BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0063–0072
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0063-0072

CLUSTERING DISTRICTS/CITIES IN EAST JAVA PROVINCE BASED ON HIV CASES USING K-MEANS, AGNES, AND ENSEMBLE

Dwi Ayu Lusia[1*], Imelda Salsabila[2], Heni Kusdarwati[3], Suci Astutik[4]

[1,2,3,4]*Department of Statistics, Faculty of Mathematics and Science, Universitas Brawijaya*
Jl. Veteran, Malang, 65145, Indonesia

*Corresponding author's e-mail: * **dwiayulusia@ub.ac.id***

Article History:

Received: 5th January 2024
Revised: 30th March 2024
Accepted: 16th October 2024
Published: 13th January 2025

Keywords:

AGNES;
Clustering;
Ensemble;
HIV;
K-means

ABSTRACT

Cluster analysis is a method of grouping data into certain groups based on similar characteristics. This research aims to group districts/cities in East Java Province in 2021 based on HIV cases using hierarchical cluster analysis (AGNES), non-hierarchical cluster analysis (K-means), and ensemble clustering. The study found that the ensemble clustering solution forms four clusters, consistent with the results of AGNES clustering. This suggests that ensemble clustering improves the quality of cluster solutions by leveraging both hierarchical and non-hierarchical methods. The grouping of districts/cities based on HIV cases provides a clear distribution pattern for more targeted interventions. The study is limited to HIV cases in East Java Province and may not be generalizable to other regions with different epidemic characteristics. Additionally, the study focuses on clustering methods without investigating temporal changes in HIV case distribution. This research is one of the few studies that applies ensemble clustering to HIV cases in East Java Province. It combines hierarchical and non-hierarchical methods to improve the clustering process and provides a practical approach for regional HIV control planning.

How to cite this article:

D. A. Lusia, I. Salsabila, H. Kusdarwati and S. Astutik., "CLUSTERING DISTRICTS/CITIES IN EAST JAVA PROVINCE BASED ON HIV CASES USING K-MEANS, AGNES, AND ENSEMBLE", *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0063-0072, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Cluster analysis is a method of grouping data into certain groups based on their characteristics [1]. Data with the same characteristics will be in the same cluster, while data with different characteristics will be in another cluster. Clusters are characterized as having high homogeneity between members in one cluster and high heterogeneity between one cluster and other clusters. Cluster analysis is divided into two methods, namely hierarchical and non-hierarchical. The hierarchical method is grouping data in a structured manner without determining the number of clusters to be formed, while the non-hierarchical method is grouping data by determining the clusters to be formed first [2]. When several methods are combined, a new method will be formed which is called a cluster ensemble. The combining the grouping results of several methods to obtain a combined solution which will be used as the final solution is called cluster ensemble [3]. Cluster ensemble has advantages compared to other methods, namely being able to improve the quality of cluster solutions [4]. Another research about grouping districts/cities in South Sulawesi Province based on regional economic development performance indicators and producing three clusters. These two studies also stated that ensemble clusters provide better grouping results compared to hierarchical and non-hierarchical methods [5].

Grouping an area can be done based on certain characteristics, one of which is the spread of the virus. Human Immunodeficiency Virus (HIV) is a virus that attacks white blood cells, causing a decrease in the body's immunity. HIV transmission occurs due to the transfer of body fluids, such as blood, breast milk, semen, or vaginal fluids from an infected person to a healthy person. This fluid can be transferred through sexual intercourse, using the same syringe, blood transfusion from an infected donor, and from mother to child during pregnancy, childbirth, or breastfeeding. People who are susceptible to infection through sexual contact are female sex workers, transgender women, men who have sex with men, customers of sex workers, and high-risk couples [6]. People will not become infected with HIV through everyday contact, such as shaking hands, hugging, or sharing personal objects.

