Journal Of Economics, Technology, and Business (JETBIS)
Volume 3, Number 5 May 2024
p-ISSN 2964-903X; e-ISSN 2962-9330

CLASSIFICATION OF INDONESIAN FROZEN SHRIMP EXPORT DATA USING K-MEDOIDS CLUSTERING

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

KEYWORDS:
machine learning; frozen shrimp export; k-medoids clustering

ABSTRACT

Indonesia is one of the countries exporting caught, cultivated, and frozen shrimp to both developed and developing countries. Previously, the frozen shrimp export data was categorized based on the countries it was exported to. However, the list of frozen shrimp exports did not include the export levels. This research aims to explore the application of machine learning to analyze the frozen shrimp export data from the main destination countries using the K-Medoids cluster. The data for this research was sourced from the Bulletin of Foreign Trade Export Statistics documents by Product Group and Country, January 2023. The data used in this study covers the period from 2022 to 2023 and includes research variables such as net weight of goods (in tons) and FOB value (free on board). The cluster analysis revealed the formation of two clusters: one with high export levels and another with medium and low export levels.

INTRODUCTION

Wild-caught shrimp, cultivated shrimp products, and frozen shrimp. According to BPS data (BPS, 2021), frozen shrimp have a higher export volume compared to shrimp in other forms. The Indonesian frozen shrimp export table is arranged based on export destination country groups and presented in tabular form. The grouping of export data is a fundamental step in understanding the complexities of a nation's trade landscape, as it enables researchers and policymakers to uncover the underlying drivers and dynamics that shape a country's export performance. (Xingyuan, 2020). One of the primary challenges in the grouping of export data is the inherent diversity of goods and services that are traded globally. (Lima et al., 2016) Countries can often exhibit comparative advantages in a wide range of products, making it difficult to identify clear-cut patterns or clusters in their export baskets.

K-Medoids is a partition-based clustering method used in this research. The K-Medoids algorithm has a simpler approach compared to K-Means. In K-Medoids, the centroid is chosen randomly but with a value from one of the dataset rows, instead of searching for or determining a new value. The centroid in this algorithm is referred to as a medoid (Saputra and Dinar, 2022; Rahman, et al., 2020). To better understand the export performance of Indonesian frozen shrimp, this study aims to classify the frozen shrimp export data using the K-Medoids clustering algorithm.

Previous studies have employed partition-based clustering methods. Tampola and Assagaf (2018) conducted a mapping of shrimp potential per province in Indonesia using the K-Means method, validating the created groupings with the Davies-Bouldin index. Erawati (2018) analyzed quality and food safety in handling frozen tiger shrimp in Indonesia. While

focusing on a different aspect of shrimp, this reference underscores the importance of quality assessment in the shrimp industry, which is crucial when considering the classification of frozen shrimp export data. Dimantara & Elida (2020) analyzed the competitiveness of Indonesian frozen shrimp exports to the United States, highlighting factors affecting export performance. They found that Indonesia's market share was the third highest after India and Thailand, with factors like domestic production and export prices significantly influencing exports. Maulana et al. (2021) used a benchmark percentage of the various types of fish products in all provinces in Indonesia. Kapita et al. (2022) also reviewed the same topic, grouping Indonesia's shrimp production capacity using the Kohonen SOM technique to assess fisheries capacity, specifically shrimp products in each province. Yolandika, et al. (2022), focused on the competitiveness of Indonesia's frozen shrimp exports in the international market. Their study aimed to analyze Indonesia's position among competing countries in the global frozen shrimp export market, shedding light on the level of competitiveness. Rindayati & Akbar (2022) discuss the competitiveness and determinants of Indonesian frozen shrimp exports to non-traditional markets, utilizing methods like Revealed Comparative Advantage (RCA) and panel data regression. Another study by Tahalea (2024), delved into clustering shrimp distribution in Indonesia using the X-Means clustering algorithm. Although not directly related to K-Medoids clustering, this reference provides insights into clustering methodologies applied to shrimp-related data, which can be relevant for the classification of Indonesian frozen shrimp export data

RESEARCH METHODS

Research Flow

This study includes a research flowchart, which is shown in Figure 1.

