pijmath
Pattimura International Journal of Mathematics (PIJMath)
Volume 02, issue 02, pages 63-70, November 2023
E-ISSN: 2830-6791
RESEARCH ARTICLE ◙ OPEN ACCESS
DOI: https://doi.org/10.30598/pijmathvol2iss2pp63-70

Application of K-Means Algorithm for Clustering Capture Fisheries Production in Maluku Province

M. Y. Matdoan [1*] , Nur A. Purnamasari[2] , Novita S. Laamena[3]

[13] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Pattimura Jl. Ir. M. Putuhena, Poka-Ambon, 97233, Indonesia*

[2] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Mataram University Jl. Maja Pahit, Mataram, 83125, Indonesia*

Corresponding author's e-mail: [1]keepyahya @gmail.com*

ABSTRACT

Article History

Received : 20th July 2023
Accepted: 21st September 2023
Published: 5th November 2023

Keywords
Clustering K-Means;
Fisheries.
;

Maluku Province has large natural resources with various potentials ranging from the ocean floor to the land. Capture fisheries products are one of the leading sectors contributing greatly to the GRDP of Maluku Province. The k-Means clustering algorithm is suitable for grouping data objects with similar identities. This research aimed to cluster regencies/cities in Maluku Province based on capture fisheries products. The type of data in this study is secondary data sourced from the publication of the Central Bureau of Statistics (BPS) of Maluku Province in 2022. The results showed 3 clusters of regencies/cities in Maluku Province based on capture fisheries products. Cluster 1 with sufficient capture fisheries product category is Tanimbar Islands, Buru, East Seram, West Seram, South Buru, Southwest Maluku, Ambon City, and Tual City. Furthermore, Cluster 2 with the category of high amount of capture fisheries products are Aru Islands Regency and Southeast Maluku Regency. Furthermore, Cluster 3, comprising very high amount of capture fisheries products, is Central Maluku Regency.

[1]***How to cite this article:***
Matdoan M. Y., Purnamasari N. A., & Laamena N. S.,, "APPLICATION OF K-MEANS ALGORITHM FOR CLUSTERING CAPTURE FISHERIES PRODUCTION IN MALUKU PROVINCE", *Pattimura Int. J. Math. (PIJMATH),* vol. 02, iss. 02, pp. 63-70, November, 2023.

1. Introduction

Maluku Province is one of the provinces where most of the area consists of ocean areas. It makes the Maluku Sea rich in natural potential that exists and is spread from the bottom of the ocean to the land. It makes the Maluku Sea known as the home or granary of various types of profitable consumption fish. Maluku marine fisheries resources are a major sector in improving community welfare and are one of the largest contributors to the GRDP of Maluku Province. Capture fisheries production is classified according to the type of fish, such as shrimp, mackerel tuna, skipjack, tuna, and others.[1]. Maluku Province has great potential for capturing fisheries due to limited or not maximized supply capacity but large domestic and international demand. Therefore, it is important to research the clustering of capture fisheries products in Maluku Province. One of the suitable techniques for clustering is by using Machine Learning technology.

Machine Learning with clustering technique is very suitable to be applied to various fields, including fisheries. There are two techniques in machine learning: supervised learning and unsupervised learning.[2]. Supervised learning is a machine learning technique for polarized data.[3]. There are several methods or algorithms in supervised learning techniques, including Support Vector Machine (SVM), Naive Bayes, Decision Tree, K-Nearest Neighbor, and others.[4]. Furthermore, Unsupervised Learning is a machine learning technique for data that has not been polarized.[5]. There are several methods or algorithms in unsupervised learning techniques, including K-Means, Single Linkage, Complete Linkage, Average Linkage, K-Medoids, X-Means, and others.[6].

The k-Means Clustering algorithm is one of the most widely used clustering algorithms because it is simple and easy to interpret. This algorithm is suitable for grouping data objects with the same identity.[7]. Based on the opinion of [8], the K-Means Clustering algorithm works by dividing data objects into various groups or clusters based on the size of the data similarity so that data objects in one cluster have the highest level of similarity and data objects outside the cluster have the smallest level of similarity. Some research on the K-Means algorithm includes the application of the K-Means algorithm for clustering in the social field.[9][10][11][12][13], industrial field[14] [15] and economic field[16].

From the studies that have been done before, there has yet to be research on applying the K-Means algorithm for clustering capture fisheries products. Therefore, this research investigated the application of the K-Means algorithm for clustering Capture Fisheries Products in Maluku Province.

2. Research Methods

2.1 Data Source

The data obtained in this research was sourced from the 2022 publication of the Central Bureau of Statistics (BPS) of Maluku Province.

