

# **Analisis Data Indeks Pembangunan Manusia Kabupaten/ Kota di Jawa Timur melalui Pengelompokan menggunakan Metode Clustering Aglomeratif Hirarki**

## ***Analysis of Regency/City Human Development Index Data in East Java Through Grouping Using Hierarchical Agglomerative Clustering Method***

**Roudlotul Jannah Alfirdausy\***, Nurissaidah Ulinnuha, Moh. Hafiyusholeh  
Matematika, Sains dan Teknologi, UIN Sunan Ampel  
Jalan Ir. H. Soekarno, Kecamatan Gunung Anyar, No. 682., Surabaya, Jawa Timur, Indonesia  
\*e-mail: [0903022048@student.uinsby.ac.id](mailto:0903022048@student.uinsby.ac.id)

(received: 13 Juni 2023, revised: 5 Agustus 2023, accepted: 9 Agustus 2023)

### **Abstrak**

Evaluasi pembangunan manusia umumnya dilakukan menggunakan Indeks Pembangunan Manusia (IPM), yang mengukur tingkat pembangunan berdasarkan beberapa aspek penting dalam kualitas hidup. Di Jawa Timur, IPM dikategorikan sebagai tinggi. Namun, distribusi IPM di antara Kabupaten/Kota di Jawa Timur masih belum merata. Oleh karena itu, diperlukan pengelompokan Kabupaten/Kota berdasarkan IPM dan pencapaian setiap indikator yang berkontribusi terhadap IPM tersebut. Pengelompokan adalah teknik analisis data yang digunakan untuk mengelompokkan data serupa. Metode pengelompokan hierarki agglomerative merupakan salah satu metode yang digunakan untuk tujuan ini. Tujuan dari penelitian ini adalah memberikan referensi kepada pemerintah untuk memahami distribusi kelompok karakteristik di antara Kabupaten/Kota berdasarkan profil IPM mereka di Jawa Timur. Analisis data IPM Jawa Timur tahun 2021 menunjukkan bahwa metode dan kelompok terbaik diperoleh menggunakan Average Linkage, dengan nilai koefisien Cophenetic sebesar 0,8105891, menghasilkan dua kelompok. Kelompok dengan nilai koefisien Silhouette tertinggi sebesar 0,6196077 terdiri dari 34 Kabupaten/Kota dan diklasifikasikan sebagai kelompok rendah, sementara kelompok tinggi terdiri dari empat Kota/Kabupaten.

**Kata kunci:** IPM, clustering agglomeratif hirarki, Average Linkage

### **Abstract**

*The evaluation of human development is typically done using the Human Development Index (HDI), which measures the level of development in terms of various essential aspects of quality of life. In the case of East Java, the HDI is categorized as high. However, the distribution of HDI among the Regencies/Cities in East Java is still uneven. Therefore, it becomes necessary to cluster the districts/cities based on their HDI and the achievement of each indicator contributing to the HDI. Clustering is a data analysis technique used to group similar data together. Hierarchical agglomerative clustering is one of the methods used for this purpose. The aim of this study is to provide a reference for the government to understand the distribution of characteristic groupings among the districts/cities based on their HDI profiles in East Java. The analysis of East Java's HDI data for 2021 revealed that the best method and cluster was obtained using Average Linkage, with a Cophenetic coefficient value of 0.8105891, resulting in two clusters. The cluster with the highest Silhouette coefficient value of 0.6196077 comprised 34 districts/cities, classified as the low cluster, while the high cluster consisted of four cities/regencies.*

**Keywords:** HDI, hierarchical agglomerative clustering, Average Linkage

## 1 Pendahuluan

Pembangunan manusia adalah sebuah konsep perkembangan yang menempatkan individu atau populasi sebagai fokus utama dan tujuan utama dari semua upaya pembangunan yang dilakukan di suatu wilayah. Hal ini mengarah pada konsep kemajuan yang bertujuan untuk meningkatkan kualitas hidup dan kesejahteraan masyarakat secara holistik[1]. Indeks Pembangunan Manusia (IPM), yang didefinisikan oleh UNDP, berfungsi sebagai alat untuk mengevaluasi kemajuan pembangunan manusia dalam suatu wilayah dengan berfokus pada berbagai aspek penting yang mengukur kualitas hidup masyarakat. IPM dirancang dengan memanfaatkan pendekatan dasar tiga dimensi yang mendasari sebagai indikator untuk mengukur kualitas hidup manusia. Aspek-aspek dimensi ini mencakup harapan hidup yang panjang dan sehat, pengetahuan, serta standar kehidupan yang memadai. [2].

