

# **Laporan Analisis Cluster: Segmentasi Negara Berdasarkan Indikator Pembangunan dengan gapminder.csv**

Disusun untuk Memenuhi Tugas Mata Kuliah Data Science

Dosen Pengampu: Khaerul Anam, M.Kom



Disusun oleh:

- |                          |            |
|--------------------------|------------|
| 1. Ayu Aulia             | (43240354) |
| 2. Dea Nova Refalina     | (43240361) |
| 3. Putri Tri Setyawati   | (43240462) |
| 4. Tiara Putri Rinjani   | (43240443) |
| 5. Tion Hermawan         | (43240445) |
| 6. Zacky Muhammad Dinata | (43240450) |

**PROGRAM STUDI SISTEM INFORMASI**

**STMIK IKMI CIREBON**

**2025**

## Kata Pengantar

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga kami dapat menyelesaikan Laporan Analisis *Cluster* ini. Laporan yang berjudul **"Laporan Analisis Cluster: Segmentasi Negara Berdasarkan Indikator Pembangunan dengan *gapminder.csv*"** ini disusun sebagai bagian dari pemenuhan tugas mata kuliah Data Science.

Tujuan utama dari laporan ini adalah untuk memetakan dan memahami keragaman kondisi antar negara di dunia. Kita tahu bahwa setiap negara memiliki tingkat perkembangan dan tantangan yang berbeda-beda. Dengan menggunakan metode Analisis *Cluster* (pengelompokan), kami berusaha mengelompokkan negara-negara tersebut ke dalam beberapa segmen berdasarkan indikator pembangunan utama seperti harapan hidup, pendapatan per kapita, dan populasi. Data yang kami gunakan bersumber dari berkas *gapminder.csv*, yang menyediakan pandangan global yang menarik.

Laporan ini akan menjelaskan langkah-langkah kami, mulai dari persiapan data, pemilihan metode K-Means, hingga interpretasi profil dari setiap kelompok negara yang terbentuk. Hasil dari analisis ini diharapkan dapat memberikan wawasan yang jelas mengenai pola-pola pembangunan global, serta memberikan dasar bagi pemahaman yang lebih baik mengenai ketimpangan dan kemajuan di berbagai kawasan dunia.

Kami menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, segala kritik dan saran yang membangun akan kami terima dengan lapang dada demi perbaikan di masa depan. Akhir kata, kami berharap laporan ini dapat memberikan manfaat, wawasan baru, dan menjadi referensi yang berguna bagi pembaca.

Cirebon, Oktober 2025

Kelompok BismillahBisaNgerjainnya

## Daftar Isi

<b>Kata Pengantar .....</b>	<b>i</b>
<b>Daftar Isi .....</b>	<b>ii</b>
<b>Daftar Gambar .....</b>	<b>iii</b>
<b>BAB 1 Pendahuluan.....</b>	<b>1</b>
<b>1.1. Latar Belakang .....</b>	<b>1</b>
<b>1.2. Rumusan Masalah.....</b>	<b>2</b>
<b>1.3. Tujuan Penelitian .....</b>	<b>2</b>
<b>1.4. Manfaat Penelitian .....</b>	<b>2</b>
<b>1.5. Ruang Lingkup dan Batasan.....</b>	<b>2</b>
<b>1.6. Sistematika Penulisan .....</b>	<b>3</b>
<b>BAB 2 Metodologi Analisis .....</b>	<b>4</b>
<b>2.1. Sumber Dataset .....</b>	<b>4</b>
<b>2.2. Pra-Pemrosesan Data.....</b>	<b>5</b>
<b>2.3. Pengembangan Model.....</b>	<b>6</b>
<b>2.4. Evaluasi Model .....</b>	<b>7</b>
<b>BAB 3 Hasil dan Pembahasan .....</b>	<b>9</b>
<b>3.1. Visualisasi Hasil Clustering.....</b>	<b>9</b>
<b>3.2. Analisis Profil Rata-Rata Cluster .....</b>	<b>10</b>
<b>3.3. Interpretasi dan Penamaan Cluster .....</b>	<b>10</b>
<b>3.3.1. Cluster 1: Negara Mega-Populasi.....</b>	<b>10</b>
<b>3.3.2. Cluster 2: Negara Berkembang / Pendapatan Rendah .....</b>	<b>11</b>
<b>3.3.3. Cluster 3: Negara Maju / Pendapatan Tinggi .....</b>	<b>11</b>
<b>BAB 5 Kesimpulan.....</b>	<b>12</b>
<b>4.1. Kesimpulan .....</b>	<b>12</b>
<b>4.2. Keterbatasan Penelitian.....</b>	<b>13</b>
<b>4.3. Saran.....</b>	<b>13</b>

