

Perbandingan Metode Elbow dan Silhouette Coefficient pada K-Means untuk Pengelompokan Wilayah Berdasarkan Indeks Pembangunan Manusia

Shinta Zahira Hayathun Nufus*, Fathir, Hilyatul Mustafidah

Fakultas Teknik dan Ilmu Komputer, Program Studi Ilmu Komputer, Universitas Muhammadiyah Bima, Bima, Indonesia

Email: ^{1,*}shintazahirahayathunnufus@gmail.com, ²fathirpuncak@gmail.com, ³hilyatulfida15@gmail.com

Email Penulis Korespondensi: shintazahirahayathunnufus@gmail.com

Submitted: 30/05/2026; Accepted: 22/06/2026; Published: 23/06/2026

Abstrak—Salah satu indikator utama untuk menilai seberapa berhasil upaya peningkatan kesejahteraan masyarakat melalui pembangunan manusia adalah Indeks Pembangunan Manusia (IPM). Meskipun IPM Indonesia terus meningkat, ketimpangan pembangunan manusia antarwilayah masih terlihat, terutama di kawasan Bali, Nusa Tenggara Barat (NTB), dan Nusa Tenggara Timur (NTT). Identifikasi kelompok wilayah dengan karakteristik IPM yang serupa penting untuk mendukung penyusunan kebijakan pembangunan yang lebih tepat sasaran. Namun, hasil pengelompokan menggunakan algoritma K-Means sangat dipengaruhi oleh jumlah cluster yang digunakan, sehingga diperlukan metode penentuan jumlah cluster yang sesuai dengan karakteristik data. Penelitian ini membandingkan metode Elbow dan Silhouette Coefficient dalam menentukan jumlah cluster optimal pada data IPM tahun 2024 yang mencakup 41 kabupaten/kota dengan variabel Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita. Hasil penelitian menunjukkan bahwa Elbow Method menghasilkan tiga cluster, sedangkan Silhouette Coefficient menghasilkan dua cluster dengan nilai silhouette sebesar 0,5312. Evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa pengelompokan dengan dua cluster memperoleh nilai DBI sebesar 0,7350, lebih rendah dibandingkan tiga cluster sebesar 1,0382. Temuan ini menunjukkan bahwa struktur data IPM di Bali, NTB, dan NTT cenderung membentuk dua kelompok utama, yaitu wilayah dengan tingkat pembangunan manusia tinggi dan wilayah dengan tingkat pembangunan manusia menengah hingga rendah. Selain memberikan gambaran pola ketimpangan pembangunan manusia antarwilayah, hasil penelitian menunjukkan bahwa Silhouette Coefficient lebih representatif dalam menentukan jumlah cluster pada data IPM dengan karakteristik wilayah yang relatif berdekatan. Hasil pengelompokan ini dapat menjadi dasar dalam penyusunan prioritas pembangunan pada sektor pendidikan, kesehatan, dan kesejahteraan masyarakat.

Kata Kunci: Davies-Bouldin Index; Elbow Method; Indeks Pembangunan Manusia; K-Means Clustering; Silhouette Coefficient

Abstract—One of the primary metrics for evaluating the effectiveness of efforts to improve people's well-being through human development is the Human Development Index (HDI). Although Indonesia's HDI has continued to improve, disparities in human development remain evident across regions, particularly in Bali, West Nusa Tenggara (NTB), and East Nusa Tenggara (NTT). Identifying regions with similar HDI characteristics is important for supporting more targeted development policies. However, the performance of K-Means clustering is highly influenced by the number of clusters used, making the selection of an appropriate cluster number essential. This study compares the Elbow Method and Silhouette Coefficient in determining the optimal number of clusters for 2024 HDI data covering 41 regencies and municipalities based on Life Expectancy, Expected Years of Schooling, Mean Years of Schooling, and Per Capita Expenditure. The results show that the Elbow Method produces three clusters, while the Silhouette Coefficient produces two clusters with a silhouette value of 0.5312. Evaluation using the Davies-Bouldin Index (DBI) indicates that the two-cluster solution achieves a lower DBI value (0.7350) than the three-cluster solution (1.0382). These findings suggest that the HDI structure in Bali, NTB, and NTT tends to form two major groups: regions with high human development and regions with medium-to-low human development. The results also indicate that the Silhouette Coefficient is more representative for determining the optimal number of clusters in HDI data with relatively similar regional characteristics. The clustering results may support policymakers in prioritizing development programs in education, health, and community welfare.

Keywords: Davies-Bouldin Index; Elbow Method; Human Development Index; K-Means Clustering; Silhouette Coefficient

1. PENDAHULUAN

Pembangunan manusia ialah tolak ukur penting untuk menilai kualitas pembangunan manusia suatu negara karena berkaitan langsung dengan kualitas hidup masyarakat. Konsep pembangunan manusia tidak hanya mencakup pertumbuhan ekonomi, namun juga memperhatikan peningkatan kualitas hidup melalui aspek tingkat kesehatan masyarakat, akses terhadap pendidikan, serta standar hidup yang layak bagi penduduk. [1]. Dalam mengukur mutu kemajuan masyarakat di suatu wilayah, Badan Pusat Statistik (BPS) bersandar pada Indeks Pembangunan Manusia (IPM). Instrumen standar global tersebut pada awalnya diformulasikan oleh United Nations Development Programme (UNDP) [2].

Secara keseluruhan, perkembangan IPM negara Indonesia terjadi kenaikan terus-menerus tiap tahunnya. Data BPS untuk tahun 2024 menunjukkan bahwa IPM Indonesia berada pada angka 75,02, artinya mengalami kenaikan sebanyak 0,63 dibandingkan tahun lalu [2]. Meski demikian, kenaikan itu belum merata di seluruh area, sehingga masih ada perbedaan pembangunan manusia antar daerah. Kondisi geografis yang berbeda, akses terhadap layanan dasar, dan tingkat kesejahteraan masyarakat merupakan faktor yang memengaruhi variasi pencapaian IPM di tiap daerah [3]. Sebab itu, analisis yang lebih komprehensif perlu dilakukan agar dapat memahami pola perbedaan perkembangan manusia di berbagai daerah.

Ketimpangan tersebut terlihat pada Nusa Tenggara Timur (NTT), Nusa Tenggara Barat (NTB), dan Bali yang berada pada satu kawasan geografis, yaitu Kepulauan Nusa Tenggara atau Kepulauan Sunda Kecil [4]. Walaupun

terletak dalam satu area yang cukup dekat, ketiga wilayah tersebut memperlihatkan variasi yang cukup besar dalam tingkat pembangunan manusia. Data BPS tahun 2024 menunjukkan bahwa Provinsi Bali memiliki nilai IPM sebesar 78,63 yang tergolong tinggi, sedangkan NTB mencapai 73,10 dan NTT sebesar 69,14 yang masih berada dalam kategori menengah ke bawah [2]. Di samping itu, beberapa wilayah di NTB dan NTT juga dikategorikan dalam zona tertinggal, terdepan, dan terluar (3T), artinya masih terdapat batasan dalam akses pendidikan, kesehatan, dan kesejahteraan ekonomi. Keadaan ini menunjukkan bahwa dalam suatu area yang sama terdapat perbedaan dalam karakteristik pembangunan manusia yang tidak seragam [5].

Analisis yang dilakukan oleh BPS umumnya bersifat deskriptif, yaitu hanya menunjukkan nilai IPM dan perubahan tiap tahunnya saja, sehingga belum dapat mengungkap pola kesamaan dan perbedaan karakteristik antar daerah secara mendalam. Padahal, dalam satu kawasan yang sama, ada kemungkinan terdapat kelompok wilayah dengan karakteristik pembangunan manusia yang bervariasi. Oleh karena itu, dibutuhkan metode analisis yang dapat mengidentifikasi pola pengelompokan daerah berdasarkan kesamaan indikator IPM [6].

Dalam konteks pengolahan data, teknik pengelompokan dengan menerapkan algoritma K-Means menawarkan solusi untuk membagi data ke dalam kelompok-kelompok berdasarkan tingkat kemiripan. Metode tersebut merupakan salah satu pendekatan dasar yang biasa dipergunakan dalam analisis data [7]. Metode ini dianggap efisien dalam menganalisis data multivariat seperti IPM yang melibatkan aspek pendidikan, kesehatan, dan ekonomi [8]. Meski demikian, menentukan jumlah kluster yang tepat menjadi salah satu kendala utama saat menerapkan algoritma K-Means, sebab akan berdampak pada hasil pengelompokan dan interpretasi data [9].

Penelitian sebelumnya telah menerapkan algoritma K-Means untuk mengelompokkan wilayah berdasarkan indikator pembangunan manusia dan sosial-ekonomi. Temuan mereka menunjukkan bahwa algoritma ini dapat menemukan kelompok wilayah dengan karakteristik yang serupa, yang akan membantu memahami pola pembangunan antarwilayah [10], [11]. Dalam penerapannya, proses clustering biasanya dimulai dengan menentukan jumlah cluster yang ideal dengan menerapkan metode seperti Elbow maupun Silhouette Coefficient [12], [13].