Pada tahun 2021, IPM Jawa Timur mencapai angka 72,14. Angka tersebut merupakan kenaikan dari IPM tahun sebelumnya, yakni 71,71. Peningkatan tersebut berdasarkan pada indikator-indikator IPM Jawa Timur, di mana harapan hidup saat lahir mencapai 71,38 tahun, harapan lama sekolah sebesar 13,36 tahun, rata-rata lama sekolah sekitar 7,88 tahun, dan pengeluaran per kapita sejumlah Rp. 11.707,- [3]. Kenaikan tersebut menjadi acuan bahwasanya proses pembangunan di Jawa Timur mengalami kemajuan.

Semakin tinggi IPM suatu daerah, dapat dikatakan bahwa daerah tersebut mengalami keberhasilan dalam proses pembangunan wilayahnya[4]. Namun, persebaran IPM Kabupaten/Kota di Jawa Timur masih belum merata. Pada tahun 2021, terdapat 17 wilayah yang berada diatas IPM Jawa Timur, sedangkan 21 wilayah lainnya masih di bawah. Untuk mewujudkan pembangunan yang merata, dibutuhkan pengelompokan wilayah yang membutuhkan fokus lebih untuk meningkatkan keberhasilan dalam proses pembangunannya [5]. Salah satu langkah yang dapat diambil untuk mencapai tujuan tersebut adalah dengan melakukan pemetaan atau pengelompokan Kabupaten/Kota berdasarkan IPM Jawa Timur dalam hal pencapaian masing-masing indikator yang membentuk IPM.

Terdapat dua metode untuk melakukan pengelompokan/*clustering* diantaranya dengan analisis *cluster* hirarki dan non-hirarki. Metode hirarki meliputi pendekatan *agglomerative* dan *divisive*. Selain itu pendekatan *partitioning* dan *graphical* dilakukan pada analisis cluster metode non-hirarki [6]. Metode yang termasuk pada pendekatan *agglomerative* yaitu *Agglomerative Hierarchical Clustering* (AHC) dengan jenis pengelompokan *Single Linkage Method*, *Complete Linkage Method* dan *Average Linkage Method* [7]

AHC memiliki beberapa kelebihan yang membuatnya menjadi metode yang populer dalam analisis pengelompokan data [8]. Salah satu kelebihannya adalah kemampuannya untuk menghasilkan pengelompokan hierarki yang memungkinkan identifikasi berbagai tingkatan kelompok yang ada dalam data. Metode ini tidak memerlukan jumlah kelompok yang ditentukan sebelumnya, sehingga tidak ada kebutuhan untuk mengestimasi atau menebak jumlah cluster yang optimal sebelum melakukan analisis. Selain itu, metode ini dapat memberikan informasi tentang tingkat kesamaan atau perbedaan antara kelompok yang ada, serta memperlihatkan hubungan antar kelompok secara visual melalui dendrogram. Dengan demikian, AHC memberikan fleksibilitas dan interpretasi yang lebih baik dalam pengelompokan data dibandingkan dengan metode *clustering* lainnya.

Terdapat pula penelitian lain dengan *clustering* kabupaten/kota di Provinsi Maluku dengan pemanfaatan algoritma k- means menurut evaluasi beberapa karakteristik daerah dengan 5 pengukuran IPM. Diperoleh hasil akhir tiga *cluster* berdasarkan IPM [9]. Penelitian lainnya mengenai perbandingan antara Metode, *K-Means*, *Fuzzy C-Means* dan *Agglomerative Hierarchical Clustering* (AGNES) diperoleh kesimpulan bahwa *Fuzzy C-Means* dan *K-Means* cocok untuk data besar, sedangkan metode hirarki yang mudah dipahami dan mudah untuk diimplementasikan [10].

Penelitian ini memiliki kontribusi dalam konteks pembangunan manusia di Jawa Timur dimana pembangunan manusia ditempatkan sebagai fokus utama dan tujuan utama dari semua upaya pembangunan di wilayah tersebut. Meskipun demikian, terdapat ketimpangan dalam persebaran IPM di berbagai Kabupaten/Kota di Jawa Timur. Oleh karena itu, penelitian ini memberikan kontribusi dengan melakukan pemetaan atau pengelompokan Kabupaten/Kota berdasarkan IPM, sehingga pemerintah dapat lebih terfokus untuk meningkatkan pembangunan di setiap wilayah secara merata.

Dengan menggunakan metode *Agglomerative Hierarchical Clustering*, penelitian ini bertujuan untuk memberikan dasar yang dapat digunakan dalam memahami karakteristik dan distribusi kelompok Kabupaten/Kota berdasarkan profil IPM di Jawa Timur. Dengan demikian, diharapkan penelitian ini

akan membantu pemerintah dalam mencapai keberhasilan dalam proses pembangunan setiap Kabupaten/Kota di Jawa Timur.

## 2 Tinjauan Literatur

Analisis Cluster adalah proses analisis data yang digunakan untuk mengidentifikasi pola tersembunyi atau mengelompokkan data. Metode kluster merupakan salah satu pendekatan dalam *unsupervised learning*, adalah jenis algoritma pembelajaran mesin yang mempelajari dataset berdasarkan jarak atau kedekatan antara objek-objeknya [11]. *Clustering* biasanya digunakan untuk mengelompokkan suatu data yang memiliki nilai persamaan karakteristik dalam suatu *cluster* yang identik serta data dengan perbedaan nilai karakteristik dikelompokkan kedalam *cluster* yang lain [7].

