang lebih jelas tentang pola dan faktor yang mempengaruhi akses listrik di berbagai wilayah di dunia. Dengan cara ini dapat mendukung pengembangan akses yang lebih baik terhadap listrik, sehingga mendorong pencapaian Tujuan Pembangunan Berkelanjutan

(SDGs), terutama dalam hal pengurangan kemiskinan dan peningkatan kualitas hidup [6].

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana tingkat akses listrik di negara-negara dunia dapat dikelompokkan menggunakan pendekatan K-Means dan Hierarchical Clustering?
2. Apa saja faktor yang mempengaruhi perbedaan tingkat akses listrik antar negara?
3. Apa implikasi dari hasil klusterisasi terhadap kebijakan distribusi listrik global?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Mengelompokkan negara-negara berdasarkan tingkat akses listrik menggunakan metode K-Means dan Hierarchical Clustering.
2. Menganalisis faktor-faktor yang mempengaruhi perbedaan tingkat akses listrik antar negara.
3. Memberikan rekomendasi kebijakan untuk meningkatkan distribusi dan akses listrik di negara-negara dengan tingkat akses rendah.

## **2. METODE PENELITIAN**

Bagian ini menjelaskan secara rinci metode yang digunakan dalam penelitian untuk mengelompokkan negara-negara berdasarkan tingkat akses listriknya. Pemilihan metode, sumber data, serta proses pengolahan dan analisis dijabarkan secara sistematis guna memberikan gambaran yang jelas mengenai pendekatan penelitian yang digunakan. Dengan demikian, pembaca dapat memahami landasan metodologis dari setiap langkah yang dilakukan dalam proses analisis, mulai dari desain penelitian, pemilihan algoritma klusterisasi, hingga teknik evaluasi yang digunakan untuk menilai kualitas hasil klaster yang terbentuk.

### **2.1 Desain Penelitian**

Penelitian ini mengambil pendekatan kuantitatif dengan menggunakan analisis klaster untuk mengelompokkan negara-negara di dunia berdasarkan tingkat akses listriknya. Analisis cluster membantu mengidentifikasi pola data yang dapat menunjukkan hubungan antar negara berdasarkan kesamaan faktor yang mempengaruhi akses listrik. Sebagaimana dijelaskan oleh peneliti sebelumnya metode analisis klaster dapat membantu mengungkap pola distribusi tersembunyi pada data terkait akses energi di berbagai negara [7]. Penelitian ini menggunakan dua metode clustering utama yaitu

clustering K-Means dan Hierarchical Clustering, masing-masing metode memiliki kelebihan dan kekurangan dalam hal pengolahan dan interpretasi data. Misalnya, metode K-Means dikenal efektif dalam mengelompokkan data berukuran besar, sedangkan Hierarchical Clustering lebih cocok untuk analisis yang memerlukan pemahaman lebih dalam mengenai struktur data [8].

Data yang digunakan dalam penelitian ini mencakup tingkat akses listrik di berbagai negara, yang dikumpulkan dari sumber terpercaya seperti Bank Dunia dan Badan Energi Internasional (IEA). Variabel yang digunakan dalam analisis ini adalah persentase penduduk yang mempunyai akses terhadap listrik, diukur sebagai persentase dari total penduduk suatu negara. Penelitian sebelumnya menunjukkan bahwa indikator akses listrik dapat berperan penting dalam menilai kemajuan pembangunan suatu negara [9]. Selain itu, faktor ekonomi, infrastruktur, dan geografis juga mungkin berperan dalam menentukan kelompok negara yang teridentifikasi dalam klaster tersebut. Misalnya, studi lain menemukan bahwa infrastruktur yang buruk dan ketergantungan pada energi terbarukan sering kali menimbulkan hambatan besar dalam meningkatkan akses listrik di negara-negara berkembang [10].

### **2.2 Algoritma Clustering**

Pada bagian ini akan dibahas lebih detail mengenai dua algoritma clustering yang digunakan dalam penelitian ini, yaitu K-Means clustering dan Hierarchical clustering. Kedua metode ini merupakan algoritma yang banyak digunakan dalam analisis klaster, namun memiliki pendekatan berbeda dalam pembentukan klaster dan interpretasi hasil. K-Means Clustering merupakan algoritma yang membagi data menjadi beberapa cluster berdasarkan kedekatannya dengan pusat cluster, yang dihitung ulang pada setiap iterasi hingga konvergensi. Keunggulan utama K-Means adalah kemampuannya dalam memproses data berukuran besar dengan cepat, meskipun kekurangannya adalah ketergantungannya pada jumlah cluster yang harus ditentukan sebelumnya. Sedangkan Hierarchical clustering merupakan suatu metode pembentukan cluster melalui pendekatan bertingkat, baik dengan menggabungkan data (kombinasi) maupun dengan membagi data (pembagian) [11]. Metode ini lebih fleksibel karena tidak memerlukan penentuan jumlah cluster terlebih dahulu, namun seringkali memerlukan waktu perhitungan yang lebih lama dibandingkan K-Means menunjukkan bahwa Hierarchical clustering

memberikan hasil yang lebih informatif ketika data memiliki struktur hierarki yang jelas, dan metode ini cenderung memberikan visualisasi yang lebih mudah diinterpretasikan sebagai dendrogram.

