Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)
Vol. 6, No. 1, March 2025, pp. 764~768
ISSN: 2721-3838, DOI: 10.30596/jcositte.v6i1.22390

Clustering of Data Monitoring Water Quality Using Mean-Shift Clustering Method

Hafizh Al Kautsar Aidilof[1], Lidya Rosnita[2], Kurniawati[3], Muhammad Ikhwani[4]
[11,2,3] Department of Informatics, Malikussaleh University, Indonesia
[4] Department of Information System, Malikussaleh University, Indonesia

ABSTRACT

This study aims to cluster water quality data from Nile tilapia ponds using the Mean Shift Clustering method. The parameters used to analyze water quality include temperature, pH, turbidity, and salinity, which are cru-cial factors for the growth and health of Nile tilapia. The data used in this research consist of water quality measurements from several Nile tilapia ponds. The clustering process seeks to identify groups of data with simi-lar water quality characteristics, providing insights into optimal environmental conditions for tilapia farming. The clustering results reveal several distinct groups of water quality based on variations in temperature, pH, turbidity, and salinity. Results of the experiment show that a bandwidth value of 400 successfully identifies a relatively simple number of clusters, specifically four clusters. The Mean Shift Clustering method proves effec-tive in grouping data without requiring assumptions about data distribution and can detect clusters with arbi-trary shapes. Consequently, the findings of this study can be used to provide recommendations for improving water quality to enhance tilapia pond productivity.

Keyword : Mean Shift Clustering; Water Quality; Temperature; pH; Turbidity; Salinity.

Corresponding Author:
Hafizh Al Kautsar Aidilof,
Department of Informatics
Universitas Malikussaleh
Jl. Batam Kampus Bukit Indah Lhokseumawe, 24355/24353, Indonesia.
Email : hafizh@unimal.ac.id

Article history:
Received Dec 23, 2024
Revised Jan 15, 2025
Accepted Mar 10, 2025

1. INTRODUCTION

Water quality is one of crucial factors in supporting the health and growth of fish, especially in aquaculture such as Nile tilapia (Oreochromis niloticus) farming. Parameters of water quality, such as temperature, turbidity, pH, and salinity, play a vital role in creating an optimal environment for fish growth. Therefore, monitoring and managing water quality are essential for the success of aquaculture and maintaining the ecological balance of the ponds.

Clustering methods offer an effective approach in analyzing water quality data. These methods enable grouping data based on similar characteristics without requiring specific labels or categories. One of the advanced clustering methods is Mean Shift Clustering, which can automatically detect cluster centers without assuming a specific data distribution. This method works by updating the cluster center positions based on data distribution, enabling it to identify groups with varying density levels.

Several studies have demonstrated the advantages of Mean Shift Clustering. For instance, Chen (2018) showed that this method addresses the limitations of K-Means Clustering, such as its dependency on the initial centroid initialization, resulting in more consistent clusters. Other studies by Zhou (2017, 2019) highlighted the application of Mean Shift in ceramic image segmentation and signal analysis with high accuracy. Moreover, research by Peng (2016) and Fei Ma et al. (2006) indicated that clustering techniques like Mean Shift can optimize water quality monitoring by managing dynamic changes, conserving resources, and improving efficiency.

In the context of water quality monitoring, previous studies have used other clustering approaches, such as K-Means. Hanifah (2019) employed this method to monitor river water quality in Cimahi City through IoT-based data acquisition. Jerom (2020) developed an IoT-based water quality monitoring prototype equipped with sensors for temperature, pH, dissolved oxygen, and dissolved carbon dioxide, installed on floating boats to ensure accurate measurements.

Based on this background, this study aims to cluster water quality data from Nile tilapia ponds using the Mean Shift Clustering method. The data include temperature, turbidity, pH, and salinity parameters collected periodically from various pond locations. The clustering results are expected to provide

insights into water quality variations at each site, aiding in more effective water quality management and identifying patterns that influence the health and productivity of Nile tilapia. This research also aims to contribute to the development of more sustainable aquaculture technology.