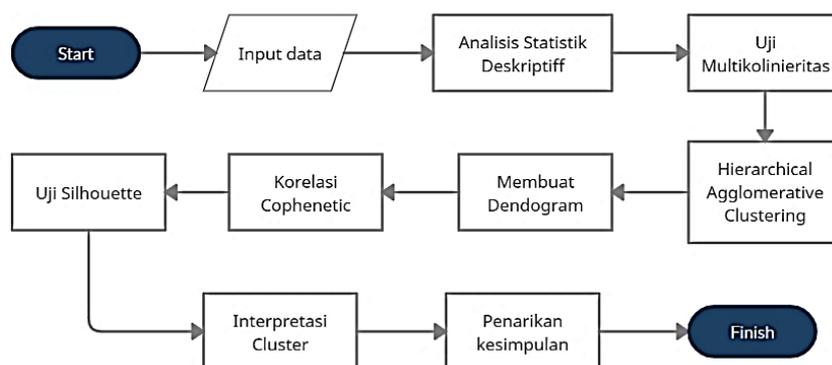
Dalam penerapan klusterisasi, setiap anggota dikelompokkan berdasarkan kesamaan atau perbedaan karakteristik dengan mempertimbangkan kluster yang terbentuk. Clustering memiliki banyak pengaplikasian untuk memecahkan masalah seperti identifikasi suatu golongan maupun sekumpulan data, deteksi anomali, pengenalan pola, dan pengolahan citra yang dapat digunakan dalam berbagai situasi [12]. Secara umum, proses clustering dengan algoritma yang berbeda memiliki dasar yang sama, namun terdapat perbedaan dalam metode pengukuran kesamaan (atau jarak) serta penentuan cluster instance [13].

*Hierarchical Clustering* merupakan metode pengelompokan berdasarkan tingkat tertentu antara data dengan menggunakan struktur hierarki untuk memisahkan atau mengumpulkan sampel untuk menarik kesimpulan [14]. Algoritma ini akan membentuk struktur hirarkis atau didasarkan pada level atau tingkatan tertentu yang mirip struktur pohon. Proses pengelompokan dilakukan secara bertahap atau bertingkat. Data yang digunakan dalam pengelompokan ini tidak terlalu besar dan jumlah *cluster* yang terbentuk tidak diketahui [7].

Umumnya, *Agglomerative (Bottom-Up)* dikenal sebagai metode penggabungan untuk pengelompokan hirarkis dengan pendekatan bawah-atas yang dimulai dengan setiap objek dalam *cluster* terpisah untuk membentuk *cluster* baru [7]. Pertama, setiap objek dalam data dianggap sebagai *cluster* individu. Kemudian, dihitung jarak antara setiap pasangan *cluster* menggunakan metrik jarak. Selanjutnya, dua *cluster* yang memiliki jarak terdekat digabungkan menjadi satu cluster baru. Jarak antara *cluster* baru dengan *cluster* lainnya dihitung menggunakan metode *linkage*, seperti Single Linkage, Complete Linkage, atau Average Linkage [15].

## 3 Metode Penelitian

Dalam analisis ini, digunakan pendekatan kuantitatif dengan memanfaatkan data yang bersumber dari Badan Pusat Statistik Provinsi Jawa Timur, yang diakses melalui situs resmi mereka di <https://jatim.bps.go.id/>. Data yang dianalisis mencakup variabel Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Angka Harapan Hidup (AHH) digunakan sebagai indikator proyeksi rata-rata tahun hidup seseorang di suatu daerah. Penghitungan AHH merujuk pada rentang nilai minimal dan maksimal harapan hidup yang telah diakui oleh standar UNDP. Rentang nilai ini memiliki batas atas 85 tahun dan batas bawah 20 tahun. Selanjutnya, Angka Melek Huruf (AMH) mengukur persentase penduduk usia 15 tahun ke atas yang memiliki kemampuan membaca dan menulis. AMH merupakan indikator penting dalam mengukur tingkat literasi dan akses pendidikan di suatu wilayah. Rata-rata Lama Sekolah (RLS) mengukur rata-rata lama waktu yang dihabiskan seseorang dalam menyelesaikan pendidikan formal. RLS mencerminkan tingkat partisipasi dan aksesibilitas pendidikan di suatu daerah. Pengeluaran per kapita (PPK) merujuk pada jumlah uang yang dihabiskan oleh individu atau rumah tangga dalam periode tertentu untuk pemenuhan kebutuhan konsumsi. PPK dapat mencerminkan tingkat kesejahteraan ekonomi masyarakat dan kemampuan mereka dalam memenuhi kebutuhan dasar.



