Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2371~2381
e-ISSN: 2548-6861

Clustering Coastal Areas Based on Aquaculture Productivity in North Aceh Regency Using K-Means Algorithm

Septia Mulya Ulfa [1]*, Rozzi Kesuma Dinata [2], Risawandi [3]***
* Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Indonesia
septia.210170118@mhs.unimal.ac.id [1], rozzi@unimal.ac.id [2], risawandi@unimal.ac.id [3]

Article Info

Article history:

Received 2025-07-06
Revised 2025-07-13
Accepted 2025-07-19

Keyword:

K-Means,
Clustering,
Aquaculture,
Coastal Areas.

ABSTRACT

This study aims to cluster coastal subdistricts in North Aceh Regency based on the productivity of seven key aquaculture commodities milkfish, vannamei shrimp, tiger shrimp, tilapia, mojarra, grouper, and crab using the K-Means algorithm. The dataset, sourced from 15 coastal subdistricts, was normalized using the Z-Score method. The optimal number of clusters was determined using the Elbow Method, and clustering performance was evaluated with the Silhouette Score, yielding a value of 0.5293, indicating a moderately well-defined structure. The resulting clusters reflect distinct productivity levels: Cluster 0 (low), Cluster 1 (moderate), and Cluster 2 (high). A two-dimensional PCA plot was used to visualize the clusters, showing clear separations among them. These findings offer valuable insights for regional planners and policymakers in developing targeted aquaculture strategies and optimizing resource allocation, particularly for underperforming areas.

I. INTRODUCTION

Indonesia is a maritime country with approximately 70% of its total area consisting of seas. The abundant natural resources further reinforce the belief that this nation is indeed a maritime country [1]. The coastal area is defined as a transitional zone between terrestrial and marine ecosystems [2].

Aquaculture is one of the fastest-growing food-producing sectors in the world [3]. Aquaculture production, including fish and shrimp, is carried out in various types of waters freshwater, brackish, and marine. The production output from aquaculture has consistently increased year by year [4]. The rise in fishery production in Indonesia is closely related to the geographical conditions and location of each region, as every area has fishing grounds with distinct characteristics. To understand the characteristics of each group in the fisheries sector based on regional data, statistical science can be used as an analytical tool [5].

K-Means is one of the algorithms used to partition data into several groups based on similarities in characteristics [6]. The K-Means algorithm is a commonly applied method in clustering processes, widely known as K-Means Clustering. This algorithm uses a distance-based approach in data mining to group data [7]. Clustering is a method of dividing data into several clusters or groups so that the data within each cluster have a high degree of similarity, while the data between clusters have minimal similarity [8].

In line with the growing interest in aquaculture analytics, several studies have explored the application of clustering algorithms in this domain.Studies on the application of clustering algorithms in the aquaculture sector have been widely conducted across various regions in Indonesia. One study applied the K-Means algorithm to cluster aquaculture production based on cultivation container types (such as ponds and floating net cages) in North Sulawesi Province, aiming to assist local agencies in designing more targeted strategies for regional production improvement [9]. Another study utilized the X-Means algorithm, a variant of K-Means, to classify shrimp aquaculture distribution across all Indonesian provinces into three clusters based on their geographical patterns [10]. A different approach employed the K-Prototypes and Two-Step Cluster algorithms to group aquaculture companies by combining numerical and categorical attributes, capturing deeper characteristics of each enterprise [11]. In addition, a study on Fuzzy C-Means clustering for aquaculture production in North Sumatra emphasized the importance of distance function selection, concluding that using the appropriate metric significantly improves clustering accuracy [12]. Collectively, these studies demonstrate the relevance of clustering methods in supporting data-driven decision-making within the aquaculture sector.

The study conducted by [13], titled “Implementation of the K-Means Clustering Algorithm in Grouping Fish Catch Results in the Riau Archipelago”, successfully demonstrated how the K-Means algorithm can be used to classify fish species based on production volume and catch location. As a result, the fish were categorized into two groups superior and non-superior which were then used by fishermen as a reference to determine more optimal fishing areas. This research highlights that K-Means is capable of simplifying complex data into more targeted groups that are practically useful in the field. Based on these findings, a similar approach is applied in the present study. However, this research focuses on clustering 15 coastal subdistricts in North Aceh Regency based on the productivity of their leading aquaculture commodities, such as milkfish, vannamei shrimp, tiger shrimp, tilapia, mojarra, grouper, and crab. These commodities represent the dominant aquaculture activities in the region and are considered vital for local economic development. The classification of coastal areas through clustering is expected to reveal spatial productivity patterns that are often overlooked in generalized statistics.

The objective of this study is to cluster the coastal areas of North Aceh Regency based on the productivity levels of their leading aquaculture commodities using the K-Means algorithm. Through this clustering process, it is expected that spatial patterns of regions with similar productivity characteristics can be identified, serving as a basis for formulating more targeted and effective strategies for regional potential development.

II. METHODOLOGY

A. *Research Flow*

To achieve the objectives of this study, a series of systematic stages were carried out, starting from the literature review to the analysis of the clustering results. Each stage was designed to ensure that the data processing and the implementation of the K-Means algorithm would produce accurate and interpretable clusters. The research workflow can be seen in Figure 1.

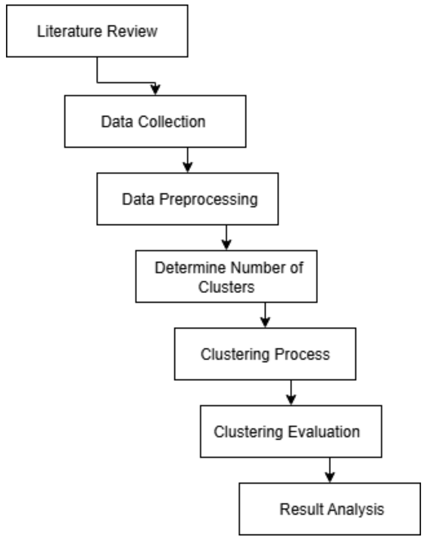


Figure 1.Research Flowchart

B. *Algorithm Flowchart*

To support the clustering process in this study, the K-Means algorithm was employed with a systematic workflow. A brief overview of the algorithm's process can be described as follows:

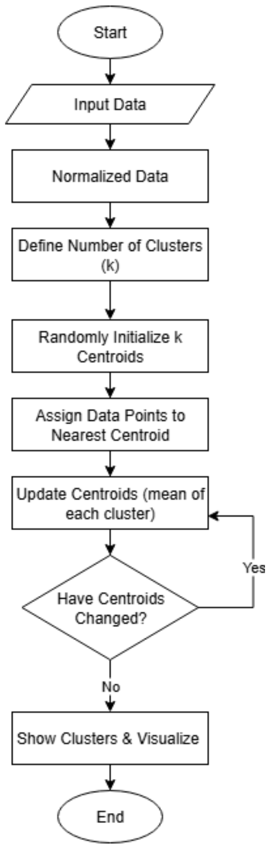


Figure 2. Algorithm Flowchart

C. Literature Review

1. Data Mining

Data mining is the process of managing data with the aim of discovering specific patterns or information. It can be defined as a set of techniques used to uncover hidden patterns within collected data [14]. Data mining is the process of extracting patterns or valuable information from large datasets [15]. The primary goal of data mining is to identify hidden patterns, relationships, and trends within the data to gain valuable insights and support better decision-making [16].

K-Means clustering is an unsupervised learning algorithm that aims to group data into k clusters based on the similarity of their characteristics. The fundamental principle of this algorithm is to minimize the sum of squared distances between data points and the nearest cluster centroids [17]. This algorithm works by partitioning the data into a number of clusters based on attribute similarity and is known for its high computational efficiency[18].

- Euclidean Distance

  The first step is to calculate the distance between each data point and each cluster centroid using the Euclidean Distance formula, as follows:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots + (x_n - c_n)^2}$$

Description:
d = Euclidean distance between the data point and the centroid
$x_1, x_2, \dots, x_n$ = The calculated value of each variable in the dataset
$c_1, c_2, \dots, c_n$ = The centroid value for each variable

- Update Centroid

  The second step is to update the position of each cluster centroid by calculating the mean of all data points assigned to that cluster.

$$C_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

Description:
C_j = New centroid value for the j-th variable
x_{ij} = Value of the i-th data point for the j-th variable that belongs to the same cluster
n = Number of data points in the cluster

2. Elbow Method

The Elbow Method performs an evaluation process by comparing the results of the total sum of clusters that form the final point and the cluster values used as the data model to determine the optimal number of clusters [19]. The results from this calculation are then used to compare different cluster quantities. The Elbow Method is commonly presented in the form of a graph to clearly identify the "elbow" point that is formed. The purpose of the Elbow Method is to select the smallest value of k that still results in a low within-cluster sum of squares (withinss) [20].

$$SSE = \sum_{i=1}^{n} (d_i)^2$$

Description:
d_i = Euclidean distance between the i-th data point and its cluster centroid
n = Total number of data points in each cluster

SSE measures the total distance between data points and their respective cluster centroids. A smaller SSE value indicates better clustering results, as it means the data points are more tightly grouped around their centroids. In the context of the Elbow Method, SSE is used to help determine the optimal number of clusters.

3. Normalization

The data normalization process is carried out before applying the K-Means clustering algorithm. The purpose of normalization is to standardize the scale of the variables so that each feature contributes equally to the clustering process. This is important because the K-Means algorithm groups data based on Euclidean distance, and differences in feature scales can lead to distorted clustering results. The normalization method used in this study is Z-Score Standardization.

Z-Score is a type of normalization based on the mean and standard deviation. It transforms the original data into a smaller scale, although it does not have a fixed range [21].

$$Z = \frac{X - \mu}{\sigma}$$

Description:
Z : Normalized value
X : Original data value
μ : Mean of the attribute
σ : Standar deviasi

Z-Score normalization is important in many data analysis applications, especially when working with algorithms that are sensitive to scale differences, such as PCA (Principal Component Analysis) [22].

In this study, snapshot data from the year 2024 were collected from 15 coastal subdistricts in North Aceh Regency. The dataset consisted of 9 attributes, including pond area, number of farmers, and production volumes of seven dominant aquaculture commodities selected based on data from the Department of Marine and Fisheries due to their significant contribution to local aquaculture. Outlier data were not removed, considering their potential to represent extreme but valid regional productivity. Clustering evaluation was conducted using the Silhouette Score, which produced a value of 0.5293, interpreted as moderate clustering performance. Additionally, the total number of iterations required for the K-Means algorithm to converge was noted as part of the evaluation. To support interpretability, a Principal Component Analysis (PCA) was performed for 2D visualization, following a prior assessment of the correlation among variables to justify its use.

III. RESULT AND DISCUSSION

The analysis begins by determining the optimal number of clusters through the Elbow Method, which identifies the point where adding more clusters does not significantly reduce the within-cluster variation. To further evaluate the quality of clustering, the Silhouette Score is calculated, providing a quantitative measure of how well each data point fits within its assigned cluster compared to other clusters. Once the optimal number of clusters is established, the distribution of each cluster is interpreted in relation to the productivity levels of leading aquaculture commodities in the coastal areas of North Aceh Regency. This analysis helps in identifying regional patterns and similarities in aquaculture performance. To enhance understanding and validate the clustering results, visual representations such as pie charts, scatter plots, and PCA based diagrams are included. These visual tools offer a clearer picture of how each area is grouped and the underlying factors influencing cluster formation, ultimately supporting data-driven decision-making for regional development.

To assess the sensitivity of the clustering parameter k, several values (k = 2, 3, and 4) were tested. The result showed that k=3 yielded the highest Silhouette Score of 0.5293, while k=2 and k=4 resulted in scores of 0.48 and 0.50, respectively. This confirms that k=3 provides the most balanced and stable clustering structure. Furthermore, for future research, it is recommended to compare clustering results with alternative algorithms such as DBSCAN or Agglomerative Clustering. Incorporating spatial data and GIS-based visualization is also suggested to enhance the interpretability of cluster distribution across coastal areas.

