International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Mapping System Model and Clustering of Fishery Products using K-Means Algorithm with Web GIS Approach

Nurdin*[1], Taufiq[2], Fajriana[3], Muhammad Zia Ulhaq[4]

Submitted: 27/04/2023 **Revised**: 26/06/2023 **Accepted**: 06/07/2023

Abstract*:* Aceh is a province that is rich in fishery product resources, however, there are several districts/cities experiencing a shortage of fishery product supplies. Therefore, a model of fishery product mapping and clustering system is needed with the Web Geographic Information System (GIS) approach. This study aims to develop an information system model for mapping and clustering capture fisheries products from 10 fishing ports located on the north coast of Aceh Indonesia. The method used for clustering uses the K-Means algorithm and its mapping using the GIS web. The variables used in the fishery product type clustering system consist of catch weight, fish price, number of fishing ports and number of months to determine fish season. This research resulted in the output of 2 clusters, namely cluster 1 is a superior fish and cluster 2 is an ordinary fish. This K-Means algorithm can be used for clustering types of fishery products, while Web GIS can be applied to mapping fishing ports on the north coast of Aceh Indonesia. This system model is a combination of data mining and a website that is visualized in the form of a GIS. The contribution of this research is to help the community and the fisheries service to obtain information on types of fishery products (high-quality fish and common fish species) every month at fishing ports and information about their selling prices.

Keywords: *K-Means Algorithms, Web GIS, Fisheries Products, Mapping and Clustering Models, Implementation and Testing*.

1. Introduction

Aceh is one of the provinces in Indonesia which is rich in potential marine and fishery resources. Many resources from capture fisheries are superior commodities, because parts of Aceh are suppliers of capture fisheries products [1]. However, until now it has been difficult for us to determine which areas are producing fishery production in the province of Aceh [2] [3], so a model application system for mapping and clustering fishery products is needed with a web geographic information system approach. The Maritime Affairs and Fisheries Service as a means of driving the community's economy must ensure the availability of fishery product data, for this reason a mapping and clustering system is needed using the K-Means Clustering Algorithm to help classify fishery products in each Regency. The Office of Maritime Affairs and Fisheries also needs a spatial data visualization facility to monitor fishery products and as a forum for publishing this information to the public in the form of geographic maps. Web GIS is a combination of mapping graphic design, digital maps with geographic analysis, computer programming and an interconnected database into one part of web design and web mapping. Geographic Information System (GIS) is a special information system that manages data that has spatial information or a computer system that has the ability to build, store, manage and display geographic-based information [4], [5].

Clustering is a method for finding and grouping data that has similar characteristics between one data and another. Clustering is a data mining method that is unsupervised which is applied without training and does not require targets or outputs. In data mining there are two types of clustering methods used in grouping data, namely hierarchical clustering and non-hierarchical clustering [6]. While the K-Means algorithm is one of the partitional algorithms, because K-Means is based on determining the initial number of groups by defining the initial centroid value. The K-Means algorithm uses an iterative process to get a cluster database [7]. It takes the desired number of initial clusters as input and produces the final number of clusters as output. The Euclidean Distance formula is used to find the shortest distance between centroid points and objects using random data. Data that has the shortest or closest distance to the centroid will form a cluster [8]. The K-Means algorithm is a non-hierarchical data grouping method that seeks to partition existing data into two or more groups, so that data with the same characteristics is included in the same group and data with other characteristics is included in another group [9].

The K-Means Algorithm has been widely applied in various case studies, such as the Application of the K-Means Clustering Algorithm for Determining Fire Prone Areas

[1]Department of Information Technology, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia
ORCID ID: 0000-0003-1887-0437
[2]Department of Electrical Engineering, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia
[3]Department of Mathematics Education, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia
[4]Student Department of Information Technology, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia
** Corresponding Author Email: nurdin@unimal.ac.id*

[10], Implementation of the K-Means Algorithm for Clustering Feasible Land for Corn Planting in South Lampung Regency [11], Application of the K-Means Clustering algorithm for NBA Guards Classification [12], Integration of the K-Means Clustering Method and the Elbow Method for Identifying the Best Customer Profile Clusters [13], and review of the K-Means Clustering algorithm approach on other data models [14]. Besides this K-Means algorithm, there are also many clustering algorithms that are applied in educational data mining [15], including the naïve Bayes classifier algorithm used for data classification, such as data classification of student scientific work using the naïve Bayes classifier method [16].

The problem in this study is that there is no model of information system for mapping and clustering capture fisheries products on the north coast of Aceh province, so it will be difficult to determine which areas are producers of capture fisheries. With this clustering system model, it is possible to find out which areas produce superior fisheries and ordinary fisheries. This study aims to develop an information system model for mapping and clustering capture fisheries products from 10 fishing ports on the north coast of Aceh using the K-Means algorithm with a WebGIS approach. The variables used in the research on fishery product mapping and clustering systems consist of catch weight, fish price, number of fishing ports and number of months to determine fish season. This research resulted in the output of 2 clusters, namely cluster 1 is a superior fish and cluster 2 is an ordinary 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 [17]. Robust Optimization Approach for Agricultural Commodity Supply Chain Planningg [18], Searching the shortest route for distribution of LPG in Medan City using ant colony algorithm [19], Lecture Scheduling System Using Welch Powell Graph Coloring Algorithm in Informatics Engineering Department of Malikussaleh University [20] and Implementation of the BFS algorithm and web scraping techniques for online shop detection in Indonesia [21].

