Original Article

Clustering Captured Fish Exports with Fuzzy C-Means

Ekka Pujo Ariesanto Akhmad[1]✉, Budi Priyono[2]
[1,2]Universitas Hang Tuah, Indonesia
Correspondence Author: eka.pujo@hangtuah.ac.id✉

Abstract:

By export destination country group, data on Indonesia's fresh/chilled captured fish exports are collated and tabulated. Nevertheless, export amounts have not been separated in the table of fresh/chilled fish exports. The clustering method will be used to tackle the fresh/chilled fish export data clustering problem. This study is to investigate the use of partition-based clustering, namely Fuzzy C-Means (FCM), to assess export data of caught fish from the primary destination nations using machine learning. The Buletin Statistik Perdagangan Luar Negeri Ekspor Menurut Kelompok Komoditi dan Negara, October 2023 is where the research's data came from. The research variables in this study, which comprise net weight of products (kg) and FOB (free on board) value, are based on data collected between 2022 and 2023. Following a cluster analysis, two clusters are produced: one for high export levels and another for medium or low export levels.

Keywords: Fresh/Chilled Captured Fish, Fuzzy C-Means, Machine Learning, Silhouette Coefficient.

Introduction

In order to fully comprehend the intricacy of a nation's trading environment, scholars and decision-makers must first cluster export data in order to identify the underlying dynamics and forces that influence a nation's export performance. For example, the Exporter Dynamics Database offers a thorough view of exporter characteristics and growth measures at various levels of aggregation, including country-year, country-product-year, and country-destination. It demonstrates that, in comparison to smaller or developing countries, larger or more developed countries have more diverse exporters and lower entry and exit rates (Cebeci et al., 2012; Olasehinde-Williams & Oshodi, 2021).

Several previous studies that used Fuzzy C-Means algorithms, namely clustering captured fish exports using the Fuzzy C-Means (FCM) algorithm can provide valuable insights into the categorization of fish species based on production data. FCM is a popular fuzzy clustering technique that assigns data points to clusters with varying degrees of membership, making it suitable for datasets with overlapping clusters and

https://jurnal.usk.ac.id/riwayat

Received: September 22, 2024 – Revised: October 19, 2024- Accepted: November 12, 2024 - Published online: January 15, 2025

inherent uncertainties (Kochuveettil & Mathew, 2022; Pratiwi & Saputro, 2020).

However, traditional FCM has limitations such as sensitivity to initial values and difficulty in handling nonlinear data and noise (Askari, 2021; Xi & Zhang, 2020). To address these issues, several enhancements have been proposed. For instance, the adaptive artificial-fish-swarm algorithm (AAFSA) improves FCM by enhancing its global and local optimization capabilities, thus reducing the number of iterations and improving clustering performance (Xi & Zhang, 2020).

Additionally, the Revised FCM (RFCM) algorithm employs adaptive exponential functions to mitigate the impact of noise and outliers, making it suitable for data with unequal cluster sizes and non-uniform mass distribution (Askari, 2021).

Furthermore, integrating topological information into FCM through persistence diagrams can capture the topological structure of data without prior knowledge, enhancing the clustering process in fields like materials science and machine learning (Davies et al., 2020).

Another approach involves the use of Kernel-based FCM (KFCM), which projects the feature space into a higher dimension to ensure linear separability of complex clusters, thereby optimizing clustering and improving the validity of clusters (Kochuveettil & Mathew, 2022).

In practical applications, such as clustering fish production data from PPN Karangantu, FCM has been compared with the Self-Organizing Map (SOM) algorithm. While SOM achieved optimal results with 200 iterations, FCM required only 88 iterations, demonstrating its efficiency in clustering fish species into low, medium, and high production groups (Fawaz, 2023).