Metode Elbow menentukan jumlah kluster berdasarkan penurunan nilai Sum of Squared Error (SSE), sementara Silhouette Coefficient menilai kualitas kluster melalui tingkat kohesi dan pemisahan antar kluster. Kedua pendekatan tersebut dapat menghasilkan jumlah cluster yang berbeda pada dataset yang sama [14]. Perbedaan ini sangat berpengaruh pada data sosial-ekonomi seperti indikator penyusun IPM, karena jumlah kluster yang dipilih akan berdampak pada pengelompokan wilayah yang terbentuk dan interpretasi kondisi pembangunan manusia yang dihasilkan.

Banyak penelitian telah menggunakan metode Elbow dan Silhouette Coefficient untuk mengelompokkan daerah. Namun, sebagian besar penelitian hanya menggunakan salah satu metode tanpa melakukan analisis perbandingan kualitas cluster yang dihasilkan [15], [13]. Akibatnya, masih ada keterbatasan dalam memahami metode untuk menentukan jumlah cluster yang lebih representatif dalam pengelompokan daerah berdasarkan indikator pembangunan manusia. Penelitian yang membandingkan metode Elbow dan Silhouette Coefficient telah dilakukan pada berbagai jenis data, seperti kumpulan data Iris, ulasan pengguna obat-obatan, serta data produksi padi [16], [17], [14]. Namun, penelitian yang secara khusus melakukan perbandingan antara kedua metode tersebut untuk menentukan jumlah cluster optimal pada data Indeks Pembangunan Manusia (IPM) belum ada. Padahal, karakteristik data IPM yang tersusun atas indikator kesehatan, pendidikan, dan standar hidup memiliki kompleksitas yang berbeda dibandingkan dataset pada penelitian sebelumnya, sehingga berpotensi menghasilkan karakteristik kluster yang berbeda pula.

Berdasarkan kondisi tersebut, penelitian ini membandingkan Elbow dan Silhouette Coefficient dengan menggunakan algoritma K-Means clustering untuk menentukan jumlah kluster yang ideal bagi data Indeks Pembangunan Manusia (IPM) kabupaten/kota yang berada pada provinsi NTT, NTB, dan Bali. Studi ini diharapkan mampu menggambarkan pola pengelompokan wilayah berdasarkan karakteristik pembangunan manusia sekaligus menjadi referensi dalam pemilihan metode penentuan jumlah cluster pada analisis data IPM. Selain itu, penelitian ini memberikan kontribusi berupa evaluasi komparatif terhadap dua metode penentuan jumlah cluster pada data IPM serta menunjukkan metode yang lebih representatif dalam menggambarkan struktur pengelompokan wilayah berdasarkan indikator pembangunan manusia.

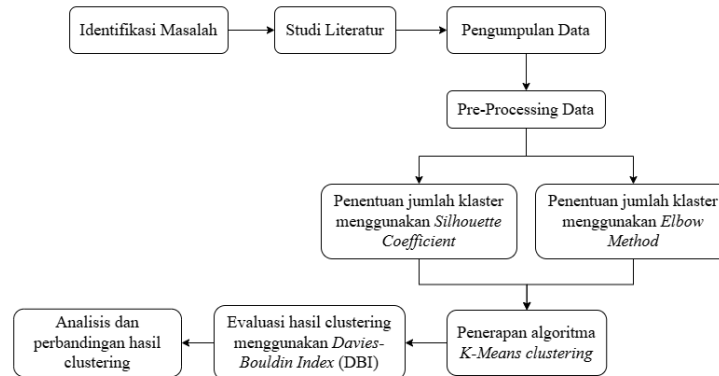
2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Alur penelitian disusun secara sistematis agar menjadi pedoman dalam melaksanakan penelitian sehingga proses analisis berjalan terarah. Adapun alur penelitian yang diterapkan ditampilkan pada Gambar 1.

Gambar 1 menunjukkan bahwa penelitian ini dimulai dari tahap identifikasi masalah untuk mengetahui permasalahan terkait ketimpangan pembangunan manusia pada wilayah Bali, NTB, dan NTT. Setelah itu dilakukan studi literatur untuk memperoleh landasan teori dan referensi penelitian yang berkaitan dengan algoritma K-Means, Elbow Method, Silhouette Coefficient, dan evaluasi clustering. Tahap berikutnya adalah pengumpulan data Indeks Pembangunan Manusia (IPM) untuk kabupaten dan kota yang berada di Nusa Tenggara Timur (NTT), Nusa Tenggara Barat (NTB), dan Bali. Selanjutnya dilakukan tahap pre-processing data yang mencakup seleksi variabel, pemeriksaan missing value dan data duplikat, serta normalisasi dengan Min-Max Normalization sehingga data siap diterapkan

dalam proses clustering. Metode Elbow dan Silhouette Coefficient kemudian digunakan sebagai metode penentuan jumlah kluster. Algoritma pengelompokan K-Means kemudian diterapkan dengan menggunakan jumlah kluster yang diperoleh untuk mengelompokkan wilayah berdasarkan karakteristik indikator-indikator yang membentuk Indeks Pembangunan Manusia (IPM). Tahap akhir penelitian adalah evaluasi hasil clustering dengan Davies–Bouldin Index (DBI) guna melihat kualitas tiap cluster serta menarik kesimpulan berdasarkan hasil pengelompokan yang diperoleh.



Gambar 1. Alur Penelitian

2.2 Pengumpulan Data

Studi ini menggunakan data sekunder dari publikasi Badan Pusat Statistik tahun 2024. Data penelitian tersebut mencakup rincian mengenai pembangunan manusia di provinsi NTT, NTB, dan Bali di tingkat kabupaten/kota.

Tabel 1. Data Indeks Pembangunan Manusia Kabupaten/Kota di Provinsi Bali, NTB, dan NTT

Kabupaten/Kota	Umur Harapan Hidup (UHH)	Harapan Lama Sekolah (HLS)	Rata-rata Lama Sekolah (RLS)	Pengeluaran Per Kapita	IPM	Pertumbuhan (%)
Jembrana	75,07	13,05	8,7	12.638	75,32	0,7
Tabanan	75,67	13,29	9,57	15.297	78,83	1,23
Badung	75,9	14,23	11,14	18.556	83,73	0,88
...
Kota Kupang	74,95	16,54	11,64	14.001	83,21	0,53

Sembilan kabupaten/kota di Provinsi Bali, sepuluh kabupaten/kota pada Provinsi Nusa Tenggara Barat (NTB), serta dua puluh dua kabupaten/kota pada Provinsi Nusa Tenggara Timur (NTT) merupakan bagian dari 41 kabupaten/kota yang tercakup dalam data yang digunakan dalam penelitian ini. Dataset mencakup data indikator pembangunan manusia serta nilai dan pertumbuhan IPM untuk setiap wilayah.

2.3 Pre-processing Data

2.3.1 Seleksi dan Pemisahan Atribut Data

Pemisahan dan pemilihan atribut dilakukan untuk memastikan bahwa hanya variabel yang relevan yang digunakan dalam proses pengolahan data. Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita yang Disesuaikan merupakan variabel dalam kajian ini. Atribut identitas wilayah berupa nama kabupaten/kota dipisahkan dari data numerik agar proses clustering hanya dilakukan terhadap variabel kuantitatif. Sementara itu, variabel nilai IPM dan pertumbuhan IPM tidak digunakan sebagai variabel input dalam proses clustering. Kedua variabel tersebut digunakan sebagai variabel pendukung pada tahap analisis hasil untuk membantu interpretasi karakteristik masing-masing kluster yang terbentuk.

2.3.2 Pemeriksaan Data

Pemeriksaan data dilakukan guna menjamin kelengkapan dan konsistensi data melalui identifikasi nilai yang hilang, data ganda, serta anomali yang mungkin terdapat dalam dataset. Pemeriksaan dilakukan terhadap seluruh variabel input clustering agar data yang digunakan valid dan sesuai untuk proses analisis. Selain pemeriksaan missing value dan data duplikat, dilakukan identifikasi outlier menggunakan visualisasi boxplot pada setiap variabel penelitian. Tahap ini bertujuan untuk mendeteksi nilai ekstrem yang berpotensi memengaruhi hasil clustering, mengingat algoritma K-Means merupakan metode berbasis jarak yang sensitif terhadap keberadaan outlier. Nilai yang teridentifikasi sebagai outlier kemudian dievaluasi berdasarkan konteks data dan karakteristik wilayah untuk menentukan kelayakannya dalam proses clustering.