## Daftar Gambar

<b>Gambar 1.</b> Metodologi Analisis .....	4
<b>Gambar 2.</b> Total Within Sum of Square.....	6
<b>Gambar 3.</b> Average Silhoutte Width .....	7
<b>Gambar 4.</b> Visualisasi Cluster K-Means (Direduksi via PCA).....	9
<b>Gambar 5.</b> Analisis Profil Rata-rata Cluster.....	10

# BAB 1

## Pendahuluan

### 1.1. Latar Belakang

Dunia saat ini ditandai oleh keragaman yang luar biasa antar negara, tidak hanya dalam hal budaya dan geografi, tetapi juga dalam tingkat pembangunan. Kesenjangan dalam kemakmuran ekonomi, kesehatan masyarakat, dampak lingkungan, dan struktur demografi merupakan isu sentral dalam studi pembangunan global. Dataset Gapminder, yang dipopulerkan oleh Hans Rosling, menyediakan data komprehensif untuk mengeksplorasi perbedaan-perbedaan ini dan memahami bagaimana negara-negara bergerak dalam lintasan pembangunan yang berbeda.

Secara tradisional, negara-negara sering dikategorikan secara sederhana sebagai "maju" atau "berkembang". Namun, realitasnya jauh lebih kompleks. Sebuah negara mungkin memiliki pendapatan tinggi namun juga emisi karbon yang tinggi. Negara lain mungkin memiliki populasi yang sangat besar yang menghadirkan tantangan unik, terlepas dari tingkat pendapatannya. Kategorisasi biner semacam itu gagal menangkap nuansa ini.

Oleh karena itu, ada kebutuhan untuk melihat melampaui metrik tunggal dan menganalisis berbagai indikator secara bersamaan. Dengan mempertimbangkan variabel-variabel kunci seperti pendapatan per kapita (*income*), angka harapan hidup (*life\_exp*), emisi CO2 per kapita (*co2*), dan total populasi (*population*), kita dapat mulai melihat gambaran yang lebih holistik. Namun, dengan banyaknya variabel, menjadi sulit untuk mengidentifikasi pola atau kelompok alami secara manual.

Di sinilah analisis *clustering* (segmentasi) menjadi alat yang ampuh. Dengan menerapkan teknik *unsupervised learning* seperti K-Means, kita dapat meminta mesin untuk menemukan struktur atau kelompok yang inheren dalam data. Analisis ini memungkinkan kita untuk mengelompokkan negara-negara berdasarkan kesamaan profil mereka di semua empat dimensi (ekonomi, kesehatan, lingkungan, dan demografi) secara bersamaan.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, penelitian ini berfokus untuk menjawab pertanyaan-pertanyaan berikut:

1. Bagaimana negara-negara di dunia dapat disegmentasi ke dalam kelompok-kelompok (cluster) yang bermakna berdasarkan empat indikator kunci: **income**, **life\_exp**, **co2**, dan **population**?
2. Apa saja karakteristik utama yang mendefinisikan dan membedakan setiap segmen negara yang terbentuk dari hasil analisis K-Means?

## 1.3. Tujuan Penelitian

Sejalan dengan rumusan masalah, tujuan dari penelitian ini adalah:

1. Menerapkan algoritma K-Means untuk melakukan segmentasi terhadap 193 negara dalam dataset Gapminder.
2. Menentukan jumlah cluster (K) yang paling optimal untuk merepresentasikan struktur data.
3. Menganalisis dan menginterpretasikan profil rata-rata dari setiap cluster yang terbentuk untuk membangun "persona" atau deskripsi kualitatif yang jelas untuk setiap segmen.

## 1.4. Manfaat Penelitian

Hasil dari analisis ini diharapkan dapat memberikan manfaat sebagai berikut:

- **Secara Akademis:** Menyajikan contoh metodologis penerapan K-Means clustering pada dataset sosial-ekonomi makro, mulai dari pra-pemrosesan hingga interpretasi hasil.
- **Secara Praktis:** Memberikan pemahaman yang lebih bernuansa tentang tipologi negara-negara di dunia. Segmentasi ini dapat membantu pembuat kebijakan, organisasi non-pemerintah (NGO), dan lembaga internasional (seperti PBB atau Bank Dunia) dalam merancang intervensi atau alokasi sumber daya yang lebih tepat sasaran, sesuai dengan profil tantangan dan kekuatan unik dari setiap kelompok negara.

