

# Pengelompokan Daerah di Jawa Timur Berbasis Indikator Kesejahteraan Masyarakat dengan Pendekatan Analisis *Cluster* Hierarki dan Nonhierarki

Muhammad Fikry Al Farizi<sup>1</sup>, Faradilla Harianto<sup>1</sup>, Maria Setya Dewanti<sup>1</sup>, Cynthia Anggelyn Siburian<sup>1</sup>, M. Fariz Fadillah Mardianto<sup>1\*</sup>, Dita Amelia<sup>1</sup>, dan Elly Ana<sup>1</sup>

Received: 30 December 2022

Revised: 23 August 2023

Accepted: 29 September 2023

<sup>1</sup>Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya, Indonesia

\*Corresponding author: [m.fariz.fadillah.m@fst.unair.ac.id](mailto:m.fariz.fadillah.m@fst.unair.ac.id)

**ABSTRAK** – Berdasarkan data Badan Pusat Statistik (BPS) pada September 2021, Jawa Timur merupakan provinsi dengan penduduk miskin terbanyak di Indonesia dengan jumlah mencapai 26,503 juta jiwa. Kemiskinan menjadi satu diantara faktor yang memengaruhi kesejahteraan masyarakat di Jawa Timur. Oleh karena itu, dilakukan penelitian untuk mengelompokkan kabupaten dan kota di Jawa Timur berdasarkan indikator kesejahteraan masyarakat melalui pendekatan analisis *cluster* menggunakan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward*, metode *centroid*, *median linkage*, dan metode *K-Means*, menentukan *cluster* optimum setiap metode menggunakan *Pseudo – F*, kemudian membandingkan ketujuh metode dan menentukan metode terbaik menggunakan nilai *icdrate*, serta mengidentifikasi karakteristik tiap kelompok *cluster* berdasarkan metode terbaik. Terdapat enam variabel yang akan digunakan pada penelitian ini. Keseluruhan data variabel merupakan data sekunder yang diperoleh dari laman resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur. Penelitian ini menghasilkan empat *cluster* dengan menggunakan metode *average linkage* sebagai metode terbaik. Penelitian ini diharapkan dapat bermanfaat sebagai pertimbangan untuk mengevaluasi pemerintah dan instansi terkait untuk mengatasi permasalahan pokok yang masih terjadi di masing-masing kabupaten dan kota. Dengan demikian, kesejahteraan masyarakat Jawa Timur dapat terwujud dan target SDGs di Indonesia dapat tercapai.

**Kata kunci**– *Clustering*, Kemiskinan, Jawa Timur, *Pseudo – F*, *icdrate*.

**ABSTRACT** – Based on Central Statistics Agency (BPS) data in September 2021, East Java is a province with the largest number of poor people in Indonesia with a total of 26,503 million people. Poverty is one of the factors that affect people's welfare in East Java. Therefore, this research was conducted to classify regencies and cities in East Java based on indicators of community welfare through a hierarchical cluster analysis approach using the *single linkage*, *complete linkage*, *average linkage*, *ward methods*, *centroid methods*, *median linkage*, and *K-Means methods*, determine the optimum cluster for each method using *Pseudo – F*, then compare the seven methods and determine the best method using the rated value, as well as identify the characteristics of each cluster group based on the best method. There are six variables that will be used in this study. All variable data is secondary data obtained from the official website of the Central Statistics Agency (BPS) of East Java Province. This study produced four clusters using the *average linkage* method as the best method. This research is expected to be useful as a consideration for evaluating the government and related agencies to overcome the main problems that still occur in each regency and city. Thus, the welfare of the people of East Java can be realized and the SDGs targets in Indonesia can be achieved.

**Keywords**– *Clustering*, *Poverty*, *East Java*, *Pseudo – F*, *icdrate*.

## I. PENDAHULUAN

Kesejahteraan masyarakat merupakan tujuan dan upaya yang dilakukan pemerintah melalui upaya-upaya seperti peningkatan pertumbuhan ekonomi [1]. Aspek tersebut merupakan satu diantara aspek terpenting dalam *Sustainable Development Goals* (SDGs) yaitu *Decent Work and Economic Growth* yang merupakan poin kedelapan dalam SDGs dan *No Poverty* yang merupakan poin pertama dalam SDGs. Pembangunan sosial ekonomi yang mencerminkan kesejahteraan masyarakat diharapkan dapat terwujud melalui upaya yang dilakukan oleh pemerintah [1]. Kesejahteraan masyarakat dapat ditinjau melalui pertumbuhan ekonomi yang dipengaruhi oleh beberapa kondisi, yaitu jumlah penduduk miskin, tingkat pengangguran terbuka, Indeks Pembangunan Manusia (IPM), rata-rata pengeluaran per kapita untuk makanan sebulan, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak, dan persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak.

Berbagai upaya dilakukan pemerintah untuk meningkatkan pertumbuhan ekonomi yang diharapkan dapat berdampak positif terhadap kesejahteraan masyarakat, satu diantaranya adalah pengentasan kemiskinan. Akan tetapi, upaya tersebut masih sulit diwujudkan oleh Indonesia hingga saat ini. Kemiskinan menjadi suatu permasalahan yang sulit diatasi karena permasalahan tersebut saling berkaitan satu sama lain. Tidak hanya berkaitan dengan masalah rendahnya tingkat pendapatan dan konsumsi, kemiskinan juga berkaitan dengan kesehatan dan ketidakberdayaan masyarakat untuk berpartisipasi dalam pembangunan, rendahnya tingkat pendidikan, serta berbagai masalah yang

berkenaan dengan pembangunan manusia. Oleh karena itu, kemiskinan dapat berdampak pada kondisi kesejahteraan masyarakat secara langsung.

Sampai dengan September 2021, jumlah penduduk miskin di Indonesia mencapai 26,503 juta jiwa dengan tingkat kemiskinan mencapai 9,71% [2], [3]. Dari jumlah tersebut, provinsi Jawa Timur menjadi provinsi dengan jumlah penduduk miskin terbanyak di Indonesia dengan jumlah penduduk miskin mencapai 4.259.600 jiwa [2], [3]. Hal ini sangat bertolak belakang dengan kondisi Jawa Timur yang menjadi provinsi potensial di Indonesia. Satu diantara potensi tersebut berada pada sektor pertanian, dimana Jawa Timur menjadi provinsi dengan produksi padi terbesar dengan produksi sebesar 9,909 juta ton yang menjadikan Jawa Timur sebagai lumbung padi terbesar Indonesia pada Tahun 2021 [2], [3]. Dengan demikian, penelitian ini diharapkan dapat bermanfaat bagi pemerintah dan instansi terkait untuk menganalisis akar permasalahan dan memberikan solusi ataupun sebagai pertimbangan kebijakan yang tepat demi mengentaskan kemiskinan yang masih terjadi di masing-masing kabupaten dan kota serta mewujudkan kesejahteraan masyarakat di Provinsi Jawa Timur.

