

Penentuan Metode Kluster Hierarki Terbaik dengan Korelasi *Cophenetic* pada Pengelompokan Kabupaten/Kota di Indonesia Berdasarkan Variabel yang Memengaruhi Indeks Pembangunan Manusia

Determination of the Best Hierarchical Clustering Method with Cophenetic Correlation in the Clustering of Districts/Cities in Indonesia Based on Variables Affecting the Human Development Index

Syarifah Andini^{1*}, Wirda Andani², Nur Asih Kurniawati³

^{1,2}Universitas Tanjungpura, Jl. Prof. Dr. H. Hadari Nawawi, Pontianak;

³BPS Provinsi Kalimantan Barat, Jl. Sutan Syahrir, Pontianak

*Penulis Korespondensi. e-mail: h1091221018@student.untan.ac.id

ABSTRACT

The Human Development Index achievement in Indonesia 2023 was 73.55 point, indicating that the figure falls into the high category. However, when viewed based on the district/city regions, there is an imbalance in development. The disparity is evident from the Human Development Index achievements in six districts/cities in the DKI Jakarta Province, which fall into the high and very high categories, and the Human Development Index in the Papua Province, which predominantly falls into the low category. This indicates that there is an uneven distribution of human development across the districts/cities in Indonesia. This research aims to help design more targeted strategies to improve the people's quality of life, both in areas with low Human Development Index and in areas with high Human Development Index, through an approach based on the similarities in the socio-economic conditions of each region. The analysis in this study utilizes data that includes Life Expectancy, Expected Years of Schooling, Mean Years of Schooling, Adjusted Real Per Capita Expenditure, Open Unemployment Rate, and Minimum Wage. This study uses hierarchical cluster analysis by determining the optimal number of clusters using the NbClust package in RStudio software. The results of study indicate that the best method used is the Average Linkage method, which is divided into 5 clusters based on the characteristics of human development region.

Keywords: cluster, Human Development Index, Cophenetic correlation, Average Linkage

ABSTRAK

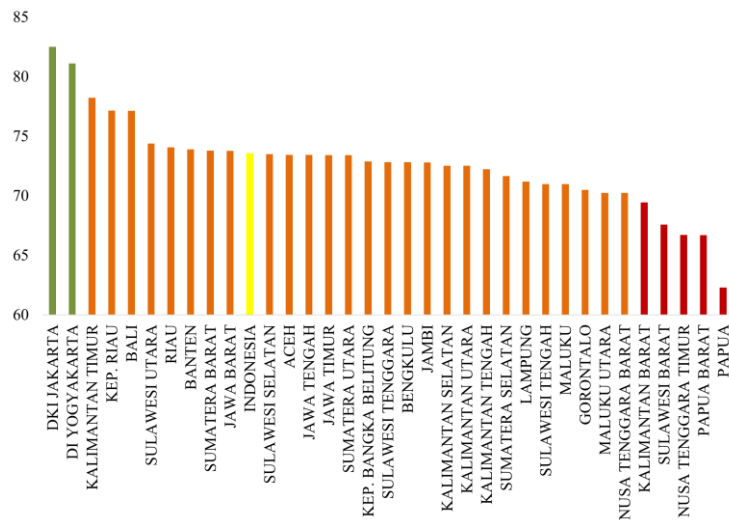
Capaian Indeks Pembangunan Manusia di Indonesia pada tahun 2023 mencapai 73,55 poin menunjukkan bahwa angka tersebut berada pada kategori tinggi. Akan tetapi, jika ditinjau berdasarkan wilayah kabupaten/kota, maka terjadi ketimpangan terhadap wilayah pembangunan. Ketimpangan ini bisa dilihat berdasarkan perbedaan capaian Indeks Pembangunan Manusia pada enam wilayah kabupaten/kota di Provinsi DKI Jakarta yang masuk dalam kategori tinggi dan sangat tinggi, sementara Provinsi Papua mendominasi pada kategori rendah. Hal tersebut menunjukkan bahwa terjadi ketimpangan pembangunan manusia pada wilayah kabupaten/kota di Indonesia. Penelitian yang dilakukan ini bertujuan untuk membantu merancang strategi yang lebih tepat sasaran untuk meningkatkan kualitas hidup masyarakat, baik di daerah dengan Indeks Pembangunan Manusia dengan kategori rendah maupun di daerah dengan Indeks Pembangunan Manusia dengan kategori tinggi

melalui pendekatan yang berbasis pada kesamaan kondisi wilayah sosial ekonomi masing-masing wilayah. Analisis dalam penelitian ini memanfaatkan data yang mencakup Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, Pengeluaran Riil per Kapita yang Disesuaikan, Tingkat Pengangguran Terbuka, dan Upah Minimum. Penelitian ini menentukan analisis *cluster* hierarki terbaik menggunakan korelasi *Cophenetic* lalu menentukan jumlah *cluster* optimum menggunakan *package NbClust* pada *software* RStudio. Hasil dari penelitian tersebut yaitu diperoleh metode terbaik yang digunakan ialah metode *Average Linkage* yang terbagi menjadi 5 *cluster* berdasarkan karakteristik wilayah pembangunan manusia.

Kata kunci: *cluster*, Indeks Pembangunan Manusia, Korelasi *Cophenetic*, *Average Linkage*

PENDAHULUAN

Pembangunan kualitas hidup masyarakat di Indonesia diukur menggunakan Indeks Pembangunan Manusia (IPM) (Mayapada et al., 2022). Pada tahun 2023, populasi penduduk Indonesia berjumlah 278,7 juta jiwa dengan IPM sebesar 73,55 poin yang termasuk dalam kategori tinggi (Badan Pusat Statistik, 2024). Namun, apabila dibandingkan dengan IPM di negara lain salah satunya negara-negara anggota ASEAN, pada tahun 2022, Indonesia berada pada posisi keenam dengan nilai 71,3 poin di bawah Thailand dan Vietnam (Badan Pusat Statistik, 2024). Meskipun berada pada kategori tinggi, jika ditinjau berdasarkan wilayah kabupaten/kota, maka terlihat bahwa terjadi ketimpangan IPM di wilayah-wilayah tersebut. Hal ini bisa dilihat dari wilayah di Provinsi DKI Jakarta yang enam kabupaten/kotanya berada pada kategori IPM yang sangat tinggi dan tinggi. Capaian indeks tersebut berada pada rentang 73,39 poin oleh Kepulauan Seribu hingga 86,07 poin oleh Kota Jakarta Selatan. Bahkan IPM di Provinsi DKI Jakarta lebih tinggi daripada Indonesia seperti yang terdapat pada Gambar 1.