### **2.2.1 K-Means Clustering**

K-Means Clustering merupakan salah satu algoritma yang paling umum digunakan dalam analisis kluster karena kesederhanaannya dan kemampuannya untuk mengelola data dalam jumlah besar. Metode ini berfungsi dengan mengelompokkan data ke dalam sejumlah kluster yang telah ditentukan sebelumnya (dikenal sebagai parameter  $k$ ), berdasarkan kedekatannya dengan pusat kluster (centroid) [12].

Tahapan K-Means dimulai dengan memilih sejumlah  $k$  titik pusat secara acak yang akan berfungsi sebagai pusat kluster sementara. Selanjutnya, setiap titik data (negara dalam konteks ini) akan dipetakan ke pusat kluster terdekat berdasarkan jarak Euclidean. Setelah setiap titik data dikelompokkan ke kluster yang sesuai, pusat kluster dihitung kembali sebagai rata-rata dari titik-titik dalam kluster tersebut. Proses ini diulang hingga posisi pusat kluster tidak mengalami perubahan yang signifikan lagi, menandakan bahwa klusterisasi telah selesai. Keunggulan utama dari K-Means adalah kemampuannya dalam menangani dataset yang besar dan kompleks secara cepat. Namun, kelemahan utama dari metode ini adalah ketergantungannya pada penentuan nilai  $k$  yang tepat. Apabila nilai  $k$  terlalu kecil atau terlalu besar, kluster yang terbentuk mungkin tidak mencerminkan pola yang ada dalam data [13].

### **2.2.2 Hierarchical Clustering**

Hierarchical Clustering berbeda dengan K-Means karena tidak memerlukan jumlah kluster yang sudah ditentukan sebelumnya. Sebaliknya, metode ini membangun kluster secara bertahap, mulai dari masing-masing titik data sebagai kluster individu dan kemudian menggabungkannya secara iteratif berdasarkan kemiripan mereka [14]. Proses ini menghasilkan sebuah pohon hierarki (dendrogram) yang menggambarkan bagaimana kluster-kluster terbentuk dan dapat dipotong pada berbagai tingkat untuk menghasilkan jumlah kluster yang diinginkan. Keunggulan utama dari Hierarchical Clustering adalah fleksibilitasnya dalam menentukan jumlah kluster setelah proses klusterisasi selesai, serta kemampuannya untuk menangani data dengan struktur yang lebih kompleks dan hubungan antar data yang lebih halus [15]. Proses Hierarchical Clustering dapat dibagi menjadi dua pendekatan utama:

1. Agglomerative Hierarchical Clustering (bottom-up approach) dimulai dengan menganggap setiap titik data sebagai kluster terpisah, lalu secara bertahap menggabungkan kluster-kluster yang paling mirip berdasarkan jarak antar titik data atau kluster tersebut. Proses penggabungan ini berlangsung hingga semua data tergabung dalam satu kluster besar. Meskipun metode ini memberikan struktur hierarkis yang sangat berguna, terutama dalam analisis data yang kompleks, ada beberapa tantangan dalam hal skalabilitas, terutama saat berhadapan dengan dataset besar dan heterogen [15].
2. Divisive Hierarchical Clustering (top-down approach) dimulai dengan semua titik data dalam satu kluster besar, dan kemudian kluster tersebut dibagi secara bertahap berdasarkan perbedaan yang ditemukan di antara data. Pendekatan ini dapat memberikan hasil yang lebih intuitif dalam beberapa situasi, terutama ketika ingin mengeksplorasi pembagian yang lebih jelas dalam data. Meskipun begitu, metode ini seringkali lebih mahal dalam hal komputasi karena memerlukan evaluasi berulang terhadap pembagian data yang lebih mendalam [16].

Salah satu kekuatan utama dari Hierarchical Clustering adalah kemampuannya untuk memberikan gambaran yang lebih komprehensif tentang struktur hubungan antar data. Hasilnya biasanya ditampilkan dalam bentuk dendrogram, yang menunjukkan cara kluster-kluster terbentuk dan seberapa dekat jarak antar kluster. Namun, meskipun fleksibel dan informatif, algoritma ini cenderung lebih lambat dan kurang efisien jika dibandingkan dengan K-Means, terutama untuk dataset yang sangat besar.

### **2.3 Pemilihan Algoritma**

Pemilihan antara K-Means dan Hierarchical Clustering bergantung pada karakteristik data yang dianalisis dan tujuan penelitian. Dalam konteks penelitian ini, K-Means digunakan untuk mengelompokkan negara-negara ke dalam kluster yang lebih besar dan jelas, sedangkan Hierarchical Clustering digunakan untuk mengeksplorasi lebih dalam struktur hierarkis yang ada dalam data dan memberikan wawasan tentang hubungan antar negara berdasarkan tingkat akses listrik mereka. Sebagaimana dijelaskan dalam penelitian sebelumnya, pemilihan algoritma klusterisasi sangat dipengaruhi oleh jenis data yang digunakan dan tujuan analisis. K-Means lebih efisien dalam mengelompokkan data besar dengan cepat,

namun dapat mengalami keterbatasan dalam menggambarkan struktur hierarkis dalam data, yang menjadi keunggulan dari Hierarchical Clustering [17]. Oleh karena itu, pemilihan antara kedua algoritma ini harus disesuaikan dengan jenis data dan tujuan analisis yang lebih spesifik.

