JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 1, January 2025, pp. 104-118

XportID: Website for Clustering Indonesian Export Commodities by Destination Continent using Gaussian Mixture Model

Angela Lisanthoni[1], Trimono[1], Dwi Arman Prasetya[1]
[1]Data Science, University of Pembangunan Nasional "Veteran" East Java, Indonesia
trimono.stat@upnjatim.ac.id

Article History:
Received : 24-10-2024
Revised : 19-12-2024
Accepted : 21-12-2024
Online : 02-01-2025

Keywords:
Gaussian Mixture Model;
Silhouette Score;
Export;
Clustering.

ABSTRACT

Exports play a crucial role in driving economic growth and increasing foreign exchange reserves. However, Indonesia's export performance has not yet reached its optimal potential, as evidenced by an 11% decline in export value in 2023. The decrease is partly attributed to the limited range of export destination markets. Therefore, this study aims to analyze export trade patterns to identify the most ideal and potential market locations. The research will employ a quantitative approach, using secondary data from the Central Bureau of Statistics and the 2022 BACI dataset, focusing on the top 5 HS2 commodity types by highest export quantity. Clustering analysis is applied to group markets based on similar characteristics, identifying countries with high, medium, and low export potential for Indonesia's export strategy. The research develops a website-based clustering system called XportID, utilizing a Gaussian Mixture Model (GMM) with the Expectation-Maximization (EM) algorithm to determine optimal cluster parameters. GMM is preferred for its flexibility and probabilistic system, providing more accurate results compared to distance-based methods. There will be 3, 4, and 5 clusters formed and then the best cluster will be selected by comparing the silhouette score obtained. Results show that the Asian continent has 5 clusters with the best value of 0.7035, the American continent has 3 clusters with the best value of 0.8165, the African continent has 3 clusters with the best value of 0.8534, the Australian continent has 3 clusters with the best value of 0.8540, and the European continent has 4 clusters with the best value of 0.8654. Overall results, the clustering system is categorized as strong structure with average value of 0.8185. Countries with high export potential include Malaysia, Philippines, South Korea, Brazil, Mexico, New Zealand, and Spain. Specifically in Africa, commodities related to HS2-15 show potential for growth.

https://doi.org/10.31764/jtam.v9i1.27500

A. INTRODUCTION

Indonesia is an active country in international trade. This activity opens opportunities for business units to sell their products to other countries (Matondang et al., 2024). Based on research Ji et al. (2022), international trade (export-import) can contribute to the improvement of a country's economy in the long term. An increase in exports can provide an increase in foreign exchange reserves, this should also apply reversely (Andriyani et al., 2020). When the value of exports exceeds imports, there is a surplus that has a positive impact on the growth of the FOB value and the country's economy (Syahrani et al., 2022).

In an effort to strengthen economic growth, the value of Indonesia's exports needs to be increased (Sani Akbar & Rezeki, 2021). However, the condition that occurs is that the value of Indonesia's exports has not yet reached an optimal level. According to data from the Central Bureau of Statistics (Central Bureau of Statistics, 2024), the value of Indonesia's exports in 2023 fell to 258,797.2 million USD, a decrease of 11%. One of the causes is limited range of export destination markets, therefore the strategy that can be implemented to increase exports is by diversifying export markets (Charles & Rangen Jaya, 2023). However, there is no effective method to determine potential export commodity market. For this reason, this study will analyze the trade patterns that occur to identify the most ideal market location. Clustering can be used as an effective method for determining market location because it allows grouping data into categories based on similar characteristics (Solikhun et al., 2022; Andi Purnomo et al., 2020; Fernando & Fianty, 2024).

Several previous studies have examined clustering methods for export commodities, including Islami & Sihombing (2021) clustered export commodities based on destination continents for 182 countries and 15 HS2 commodities in 2017 using K-medians algorithm. The best silhouette score for the Asian continent was 0.6 for 4 clusters, the African continent was 0.88 for 3 clusters, the American continent was 0.71 for 3 clusters, and the European continent was 0.36 with 3 clusters. However, this study is lack of analysis related to the Australian continent. Akhmad & Priyono (2024) clustered frozen shrimp export products with Central Bureau of Statistics data from 2022 to 2023 using K-Medoids. The results obtained two clusters and evaluated using a silhouette score of 0.725 for 2022 data and 0.755 for 2023 data. Annisa Octaria et al. (2023) clustered crumb rubber exports using k-Means algorithm. The results obtained two clusters and evaluated using a silhouette score of 0.732. Both studies used distance-based clustering algorithms, which have limitations. To improve silhouette scores, more advanced algorithms are necessary. There are other studies such as clustering pretroleum materials export using K-Medoids algorithm (Rahman et al. 2020), and clustering of export products using the K-means and K-Medoids algorithms (Ulvi & Ikhsan, 2024).

K-means is one of the most commonly used methods, but this algorithm has the disadvantage that it is less effective in clustering data on a large scale (Ahmed et al., 2020), and ineffective in clustering data that has a non-convex structure (Adiwidyatma et al., 2024). Therefore, this study will use Gaussian Mixture Model (GMM) that offers greater flexibility by being able to cope with elliptical cluster-shaped data patterns as well as, being able to capture the complexity of data patterns because it considers the probability of data into several groups (Santoso & Haw, 2023). Based on the research by Patel & Kushwaha (2020), GMM can identify complex non-linear patterns and can manage overlapping clusters so it is said to be better than K-means.

This research utilizes GMM-based clustering to analyze Indonesian export commodities by destination continent, presented through the XportID website. The novelty lies in applying GMM, which has limited prior use in export clustering. This study addresses a research gap by exploring data from unrepresented countries across Asia, America, Africa, Australia, and Europe, aiming to uncover potential trade patterns. Using GMM, this study provides more accurate results through a probability-based approach rather than distance-based methods. The findings offer valuable insights for the Indonesian government and businesses to enhance export strategies, target marketing, and identify high-potential trade partnerships.