Gambar 1. Indeks Pembangunan Manusia tahun 2023 berdasarkan Provinsi di Indonesia
Sumber: Badan Pusat Statistik (diolah), 2024

Angka tersebut sangat timpang dibandingkan dengan IPM yang kategorinya rendah. Kategori tersebut didominasi oleh wilayah di Provinsi Papua seperti Kabupaten Nduga dengan indeks sebesar 35,19 poin, Kabupaten Puncak sebesar 44,59 poin, dan Kabupaten Pegunungan Bintang sebesar 48,18 poin. Selain itu, terdapat pula Kabupaten Sabu Raijua yang merupakan bagian dari wilayah Provinsi Nusa Tenggara Timur dengan indeks sebesar 58,89 poin (Badan Pusat Statistik, 2024). Uraian tersebut mengindikasikan adanya ketimpangan IPM di Indonesia yang terjadi pada level kabupaten/kota.

Perbedaan nilai IPM di wilayah kabupaten/kota di Indonesia dapat disebabkan oleh ketimpangan ketiga dimensi yang mencerminkan aspek-aspek terkait dengan kesehatan, pendidikan, dan pertumbuhan ekonomi berkelanjutan (Sukmawati, 2022). IPM berfungsi sebagai indikator kinerja pembangunan berkelanjutan yang mencerminkan seberapa besar akses masyarakat terhadap hasil pembangunan dari segi kesehatan, edukasi, dan pendapatan (Badan Pusat Statistik Kabupaten Aceh Barat Daya, 2022). Kesehatan dan edukasi merupakan elemen fundamental dalam meningkatkan kualitas hidup, dengan kesehatan yang memanfaatkan indikator Angka Harapan Hidup (AHH) dan edukasi yang memanfaatkan indikator Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). Indikator-indikator ini memiliki peran krusial dalam membentuk kemampuan manusia dalam skala yang lebih luas (Matondang, 2018). Pengeluaran Riil per Kapita yang Disesuaikan juga mencerminkan tingkat penghasilan dan kesejahteraan masyarakat (Badan Pusat Statistik, 2024). Di sisi lain, terdapat isu besar yang terjadi di negara berkembang termasuk Indonesia, yaitu tingginya angka pengangguran yang disebabkan oleh pertumbuhan penduduk yang pesat tetapi tidak diiringi dengan peningkatan kualitas sumber daya manusia, sehingga menciptakan masalah ketenagakerjaan yang serius (Mahroji & Nurkhasanah, 2019). Masalah ketenagakerjaan lainnya yang perlu diperhatikan adalah upah minimum (Mukaromah et al., 2023). Peningkatan Upah Minimum Kabupaten/Kota (UMK) berkontribusi pada peningkatan kualitas hidup yang layak dan daya beli masyarakat yang pada akhirnya berdampak positif terhadap IPM (Zamharir, 2016).

Penelitian sebelumnya menunjukkan bahwa Tingkat Pengangguran Terbuka (TPT) dan Upah Minimum Regional (UMR) memiliki pengaruh yang signifikan terhadap IPM. Penelitian oleh Sania et al. (2021) menemukan bahwa tingginya angka TPT berkontribusi pada berkurangnya pendapatan masyarakat, yang berdampak pada kemampuan mereka dalam memenuhi kebutuhan dasar seperti pendidikan dan kesehatan, serta mengurangi kualitas hidup secara keseluruhan. Selain itu, Kadri et al. (2020) dalam penelitiannya menemukan bahwa Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan pengeluaran per kapita memiliki pengaruh signifikan terhadap IPM, sehingga kebijakan yang mendukung peningkatan indikator-indikator tersebut sangat diperlukan. Berdasarkan uraian di atas, penting untuk melakukan analisis yang lebih mendalam guna memahami pola-pola yang ada di wilayah kabupaten/kota di Indonesia yang mengalami ketimpangan Indeks Pembangunan Manusia (IPM).

Salah satu analisis pada Statistika yaitu analisis *cluster* dapat memberikan wawasan tentang kelompok-kelompok daerah yang memiliki karakteristik serupa dalam hal pembangunan manusia pada kabupaten/kota di Indonesia. Dengan melakukan analisis ini, kita dapat mengidentifikasi wilayah-wilayah yang membutuhkan perhatian dalam hal kebijakan dan intervensi pembangunan. Selain itu, analisis *cluster* dapat membantu merancang strategi yang lebih tepat sasaran untuk meningkatkan kualitas hidup masyarakat, baik di daerah dengan IPM yang rendah maupun tinggi, melalui pendekatan yang berbasis pada kesamaan kondisi wilayah sosial ekonomi masing-masing wilayah. Dengan demikian, analisis *cluster* tidak hanya berfungsi sebagai alat untuk memetakan ketimpangan tetapi juga sebagai dasar dalam merumuskan kebijakan pembangunan yang lebih efektif dan efisien di tingkat lokal.

METODOLOGI

Data yang digunakan pada penelitian ini berupa data sekunder yang sebagian besarnya bersumber dari publikasi resmi Badan Pusat Statistik (BPS) berjudul “Indeks Pembangunan Manusia 2023”. Variabel yang dipakai berupa Umur Harapan Hidup (X_1), Harapan Lama Sekolah (X_2), Rata-rata Lama Sekolah (X_3), Pengeluaran Riil per Kapita yang Disesuaikan (X_4), Tingkat Pengangguran Terbuka (X_5), dan Upah Minimum (X_6). Penelitian ini menggunakan analisis *cluster* hierarki yang terdiri dari beberapa metode, diantaranya yaitu metode *Single Linkage*, *Complete Linkage*, *Average Linkage*,

Centroid Linkage, dan *Ward's* (Asiska et al., 2019). Kelima metode tersebut akan dibandingkan menggunakan korelasi *Cophenetic* sehingga diperoleh metode *cluster* hierarki yang terbaik. Langkah selanjutnya yaitu menentukan jumlah *cluster* optimum dengan tujuan memiliki kualitas *cluster* yang semakin baik. Pada penelitian ini, jumlah *cluster* optimum akan ditentukan dengan menggunakan *package NbClust* pada *software* RStudio hingga mendapatkan kesimpulan banyaknya *cluster* yang diusulkan dan anggota-anggota dari *cluster* berdasarkan *output* analisis tersebut.