The United Nations Program on HIV and AIDS (UNAIDS) noted that in 2019, the number of HIV sufferers in Southeast Asia reached 3.8 million people. The high number of HIV sufferers in Southeast Asia requires Indonesia to be more vigilant about the spread of the HIV virus. Based on data from the Indonesian Ministry of Health's HIV/AIDS Development Situation Report in 2019, the number of HIV sufferers in Indonesia reached 50,282 cases spread across various provinces. East Java is the province with the highest number of HIV cases, namely 8,935 cases, followed by DKI Jakarta and West Java. HIV sufferers in East Java will likely continue to increase because it is an area with developed industry and transportation. HIV control needs to be carried out effectively so that it does not become a big problem in the future.

The World Health Organization (WHO) determines the epidemic level to describe the situation of HIV/AIDS development, namely low, concentrated and widespread. Since 2003, East Java has been designated as a concentrated epidemic area [7]. Determining the epidemic level is carried out to maximize HIV control, but this method is only applied to provincial areas. It is necessary to carry out an analysis to group district/city areas so that HIV control in East Java becomes faster because it can be handled by the district/city government itself. Previous research about clustering districts/cities based on HIV cases using hierarchical clustering in China [8], South Carolina [9], and two sub-Saharan African region [10]. Apart from using hierarchical clustering, clustering districts/cities based on HIV cases can use non-hierarchical clustering (using K-means) i.e. Dewi and Voutama [11], Rosida and Wijaya [12], Munawar and Purnamasari [13]. Ensemble clustering on HIV cases was studied by Lynch and DeGruttola in Botswana [14]. However, none of these studies used East Java Province as a research case. Therefore, this research was conducted to group districts/cities in East Java Province based on HIV cases using hierarchical cluster (AGNES), non-hierarchical cluster (K-means) and ensemble cluster analysis.

2. RESEARCH METHODS

The research used quantitative methods, i.e., K-means, AGNES, and ensemble clustering. The data used is secondary data sourced from the East Java Provincial Health Service regarding the 2021 HIV Development Situation Report data. The data consists of 38 districts/cities in East Java Province with variables namely the number of HIV cases in Female sex workers ($\boldsymbol{x_1}$), transvestites ($\boldsymbol{x_2}$), Male sex men ($\boldsymbol{x_3}$), High risk couples ($\boldsymbol{x_4}$), Customers of public sex workers ($\boldsymbol{x_5}$), IDU ($\boldsymbol{x_6}$), and Pregnant women ($\boldsymbol{x_7}$). The steps of analysis were as follows:

1. Exploration data in the form of tables or images that describe the data
2. Pearson correlation test

Pearson correlation is used to see the closeness of the linear relationship between 2 variables that have an interval or ratio data scale. The correlation calculation is formulated in **Equation (1)**.

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad (1)$$

The correlation coefficient can be positive or negative in the range -1 to 1. If the value is close to -1 or 1 then the correlation between the two variables is stronger and vice versa.

3. K-means

The grouping process with K-means is as follows **[15]**

a. Determine the number of clusters.

b. Divide the data randomly for each cluster and determine the centroid value in each cluster with equation:

$$\bar{x}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ik} \quad (2)$$

Where $\bar{x}_k$ is Average of the kth cluster, n_k is Number of objects in the kth cluster, and x_{ik} is Value of the i-th object in the kth cluster

c. Calculate the distance of each data to the centroid value. Distance using Mahalanobis as follow:

$$d_{ij} = \sqrt{(x_i - x_j)^T S^{-1}(x_i - x_j)} \quad (3)$$

Grouping data by taking into account the minimum distance.

d. Recalculate the centroid value for the newly formed cluster.

e. Repeat step c. until the resulting centroid value remains constant and cluster members do not move to other clusters.

4. AGNES

AGNES or agglomerative is bottom-up approach of hierarchical clustering with each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. The AGNES algorithm as follow: **[16]**

a. Assume each tuple is a cluster.

b. Calculate the distance between objects using Mahalanobis distance

c. Choose the smallest distance, then combine the two objects.

d. Calculate the centroid: $\bar{x}_{ij} = \frac{x_i + x_j}{2}$

e. Calculate the new distance containing the combined cluster.

f. Repeat the second step until all data is combined into one single cluster.

