Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2382~2390
e-ISSN: 2548-6861

Clustering of Aquaculture Productivity Villages in East Aceh Using the K-Means Algorithm

M. Arif Saputra [1]*, Rozzi Kesuma Dinata [2], Yesy Afrillia [3]***
* Department of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia
* Department of Information Technology, Universitas Malikussaleh, Lhokseumawe, Indonesia
arif.210170044@mhs.unimal.ac.id [1], rozzi@unimal.ac.id [2], yesy.afrillia@unimal.ac.id [3]

Article Info

Article history:

Received 2025-07-07
Revised 2025-08-23
Accepted 2025-09-03

Keyword:

Clustering,
Silhouette Score,
Aquaculture,
K-Means,
East Aceh.

ABSTRACT

This study aims to classify villages based on the level of pond utilization and to develop a web-based application for categorizing aquaculture areas in East Aceh Regency. In contrast to traditional definitions based on harvest volume, this research defines productivity functionally—whether the pond area is actively managed or abandoned. The dataset consists of 146 villages and includes five primary variables: number of fish farmers, total pond area, number of pond plots, productive pond area, and abandoned pond area. Clustering was conducted using the K-Means algorithm, resulting in two main groups: productive and non-productive villages. Validation through the Silhouette Score revealed that using $k = 2$ yielded the highest score of 0.7576, indicating the most optimal clustering structure. The analysis showed that 92% of villages were categorized as productive, while 8% fell into the non-productive cluster. These two clusters differ significantly in terms of land utilization ratios and the number of active aquaculture workers. The findings not only offer a more refined spatial insight but also serve as a basis for the Department of Marine Affairs and Fisheries in formulating aquaculture zoning, revitalization programs, and more targeted resource allocation.

I. Introduction

Aquaculture plays a crucial role in the development of Indonesia's fisheries sector, serving as a key contributor to both national food security and export commodities. High-value species such as shrimp and milkfish are among the main aquaculture products that continue to experience rising demand at both domestic and global levels [1]. In response, efforts to utilize coastal and brackish water areas for aquaculture expansion have intensified, particularly in provinces with significant marine potential like Aceh [2].

East Aceh Regency is recognized as a coastal region with substantial potential for aquaculture, encompassing approximately 18,835.59 hectares of pond area. Supported by favorable geographic characteristics, the region is highly suitable for both intensive and extensive aquaculture development. However, data from the East Aceh Department of Marine Affairs and Fisheries (2023) indicate that around 1,371.93 hectares—or about 7.3% of the total pond area—are classified as inactive or abandoned, meaning they have not been utilized productively in recent years. This imbalance reflects a significant gap in land utilization and suggests underlying issues in the management and spatial distribution of aquaculture resources at the village level [3].

Several factors contribute to the underutilization of aquaculture ponds in the region. These include limited technical capacity among fish farmers, inadequate access to financing and modern aquaculture technologies, and deteriorating infrastructure such as irrigation canals and farm access roads. Environmental challenges—such as seawater intrusion, sediment buildup, and climate variability—also negatively affect water quality and pond suitability. Furthermore, some villages possess extensive pond areas but lack sufficient labor or active aquaculture institutions, leading to many ponds being left idle or managed inefficiently [4].

Despite these initiatives, the utilization of aquaculture ponds in East Aceh remains uneven. While some villages exhibit active and productive farming activity, many others have abandoned or underutilized ponds. This disparity in land use highlights the need for spatial classification to identify regions that require policy intervention and targeted development. Accurate mapping of aquaculture productivity can serve not only as an evaluation tool but also as a basis for more focused and data-driven planning in coastal resource management [5].

Previous studies have generally assessed aquaculture productivity based on harvest volume, such as the work conducted by S. Budi et al. (2024), which utilized production output per hectare as the primary indicator. Their clustering approach resulted in three distinct groups, with the main focus placed on measurable yields from the aquaculture system. However, such methods often overlook whether the available pond areas are actively utilized or left idle [6]. In contrast, this study introduces a different approach by defining productivity in terms of functional land use specifically, whether a pond is being actively managed or abandoned. Under this perspective, a pond is considered productive as long as it is actively used, regardless of its output level, while ponds that are left unmanaged or neglected are classified as non-productive. [7].

To achieve this classification, a clustering approach using the K-Means algorithm was employed [8]. The data used in this study were sourced from the East Aceh Department of Marine Affairs and Fisheries for the year 2023, covering 146 villages. The clustering results indicate that the majority of the villages fall into the productive category, while the rest are classified as non-productive [9][10].

