Variance: Journal of Statistics and Its Applications
Volume 6 Number 1 | April 2024
P-ISSN: 2685-8738 E-ISSN: 2685-872X
10.30598/variancevol6iss1page49-54

CLUSTERING SHRIMP DISTRIBUTION IN INDONESIA USING THE X-MEANS CLUSTERING ALGORITHM

Rahmi Fadhilah[1], M. Y. Matdoan[2], Dinda Ayu Safira[3], S. P. Tahalea[4]
[1,3]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember
Jl. Arief Rachman Hakim, Kampus ITS Sukolilo, Surabaya, 60111, East Java, Indonesia,
[2]Statistics Study Program, FMIPA, Pattimura University
Jl. Ir. M. Putuhena, kampus Unpatti-Poka, Ambon, 97233, Indonesia
[4]Institute of Informatics, University of Szeged
Arpad Ter 2, 6720, Hungary

E-mail Corresponding Author: *6003231016@student.its.ac.id*

Abstract: Shrimp is one of the marine biological resources available in almost all Indonesian waters and is one of the mainstay export commodities from the fisheries sub-sector. This is expected to improve the welfare of the community, so it is necessary to cluster the distribution of shrimp in Indonesia. Clustering is a data mining technique used to group data or partition datasets into subsets. One of the best clustering algorithms is X-means. X-means clustering is used to solve one of the main disadvantages of K-means clustering, namely the need for prior knowledge of the number of clusters (K). The purpose of this research is to obtain the results of clustering the distribution of shrimp in Indonesia using the X-means clustering algorithm. The data used in this study comes from the publication of Marine and Coastal Resources Statistics 2022 by the Central Bureau of Statistics of the Republic of Indonesia. This study obtained the results that there are 3 clusters in the clusterization of shrimp distribution in Indonesia. Cluster 0 consists of 1 province, cluster 1 consists of 27 provinces, and cluster 2 consists of 6 provinces.

Keywords: Fisheries, Indonesia, Shrimp, X-Means Clustering

1. INTRODUCTION

Indonesia is an archipelago that has two-thirds of its territory consisting of oceans. This makes marine resources so abundant that it becomes one of the mainstay sectors that can be developed and bring income for the progress and welfare of the Indonesian people[1].

Fisheries is an activity that utilizes marine natural resources using science and technology for human welfare. Shrimp is an excellent Indonesian fishery commodity. Shrimp is one of the marine biological resources available in almost all Indonesian waters and is one of the main export commodities of the fisheries sub-sector [2]. Every year, there is an increase in shrimp export market share to export destination countries such as Japan, the United States, and the European Union.

Shrimp is Indonesia's largest fishery export commodity above the tuna commodity, which ranks second. Based on data on the volume of Indonesian shrimp exports to foreign countries from 2017 to 2021, it has always increased. In 2017, Indonesian exports amounted to 180,592 tons, then in 2018, it was 197,434; in 2019, it was 207,703; in 2020, it was 239,282, and continued to increase until 2021 to 250,715 [3]. Along with the increase in shrimp exports from year to year, this is also expected to be one indicator of increasing the income and welfare of fishing communities in Indonesia. The target of poverty reduction in fishing communities is increasing access to capital and production infrastructure so that it impacts community welfare [4]. However, data from BPS-Statistics Indonesia shows that most of the biggest contributors to poverty come from rural areas, especially in communities where fishermen work. Therefore, it is necessary to conduct an in-depth study and analysis of the distribution of shrimp in Indonesia and then associated with the socio-economic community in the area.

Clustering is a process for grouping data into several clusters or groups so that the data in one cluster has the maximum level of similarity and the data between clusters has minimum similarity [5][6]. Clustering can also be said to be a data mining technique used to analyze data to solve problems in grouping data or, rather, partitioning datasets into subsets [7]. In clustering techniques, the goal is to distribute cases (objects, people, events, and others) into a group so that the level of connection between members of the same group is

https://ojs3.unpatti.ac.id/index.php/variance/ *Submitted: 2023-11-29* *Accepted: 2024-02-05*

strong and weak between different members [8]. The development of clustering techniques has resulted in various studies being conducted to produce clusters with an increasingly good level of accuracy[9].

The X-Means Clustering algorithm was developed by Dan Pelleg and Andre Moore in 2000. In this algorithm, the number of clusters is calculated dynamically using upper and lower limits provided by the user [10]. X-means clustering is used to solve one of the main drawbacks of K-means clustering, namely the need for prior knowledge of the number of clusters (K) [11]. In this method, the actual value of K is estimated in an unsupervised manner and based solely on the data set itself.