## 1.5. Ruang Lingkup dan Batasan

Analisis ini memiliki ruang lingkup dan batasan sebagai berikut:

1. **Data:** Analisis terbatas pada dataset **gapminder.csv** yang terdiri dari 193 negara.

2. **Variabel:** Model clustering secara eksklusif didasarkan pada empat variabel kuantitatif: *income*, *life\_exp*, *co2*, dan *population*. Variabel kualitatif seperti *region* atau *income\_level* yang ada di dataset tidak digunakan dalam pembentukan model, tetapi dapat digunakan untuk analisis kontekstual tambahan.
3. **Metodologi:** Penelitian ini hanya menggunakan algoritma K-Means. Metode clustering lain (misalnya, Hierarchical Clustering, DBSCAN) tidak dieksplorasi.
4. **Sifat Data:** Data bersifat *cross-sectional* (satu titik waktu). Analisis ini tidak melihat tren atau perubahan dari waktu ke waktu (*time-series*).

## 1.6. Sistematika Penulisan

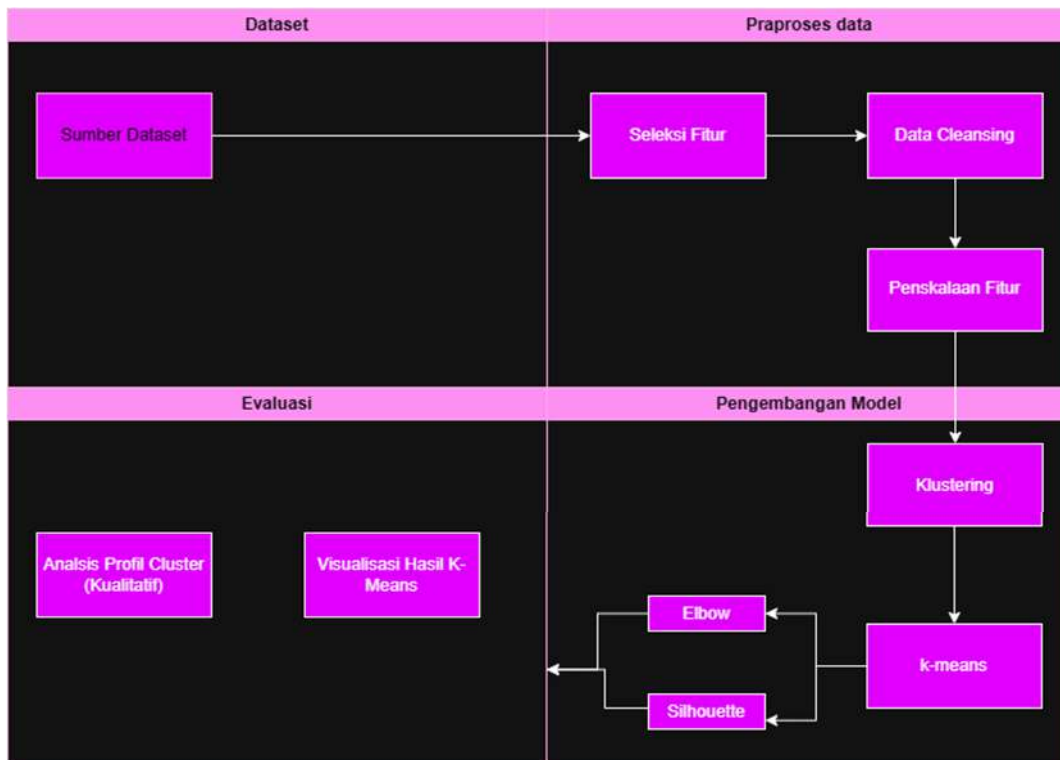
Laporan ini disusun ke dalam beberapa bab dengan sistematika sebagai berikut:

- **BAB 1: PENDAHULUAN** Bab ini berisi pengantar umum, mencakup latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup, serta sistematika penulisan laporan.
- **BAB 2: METODOLOGI ANALISIS** Bab ini menguraikan secara rinci langkah-langkah teknis analisis, mulai dari deskripsi sumber data, proses pra-pemrosesan data (termasuk seleksi fitur, penanganan data hilang, dan standardisasi), penentuan jumlah K optimal, hingga proses pelatihan model K-Means dan metode evaluasi yang digunakan.
- **BAB 3: HASIL DAN PEMBAHASAN** Bab ini menyajikan temuan dari analisis. Ini mencakup visualisasi cluster yang terbentuk melalui PCA, penyajian tabel profil rata-rata untuk setiap cluster, serta interpretasi dan penamaan kualitatif untuk setiap segmen negara yang teridentifikasi.
- **BAB 4: PENUTUP** Bab terakhir ini merangkum kesimpulan utama dari penelitian dan memberikan saran untuk penelitian lebih lanjut yang dapat dikembangkan berdasarkan temuan yang ada.