2. Literature Review

There are several related previous studies which became a literature review in this study, as 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 [22]. 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 [23]. Research conducted by (Nurdin et al., 2020) 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 [24]. Research conducted by (Saifullah et al., 2021) 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 [25].

Research conducted by (Annas & 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 [26]. Another study using the K-Means Clustering algorithm is the grouping of provinces based on population density, school enrollment rates of 13-15, human development index and open unemployment rate of 5 clusters centered in South Sumatra, Lampung, DKI Jakarta, Central Java and West Kalimantan [27]. Research conducted by [28] clustering Information and Communication Technology competencies for SMK students used the K-Means Clustering Algorithm, this study used 3 clusters namely very competent, competent and less competent clusters with the results of testing 10 students in cluster 1 with predicate very competent, 64 students in cluster 2 with a competent predicate, and 10 students in cluster 3 with a less competent predicate. Research conducted by [29] this study discusses the analysis of student academic performance using the Fuzzy C-Means Clustering method [30]. Based on the results of the clustering process of 4255 student academic data, 4 clusters were obtained. The clusters with the best achievement category were cluster 3 with 1753 students, cluster 4 with 1496 students, cluster 2 with 676 students and cluster 1 with 330 students.

Research conducted by [31] research results the number of clusters formed there are 3. In the first cluster the category of satisfied with online learning is categorized as "high" with a value of 9.33 while the dissatisfied category is categorized as "low" with a value of 0.67. In the second cluster, the happy category with online learning was classified as "low" with a value of 4.73, while the dissatisfied category was rated as "high" with a value of 5.27

and in the third cluster, those who were satisfied or dissatisfied had no value (0.0). Research conducted by [32] This research uses K-Means and Fuzzy C-Means cluster analysis to group students into three groups based on their learning outcomes. In the first cluster 2.63% got low scores, in the second cluster 23.68% got scores around the average. And in the third cluster 73.68% get high scores.

3. Research Methodology

3.1. Description of Problem Formulation

It is difficult for fisheries stakeholders to obtain information on the types of fishery products available at each fishing port in Aceh province and the community also has difficulty getting the types of fish they are looking for and the need for data collection on superior and common fish species at each fishing port, so researchers create a mapping application model and Clustering of capture fisheries products on the north coast of Aceh uses the K-Means algorithm with a Web GIS approach to be able to help the community and marine and fisheries services in obtaining information on mapping fish catches at every port on the north coast of Aceh, where this application model combines data mining with the website and will be visualized in the form of a Geographic Information System, so that it can be accessed anywhere and anytime just by accessing the website address. This study was to classify capture fisheries products from 10 ports on the north coast of Aceh using the K-Means Clustering algorithm to obtain superior fish clusters and ordinary fish clusters.

3.2. Research Stage

This research method was carried out by building and developing a software model for mapping information systems and clustering fishery products using the K-Means algorithm with a Web GIS 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.

6. System testing

   System testing is a stage that discusses the results of system trials from research that has been done, namely in the form of software application testing results. The following are the stages of this research

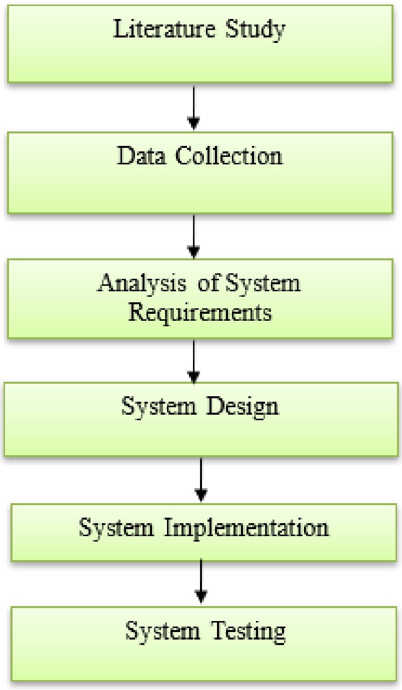

Fig. 1. Research stages

3.3. K-Means Clustering Algorithm

The k-means algorithm is used to determine the number of clusters [33], 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 [34]. 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 [35]. 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 [36].

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

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 [37] :

a. Determine the value of k as the number of clusters to be formed;

b. Initialize k cluster centers in a random way from the dataset;

c. Calculating the distance of each input data to each centroid using the Euclidean Distance formula;

d. Classify each data based on the closest distance to the centroid;

e. Update the centroid value, the new centroid value is obtained from the cluster average;

f. Repeat from step 2 to 5, until nothing changes in the members of each cluster.

The following is a Flowchart of Data Clusterization Stages using the K-Means Algorithm.