2.2 Research Variables

The data taken regarding the indicator of capture fisheries production consists of variables of skipjack volume (X1), tuna volume (X2), tuna volume (X3), shrimp (X4), and others (X5). The sample units in this study consisted of 11 (eleven) regencies/cities, namely West Southeast Maluku, Southeast Maluku, Central Maluku, East Seram, West Seram, Southwest Maluku, Aru Islands, South Buru, Buru, Ambon City, and Tual City.

3. Results And Discussion

3.1 Overview of Research Variables

Before the clustering process, a general description of the research variables sourced from the Central Bureau of Statistics (BPS) was first described. Data was collected at the Central Bureau of Statistics of Maluku Province in 2022. The data described are skipjack volume data (X1), mackerel tuna volume (X2), tuna volume (X3), shrimp (X4), and others (X5).

Table 1. Descriptive Statistics

	N	Minimum	Maximum	Mean
Volume of Skipjack	11	1.30	15380.80	2576.4727
The volume of Mackarel Tuna	11	408.40	27908.10	5632.3455
Volume of Tuna	11	17.60	29228.80	4145.9545
Shrimp	11	.00	8174.10	817.1909
Others	11	2110.50	115239.60	36141.0455

Table 1 shows that this study's sample units are 11 regencies/cities located in Maluku Province. The highest volume of skipjack is 15380.80, and the least is 1.30. Furthermore, the highest volume of tuna is 27908.10, and the lowest is 27908.10. Furthermore, the highest volume of tuna is 29228.80, and the lowest is 17.60. Furthermore, the highest volume of shrimp was 8174.10, and the lowest was .00. Then, the highest volume of other fish was 115239.60, and the lowest was 2110.50.

3.2 Designing the Research

3.2.1 Data Standardization

The data is standardized if there are significant unit differences among the research variables. The following are the results of standardizing the research variables.

Table 2. Standardization Results

Regency/City	X1	X2	X3	X4	X5
Southeast West Maluku	-.40674	-.51670	-.45379	-.27000	-.64698
Southeast Maluku	-.24918	.58753	-.34836	-.33400	1.15195
Central Maluku	1.91494	1.67086	1.94360	-.3400	1.93871
Buru	-.45692	-.57132	-.32136	-.31168	-.68503
Aru Islands	-.10824	.63725	-.39097	1.00690	1.41383
West Seram	.19699	-.09096	.03666	-.33400	-.41694
East Seram	-.40385	-.44892	-.32921	-.08722	-.51272
Southwest Maluku	-.56516	-.61174	-.47724	-.33400	-.83409
South Buru	.09086	-.60257	.19923	-.33400	-.82047
Ambon	-.42646	-.62635	-.37405	-.33400	-.45069
Tual	-.58624	-.42708	-.48448	-.33400	-.13757

3.2.2 *Outlier* Detection

Based on the standardized data, it can be seen in **Table 1** that no values exceed ±2.5, so it is concluded that the data used in this study does not contain outliers.

3.3 Assumption Test

3.3.1 Sample Sufficiency Assumption

The sample adequacy assumption test used to test the sample in the study is sufficient for use. The sample adequacy test was carried out by calculating the *Kaiser Meyer Olkin* (KMO) value.

H_0 : Not enough samples for analysis
H_1 : Enough samples for analysis

Table 3. *KMO and Barlett's Test*

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.577
Bartlett's Test of Sphericity	Approx. Chi-Square	79.874
	Df	10
	Sig.	.000

From **Table 3**, it can be shown that the KMO MSA is 0.577> 0.50, and *Barlett's Test of Sphericity* (Sig.) value is 0.000 < 0.05, so clustering can be done using the *K-Means* algorithm.

3.3.2 Multicollinearity Assumption

Determining the presence or absence of multicollinearity can be seen from the correlation values in the correlation matrix. If the correlation value is < 0.80, the independent variables do not contain cases of multicollinearity and vice versa.

Table 4. Multicollinearity Test

	X1	X2	X3	X4	X5
X1	1	.701**	.691**	-.050	.658*
X2	.701**	1	.758**	.197	.710**
X3	.691**	.758**	1	-.142	.589
X4	-.050	.197	-.142	1	.454
X5	.658*	.710**	.589	.454	1

Table 4 shows that the correlation value for all variables is <0.80, so it can be concluded that the variables used in this study do not have a case of multicollinearity**.**

3.4 *K-Means* Cluster Analysis

K-Means Cluster is part of a non-hierarchical *cluster* that aims to divide all existing objects into one or more groups based on their characteristics. It is done so that objects with similar characteristics are grouped in one group, and objects with different characteristics are grouped in different groups. Before iteration is performed, an initial look at the data clustering process.

Table 5. *Initial Cluster Centers*

	Cluster		
	1	2	3
X1	93.90	1481.90	15380.80
X2	530.30	10532.50	27908.10
X3	79.30	1177.50	29228.80
X4	.00	.00	.00
X5	2110.50	83140.30	115239.60

From **Table 5**, it can be seen the results of the temporary clustering process. Then the iteration process is carried out.

