Journal of Advanced Zoology

ISSN: 0253-7214
Volume 44 Issue 04 Year 2023 Page 39:49

Application of the K-Means Method for Clustering Capture Fisheries Products in North Aceh with A Data Mining Approach

Nurdin[1*], Bustami[2], Rini Meiyanti[3], Amalia Fahada[4], Marleni[5]

[1,5]Department of Information Technology, Universitas Malikussaleh, Lhokseumawe, Indonesia.
[2,3,4]Department of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia.

**Corresponding author's E-mail: nurdin@unimal.ac.id*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Oct 2023

Abstract

North Aceh is one of the districts that has great potential in the maritime and fisheries sector, because part of the North Aceh region is a supplier of captured fisheries products. The problem with this research is that there is no model of a clustering system for captured fisheries products in North Aceh District, so it will be difficult to determine which areas produce superior and non-superior types of fish. This research aims to classify captured fishery results for the 2021-2022 period using the K-Means algorithm with a Data Mining approach to determine superior and non-superior fish types. Therefore, this research can help a little in providing information about the types of fish that are superior and not superior or the types of fish that are most numerous and the types of fish that appear the least at fishing ports in North Aceh District. The stages carried out in this research started from compiling research instruments and literature review, data collection and analysis, designing a clustering system model and implementing the system. This research produces 2 clusters, namely cluster 1 is superior fish and cluster 2 is non-superior fish. The K-Means algorithm with a Data Mining approach can be used to group types of capture fishery products at fishing ports in North Aceh District.

Keywords: *K-Means Method, Fisheries Products, Clustering, Data Mining*

1. Introduction

North Aceh is one of the districts that has great potential in the marine and fisheries sector, because part of the North Aceh region is a supplier of capture fisheries products (Fajriana, 2021). However, until now it has been difficult for us to determine which areas are producing fishery production in the province of Aceh (Nurdin et al., 2019; Nurdin et al., 2020). So we need a clustering application system for capture fisheries products using the K-Means Clustering method with a data mining approach.

The K-Means method is one of the most widely used clustering methods and has been applied in many fields of science and technology. One of the main problems of the k-means algorithm is that it can produce empty clusters depending on the initial center vector (Malki et al., 2016). The K-Means method is a partition algorithm, because K-Means is based on determining the initial number of groups by determining the initial centroid value. The K-Means algorithm uses an iterative process to obtain a cluster database (Arora & Varshney, 2016). It takes the desired initial number of clusters as input and produces the final number of clusters as output. The Euclidean Distance Formula is used to find the shortest distance between the center of mass and an object using random data. Data that has the shortest or closest distance to the centroid will form a cluster (Mardalius, 2017). The K-Means algorithm is a non-hierarchical data grouping method that attempts to partition existing data into two or more groups, so that data that has the same characteristics is put into the same group and data that has other characteristics is put into another group (Xiao et al., 2017).

Clustering is a method for searching and grouping data that has similar characteristics between one data and other data. Clustering is an unsupervised data mining method (Zhou et al., 2017) that is applied without training and does not require targets or outputs. In data mining, there are two types of clustering methods used to group data, namely hierarchical clustering and non-hierarchical clustering (Velmurugan & Santhanam, 2010). Data mining is the process of discovering useful patterns from large amounts of data, and due to the high availability of data, there is a need for powerful and automated tools to uncover valuable information from large amounts of data, which has become very pervasive in each data collection every area of daily life (Bhatnagar et al., 2012). Data mining is the process of extracting valid, previously unknown, understandable, and actionable information from large databases and using it to make important business decisions (Ha & Park, 1998). Furthermore, thzis framework strengthens the theoretical foundation of data mining by identifying systemic factors that are important at each stage of the data mining process (Debuse, 2007). The research results show that data mining technology can analyze statistical data from various angles and perspectives by modeling, classifying and grouping large amounts of data, as well as finding correlations between data (Guo et al., 2020). The data set is divided into several clusters through certain processes (Bulut et al., 2020; Nurdin et al., 2022).

The K-Means algorithm has been widely used because of its simple algorithm idea, easy to implement and high efficiency when processing large-scale data (Ismkhan, 2018; Sinaga et al., 2021) . The K-Means method and data mining have been widely applied in various case studies, one of which is the application of the K-Means method for clustering captured fisheries results using the data mining approach that this researcher carried out. The problem in this research is that there is no information system model for mapping and grouping capture fisheries products in North Aceh Regency, Aceh Province, so it will be difficult to determine which areas are producers of capture fisheries types. With this clustering system model, it is possible to identify areas that produce superior fisheries and non-superior fisheries. This research aims to develop an information system model for clustering captured fisheries results at fishing ports in North Aceh Regency using the K-Means algorithm with a data mining approach. The variables used in this research on the capture fisheries results clustering system consist of the type of fishery and the total fish catch in the 8 capture fisheries producing sub-districts. This research produces 2 clusters, namely cluster 1 (C1) is superior fish and cluster 2 (C2) is non-excellent fish.