5. Ensemble

Cluster ensemble is a method of grouping data by combining the results of grouping several different methods so that a combined solution is obtained which is used as the final solution **[3]**. Grouping with ensemble clusters is carried out in two stages, namely as follows.

a. Form ensemble members from several different grouping methods.

b. Combining all ensemble members to obtain one final solution called the consensus function (one of them is pairwise similarity where a similarity matrix will be formed). In forming the similarity matrix, the Link-Based Cluster Ensemble algorithm is used as follow:

i. Label ensemble cluster members.

ii. Form a Binary Association Matrix (a matrix of size $\boldsymbol{n \times p}$, where n is the row that describes the number of objects and q is the column that describes the number of labels) If the member corresponds to the label, then the object has a value of 1, and otherwise it has a value of 0.

iii. Form a weighting matrix which is as follows.

1. Form a matrix w with **Equation (4)**

$$w_{C_{ki}^i C_{kj}^j} = \frac{\left| X_{C_{ki}^i} \cap X_{C_{kj}^j} \right|}{\left| X_{C_{ki}^i} \cup X_{C_{kj}^j} \right|} \tag{4}$$

with $\boldsymbol{X_{C_{ki}^i}}$ is the set of objects in the ith cluster and $\boldsymbol{X_{C_{kj}^j}}$ is the set of objects in the jth cluster, with $\boldsymbol{i \neq j}$

2. Form a WCT (Weighted Connected-Triple) matrix with **Equation (5), (6),** and **(7).**

$$WCT_{C_{ki}^i C_{ki*}^{i*}}^{C_{kj}^j} = \min\left(w_{C_{ki}^i C_{kj}^j}, w_{C_{ki*}^{i*} C_{kj}^j}\right) \tag{5}$$

$$WCT_{C_{ki}^i C_{ki*}^{i*}} = \sum_{kj=1}^{q} WCT_{C_{ki}^i C_{ki*}^{i*}}^{C_{kj}^j} \tag{6}$$

$$Sim^{WCT}\left(C_{ki}^i, C_{ki*}^{i*}\right) = \frac{WCT_{C_{ki}^i C_{ki*}^{i*}}}{WCT_{max}} \times DC \tag{7}$$

where $\boldsymbol{k}$ is Label index, $\boldsymbol{k = 1, 2, \ldots, q}$, $\boldsymbol{WCT_{max}}$ is Highest value on the WCT matrix, and $\boldsymbol{DC}$ is Confidence level accepts two non-identical objects as two similar objects, $\boldsymbol{DC \in (0, 1]}$

3. Form a similarity value between labels with **Equation (8)**.

$$RM(x_i, C_l) = \begin{Bmatrix} 1, & \text{if } C_l = C(x_i) \\ Sim^{WCT}\left(C_{ki}^i, C_{ki*}^{i*}\right), & \text{others} \end{Bmatrix} \tag{8}$$

where $\boldsymbol{C(x_i)}$ is Label of the $\boldsymbol{i}$-th object, $\boldsymbol{C_l}$ is other labels, and $\boldsymbol{Sim^{WCT}\left(C_{ki}^i, C_{ki*}^{i*}\right)}$ is the similarity value between the ith object label and other labels.

4. The CTS (Connected-Triple Based Similarity) similarity matrix can be obtained as follows.

$$CTS_{(x_i, x_j)} = \frac{1}{M} \sum_{m=1}^{M} RM(x_i, C_l) \tag{9}$$

with $\boldsymbol{M}$ is the number of cluster methods used.