Sejalan dengan Era Revolusi Industri 4.0, banyak metode statistik yang telah dikembangkan, satu diantaranya adalah analisis *cluster*. Analisis *cluster* merupakan metode statistik yang bertujuan untuk mengelompokkan objek berdasarkan kesamaan karakteristik antara objek-objek tersebut [4]. Analisis *cluster* memiliki dua metode, yaitu metode hierarki dan nonhierarki. Beberapa penelitian sebelumnya tentang analisis *cluster* telah dilakukan, diantaranya adalah penelitian yang dilakukan oleh Niswatin dkk. yang mengelompokkan kabupaten dan kota di Indonesia berdasarkan indikator kemiskinan menggunakan metode *k-means* [5], penelitian oleh Khikmah dan Sofro [6] yang mengelompokkan kabupaten dan kota di Jawa Timur berdasarkan kepadatan penduduk dan kumulatif kasus terkonfirmasi Covid-19, dan penelitian oleh Mardianto dkk. yang membandingkan metode analisis *cluster* hierarki untuk mengklasifikasikan kabupaten di Jawa Timur berdasarkan potensi pangannya [7]. Oleh karena itu, pengelompokan kabupaten dan kota di Jawa Timur berbasis indikator kesejahteraan masyarakat dapat dilakukan melalui analisis *cluster* baik dengan pendekatan hierarki maupun pendekatan nonhierarki. Metode analisis cluster terbaik diperoleh dengan meninjau *icdrate* terkecil dan  $R^2$  terbesar [8].

Pengembangan yang dilakukan dalam penelitian ini adalah penggunaan metode *cluster* hierarki yang meliputi enam metode, yaitu *single linkage*, *complete linkage*, *average linkage*, metode *ward*, metode *centroid*, dan *median linkage* yang merupakan metode analisis *cluster* hierarki dengan prinsip penggabungan (*agglomerative*), serta metode *K-Means* sebagai metode analisis *cluster* nonhierarki. Hal ini dilakukan untuk mengetahui metode *cluster* terbaik dalam pengelompokan kabupaten dan kota di Jawa Timur berbasis indikator kesejahteraan masyarakat. Dengan demikian, kesejahteraan masyarakat Jawa Timur dapat terwujud dan target SDGs di Indonesia dapat tercapai melalui beberapa rekomendasi yang diajukan untuk meningkatkan kesejahteraan masyarakat di Jawa Timur meliputi berbagai aspek kehidupan.

## II. TINJAUAN PUSTAKA

### A. Uji Multikolinieritas

Multikolinieritas adalah adanya hubungan linear sempurna ataupun pasti di antara beberapa ataupun seluruh variabel. Terjadinya multikolinieritas merupakan suatu pelanggaran dalam analisis *cluster* karena dapat mempengaruhi hasil *cluster* [9]. Satu diantara cara mengidentifikasi adanya multikolinieritas adalah menghitung nilai *Variance Inflation Factor* (VIF) dengan formulasi statistik dalam persamaan (1) [10].

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

dengan  $R^2$  adalah nilai koefisien determinasi variabel dependen dengan variabel independen. Multikolinieritas terindikasi apabila nilai  $VIF > 10$ . Jika terjadi multikolinieritas antarvariabel, maka cara yang dapat dilakukan adalah dengan mengeliminasi variabel yang berkorelasi dalam model [10]. Di samping itu, jika terjadi multikolinieritas antarvariabel maka ukuran jarak yang digunakan adalah jarak Mahalanobis karena Jarak Mahalanobis dapat menyesuaikan korelasi antarvariabel [11]. Apabila tidak terdapat multikolinieritas, maka ukuran jarak yang digunakan adalah jarak *Euclidean*.

### B. Analisis Cluster Hierarki

Analisis *cluster* dimanfaatkan untuk mengelompokkan suatu objek observasi menjadi kelompok-kelompok berdasarkan pengukuran karakteristik variabel yang diamati. Hasilnya adalah tiap objek pada kelompok yang sama mempunyai karakteristik yang cenderung sama atau mirip, sedangkan pada antarkelompok akan memiliki karakteristik yang berbeda. Satu di antara metode dalam analisis *cluster*, yaitu *cluster* hierarki. Metode hierarki memiliki hasil pengelompokan yang membentuk suatu tingkatan seperti struktur pohon, karena proses pengelompokan pada metode ini dilakukan secara bertingkat. Hasil dari *cluster* hierarki direpresentasikan sebagai dendrogram, yang menunjukkan bagaimana pengelompokan terbentuk dan nilai koefisien jarak pada tiap langkah [12]. Fungsi jarak yang seringkali digunakan adalah *Euclidean*, didefinisikan sebagai jarak antara vektor  $\mathbf{x}$  dengan vektor  $\mathbf{y}$  [13]. Jarak Euclid antara vektor  $\mathbf{x}$  dengan vektor  $\mathbf{y}$  dirumuskan pada persamaan (2).

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (2)$$

dengan  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_p)'$ ;  $\mathbf{y} = (y_1, y_2, y_3, \dots, y_p)'$ .

Terdapat beberapa metode aglomerasi yang bisa digunakan dalam analisis *cluster* hierarki, yaitu sebagai berikut.

1. *Single Linkage*

*Single linkage* membentuk kelompok-kelompok dari individu dengan menggabungkan jarak terdekat atau terpendek. *Single linkage* didefinisikan sebagai jarak minimum antara suatu titik di *A* dan titik di *B* yang dirumuskan dalam persamaan (3).

$$D(A, B) = \min\{d(\mathbf{y}_i, \mathbf{y}_j), \text{ untuk } \mathbf{y}_i \text{ dalam } A \text{ dan } \mathbf{y}_j \text{ dalam } B\} \tag{3}$$

dengan  $d(\mathbf{y}_i, \mathbf{y}_j)$  adalah jarak *Euclidean* pada persamaan (2), serta *A* dan *B* masing-masing adalah sebuah *cluster*.

2. *Complete Linkage*

*Complete linkage* mengelompokkan seluruh objek dalam *cluster* yang berada paling jauh satu sama lain. *Complete linkage* didefinisikan sebagai jarak maksimum antara suatu titik di *A* dan titik di *B* yang dirumuskan dalam persamaan (4).

$$D(A, B) = \max\{d(\mathbf{y}_i, \mathbf{y}_j), \text{ untuk } \mathbf{y}_i \text{ dalam } A \text{ dan } \mathbf{y}_j \text{ dalam } B\} \tag{4}$$

dengan  $d(\mathbf{y}_i, \mathbf{y}_j)$  adalah jarak *Euclidean* pada persamaan (2), serta *A* dan *B* masing-masing adalah sebuah *cluster*.