B. METHODS

1. Theoritical Framework

a. Gaussian Mixture Model (GMM)

GMM has parameters that use the EM algorithm that works iteratively, with the aim of grouping data based on similar characteristics (Faidah et al., 2024). GMM has several stages as follows (Patel & Kushwaha, 2020):

1) GMM for the multivariate case determines clusters with three parameters: mean (μ), covariance (Σ), and weights (w). So, the first step is initialize the number k of clusters to be used. Then, initialize each GMM parameter such as weight (w_j), average (μ_j) which has vector dimension (d), and covariance (Σ_j) for j = 1, 2, ..., k. GMM has the following formula:

$$p(x) = \sum_{j=1}^{k} w_j P\left(\boldsymbol{X} \middle| \mu_j, \Sigma_j\right) \tag{1}$$

Where, k is Number of clusters; $\boldsymbol{X}$ is data used; $P\left(X \middle| \mu_j, \Sigma_j\right)$ is probability that data **x** belongs to a cluster; and $p(x)$ is joint probability function of the data **x** over all Gaussian distributions.

2) The EM algorithm is implemented to optimize the estimation of the three GMM parameters. The EM algorithm has the following process:

a) E-step: focuses on calculating the probability of how the data ($\boldsymbol{x}_i$) is distributed in the Gaussian component (C_j) based on the current parameters. Then calculate the probability of the data $\boldsymbol{x}_i$ belongs to a cluster C_j using:

$$Z_{ij} = \frac{\rho\left(\boldsymbol{x}_i \middle| C_j\right).w_j}{\rho(x_i)} \tag{2}$$

Where, $\rho\left(\boldsymbol{x}_i \middle| C_j\right)$ is probability $\boldsymbol{x}_i$ generated by the Gaussian distribution of the cluster C_j; w_j is weight of the cluster C_j. Then, calculate the value of the likelihood $\rho\left(\boldsymbol{x}_i \middle| C_j\right)$ and total probability $\rho(\boldsymbol{x}_i)$:

$$\rho\left(x_i \middle| C_j\right) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_j|^{\frac{1}{2}}} \, exp\left(-\frac{1}{2}\left(\boldsymbol{x}_i - \mu_j\right)^T \Sigma_j^{-1}\left(\boldsymbol{x}_i - \mu_j\right)\right) \tag{3}$$

$$\rho(x_i) = \sum_{j=1}^{k} \rho\left(x_i \middle| C_j\right).w_j \tag{4}$$

b) M-step: focuses on updating all three model parameters to maximize the likelihood to better match the data distribution.

$$\mu_j^{new} = \frac{\sum_{i=1}^{n} z_{ij}.\boldsymbol{x}_i}{\sum_{i=1}^{n} z_{ij}} \tag{5}$$

$$\Sigma_j^{new} = \frac{\sum_{i=1}^{n} z_{ij}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{n} z_{ij}} \tag{6}$$

$$w_j^{new} = \frac{1}{n}\sum_{i=1}^{n} z_{ij} \tag{7}$$

c) Reiterate steps a and b until the convergence criterion is met.

b. Silhouette Score
Silhouette score is a clustering model evaluation technique that has a range of values between -1 to 1 where the greater the value, indicating that the object has been well clustered (Januzaj et al., 2023). There are four categories of silhouette score assessment seen in Table 1 (Yohansa et al., 2022).

Table 1. Silhouette score categories

Category	Score	Criteria
1	0,71 – 1,00	Strong Structure
2	0,51 – 0,70	Good Structure
3	0,26 – 0,50	Weak Structure
4	≤ 0,25	Bad Structure

Silhouette score will measure the average distance between objects in the same cluster and the average distance of objects to other objects in other adjacent clusters (Myagmarsuren, 2024; Renaldi et al., 2024) with formula shown in (8).

$$silhoette(x) = \frac{b(x)-a(x)}{\max(a(x),b(x))} \tag{8}$$

Where, *a(x)* is average distance of object *x* and all other objects in the same cluster; and *b(x)* : the average distance of object *x* and all other objects in different nearby clusters.

2. Research Methodology

This research using quantitative approach with secondary data from the Central Bureau of Statistics (Directorate of Statistical Distribution, 2023)and the BACI database from CEPII in 2022 (Gaulier & Zignago, 2010). Python and Jupyter Notebook are used as the primary language and software tools. The research flow is illustrated in Figure 1.

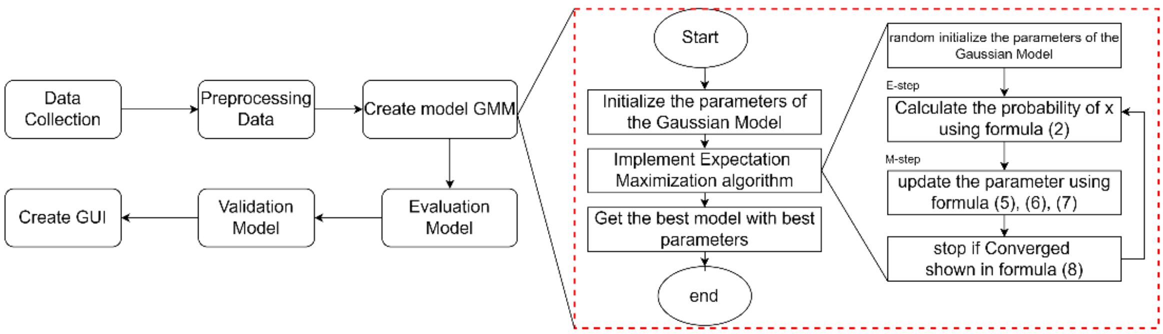

Figure 1. Research Flow

a. Data Collection

Data is stored in the form of *Harmonized Commodity* (HS) which is an international standard code of trade product classification. The data has 15 variables including *"Section ID", "Section", "HS2 ID", "HS2", "HS4 ID", "HS4", "HS6 ID", "HS6", "Country ID", "Country", "Unit ID", "Unit", "Year", "Trade Value",* and *"Quantity"* with each variable having 122,904 data.

b. Preprocessing Data

Data preprocessing involves the use of various techniques to improve the quality of the raw data (Fan et al., 2021). Therefore, various types of data pre-processing are carried out, namely:

1) Handling null values aims to avoid analysis errors so that imputation is by median value (Hameed & Ali, 2022).
2) Handling data duplication aims to maintain integrity and efficiency in data analysis by using the *drop_duplicates()* function.
3) Data type handling aims to ensure that each column in the dataset has the appropriate data type. Checking the data type uses the *info()* function and changing the data type uses the *astype()* function.
4) Data scaling aims to bring data points closer together to make the data more uniform. In this study, *MinMaxScaler* is the data scaling method used (Sharma, 2022).
5) The separation of data aims to facilitate analysis focused on each continent by retaining the top five commodities by export volume at the HS2 level listed as shown in Table 2.

