

Kelompok 1



Andiko Ramadani 3337230003



Ussy Cantika 3337230008



Ismet Maulana Azhari 3337230014

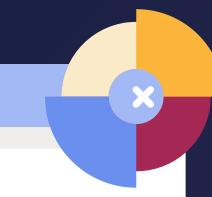
Latar Belakang



Indeks Pembangunan Manusia (IPM) adalah salah satu indikator yang esensial dalam menilai keberhasilan suatu wilayah dalam membangun kualitas hidup masyarakatnya secara menyeluruh. Namun realitas di Indonesia menunjukkan bahwa pencapaian IPM antar provinsi tidak merata. Ketimpangan ini menjadi sorotan penting dalam diskusi mengenai pemerataan pembangunan nasional, dan secara khusus mendorong perlunya pendekatan analitis yang mampu memetakan kondisi ini secara lebih objektif dan terukur. Salah satu pendekatan yang dapat digunakan untuk memahami kondisi tersebut adalah melalui analisis clustering menggunakan algoritma K-Means.



- Bagaimana proses pengelompokan provinsi-provinsi di Indonesia dapat dilakukan menggunakan algoritma K-Means berdasarkan
- paling optimal untuk merepresentasikan variasi pembangunan manusia di seluruh
- Apa saja karakteristik utama yang membedakan masing-masing klaster hasil pengelompokan, baik dari sisi IPM maupun dari komponen-komponen penyusunnya (AHH, RLS, dan PPP)?
- Bagaimana hasil pengelompokan tersebut dapat membantu dalam memahami struktur perbedaan pembangunan manusia antar provinsi secara lebih sistematis dan menyeluruh?



Tujuan Penelitian



Menerapkan algoritma
K-Means untuk
mengelompokkan
provinsi-provinsi di
Indonesia berdasarkan
empat variabel



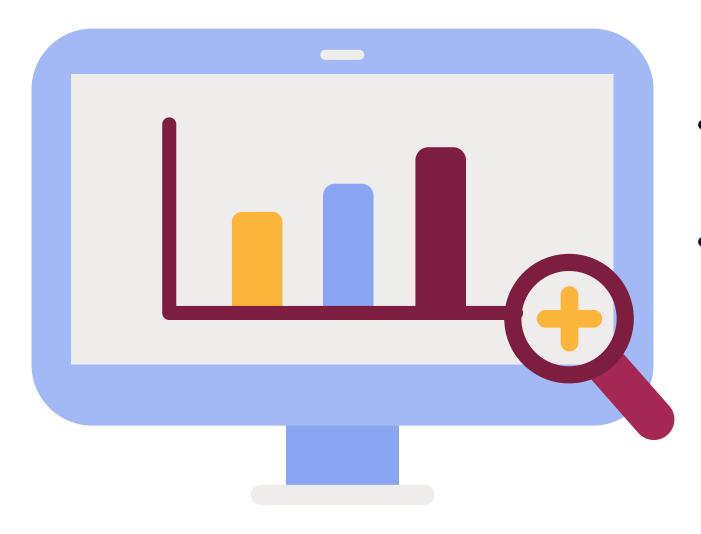
Menentukan jumlah klaster yang paling optimal dengan mempertimbangkan metrik evaluasi



Mengidentifikasi karakteristik utama dari setiap klaster



Memberikan
pemahaman yang
lebih dalam dan
menyeluruh mengenai
struktur perbedaan
pembangunan
manusia antar provinsi



Batasan Masalah

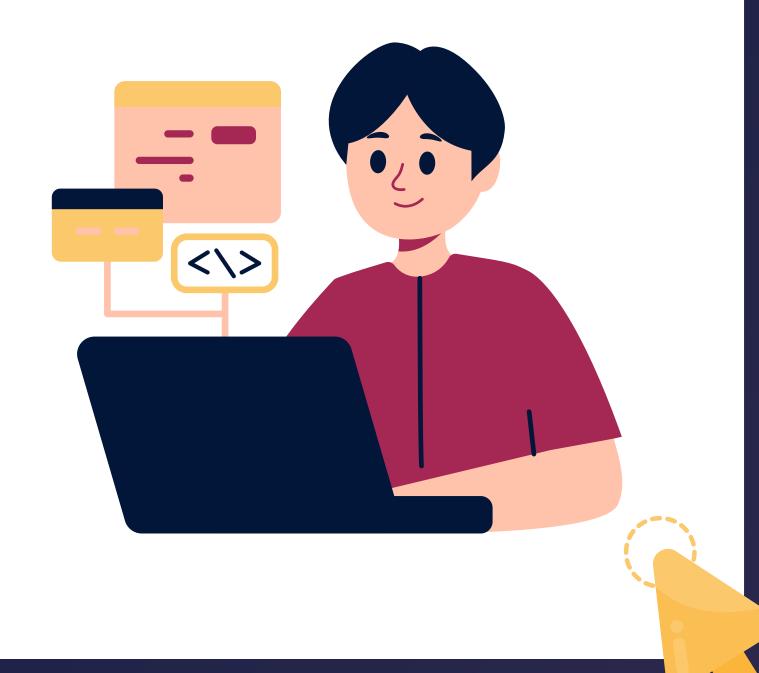
- Data yang digunakan dalam penelitian ini terbatas pada data provinsi di Indonesia dan tidak mencakup tingkat kabupaten atau kota.
- Variabel yang dianalisis meliputi empat indikator utama, yaitu: Angka Harapan Hidup (AHH), Rata-Rata Lama Sekolah (RLS), Pengeluaran per Kapita (PPP), serta nilai IPM sebagai indeks komposit. Variabel lain yang juga berpengaruh terhadap pembangunan manusia, seperti tingkat pengangguran, akses sanitasi, atau ketimpangan pendapatan, tidak dimasukkan dalam ruang lingkup analisis ini.



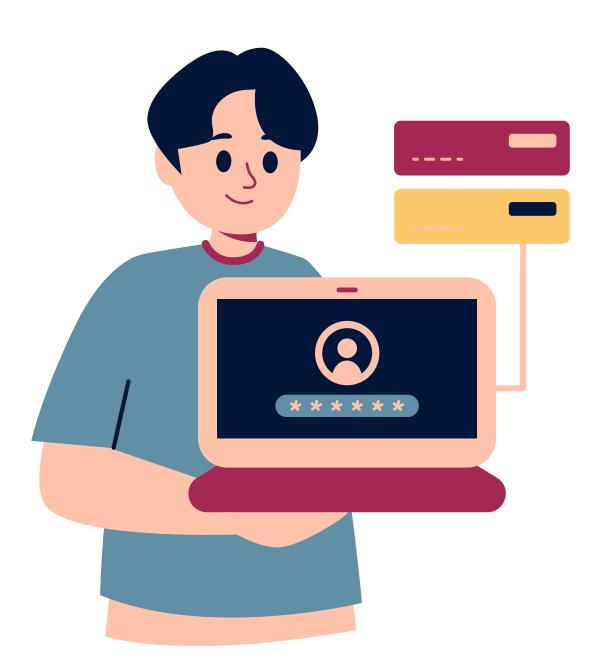


Batasan Masalah

- Seluruh data diambil dari sumber resmi, yaitu Badan Pusat Statistik (BPS), dan dibatasi hanya pada tahun 2022-2024.
- Pengelompokan dilakukan menggunakan algoritma K-Means, sehingga interpretasi hasil bersifat numerik dan berbasis pada pendekatan jarak antar titik dalam ruang fitur yang telah dinormalisasi. Potensi variasi hasil akibat sensitivitas algoritma terhadap inisialisasi centroid juga menjadi bagian dari keterbatasan yang disadari dalam penelitian ini.