3. *Average Linkage*

Prosedur pada *average linkage* hampir sama dengan *single linkage* maupun *complete linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu *cluster* dengan jarak seluruh individu dalam *cluster* yang lain [14]. *Average linkage* dirumuskan pada persamaan (5).

$$D(A, B) = \frac{(\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(\mathbf{y}_i, \mathbf{y}_j))}{n_A n_B} \tag{5}$$

dengan jumlahan seluruh  $\mathbf{y}_i$  di *A* dan seluruh  $\mathbf{y}_j$  di *B*. Untuk setiap langkah penghitungan, kedua *cluster* dengan jarak terkecil yang diukur dalam persamaan (5) digabungkan.

4. Metode *Ward*

Pengelompokan metode *Ward* dilakukan dengan memperkecil total jumlah kuadrat dalam kelompok. Masing-masing kelompok dibentuk sedemikian hingga menghasilkan kelompok dengan jumlah kuadrat terkecil dan dikenal dengan *Error Sums of Squares* (ESS). Dua *cluster* yang memiliki peningkatan ESS paling minimum akan berkelompok. Jika *cluster* sebanyak *c* maka ESS merupakan jumlahan dari  $ESS_c$ .

$$ESS = ESS_1 + ESS_2 + \dots + ESS_c. \tag{6}$$

Saat semua *cluster* bergabung menjadi satu kelompok dari *N* item, maka nilai SSE dirumuskan dalam persamaan (7).

$$ESS = \sum_{j=1}^n (x_j - \bar{x})'(x_j - \bar{x}) \tag{7}$$

dengan  $x_j$  matriks multivariat objek ke - *j* dan  $\bar{x}$  adalah *mean* dari semua objek [13].

5. Metode *Centroid*

Dalam metode *Centroid*, jarak antara dua *cluster* didefinisikan sebagai jarak *Euclidean* antara vektor rata-rata (atau sering disebut sebagai *Centroid*) antara kedua *cluster* sebagai berikut [15].

$$D(A, B) = d(\bar{\mathbf{y}}_A, \bar{\mathbf{y}}_B) \tag{8}$$

dengan  $\bar{\mathbf{y}}_A$  dan  $\bar{\mathbf{y}}_B$  masing-masing adalah vektor rata-rata untuk vektor observasi di *A* dan vektor observasi di *B*, serta  $d(\bar{\mathbf{y}}_A, \bar{\mathbf{y}}_B)$  didefinisikan dalam persamaan (2).  $\bar{\mathbf{y}}_A$  didefinisikan dalam persamaan (9), serta dua *cluster* dengan jarak terkecil antara *centroid* digabungkan pada setiap langkah [15]. Setelah dua *cluster* *A* dan *B* digabungkan, terdapat *centroid* dari *cluster* yang baru *AB* dengan rata-rata terboboti yang dituliskan dalam persamaan (10).

$$\bar{\mathbf{y}}_A = \frac{(\sum_{i=1}^{n_A} \mathbf{y}_i)}{n_A} \tag{9}$$

$$\bar{\mathbf{y}}_{AB} = \frac{n_A \bar{\mathbf{y}}_A + n_B \bar{\mathbf{y}}_B}{n_A + n_B} \tag{10}$$

6. *Median Linkage*

Jika dua *cluster* *A* dan *B* digabungkan menggunakan metode *Centroid*, dan jika *A* berisi jumlah item lebih besar dari *B*, maka *centroid* baru dalam persamaan (10) dapat lebih dekat ke  $\bar{\mathbf{y}}_A$  daripada  $\bar{\mathbf{y}}_B$  [15]. Untuk menghindari pembobotan vektor rata-rata berdasarkan ukuran *cluster*, dapat digunakan median dari garis yang menghubungkan *A* dan *B* sebagai titik untuk menghitung jarak baru ke *cluster* lain dengan persamaan (11).

$$\mathbf{m}_{AB} = \frac{1}{2}(\bar{\mathbf{y}}_A + \bar{\mathbf{y}}_B) \tag{11}$$

Dua *cluster* dengan jarak antarmedian terkecil digabungkan pada setiap langkah berikutnya.

**C. Analisis Cluster Nonhierarki K-Means**

*K-Means* merupakan metode yang mencoba mempartisi data menjadi dua atau lebih kelompok menggunakan nilai rata-rata sebagai pusat kluster [16]. Metode ini memungkinkan suatu observasi berpindah dari satu *cluster* ke *cluster* lainnya, sebagai bentuk suatu realokasi yang tidak terdapat dalam analisis *cluster* hierarki. Urutan langkah melakukan analisis *cluster* nonhierarki *K-Means* adalah sebagai berikut [17].

1. Menentukan besar  $K$  (banyaknya *cluster*);
2. Menentukan *centroid* secara acak pada awal tahap;
3. Menghitung jarak terdekat pada setiap data dengan *centroid*. Untuk menghitung jarak *centroid* digunakan jarak *Euclidean* yang didefinisikan dalam persamaan (3);
4. Menghitung kembali pusat *cluster* ( $C$ ) dengan keanggotaan *cluster* yang baru terbentuk. Jika  $M$  menyatakan jumlah data dalam satu *cluster*,  $i$  menyatakan data atau observasi ke  $i$  dalam sebuah *cluster*, dan  $p$  menyatakan dimensi data, maka diperoleh persamaan (12) sebagai berikut.

$$C = \frac{1}{M} \sum_{i=1}^M X_i \tag{12}$$

5. Menghitung kembali setiap objek menggunakan *cluster* baru menggunakan *centroid* baru, jika anggota *cluster* masih mengalami perpindahan atau menempati lebih dari satu *cluster* maka proses *clustering* dapat dimulai dari langkah  $c$  kembali. Proses selesai apabila tidak terdapat anggota yang mengalami perpindahan *cluster* kembali.

**D. Pseudo – F dan Internal Cluster Dispersion Rate (icdrate)**

Jumlah *cluster* atau kelompok optimal ditentukan dengan meninjau nilai statistik *Pseudo – F*. Statistik *Pseudo – F* secara komputasi memberikan hasil terbaik untuk menentukan berapa *cluster* yang optimal [18]. Persamaan (13) merupakan formulasi statistik *Pseudo – F*.

$$Pseudo - F = \frac{\left(\frac{R^2}{c - 1}\right)}{\left(\frac{1 - R^2}{n - c}\right)} \tag{13}$$

dengan,

$$R^2 = \frac{SST - SSW}{SST} \tag{14}$$

$$SST = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_k)^2 \tag{15}$$

$$SSW = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \tag{16}$$

$R^2$  adalah koefisien determinasi yang menunjukkan kontribusi dari variabel-variabel,  $SST$  adalah total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan,  $SSW$  adalah total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya,  $n$  adalah banyaknya sampel,  $c$  adalah banyak kelompok,  $n_c$  adalah banyak data pada kelompok ke  $i$ ,  $p$  adalah banyak variabel,  $x_{ijk}$  adalah sampel ke  $i$  pada kelompok ke  $j$  dan variabel ke  $k$ ,  $\bar{x}_k$  adalah rata-rata seluruh sampel pada variabel  $k$ ,  $\bar{x}_{jk}$  adalah rata-rata sampel pada kelompok ke  $j$  dan variabel ke  $k$  [13].