2.3.3 Normalisasi Data

Tahapan terakhir pre-processing yaitu normalisasi data melalui metode Min-Max. Normalisasi dilakukan agar

menyamakan rentang nilai antar variabel tiap setiap indikator dapat berkontribusi secara merata dalam proses perhitungan jarak pengelompokan [18]. Data dikonversi menjadi rentang nilai 0 hingga 1 menggunakan Min-Max Normalization. Rumus normalisasi yang digunakan ditunjukkan pada Persamaan (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Di mana x adalah nilai data asli, x' merupakan nilai hasil yang telah dinormalisasi, x_{min} adalah nilai minimum variabel, dan x_{max} adalah nilai maksimumnya. Hasil normalisasi menghasilkan data dengan skala yang seragam sehingga proses clustering dapat dilakukan dengan baik.

Normalisasi Min-Max dipilih karena memiliki kemampuan untuk mengubah seluruh variabel ke dalam rentang 0–1. Ini memastikan bahwa setiap indikator memiliki kontribusi yang seimbang dalam perhitungan jarak. Pendekatan ini sesuai digunakan pada data IPM karena variabel yang digunakan memiliki rentang nilai yang berbeda. Selain itu, Min-Max Normalization tetap mempertahankan proporsi perbedaan nilai antarwilayah sehingga karakteristik pembangunan manusia yang menjadi fokus penelitian dapat direpresentasikan dengan baik dalam proses clustering.

2.4 Penentuan Jumlah Kluster dengan *Elbow Method*

Karena jumlah kluster dipilih akan memengaruhi hasil pengelompokan data, menentukan jumlah cluster merupakan langkah penting dalam proses pengelompokan menggunakan K-Means. Pada penelitian ini, nilai SSE dihitung untuk beberapa variasi jumlah kluster, yaitu $k = 2$ hingga $k = 10$. Jumlah kluster optimal ditentukan berdasarkan titik siku (elbow point) yang menunjukkan perubahan penurunan SSE paling signifikan sebelum grafik mulai melandai.

Jumlah kuadrat jarak antara data dan pusat kluster data berada ditunjukkan oleh nilai SSE. Beberapa variasi jumlah kluster (k) dihitung menggunakan SSE. Semakin kecil nilai SSE menunjukkan bahwa data dalam suatu kluster semakin dekat dengan centroid [19]. Rumus SSE ditunjukkan pada Persamaan (2).

$$SSE = \sum_{i=1}^k \sum_{x \in c_i} (x - c_i)^2 \quad (2)$$

Di mana k merupakan jumlah kluster, x adalah data pada kluster c_i , dan c_i adalah centroid pada kluster ke- i .

Grafik yang menggambarkan korelasi antara jumlah kluster dan nilai SSE dihasilkan setelah perhitungan SSE dilakukan pada berbagai nilai k . Seiring bertambahnya jumlah kluster, nilai SSE cenderung terus menurun. Namun, pada akhirnya, penurunan ini akan melambat, sehingga membentuk pola siku. Karena dapat menggambarkan struktur data dengan akurat, titik ini dianggap sebagai jumlah kluster yang ideal [20].

2.5 Penentuan Jumlah Kluster dengan *Silhouette Coefficient*

Selain menggunakan *Elbow Method*, jumlah cluster dalam studi ini juga ditentukan dengan *Silhouette Coefficient*. Rumus *Silhouette Coefficient* berada pada Persamaan (3).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Di mana $b(i)$ adalah jarak terdekat rata-rata antara data ke- i dan data dalam kluster lain, sedangkan $a(i)$ adalah jarak rata-rata antara data ke- i dan data lain pada cluster yang sama. Nilai $s(i)$ adalah nilai silhouette untuk data ke- i .

Nilai Silhouette Coefficient berada di antara -1 dan 1. Nilai yang lebih rendah berarti data berada di antara dua kelompok, sedangkan nilai yang lebih tinggi artinya data telah dikategorikan dengan baik ke dalam kelompoknya. Jika nilainya negatif, berarti data tersebut lebih dekat ke kelompok lain daripada ke kelompok asalnya [16].

Pada penelitian ini, nilai silhouette rata-rata dihitung untuk setiap variasi jumlah kluster yang diuji. Jumlah kluster optimal dipilih berdasarkan nilai silhouette rata-rata tertinggi karena menunjukkan tingkat kohesi dalam kluster dan separasi antar kluster yang lebih baik.

2.6 Penerapan Algoritma K-Means Clustering

Pada penelitian ini, proses clustering dilakukan terhadap 41 kabupaten/kota di Provinsi Bali, NTB, dan NTT menggunakan empat indikator pembentuk IPM, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita yang Disesuaikan. Clustering dilakukan menggunakan jumlah kluster yang diperoleh dari metode Elbow dan Silhouette Coefficient yang selanjutnya dibandingkan kualitas pengelompokannya menggunakan Davies–Bouldin Index (DBI). Pengelompokan dengan algoritma K-Means diterapkan melalui metode inisialisasi centroid k-means++, yang dipilih sebagai titik pusat awal cluster sebelum proses iterasi dilakukan.

Tahapan K-Means diawali dengan menentukan pusat data awal sebanyak jumlah cluster sebagaimana telah ditetapkan. Selanjutnya dilakukan perhitungan jarak setiap data dengan pusat data menggunakan Euclidean Distance. Data kemudian dikelompokkan ke dalam kluster dengan jarak terdekat terhadap pusat data. Setelah proses pengelompokan selesai, pusat diperbarui berdasarkan rata-rata data pada masing-masing kluster. Tahap tersebut dilaksanakan terus-menerus hingga tidak terjadi perubahan centroid atau hasil kluster telah memusat [21].

Jarak Euclidean digunakan untuk menghitung jarak dalam algoritma K-Means, sebagaimana terlihat pada Persamaan (4).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{4}$$

Di mana n adalah jumlah variabel, x_i adalah nilai data untuk variabel ke- i , y_i adalah nilai centroid untuk variabel ke- i , dan $d(x, y)$ adalah jarak antara data dan centroid.

2.7 Evaluasi Clustering menggunakan Davies–Bouldin Index (DBI)

Kualitas hasil pengelompokan yang dihasilkan oleh K-Means dievaluasi. Dalam penelitian ini, digunakan Indeks Davies-Bouldin (DBI), sebuah teknik evaluasi internal yang mengukur tingkat kepadatan data di dalam setiap kelompok serta jarak antar kelompok [22].

Nilai DBI diperoleh dari perbandingan antara jarak rata-rata data dalam kluster terhadap centroid dengan jarak antar centroid kluster. Rumus Davies–Bouldin Index ditunjukkan pada Persamaan (5).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \tag{5}$$

Di mana M_{ij} adalah jarak antara pusat kluster ke- i dan ke- j , k adalah jumlah kluster, serta S_i dan S_j masing-masing merupakan jarak rata-rata data ke pusat kluster ke- i dan ke- j .

Karena memiliki tingkat kepadatan tinggi di setiap kluster dan perbedaan yang jelas tiap kluster, kualitas pengelompokan yang lebih baik ditunjukkan oleh nilai DBI yang lebih rendah. Dalam studi ini, DBI digunakan guna mengevaluasi hasil clustering yang diperoleh dari jumlah kluster hasil Elbow Method dan Silhouette Coefficient. Metode yang menghasilkan nilai DBI lebih kecil dianggap memberikan kualitas clustering yang lebih baik dan digunakan sebagai dasar dalam menentukan pendekatan yang paling sesuai untuk data IPM wilayah Bali, NTB, dan NTT.

2.8 Perbandingan Hasil Clustering

Tahap perbandingan hasil clustering dilakukan untuk membandingkan kualitas pengelompokan yang diperoleh dari jumlah kluster berdasarkan Elbow Method dan Silhouette Coefficient. Perbandingan dilakukan terhadap struktur kluster yang terbentuk serta nilai evaluasi clustering menggunakan Davies–Bouldin Index (DBI).

Berdasarkan nilai rata-rata indikator-indikator yang membentuk Indeks Pembangunan Manusia (IPM), yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita, karakteristik masing-masing kluster dianalisis. Selain itu, dilakukan perbandingan nilai DBI untuk menentukan metode penentuan jumlah kluster yang menghasilkan kualitas clustering terbaik.

3. HASIL DAN PEMBAHASAN

3.1 Pre-processing Data

Tahap berikut kumpulan data diproses terlebih dahulu sebelum teknik K-Means digunakan untuk pengelompokan. Langkah-langkah yang diambil meliputi seleksi atribut data, pemeriksaan kelengkapan data, dan normalisasi data.