2. Research Stages

2.1 Water Quality Data Acquisition

The research conducted by Wongmeekaew et al. (2019) utilized parameters such as temperature, pH, dissolved oxygen levels, and water level to monitor water quality in Nile tilapia ponds. In the study by Lidya et al. (2024), the parameters used were temperature, pH, water turbidity, and salinity levels. The sensors employed in their system included the DS18B20 sensor for temperature readings, the Turbidity sensor for measuring water turbidity, the SKU SEN0161 sensor for pH monitoring, and the Salinity sensor for salinity level measurement. All these sensors were connected to an Arduino Uno microcontroller, which transmitted data in real-time to a web-based platform using the Node MCU ESP32 module. The complete design of the data acquisition device is as follows:

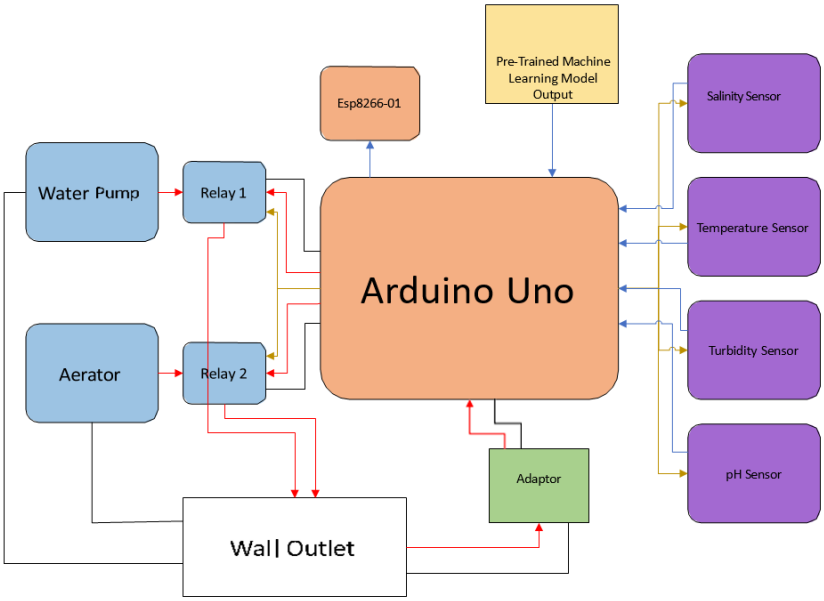

Figure 1. Framework Design of the Data Acquisition Device

This study utilized unlabeled data as the clustering objects, consisting of 115 data points collected over a 24-hour monitoring period with a data acquisition interval of every 10 minutes. Monitoring was conducted at a Nile tilapia pond owned by a local resident in Kuala Kerto, Lapang Subdistrict, North Aceh Regency.

2.2 The Implementation of Mean Shift Clustering

This research method uses the Mean Shift Clustering algorithm to perform water quality data clustering. Mean Shift Clustering is a representative method that seeks the cluster center in the data distribution by shifting the centroid toward data concentrations within a specified range (Baek, 2018). This process continues until the variance value is lower than the previous iteration, resulting in convergence to a local minimum value rather than a global minimum.

The Mean Shift Clustering algorithm has the advantage of automatically determining the number of clusters based on the bandwidth radius range. This bandwidth radius plays a crucial role in influencing the clustering results because it determines the data density threshold used to form clusters. Mean Shift groups data based on density in a high-dimensional space, resulting in clusters that align with the data characteristics (Zhou, 2019; Utari, 2024). Technically, the Mean Shift algorithm consists of two main tasks:

1. **Removing Exceptional Data – Data that does not meet consistency patterns are excluded from the clustering process.**
2. **Combining Normal Data** – Data with similar characteristics are grouped into a specific cluster based on unique points that are converged after a limited number of iterations (Chen, 2021; Cariou, 2022).

In this study, the Mean Shift Clustering algorithm is used to partition the water quality data set from the pond, assigning labels to each data point based on the convergence results. The data used includes parameters such as temperature, turbidity, pH, and salinity, which are collected periodically. The determination of the bandwidth radius and iterative processing are performed to generate optimal clusters that align with the observed data distribution patterns. The Mean Shift update equation is written as follows:

$$y_i^{(t+1)} = \frac{\sum_{k=1}^{N} f\left(x_k, y_i^{(t)}\right) \cdot x_k}{\sum_{j=1}^{N} f\left(x_j, y_i^{(t)}\right)} \qquad \textbf{(Equation 1, in the paper A Novel ...)}$$

Steps of the Mean Shift Clustering Algorithm (taken from Utari, 2024):