The clustering results are expected to provide a clearer spatial representation of villages categorized as either productive or non-productive, thereby serving as a foundation for more targeted, data-driven fisheries resource management policies. Practically, the outcomes of this clustering analysis can support the Department of Marine Affairs and Fisheries (DKP) in designing aquaculture zoning plans based on actual land utilization levels, identifying priority villages for pond revitalization programs, and allocating budgets and technical assistance more efficiently. This spatial-based mapping approach enables more informed decision-making by relying on functional land-use indicators and geographic distribution, rather than general assumptions or land area alone.

II. METHODOLOGY

This research is a descriptive quantitative study that applies a data mining approach to classify aquaculture productivity areas in East Aceh Regency. One of the data mining techniques used in this study is clustering, which aims to group data into specific clusters based on characteristic similarities. The selected algorithm is K-Means, as it efficiently and systematically partitions data into homogeneous groups based on centroid values and the distance between data points.

A. Basic Concept

1) Data Mining: Data mining is the process of extracting hidden information or patterns from large datasets using various analytical methods and techniques [11].

2) Clustering: Clustering is a technique used to group data based on specific characteristics [12].

B. K-Means Method

K-Means is one of the data analysis methods used to divide a dataset into several clusters based on similar characteristics [13].The K-Means algorithm scheme is illustrated in Figure1.

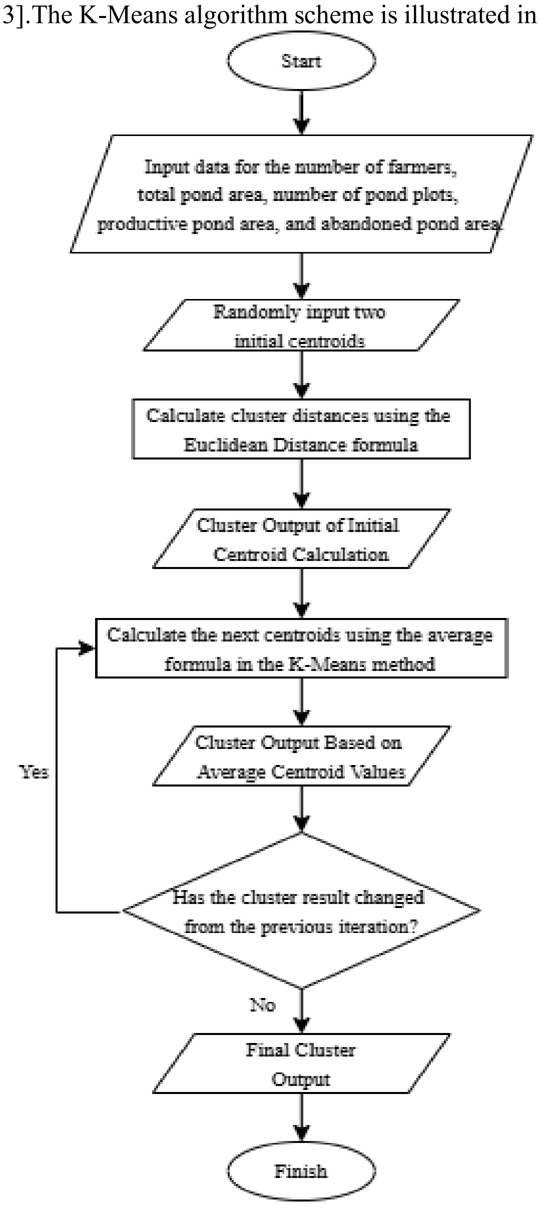

Figure 1. Schematic of Clustering Algorithm

The steps involved in the K-Means algorithm calculation are as follows:

- Input data for the number of farmers, total pond area, number of pond plots, productive pond area, and abandoned pond area.
- Randomly input two initial centroids.

- Calculate cluster distances using the Euclidean Distance formula:

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

- Calculate the next centroids using the average formula in the K-Means method:

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

- Repeat steps 3 onward until there are no changes in cluster membership.

C. Preprocessing and Data Normalization

Prior to the clustering process, the numerical dataset underwent a preprocessing stage through normalization. Data normalization is the process of transforming variable values into a consistent scale to prevent disproportionate influence from any single attribute [14]. In this study, min-max normalization was applied to ensure that all features were within the same range, allowing the K-Means clustering algorithm to treat each variable fairly during distance calculations [15].

$$X' : \frac{X - X_{min}}{X_{max} - X_{min}}$$

D. Silhouette Score

To determine the most appropriate number of clusters (*k*) in this study, an internal evaluation method known as the Silhouette Score was applied. This metric assesses how well each data point fits within its assigned cluster by considering its proximity to other members of the same cluster and its distance from neighboring clusters. The Silhouette Score ranges from 1 to 1, higher values indicate better separation between clusters and stronger cohesion within them. This evaluation ensures that the clustering results generated by the K-Means algorithm are not only numerically formed, but also statistically valid and meaningfully interpretable [16].