Besides that, there are several other studies that have been conducted by researchers including, research Optimization and Computing Model of Fish Resource Supply Chain Distribution Network (Nurdin et al., 2021). Robust Optimization Approach for Agricultural Commodity Supply Chain Planningg (Nurdin et al., 2021), Lecture Scheduling System Using Welch Powell Graph Coloring Algorithm in Informatics Engineering Department of Malikussaleh University (Abdullah et al., 2019).

2. Literature Review

There are several related previous studies which became a literature review in this study, such as research conducted by Selvaraj et al., (2020) discussing time-series modeling of fish landings in the Colombian Pacific Ocean using the ARIMA model. The results of his research show that the ARIMA model is a suitable method for analyzing statistics. In data-deficient fisheries situations, this method can support the evaluation of fisheries production potential for decision making and management. Recreational fisheries are vulnerable to drastic declines, with analysis showing an overall decline in bonefish catches of 42%. The trend in the proportion of positive trips follows a pattern of declining capture, indicating major population changes beginning in 1999–2000. When bonefish catches decreased, species richness in bonefish runs increased by 34%, indicating a decline in bonefish abundance and a shift in fishing effort. Data limitations in most recreational fisheries, and the increasing use of catch and release as a fisheries management strategy indicate the need to develop further data integration tools to assess population trends and sustainability of fisheries resources (Santos et al., 2017). Management of fishery product resources requires mapping and clustering of fishery products by utilizing the Geographic Information System (GIS) model (Nurdin et al., 2023; Selvaraj et al., 2014).

Other related research carried out by Fitrianah et al., (2016) aims to analyze the determination of potential fishing zones based on mining approach data. The algorithm used in this research is AGRID+. The case study area is in the eastern Indian Ocean. The results of his research showed that the best clusters were formed in daily temporal aggregates with the number of cells (m) = 14 and the number of clusters formed was 50 clusters. To determine the fishing potential of zones for different temporal aggregates can be achieved by applying threshold techniques to the cluster results. Utilizing a data mining approach resulted in 22 daily prominent groups identified as potential fishing zones. The use of data mining can provide better input into the data provided, where further research can further explore various useful data mining tools to gain a better understanding of the available data to help improve the welfare of fishermen and aquaculture farmers in Indonesia (Anselmus Teniwut et al., 2019).

The following is some research related to the application of the K-Means method, such as research conducted by Sunarmo et al., (2020) implement machine learning techniques for partition grouping using the K-Means algorithm. This grouping is used for spatial representation of distribution VMS data in the WPPNRI-711 Fishery Management Area. Based on the Elbow Method, the optimal number of clusters obtained is 7, and the results of clustering with the K-Means Algorithm show that the data distribution in each cluster has a compactness value in percentage (90.7%), namely the level of similarity of members in the same cluster. The spatial distribution of VMS data in the WPPNRI-711 Fisheries Management Area is uneven. Research conducted by Saifullah et al., (2021) this research aims to detect fish using segmentation, namely segmenting fish images using K-Means clustering. The initial process carried out is preprocessing to perfect the image. Preprocessing is carried out twice, namely before segmentation using K-Means and after. Preprocessing stage 1 uses resize and reshape. The preprocessing results are segmented using K-Means clustering. The processed object provides a clear picture of the fish object so that K-Means segmentation can help detect fish objects. A new fish image segmentation method combining K-means clustering segmentation algorithm and mathematical morphology has been proposed, which is more accurate and stable than Otsu and other segmentation algorithms (Yao et al., 2012).

Research conducted by Hablum et al., (2019) the K-Means algorithm succeeded in classifying fish catches for the 2015-2017 period using 2 groups, namely group one was categorized as the few catches, and group two was categorized as the most catches. The initial cluster center or centroid is adjusted to the number of variables present. Research conducted by Sugiarto et al., (2020) Geographic Information System for fisheries and livestock areas in Pasuruan Regency. This study explains that the geographic information system of Pasuruan Regency has been able to provide information including fishery areas, livestock, the amount of production or yields per year.

Research conducted by Nurdin et al., (2022) information System for Predicting Fisheries Outcomes Using Multiple Linear Regression Algorithms. This information system can predict the yield of capture fisheries in Bireun Regency in 2021 of 12,813.870305238 Tons. This system is an alternative to a manual prediction system. The variables used in this study consisted of the number of fish caught, the number of motor boats and the number of rainy days. Research conducted by Saifullah et al., (2018) aims to detect fish using segmentation, namely segmenting fish images using the K-Means algorithm. Image segmentation is a concept that is often used for object detection.