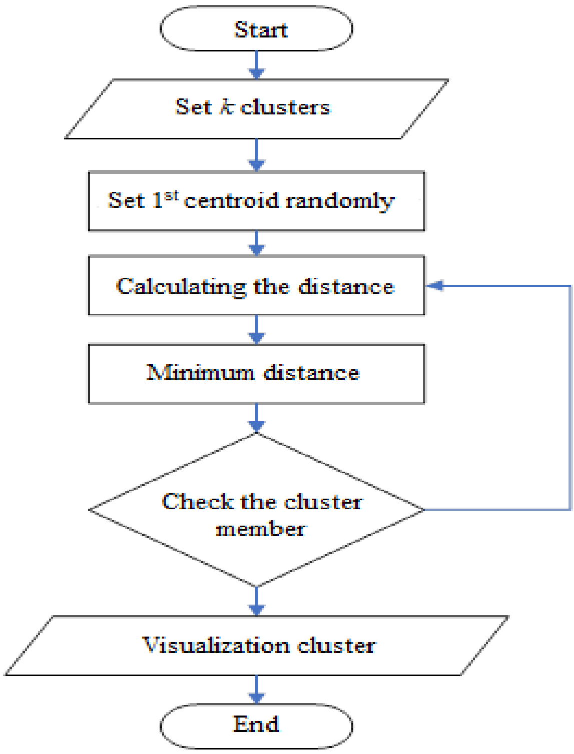

Fig. 2. Data clustering flowchart using K-Means Algorithm

3.4. Data Collection and Variable Type

This research was conducted at the Department of Maritime Affairs and Fisheries of the Province of Aceh, Indonesia, covering areas of capture fisheries products on the north coast of Aceh. The data used in this study were obtained from the Aceh Maritime Affairs and Fisheries Service consisting of 10 fishing ports in the north coast of Aceh in Table 1.

Table 1. Fishing Port Data on the North Coast of Aceh Indonesia

Fishing Port Name	*Fishing Port Address*
PPN. Idi	Gampong Baro, Idi Rayeuk, East Aceh District
PPI. Blang Mee	Meucat, Samudera, North Aceh District
PPI. Krueng Mane	Tanoh Anoue, Muara Batu, North Aceh District
PPI. Kuala Cangkoy	Kuala Cangkoi, Lapang, North Aceh District
PPS. Kutaraja	Lampulo, Kec. Kuta Alam, Banda Aceh City
PPI. Pusong	Pusong Lama, Kec. Banda Sakti, Lhokseumawe City
PPI. Kuala Gigieng	Mesjid Gigieng, Simpang Tiga, Pidie District
PPI. Kuala Peukan Baro	Desa Kuala Peukan Baro, Kec. City of Sigli
PPI. Kuala Tari	Jeumeurang, Kembang Tj., Pidie District
PPI. Meuredu	Kota Meureudu, Meureudu, Pidie Jaya Regency

The variables or parameters used to determine the criteria for superior fish and ordinary fish are catch weight, fish price, number of fishing ports, and number of months (to determine the fishing season). This research resulted in 2 clusters, namely Cluster one (C1) superior fish and Cluster two (C2) ordinary fish, then visualized using the Geographic Information System.

4. Result and Discussion

4.1. Variable Attribute Values Used

In this study, to obtain the results of calculating superior fish clusters and ordinary fish clusters, the K-Means algorithm uses four variable attributes consisting of: Catch Weight

(X1), Fish Price (X2), Number of Fishing Ports (X3) and Number of Months (X4). From the data previously obtained, the value for each attribute has different characteristics, for the value for the catch weight attribute it has a value of hundreds to tens of thousands of tons, for the value for the fish price attribute it has a nominal value in rupiah currency, while the value for the number of fishing ports attribute and the number of months attribute has a unit value which can be directly used for the calculation process, so the authors reduce the value for the number of catches attribute and the fish price attribute to a unit value with a range of 1 to 10 (according to the number of 10 ports in this study). The following is Table 2. Fisheries data and attribute values that will be used in the clustering calculation process using the K-Means algorithm.

Table 2. Fisheries Data with Reduced Attribute Values

Fish Name	***X1***	***X2***	***X3***	***X4***
Ayam-ayam; Kambing-kambing (*Abalistes stellaris*)	2	2	2	6
Bawal Hitam (*Parastromateus niger*)	1	1	1	6
Bawal Putih (*Pampus argenteus*)	3	3	3	6
Biji Nangka Karang (*Parupeneus indicus*)	3	3	3	6
Bilis (*Herklotsichthys dispilonotus*)	2	3	3	6
Botana Biru Palsu (*Acanthurus nigricauda*)	1	1	1	6
Cakalang (*Katsuwonus pelamis*)	10	4	10	6
Cendro (*Tylosurus crocodilus*)	1	1	1	6
Hiu Botol (*Centroscymnus crepidater*)	1	1	1	1
Japuh (*Dussumieria acuta*)	1	1	1	1
Kakap (*Liopropoma randalli*)	3	3	3	6
Kakap Merah (*Lutjanus bitaeniatus*)	1	2	2	6
Kakap Putih (*Plectorhinchus gibbosus*)	2	2	3	6
Kembung (*Rastrelliger faughni*)	8	3	7	6
Kembung Lelaki (*Rastrelliger kanagurta*)	2	2	2	6
Kerapu Balong (*Epinephelus coioides*)	1	1	1	1
Kerapu Karang (*Chephalopholis boenack*)	1	2	2	6
Kerapu Macan (*Epinephelus maculatus*)	1	1	1	6
Kerapu Sunu (*Plectropomus leopardus*)	1	1	1	6
Kerapu Tutul (*Epinephelus quoyanus*)	2	3	3	6
Kuniran; Biji Nangka (*Upeneus tragula*)	1	1	1	6
Kurisi (*Nemipterus japonicus*)	3	3	3	6
Kuwe (*Caranx lugubris*)	5	4	4	6
Kuwe Lilin (*Caranx tille*)	1	1	1	6
Kuwe Sirip kuning (*Seriola nigrofasciatus*)	2	3	3	6
Layang (*Decapterus maruadsi*)	8	4	7	6
Layang Benggol (*Decapterus russelli*)	3	3	4	6
Layang Lidi; Layar Deles (*Decapterus macrosoma*)	1	1	1	6
Layaran (*Istiophorus platypterus*)	1	1	1	6
Layur (*Trichiurus lepturus*)	4	2	4	6
Lemadang (*Coryphaena hippurus*)	3	2	3	6
Lemuru (*Sardinella lemuru*)	2	1	2	6