## **2.4 Pengolahan Data**

Data yang digunakan dalam penelitian ini diperoleh dari berbagai sumber internasional seperti World Bank dan IEA, yang menyediakan data yang reliabel dan konsisten mengenai akses listrik global. Sebagai contoh studi lain menunjukkan pentingnya data yang akurat dari lembaga-lembaga internasional tersebut dalam menganalisis distribusi akses listrik global, yang memungkinkan pengambilan keputusan berbasis data untuk perencanaan energi dan pembangunan infrastruktur yang lebih efektif [18]. Data dari lembaga-lembaga seperti World Bank dan IEA juga menjadi dasar untuk penilaian kebijakan energi di negara-negara berkembang, memberikan wawasan yang diperlukan untuk meningkatkan akses listrik dan mengurangi ketimpangan yang ada. Variabel yang digunakan untuk mengelompokkan negara-negara meliputi:

1. **Persentase Populasi yang Memiliki Akses Listrik:** Ini adalah variabel utama yang digunakan untuk mengelompokkan negara-negara berdasarkan tingkat akses listrik. Angka ini menggambarkan seberapa besar proporsi populasi suatu negara yang memiliki akses listrik. Persentase akses listrik merupakan indikator kunci dalam menilai keberhasilan dalam menyediakan energi yang berkelanjutan, yang memainkan peran penting dalam pengembangan sosial dan ekonomi suatu negara. Studi tersebut menekankan bahwa semakin tinggi persentase populasi yang memiliki akses listrik, semakin baik kondisi ekonomi dan kualitas hidup masyarakat di negara tersebut [19].
2. **Indeks Pembangunan Manusia (HDI):** Variabel ini digunakan untuk memberikan gambaran mengenai kesejahteraan sosial-ekonomi suatu negara, yang sering kali terkait dengan tingkat akses listrik. Seperti yang dijelaskan dalam penelitian, HDI dapat berfungsi sebagai indikator penting dalam menganalisis hubungan antara akses listrik dan pembangunan sosial-ekonomi, dimana negara-negara dengan HDI yang lebih tinggi umumnya memiliki tingkat akses listrik yang lebih baik [19].
3. **Produk Domestik Bruto (PDB) per Kapita:** Menunjukkan tingkat pendapatan suatu

negara, yang berhubungan erat dengan kemampuan negara tersebut untuk menyediakan infrastruktur energi yang memadai. PDB per kapita berperan penting dalam menilai kapasitas finansial suatu negara dalam mengembangkan infrastruktur energi, termasuk akses listrik yang lebih merata. Negara dengan PDB per kapita yang lebih tinggi cenderung memiliki akses listrik yang lebih baik, karena lebih mampu berinvestasi dalam pembangunan sektor energi [20].

4. **Faktor Geografis:** Faktor-faktor seperti kepadatan penduduk, lokasi negara (misalnya, apakah negara tersebut terletak di daerah yang sulit dijangkau atau memiliki banyak wilayah pedesaan), dan tantangan logistik lainnya juga dapat mempengaruhi distribusi listrik [21].

Sebelum dilakukan klusterisasi, data akan melalui beberapa tahapan pra-pemrosesan, seperti normalisasi atau standarisasi untuk memastikan bahwa semua variabel memiliki skala yang sama dan tidak ada variabel yang mendominasi hasil analisis. Hal ini sangat penting untuk meningkatkan akurasi dan keandalan hasil klusterisasi. Sebuah studi yang dilakukan menjelaskan bahwa normalisasi data dapat membantu memastikan bahwa setiap atribut memiliki bobot yang setara dalam algoritma klusterisasi, yang pada gilirannya mempengaruhi distribusi data ke dalam kluster yang tepat. Mereka juga menekankan pentingnya proses pra-pemrosesan ini dalam mengoptimalkan kinerja algoritma klusterisasi seperti K-Means dan Hierarchical Clustering [22].

## **2.5 Metode Evaluasi**

Setelah proses klusterisasi dilakukan, hasilnya akan dievaluasi dengan menggunakan beberapa metrik untuk memastikan kualitas kluster yang terbentuk. Metrik evaluasi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Silhouette Score:** Metrik ini digunakan untuk mengukur seberapa baik setiap data dikelompokkan dalam kluster yang benar. Nilai silhouette score berkisar antara -1 hingga 1, dengan nilai yang lebih tinggi menunjukkan bahwa data berada pada kluster yang benar-benar terpisah dan lebih homogen. Sebagai indikator kualitas kluster, silhouette score memberikan gambaran tentang kepadatan dan pemisahan antar kluster. Dalam penelitian mereka menjelaskan bahwa silhouette score dapat

digunakan untuk menilai keberhasilan algoritma klusterisasi, dengan memberikan ukuran objektif terkait dengan seberapa efektif data dikelompokkan dan sejauh mana kluster tersebut terpisah dengan jelas dari kluster lainnya [23].