5. Transform the similarity matrix into a distance matrix with **Equation (10)**.

$$d_{ij} = 1 - CTS_{(x_i, x_j)} \tag{10}$$

6. Group to get the final solution.

6. C-index

C-Index is a validation index for cluster results with internal criteria [17]. The C-Index value is in the range 0 to 1, the optimal cluster is indicated by the minimum value of the C-Index. The C-Index calculation is formulated in Equation (11).

$$C_{Index} = \frac{T_w - T_{min}}{T_{max} - T_{min}} \tag{11}$$

where $\boldsymbol{T_w}$ is the number of distances between objects in each cluster, $\boldsymbol{T_{min}}$ is The smallest number of distances between objects in each cluster, and $\boldsymbol{T_{max}}$ is The largest number of distances between objects in each cluster

3. RESULTS AND DISCUSSION

A graph that can describe the distribution of each type of HIV/AIDS cases can be depicted as Figure 1, showed that there are outliers for each type of HIV/AIDS. The variable female sex workers (x_1) contains 3 outliers, namely in Banyuwangi, Kediri and Situbondo. The transvestites variable (x_2) contains 2 outliers, namely in Banyuwangi and Malang. The male sex men variable (x_3) contains 6 outliers, namely in Banyuwangi, Kediri, Kediri City, Malang City, Surabaya City, and Sidoarjo. The high risk couple variable (x_4) contains 2 outliers, namely in Banyuwangi and Malang City. The Customers of public sex workers variable (x_5) contains 1 outlier, namely in Kediri. The IDU variable (x_6) contains 8 outliers, namely in Banyuwangi, Jombang, Kediri, Malang City, Surabaya City, Madiun, Pasuruan, and Sidoarjo. The pregnant women variable (x_7) contains 2 outliers, namely in Pasuruan and Tuban.

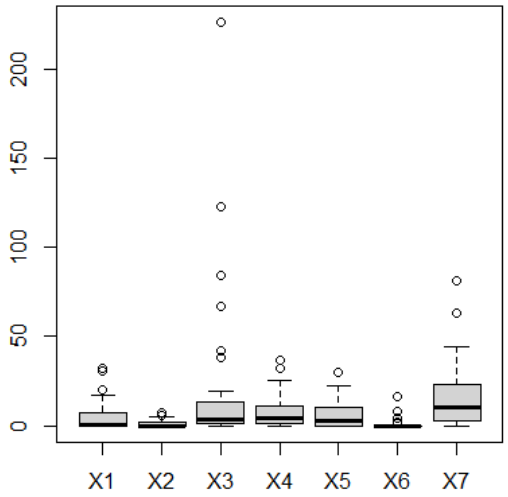

Figure 1. Boxplot of 7 Type of Cases HIV/AIDS

Correlation value testing is used to fulfill multicollinearity assumptions before conducting cluster analysis. The correlation for each combination of variables is calculated using Equation (1). The following hypothesis and correlation test results can be seen in Table 1.

$H_0 : \rho = 0$, There is no relationship between the two variables.
$H_1 : \rho \neq 0$, There exists a relationship between the two variables.

Table 1. Correlation Between Variables

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1	0.371	0.487	0.621	0.681	0.294	-0.002
	0	0.022	0.002	3.13×10^{-5}	2.54×10^{-10}	0.073	0.99
X_2	0.371	1	0.436	0,692	0.457	0.351	0.145
	0.022	0	0.006	1.52×10^{-6}	0.004	0.031	0.384
X_3	0.487	0.436	1	0.705	0.517	0.454	-0.002
	0.002	0.006	0	7.78×10^{-7}	0.001	0.004	0.991
X_4	0.621	0,692	0.705	1	0.756	0.546	0.144
	3.13×10^{-5}	1.52×10^{-6}	7.78×10^{-7}	0	3.99×10^{-8}	0.001	0.387

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_5	0.681	0.457	0.517	0.756	1	0.336	0.114
	2.54×10^{-10}	0.004	0.001	3.99×10^{-8}	0	0.039	0.496
X_6	0.294	0.351	0.454	0.546	0.336	1	0.185
	0.073	0.031	0.004	0.001	0.039	0	0.265
X_7	-0.002	0.145	-0.002	0.144	0.114	0.185	1
	0.99	0.384	0.991	0.387	0.496	0.265	0

Description: Row 1 shows the correlation and row 2 shows the p-value

Based on **Table 1**, we reject H_0 and Highly correlated so Mahalanobis distance is used. If the number of clusters is 2, 3, and 4, and the c-index value (**Table 1**) is used to determine the number of clusters, then the AGNES method has 4 clusters and K-means has 3 clusters.