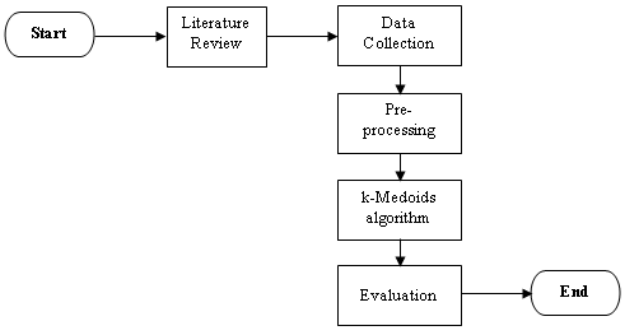

Figure 1
Research Flow

Study of Literature

To serve as a research reference, researchers conducted a literature review of previous studies that explored the research topic. The literature analysis drew from academic books and machine-learning journals that focused on clustering.

Data Collection

The collection of data was conducted by utilizing the export data of Indonesian frozen shrimp acquired from BPS for the years 2022-2023. The four attributes employed were country, year, weight in tons, and free on board (US dollars).

Pre-processing

First, perform a data cleaning process that involves deleting duplicate data, checking for irrelevant data, and correcting data errors. Implement data unification to merge separate data into a new database to prevent data duplication. Next, select and minimize the amount of data used, then proceed with the clustering process while preserving the original data. After selecting the data, the subsequent step is to transform or combine it into a standardized format. Lastly, perform the clustering process using the K-Medoids clustering method.

K-Medoids Algorithm

Different from k-means, which only require calculating the average distance, k-medoids typically utilize the Partitioning Around Medoids (PAM) algorithm, which is essentially similar to the k-means algorithm. The steps of PAM are as follows: First, choose a number of representative objects. Second, evaluate whether substituting a representative object with a non-representative object would enhance the clustering quality. This evaluation is performed for all potential non-representative objects. Third, repeat the second step until convergence is achieved, indicating that the clustering quality cannot be further improved. The clustering quality is assessed using a cost function in the form of the average dissimilarity between a non-representative object and a representative object in its cluster.

Evaluation

Evaluation of k-medoid clustering results using the silhouette coefficient method. The coefficient calculation follows the method outlined by M. Fornalski, et al. (2014) as follows.

a. Use the equation below to calculate the average distance of an object from all other objects in a cluster.

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

b. Calculate the average distance between that object and all the other objects in the other clusters, and then use that equation to determine the minimum value.

$$\boldsymbol{d(i,C) = \frac{1}{[A]} \sum j \in C\; d(i,j)} \quad (2)$$

c. Calculate the silhouette coefficient value using the given equation. (3) [7].

$$\boldsymbol{s(i) = \frac{b[i]-a[i]}{Max(a[i],b[i])}} \quad (3)$$

Where

s(i) =silhouette at point i

a(i)= average dissimilarity between the ith point and all other points in the cluster

b(i)= minimum difference between the ith point and other points in different clusters

d. The silhouette coefficient value is in the range of -1 to 1. The higher the coefficient value is close to 1, the better the cluster data grouping.

RESULTS AND DISCUSSION

Data collection

Data on Indonesian frozen shrimp exports were obtained from the Statistical Bulletin on

Foreign Trade Exports by Commodity Group and Country published by the Central Bureau of Statistics. The table of Indonesian frozen shrimp exports contains five columns, namely commodity/country of destination, net weight (kg), net weight change (%), FOB value (US dollars), and FOB value change (%).

Initial Processing (Preprocessing)

The first stage in the initial work is to carry out data repair operations. The cleaning process is carried out by deleting the percentage change column in net weight and FOB value in frozen shrimp export data for January 2022 and 2023, as well as deleting countries that have empty data for net weight and FOB value for January 2022 and 2023. To combine or unify different data into a New database, data integration and data selection are used, which are ways to reduce the amount of data required. To work on data integration and sorting, the frozen shrimp export table from January 2022 and January 2023 was split into two separate tables. After selection, the data is transformed or normalized. Table 1 and Table 2 show the results of the initial processing of Frozen Shrimp Export data by destination country in January 2022 and 2023.

