VOLUME 28, NOMOR 1, APRIL 2024 ISSN 1410-9883

CAKRAWALA PENDIDIKAN

FORUM KOMUNIKASI ILMIAH DAN EKSPRESI KREATIF ILMU PENDIDIKAN

Terbit 30 April 2024

ISSN 1410-9883

CAKRAWALA PENDIDIKAN
Forum Komunikasi Ilmiah dan Ekspresi Kreatif Ilmu Pendidikan

Terbit dua kali setahun pada bulan April dan Oktober
Terbit pertama kali April 1999

Alamat Penerbit/Redaksi : Universitas PGRI Adi Buana Kampus Blitar: Jl. Kalimantan No. 111 Blitar, Telp. (0342) 801493. Langganan 2 Nomor setahun Rp. 200.000,00 ditambah ongkos kirim Rp. 50.000,00.

CAKRAWALA PENDIDIKAN diterbitkan oleh Universitas PGRI Adi Buana Kampus Blitar. **Direktur Operasional** : Dra. Riki Suliana RS., M.Pd.

Penyunting menerima artikel yang belum pernah diterbitkan di media cetak yang lainnya. Syarat-syarat, format dan aturan tata tulis artikel dapat diperiksa pada *Petunjuk bagi Penulis* di sampul belakang dalam jurnal ini. Artikel yang masuk akan ditelaah oleh Tim Penyunting dan Mitra Bestari untuk dinilai kelayakannya. Tim akan melakukan perubahan tata letak dan tata bahasa yang diperlukan tanpa mengubah maksud dan isinya.

Petunjuk Penulisan Cakrawala Pendidikan

1. Artikel belum pernah diterbitkan di media cetak yang lainnya.
2. Artikel diketik dengan memperhatikan aturan tentang penggunaan tanda baca dan ejaan yang baik dan benar sesuai *Pedoman Umum Ejaan Bahasa Indonesia yang Disempurnakan (Depdikbud, 1987)*
3. Pengetikan Artikel dalam format Microsoft Word, ukuran kertas A4, spasi 1.5, jenis huruf *Times New Roman*; ukuran huruf 12. Dengan jumlah halaman; 10 – 20 halaman.
4. Artikel yang dimuat dalam Jurnal ini meliputi tulisan tentang hasil penelitian, gagasan konseptual, kajian dan aplikasi teori, tinjauan kepustakaan, dan tinjauan buku baru.
5. Artikel ditulis dalam bentuk esai, disertai judul sub bab (heading) masing-masing bagian, kecuali bagian pendahuluan yang disajikan tanpa judul sub bab. Peringkat judul sub bab dinyatakan dengan jenis huruf yang berbeda, letaknya rata tepi kiri halaman, dan tidak menggunakan nomor angka, sebagai berikut:

   PERINGKAT 1 (HURUF BESAR SEMUA TEBAL, RATA TEPI KIRI)
   Peringkat 2 (Huruf Besar-kecil Tebal, Rata Tepi Kiri)
   Peringkat 3 (*Huruf Besar-kecil Tebal, Miring, Rata Tepi Kiri*)

6. Artikel konseptual meliputi; (a) judul, (b) nama penulis, (c) abstrak dalam bahasa Indonesia dan Inggris (maksimal 200 kata), (d) kata kunci, (e) identitas penulis (tanpa gelar akademik), (f) pendahuluan yang berisi latar belakang dan tujuan atau ruang lingkup tulisan, (g) isi/pembahasan (terbagi atas sub-sub judul), (h) penutup, dan (i) daftar rujukan. Artikel hasil penelitian disajikan dengan sistematika: (a) judul, (b) nama- nama peneliti, (c) abstrak dalam bahasa Indonesia dan Inggris (maksimal 200 kata), (d) kata kunci, (e) identitas penulis (tanpa gelar akademik), (f) pendahuluan yang berisi pembahasan kepustakaan dan tujuan penelitian, (g) metode, (h) hasil, (i) pembahasan (j) kesimpulan dan saran, dan (k) daftar rujukan.
7. Daftar rujukan disajikan mengikuti tata cara seperti contoh berikut dan diurutkan secara alfabetis dan kronologis.

   Anderson, D.W., Vault, V.D., dan Dickson, C.E. 1993. *Problem and Prospects for the Decades*
   Ahead: Competency Based Teacher Education. Barkeley: McCutchan Publishing Co.
   Huda, N. 1991. *Penulisan Laporan Penelitian untuk Jurnal.* Makalah disajikan dalam Loka
   Karya Penelitian Tingkat Dasar bagi Dosen PTN dan PTS di Malang Angkatan XIV, Pusat Penelitian IKIP MALANG, Malang, 12 Juli.
   Prawoto, 1998. *Pengaruh Pengirformasian Tujuan Pembelajaran dalam Modul terhadap Hasil*
   Belajar Siswa SD PAMONG Kelas Jauh. Tesis tidak diterbitkan. Malang: FPS IKIP MALANG.
   Russel, T. 1993. An Alternative Conception: Representing Representation. Dalam P.J. Nlack & A. Lucas (Eds.) *Children's Informal Ideas in Science* (hlm.62-84). London:Routledge.
   Sihombing, U. 2003. *Pendataan Pendidikan Berbasis Masyarakat.* http://www.puskur.or.id.Diakses pada 21 April 2006.
   Zainuddin, M.H. 1999. Meningkatkan Mutu Profesi Keguruan Indonesia. *Cakrawala Pendidikan.* 1 (1):45-52.

8. Pengiriman Artikel via email ke hudaferi@gmail.com paling lambat 3 bulan sebelum bulan penerbitan.

CAKRAWALA PENDIDIKAN

Forum Komunikasi Ilmiah dan Ekspresi Kreatif Ilmu Pendidikan

Volume 28, Nomor 1, April 2024

Daftar Isi

COMPARATIVE ANALYSIS OF DISADVANTAGED AREAS IN REGENCIES/CITIES IN EASTERN INDONESIA IN 2021 USING THE K-MEANS CLUSTERING, K-MEDOIDS CLUSTERING, AND FUZZY C-MEANS CLUSTERING METHODS

Bayu Aji Bachtiar[1)], Udin Erawanto[2)]
212011594@stis.ac.id[1)] erawantoudin@gmail.com[2)]
Politeknik Statistika STIS[1]
Universitas PGRI Adi Buana Kampus Blitar[2]

Abstrak: Tujuan dari penelitian ini adalah untuk membandingkan penggunaan metode K-Means Clustering, K-Medoids Clustering dan Fuzzy C-Means Clustering dalam klasifikasi daerah tertinggal kabupaten/kota di Indonesia Timur. Data yang digunakan dalam penelitian ini adalah data sekunder Statistik Keuangan Pemerintah Kabupaten/Kota edisi 2020/2021 di website BPS dengan menggunakan unit observasi Sulawesi Utara, Sulawesi Selatan, Maluku, Maluku Utara, NTB dan unit observasi provinsi kabupaten/kota. Daerah Beda Persentase Penduduk Miskin di Papua Barat, Dana Alokasi Khusus Daerah, Angka Harapan Hidup, Jumlah Puskesmas. Hasil penelitian ini menunjukkan bahwa clustering singular dengan menggunakan algoritma k-means lebih optimal dibandingkan dengan clustering dengan algoritma k-medoids atau fuzzy c-means untuk indikator daerah tertinggal di Indonesia bagian timur. Hasil klasterisasi k-means terhadap 176 daerah di Indonesia bagian timur berdasarkan indikator daerah tertinggal terbagi menjadi 3 klaster yaitu klaster sangat tertinggal dan tidak tertinggal. Secara total, terdapat 91 kabupaten/kota yang tidak tertinggal jauh, 48 kabupaten/kota tertinggal sedang, dan 37 kabupaten/kota tertinggal.