Table 2. Top 5 Commodity in HS2 based on Export Volume

HS2 ID	HS2
15	Animal or vegetable fats and oils and their cleavage products; prepared animal fats; animal or vegetable waxes
72	Iron and steel
27	Mineral fuels, mineral oils and products of their distillation; bituminous substances; mineral waxes
26	Ores, slag and ash
25	Salt; sulphur; earths, stone; plastering materials, lime and cement.

6) Handling outliers aims to reduce errors in analysis (Dash et al., 2023). In this study, Outlier detection uses the z-score method and outliers are removed.

c. Creating Model

Clustering will use quantity and trade values as the main factors with k=3, 4, and 5 for comparison. The clustering method used in the research is *GMM* which utilizes the EM algorithm. The implementation of GMM is done using the sklearn library, which provides a built-in function for GMM, with the default parameters defined.

d. Evaluation and Validation Model

The model is evaluated using the silhouette score method and the best k model with the highest silhouette score for each continent is selected. To validate the clustering results, a comparison will be made with the 2023 export data from the Central Bureau of Statistics.

e. Creating GUI

This application will be built using flask, HTML and CSS called XportID (*Export of Indonesia*). This application provides several pages, namely home page, data page, method page, about us page and clustering page.

C. RESULT AND DISCUSSION

1. Result of Pre-processing Data

Table 3 shows the data results after preprocessing for Asian continent and other continents have the same data format. The amount of data is 478 rows with 5 columns namely "Continent", "Country", "HS2", "Trade Value", and "Quantity".

Table 3. Part of the data preprocessing results for the Asian continent

Continent	Country	HS2	Trade Value	Quantity
Asia	Afghanistan	15	0.000008	0.000001
Asia	Armenia	15	0.000037	0.000006
⋮	⋮	⋮	⋮	⋮

Asia continent has 151 data from 42 countries, America continent has 71 data from 30 countries, Africa continent has 82 data from 43 countries, Australia continent has 31 data from 13 countries, and Europe continent has 143 data from 41 countries.

2. GMM based Clustering Model

a. Asia

The best cluster for the Asian continent is 5 clusters with a value of 0.7035 categorized as good structure with the most optimal parameters shown in Table 4.

Table 4. Optimal parameters for Cl.ustering the Asian continent

k	μ	Σ	w
0	$[6.58\ x\ 10^{-4}, 1.76\ x\ 10^{-4}]$	$\begin{bmatrix} 2.57\ x\ 10^{-6} & 1.91\ x\ 10^{-7} \\ 1.91\ x\ 10^{-7} & 1.16\ x\ 10^{-6} \end{bmatrix}$	0.7188
1	$[3.01\ x\ 10^{-1}, 3.60\ x\ 10^{-2}]$	$\begin{bmatrix} 1.18\ x\ 10^{-3} & -1.26\ x\ 10^{-4} \\ -1.26\ x\ 10^{-4} & 5.47\ x\ 10^{-5} \end{bmatrix}$	0.0199
2	$[3.76\ x\ 10^{-2}, 9.99\ x\ 10^{-3}]$	$\begin{bmatrix} 1.02\ x\ 10^{-3} & 9.86\ x\ 10^{-5} \\ 9.86\ x\ 10^{-5} & 2.31\ x\ 10^{-4} \end{bmatrix}$	0.2217
3	$[2.83\ x\ 10^{-1}, 1.67\ x\ 10^{-1}]$	$\begin{bmatrix} 5.51\ x\ 10^{-4} & 5.68\ x\ 10^{-5} \\ 5.68\ x\ 10^{-5} & 6.60\ x\ 10^{-5} \end{bmatrix}$	0.0199
4	$[1.72\ x\ 10^{-1}, 7.91\ x\ 10^{-2}]$	$\begin{bmatrix} 2.30\ x\ 10^{-5} & 3.59\ x\ 10^{-5} \\ 3.59\ x\ 10^{-5} & 2.10\ x\ 10^{-3} \end{bmatrix}$	0.0198

Table 5 shows the clustering results for the Asian continent. The 5 clusters consist of very high market opportunities (cluster 4) with total of 3 data, high market opportunities (cluster 3) with total of 3 data, medium market opportunities (cluster 2) with total of 3 data, low market opportunities (cluster 1) with total of 33 data, and very low market opportunities (cluster 0) with total of 104 data.

Table 5. Clustering results of the Asian continent

k	HS2-15	HS2-72	HS2-27	HS2-26	HS2-25
0	Afghanistan, Armenia, ...	Azerbaijan, Bahrain, ...	Armenia, Azerbaijan, ...	Cambodia, Georgia, ...	Armenia,Bahrain, Brunei, ...
1	Bangladesh, Burma, Iraq, ...	India, Malaysia, Philippines, ...	Bangladesh, Brunei, ...	India, Japan, Malaysia, ...	Bangladesh, China, ...
2	Pakistan	-	Thailand	China	-
3	-	-	Malaysia, Philippines, South Korea	-	-
4	China, India	-	Singapore	-	-

Figure 2 shows the visualization of the clustering results for the Asian continent with each icon shape and colour drawing a different cluster.