For high-dimensional data, the Fractional Sail Fish-Sparse FCM combined with Particle Whale Optimization (FSF-Sparse FCM + PWO) within a MapReduce framework has been proposed. This method efficiently processes large datasets by computing cluster centroids in the mapper phase and optimizing them in the reducer phase, achieving high accuracy and low DB-Index (Kulkarni & Vadali, 2022).

The phases of the study process, which include gathering data, cleaning it up beforehand to remove unwanted qualities, clustering using the Fuzzy C-Means algorithm, testing with MATLAB, and assessing validity with the Davies Bouldin Index (DBI). 569 of the 894 data that are part of the good cluster and 325 of the unfavorable group are based on the clustering results that were obtained (Widiastuti et al., 2023).

These forecasts can be made utilizing information on fish productivity, water temperature, and quality using the fuzzy c-means (FCM) technique. Fuzzy c-means analysis is used to forecast fish production for the upcoming several years. Based on data on water quality and temperature, these projections offer information on the volume of fish production that will occur in the future. The study's findings show that the approach can produce precise forecasts that may be helpful for continuing fish production (Ngii et al., 2023).

Overall, these advancements in FCM and its variants demonstrate their applicability and effectiveness in clustering captured fish exports, offering improved accuracy, efficiency, and robustness in handling complex and noisy datasets.

This study attempts to categorize the captured fish export data using the Fuzzy C-Means clustering algorithm in order to have a better understanding of the export performance of Indonesian captured fish.

Methods

Planning, data collection, pre processing, fuzzy c-means algorithm, and evaluation phases comprised the research process. Figure 1 illustrates the research phases.

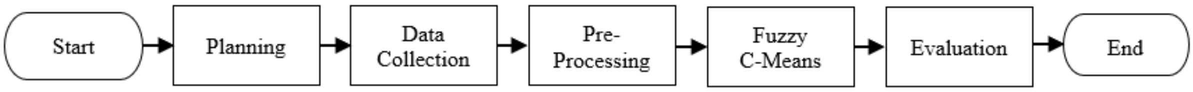

Figure 1. Research Phases

Planning

The scope of the system being designed and its requirements must be ascertained during the planning phase (Wulandari & Yogantara, 2022). In this work, a system for data clustering analysis was created. This clustering study groups together countries those having medium or low export levels and those with high export levels. The analysis's coverage includes 26 different countries worldwide, as well as weight in kilograms, year, and free on board (US dollars) data. The Python programming language and the Fuzzy C-Means techniques are used in this clustering research.

Data Collection

The BPS data for Indonesian fresh/chilled fish exports for the years 2022–2023 was used to collect the data. The four criteria that were considered were year, weight, free on board, and country.

Pre-ProStart by cleaning the data, which entails removing redundant information, looking for irrelevant information, and fixing any errors (Muchtar & Afiyati, 2024). To avoid data duplication, Utilize data unification to combine disparate data into a new database. After determining which data should be used and minimizing the amount of each, cluster the data while maintaining the original. Following selection, the data needs to be integrated or converted into a common format. Finally, use the Fuzzy C-Means clustering technique to complete the clustering process.

Fuzzy C-Means Algorithm

At this point, the Fuzzy C-Means technique is used to create a clustering model. As seen in Figure 2, the use of Fuzzy C-Means is done in multiple steps (Handayanto, 2020). The steps involved in using the Fuzzy C-Means algorithm are explained as follows:

1. Ascertain the ideal number of groupings.
2. Designates the value constraint at which the loop terminates as the minimal error value.
3. As the elements of the initial membership matrix, randomly initialize the membership matrix U = $\{uij\}$.
4. The following formula can be used to get the cluster center (V), where the matrix element V is vkj.

$$v_{kj} = \frac{\sum_{j=1}^{n}(\mu_{ik})^{m} X_{ij}}{\sum_{j=1}^{n}(\mu_{ik})^{m}} \qquad (1)$$

5. The following formula can be employed to determine the value of the goal function that was utilized to determine the error value.

$$\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^{2}\right]^{\mu_{ik}{}^{w}}\right) \qquad (2)$$

6. Use the following formula to determine the membership matrix's change.

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^{2}\ \right]^{\frac{-1}{w-1}}}{\left[\sum_{k=1}^{c}(X_{ij}-V_{kj})^{2}\ \right]^{\frac{-1}{m-1}}} \qquad (3)$$

7. Go back to step 4 and so forth if any prerequisites have not been met. Should the least amount of mistake have been Repetition of step 4 and subsequent steps should be done if specific conditions are not met. The process can be terminated and the final cluster result received if the minimum error value has been achieved.