Kata kunci: Analisis Komparatif, Daerah Tertinggal, Indonesia Bagian Timur, K-Means Clustering, K-Medoids Clustering, Metode Fuzzy C-Means Clustering

Abstract: The purpose of this study is to compare the use of K-Means Clustering, K-Medoids Clustering and Fuzzy C-Means Clustering methods in the classification of districts/urban underdeveloped areas in Eastern Indonesia. The data used in this study is secondary data from the Regency/City Government Financial Statistics 2020/2021 edition on the BPS website using North Sulawesi, South Sulawesi, Maluku, North Maluku, NTB and provincial districts/cities observation units. area Different percentage of poor people in West Papua, Regional Special Allocation Fund, Life Expectancy, number of health centers. The results of this study show that singular clustering using k-means algorithms is more optimal than clustering with k-medoids or fuzzy c-means algorithms for indicators of disadvantaged areas in eastern Indonesia. The k-means clustering results of 176 regions in eastern Indonesia, based on the indicators of underdeveloped areas, are divided into 3 clusters, namely very underdeveloped clusters and not backward. In total, 91 districts are not far behind, 48 are moderately behind and 37 are behind.

Keywords: Comparative Analysis, Disadvantaged Areas, Eastern Indonesia, K-Means Clustering, K-Medoids Clustering, Fuzzy C-Means Clustering Methods

INTRODUCTION

Development in a country is very important, because the main purpose of implementing the development itself is to improve people's well-being by satisfying their needs and desires [3]. Development is a series of growth and change efforts that the nation, state and government consciously plan and implement towards modernity as part of nation building [23]. Economic development leads to strong growth, which has not completely overcome the problem of regional disparities. Regional differences in development rates lead to differences in wealth and development between regions, especially between Java and non-Java, West Indonesia and East Indonesia, and between urban and rural areas.

The development gap between regions is one of the development problems that often occur in developing countries. The reason for the underdevelopment of the region can be the underdevelopment of the economic activities of the community, the poor quality of human resources, the lack of adequate facilities and infrastructure, the increase of natural disasters and conflicts, and the poor availability of transport, telecommunication and data connection information [20].

Disadvantaged regions are regions whose areas and people are less developed compared to other regions nationwide. A region is defined as a disadvantaged area based on community economy, human resources, facilities and infrastructure, regional economic capacity, accessibility and regional characteristics. Based on the presidential decree, there are 64 districts in the list of underdeveloped regions from 2020 to 2024, with data from 7 districts in the western region of Indonesia and 54 districts in the eastern region of Indonesia [19]. Based on the National Policy Guidelines of 1993, the Indonesian region is divided into two development regions, namely the Western Region of Indonesia, which consists of the islands of Java, Sumatra, Kalimantan, and Bali, and the Eastern Region of Indonesia, which consists of the islands Sulawesi, Maluku, Papua, West and East Nusa Tenggara.

The Minister evaluates underdeveloped areas periodically in accordance with the planning period for accelerating the development of underdeveloped areas using the composite index calculation method. The use of composite indices in the process of classifying underdeveloped areas in Indonesia is still not optimal. This is because there are still areas that are not included in the category of underdeveloped but tend to have the same characteristics as underdeveloped areas. As well as calculations using composite index calculations have the following drawbacks: 1) Can give rise to the impression of a misinterpreted policy, if the process is not according to standards. 2) Can create an impression of simplification/simplification. 3) It is easy to misuse if the process of compiling the index is not transparent, not based on theory, and does not use generally accepted statistical methods. 4) Selection of indicators and weights can lead to rejection. 5) Can disguise the failure of a program or conversely exaggerate a success, if the index is not compiled transparently. 6) Can lead to inappropriate policies due to neglect of hard-to-measure realities [5].

Cluster analysis is an approach to look for similarities in data and place similar data into groups. Cluster analysis divides a data set into several groups where the similarities within a group are greater than between groups [1]. In clustering, a measure is used that describes the similarity between objects (data) to explain a simpler group structure and originates from more cmplex data, namely a measure of similarity using Euclidean distance [10]. Cluster analysis is divided into several methods, namely hard clustering which consists of hierarchical and non-hierarchical methods, and soft clustering using Fuzzy C Means [27].

Several studies have been conducted to select the best grouping for cases of underdeveloped areas using various methods. Research conducted [22] found that grouping mixed data using the ROCK ensemble produced four optimum groups with an accuracy of 0.724. Where for numerical data, the best grouping results were obtained using the ward method with three variables resulting from factor analysis resulting in five optimum groups with pseudo-f and icdrate of 163.408 and 0.048 respectively. While grouping categorical data using the ROCK method produces four optimum groups with an accuracy of 0.711. 3. The results of grouping using the SWFM ensemble are five optimum groups with an accuracy of 0.618. 4. Comparison of mixed data grouping with ROCK and SWFM ensembles shows that the ROCK ensemble method. In addition, research conducted by [21] in Bangkalan Regency obtained the results of a cluster analysis of the existing conditions of each of these factors, underdeveloped villages in Bangkalan Regency are divided into 3 clusters, namely cluster I (underdeveloped villages with very high agricultural potential development). less developed), cluster II (underdeveloped villages with underdeveloped agricultural potential) of 16 villages and cluster III (underdeveloped villages with fairly developed agricultural potential)

This study aims to classify regencies/cities in Eastern Indonesia based on the characteristics of disadvantaged areas. With this grouping, it is hoped that the government will be able to apply appropriate policies to create equal distribution of wealth. In addition, thisSstudy also aims to compare several clustering methods that can be used to classify regions based on welfare levels in order to obtain the best clustering method for grouping districts/cities in Eastern Indonesia.

RESEARCH METHOD

Data and Data Sources

The data used in this study is secondary data sourced from the Indonesian Central Bureau of Statistics, both in the form of publications and data available on the website of the Indonesian Central Bureau of Statistics. The publication used is Regency/City Government Financial Statistics for 2020/2021

This study used district/city observation units in the provinces of North Sulawesi, South Sulawesi, Maluku, North Maluku, NTB and West Papua. The variables used in this study are as follows:

Table 1. Variable and Operational Definitions

Variable	Operational definitions
Percentage of poor people (P0)	Percentage of poor people who are below the poverty line.
Regional Special Allocation Fund	Funds provided to regions to meet special needs in accordance with applicable laws
Life expectancy	The estimated average number of years that a person can live from birth.
Number of health centers	The number of community health centers in each province that are the observation units.

