

CLUSTERING OF PROVINCE IN INDONESIA BASED ON AQUACULTURE PRODUCTIVITY USING AVERAGE LINKAGE METHOD

Fachruddin Hari Anggara Putera^{1*}, Septina F. Mangitung², Madinawati³, Lilies Handayani⁴

^{1, 2, 3}Fisheries and Marine Department, Tadulako University, Indonesia

⁴Statistics Department, Tadulako University, Indonesia

*e-mail: fachruddinharianggaraputera@gmail.com

ABSTRACT

Fishery is an agricultural sector that plays a role in contributing to income for the country because most of Indonesia's territory is water so it is feasible to be developed, one of which is through aquaculture. One of the efforts that can increase and maintain productivity in the aquaculture sector is to classify provinces that produce aquaculture production into groups based on the similarity of characteristics possessed by each province in Indonesia. In this study, clustering was carried out using cluster analysis using the average linkage method and based on the analysis results obtained showed that cluster 1 consists of 25 provinces, cluster 2 consists of 5 provinces, cluster 3 consists of 2 provinces, cluster 4 consists of 1 province, and cluster 5 consists of 1 province with a standard deviation value within a cluster of 11.729 and a standard deviation between clusters of 118.745.

Keywords: Average Linkage, Aquaculture, Clustering, Productivity, Indonesia

Cite: Putera, F. H. A., Mangitung, S. F., Madinawati, & Handayani, L. (2021). Clustering of Province in Indonesia Based on Aquaculture Productivity Using Average Linkage Method. *Parameter: Journal of Statistics*, 2(1), 16-20.



Copyright © 2021 Putera et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Based on the Decree of the Minister of Maritime Affairs and Fisheries Number 45 of 2011 concerning the Estimation of Potential Fish Resources in the Fisheries Management Area of the Republic of Indonesia, Indonesia is a maritime country that has a water area of 3.25 million km² or about 63% of Indonesia's territory. The Indonesian seas has a large enough potential for sustainable marine fish production, assuming around 6.51 million tons/year or 8.2% of the total world marine fish production potential. Fisheries are part of the development of agricultural science with a crucial role, namely advancing community income significantly and developing the potential of each region owned by relying on the fishery sector as the basis for advancing the community (Nurlia, 2009).

It is the duty of the government to continue to increase the fishery population as much as possible. Therefore, as actors and policy makers, fish farmers and the government must work together in order to increase and maintain productivity in the fishery sector. In order to make information related to the types of aquacultures spread in each existing region more efficient and specific, efforts are needed to cluster the provinces in Indonesia that produce aquaculture production into clusters based on their similarity of characteristics, so that in conducting guidance in each regional group in the field of fisheries, it will more focused and on target.

In the hierarchical cluster analysis there are several methods, namely complete linkage, single linkage, average linkage, median linkage, centroid linkage, mcquitty linkage, and ward linkage. Several studies that apply cluster analysis using the average linkage and median linkage methods include Fadliana and Rozi (2015), using the agglomerative hierarchical clustering method. In his research, the average linkage method provides a better cluster solution than the single linkage, complete linkage and ward methods based on the results of the cluster validity test. Based on these considerations, this study conducted clustering of provinces in Indonesia based on aquaculture productivity using average linkage cluster analysis.

MATERIALS AND METHODS

1. Data

This study uses secondary data sourced from the One Data Marine and Fisheries website, Ministry of Maritime Affairs and Fisheries of Indonesia (Pusdatin, 2013). The research variable used is aquaculture productivity (tons) in each province in Indonesia, which includes 34 provinces with 11 variables in this study in the form of aquaculture productivity data with various types of commodities, namely milkfish, gourami, snapper, grouper, catfish, goldfish, tilapia, catfish, other fish, seaweed, and shrimp.