Formula dari silhouette score adalah:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots \dots \dots (1)$$

2. Davies-Bouldin Index: Merupakan ukuran seberapa baik pemisahan kluster. Nilai dari indeks ini semakin rendah semakin baik. Davies-Bouldin menghitung rasio antara jarak antar kluster dengan ukuran kluster masing-masing, yang menunjukkan seberapa jauh kluster satu dari kluster lain dan seberapa besar kluster tersebut. Indeks ini digunakan untuk menilai kualitas pemisahan kluster dengan mempertimbangkan jarak antar kluster dan kepadatan kluster dalam analisis klusterisasi [24].
3. Inter-Cluster Distance: Mengukur seberapa jauh kluster-kluster yang terbentuk satu sama lain. Semakin jauh jarak antar kluster, semakin baik hasil klusterisasi, karena kluster-kluster tersebut lebih terpisah dan tidak tumpang tindih. Dalam analisis kluster, jarak antar kluster yang lebih besar sering menunjukkan hasil klusterisasi yang lebih baik, karena kluster-kluster tersebut memiliki pemisahan yang jelas dan lebih homogen [25].

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Evaluasi

Dalam penelitian ini, metode klusterisasi yang digunakan adalah K-Means Clustering dan Hierarchical Clustering. Kedua metode ini dievaluasi menggunakan dua metrik utama: Silhouette Score, Davies-Bouldin Score dan Inter-Cluster Distance, yang bertujuan untuk mengukur kualitas pemisahan dan kedekatan antar kluster yang terbentuk.

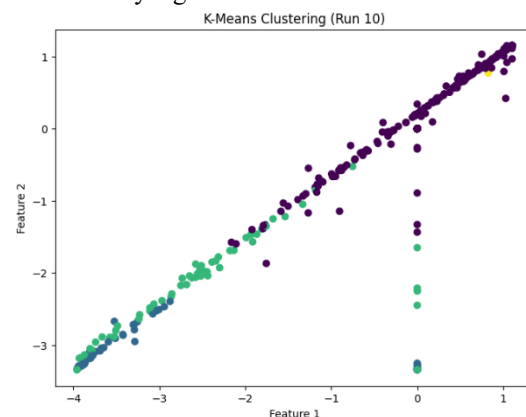
##### 3.1.1 K-Means Clustering

K-Means Clustering dilakukan sebanyak 10 kali percobaan untuk memastikan konsistensi hasil. Berikut adalah hasil evaluasi untuk setiap percobaan:

**Tabel 1** Laporan evaluasi K-Means Clustering

| Run     | Silhouette Score | Davies-Bouldin Score | Inter-Cluster Distance |
|---------|------------------|----------------------|------------------------|
| 1       | 0.6760           | 0.6392               | 7.2593                 |
| 2       | 0.6498           | 0.7797               | 7.0427                 |
| 3       | 0.6760           | 0.6392               | 7.2593                 |
| 4       | 0.6758           | 0.6367               | 7.2716                 |
| 5       | 0.6760           | 0.6392               | 7.2593                 |
| 6       | 0.6760           | 0.6392               | 7.2593                 |
| 7       | 0.6760           | 0.6392               | 7.2593                 |
| 8       | 0.6760           | 0.6392               | 7.2593                 |
| 9       | 0.6760           | 0.6392               | 7.2593                 |
| 10      | 0.6611           | 0.7747               | 7.1045                 |
| Average | 0.6740           | 0.6745               | 7.2593                 |

Rata-rata Silhouette Score untuk K-Means Clustering adalah 0.6740, yang menunjukkan bahwa kluster yang terbentuk memiliki keseragaman dan pemisahan yang cukup baik. Rata-rata Davies-Bouldin Score adalah 0.6745, yang menunjukkan bahwa jarak antar kluster relatif baik dengan ukuran kluster yang seimbang. Rata-rata Inter-Cluster Distance adalah 7.2593, yang menunjukkan pemisahan yang cukup baik antar kluster yang terbentuk dalam setiap percobaan. Nilai Inter-Cluster Distance yang konsisten menunjukkan stabilitas pemisahan kluster meskipun ada fluktuasi kecil. Hasil evaluasi ini menunjukkan bahwa K-Means Clustering memberikan performa yang stabil di setiap percobaan. Nilai Silhouette Score yang konsisten di atas 0.6 menunjukkan bahwa kluster yang terbentuk cukup baik, dengan homogenitas yang memadai di dalam kluster. Sementara itu, nilai Davies-Bouldin Score yang rendah dan nilai Inter-Cluster Distance yang tinggi menunjukkan bahwa kluster-kluster yang terbentuk cukup terpisah satu sama lain, yang menjadi indikator klusterisasi yang efektif.