3.1.1 Seleksi dan Pemisahan Atribut Data

Pada penelitian ini, atribut yang digunakan sebagai variabel input clustering terdiri atas Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran per Kapita. Sementara itu, atribut wilayah dipisahkan dari data numerik karena hanya digunakan sebagai identitas kabupaten/kota pada tahap interpretasi hasil clustering.

```

=== Data Wilayah ===
Wilayah
0  Jembrana
1  Tabanan
2  Badung
3  Gianyar
4  Klungkung

dtype: object

=== Data Fitur (X) ===
   UHH  HLS  RLS  Pengeluaran
0  75.07  13.05  8.70  12638
1  75.67  13.29  9.57  15297
2  75.90  14.23  11.14  18556
3  75.72  14.11  9.84  15577
4  74.90  13.13  8.74  12358

Shape Data Fitur: (41, 4)
    
```

Gambar 2. Hasil Seleksi dan Pemisahan Atribut Data

Terlihat jelas dari Gambar 2 bahwa data fitur yang akan digunakan dalam proses pengelompokan telah dipisahkan dari data wilayah. Data wilayah berisi nama kabupaten/kota sebagai identitas objek penelitian, sedangkan data fitur (X) berisi empat variabel numerik, yaitu UHH, HLS, RLS, dan Pengeluaran.

Selain itu, output Shape Data Fitur: (41, 4) menunjukkan bahwa data yang diterapkan dalam proses clustering mencakup 41 kabupaten/kota dan 4 variabel numerik. Data fitur tersebut selanjutnya akan menjadi input dalam proses normalisasi dan pengelompokan dengan K-Means.

3.1.2 Pemeriksaan Data

Tahap pemeriksaan kelengkapan dan konsistensi data dilaksanakan untuk memastikan bahwa tidak ada data yang berulang atau hilang (*missing value*) dalam kumpulan data yang digunakan. Pemeriksaan dilakukan pada semua variabel numerik yang digunakan sebagai input clustering, yaitu UHH, HLS, RLS, serta Pengeluaran per Kapita.

```
# Cek missing value
print("=== Missing Value ===")
print(X.isnull().sum())

# Cek duplikasi
print("\n=== Jumlah Data Duplikat ===")
print(X.duplicated().sum())

=== Missing Value ===
UHH      0
HLS      0
RLS      0
Pengeluaran  0
dtype: int64

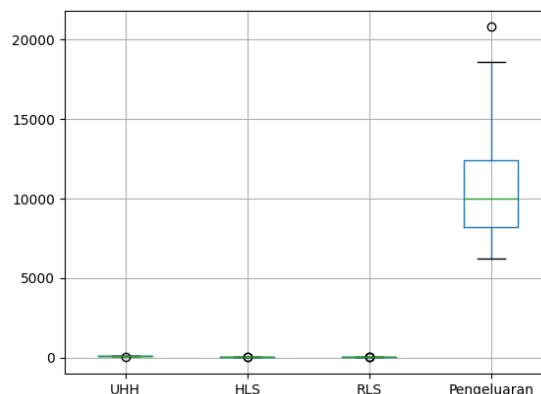
=== Jumlah Data Duplikat ===
0
```

Gambar 3. Hasil Pemeriksaan Missing Value dan Data Duplikat

Berdasarkan Gambar 3, terlihat bahwa semua variabel menunjukkan nilai yang hilang sebesar 0, yang mengindikasikan bahwa tidak ada data yang hilang dalam kumpulan data penelitian. Di samping itu, hasil pemeriksaan juga memperlihatkan bahwa total data duplikat adalah 0, sehingga tidak ada data yang terulang dalam dataset.

3.1.3 Identifikasi Outlier

Identifikasi outlier dilakukan menggunakan visualisasi boxplot pada setiap variabel penelitian untuk mendeteksi nilai ekstrem yang berpotensi memengaruhi hasil clustering. Tahap ini penting karena algoritma K-Means ialah pendekatan berbasis jarak yang sensitif terhadap keberadaan outlier.



Gambar 4. Boxplot Variabel Penelitian

Berdasarkan hasil identifikasi menggunakan boxplot, ditemukan adanya nilai ekstrem pada variabel pengeluaran, sedangkan variabel Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), dan Rata-rata Lama Sekolah (RLS) tidak menunjukkan outlier yang signifikan. Nilai ekstrem tersebut tidak dihapus dari dataset karena berasal dari data resmi Badan Pusat Statistik (BPS) dan merepresentasikan kondisi aktual masing-masing kabupaten/kota. Dengan demikian, seluruh data tetap digunakan dalam proses clustering agar variasi karakteristik pembangunan manusia antarwilayah dapat dipertahankan.

3.1.4 Normalisasi Data

Metode Normalisasi Min-Max digunakan pada tahap normalisasi untuk mengubah rentang nilai setiap variabel menjadi skala antara 0 dan 1. Untuk mencegah sehingga tidak terdapat variabel yang mendominasi perhitungan jarak dalam pengelompokan, prosedur tersebut dilakukan sehingga tiap variabel memiliki rentang nilai yang seragam. Tahap normalisasi menjadi penting karena variabel yang digunakan dalam penelitian ini memiliki satuan dan rentang nilai yang berbeda, seperti Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah

(RLS), dan Pengeluaran Per Kapita. Tanpa normalisasi, variabel dengan rentang nilai yang lebih besar berpotensi memberikan pengaruh yang lebih dominan terhadap hasil clustering. Oleh karena itu, normalisasi dilakukan untuk memastikan bahwa proses pengelompokan mencerminkan karakteristik seluruh variabel secara proporsional.

```

=== Data Setelah Normalisasi Min-Max ===
      UHH      HLS      RLS  Pengeluaran
0  0.910270  0.172986  0.440000  0.441811
1  0.975135  0.229858  0.605714  0.624485
2  1.000000  0.452607  0.904762  0.848379
3  0.980541  0.424171  0.657143  0.643721
4  0.891892  0.191943  0.447619  0.422575

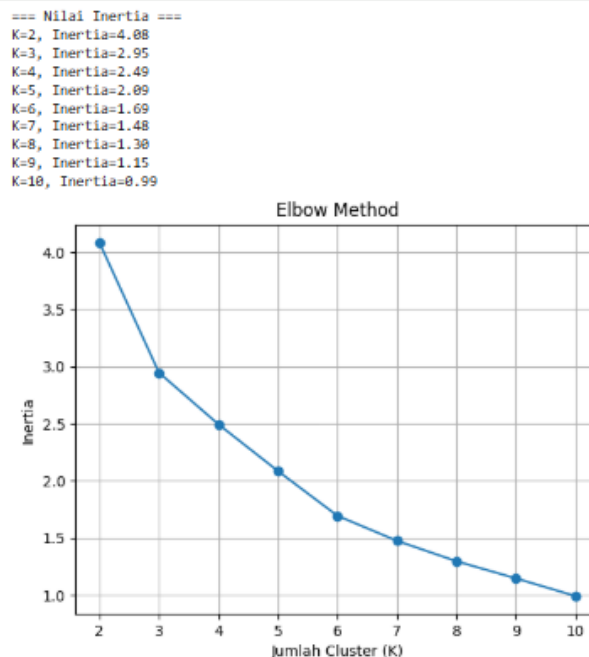
Shape X_scaled: (41, 4)
    
```

Gambar 5. Hasil Normalisasi Data

Hasil normalisasi menunjukkan bahwa setiap variabel telah berhasil diubah menjadi rentang nilai yang seragam tanpa mengubah pola distribusi data. Variabel yang memiliki nilai awal dengan rentang luas, seperti pengeluaran per kapita, menjadi lebih proporsional dibandingkan variabel yang lain. Selain itu, hasil normalisasi menunjukkan bahwa jumlah data dan jumlah variabel tetap tidak berubah, yaitu masih ada 41 data wilayah dan 4 variabel numerik. Hasil normalisasi tersebut kemudian dimasukkan ke dalam metode K-Means untuk pengelompokan, serta Metode Elbow dan Silhouette Coefficient untuk menentukan jumlah kelompok yang ideal.

3.2 Penentuan Jumlah Kluster dengan Elbow Method

Pada tahap ini, Elbow Method diterapkan sebagai pendekatan untuk menentukan jumlah kluster dengan menghitung nilai inerti atau Within Cluster Sum of Squares (WCSS) berbagai jumlah kluster, yaitu k=2 hingga k=10. Hasil perhitungan inerti dan grafik Elbow Method disajikan di Gambar 6.



Gambar 6. Jumlah Cluster *Elbow Method*

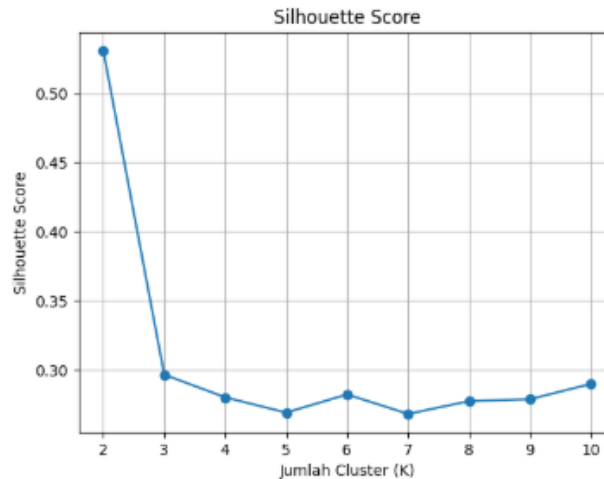
Grafik Elbow Method menunjukkan bahwa nilai inerti menurun cukup tajam dari K=2 ke K=3 sebelum akhirnya stabil pada tingkat K berikutnya. Hal tersebut berarti titik siku berada pada K=3 karena penurunan inerti setelahnya tidak terlalu signifikan dibandingkan sebelumnya. Oleh karena itu, tiga kelompok merupakan jumlah yang ideal menurut Elbow Method.