2. METHOD

This study uses secondary data obtained from the Central Bureau of Statistics (BPS). The sample unit used in this study is all provinces in Indonesia, which amounted to 34 provinces. The variables used in this study consisted of Donggol Shrimp (X1), White Shrimp / Jerbung (X2), Krosok Shrimp (X3), Queen / King Shrimp (X4), Windu Shrimp (X5), Barong Shrimp / Coral Shrimp (X6) and Other Shrimp (X7). The data analysis in this study consists of descriptive statistics and continues with the x-means algorithm.

The X-means algorithm is one of the clustering algorithms; the x-means algorithm starts with the assumption of having a minimum number of clusters and then dynamically increases it. X-means uses certain separation criteria to control the clustering process [12]. The x-means algorithm is a variation of k-means clustering that refines the cluster assignment by repeatedly attempting splits and keeping the best-split result until some criterion is reached. Pelleg and Moore [13] define the X-means algorithm adapted from K-means; x-means acts after each run of K-means, deciding which part of the stream the centers should split off to get better data. The splitting decision is done by calculating the BIC criterion. This new approach proposes an efficient solution to one major drawback of K-means, which is the search for the number of clusters K. In addition, X-means has a low computational cost [14].

The following are the completion steps in the *X-Means Algorithm* [15].

a. Initialize the number of clusters.

b. Calculate the distance of each data to each cluster center using the Euclidean distance equation, which is as follows:

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

c. Determined the new cluster center by calculating the average attribute value of each data belonging to the cluster.

d. Return to step 2 until No. data moves to another cluster or until the maximum iteration limit.

3. RESULTS AND DISCUSSION

Overview of Shrimp Distribution in Indonesia

Table 1. Descriptive Statistics

Attribute	N	Minimum	Maximum
Donggol Shrimp (X1)	34	.00	10769.00
White Shrimp/Jerbung (X2)	34	.00	3094.00
Krosok Shrimp (X3)	34	.00	5424.00
Queen/King Shrimp (X4)	34	.00	1487.00
Tiger Prawns (X5)	34	.00	5224.00
Barong Shrimp/Coral Shrimp (X6)	34	.00	3.00
Other Shrimp (X7)	34	.00	19256.00

a. Testing Using Rapdminer

The data testing process is carried out using the Rapidminer application. Testing is done using the X-means clustering algorithm as shown in Figure 1.