Marlin Hitam (*Istiompax indica*)	5	2	4	6

4.2. Application of the K-Means Algorithm

The following are the stages for the clustering process of capture fisheries products on the north coast of Aceh using the K-Means algorithm

1. Determine the number of clusters
   In this capture fisheries product clustering study, 2 clusters were used, consisting of: Superior Fish Cluster (C1) and Common Fish Cluster (C2).
2. Choose a starting centroid randomly
   The researcher chose the following random centroids to be used in the manual clustering calculation process in this study. The following is Table 3. Random Centroid for Iteration-1, two types of fisheries were taken as data samples.

Table 3. Random centroid for Iteration-1

Centroid Point	***X1***	***X2***	***X3***	***X4***
Kakap Merah (*Lutjanus bitaeniatus*)	1	2	2	6
Selar Kuning (*Selaroides leptolepis*)	7	2	6	6

3. Calculation Process Iteration-1

It is known: the number of fishery data (I) = 57 and the number of clusters that have been determined = 2. Calculate the distance between the data and the centroid using the Euclidean Distance formula. The following is the formula for the Euclidian Distance equation:

$$D(A_n, C_X) = \sqrt{(A_n - C_x)^2 + (A_n - C_x)^2 + \ldots} \quad (2)$$

Calculating the data distance with the C-1 centroid, the following is Table 4. C-1 centroid for Iteration-1.

Table 4. Centroid C-1 for Iteration-1

Centroid	***X1***	***X2***	***X3***	***X4***
Kakap Merah (*Lutjanus bitaeniatus*)	1	2	2	6

$Euclidian\ Distance: D(I_1, C_1)$

$$= \sqrt{(I_{1a} - C_{1a})^2 + (I_{1b} - C_{1b})^2 + (I_{1c} - C_{1c})^2 + (I_{1d} - C_{1d})^2}$$
$$= \sqrt{(2-1)^2 + (2-2)^2 + (2-2)^2 + (6-6)^2}$$
$$= \sqrt{1}$$
$$= 1$$

Then calculate the data distance with Centroid C-2, The following is Table 5. Centroid C-2 for Iteration-1.

Table 5. Centroid C-2 for Iteration-1

Centroid	***X1***	***X2***	***X3***	***X4***
Selar Kuning (*Selaroides leptolepis*)	7	2	6	6

$Euclidian\ Distance: D(I_1, C_2)$

$$= \sqrt{(I_{1a} - C_{2a})^2 + (I_{1b} - C_{2b})^2 + (I_{1c} - C_{2c})^2 + (I_{1d} - C_{2d})^2}$$
$$= \sqrt{(2-7)^2 + (2-2)^2 + (2-6)^2 + (6-6)^2}$$
$$= \sqrt{41}$$
$$= 6.403124$$

Based on the results of the calculation above, the shortest distance for fisheries data that has a calculated value of *D(I, C1)* is smaller than the calculated value of *D(I, C2)* will enter into cluster 1, on the other hand fisheries data that has a calculated value of *D(I, C2)* is smaller than the value of *D(I, C1)* will enter into cluster 2. The following table 6. Iteration-1 calculation results.

Table 6. Iteration Calculation Results-1

Fishery Data (represented by serial number of fish species)	***D(I, C1)***	***D(I, C2)***	***Cluster***
1	1	6.403124	1
2	1.414214	7.874008	1
3	2.44949	5.09902	1
4	2.44949	5.09902	1
5	1.732051	5.91608	1
6	1.414214	7.874008	1
7	12.20656	5.385165	2
8	1.414214	7.874008	1
9	5.196152	9.327379	1
10	5.196152	9.327379	1
11	2.44949	5.09902	1

12	0	7.211103	1
…..	…….	…….	….
57	8.660254	1.732051	2

Based on the calculation results of Iteration-1 in Table 6 above, it is obtained that the number of data that enters the Superior Cluster (C1) is 44 data and the amount of data that enters the Ordinary Cluster (C2) is 13 data, out of a total of 57 data.

4. Calculation Process Iteration-2

If the initial centroid point was previously determined randomly, then to determine the next centroid point it will use the K-Means formula utilizing the data from cluster calculations using the previous Euclidean Distance. Here is the average K-Means formula:

$$C_X, A_n = (dC_X + dC_X + \cdots) \,/\, jC_1 \quad (3)$$

It is known that cluster 1 (C1) has 44 data, so the average value in cluster 1 is:

$$\begin{aligned} C_1, A_1 &= (dC_1 + dC_2 + dC_3 + dC_4 + dC_5 + dC_6 + dC_8 + dC_9 \\ &\quad + dC_{10} + dC_{11} + dC_{12} + dC_{13} + dC_{15} \\ &\quad + dC_{16} + dC_{17} + dC_{18} + dC_{19} + dC_{20} \\ &\quad + dC_{21} + dC_{22} + dC_{24} + dC_{25} + dC_{27} \\ &\quad + dC_{28} + dC_{29} + dC_{31} + dC_{32} + dC_{34} \\ &\quad + dC_{35} + dC_{36} + dC_{37} + dC_{38} + dC_{39} \\ &\quad + dC_{40} + dC_{43} + dC_{44} + dC_{45} + dC_{46} \\ &\quad + dC_{48} + dC_{49} + dC_{50} + dC_{52} + dC_{53} \\ &\quad + dC_{55}) \,/\, jC_1 \\ &= (2 + 1 + 3 + 3 + 2 + 1 + 1 + 1 + 1 + 1 + 3 + 1 + 2 + 2 \\ &\quad + 1 + 1 + 1 + 1 + 2 + 1 + 3 + 1 + 2 + 3 \\ &\quad + 1 + 1 + 3 + 2 + 1 + 2 + 2 + 2 + 1 + 1 \\ &\quad + 1 + 3 + 2 + 3 + 4 + 2 + 1 + 3 + 3 + 4 \\ &\quad + 3) \,/\, 44 \\ &= (84) \,/\, 44 \\ &= 1.909091 \end{aligned}$$

Whereas for cluster 2 (C2) there are 13 data, then the average value in cluster 2 is:

$$\begin{aligned} C_2, A_1 &= (dC_7 + dC_{14} + dC_{23} + dC_{26} + dC_{30} \\ &\quad + dC_{33} + dC_{41} + dC_{42} + dC_{47} \\ &\quad + dC_{51} + dC_{54} + dC_{56} + dC_{57}) \\ &\quad / \, jC_2 \\ &= (10 + 8 + 5 + 8 + 4 + 5 + 5 + 7 + 7 \\ &\quad + 10 + 5 + 4 + 7) \,/\, 13 \\ &= (85) \,/\, 13 \\ &= 6.538462 \end{aligned}$$

And so on for the calculation of attribute variables X2, X3 and X4, then the latest centroid calculation results for Iteration-2 are obtained in Table 7.

Table 7. New Centroid Iteration-2

New Centroid Iteration-2	***X1***	***X2***	***X3***	***X4***
1	1.909	1.863	2.090	5.659
2	6.538	3.230	5.923	6

Then recalculate the distance between the data and the latest Iteration-2 centroid using the Euclidean Distance formula. Calculating the data distance with the C-1 centroid for Iteration-2.

Table 8. Centroid C-1 For Iteration-2

Centroid	***X1***	***X2***	***X3***	***X4***
1	1.909	1.863	2.090	5.659

$$\begin{aligned} &Euclidian\ Distance: D(I_2, C_1) \\ &= \sqrt{(I_{1a} - C_{1a})^2 + (I_{1b} - C_{1b})^2 + (I_{1c} - C_{1c})^2 + (I_{1d} - C_{1d})^2} \\ &= \sqrt{(2 - 1.909091)^2 + (2 - 1.863636)^2 + (2 - 2.090909)^2 + \ldots} \\ &\qquad \sqrt{\ldots\ (6 - 5.659091)^2} \\ &= \sqrt{0.1513429793} \\ &= 0.389028 \end{aligned}$$

Then calculate the data distance with the C-2 centroid for Iteration-2. The following is Table 8. Centroid C-2 for Iteration-2.

Table 9. Centroid C-2 For Iteration-2

Centroid	***X1***	***X2***	***X3***	***X4***
2	6.538	3.230	5.923	6

$$\begin{aligned} &Euclidian\ Distance: D(I_2, C_2) \\ &= \sqrt{(I_{1a} - C_{2a})^2 + (I_{1b} - C_{2b})^2 + (I_{1c} - C_{2c})^2 + (I_{1d} - C_{2d})^2} \\ &= \sqrt{(2 - 6.538462)^2 + (2 - 3.230769)^2 + (2 - 5.923077)^2 + \cdots} \\ &\qquad \sqrt{\ldots\ (6 - 6)^2} \\ &= \sqrt{37.5029628} \\ &= 6.123966 \end{aligned}$$

The following is Table 10. The results of iteration-2 calculations.

Table 10. Iteration Calculation Results-2

Fishery Data (represented by fish serial number)	***D(I2, C1)***	***D(I2, C2)***	***Cluster***
1	0.389028	6.123966	1
2	1.696648	7.738706	1

3	1.850424	4.59547	1
4	1.850424	4.59547	1
5	1.497415	5.403264	1
6	1.696648	7.738706	1
7	11.51941	5.403264	2
8	1.696648	7.738706	1
9	4.946668	9.213445	1
10	4.946668	9.213445	1
11	1.850424	4.59547	1
12	0.984645	6.897817	1
13	0.984645	5.536859	1
14	7.912388	1.830058	2
…………….	……..	………	1
16	4.946668	9.213445	1

Based on the calculation results of Iteration-2 in Table 10 above, the amount of data that enters the Cluster (C1) of Superior Fish is 45 data and the amount of data that enters the Cluster (C2) of Common Fish is 12 data, with a total of 57 data. Following are the results of a comparison of the Iteration-1 and Iteration-2 calculations in Table 11.