3.3 Penentuan Jumlah Kluster dengan Silhouette Coefficient

Setelah mendapatkan kandidat jumlah kluster dari Elbow Method, tahap selanjutnya adalah menentukan jumlah kluster optimal menggunakan Silhouette Coefficient. Dengan menggunakan metode ini, tingkat kohesi (kedekatan data di dalam suatu kluster) dan pemisahan (jarak antar kluster) dimanfaatkan untuk mengukur kualitas pengelompokan. Kualitas pengelompokan yang lebih baik ditandai dengan hasil yang mendekati angka 1 pada skala Silhouette Score, yang berkisar antara -1 hingga 1. Hasil perhitungan Silhouette Score ditunjukkan dalam Gambar 7.

```

=== Silhouette Score ===
K=2, Silhouette Score=0.5312
K=3, Silhouette Score=0.2962
K=4, Silhouette Score=0.2798
K=5, Silhouette Score=0.2687
K=6, Silhouette Score=0.2819
K=7, Silhouette Score=0.2678
K=8, Silhouette Score=0.2778
K=9, Silhouette Score=0.2785
K=10, Silhouette Score=0.2895
    
```



Gambar 7. Hasil Perhitungan Jumlah Cluster Silhouette Coefficient

Berdasarkan hasil pengujian, K=2 memiliki nilai Silhouette Score tertinggi, yaitu 0,5312. Nilai tersebut menunjukkan bahwa pembagian data menjadi 2 kluster menghasilkan struktur kluster yang optimal, yaitu memiliki tingkat kohesi yang tinggi dan separasi yang lebih jelas daripada jumlah kluster lainnya. Sementara itu, pada jumlah kluster hasil penentuan dengan Elbow Method, yaitu K=3, diperoleh nilai Silhouette Score sebesar 0,2962 yang lebih rendah dibandingkan K=2.

Perbedaan hasil tersebut menunjukkan bahwa metode Elbow dan Silhouette Coefficient memperoleh rekomendasi jumlah kluster yang berbeda. Elbow Method berada pada titik siku pada K=3, sedangkan Silhouette Coefficient menunjukkan kualitas clustering terbaik pada K=2. Kondisi ini menjadi dasar untuk melakukan perbandingan hasil clustering pada kedua jumlah kluster tersebut pada tahap selanjutnya pengelompokan dengan K-Means dan dilakukan evaluasi pengelompokan dengan Davies-Bouldin Index (DBI).

3.4 Penerapan Algoritma K-Means Clustering

3.4.1 Penerapan K-Means dengan Jumlah Cluster Hasil Elbow Method

Hasil pengujian Metode Elbow menunjukkan bahwa tiga kelompok merupakan jumlah yang ideal. Langkah selanjutnya adalah mengelompokkan wilayah kabupaten/kota berdasarkan kesamaan indikator pembangunan manusia, seperti Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita, dengan K-Means dimana nilai K=3.

Data yang telah dinormalisasikan digunakan untuk proses pengelompokan. Data tersebut kemudian dibagi menjadi tiga kluster dengan teknik K-Means berdasarkan seberapa dekat data tersebut dengan pusat kluster. Tabel 2 menampilkan hasil pengelompokan beserta jumlah kluster yang diperoleh dari Metode Elbow.

Tabel 2. Hasil *clustering* dengan jumlah cluster yang diperoleh dari Elbow Method

Kabupaten/Kota	Cluster	Jumlah
Lombok Utara, Sumba Barat, Kupang, Timor Tengah Selatan, Timor Tengah Utara, Alor, Rote Ndao, Sumba Tengah, Sumba Barat Daya, Sabu Raijua, dan Malaka	1	11 Kabupaten
Tabanan, Badung, Gianyar, Kota Denpasar, Kota Mataram, Kota Bima, dan Kota Kupang	2	3 Kabupaten dan 4 Kota
Jembrana, Klungkung, Bangli, Karangasem, Buleleng, Lombok Barat, Lombok Tengah, Lombok Timur, Sumbawa, Dompu, Bima, Sumbawa Barat, Sumba Timur, Belu, Lembata, Flores Timur, Sikka, Ende, Ngada, Manggarai, Manggarai Barat, Nagekeo, dan Manggarai Timur	3	23 Kabupaten

Berdasarkan hasil clustering, Cluster 1 terdiri dari wilayah dengan tingkat pembangunan manusia yang relatif rendah dan sebagian besar berasal dari Provinsi Nusa Tenggara Timur. Rendahnya capaian pembangunan manusia pada kelompok ini dapat dikaitkan dengan keterbatasan akses terhadap pendidikan, layanan kesehatan, serta tingkat kesejahteraan masyarakat yang relatif lebih rendah dibandingkan cluster lainnya. Selain itu, karakteristik geografis

berupa wilayah kepulauan dan sebaran penduduk yang tidak merata turut menjadi tantangan dalam pemerataan pembangunan manusia.

Cluster 2 merupakan kelompok wilayah dengan tingkat pembangunan manusia yang tinggi Cluster ini didominasi oleh daerah perkotaan dan pusat pertumbuhan ekonomi seperti Badung, Kota Denpasar, Kota Mataram, dan Kota Kupang. Tingginya capaian indikator pembangunan manusia pada wilayah tersebut dapat dikaitkan dengan aktivitas ekonomi yang lebih berkembang, akses pendidikan yang lebih baik, ketersediaan fasilitas kesehatan yang lebih memadai, serta tingkat pengeluaran per kapita yang relatif tinggi.

Cluster 3 terdiri dari wilayah dengan tingkat pembangunan manusia sedang dan menjadi kelompok dengan jumlah anggota terbanyak. Wilayah dalam cluster ini menunjukkan kondisi transisi antara daerah dengan pembangunan manusia tinggi dan rendah. Meskipun memiliki capaian indikator yang lebih baik dibandingkan Cluster 1, wilayah-wilayah dalam cluster ini masih memerlukan peningkatan pada aspek pendidikan, kesehatan, dan kesejahteraan agar mampu mencapai tingkat pembangunan manusia seperti yang terdapat pada Cluster 2.

Secara keseluruhan, hasil pengelompokan menunjukkan adanya pola spasial yang cukup jelas di kawasan Bali, NTB, dan NTT. Wilayah Bali cenderung berada pada kelompok dengan tingkat pembangunan manusia yang lebih tinggi, sedangkan sebagian besar kabupaten di NTT berada pada kelompok dengan tingkat pembangunan manusia yang lebih rendah. Temuan ini menunjukkan bahwa meskipun ketiga provinsi berada dalam satu kawasan geografis yang sama, masih terdapat ketimpangan pembangunan manusia yang cukup signifikan antarwilayah.

Nilai rata-rata setiap variabel dalam setiap kelompok dihitung untuk memberikan pemahaman mendalam terkait karakteristik masing-masing kluster. Tabel 3 di bawah ini menunjukkan nilai rata-rata setiap kelompok.

Tabel 3. Rata-rata tiap Cluster (Elbow)

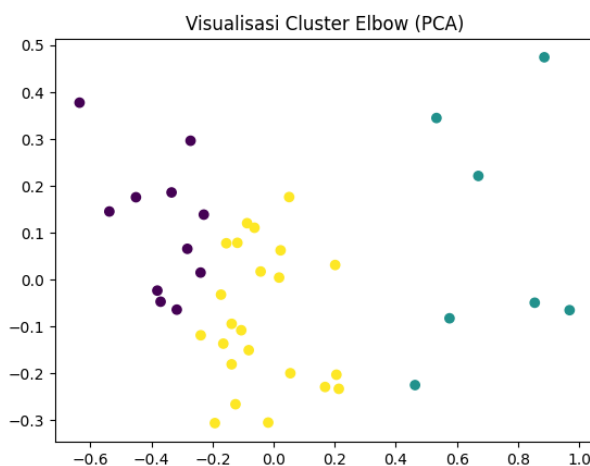
Cluster <i>Elbow</i>	UHH	HLS	RLS	Pengeluaran Per Kapita
1	70.193636	13.116364	7.341818	7670.727273
2	75.145714	14.722857	10.608571	16152.857143
3	73.217391	13.212174	7.970435	10241.304348

Berdasarkan nilai rata-rata pada Tabel 3, terlihat bahwa Cluster 1 memiliki nilai rata-rata terendah pada seluruh indikator IPM, yaitu UHH, HLS, RLS, dan pengeluaran per kapita. Kondisi ini menunjukkan bahwa rendahnya pembangunan manusia pada cluster tersebut tidak hanya terjadi pada satu aspek, tetapi mencakup aspek kesehatan, pendidikan, dan kesejahteraan ekonomi secara bersamaan.