Kebaikan dalam *cluster* adalah homogenitas yang tinggi antaranggota dalam satu *cluster* (*within-cluster*) dan heterogenitas yang tinggi antar-*cluster* (*between-cluster*). *Icdrate* digunakan untuk menilai homogenitas dalam kelompok dan heterogenitas antarkelompok. Semakin kecil nilai *icdrate*, semakin baik hasil pengelompokannya [13]. Persamaan (17) merupakan formulasi statistik *internal cluster dispersion rate (icdrate)*.

$$icdrate = 1 - \frac{SST - SSW}{SST} = 1 - R^2 \tag{17}$$

**III. METODOLOGI PENELITIAN**

**A. Sumber Data**

Penelitian ini menggunakan objek sebanyak 38 kabupaten dan kota di Jawa Timur dengan menggunakan data sekunder. Seluruh data pada penelitian ini diperoleh dari publikasi berjudul Provinsi Jawa Timur dalam Angka 2022 yang diterbitkan oleh Badan Pusat Statistik (BPS) Provinsi Jawa Timur.

**B. Variabel Penelitian**

Dalam penelitian ini, data dikelompokkan menurut 38 kabupaten dan kota di Jawa Timur. Variabel-variabel yang digunakan merupakan indikator kesejahteraan masyarakat yang didasarkan pada penelitian sebelumnya berjudul, “*Clustering of Districts and Cities in Indonesia Based on Poverty Indicators Using The K-Means Method*” yang dilakukan oleh Niswatin dkk [5]. Variabel yang digunakan beserta definisi operasional variabel disajikan dalam Tabel 1.

**Tabel 1** Variabel, Definisi Operasional Variabel, dan Satuan Variabel Penelitian

Variabel Penelitian	Definisi Operasional	Satuan
$X_1$	<b>Jumlah Penduduk Miskin</b> Jumlah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan di bawah garis kemiskinan.	Ribuan
$X_2$	<b>Tingkat Pengangguran Terbuka (TPT)</b> Persentase jumlah pengangguran terhadap jumlah angkatan kerja.	Persentase (%)
$X_3$	<b>Indeks Pembangunan Manusia (IPM)</b> Indeks komposit yang mengukur pembangunan manusia dari tiga aspek dasar yaitu umur panjang dan hidup sehat, pengetahuan, dan standar hidup layak.	Desimal
$X_4$	<b>Rata-rata Pengeluaran per Kapita untuk Makanan Sebulan</b> Biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut.	Rupiah
$X_5$	<b>Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak</b> Perbandingan antara jumlah rumah tangga yang memiliki akses terhadap sumber air minum layak terhadap jumlah rumah tangga, dinyatakan dalam persen (%).	Persentase (%)
$X_6$	<b>Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Sanitasi Layak</b> Perbandingan antara jumlah rumah tangga yang memiliki akses terhadap layanan sanitasi layak terhadap jumlah rumah tangga, dinyatakan dalam persen (%).	Persentase (%)

**C. Teknik Analisis Data**

Pada penelitian ini digunakan analisis *cluster* dengan metode pengklasteran *single linkage*, *complete linkage*, *average linkage*, metode *ward*, metode *centroid*, *median linkage*, dan metode *K-Means*. Tahapan analisis data yang digunakan adalah sebagai berikut [9].

1. Mengumpulkan data yang akan digunakan dalam penelitian.
2. Mendeskripsikan variabel-variabel yang merupakan indikator kesejahteraan dengan analisis statistika deskriptif yang disajikan dengan nilai *mean* (rata-rata), nilai minimum, dan nilai maksimum.
3. Melakukan standarisasi pada data.
4. Melakukan pengecekan asumsi analisis *cluster*, yaitu terjadinya kondisi multikolinieritas pada data. Untuk mengetahui apakah data terindikasi multikolinieritas dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF).
5. Melakukan analisis *cluster* dengan pendekatan analisis *cluster* hierarki menggunakan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward*, metode *centroid*, dan *median linkage* sebagai metode analisis *cluster* hierarki berbasis prinsip penggabungan (*agglomerative*), serta metode *K-Means* sebagai metode analisis *cluster* nonhierarki, dengan jumlah *cluster* pada tiap metode sebanyak 2 hingga 6 *cluster*.
6. Menentukan jumlah *cluster* optimum pada setiap metode dengan cara menghitung nilai *Pseudo - F*.
7. Membandingkan ketujuh metode dan menentukan metode terbaik dengan menghitung nilai *icdrate*.
8. Melakukan identifikasi karakteristik masing-masing *cluster* dengan menghitung *centroid* berdasarkan metode terbaik.

**IV. HASIL DAN PEMBAHASAN**

**A. Deskripsi Data**

Data pada penelitian ini diperoleh dari publikasi berjudul Provinsi Jawa Timur dalam Angka 2022 yang diterbitkan oleh Badan Pusat Statistik Provinsi Jawa Timur. Berdasarkan data yang diperoleh, akan dilakukan pengelompokan kondisi karakteristik wilayah di 38 kabupaten dan kota pada Provinsi Jawa Timur menurut indikator kesejahteraan masyarakat. Ringkasan deskripsi data untuk setiap variabel penelitian dapat ditinjau pada Tabel 2.

**Tabel 2** Gambaran umum Kesejahteraan Masyarakat pada Jawa Timur

Variabel Penelitian	Rata-Rata ( <i>Mean</i> )	Minimum		Maksimum	
		Nilai	Kabupaten/Kota	Nilai	Kabupaten/Kota
Jumlah Penduduk Miskin ( $X_1$ )	120,34	8,37	Kota Mojokerto	276,58	Kabupaten Malang
Tingkat Pengangguran Terbuka (TPT) ( $X_2$ )	5,52	2,04	Kabupaten Pacitan	10,87	Kabupaten Sidoarjo
Indeks Pembangunan Manusia (IPM) ( $X_3$ )	72,23	62,8	Kabupaten Sampang	82,31	Kota Surabaya
Rata-rata Pengeluaran per Kapita untuk Makanan Sebulan ( $X_4$ )	550.879,03	385.890	Kabupaten Pacitan	810.743	Kota Surabaya
Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak ( $X_5$ )	95,18	75,78	Kabupaten Trenggalek	100	Kota Batu
Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Sanitasi Layak ( $X_6$ )	81,97	39,44	Kabupaten Bangkalan	97,31	Kota Madiun

**B. Pengujian Asumsi**

Analisis *cluster* harus memenuhi asumsi bahwa data yang akan digunakan tidak terjadi multikolinieritas, atau dengan kata lain, asumsi nonmultikolinieritas antarvariabel harus terpenuhi agar tidak ada variabel yang bersifat duplikatif atau merepresentasikan hal yang sama dalam analisis *cluster*. Hasil dari pengujian asumsi multikolinieritas disajikan dalam Tabel 3.