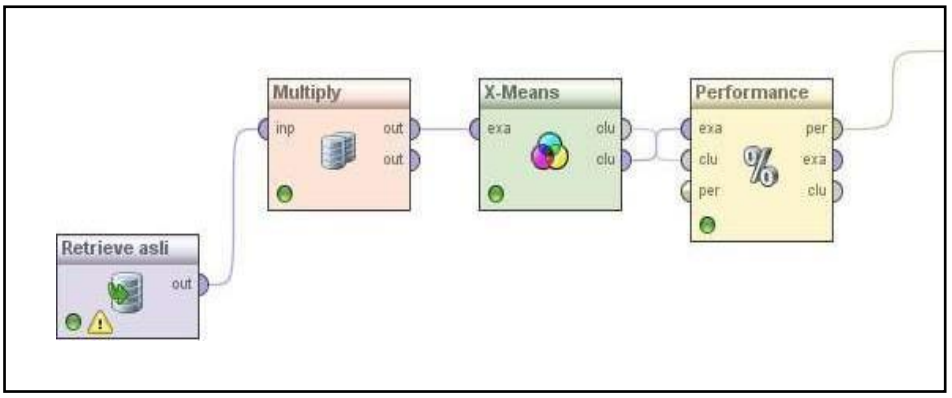

Figure 1. X-Means Clustering Modeling

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -7127946.832
Avg. within centroid distance_cluster_0: -0.000
Avg. within centroid distance_cluster_1: -4688132.060
Avg. within centroid distance_cluster_2: -19295104.444
Davies Bouldin: -0.431
```

Figure 2. Davies Buldin

a. X-Means Clustering Testing

Furthermore, the clustering of Indonesian shrimp distribution was tested using the x-means clustering algorithm.

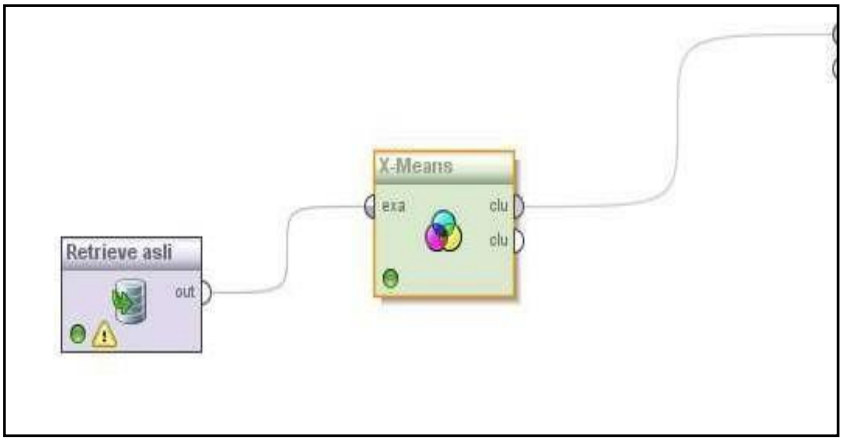

Figure 3. Testing the X-Means Algorithm

Based on the x-means clustering modeling in Figure 3, the clusters formed are 3 clusters. For cluster 0, there is 1 item, cluster 1 has 27 items, and cluster 2 has 6 items with a total of 34 items. The explanation can be seen in Figure 4.

Cluster Model