Preprocessing

In data analysis, the next initial steps are data collection, data structure visualization, and summarization. A data structure is a way of storing, organizing and managing data on computer storage media so that the data can be used efficiently [26]. Summary data is a summary that summarizes the description of the data which consists of 5 measures, namely the average (mean), median, quartiles, minimum value, and maximum value. After that, pre-processing is done. Preprocessing is the stage before the clustering analysis process is carried out for the purpose of cleaning, eliminating, changing data types, and adjusting data types to analysis requirements. In this step, also checking for missing values and outlier data. Missing value refers to missing data in a data set. If there is a missing value then imputation is carried out with various methods. However, in this study there is no missing value. Meanwhile, outliers can be defined as extreme points of data that are far from the data center and deviate from other data and are likely to affect the regression analysis. Outlier data is not deleted in this study, for two reasons:Comparing the k-means, k-Medoids, and Fuzzy c-Means clustering methods in modeling outlier data, and all district/city observation units in Eastern Indonesia are considered important, deleting them from the dataset will destroy the information contained therein. Thus affecting the results of regional classification according to the level of regional wealth.

Determining Number of Clusters (K)

One challenge of the non-hierarchical clustering method is determining the number of clusters. To determine the number of clusters can be used according to the needs of researchers and/or by considering the most optimal k. This k value must be known in the output so that the computational algorithm can be started. The optimal k value can define a cluster where the objects in the cluster are homogeneous and between clusters are heterogeneous. In this study, the optimal number of clusters is determined by three methods as follows:

1. ***Elbow method***

The basic idea behind cluster partitioning is to determine k such that the within-cluster variation (known as the total within-cluster variation or number of squares within a cluster) is minimized.

$$minimize \left(\sum_{k=1}^{k} W(C_k)\right) \quad \textbf{(1)}$$

Where C_k is the kth cluster and $W(C_k)$ is the within-cluster variation. The total within a cluster measures the cohesiveness of the grouping and is expected to be as small as possible [12].

2. ***Average Silhouette Method***

The average silhouette approach measures clustering quality by determining how well each object is contained within its cluster. High average silhouette width indicates good grouping. The silhouette method is used to select the optimal number of clusters based on ratio scale data [4].

3. ***Gap Statistics Method***

The gap statistical method compares the sum of within-cluster variation for different values of k to the expected value under the zero-reference distribution of the data [6].A reference dataset is generated using Monte Carlo simulations in the sampling process. This means determining the range $[min(x_i), max\ (x_j)]$ for each variable (x_i) in the dataset and generating the values for point n uniformly from the minimum interval to the maximum interval. increase. For observed and reference data, the total within-cluster variation is computed using different k values. Gap statistics are defined as:

$$d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \quad \textbf{(2)}$$

Where x and y are objects; p is the number of variables. Combination x and y as many objects. So for example the dataset consists of 50 observations. So the Euclidean distance matrix has an order of 50 x 50, where the distance to the object itself is 0 [25].

Assumptions in Cluster Analysis

Cluster analysis requires the following prerequisites: The samples used can be representative of the population or can be described as representative. A representative sample is a sample selected to be representative of the study population. In this case, the Kaiser-Meier-Orkin test (KMO) is used to measure the adequacy of both the overall sample and each indicator. The KMO test results show that the KMO value is in the range of 0.5.

Besides that, Non-multicollinearity where Multicollinearity is the relationship between the independent variables, both positively and negatively related. If the value of the correlation coefficient is > 0.8 then there is a problem of multicollinearity between one or more independent variables [7].

Cluster Analysis

According to [15] ,clustering is a

method for grouping instances (samples) into several groups or subsets based on similarity to other instances. In multivariate analysis, clustering is basically grouping objects based on the similarity of the variable characteristics. The measure of this similarity is at a distance. Each object will be measured its distance matrix in the variable range to be grouped based on its proximity. Therefore a cluster is said to be good if the cluster is homogeneous and between clusters is heterogeneous. The following is the cluster analysis method used in this study:

K-Means Clustering

K-means clustering requires k input parameters and divides a set of n objects into k A non-hierarchical clustering method that divides into clusters. High ones are very low ones. The similarity or similarity between items (within a cluster) is determined based on their "closeness" calculated by averaging the midpoints to the given data (nearest mean). The center of the cluster is called the centroid. The number of foci corresponds to the number of clusters. However, the number of clusters cannot be determined initially [9].

The k-means algorithm is an algorithm that takes k input parameters and partitions a set of n objects into k clusters [28]. The basic concept of K-Means is a cluster-centric iterative search. The cluster center is determined based on the distance of each datum from the cluster center.

K-Medoids (Partitioning Around Medoids/PAM)

K-medoids uses k as the number of initial cluster centers randomly generated at the start of the clustering process [8]. The algorithm is similar to K-Means. Objects close to the center of the cluster are grouped together to form a new cluster. K-Medoids uses a partition clustering technique to group a set of n objects into k clusters. This algorithm uses objects in a collection of objects that represent clusters. Objects representing clusters are called medoids.

Fuzzy C-Means

According to [18], fuzzy C-means clustering is a technique that determines the best clusters in a vector space based on the Euclidean normal form of the distance between vectors. The fuzzy C-means clustering algorithm is often applied to group data based on region similarity. Fuzzy C-Means (FCM) is one of the most commonly used fuzzy algorithms. FCM groups objects based on fuzzy logic.

MODEL EVALUATION

Silhouette Coefficient

The Silhoutte Coefficient is an evaluation method used to see the quality and strength of the cluster results. The Silhoutte coefficient method combines cohesion and separation methods on individual data and is formulated as follows [2].

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

where a(i) is the average distance from the i-th object to other objects in one cluster and b(i) is the average distance from the i-th object to other objects in another cluster. Silhouette values range from -1 to +1. A value of +1 indicates that the object is in the correct cluster, a value of 0 indicates that the object is in her two clusters or ambiguous. On the other hand, a value of -1 indicates that the objects are clustered more precisely into another cluster.

Dunn Index

The Dunn index is a metric used to evaluate the results of the clustering algorithm [16]. The Dunn index calculates the compactness value and separation value which is formulated as following :

$$(C) = \min_{1\le j\le c, j\ne i} \left\{ \frac{dist(x_i, x_j)}{\max_{1\le k\le c}\{dian(x_k)\}} \right\} \quad \textbf{(4)}$$

where x is an object in cluster C. Good clustering results are indicated by the maximum dunnindex value.

Davies and Bouldin Index (DB Index)

The Davies-Bouldin Index is one way to measure the effectiveness of clusters, which allows you to look at how good or bad a cluster is in terms of cohesion and segregation. Cohesion is defined as the sum of object proximity to the cluster center of the tracked clusters, and separation is defined as the sum of object proximity based on the cluster center-to-cluster distance [14].The Davies-Bouldin exponent is formulated as

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i\neq j} (R_{i,j}) \quad \textbf{(5)}$$

where $R_{i,j}$ is the ratio value owned by each cluster i,j. The smaller the DBI value obtained DBI ≥ 0, the better the cluster obtained from grouping using the clustering algorithm.

Calinski and Harabasz Index (CH Index)

The Calinski-Harabasz (CH) validity index is an index that calculates a comparison of between-cluster sum of squares (SSB) and within-cluster sum of squares (SSW) values as separation. Compactness multiplied by a normalization factor, i.e. the difference beteen the number of data times the number of clusters and the number of clusters minus 1 [13]. A maximum value of the CH index indicates a good clustering result. Suppose you have a set with k clusters of data and at most N data points. For example, Ci is the lth cluster, xi is the ith point in the lth cluster, Ni is the number of points in the lth cluster, $\bar{x}_i$ if the point is central. The formula for the CH validity index for the lth cluster is:

$$CH = \frac{trace(SSB)}{trace(SSW)} \times \frac{N-k}{k-1} \quad \textbf{(6)}$$

$$SSW = \sum_{i=1}^{k} \sum_{x_i \in c_i} (x_i - \bar{x}_i)(x_i - \bar{x}_i)^T$$

$$SSB = \sum_{i=1}^{k} N_i (x_i - \bar{x}_i)(x_i - \bar{x}_i)^T$$

Good clustering will produce the maximum CH Index value.