Research conducted by Anas & Rais (2020) aims to map risk areas based on the number of natural disasters that have occurred. One of the clustering methods used in this study is the K-Means method. Although the K-Means method can analyze data well, this method has not been able to provide detailed information regarding disaster-prone areas. To overcome this weakness, the grouping carried out by the K-Means method is then applied to a Geographical Information System (GIS) to map the type of disaster that is used as an identifier variable for a disaster area.

3. Materials And Methods

This research method was carried out by building and developing a software model for a clustering information system for captured fisheries products using the K-Means algorithm with a data mining approach with the following research stages:

1. Literature study

Literature study is a stage to discuss the theory or method used to support this research which contains literature from other relevant research journals.

2. Data collection

Data collection is a stage to collect the data needed to be input into the system, namely data on capture fisheries products on the north coast of Aceh.

3. Analysis of system requirements

System requirements analysis is the stage for analyzing the system to be built, after the system analysis results are obtained, the next step is system design. The results of the system analysis will be a reference for the design of the system to be built.

4. System design

System design is a stage for designing software or user interface design and database design using a programming language which is described in the form of a model diagram.

5. System implementation

System implementation is a stage that discusses the results of research implementation from the resulting system design that explains the features that exist in software applications.

K-Means Clustering Algorithm

The k-means algorithm is used to determine the number of clusters (Escorcia Guzman, 2021), formed through the use of a specific condition known as criterion, which is involved in the optimal, the splitting method utilizes a condition called as criterion, which is involved in the optimal division of the dataset set by appropriate optimization problems (Borlea et al., 2021). K-means provides a more comprehensive view of applicant characteristics and needs; using K-means clustering, it is possible to identify the key characteristics of each potential data cluster (Abdullah et al., 2022). Data that has a representative value similarity in one group and data that has a difference in another group so that it allows grouping different data that has a small level of variation. The main principle of this technique is to construct K centroid mass partitions from a set of data, Using the Euclidean Distance formula, calculate the distance between each input data point and each centroid (Sulistiyawati & Supriyanto, 2021).

$$D_{(x,y)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

Where, $D_{(x,y)}$ represents *Euclidean Distance*, X_1 represents first data training, Y_1 first data testing, X_2 represents second data training, Y_2 represents second data testing, X_n represents *n* data training, and Y_n represents *n* data testing.

The process stages in implementing the K-Means Clustering algorithm are as follows (Kane et al., 2016):

1. Determine the value of k as the number of clusters to be formed;
2. Initialize k cluster centers in a random way from the dataset;
3. Calculating the distance of each input data to each centroid using the Euclidean Distance formula;
4. Classify each data based on the closest distance to the centroid;
5. Update the centroid value, the new centroid value is obtained from the cluster average;
6. Repeat from step 2 to 5, until nothing changes in the members of each cluster.

System Scheme for Fishery Clustering

The scheme of the capture fisheries product clustering system using the K-Means method starts from inputting fishery clustering variable data and selecting the initial centroid randomly. Then the system will perform an initial centroid calculation using the Euclidean Distance formula for Iteration-1. Then proceed with the next centroid calculation using the K-Means Clustering Method for Iteration-2. If the calculation results do not change from the previous iteration, then the process is complete and results in a clustering of fishery products. The scheme of the clustering system for capture fisheries products in North Aceh Regency using the K-Means Clustering method can be seen in Figure 1.