```
Cluster 0: 1 items
Cluster 1: 27 items
Cluster 2: 6 items
Total number of items: 34
```

Figure 4. X-Means Clustering Results

Based on Figure 4, the clusterization results can be detailed as follows.

Table 2. Clusterization Results

Cluster 0	Cluster 1	Cluster 2
Papua Barat	Aceh	North Sumatra
	West Sumatra	Jambi
	Riau	Bengkulu
	South Sumatra	West Kalimantan
	Lampung	South Kalimantan
	Bangka Belitung Islands	East Kalimantan
	Riau Islands	
	DKI Jakarta	
	West Java	
	Central Java	
	DI Yogyakarta	
	East Java	
	Banten	
	Bali	
	West Nusa Tenggara	
	East Nusa Tenggara	
	Central Kalimantan	
	North Kalimantan	
	North Sulawesi	
	Central Sulawesi	
	South Sulawesi	
	Southeast Sulawesi	
	Gorontalo	
	West Sulawesi	
	Maluku	
	North Maluku	
	Papua	

Based on Table 2, it shows that the test results of the x-means clustering algorithm can be detailed as follows.

Cluster 0 : West Papua Province.

Cluster 1 : Province Aceh, West Sumatra, Riau, South Sumatra, Lampung, Bangka Belitung Islands, Riau Islands, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, Central Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku and Papua Province.

Cluster 2 : Furthermore, Table 3 provides information about the centroid points in each cluster, from the results of data processing carried out using the rapid miner application

Table 3. Centroid results

Attribute	Cluster 0	Cluster 1	Cluster 2
Udang Donggol (X1)	10769	268.481	809.833
Udang Putih/Jerbung (X2)	590	564.889	1115.500
Udang Krosok (X3)	0	307.111	2661.667
Udang Ratu/Raja (X4)	0	17.741	247.833
Udang Windu (X5)	723	487.741	607.333
Udang Barong/Udang Karang (X6)	0	0.111	0
Udang Lainnya (X7)	339	1551.407	12555.167

Based on Table 3. shows that the highest donggol shrimp (X1) is in cluster 2. and the lowest is in cluster 0. White shrimp / Jerbung (X2) the highest is in cluster 2 and the lowest is in cluster 0. Krosok shrimp (X3) the highest is in cluster 2 and the lowest is in cluster 0. Queen/king shrimp (X4) the highest is in cluster 2 and the lowest is in cluster 0. Tiger shrimp (X5) is highest in cluster 2 and lowest in cluster 0. Barong shrimp/crayfish (X6) is highest in cluster 1 and lowest in clusters 0 and 2. Other shrimp (X7) is highest in cluster 2 and lowest in cluster 0. Furthermore, the visualization can be seen in Figure 5 below.

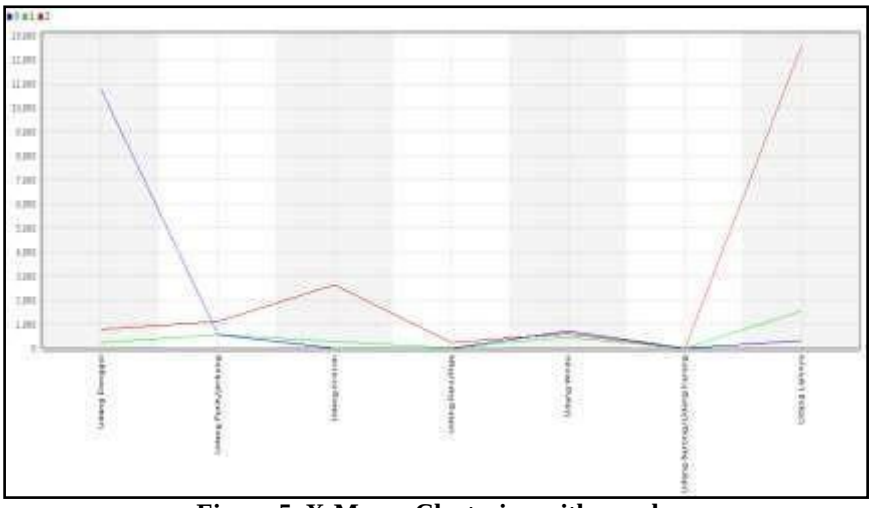

Figure 5. X-Means Clustering with graphs

4. CONCLUSIONS

Based on the results and discussion, it can be concluded that there are 3 clusters in the clusterization of shrimp distribution in Indonesia. Cluster 0 consists of 1 province, cluster 1 consists of 27 provinces, and cluster 2 consists of 6 provinces. Furthermore, the highest donggol shrimp (X1) is in cluster 2. and the lowest is in cluster 0. White shrimp / Jerbung (X2), the highest is in cluster 2, and the lowest is in cluster 0. Krosok shrimp (X3), the highest is in cluster 2, and the lowest is in cluster 0. Queen/king shrimp (X4) is highest in cluster 2 and lowest in cluster 0. Tiger shrimp (X5) is highest in cluster 2 and lowest in cluster 0. Barong shrimp/crayfish (X6) is highest in cluster 1 and lowest in clusters 0 and 2. Other shrimp (X7) are highest in cluster 2 and lowest in cluster 0.

REFERENCES

[1] Sahri, A. (2023). Mengenal potensi rumput laut: kajian pemanfaatan sumber daya rumput laut dari aspek industri dan kesehatan. *Majalah Ilmiah Sultan Agung*, *44*(118), 95-116.

[2] Indrayani, L., Soetjipto, H., & Sihasale, L. (2006). Skrining fitokimia dan uji toksisitas ekstrak daun pecut kuda (Stachytarpheta jamaicensis L. Vahl) terhadap larva udang Artemia salina Leach. *Berkala Penelitian Hayati*, *12*(1), 57-61.

[3] Statistik, B. P. (2017). Badan pusat statistik. *Badan Pusat Statistik*.

[4] Lamia, K. A. (2013). Faktor-faktor yang mempengaruhi tingkat pendapatan nelayan kecamatan tumpaan, kabupaten minahasa selatan. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, *1*(4).

[5] Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan rapidminer dengan K-means cluster pada daerah terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. *CESS (Journal of Computer Engineering, System and Science)*, *3*(2), 173-178.

[6] Rokach, L., & Maimon, O. (2005). Clustering methods. *Data mining and knowledge discovery handbook*, 321-352.

[7] Putra, A. I. (2017). *Penerapan Metode Clustering K-Means Untuk Penerimaan Beasiswa SMA Negeri 3 Metro* (Doctoral dissertation, IIB DARMAJAYA).

[8] Anwar, S., Nuris, N. D., & Wijaya, Y. A. (2019). Pengelompokkan Tingkat Pemahaman Kurikulum Berbasis KKNI Menggunakan Metode X-Means Clustering. *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, *4*(2-2).

[9] Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021). Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari. *Jurnal Sains Teknologi Transportasi Maritim*, *3*(1), 1-7.

[10] Putra, P. H., Hasibuan, A., & Marpaung, E. A.(2022). Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, *19*(2), 424-429.

[11] Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi*. Deepublish.

[12] Mustofa, Z., & Suasana, I. S. (2018). Algoritma clustering K-medoids Pada e-government Bidang information and communication technology Dalam Penentuan status Edgi. *Jurnal Teknologi Informasi dan Komunikasi*, *9*(1), 1-10.

[13] Pelleg, D., & Moore, A. W. (2000, June). X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml* (Vol. 1, pp. 727-734).

[14] Patimah, E., Ermatita, E., & Chamidah, N. (2021). Analisis Cluster Kepuasan Pengguna Terhadap Layanan Shopee Menggunakan Algoritma K-Means. *Informatik: Jurnal Ilmu Komputer*, *17*(3), 209-217.

[15] Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai. *IT-EXPLORE: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, *1*(3), 179-193.