

# CLUSTERING NEGARA BERDASARKAN GDP MENGGUNAKAN ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING

**Dosen Pengampu : Dr. Atik Wintarti, M.Kom.**

**Aufatir Diaul Haq (23031554127)**

Program Studi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya

**Muhammad Taufiqulhakim (23031554111)**

Program Studi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya

## Abstrak

Dalam era globalisasi, pengelompokan negara berdasarkan indikator ekonomi seperti GDP menjadi penting untuk memahami pola perkembangan ekonomi global. Proyek ini bertujuan untuk mengelompokkan negara-negara berdasarkan GDP dan indikator ekonomi lainnya menggunakan algoritma K-Means dan Hierarchical Clustering. Penelitian ini dilakukan untuk memberikan wawasan tentang hubungan ekonomi antarnegara, serta mengevaluasi keunggulan dan kelemahan masing-masing algoritma clustering dalam menganalisis data skala besar. Hasil pengelompokan menunjukkan pembentukan klaster negara dengan karakteristik ekonomi serupa, yang dapat dimanfaatkan untuk pengambilan keputusan strategis oleh pemerintah, ekonom, dan investor. Selain itu, evaluasi performa algoritma memberikan panduan praktis bagi implementasi metode clustering di masa depan.

**Kata Kunci :** Clustering, GDP, K-Means, Hierarchical Clustering, Analisis Ekonomi, Data Skala Besar

## I. PENDAHULUAN

Dalam era globalisasi, pengelompokan negara berdasarkan indikator ekonomi seperti GDP (Gross Domestic Product) menjadi sangat penting untuk memahami pola perkembangan ekonomi di seluruh dunia. Indikator ekonomi seperti GDP mencerminkan tingkat kesejahteraan, produktivitas, dan pertumbuhan suatu negara, sehingga analisis data ini menjadi relevan bagi berbagai pihak, termasuk pemerintah, ekonom, dan investor. Clustering adalah salah satu metode analisis data yang memungkinkan pengelompokan objek berdasarkan kesamaan karakteristik tertentu.

Proyek ini bertujuan untuk mengelompokkan negara-negara berdasarkan GDP dan indikator tambahan lainnya menggunakan dua algoritma clustering yang populer, yaitu K-Means dan Hierarchical Clustering. Kedua algoritma ini dipilih karena memiliki pendekatan yang berbeda

dalam mengelompokkan data. Hasil dari penelitian ini diharapkan dapat memberikan informasi tentang pola ekonomi global yang mendalam dan membantu berbagai pihak dalam pengambilan keputusan strategis. Selain itu, penelitian ini juga akan mengevaluasi kelebihan dan kekurangan dari kedua metode clustering dalam menganalisis data ekonomi skala besar.

## II. TUJUAN

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengelompokkan negara-negara berdasarkan GDP dan indikator ekonomi lainnya untuk mengidentifikasi pola perkembangan ekonomi global.
2. Mengevaluasi kelebihan dan kekurangan algoritma K-Means dan Hierarchical

Clustering dalam pengelompokan data ekonomi skala besar.

3. Memberikan wawasan tentang kelompok negara dengan karakteristik ekonomi serupa yang dapat dimanfaatkan oleh pemerintah, ekonom, dan investor.
4. Mengembangkan panduan untuk implementasi algoritma clustering dalam analisis data ekonomi.

### III. METODE

#### 3.1. Pembersihan Data

Tahapan ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis bebas dari anomali, kesalahan, atau data yang tidak relevan. Proses pembersihan melibatkan penanganan nilai yang hilang, duplikasi data, dan penyesuaian format data agar sesuai untuk analisis. Hasil akhir dari tahap ini adalah data yang bersih dan siap digunakan untuk proses lebih lanjut.

#### 3.2. Inisialisasi Fungsi

Sebelum melakukan analisis clustering, beberapa fungsi penting diinisialisasi untuk mendukung perhitungan dan pengelompokan data. Fungsi-fungsi yang dibuat meliputi:

- **Fungsi Euclidean Distance:** Menghitung jarak antara dua titik dalam ruang multi-dimensi, yang menjadi dasar untuk menentukan kesamaan atau perbedaan antar data.
- **Fungsi K-Means:** Mengimplementasikan algoritma K-Means untuk mengelompokkan data ke dalam beberapa cluster berdasarkan nilai centroids.
- **Fungsi WCSS (Within-Cluster Sum of Squares):** Mengukur total jarak kuadrat data dari pusat clusternya, yang digunakan untuk menentukan optimalitas jumlah cluster.
- **Fungsi Elbow Method:** Menggunakan hasil WCSS untuk mengidentifikasi jumlah cluster optimal melalui grafik WCSS versus jumlah cluster.