Separation Index

The Separation Index is a partition validation index with a minimum distance separation measurement [24]. Separation index can be formulated as follows:

$$SI = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ik})^2 \|x_j - v_i\|^2}{N \min_{i,k} \|x_j - v_i\|^2} \quad \textbf{(7)}$$

The minimum SI index value indicates a good number of clusters.

Ratio of Average Within Cluster Standard Deviation to Betsialween Cluster Standard Deviation (S_w/S_b)

To check the homogeneity of the formed clusters and find out which method is most effective, you can calculate the ratio of the within-cluster mean standard deviation (s_w) to the between-cluster standard deviation (s_b) [11].The formula for the within-cluster mean standard deviation (s_w) is:

$$S_w = \frac{1}{K}\sum_{k=1}^{K} S_k \quad \textbf{(8)}$$

Where:

K : the number of clusters formed

Sk : standard deviation of the kth cluster

The standard deviation formula between clusters (sb) is as follows:

$$S_B = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\bar{X}_k - \bar{X})^2} \quad \textbf{(9)}$$

where

$\bar{X}_k$: rataan cluster ke-k

$\bar{X}$: rataan keseluruhan cluster

A good cluster is a cluster that is homogeneous among members in one cluster or has a minimum Sw value and is heterogeneous between clusters with one another or has a maximum Sb. This means that a good cluster is indicated by a minimum Sw/Sb ratio. The Sw/Sb ratio value indicates homogeneity within a cluster [17].

RESULT AND DISCUSSION

Descriptive Statistical Analysis

Table 2. Five Number Summary

	Percentage of poor people	Regional Special Allocated Fund	Life Expectancy	Number of health centers

Min	3.460	52.870	55.270	5.000
Median	14.445	162.650	65.650	17.000
Mean	17.310	179.700	67.280	21.470
Max	41.760	502.580	73.770	78.930

It can be seen in the results of **Table 2**, the median and mean values for the life expectancy attribute have almost the same value. So it can be said that the life expectancy attribute has a symmetric distribution. While the other 3 attributes, such as the number of poor people, the Regional Special Allocation Fund, and the number of Community Health Centers, are not symmetrical. The descriptive analysis table shows that on average, in eastern Indonesia there are as many as 17.31 thousand poor people living in each city-district. This Figure is quite high considering the large amount of regional budget issued by each district/city each year reaching 179.70 billion rupiah. With the high budget, the local government should be able to empower the community, starting from education, health and decent work. For the availability of infrastructure, on average, eastern Indonesia provides 21 health centers for each district/city. This number is of course very large, unfortunately there are still areas in eastern Indonesia with only 5 Number of health centers. For quality of life, eastern Indonesia has an average life span of around 50 to 70 years, this indicates that the health services provided at the Number of health centers are quite good.

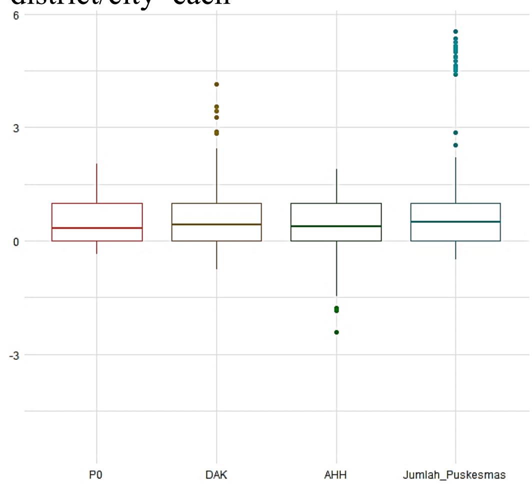


Figure 1. boxplot of research variables

In the **Figure.1** above, the boxplot output for all attributes is displayed. It can be seen that the Regional Special Allocation Fund attributes, life expectancy, and the number of Community Health Centers have outliers because they are outside the whisker line. For the life expectancy attribute, the outlier it has is a mild outlier because it is still in the range -1.5*IQR to -3*IQR. While the Regional Special Allocation Fund attributes, and the number of Community Health Centers have outliers with the types of mild outliers and extreme outliers because they are in the range of more than 3*IQR and less than -3*IQR.

KMO and MSA Assumptions Test

Before cluster analysis uses the characteristics of the factors that cause stunting, it is necessary to test the assumption of adequacy of the sample and that there is no multicollinearity. KMO and MSA tests can be performed to check sample adequacy.

Table 3. KMO and MSA Assumptions Test

Test Name	**Value**
Kaiser Meyer Okin (KMO)	0.507
Measures of Sampling Adequacy (MSA)	
Percentage of Poor People	0.509

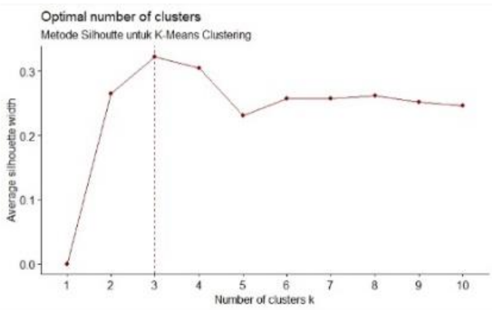
Regional Special Allocation Fund	0.508
Life Expectancy	0.508
Number of health centers	0.502

Based on the processing results, the KMO value is 0.507. KMO value > 0.5, so it can be concluded that the sample used is representative enough so that it is suitable for cluster analysis. In addition, in the MSA test, values are obtained for each

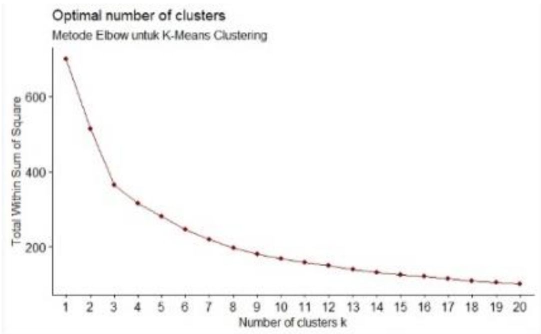
Grouping of Disadvantaged Regencies/Cities in Eastern Indonesia Region

K-means Method

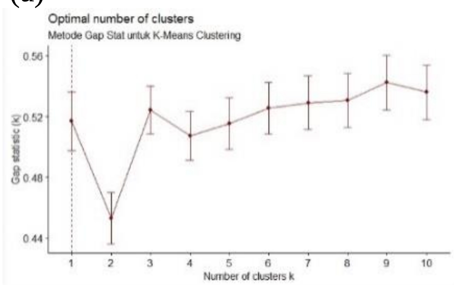
K-Means clustering method using elbows, silhouette scores, and gap statistics is used to determine the optimal number of clusters for grouping undeveloped districts/cities. The following table shows graphs of Elbow, Silhouette Score, and Gap Statistics by number of clusters of disadvantaged districts/cities in Eastern Indonesia formed using the K-means clustering method.