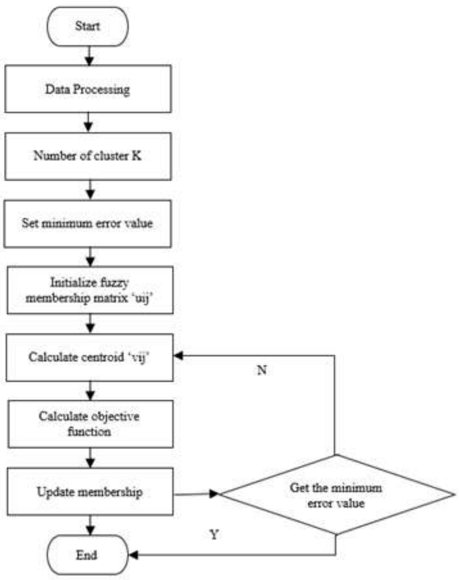

Figure 2. Stages of the Fuzzy C-Means Algorithm

Evaluation

Evaluate fuzzy c-means clustering results with the silhouette coefficient approach. As stated by (Handoyo et al., 2014), the coefficient calculation is performed using their method.

1. To find an object's average distance from every other object in a cluster, use the formula below.

$$a(i) = \frac{1}{[A]-1}\sum j\epsilon_{A,j\neq i}\, d(i,j) \qquad (4)$$

2. Using the equation to find the lowest value, calculate the average distance between that object and every other object in the other clusters.

$$d(i, C) = \frac{1}{[A]} \sum j \in C \; d(i, j) \qquad (5)$$

3. Using the provided equation, get the silhouette coefficient value (6).

$$s(i) = \frac{b[i]-a[i]}{Max(a[i],b[i])} \qquad (6)$$

   Where

   s(i) = silhouette at point i

   a(i) = the average difference between each point in the cluster and the i-th point

   b(i) = minimal distinction between other points in various clusters and the i-th point

4. The value of the silhouette coefficient ranges from -1 to 1. The better the data clustering in clusters, the closer the coefficient value is to 1.

Results

Data collection

The Central Bureau of Statistics' Buletin Statistik Perdagangan Luar Negeri Ekspor Menurut Kelompok Komoditi dan Negara provided the information on Indonesian captured fish exports. The five columns in the table of Indonesian captured fish exports are commodity/country of destination, FOB value (US dollars), FOB weight (kg), net weight change (%), and FOB value change (%).

Initial Processing

Doing data repair procedures is the first step in the starting work. In the gathered fish export data for January 2022 and 2023, the net weight and FOB value % change column is removed, along with the nations whose net weight and FOB value data for those same months are empty. These actions constitute the cleaning procedure. Data integration and data selection are methods to limit the quantity of data required to join or unify diverse data into a new database. The captured fish export database for January 2022 and January 2023 was divided into two distinct tables in order to work on data integration and sorting. The data is converted or normalized after selection.

Tables 1 and 2 present the results of the preprocessing of the exported fish data that was caught by destination country in January 2022 and January 2023.