A. *Training Dataset Analysis*

Before the clustering process was carried out, the data were first collected and organized into a table to facilitate processing and analysis. The dataset used in this study consists of productivity data of leading aquaculture commodities from coastal areas in North Aceh Regency. The details of the data are presented in the following table:

TABLE I
RESEARCH DATASET

No	Year	Sub-district	Pond Area (ha)	Farmers	Milk-fish (ton)	Vannamei(ton)	Tiger Prawn(ton)	Parrot Fish (ton)	Tilapia (ton)	Grouper (ton)	Crab (ton)
1	2017	Baktiya	125.0	103	67.5	2.4	3.8	0.8	0.6	2.0	1.0
2	2017	Dewantara	210.0	189	115.4	7.2	5.6	2.5	1.1	1.0	0.8
3	2017	Lapang	100.0	98	70.2	1.9	4.2	1.4	0.3	3.0	2.0
4	2017	Meurah Mulia	45.0	42	33.1	0.5	1.2	0.2	0.1	2.0	1.0
5	2017	Muara Batu	165.0	120	84.8	3.6	4.7	1.9	0.7	3.0	0.7
…	…	…	…	…	…	…	…	…	…	…	…
92	2024	Nibong	27.0	23	15.0	0.6	1.1	0.3	0.2	1.0	0.2
93	2024	Seunuddon	340.0	230	192.0	6.0	7.1	3.9	1.8	0.3	1.7
94	2024	Syamtalira Aron	25.0	24	14.0	0.4	1.0	0.2	0.2	1.0	0.2
95	2024	Syamtalira Bayu	67.0	56	35.2	1.1	2.2	0.7	0.3	4.0	0.3
96	2024	Tanah Jambo Aye	206.0	157	123.0	3.7	4.9	1.8	0.7	2.0	1.0

Table 1 contains aquaculture data from 96 coastal subdistrict entries in North Aceh Regency, spanning from 2017 to 2024. The dataset includes attributes such as year, subdistrict name, pond area (ha), number of farmers, and production volumes (in tons) of various commodities including milkfish, vannamei shrimp, tiger shrimp, tilapia, carp, grouper, and crab. The diversity of productivity values across subdistricts reflects significant regional variations in aquaculture potential. To identify patterns in this variation, the data were analyzed using the K-Means clustering algorithm. Before clustering, Z-Score normalization was applied to standardize the scales of all attributes and improve the accuracy of the clustering process.

TABLE II
EXAMPLE CALCULATION TABLE FOR CLUSTERING

Year	Sub-district	Pond Area (ha)	Farmers	Milkfish (ton)	Vannamei (ton)	Tiger Prawns (ton)	Parrot Fish (ton)	Tilapia (ton)	Grouper (ton)	Crab (ton)
2024	Baktiya Barat	63.0	57	38.5	1.3	2.4	0.7	0.3	1.0	0.3
2024	Dewantara	234.0	208	130.0	9.1	7.1	3.3	1.6	0.7	1.2
2024	Geureudong Pase	13.5	12	5.9	0.2	0.7	0.1	0.1	4.0	0.1
2024	Lapang	114.0	106	79.3	2.6	5.3	2.1	0.5	2.0	0.8
2024	Langkahan	31.0	28	20.0	0.7	1.3	0.2	0.1	1.0	0.2
2024	Lhoksukon	101.0	80	56.9	1.7	3.1	1.1	0.6	3.0	0.5
2024	Matangkuli	23.0	21	12.2	0.5	1.0	0.3	0.2	2.0	0.2
2024	Meurah Mulia	53.0	49	38.0	1.1	1.8	0.4	0.2	0.0	0.3
2024	Muara Batu	184.0	138	95.0	4.8	5.7	2.6	1.1	0.4	1.0
2024	Nibong	27.0	23	15.0	0.6	1.1	0.3	0.2	1.0	0.2

As an illustration of the clustering process, the first 10 data entries were analyzed using the K-Means algorithm. Prior to the modeling phase, all numerical data were normalized using the Z-Score method to ensure that each attribute contributes equally to the clustering process. The Elbow Method was used to determine the optimal number of clusters, while the validity of the results was assessed using the Silhouette Score. This approach ensures meaningful cluster formation with distinct boundaries and internal consistency within each group.

1. Z-Score Normalization

$$Z = \frac{X - \mu}{\sigma}$$

$$\text{Mean} = \frac{63.0+234.0+13.5+114.0+31.0+101.0+23.0+53.0+184.0+27.0}{10}$$

$$= 84.35$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum(x_i - \mu)^2} = \sqrt{\frac{49741.45}{9}} \approx \sqrt{5526.83} = 74.32$$

$$Z = \frac{63.0-84.35}{74.32} = \frac{-21.35}{74.32} = -0.287$$

All data has been normalized using the Z-Score formula the results are shown below.

TABLE III
NORMALIZATION RESULTS

No	Sub-District	Pond Area	Farmer	Milkfish	Vannamei	Tiger Prawns	Parrot Fish	Tilapia	Grouper	Crab
1	Baktiya Barat	-0.29	-0.24	-0.26	-0.35	-0.24	-0.36	-0.38	-0.41	-0.47
2	Dewantara	2.01	2.17	1.98	2.48	1.82	1.91	2.24	-0.65	1.87
3	Geureudong Pase	-0.95	-0.96	-1.05	-0.75	-0.98	-0.88	-0.79	2.01	-0.99
4	Lapang	0.4	0.54	0.74	0.12	1.03	0.86	0.02	0.4	0.83
5	Langkahan	-0.72	-0.71	-0.71	-0.57	-0.72	-0.79	-0.79	-0.41	-0.73
6	Lhoksukon	0.22	0.12	0.19	-0.2	0.07	-0.01	0.22	1.2	0.05
7	Matangkuli	-0.82	-0.82	-0.9	-0.64	-0.85	-0.71	-0.59	0.4	-0.73
8	Meurah Mulia	-0.42	-0.37	-0.27	-0.42	-0.5	-0.62	-0.59	-1.22	-0.47
9	Muara Batu	1.34	1.05	1.12	0.92	1.2	1.3	1.23	-0.89	1.35
10	Nibong	-0.77	-0.78	-0.83	-0.6	-0.81	-0.71	-0.59	-0.41	-0.73

2. K-Means Clustering

After all the data were normalized using the Z-Score method, the process continued with the application of the K-Means algorithm to perform clustering. In the initial stage, the number of clusters was set to three (k = 3), based on domain knowledge and the evaluation results from the Elbow Method. Subsequently, three initial centroids were manually selected from three different data rows for the purpose of manual calculation in the first iteration, namely:

- C0: Baktiya Barat Sub-district
- C1: Langkahan Sub-district
- C2: Muara Batu Sub-district

Each data point was then measured for its distance to the three centroids using the Euclidean distance formula, taking into account all normalized numerical features. The data were classified into clusters based on the nearest centroid.

To illustrate the algorithm process more clearly, a manual calculation was performed for the first iteration, starting with the computation of Euclidean distances between each data point and the three initial centroids. The results of this calculation demonstrate how the data are clustered based on the similarity of feature values to the nearest centroid.