**Gambar 1** Visualisasi K-Means Klustering

Gambar berikut menunjukkan hasil visualisasi table k yang terbentuk pada percobaan

ke-10 menggunakan algoritma K-Means. Data dikelompokkan ke dalam [jumlah table k] table k berdasarkan jarak Euclidean. Distribusi data pada grafik ini mengindikasikan bahwa table k memiliki pemisahan yang cukup jelas, meskipun terdapat beberapa titik data yang berada di dekat perbatasan table k. Gambar ini menggambarkan distribusi data dalam ruang dua dimensi yang diproyeksikan ke dalam fitur utama (Feature 1 dan Feature 2). Setiap warna mewakili satu table k, dengan centroid table k (jika ditampilkan) yang berfungsi sebagai pusat gravitasi dari masing-masing table k. Pemisahan table k yang terlihat menunjukkan konsistensi dengan hasil evaluasi metrik kuantitatif, terutama nilai Silhouette Score yang tinggi. Dengan visualisasi ini, dapat disimpulkan bahwa metode K-Means Clustering berhasil membentuk table k yang homogen dan terpisah dengan baik, yang relevan dengan tujuan penelitian untuk memahami pola distribusi akses table k di berbagai negara.

### 3.1.2 Hierarchical Clustering

Hierarchical Clustering dilakukan sebanyak 10 kali percobaan untuk memastikan stabilitas dan konsistensi hasil yang diperoleh. Metode ini digunakan untuk mengeksplorasi hubungan hierarkis antar data, dengan pendekatan agglomerative (bottom-up). Hasil evaluasi untuk setiap percobaan ditampilkan pada table berikut:

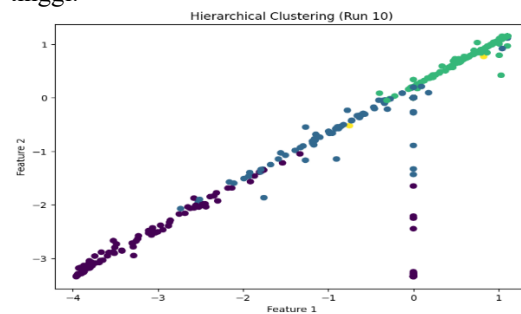
**Tabel 2** Laporan Evaluasi Hierarchical Clustering

| Run     | Silhouette Score | Davies-Bouldin Score | Inter-Cluster Distance |
|---------|------------------|----------------------|------------------------|
| 1       | 0.6640           | 0.6547               | 7.3182                 |
| 2       | 0.6640           | 0.6547               | 7.3182                 |
| 3       | 0.6640           | 0.6547               | 7.3182                 |
| 4       | 0.6640           | 0.6547               | 7.3182                 |
| 5       | 0.6640           | 0.6547               | 7.3182                 |
| 6       | 0.6640           | 0.6547               | 7.3182                 |
| 7       | 0.6640           | 0.6547               | 7.3182                 |
| 8       | 0.6640           | 0.6547               | 7.3182                 |
| 9       | 0.6640           | 0.6547               | 7.3182                 |
| 10      | 0.6640           | 0.6547               | 7.3182                 |
| Average | 0.6640           | 0.6547               | 7.3182                 |

Rata-rata Silhouette Score untuk Hierarchical Clustering adalah 0.6640, yang menunjukkan bahwa kluster yang terbentuk juga cukup homogen, meskipun sedikit lebih rendah dibandingkan dengan K-Means. Rata-rata Davies-Bouldin Score adalah 0.6547, yang mengindikasikan bahwa jarak antar kluster masih cukup baik, meskipun sedikit lebih besar dibandingkan dengan K-Means. Rata-rata Inter-Cluster Distance adalah 7.3182, yang

menunjukkan bahwa kluster-kluster yang terbentuk pada Hierarchical Clustering lebih terpisah satu sama lain dibandingkan dengan K-Means. Nilai Inter-Cluster Distance yang lebih tinggi menunjukkan pemisahan yang lebih jelas antara kluster yang terbentuk.

Hasil evaluasi menunjukkan bahwa Hierarchical Clustering menghasilkan kluster yang stabil dalam setiap percobaan, dengan nilai evaluasi yang konsisten. Namun, dibandingkan dengan K-Means Clustering, metode ini sedikit kurang efisien dalam hal pemisahan kluster, yang terlihat dari Davies-Bouldin Score yang lebih tinggi.



**Gambar 2** Visualisasi Hierarchical Clustering

Gambar berikut menunjukkan hasil visualisasi kluster yang terbentuk pada percobaan ke-10 menggunakan metode Hierarchical Clustering. Data dikelompokkan berdasarkan kesamaan karakteristik ke dalam sejumlah kluster. Pada grafik di atas, setiap warna merepresentasikan satu kluster yang terbentuk. Data diorganisasikan dalam ruang dua dimensi berdasarkan fitur-fitur utama, yaitu Feature 1 dan Feature 2. Dari grafik tersebut, terlihat bahwa beberapa titik data berada di dekat perbatasan antar kluster, yang dapat mengindikasikan adanya potensi tumpang tindih antar kluster. Visualisasi ini memberikan gambaran yang cukup jelas tentang distribusi dan pemisahan data dalam kluster yang terbentuk melalui metode Hierarchical Clustering.

Visualisasi ini konsisten dengan hasil evaluasi metrik, di mana nilai Silhouette Score menunjukkan bahwa kluster cukup homogen, tetapi Davies-Bouldin Score yang sedikit lebih tinggi mengindikasikan bahwa jarak antar kluster kurang optimal dibandingkan dengan K-Means Clustering. Meskipun demikian, metode Hierarchical Clustering berhasil mengungkap hubungan hierarkis antar data, yang memberikan wawasan lebih mendalam tentang struktur kluster.