Table 1
Results of Initial Processing of Frozen Shrimp Export Data by Destination Country January 2022

No	Country of Destination	Net Weight (Kg)	FOB Value (US $)
1.	JAPAN	2.130.162	23.575.900
2.	HONGKONG	60.928	589.623
3.	KOREA, REPUBLIC OF	66.276	525.925
4.	TAIWAN	139.706	1.386.822
5.	CHINA	552.292	2.751.421
6.	SINGAPURE	73.586	416.829
7.	MALAYSIA	112.939	492.107
8.	CYPRUS	1.064	11.438
9.	AUSTRALIA	6.360	83.099
10.	EAST TIMOR	2.000	2.000
11.	UNITED STATES	12.440.874	117.737.407
12.	CANADA	258.151	2.661.737
13.	PUERTO RICO	72.204	771.304
14.	DOMINICAN REPUBLIC	3.629	37.003
15.	UNITED KINGDOM	49.147	453.674
16	NETHERLANDS	228.005	2.306.043
17.	FRANCE	24.734	44.385
18.	GERMANY, FED, REP	42.586	447.995
19.	BELGIUM	99.095	921.897
20.	ITALY	3.600	24.600
21.	GREECE	2.000	16.500
22.	RUSSIA FEDERATION	40.050	372.260

Table 2
Initial Processing Results of Frozen Shrimp Export Data by Destination Country January 2023

No	Country of Destination	Net Weight (Kg)	FOB Value (US $)
1.	JAPAN	1.917.059	18.625/019
2.	HONGKONG	13.256	126.429
3.	KOREA, REPUBLIC OF	56.929	584.532
4.	TAIWAN	138.192	1.266.519

5.	CHINA	1.367.770	6.618.933
6.	THAILAND	129.972	1.179.785
7.	SINGAPURE	49.089	333.883
8.	MALAYSIA	112.959	699.826
9.	VIETNAM	50.220	405.453
10.	MOROCCO	16.938	36.197
11.	AUSTRALIA	14.988	199.361
12.	UNITED STATES	7.509.788	58.971.768
13.	CANADA	71.513	654.479
14.	TRINIDAD AND TOBAGO	17.207	133.979
15.	PUERTO RICO	77.465	636.778
16.	UNITED KINGDOM	28.225	261.954
17.	NETHERLANDS	39.228	451.227
18.	FRANCE	70.930	688.651
19.	GERMANY, FED, REF, OF	44.914	477.848
20.	BELGIUM	88.094	595.050
21.	GREECE	19.132	229.705
22.	GREECE	7.500	57.375
23.	RUSSIA FEDERATION	115.550	982.687

Python k-Medoids calculation

Python k-Medoids calculations were carried out on frozen shrimp export data, according to the main destination countries for 2022 and 2023. The results are as follows.

a. import pandas as pd is commonly used to check, read, change dimensions, and create tables.
b. import numpy as an np function that enables arithmetic operations on numeric data types such as addition, multiplication, subtraction, exponentiality, and other mathematical operations.
c. import Seaborn as SNS, Seaborn is a library to brings up data visualization.
d. import matplotlib. pyplot as plt, call library matplotlib to createplots.
e. from sklearn_extra.cluster import KMedoids, calls the KMedoids algorithm built into Sklearn.
f. Read the dataset that has been provided in xlsx format. Save the dataset in the same folder as the project folder you are working on.
g. Head() is used to display the top 5 data.
h. shrimp22.info ( ), to get information about the dataset features, the data type used, the number of tuples for each feature, and the memory used.
i. shrimp22_x = shrimp22.iloc[:, 2:4], to select which feature locations will be processed. In this case, the net weight features Jan 22 and fob Jan 22. [:, 2:4] means we will take features starting from the number 2 to the number before 4.
j. Head() is used to display the top 5 data.
k. plt. scatter(shrimp22.BBJAN22, shrimp22.FOBJAN22, s=100,c= "c",marker= "o",alpha= 1) to display the visualization in scatter form and call the features that will be used as the x-axis and y-axis.
l. plt show( ) displays a scatter plot visualization.
m. converts net weight and fob feature frame data into an array. Functions to make it easier to normalize data in subsequent code.

n. print (x_array), displays the resulting array.
o. kMedoids = KMedoids (random_state = 0, n_clusters = 2), used to randomly select two clusters.
p. Print(KMedoids.cluster_centers_), to print the cluster center value.
q. shrimp22["cluster"] = KMedoids.labels_ to add a new variable named 'cluster'.
r. shrimp22.head(22) is used to display 22 data. If the number 22 is omitted, it will only output the top 5 data.

The cluster results for k-medoids frozen shrimp exports, categorized by the main destination countries in 2022, can be observed in Table 3.