The centroids were then updated by averaging the members of each cluster. While the process may iterate several times until the results stabilize, this paper only presents the manual calculation for the first iteration. The remaining iterations were performed automatically until convergence in the seventh iteration. Manual Calculation Example:

TABLE IV
Z-SCORE VALUES OF LAPANG SUB-DISTRICT AND CENTROID

Feature	Z(Lapang)	Z(C0)	Z(C1)
Pond Area	0.398	-0.287	-0.717
Farmers	0.512	-0.236	-0.646
Milkfish	0.635	-0.395	-0.835
Vaname Shrimp	0.093	-0.414	-0.692
Windu Shrimp	1.122	-0.287	-0.692
Tilapia	0.691	-0.531	-0.969
Mujair	-0.008	-0.592	-1.006
Grouper	0.327	-0.484	-0.484
Crab	0.332	-0.486	-0.856

Euclidean Distance Calculation

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots + (x_n - c_n)^2}$$

Distance results:

- d(Lapang, C0) = √7.497 ≈ 2.739
- d(Lapang, C1) = √14.474 ≈ 3.804
- d(Lapang, C2) = √6.304 ≈ 2.511

TABLE V
CLUSTERING RESULT

Sub-district	Dist. to C0	Dist. to C1	Dist. to C2	Cluster
Baktiya Barat	0.0	1.141	4.336	0
Dewantara	6.771	7.867	2.637	2
Geureudong Pase	2.976	2.503	6.656	1
Lapang	2.81	3.86	2.356	2
Langkahan	1.141	0.0	5.439	1
Lhoksukon	2.017	2.824	3.786	0
Matangkuli	1.54	0.884	5.676	1
Meurah Mulia	0.932	1.118	4.707	0
Muara Batu	4.336	5.439	0.0	2
Nibong	1.215	0.285	5.466	1

The Elbow Method is used to evaluate the optimal number of clusters based on the SSE (Sum of Squared Error) or WCSS (Within-Cluster Sum of Squares). The formula is as follows:

$$SSE = \sum_{i=1}^{n} (d_i)^2$$

Where d_i is the distance of each data point to its centroid. The manual calculation results for the first iteration are presented as follows:

- Cluster 0
  $0.000^2 + 2.017^2 + 0.932^2 = 0 + 4.068 + 0.869 = 4.937$
- Cluster 1
  $0.000^2 + 2.503^2 + 0.884^2 + 0.285^2 = 0 + 6.265 + 0.781 + 0.081 = 7.127$
- Cluster 2
  $2.637^2 + 2.356^2 + 0.000^2 = 6.957 + 5.551 + 0 = 12.508$

Total SSE (Iteration 1): 4.937 + 7.127 + 12.508 = 24.572

After identifying k = 3 as the optimal number of clusters using the Elbow Method, the clustering quality was validated through a Silhouette Score of 0.5293, indicating a moderately good structure. To improve interpretability, PCA was applied, and the 2D visualization confirmed clear cluster separation, reinforcing the model's validity.

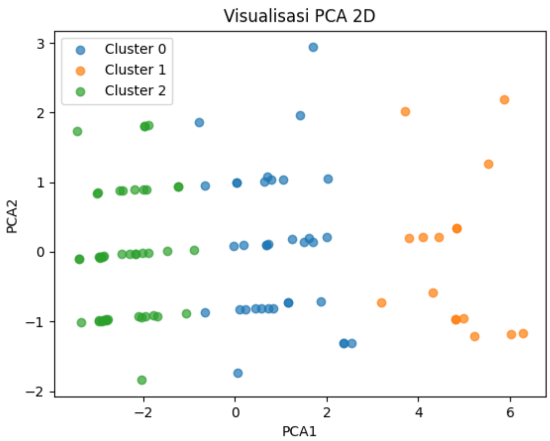


Figure 3. PCA visualization

These additional evaluations help confirm that the use of the K-Means algorithm provides a reasonable basis for segmenting coastal subdistricts based on aquaculture productivity, which can support further policy formulation and decision-making.

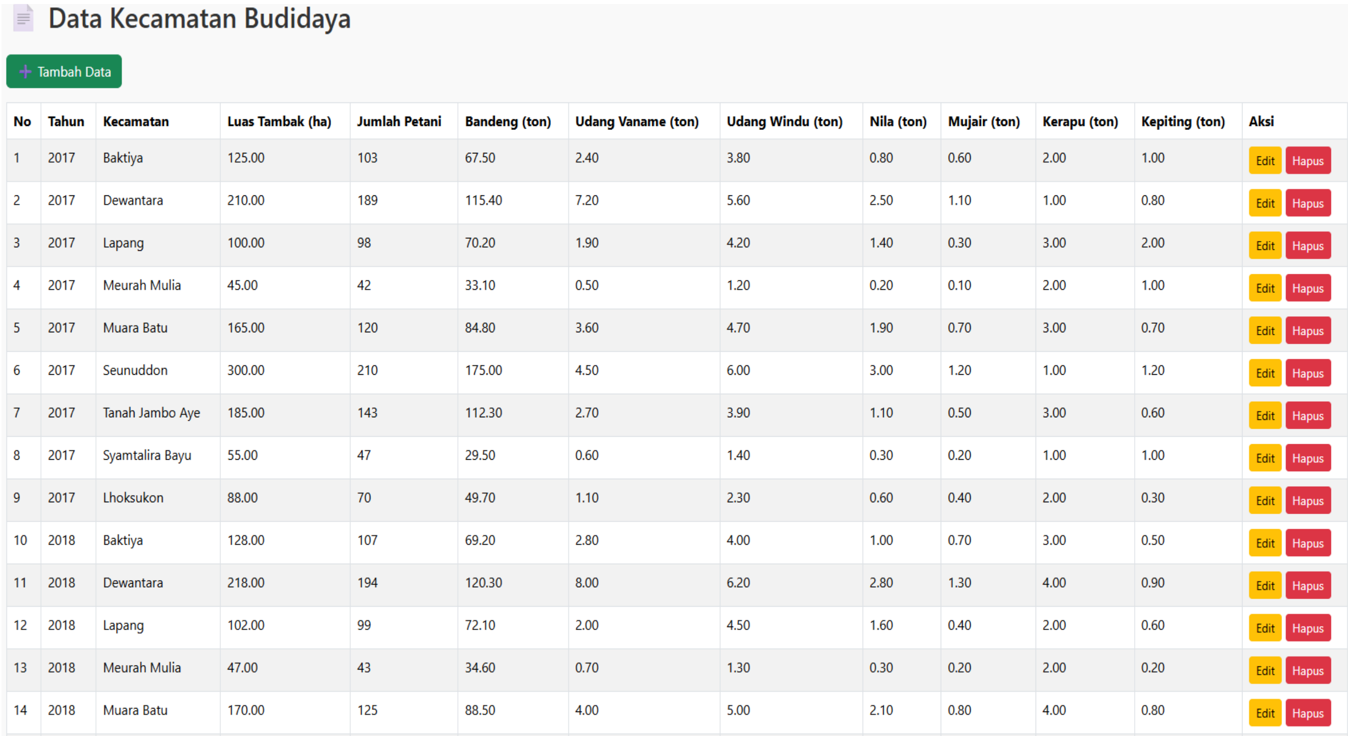
The final clustering result divided the coastal subdistricts into three distinct clusters based on aquaculture productivity levels and pond area utilization:

- Cluster 0 represents areas with narrow pond areas and low production volumes. These regions generally face limitations in infrastructure, technical capacity, or natural resources. Such areas are suitable targets for intensive training, technical assistance, and improvement programs to boost their aquaculture performance.
- Cluster 1 consists of subdistricts with moderate pond sizes and average productivity levels. These regions may benefit from infrastructure strengthening, input subsidies, or market access support to further improve their performance.
- Cluster 2 includes areas with large pond areas and high production volumes, indicating strong aquaculture potential. These regions may serve as export centers or pilot zones for advanced aquaculture practices. Policy efforts in these areas can focus on scaling production, improving export infrastructure, and ensuring sustainability.

The average production volume and number of farmers were also summarized per cluster to support this classification and provide a clearer picture of regional disparities. These insights may help local governments allocate resources more efficiently, design targeted interventions, and develop data-driven development strategies based on the actual condition of each coastal area.

B. *Website Implementation*

The Aquaculture Data Page displays detailed information on coastal aquaculture in North Aceh Regency, including sub-district names, pond area, number of farmers, and production volumes of key commodities such as milkfish, vannamei shrimp, tiger shrimp, tilapia, mojarra, grouper, and crab. The interface is designed to be user-friendly and responsive, featuring a data search function and an "Add Data" button to input new entries. This dataset serves as the foundation for the clustering process using the K-Means algorithm, aiming to group regions with similar aquaculture characteristics. The clustering results are expected to support more focused and effective development strategies for coastal resource management.



Data Kecamatan Budidaya

Tambah Data

No	Tahun	Kecamatan	Luas Tambak (ha)	Jumlah Petani	Bandeng (ton)	Udang Vaname (ton)	Udang Windu (ton)	Nila (ton)	Mujair (ton)	Kerapu (ton)	Kepiting (ton)	Aksi
1	2017	Baktiya	125.00	103	67.50	2.40	3.80	0.80	0.60	2.00	1.00	Edit Hapus
2	2017	Dewantara	210.00	189	115.40	7.20	5.60	2.50	1.10	1.00	0.80	Edit Hapus
3	2017	Lapang	100.00	98	70.20	1.90	4.20	1.40	0.30	3.00	2.00	Edit Hapus
4	2017	Meurah Mulia	45.00	42	33.10	0.50	1.20	0.20	0.10	2.00	1.00	Edit Hapus
5	2017	Muara Batu	165.00	120	84.80	3.60	4.70	1.90	0.70	3.00	0.70	Edit Hapus
6	2017	Seunuddon	300.00	210	175.00	4.50	6.00	3.00	1.20	1.00	1.20	Edit Hapus
7	2017	Tanah Jambo Aye	185.00	143	112.30	2.70	3.90	1.10	0.50	3.00	0.60	Edit Hapus
8	2017	Syamtalira Bayu	55.00	47	29.50	0.60	1.40	0.30	0.20	1.00	1.00	Edit Hapus
9	2017	Lhoksukon	88.00	70	49.70	1.10	2.30	0.60	0.40	2.00	0.30	Edit Hapus
10	2018	Baktiya	128.00	107	69.20	2.80	4.00	1.00	0.70	3.00	0.50	Edit Hapus
11	2018	Dewantara	218.00	194	120.30	8.00	6.20	2.80	1.30	4.00	0.90	Edit Hapus
12	2018	Lapang	102.00	99	72.10	2.00	4.50	1.60	0.40	2.00	0.60	Edit Hapus
13	2018	Meurah Mulia	47.00	43	34.60	0.70	1.30	0.30	0.20	2.00	0.20	Edit Hapus
14	2018	Muara Batu	170.00	125	88.50	4.00	5.00	2.10	0.80	4.00	0.80	Edit Hapus

Figure 4.Data Page

The web application developed in this study provides an interactive platform for managing aquaculture data in the coastal regions of North Aceh Regency. As shown in the "Aquaculture Data" page, users can view detailed information for each subdistrict, including the year, subdistrict name, pond area (ha), number of farmers, and production volumes (in tons) for various aquaculture commodities such as milkfish, vannamei shrimp, tiger shrimp, tilapia, mojarra, grouper, and crab.

The interface is designed to be simple and responsive, with key functionalities such as "Add Data", "Edit", and "Delete", allowing users to update the database directly through the browser. The sidebar menu enables easy navigation between different modules, including Dashboard, Data, Analysis, Clustering, and Results. This structured design improves usability and ensures that the collected data can be efficiently processed and used as the basis for clustering analysis using the K-Means algorithm.

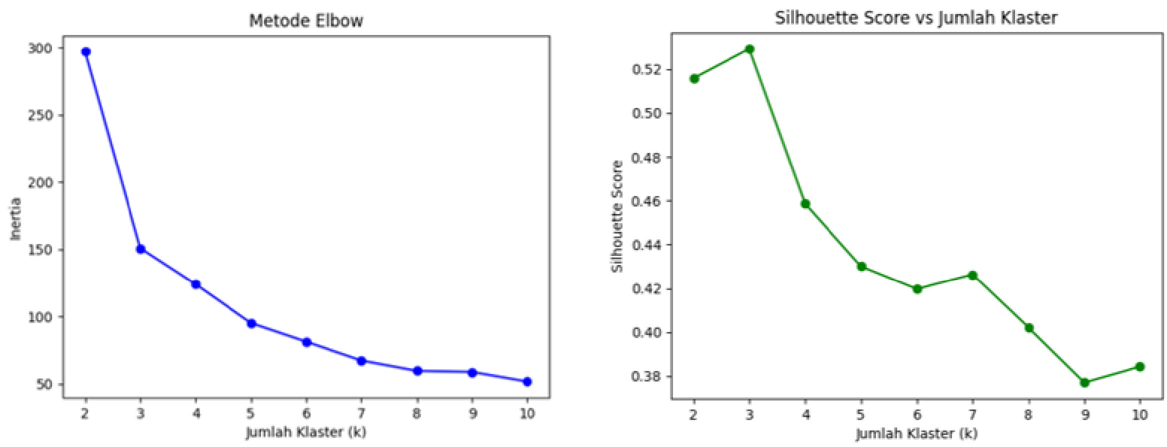


Figure 5.Clustering Evaluation Page

The "Model Evaluation" page displays the results of determining the optimal number of clusters for the K-Means algorithm. Two key evaluation methods are visualized: the Elbow Method and the Silhouette Score.