## BAB 2

### Metodologi Analisis

Bab ini menguraikan secara rinci tahapan metodologi yang digunakan untuk melakukan segmentasi negara-negara dalam dataset Gapminder. Mengikuti kerangka kerja yang sistematis sangat penting untuk memastikan bahwa hasil analisis valid, dapat direplikasi, dan dapat diinterpretasikan dengan benar. Proses ini mencakup empat tahapan utama: penentuan dan pemahaman sumber data, pra-pemrosesan data untuk persiapan model, pengembangan model clustering, dan evaluasi hasil model.



Gambar 1. Metodologi Analisis

### 2.1. Sumber Dataset

Analisis ini menggunakan data sekunder yang bersumber dari dataset Gapminder ([gapminder.csv](#)). Data ini bersifat *cross-sectional*, yang berarti data dikumpulkan pada satu titik waktu tertentu (dalam hal ini, data tahunan terbaru yang tersedia) untuk 193 unit observasi (negara), di mana setiap negara bertindak sebagai satu titik data independen.

Variabel kunci (fitur) yang digunakan dalam analisis ini dibagi menjadi dua jenis. Pertama adalah fitur **kuantitatif**, yang mencakup **income** (pendapatan per kapita),



`life_exp` (angka harapan hidup), `co2` (emisi CO2 per kapita), dan `population` (jumlah populasi). Keempat variabel ini dipilih karena secara kolektif mewakili pilar-pilar utama pembangunan: ekonomi, kesehatan, lingkungan, dan demografi. Kedua adalah fitur **kualitatif**, seperti `country`, `region`, `income_level`, dan `co2_change`, yang meskipun penting untuk konteks, tidak digunakan secara langsung dalam perhitungan model clustering ini.

## 2.2. Pra-Pemrosesan Data

Tujuan dari tahap ini adalah untuk membersihkan, mentransformasi, dan menyiapkan data mentah agar sesuai dengan persyaratan teknis algoritma K-Means, sehingga memastikan model yang dihasilkan akurat.

- **Seleksi Fitur:**

- Hanya fitur kuantitatif (`income`, `life_exp`, `co2`, `population`) yang dipilih untuk analisis.
- Algoritma K-Means secara fundamental beroperasi dengan mengukur jarak matematis (seperti jarak Euclidean) antar titik data dalam ruang multidimensi, sehingga perhitungan ini hanya dapat dilakukan pada data numerik.

- **Penanganan Data Hilang (Missing Values):**

- Integritas data diperiksa menggunakan metode `na.omit()` untuk mendeteksi dan menghapus baris mana pun yang memiliki data hilang pada kolom yang dipilih.
- Hasil pengecekan menunjukkan tidak ada data yang hilang, sehingga 193 observasi (negara) asli tetap utuh untuk analisis.

- **Penskalaan Fitur (Standardisasi):**

- Ini adalah langkah **wajib** untuk K-Means karena algoritma ini sangat sensitif terhadap skala variabel.
- Tanpa penskalaan, variabel dengan rentang nilai yang besar (seperti `population` dalam jutaan) akan secara tidak proporsional mendominasi perhitungan jarak dibandingkan dengan variabel dengan rentang nilai kecil (seperti `co2`). Hal ini akan menyebabkan model clustering menjadi bias dan secara efektif hanya mengelompokkan berdasarkan populasi.
- Metode standardisasi (Z-score) diterapkan menggunakan fungsi `scale()` di R. Metode ini mengubah setiap nilai dalam suatu fitur dengan

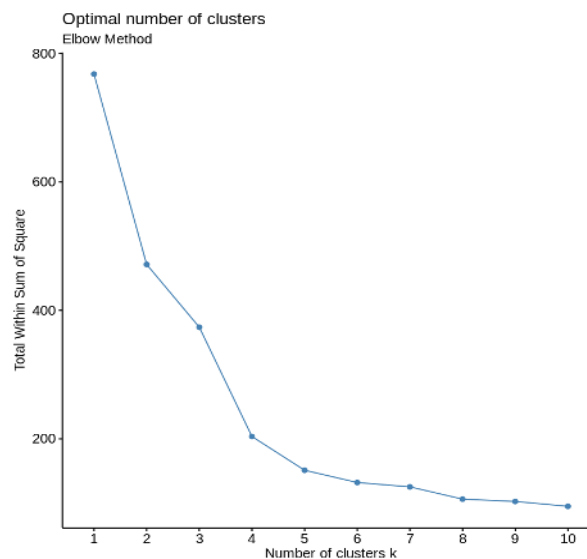
mengurangkan rata-rata fitur dan membaginya dengan standar deviasi fitur tersebut.