(a)



(b)



(c)

Figure 2: Silhoutte Score (a), Gap Statistics (b), and Elbow Method(c) for grouping of underdeveloped areas in Eastern Indonesia using K-means method

The silhouette score values **in Figure 2(a)** indicate that the highest silhouette score is cluster number 2. In contrast, in the silhouette method in **Figure. 2(b)** above, the lines form arcs in the number of clusters 3. The gap statistics shown in **Figure. 2(c)** show the gap values for cluster 2. From **Figure 2** above, we can conclude that LFA clustering of districts/cities in eastern Indonesia using the K-means clustering method works best when the number of clusters is 3.

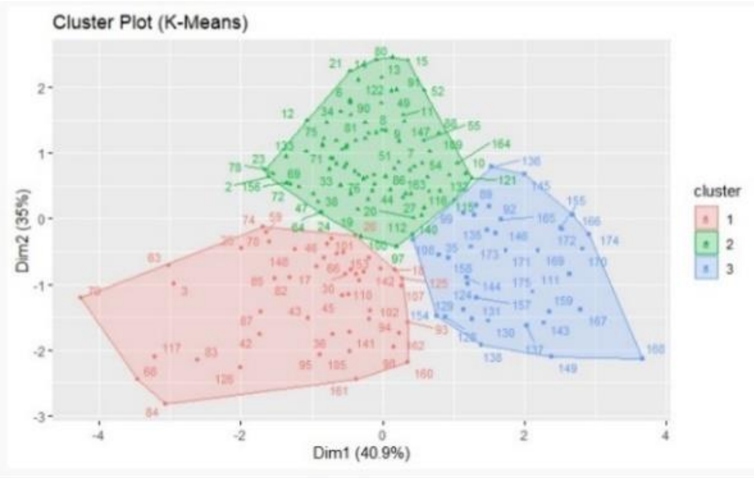


Figure 3. K-means Cluster Plot for grouping of underdeveloped areas in Eastern Indonesia

Figure 3 shows the visualization of the results of the k-means clustering algorithm on the distribution of district and city objects. The cluster that is formed is one of the very large clusters (cluster 1), followed by two smaller clusters, namely cluster 3 and cluster 2. The larger the cluster area indicates that there are more and more district and city objects in the cluster. However, in Figure 2, many regency and city objects are actually included in clusters 2 and 3. This is because there are outlier observation data in cluster 2, thus pulling the coverage of cluster 1 into a wider area.

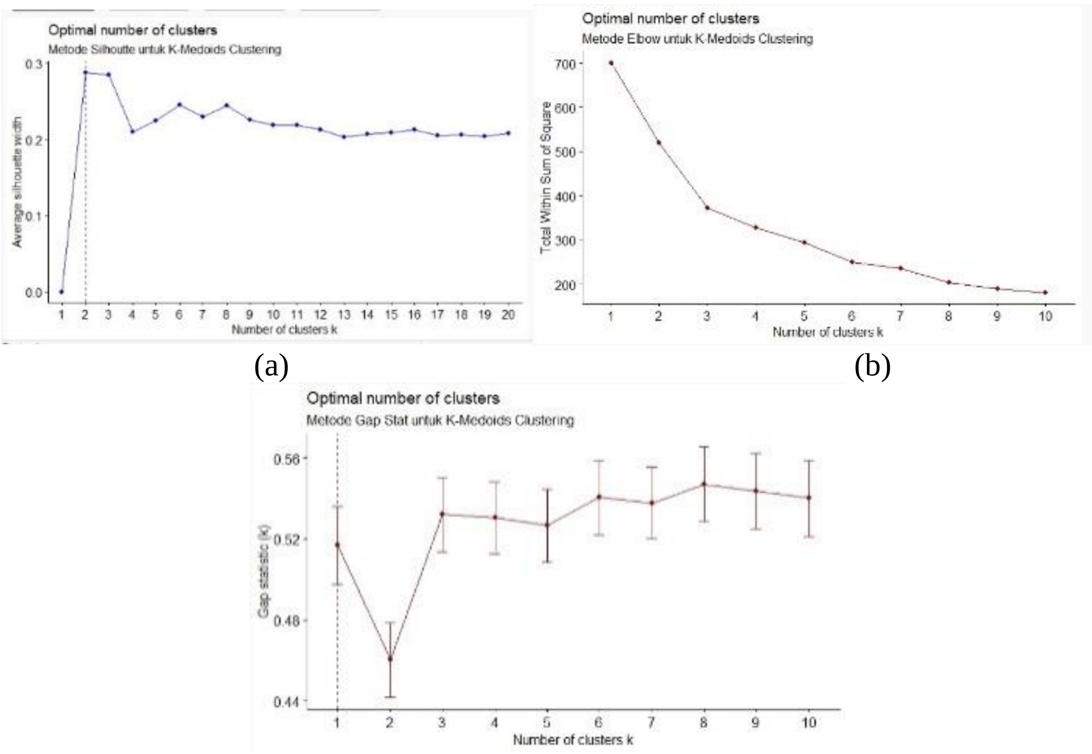
Table 4. Mean and Standard Deviation of each K-Means Clustering

Cluster	Variabel	Mean	Standard deviation
1	Percentage of Poor People	20.833	86.437
	Regional Special Allocation Fund	116.212	103.430
	Life Expectancy	70.032	68.300
	Number of health centers	10.794	100.350
2	Percentage of Poor People	63.559	49.827
	Regional Special Allocation Fund	35.084	51.078
	Life Expectancy	50.768	72.627
	Number of health centers	32.963	61.299
3	Percentage of Poor People	12.295	68.392
	Regional Special Allocation Fund	64.473	52.259
	Life Expectancy	41.847	35.069
	Number of health centers	54.876	74.737

Based on **Table 4**, it can be interpreted that cluster 3 is the district and city area that has the lowest level of underdevelopment or it can be interpreted as having the most rapid development between clusters 1 and 2. cluster 1 is the regency and city area which has a moderate level of regional development, and cluster 1 This is an area with a high level of development. Clusters with the k-means algorithm are depicted in **Figure 3**, where the distribution of this data is in a 2-dimensional plot.

K-medoids Method

The K-Medoids clustering method with elbow method, silhouette score, and gap statistics is used to determine the optimal number of clusters for grouping undeveloped districts/cities. The following table shows graphs of elbow, silhouette score and gap statistics by number of clusters of disadvantaged districts/cities in eastern Indonesia formed using the K-Medoids clustering method.



(a) (b)

(c)

Figure 4. Silhoutte Score (a), Gap Statistics (b), and Elbow Method(c) for grouping of underdeveloped areas in Eastern Indonesia using K-medoids method

From **Figure 4(a)** for the value of the Silhoutte Score, the highest Silhoutte Score is in the number of clusters 2. Meanwhile, in the Silhoutte Method in **Figure 4(b)** above, the line forms an elbow in the number of clusters 3. The Gap statistics shown in **Figure 4(c)** there is a gap value in cluster 2. From Figure 3 above, it can be concluded that grouping disadvantaged areas for regencies/cities in eastern Indonesia using the K-medoids Clustering method will be optimal if the number of clusters is 3.