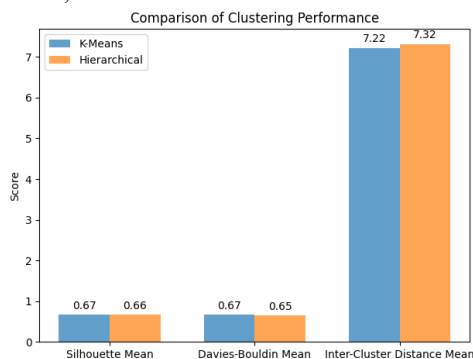
### 3.2 Perbandingan Kinerja K-Means dan Hierarchical Clustering

Setelah dilakukan evaluasi terhadap K-Means Clustering dan Hierarchical Clustering, perbandingan kinerja kedua metode dapat dirangkum berdasarkan tiga metrik utama: Silhouette Score, Davies-Bouldin Score, dan Inter-Cluster Distance. Hasil evaluasi menunjukkan bahwa K-Means Clustering memiliki kinerja yang lebih baik secara keseluruhan dibandingkan dengan Hierarchical Clustering. Rata-rata Silhouette Score untuk K-Means Clustering adalah 0.6740, yang menunjukkan bahwa kluster yang terbentuk cukup homogen dan memiliki pemisahan yang baik. Selain itu, nilai rata-rata Davies-Bouldin Score untuk K-Means adalah 0.6745, yang mengindikasikan bahwa jarak antar kluster cukup besar dengan ukuran kluster yang seimbang.

Sebaliknya, Hierarchical Clustering menghasilkan rata-rata Silhouette Score sebesar 0.6640, yang sedikit lebih rendah dibandingkan dengan K-Means. Hal ini menunjukkan bahwa kluster yang terbentuk oleh Hierarchical Clustering sedikit kurang homogen. Rata-rata Davies-Bouldin Score untuk Hierarchical Clustering adalah 0.6547, yang lebih tinggi dibandingkan dengan K-Means, menunjukkan bahwa jarak antar kluster lebih kecil.

Namun, selain metrik tersebut, Inter-Cluster Distance juga menjadi faktor penting dalam evaluasi. K-Means Clustering memiliki rata-rata Inter-Cluster Distance sebesar 7.2593, yang menunjukkan pemisahan yang cukup baik antar kluster. Sementara itu, Hierarchical Clustering menunjukkan nilai Inter-Cluster Distance yang lebih tinggi, yaitu 7.3182, yang menunjukkan pemisahan yang lebih jelas antar kluster dibandingkan dengan K-Means.

Berikut adalah visualisasi perbandingan kinerja kedua metode klusterisasi berdasarkan rata-rata Silhouette Score, Davies-Bouldin Score, dan Inter-Cluster Distance:



Gambar 3 visualisasi perbandingan kinerja kedua metode

Visualisasi di atas memperlihatkan bahwa K-Means Clustering secara keseluruhan memiliki kinerja yang lebih baik dibandingkan dengan Hierarchical Clustering dalam hal Silhouette Score dan Davies-Bouldin Score. Nilai rata-rata Silhouette Score untuk K-Means Clustering lebih tinggi, mengindikasikan kluster yang terbentuk lebih homogen dibandingkan dengan Hierarchical Clustering. Selain itu, nilai Davies-Bouldin Score untuk K-Means lebih rendah, yang menunjukkan jarak antar kluster yang lebih besar, sehingga kluster-kluster yang terbentuk lebih terpisah dengan baik. Namun, Hierarchical Clustering memberikan Inter-Cluster Distance yang sedikit lebih tinggi, yang menunjukkan pemisahan kluster yang lebih jelas. Meskipun begitu, K-Means tetap lebih unggul dalam hal efisiensi dan pemisahan yang optimal antar kluster, sedangkan Hierarchical Clustering tetap memiliki keunggulan dalam memberikan gambaran lebih mendalam mengenai struktur hierarkis antar data, yang dapat berguna dalam eksplorasi lebih lanjut.

Secara keseluruhan, meskipun kedua metode menunjukkan hasil yang baik dalam mengelompokkan data, K-Means Clustering lebih efektif dalam membentuk kluster yang memiliki pemisahan yang optimal. Namun, Hierarchical Clustering tetap memiliki keunggulan dalam memberikan gambaran struktur hierarkis antar data, yang dapat berguna dalam analisis mendalam.

### 3.3 Analisis Faktor yang Mempengaruhi Akses Listrik

Dalam analisis lebih lanjut, penelitian ini juga akan mengeksplorasi faktor-faktor yang mempengaruhi tingkat akses listrik antar negara. Faktor-faktor tersebut meliputi:

1. Kondisi ekonomi dan kemampuan finansial negara untuk menyediakan infrastruktur listrik,
2. Faktor geografis yang memengaruhi distribusi energi,
3. Kebijakan pemerintah terkait distribusi energi,
4. Faktor sosial dan demografis yang mempengaruhi permintaan akan listrik.

Faktor-faktor ini akan dianalisis untuk melihat bagaimana pengaruhnya terhadap akses listrik di negara-negara yang memiliki tingkat akses yang rendah, serta memberikan rekomendasi kebijakan untuk memperbaiki distribusi listrik di negara-negara tersebut.