Table 1. Outcomes of Captured Fish Export Data Pre Processing by Destination Nation, January 2022

Destination Country	BB2022	FOB2022
Japan	1.194.666	6.475.259
Hong Kong	803.016	2.691.740
Korea, Republic Of	77.340	159.161
Taiwan	1.056.386	3.822.883
China	950.644	3.315.086
Thailand	118.384	343.403
Singapore	12.129.076	28.465.644
Malaysia	25.872.491	42.068.563
Brunei Darussalam	139	5.894
Saudi Arabia	118.724	414.766
United Arab Emirates	410	1.942
Australia	174.615	1.926.780
United States	138.441	1.199.075
Canada	13.700	138.883
Guatemala	868	2.324
Aruba	3.006	7.304
Netherlands	8.670	30.345

Table 2. Outcomes of the Captured Fish Export DataPre Processing by Destination Country, January 2023

Destination Country	BB2023	FOB2023
Japan	1.281.682	8.285.991
Hong Kong	535.604	2.358.789
Korea, Republic Of	10.315	45.129
Taiwan	1.102.920	3.951.298
China	343.311	1.809.784
Thailand	7.365	56.703
Singapore	12.817.012	30.680.292
Philippines	13.385	79.345
Malaysia	21.555.060	31.896.371
Viet Nam	25.047	157.464
Saudi Arabia	201.191	788.335
Kuwait	1.709	28.461
United Arab Emirates	4.224	27.267
Qatar	129	1.189
Australia	135.698	1.539.574
East Timor	1.080	1.750
United States	325.514	3.157.133
Canada	26.679	344.855
United Kingdom	12.710	54.161
France	389	4.463
Germany, FED. REP. OF	2.659	30.504
Italy	1.922	30.746

Fuzzy C-Means Clustering Results

Table 3 displays the outcomes of data clustering using the Fuzzy C-Means technique, arranged according to the primary destination nations in 2022.

Table 3. Fuzzy c-means cluster results in captured fish exports in 2022

Destination Country	BB2022	FOB2022	Cluster
Japan	1.194.666	6.475.259	0
Hong Kong	803.016	2.691.740	0
Korea, Republic Of	77.340	159.161	0
Taiwan	1.056.386	3.822.883	0
China	950.644	3.315.086	0
Thailand	118.384	343.403	0
Singapore	12.129.076	28.465.644	1
Malaysia	25.872.491	42.068.563	1
Brunei Darussalam	139	5.894	0
Saudi Arabia	118.724	414.766	0
United Arab Emirates	410	1.942	0
Australia	174.615	1.926.780	0
United States	138.441	1.199.075	0
Canada	13.700	138.883	0
Guatemala	868	2.324	0
Aruba	3.006	7.304	0
Netherlands	8.670	30.345	0

Table 4. Fuzzy c-means cluster results in captured fish exports in 2023

Destination Country	BB2023	FOB2023	Cluster
Japan	1.281.682	8.285.991	1
Hong Kong	535.604	2.358.789	1
Korea, Republic Of	10.315	45.129	1
Taiwan	1.102.920	3.951.298	1
China	343.311	1.809.784	1
Thailand	7.365	56.703	1
Singapore	12.817.012	30.680.292	0
Philippines	13.385	79.345	1
Malaysia	21.555.060	31.896.371	0
Viet Nam	25.047	157.464	1
Saudi Arabia	201.191	788.335	1
Kuwait	1.709	28.461	1
United Arab Emirates	4.224	27.267	1
Qatar	129	1.189	1
Australia	135.698	1.539.574	1
East Timor	1.080	1.750	1
United States	325.514	3.157.133	1
Canada	26.679	344.855	1
United Kingdom	12.710	54.161	1
France	389	4.463	1
Germany, FED. REP. OF	2.659	30.504	1
Italy	1.922	30.746	1

Table 3 shows that fifteen countries have a medium or low export level cluster (0), whereas two countries (Malaysia and Singapore) have a high export level cluster (1). The cluster results for exported captured fish are shown in Table 4, with a breakdown by the main destination countries for 2023. Table 4 shows that twenty countries have a medium or low export level cluster (1), Conversely, Singapore and Malaysia are the only two nations with a high export level cluster (0).