- Hasilnya adalah sebuah dataset baru (*df\_scaled*) di mana semua fitur memiliki "bobot" yang setara, dengan nilai rata-rata 0 dan standar deviasi 1.

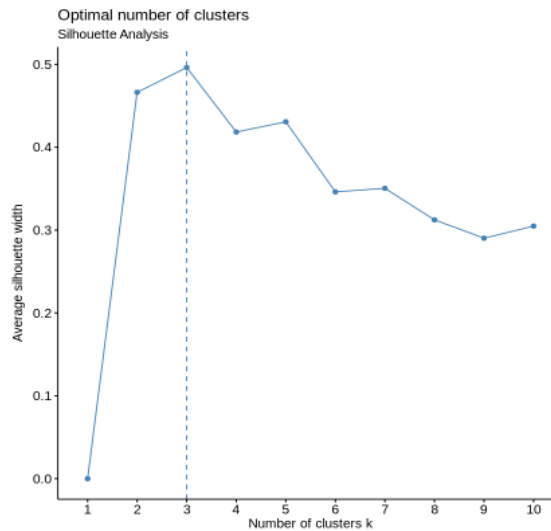
### 2.3. Pengembangan Model

Model clustering dikembangkan menggunakan algoritma *unsupervised learning* K-Means, yang bertujuan untuk mempartisi data ke dalam  $K$  kelompok di mana variasi di dalam kelompok diminimalkan.

Langkah krusial pertama dalam K-Means adalah **penentuan 'K' Optimal**, yaitu menentukan jumlah cluster ( $K$ ) yang paling optimal dan bermakna secara inheren dalam data. Dua metode digunakan untuk ini. Pertama, **Metode Elbow (WSS)**, yang memvisualisasikan *trade-off* antara jumlah cluster ( $K$ ) dan *Within-Cluster Sum of Squares* (WSS). Kami mencari titik "siku" (elbow), di mana penambahan  $K$  baru tidak lagi menghasilkan penurunan WSS yang substansial. Kedua, **Metode Silhouette Analysis**, yang memberikan metrik lebih kuat dengan mengukur seberapa baik setiap titik data cocok di dalam clusternya sendiri (kohesi) dibandingkan dengan cluster terdekat (separasi). Berdasarkan skor Silhouette Analysis yang tertinggi, yang menunjukkan keseimbangan terbaik antara kohesi dan separasi, jumlah cluster optimal ditetapkan pada  **$K = 3$** .



**Gambar 2.** *Total Within Sum of Square*



**Gambar 3.** *Average Silhoutte Width*

Setelah K ditentukan, dilakukan **pelatihan model final**. Model K-Means dilatih menggunakan data yang telah distandardisasi (`df_scaled`) dengan parameter yang telah ditentukan untuk memastikan robustisitas dan reproduktifitas. Parameter tersebut adalah `centers = 3` (menggunakan K optimal), `nstart = 50`, dan `set.seed(123)`. Pengaturan `nstart = 50` menjalankan algoritma sebanyak 50 kali dengan titik awal acak yang berbeda untuk memastikan model menemukan solusi global optimum yang paling stabil, bukan hanya *local optimum*. Pengaturan `set.seed(123)` digunakan untuk memastikan bahwa proses inisialisasi acak dapat direplikasi, sehingga siapa pun yang menjalankan skrip ini akan mendapatkan hasil yang sama persis.

## 2.4. Evaluasi Model

Setelah model terbentuk dan negara-negara dikelompokkan, evaluasi dilakukan untuk menginterpretasi dan memberi makna pada ketiga cluster yang abstrak tersebut.

Evaluasi pertama adalah **visualisasi cluster**. Menggunakan fungsi `fviz_cluster` dari paket `factoextra`, cluster divisualisasikan. Karena data input memiliki 4 dimensi (4 fitur) yang tidak mungkin divisualisasikan secara langsung, fungsi ini secara otomatis menerapkan Principal Component Analysis (PCA) di latar belakang. PCA mereduksi 4 dimensi tersebut menjadi 2 "Komponen Utama" yang menangkap sebagian besar varians data, memungkinkan kita untuk memplot cluster pada grafik 2D dan menginspeksi seberapa baik cluster tersebut terpisah secara visual.