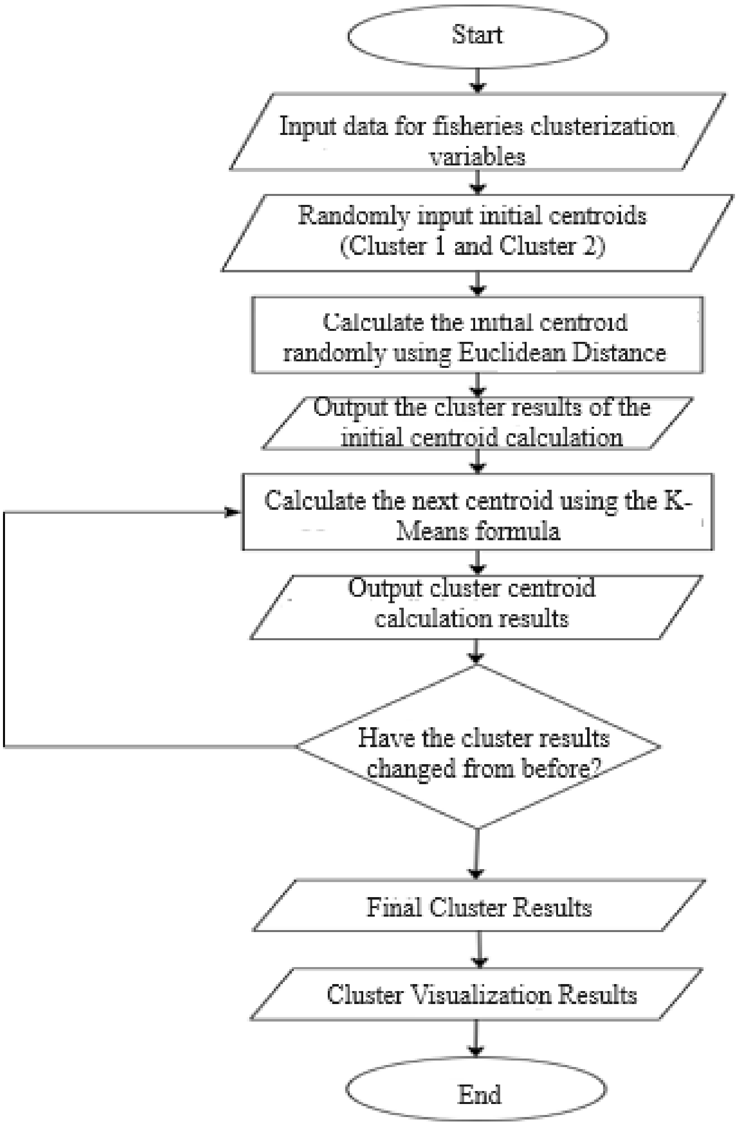

Figure 1. Fisheries Clusterization System Scheme.

3. Results and Discussion

Variable Attribute Values and Dataset Used

In this study, to obtain the results of calculating superior fish clusters and non- superior fish clusters in North Aceh Regency using a dataset of capture fisheries production results in 2022 using 2 variables, namely: type of fishery and number of fish catches in 8 sub-districts producing capture fisheries and each sub-district was given a value variables with attributes (X1: Dewantara District, X2: Lapang District, X3: Muara Batu District, X4: Samudera District, X5: Seuneudon District, X6: Syamtalira Bayu District, X7: Tanah Jambo Aye District and X8: Tanah Pasir District). The number of datasets used in this research was 78 fish species data. In this dataset there are types of fish and total fish catches in each sub-district in North Aceh Regency as in Table 1 below.

Table 1: Capture Fisheries Production Dataset in 2022

No.	Types of Fish	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	Ikan Terbang (*Flying Fish*)	0,00	12,73	368,76	26,48	0,00	0,00	0,00	0,00
2	Belanak (*Mullet*)	28,20	20,37	84,62	16,10	62,32	20,48	39,91	13,05
3	Bentong (*Oxeye/Bigeyo Scad*)	5,33	6,08	9,50	8,24	3,04	12,04	2,66	3,17
4	Kerapuh balong (*Honeycom grouper*)	7,79	4,90	15,08	10,05	8,17	2,39	3,35	9,93
5	Bawal putih (*Silver pomfred*)	3,02	5,03	14,57	18,34	11,43	2,01	2,58	5,90
6	Julung-Juling (*Gerfish*)	20,53	16,97	91,10	14,32	79,43	26,35	27,11	22,80
7	Banyar (*Indian Mackerel*)	13,94	23,18	68,92	21,54	35,73	17,01	35,73	15,08
8	Kembung (*Short Body*)	45,45	58,84	102,10	37,08	49,67	36,62	49,67	32,68
9	Layang (*Scad*)	28,46	17,13	39,91	17,61	53,09	11,14	53,09	15,07
10	Tembang (*Frigescole*)	29,40	29,26	154,18	19,64	124,28	60,94	32,56	20,39

11	Siro (*Bali Sarilla*)	12,79	17,23	74,24	20,46	21,28	5,57	9,13	14,32
12	Selar (*Trevalles*)	11,42	5,71	53,58	16,83	19,89	8,61	4,81	12,41
13	Sunglir (*Rainbow*)	23,57	14,32	73,60	22,04	22,42	3,80	11,53	15,33
14	Cakalang (*Skipjack tuna*)	37,03	44,45	406,90	23,52	154,24	21,28	5,95	88,47
15	Tongkol Krai (*Frigate Tuna*)	32,56	27,74	33,05	12,41	27,62	10,84	25,38	33,71
16	Kenyar (*Stripped Bonita*)	8,61	18,49	11,46	11,14	5,95	3,97	11,28	11,68
17	Terubuk (*Hilso Shad*)	39,02	105,28	61,70	15,33	19,64	11,66	10,52	6,46
18	Kapas kapas (*Fels cravelly*)	32,92	15,70	71,63	30,79	13,95	10,68	6,66	22,87
19	Selanget (*Chacusda*)	14,70	8,61	69,05	18,11	19,51	5,33	10,77	12,55
…	………………	……	……	……	……	…..	…..	……	…..
78	Bawal Putih (*Silver pomfred*)	3,02	5,03	14,57	18,34	11,43	2,01	2,58	5,90

Application of the K-Means Method

The following are the stages of the clustering process for captured fisheries products in North Aceh Regency using the K-Means Method 1. Determine the number of clusters In this clustering study of captured fishery products, 2 clusters were used, consisting of: Superior Fish Cluster (C1) and Non-Superior Fish Cluster (C2). 2. Data normalization and initialization of cluster center values Researchers normalized the data and chose the following random centroids to be used in the manual clustering calculation process in this study.