Table 2. C-Index for AGNES and K-means

Cluster	AGNES	K-means
2	0.236	0.2865
3	0.211	0.2309
4	0.205	0.2562

Dendrogram for AGNES show in **Figure 2**. That figure show that cluster 1 have 35 member, i.e. Bangkalan, Banyuwangi, Blitar, Bojonegoro, Bondowoso, Gresik, Jember, Jombang, Kota Batu, Kota Blitar, Kota Kediri, Kota Madiun, Kota Malang, Kota Mojokerto, Kota Pasuruan, Kota Probolinggo, Lamongan, Lumajang, Madiun, Magetan, Malang, Mojokerto, Nganjuk, Ngawi, Pacitan, Pamekasan, Ponorogo, Probolinggo, Sampang, Sidoarjo, Situbondo, Sumenep, Trenggalek, Tuban, Tulungagung, cluster 2 have 1 member (Surabaya city), cluster 3 have 1 member (pasuruan), and cluster 4 have 1 member (Kediri).

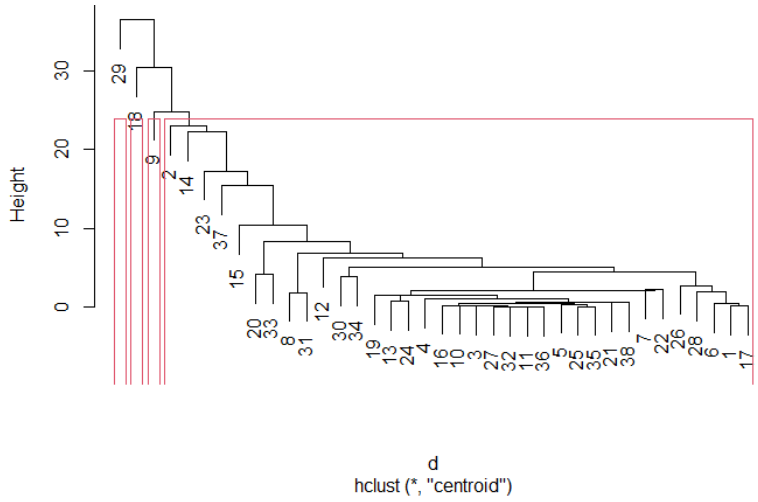

Figure 2. Dendogram of AGNES

K-means has 3 cluster. First cluster have 24 members, i.e. Banyuwangi, Blitar, Bondowoso, Jember, Jombang, Kediri, Kota Batu, Kota Blitar, Kota Kediri, Kota Pasuruan, Magetan, Nganjuk, Pacitan, Sampang, Situbondo, Sumenep, Trenggalek, Bojonegoro, Lamongan, Madiun, Tulungagung, Ponorogo, Kota Madiun, Mojokerto. Second cluster have 6 members i.e. Kota Malang, Kota Mojokerto, Kota Surabaya, Lumajang, Malang, Sidoarjo. Third cluster have 8 members, i.e. Bangkalan, Gresik, Kota Probolinggo, Ngawi, Pamekasan, Pasuruan, Probolinggo, Tuban. Center value calculated using **Equation (2)** for K-means show in **Table 2**.