1. Determine the type of kernel to be used and the bandwidth width (h) that will control the range for neighbor search. Kernel function :

$$(k(x)) = \left(\frac{1}{h}\right) \varphi\left(\frac{x}{h}\right)$$

   where *h* is the bandwidth that influences the distance or radius from the cluster center when calculating the mean shift. The Gaussian or Epanechnikov (φ) functions allow the determination of weights for each data point based on its distance from the cluster center.
2. Initialize the initial mass centers that will serve as the centroids.
3. For each centroid, recalculate the new mass center using the Mean-Shift formula.

$$m(x) = \frac{\sum(k(\,x - x_i)x_i)}{\sum k(x - x_i)}$$

   The calculation of the new mass center ((x)) is based on the total of the kernel function ($(x-\,)$) divided by the total of the kernel function values ($K(x-xi\,)$) within the given kernel radius, where x is the current mass center and xi is a data point within the kernel radius.
3. Repeat the calculation of the new mass center for each centroid until the centroid no longer changes
4. or meets the stopping criteria that have been set, such as reaching the specified maximum number of iterations.
5. After the convergence process, each centroid will represent the mode or center of the identified data cluster.

3. RESULTS AND DISCUSSION

The results of the research show that 96 clusters were formed using a bandwidth value of 2 and 4 clusters were formed using a bandwidth value of 400. The following displays the number of clusters formed with various bandwidth value trials.:

Table 1. Number of Clusters Formed Based on Bandwidth

No.	Bandwith	Number of Clusters Formed
1	2	96
2	3	87
3	4	79
4	5	76
5	6	72
6	7	70
7	8	67
8	9	66
9	50	23
10	100	17
11	200	10

12	400	4

From the experiment results, it can be seen that the higher the specified bandwidth value, the fewer the number of clusters formed. This occurs because the bandwidth radius range becomes larger, allowing it to accommodate more scattered data into a single cluster. If assumed that only a few clusters are needed, then the most optimal bandwidth value in this study is 400. The membership of each data point to each cluster can be seen in the following table:

Table 2. The Result of Water Qaulity Data Clustering Process

No.	Formed Cluster	Data of Cluster Member
1	1	1 – 45, 50-54, 56-65, 67-69, 71, 72, 74, 75, 79-81, 84-89, 91, 92, 95, 97, 102, 104-108, 110, 111, 114, 115
2	2	73, 76, 78, 94, 96, 98, 100, 101, 113
3	3	66, 70, 77, 82, 83, 93, 99, 109, 112
4	4	46, 47, 48, 49, 55, 90, 103

The distribution of the members of each cluster can be seen in the following figure :

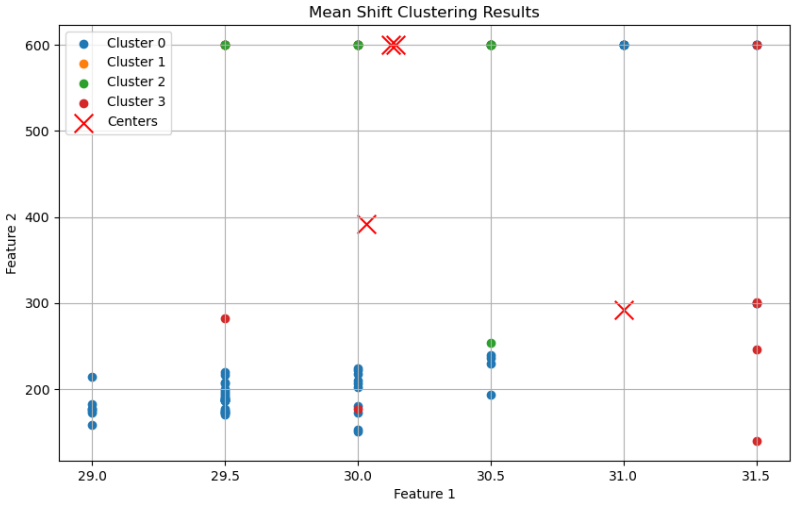

Figure 2. Distribution of Water Quality Data in Each Cluster

4. CONCLUSION

Global warming and awareness of the need to reduce greenhouse gas emissions have prompted greater attention to energy efficiency in the HVAC industry, where air conditioning (AC) systems play an important role in providing thermal comfort. However, selecting the right air conditioning system is often challenging due to the various factors that need to be considered, such as energy efficiency, operational costs, system reliability and environmental impact. To address this complexity, Complex Proportional Assessment (COPRAS) emerges as an effective multi-criteria analysis method in evaluating AC system alternatives by considering a number of criteria simultaneously. Although the potential of COPRAS has been demonstrated in various contexts, its application in determining AC system traffickers is still limited. Therefore, this study aims to explore the possible application of COPRAS in this context and identify key factors to consider in determining the optimal AC system trafficker. The evaluation results using the COPRAS method show that Medan Elektronik with a score of 100 and Citra Inovasi Prima with a score of 100 are the top choices in the selection of AC system traffickers that can be used as a reference in the selection. It is hoped that this research will contribute to the development of more sophisticated and applicable analysis methods in the HVAC industry and assist decision makers in making more informed and sustainable decisions regarding the selection of air conditioning systems.