Cluster Center Data Normalization

In Table 1 is data that has not yet been processed and must be normalized. The purpose of this data normalization is for harmony between attributes with different numerical magnitudes. The equations used in data normalization use minimum and maximum functions on the attributes to be recalculated. The data normalization is carried out using the equation:

$$Normalized\ Value = \frac{(Initial\ Value) - (Min\ Value)}{(Max\ Value) - (Min\ Value)}$$

The following is a normalization calculation for data on the mullet fish type as follows.

Dewantara District (X_1)

$$Normalized\ Value = \frac{(28{,}37) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 0{,}211681$$

Lapang District (X_2)

$$Normalized\ Value = \frac{(20{,}73) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 0{,}240625$$

Muara Batu District (X_3)

$$Normalized\ Value = \frac{(84{,}62) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 1$$

Samudra District (X_4)

$$Normalized\ Value = \frac{(16{,}1) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 0{,}258214$$

Seunuddon District (X_5)

$$Normalized\ Value = \frac{(62{,}32) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 1$$

Syamtalira Bayu District (X_6)

$$Normalized\ Value = \frac{(20{,}48) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 0{,}513022$$

Tanah Jambo Aye District (X_7)

$$Normalized\ Value = \frac{(39{,}91) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 1$$

Tanah Pasir District (X_8)

$$Normalized\ Value = \frac{(13{,}05) - (13{,}05)}{(84{,}62) - (13{,}05)}$$

$$= 1$$

Based on the data normalization results above, the next step is to initialize the cluster center values after all the attributes of the fish species data have been calculated and obtain data that has the same numerical magnitude. The number of clusters in this study is 2 clusters consisting of cluster (C1) of superior fish and cluster (C2) of non-superior fish. Then, for each cluster, the central point value for each available attribute is determined. The initial cluster value is chosen randomly as in Table 2 below.

Table 2: Cluster Center Values (Initial Centroid)

Cluster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
C1	1	0,186828	1	0,258891	1	0,87964	1	1
C2	0	0,150446	1	0,602855	1	0,476071	1	1

Euclidean Distance Calculation

After finding the centroid, the initial calculation will be used to calculate the distance between the data and the centroid. The equation used to calculate distance in this research uses the Euclidean Distance equation:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

This equation can be implemented into the data as follows.

$$D_{C1} = \sqrt{\begin{array}{c}(0-1)^2 + (0{,}034-0{,}02)^2 + (1-0{,}023)^2 + (0{,}071-0)^2 \\ +(0-0{,}005)^2 + (0-0{,}004)^2 + (0-0{,}005)^2 + (0-0{,}003)^2\end{array}}$$

$$D_{C1} = \sqrt{1{,}960}$$

$$D_{C1} = 1{,}4$$

$$D_{C2} = \sqrt{\begin{array}{c}(0-0)^2 + (0{,}034-0{,}168)^2 + (1-1)^2 + (0{,}071-0{,}138)^2 \\ +(0-0{,}396)^2 + (0-0{,}055)^2 + (0-0{,}396)^2 + (0-0{,}02)^2\end{array}}$$

$$D_{C2} = \sqrt{0{,}339}$$

$$D_{C2} = 0{,}583$$

This calculation was carried out on all sample data so that data from the first clustering iteration was obtained in Table 3 below.

Table 3: Iteration Calculation Results-1

No	C1	C2	Distance	Cluster
1	1.40015	0.58305	0.58305	Non-superior
2	1.48327	0.38311	0.38311	Non-superior
3	1.57814	1.25609	1.25609	Non-superior
4	1.48514	0.85552	0.85552	Non-superior
5	0	1.51956	0	Superior
6	1.4914	0.42257	0.42257	Non-superior

7	1.45648	0.4978	0.4978	Non-superior
8	1.49768	0.91478	0.91478	Superior
…	….	….	….	……..
78	1.68876	1.00682	1.00682	Non-superior

Based on the calculation above, a new centroid is generated for the next iteration of calculation with the new centroid value in Table 4 below.

Table 4: New centroid in second iteration

C1	0.945502	0.430821	0.038842	0.038842	0.231357	0.224862	0.274601	0.001545
C2	0.230924	0.22096	0.862739	0.244813	0.493849	0.143785	0.21749	0.162105

After the new centroid is calculated, the next step is to compare it with the initial centroid value. The calculated values for the centroid values for the second iteration can be seen in Table 5 below.