Metodologi Penelitian

- Data Science Methodology
- Data Preparation
- Modeling with K-Means
- Evaluation and Interpretation



Pengumpulan Data

Berikut adalah tiga tautan resmi dari situs BPS tempat data diambil:

- Angka Harapan Hidup (AHH):
- https://www.bps.go.id/id/statisticstable/2/Mjl3MyMy/angka-harapan-hidup--ahh-menurut-provinsi-dan-jenis-kelamin--menggunakanuhh-hasil-sp2020-lf-.html
- Rata-rata Lama Sekolah (RLS):
 https://www.bps.go.id/id/statistics table/2/MTQyOSMy/rata-rata-lama-sekolah-penduduk-umur-15-tahun-ke-atas-menurut-provinsi.html
- Pengeluaran per Kapita Disesuaikan (PPP):
 https://www.bps.go.id/id/statistics-table/2/NDE2IzI=/-
 metode-baru-pengeluaran-per-kapita-disesuaikan.html





```
# Import library yang digunakan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from functools import reduce
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, davies_bouldin_score
```

```
# --- 1. DATA PREPARATION ---

daftarProvinsi = [
    "ACEH", "SUMATERA UTARA", "SUMATERA BARAT", "RIAU", "JAMBI", "SUMATERA SELATAN",
    "BENGKULU", "LAMPUNG", "KEP. BANGKA BELITUNG", "KEPULAUAN RIAU", "DKI JAKARTA",
    "JAWA BARAT", "JAWA TENGAH", "DI YOGYAKARTA", "JAWA TIMUR", "BANTEN", "BALI",
    "NUSA TENGGARA BARAT", "NUSA TENGGARA TIMUR", "KALIMANTAN BARAT", "KALIMANTAN TENGAH",
    "KALIMANTAN SELATAN", "KALIMANTAN TIMUR", "KALIMANTAN UTARA", "SULAWESI UTARA",
    "SULAWESI TENGAH", "SULAWESI SELATAN", "SULAWESI TENGGARA", "GORONTALO",
    "SULAWESI BARAT", "MALUKU", "MALUKU UTARA", "PAPUA", "PAPUA BARAT",
    "PAPUA TENGAH", "PAPUA PEGUNUNGAN", "PAPUA SELATAN", "PAPUA BARAT DAYA"

tahunProses = [2022, 2023, 2024]
pathDasar = './Raw Data/'
provinsiIndukPapua = ["PAPUA", "PAPUA BARAT"]
provinsiBaruPapua = ["PAPUA TENGAH", "PAPUA PEGUNUNGAN", "PAPUA SELATAN", "PAPUA BARAT DAYA"]
```

```
listDataPerTahun = []
for tahun in tahunProses:
    dataPengeluaran = (
        pd.read_csv(f'{pathDasar}[Metode Baru] Pengeluaran per Kapita Disesuaikan, {tahun}.csv', skiprows=1)
        .rename(columns={'Unnamed: 0': 'Provinsi','[Metode Baru] Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang/Tahun)': 'Pengeluaran per Kapita Disesuaikan (PPP)'})
        .assign(Provinsi=lambda df: df['Provinsi'].str.upper().replace({"D I YOGYAKARTA": "DI YOGYAKARTA"}))
        .drop_duplicates(subset=['Provinsi'], keep='first').query("Provinsi in @daftarProvinsi").assign(Tahun=tahun)
    dataPendidikan = (
        pd.read csv(f'{pathDasar}Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Provinsi, {tahun}.csv', skiprows=1)
        .rename(columns={'Unnamed: 0': 'Provinsi', 'Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Provinsi': 'Rata-Rata Lama Sekolah (RLS)'})
        .assign(Provinsi=lambda df: df['Provinsi'].str.upper().replace({"KEP. RIAU": "KEPULAUAN RIAU"}))
        .drop duplicates(subset=['Provinsi'], keep='first').query("Provinsi in @daftarProvinsi").assign(Tahun=tahun)
    dataAhh = pd.read_csv(f'{pathDasar}Angka Harapan Hidup (AHH) Menurut Provinsi dan Jenis Kelamin (menggunakan UHH hasil SP2020 LF), {tahun}.csv', skiprows=2)
    dataAhh = dataAhh.rename(columns={'Unnamed: 0': 'Provinsi'})
    dataAhh['Angka Harapan Hidup (AHH)'] = (pd.to numeric(dataAhh['Laki-laki'], errors='coerce') + pd.to numeric(dataAhh['Perempuan'], errors='coerce')) / 2
    dataAhh = dataAhh.assign(Provinsi=lambda df: df['Provinsi'].str.upper().replace({"KEP. RIAU": "KEPULAUAN RIAU"}))
    dataAhh = dataAhh.drop duplicates(subset=['Provinsi'], keep='first').query("Provinsi in @daftarProvinsi").assign(Tahun=tahun)
    listDfTahunan = [
        dataPengeluaran[['Provinsi', 'Tahun', 'Pengeluaran per Kapita Disesuaikan (PPP)']],
        dataPendidikan[['Provinsi', 'Tahun', 'Rata-Rata Lama Sekolah (RLS)']],
        dataAhh[['Provinsi', 'Tahun', 'Angka Harapan Hidup (AHH)']]
    dataTahunGabung = reduce(lambda kiri, kanan: pd.merge(kiri, kanan, on=['Provinsi', 'Tahun'], how='outer'), listDfTahunan)
    listDataPerTahun.append(dataTahunGabung)
dataGabungan = pd.concat(listDataPerTahun, ignore index=True).drop duplicates()
dataGabungan['Pengeluaran per Kapita Disesuaikan (PPP)'] = pd.to numeric(
    dataGabungan['Pengeluaran per Kapita Disesuaikan (PPP)'], errors='coerce'
dataGabungan['Rata-Rata Lama Sekolah (RLS)'] = pd.to_numeric(
    dataGabungan['Rata-Rata Lama Sekolah (RLS)'], errors='coerce'
```

```
--- Pengecekan Missing Value (Sebelum Imputasi) ---
Jumlah missing value per kolom:
Pengeluaran per Kapita Disesuaikan (PPP)
Rata-Rata Lama Sekolah (RLS)
Angka Harapan Hidup (AHH)
dtype: int64
--- Pengecekan Missing Value (Setelah Imputasi) ---
Jumlah missing value di kolom target setelah diisi:
Pengeluaran per Kapita Disesuaikan (PPP)
Rata-Rata Lama Sekolah (RLS)
Angka Harapan Hidup (AHH)
dtype: int64
```

```
# --- Pengecekan Missing Value (Sebelum Imputasi) ---
print("\n--- Pengecekan Missing Value (Sebelum Imputasi) ---")
missingValuesSebelum = dataGabungan.isnull().sum()
print("Jumlah missing value per kolom:")
# Hanya tampilkan kolom yang memiliki missing value
print(missingValuesSebelum[missingValuesSebelum > 0])
print("-" * 50)
# --- Imputasi Nilai Kosong Papua ---
kolomUntukImputasi = ['Pengeluaran per Kapita Disesuaikan (PPP)', 'Rata-Rata Lama Sekolah (RLS)', 'Angka Harapan Hidup (AHH)']
for kolom in kolomUntukImputasi:
    for tahun in tahunProses:
        rerataInduk = dataGabungan[(dataGabungan['Provinsi'].isin(provinsiIndukPapua)) & (dataGabungan['Tahun'] == tahun)][kolom].mean()
        kondisi = ((dataGabungan['Provinsi'].isin(provinsiBaruPapua)) & (dataGabungan['Tahun'] == tahun) & (dataGabungan[kolom].isnull()))
        dataGabungan.loc[kondisi, kolom] = rerataInduk
# --- Pengecekan Missing Value (Setelah Imputasi) ---
print("\n--- Pengecekan Missing Value (Setelah Imputasi) ---")
missingValuesSetelah = dataGabungan[kolomUntukImputasi].isnull().sum()
print("Jumlah missing value di kolom target setelah diisi:")
print(missingValuesSetelah)
print("-" * 50)
```

```
# --- Menampilkan Tabel Sebelum Menghitung IPM ---
dataTampil = dataGabungan.copy()
tipeKategoriProvinsi = pd.CategoricalDtype(categories=daftarProvinsi, ordered=True)
dataTampil['Provinsi'] = dataTampil['Provinsi'].astype(tipeKategoriProvinsi)
dataTampil = dataTampil.sort_values(by=['Provinsi', 'Tahun']).reset_index(drop=True)
dataTampil['Provinsi'] = dataTampil['Provinsi'].astype(str) + ' ' + dataTampil['Tahun'].astype(str)
dataTampil = dataTampil[['Provinsi', 'Pengeluaran per Kapita Disesuaikan (PPP)', 'Rata-Rata Lama Sekolah (RLS)', 'Angka Harapan Hidup (AHH)']]
for kolom in ['Rata-Rata Lama Sekolah (RLS)', 'Angka Harapan Hidup (AHH)']:
    dataTampil[kolom] = pd.to_numeric(dataTampil[kolom], errors='coerce').round(2)

print("\n--- Tabel Setelah Imputasi (Sebelum Perhitungan IPM) ---\n")
#print(dataTampil.to_string(index=False))
display(dataTampil.head(9).style.hide(axis="index"))
```