E. System Design

To automate the clustering process, a web-based system was developed using PHP as the programming language and MySQL as the database management system. The user interface was designed using HTML, CSS, and JavaScript to facilitate data entry, execute the clustering process, and present the results in both tabular and visual formats.

This system includes features such as data upload, automated clustering, result visualization, and display of centroids and iteration processes. The system workflow is illustrated in the following flowchart:

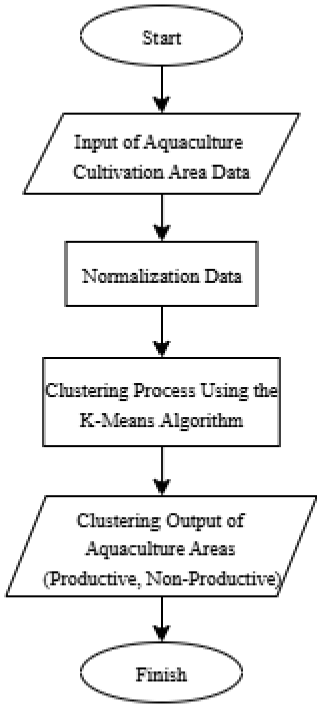

Figure 2. Schematic of Clustering System

F. Needs Analysis

1) Hardware Requirement Analysis: The hardware specifications were determined to ensure that the system operates optimally. The hardware used in the design of this system includes the following components:

- Laptop Lenovo Ideapad Slim 3
- Processor 11th Gen Intel Core i3-1115G4 3.00 GHz
- RAM 8GB
- SSD 512GB

2) Software Requirement Analysis: The design and development of a system begin with software specification as a primary step. The following is the general specification of the software used in this study:

- System operation windows 11
- Microsoft office
- XAMPP
- Visual studio code

G. Validation of Results

System validation was carried out by comparing the clustering results generated automatically by the system with those obtained through manual calculations using Microsoft Excel. The manual process followed the stages of the K-Means algorithm, including distance calculation using the Euclidean Distance formula and centroid determination based on the average values. The purpose of this validation was to ensure that the system produces consistent output that aligns with the intended algorithmic logic.

III. Result And Discussion

All results presented in this section were obtained through the clustering process using the K-Means algorithm applied to aquaculture data from East Aceh Regency. This study does not compare the performance of other clustering algorithms, as the primary focus is on implementing the K-Means method to group sub-districts based on productivity indicators. Evaluation was carried out through iterative stages, including centroid initialization and updating, cluster membership identification, and validation of results through comparison with manual calculations conducted using Microsoft Excel.

A. Data Description

The data used in this study covers information from 146 regions in East Aceh Regency for the year 2023, consisting of five main variables: number of fish farmers, total pond area, number of pond plots, productive pond area, and abandoned pond area. All data were obtained from the Department of Marine Affairs and Fisheries of East Aceh Regency. Each region is expected to be classified into one of two categories (productive or non-productive) based on the combination of these five variables.

1) Data Criteria: To facilitate formatting in a two-column layout, each variable was assigned a specific code, which is used during the normalization and clustering stages. The explanation of each code and its corresponding description is provided in Table I.

TABLE I
DATA CRITERIA

No	Criteria Code	Criteria (Unit)
1	X1	Number of Farmers
2	X2	Total Pond Area (Ha)
3	X3	Number of Pond Plots (Unit)
4	X4	Area of Productive Ponds (Ha)
5	X5	Area of Abandoned Ponds (Ha)

Based on Table I, five criteria were used in this study: X1 represents the number of farmers, X2 the total pond area, X3 the number of pond plots, X4 the area of productive ponds, and X5 the area of abandoned ponds.

2) Research Dataset: The following dataset was obtained in 2023 from the Department of Marine Affairs and Fisheries of East Aceh Regency.

TABLE II
RESEARCH DATASET

No	Name of Village	X1	X2	X3	X4	X5
1	Paya Peulawi	113	345	78	345	0
2	Aramiah	55	330	169	307,5	22,5
3	Keude Birem	56	125	114	118	7
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
146	Alue Ie Itam	30	0,5	0	0,5	0

B. Data Normalization

To ensure equal scale across variables, the data was normalized using the Min-Max method before the clustering process was carried out. This technique transforms the original values into a standardized range between 0 and 1.