Table 6. *Iteration History*

	Change in Cluster Centers		
Iteration	1	2	3
1	11291.002	6739.054	.000
2	.000	.000	.000

From **Table 6**, it shows that there are two iteration processes. There are some insignificant *centroids* in the first iteration, and in the second iteration, all *centroids are* significant.

Table 7. *Number of Cases in Each* Cluster

Cluster	1	8.000
	2	2.000
	3	1.000
Valid		11.000
Missing		.000

Table 7 shows that **Cluster 1** consists of 8 (eight) regencies/cities, **Cluster 2 consists** of 2 (two) regencies/cities, and **Cluster 3** consists of 1 regency/city.

Table 8. *Final Cluster Centers*

	Cluster		
	1	**2**	**3**
X1	1172.19	1791.45	15380.80
X2	1571.00	10739.85	27908.10
X3	1798.10	995.95	29228.80
X4	101.88	4087.05	.00
X5	13168.36	88482.50	115239.60

Table 8 shows that there are 3 clusters formed for each variable. Cluster 1 shows that the most dominant variable in cluster 1 is other types of fish (X5), then tuna (X3). Furthermore, Cluster 2 shows that the most dominant variables are other types of fish (X5) and mackerel tuna (X2). Furthermore, cluster 3 shows that the most dominant variables are other types of fish (X5) and tuna fish (X3), and shrimp (X4) are not found in Cluster 3.

Table 9. Regency/City Clusters Formed

Regency/City	Cluster	Distance
Tanimbar Islands	1	3773.539
Southeast Maluku	2	6739.054
Central Maluku	3	.000
Buru	1	5077.234
Aru Islands	2	6739.054
West Seram	1	7660.831
East Seram	1	2217.543
Southwest Maluku	1	11291.002
South Buru	1	11439.329
Ambon	1	4827.495
Tual	1	17497.767

Table 9 shows indicators of capture fisheries production in each regency/city in Maluku Province; three clusters were formed, which are then visualized in **Figure 1** below.

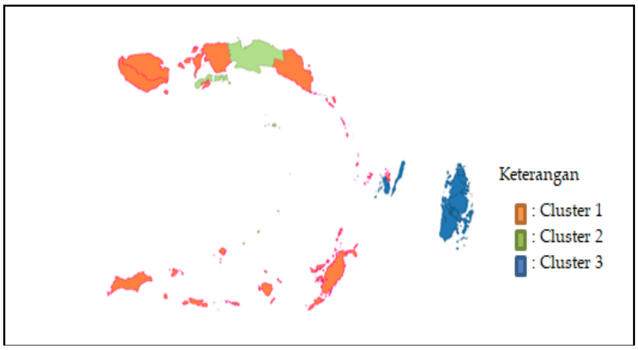

Figure 1. Visualization of clustering the distribution of capture fishery products.

Based on **Figure 1**, it can be explained that

Cluster 1 consists of the regencies of Tanimbar Islands, Buru, East Seram, West Seram, South Buru, Southwest Maluku, Ambon City, and Tual City.
Cluster 2 consists of the Aru Islands and Southeast Maluku regencies.
Cluster 3 consists of Central Maluku Regency.

Furthermore, an ANOVA test was conducted to determine whether each cluster had a difference. If the $F_{count} > F_{tabel}$ value or, in other words, if the significance value (sig) < 0.05, a difference exists between the three clusters formed.

Table 10. ANOVA

	Cluster Mean Square	Df	Error Mean Square	Df	F	Sig.
X1	90479727.134	2	1499369.064	8	60.345	.000
X2	340169330.571	2	1908125.981	8	178.274	.000
X3	346546780.551	2	4125866.881	8	83.994	.000
X4	83072586.365	2	4214625.805	8	43.102	.001
X5	7978895178.044	2	86036382.115	8	92.739	.000

From **Table 10**, it can be seen that the significance value (Sig.) for the five variables, namely skipjack volume (X1), mackerel tuna volume (X2), tuna volume (X3), shrimp (X4 and others (X5) < 0.05 so it can be concluded that the five variables have significant differences.

4. Conclusions

From the results and discussion, it is concluded that 3 clusters in the clusterization of capture fisheries products in regencies/cities in Maluku Province. Cluster 1 with sufficient capture fisheries product category is Tanimbar Islands, Buru, East Seram, West Seram, South Buru, Southwest Maluku, Ambon City, and Tual City. Furthermore, Cluster 2, comprising high amount of fishery products, is the Aru Islands Regency and Southeast Maluku. Then, Cluster 3, with the category of very high amount of fishery products, is Central Maluku Regency.