Table 2. Center Value for K-means

cluster	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	2.6	1.6	20.3	6.5	4.4	1.7	23.3
2	8.0	0.0	3.0	8.0	7.0	0.0	13.0
3	5.3	1.3	13.7	6.3	5.7	0.3	21.0

Based on Table 2 show that cluster 1 have highest variable in transvestites (x_2), male sex men (x_3), IDU (x_6), and pregnant women (x_7), cluster 2 have highest variable in female sex workers (x_1), high risk couples (x_4) and Customers of public sex workers variable (x_5), cluster 3 have lower center value than other clusters.

AGNES method has 4 clusters and K-means has 3 clusters are used as ensemble members. Member of each cluster for AGNES and K-means show in Table 3

Table 3. Crosstabulation Member AGNES and K-means

K-means	AGNES			
	1	2	3	4
1	23	0	0	1
2	5	1	0	0
3	7	0	1	0

Table 3 show that AGNES cluster 2 is cluster 2 in K-means, AGNES cluster 3 is cluster 3 in K-means, and AGNES cluster 4 is cluster 1 in K-means.

The relationship between AGNES and K-means show in WCT matrix calculated using Equation (6) in Figure 2. It can be seen that for the relationship between clusters in the AGNES there is a relationship between cluster 1 and clusters 2 (0.139), 1 and 3 (0.125), 1 and 4 (0.042), but there is no relationship between cluster 2 and clusters 3 and 4, and cluster 3 and cluster 4. Meanwhile, in the K-means there is the relationship between cluster 1 with clusters 2 (0.139), 1 and 2 (0.194), 2 and 3 (0.139). For relationships between cluster methods calculated using Equation (4) that have relationships, namely AGNES cluster 1 with clusters 1 in K-means (0.639), AGNES cluster 1 with cluster 2 in K-mean (0.139), AGNES cluster 1 and cluster 3 in K-means (0.194), AGNES cluster 2 with cluster 2 in K-means (0.167), AGNES cluster 3 with cluster 3 in K-means (0.125), AGNES cluster 4 with cluster 1 in K-means (0.042). Meanwhile, the others have no relationship because they have the value 0 in the w matrix calculation.

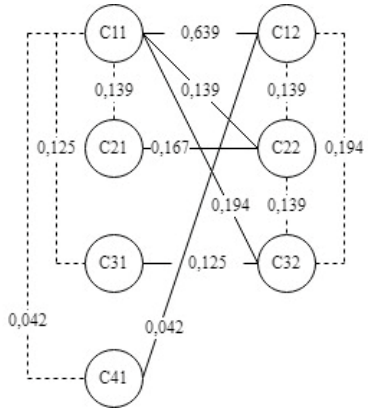

Figure 3. The WCT Matrix between AGNES and K-means

Ensemble cluster with AGNES method has 4 clusters and K-means has 3 clusters as members yield c-index for 2, 3, and 4 cluster respectively 0.254, 0.226, and 0.205. It shows that ensemble cluster have 4 cluster and the member of each cluster is same with AGNES. The characteristics of each cluster can be seen from the average value of each variable show in Table 4

Table 4. Center Value for Ensemble

Variable	*Cluster* 1	*Cluster* 2	*Cluster* 3	*Cluster* 4
female sex workers (x_1)	4.222	17.000	10.000	3.000
transvestites (x_2)	1.083	3.000	3.000	1.000
Male sex men (x_3)	13.306	226.000	15.000	42.000
High risk couples (x_4)	6.500	25.000	21.000	14.000
Customers of public sex workers (x_5)	5.389	14.000	11.000	30.000
IDU (x_6)	0.278	8.000	16.000	1.000
Pregnant women (x_7)	16.028	23.000	40.000	17.000

Table 4 show that cluster 1 have lower center value than other cluster, cluster 2 have highest variable in transvestites (x_2), male sex men (x_3), and high risk couples (x_4), cluster 3 have highest variable in transvestites (x_2), IDU (x_6), and pregnant women (x_7), cluster 4 have highest variable in female sex workers (x_1), and Customers of public sex workers (x_5). 4 cluster in ensemble or AGNES can mapping in East Java as show in **Figure 4**