Evaluasi kedua dan yang paling utama adalah **analisis profil cluster (evaluasi kualitatif)**. Ini adalah langkah untuk menerjemahkan hasil matematis model menjadi

wawasan yang dapat dipahami manusia. Proses ini melibatkan dua langkah: pertama, label cluster (1, 2, atau 3) yang dihasilkan oleh model digabungkan kembali dengan data **asli** (data *sebelum* di-scale). Kedua, nilai rata-rata (mean) dari setiap fitur asli (*income*, *life\_exp*, *co2*, *population*) dihitung untuk masing-masing dari 3 cluster. Outputnya adalah sebuah tabel profil (*kmeans\_profile*) yang merangkum karakteristik rata-rata setiap cluster dalam satuan yang dapat dipahami (misal: pendapatan dalam dolar, harapan hidup dalam tahun). Tujuan akhirnya adalah untuk membangun "persona" atau deskripsi kualitatif untuk setiap cluster, yang merupakan hasil akhir dari analisis segmentasi ini.

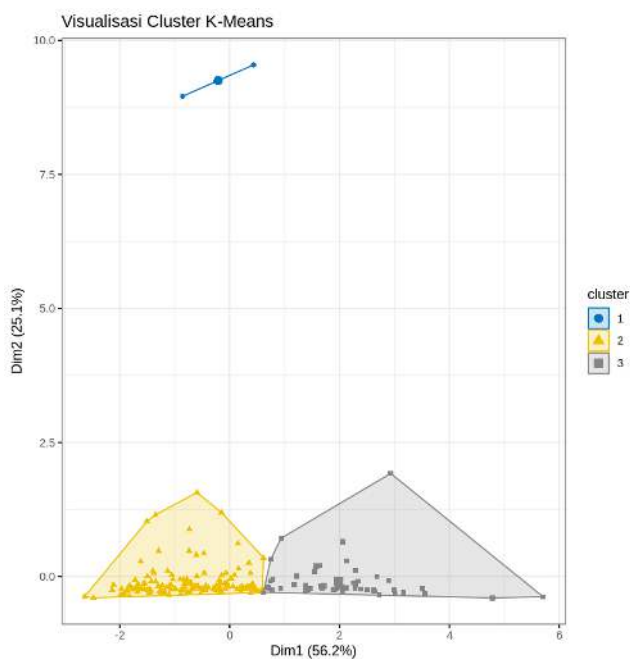
## BAB 3

### Hasil dan Pembahasan

Bab ini menyajikan temuan-temuan kunci dari penerapan model K-Means yang telah dijelaskan dalam Bab 2. Proses ini dimulai dengan visualisasi cluster yang terbentuk, diikuti oleh analisis kuantitatif terhadap profil masing-masing cluster, dan diakhiri dengan interpretasi kualitatif untuk memberikan makna pada setiap segmen negara yang teridentifikasi.

#### 3.1. Visualisasi Hasil Clustering

Sesuai dengan metodologi, model K-Means dijalankan dengan  $K=3$ . Untuk memvisualisasikan hasil segmentasi dari data 4-dimensi (*income*, *life\_exp*, *co2*, *population*), teknik reduksi dimensi Principal Component Analysis (PCA) diterapkan secara otomatis oleh fungsi *fviz\_cluster*. Hasilnya disajikan dalam Gambar 3.1.



**Gambar 4.** *Visualisasi Cluster K-Means (Direduksi via PCA)*

Grafik di atas memetakan 193 negara ke dalam ruang 2-dimensi.

- Sumbu horizontal (Dim1) menjelaskan **56.2%** dari total varians dalam data.
- Sumbu vertikal (Dim2) menjelaskan **25.1%** dari total varians.

Secara kumulatif, kedua sumbu ini berhasil menangkap **81.3%** dari total informasi asli, yang menunjukkan bahwa visualisasi ini sangat representatif.

Dari plot tersebut, kita dapat mengamati tiga kelompok yang terpisah dengan jelas:

1. **Cluster 1 (Biru, Lingkaran):** Terletak di kuadran kiri atas, sangat terisolasi dari kelompok lain, dan hanya terdiri dari dua titik data.
2. **Cluster 2 (Kuning, Segitiga):** Membentuk kelompok besar dan padat di kuadran kiri bawah.
3. **Cluster 3 (Abu-abu, Kotak):** Membentuk kelompok besar lainnya di sisi kanan, tampak sedikit lebih tersebar daripada Cluster 2.