TABLE III
NORMALIZATION PROCESS DATASET

No	Name of Village	X1	X2	X3	X4	X5
1	Paya Peulawi	113	345	78	345	0
2	Aramiah	55	330	169	307,5	22,5
3	Keude Birem	56	125	114	118	7
4	Birem Rayek	53	266	263	256	10
5	Seuneubok Dalam	43	146,5	84	124,5	22
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
146	Alue Ie Itam	30	0,5	0	0,5	0
MIN		0	0,0016	0	0	0
MAX		333	2269	526	2122	173

$$X' : \frac{113-0}{333-0} = \frac{113}{333} = 0{,}339$$

The results of data normalization can be seen in table IV:

TABLE IV
NORMALIZATION DATASET

No	Name of Village	X1	X2	X3	X4	X5
1	Paya Peulawi	0,339	0,152	0,148	0,163	0,000
2	Aramiah	0,165	0,145	0,321	0,145	0,130
3	Keude Birem	0,168	0,055	0,217	0,056	0,040
4	Birem Rayek	0,159	0,177	0,500	0,121	0,058
5	Seuneubok Dalam	0,129	0.065	0,161	0,059	0,127
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
146	Alue Ie Itam	0,090	0,000	0,000	0,000	0,000

C. Evaluation of Optimal Cluster Number Using Silhouette Score

To validate the optimal number of clusters (k) in the K-Means clustering process, this study employed the Silhouette Score method. This metric measures how well each data point fits within its assigned cluster compared to other clusters, based on intra-cluster cohesion and inter-cluster separation. A higher Silhouette Score indicates a more appropriate clustering structure. Based on the evaluation of multiple k values (from 2 to 10), the highest Silhouette Score (0.7576) was obtained when $k = 2$, suggesting that dividing the dataset into two clusters—productive and non-productive villages—is both statistically and structurally valid. This finding further supports the interpretability of the clustering results for decision-making purposes in spatial planning and aquaculture policy development.

TABLE V
RESULT OF SILHOUETTE SCORE

Number of Clusters (k)	*Silhouette Score*	*Interpretation*
2	0,7576	Very Good
3	0,6932	Good
4	0,5289	Fair
5	0,5244	Fair
6	0,5368	Fair
7	0,5314	Fair
8	0,5386	Fair
9	0,4056	Sufficient
10	0,3923	Sufficient

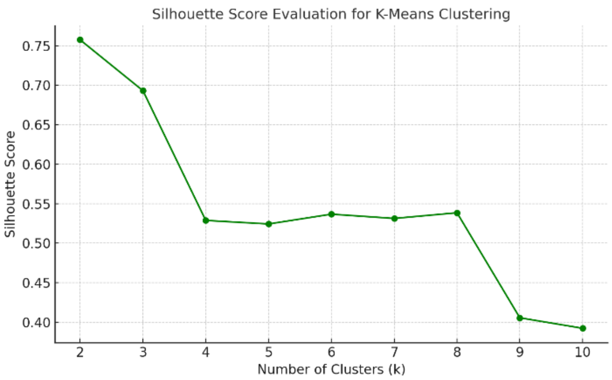

Figure 3. Graphic of Silhouette Score

D. Manual Calculation Process of K-Means Clustering

After the normalization process is completed, the next step is the implementation of the K-Means algorithm. This process begins with the random selection of initial centroids, followed by calculating the distance of each data point to the centroids using the Euclidean Distance formula. Each data point is then classified into one of two clusters—productive or non-productive—based on its proximity to the nearest centroid. The centroid values are then updated by calculating the average of all members within each cluster. This step is repeated iteratively until the cluster membership composition no longer changes.

1) Initial Centroid Determination:

TABLE VI
INITIAL CENTROID

Initial Centroid	X1	X2	X3	X4	X5
C1	0,018	0,000	0,000	0,000	0,001
C2	0,586	0,433	0,494	0,463	0,000

2) Distance Calculation Using Euclidean Distance Formula:

$$d(1) = \sqrt{\begin{matrix}\{(0{,}339 - 0{,}018)^2 + (0{,}152 - 0{,}000)^2 \\ + (0{,}148 - 0{,}000)^2 + (0{,}163 - 0{,}000)^2 \\ + (0{,}000 - 0{,}001)^2\}\end{matrix}}$$

$$d(1) = 0{,}418043136$$

$$d(2) = \sqrt{\begin{matrix}\{(0{,}339 - 0{,}586)^2 + (0{,}152 - 0{,}433)^2 \\ + (0{,}148 - 0{,}494)^2 + (0{,}163 - 0{,}463)^2 \\ + (0{,}000 - 0{,}000)^2\}\end{matrix}}$$

$$d(2) = 0{,}583545822$$

The results of the closest distance calculation for each data are presented in table VII.

TABLE VII
RESULT OF ITERATION 1

No	Name of Village	C1	C2	Cluster
1	Paya Peulawi	0,413043136	0,583545822	C1
2	Aramiah	0,428455682	0,626225298	C1
3	Keude Birem	0,277784804	0,7391566	C1
4	Birem Rayek	0,548984542	0,621740049	C1
5	Seuneubok Dalam	0,24816558	0,786337057	C1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
146	Alue Ie Itam	0,072084816	0,93436745	C1

Based on the results in Table VII, the first iteration showed that 139 out of 146 data points were classified into cluster C1 (productive), while the remaining 7 were grouped into cluster C2 (non-productive).