Cluster Hierarki

Pada metode hierarki, pengelompokan dimulai dari tahap mengelompokkan beberapa objek yang memiliki kriteria berupa kemiripan objek yang paling dekat dan berlanjut ke objek yang lain dengan kemiripan kedua. Tahap ini dilanjutkan hingga terbentuk *cluster* seperti “pohon” dengan terdapat tingkatan atau hierarki yang jelas antar objek dimulai dari objek yang paling mirip hingga paling tidak mirip. Umumnya metode pengelompokan hierarki ini ditampilkan dengan *dendogram* yang memiliki kemiripan dengan “struktur diagram pohon” (Iis et al., 2022). *Dendogram* merupakan representatif visual dari tahap-tahap terbentuknya *cluster* dan nilai koefisien jarak setiap prosesnya (Ramadhani et al., 2018).

Terdapat dua macam metode hierarki, yaitu metode agglomeratif dan metode divisif. Pada metode agglomeratif, setiap objek pengamatan dianggap *cluster* sehingga diperoleh *cluster* sebanyak jumlah objek. Lalu, dua *cluster* yang memiliki kemiripan akan digabungkan menjadi *cluster* yang baru dan dilanjutkan hingga seterusnya. Sebaliknya, metode divisif dimulai dengan satu *cluster* dalam jumlah besar yang berisi seluruh observasi. Kemudian observasi-observasi dengan kemiripan yang paling kecil dipisah dan dibentuk menjadi *cluster* yang berukuran lebih kecil (Hanada & Yanti, 2021).

Metode Single Linkage

Dalam menggunakan metode *Single Linkage*, *cluster* akan dibentuk berdasarkan objek individu dengan melakukan penggabungan jarak terdekat. Tahap pertama yang dilakukan pada metode ini adalah menentukan objek-objek yang mempunyai jarak terkecil yang terdapat pada matriks *proximity* $D = \{d_{ik}\}$. Objek-objek dengan jarak paling minimum dikelompokkan dalam *cluster* yang sama, seperti U dan V untuk dikelompokkan dan dengan demikian diperoleh *cluster* (UV). Berikutnya, guna menentukan jarak antara *cluster* (UV) dan *cluster* W atau pada *cluster* lainnya, maka digunakan Persamaan (1) berikut (Muflihan et al., 2022).

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1)$$

dengan:

d_{UW} : jarak minimum dari *cluster* U dan W

d_{VW} : jarak minimum dari *cluster* V dan W

Metode Complete Linkage

Metode *Complete Linkage* merupakan metode yang berkebalikan dengan *Single Linkage*. Metode ini melakukan pengukuran terhadap antar *cluster* yang mengacu pada objek dengan jarak terjauh. Objek-objek pada tiap *cluster* disambungkan antara satu dan yang lainnya dengan kesamaan jarak minimum maupun maksimum. Tahap awal yang dilakukan adalah menghitung nilai minimum dari $D = d_{ij}$ dan menggabungkannya dengan objek lain yang mempunyai kemiripan seperti U dan V guna memperoleh *cluster* (UV). Kemudian, untuk mengukur jarak antara *cluster* (UV) dan *cluster* W , maka digunakan Persamaan (2) (Wulandari, 2024).

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (2)$$

dengan:

d_{UW} : jarak maksimum dari *cluster U* dan *W*

d_{VW} : jarak maksimum dari *cluster V* dan *W*

Metode Average Linkage

Metode *Average Linkage* merupakan metode yang pengelompokannya berawal dari sepasang objek dengan jarak terdekat dan jarak rata-rata. Pengelompokan metode *Average Linkage* dilakukan berdasarkan dua objek dengan kesamaan terdekat pada matriks similaritas. Dimisalkan objek terdekat yaitu objek *X* dan objek *Y*, maka kedua objek tersebut bisa digabung sehingga terbentuk menjadi satu *cluster*. Tahap tersebut dilanjutkan hingga $N - 1$ kali dengan N berupa banyak objek. Persamaan (3) yang digunakan dalam metode ini ialah sebagai berikut (Amelia, 2022).

$$d_{(xy)} = \frac{\sum_a \sum_b d_{ab}}{N_{XY} N_Z} \quad (3)$$

dengan:

d_{ab} : jarak antara objek *x* pada *cluster XY* dan objek *b* pada *cluster Z*

N_{XY} : banyak *item* dalam *cluster XY*

N_Z : banyak *item* dalam *cluster Z*

Metode Centroid Linkage

Metode *Centroid Linkage* mendefinisikan jarak antara *cluster* sebagai jarak antara pusat/*centroid*-nya. Setelah menghitung *centroid* untuk setiap *cluster*, jarak antar *centroid* tersebut dihitung dengan menggunakan fungsi jarak. Metode ini dibentuk dengan memusatkan pada minimalisasi nilai dari standar deviasi *cluster* terkecil. Metode *Centroid Linkage* ini mengelompokkan dua *cluster* dengan jarak paling minimum antar *cluster* diantara titik pusatnya. Misalkan *cluster U* dan *cluster V* yang memiliki jarak paling minimum, maka terbentuklah satu *cluster*. Persamaan (4) yang digunakan untuk menghitung *centroid* dari *U* dan *V* ialah sebagai berikut (Raja et al., 2024).

$$X_{(UV)} = \frac{(N_U \times \bar{x}_U) + (N_V \times \bar{x}_V)}{N_U + N_V} \quad (4)$$

dengan:

N_U, N_V : banyak objek pada *cluster* ke *U* dan *cluster* ke *V*

\bar{x}_U, \bar{x}_V : rata-rata objek pada *cluster* ke *U* dan *cluster* ke *V*