Sebaliknya, Cluster 2 menunjukkan nilai rata-rata tertinggi pada seluruh indikator IPM. Hal ini mengindikasikan bahwa wilayah-wilayah dalam cluster tersebut memiliki capaian pembangunan manusia yang lebih baik secara menyeluruh, baik dari aspek kesehatan, pendidikan, maupun kesejahteraan ekonomi.

Adapun Cluster 3 menunjukkan nilai indikator yang berada di antara Cluster 1 dan Cluster 2. Kondisi ini menunjukkan bahwa kelompok wilayah tersebut memiliki tingkat pembangunan manusia yang relatif lebih baik dibandingkan Cluster 1, namun masih memerlukan peningkatan pada beberapa indikator agar dapat mencapai karakteristik pembangunan manusia yang setara dengan Cluster 2.

Principal Component Analysis (PCA) digunakan untuk visualisasi kluster guna membuat distribusi hasil pengelompokan menjadi lebih mudah dipahami. Hasil visualisasi cluster ditampilkan pada Gambar 8.



Gambar 8. Visualisasi Cluster Elbow

Visualisasi PCA menunjukkan bahwa hasil clustering dengan K=3 mampu membentuk tiga kelompok wilayah berdasarkan kemiripan indikator pembangunan manusia. Setiap warna merepresentasikan cluster yang terbentuk berdasarkan karakteristik data yang serupa. Cluster 2 terlihat memiliki pemisahan yang relatif lebih jelas dibandingkan cluster lainnya. Kondisi ini menunjukkan bahwa wilayah-wilayah dalam cluster tersebut memiliki karakteristik pembangunan manusia yang berbeda secara signifikan dibandingkan kelompok lainnya.



Di sisi lain, terlihat adanya tumpang tindih (*overlapping*) antara Cluster 1 dan Cluster 3. Kondisi ini menunjukkan bahwa sebagian wilayah dalam kedua cluster memiliki karakteristik pembangunan manusia yang relatif mirip sehingga batas antar cluster tidak terbentuk secara tegas. *Overlapping* tersebut mengindikasikan bahwa pembagian data menjadi tiga cluster masih menghasilkan kelompok yang memiliki tingkat kemiripan cukup tinggi, terutama antara wilayah dengan tingkat pembangunan manusia rendah dan menengah.

3.4.2 Penerapan K-Means dengan Jumlah Cluster Hasil Silhouette Coefficient

Berdasarkan jumlah kluster yang diperoleh dengan Silhouette Coefficient, dua kluster merupakan jumlah yang ideal. Setelah itu, data dari 41 kabupaten/kota di wilayah penelitian diproses melalui pengelompokan menggunakan algoritma K-Means dengan nilai K=2.

Berdasarkan kesamaan indikator pembangunan manusia, data tersebut dibagi menjadi dua kelompok, sesuai dengan hasil pengelompokan. Tabel 4 menampilkan distribusi anggota kelompok yang diperoleh dengan metode K-Means.

Tabel 4. Hasil *clustering* dengan jumlah cluster berdasarkan Silhouette Coefficient

Kabupaten/Kota	Cluster	Jumlah
Tabanan, Badung, Gianyar, Kota Denpasar, Kota Mataram, Kota Bima, dan Kota Kupang	1	3 Kabupaten dan 4 Kota
Jembrana, Klungkung, Bangli, Karangasem, Buleleng, Lombok Barat, Lombok Tengah, Lombok Timur, Sumbawa, Dompu, Bima, Sumbawa Barat, Lombok Utara, Sumba Barat, Sumba Timur, Kupang, Timor Tengah Selatan, Timor Tengah Utara, Belu, Alor, Lembata, Flores Timur, Sikka, Ende, Ngada, Manggarai, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, dan Malaka	2	34 Kabupaten

Berdasarkan hasil *clustering*, Cluster 1 terdiri atas wilayah dengan tingkat pembangunan manusia yang relatif tinggi, yaitu Tabanan, Badung, Gianyar, Kota Denpasar, Kota Mataram, Kota Bima, dan Kota Kupang. Cluster ini didominasi oleh wilayah perkotaan dan pusat pertumbuhan ekonomi yang memiliki akses pendidikan, layanan kesehatan, serta tingkat kesejahteraan masyarakat yang relatif lebih baik dibandingkan wilayah lainnya. Tingginya capaian pembangunan manusia pada cluster ini menunjukkan bahwa faktor pendidikan, kesehatan, dan ekonomi berkembang secara lebih merata dibandingkan wilayah lain di kawasan penelitian.

Sementara itu, Cluster 2 terdiri atas sebagian besar kabupaten di Bali, NTB, dan NTT dengan tingkat pembangunan manusia menengah hingga rendah. Wilayah-wilayah dalam cluster ini masih menghadapi berbagai tantangan pembangunan manusia, seperti keterbatasan akses terhadap pendidikan dan layanan kesehatan, serta tingkat kesejahteraan masyarakat yang relatif lebih rendah dibandingkan wilayah pada Cluster 1. Selain itu, beberapa wilayah dalam cluster ini juga memiliki karakteristik geografis yang dapat memengaruhi pemerataan pembangunan dan akses terhadap layanan publik.

Hasil pengelompokan menunjukkan adanya pemisahan yang cukup jelas antara wilayah dengan tingkat pembangunan manusia tinggi dan wilayah dengan tingkat pembangunan manusia menengah hingga rendah. Temuan ini menunjukkan bahwa ketimpangan pembangunan manusia di kawasan Bali, NTB, dan NTT masih cukup nyata. Wilayah dengan capaian pembangunan manusia yang lebih tinggi cenderung terkonsentrasi pada daerah perkotaan dan pusat pertumbuhan ekonomi, sedangkan sebagian besar kabupaten lainnya berada pada kelompok dengan capaian pembangunan manusia yang lebih rendah.

Nilai rata-rata setiap variabel dalam setiap kelompok dihitung untuk mendapatkan pemahaman yang lebih mendalam mengenai karakteristik masing-masing kelompok. Tabel 5 di bawah ini menunjukkan nilai rata-rata setiap kelompok.

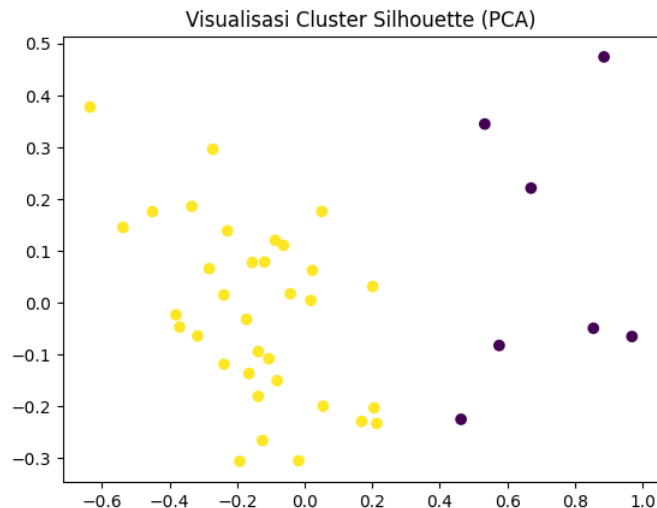
Tabel 5. Rata-rata tiap Cluster (Silhouette Coefficient)

Cluster	Silhouette Coefficient	UHH	HLS	RLS	Pengeluaran Per Kapita
1		75.145714	14.722857	10.608571	16152.857143
2		72.239118	13.181176	7.767059	9409.647059

Berdasarkan nilai rata-rata pada Tabel 5, terlihat bahwa Cluster 1 memiliki nilai rata-rata yang lebih tinggi pada seluruh indikator pembangunan manusia dibandingkan Cluster 2. Perbedaan terbesar terlihat pada variabel pengeluaran per kapita, sedangkan indikator pendidikan dan kesehatan juga menunjukkan nilai yang lebih baik pada Cluster 1. Kondisi ini menunjukkan adanya kesenjangan pembangunan manusia antara kedua kelompok wilayah.

Sementara itu, Cluster 2 memiliki nilai rata-rata yang lebih rendah pada seluruh indikator, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan pengeluaran per kapita. Kondisi ini menunjukkan bahwa wilayah-wilayah dalam cluster tersebut masih menghadapi tantangan pembangunan manusia pada berbagai aspek secara bersamaan.

Diterapkan visualisasi dengan Principal Component Analysis (PCA) agar distribusi hasil pengelompokan lebih mudah dipahami. Hasil visualisasi cluster ditunjukkan pada Gambar 9.