Table 5: Iteration Calculation Results-2

No	C1	C2	Distance	Cluster
1	1.46814	0.6892	0.6892	Non-superior
2	1.34312	0.40598	0.40598	Non-superior
3	1.39369	1.0777	1.0777	Non-superior
4	1.26905	0.54231	0.54231	Non-superior
5	0.60015	1.31599	0.60015	Superior
6	1.44173	0.34271	0.34271	Non-superior
7	1.35478	0.4161	0.4161	Non-superior
8	1.43494	0.64416	0.64416	Superior
…	……	…….	…..	………
78	1.60681	0.82198	0.82198	Non-superior

If the values are not the same, then the next iteration is carried out until the new centroid value is the same. So, based on the results in Table 3 and Table 5, the results were not the same, and the next iteration was carried out by inserting a new centroid again as seen in Table 6 below.

Table 6: New centroid in third iteration

C1	0.945502	0.430821	0.038842	0.038842	0.231357	0.224862	0.274601	0.001545
C2	0.230924	0.22096	0.862739	0.244813	0.493849	0.143785	0.21749	0.162105

After calculating the new centroid, the centroid value was calculated again and the results obtained were as shown in Table 7 below.

Table 7: Iteration Calculation Results-3

No	C1	C2	Distance	Cluster
1	1.46814	0.6892	0.6892	Non-superior
2	1.34312	0.40598	0.40598	Non-superior
3	1.39369	1.0777	1.0777	Non-superior
4	1.26905	0.54231	0.54231	Non-superior
5	0.60015	1.31599	0.60015	Superior
6	1.44173	0.34271	0.34271	Non-superior
7	1.35478	0.4161	0.4161	Non-superior
8	1.43494	0.64416	0.64416	Superior
…	…..	…..	…..	…….
78	1.60681	0.82198	0.82198	Non-superior

After checking the last iteration, namely the third iteration, the iteration is the same as the second iteration, which indicates that the calculation can be stopped in the third iteration. So it can be concluded that calculations using the K-Means method produce results for clustering capture fisheries results in North Aceh Regency, there are 30 types of superior fish (C1) and 48 types of fish that are not superior (C2). The following is a graph of the results of capture fisheries clustering using the K-Means algorithm.

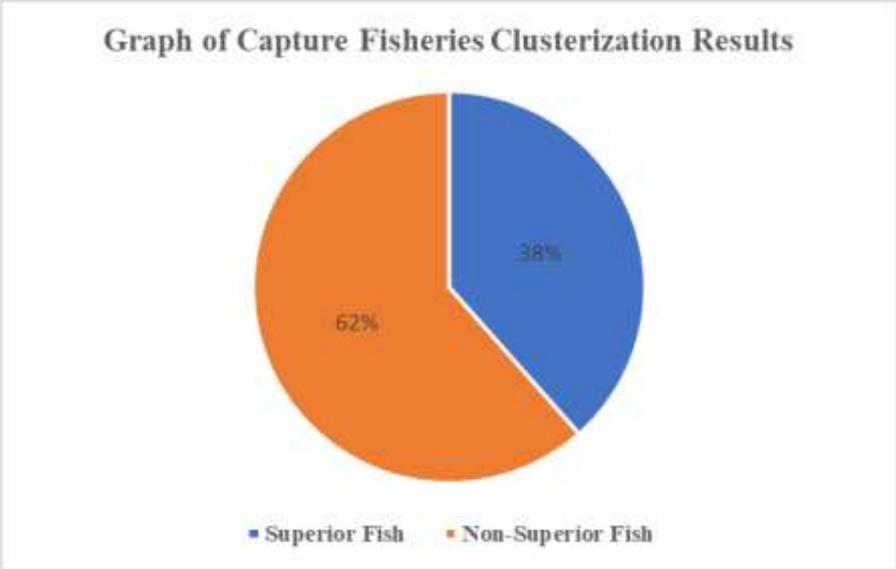

Figure 2. Graph of Capture Fisheries Clusterization Results

4. Conclusion

Research on the application of the K-Means method for clustering capture fishery results in North Aceh Regency using a dataset of capture fishery production results in 2022 using 2 variables, namely: type of fishery and total fish catch in 8 capture fishery producing sub-districts, and each sub-district is given a variable value with attributes (X1: Dewantara District, X2: Lapang District, X3: Muara Batu District, X4: Samudera District, X5: Seuneudon District, X6: Syamtalira Bayu District, X7: Tanah Jambo Aye District and X8: Tanah Pasir District). Based on the clustering results using the K-Means algorithm, the results showed that there were 30 superior fish clusters (C1) and 48 non-superior fish clusters (C2), data on these types of fish were spread across 8 sub-districts in North Aceh Regency. The results of this clustering are visualized on the Geographic Information System Website. The results of the clusters formed using the K-Means method are very dependent on the initial cluster value determined. In this research, the K-Means method with a data mining approach can be used to cluster capture fisheries results in North Aceh Regency. This application system was built and designed using the PHP programming language and MYSQL database. The contribution of this research is to help the community and the fisheries department to obtain information on the types of captured fishery products (superior fish and non-superior fish) at fishing ports in North Aceh Regency, Aceh, Indonesia. Further research is recommended to use an approach with algorithms or other clustering methods as a comparison to find out which method is better for clustering.