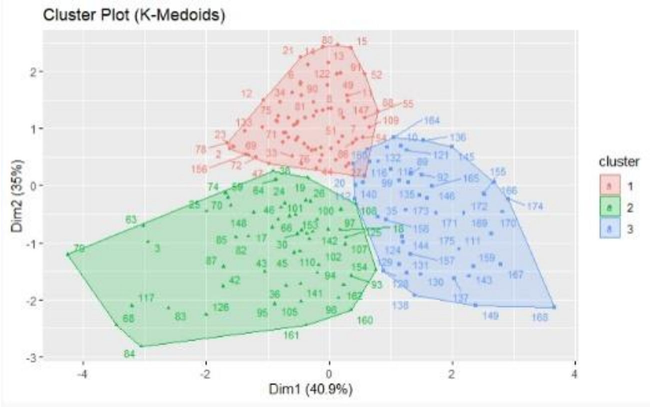


Figure 5. K-medoids Cluster Plot for grouping of underdeveloped areas in Eastern Indonesia

Figure 5 shows the visualization of the results of the k-medoids clustering algorithm on the distribution of district and city objects. Cluster 2 has a wider area, then the first cluster and the third cluster. The larger the cluster area indicates that there are more district and city objects in the cluster. However, in Figure 4, many district and city objects are included in clusters 1 and 3. This is because there are outlier observation data in cluster 2, thus pulling the coverage of cluster 2 into a wider area.

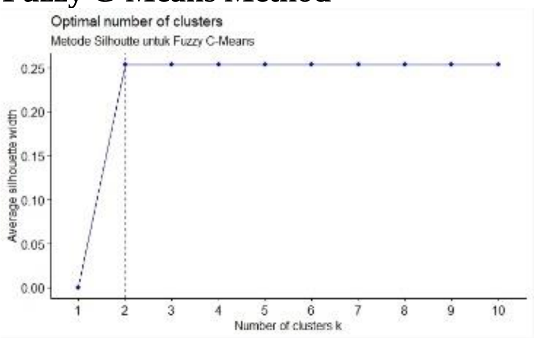
Table 5. Mean and Standard Deviation of each K-Medoids Clustering

Cluster	Variabel	Mean	Standard deviation
1	Percentage of Poor People	69.674	51.376
	Regional Special Allocation Fund	32.675	51.740
	Life Expectancy	76.017	53.679
	Number of health centers	36.759	59.498
2	Percentage of Poor People	17.407	88.983
	Regional Special Allocation Fund	101.309	103.342
	Life Expectancy	91.228	65.719
	Number of health centers	94.948	97.126
3	Percentage of Poor People	84.202	94.562
	Regional Special Allocation Fund	69.180	44.047
	Life Expectancy	103.258	89.352
	Number of health centers	55.633	69.817

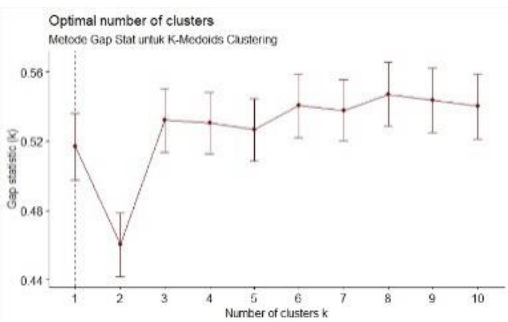
Based on **Table 5**, it can be interpreted that cluster 2 is a district and city area that has the lowest level of underdevelopment or it can be said to have the highest level of development, cluster 1 is a district and city area that has a moderate level of underdevelopment, and cluster 3 is an area with a high level of underdevelopment. low.

Fuzzy C-Means Method

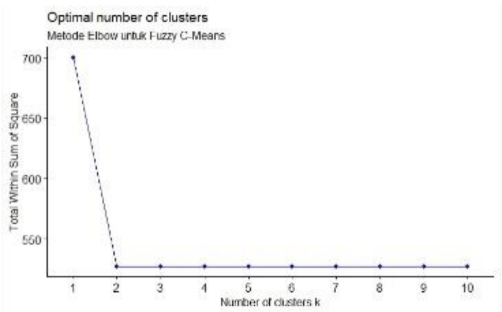
A fuzzy C-means clustering technique with elbows, silhouette scores, and gap statistics is used to determine the optimal number of clusters for grouping undeveloped districts/cities. The following table shows plots of elbow, silhouette score, and gap statistics by number of clusters of disadvantaged districts/cities in eastern Indonesia formed using the fuzzy C-means clustering method.



(a)



(b)



(c)

Figure 6 Silhoutte Score (a), Gap Statistics (b), and Elbow Method(c) for grouping of underdeveloped areas in Eastern Indonesia using K-medoids method

From **Figure 5(a)** for the value of the Silhoutte Score, the highest Silhoutte Score is found in the number of clusters 2. Meanwhile, in the Silhoutte Method in **Figure 5(b)** above, the line forms an elbow in the number of clusters 2. The Gap statistics shown in the **Figure 5(c)** there is a gap value in cluster 2. From Figure 1 above, it can be concluded that grouping disadvantaged areas for regencies/cities in eastern Indonesia using the Fuzzy C-Means Clustering method will be optimum if the number of clusters is 2.

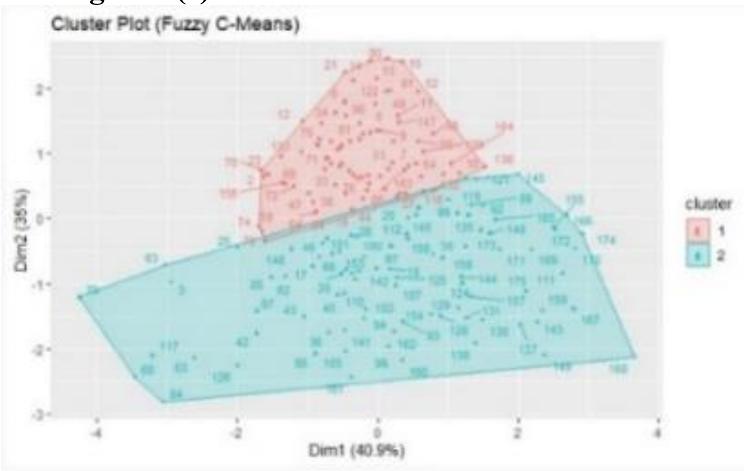


Figure 7. Fuzzy C-Means Cluster Plot for grouping of underdeveloped areas in Eastern Indonesia

Figure 6 shows a visualization of the results of the k-medoids clustering algorithm on the distribution of district and city objects. Cluster 2 (blue color) has a wider area and more coverage compared to cluster 1 (red color) which only covers a few areas. The larger the area clusters show that there are more and more district and city objects in the cluster. However, in **Figure 6,** many district and city objects are included in clusters 1 and 3.

Table 6. Mean and Standard Deviation of each Fuzzy C-Means Clustering

Cluster	Variabel	Mean	Standard deviation
1	Percentage of Poor People	68.526	46.736
	Regional Special Allocation Fund	29.387	60.233
	Life Expectancy	29.387	63.823
	Number of health centers	32.237	63.441
2	Percentage of Poor People	65.481	93.291
	Regional Special Allocation Fund	28.081	120.689
	Life Expectancy	57.967	120.689
	Number of health centers	30.804	117.7

Based on **Table 6,** it can be interpreted that cluster 1 is a district and city area that has the highest level of underdeveloped area indicators and cluster 2 is an area with low indicators of underdeveloped areas.