The Elbow Method graph shows the relationship between the number of clusters (k) and the total within-cluster sum of squares (inertia). A noticeable "elbow" occurs at k = 3, where the rate of decrease in inertia slows down significantly, suggesting this as the optimal cluster count. The Silhouette Score graph measures how well each data point fits within its assigned cluster. A higher silhouette score indicates better-defined clusters. The highest silhouette score of 0.5293 is also achieved at k = 3, validating the result of the Elbow Method. A summary message at the bottom of the page confirms that the optimal number of clusters is 3, with a silhouette score of 0.5293, indicating moderately good clustering quality.

Unduh Hasil Klasterisasi (Excel)

ID	Tahun	Kecamatan	Luas Tambak (ha)	Jumlah Petani	Bandeng (ton)	Udang Vaname (ton)	Udang Windu (ton)	Nila (ton)	Mujair (ton)	Kerapu (ton)	Kepiting (ton)	Cluster
1	2017	Baktiya	125.00	103	67.50	2.40	3.80	0.80	0.60	2.00	1.00	Cluster 0
2	2017	Dewantara	210.00	189	115.40	7.20	5.60	2.50	1.10	1.00	0.80	Cluster 1
3	2017	Lapang	100.00	98	70.20	1.90	4.20	1.40	0.30	3.00	2.00	Cluster 0
4	2017	Meurah Mulia	45.00	42	33.10	0.50	1.20	0.20	0.10	2.00	1.00	Cluster 2
5	2017	Muara Batu	165.00	120	84.80	3.60	4.70	1.90	0.70	3.00	0.70	Cluster 0
6	2017	Seunuddon	300.00	210	175.00	4.50	6.00	3.00	1.20	1.00	1.20	Cluster 1
7	2017	Tanah Jambo Aye	185.00	143	112.30	2.70	3.90	1.10	0.50	3.00	0.60	Cluster 0
8	2017	Syamtalira Bayu	55.00	47	29.50	0.60	1.40	0.30	0.20	1.00	1.00	Cluster 2
9	2017	Lhoksukon	88.00	70	49.70	1.10	2.30	0.60	0.40	2.00	0.30	Cluster 2
10	2018	Baktiya	128.00	107	69.20	2.80	4.00	1.00	0.70	3.00	0.50	Cluster 0
11	2018	Dewantara	218.00	194	120.30	8.00	6.20	2.80	1.30	4.00	0.90	Cluster 1
12	2018	Lapang	102.00	99	72.10	2.00	4.50	1.60	0.40	2.00	0.60	Cluster 0
13	2018	Meurah Mulia	47.00	43	34.60	0.70	1.30	0.30	0.20	2.00	0.20	Cluster 2
14	2018	Muara Batu	170.00	125	88.50	4.00	5.00	2.10	0.80	4.00	0.80	Cluster 0
15	2018	Seunuddon	310.00	215	180.70	5.00	6.30	3.30	1.40	2.00	1.30	Cluster 1
16	2018	Tanah Jambo Aye	190.00	146	115.60	3.00	4.20	1.30	0.60	1.00	0.70	Cluster 0
17	2018	Syamtalira Bayu	58.00	49	31.20	0.70	1.60	0.40	0.30	2.00	0.30	Cluster 2
18	2018	Lhoksukon	91.00	72	51.00	1.30	2.50	0.70	0.50	3.00	0.40	Cluster 2
19	2019	Baktiya	130.00	109	70.80	2.90	4.10	1.10	0.70	1.00	0.50	Cluster 0
20	2019	Dewantara	220.00	197	122.60	8.30	6.50	2.90	1.40	2.00	1.00	Cluster 1
21	2019	Lapang	104.00	100	73.50	2.10	4.60	1.70	0.40	1.00	0.60	Cluster 0
22	2019	Meurah Mulia	48.00	44	35.20	0.80	1.40	0.30	0.20	3.00	0.20	Cluster 2

Figure 6. K-Means Clustering Result

This web-based display presents the results of a clustering analysis of aquaculture productivity data in the coastal areas of North Aceh Regency from 2017 to 2024**.** The analysis was conducted using the K-Means algorithm, a widely used method for grouping data into clusters based on attribute similarity. The displayed data includes key variables such as pond area (in hectares), number of farmers, and production volume (in tons) of several leading aquaculture commodities, including milkfish, vannamei shrimp, tiger shrimp, tilapia, mujair, grouper, and crab. Each row represents subdistrict-level data for a specific year and is grouped into one of three clusters—Cluster 0, Cluster 1, or Cluster 2—based on similarities in production profiles. To support visual interpretation, each cluster is marked with a distinct color label.

This clustering implementation offers insights into regional productivity disparities and helps identify subdistricts with high, moderate, or low production levels. Such findings are essential for policymakers and stakeholders in designing more targeted and effective interventions. Areas with consistently low productivity can be prioritized for development support, while high-performing areas may serve as models of best practices.

Although the screenshot only captures part of the full dataset due to display limitations, the system encompasses eight years of data and supports dynamic information presentation. Additionally, the dashboard includes an export-to-Excel feature to facilitate further analysis and reporting. Overall, this system enables data-driven decision-making by transforming raw aquaculture data into meaningful cluster insights, contributing to the sustainable development of the regional aquaculture sector.