#### 3.3. Normalisasi Data

Data dinormalisasi menggunakan *Standard Scaler* agar setiap variabel memiliki skala yang sama, dengan mean nol dan standar deviasi satu. Langkah ini penting untuk mencegah dominasi variabel tertentu dalam proses clustering, sehingga menghasilkan pengelompokan yang lebih akurat.

#### 3.4. Metode Clustering

Dua pendekatan clustering diterapkan dalam penelitian ini untuk mengelompokkan data:

##### a. K-Means Clustering

K-Means Clustering adalah salah satu metode *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam  $k$  cluster berdasarkan kesamaan atau kedekatan antar data. Algoritma ini berusaha meminimalkan variasi dalam cluster (Within-Cluster Sum of Squares / Within-Cluster Sum of Squares, WCSS) dan memaksimalkan perbedaan antar cluster.

**Penentuan Jumlah Cluster Optimal:** Dilakukan menggunakan metode Elbow, yang menghitung WCSS untuk setiap jumlah cluster. Jumlah cluster optimal ditentukan dari titik "siku" pada grafik Elbow.

**Pengelompokan Data:** Setelah jumlah cluster optimal ditemukan, data dikelompokkan menggunakan algoritma K-Means, yang menghasilkan label cluster untuk setiap data.

##### b. Hierarchical Clustering

Hierarchical Clustering adalah metode pengelompokan data yang membangun hubungan hierarkis antar data dalam bentuk struktur pohon yang disebut *dendrogram*. Metode ini tidak memerlukan penentuan jumlah cluster di awal proses, karena jumlah cluster dapat ditentukan dengan memotong dendrogram pada tingkat tertentu.

Dalam penelitian ini, tiga metode linkage diterapkan:

- 1) **Complete Linkage:** Mengukur jarak maksimum antara elemen dari dua cluster. Cocok untuk data yang ingin memastikan elemen dalam cluster memiliki hubungan erat.

- 2) **Single Linkage:** Mengukur jarak minimum antara elemen dari dua cluster. Biasanya menghasilkan cluster yang lebih panjang dan ramping.
- 3) **Average Linkage:** Mengukur rata-rata jarak antara elemen dari dua cluster. Metode ini memberikan keseimbangan antara *complete* dan *single linkage*.

Setelah dendrogram dibuat, jumlah cluster ditentukan dengan memotong dendrogram pada tingkat tertentu berdasarkan jarak antar cluster (*threshold*). Misalnya:

- Pada *Complete Linkage*, cluster dibagi sedemikian rupa sehingga elemen dalam cluster memiliki jarak maksimum tertentu.
- Pada *Single Linkage*, cluster dibentuk dengan menjaga jarak minimum antara anggota cluster.
- Pada *Average Linkage*, cluster dipilih berdasarkan jarak rata-rata antar anggota.

### 3.5. Evaluasi dan Visualisasi

Tahapan ini digunakan untuk mengevaluasi dan memvisualisasikan hasil clustering:

- **Grafik Elbow Method:** Digunakan untuk menentukan jumlah cluster optimal pada K-Means Clustering.
- **Scatter Plot:** Memvisualisasikan hasil clustering dari algoritma K-Means, menunjukkan distribusi data pada masing-masing cluster.
- **Dendrogram:** Digunakan untuk menampilkan hasil Hierarchical Clustering, memberikan gambaran hubungan antar data secara hierarkis.

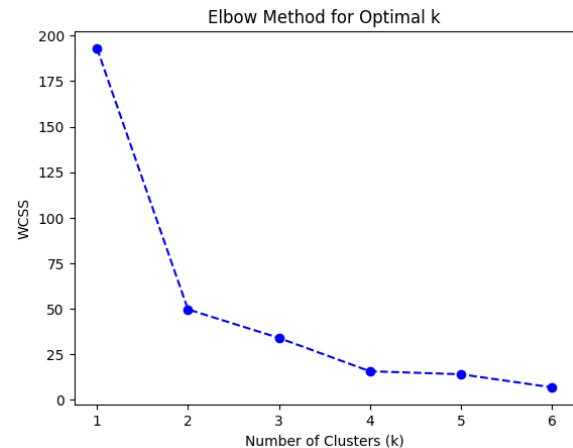
## IV. HASIL DAN DISKUSI

### 4.1. K – Means (Tidak Termasuk China dan Amerika Serikat)

#### 4.1.1. Elbow Method Execution

```
Running Elbow Method...
Converged after 2 iterations.
Converged after 5 iterations.
Converged after 9 iterations.
Converged after 14 iterations.
Converged after 17 iterations.
Converged after 20 iterations.
Elbow Method executed in 0.0588 seconds
WCSS Values: [193.0, 49.692216880429134, 34.02956814788351, 15.717374996982404, 14.07506542891452, 6.989478619227872]
```