3) Centroid Update: The centroid values were updated by calculating the average of each attribute (X1 to X5) based on the members assigned to each cluster. In this iteration, cluster C1 contained 139 members, so the centroid values were obtained by summing all normalized values of each attribute and dividing them by the number of data points in the cluster. The updated centroid values are presented in Table VIII.

TABLE VIII
NEW CENTROID VALUE

Variable	C1	C2
X1	0,095512779	0,555984556
X2	0,038804831	0,415330445
X3	0,049450174	0,541825095
X4	0,038185772	0.41734213
X5	0,040510172	0,328488852

4) Final Iteration and Clustering Results: After the first iteration, the centroid update process continued iteratively using the same steps until there were no further changes in cluster membership. This process occurred over several iterations, and the final result indicated that the cluster composition had stabilized, signifying that the algorithm had reached convergence. At this stage, the centroids no longer experienced significant changes, and all data points remained within their respective clusters.

TABLE IX
RESULTS OF ITERATION 6 (CONVERGED)

No	Name of Village	C1	C2	Cluster
1	Paya Peulawi	0,323514032	0,534175709	C1
2	Aramiah	0,343739152	0,471188306	C1
3	Keude Birem	0,191331813	0,615939942	C1
4	Birem Rayek	0,47599087	0,554119575	C1
5	Seuneubok Dalam	0,16546819	0,598415591	C1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
146	Alue Ie Itam	0,069341459	0,798263592	C1

Based on the final clustering results, the visual distribution of the two clusters can be seen in Figure 3.

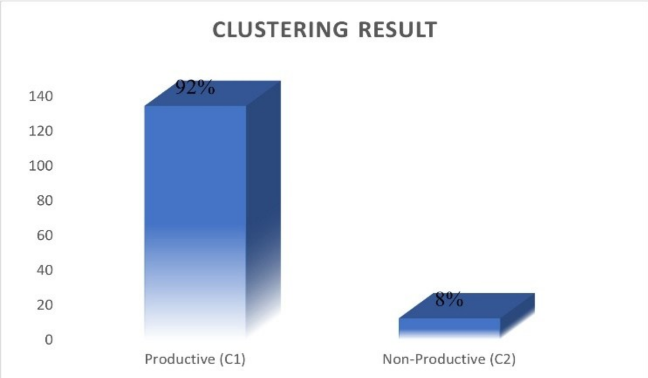

Figure 4. Clustering Result Graph

The clustering results revealed that 92% of the villages were categorized as productive, while the remaining 8% fell into the non-productive group. Although this distribution may appear imbalanced, it reflects the actual conditions as documented in the 2023 dataset from the Department of Marine Affairs and Fisheries of East Aceh Regency.

The villages identified as productive (Cluster C1) include Paya Peulawi, Aramiah, Keude Bireum, Birem Rayeuk, Seuneubok Dalam, Alue Kumba, Simpanga Aneuh, Sarah Teube, Rantau Panjang, Bukit Selamat, Krueng Lingka, Alue Rangan, Kuala Parek, Labuhan Keude, Seuneubok Punti, Alue Bugeng, Seuneubok Teupin, Aluebu Alue Nireh, Seuneubok Paya, Seuneubok Lapang, Geulanggang Meurak, seumatang Keude, Babah Krueng, Kruet Lintang, Jeungki, Seuneubok Dalam, Seuneubok Pidie, Kuala Luege, Seuneubok Peusangan, Dama Tutong, Bale Buya, Matang Gleum, Alue Nibong, Seuneubok Aceh, Bangka Rimueng, Leuge, Cot Muda Itam, Paya Lipah, Matang Peulawi, Kuala Bugak, Cot Keh, Seumatang Muda Itam, Pasir Putih, Alue Bu Jalan, Alue Bu Jalan Baroeh, Beusa Baroeh, Gampong Keude, Paya Due, Alue Ie Itam, Kuala Peudawa, Matang Rayeuk, Matang Bungong, Matag Rayeuk SMK, Matag Rayeuk PP. Seuneubok Rambong, Kuala Peudawa Puntong, Alue Dua Muka S, Alue Dua Muka O, Tanjong Kapai, Keutapang Mameh, Ule Blang, Calok Geulima, Kampung Beunot, Teupin Drum, Kuala Idi Cut, Gampong Baro, Gampong Keude, Seuneubok Baroh, Meunasah Blang, Seneubok Aceh, Matang Pineung, Bagok Panah Sa, Matang Kunyet, Teupin Pukat, Mudak Ara, Keude Bagok Sa, Keude Bagok Dua, Kuala Bagok, Asan Tanjung, Baroh Bugeng, Peulawi, Matang Seuleumak, Kuala Geulumpang, Blang Uyok, Ulee Blang, Labuhan, Gampong Baro, Lhok Seuntang, Simpang Lhee, Teupin Breuh, Keude Tuha, Lampoh Rayeuk, Teupin Mamplam, Alue Buloh Dua, Alue Buloh Sa, Peulalu, Matang Rayeuk, Ulee Ateung, Matang Kupula Lhee, Leung Peut, Matang Guru, Meunasah Tingkeum, Lueng Dua, Pante Bayam, Seuneubok Tuha, Pantee Panah, Seuneubok Saboh, Buket Bata, Alue Ie Mirah, Matang Peureulak, Beurandang, Seuneubok Dalam, Tampak, Blang Barom, Pulo Blang, Simpang Jernih, Batu Sumbang, Nalon, Lot, Jeuring, Bukit Tiga, Paya Awe, Uram Jalam, Panton Rayeuk A, Paya Laman, Panton Rayeuk M, Seuneubok Pango, Dusun Buket Panyang, Keude Blang, Buket Teumpeun, Paya Kruep, Seuneubok Teungeh, Keude Dua, and Alur Ie Itam. And the villages identified as non-productive (Cluster C2) include Alue Raya, Bayeun, Geulumpang Payong, Seuneubok Rawang, Paya Gajah, Matang Neuheun, Naleung, Kuala SP. Ulim, Bantayan, Abeuk Geulanteu, Lueng Sa, and Meunasah Asan.