**Tabel 3** Nilai VIF masing-masing Variabel Penelitian

Variabel Penelitian	Nilai VIF
Jumlah Penduduk Miskin ( $X_1$ )	2,082
Tingkat Pengangguran Terbuka (TPT) ( $X_2$ )	3,492
Indeks Pembangunan Manusia (IPM) ( $X_3$ )	9,198
Rata-rata Pengeluaran per Kapita untuk Makanan Sebulan ( $X_4$ )	3,267
Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak ( $X_5$ )	1,324
Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Sanitasi Layak ( $X_6$ )	2,867

Berdasarkan Tabel, diketahui bahwa nilai VIF pada tiap variabel bernilai kurang dari 10. Maka dapat diketahui bahwa tidak terjadi multikolinieritas antar variabel-variabel tersebut. Dengan demikian, ukuran jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean*.

**C. Proses Pengklasteran**

Proses pengklasteran pada penelitian ini dilakukan dengan mengelompokkan kabupaten dan kota di Provinsi Jawa Timur berdasarkan indikator kesejahteraan masyarakat menggunakan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward*, metode *centroid*, dan *median linkage* yang merupakan metode analisis *cluster* hierarki dengan prinsip penggabungan (*agglomerative*), serta metode *K-Means* sebagai metode analisis *cluster* nonhierarki. Jarak tiap objek (kabupaten dan kota) dihitung dengan jarak *Euclidean* menggunakan persamaan (3), di mana  $i = 1, 2, 3, \dots, 38$ ,  $h = 1, 2, 3, \dots, 38$ , dengan kondisi  $i \neq h$  dan  $k = 1, 2, 3, \dots, 6$ . Setelah memperoleh jarak *Euclidean* dari satu objek ke objek yang lain, maka dilanjutkan dengan membuat *cluster*.

1. *Single Linkage*

Pembentukan kelompok menggunakan metode ini dilakukan dengan menggabungkan objek yang memiliki jarak terdekat ataupun tingkat kemiripan tertinggi. Pada tahap ini, untuk menentukan penggabungan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *single linkage* disajikan dalam Tabel 4.

**Tabel 4** Nilai *Pseudo - F* untuk metode *Single Linkage*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	3,071
3	2,388
<b>4</b>	<b>4,050</b>
5	3,431
6	3,050

Dapat ditinjau berdasarkan Tabel 4, Banyak *cluster* optimal berdasarkan metode *single linkage* adalah 4 *cluster*.

2. *Complete Linkage*

Pembentukan kelompok menggunakan metode ini dilakukan dengan menggabungkan objek yang memiliki jarak terjauh satu sama lain. Pada tahap ini, untuk menentukan penggabungan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *complete linkage* disajikan dalam Tabel 5.

**Tabel 5** Nilai *Pseudo - F* untuk metode *Complete Linkage*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	<b>26,566</b>
3	20,883
4	17,306
5	19,346
6	20,869

Dapat ditinjau berdasarkan Tabel 5, Banyak *cluster* optimal berdasarkan metode *complete linkage* adalah 2 *cluster*.

3. *Average Linkage*

Pembentukan kelompok menggunakan metode ini dilakukan dengan mengelompokkan objek berdasarkan rata-rata yang diperoleh dengan melakukan rata-rata objek secara keseluruhan terlebih dahulu. Pada tahap ini, untuk menentukan pengelompokan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *average linkage* disajikan dalam Tabel 6.

**Tabel 6** Nilai *Pseudo - F* untuk metode *Average Linkage*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	3,071
3	16,669
4	<b>17,306</b>
5	14,267
6	12,344

Dapat ditinjau berdasarkan Tabel 6, Banyak *cluster* optimal berdasarkan metode *average linkage* adalah 4 *cluster*.

4. *Metode Ward*

Pengelompokan metode *ward* dilakukan dengan memperkecil total jumlah kuadrat dalam kelompok. Masing-masing kelompok dibentuk sedemikian hingga menghasilkan kelompok dengan jumlah kuadrat terkecil dan dikenal dengan *Error Sums of Squares* (ESS). Pada tahap ini, untuk menentukan pengelompokan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *ward* disajikan dalam Tabel 7.

**Tabel 7** Nilai *Pseudo - F* untuk metode *Ward*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	<b>23,730</b>
3	19,961
4	19,591
5	19,806
6	19,889

Dapat ditinjau berdasarkan Tabel 7, Banyak *cluster* optimal berdasarkan metode *ward* adalah 2 *cluster*.

5. *Metode Centroid*

Pengelompokan metode *centroid* dilakukan berdasarkan dua *cluster* dengan jarak terkecil antara centroid yang digabungkan pada setiap langkah. Pada tahap ini, untuk menentukan pengelompokan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *centroid* disajikan dalam Tabel 8.

**Tabel 8** Nilai *Pseudo - F* untuk metode *Centroid*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	3,071
3	4,983
4	<b>6,662</b>
5	5,546
6	4,574

Dapat ditinjau berdasarkan Tabel 8, Banyak *cluster* optimal berdasarkan metode *centroid* adalah 4 *cluster*.

6. *Median Linkage*

Pembentukan kelompok dengan metode *median linkage* dilakukan untuk menghindari pembobotan vektor rata-rata berdasarkan ukuran *cluster*. Metode *median linkage* berbasis penggunaan median dari garis yang menghubungkan A dan B sebagai titik untuk menghitung jarak baru ke *cluster* lain. Pada tahap ini, untuk menentukan pengelompokan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo - F* dilakukan secara komputasi. Nilai *Pseudo - F* untuk *cluster* 2 hingga 6 berdasarkan metode *median linkage* disajikan dalam Tabel 9.

**Tabel 9** Nilai *Pseudo - F* untuk metode *Median Linkage*

Banyak Cluster	Nilai <i>Pseudo - F</i>
2	3,071
3	4,983
4	<b>9,362</b>
5	7,730
6	6,930

Dapat ditinjau berdasarkan Tabel 9, Banyak *cluster* optimal berdasarkan metode *median linkage* adalah 4 *cluster*.

7. Analisis Cluster Nonhierarki K-Means

Pembentukan kelompok dengan metode *K-Means* dilakukan untuk memungkinkan suatu observasi berpindah dari satu *cluster* ke *cluster* lainnya, sebagai bentuk suatu realokasi yang tidak terdapat dalam analisis *cluster* hierarki. Pada tahap ini, untuk menentukan pengelompokan objek dari tiap *cluster* serta penentuan *cluster* optimum melalui nilai *Pseudo – F* dilakukan secara komputasi. Nilai *Pseudo – F* untuk *cluster* 2 hingga 6 berdasarkan metode *K-Means* disajikan dalam Tabel 10.