**Waktu Eksekusi:** Elbow Method memakan waktu 0.0508 detik dengan 6 iterasi untuk menentukan jumlah cluster optimal.



**Hasil Elbow Method:** Menunjukkan hubungan antara jumlah cluster (kk) dan nilai WCSS (Within-Cluster Sum of Squares). Pada grafik, terlihat penurunan WCSS yang signifikan pada awalnya, dan perlambatan terjadi di sekitar  $k=2$ . Ini menunjukkan bahwa jumlah cluster optimal adalah 2, di mana "titik siku" (*elbow point*) berada.

#### 4.1.2. K-Means Execution

**Waktu Eksekusi:** K-Means memakan waktu **0.0014** detik dengan  $k=2$ , yang menunjukkan bahwa algoritma clustering dijalankan dengan cepat pada dataset yang diberikan.

#### 4.1.3. Hasil Cluster K-Means



**Cluster 1:** Semua negara selain yang termasuk Cluster 2

**Cluster 2:** ['Australia', 'Brazil', 'Canada', 'France', 'Germany', 'India', 'Italy', 'Japan', 'Mexico', 'Russia', 'South Korea', 'Spain', 'United Kingdom']

**Cluster 1** : mencakup mayoritas negara yang memiliki GDP lebih rendah dibandingkan negara-negara dalam Cluster 2. Negara-negara ini meliputi sebagian besar negara berkembang, negara-negara dengan ekonomi kecil, atau negara yang berada pada tahap awal industrialisasi.

**Cluster 2** : Cluster ini terdiri dari negara-negara yang memiliki GDP tinggi atau merupakan ekonomi besar di dunia. Negara-negara dalam cluster ini memiliki peran penting dalam perekonomian global dan mencakup ekonomi maju serta negara berkembang besar.

#### 4.1.4. Silhouette Score

Nilai Silhouette Score : 0.8858

Menunjukkan bahwa hasil clustering yang dihasilkan oleh algoritma K-Means sangat baik. Nilai ini berada dalam rentang -1 hingga 1, di mana nilai yang mendekati 1, seperti 0.8858, menunjukkan bahwa data dalam cluster sangat terpisah dengan jelas dan terkelompok secara kohesif. Dengan demikian, dapat disimpulkan bahwa algoritma K-Means berhasil mengelompokkan negara-negara berdasarkan GDP mereka dengan sangat baik, mencerminkan tingkat kemiripan yang tinggi di dalam cluster dan perbedaan yang signifikan antarcluster.

#### 4.1.5 Diskusi Hasil K- Means

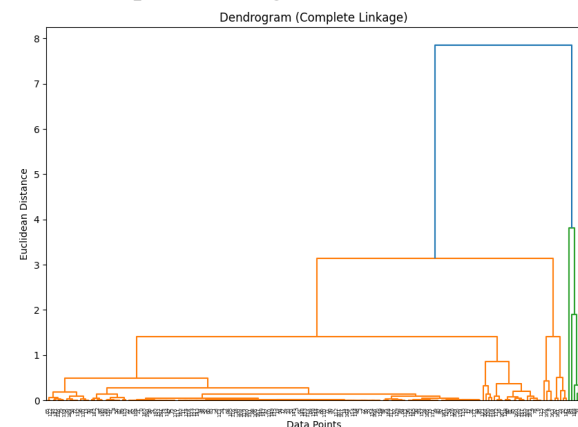
Algoritma K-Means berhasil mengelompokkan negara berdasarkan GDP menjadi dua cluster yang jelas, dengan hasil yang sangat baik. Berdasarkan Elbow Method, jumlah cluster optimal ditemukan pada  $k = 2$ , yang terlihat dari titik siku pada grafik penurunan Within-Cluster Sum of Squares (WCSS) yang signifikan di awal dan melambat pada  $k=2$ . Dengan waktu eksekusi yang sangat cepat, hanya memakan 0.0014 detik untuk menentukan pengelompokan, K-Means menunjukkan efisiensi dalam clustering. Hasil clustering menunjukkan bahwa Cluster 1 terdiri dari negara-negara dengan GDP rendah, yang umumnya adalah negara berkembang atau negara dengan ekonomi kecil, sedangkan Cluster 2 mencakup negara-negara dengan GDP tinggi, termasuk negara

maju dan beberapa negara berkembang besar seperti India dan Brasil.