Most villages remain involved in aquaculture activities, albeit with varying levels of land utilization. All input variables were normalized prior to clustering to ensure equal scaling and reduce the influence of dominant attributes. Furthermore, the decision to use two clusters was validated through the Silhouette Score method, where the highest score was obtained at $k = 2$, indicating a compact and well separated clustering structure. Nevertheless, further validation through field observations or stakeholder interviews is recommended to strengthen the interpretation and reliability of the clustering outcomes.

E. System Implementation

The system was developed using PHP and MySQL, with the capability to receive numerical data input directly from users. The clustering process implemented within the system includes Min-Max normalization, Euclidean distance calculation, centroid updating, and cluster assignment based on proximity. Thus, the system functions not only as a visualization platform but also as an engine that executes the core logic of the K-Means algorithm applied in this research. In addition, it provides interactive visualization features that allow stakeholders to explore, compare, and analyze the status of each village spatially. This supports more informed decision-making in aquaculture policy planning.

1) Main Page: The main page serves as the system's landing view, displaying the application name, a brief description of its functions, and statistical information such as the total number of regions, total number of fish farmers, and overall pond area. The interface is designed to be interactive and informative to help users easily understand the system's purpose.

Figure 5. Main Page

2) Login Page: To ensure secure access, the system provides a login page that requires users to enter their credentials before accessing all available features. The simple and user-friendly interface facilitates the authentication process.

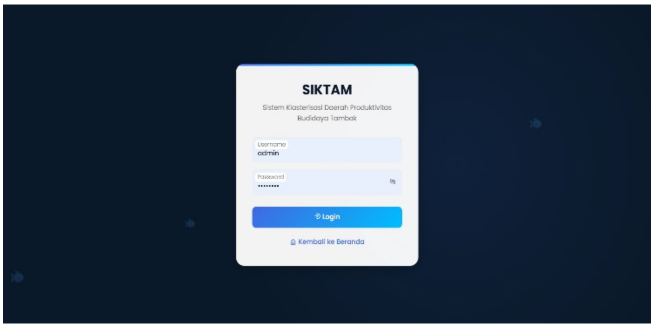

Figure 6. Login Page

3) Dashboard Page: After successfully logging in, users are directed to the dashboard page, which displays statistical summaries and a general overview of the system workflow. The dashboard also serves as an access point to key features such as data management and the clustering process.

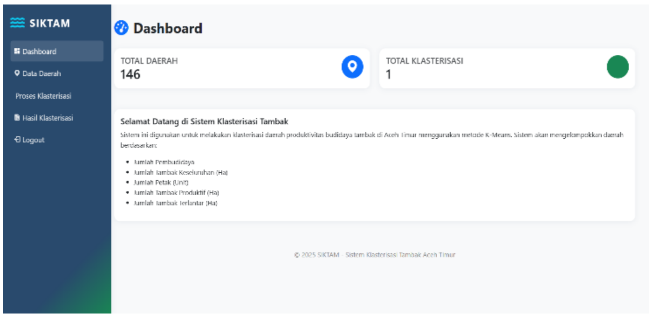

Figure 7. Dashboard Page

4) Data Input Page (Regional Data): This page is used to manage aquaculture area data, including adding, editing, and deleting records. The system accepts input for five key variables: number of fish farmers, total pond area, number of pond plots, productive pond area, and abandoned pond area.