Gambar 1. Flowchart Hierarchical Agglomerative Clustering

Penelitian ini menggunakan objek 9 Kota dan 29 Kabupaten di Provinsi Jawa Timur. Teknik dalam pemrosesan data menggunakan perangkat lunak RStudio. *Clustering* data diproses berdasarkan *flowchart* pada Gambar 1 menggunakan metode *Agglomerative Hierarchical Clustering* dengan langkah dan analisis sebagai berikut:

1. Mengumpulkan data masukan
2. Melakukan analisis deskriptif terhadap data.
3. Melakukan uji multikolinieritas pada data. Dalam konteks clustering, multikolinearitas dapat menjadi penting karena dapat mempengaruhi hasil pengelompokan dan interpretasi hasil. Hal ini karena multikolinearitas menyebabkan redundansi informasi. Jika terdapat variabel-variabel yang memiliki korelasi tinggi, maka pembentukan cluster menjadi tidak informatif atau tidak memisahkan dengan baik antara anggota *cluster* yang berbeda. Uji multikolinearitas dapat dilihat pada nilai *Variance Inflation Factor* (VIF):

$$VIF = \frac{1}{(1-R_j^2)} \quad (1)$$

Dengan

R : Koefisien korelasi variabel bebas

Keputusan : Terdapat multikolinearitas apabila nilai  $VIF \geq 10$

4. Melakukan *Hierarchical Agglomerative Clustering* dengan teknik penggabungan *cluster*:
  - a. *Single linkage*

Teknik ini menggabungkan *cluster* sesuai dengan jarak antara anggota terdekat. Jika dua objek dipisahkan oleh jarak terpendek, kedua objek tersebut digabungkan menjadi sebuah *cluster* yang diperoleh berdasarkan rumus:

$$d_{(xy)z} = \min\{d_{xz}, d_{yz}\} \quad (2)$$

- b. *Average linkage*

Teknik ini menggabungkan *cluster* sesuai dengan jarak rata-rata antara anggota. Jika dua objek dipisahkan oleh jarak rata-rata, kedua objek tersebut digabungkan menjadi sebuah *cluster* yang diperoleh berdasarkan rumus:

$$d_{(xy)z} = \text{average}\{d_{xz}, d_{yz}\} \quad (3)$$

- c. *Complete linkage*

Teknik ini menggabungkan *cluster* sesuai dengan jarak antara anggota terjauh. Jika dua objek dipisahkan oleh jarak yang panjang atau jauh, kedua objek tersebut digabungkan menjadi sebuah *cluster* yang diperoleh berdasarkan rumus:

$$d_{(xy)z} = \max\{d_{xz}, d_{yz}\} \quad (4)$$

Keterangan:

$d_{xz}$  adalah jarak dari klaster  $x$  dan  $z$

$d_{yz}$  adalah jarak dari klaster  $y$  dan  $z$

5. Membuat dendrogram. Dendrogram adalah representasi grafis dari hierarki pengelompokan dalam analisis clustering. Dendrogram digunakan untuk memvisualisasikan hubungan antara objek-objek yang dikelompokkan berdasarkan jarak
6. Menghitung nilai korelasi *Cophenetic* adalah

$$r_{\text{Coph}} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{c_{ik}} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2]}} \quad (5)$$

Dengan:

$r_{\text{Coph}}$  : koefisien korelasi *cophenetic*

$\bar{d}$  : rata-rata  $d_{ik}$

$d_{ik}$  : jarak asli (jarak *Euclidean*) antara objek  $i$  dan  $k$

$d_{c_{ik}}$  : jarak *cophenetic* objek  $i$  dan  $k$

$\bar{d}_c$  : rata-rata  $d_{c_{ik}}$

7. Melakukan uji *Shilhouette* adalah

$$s(i) = (b(i) - a(i)) / m \quad (a(i), b(i)) \quad (6)$$

Dengan  $s(i)$  merupakan semua rata-rata seluruh kumpulan data.

8. Melakukan Interpretasi cluster
9. Penarikan kesimpulan

## 4 Hasil dan Pembahasan

Pendekatan clustering aglomeratif hirarki muncul sebagai alat analisis yang kuat untuk mengelompokkan kabupaten dan kota berdasarkan kesamaan karakteristik pembangunan manusia. Metode ini memungkinkan kita untuk mengidentifikasi kelompok-kelompok yang memiliki tingkat kesamaan yang tinggi dalam hal aspek-aspek yang relevan dengan pembangunan manusia. Berikut hasil analisis *agglomerative* yaitu *Agglomerative Hierarchical Clustering* (AHC) dengan jenis pengelompokan *Single Linkage Method*, *Complete Linkage Method* dan *Average Linkage Method*.