Acknowledgments

We acknowledge the support grant and research funding received from The LPPM Universitas Malikussaleh in 2023.

References:

Abdullah, D., Nurdin, N., Yaton, M., et al. (2019). Lecture Scheduling System Using Welch Powell Graph Coloring Algorithm in Informatics Engineering Departement of Universitas Malikussaleh. IOP Conf. Series: Journal of Physics: Conf. Series, 1363, 012074.

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Qual. Quant, 56(3), 1283–1291. https://doi.org/10.1007/s11135-021-01176-w

Annas, S., & Rais, Z. (2020). K-Means and GIS for Mapping Natural Disaster Prone Areas in Indonesia. Pervasive Health: Pervasive Computing Technologies for Healthcare, 1, 340–462.

Anselmus Teniwut, W., Pentury, F., & Anastasia Ngamel, Y. (2019). Forecasting the welfare of fishermen and aquaculture farmers in Indonesia: Data Mining Approach. J. Phys. Conf. Ser, vol. 1175(1). https://doi.org/10.1088/1742-6596/1175/1/012066.

Arora, P., & Varshney, S. (2016). Analysis of K-means and K-medoids algorithm for big data. Procedia Computer Science, 78, 507-512.

Bhatnagar, A., Jadye, S., & Nagar, M. (2012). Data Mining Techniques and Distinct Applications: A Literature Review. International journal of engineering research and technology, 1(9).

Borlea, I. D., Precup, R. E., & Borlea, A. B. (2021). Improvement of K-means Cluster Quality by Post Processing Resulted Clusters. Procedia Comput. Sci, 199, 63–70. https://doi.org/10.1016/j.procs.2022.01.009.

Bulut, H., Onan, A., & Korukoglu, S. (2020). An improved ant-based algorithm based on heaps merging and fuzzy c-means for clustering cancer gene expression data. S¯adhan¯a, 45(160), 1–17.

Debuse, J. (2007). Extending data mining methodologies to encompass organizational factors. Systems Research and Behavioral Science, 24, 183-190. https://doi.org/10.1002/SRES.823.

Escorcia Guzman, J. H., Zuluaga-Ortiz, R. A., Barrios-Miranda, D. A., & Delahoz-Dominguez, E. J. (2021). Information and Communication Technologies (ICT) in the processes of distribution and use of knowledge in Higher Education Institutions (HEIs). Procedia Comput. Sci, 198, 644–649, https://doi.org/10.1016/j.procs.2021.12.300.

Fajriana, F. (2021). Analysis of the K-Medoids Algorithm in the Capture Fisheries Production Clustering System in North Aceh District. JEPIN (Jurnal Edukasi dan Penelitian Informatika), 7(2), 263-269.

Fitrianah, D., Fahmi, H., Hidayanto, A. N., & Arymurthy, A. M. (2016). A Data Mining Based Approach for Determining the Potential Fishing Zones. Int. J. Inf. Educ. Technol, 6(3), 187–191. https://doi.org/10.7763/ijiet.2016.v6.682.

Guo, Y., Wang, N., Xu, Z., & Wu, K. (2020). The internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. Mechanical Systems and Signal Processing, 142. https://doi.org/10.1016/j.ymssp.2020.1066

Ha, S., & Park, S. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. Expert Systems With Applications, 15, 1-31. https://doi.org/10.1016/S09574174(98)00008-6.

Hablum, R. J., Khairan, A., & Rosihan. (2019). Clustering Hasil Tangkap Ikan di Pelabuhan Perikanan Nusantara (PPN) Ternate Menggunakan Algoritma K-Means. JIKO (Jurnal Informatika dan Komputer) Ternate, 2(1), 26-33. http://dx.doi.org/10.33387/jiko.v2i1.1053

https://doi.org/10.1093/ICESJMS/FST204.

Ismkhan, H. (2018). I-k-means-plus: An iterative clustering algorithm based on an enhanced version of the k-means. Pattern Recognition, 79, 402–413. https://doi.org/10.1016/j.patcog.2018.02.01

Kane, S. N., Mishra, A., & Dutta, A. K. (2016). Preface: International Conference on Recent Trends in Physics (ICRTP 2016). J. Phys. Conf. Ser, 755(1). https://doi.org/10.1088/1742-6596/755/1/011001

Léopold, M., Guillemot, N., Rocklin, D., & Chen, C. (2014). A framework for mapping small-scale coastal fisheries using fishers' knowledge. Ices Journal of Marine Science, 71, 1781-1792.

Malki, A., Rizk, M., El-Shorbagy, M., & Mousa, A. (2016). Hybrid Genetic Algorithm with K-Means for Clustering Problems. Open Journal of Optimization, 5(2), 71-83. https://doi.org/10.4236/OJOP.2016.52009.