Table 11. Comparison of Calculation Results Iteration-1 with Iteration-2

No	***Fishery Name***	***Iteration 1***	***Iteration 2***
1	Ayam-ayam; Kambing-kambing (*Abalistes stellaris*)	1	1
2	Bawal Hitam (*Parastromateus niger*)	1	1
3	Botana Biru Palsu (*Acanthurus nigricauda*)	1	1
4	Cakalang (*Katsuwonus pelamis*)	2	2
5	Cendro (*Tylosurus crocodilus*)	1	1
6	Kakap Merah (*Lutjanus bitaeniatus*)	1	1
7	Kakap Putih (*Plectorhinchus gibbosus*)	1	1
8	Kembung (*Rastrelliger faughni*)	2	2
9	Kembung Lelaki (*Rastrelliger kanagurta*)	1	1
10	Kerapu Balong (*Epinephelus coioides*)	1	1
11	Kuniran; Biji Nangka (*Upeneus tragula*)	1	1
12	Kurisi (*Nemipterus japonicus*)	1	1
13	Kuwe (*Caranx lugubris*)	2	2
14	Kuwe Lilin (*Caranx tille*)	1	1
15	Kuwe Sirip kuning (*Seriola nigrofasciatus*)	1	1
16	Layang (*Decapterus maruadsi*)	2	2
17	…………………… ……	….	….
57	Layang Lidi; Layar Deles (*Decapterus macrosoma*)	1	1

After completing the calculation of Iteration 2, it is continued with the calculation of Iteration-3. The results of Iteration-3 are obtained the same as the results of the calculation of Iteration-2 and there is no change in data, so the calculation is over and finished in Iteration-3, so that the final results of the calculation are obtained to determine the clusters (C1) Superior Fish as many as 45 and clusters (C2) Common Fish as many as 12 in Tables 12 and 13 below.

Table 12. Cluster Calculation Results for Superior Fish (C1)

Fishery Name	***X1***	***X2***	***X3***	***X4***	***Cluster (C1)***
Ayam-ayam; Kambing-kambing	2	2	2	6	Superior

(*Abalistes stellaris*)					
Bawal Hitam (*Parastromateus niger*)	1	1	1	6	Superior
Bawal Putih (*Pampus argenteus*)	3	3	3	6	Superior
Biji Nangka Karang (*Parupeneus indicus*)	3	3	3	6	Superior
Bilis (*Herklotsichthys dispilonotus*)	2	3	3	6	Superior
Botana Biru Palsu (*Acanthurus nigricauda*)	1	1	1	6	Superior
Cendro (*Tylosurus crocodilus*)	1	1	1	6	Superior
Hiu Botol (*Centroscymnus crepidater*)	1	1	1	1	Superior
Japuh (*Dussumieria acuta*)	1	1	1	1	Superior
……………	..	..	..	..	…..
Tongkol Lisong (*Auxis rochei*)	3	2	3	6	Superior

Table 13. Cluster Calculation Results for Common Fish (C2)

Fishery Name	***X1***	***X2***	***X3***	***X4***	***Cluster (C1)***
Cakalang (*Katsuwonus pelamis*)	10	4	10	6	Normal
Kembung (*Rastrelliger faughni*)	8	3	7	6	Normal
Kuwe (*Caranx lugubris*)	5	4	4	6	Normal
Layang (*Decapterus maruadsi*)	8	4	7	6	Normal
Marlin Hitam (*Istiompax indica*)	5	2	4	6	Normal
Selar Hijau (*Atule mate*)	5	2	4	6	Normal
Selar Kuning (*Selaroides leptolepis*)	7	2	6	6	Normal
Tenggiri (*Scomberomorus commerson*)	7	5	7	6	Normal
Tongkol (*Auxis thazard*)	10	3	9	6	Normal
Tongkol Komo (*Euthynnus affinis*)	5	2	4	6	Normal
Tuna Mata Besar (*Thunnus obesus*)	4	4	5	6	Normal
Tuna Sirip Kuning; Madidihang (*Thunnus albacares*)	7	5	6	6	Normal

The following is a graph of the results of capture fisheries clustering using the K-Means algorithm.

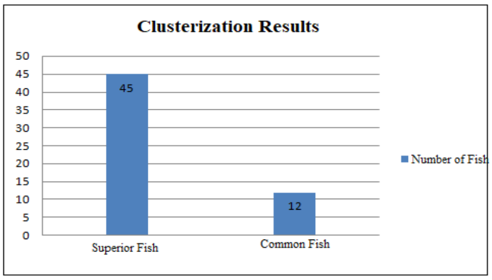

Fig. 3. Fishery Product Clusterization Graph

4.3. Implementation of System

The following page displays the distribution map of capture fisheries products. This page contains map data that displays fishing port point icons and types of fishery product clustering in Figure 4.

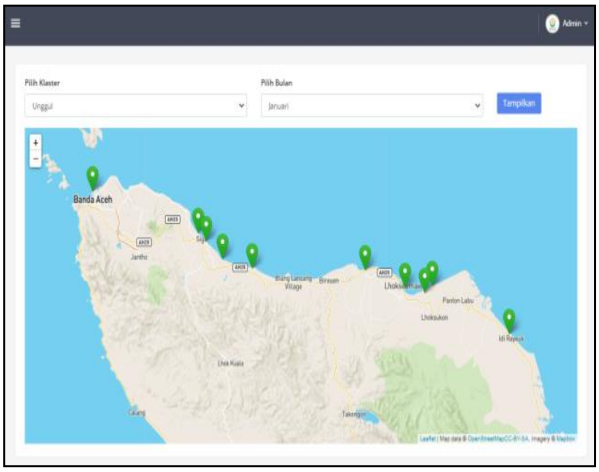

Fig. 4. Map of Distribution of Capture Fishery Products

The following is a display of the search results output page based on Cluster Type and Month. This page also displays a model on the map containing information on the name and address of the port, and there is a fish list button which when clicked will display a model containing information on the list of fish along with their prices, as shown in Figure 5.