### 4.1 Statistik Deskriptif

Berdasarkan data yang digunakan dalam analisis cluster, mulanya dilakukan analisis statistik deskriptif agar diperoleh suatu gambaran umum data yang akan digunakan dalam analisis cluster. Terdiri dari jumlah data, minimum, mean, maksimum, dan standar deviasi.

**Tabel 1. Analisis Deskriptif Indikator IPM**

Variabel	N	Min	Mean	Max	St. Deviasi
Umur Harapan Hidup (UHH)	38	66,89	71,72	74.18	1,97
Harapan Lama Sekolah (HLS)	38	11,73	13,40	15,75	0,90
Rata-Rata Lama Sekolah (RLS)	38	4,860	8,061	11,370	1,58
Pengeluaran Per Kapita (PPK) dalam rupiah	38	8.673	11.569	17.862	2.216,22

Pada tabel 1 diatas, kabupaten dengan umur harapan hidup terkecil yaitu Kabupaten Bondowoso, sedangkan nilai tertinggi Kota Surabaya. Harapan lama sekolah tertinggi pada Kota Malang, sedangkan nilai terendah pada Kabupaten Bangkalan. Rata-rata lama sekolah tertinggi terletak di Kota Madiun dan rata-rata terendah pada Kabupaten Sampang. Pengeluaran perkapita Kota Surabaya merupakan Pengeluaran Perkapita tertinggi di Jawa Timur, sebaliknya Kabupaten Bangkalan termasuk kabupaten dengan pengeluaran perkapita terendah.

### 4.2 Uji Multikolinieritas

Hasil uji asumsi multikolinieritas dengan memperhatikan nilai VIF pada setiap variabel yang disajikan pada Tabel 2.

**Tabel 2. Nilai VIF Indikator IPM**

Variabel	VIF
Umur Harapan Hidup (UHH)	2,39
Harapan Lama Sekolah (HLS)	2,77
Rata-rata Lama Sekolah (RLS)	8,34
Pengeluaran Per Kapita (PPK) dalam rupiah	4,72

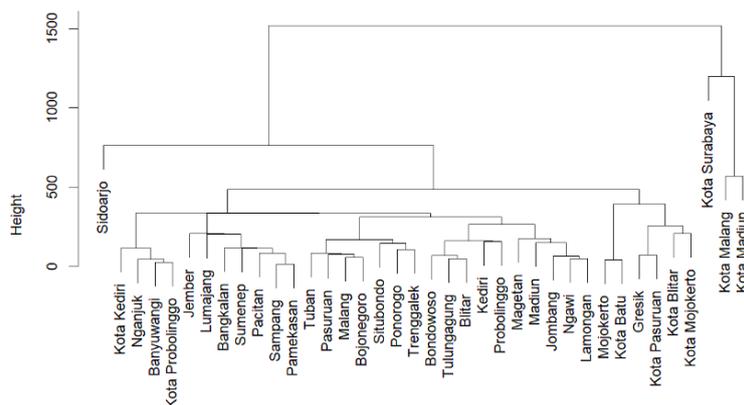
Berdasarkan tabel 2 didapatkan nilai VIF setiap variabel bernilai < 10 yang berarti setiap variabel tidak terindikasi adanya multikolinieritas.

### 4.3 Hierarchical Agglomerative Clustering

Visualisasi hasil *clustering* disajikan dengan dendrogram sebanyak jumlah objek berupa 38 Kabupaten/Kota di Jawa Timur.

#### 4.3.1 Single linkage

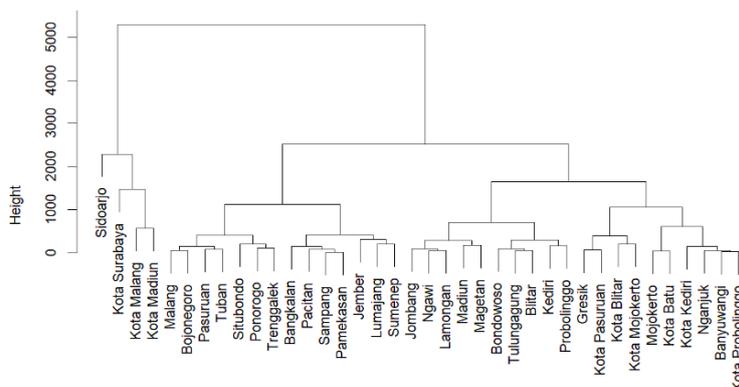
*Clustering* dengan memanfaatkan metode *Single Linkage*. Proses *clustering* menggabungkan *cluster* menurut jarak antara data atau objek terdekat dari dua *cluster*. Hasil pengelompokan dengan metode *Single Linkage* berupa dendrogram pada gambar 2 dibawah ini.