Metode Ward's

Metode *Ward's* ialah metode yang melakukan pembentukan *cluster* dengan meminimalisir hilangnya informasi ketika melakukan pembentukan kelompok dari objek-objek. Pengukuran tersebut berdasarkan banyaknya jumlah kuadrat deviasi yang berada pada rata-rata kelompok di setiap pengamatan. Jarak pada dua *cluster* berdasarkan nilai *Sum of Square Error* (SSE) sehingga menyebabkan kecilnya ragam dalam *cluster* pada kelompok yang dibentuk. Persamaan (5) yang digunakan pada metode ini ialah sebagai berikut (Sri et al., 2024).

$$SSE = \sum_{j=1}^p \left(X_{ij}^2 - \frac{1}{n} \left(\sum_{i=1}^n X_{ij} \right)^2 \right) \quad (5)$$

dengan:

X_{ij} : nilai variabel ke-*j* pada objek ke-*i*

p : banyak variabel yang diukur

n : banyak objek dalam *cluster* yang terbentuk

Uji Multikolinearitas

Asumsi pada analisis *cluster* yang harus terpenuhi yaitu tidak terjadi multikolinearitas atau tidak terdapat korelasi antar variabel. Guna mengetahui bahwa asumsi tersebut telah terpenuhi atau tidak, maka digunakan persamaan korelasi *Pearson* untuk memperoleh koefisien korelasi *Pearson* antar variabel bebas. Apabila hasil uji tersebut mencapai atau melebihi 0,8, maka terjadi multikolinearitas (Raja et al., 2024). Adapun Persamaan (6) korelasi *Pearson* sebagai berikut (Ningsih et al., 2016).

$$r = \frac{\sum_{j=1}^p [(x_j - \bar{x})(y_j - \bar{y})]}{\sqrt{\sum_{j=1}^p (x_j - \bar{x})^2 (y_j - \bar{y})^2}} \quad (6)$$

dengan:

- r : koefisien korelasi objek x_j dengan objek y_j
- \bar{x} : rata-rata dari objek x
- \bar{y} : rata-rata dari objek y
- x_j : data ke- j dari objek x
- y_j : data ke- j dari objek y
- p : jumlah variabel penelitian

Standardisasi

Apabila teridentifikasi bahwa adanya perbedaan terhadap ukuran satuan yang besar antar variabel, maka dilakukan standardisasi data. Standardisasi data tersebut berupa transformasi ke bentuk *Z-Score* menggunakan Persamaan (7) sebagai berikut (Puspita, 2021).

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (7)$$

dengan:

- Z_i : Nilai *Z-Score* ke- i
- x_i : Nilai data ke- i
- \bar{x} : Nilai rata-rata
- S : Nilai standar deviasi
- i : 1,2,3, ..., n

Jarak Euclidean

Jarak *Euclidean* didefinisikan sebagai besarnya jarak yang terdapat pada garis lurus yang menjadi penghubung antar objek. Jarak *Euclidean* ini yang kerap kali digunakan sebagai metode pengukuran jarak dari objek data menuju pusat *cluster* (Dani et al., 2020). Contoh ukuran jarak antara objek ke- i dan objek ke- j dapat dinyatakan sebagai d_{ij} pada variabel ke- k sebanyak $k = 1, \dots, p$. Sehingga setiap objek dihitung lewat perhitungan jarak *Euclidean* berdasarkan rumus berikut (Ls et al., 2021).

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (8)$$

dengan:

- d_{ij} : jarak antara objek ke- i dan objek ke- j
- p : banyak variabel *cluster*
- X_{ik} : data dari subjek ke- i pada variabel ke- k
- X_{jk} : data dari subjek ke- j pada variabel ke- k
- k : 1, 2, ..., p

Korelasi *Cophenetic*

Koefisien korelasi *Cophenetic* didefinisikan sebagai ukuran yang digunakan sebagai acuan untuk mengetahui tingkat ketelitian pada metode *cluster* yang digunakan. Koefisien korelasi *Cophenetic* ialah koefisien korelasi yang terdapat antara elemen matriks jarak Euclidean dan elemen yang dibentuk oleh matriks *Cophenetic* (Widodo et al., 2018). Koefisien korelasi *Cophenetic* yang menunjukkan nilai yang lebih tinggi diantara metode-metode *cluster* hierarki tersebut dimaknai sebagai solusi yang dihasilkan semakin baik (Hanada & Yanti, 2021). Adapun Persamaan (9) korelasi *Cophenetic* sebagai berikut (Iis et al., 2022).

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{cik} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{cik} - \bar{d}_c)^2]}} \quad (9)$$

dengan:

r_{coph} : koefisien korelasi *Cophenetic*

d_{ik} : jarak *Euclidean* objek ke- i dan ke- k

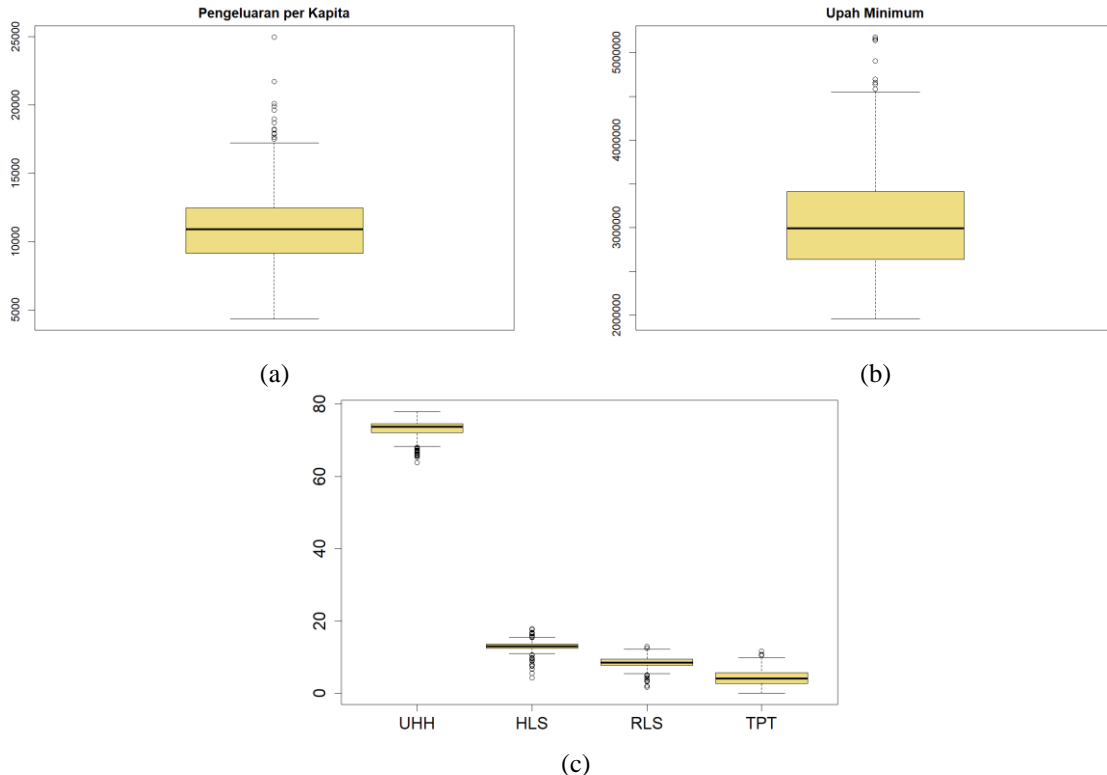
\bar{d} : nilai rata-rata d_{ik}

d_{cik} : jarak dari *Cophenetic* objek ke- i dan ke- k

\bar{d}_c : nilai rata-rata d_{cik}

HASIL DAN PEMBAHASAN

Berdasarkan data kabupaten/kota di Indonesia pada tahun 2023, dilakukan analisis deskriptif terhadap karakteristik terlebih dahulu pada variabel terkait sebelum dilakukan analisis *cluster*. Berikut disajikan analisis deskriptif melalui Gambar 2.



Gambar 2. Analisis Deskriptif: (a) Variabel Pengeluaran Riil per Kapita yang Disesuaikan, (b) Variabel Upah Minimum, (c) Variabel Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Tingkat Pengangguran Terbuka (TPT)

Sumber: Badan Pusat Statistik (diolah), 2024

Berdasarkan *boxplot* yang terdapat pada Gambar 2, diperoleh nilai tertinggi, terendah, dan *outlier* pada data tersebut. Terlihat bahwa keenam variabel di atas mengandung *outlier* yang menggambarkan adanya penyebaran data yang cukup luas. Hal ini mengindikasikan ketidakmerataannya enam variabel tersebut pada kabupaten/kota di Indonesia. Diantaranya yaitu Kabupaten Nduga merupakan daerah dengan nilai Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), Pengeluaran Riil per Kapita yang Disesuaikan, dan Tingkat Pengangguran Terbuka terendah, sedangkan wilayah dengan Upah Minimum terendah adalah Kabupaten Banjarnegara. Kemudian, wilayah dengan nilai tertinggi pada variabel Umur Harapan Hidup ialah Kota Salatiga, variabel Harapan Lama Sekolah dan Rata-rata Lama Sekolah ialah Kota Banda Aceh, variabel Pengeluaran Riil per Kapita yang Disesuaikan ialah Kota Jakarta Selatan, variabel Tingkat Pengangguran Terbuka ialah Kota Ambon, dan variabel Upah Minimum ialah Kabupaten Karawang.

Tabel 1. Variabel dan satuannya

Variabel	Satuan
(1)	(2)
Umur Harapan Hidup (X_1)	Tahun
Harapan Lama Sekolah (X_2)	Tahun
Rata-rata Lama Sekolah (X_3)	Tahun
Pengeluaran Riil per Kapita yang Disesuaikan (X_4)	Rupiah
Tingkat Pengangguran Terbuka (X_5)	Persen
Upah Minimum (X_6)	Rupiah

Sumber: Badan Pusat Statistik, (2024)

Berdasarkan Tabel 1, terdapat variasi pada satuan dan skala data yang dimanfaatkan dalam penelitian ini. Oleh karena itu, diperlukan standardisasi data menggunakan *Z-Score* untuk mengonversikan variabel terhadap nilai standar.

Uji Multikolinearitas

Multikolinearitas merupakan kondisi korelasi yang tinggi antara dua atau lebih variabel dalam suatu model regresi. Kondisi ini muncul apabila variabel yang dilibatkan dalam penelitian saling memengaruhi (Laksmi et al., 2023).

Tabel 2. Uji Multikolinearitas

	X_1	X_2	X_3	X_4	X_5	X_6
	(1)	(2)	(3)	(4)	(5)	(6)
X_1	1	0,43973	0,48120	0,67701	0,41454	-0,11055
X_2	0,43973	1	0,77759	0,51642	0,40180	-0,12197
X_3	0,48120	0,77759	1	0,66719	0,48859	0,16315
X_4	0,67701	0,51642	0,66719	1	0,47168	0,21800
X_5	0,41454	0,40180	0,48859	0,47168	1	0,22098
X_6	-0,11055	-0,12197	0,16315	0,21800	0,22098	1

Sumber: Badan Pusat Statistik (diolah), 2024

Berdasarkan Tabel 2, diketahui nilai korelasi antar variabel dengan nilai koefisien korelasi yang telah dihasilkan tidak melebihi 0,8 (Raja et al., 2024). Maknanya, pada variabel-variabel penelitian tersebut tidak terjadi multikolinearitas sehingga tidak akan dilakukan langkah penanganan terhadap masalah tersebut.

Metode Cluster Terbaik dengan Koefisien Korelasi *Cophenetic*

Berdasarkan nilai korelasi *Cophenetic* yang telah diperoleh maka ditentukan metode *cluster* terbaik. Pada informasi yang terdapat dalam Tabel 3, bisa dilihat bahwa nilai korelasi *Cophenetic* pada

metode *Average Linkage* merupakan tertinggi dibandingkan metode lainnya dengan nilai sebesar 0,7611228 sehingga metode *cluster* terbaik yang digunakan untuk mengelompokkan data ialah metode *Average Linkage*.