REFERENCES

Baek, Jiyeon., Chung, Byungjin. & Yim, Changhoon. (2018). Linear Spectral Clustering with Mean Shift Filtering for Superpixel Segmentation. In *2018 Internation Conference on Electronics, Information, and Communication.*

Cariou, Claude., Le Moan, Steven. & Chehdi, Kacem. (2022). A Novel Mean Shift Algorithm for Data Clustering. HAL Open Scince, 10(1), 14575-14585.

Chakraborty, Saptarshi., Paul, Debolina. & Das, Swagatam. (2021). Automated Clustering of High-Dimensional Data with a Feature Weighted Mean Shift Algorithm. In 35th *AAAI Conference on Artificial Intelligence.*

Chen, Jingxue., Yang, Jingkang., Huang, Juan. & Liu, Yining. (2021). Robust Truth Discovery Cheme Based on Mean Shift Clustering Algorithm. *Journal of Internet Technology,* 22(4), 835-842.

Chen, Yang., Hu, PengFei. & Wang, Weilan. (2018). Improve K-Means Algorithm and Its Implementation based on Mean Shift. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineeering and Informatics.*

Hanifah, Hani Purwati. & Supangkat, Suhono Harso. (2019). IoT based River Water Quality Monitoring Design for Smart Environments in Cimahi City. In *2019 International Conference on Electrical Engineering and Informatics.*

Huang, Leijun., Feng, Hailin. & Le, Ying. (2019). Finding Water Quality Trend Patterns using Time Series Clustering: A Case Study. In *2019 IEEE 4th International Conference on Data Science in Cyberspace.*

Jerom, Ajith., R, Manimegalai. & V, Ilayaraja. (2020). An IoT based Smart Water Quality Monitoring System using Cloud. In *2020 International Conference on Emerging Trends in Information Technology and Engineering.*

Peng, Sen., Lian, Xiaofeng., Wang, Xiaoyi. & Xu, Jiping. (2015). Optimization of Water Quality Monitoring Section Based on Comprehensive Hierarchical Clustering. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery.*

Rosnita, Lidya., Ikhwani, Muhammad., Aidilof, Hafizh Al Kautsar. & Salamah. (2024). Water Quality Monitoring and Control System for Tilapia Cultivation based on Internet of Things. *International Journal of Engineering, Science, and Information Technology*, 4(4), 38-43.

Sadewo, H., Satria, Y. & Burhan, H. (2020). Application of Mean Shift Clustering to Optimize Amtching Problems in Ridesharing for Maximize the Total Number of Match. *Journal of Physics : Conference Series,* 1-7.

Utari, Roid Fitrah., Insani, Fitri., Agustian, Surya. & Afriyanti, Liza. (2024). Pengelompokan Data Pendistribusioan Listrik menggunakan Algoritma Mean Shift. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1015-1023.

Virupakshappa, Kushal & Oruklu, Erdal. (2019). Unsupervised Machine Learning for Ultrasonic Flaw Detection using Gaussian Mixture Modelling, K-Means Clustering and Mean Shift Clustering. In *2019 IEEE International Ultrasonics Symposium*.

Wongmeekawe, Tanomsak., Boonkirdram, Sarawoot. & Phimpisan, Songgrod. (2019). Wireless Sensor Network for Monitoring of Water Quality for Pond Tilapia. In *2019 Twelfth International Conference on Ubi-Media Computing.*

Zhou, Pengbo. & Wang, Kegang. (2017). Porcelain Image Classification based on Semi-Supervised Mean Shift Clustering. In *2017 8th IEEE International Conference on Software Engineering and Service Science.*

Zhou, Yi., Feng, Yi., Tarokh, Vahid., Gintautas, Vadas., Mc Clelland, Jessee. & Garagic, Denis. (2019). Multi-Level Mean Shift Clustering for Single Channel Radio Frequency Signal Separation. In *29th International Workshop on ML for Signal Processing.*