**Gambar 2. Dendrogram Metode Single Linkage**

#### 4.3.2 Average linkage

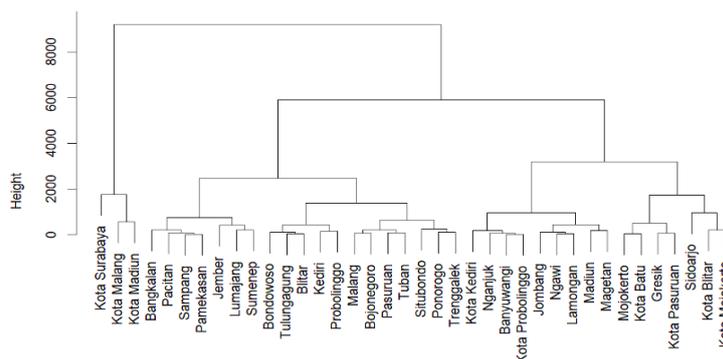
Metode *Average linkage* digunakan dalam proses *clustering* dengan menggabungkan jarak rata-rata antara data atau objek antara 2 *cluster*. Hasil pengelompokan dengan metode *average linkage* berupa dendrogram pada Gambar 3 di bawah ini.



**Gambar 3. Dendrogram Metode Average Linkage**

### 4.3.3 Complete linkage

*Complete linkage* digunakan dalam proses *clustering* dengan menggabungkan jarak terjauh antara data atau objek antara 2 *cluster*. Hasil pengelompokan dengan metode *complete linkage* berupa dendrogram pada Gambar 4 di bawah ini.



**Gambar 4. Dendrogram Metode Complete Linkage**

### 4.4 Uji Korelasi Cophenetic

Pada penelitian ini menggunakan analisis *cluster* berupa *Agglomerative Hierarchical Clustering* dengan metode *Single Linkage*, *Average Linkage*, dan *Complete Linkage*. Kemudian membandingkan nilai korelasi *Cophenetic* ketiga metode tersebut pada tabel 3.

**Tabel 3. Nilai Koefisien Korelasi Cophenetic**

Metode	Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0.7951813
<i>Average Linkage</i>	0.8105901
<i>Complete Linkage</i>	0.8008910

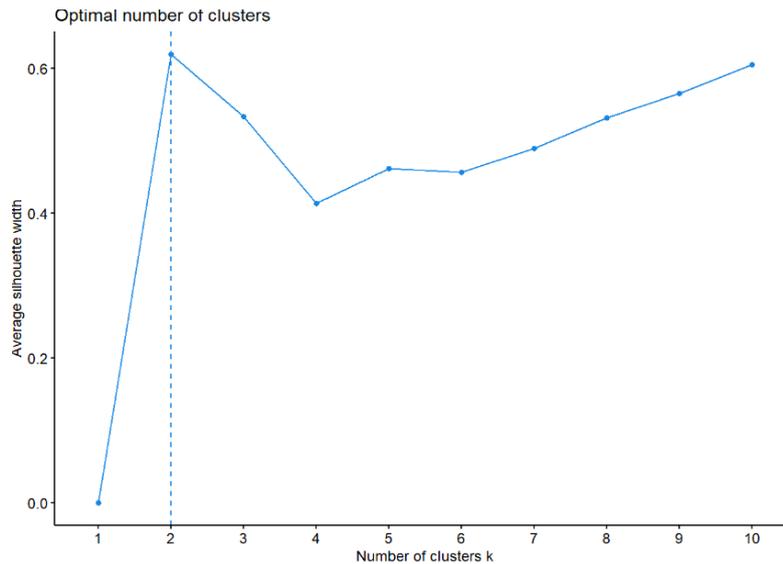
Berdasarkan tabel 3, metode *Average Linkage* bernilai lebih besar dibandingkan dua metode lainnya yaitu sebesar 0,8105891. Dengan demikian metode *Average Linkage* merupakan metode *clustering* terbaik untuk digunakan dalam penelitian ini

### 4.5 Uji Silhouette

Nilai rata-rata koefisien *Silhouette* setiap nilai dari jumlah *cluster* yang berbeda untuk mengevaluasi *cluster*. Pengujian dijalankan dengan nilai *cluster* 2 sampai 10. Hasil uji berisi nilai koefisien *silhouette* pada tabel 4.

**Tabel 4. Nilai Koefisien Silhouette**

Jumlah Cluster	Koefisien <i>Silhouette</i>
2	0.6196091
3	0.5339319
4	0.4139281
5	0.4619507
6	0.4566029
7	0.4900706
8	0.5316800
9	0.5654363
10	0.6050196



**Gambar 5. Diagram nilai koefisien Shilhoute**

Hasil pada Gambar 5 didapatkan bahwa nilai koefisien *Silhouette 2 cluster* merupakan nilai tertinggi dengan skor 0.6196091 yang menunjukkan bahwa pengelompokan yang dilakukan memiliki tingkat pemisahan (*separation*) yang relatif baik, dengan angka yang mendekati 1 menunjukkan tingkat keseragaman dan kekompakan yang tinggi dalam setiap cluster. Dengan demikian, hasil pengelompokan dapat dianggap cukup baik dalam memisahkan data ke dalam kelompok yang berbeda berdasarkan karakteristik IPM di Jawa Timur. Sedangkan nilai koefisien *Shilhoute* terendah ditemukan pada 4 cluster.