Tabel 3. Koefisien Korelasi *Cophenetic*

Metode	Korelasi <i>Cophenetic</i>
(1)	(2)
<i>Single Linkage</i>	0,7411818
<i>Average Linkage</i>	0,7611228
<i>Complete Linkage</i>	0,7449612
<i>Centroid Linkage</i>	0,7552831
<i>Ward's</i>	0,5263969

Sumber: Badan Pusat Statistik (diolah), 2024

Penentuan *Cluster* Optimum

Pada tahap ini, ditentukan *cluster* optimum dengan menggunakan fungsi *NbClust* pada *software* RStudio untuk memperoleh *k cluster* optimum. Diantara 30 indeks yang terdapat dalam paket tersebut, diperoleh 10 indeks yang mengusulkan 5 sebagai jumlah *cluster* terbaik sehingga *cluster* optimum dari *output* analisis tersebut ialah saat $k = 5$. Oleh karena itu, *cluster* akan dibentuk menjadi lima kelompok dengan menggunakan metode *Average Linkage*. Berikut disajikan *output* dari 10 indeks tersebut dengan nilai indeksnya masing-masing.

Tabel 4. *Cluster* Optimum

Indeks	Value Index
(1)	(2)
CH	125,8236
Friedman	3,9904
Hartigan	167,9403
KL	7,0489
Marriot	2,194273e+15
Rubin	-0,4703
Ratkowsky	0,3137
Scott	471,9834
TrCovW	18717,57
TraceW	510,5249

Sumber: Badan Pusat Statistik (diolah), 2024

Pembentukan *Cluster* dan Interpretasi

Tabel 5 memperlihatkan *output* dari analisis *cluster* yang telah dilakukan dengan metode *Average Linkage* yang menunjukkan jumlah anggota pada lima *cluster* dari 514 kabupaten/kota yang menjadi observasi. Banyaknya anggota dari setiap klaster ditunjukkan pada Tabel 5 di bawah ini.

Tabel 5. Pembentukan *Cluster*

Cluster	Jumlah Anggota
(1)	(2)
1	385
2	80
3	27
4	20
5	2

Sumber: Badan Pusat Statistik (diolah), 2024

Setelah menentukan banyak *cluster* dan anggota yang terbentuk, langkah yang dilakukan selanjutnya ialah melakukan interpretasi terhadap *cluster*. Hal tersebut dilakukan untuk mengetahui

karakteristik pada setiap *cluster* yang terbentuk dan memberi ciri yang spesifik terhadap setiap variabel masing-masing *cluster* sehingga menggambarkan isi *cluster*. Pada Tabel 6 akan ditunjukkan nilai rata-rata dari tiap variabel dalam lima *cluster*.

Tabel 6. Rata-rata Variabel Berdasarkan *Cluster*

<i>Cluster</i>	X_1	X_2	X_3	X_4	X_5	X_6
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	72,91	13,00	8,30	10.334.000	3,89	2.887.446
2	74,70	14,68	10,95	14.192.000	6,32	3.039.216
3	75,20	13,40	10,08	15.512.000	7,30	4.634.821
4	67,21	10,29	4,99	5.929.000	1,63	3.661.122
5	65,33	4,96	1,95	5.164.000	0,15	3.864.696

Sumber: Badan Pusat Statistik (diolah), 2024

Seperti yang ditunjukkan pada Tabel 6, hasil dari pengelompokan dideskripsikan berdasarkan nilai rata-rata masing-masing *cluster* sehingga diperoleh karakteristik dari kelima *cluster*. Adapun interpretasinya sebagai berikut.

1. Rata-rata paling tinggi pada variabel Umur Harapan Hidup (UHH) mencapai 75,20 yang terdapat pada *cluster* 3 dan rata-rata paling rendah yaitu 65,33 yang terdapat pada *cluster* 5.
2. Rata-rata paling tinggi pada variabel Harapan Lama Sekolah (HLS) mencapai 14,68 yang terdapat pada *cluster* 2 dan rata-rata paling rendah yaitu 4,96 yang terdapat pada *cluster* 5.
3. Rata-rata paling tinggi pada variabel Rata-rata Lama Sekolah (RLS) mencapai 10,95 yang terdapat pada *cluster* 2 dan rata-rata paling rendah yaitu 1,95 yang terdapat pada *cluster* 5.
4. Rata-rata paling tinggi pada variabel Pengeluaran per Kapita yang Disesuaikan mencapai 15.512.000 rupiah yang terdapat pada *cluster* 3 dan rata-rata paling rendah yaitu 5.164.000 rupiah yang terdapat pada *cluster* 5.
5. Rata-rata paling tinggi pada variabel Tingkat Pengangguran Terbuka (TPT) mencapai 7,30 yang terdapat pada *cluster* 3 dan rata-rata paling rendah yaitu 0,15 yang terdapat pada *cluster* 5.
6. Rata-rata paling tinggi pada variabel upah minimum mencapai 4.634.821 rupiah yang terdapat pada *cluster* 3 dan rata-rata paling rendah yaitu 2.887.446 rupiah yang terdapat pada *cluster* 1.

Merujuk pada hasil penelitian ini, maka diperoleh temuan bahwa pengelompokan pada level kabupaten/kota terbagi dalam lima *cluster*. Terdapat beberapa karakteristik wilayah pembangunan pada setiap *cluster*, diantaranya adalah karakteristik paling baik, baik, cukup, buruk, dan paling buruk. Karakteristik tersebut dikategorikan berdasarkan capaian variabel-variabel setiap *cluster*. *Cluster* dengan karakteristik wilayah pembangunan manusia paling baik memiliki capaian Umur Harapan Hidup, Pengeluaran Riil per Kapita yang Disesuaikan, dan Upah Minimum yang paling tinggi diantaranya *cluster* lainnya. *Cluster* dengan karakteristik wilayah pembangunan yang baik memiliki capaian Harapan Lama Sekolah, dan Rata-rata Lama Sekolah yang paling tinggi diantara *cluster* lainnya. *Cluster* dengan karakteristik wilayah pembangunan manusia yang cukup memiliki capaian Umur Harapan Hidup, Rata-rata Lama Sekolah, Pengeluaran Riil per Kapita yang Disesuaikan, dan Tingkat Pengangguran Terbuka yang berada di atas *cluster* dengan karakteristik wilayah pembangunan buruk dan *cluster* dengan karakteristik wilayah pembangunan paling buruk. Kemudian, *cluster* dengan karakteristik wilayah pembangunan manusia yang buruk memiliki capaian Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran Riil per Kapita yang Disesuaikan yang tidak lebih baik dari capaian pada *cluster* dengan karakteristik cukup, tetapi memiliki angka Tingkat Pengangguran Terbuka yang lebih baik dibandingkan *cluster* 1, *cluster* 2, dan *cluster* 3. Selanjutnya, *cluster* dengan karakteristik paling buruk memiliki capaian Umur Harapan Hidup, Harapan

Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran Riil per Kapita yang Disesuaikan yang paling rendah diantara *cluster* lainnya. Namun, pada *cluster* tersebut memiliki capaian Tingkat Pengangguran Terbuka yang paling baik dibandingkan *cluster* yang lain.