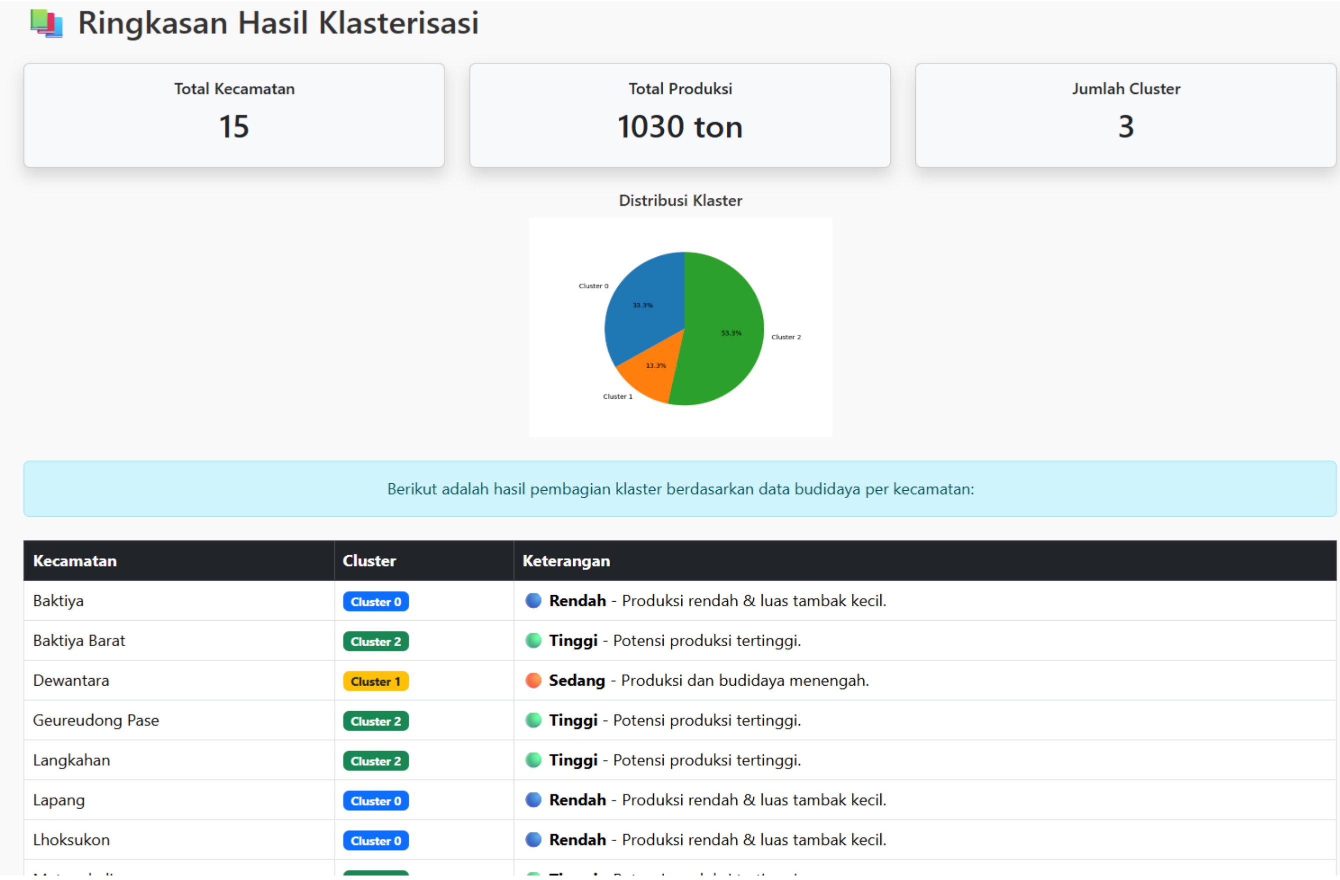


Figure 7. Clustering Summary Result

The Clustering Summary page presents an overview of the final results from the K-Means clustering process applied to aquaculture data in North Aceh Regency. This page summarizes three key indicators:

- Total Sub-districts: 15
- Total Production: 1030 tons
- Number of Clusters: 3

The pie chart at the center visualizes the distribution of sub-districts across three clusters.

A table below provides a detailed cluster assignment for each sub-district along with its interpretation:

- Sub-districts in Cluster 0 are categorized as Low, indicating small-scale production areas.
- Cluster 1 is labeled as Medium, reflecting moderate aquaculture activities.
- Cluster 2 represents High potential regions with large pond areas and high yields.

This page helps decision-makers quickly identify regions with varying levels of aquaculture productivity, enabling more effective and targeted planning, especially for infrastructure investment, training, or resource allocation.

Based on these results, an additional evaluation was conducted to assess the model's sensitivity and explore alternative clustering approaches. To assess the sensitivity of the clustering parameter K, several values (K=2, 3, and 4) were tested. The result showed that K = 3 yielded the highest Silhouette Score of 0.5293, while K = 2 and K = 4 resulted in scores of 0.48 and 0.50, respectively. This confirms that K = 3 provides the most balanced and stable clustering structure. Furthermore, for future research, it is recommended to compare clustering results with alternative algorithms such as DBSCAN or Agglomerative Clustering. Incorporating spatial data and GIS-based visualization is also suggested to enhance the interpretability of cluster distribution across coastal areas.

IV. Conclusion

This study successfully clustered coastal subdistricts in North Aceh Regency based on the productivity of dominant aquaculture commodities using the K-Means algorithm. The dataset was normalized using the Z-Score method, and the optimal number of clusters was determined through the Elbow Method, resulting in three clusters. Clustering quality was evaluated using the Silhouette Score, which yielded a value of 0.5293, indicating a moderately good clustering structure.

The three clusters represent different levels of aquaculture performance: Cluster 0 includes low-productivity areas with limited pond area; Cluster 1 represents moderate productivity; and Cluster 2 covers subdistricts with high productivity and large aquaculture areas. These distinctions are valuable for policymakers to design targeted strategies. For instance, Cluster 0 areas may benefit from technical training and assistance, while Cluster 2 has the potential to be developed into aquaculture export hubs.

To enhance interpretability, the clustering results were visualized using PCA and implemented into an interactive web-based application using the Flask framework. This system enables local governments and stakeholders to explore the results and support evidence-based planning for regional aquaculture development.

However, this study has limitations, particularly regarding external validation. The clustering results have not yet been cross-checked with official maps or expert feedback. Future research should consider spatial validation, expert involvement, and comparisons with alternative algorithms such as DBSCAN or Agglomerative Clustering. The inclusion of environmental, infrastructure, or socioeconomic variables, along with GIS-based system enhancements, is also recommended to increase the system's practical utility for policy planning.