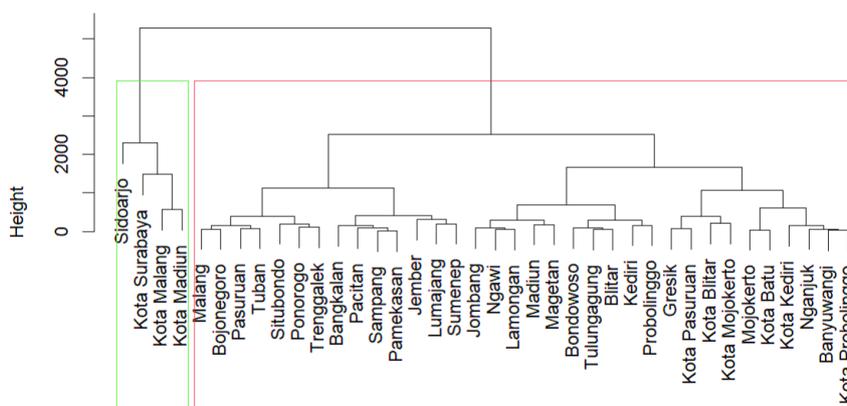
#### 4.6 Interpretasi Cluster

*Clustering* atau pengelompokan dengan *Agglomerative Hierarchical Clustering* menggunakan metode *Average Linkage* didapatkan 2 Cluster dimana berdasarkan Tabel 5, Cluster 1 merupakan cluster yang nilai seluruh variabelnya lebih rendah dibandingkan cluster 2.

**Tabel 5. Rata-Rata Cluster Tiap Indikator IPM**

Average	RLS	HLS	UHH	PPP
Cluster 1	71,4956	13,2132	7,744	11011,94
Cluster 2	73,6075	14,9775	10,750	16299,50

Hasil pengelompokan indikator IPM Jawa Timur berupa dendogram pada gambar 6.



**Gambar 6. Dendogram Pengelompokan dengan metode Average Linkage**

Berdasarkan gambar 6 dapat diketahui bahwa terdapat dua *cluster*, dimana *cluster 1* atau *cluster* rendah terdiri dari 34 Kabupaten/Kota diantaranya; Bangkalan, Pacitan, Sampang, Pamekasan, Jember, Lumajang, Sumenep, Bondowoso, Tulungagung, Blitar, Kediri, Probolinggo, Malang, Bojonegoro, Pasuruan, Tuban, Situbondo, Ponorogo, Trenggalek, Kota Kediri, Nganjuk, Banyuwangi, Kota Probolinggo, Jombang, Ngawi, Lamongan, Madiun, Magetan, Mojokerto, Kota Batu, Gresik, Kota Pasuruan, Kota Blitar, Kota Mojokerto. Sedangkan *cluster 2* atau *cluster* tinggi hanya ada 4 Kabupaten/Kota, yaitu Kota Surabaya, Kota Malang, Kota Madiun, dan Kabupaten Sidoarjo.

Pada dasarnya, keberadaan empat Kabupaten/Kota yang masuk dalam *cluster 2*, yaitu Kota Surabaya, Kota Malang, Kota Madiun, dan Kabupaten Sidoarjo, menunjukkan prestasi yang baik dalam pembangunan manusia. Kabupaten/Kota ini memiliki IPM dengan kategori "sangat tinggi" dengan nilai  $IPM \geq 80$  yang menunjukkan bahwa daerah telah mencapai tingkat kualitas hidup yang tinggi dalam berbagai aspek yang diukur melalui IPM. Namun, tetap penting untuk memastikan bahwa kesenjangan antara anggota *cluster* tinggi dan *cluster* rendah perlu dikurangi, sehingga keberhasilan pembangunan tidak hanya terkonsentrasi pada beberapa wilayah saja.

*Cluster* pertama, atau *cluster* rendah, terdiri dari sejumlah Kabupaten/Kota di Jawa Timur yang memiliki nilai IPM yang relatif rendah. Kabupaten/Kota dalam klaster ini, seperti yang disebutkan sebelumnya, menghadapi tantangan dalam meningkatkan kualitas hidup dan kesejahteraan masyarakatnya. Dalam konteks ini, perhatian khusus dari pemerintah sangat diperlukan untuk membantu memperbaiki kondisi di wilayah-wilayah tersebut. Dalam upaya meningkatkan pembangunan di Kabupaten/Kota dalam klaster rendah, pemerintah dapat mengarahkan sumber daya dan program-program pembangunan yang tepat untuk meningkatkan IPM. Dengan memberikan perhatian yang lebih pada Kabupaten/Kota dalam klaster rendah, pemerintah dapat mempercepat kemajuan pembangunan di wilayah tersebut, mengurangi kesenjangan antar daerah, dan meningkatkan kesejahteraan masyarakat secara keseluruhan.