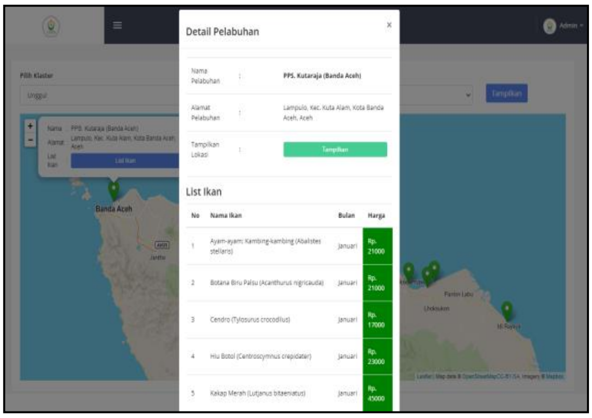

Fig. 5. Cluster Type Search Result Output Map

The following is a page view of the number of fishery products on the north coast of Aceh. This page displays a graph of capture fisheries results based on 10 fishing ports, which can be seen in Figure 6.

Fig. 6. Number of Fishery Products on the North Coast of Aceh

5. Conclusion

This study of clustering capture fisheries products on the north coast of Aceh uses 4 attribute variables, namely: catch weight, fish price, number of fishing ports and number of months. Based on the results of clustering using the K-Means algorithm, 45 Clusters of Superior Fish (C1) were obtained and 12 Clusters of Common Fish (C2), the data on fish species were spread across 10 fishing ports. The results of this clustering are visualized on the Geographic Information System Website. The results of the clusters formed using the K-Means algorithm are very dependent on the initial initial value of the cluster that is determined, this makes it very difficult to obtain unique initial centroid results. In this study the K-Means Algorithm can be used to cluster capture fisheries products. 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 service to obtain information on types of fishery products (high quality fish and common fish) every month at fishing ports on the north coast of Aceh Indonesia.

Further research is recommended to use other clustering methods as a comparison to find out which method is better. This application system is still web-based can be developed again into an android-based system to make it easier to use.

Acknowledgements

Thank you to the Directorate of Research, Technology and Community Service. Ministry of Education, Culture, Research and Technology for research funding support for the Fundamental-Regular Research scheme.

References

[1] F. Fajriana, "Analysis of the K-Medoids Algorithm in the Capture Fisheries Production Clustering System in North Aceh District," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol.*7*, *no.*2, pp. 263-269, 2021.

[2] N. Nurdin, M. Zarlis, Tulus and S. Efendi, "Data

Driven Optimization Approach to Fish Resources Supply Chain Planning in Aceh Province", *IOP Conf. Series: Journal of Physics: Conf. Series* 1255 (2019) 012081, 2019.

[3] N. Nurdin, M. Zarlis, Tulus and S. Efendi, "Mixed Integer Linear Programming Model for Integrated Fish Supply Chain Planning" *Journal of Theoretical and Applied Information Technology*, Vol.98, No.12, 2020, pp. 2017-2028.

[4] S. Zhu and P. Bao, "The use of Geographic Information System in the development and utilization of ancient local chronicles" *Library Hi Tech*, vol. 33, no. 3, pp. 356–368, 2015. https://doi: 10.1108/LHT-03-2015-0028.

[5] N. Ahmed and A. Razak T, "An Overview of Various Improvements of DBSCAN Algorithm in Clustering Spatial Databases" *International Journal of Advanced Research in Computer and Communication Engineering"*, vol. 5, no. 2, pp. 360–363, 2016.

[6] T. Velmurugan and T. Santhanam, "Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points" *Journal of Computer Science*, vol. 6, no. 3, pp. 363-368, 2010. https://doi.org/10.3844/jcssp.2010.363.

[7] P. Arora and S. Varshney, "Analysis of K-means and K-medoids algorithm for big data", *Procedia Computer Science*, 78: 507-512.

[8] Mardalius, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan (Studi Kasus : Siswa Sma Negeri 1 Ranah Pesisir)", *Proceding SEMILOKA R. 2017 "Teknologi Mobile,"* vol. 1, p. 105, 2017, doi: 10.31219/osf.io/6mec3.

[9] J. Xiao, J. Lu, and X. Li. "Davies Bouldin Index based hierarchical initialization K-means", Intelligent Data Analysis, vol. 21, no. 6, pp. 1327-1338, 2017. https://doi.org/10.3233/IDA-163129.

[10] N. A. Khairani and E. Sutoyo, "Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province," *Int. J. Adv. Data Inf. Syst.*, vol. 1, no. 1, pp. 9–16, 2020, doi: 10.25008/ijadis.v1i1.13.

[11] A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.

[12] L. ZHANG, F. L. LU, A. LIU, P. GUO, and C. LIU, "Application of K-Means Clustering Algorithm for Classification of NBA Guards," *Int. J. Sci. Eng. Appl.*, vol. 5, no. 1, pp. 1–6, 2016, doi: 10.7753/ijsea0501.1001.

[13] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.

[14] S. Shukla, "A Review ON K-means DATA Clustering APPROACH," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014.

A. Dutt, "Clustering Algorithms Applied in Educational Data Mining," *Int. J. Inf. Electron. Eng.*, vol. 5, no. 2, pp. 112–116, 2015, doi: 10.7763/ijiee.2015.v5.513.

[15] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC)," *Sistemasi*, vol. 10, no. 2, p. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.

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

[17] N. Nurdin, B. Bustami and M. Maryana, "Robust Optimization Approach for Agricultural Commodity Supply Chain Planningg," *Journal of Theoretical and Applied Information Technology*., vol. 99, no. 2, pp. 304–315, 2021.