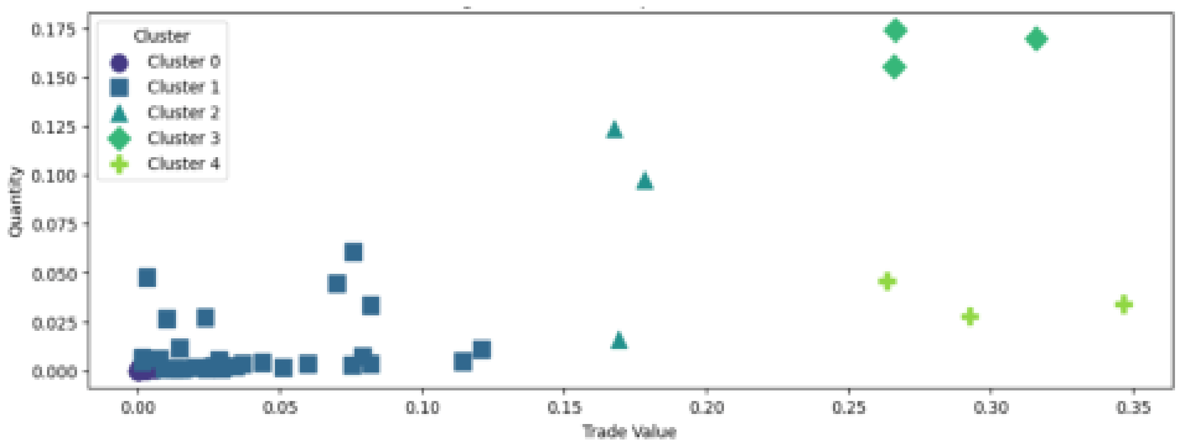

Figure 2. Clustering of Indonesian Export Commodities to Asia

It can be seen from the pattern that Malaysia, the Philippines, and South Korea are countries that are in the same cluster for 4 out of 5 commodities, indicating that their preferences are similar. This indicates that a marketing strategy that is successful in one country is likely to be implemented effectively in other countries in this cluster.

b. America

The best cluster for the continent of America is 3 clusters with a value of 0.8534 categorized as strong structure with the most optimal parameters shown in Table 6.

Table 6. Optimal parameters for Clustering the American continent

k	μ	Σ	w
0	$[2.58\ x\ 10^{-4}, 3.12\ x\ 10^{-5}]$	$\begin{bmatrix} 1.38\ x\ 10^{-6} & 4.15\ x\ 10^{-8} \\ 4.15\ x\ 10^{-8} & 1.01\ x\ 10^{-6} \end{bmatrix}$	0.8821
1	$[3.02\ x\ 10^{-2}, 2.15\ x\ 10^{-3}]$	$\begin{bmatrix} 1.00\ x\ 10^{-6} & 3.14\ x\ 10^{-34} \\ 3.14\ x\ 10^{-34} & 1.00\ x\ 10^{-6} \end{bmatrix}$	0.0140
2	$[6.34\ x\ 10^{-3}, 7.03\ x\ 10^{-4}]$	$\begin{bmatrix} 1.78x\ 10^{-5} & 1.21\ x\ 10^{-6} \\ 1.21\ x\ 10^{-6} & 1.29\ x\ 10^{-6} \end{bmatrix}$	0.1038

Table 7 shows the clustering results for the American continent separated by cluster and commodity type. The 3 clusters consist of high market opportunities (cluster 2) with total of 1 data, medium market opportunities (cluster 1) with total of 7 data, and low market opportunities (cluster 0) with total of 64 data.

Table 7. Clustering results of the American Continent

k	HS2-15	HS2-72	HS2-27	HS2-26	HS2-25
0	Argentina, Barbados, Bolivia, Canada, ...	Argentina, Canada, ...	Argentina,Brazil, Canada, ...	Brazil, Canada, Chile, ...	Argentina, Brazil, ...
1	Colombia, Haiti, Mexico	Brazil, USA	Mexico, USA	-	-
2	Brazil	-	-	-	-

Figure 3 shows the visualization of the clustering results for the American continent with each icon shape and colour drawing a different cluster.

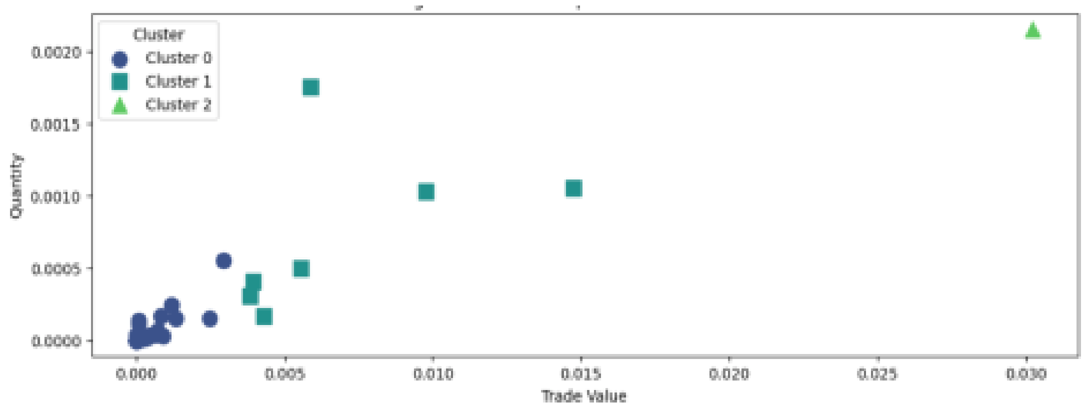

Figure 3. Clustering of Indonesian Export Commodities to America

From the pattern, it can be seen that Indonesia's export activities to America are still only focused on the United States, which has always been the main destination. However, Brazil and Mexico can be new market opportunities because they are located in clusters with high and medium potential.

c. Africa

The best cluster for the African continent is 3 clusters with a value of 0.8165 categorized as strong structure with the most optimal parameters shown in Table 8.

Table 8. Optimal parameters for Clustering the African continent

k	μ	Σ	w
0	$[3.83\ x\ 10^{-4}, 4.45\ x\ 10^{-5}]$	$\begin{bmatrix} 1.58\ x\ 10^{-6} & 5.22\ x\ 10^{-8} \\ 5.22\ x\ 10^{-8} & 1.01\ x\ 10^{-6} \end{bmatrix}$	0.8511
1	$[5.92\ x\ 10^{-3}, 5.42\ x\ 10^{-4}]$	$\begin{bmatrix} 7.78\ x\ 10^{-6} & 5.98\ x\ 10^{-7} \\ 5.98\ x\ 10^{-7} & 1.05\ x\ 10^{-6} \end{bmatrix}$	0.0658
2	$[1.06\ x\ 10^{-2}, 9.47\ x\ 10^{-4}]$	$\begin{bmatrix} 5.64\ x\ 10^{-6} & 4.59\ x\ 10^{-7} \\ 4.59\ x\ 10^{-7} & 1.05\ x\ 10^{-6} \end{bmatrix}$	0.0831

Table 9 shows the clustering results for the American continent. The 3 clusters are categorized as high market opportunities (cluster 2) with total of 7 data, medium market opportunities (cluster 1) with total of 5 data, and low market opportunities (cluster 0) with total of 70 data.