Penelitian ini memiliki perbedaan dengan penelitian sebelumnya yang telah dilakukan, karena dalam penelitian ini metode atau jumlah klaster tidak ditentukan terlebih dahulu [8], sedangkan dalam penelitian ini menguji 3 metode *Agglomerative Hierarchical Clustering* berdasarkan nilai koefisien *Cophenetic* serta jumlah *cluster* dengan uji *Shillhoutte*. Metode *Agglomerative Hierarchical Clustering* mudah dipahami serta mudah untuk diimplementasikan [10].

## 5 Kesimpulan

Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan nilai indikator IPM tahun 2021 menggunakan metode *Agglomerative Hierarchical Clustering* menghasilkan beberapa temuan. Dalam analisis ini, Metode *Average Linkage* terbukti menjadi yang paling optimal dengan nilai koefisien *Cophenetic* tertinggi dibandingkan metode *Single Linkage* dan *Complete Linkage*, mencapai korelasi *Cophenetic* sebesar 0,8105891. Dengan penerapan metode *Average Linkage Clustering*, terjadi pemisahan menjadi 2 kelompok berdasarkan uji *Silhouette* dengan nilai koefisien tertinggi mencapai 0,6196077. Dalam penelitian ini, teridentifikasi bahwa terdapat 34 kabupaten/kota yang termasuk dalam *cluster 1* (*cluster* rendah) dan 4 Kabupaten/Kota yang termasuk dalam *cluster 2* (*cluster* tinggi).

## Referensi

- [1] M. R. Serang, "Pengaruh Pengeluaran Pemerintah, Produktivitas Tenaga Kerja dan Faktor Demografi terhadap Kinerja Pembangunan Manusia di Kabupaten/kota Provinsi Maluku," *Cita Ekon.*, vol. 11, no. 2, 2017.
- [2] K. C. Nawaji, *Indeks Pembangunan Manusia Kabupaten Bojonegoro 2021*. 2021. [Online]. Available: Uny.ac.id
- [3] BPS, "Indeks Pembangunan Manusia Provinsi Jawa Timur 2021," pp. 1–77, 2022.
- [4] J. A. Rosyadah, "Determinan Indeks Pembangunan Manusia (IPM) Provinsi Nusa Tenggara Timur (NTT)," *Indones. J. Dev. Econ.*, vol. 4, no. 1, pp. 1080–1092, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/efficient>
- [5] M. B. Setiawan and A. Hakim, "Indeks Pembangunan Manusia Indonesia," *J. Econ.* 9(1), 18–26, vol. 9(1), pp. 18–26, 2008, [Online]. Available: Uny.ac.id

- [6] D. Herawatie, “Perbandingan Algoritma Pengelompokan Non-Hierarki untuk Dataset Dokumen,” *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, pp. 11–16, 2014.
- [7] J. Gifari, T. A. Adinegara, S. F. H. Arildan, and D. Dewanti, *Modul Belajar Bootcamp Data Science & Machine Learning*. Bogor: Inspira Pustaka Aksara, 2021.
- [8] K. P. Simanjuntak and U. Khaira, “Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 7–16, 2021, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/6>
- [9] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, “Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017, doi: 10.30598/barekengvol11iss2pp119-128.
- [10] R. Rahmati, A. W. Wijayanto, P. Studi, K. Statistik, and P. Sains, “Rizqina Rahmati 1 , Arie Wahyu Wijayanto 2 Program Studi Komputasi Statistik Peminatan Sains Data, Politeknik Statistika STIS,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 73–80, 2021.
- [11] N. Buslim and R. P. Iswara, “Pengembangan Algoritma Unsupervised Learning Technique pada Big data Analysis di media sosial sebagai media promosi usaha online bagi masyarakat,” *J. Tek. Inform.*, vol. 12, no. 1, pp. 79–96, 2019, doi: 10.15408/jti.v12i1.11342.
- [12] T. Li, A. Rezaeipannah, and E. S. M. Tag El Din, “An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3828–3842, 2022, doi: 10.1016/j.jksuci.2022.04.010.
- [13] S. T. M. A. K. Wierzchoń, *Modern Algorithms of Cluster Analysis*. Cham: Springer, 2018.
- [14] D. Si, W. Hu, Z. Deng, and Y. Xu, “Fair hierarchical clustering of substations based on Gini coefficient,” *Glob. Energy Interconnect.*, vol. 4, no. 6, pp. 576–586, 2021, doi: 10.1016/j.gloi.2022.01.009.
- [15] N. A. N. S. Hendra Perdana, “Pencarian Cluster Optimum Pada Single Linkage, Complete Linkage Dan Average Linkage,” *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 8, no. 3, pp. 393–398, 2019, doi: 10.26418/bbimst.v8i3.33173.