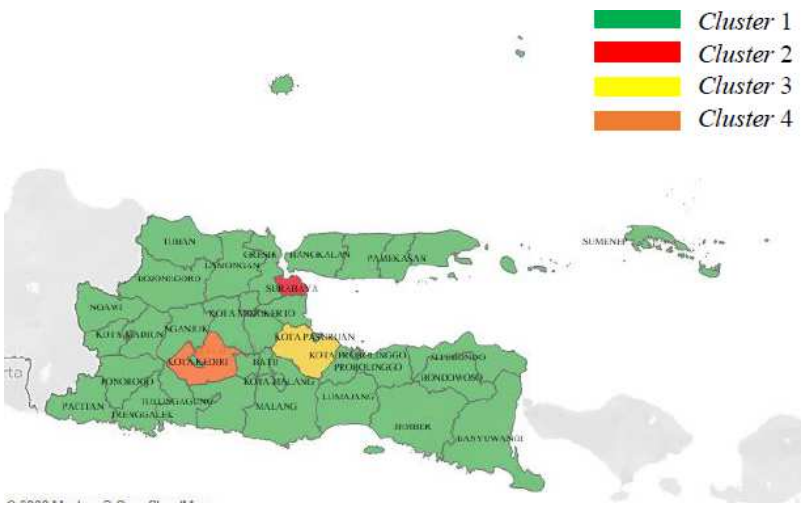

Figure 4. HIV/AIDS Cases in East Java based on Ensemble or AGNES Clustering

The ensemble clustering solution results in four clusters, outperforming its individual components (AGNES and K-means). However, the cluster members in the ensemble solution are identical to those of AGNES. The performance of ensemble clustering is influenced by two key factors: diversity and quality. The selection of a subset of ensemble members, based on these factors, often results in a more accurate and reliable solution [18]. This research problem is addressed by enhancing the clustering process through the integration of AGNES and K-means, where the ensemble method is shown to optimize clustering outcomes by balancing between diversity and quality. By leveraging these principles, the solution not only ensures accuracy but also provides a novel approach to clustering in the context of HIV case analysis in East Java.

4. CONCLUSIONS

The Mahalanobis distance is used in cluster analysis because there is multicollinearity between variables. If the number of clusters is 2, 3, and 4, and the c-index value is used to determine the number of clusters, then the AGNES method has 4 clusters and K-means has 3 clusters. AGNES and K-means results are used as ensemble members. Based on the c-index value, the number of clusters formed in the AGNEs-K-means ensemble is 4. The ensemble cluster members are the same as AGNES.

ACKNOWLEDGMENT

Researchers would like to thank all parties involved in this research, including Laboratorium Simulasi dan Komputasi Statistika as a place to conduct this research, and also colleagues and friends for supporting this research. Hopefully, this research can give more benefits for academia and institution.

REFERENCES

[1] M. R. Ridlo, S. Defiyanti, dan A. Primajaya, "Implementasi Algoritme K-means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang," dipresentasikan pada Conference on Information Technology and Electrical Engineering (CITEE), Yogyakarta, Jul 2017. [Online]. Tersedia pada: https://docplayer.info/53002307-Implementasi-algoritme-K-means-untuk-pemetaan-produktivitas-panen-padi-di-kabupaten-karawang.html [Diakses: 14 Oktober 2023].

[2] M. W. Talakua, Z. A. Leleury, dan A. W. Taluta, "Analisis Cluster dengan Menggunakan Metode K-means untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," BAREKENG: Jurnal Ilmu Matematika dan Terapan, vol. 11, no. 2, Art. no. 2, Des 2017.