Mardalius. (2017). Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan (Studi Kasus : Siswa Sma Negeri 1 Ranah Pesisir). Proceding SEMILOKA, 1, 105, https://doi.org/10.31219/osf.io/6mec3

Nurdin, N., Bustami, B., & Maryana, M. (2021). Robust Optimization Approach for Agricultural Commodity Supply Chain Planningg. Journal of Theoretical and Applied Information Technology, 99(2), 304–315.

Nurdin, N., Fajriana, F., Maryana, M., & Zanati, A. (2022). Information System for Predicting Fisheries Outcomes Using Regression Algorithm Multiple Linear. JITE (Journal of Informatics and Telecommunication Engineering), 5(2), 247-258.

Nurdin, N., Fitriani, S., & Yunizar, Z. (2022). Clustering the Distribution of COVID-19 in Aceh Province Using the Fuzzy C-Means Algorithm. JTAM (Jurnal Teori dan Aplikasi Matematika), 6(3), 665–677. https://doi.org/10.31764/jtam.v6i3.8576

Nurdin, N., Taufiq, T., Fajriana, F., & Ulhaq, M. Z. (2023). Mapping System Model and Clustering of Fishery Products using K-Means Algorithm with Web GIS Approach. Int J Intell Syst Appl Eng, 11(3), 738–749.

Nurdin, N., Zarlis, M., Tulus., & Efendi, S. (2021). Optimization and Computing Model of Fish Resource Supply Chain Distribution Network. IOP Conf. Series: Journal of Physics: Conf. Series, 1898, 012022.

Nurdin, N., Zarlis, M., Tulus., & Efendi, S. (2019). Data Driven Optimization Approach to Fish Resources Supply Chain Planning in Aceh Province. IOP Conf. Series: Journal of Physics: Conf. Series 1255 012081.

Nurdin, N., Zarlis, M., Tulus., & Efendi, S. (2020). Mixed Integer Linear Programming Model for Integrated Fish Supply Chain Planning. Journal of Theoretical and Applied Information Technology, 98(12), 2017-2028.

Saifullah, S., Andiko, P. S., & Bambang, Y. (2018). Fish Detection Using Morphological Aproach Based-On K-Means Segmentation. Image and Video Processing (Eess.IV) 1, 1–9.

Saifullah, S., Suryotomo, A. P., & Yuwono, B. (2021). Fish detection using morphological approach based-on k-means segmentation. Compiler, 10(1). https://doi.org/10.28989/compiler.v10i1.946.

Santos, R., Rehage, J., Adams, A., Black, B., Osborne, J., & Kroloff, E. (2017). Quantitative assessment of a data-limited recreational bonefish fishery using a time-series of fishing guides reports. PLoS ONE, 12. https://doi.org/10.1371/journal.pone.0184776

Selvaraj, J., ARUNACHALAM, V., Coronado-Franco, K., Romero-Orjuela, L., & Ramírez-Yara, Y. (2020). Time-series modeling of fishery landings in the Colombian Pacific Ocean using an ARIMA model. Regional Studies in Marine Science, 39, 101477. https://doi.org/10.1016/J.RSMA.2020.101477

Sinaga, K. P., Hussain, I., & Yang, M. S. (2021). Entropy K-Means Clustering with Feature Reduction Under Unknown Number of Clusters. IEEE Access, 9, 67736–67751.

Sugiarto, A. S., Hani, Z. Z., & Yosep, A. P. (2020). Sistem Informasi Geografis Daerah Perikanan Dan Peternakan Kabupaten Pasuruan. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1). https://doi.org/10.36040/jati.v4i1.232.

Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan. J. Tekno Kompak, 15(2), 25-36. https://doi.org/10.33365/jtk.v15i2.1162

Sunarmo., Affandi, A., & Sumpeno, S. (2020). Clustering Spatial Temporal Distribution of Fishing Vessel Based lOn VMS Data Using K-Means. 2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020, 1–6. https://doi.org/10.1109/ICOIACT50329.2020.9331994.

Velmurugan, T., & Santhanam, T. (2010). Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. Journal of Computer Science, 6(3), 363-368. https://doi.org/10.3844/jcssp.2010.363

Xiao, J., Lu, J., & Li, X. (2017). Davies Bouldin Index based hierarchical initialization K-means. Intelligent Data Analysis, 21(6), 1327-1338. https://doi.org/10.3233/IDA-163129

Yao, H., Duan, Q., Li, D., & Wang, J. (2012). An improved KK-means clustering algorithm for fish image segmentation. *Math. Comput. Model,* 58, 790-798. https://doi.org/10.1016/j.mcm.2012.12.025.

Zhou, J., Wang, Q., Hung, C.C., & Yang, F. (2017). Credibilistic clustering algorithms via alternating cluster estimation. J. Intell. Manuf, 28, 727–738.