Best Model Selection

Table 7. Best Selection Model

Metode Evaluasi	Metode Clustering		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
Silhoutte Coefficients	0.322*	0.284	0.253
Dunn Index	0.080*	0.077	0.046
DB Index	1.251*	1.289	1.666
CH Index	79.880*	75.898	57.061
Separation Index	0.677	0.545*	0.6136
Sw/Sb Ratio	0.608*	0.635	0.732

The K-Means, K-Medoids, and Fuzzy clustering methods that have been carried out give different cluster results. Therefore, a comparison of the three methods was carried out to determine the best grouping results to be used. The evaluation method used is silhouette coefficient, Dunn index, DB index, CH index, separation index, and Sw/Sb comparison.

Table 7 displays the evaluation methods of the K-Means, K-Medoids, and Fuzzy C-Means methods. The results above show that the best method in this study is more dominant in K-Means. The silhouette coefficient value of the third method is above 0 so that it can be said that in general the clusters on objects from the third method tend to be clear.

In the Dunn index, K-Means produces the maximum Dunn index value of the other two methods so that the clustering results obtained can be said to be better than the other two methods. This is also similar to DBIndex where the K-Means method provides a minimum value so that the resulting grouping is better than the others. Based on the value of the CH Index, K-Means gives the maximum result which indicates that the clustering results obtained are better than the K-Medoids and Fuzzy C-Means methods.

When viewed from the separation index which is the minimum distance separation or partition validity, the K-Medoids method gives the minimum value of the other two methods or the groupings obtained are better. Different things are shown by the validation method with the average standard deviation within the cluster against the standard deviation between clusters where the minimum value is obtained from the K-Means method so that it can be said that with this method the homogeneity within the cluster is higher than the homogeneity within the cluster. obtained from the K-Medoids and Fuzzy C-Means methods. Based on this explanation, it can be said that the homogeneity of the cluster members formed using the K-Means method is greater than the K-Medoids and Fuzzy C-Means methods. In addition, the dissimilarity between clusters formed is greater than the other two methods. The data used has many outliers. However, based on this, even though the K-Means method is not robust against outliers, the resulting clustering is better than the K-Medoids and Fuzzy C-Means methods in grouping regencies/cities in eastern Indonesia according to grouping underdeveloped regions.

Interpretation of The Best Clustering Results

The results of clustering using the k-means method produce three clusters with districts/cities which are grouped in **table 8.** Based on the output of **table 8.** It is concluded that cluster 1 consists of 91 districts/cities with very not lagging status (areas with high development), clusters 2 consists of 48 regencies/cities with medium underdeveloped status (regions with a moderate level of development), and cluster 3 consists of 37 regencies/cities with underdeveloped status (regions with a low level of development). Based on visual observations of the clustering results in Figure 1, it shows that cluster 2 is the largest cluster, this shows that most of the regions in Eastern Indonesia are classified as having middle lagging status.

Based on **table 4** ,the first cluster is an area with a group of regencies/cities that are not left behind (regions with a high level of development), characterized by a low poverty rate, high realization of the regional development budget, and long life expectancy. However, the problem in these districts/cities is the lack of basic health facilities such as health centers. On average, the areas categorized as the first cluster only have ten to eleven health centers in each district/city. This is important to be followed up as a regional government material regarding the realization of the regional budget for health facilities to support life expectancy.

The second cluster is a group of regions with a middle level of underdevelopment (regions with a moderate level of development), characterized by high levels of poverty, low realization of regional budget funds, adequacy of health centers starting to be reached, but life expectancy is still at the age of fifty years. The area belonging to the second cluster has a relatively high number of poor people, namely around 63 thousand people for each urban district. This could be due to the lack of realization of regional budget funds to create real development programs to improve people's welfare. In areas covered in the second cluster, life expectancy is only around 50 years old, this Figure is quite low when compared to the life expectancy in the first and third clusters. namely almost 32 health centers in each district/city, the health services provided in the second cluster are quite low and need to be reviewed by the local government.

The third cluster is an area that is classified as a disadvantaged area (a region with a low level of development), characterized by various complex problems such as high poverty, a high number of health centers but not matched by good quality health services which is illustrated by low life expectancy and low regional budget realization.

The first cluster is a group of districts/cities with a high expectation of school years and low poverty. There are no other dominant indicators in this cluster. Based on this, it can be said that this cluster is a group of regencies/cities with a moderate level of community welfare.

In this study, overlapping between clusters was successfully handled by the k-means method. In the k-means method there is a clear dividing line between clusters, meaning that there is no intersection between clusters and you are able to get high optimum clusters with few variables. In contrast to k-medoids, even though it consists of three classes, there are still overlapping clusters, namely cluster one and cluster three, this indicates that k-medoids has not been able to form clusters with outlier data problems. Outlier data has a tendency to draw the outer boundaries of a cluster so that it coincides with other clusters. A different problem is found in the fuzzy c-means method, namely the optimum number of clusters is only two. This can be a problem because it is unable to explain the problems that occur in certain areas with relatively moderate development and tends to classify areas with medium development to areas with high or short development. This can later lead to confusion in establishing policies regarding the areas where the problem must be resolved.

Table 7. Result of clustering

Cluster	Region
1	Kepulauan Sangihe, Gorontalo, Pohuwato, Poso, Donggala, Parigi, Moutong ,Tojo, Una-Una, Polewali Mandar, Muna, Konawe, Konawe Selatan, Bombana, Kabupaten Lombok Barat, Kabupaten Lombok Tengah, Kabupaten Lombok Timur, Kabupaten Sumbawa, Kabupaten Bima, Bulukumba, Jeneponto, Gowa, Pangkeajene ,Bone, Wajo, Luwu ,Makassar, Sumba Timur, Kupang ,Timor Tengah, Selatan Timor, Tengah Utara, Alor ,Sikka, Ende ,Manggarai ,Rote, Ndao, Manggarai Barat, Sumba,Tengah, Manggarai Timur, Halmahera Selatan, Maluku Tenggara, Maluku Tengah ,Sorong, Raja Ampat, Merauke ,Nabire, Biak, Numfor, Yahukimo, Pegunungan, Bintang ,Tolikara
2	Halmahera Timur, Kepulauan Muna Barat, Halmahera Barat,Luwu Utara, Luwu Timur,Kolaka, Kota,Manado, Kota Gorontalo,Polopo, Manokwari ,Pulau Talibu,Gorontalo Utara, Kota Mataram, Sumbawa Barat,Kota Jayapura,Ambon, Bolaang Mongondow,Morowali, Konawe,Ternate, Kota Sorong, Kepulauan Talaud,Kolaka, Konawe Utara, Bolaang Mongondow Timur, Kota Bima,Yapen,Minahasa Selatan,Enrekang, Ngada, Pulau Morotai, Minahasa,Sorong Selatan, Kepulauan Sitaro, Tidore Kepulauan. Kolaka Timur, Keerom, Barru,Buol, Bolaang Mongondow Utara, Tolitoli, Halmahera Utara, Buru-Buru, Kota Bau-Bau, Sami, Malaka, Kota Bitung, Buton Utara, Soppeng, Kabupaten Belu, Kepulauan Mimika, Minahasa Tenggara , Tana Toraja, Halmahera Tengah, Kepulauan Sula Selatan, Kepulauan Banggai , Bolaang Mongondow Selatan, Kolaka Utara, Pinrang, Kepulauan Selayar, Minahasa Utara, Bantaeng, Maros, Nagekeo, Mamuju, Buton Selatan, Bone , Mamuju Tengah, Kolaka, Boalemo, Kabupaten Dompu, Pasangkayu, Banggai, Kota Kendari , Kota Kupang, Buton Tengah, Sidenreng, Kota Kotamobagu, Jayapura, Palu , Morowali Utara, Pare Pare, Toraja Utara, Flores Timur , Banggai, Buton, Sinjai, Barru Rappang, Sigi, Laut Mamasa, Kota Tomohon, Wakatobi Takalar
3	Kabupaten Lembata, Intan Jaya Deiyai, Sumba Barat, Mamberamo Raya, Nduga, Mamberamo Tengah, Raijua, Maluku Barat Daya, Waropen , Seram Bagian Barat, Teluk Wondama, Boven Digoel , Puncak Jaya, Kepulauan Aru, Pegunungan Arfak, Lombok Utara , Teluk Bintuni, Supiori, Fakfak, Sumba Barat Daya, Manokwari Selatan, Mappi, Asmat, Lombok Utara, Maluku Tenggara Barat, Maybrat, Jayawijaya, Lanny Jaya, Sabu, Puncak Dogiyai, Tambrauw, Maybrat, Seram Bagian Timur,Tual ,Kaimana ,Paniai Yalimo