Tabel Setelah Imputasi (Sebelum Perhitungan IPM)						
Provinsi	Pengeluaran per Kapita Disesuaikan (PPP)	Rata-Rata Lama Sekolah (RLS)	Angka Harapan Hidup (AHH)			
ACEH 2022	9963.000000	9.790000	72.970000			
ACEH 2023	10334.000000	9.890000	73.110000			
ACEH 2024	10811.000000	9.950000	73.260000			
SUMATERA UTARA 2022	10848.000000	9.990000	73.440000			
SUMATERA UTARA 2023	11049.000000	10.070000	73.720000			
SUMATERA UTARA 2024	11460.000000	10.180000	73.960000			
SUMATERA BARAT 2022	11130.000000	9.510000	73.940000			
SUMATERA BARAT 2023	11380.000000	9.590000	74.200000			
SUMATERA BARAT 2024	11718.000000	9.720000	74.440000			

```
# --- Perhitungan IPM ---
def hitungIpm(ppp, rls, ahh):
    pppMin, pppMaks, rlsMin, rlsMaks, ahhMin, ahhMaks = 100, 75000, 0, 15, 20, 85
    indeksPendapatan = (np.log(ppp) - np.log(pppMin)) / (np.log(pppMaks) - np.log(pppMin))
    indeksPendidikan = (rls - rlsMin) / (rlsMaks - rlsMin)
    indeksKesehatan = (ahh - ahhMin) / (ahhMaks - ahhMin)
    return ((indeksPendapatan * indeksPendidikan * indeksKesehatan) ** (1/3) * 100)
# Hitung dan tambahkan kolom IPM
dataGabungan['Indeks Pembangunan Manusia (IPM)'] = hitungIpm(
    dataGabungan['Pengeluaran per Kapita Disesuaikan (PPP)'],
    dataGabungan['Rata-Rata Lama Sekolah (RLS)'],
    dataGabungan['Angka Harapan Hidup (AHH)']
# Format nama provinsi + tahun
dataGabungan['Provinsi'] = dataGabungan['Provinsi'].astype(tipeKategoriProvinsi)
dataGabungan = dataGabungan.sort_values(by=['Provinsi', 'Tahun']).reset_index(drop=True)
dataGabungan['Provinsi'] = dataGabungan['Provinsi'].astype(str) + ' ' + dataGabungan['Tahun'].astype(str)
# Ambil kolom yang dibutuhkan untuk klasterisasi
dataFinal = dataGabungan[['Provinsi',
                          'Angka Harapan Hidup (AHH)',
                          'Rata-Rata Lama Sekolah (RLS)',
                          'Pengeluaran per Kapita Disesuaikan (PPP)',
                          'Indeks Pembangunan Manusia (IPM)']].copy()
# Pembulatan angka
dataFinal[['Angka Harapan Hidup (AHH)',
           'Rata-Rata Lama Sekolah (RLS)',
           'Pengeluaran per Kapita Disesuaikan (PPP)',
           'Indeks Pembangunan Manusia (IPM)']] = \
    dataFinal[['Angka Harapan Hidup (AHH)',
               'Rata-Rata Lama Sekolah (RLS)',
               'Pengeluaran per Kapita Disesuaikan (PPP)',
               'Indeks Pembangunan Manusia (IPM)']].round(2)
# Simpan ke file CSV
dataFinal.to_csv('dataPreprocessing.csv', index=False)
print(f"\nData preparation selesai. File disimpan ke dataPreprocessing.csv")
print("\n--- Tabel Final untuk Preprocessing ---\n")
display(dataFinal.head(9).style.hide(axis="index"))
```

Output:

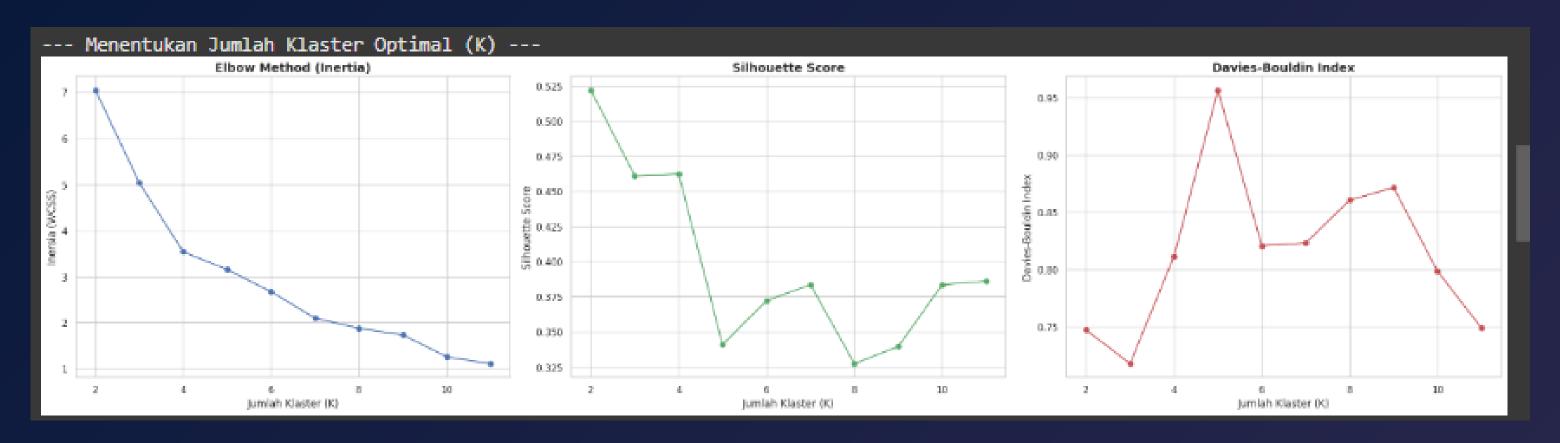
Data preparation selesai. File disimpan ke dataPreprocessing.csv
--- Tabel Final untuk Preprocessing --Provinsi Angka Harapan Hidup (AHH) Rata-Rata Lama Sekolah (RLS) Pengeluaran per Kapita Disesuaikan (PPP) Indeks Pembangunan Manusia