[18] Nurdin, Taufiq, and Fajriana. "Searching the shortest route for distribution of LPG in Medan City using ant colony algorithm". IOP Conf. Series: Materials Science and Engineering, 725 (2020) 012121, 2020.

[19] D. Abdullah, N. Nurdin, M. Yaton, H. Sujatmiko and …, "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 (2019) 012074, 2019.

[20] N. Nurdin, B. Bustami, M. Hutomi, M. Elveny and R. Syah, "Implementation of the BFS algorithm and web scraping techniques for online shop detection in Indonesia," *Journal of Theoretical and Applied Information Technology.*, vol. 99, no. 12, pp. 2878-2889, 2021.

[21] R. J. Hablum, A. Khairan and Rosihan. "Clustering Hasil Tangkap Ikan di Pelabuhan Perikanan Nusantara (PPN) Ternate Menggunakan Algoritma K-Means" *JIKO (Jurnal Informatika dan Komputer) Ternate*. vol. 2, no. 1, pp. 26-33, 2019. http://dx.doi.org/10.33387/jiko.v2i1.1053

[22] S. Sugiarto, Z.Z. Hani, and A. P. Yosep. "Sistem Informasi Geografis Daerah Perikanan Dan Peternakan Kabupaten Pasuruan." *JATI (Jurnal Mahasiswa Teknik Informatika)*. Vol.8., no. 1, 2020. https://doi.org/10.36040/jati.v4i1.232

[23] N. Nurdin, F. Fajriana, M. Maryana, and A. Zanati. "Information System for Predicting Fisheries Outcomes Using Regression Algorithm Multiple

Linear." *JITE (Journal of Informatics and Telecommunication Engineering*), vol. 5, no. 2, pp. 247-258, 2022.

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

[25] S. Annas and Z. Rais. "K-Means and GIS for Mapping Natural Disaster Prone Areas in Indonesia." *Pervasive Health: Pervasive Computing Technologies for Healthcare*, 1: 340–462, 2020.

[26] S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha and M. Azmi. "Using K-Means Clustering to Cluster Provinces in Indonesia."*Journal of Physics: Conference Series*,*1028*(1).https://doi.org/10.1088/1742-6596/1028/1/012006.

[27] M. Faisal, N. Nurdin, F. Fajriana, and Z. Fitri, "Information and Communication Technology Competencies Clustering For Students For Vocational High School Students Using K-Means Clustering Algorithm," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 111–120, 2022, doi: 10.52088/ijesty.v2i3.318.

[28] R. Rosadi, Akamal, R. Sudrajat, B. Kharismawan, and Y. A. Hambali, " Student academic performance analysis using fuzzy C-means clustering." *IOP Conf. Ser.: Mater. Sci. 2017.* https://doi.org/10.1088/1757899X/166/1/012036.

[29] Y. Shan, and Z. Zhi, "A novel fuzzy SVM based on fuzzy c-means for credit scoring. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 1349–1353. 2016, https://doi.org/10.1109/CompComm.2016.7924924

[30] T. S. J. Arifin, N. Natassa, D. Khoirunnisa, and R. Hendrowati, "The Level of Student Satisfaction with the Online Learning Process During a Pandemic Using the K-means Algorithm," *Inf. Sci. J. Inf. Commun. Technol.*, vol. 6, no. 2, pp. 123–126, 2021.

[31] P. Sari, and I. H. Batubara, "Cluster Analysis Using K-Means Algorithm and Fuzzy C- Means Clustering for Grouping Students Abilities in Online Learning Process." *J. Comput. Sci. Inf. Technol. Telecommun. Eng*, vol. 2, no.1, pp. 139–144, 2021, https://doi: 10.30596/jcositte.v2i1.6504

[32] H. Escorcia Guzman, R. A. Zuluaga-Ortiz, D. A. Barrios-Miranda, and E. J. Delahoz-Dominguez, "Information and Communication Technologies (ICT) in the processes of distribution and use of knowledge in Higher Education Institutions (HEIs)," *Procedia Comput. Sci.*, vol. 198, no. 2021, pp. 644–649, 2021, doi: 10.1016/j.procs.2021.12.300.

[33] D. Borlea, R. E. Precup, and A. B. Borlea, "Improvement of K-means Cluster Quality by Post Processing Resulted Clusters," *Procedia Comput. Sci.*, vol. 199, pp. 63–70, 2021, doi: 10.1016/j.procs.2022.01.009.

[34] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "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.*, vol. 56, no. 3, pp. 1283–1291, 2022, doi: 10.1007/s11135-021-01176-w.

[35] Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.

[36] S. N. Kane, A. Mishra, and A. K. Dutta, "Preface: International Conference on Recent Trends in Physics (ICRTP 2016)," *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016, doi: 10.1088/1742-6596/755/1/011001.

[37] Ana Oliveira, Yosef Ben-David, Susan Smit, Elena Popova, Milica Milić. Improving Decision Quality through Machine Learning Techniques. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/202

[38] Shukla, A., Juneja, V., Singh, S., Prajapati, U., Gupta, A., & Dhabliya, D. (2022). Role of hybrid optimization in improving performance of sentiment classification system. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 541-546. doi:10.1109/PDGC56933.2022.10053333 Retrieved from www.scopus.com

[39] Juan Garcia, Guðmundsdóttir Anna, Johansson Anna, Maria Jansen, Anna Wagner. Intelligent Decision Support using Neural Networks in Decision Science. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/203