References

[1] K. Kusdiantoro, A. Fahrudin, S. H. Wisudo, and B. Juanda, “Perikanan Tangkap di Indonesia: Potret Dan Tantangan Keberlanjutannya,” [Capture Fisheries in Indonesia: Portrait and Sustainability Challenges] *J. Sos. Ekon. Kelaut. Dan Perikan.*, vol. 14, no. 2, p. 145, Dec. 2019, doi: 10.15578/jsekp.v14i2.8056.

[2]h I. M. Faiza, G. Gunawan, and W. Andriani, "Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir," [Systematic Literature Review: Application of Machine Learning Methods for Flood Disaster Detection] *J. Minfo Polgan*, vol. 11, no. 2, pp. 59–63, Aug. 2022, doi: 10.33395/jmp.v11i2.11657.

[3] F. S. Pamungkas, B. D. Prasetya, and I. Kharisudin, “Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python,” [Comparison of Supervised Learning Classification Methods on Bank Customers Data Using Python] vol. 3, 2020.

[4] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” [Comparison of Naïve Bayes and Support Vector Machine Methods on Twitter Sentiment Analysis*] SMATIKA J.*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.

[5] Desta Yolanda, Mohammad Hafiz Hersyah, and Eno Marozi, “Implementasi Metode Unsupervised Learning Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP,” [Implementation of Unsupervised Learning Method on Security System with IP Camera Storage Optimization] *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 6, pp. 1099–1105, Dec. 2021, doi: 10.29207/resti.v5i6.3552.

[6] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” [Getting to Know Machine Learning With Supervised and Unsupervised Learning Techniques Using Python*] BINA INSANI ICT J.*, vol. 7, no. 2, p. 156, Dec. 2020, doi: 10.51211/biict.v7i2.1422.

[7] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan,” [Implementation of K-means Clustring Algorithm in Determining Excellent Class Students*] J. Tekno Kompak*, vol. 15, no. 2, p. 25, Aug. 2021, doi: 10.33365/jtk.v15i2.1162.

[8] N. Sepriyanti, R. S. Nahampun, M. H. Zikri, I. Ambarani, and A. Rahmadeyan, “Implementation of K-Means Clustering to Group Poverty Levels in Riau Province,” [Implementation of K-Means Clustering to Group Poverty Levels in Riau Province] 2022.

[9] M. A. Mutaqin and W. Andriyani, “Klasterisasi Data Disabilitas Menggunakan Algoritma K-Means,” [Disability Data Clustering Using K-Means Algorithm] vol. 3, no. 1, 2022.

[10] D. Bahtiar, “Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering,” [Mapping of Social Assistance Recipients in Waru Jaya Village Using K-Means Clustering Algorithm] vol. 3, no. 2, 2023.

[11] M. Y. Matdoan, U. A. Matdoan, and M. S. Far-far, “Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting,” [K-Means Algorithm for Classification of Provinces in Indonesia Based on Stunting Service Package] vol. 1, no. 2, 2022.

[12] F. Febriansyah and S. Muntari, “Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam,” [Application of K-Means Algorithm for Clustering the Poor in Pagar Alam City] *JISKA J. Inform. Sunan Kalijaga*, vol. 8, no. 1, pp. 66–77, Jan. 2023, doi: 10.14421/jiska.2023.8.1.66-77.

[13] N. Nurahman and J. Susanto, "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means," [Clustering of Direct Cash Transfer Recipient Data Using the K-Means Algorithm] *JURIKOM J. Ris. Komput.*, vol. 10, no. 2, p. 461, Apr. 2023, doi: 10.30865/jurikom.v10i2.5807.

[14] A. Y. Permana, "Analisis dan Penerapan Metode Klasterisasi Menggunakan Algoritma K-Means Untuk Pengelompokkan Order Produk Pada PT. Arthautama Plasindo," [Analysis and Application of Clustering Methods Using the K-Means Algorithm for Product Order Grouping at PT Arthautama Plasindo] Vol. 10, 2019.

[15] Y. Prastyo, "Pembagian Tingkat Kecanduan Game Online Menggunakan K-Means Clustering serta Korelasinya Terhadap Prestasi Akademik," [Division of Online Game Addiction Levels Using K-Means Clustering and its Correlation with Academic Achievement*] Elinvo Electron. Inform. Vocat. Educ.*, vol. 2, no. 2, pp. 138–148, Dec. 2017, doi: 10.21831/elinvo.v2i2.17307.

[16] M. Y. Matdoan, F. Y. Risdiana, and G. Haumahu, "Application of the K-Means Cluster for the Classification of Disadvantaged Regencies/Cities in Maluku Province," *JRST J. Ris. Sains Dan Teknol.*, vol. 6, no. 1, p. 61, Nov. 2022, doi: 10.30595/jrst.v6i1.11637.