Cluster 1 didefinisikan sebagai *cluster* wilayah pembangunan dengan karakteristik yang cukup. *Cluster* ini merupakan *cluster* paling dominan dibandingkan dengan yang lainnya karena sebagian besar wilayah kabupaten/kota di Indonesia berada pada *cluster* tersebut. Kemudian, *cluster* 2 merupakan *cluster* wilayah pembangunan dengan karakteristik yang baik. Pada *cluster* ini didominasi oleh wilayah perkotaan seperti Kota Yogyakarta dari Provinsi D.I. Yogyakarta, Kota Pontianak dari Provinsi Kalimantan Barat, Kota Makassar dari Provinsi Sulawesi Selatan, dan Kota Jayapura dari Provinsi Papua. *Cluster* 3 merupakan *cluster* wilayah pembangunan dengan karakteristik yang paling baik. Pada *cluster* ini, semua wilayah kota di Provinsi DKI Jakarta tergabung. Hal demikian membuktikan bahwa pembangunan manusia di wilayah tersebut sangat baik dan merata. Selain itu, Kota Batam dan Kabupaten Bintan dari Provinsi Kepulauan Riau, dan beberapa kabupaten/kota dari Provinsi Banten yang lebih mendominasi di *cluster* 3 daripada di *cluster* lainnya. *Cluster* 4 merupakan *cluster* wilayah pembangunan dengan karakteristik yang buruk, dan *cluster* 5 merupakan *cluster* wilayah pembangunan dengan karakteristik yang paling buruk. Pada *cluster* 4 dan *cluster* 5 didominasi oleh kabupaten/kota di Provinsi Papua dan Papua Barat. Selain itu, terdapat pula satu kabupaten dari Provinsi Nusa Tenggara Timur yang berada pada *cluster* 4. Ketiga provinsi tersebut merupakan wilayah yang tidak satu pun kabupaten/kotanya berada pada *cluster* dengan karakteristik wilayah pembangunan paling baik. Hanya sebagian kecil pula wilayahnya yang berada pada *cluster* 2 dengan karakteristik wilayah pembangunan yang baik.

KESIMPULAN DAN SARAN

Berdasarkan hasil dari analisis yang telah dilakukan pada pengelompokan kabupaten/kota di Indonesia menggunakan metode *Average Linkage*, diperoleh kesimpulan bahwa dari 514 kabupaten/kota, terbentuk 5 *cluster* berdasarkan karakteristik wilayah pembangunan manusia pada tahun 2023. *Cluster* yang paling dominan ialah *cluster* 1 sebanyak 385 kabupaten/kota dengan karakteristik wilayah pembangunan manusia yang cukup. *Cluster* dengan karakteristik wilayah pembangunan manusia paling baik berada pada *cluster* 3 sebanyak 27 kabupaten/kota, sedangkan *cluster* dengan wilayah pembangunan manusia paling buruk berada pada *cluster* 5 oleh Kabupaten Nduga dan Kabupaten Puncak dari Provinsi Papua. Analisis *cluster* tersebut memperlihatkan terjadi ketimpangan antar wilayah pembangunan sehingga wilayah yang berada pada *cluster* 4 dan utamanya pada *cluster* 5 harus mendapatkan fokus perhatian khusus dari pemerintah untuk meningkatkan dan mengembangkan IPM. Hal tersebut dilakukan untuk mengatasi kesenjangan antar wilayah provinsi menurut karakteristik masing-masing daerah.

Saran pada penelitian ini adalah pemerintah dapat melihat hasil *cluster* dan variabel-variabel yang menjadi acuan perkembangan pembangunan sumber daya masyarakat, khususnya pada *cluster* 4 dan *cluster* 5 dengan persentase Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran Riil per Kapita yang Disesuaikan yang berada pada nilai yang rendah. Sehingga pemerintah dapat menjadikan *cluster* 5 sebagai prioritas untuk meningkatkan angka pembangunan manusia dan menjadikan *cluster* 3 sebagai tolak ukur pembangunan Indeks Pembangunan Manusia.