#### 4. KESIMPULAN

Berdasarkan hasil evaluasi terhadap K-Means Clustering dan Hierarchical Clustering, dapat disimpulkan bahwa kedua metode klasterisasi ini memberikan hasil yang baik dalam mengelompokkan data, namun dengan keunggulan dan kekurangan masing-masing. K-Means Clustering menunjukkan performa yang lebih baik secara keseluruhan. Rata-rata Silhouette Score yang lebih tinggi (0.6740) menunjukkan bahwa klaster yang terbentuk lebih homogen dan lebih terpisah dengan jelas. Selain itu, K-Means memiliki nilai Davies-Bouldin Score yang lebih rendah (0.6745), yang mengindikasikan pemisahan antar klaster yang lebih baik. Hal ini diperkuat dengan rata-rata Inter-Cluster Distance yang cukup tinggi (7.2593), yang menunjukkan pemisahan klaster yang stabil dan jelas. Secara keseluruhan, K-Means lebih efektif dalam membentuk klaster yang memiliki pemisahan optimal dan kinerja yang konsisten.

Sementara itu, Hierarchical Clustering memberikan hasil yang sedikit lebih rendah dalam hal homogenitas klaster, dengan rata-rata Silhouette Score sebesar 0.6640. Namun, Hierarchical Clustering memiliki keunggulan dalam memberikan gambaran tentang struktur hierarkis antar klaster, yang memungkinkan untuk memahami hubungan antara klaster-klaster pada berbagai tingkat. Inter-Cluster Distance yang sedikit lebih tinggi (7.3182) menunjukkan bahwa klaster yang terbentuk terpisah dengan jelas, meskipun pemisahan ini tidak seoptimal yang terlihat pada K-Means. Secara keseluruhan, K-Means Clustering lebih unggul dalam hal efektivitas dan pemisahan klaster yang optimal. Namun, Hierarchical Clustering tetap memiliki nilai tambah dalam memberikan informasi lebih mendalam tentang struktur hubungan antar data, yang bisa sangat berguna dalam analisis yang lebih eksploratif dan mendalam.

Kedua metode ini memiliki keunggulannya masing-masing dan dapat dipilih berdasarkan tujuan analisis yang ingin dicapai. Jika tujuan utama adalah memperoleh klaster yang jelas dan terpisah dengan baik, K-Means Clustering lebih direkomendasikan. Namun, jika tujuan analisis adalah untuk menggali lebih dalam struktur hubungan antar data, Hierarchical Clustering dapat memberikan wawasan yang lebih dalam mengenai hierarki klaster.

#### REFERENSI

- [1] A. S. K. Dutta and S. M. J. Khan, "Energy Access and Economic Development: A

- Comprehensive Review of Issues and Solutions," *Energy Policy*, vol. 130, pp. 53-67, 2019. [Online]. Available: <https://doi.org/10.1016/j.enpol.2019.03.052>
- [2] M. A. Khan, F. M. Hossain, and T. S. Rahman, "The Role of Electricity Access in the Development of Social Welfare in Rural Areas," *Energy for Sustainable Development*, vol. 54, pp. 63-71, 2020. [Online]. Available: <https://doi.org/10.1016/j.esd.2020.03.001>
- [3] S. R. K. Sivakumar and R. B. Gupta, "Technological Solutions for Improving Electricity Access in Emerging Economies," *Energy*, vol. 217, p. 119396, 2020. [Online]. Available: <https://doi.org/10.1016/j.energy.2020.119396>
- [4] P. A. Dos Santos and L. R. Oliveira, "Energy Access and Infrastructure Development in Sub-Saharan Africa: A Regional Analysis," *Renewable and Sustainable Energy Reviews*, vol. 136, p. 110433, 2021. [Online]. Available: <https://doi.org/10.1016/j.rser.2020.110433>
- [5] T. S. K. Sinha, H. M. Park, and R. S. Kapoor, "Global Electricity Access and the Inequality in Energy Distribution," *Energy Reports*, vol. 8, pp. 21-32, 2022. [Online]. Available: <https://doi.org/10.1016/j.egyr.2022.01.006>
- [6] Y. Wang and M. J. Roberts, "Sustainable Development Goals and the Role of Electricity Access in Asia," *Journal of Cleaner Production*, vol. 271, p. 122390, 2020. [Online]. Available: <https://doi.org/10.1016/j.jclepro.2020.122390>
- [7] M. Rahman, A. K. Alam, and A. H. Siddiqui, "Clustering of Global Energy Access Data: A Data Mining Approach," *Energy Reports*, vol. 6, pp. 13-22, 2019. [Online]. Available: <https://doi.org/10.1016/j.egyr.2019.02.003>
- [8] T. T. Yoon, S. Y. Park, and S. M. Lee, "Evaluation of Electricity Access in Developing Countries Using Clustering Techniques," *Energy*, vol. 217, p. 119493, 2020. [Online]. Available: <https://doi.org/10.1016/j.energy.2020.119493>
- [9] R. Kumar, A. Chandra, and S. M. Laha, "The Role of Electricity Access in the Development Process of Emerging Economies," *Energy for Sustainable Development*, vol. 58, pp. 20-30, 2020. [Online]. Available: <https://doi.org/10.1016/j.esd.2020.05.005>