Nilai Silhouette Score sebesar 0.8858 menunjukkan bahwa hasil clustering sangat baik, di mana data dalam setiap cluster sangat kohesif dan terpisah dengan jelas antar cluster. Hal ini mengindikasikan bahwa K-Means mampu memisahkan negara-negara dengan GDP tinggi dan rendah dengan jelas. Secara keseluruhan, algoritma K-Means memberikan hasil yang memadai dalam pengelompokan negara berdasarkan GDP, dengan kualitas cluster yang tinggi dan waktu eksekusi yang sangat efisien. Dengan skor Silhouette yang tinggi, dapat disimpulkan bahwa K-Means sangat efektif dalam menghasilkan cluster yang bermakna dan terpisah dengan baik.

## 4.2 Hierarchical Clustering (Tidak Termasuk China dan Amerika Serikat)

### 4.2.1 Complete Linkage



Waktu Eksekusi: 5.4517 detik

Silhouette Score: 0.8654

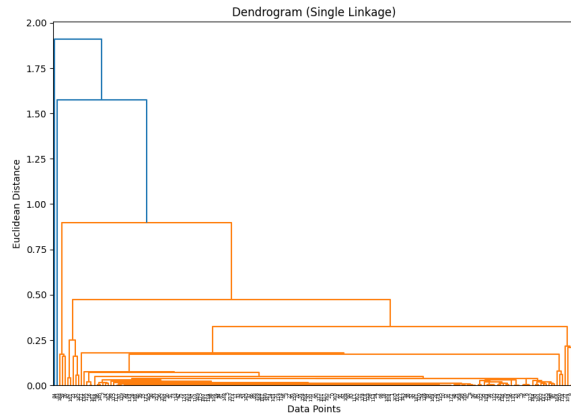
#### Hasil Cluster

Cluster 1:		
	Country	GDP
0	Afghanistan	1.910135e+10
1	Albania	1.527808e+10
2	Algeria	1.699802e+11
3	Andorra	3.154058e+09
4	Angola	9.463542e+10
..	...	...
188	Venezuela	4.823593e+11
189	Vietnam	2.619212e+11
190	Yemen	2.691440e+10
191	Zambia	2.306472e+10
192	Zimbabwe	2.144076e+10
[188 rows x 2 columns]		
Cluster 2:		
	Country	GDP
59	France	2.715518e+12
63	Germany	3.845630e+12
76	India	2.611000e+12
184	United Kingdom	2.827113e+12
Cluster 3:		
	Country	GDP
84	Japan	5.881770e+12

Hasil dari complete linkage menunjukkan waktu eksekusi yang relatif cepat, dan skor Silhouette Score yang sangat baik, yang mengindikasikan bahwa negara-negara dalam masing-masing cluster memiliki kesamaan yang tinggi dan jarak antar cluster relatif jauh.

Berdasarkan hasil cluster, kita memperoleh tiga cluster, di mana cluster pertama berisi sejumlah besar negara dengan GDP rendah hingga menengah, cluster kedua terdiri dari negara-negara dengan GDP sangat tinggi seperti **Prancis, Jerman, India, dan Inggris**, sementara cluster ketiga hanya mencakup **Jepang** dengan GDP yang tertinggi.

#### 4.2.2. Single Linkage



Waktu Eksekusi: 6.0642 detik

Silhouette Score: 0.8685

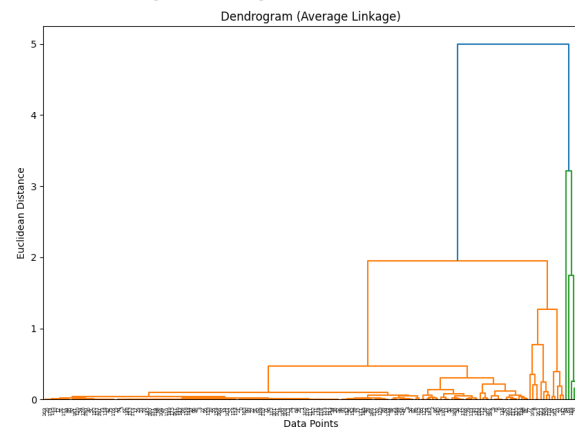
##### Hasil Cluster

Cluster 1:		
	Country	GDP
0	Afghanistan	1.910135e+10
1	Albania	1.527808e+10
2	Algeria	1.699882e+11
3	Andorra	3.154058e+09
4	Angola	9.463542e+10
..	...	...
188	Venezuela	4.823593e+11
189	Vietnam	2.619212e+11
190	Yemen	2.691440e+10
191	Zambia	2.306472e+10
192	Zimbabwe	2.144076e+10
[191 rows x 2 columns]		
Cluster 2:		
	Country	GDP
63	Germany	3.845630e+12
Cluster 3:		
	Country	GDP
84	Japan	5.081770e+12

Hasil pada single linkage sedikit lebih lama dibandingkan dengan complete linkage, namun memiliki skor Silhouette Score yang sedikit lebih tinggi, yang menunjukkan bahwa struktur cluster yang terbentuk relatif baik.