CONCLUSION AND SUGGESTION

Clustering indicator data for disadvantaged areas in eastern Indonesia with outliers using the k-means algorithm is more optimal than grouping using the k-medoids or fuzzy c-means algorithms. The k-means algorithm can obtain k optimum or in this case the optimum number of clusters is quite high, namely as many as three clusters. In addition, k-means is able to avoid outlier problems which can cause an expansion of boundaries between clusters which results in overlapping between clusters.

The results of grouping using the k-means method from 176 regencies in eastern Indonesia based on indicators of underdeveloped areas into 3 clusters with different characteristics and problems for each cluster, namely the first cluster, namely the group of areas not lagging behind (regions with different levels of development high) of 91 districts/cities, characterized by high realization of regional funds, high life expectancy, low poverty rate, and few health centers in each district/city. The second cluster is a group of middle-disadvantaged regions (regions with a moderate level of development) of 48 regencies/cities, characterized by low realization of regional funds, a high number of poor people, and life expectancy that is disproportionate to the number of basic health services such as Number of health centers. namely the group of underdeveloped regions (areas with low levels of development) of 37 districts/cities which are characterized by various problems of low realization of regional funds, high levels of poverty, and low life expectancy.

With the formation of clusters of underdeveloped regions in Eastern Indonesia, it is hoped that it will become a picture for the government to make policies that are right on target in accordance with the problems that exist in each region. In addition, the government can focus on areas that must be assisted first in order to avoid various kinds of crises that might occur.

REFERENCES

Afira, N., & Wijayanto, A. W. (2021). Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019. *Komputika: Jurnal Sistem Komputer*, *10*(2), 101-109.

Aranganayagi, S., & Thangavel, K. (2007, December). Clustering categorical data using silhouette coefficient as a relocating measure. In *International conference on computational intelligence and multimedia applications (ICCIMA 2007)* (Vol. 2, pp. 13-17). IEEE.

Badan Pusat Statistik. (2015). Indikator Pembangunan Berkelanjutan. Jakarta: Badan PusatStatistik.

Batool, F., & Hennig, C. (2021). Clustering with the average silhouette width. *Computational Statistics & Data Analysis*, *158*, 107190.

BPKP. (2021). *Pedoman Praktis Membangun Indeks Komposit.* Jakarta: BPKP.

El-Mandouh, A. M., Abd-Elmegid, L. A., Mahmoud, H. A., & Haggag, M. H. (2019). Optimized K-means clustering model based on gap statistic. *International Journal of Advanced Computer Science and Applications*, *10*(1).

Gujarati, D. N. (1995). Basic Econometrics (4th ed.). United State Military Academy.

Hardiyanti, F., Tambunan, H. S., & Saragih, I. S. (2019). Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, *3*(1).

J.F. Hair, Black, W. C., Babin, W. J., Anderson, R. E., & Tatham, R. L. (2006).Multivariate Data Analysis 6th Edition. Pearson Prentice Hall.

Johnson, R. A., Wichern, D. W., & others. (2002). Applied multivariate statistical analysis(Vol.Prentice hall Upper Saddle River, NJ.

Laeli, S. (2014). *Analisis Cluster dengan Average Linkage Method dan Wad's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.

Liu, F., & Deng, Y. (2020). Determine the number of unknown targets in open world based on elbow method. *IEEE Transactions on Fuzzy Systems*, *29*(5), 986-995.

Luthfi, E., & Wijayanto, A. W. (2021). Analisis Perbandingan Metode Hirearchical, K-Means, dan K-Medoids Clustering dalam Pengelompokkan Indeks Pembangunan Manusia Indonesia. *Jurnal FEB Unmul*, 761-773.

Medela, A., & Picon, A. (2020). Constellation loss: Improving the efficiency of deep metric learning loss functions for the optimal embedding of histopathological images. *Journal of Pathology Informatics*, *11*(1), 38.

Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal media infotama*, *11*(2).

Ncir, C. E. B., Hamza, A., & Bouaguel, W. (2021). Parallel and scalable Dunn

Index for the validation of big data clusters. *Parallel Computing*, *102*, 102751.

Qiu, W., & Joe, H. (2006). Separation index and partial membership for clustering. *Computational statistics & data analysis*, *50*(3), 585-603.

Rahakbauw, D. L., Ilwaru, V. Y., & Hahury, M. H. (2017). Implementasi fuzzy c-means clustering dalam penentuan beasiswa. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, *11*(1), 1-12.

Republik Indonesia. (2020). *Peraturan Presiden Republik Indonesia Nomor 63Tahun 2020 Tentang Penetapan Daerah Tertinggal Tahun 2020-2024.* Jakarta: Sekretariat Negara.

Sari, E. A. (2020). Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vektor Machine. *Jurnal Informatika dan Komputer*, 188-195.

Santoso, E. B., & Rohmawati, A. N. (2017). Penentuan Cluster Pengembangan Komoditas Unggulan Desa-Desa Tertinggal Di Kabupaten Bangkalan Berdasarkan Faktor-Faktor Yang Mempengaruhi. *Jurnal Teknik ITS*, *6*(2), C334-C338.

Setiadi, D. H. (2018). Pemetaan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Daerah Tertinggal Dengan Metode Data Campuran Ensamble ROCK dan SWFM.

Siagian, S. (2008). Administrasi Pembangunan. Jakarta: PT. Bumi Aksara

Siallagan, J., Nasiha, L. L., & Wijaya, Y. T. (2021). Analisis Cluster dengan Data Outlier untuk Pengelompokkan Daerah di Provinsi Jawa Barat Berdasarkan Risiko Bencana Banjir. Politeknik Statistika STIS.

Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. JURNAL MATEMATIKA MANTIK, 4(1), 22–31.

Sjukani, M. (2012). Struktur Algoritma dan Struktur Data 2 Edisi 3. *Jakarta: Mitra Wacana Media*.

Thamrin, N., & Wijayanto, A. W. (2021). Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island: Analisis cluster dengan menggunakan hard clustering dan soft clustering untuk pengelompokkan tingkat kesejahteraan kabupaten/kota di pulau Jawa. *Indonesian Journal of Statistics and Its Applications*, *5*(1), 141-160.

Wakhidah, N. (2010). Clustering menggunakan k-means algorithm. *Jurnal Transformatika*, *8*(1)