Pemisahan visual yang jelas ini mengindikasikan bahwa model K-Means berhasil menemukan struktur yang kuat dan berbeda dalam data.

### 3.2. Analisis Profil Rata-Rata Cluster

Untuk memahami *apa* yang diwakili oleh ketiga cluster visual ini, kita perlu menganalisis karakteristik rata-rata dari variabel aslinya (sebelum penskalaan). Tabel 3.1 merangkum profil rata-rata untuk setiap cluster.

```
### Profil Rata-Rata Cluster (K-Means) ###
# A tibble: 3 x 6
  cluster income life_exp co2 population hierarchical_cluster
  <int>   <dbl>   <dbl> <dbl>   <dbl>             <dbl>
1     1     10.8    74.0  4.51    1390              3
2     2     9.27    69.6  2.17     24.6              1
3     3    47.6    79.9 10.4    27.3              1.70
```

Gambar 5. Analisis Profil Rata-rata Cluster

### 3.3. Interpretasi dan Penamaan Cluster

Berdasarkan data kuantitatif pada Tabel 3.1 dan didukung oleh visualisasi pada Gambar 3.1, kita dapat membangun "persona" atau interpretasi kualitatif untuk setiap cluster:

#### 3.3.1. Cluster 1: Negara Mega-Populasi

- **Profil:** Cluster ini memiliki nilai rata-rata pendapatan (10.8k), harapan hidup (74.0), dan CO2 (4.51) yang "menengah" (berada di antara Cluster 2 dan 3). Namun, yang paling menonjol adalah **populasi rata-rata yang ekstrem (1.39 Miliar)**.
- **Analisis:** Ini adalah cluster "outlier" yang didominasi oleh variabel populasi. Seperti yang terlihat pada Gambar 3.1, cluster ini hanya berisi dua titik data (India dan Tiongkok). Model K-Means mengisolasi mereka ke dalam cluster

tersendiri karena skala populasi mereka secara fundamental membuat mereka berbeda dari semua negara lain di dunia.

### 3.3.2. Cluster 2: Negara Berkembang / Pendapatan Rendah

- **Profil:** Cluster ini memiliki nilai **terendah** di hampir semua metrik: pendapatan rata-rata terendah (9.27k), harapan hidup rata-rata terendah (69.6 tahun), dan emisi CO2 rata-rata terendah (2.17 ton). Rata-rata populasinya (24.6 juta) relatif normal.
- **Analisis:** Ini adalah segmen terbesar yang mewakili negara-negara berkembang atau berpendapatan rendah. Mereka dicirikan oleh tingkat ekonomi dan industrialisasi yang lebih rendah, yang berkorelasi dengan angka harapan hidup yang lebih rendah dan jejak karbon yang lebih kecil.

### 3.3.3. Cluster 3: Negara Maju / Pendapatan Tinggi

- **Profil:** Cluster ini secara konsisten menunjukkan nilai **tertinggi** untuk metrik pembangunan: pendapatan rata-rata tertinggi (47.6k), harapan hidup rata-rata tertinggi (79.9 tahun), dan emisi CO2 rata-rata tertinggi (10.4 ton). Rata-rata populasinya (27.3 juta) mirip dengan Cluster 2.
- **Analisis:** Ini adalah segmen negara-negara maju, kaya, dan terindustrialisasi. Tingkat pendapatan yang tinggi berkorelasi kuat dengan harapan hidup yang tinggi, tetapi juga (dalam konteks data ini) dengan emisi CO2 per kapita yang tertinggi, yang mencerminkan gaya hidup dan ekonomi berbasis industri.

## BAB 5

### Kesimpulan

Bab terakhir ini menyajikan rangkuman kesimpulan yang ditarik dari hasil analisis dan pembahasan, serta mengidentifikasi keterbatasan penelitian dan memberikan saran untuk pengembangan penelitian di masa mendatang.

#### 4.1. Kesimpulan

Berdasarkan analisis yang telah dilakukan, penelitian ini berhasil menjawab rumusan masalah yang telah dipaparkan pada Bab 1.