Gambar 9. Visualisasi Cluster Silhouette Coefficient

Visualisasi PCA pada Gambar 9 menunjukkan bahwa hasil clustering dengan $K=2$ mampu membentuk dua kelompok wilayah yang relatif terpisah. Setiap warna merepresentasikan cluster yang terbentuk berdasarkan kemiripan indikator pembangunan manusia. Cluster 1 terlihat terkonsentrasi pada sisi kanan grafik, sedangkan Cluster 2 mendominasi sisi kiri grafik.

Pemisahan yang cukup jelas antara kedua cluster menunjukkan bahwa wilayah dalam masing-masing kelompok memiliki karakteristik pembangunan manusia yang berbeda. Cluster 1 berisi wilayah dengan capaian indikator pembangunan manusia yang lebih tinggi, sedangkan Cluster 2 mencakup wilayah dengan capaian indikator menengah hingga rendah.

Meskipun terdapat beberapa titik yang berada di dekat batas pemisahan cluster, secara umum persebaran data menunjukkan bahwa kedua cluster memiliki tingkat pemisahan yang baik. Kondisi ini mengindikasikan bahwa pembagian data menjadi dua cluster mampu menghasilkan kelompok yang relatif kompak dan mudah dibedakan.

3.5 Evaluasi Clustering menggunakan Davies–Bouldin Index (DBI)

Hasil pengelompokan yang diperoleh dari metode K-Means selanjutnya dievaluasi kualitasnya. Indeks Davies-Bouldin (DBI) merupakan metode evaluasi yang digunakan dalam studi ini. Pendekatan tersebut digunakan sehingga dapat mengukur seberapa mirip data di dalam suatu kluster dan seberapa berbeda kluster-kluster tersebut satu sama lain.

Pengelompokan yang dihasilkan memiliki kualitas yang lebih baik jika nilai DBI lebih rendah, karena hal itu menunjukkan kelompok yang lebih padat dan pemisahan antar kelompok yang lebih jelas. Evaluasi dilakukan pada dua hasil clustering, yaitu clustering dengan jumlah cluster berdasarkan Elbow Method ($K=3$) dan clustering dengan jumlah cluster menurut Silhouette Coefficient ($K=2$). Hasil penilaian DBI disajikan dalam Gambar 10.

```
=== Davies-Bouldin Index ===
DBI (Elbow, K=3): 1.0382
DBI (Silhouette, K=2): 0.7350
```

Gambar 10. Hasil evaluasi DBI

Berdasarkan hasil perhitungan, pengelompokan dengan jumlah cluster hasil Elbow Method ($K=3$) menghasilkan nilai Davies-Bouldin Index (DBI) sebesar 1,0382, sedangkan pengelompokan dengan jumlah cluster hasil Silhouette Coefficient ($K=2$) menghasilkan nilai DBI sebesar 0,7350. Nilai DBI yang lebih rendah menunjukkan bahwa cluster yang terbentuk memiliki tingkat homogenitas yang lebih tinggi di dalam cluster serta pemisahan yang lebih baik antar cluster.

Hasil ini sejalan dengan visualisasi PCA yang menunjukkan bahwa pengelompokan dengan $K=2$ menghasilkan dua kelompok yang relatif terpisah dengan batas cluster yang lebih jelas. Sebaliknya, pada pengelompokan $K=3$ masih terlihat adanya kedekatan distribusi data antara Cluster 1 dan Cluster 3, yang mengindikasikan bahwa kedua kelompok memiliki karakteristik pembangunan manusia yang relatif mirip. Kondisi tersebut menyebabkan pemisahan antar cluster menjadi kurang optimal sehingga nilai DBI yang dihasilkan lebih tinggi.

Dari sisi karakteristik data, wilayah dengan tingkat pembangunan manusia tinggi membentuk kelompok yang cukup berbeda dibandingkan wilayah lainnya, terutama pada indikator pendidikan dan pengeluaran per kapita. Namun, karakteristik wilayah dengan tingkat pembangunan manusia rendah dan menengah cenderung saling berdekatan sehingga pemisahan keduanya ke dalam cluster yang berbeda menghasilkan batas cluster yang kurang tegas. Oleh karena itu, pengelompokan dengan dua cluster mampu menghasilkan struktur cluster yang lebih kompak dan terpisah dibandingkan pengelompokan dengan tiga cluster.

3.6 Perbandingan Hasil Clustering

Untuk mengetahui perbedaan kualitas pengelompokan yang dihasilkan dari perhitungan jumlah cluster menggunakan Metode Elbow dan Silhouette Coefficient, dilakukan perbandingan terhadap hasil pengelompokan tersebut. Metode Elbow menghasilkan jumlah cluster optimal sebanyak tiga cluster ($K=3$), sedangkan Silhouette Coefficient menghasilkan jumlah cluster optimal sebanyak dua cluster ($K=2$).

Hasil clustering menggunakan $K=3$ menghasilkan pembagian wilayah yang lebih rinci karena data terbagi ke dalam tiga kelompok, yaitu cluster dengan tingkat pembangunan manusia rendah, sedang, dan tinggi. Pengelompokan ini mampu memberikan gambaran yang lebih detail mengenai variasi tingkat pembangunan manusia antarwilayah. Namun, hasil visualisasi PCA menunjukkan bahwa masih terdapat kedekatan distribusi data antara cluster rendah dan cluster sedang, yang mengindikasikan bahwa kedua kelompok tersebut memiliki karakteristik yang relatif mirip. Kondisi ini menyebabkan pemisahan antar cluster belum terbentuk secara optimal dan menghasilkan nilai DBI sebesar 1,0382.

Sementara itu, clustering dengan $K=2$ menghasilkan dua kelompok utama, yaitu wilayah dengan tingkat pembangunan manusia tinggi dan wilayah dengan tingkat pembangunan manusia menengah hingga rendah. Hasil visualisasi PCA menunjukkan bahwa kedua kelompok tersebut memiliki pemisahan yang lebih jelas dengan tingkat tumpang tindih yang lebih rendah. Hal ini menunjukkan bahwa karakteristik data IPM kabupaten/kota di Bali, NTB, dan NTT lebih mudah dipisahkan ke dalam dua kelompok utama dibandingkan tiga kelompok yang berbeda secara jelas. Kondisi tersebut juga tercermin pada nilai DBI sebesar 0,7350 yang lebih rendah dibandingkan hasil clustering $K=3$.

Berdasarkan hasil evaluasi dan karakteristik data yang terbentuk, clustering dengan jumlah cluster hasil Silhouette Coefficient ($K=2$) menghasilkan kualitas pengelompokan yang lebih baik dibandingkan clustering dengan jumlah cluster hasil Elbow Method ($K=3$). Meskipun $K=3$ mampu memberikan informasi yang lebih rinci mengenai tingkat pembangunan manusia, pengelompokan $K=2$ menghasilkan cluster yang lebih homogen dan memiliki pemisahan yang lebih jelas. Oleh karena itu, jumlah cluster yang diperoleh dari Silhouette Coefficient dinilai lebih representatif untuk menggambarkan pola pengelompokan kabupaten/kota berdasarkan indikator pembangunan manusia di Provinsi Bali, Nusa Tenggara Barat, dan Nusa Tenggara Timur.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan kabupaten/kota di Bali, Nusa Tenggara Barat (NTB), dan Nusa Tenggara Timur (NTT) berdasarkan kemiripan indikator pembangunan manusia. Temuan ini sejalan dengan penelitian [10] dan [11] yang menyatakan bahwa K-Means efektif digunakan untuk mengidentifikasi kelompok wilayah yang memiliki karakteristik sosial-ekonomi yang serupa.

Dalam menentukan jumlah cluster, penelitian ini menemukan bahwa Metode Elbow menghasilkan tiga cluster, sedangkan Silhouette Coefficient menghasilkan dua cluster. Hasil evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa pengelompokan dengan dua cluster menghasilkan kualitas yang lebih baik dibandingkan tiga cluster. Temuan ini sejalan dengan penelitian [14], [16], dan [17] yang menunjukkan bahwa Silhouette Coefficient mampu menghasilkan jumlah cluster yang lebih representatif karena mempertimbangkan tingkat kohesi dan separasi cluster secara bersamaan.

Perbedaan hasil tersebut menunjukkan bahwa karakteristik data IPM pada wilayah Bali, NTB, dan NTT cenderung membentuk dua kelompok utama, yaitu wilayah dengan tingkat pembangunan manusia tinggi dan wilayah dengan tingkat pembangunan manusia menengah hingga rendah. Dengan demikian, penelitian ini memperkuat temuan sebelumnya bahwa pemilihan metode penentuan jumlah cluster dapat memengaruhi kualitas hasil clustering dan interpretasi karakteristik wilayah yang terbentuk.