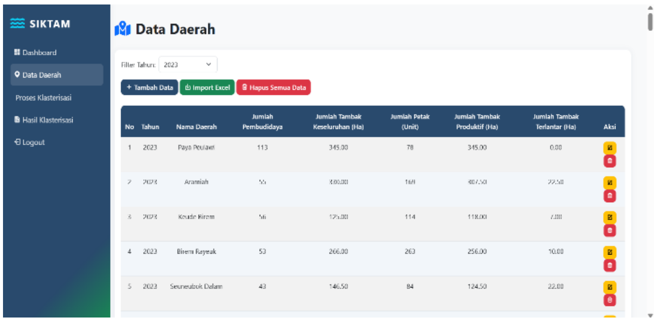

Figure 8. Regional Data Page

5) Data Normalization Page: All raw input data is first normalized using the Min-Max method. This page displays the normalization results, which serve as the basis for calculating centroids and distances between data points.

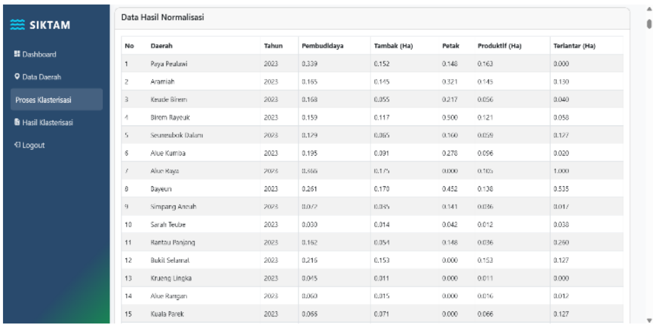

Figure 9. Data Normalization Page

6) Clustering Process Page: This page serves as the main interface for executing data processing within the system. Users can select two initial regions as the starting centroids to initiate the clustering process using the K-Means algorithm. Once selected, the system automatically carries out a sequence of calculations, including Min-Max normalization, Euclidean distance computation between data points, and iterative centroid updates. These steps are repeated until convergence is reached, meaning the centroids no longer change significantly. As such, this page is not merely a display of final results, but actively performs the complete clustering logic embedded within the system.

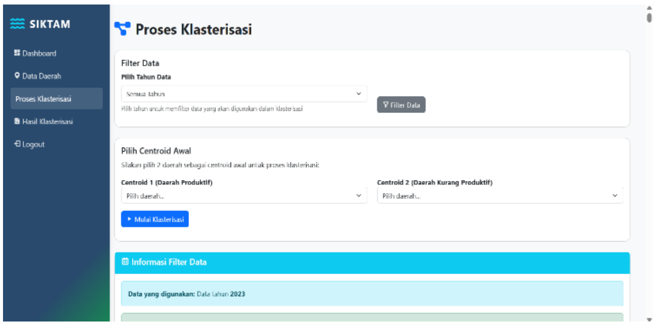

Figure 10. Clustering Process Page

7) *Clustering Result Page:* After the process is complete, the clustering results are displayed in the form of a table that contains details of the status of each region (productive or non-productive), the amount of data for each cluster, and the time of clustering.

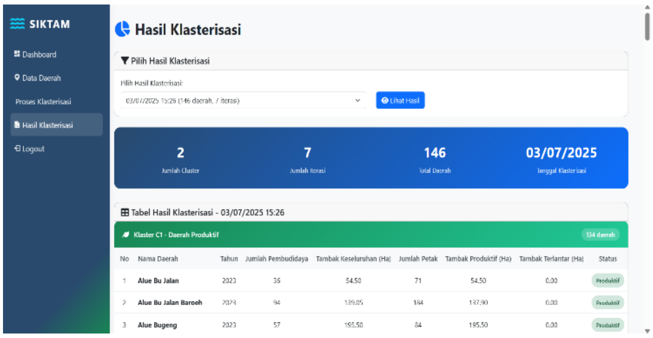

Figure 11. Clustering Results Page

IV. CONCLUSION

This study demonstrates that the application of the K-Means algorithm can be effectively used to cluster pond farming areas in East Aceh District based on their productivity levels. Using five main variables-number of farmers, total pond area, number of plots, productive pond area, and abandoned pond area-the system was able to classify 146 villages into two main clusters, namely productive and unproductive villages. The final results showed that 134 villages (92%) were classified as productive, while 12 villages (8%) were categorized as unproductive.

The implementation of a web-based clustering system using the PHP programming language allows the analysis process to be done automatically, quickly, and accurately. The system also facilitates visualization of clustering results, thus supporting more targeted decision-making. Validation of the results through manual comparison using Excel shows that the system produces consistent output and in accordance with algorithmic logic.