Tabel 10 Nilai *Pseudo – F* untuk metode *K-Means*

Banyak Cluster	Nilai <i>Pseudo – F</i>
2	28,067
3	18,258
4	18,768
5	22,234
6	20,104

Dapat ditinjau berdasarkan Tabel 10, Banyak *cluster* optimal berdasarkan metode *K-Means* adalah 2 *cluster*.

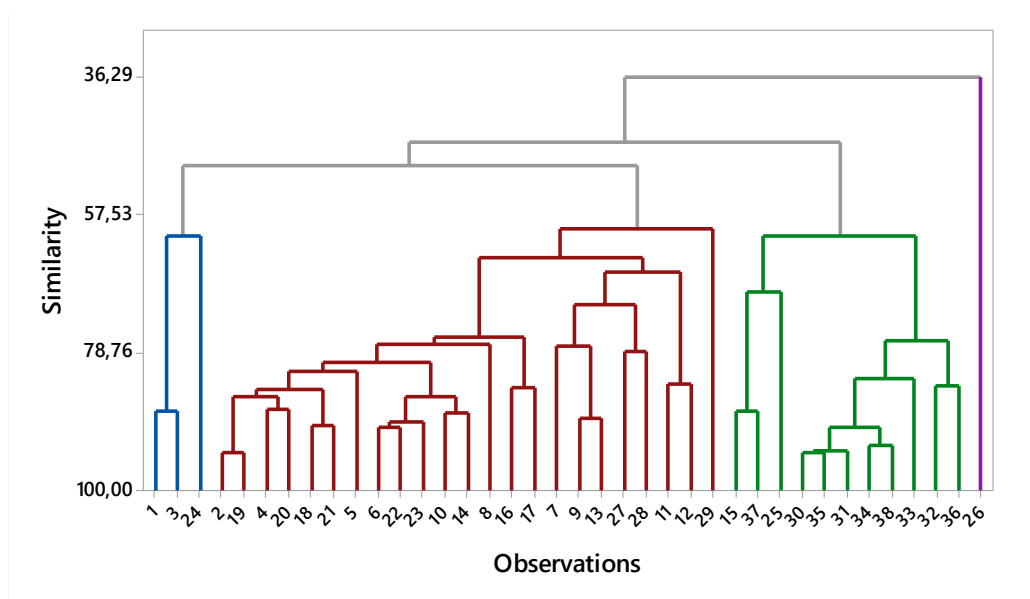
D. Penentuan Metode Analisis Terbaik

Untuk memperoleh hasil pengelompokan optimal diantara ketujuh metode yang telah digunakan dalam mengategorisasi kabupaten dan kota di Provinsi Jawa Timur, maka ketujuh metode perlu dibandingkan berdasarkan tingkat kebaikan hasil pengelompokan. Kebaikan hasil pengelompokan tersebut dapat ditinjau berdasarkan nilai *internal cluster dispersion rate (icdrate)*. Semakin kecil nilai *icdrate*, maka semakin baik metode tersebut dalam melakukan kategorisasi. Nilai  $R^2$  dan *icdrate* yang digunakan untuk kebaikan metode *cluster* disajikan dalam Tabel 11.

Tabel 11 Nilai  $R^2$  dan *icdrate* masing-masing metode *cluster*

Metode Cluster	Banyak Cluster Optimum	$R^2$	<i>icdrate</i>
Single Linkage	4	0,263	0,737
Complete Linkage	2	0,425	0,425
Average Linkage	4	0,604	0,396
Ward	2	0,397	0,603
Centroid	4	0,37	0,630
Median Linkage	4	0,452	0,548
K-Means	2	0,438	0,562

Berdasarkan Tabel 11, nilai *icdrate* terkecil dimiliki oleh metode *Average Linkage* yaitu sebesar 0,396 dengan koefisien determinasi ( $R^2$ ) terbesar yaitu 0,604. Dengan demikian, dalam penelitian ini metode *average linkage* dengan empat *cluster* optimal dipilih sebagai metode terbaik untuk mengelompokkan atau mengategorisasi kabupaten dan kota di Jawa Timur berdasarkan indikator kesejahteraan rakyat. Gambar 1 menyajikan dendrogram untuk pembentukan 4 *cluster* berbasis metode *average linkage* sebagai metode pengelompokan terbaik.



Gambar 1 Dendrogram 4 *cluster* metode *average linkage* sebagai metode pengelompokan terbaik

E. Identifikasi Karakteristik Cluster

Setelah diketahui metode terbaik dan jumlah *cluster* optimum untuk pengelompokan kabupaten dan kota di Jawa Timur, selanjutnya dilakukan identifikasi karakteristik untuk setiap *cluster*. Hasil identifikasi karakteristik tiap kelompok *cluster* disajikan dalam Tabel 12.