Table 9. Africa continent clustering results

k	HS2-15	HS2-72	HS2-27	HS2-26	HS2-25
0	Burkina Faso, Burundi, Chad, Comoros, Gabon...	Algeria,Cameroon, Egypt, ...	Burkina Faso, Egypt, Kenya, ...	South Africa	Egypt, Ghana, Kenya, ...
1	Mauritania, Nigeria, Senegal, Sudan, Tunisia	-	-	-	-
2	Algeria, Angola, Benin, Djibouti, Kenya, Tanzania, Togo	-	-	-	-

Figure 4 shows the visualization of the clustering results for the African continent with each icon shape and colour drawing a different cluster.

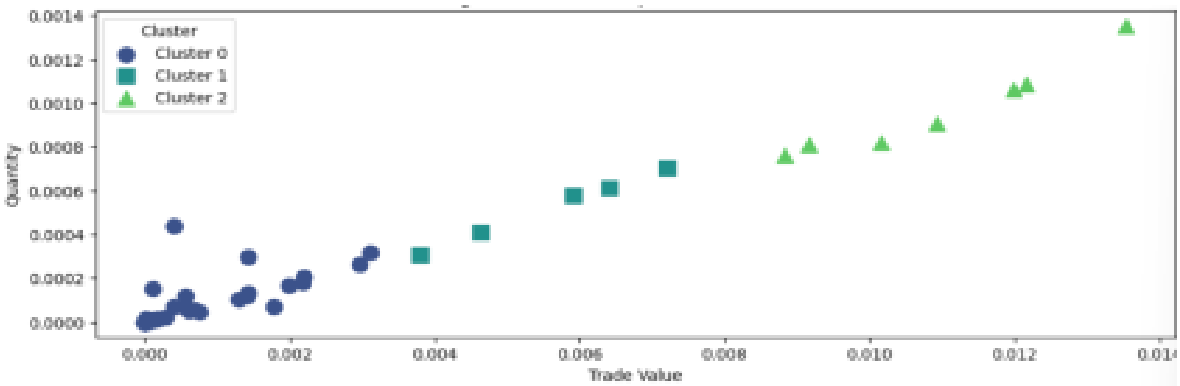

Figure 4. Clustering of Indonesian Export Commodities to Africa

The destination countries for exports in Africa are quite diverse. Therefore, the focus should be directed at the product side. There is a high correlation in HS2-15, which is a commodity related to animal and plant products. This shows that a strategy that focuses on the development of these commodities can be more effective.

d. Australia

The best cluster for the Australian continent is 3 clusters with a value of 0.8540 categorized as strong structure with the most optimal parameters shown in Table 10.

Table 10. Optimal parameters for Clustering the Australian continent

k	μ	Σ	w
0	$[1.98\ x\ 10^{-4}, 5.40\ x\ 10^{-5}]$	$\begin{bmatrix} 1.09\ x\ 10^{-6} & 1.49\ x\ 10^{-8} \\ 1.49\ x\ 10^{-8} & 1.02\ x\ 10^{-6} \end{bmatrix}$	0.9369
1	$[4.23\ x\ 10^{-3}, 3.77\ x\ 10^{-3}]$	$\begin{bmatrix} 1.00\ x\ 10^{-6} & 5.04\ x\ 10^{-9} \\ 5.04\ x\ 10^{-9} & 1.06\ x\ 10^{-6} \end{bmatrix}$	0.0323
2	$[3.92\ x\ 10^{-3}, 4.50\ x\ 10^{-4}]$	$\begin{bmatrix} 1.02\ x\ 10^{-6} & 7.35\ x\ 10^{-9} \\ 7.35\ x\ 10^{-9} & 1.05\ x\ 10^{-6} \end{bmatrix}$	0.0308

Table 11 shows the clustering results for the Australian continent. The 3 clusters consist of high market opportunities (cluster 2) with total of 1 data, medium market opportunities (cluster 1) with total of 1 data, and low market opportunities (cluster 0) with total of 29 data.

Table 11. Clustering results of the Australian continent

k	HS2-15	HS2-72	HS2-27	HS2-26	HS2-25
0	Australia, Fiji, Kiribati, Marshall Islands, ...	Fiji, New Zealand, ...	Fiji, Papua New Guinea,...	Australia ,...	Fiji, New Zealand, ...
1	-	Australia	-	-	-
2	-	-	New Zealand	-	-

Figure 5 shows the visualization of the clustering results for the Australian continent with each icon shape and colour drawing a different cluster.

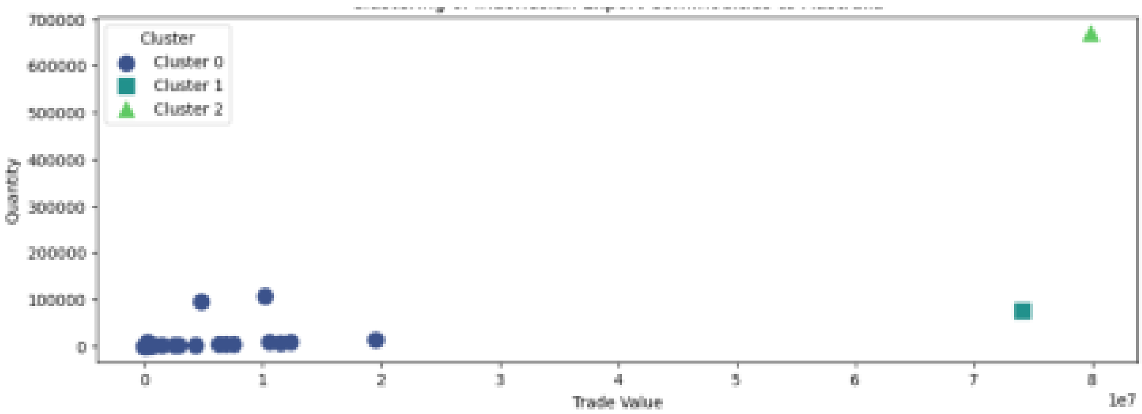

Figure 5. Clustering of Indonesian Export Commodities to Australia

The focus in the Australian continent remains primarily on Australia, which has always been the main destination for exports. However, New Zealand may become a new market opportunity, particularly for HS2-27 commodities.

e. Europe

The best cluster for the European continent is 3 clusters with a value of 0.8654 categorized as strong structure with the most optimal parameters shown in Table 12.