Cluster yang terbentuk menunjukkan hasil yang serupa dengan complete linkage, dengan cluster pertama mencakup banyak negara dengan GDP rendah hingga menengah, cluster kedua berisi **Jerman** sebagai negara dengan GDP yang sangat tinggi, dan cluster ketiga sebagai negara dengan GDP tertinggi hanya mencakup **Jepang**.

#### 4.2.3. Average Linkage



Waktu Eksekusi: 40.1771 detik

Silhouette Score: 0.8654

##### Hasil Cluster

Cluster 1:		
	Country	GDP
0	Afghanistan	1.910135e+10
1	Albania	1.527808e+10
2	Algeria	1.699882e+11
3	Andorra	3.154058e+09
4	Angola	9.463542e+10
..	...	...
188	Venezuela	4.823593e+11
189	Vietnam	2.619212e+11
190	Yemen	2.691440e+10
191	Zambia	2.306472e+10
192	Zimbabwe	2.144076e+10
[188 rows x 2 columns]		
Cluster 2:		
	Country	GDP
59	France	2.715518e+12
63	Germany	3.845630e+12
76	India	2.611000e+12
184	United Kingdom	2.827113e+12
Cluster 3:		
	Country	GDP
84	Japan	5.081770e+12

Waktu eksekusi pada average linkage jauh lebih lama dibandingkan kedua metode lainnya, meskipun skor Silhouette Score tetap tinggi dan menunjukkan kualitas cluster yang baik. Struktur cluster yang terbentuk pada average linkage sangat mirip dengan yang dihasilkan oleh complete linkage, dengan adanya tiga cluster utama yang serupa.

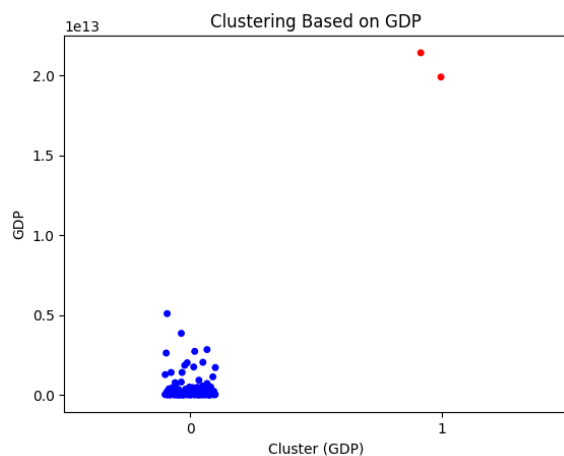
#### 4.2.4. Diskusi Hasil Hierarchical

Hasil clustering menunjukkan bahwa perbedaan metode linkage memiliki dampak pada waktu eksekusi dan dalam beberapa kasus pada struktur cluster, namun perbedaan tersebut tidak terlalu signifikan pada kualitas cluster yang terbentuk berdasarkan Silhouette Score. Complete linkage dan single linkage menghasilkan cluster dengan distribusi negara-negara yang relatif sama, dengan perbedaan utama terletak pada pengelompokan negara-negara besar seperti Jerman, Prancis, dan Jepang.

Meskipun average linkage memberikan hasil yang serupa dengan complete linkage dalam hal struktur cluster, waktu eksekusinya jauh lebih lama, yang dapat menjadi pertimbangan dalam memilih algoritma yang lebih efisien, terutama jika ukuran dataset meningkat.

### 4.3. Hasil K-Means dan Hierarchical Clustering (Termasuk Negara China dan Amerika Serikat)

#### 4.3.1. K-Means



#### Waktu Eksekusi

**Elbow Method:** 0.0621 detik

**K-means:** 0.0025 detik

**Hasil Clustering:**

Cluster 1: Semua negara kecuali China dan Amerika Serikat

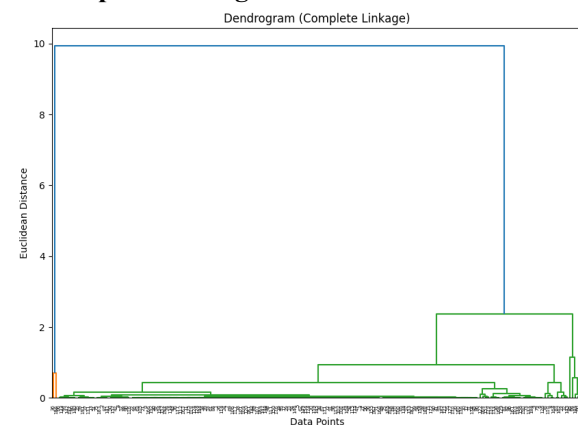
Cluster 2: China dan Amerika Serikat

**Silhouette Score (k=2):** 0.9772, yang menunjukkan bahwa hasil clustering sangat baik dengan pemisahan yang jelas antara kedua cluster.