[3] S. Nuraidah, "Cluster Ensemble Dalam Penggerombolan Kabupaten/Kota Provinsi Jawa Barat Berdasarkan Indikator Penddikan SMA/SMK/MA.," Skripsi, Institut Pertanian Bogor, Bogor, 2014.
[4] C. Suhaeni, A. Kurnia, dan R. Ristiyanti, "Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhirarki, K-means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)," JURNAL MEDIA INFOTAMA, vol. 14, no. 1, hlm. 31–38, Feb 2018.
[5] A. A. Yusfar, M. A. Tiro, dan S. Sudarmin, "Analisis Cluster Ensemble dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah," VARIANSI: Journal of Statistics and Its application on Teaching and Research, vol. 3, no. 1, Art. no. 1, Sep 2020.
[6] D. Y. Nugroho dan A. Wibowo, "Analisis Cluster K-means Kabupaten/Kota di Jawa Timur Berdasarkan Level Epidemik HIV," Jurnal Biometrika dan Kependudukan (Journal of Biometrics and Population), vol. 8, no. 2, Art. no. 2, Des 2019.
[7] Dinkes Jatim, "Program P2 HIV/AIDS." 2012. [Daring]. Tersedia pada: https://dinkes.jatimprov.go.id/userimage/Data%20P2%20sd%20September%20WEB.pdf [Diakses: 14 Oktober 2023].
[8] S. Qian dkk., "Diversity of HIV/AIDS epidemic in China: a result from hierarchical clustering analysis and spatial autocorrelation analysis," AIDS, vol. 28, no. 12, hlm. 1805, Jul 2014.
[9] X. Yang, J. Zhang, S. Chen, S. Weissman, B. Olatosi, dan X. Li, "Comorbidity patterns among people living with HIV: a hierarchical clustering approach through integrated electronic health records data in South Carolina," AIDS Care, vol. 33, no. 5, hlm. 594–606, Mei 2021.
[10] K. Moens dkk., "Symptom Clusters in People Living with HIV Attending Five Palliative Care Facilities in Two Sub-Saharan African Countries: A Hierarchical Cluster Analysis," PLOS ONE, vol. 10, no. 5, hlm. 1–12, Mei 2015.
[11] D. P. S. Dewi dan A. Voutama, "Penerapan Algoritma K-means Pada Kasus HIV Di Jawa Barat Untuk Pengelompokan Berdasarkan Tingkat Penyebaran Di Setiap Kabupaten/Kota Menggunakan Rapidminer," Jurnal Teknologi Sistem Informasi, vol. 4, no. 2, Art. no. 2, Sep 2023.
[12] W. Rosida dan Y. A. Wijaya, "Klasterisasi Penyakit HIV/AIDS di Jawa Barat Menggunakan Algoritma K-means Clustering," Blend Sains Jurnal Teknik, vol. 1, no. 4, Art. no. 4, Mar 2023.
[13] K. K. Munawar dan A. I. Purnamasari, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA KLASTERISASI KASUS HIV DI JAWA BARAT," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 2, Art. no. 2, Agu 2023.
[14] M. L. Lynch dan V. DeGruttola, "Ensemble clustering of longitudinal bivariate HIV biomarker profiles to group patients by patterns of disease progression," Int J Data Sci Anal, vol. 14, no. 3, hlm. 305–318, Sep 2022.
[15] N. Wakhidah, "CLUSTERING MENGGUNAKAN K-MEANS ALGORITHM," Jurnal Transformatika, vol. 8, no. 1, Art. no. 1, Jul 2010.
[16] R. Silvi, "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia," JMM, vol. 4, no. 1, hlm. 22–31, Mei 2018.
[17] D. A. Saidah, R. Santoso, dan T. Widiharih, "PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KESEHATAN LINGKUNGAN MENGGUNAKAN METODE PARTITIONING AROUND MEDOIDS DENGAN VALIDASI INDEKS INTERNAL," Jurnal Gaussian, vol. 11, no. 2, Art. no. 2, Agu 2022.
[18] E. Akbari, H. Mohamed Dahlan, R. Ibrahim, dan H. Alizadeh, "Hierarchical cluster ensemble selection," Engineering Applications of Artificial Intelligence, vol. 39, hlm. 146–156, Mar 2015.