	PLOATURE	Angka narapan nioup (Ann)	Kata-Kata Lama Sekutan (KLS)	rengetuaran per kapita bisesuatkan (PPP)	Thucks Pembangunan Manusta (1PM)
	ACEH 2022	72.970000	9.790000	9963.000000	71.770000
	ACEH 2023	73.110000	9.890000	10334.000000	72.270000
	ACEH 2024	73.260000	9.950000	10811.000000	72.710000
SU	JMATERA UTARA 2022	73.440000	9.990000	10848.000000	72.920000
SU	JMATERA UTARA 2023	73.720000	10.070000	11049.000000	73.330000
SL	JMATERA UTARA 2024	73.960000	10.180000	11460.000000	73.890000
SL	JMATERA BARAT 2022	73.940000	9.510000	11130.000000	72.080000
SL	JMATERA BARAT 2023	74.200000	9.590000	11380.000000	72.510000
SU	JMATERA BARAT 2024	74.440000	9.720000	11718.000000	73.090000

```
# --- 2. DATA PREPROCESSING ---
# Baca data
dataUntukScaling = pd.read_csv('dataPreprocessing.csv')
# Inisialisasi scaler
scaler = MinMaxScaler()
# Pilih semua kolom numerik untuk di-scaling
fiturNumerik = ['Angka Harapan Hidup (AHH)',
                'Rata-Rata Lama Sekolah (RLS)',
                'Pengeluaran per Kapita Disesuaikan (PPP)',
                'Indeks Pembangunan Manusia (IPM)']
# Simpan hasil scaling dalam DataFrame baru
dataScaled = scaler.fit transform(dataUntukScaling[fiturNumerik])
dataScaledDF = pd.DataFrame(dataScaled, columns=[col + '_Scaled' for col in fiturNumerik])
# Gabungkan dengan kolom Provinsi
dataUntukScalingScaled = pd.concat([dataUntukScaling[['Provinsi']], dataScaledDF], axis=1)
# Simpan hasilnya
namaFileOutput = 'dataScaled.csv'
dataUntukScalingScaled.to_csv(namaFileOutput, index=False)
print(f"Scaling dengan MinMaxScaler selesai. Data disimpan ke {namaFileOutput}\n")
print("--- Tabel Dengan Semua Fitur yang Sudah di-Scaling ---\n")
display(dataUntukScalingScaled.head(9).style.hide(axis="index"))
Scaling dengan MinMaxScaler selesai. Data disimpan ke dataScaled.csv
--- Tabel Dengan Semua Fitur yang Sudah di-Scaling ---
            Provinsi Angka Harapan Hidup (AHH)_Scaled Rata-Rata Lama Sekolah (RLS)_Scaled Pengeluaran per Kapita Disesuaikan (PPP)_Scaled Indeks Pembangunan Manusia (IPM)_Scaled
           ACEH 2022
                                              0.641531
                                                                                                                                   0.298751
                                                                                                                                                                           0.670178
           ACEH 2023
                                              0.657773
                                                                                   0.749609
                                                                                                                                   0.324793
                                                                                                                                                                           0.688340
           ACEH 2024
                                                                                                                                   0.358276
                                                                                                                                                                           0.704323
                                              0.675174
                                                                                   0.758998
 SUMATERA UTARA 2022
                                                                                                                                   0.360873
                                                                                   0.765258
                                                                                                                                                                           0.711951
                                              0.696056
                                                                                   0.777778
                                                                                                                                   0.374982
                                                                                                                                                                           0.726843
 SUMATERA UTARA 2023
                                              0.728538
 SUMATERA UTARA 2024
                                              0.756381
                                                                                   0.794992
                                                                                                                                   0.403833
                                                                                                                                                                           0.747185
 SUMATERA BARAT 2022
                                              0.754060
                                                                                   0.690141
                                                                                                                                   0.380668
                                                                                                                                                                           0.681438
 SUMATERA BARAT 2023
                                              0.784223
                                                                                   0.702660
                                                                                                                                   0.398217
                                                                                                                                                                           0.697058
 SUMATERA BARAT 2024
                                              0.812065
                                                                                   0.723005
                                                                                                                                   0.421943
                                                                                                                                                                           0.718126
```

```
# Visualisasi hasil evaluasi untuk berbagai nilai K
sns.set theme(style="whitegrid")
fig, axes = plt.subplots(1, 3, figsize=(24, 6))
# Plot Elbow Method (Inertia/WCSS)
axes[0].plot(rentangK, inersia, 'bo-')
axes[0].set title('Elbow Method (Inertia)', fontsize=14, weight='bold')
axes[0].set xlabel('Jumlah Klaster (K)')
axes[0].set_ylabel('Inersia (WCSS)')
# Plot Silhouette Score
axes[1].plot(rentangK, skorSiluet, 'go-')
axes[1].set_title('Silhouette Score', fontsize=14, weight='bold')
axes[1].set_xlabel('Jumlah Klaster (K)')
axes[1].set_ylabel('Silhouette Score')
# Plot Davies-Bouldin Index
axes[2].plot(rentangK, skorDaviesBouldin, 'ro-')
axes[2].set_title('Davies-Bouldin Index', fontsize=14, weight='bold')
axes[2].set xlabel('Jumlah Klaster (K)')
axes[2].set_ylabel('Davies-Bouldin Index')
plt.tight_layout()
plt.show()
```

```
# --- Memuat Data Hasil Preprocessing ---
dataKlaster = pd.read_csv('dataScaled.csv')
# Siapkan data untuk clustering (4 fitur hasil scaling)
fiturUntukKlaster = [
    'Angka Harapan Hidup (AHH)_Scaled',
    'Rata-Rata Lama Sekolah (RLS)_Scaled',
    'Pengeluaran per Kapita Disesuaikan (PPP)_Scaled',
    'Indeks Pembangunan Manusia (IPM)_Scaled'
dataUntukKlaster = dataKlaster[fiturUntukKlaster].values
# --- Penentuan Jumlah Klaster (K) Optimal ---
print("--- Menentukan Jumlah Klaster Optimal (K) ---")
rentangK = range(2, 12)
inersia = []
skorSiluet = []
skorDaviesBouldin = []
for k in rentangK:
    # Inisialisasi dan latih model KMeans
    modelKMeans = KMeans(n_clusters=k, n_init='auto', random_state=42)
    modelKMeans.fit(dataUntukKlaster)
    # Dapatkan label klaster
    label = modelKMeans.labels_
    # Simpan metrik evaluasi
    inersia.append(modelKMeans.inertia_)
    skorSiluet.append(silhouette_score(dataUntukKlaster, label))
    skorDaviesBouldin.append(davies_bouldin_score(dataUntukKlaster, label))
```



```
# --- Evaluasi Hasil Klasterisasi ---
# Cetak semua skor evaluasi
print("\n--- Evaluasi K untuk Setiap Metrik ---")
print(f"{'K':<5}{'Inertia':<15}{'Silhouette':<15}{'Davies-Bouldin':<20}")</pre>
for i, k in enumerate(rentangK):
   print(f"{k:<5}{inersia[i]:<15.2f}{skorSiluet[i]:<15.4f}{skorDaviesBouldin[i]:<20.4f}")</pre>
# Tentukan K optimal berdasarkan dua metrik:
# 1. Silhouette Score maksimum
# 2. Davies-Bouldin Index minimum
# K dengan silhouette tertinggi
k silhouette = rentangK[np.argmax(skorSiluet)]
# K dengan Davies-Bouldin Index terendah
k_dbi = rentangK[np.argmin(skorDaviesBouldin)]
print(f"\nBerdasarkan Silhouette Score tertinggi → K optimal = {k_silhouette}")
print(f"Berdasarkan Davies-Bouldin Index terendah → K optimal = {k_dbi}")
# --- Tentukan K final ---
# Jika keduanya sama, langsung pakai
# Jika berbeda, pilih salah satu atau kompromi (misalnya pakai Silhouette karena lebih stabil)
kOptimal = k_dbi # atau bisa pakai k_dbi atau rumus kompromi jika kamu mau
print(f"\nJumlah klaster yang digunakan (berdasarkan evaluasi) adalah: {kOptimal}")
# --- Penerapan Model K-Means dengan K Optimal ---
print(f"\nMenerapkan clustering dengan K={kOptimal}...")
# Latih model final dengan K optimal
modelFinal = KMeans(n clusters=kOptimal, n init='auto', random state=42)
labelPrediksi = modelFinal.fit_predict(dataUntukKlaster)
# Tambahkan hasil label klaster ke DataFrame
dataKlaster['Klaster'] = labelPrediksi
```

```
--- Evaluasi K untuk Setiap Metrik ---
    Inertia
                                  Davies-Bouldin
                   Silhouette
    7.05
                   0.5219
                                  0.7470
    5.04
                   0.4610
                                  0.7177
    3.54
                   0.4622
                                  0.8114
    3.15
                   0.3410
                                  0.9564
     2.67
                   0.3722
                                  0.8204
     2.10
                   0.3835
                                  0.8230
                   0.3274
                                  0.8603
    1.88
9
    1.73
                   0.3396
                                  0.8713
    1.26
                   0.3836
                                  0.7983
11 1.11
                   0.3861
                                  0.7487
Berdasarkan Silhouette Score tertinggi → K optimal = 2
Berdasarkan Davies-Bouldin Index terendah → K optimal = 3
Jumlah klaster yang digunakan (berdasarkan evaluasi) adalah: 3
Menerapkan clustering dengan K=3...
```

```
# --- Analisis dan Pengurutan Hasil Klaster ---
# Agar label klaster lebih bermakna (0=terendah, 1=menengah, dst.), kita urutkan
rerataIpmPerKlaster = dataKlaster.groupby('Klaster')['Indeks Pembangunan Manusia (IPM)_Scaled'].mean().sort_values()

# Buat pemetaan dari label acak ke label berurutan
pemetaanKlaster = {labelAsli: labelBaru for labelBaru, labelAsli in enumerate(rerataIpmPerKlaster.index)}

# Terapkan pemetaan ke kolom Klaster
dataKlaster['Klaster'] = dataKlaster['Klaster'].map(pemetaanKlaster)

# Urutkan hasil akhir agar mudah dibaca
dataKlaster = dataKlaster.sort_values(by=['Klaster', 'Provinsi']).reset_index(drop=True)

dataKlaster.to_csv('hasilClustering.csv', index=False)

print(f"\nClustering selesai. Data hasil clustering disimpan ke hasilClustering.csv\n")

#print(dataKlaster.to_string(index=False))

display(dataKlaster.head(150).style.hide(axis="index"))
```