2. Data Analysis

- a. Collecting data and conducting data feasibility testing using the Kaiser Meyer Olkin (KMO) test with KMO test statistics (Amaliyah and Wibawati, 2012), namely:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (1)$$

- b. Determine the size of the similarity and dissimilarity between two objects using the Euclidean distance formula as follows (Gudono, 2015):

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

- c. Perform cluster analysis procedures using the hierarchical method, namely the average linkage method as follow (Hardius dan Nurdin, 2013):

$$d_{(ij)k} = \frac{n_i}{n_i + n_j} d_{ik} + \frac{n_j}{n_i + n_j} d_{jk} \quad (3)$$

- d. Determine the value of the standard deviation within the cluster (S_W) and between clusters (S_B) with the following equation (Goreti et al, 2016):

$$S_W = K^{-1} \sum_{k=1}^K S_k \quad (4)$$

$$S_B = \left[(K - 1)^{-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2 \right]^{1/2} \quad (5)$$

e. Draw a conclusion.

RESULTS AND DISCUSSION

1. Data Feasibility Test

The data feasibility test is needed to ensure that the data used in this study is objectively feasible. This study used the Kaiser Meyer Olkin (KMO) test and obtained a KMO value of 0.622 which is greater than 0.5, so it can be concluded that the sample used was sufficient for further analysis.

2. Cluster Analysis of Average Linkage

The analysis process using the average linkage method is a complex agglomeration process, especially in the calculation of coefficients involving several provinces. the grouping process using the average linkage method starting from the closest distance. The calculation is carried out until the calculation of the Euclidean distance matrix of all objects has been combined which produces a new matrix. The following is an optimal cluster scree plot from the average linkage method.

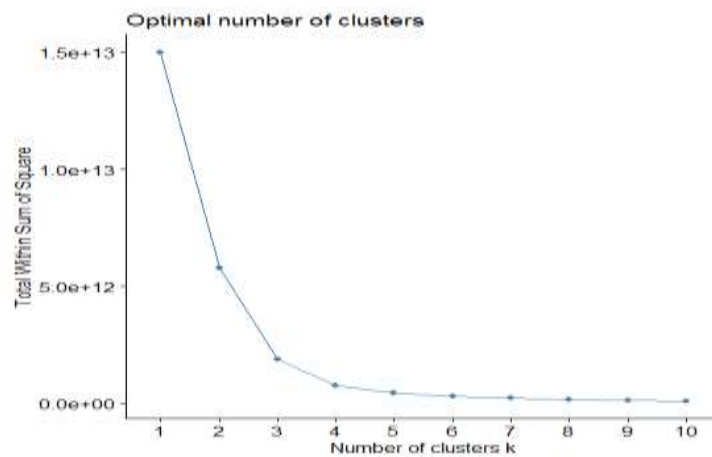


Figure 1. Scree Plot Cluster Optimal of Average Linkage

Based on the Figure 1 above, information can be obtained that the scree plot shows the optimal value of the cluster, after the fifth cluster the decline occurs slowly. So that, it can be predicted that there are 5 clusters in this grouping process where the grouping process can be illustrated in the form of a dendrogram from the average linkage method as follows.

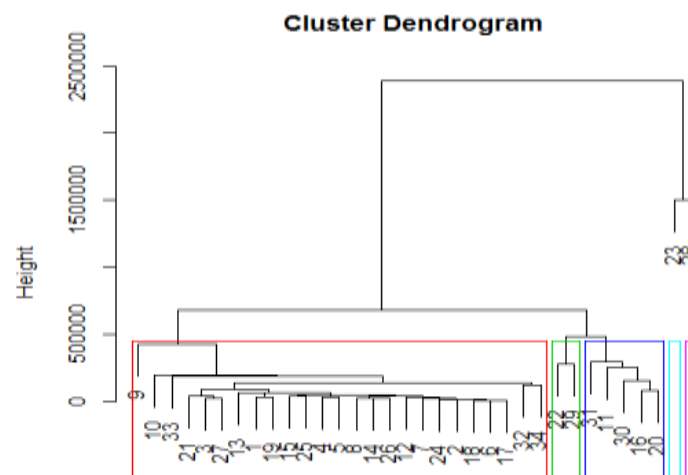


Figure 2. Cluster Dendrogram of Average Linkage

Based on the dendrogram above, it shows that there are 5 clusters. The distribution of the members of each cluster can be seen in the following table.