REFERENCES

[1] P. H. Merdeka, "Manajemen Peningkatan Kesejahteraan Masyarakat Pesisir Melalui Pemberdayaan Usaha Lokal Masyarakat: a Review," *J. Accounting, Manag. Econ. Bus.*, vol. 1, no. 1, pp. 1–9, 2023, doi: 10.56855/analysis.v1i1.180.

[2] Nabila Afifah Azuga, "Kajian Kerentanan Kawasan Pesisir Terhadap Bencana Kenaikan Muka Air Laut (Sea Level Rise) Di Indonesia," *J. Ris. Kelaut. Trop. (Journal Trop. Mar. Res.*, vol. 3, no. 2, pp. 65–76, 2021, doi: 10.30649/jrkt.v3i2.41.

[3] Estu Nugroho and T. H. dan M. N. , Raden Roro Sri Pudji Sinarni Dewi , Aisyah, "Pemanfaatan Sumberdaya Kelautan Dan Perikanan Melalui Budidaya Perikanan Berkelanjutan Menuju Masyarakat Pembudidaya 5.0 Utilization," *J. Kebijak. Perikan. Indones.*, vol. 14, no. 2, pp. 111–119, 2022, [Online]. Available: http://ejournal-balitbang.kkp.go.id/index.php/jkpi

[4] Y. Yunarty, A. Kurniaji, and K. Kasmatang, "Pemeriksaan Ektoparasit pada Berbagai Komoditas Budidaya Perikanan Payau," *Biosci. J. Ilm. Biol.*, vol. 11, no. 1, p. 579, 2023, doi: 10.33394/bioscientist.v11i1.7699.

[5] E. Sulistiyawan, A. Hapsery, and L. J. A. Arifahanum, "Perbandingan Metode Optimasi Untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan Di Indonesia (Studi Kasus Dinas Kelautan dan Perikanan Indonesia)," *J. Gaussian*, vol. 10, no. 1, pp. 76–84, 2021, doi: 10.14710/j.gauss.v10i1.30936.

[6] F. A. Maresti, W. I. Rahayu, M. B. C. Lustin, and T. H. Pakpahan, "Implementasi K-Means untuk Melakukan Segmnetasi Produk Berdasarkan Data Transaksi Retail," *J. Ilm. Sains dan Teknol.*, vol. 9, no. 1, pp. 20–32, 2025, doi: 10.47080/saintek.v9i1.3856.

[7] A. Rahmadani, "Implementasi Algoritma K-Means Untuk Clustering Data Inventori," *Indones. J. Inform. Res. Softw. Eng.*, vol. 5, no. 1, pp. 1–11, 2025, [Online]. Available: https://journal.irpi.or.id/index.php/ijirse/article/view/1832

[8] H. Hamsiah, "Implementasi Data Mining Dalam Penerapan Clustering Algoritma K-Medoid Sebaran Mahasiwa Baru Pada STIE-SAK," *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 8, no. 1, pp. 16–24, 2025, doi: 10.47324/ilkominfo.v8i1.311.

[9] S. B. H. Sakur, M. Silangen, and D. Tuwohingide, "Penerapan Algoritma K-Means untuk Klasterisasi Produksi Budidaya Perikanan Provinsi Sulawesi Utara," *J. SAINTEKOM*, vol. 14, no. 1, pp. 38–47, 2024, doi: 10.33020/saintekom.v14i1.528.

[10] S. Tahalea, R. Fadhilah, M. Y. Matdoan, and D. A. Safira, "Clustering Shrimp Distribution in Indonesia Using the X-Means Clustering Algorithm," *Var. J. Stat. Its Appl.*, vol. 6, no. 1, pp. 49–54, 2024, doi: 10.30598/variancevol6iss1page49-54.

[11] S. Sulastri, B. Susetyo, and I. M. Sumertajaya, "The Clustering of the Aquaculture Fisheries Companies in Indonesia Using the K-Prototypes and Two Step Cluster (TSC) Algorithm," *Int. J. Sci. Basic Appl. Res.*, vol. 38, no. 1, pp. 171–186, 2021.

[12] N. Hafni, Y. M. Rangkuti, and I. M. K. Karo, "Distance Function Analysis on Fuzzy C-Means for Clustering Aquaculture Production in North Sumatra," *J. Math. Comput. Stat.*, vol. 8, no. 1, pp. 154–162, 2025, doi: 10.35580/jmathcos.v8i1.6500.

[13] O. Nadia, W. Nasution, N. Ilmi, and J. E. Candra, "Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Hasil Tangkap Ikan Di Kepulauan Riau," *Cetak) J. Innov. Res. Knowl.*, vol. 5, no. 1, pp. 693–704, 2025.

[14] M. Marclyna Nau, V. Nurul Fathya, and O. Pratama Martadireja, "Implementasi Data Mining Pada Analisis Karakteristik Pelanggan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 5209–5215, 2025, doi: 10.36040/jati.v9i3.13725.

[15] M. Alta Melfia, K. Athaya Ramadhini, M. Ajmal Maulana, A. Laura Dasiva, A. Aria Briantoro, and B. Oloan Lubis, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Sembako Terlaris Dengan Algoritma C4.5," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4616–4621, 2025, doi: 10.36040/jati.v9i3.13450.

[16] S. A. Wardani, N. Bahri, A. Varuq, and H. T. Santoso, "Implementasi Data Minning Clustering Dalam Mengelompokan Kasus Perceraian di Provinsi Jawa Timur Menggunakan Algoritma K-Means," vol. 6, no. June, pp. 68–83, 2025.

[17] T. Hartini *et al.*, "Implementasi Algoritma K-Means Clustering," vol. 9, no. 1, pp. 76–83, 2025.

[18] R. Pajri, N. Suarna, I. Ali, and D. Indriya Efendi, "Penerapan Algoritma K-Means Untuk Optimalisasi Klasterisasi Penjualan Obat Di Apotek Perjuagan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1594–1599, 2025, doi: 10.36040/jati.v9i1.12472.

[19] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.

[20] M. Qusyairi, Zul Hidayatullah, and Arnila Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 500–510, 2024, doi: 10.29408/jit.v7i2.26375.

[21] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, doi: 10.14421/jiska.2024.9.3.178-191.

[22] R. Saputra and A. Ramadhanu, "Implementasi Hybrid Intelligent System untuk KlasifikasiJenis Bola Berbasis Computer Vision," *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 318–325, 2025.