Dalam K-means, metode elbow digunakan untuk menentukan jumlah cluster optimal. Setelah melakukan eksekusi, jumlah cluster yang ditemukan adalah dua, dengan China dan Amerika Serikat membentuk cluster tersendiri karena GDP mereka yang sangat besar dibandingkan negara lain.

#### 4.3.2. Hierarchical Clustering

##### a. Complete Linkage



**Waktu Eksekusi:** 6.3212 detik

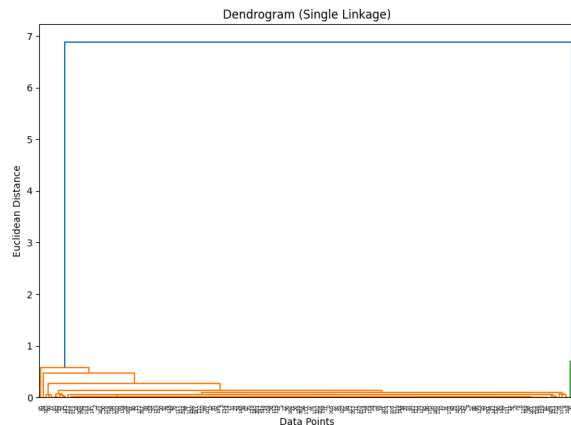
**Silhouette Score:** 0.8868

**Hasil Clustering:**

Cluster 1: China dan Amerika Serikat

Cluster 2: Negara-negara lainnya

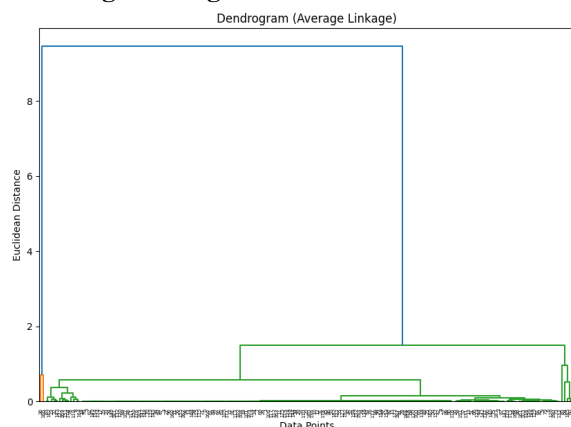
Cluster 3: Negara dengan GDP yang lebih besar, seperti Perancis, Jerman, India, Jepang, dan Inggris.

**b. Single Linkage****Waktu Eksekusi:** 4.8051 detik**Silhouette Score:** 0.9668**Hasil Clustering:**

Cluster 1: Negara-negara dengan GDP rendah

Cluster 2: China

Cluster 3: Amerika Serikat

**c. Average Linkage****Waktu Eksekusi:** 39.1761 detik**Silhouette Score:** 0.8868**Hasil Clustering:**

Cluster 1: China dan Amerika Serikat

Cluster 2: Negara-negara lainnya

Cluster 3: Negara dengan GDP besar, seperti Perancis, Jerman, India, Jepang, dan Inggris.

**4.3.3. Kesimpulan Hasil**

Berdasarkan hasil analisis clustering menggunakan K-means dan Hierarchical Clustering, kedua algoritma memberikan pemisahan yang cukup jelas antara negara-negara dengan GDP besar dan kecil. K-means dengan dua

cluster menempatkan China dan Amerika Serikat dalam satu cluster terpisah, sementara negara lainnya tergabung dalam cluster yang berbeda, dengan skor Silhouette yang sangat tinggi (0.9772), menunjukkan kualitas clustering yang sangat baik. Di sisi lain, Hierarchical Clustering, baik dengan complete, single, maupun average linkage, menghasilkan lebih banyak cluster dan pemisahan yang lebih terperinci, meskipun dengan waktu eksekusi yang lebih lama. Skor Silhouette untuk metode hierarchial juga cukup tinggi, terutama pada single linkage (0.9668). Secara keseluruhan, K-means lebih efisien dan lebih cepat dalam hal eksekusi, sementara Hierarchical Clustering memberikan pemahaman yang lebih rinci dengan mengidentifikasi lebih banyak cluster, meskipun membutuhkan lebih banyak waktu.