Table 12. Optimal parameters for Clustering the European continent

k	μ	Σ	w
0	$[2.58\ x\ 10^{-4}, 5.07\ x\ 10^{-5}]$	$\begin{bmatrix} 1.36\ x\ 10^{-6} & 6.59\ x\ 10^{-8} \\ 6.59\ x\ 10^{-8} & 1.03\ x\ 10^{-6} \end{bmatrix}$	0.8825
1	$[1.82\ x\ 10^{-2}, 1.21\ x\ 10^{-3}]$	$\begin{bmatrix} 5.01\ x\ 10^{-5} & 4.00\ x\ 10^{-6} \\ 4.00\ x\ 10^{-6} & 1.54\ x\ 10^{-6} \end{bmatrix}$	0.0259
2	$[8.00\ x\ 10^{-3}, 1.10\ x\ 10^{-3}]$	$\begin{bmatrix} 1.45\ x\ 10^{-5} & 7.07\ x\ 10^{-7} \\ 7.07\ x\ 10^{-7} & 2.24\ x\ 10^{-6} \end{bmatrix}$	0.0916

Table 13 shows the clustering results for the European continent. The 3 clusters consist of high market opportunities (cluster 2) with total of 3 data, medium market opportunities (cluster 1) with total of 14 data, and low market opportunities (cluster 0) with total of 126 data.

Table 13. Clustering results of the European continent

k	HS2-15	HS2-72	HS2-27	HS2-26	HS2-25
0	Albania, Austria, Belgium, ...	Andorra, Austria, ...	Andorra, Denmark, Estonia, Hungary, ...	Belgium, Czechia, ...	Austria, Belgium, ...
1	Denmark, Estonia, France, ...	Belgium, Netherland, Spain	Croatia, Slovenia, Spain	Bulgaria	-
2	Turkey	-	-	Germany, Spain	-

Figure 6 shows the visualization of the clustering results for the European continent with each icon shape and colour drawing a different cluster.

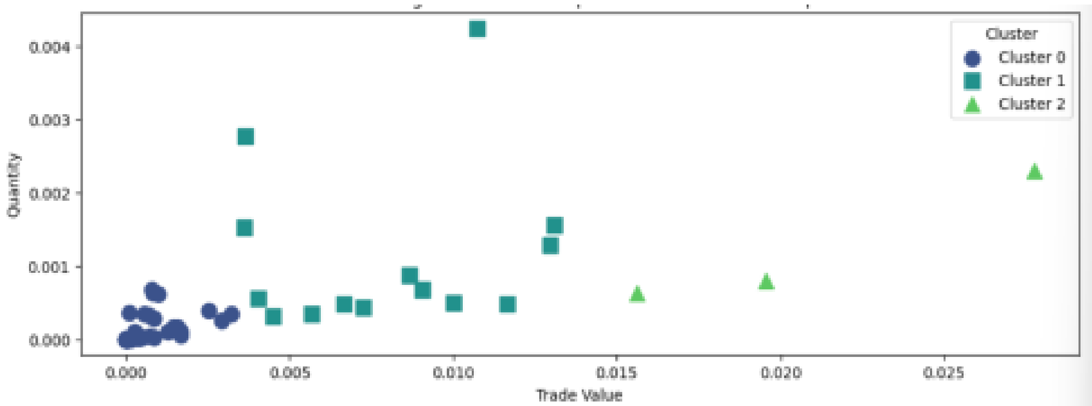

Figure 6. Clustering of Indonesian Export Commodities to Europe

European countries have high market opportunities due to their country destinations and fairly diverse range of commodities. However, Spain may have the most market potential due to 3 out of 5 commodities that have high and medium demand.

3. Model Evaluation

Table 14 shows the comparison of model evaluation values between the proposed model and previous research using Silhouette Score.

Table 14. Comparison of evaluation scores with previous research

Model	K	Asia	America	Africa	Australia	Europe	Avg./Scr.
Proposed	3	0.6530	**0.8534**	**0.8165**	**0.8540**	**0.8654**	0.819
	4	0.6552	0.8446	0.8127	0.8539	0.8528	
	5	**0.7035**	0.8358	0.8127	0.8539	0.8445	
Islami & Sihombing (2021)	3	0.1346	**0.7140**	**0.8883**	-	**0.3688**	0.643
	4	**0.6014**	0.3464	0.1059	-	0.3599	
	5	0.1039	0.3640	0.3543	-	0.1291	
Akhmad & Priyono (2024)	2	-	-	-	-	-	0.740
Annisa Octaria et al. (2023)	2	-	-	-	-	-	0.730

On the silhouette score value, the proposed model has a high performance for each continent and gets an average value of 0.819. Comparison results with Islami & Sihombing (2021); Akhmad & Priyono (2024); Annisa Octaria et al. (2023) shows that the proposed model has a higher SC value which indicates that the proposed model is better at clustering data.

4. Model Validation

Table 15 presents a comparison of countries with medium and high potential based on the clustering results and data from 2023. The selection of countries for 2023 is based on the highest trade value multiplied by quantity.