DAFTAR PUSTAKA

- Amelia, S. (2022). Klasterisasi Kabupaten/Kota Berdasarkan Indikator Indeks Pembangunan Manusia dengan Metode Average Linkage (Studi Kasus: Provinsi Banten). *MathVision : Jurnal Matematika*, 4(2). <https://doi.org/10.55719/mv.v4i2.416>
- Asiska, N., Satyahadewi, N., & Perdana, H. (2019). Pencarian Cluster Optimum pada Single Linkage, Complete Linkage dan Average Linkage. *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, 8(3). <https://doi.org/10.26418/bbimst.v8i3.33173>
- Badan Pusat Statistik. (2024). *Indeks Pembangunan Manusia 2023* (Vol. 18). Badan Pusat Statistik.
- Badan Pusat Statistik. (2024). *Statistik Indonesia 2024* (Vol. 52). Badan Pusat Statistik.
- Badan Pusat Statistik Kabupaten Aceh Barat Daya. (2022). *Indeks Pembangunan Manusia Kabupaten Aceh Barat Daya*. Badan Pusat Statistik Kabupaten Aceh Barat Daya.
- Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2020). Pengelompokan Data Runtun Waktu Menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016). *Jurnal EKSPONENSIAL*, 11(1).
- Hanada, S., & Yanti, T. S. (2021). Penggunaan Analisis Cluster dalam Pengelompokan Kecamatan di Kabupaten Karawang Berdasarkan Metode Kontrasepsi Peserta KB Aktif. *Prosiding Statistika*.
- Iis, I., Yahya, I., A, G. N., Wibawa, W., Baharuddin, B., Ruslan, R., & Laome, L. (2022). Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020. *Prosiding Seminar Nasional Sains dan Terapan (Sinta)*, VI(April).
- Kadri, I. A., Susilawati, M., & Sari, K. (2020). Faktor-Faktor yang Berpengaruh Signifikan Terhadap Indeks Pembangunan Manusia di Provinsi Papua. *E-Jurnal Matematika*, 9(1). <https://doi.org/10.24843/mtk.2020.v09.i01.p275>
- Laksmi, A. D., Kusnandar, D., & Imro'ah, N. (2023). Hierarchical Clustering Multiscale Bootstrap pada Capaian Dimensi IPM Tahun 2020 yang Terdampak COVID-19. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 12(2).
- Ls, D., Lesnussa, Y. A., Talakua, M. W., & Matdoan, M. Y. (2021). Analisis Kluster untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward. *Jurnal Statistika dan Aplikasinya*, 5(1). <https://doi.org/10.21009/jsa.05105>
- Mahroji, D., & Nurkhasanah, I. (2019). Pengaruh Indeks Pembangunan Manusia Terhadap Tingkat Pengangguran di Provinsi Banten. *Jurnal Ekonomi-Qu*, 9(1). <https://doi.org/10.35448/jequ.v9i1.5436>
- Matondang, Z. (2018). Pengaruh Pertumbuhan Ekonomi, Realisasi Pengeluaran Pendidikan, dan Realisasi Pengeluaran Kesehatan Terhadap Indeks Pembangunan Manusia dalam Pandangan Islam di Provinsi Sumatera Utara. *Al-Masharif: Jurnal Ilmu Ekonomi dan Keislaman*, 6(1). <https://doi.org/10.24952/masharif.v6i1.1131>
- Mayapada, R., Yanti, R. W., & Syarifuddin, S. (2022). Analisis Tingkat Kepentingan terhadap Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Indonesia. *Journal of Mathematics: Theory and Applications*. <https://doi.org/10.31605/jomta.v4i2.2030>
- Muflihan, Y., Retnawati, H., & Kistian, A. (2022). Analisis cluster dengan Metode Hierarki Untuk Pengelompokan Sekolah Menengah Atas Berdasarkan Raport Mutu Sekolah di Kabupaten Nagan Raya. *Measurement In Educational Research (Meter)*, 2(1). <https://doi.org/10.33292/meter.v2i1.155>
- Mukaromah, L., Hanifatuzzahra, Z., Nasrullah, A., Latifah, T. M., Purwaningsih, V. T., & Suparta, I. W. (2023). Pengaruh Indeks Pembangunan Manusia, Tingkat Upah Minimum, dan Tingkat Pengangguran Terhadap Pertumbuhan Ekonomi Indonesia Tahun 2022. *ANALISIS*, 13(2). <https://doi.org/10.37478/als.v13i2.2874>

- Munthe, A. D. (2019). Penerapan Clustering Time Series Untuk Menggerombolkan Provinsi di Indonesia Berdasarkan Nilai Produksi Padi. *Jurnal Litbang Sukowati: Media Penelitian dan Pengembangan*, 2(2). <https://doi.org/10.32630/sukowati.v2i2.61>
- Ningsih, S., Wahyuningsih, S., & Nasution, Y. N. (2016). Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Hasil Analisis Cluster (Studi Kasus : Produksi Palawija Provinsi Kalimantan Timur 2014 / 2015). *Prosiding Seminar Sains dan Teknologi FMIPA Unmul*, 1(1).
- Puspita, R. N. (2021). Analisis K-Means Cluster pada Kabupaten/Kota di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 2(3). <https://doi.org/10.46306/lb.v2i3.85>
- Raja, N. A., Tinungki, G. M., & Sirajang, N. (2024). Implementasi Algoritma Centroid Linkage dan K-Medoids dalam Mengelompokkan Kabupaten/Kota di Sulawesi Selatan Berdasarkan Indikator Pendidikan. *ESTIMASI: Journal of Statistics and Its Application*. <https://doi.org/10.20956/ejsa.v5i1.13605>
- Ramadhani, L., Purnamasari, I., & Amijaya, F. D. T. (2018). Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016). *Eksponensial*, 9(2016).
- Sania, L., Balafif, M., & Imamah, N. (2021). Pengaruh PDRB, Tingkat Pengangguran Terbuka dan UMR Terhadap Indeks Pembangunan Manusia di Kabupaten dan Kota Provinsi Jawa Timur. *Bharanomics*, 2(1). <https://doi.org/10.46821/bharanomics.v2i1.189>
- Sri, I. M., Dwipurwani, O., & Aisyah, N. (2024). Pengelompokan Provinsi di Indonesia Berdasarkan Konsumsi Kalori Per Kapita Sehari Menurut Kelompok Komoditas / Makanan Menggunakan Average Linkage dan Ward's Method. *Azizah, dan Nurul Aisyah*, 2, 1698–1713. <https://journal.institercom-edu.org/index.php/multiple>
- Sukmawati, A. (2022). Analisis Determinan Indeks Pembangunan Manusia di Indonesia Tahun 2019 dengan Spatial Error Model (SEM). *Seminar Nasional Official Statistics*, 2022(1), 1305–1314. <https://doi.org/10.34123/semnasoffstat.v2022i1.1532>
- Widodo, E., Sari, N. N., Hidayati, I., Yubinas, F., Yuniarti, M., & Novyantika, R. D. (2018). Analisis Cluster Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016. *Program Studi Pendidikan Matematika FKIP UMS, March*.
- Wulandari, S. (2024). Analisis Clustering Provinsi Berdasarkan Prevalensi Stunting Balita Menggunakan Algoritma Single dan Complete Linkage. *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, 8(01). <https://doi.org/10.30998/semnasristek.v8i01.7128>
- Zamharir, A. (2016). Pengaruh Pertumbuhan Ekonomi, Kemiskinan, PDRB Per Kapita, dan Upah Minimum Terhadap Human Development Index: Studi Kasus 12 Provinsi dengan Kategori Lower Medium di Indonesia. *IOSR Journal of Economics and Finance*, 3(1).