**4.3.4. Penjelasan Mengenai Penghilangan China dan Amerika Serikat**

Dalam analisis ini, kami sengaja menghilangkan negara China dan Amerika Serikat dari dataset untuk fokus pada negara-negara dengan GDP lebih rendah dan melihat pola clustering yang lebih jelas tanpa dominasi dua negara dengan GDP terbesar. Penghilangan ini membantu kami untuk memeriksa distribusi negara-negara dengan ekonomi lebih kecil secara lebih rinci dan mencegah bias yang mungkin timbul akibat kontribusi besar GDP China dan Amerika Serikat terhadap distribusi keseluruhan.

Secara keseluruhan, baik algoritma K-means maupun Hierarchical Clustering memberikan hasil yang serupa dalam mengelompokkan negara berdasarkan GDP. Namun, masing-masing algoritma memiliki kelebihan dalam hal interpretasi dan kecepatan eksekusi. K-means dengan jumlah cluster yang lebih sedikit (2 cluster) menunjukkan pemisahan yang jelas antara negara dengan GDP besar dan kecil, sementara Hierarchical Clustering memberikan lebih banyak detail dengan tiga cluster, mencerminkan kedekatan antar negara-negara dengan GDP yang serupa

#### 4.4. Perbandingan Hasil Antar 2 Algoritma Clustering

Berdasarkan hasil yang diperoleh dari kedua algoritma clustering, Hierarchical Clustering dan K-Means, dapat dilakukan perbandingan pada beberapa aspek utama seperti waktu eksekusi, kualitas clustering, dan struktur cluster yang terbentuk.

- **Waktu Eksekusi**

**Hierarchical Clustering:** Algoritma ini membutuhkan waktu yang lebih lama dibandingkan dengan K-Means. Untuk metode **Complete Linkage**, waktu eksekusi adalah 5.4517 detik, sedangkan untuk **Average Linkage**, waktu eksekusi mencapai 40.1771 detik. Hal ini menunjukkan bahwa Hierarchical Clustering lebih memakan waktu, terutama pada metode Average Linkage, yang mungkin menjadi kendala pada dataset yang lebih besar.

**K-Means:** Sebaliknya, K-Means lebih efisien dengan waktu eksekusi hanya 0.0014 detik untuk  $k=2$ , yang menunjukkan kecepatan algoritma ini dalam menjalankan clustering, bahkan dengan dataset yang lebih besar.

- **Kualitas Clustering (Silhouette Score)**

**Hierarchical Clustering:** Skor Silhouette Score untuk metode **Complete Linkage** dan **Average Linkage** adalah 0.8654, sedangkan **Single Linkage** menghasilkan skor sedikit lebih tinggi yaitu 0.8685. Skor ini menunjukkan bahwa hasil clustering dengan **Hierarchical Clustering** cukup baik, dengan pemisahan antar cluster yang jelas dan data dalam cluster yang kohesif.

**K-Means:** Hasil dari K-Means menunjukkan skor Silhouette Score yang lebih tinggi, yaitu 0.8858. Hal ini menunjukkan bahwa clustering dengan K-Means lebih terstruktur dengan pemisahan yang lebih jelas antar cluster, dan data dalam setiap cluster sangat terpisah dengan baik. Skor yang lebih tinggi ini menandakan kualitas clustering yang lebih baik dibandingkan dengan Hierarchical Clustering dalam hal kohesi dan pemisahan antar cluster.

- **Struktur Cluster yang Terbentuk**

**Hierarchical Clustering:** Pada metode **Complete Linkage** dan **Average Linkage**, dua cluster utama terbentuk: satu cluster besar yang mencakup negara-negara dengan GDP rendah hingga menengah, dan satu cluster kecil yang terdiri dari negara-negara dengan GDP sangat tinggi. Meskipun ada kemiripan antara ketiga metode linkage, struktur cluster menunjukkan bahwa negara-negara dengan GDP rendah dan tinggi terpisah dengan jelas.

**K-Means:** Hasil K-Means juga membentuk dua cluster utama: Cluster pertama berisi negara-negara dengan GDP rendah, sedangkan Cluster kedua mencakup negara-negara dengan GDP tinggi. Perbedaan utama dengan Hierarchical Clustering adalah bahwa K-Means lebih fokus pada pemisahan negara-negara berdasarkan GDP yang lebih ekstrem (tinggi dan rendah), dengan negara-negara ekonomi besar dikelompokkan secara lebih kohesif.