Table 15. Country High and Medium Potential Comparison

Continent	Clustering Result	Data from 2023
Asia	China, India, Pakistan, Singapore, Malaysia, Philippines, South Korea, Thailand	China, India, Pakistan, Malaysia, Vietnam, Philippines, South Korea, Japan
America	Brazil, Haiti, Mexico, Colombia, USA	USA, Brazil, Mexico, Haiti, Canada
Africa	Algeria, Angola, Benin, Djibouti, Kenya, Tanzania, Togo, Mauritania, Nigeria, Senegal, Sudan, Tunisia	Egypt, South Africa, Djibouti, Kenya, Algeria, Tanzania, Benin, Nigeria, Senegal, Togo, Mauritania
Australia	Australia, New Zealand	Australia, New Zealand
Europe	Turkey, Denmark, Estonia, France, Germany, Greece, Sweden, United Kingdom, Belgium, Netherland, Spain, Croatia, Slovernia, Bulgaria	Spain, Russia, Netherland, Turkey, Belgium, Italy, Germany, Bulgaria, Poland, Greece

Overall, the comparison across five continents shows that countries with medium and high market potential maintained strong export trends in 2023, especially for key commodities like oil and vegetable fats. While some high-exporting countries were not classified as medium or high potential, the clustering results overall accurately identified potential markets.

5. XportID: Web-based Application

Figure 7 shows a display of a website-based application created by Flask, HTML, and CSS that has 5 pages which are home, method, about us, data, and clustering page.

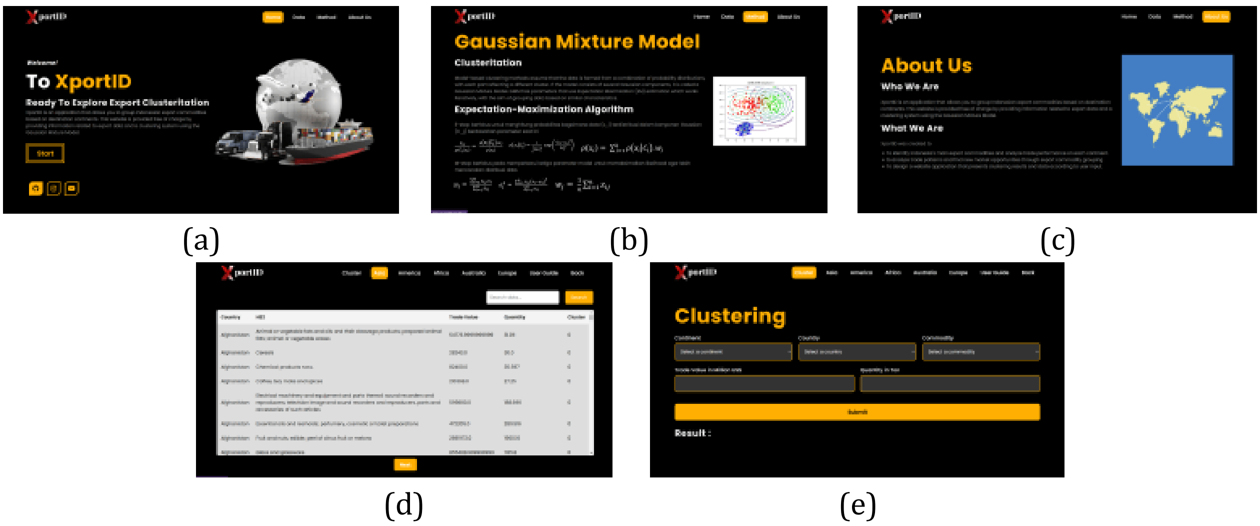

Figure 7. Page views of (a) home, (b) method, (c) about us, (d) data, and (e) clustering

In the clustering page, users are required to input the required parameters, namely continent name, country name, selected commodity, trade value, and quantity. Users are required to select the continent name from the *dropdown* provided first because it will affect the country name that can be selected. Then, the user can click the *submit* button to start the clustering process. If there is one column that is not filled in, then when the user clicks the *submit* button, a warning will appear to fill in the empty column.

D. CONCLUSION AND SUGGESTIONS

XportID is a web-based application designed to cluster Indonesian export commodities based on their destination continent. The dataset used, after preprocessing, consists of 478 rows and 5 columns: "Continent", "Country", "HS2", "Trade Value," and "Quantity." According to the findings of this research, the Asian continent has 5 clusters with the best value of 0.7035, the American continent has 3 clusters with the best value of 0.8165, the African continent has 3 clusters with the best value of 0.8534, the Australian continent has 3 clusters with the best value of 0.8540, and the European continent has 4 clusters with the best value of 0.8654. Based on these results, the clustering can be categorized as a strong structure, as the silhouette scores obtained are close to 1, indicating well-defined clusters. A higher silhouette score reflects better-defined clusters. The clustering results can be used to enhance export strategies by identifying regions with high export potential. Countries within the same cluster may share similar preferences, making it more efficient to target them with specific export products. For future research, it is suggested to incorporate a time dimension, such as seasonality or economic trends, as these factors can impact trade patterns and clustering outcomes, since countries may experience shifts in demand from year to year.

REFERENCES

Adiwidyatma, A. R., Diyasa, I. G. S. M., & Trimono. (2024). Analysis of Clustering Methods on the Causal Factors of Diabetes Mellitus with Fuzzy C Means Method. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *5*(2), 983–996. https://doi.org/10.46306/lb.v5i2

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. In *Electronics (Switzerland)* (Vol. 9, Issue 8, pp. 1–12). MDPI AG. https://doi.org/10.3390/electronics9081295

Akhmad, E. P. A., & Priyono, B. (2024). Classification of Indonesian Frozen Shrimp Export Data Using K-Medoids Clustering. *Technology, and Business (JETBIS)*, *3*(5). https://doi.org/https://doi.org/10.57185/jetbis.v3i5.106

Andi Purnomo, M. R., Azzam, A., & Uswatun Khasanah, A. (2020). Effective Marketing Strategy Determination Based on Customers Clustering Using Machine Learning Technique. *Journal of Physics: Conference Series*, *1471*(1). https://doi.org/10.1088/1742-6596/1471/1/012023

Andriyani, K., Marwa, T., Adnan, N., & Muizzuddin. (2020). The Determinants of Foreign Exchange Reserves: Evidence from Indonesia. *The Journal of Asian Finance, Economics and Business*, *7*(11), 629–636. https://doi.org/10.13106/jafeb.2020.vol7.no11.629

Annisa Octaria, E., Maulani, D., & Daris Syafiq, M. (2023). Clustering Crumb Rubber Exports Based on Destination Countries Using the K-Means Method. *Journal of International Trade*, *2*(2), 46–51. https://doi.org/10.32832/jit

Central Bureau of Statistics. (2023). *Indonesian Foreign Trade Statistics Exports 2022, Volume II*.