**Tabel 12** Identifikasi Karakteristik *Cluster*

<b>Cluster 1 (Cluster Daerah Sejahtera)</b>			
Anggota <i>Cluster</i>	Kabupaten Pacitan	Kabupaten Trenggalek	Kabupaten Lamongan.
Karakteristik	<ul style="list-style-type: none"> <li>- Daerah dengan angka kemiskinan tinggi dan daerah ketersediaan air rendah.</li> <li>- Meskipun angka kemiskinan pada <i>cluster</i> ini tinggi, namun masih berada di bawah angka kemiskinan <i>cluster</i> 3 dan 4.</li> <li>- Ketersediaan air bersih dan tingkat pengangguran pada <i>cluster</i> ini menjadi yang paling rendah jika dibandingkan dengan <i>cluster</i> yang lain.</li> <li>- Indeks Pembangunan Manusia (IPM) dan kepemilikan sanitasi pada <i>cluster</i> ini tergolong tinggi jika dibandingkan dengan <i>cluster</i> 2 dan 4.</li> <li>- Pengeluaran per kapita pada <i>cluster</i> ini tergolong rendah jika dibandingkan dengan <i>cluster</i> 2 dan 3.</li> </ul>		
<b>Cluster 2 (Cluster Daerah Cukup Sejahtera)</b>			
Anggota <i>Cluster</i>	Kabupaten Ponorogo Kabupaten Tulungagung Kabupaten Blitar Kabupaten Kediri Kabupaten Malang Kabupaten Lumajang Kabupaten Jember Kabupaten Banyuwangi	Kabupaten Bondowoso Kabupaten Situbondo Kabupaten Probolinggo Kabupaten Pasuruan Kabupaten Mojokerto Kabupaten Jombang Kabupaten Nganjuk Kabupaten Madiun	Kabupaten Magetan Kabupaten Ngawi Kabupaten Bojonegoro Kabupaten Tuban Kabupaten Sampang Kabupaten Pamekasan Kabupaten Sumenep
Karakteristik	<ul style="list-style-type: none"> <li>- Daerah dengan angka kemiskinan tinggi dan IPM rendah.</li> <li>- Meskipun angka kemiskinan pada <i>cluster</i> ini tinggi, namun masih berada di bawah angka kemiskinan <i>cluster</i> 4.</li> <li>- Tingkat pengangguran, IPM, ketersediaan air bersih, dan kepemilikan sanitasi pada <i>cluster</i> ini tergolong rendah jika dibandingkan <i>cluster</i> lain.</li> <li>- Pengeluaran per kapita pada <i>cluster</i> ini tergolong tinggi jika dibandingkan dengan <i>cluster</i> lain.</li> </ul>		
<b>Cluster 3 (Cluster Daerah Sangat Sejahtera)</b>			
Anggota <i>Cluster</i>	Kabupaten Sidoarjo Kabupaten Gresik Kota Kediri Kota Blitar	Kota Malang Kota Probolinggo Kota Pasuruan Kota Mojokerto	Kota Madiun Kota Surabaya Kota Batu
Karakteristik	<ul style="list-style-type: none"> <li>- IPM, pengeluaran per kapita, ketersediaan air bersih, dan kepemilikan sanitasi pada <i>cluster</i> ini merupakan yang paling tinggi dibandingkan dengan seluruh <i>cluster</i> lain.</li> <li>- Angka kemiskinan pada <i>cluster</i> 3 merupakan yang paling rendah diantara seluruh <i>cluster</i> lain.</li> <li>- Meskipun tingkat pengangguran pada <i>cluster</i> ini tergolong tinggi, namun masih berada di bawah tingkat pengangguran <i>cluster</i> 4.</li> </ul>		
<b>Cluster 4 (Cluster Daerah Kurang Sejahtera)</b>			
Anggota <i>Cluster</i>	Kabupaten Bangkalan		
Karakteristik	<ul style="list-style-type: none"> <li>- Daerah dengan kepemilikan sanitasi rendah.</li> <li>- Angka kemiskinan dan tingkat pengangguran pada <i>cluster</i> 4 merupakan yang paling tinggi diantara seluruh <i>cluster</i> lain.</li> <li>- IPM dan pengeluaran per kapita pada <i>cluster</i> ini merupakan yang paling rendah dibandingkan dengan <i>seluruh cluster</i> lain.</li> <li>- Ketersediaan air bersih pada <i>cluster</i> ini tergolong tinggi, meskipun masih berada di bawah <i>cluster</i> 3.</li> </ul>		

Berdasarkan hasil pengelompokan yang telah diperoleh pada Tabel 12, dapat diketahui bahwa *Cluster* 4 hanya terdiri dari satu anggota, yaitu Kabupaten Bangkalan. Hal ini disebabkan karena Kabupaten Bangkalan merupakan kabupaten dengan kondisi data *outlier* pada variabel Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Sanitasi Layak ( $X_6$ ). Hal yang demikian juga terjadi dalam *Cluster* 1 dengan tiga anggota, yaitu Kabupaten Pacitan, Kabupaten Trenggalek, dan Kabupaten Lamongan. Ketiga kabupaten tersebut merupakan kabupaten dengan kondisi data *outlier* pada variabel Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak ( $X_5$ ) sehingga ketiga daerah tersebut tergabung ke dalam satu *cluster* yang sama.

Pengelompokan daerah berdasarkan hasil analisis *cluster* dengan menggunakan indikator kesejahteraan masyarakat divisualisasikan melalui peta Provinsi Jawa Timur pada Gambar 2.



Gambar 2 Pemetaan cluster daerah berdasarkan Indikator Kesejahteraan Masyarakat Provinsi Jawa Timur

**F. Rekomendasi**

Berdasarkan hasil identifikasi karakteristik tiap kelompok cluster yang disajikan dalam Tabel 12, berikut merupakan rekomendasi yang dapat dilakukan pemerintah agar mampu mengoptimalkan kebijakan yang tepat bagi setiap wilayah berdasarkan kepentingan wilayah tersebut sekaligus meminimalisasi terjadinya kebijakan salah sasaran.

1. Untuk daerah dengan angka kemiskinan yang tinggi, pemerintah dapat memberikan bantuan sosial dan pelatihan agar tidak hanya kebutuhan masyarakat yang terpenuhi namun kemampuan dan keterampilan pun juga meningkat.
2. Untuk daerah dengan tingkat pengangguran tinggi, pemerintah dapat menyediakan lowongan pekerjaan lebih banyak atau memberikan pelatihan agar masyarakat di daerah tersebut dapat membuka usaha sendiri, sehingga memiliki pekerjaan dan penghasilan yang tetap.
3. Untuk daerah dengan IPM rendah, pemerintah dapat memperbaiki sarana kesehatan dan mendorong masyarakat untuk selalu mengonsumsi makanan sehat dan bergizi. Pemerintah juga dapat memperbaiki sarana pendidikan dan mendorong masyarakat setempat supaya dapat mengenyam pendidikan setinggi mungkin.
4. Untuk daerah dengan pengeluaran per kapita rendah, pemerintah dapat lebih fokus memperbaiki keadaan masyarakat setempat dari sisi ekonomi. Mengurangi tingkat pengangguran juga menjadi satu diantara cara agar pengeluaran per kapita masyarakat bisa naik.
5. Untuk daerah dengan ketersediaan air bersih rendah, pemerintah dapat meningkatkan peningkatan kualitas infrastruktur manajemen air yang tahan terhadap perubahan iklim. Daerah dengan kendala berasal dari sumber air ataupun mata air, maka pemerintah dapat melakukan sinergi antardaerah dalam rangka penyediaan air bersih sebagai alternatif solusi jangka pendek sekaligus melakukan kajian dan riset ilmiah mendalam untuk menemukan solusi dari masalah tersebut agar tidak terjadi ketergantungan bantuan.
6. Untuk daerah dengan kepemilikan sanitasi rendah, pemerintah dapat memfasilitasi masyarakat dengan membangun sanitasi yang lebih layak untuk setiap tempat tinggal.

**V. KESIMPULAN DAN SARAN**

**A. Kesimpulan**

Perbandingan dan penentuan jumlah cluster optimum dalam pengelompokan kabupaten dan kota di Jawa Timur ditentukan melalui nilai *Pseudo - F*. Selain itu, peneliti juga menentukan metode terbaik melalui nilai *icdrate*. Diperoleh bahwa metode *average linkage* merupakan metode terbaik dengan empat cluster sebagai jumlah optimal dalam penelitian ini. Kesimpulan tersebut didasarkan pada nilai *icdrate* sebesar 0,396 dan  $R^2$  sebesar 0,604.