#### 4.5. Analisis Kelebihan dan Kekurangan

##### 4.5.1. Kelebihan Hierarchical Clustering

1. **Tanpa Memerlukan Jumlah Cluster yang Ditetapkan:** Salah satu kelebihan utama dari Hierarchical Clustering adalah bahwa jumlah cluster tidak perlu ditentukan sebelumnya, memungkinkan algoritma ini untuk membentuk struktur dendrogram yang dapat dianalisis lebih lanjut.
2. **Visualisasi:** Hasil dari Hierarchical Clustering lebih mudah untuk divisualisasikan dalam bentuk dendrogram, yang memungkinkan pengguna untuk memilih jumlah cluster yang optimal secara lebih fleksibel dan berdasarkan analisis visual.

##### 4.5.2. Kekurangan Hierarchical Clustering

1. **Waktu Eksekusi yang Lama:** Terutama pada metode Average Linkage, waktu eksekusi bisa sangat lama. Ini menjadi kendala jika dataset yang digunakan sangat besar.
2. **Kurang Efisien pada Skala Besar:** Hierarchical Clustering cenderung lebih



lambat dibandingkan dengan algoritma lain seperti K-Means pada dataset besar, membuatnya kurang efisien untuk aplikasi dengan data besar.

#### 4.5.3. Kelebihan K-Means

1. **Kecepatan Eksekusi yang Sangat Cepat:** K-Means menunjukkan kecepatan yang sangat tinggi dalam eksekusi, dengan waktu yang hanya memakan beberapa detik bahkan untuk dataset besar. Ini membuat K-Means lebih efisien dalam penggunaan waktu.
2. **Skor Silhouette yang Lebih Baik:** Berdasarkan hasil Silhouette Score, K-Means memberikan hasil clustering yang lebih baik dengan pemisahan antar cluster yang lebih jelas dan kohesif.
3. **Skalabilitas yang Baik:** K-Means lebih cocok untuk dataset besar karena kecepatannya dalam eksekusi dan kemampuannya untuk menangani data dalam jumlah besar.

#### 4.5.4. Kekurangan K-Means

1. **Memerlukan Penentuan Jumlah Cluster:** Salah satu kelemahan terbesar K-Means adalah bahwa jumlah cluster harus ditentukan sebelumnya, yang dapat menjadi masalah jika jumlah cluster yang optimal tidak diketahui.
2. **Sensitif terhadap Inisialisasi Centroid:** Hasil clustering K-Means dapat dipengaruhi oleh pemilihan awal centroid. Jika centroid awal dipilih dengan buruk, hal ini bisa menghasilkan cluster yang suboptimal.
3. **Kesulitan dengan Data yang Tidak Terpisah Secara Linear:** K-Means tidak bekerja dengan baik pada dataset yang memiliki cluster yang tidak berbentuk bulat atau tidak terpisah secara linear.

## V. KESIMPULAN

Berdasarkan hasil eksperimen, kedua algoritma clustering, Hierarchical Clustering dan K-Means, memiliki kelebihan dan kekurangan masing-masing. Hierarchical Clustering menawarkan fleksibilitas karena tidak memerlukan jumlah cluster yang ditentukan sebelumnya dan

dapat divisualisasikan dalam bentuk dendrogram, yang memungkinkan pengguna untuk menentukan jumlah cluster optimal secara visual. Namun, algoritma ini memerlukan waktu eksekusi yang lebih lama, terutama pada metode Average Linkage, yang mengurangi efisiensinya pada dataset besar. Sebaliknya, K-Means lebih efisien dan cepat, dengan waktu eksekusi yang sangat singkat bahkan pada dataset besar, serta menghasilkan kualitas clustering yang sangat baik, yang tercermin dari nilai Silhouette Score yang lebih tinggi. K-Means juga lebih cocok untuk aplikasi dengan dataset besar karena skalabilitasnya.

Namun, K-Means memerlukan penentuan jumlah cluster sebelumnya dan sensitif terhadap pemilihan centroid awal, yang bisa mempengaruhi hasil clustering. Dalam hal ini, pemilihan algoritma tergantung pada karakteristik dataset dan kebutuhan aplikasi. Jika jumlah cluster sudah diketahui dan efisiensi waktu adalah prioritas, maka K-Means adalah pilihan yang lebih baik. Namun, jika fleksibilitas dan analisis visual lebih penting, dan waktu eksekusi bukan masalah besar, maka **Hierarchical Clustering dapat menjadi pilihan yang tepat**. Kedua algoritma ini memberikan wawasan yang berbeda mengenai pengelompokan negara berdasarkan GDP, masing-masing dengan pendekatan yang sesuai untuk tujuan analisis tertentu.

## VI. REFERENSI

Kumar, S, Rani, R, Pippal, SK, & ... (2025). Customer segmentation in e-commerce: K-means vs hierarchical clustering. ... *Electronics and Control*), telkomnika.uad.ac.id, <https://telkomnika.uad.ac.id/index.php/TELKOMNIKA/article/view/26384>