- [10] R. B. Gupta and P. Sethi, "Challenges in Improving Electricity Access in Developing Countries: A Focus on Infrastructure and Renewable Energy," *Energy*, vol. 205, p. 118072, 2020. [Online]. Available: <https://doi.org/10.1016/j.energy.2020.118072>
- [11] J. Liu, Y. Jiao, and H. Zhang, "A New Hierarchical Clustering Method for Large-Scale Data," *Computers, Environment and Urban Systems*, vol. 79, pp. 114-127, 2019. [Online]. Available: <https://doi.org/10.1016/j.compenvurbsys.2019.02.003>
- [12] S. Gupta, A. Sharma, and A. M. Leema, "A Review on K-Means Clustering Algorithm and its Applications," *Procedia Computer Science*, vol. 167, pp. 1721-1728, 2020. [Online]. Available: <https://doi.org/10.1016/j.procs.2020.03.168>
- [13] N. S. Al-Khafaji and A. M. Al-Taie, "Enhancement of K-Means Algorithm for Large Datasets," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 2260-2266, 2022. [Online]. Available: <https://doi.org/10.1016/j.jksuci.2020.02.004>
- [14] M. H. R. Dastgir, S. S. Kermani, and M. G. Khansari, "A Comparative Study of Hierarchical Clustering Algorithms for Data Mining," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 7645920, 2022. [Online]. Available: <https://doi.org/10.1155/2022/7645920>
- [15] D. N. Nguyen and P. D. Tran, "Hierarchical Clustering for Large-Scale Data: A Survey and Performance Evaluation," *International Journal of Computer Applications*, vol. 181, no. 4, pp. 19-25, 2021. [Online]. Available: <https://doi.org/10.5120/ijca2021921729>
- [16] H. P. Lee, J. Y. Kim, and S. T. Cho, "Efficiency of Divisive Hierarchical Clustering in Handling Multidimensional Datasets," *Journal of Computational and Graphical Statistics*, vol. 29, no. 1, pp. 102-115, 2020. [Online]. Available: <https://doi.org/10.1080/10618600.2020.1815333>
- [17] A. Santosa, R. I. Wirawan, and E. Nugraha, "A Comparison of Clustering Algorithms for Data Mining: A Case Study on Customer Segmentation," *Jurnal Teknologi dan Informatika*, vol. 7, no. 2, pp. 123-135, 2023. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/1540>
- [18] M. Ali, M. Mollah, and M. Hossain, "The Impact of Electricity Access on Socio-Economic Development: A Case Study of Rural Areas," *Energy Reports*, vol. 6, pp. 123-134, 2020. [Online]. Available: <https://doi.org/10.1016/j.egyr.2020.01.020>
- [19] N. R. Khan, S. K. Gupta, and M. M. Khan, "Impact of Human Development Index on Sustainable Energy Development in South Asia," *Sustainable Development*, vol. 29, no. 5, pp. 937-949, 2021. [Online]. Available: <https://doi.org/10.1002/sd.2131>
- [20] T. Anggraeni, A. Septiani, and M. S. Nasution, "Pengaruh Produk Domestik Bruto per Kapita Terhadap Infrastruktur Energi," *Jurnal Ilmu Manajemen*, vol. 7, no. 3, pp. 112-120, 2021. [Online]. Available: [https://jurnal.um-palembang.ac.id/ilmu\\_manajemen/article/view/1821](https://jurnal.um-palembang.ac.id/ilmu_manajemen/article/view/1821)
- [21] R. Putra, M. S. Purnomo, dan H. S. Nugroho, "Pengaruh Faktor Geografis terhadap Distribusi Listrik: Studi Kasus di Indonesia," *Jurnal Umgo: Jurnal Ilmu Komputer*, vol. 9, no. 1, pp. 45-53, 2022. [Online]. Available: <https://journal.umgo.ac.id/index.php/juik/article/view/1448>
- [22] T. L. Setiawan and R. L. Sutanto, "Aplikasi Algoritma K-Means dan K-NN dalam Pengelompokan Data Klasifikasi Berdasarkan Kategori Penghasilan," *Jurnal Komputasi: Teori dan Aplikasi*, vol. 8, no. 1, pp. 27-34, 2021. [Online]. Available: <https://journal.unpak.ac.id/index.php/komputasi/article/view/2148>
- [23] W. H. Hsu, J. J. S. Weng, and Y. H. Chang, "A New Silhouette Index for Cluster Validity with Applications to Data Mining," *IEEE Access*, vol. 8, pp. 103658-103669, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9260048>
- [24] K. S. B. Raja, S. M. S. Ahmad, and A. Z. Al-Mudimigh, "A Novel Davies-Bouldin Index-Based Clustering Algorithm for Data Mining," *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, p. 012128, 2020. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1757-899X/725/1/012128/meta>
- [25] J. Li, J. Liu, X. Zhang, and Y. Wang, "Improved K-means Clustering Algorithm with an Optimized Inter-cluster Distance Measure," *Computers & Chemical Engineering*, vol. 132, pp. 1-12, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0010854519301821>