Output:

Clustering selesai. Data hasil clustering disimpan ke hasilClustering.csv

--- Tabel Hasil Akhir Clustering ---

Provinsi	Angka Harapan Hidup (AHH)_Scaled	Rata-Rata Lama Sekolah (RLS)_Scaled	Pengeluaran per Kapita Disesuaikan (PPP)_Scaled	Indeks Pembangunan Manusia (IPM)_Scaled	Klaster
GORONTALO 2022	0.328306	0.514867	0.349572	0.508899	0
GORONTALO 2023	0.359629	0.528951	0.376386	0.528151	0
GORONTALO 2024	0.386311	0.553991	0.409378	0.554668	0
MALUKU 2022	0.321346	0.824726	0.222448	0.651653	0
MALUKU 2023	0.354988	0.826291	0.250667	0.665819	0
MALUKU 2024	0.381671	0.835681	0.279166	0.683255	0
MALUKU UTARA 2022	0.357309	0.704225	0.188895	0.580458	0
MALUKU UTARA 2023	0.390951	0.705790	0.219500	0.596077	0
MALUKU UTARA 2024	0.424594	0.719875	0.253615	0.618598	0

NUSA TENGGARA BARAT 2022	0.495360	0.502347	0.349151	0.524155	0
NUSA TENGGARA TIMUR 2022	0.453596	0.492958	0.152323	0.458046	0
NUSA TENGGARA TIMUR 2023	0.484919	0.502347	0.178366	0.476571	0
NUSA TENGGARA TIMUR 2024	0.515081	0.528951	0.198442	0.503088	o
PAPUA 2022	0.068445	0.345853	0.101011	0.295677	0
PAPUA 2023	0.090487	0.350548	0.130212	0.311660	0
PAPUA 2024	0.357309	0.876369	0.374140	0.725754	0
PAPUA BARAT 2022	0.097448	0.788732	0.168047	0.581547	0
PAPUA BARAT 2023	0.129930	0.788732	0.189316	0.593534	0
PAPUA BARAT 2024	0.125290	0.746479	0.217465	0.579005	0
PAPUA BARAT DAYA 2022	0.082367	0.566510	0.134529	0.445332	0
PAPUA BARAT DAYA 2023	0.110209	0.569640	0.159764	0.459135	0
PAPUA BARAT DAYA 2024	0.305104	0.841941	0.212411	0.655285	0

Output:

JAWA TENGAH 2022	0.874710	0.513302	0.398006	0.593171	- 1
JAWA TENGAH 2023	0.897912	0.522692	0.430156	0.609517	1
JAWA TENGAH 2024	0.923434	0.527387	0.461112	0.622230	1
JAWA TIMUR 2022	0.834107	0.532081	0.441176	0.609517	1
JAWA TIMUR 2023	0.870070	0.536776	0.471290	0.623320	1
JAWA TIMUR 2024	0.892111	0.561815	0.501544	0.648384	1
KALIMANTAN BARAT 2022	0.706497	0.469484	0.256072	0.507810	1
KALIMANTAN BARAT 2023	0.734339	0.480438	0.288011	0.527061	1
KALIMANTAN BARAT 2024	0.761021	0.497653	0.323880	0.550672	1
KALIMANTAN SELATAN 2022	0.733179	0.593114	0.474660	0.640756	1
KALIMANTAN SELATAN 2023	0.764501	0.602504	0.508634	0.657828	1
KALIMANTAN SELATAN 2024	0.788863	0.613459	0.539941	0.673810	1
KALIMANTAN TENGAH 2022	0.691415	0.615023	0.403692	0.633127	1
KALIMANTAN TENGAH 2023	0.714617	0.621283	0.433174	0.646567	1
KALIMANTAN TENGAH 2024	0.736659	0.636933	0.463007	0.665819	1

KALIMANTAN TIMUR 2022	0.820186	0.787167	0.486733	0.770432	1
KALIMANTAN UTARA 2022	0.709977	0.680751	0.255721	0.637486	1
KALIMANTAN UTARA 2023	0.714617	0.693271	0.282676	0.652742	1
KALIMANTAN UTARA 2024	0.716937	0.701095	0.315176	0.666546	1
KEP. BANGKA BELITUNG 2022	0.730858	0.543036	0.537063	0.621141	1
KEP. BANGKA BELITUNG 2023	0.756381	0.557121	0.553278	0.636397	1
KEP. BANGKA BELITUNG 2024	0.781903	0.575900	0.558753	0.653106	1
LAMPUNG 2022	0.761021	0.549296	0.324933	0.583727	1
LAMPUNG 2023	0.787703	0.566510	0.355328	0.605521	1
LAMPUNG 2024	0.813225	0.579030	0.389653	0.624773	1
NUSA TENGGARA BARAT 2023	0.537123	0.514867	0.378211	0.544497	1
NUSA TENGGARA BARAT 2024	0.564965	0.532081	0.414081	0.567018	1
RIAU 2022	0.761021	0.694836	0.382634	0.685797	1
RIAU 2023	0.788863	0.704225	0.402990	0.699600	1
RIAU 2024	0.815545	0.718310	0.431700	0.718126	1

Output:

	SULAWESI SELATAN 2022	0.698376	0.624413	0.401727	0.639303	1
	SULAWESI SELATAN 2023	0.723898	0.629108	0.430577	0.652379	1
	SULAWESI SELATAN 2024	0.748260	0.644757	0.461042	0.671631	1
	SULAWESI TENGGARA 2022	0.500000	0.702660	0.280851	0.627316	1
	SULAWESI TENGGARA 2023	0.510441	0.707355	0.309561	0.639303	1
	SULAWESI TENGGARA 2024	0.522042	0.726135	0.343886	0.660371	1
	SULAWESI UTARA 2022	0.720418	0.746479	0.384108	0.710134	1
	SULAWESI UTARA 2023	0.750580	0.757433	0.406430	0.725754	1
	SULAWESI UTARA 2024	0.777262	0.769953	0.441598	0.745005	1
	SUMATERA BARAT 2022	0.754060	0.690141	0.380668	0.681438	1
	SUMATERA BARAT 2023	0.784223	0.702660	0.398217	0.697058	1
	SUMATERA BARAT 2024	0.812065	0.723005	0.421943	0.718126	1
7	SUMATERA SELATAN 2022	0.740139	0.582160	0.379194	0.614239	1

SUMATERA SELATAN 2022	0.740139	0.582160	0.3791
SUMATERA SELATAN 2023	0.772622	0.594679	0.404€
SUMATERA SELATAN 2024	0.798144	0.607199	0.4427
SUMATERA UTARA 2022	0.696056	0.765258	0.3608
SUMATERA UTARA 2023	0.728538	0.777778	0.3749
SUMATERA UTARA 2024	0.756381	0.794992	0.4038
BALI 2022	0.837587	0.726135	0.5780
BALI 2023	0.870070	0.726135	0.6089
BALI 2024	0.896752	0.746479	0.6467

Output:

SUMATERA BARAT 2022	0.754060	0.690141	0.380668	0.681438	1
SUMATERA BARAT 2023	0.784223	0.702660	0.398217	0.697058	1
SUMATERA BARAT 2024	0.812065	0.723005	0.421943	0.718126	1
SUMATERA SELATAN 2022	0.740139	0.582160	0.379194	0.614239	1
SUMATERA SELATAN 2023	0.772622	0.594679	0.404675	0.632401	1
SUMATERA SELATAN 2024	0.798144	0.607199	0.442791	0.651653	1
SUMATERA UTARA 2022	0.696056	0.765258	0.360873	0.711951	1
SUMATERA UTARA 2023	0.728538	0.777778	0.374982	0.726843	1
SUMATERA UTARA 2024	0.756381	0.794992	0.403833	0.747185	1
BALI 2022	0.837587	0.726135	0.578057	0.755903	2
BALI 2023	0.870070	0.726135	0.608943	0.766073	2
BALI 2024	0.896752	0.746479	0.646708	0.788231	2
DI YOGYAKARTA 2022	0.894432	0.777778	0.615962	0.800581	2
DI YOGYAKARTA 2023	0.916473	0.791862	0.646989	0.817653	2
DI YOGYAKARTA 2024	0.957077	0.802817	0.677664	0.835089	2
DKI JAKARTA 2022	0.947796	0.970266	0.927980	0.966219	2
DKI JAKARTA 2023	0.979118	0.989045	0.959287	0.985470	2
DKI JAKARTA 2024	1.000000	1.000000	1.000000	1.000000	2
KALIMANTAN TIMUR 2023	0.852668	0.793427	0.526113	0.786778	2
KALIMANTAN TIMUR 2024	0.882831	0.801252	0.567598	0.803850	2

KEPULAUAN RIAU 2022	0.839907	0.838811	0.615050	0.827824	2
KEPULAUAN RIAU 2023	0.873550	0.848200	0.652183	0.844533	2
KEPULAUAN RIAU 2024	0.897912	0.868545	0.692545	0.866328	2

Evaluasi Clustering

```
# --- 4. EVALUASI CLUSTERING ---
# --- Memuat Data Hasil Clustering ---
dataHasil = pd.read_csv('hasilClustering.csv')
# Ambil semua fitur yang digunakan dalam clustering (harus yang scaled)
fiturYangDigunakan = [
    'Indeks Pembangunan Manusia (IPM)_Scaled',
    'Angka Harapan Hidup (AHH)_Scaled',
    'Rata-Rata Lama Sekolah (RLS)_Scaled',
    'Pengeluaran per Kapita Disesuaikan (PPP)_Scaled'
dataUntukMetrik = dataHasil[fiturYangDigunakan].values
labelKlaster = dataHasil['Klaster'].values
# --- Perhitungan Metrik Kuantitatif ---
print("--- Menghitung Metrik Evaluasi ---")
# 1. Inertia (Within-Cluster Sum of Squares)
kOptimal = dataHasil['Klaster'].nunique()
modelFinal = KMeans(n_clusters=kOptimal, n_init='auto', random_state=42)
modelFinal.fit(dataUntukMetrik)
nilaiInersia = modelFinal.inertia_
print(f"Inertia (WCSS): {nilaiInersia:.2f}")
# 2. Silhouette Score
skorSiluet = silhouette_score(dataUntukMetrik, labelKlaster)
print(f"Silhouette Score: {skorSiluet:.4f}")
# 3. Davies-Bouldin Index
skorDaviesBouldin = davies_bouldin_score(dataUntukMetrik, labelKlaster)
print(f"Davies-Bouldin Index: {skorDaviesBouldin:.4f}")
```

Output:

--- Menghitung Metrik Evaluasi ---

Inertia (WCSS): 5.45 Silhouette Score: 0.4610 Davies-Bouldin Index: 0.7177

Evaluasi Clustering

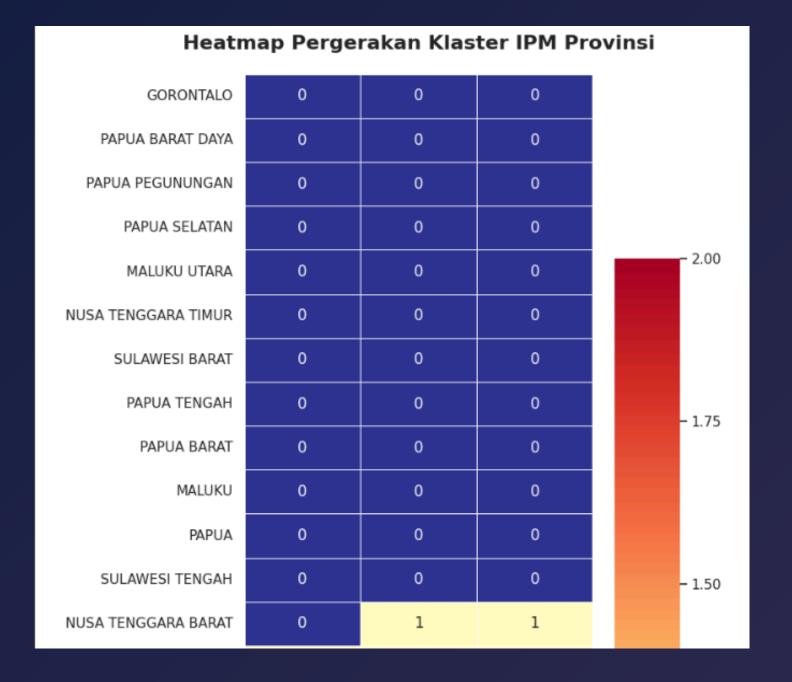
```
# --- ANALISIS PERGERAKAN KLASTER ---
# Muat data hasil clustering
dataHasil = pd.read_csv('hasilClustering.csv')
dataHasil['Tahun'] = pd.to_numeric(dataHasil['Provinsi'].str[-4:])
dataHasil['NamaProvinsi'] = dataHasil['Provinsi'].str[:-5]
tabelPerubahan = dataHasil.pivot_table(
    index='NamaProvinsi',
    columns='Tahun',
    values='Klaster'
).rename_axis(index='Provinsi', columns=None)
# --- Analisis Tren Kenaikan, Stagnan, atau Penurunan ---
def tentukanTren(row):
    # Periksa apakah data untuk tahun awal dan akhir tersedia
    if pd.notna(row[2022]) and pd.notna(row[2024]):
        if row[2024] > row[2022]:
            return 'Peningkatan'
        elif row[2024] < row[2022]:
            return 'Penurunan'
```