Cluster	Province
1	Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Jambi, West Java, Center Java, West Kalimantan, South Kalimantan, Center Kalimantan, East Kalimantan, Bangka Belitung Islands, Riau Islands, Lampung, North Maluku, Papua, West Papua, Riau, West Sulawesi, West Sumatera, South Sumatera, and North Sumatera
2	East Java, North Kalimantan, Maluku, Southeast Sulawesi, and North Sulawesi
3	West Nusa Tenggara and Center Sulawesi
4	East Nusa Tenggara
5	South Sulawesi

The results of the cluster analysis above indicate that there is a closeness between objects that can be used as a guide for which objects have similar characteristics. Two objects with the same characteristics will be described as two points that are close together. The analysis obtained shows that cluster 1 consists of 25 provinces, cluster 2 consists of 5 provinces, cluster 3 consists of 2 provinces, cluster 4 consists of 1 province, and cluster 5 consists of 1 province. The following is a map of the distribution of these clusters.



Figure 3. Distribution Map of Average Linkage

3. Standard Deviation of Cluster Result

- a. The following is the calculation of the standard deviation value in the cluster.

$$\begin{aligned}
 S_W &= K^{-1} \sum_{k=1}^K S_k \\
 S_W &= 5^{-1} (s_1 + s_2 + s_3 + s_4 + s_5) \\
 S_W &= \frac{(s_1 + s_2 + s_3 + s_4 + s_5)}{5} \\
 S_W &= \frac{(22,126 + 26,245 + 10,278 + 0 + 0)}{5} \\
 S_W &= \frac{58,649}{5} = 11,729
 \end{aligned}$$

- b. To calculate the standard deviation value between clusters, it is necessary to know the average value of each cluster. The following is a calculation of the standard deviation between clusters.

$$S_B = \left[(K - 1)^{-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2 \right]^{1/2}$$

$$\bar{X} = \frac{\bar{X}_A + \bar{X}_B + \bar{X}_C + \bar{X}_D + \bar{X}_E}{5}$$

$$\bar{X} = \frac{15,834 + 65,574 + 108,237 + 172,466 + 322,832}{5} = \frac{684,943}{5} = 136,988$$

$$S_B = \left(\frac{(15,834 - 136,988)^2 + (65,574 - 136,988)^2 + \dots + (322,832 - 136,988)^2}{5 - 1} \right)^{1/2}$$

$$S_B = \left(\frac{56,401,551,933}{4} \right)^{1/2} = (14,100,387,983)^{1/2} = 118,745$$

CONCLUSION

There are 5 clusters obtained from the analysis using the average linkage method which indicates the closeness between objects in the cluster, in other words the provinces in each cluster have similar characteristics based on aquaculture productivity where cluster 1 has the most members consisting of 25 provinces. which are mostly spread over the islands of Sumatra and Java, cluster 2 consists of 5 provinces, namely East Java, North Kalimantan, Maluku, Southeast Sulawesi, and North Sulawesi, cluster 3 consists of the provinces of West Nusa Tenggara and Central Sulawesi, cluster 4 is the province of Nusa Tenggara. East, and cluster 5 is the province of South Sulawesi.

REFERENCES

- Amaliyah, F., and Wibawati. (2012). Grouping of Regencies/Cities in East Java Based on Healthy Indonesia Indicator 2010. *Journal of Science and Art ITS*, 1(1), 188-193
- Fadliana, A., and Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *Journal of Cauchy*, 4(1), 25-40
- Goreti, M., Novia, Y., and Wahyuningsih, S. (2016). Comparison of Cluster Analysis Results Using Single Linkage Method and C-Means Method. *Journal of Exponential*, 7(1), 9-16
- Gudono. (2015). *Multivariate Data Analysis, Four Edition*. Yogyakarta: BPFE.
- Hardius, U., and Nurdin, S. (2013). *Application of Multivariate Techniques for Marketing Research*. Jakarta: Raja Grafindo Persada.
- Nurlia. (2009). Peranan Sub Sektor Perikanan terhadap Produk Domestik Regional Bruto (PDRB) dan Kesempatan Kerja di Kabupaten Pinrang Periode 2005-2009. Makassar: Hasanuddin University.
- Pusdatin [KKP]. (2013). *Marine and Fisheries in Figures 2013*. Jakarta: Center for Data, Statistics and Information, Ministry of Marine Affairs and Fisheries.