Although the clustering process numerically separates villages into two categories—productive and non-productive—further analysis was conducted to ensure that this classification also carries substantive and policy relevance. The two clusters demonstrate clearly distinguishable characteristics: productive villages tend to have a higher pond utilization rate, averaging around 75% of their total pond area, and involve an average of 70 active fish farmers per village. In contrast, non-productive villages typically exhibit lower levels of land utilization, with less than 30% of their pond areas being actively used, and face limitations in human resources for aquaculture activities. These disparities are not merely statistical in nature, but reflect deeper structural gaps that require targeted interventions. Therefore, the clustering results can be utilized by the Department of Marine Affairs and Fisheries to establish aquaculture zoning plans, identify priority villages for pond revitalization programs, and direct technical assistance and development budgets more effectively. In this way, the clustering model serves not only as a data classification tool but also as a practical foundation for evidence-based decision-making and spatially informed policy planning.

REFERENCES

[1] Rome, *The State of World Fisheries and Aquaculture 2022*. 2022. doi: 10.4060/cc0461en.

[2] T. Salsabila, N. Nurdin, and S. Retno, "Comparison of K-Medoids and K-Means Result for Regional Clustering of Capture Fisheries in Aceh Province," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 282–289, 2025, doi: 10.52088/ijesty.v5i2.829.

[3] N. Yulinar, M. Akmal, I. Zulkarnaen, and ..., "Implementasi Kebijakan Minapolitan di Kabupaten Aceh Timur," *J. Transparansi ...*, vol. 3, no. 2, pp. 74–79, 2023, [Online]. Available: https://ojs.unimal.ac.id/jtp/article/view/15192%0Ahttps://ojs.unimal.ac.id/jtp/article/download/15192/5967

[4] W. Zam and N. A. Kasim, "Faktor - Faktor yang Mempengaruhi Produktivitas Usaha Budidaya Udang Vaname (Litopenaeus vannamei) Di Kabupaten Pangkep Factors Affecting the Productivity of Vaname Shrimp (Litopenaeus vannamei) Culture in Pangkep District," vol. 13, no. 3, pp. 303–310, 2024.

[5] R. Rahmawati, M. Arif, R. Rahayu, and A. Akbardiansyah, "Persepsi Masyarakat Kabupaten Aceh Timur dalam Pengelolaan Ekosistem Mangrove Berkelanjutan Community Perceptions of East Aceh District in Sustainable Management of Mangrove Ecosystem," *J. Kelaut. dan Perikan. Indones.*, vol. 3, no. 1, pp. 45–58, 2023, [Online]. Available: http://jurnal.usk.ac.id/JKPIhttps://doi.org/10.24815/jkpi.v3i1.31709

[6] 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.

[7] & B. N. et al. Pria Wibawa Utama, Vincentius P. Siregar, "Klasifikasi Habitat Dasar Berbasis Objek Di Perairan Dangkal Karang Lebar Dan Pulau Lancang," vol. 15, no. August, pp. 167–184, 2023.

[8] R. K. Dinata, N. Hasdyna, S. Retno, and M. Nurfahmi, "K-means algorithm for clustering system of plant seeds specialization areas in east Aceh," *Ilk. J. Ilm.*, vol. 13, no. 3, pp. 235–243, 2021, doi: 10.33096/ilkom.v13i3.863.235-243.

[9] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.

[10] Nurdin, Bustami, Rini Meiyanti, Amalia Fahada, and Marleni, "Application of the K-Means Method for Clustering Capture Fisheries Products in North Aceh with A Data Mining Approach," *J. Adv. Zool.*, vol. 44, no. 4, pp. 39–49, 2023, doi: 10.17762/jaz.v44i4.1358.

[11] A. W. Fernando, Z. Fatah, I. F. Sains, and U. Ibrahimy, "Jurnal Ilmiah Multidisiplin Nusantara Implementasi Data Mining Dalam Prediksi Penjualan Sembako Menggunakan Metode Apriori Jurnal Ilmiah Multidisiplin Nusantara," vol. 2, no. November, pp. 124–130, 2024.

[12] N. Nur Afidah, "Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang," *Prism. Pros. Semin. Nas. Mat.*, vol. 6, pp. 729–738, 2023, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[13] V. Alvianatinova, I. Ali, N. Rahaningsih, and A. Bahtiar, "Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Penjualan Supermarket Berdasarkan Cabang (Branch)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1529–1535, 2024, doi: 10.36040/jati.v8i2.8993.

[14] 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.

[15] M. R. Kusnaidi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 330–338, 2022, doi: 10.47065/josyc.v3i4.2112.

[16] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 2020, doi: 10.1016/0377-0427(87)90125-7.