Tabel Perubahan Kla	ster Provinsi	(2022-2024) -		
	Klaster 2022	Klaster 2023	Klaster 2024	Tren
Provinsi				
ACEH	1.0	1.0	1.0	Stagnan
BALI	2.0	2.0	2.0	Stagnan
BANTEN	1.0	1.0	1.0	Stagnan
BENGKULU	1.0	1.0	1.0	Stagnan
DI YOGYAKARTA	2.0	2.0	2.0	Stagnan
DKI JAKARTA	2.0	2.0	2.0	Stagnan
GORONTALO	0.0	0.0	0.0	Stagnan
JAMBI	1.0	1.0	1.0	Stagnan
JAWA BARAT	1.0	1.0	1.0	Stagnan
JAWA TENGAH	1.0	1.0	1.0	Stagnan
JAWA TIMUR	1.0	1.0	1.0	Stagnan
KALIMANTAN BARAT	1.0	1.0	1.0	Stagnan
KALIMANTAN SELATAN	1.0	1.0	1.0	Stagnan
KALIMANTAN TENGAH	1.0	1.0	1.0	Stagnan
KALIMANTAN TIMUR	1.0	2.0	2.0	Peningkatan
KALIMANTAN UTARA	1.0	1.0	1.0	Stagnan
KEP. BANGKA BELITUNG	1.0	1.0	1.0	Stagnan
KEPULAUAN RIAU	2.0	2.0	2.0	Stagnan
LAMPUNG	1.0	1.0	1.0	Stagnan
MALUKU	0.0	0.0	0.0	Stagnan
MALUKU UTARA	0.0	0.0	0.0	Stagnan
NUSA TENGGARA BARAT	0.0	1.0	1.0	Peningkatan
NUSA TENGGARA TIMUR	0.0	0.0	0.0	Stagnan

0.0	0.0	0.0	Stagnan
0.0	0.0	0.0	Stagnan
0.0	0.0	0.0	Stagnan
0.0	0.0	0.0	Stagnan
0.0	0.0	0.0	Stagnan
0.0	0.0	0.0	Stagnan
1.0	1.0	1.0	Stagnan
0.0	0.0	0.0	Stagnan
1.0	1.0	1.0	Stagnan
0.0	0.0	0.0	Stagnan
1.0	1.0	1.0	Stagnan
1.0	1.0	1.0	Stagnan
1.0	1.0	1.0	Stagnan
1.0	1.0	1.0	Stagnan
1.0	1.0	1.0	Stagnan
	0.0 0.0 0.0 0.0 0.0 1.0 0.0 1.0 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0

Visualisasi Heatmap Pergerakan Klaster

```
# --- VISUALISASI HEATMAP PERGERAKAN KLASTER ---
# Ekstrak tahun dan nama provinsi
dataHasil['Tahun'] = pd.to_numeric(dataHasil['Provinsi'].str[-4:])
dataHasil['NamaProvinsi'] = dataHasil['Provinsi'].str[:-5]
# Buat pivot table untuk pergerakan klaster
tabelPergerakan = dataHasil.pivot_table(
    index='NamaProvinsi',
   columns='Tahun',
    values='Klaster'
).rename_axis(index='Provinsi', columns=None)
# Urutkan berdasarkan rata-rata klaster atau klaster di tahun terakhir
tabelPergerakan['RataRata'] = tabelPergerakan.mean(axis=1)
tabelPergerakan = tabelPergerakan.sort_values('RataRata').drop('RataRata', axis=1)
# --- HEATMAP PERGERAKAN KLASTER ---
plt.figure(figsize=(8, 20))
# Buat heatmap
sns.heatmap(tabelPergerakan[[2022, 2023, 2024]],
           annot=True, fmt='.0f', cmap='RdYlBu_r',
           cbar_kws={'label': 'Klaster IPM'},
           linewidths=0.5)
plt.title('Heatmap Pergerakan Klaster IPM Provinsi',
          fontsize=16, fontweight='bold', pad=20)
plt.xlabel('Tahun', fontsize=12, fontweight='bold')
plt.ylabel('Provinsi', fontsize=12, fontweight='bold')
plt.tight_layout()
plt.show()
```



Visualisasi Heatmap Pergerakan Klaster

	SOLAWESI TENGAH	·	·	<u> </u>	- 1.50
	NUSA TENGGARA BARAT	0	1	1	
	JAMBI	1	1	1	
	BENGKULU	1	1	1	
	ACEH	1	1	1	- 1.25
	JAWA BARAT	1	1	1	
	JAWA TENGAH	1	1	1	
insi	JAWA TIMUR	1	1	1	- 1.00 à
Provinsi	KEP. BANGKA BELITUNG	1	1	1	- 1.00 \$
	KALIMANTAN BARAT	1	1	1	
	KALIMANTAN UTARA	1	1	1	
	KALIMANTAN TENGAH	1	1	1	- 0.75
	KALIMANTAN SELATAN	1	1	1	
	RIAU	1	1	1	
	LAMPUNG	1	1	1	



Hasil Klasterisasi K-Means

Pada tahap ini, dilakukan klasterisasi terhadap data gabungan seluruh provinsi di Indonesia dari tahun 2022 hingga 2024. Setiap entitas data merupakan kombinasi nama provinsi dan tahun, sehingga analisis mencakup dimensi wilayah dan waktu secara bersamaan. Empat variabel utama yang digunakan adalah Angka Harapan Hidup (AHH), Rata-Rata Lama Sekolah (RLS), Pengeluaran per Kapita (PPP), dan Indeks Pembangunan Manusia (IPM), yang telah dinormalisasi dengan Min-Max Scaling.

Algoritma yang digunakan adalah K-Means dengan metode inisialisasi K-Means++. Evaluasi jumlah klaster terbaik dilakukan menggunakan tiga metode: Elbow Method, Silhouette Score, dan Davies-Bouldin Index. Hasil evaluasi menunjukkan bahwa jumlah klaster optimal adalah tiga (K = 3). Secara substansi, pembagian ini dianggap paling representatif untuk menggambarkan variasi pembangunan manusia: dari yang tertinggal, sedang berkembang, hingga yang mapan.

Komposisi Klaster

Setelah model K-Means diterapkan dengan K = 3, terbentuk tiga kelompok utama yang menggambarkan variasi tingkat capaian pembangunan manusia. Masing-masing klaster memiliki ciri khas berdasarkan rata-rata indikator IPM. Klaster 0 mencerminkan provinsi dengan IPM rendah, klaster 1 mewakili provinsi dengan capaian sedang, dan klaster 2 terdiri dari provinsi dengan IPM tinggi.

Karena data yang diklaster mencakup tiga tahun, hasil ini juga menunjukkan tren perubahan. Provinsi yang tetap berada dalam satu klaster selama tiga tahun menunjukkan konsistensi pembangunannya, sedangkan yang berpindah klaster menggambarkan adanya perubahan, baik kemajuan maupun penurunan.

a. Klaster 0 – Pembangunan Rendah

Kelompok ini mencakup provinsi dengan capaian pembangunan manusia paling rendah. Indikator AHH, RLS, dan PPP semuanya berada di bawah rata-rata nasional. Provinsi seperti Papua, Papua Selatan, Papua Pegunungan, Papua Barat Daya, dan NTT secara konsisten masuk klaster ini. Klaster ini didominasi wilayah timur Indonesia, yang masih menghadapi hambatan geografis dan keterbatasan infrastruktur.

b. Klaster 1 – Pembangunan Menengah

Klaster ini terdiri dari provinsi yang sedang berada dalam transisi pembangunan. Indikator AHH dan RLS menunjukkan peningkatan, namun PPP dan IPM belum cukup tinggi. Contoh provinsi: Sumatera Utara, Jawa Barat, Kalimantan Selatan, Sulawesi Selatan. Wilayah-wilayah ini memiliki potensi besar untuk naik ke klaster tinggi jika pertumbuhan terus dikawal secara konsisten.

c. Klaster 2 – Pembangunan Tinggi

Provinsi dalam klaster ini mencatat IPM yang tinggi dan stabil. AHH dan RLS di atas rata-rata nasional, PPP kuat, serta sistem pembangunan relatif mapan. Provinsi seperti DKI Jakarta, DI Yogyakarta, Bali, dan Kepulauan Riau termasuk di dalamnya. Wilayah-wilayah ini umumnya sudah memiliki infrastruktur baik, akses layanan publik merata, dan kebijakan pembangunan yang berkelanjutan.

d. Distribusi Geografis dan Catatan Umum

Secara geografis:

- Klaster 2: Umumnya wilayah barat dan pusat ekonomi nasional.
- Klaster 1: Wilayah tengah dan sebagian timur Indonesia.
- Klaster 0: Didominasi wilayah timur dan daerah tertinggal.