Central Bureau of Statistics. (2024). *Indonesian Foreign Trade Statistics Exports 2023, Volume I*.

Charles, G. E., & Rangen Jaya, M. H. (2023). Credit For Export Bow Penetration: Catalyzing A Breakthrough In The International Trade Sector. *Journal of Sustainable Economics*, *1*(1), 24–28. https://doi.org/10.32734/jse.v1i1.12066

Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, *6*. https://doi.org/10.1016/j.dajour.2023.100164

Faidah, D. Y., Hudzaifa, A. M., & Pontoh, R. S. (2024). Clustering of Childhood Diarrhea Diseases using Gaussian Mixture Model. *Communications in Mathematical Biology and Neuroscience*. https://doi.org/10.28919/cmbn/8365

Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Frontiers in Energy Research*, *9*. https://doi.org/10.3389/fenrg.2021.652801

Fernando, L., & Fianty, M. I. (2024). Optimizing Motorcycle Sales: Enhancing Customer Segmentation with K-Means Clustering and Data Mining Techniques. *Journal of Information Systems and Informatics*, *6*(3), 1484–1498. https://doi.org/10.51519/journalisi.v6i3.799

Gaulier, G., & Zignago, S. (2010). *BACI: International Trade Database at the Product-level The 1994-2007 Version*. http://www.cepii.fr/anglaisgraph/bdd/baci.htm.

Hameed, W. M., & Ali, N. A. (2022). Comparison of Seventeen Missing Value Imputation Techniques. *Journal of Hunan University Natural Sciences*, *49*(7), 26–36. https://doi.org/10.55463/issn.1674-2974.49.7.4

Islami, R. L., & Sihombing, P. R. (2021). Application Biplot and K-Medians Clustering to Group Export Destination Countries of Indonesia's Product. *Advance Sustainable Science, Engineering and Technology*, *3*(1). https://doi.org/10.26877/asset.v3i1.8451

Januzaj, Y., Beqiri, E., & Luma, A. (2023). Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique. *International Journal of Online and Biomedical Engineering*, *19*(4), 174–182. https://doi.org/10.3991/ijoe.v19i04.37059

Ji, X., Dong, F., Zheng, C., & Bu, N. (2022). The Influences of International Trade on Sustainable Economic Growth: An Economic Policy Perspective. *Sustainability (Switzerland)*, *14*(5). https://doi.org/10.3390/su14052781

Matondang, K. A., Silalahi, H. H. B., Naibaho, H. S. D., Safitri, L., & Siregar, A. N. (2024). The Role of International Trade in Increasing The Rate of Economic Growth in Indonesia: A Literature Review. *International Journal Of Education, Social Studies, And Management (IJESSM)*, *4*(2), 278–290. https://doi.org/10.52121/ijessm.v4i2.238

Myagmarsuren, S. (2024). Exploring the Use of Silhouette Score in K-Means Clustering for Image Segmentation (Exploring the Use of Silhouette Score in K-Means Clustering for Image Segmentation). *International Journal of Engineering Research & Technology (IJERT)*, *13*(4). http://www.ijert.org

Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*, *171*, 158–167. https://doi.org/10.1016/j.procs.2020.04.017

Rahman, F., Ridho, I. I., Muflih, M., Pratama, S., Raharjo, M. R., & Windarto, A. P. (2020). Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country. *IOP Conference Series: Materials Science and Engineering*, *835*(1). https://doi.org/10.1088/1757-899X/835/1/012058

Renaldi. S, S., Prasetya, D. A., & Muhaimin, A. (2024). Analisis Klaster Partitioning Around Medoids dengan Gower Distance untuk Rekomendasi Indekos (Studi Kasus: Indekos di Sekitar Kampus UPNVJT). *G-Tech: Jurnal Teknologi Terapan*, *8*(3), 2060–2069. https://doi.org/10.33379/gtech.v8i3.4898

Sani Akbar, J., & Rezeki, N. S. (2021). Analysis of Economic Growth and Export Development of Bangka Belitung Islands. *Business and Accounting Research (IJEBAR)* , *5*(4), 625–633.

Santoso, H. A., & Haw, S. C. (2023). Improvement of k-Means Clustering Performance on Disease Clustering using Gaussian Mixture Model. *Journal of System and Management Sciences*, *13*(5), 169–179. https://doi.org/10.33168/JSMS.2023.0511

Sharma, V. (2022). A Study on Data Scaling Methods for Machine Learning. *International Journal for Global Academic & Scientific Research*, *1*(1). https://doi.org/10.55938/ijgasr.v1i1.4

Solikhun, S., Yasin, V., & Donni Nasution. (2022). Optimization of the Number of Clusters of the K-Means Method in Grouping Egg Production Data in Indonesia. *International Journal of Artificial Intelligence & Robotics (IJAIR)*, *4*(1), 39–47. https://doi.org/10.25139/ijair.v4i1.4328

Syahrani, D., Sitorus, H. N. S., & Sitompul, R. S. M. (2022). The Influence of International Trade on Indonesia's Economic Growth. *International Journal of Business and Applied Economics*, *1*(1), 27–30. https://doi.org/10.55927/ijbae.v1i1.2147

Ulvi, H. A., & Ikhsan, M. (2024). Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia. *Sinkron*, *8*(3), 1671–1685. https://doi.org/10.33395/sinkron.v8i3.13815

Yohansa, M., Notodiputro, K. A., & Erfiani, E. (2022). Dynamic Time Warping Techniques for Time Series Clustering of Covid-19 Cases in DKI Jakarta. *ComTech: Computer, Mathematics and Engineering Applications*, *13*(2), 63–73. https://doi.org/10.21512/comtech.v13i2.7413