Karakteristik tingkat kemiskinan yang terjadi pada setiap kelompok daerah sangat beragam. Pada daerah cluster 1, angka kemiskinan dan IPM tinggi. Namun ketersediaan air bersih dan tingkat pengangguran rendah. Untuk wilayah cluster 2 merupakan daerah dengan angka kemiskinan tinggi dan IPM rendah. Cluster 3 merupakan kebalikan dari daerah di cluster 2 yaitu memiliki IPM tinggi dan angka kemiskinan rendah. Daerah di cluster 4 merupakan daerah dengan angka kemiskinan tinggi dan kepemilikan sanitasi rendah. Dengan mengetahui pengelompokan kabupaten dan kota di Jawa Timur dan melalui perumusan rekomendasi setiap cluster, penelitian ini diharapkan dapat bermanfaat sebagai pertimbangan untuk mengevaluasi pemerintah dan instansi terkait untuk mengatasi permasalahan pokok yang masih terjadi di masing-masing kabupaten dan kota. Dengan demikian, kesejahteraan masyarakat Jawa Timur dapat terwujud dan target SDGs di Indonesia dapat tercapai.

## B. Saran

Pengelompokan ini merupakan penelitian berdasarkan data tahun 2021. Penelitian daerah dengan tingkat kesejahteraan rakyat rendah perlu dilakukan berkala tiap tahun. Untuk penelitian selanjutnya jumlah *cluster* optimal juga bisa berubah sesuai nilai *Pseudo – F* data *cluster*. Metode terbaik yang digunakan juga dapat berubah sesuai dengan nilai *icdrate* terkecil. Selain itu keanggotaan *cluster* dapat berubah, daerah yang masuk kategori tingkat kesejahteraan rendah bisa saja berubah ke kategori tinggi ataupun sebaliknya.

Berdasarkan variabel dalam penelitian yang digunakan, dapat diketahui bahwa terdapat kabupaten sebagai observasi mengalami kondisi *outlier*, sehingga untuk penelitian selanjutnya peneliti dapat menggunakan metode analisis *cluster* yang mampu mengakomodasi kemunculan data *outlier*, misalnya dengan menggunakan *K-Medoids*. Penelitian ini menerapkan pengelompokan optimal dari kumpulan data multivariat dengan menggunakan metode *cluster* terbaik. Hasil penelitian ini logis dan sesuai jika dibandingkan dengan kondisi nyata kesejahteraan kabupaten dan kota di Jawa Timur saat ini.

## UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada Program Studi S1 Statistika Universitas Airlangga dan Badan Pusat Statistik yang turut memberikan kontribusi dalam penelitian dan penyusunan artikel ilmiah ini.

## DAFTAR PUSTAKA

- [1] F. Sellvy and W. Rolia, "Pengaruh Pendapatan Asli Daerah (PAD), Dana Bagi Hasil (DBH), Dana Alokasi Umum (DAU), dan Dana Alokasi Khusus (DAK), Terhadap Pengalokasian Belanja Modal pada Kota/ Kabupaten di Provinsi Sumatera Selatan," Diploma, Universitas Bina Darma, 2020. Accessed: Dec. 30, 2022. [Online]. Available: <http://repository.binadarma.ac.id/1201/>
- [2] "Badan Pusat Statistik." <https://www.bps.go.id/publication/2022/02/25/0a2afea4fab72a5d052cb315/statistik-indonesia-2022.html> (accessed Dec. 30, 2022).
- [3] "BPS Provinsi Jawa Timur." <https://jatim.bps.go.id/publication/2022/02/25/33699f6fcd84e0e2a0ad96f0/provinsi-jawa-timur-dalam-angka-2022.html>. (accessed Dec. 30, 2022).
- [4] A. Afifi, S. May, R. A. Donatello, V. A. Clark, *Practical Multivariate Analysis*. Chapman and Hall/CRC, 2019. doi: 10.1201/9781315203737.
- [5] K. Niswatin, C. Andreas, P. F. A. O. Hans, and M. F. F. Mardianto, "Clustering of Districts and Cities in Indonesia Based on Poverty Indicators Using The K-Means Method," Aug. 2021.
- [6] K. N. Khikmah and A. Sofro, "Clustering Regency and City in East Java Based on Population Density and Cumulative Confirmed COVID-19 Cases," *ComTech Comput. Math. Eng. Appl.*, vol. 12, no. 2, pp. 111–121, Nov. 2021, doi: 10.21512/comtech.v12i2.6891.
- [7] M. F. F. Mardianto *et al.*, "Classification of food menu and grouping of food potential to support the food security and nutrition quality," *Commun Math Biol Neurosci*, vol. 2022, no. 0, Art. no. 0, Mar. 2022, doi: 10.28919/cmbn/6801.
- [8] S. Wierzchoń and M. Kłopotek, *Modern Algorithms of Cluster Analysis*, vol. 34. in *Studies in Big Data*, vol. 34. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-69308-8.
- [9] A. N. Fathia and R. Rahmawati, "Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage," vol. 5, no. 4, 2016.
- [10] M. Sriningsih, D. Hatidja, and J. D. Prang, "Penanganan Multikolinearitas dengan Menggunakan Analisis Regresi Komponen Utama pada Kasus Impor Beras di Provinsi Sulut," *J. Ilm. SAINS*, vol. 18, no. 1, p. 18, Jul. 2018, doi: 10.35799/jis.18.1.2018.19396.
- [11] J. F. Hair, W. C. Black, and B. J. Babin, *Multivariate Data Analysis: A Global Perspective*. Pearson Education, 2010.
- [12] B. Simamora, "Analisis Multivariat Pemasaran," *Universitas Indonesia Library*, 2005. <https://lib.ui.ac.id> (accessed Dec. 30, 2022).
- [13] M. F. F. Mardianto, K. Kuzairi, T. Yulianto, R. Amalia, and F. Faisol, "Pengelompokan Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik," *Zeta - Math J.*, vol. 1, no. 1, Art. no. 1, May 2015, doi: 10.31102/zeta.2015.1.1.22-29.
- [14] R. W. Maya, A. Ramadhan, I. M. Ayuputri, and B. W. Otok, "Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan Faktor-Faktor yang Mempengaruhi AIDS," *Inferensi*, vol. 2, no. 1, p. 19, Mar. 2019, doi: 10.12962/j27213862.v2i1.6807.
- [15] A. C. Rencher, *Methods of Multivariate Analysis*. John Wiley & Sons, 2003.
- [16] C. Oktarina, K. A. Notodiputro, and I. Indahwati, "Comparison of K-Means Clustering Method and K-Medoids on Twitter Data," *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, Art. no. 1, Feb. 2020, doi: 10.29244/ijsa.v4i1.599.
- [17] A. Afnanda and A. Arnellis, "Analisis Cluster Covid-19 di Sumatera Barat dengan Metode Non-Hirarki (K-Means)," *J. Math. UNP*, vol. 6, no. 3, Art. no. 3, Dec. 2021, doi: 10.24036/unpjomath.v6i3.11818.
- [18] A. Sirojuddin, "Analisis Cluster pada Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Indeks Pembangunan Manusia," Undergraduate, Universitas Islam Negeri Maulana Malik Ibrahim, 2016. Accessed: Dec. 30, 2022. [Online]. Available: <http://etheses.uin-malang.ac.id/5762/>



© 2023 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).