Namun, pola ini tidak mutlak. Beberapa provinsi menunjukkan dinamika yang tidak mengikuti pola umum, menandakan pentingnya kebijakan berbasis data dalam merespons kebutuhan nyata di tiap daerah.

Perubahan IPM dan Dampaknya terhadap Klaster

Setelah proses klasterisasi selesai, analisis dilanjutkan dengan melacak perubahan posisi provinsi dalam klaster dari tahun 2022 hingga 2024. Visualisasi dilakukan dalam bentuk heatmap, di mana setiap baris menunjukkan satu provinsi dan setiap kolom mewakili tahun. Warna menunjukkan jenis klaster:

- Klaster 0: IPM rendah
- Klaster 1: IPM menengah
- Klaster 2: IPM tinggi

Sebagian besar provinsi tetap berada dalam klaster yang sama, menandakan bahwa meskipun nilai IPM meningkat, perubahannya belum cukup besar untuk mengubah posisi relatif terhadap provinsi lain.

Perubahan IPM dan Dampaknya terhadap Klaster

multidimensi pembangunan manusia.

Namun, terdapat beberapa provinsi yang mengalami peningkatan klaster, seperti:

- Nusa Tenggara Barat (NTB): naik dari klaster 0 ke klaster 1 (2022 → 2023), dan bertahan.
- Kalimantan Timur: naik dari klaster 1 ke klaster 2, didorong oleh peningkatan IPM dan PPP.

Tidak ada provinsi yang turun ke klaster lebih rendah, yang menunjukkan bahwa secara nasional, pembangunan manusia cenderung mengalami tren positif meskipun dengan kecepatan yang berbeda-beda antar daerah.

Penting dipahami bahwa perpindahan klaster tidak hanya bergantung pada IPM saja, melainkan kombinasi seluruh indikator (AHH, RLS, PPP, dan IPM) secara proporsional. Artinya, meskipun IPM naik, jika provinsi lain juga naik secara serupa, maka posisi relatif bisa tetap. Selain itu, beberapa provinsi menunjukkan sinyal mendekati batas atas klaster, artinya mereka memiliki potensi kuat untuk naik ke klaster berikutnya dalam waktu dekat. Secara keseluruhan, analisis ini lebih informatif dibanding sekadar melihat ranking IPM,

karena mencerminkan perbandingan struktur dan dinamika antar daerah dalam konteks

Perubahan IPM dan Dampaknya terhadap Klaster

multidimensi pembangunan manusia.

Namun, terdapat beberapa provinsi yang mengalami peningkatan klaster, seperti:

- Nusa Tenggara Barat (NTB): naik dari klaster 0 ke klaster 1 (2022 → 2023), dan bertahan.
- Kalimantan Timur: naik dari klaster 1 ke klaster 2, didorong oleh peningkatan IPM dan PPP.

Tidak ada provinsi yang turun ke klaster lebih rendah, yang menunjukkan bahwa secara nasional, pembangunan manusia cenderung mengalami tren positif meskipun dengan kecepatan yang berbeda-beda antar daerah.

Penting dipahami bahwa perpindahan klaster tidak hanya bergantung pada IPM saja, melainkan kombinasi seluruh indikator (AHH, RLS, PPP, dan IPM) secara proporsional. Artinya, meskipun IPM naik, jika provinsi lain juga naik secara serupa, maka posisi relatif bisa tetap. Selain itu, beberapa provinsi menunjukkan sinyal mendekati batas atas klaster, artinya mereka memiliki potensi kuat untuk naik ke klaster berikutnya dalam waktu dekat. Secara keseluruhan, analisis ini lebih informatif dibanding sekadar melihat ranking IPM, karena mencerminkan perbandingan struktur dan dinamika antar daerah dalam konteks

Evaluasi dan Keterbatasan Model

Evaluasi Klasterisasi (K = 3):

Pemilihan jumlah klaster dievaluasi dengan tiga metrik:

- Elbow Method: menunjukkan titik optimal di K = 3.
- Silhouette Score: tertinggi di K = 2, namun K = 3 masih cukup baik.
- Davies-Bouldin Index: nilai terendah juga ditemukan pada K = 3.

K = 3 dipilih karena secara statistik efisien, dan secara substansi mewakili tiga kategori pembangunan: rendah, sedang, dan tinggi.

Keterbatasan K-Means:

- Mengasumsikan klaster berbentuk bulat dan seimbang.
- Sensitif terhadap inisialisasi centroid (diatasi dengan K-Means++).
- Jumlah klaster harus ditentukan di awal.
- Bergantung pada skala data, sehingga perlu normalisasi.

Refleksi:

Meski K-Means memberikan hasil yang solid dan mudah ditafsirkan, metode lain seperti DBSCAN atau GMM bisa dipertimbangkan untuk data yang lebih kompleks.

Pola Pembangunan Wilayah Berdasarkan Klaster

- Klaster 2 Wilayah Mapan: DKI Jakarta, Yogyakarta, Bali, dan Kalimantan Timur tergolong provinsi dengan IPM tinggi, ditopang oleh urbanisasi, infrastruktur memadai, dan tata kelola yang kuat. Namun dominasi wilayah barat menunjukkan ketimpangan yang masih terjadi.
- Klaster 1 Wilayah Transisi: Provinsi seperti Jawa Barat, Riau, dan Sulawesi Selatan menunjukkan indikator pembangunan yang sedang tumbuh namun belum merata. Provinsi di klaster ini berpotensi naik kelas, tapi berisiko stagnasi jika tak didukung kebijakan yang tepat.
- Klaster 0 Ketertinggalan Struktural: Papua, NTT, dan Maluku tergolong wilayah dengan IPM rendah dan tantangan historis serta geografis. Sebagian besar tidak mengalami perpindahan klaster selama 3 tahun terakhir, menunjukkan perlunya intervensi afirmatif dan terstruktur.





Kesimpulan



Penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam memetakan 38 provinsi di Indonesia ke dalam tiga klaster berdasarkan capaian Indeks Pembangunan Manusia (IPM) tahun 2022–2024. Klaster pertama mencakup provinsi dengan IPM rendah, umumnya di wilayah timur dengan tantangan struktural; klaster kedua berisi provinsi dengan IPM sedang dalam tahap transisi; dan klaster ketiga terdiri atas provinsi dengan IPM tinggi, didominasi wilayah barat dan urban. Pemilihan K=3 didasarkan pada evaluasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index, yang menunjukkan hasil numerik dan interpretasi geografis terbaik. Meskipun IPM meningkat dari tahun ke tahun, banyak provinsi tetap di klaster yang sama karena kemajuan bersifat relatif. Visualisasi heatmap membantu melihat dinamika spasial dan klasterisasi mencerminkan realitas sosial-ekonomi yang relevan, memberikan dasar kuat bagi perumusan kebijakan pembangunan yang lebih kontekstual dan berkeadilan.





Terima Kasih