4. KESIMPULAN

Berdasarkan hasil penelitian, algoritma K-Means clustering mampu mengelompokkan kabupaten/kota di Provinsi Bali, Nusa Tenggara Barat (NTB), dan Nusa Tenggara Timur (NTT) berdasarkan indikator Indeks Pembangunan Manusia (IPM), yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita. Metode Elbow menghasilkan jumlah cluster optimal sebanyak tiga cluster ($K=3$), sedangkan Silhouette Coefficient menghasilkan dua cluster ($K=2$). Hasil evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa pengelompokan dengan $K=2$ memiliki kualitas yang lebih baik dibandingkan $K=3$, ditunjukkan oleh nilai DBI yang lebih rendah serta pemisahan cluster yang lebih jelas pada visualisasi PCA. Temuan ini menunjukkan bahwa karakteristik data IPM kabupaten/kota di Bali, NTB, dan NTT cenderung membentuk dua kelompok utama, yaitu wilayah dengan tingkat pembangunan manusia tinggi dan wilayah dengan tingkat pembangunan manusia menengah hingga rendah. Hasil penelitian juga menunjukkan bahwa Silhouette Coefficient lebih representatif dalam menentukan jumlah cluster pada data IPM karena mempertimbangkan tingkat kekompakan dan keterpisahan cluster secara bersamaan. Sebaliknya, Metode Elbow menghasilkan pemisahan tambahan antara kelompok wilayah yang masih memiliki karakteristik relatif berdekatan, sehingga kualitas pengelompokan menjadi kurang optimal. Selain memberikan gambaran mengenai pola ketimpangan pembangunan manusia antarwilayah, hasil penelitian ini dapat menjadi dasar bagi pemerintah daerah dalam menetapkan prioritas pembangunan pada wilayah

dengan tingkat pembangunan manusia menengah hingga rendah, terutama melalui peningkatan akses pendidikan, layanan kesehatan, dan kesejahteraan ekonomi masyarakat.

REFERENCES

- [1] D. Azfirmawarman, L. Magriasti, and Yulhendri, “Indeks Pembangunan Manusia Di Indonesia (Kajian Perubahan Metodologi Penghitungan) Dony,” *J. Pendidik. dan Konseling*, vol. 5, no. 5, pp. 117–125, 2023, doi: <https://doi.org/10.31004/jpdk.v5i5.22864>.
- [2] B. P. Statistik, “Indeks Pembangunan Manusia 2024,” 2025. [Online]. Available: <https://www.bps.go.id/id/infographic?id=1029>
- [3] M. Andreanto, B. C. Amanullah, and N. Wakhidah, “Analisis K-Means Clustering pada Kabupaten/Kota di Indonesia Berdasarkan Indeks Pembangunan Manusia (IPM),” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 11, no. 1, pp. 227–239, 2026, doi: <https://doi.org/10.29100/jipi.v11i1.7455>.
- [4] Wikipedia, “Kepulauan Nusa Tenggara.” Accessed: Mar. 15, 2026. [Online]. Available: https://id.wikipedia.org/wiki/Kepulauan_Nusa_Tenggara
- [5] *Peraturan Presiden (Perpres) Nomor 63 Tahun 2020 tentang Penetapan Daerah Tertinggal Tahun 2020-2024*. 2020. [Online]. Available: <https://peraturan.bpk.go.id/Details/136563/perpres-no-63-tahun-2020>
- [6] I. Julistia, M. Kurniawan, and D. M. Hasimi, “The relationship between education expenditure, health expenditure and final education level towards the human development index from a shariah economic perspective: Evidence from 10 provinces with the lowest HDI in Indonesia,” *J. Ilm. Manajemen, Ekon. Akunt.*, vol. 9, no. 1, pp. 46–75, 2025, doi: <https://doi.org/10.31955/mea.v9i1.4871>.
- [7] K. Yulianawan, “Metode K-Means Clustering Untuk Mengelompokkan Indeks Pembangunan Manusia (IPM) Provinsi Papua,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 2911–2917, 2025, doi: <https://doi.org/10.36040/jati.v9i2.13243>.
- [8] S. Annas, B. Poerwanto, and M. F. S., “Implementation of K-Means Clustering on Poverty Indicators in Indonesia,” *J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 257–266, 2022, doi: [10.30812/matrik.v21i2.1289](https://doi.org/10.30812/matrik.v21i2.1289).
- [9] E. Merliana, N. Putu, E. Ernawati, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering,” in *Seminar Nasional Multi Disiplin Ilmu Unisbank 2015*, 2015.
- [10] K. D. R. Sianipar and I. Gunawan, “Algoritma K-Means Dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia Di Sumatera Utara,” *J. Infomedia Tek. Inform. Multimed. Jar.*, vol. 6, no. 2, pp. 57–62, 2021, doi: <http://dx.doi.org/10.30811/jim.v6i2.2426>.
- [11] D. Fariha, D. Hartawan, M. Al-Abrou, Muhammad Raffi Shafwan, and M. Ula, “Pengelompokan Wilayah Indonesia Berdasarkan Indikator Sosio-Ekonomi Tahun 2021 dengan Menggunakan Metode K-Means,” *Semin. Nas. Teknol. Inf. dan Tek. Inform.*, vol. 1, no. 1, pp. 1–6, 2024, [Online]. Available: <https://proceedings.unimal.ac.id/senastika/article/view/835/570>
- [12] D. Fitrianiingsih and M. D. Kartikasari, “Penerapan K-Means Clustering dengan Metode Elbow untuk Mengelompokkan Kabupaten / Kota Berdasarkan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Barat,” *Emerg. Stat. Data Sci. J.*, vol. 2, no. 2, pp. 250–257, 2024, doi: <https://doi.org/10.20885/esds.vol2.iss.2.art18>.
- [13] M. R. Putri, G. S. Nugraha, and R. Dwiyanaputra, “Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering,” *J-COSINE (Journal Comput. Sci. Informatics Eng.)*, vol. 7, no. 1, pp. 1–8, 2023, doi: <https://doi.org/10.29303/jcosine.v7i1.509>.
- [14] P. Vania and B. N. Sari, “Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Kluster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. November, pp. 547–558, 2023, doi: <https://doi.org/10.5281/zenodo.10081332>.
- [15] A. M. Putri, A. A. Picaulima, A. A. N. Subakti, and H. M. Winata, “Analisis Clustering Provinsi Di Indonesia Berdasarkan Indikator Pembangunan Manusia Untuk Mendukung Perencanaan Kebijakan Publik,” *DISTINGSI J. Digit. Soc.*, vol. 4, no. 1, pp. 57–68, 2025, [Online]. Available: <https://jurnal.panengeninsani.com/index.php/djods/article/view/114>
- [16] P. K. Ritonga and M. S. Hasibuan, “Analisis Perbandingan Silhouette dengan Elbow pada Algoritma K-Means dan DBSCAN,” *Media Teknol. Indormasi dan Komput. J.*, vol. 9, no. 1, pp. 64–71, 2025, doi: [10.47002/metik.v9i1.1027](https://doi.org/10.47002/metik.v9i1.1027).
- [17] S. Juanita and R. D. Cahyono, “K-means clustering with comparison of elbow and silhouette methods for medicines clustering based on user reviews,” *J. Tek. Inform.*, vol. 5, no. 1, pp. 283–289, 2024, doi: <https://doi.org/10.52436/1.jutif.2024.5.1.1349>.
- [18] M. Naufal *et al.*, “Analisis K-Means Cluster Kabupaten / Kota di Provinsi Kalimantan Selatan berdasarkan Indikator Indeks Pembangunan Manusia,” *Komputika J. Sist. Komput.*, vol. 14, no. 2, pp. 125–131, 2025, doi: <https://doi.org/10.34010/komputika.v14i2.16301>.
- [19] M. W. Diantika, A. Rusgiyono, and B. A. Saputra, “Perbandingan Metode Optimasi Silhouette, Elbow, dan Gap Statistics dalam Menentukan Banyaknya Cluster Terbaik Pada Analisis K-Means Clustering,” *J. GAUSSIAN*, vol. 14, no. 2012, pp. 335–344, 2025, doi: [10.14710/j.gauss.14.2.335-344](https://doi.org/10.14710/j.gauss.14.2.335-344).
- [20] R. Y. Sari, H. Oktavianto, and H. W. Sulistyono, “Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia,” *J. Smart Teknol.*, vol. 3, no. 2, pp. 104–108, 2022, [Online]. Available: <https://jurnal.unmuhjembar.ac.id/index.php/JST/article/view/6928/0>
- [21] W. S. Agustin, R. Armadhan, S. K. Dini, and H. Murda, “Comparison of K-Means and K-Medoids Methods in Grouping Provinces in Indonesia Based on Economic Development Indicators,” *J. Math. Comput. Stat.*, vol. 9, no. 1, pp. 131–147, 2026, doi: <https://doi.org/10.35580/jmathcosv9n19866>.
- [22] I. W. Septiani, A. C. Fauzan, and M. M. Huda, “Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru,” *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 556, 2022, doi: [10.30865/json.v3i4.4055](https://doi.org/10.30865/json.v3i4.4055).