Evaluation

The silhouette coefficient approach was used to assess the cluster output. There is a range of -1 to 1 for the silhouette coefficient. An algorithm's performance improves with a greater value. The silhouette coefficient results for captured fish exports in 2022 and 2023 are shown in Table 5, which is organized by the primary destination nations.

Table 5. The silhouette coefficient for exported caught fish in 2022 and 2023

Parameter Performance Test	2022	2023
Silhouette Coefficient	0.8904	0.9263

.

Conclusion

For the years 2022 and 2023, Indonesian-captured fish export data is clustered at the high, medium, and low export levels using the Fuzzy C-Means clustering technique. Using the Silhouette Coefficient (SC) value, the performance of the Fuzzy C-Means cluster captured fish export data was evaluated by primary destination nations in 2022 and 2023. The clustering is good, as evidenced by the SC values of 0.8904 and 0.9263 for the captured fish export data group in 2022 and 2023, respectively.

References

Askari, S. (2021). Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, *165*, 113856.

Azis, A., & Nurasiah, N. (2024). Development of History Problems Based on Higher Order Thinking Skills (HOTS) Using Anderson Krathwohl Taxonomy. *Riwayat: Educational Journal of History and Humanities*, *7*(1), 111-118.

Cebeci, T., Fernandes, A. M., Freund, C. L., & Pierola, M. D. (2012). Exporter dynamics database. *World Bank Policy Research Working Paper*, *6229*.

Davies, T., Aspinall, J., Wilder, B., & Tran-Thanh, L. (2020). Fuzzy c-means clustering for persistence diagrams. *ArXiv Preprint ArXiv:2006.02796*.

Fauzan, R., Harsono, K., Meisandy, R. P., Barokah, M., & Muhaimin, M. I. (2024). Optimising Human Resource Management as an Effort to Improve Employee Performance through Digital Attendance. *Riwayat: Educational Journal of History and Humanities*, *7*(1), 16-25.

Fawaz, F. (2023). *Perbandingan Algoritma Self Organizing Map dan Fuzzy C-Means dalam clustering hasil produksi ikan PPN Karangantu*. Universitas Pendidikan Indonesia.

Imran, M. F. (2024). Criminological Examination of Physical and Psychological Violence Committed Against Children in the School Environment. *Riwayat: Educational Journal of History and Humanities*, *7*(1), 41-47.

Kochuveettil, A. D., & Mathew, R. (2022). A novel approach to fuzzy c-Means clustering using kernel function. *Intelligent Decision Technologies*, *16*(4), 643–651.

Kulkarni, O., & Vadali, R. S. (2022). Big data clustering using fractional sail fish-sparse fuzzy C-means and particle whale optimization based MapReduce framework. *Web Intelligence*, *20*(3), 153–171.

Ngii, K., Sabrina, D., & Saputra, R. (2023). Implementasi Algoritma Fuzzy C-Means (FCM) Dalam Memprediksi Hasil Tangkapan Ikan Di Kota Kendari. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, *7*(2), 319–324.

Olasehinde-Williams, G., & Oshodi, A. F. (2021). Can Africa raise export competitiveness through economic complexity? Evidence from (non)-parametric panel techniques. *African Development Review*, *33*(3), 426–438.

Pratiwi, N. B. I., & Saputro, D. R. S. (2020). Fuzzy c-shells clustering algorithm. *Journal of Physics: Conference Series*, *1613*(1), 12006.

Widiastuti, N. A., Kusumodestoni, R. H., Wahono, B. B., & Chumaisaroh, D. A. (2023). Implementasi Algoritma Fuzzy C-Means dalam Mengelompokkan Produktivitas Nauplius Vanammei di BBPBAP Jepara. *JSINBIS (Jurnal Sistem Informasi Bisnis)*, *13*(1), 37–44.

Xi, L., & Zhang, F. (2020). An adaptive artificial-fish-swarm-inspired fuzzy C-means algorithm. *Neural Computing and Applications*, *32*(22), 16891–16899.