Table 3
K-medoids cluster results in frozen shrimp exports in 2022

Country of Destination	BBJAN22	FOBJAN22	Cluster
JAPAN	2130162	23575900	0
HONG KONG	60928	589623	1
KOREA, REPUBLIC OF	66276	525925	1
TAIWAN	139706	1386822	1
CHINA	552292	2751421	0
SINGAPORE	73586	416829	1
MALAYSIA	112939	492107	1
CYPRUS	1064	11438	1
AUSTRALIA	6360	83099	1
EAST TIMOR	2000	2000	1
UNITED STATES	1,2E+07	1,18E+08	0
CANADA	258151	2661737	0
PUERTO RICO	72204	771304	1
DOMINICAN REPUBLIC	3629	37003	1
UNITED KINGDOM	49137	453674	1
NETHERLANDS	228055	2306043	0
FRANCE	24734	444385	0
GERMANY, FED, REP, OF	42596	447995	1
BELGIUM	99095	921897	1
ITALY	3600	24600	1
GREECE	2000	16500	1
RUSSIA FEDERATION	40050	372260	1

Table 3 shows that there are five countries with a high export level cluster (0) and seventeen countries with a low export level cluster (1).

Cluster results for frozen shrimp exports, categorized by the main destination countries in 2023, can be seen in Table 4.

Table 4
K-medoids cluster results in frozen shrimp exports in 2023

Country of Destination	BBJAN22	FOBJAN22	Cluster
JAPAN	1917059	18625019	1
HONG KONG	13256	126429	0
KOREA, REPUBLIC OF	56929	584532	0
TAIWAN	138192	1266519	1
CHINA	1367770	6618933	1

THAILAND	129972	1179785	1
SINGAPORE	49089	333883	0
MALAYSIA	112959	699826	1
VIETNAM	50220	405453	0
MOROCCO	16938	36197	0
AUSTRALIA	14988	199361	0
UNITED STATES	7509788	58971768	1
CANADA	71513	654479	1
TRINIDAD AND TOBAGO	17207	133979	0
PUERTO RICO	77465	636778	1
UNITED KINGDOM	28225	261954	0
NETHERLANDS	39228	451227	0
FRANCE	70730	688651	1
GERMANY, FED, REP, OF	44914	477848	0
BELGIUM	88094	595050	0
SPAIN	19132	229705	0
GREECE	7500	57375	0
RUSSIA FEDERATION	115550	982687	1

Table 4 shows that there are ten countries with a high export level cluster (1) and thirteen countries with a low export level cluster (0).

Evaluation

The cluster output was evaluated using the silhouette coefficient method. Table 5 presents the results of the silhouette coefficient for frozen shrimp exports, categorized by the main destination countries in 2022.

Table 5
Silhouette coefficient for frozen shrimp exports in 2022

a (i)	b (i)	s (i)
39490255,93	2,3E+07	-0,41105047
355845,8936	2,9E+07	0,987888348
321607,9649	2,9E+07	0,989076922
1041001,566	2,9E+07	0,963577579
34335380,7	1E+07	-0,69803103
297784,3423	3E+07	0,989922832
319604,9114	2,9E+07	0,989155429
429497,3788	3E+07	0,98566481
392646,7633	3E+07	0,986863312
438193,2356	3E+07	0,985379088
110529415,3	1,3E+09	0,915347581
34345249,81	2,6E+07	-0,23951933
480159,5337	2,9E+07	0,983555639
412748,3138	3E+07	0,986212026
295959,1875	3E+07	0,989973151
34584559,22	2,2E+07	-0,37077669
298357,622	3E+07	0,989896074
295260,0518	3E+07	0,989999013
603687,9698	2,9e+07	0,9792165112
41962,0752	3E+07	0,985986498
425320,9546	3E+07	0,985801773
302821,537	3E+537	0,989769022
Nilai Silhouette		0,72517764

The results for the mark silhouette coefficient for frozen shrimp exports according to the main destination countries in 2023 can be found in Table 6.

Table 6
Silhouette coefficient for frozen shrimp exports in 2023

a (i)	b (i)	s (i)
19740577,84	18422024,89	-0,066794942
215522,5326	8981135,778	0,976002753
313714,8202	8522292,438	0,963188528
9320517,213	972679,2402	-0,895641066
11768754,35	6459662,647	-0,451202527
9301088,817	885570,9085	-0,904788469
180188,2804	8771011,872	0,979459386
9368198,788	408664,7255	-0,956377449
197741,8575	8999873617	0,977270721
286291,737	9070213,251	0,968436052
184065,5407	8908952,322	0,979335198
55936694,42	59146621,5	0,054270675
9394237,771	357399,089	-0,961955499
210901,8182	8973151,43	0,976496349
9408652,958	340550,6389	-0,963804528
173535,9897	8844890,294	0,980380086
216506,947	8656160,082	0,974988107
9373516,664	391237,0383	-0,958261445
232058,4169	8629015,439	0,973107196
32717,4509	8509322,613	0,961529732
176415,0448	8878044,273	0,980129065
267579,0438	9050328,126	0,970434327
9300906,941	688128,9971	-0,92601485
Nilai Silhouette		0,244790665
		0,755209335