1. **Segmentasi Negara:** Penelitian ini berhasil menerapkan algoritma K-Means untuk disegmentasi 193 negara dalam dataset Gapminder. Berdasarkan metode Elbow dan Silhouette Analysis (seperti dijelaskan di Bab 2), jumlah cluster (K) yang paling optimal untuk merepresentasikan struktur data adalah **K = 3**.
2. **Karakteristik Cluster:** Analisis profil rata-rata (dijelaskan di Bab 3) berhasil mengidentifikasi tiga segmen negara yang unik dan bermakna, yang melampaui kategorisasi biner "maju" atau "berkembang":
  - **Cluster 1: Negara Mega-Populasi.** Cluster ini (terdiri dari 2 negara) secara fundamental didefinisikan oleh variabel **population** yang ekstrem (rata-rata 1.39 Miliar). Meskipun memiliki profil ekonomi, kesehatan, dan lingkungan yang menengah, skala populasinya yang masif menempatkan mereka dalam kategori tersendiri dengan tantangan pembangunan yang unik.
  - **Cluster 2: Negara Berkembang / Pendapatan Rendah.** Cluster ini merupakan kelompok terbesar, ditandai dengan profil rata-rata terendah pada tiga metrik kunci: **income** (9.27k), **life\_exp** (69.6 tahun), dan **co2** (2.17 ton/kapita).
  - **Cluster 3: Negara Maju / Pendapatan Tinggi.** Cluster ini adalah kebalikan dari Cluster 2, menunjukkan profil rata-rata tertinggi pada tiga metrik yang sama: **income** (47.6k), **life\_exp** (79.9 tahun), dan **co2** (10.4 ton/kapita).

Secara keseluruhan, analisis ini menunjukkan bahwa K-Means clustering adalah metode yang efektif untuk menemukan pola yang kompleks dalam data sosial-



ekonomi makro, memberikan wawasan yang lebih bernuansa tentang tipologi negara-negara di dunia.

## 4.2. Keterbatasan Penelitian

Meskipun temuan ini memberikan wawasan yang berharga, penelitian ini memiliki beberapa keterbatasan yang perlu diakui, sebagian besar berasal dari ruang lingkup yang ditetapkan di Bab 1:

1. **Sifat Data:** Analisis ini bersifat *cross-sectional*, artinya hanya melihat potret data pada satu titik waktu. Hal ini tidak memungkinkan untuk menganalisis bagaimana negara-negara bergerak antar cluster dari waktu ke waktu (misalnya, apakah negara dari Cluster 2 bermigrasi ke Cluster 3).
2. **Cakupan Variabel:** Model ini hanya didasarkan pada empat variabel kuantitatif. Faktor-faktor penting lainnya yang memengaruhi pembangunan, seperti tingkat pendidikan, stabilitas politik, atau indeks tata kelola (governance), tidak dimasukkan ke dalam model.
3. **Metodologi:** Penelitian ini secara eksklusif menggunakan algoritma K-Means. Algoritma clustering lain (seperti Hierarchical Clustering atau DBSCAN) mungkin dapat menghasilkan struktur cluster yang berbeda.

## 4.3. Saran

Berdasarkan kesimpulan dan keterbatasan di atas, berikut adalah beberapa saran untuk penelitian selanjutnya dan implikasi praktis:

- **Untuk Penelitian Selanjutnya:**
  - **Analisis Longitudinal (Time-Series):** Penelitian di masa depan dapat menggunakan data panel (time-series) Gapminder untuk menganalisis *lintasan pembangunan*. Ini dapat menjawab pertanyaan seperti: "Seberapa cepat negara-negara bergerak dari cluster 'Berkembang' ke 'Maju'?" atau "Apakah 'Mega-Populasi' tetap menjadi cluster yang stabil dari waktu ke waktu?"
  - **Penambahan Fitur:** Model dapat diperkaya dengan memasukkan variabel kualitatif (seperti **region**) atau variabel kuantitatif tambahan (seperti Indeks Pembangunan Manusia/IPM, indeks pendidikan, atau Gini ratio) untuk melihat apakah cluster yang lebih bernuansa dapat ditemukan.

- **Perbandingan Metode:** Melakukan perbandingan hasil antara K-Means dengan metode clustering lain (misalnya DBSCAN, yang dapat mengidentifikasi *outlier* secara lebih formal) akan bermanfaat untuk menguji robustitas temuan.
- **Untuk Implikasi Praktis:**
  - Segmentasi ini dapat digunakan oleh pembuat kebijakan dan organisasi internasional. Alih-alih menerapkan kebijakan "satu ukuran untuk semua", intervensi dapat dirancang khusus:
    - **Cluster 3** mungkin memerlukan fokus pada kebijakan dekarbonisasi dan lingkungan.
    - **Cluster 2** mungkin memerlukan fokus pada bantuan pembangunan ekonomi dan program kesehatan dasar.
    - **Cluster 1** memerlukan strategi unik yang memperhitungkan skala populasi mereka yang ekstrem dalam segala hal, mulai dari distribusi sumber daya hingga dampak lingkungan.