CONCLUSION

The high and low export level clusters for Indonesian frozen shrimp export data in 2022 and 2023 are generated from the application of the K-Medoids clustering algorithm.

Performance assessment of K-medoids cluster frozen shrimp export data by main destination countries in 2022 and 2023 using the Silhouette Coefficient (SC) value. SC value for the frozen shrimp export data group in 2022 is 0.725 and SC value for the frozen shrimp export data group in 2023 is 0.755, indicating that the clustering is also quite good.

BIBLIOGRAPHY

BPS, 2021, Bulletin of Foreign Trade Export Statistics by Commodity Group and Country, Jakarta.

Dimantara, R. W. and Elida, S. 2020. "Analisis daya saing ekspor udang beku indonesia di pasar amerika serikat". Dinamika Pertanian, 36(1), 79-90. https://doi.org/10.25299/dp.2020.vol36(1).5373

Erawati, N. M. R. 2018. "The analysis of quality and food safety on frozen tiger shrimp (penaeus monodon) handling in central Java, Indonesia". Russian Journal of Agricultural and Socio-Economic Sciences, 77(5), 288-296. https://doi.org/10.18551/rjoas.2018-05.34

Saputra, I. and Dinar AK, 2022, Machine Learning for Beginners, Informatics Publishers, Bandung.

Tempola, F. and Assagaf, 2018, "Clustering of Potency of Shrimp in Indonesia with K-Means Algorithm and Validation of Davies-Bouldin Index". AHE, volume 1, ICST.

Capita, S., A. Mubarak, S. Do. Abdullah, and M. Fhadli, 2022, "Application of the Kohonen-Som Clustering Algorithm with Validation of the Davies Bouldin Index in Grouping Shrimp Potential in Indonesia". Indonesian Journal on Information Systems (IJIS), Volume 7 Number 2.

Lima, J.E.D., Alvarez, M., Cracau, D., 2016, "Manual on Foreign Trade and Trade Policy: Basics, Classification and Indicators of Trade Patterns and Trade Dynamics", Documentos de Proyectos.

Maulana, A., KN Akbar, and Nurahman, 2021, "Application of Clustering Using the K-Means Algorithm as an Analysis of Provincial Fishery Commodity Production in Indonesia." EJECTS: E-Journal Computer, Technology and Informations Systems, Vol.01, No.01, September 2021 LPPM Darwan Ali University.

M. Fornalski, et al., 2014, "Comparison of Clustering Using Single Linkage and K-Means Document Grouping". Int. J. Comput. Appl. Technol., Vol. 7, no. 2.

Kaufman and Rousseeuw, 2005, Finding Groups in Data: an Introduction to Cluster Analysis. Hoboken, NJ, JW & Sons, Inc.

Rahman, F., Ridho, I. I., Muflih, M., Pratama, S., 2020, "Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the ", IOP conference series. Materials science and engineering.

Rindayati, W. and Akbar, R. 2022. "Competitiveness and determinants of Indonesian frozen shrimp exports to non-traditional markets." Jurnal Manajemen dan Agribisnis. https://doi.org/10.17358/jma.19.3.367

Tahalea, S., Fadhilah, R., Matdoan, M. Y., & Safira, D. A. 2024. "Clustering shrimp distribution in Indonesia using the x-means clustering algorithm." VARIANCE: Journal of Statistics and Its Applications, 6(1), 49-54. https://doi.org/10.30598/variancevol6iss1page49-54

Xingyuan, Natasha, 2020, "Intelligent Export Diversification: An Export Recommendation System with Machine Learning". Social Science Research Network.

Yolandika, C., Sofyani, T., Mursyid, H., Anggraini, N., & Berliana, D. 2022. "The competitiveness of Indonesia's frozen shrimp exports in the international market". IOP Conference Series: Earth and Environmental Science, 1118(1), 012074. https://doi.org/10.1088/1755-1315/1118/1/012074