
Cluster Analysis of Indonesia's Electronic Integrated Circuit Export Destination Using K-Means and K-Medoids

Christian Andrew¹, Rahmatina Hidayati^{1*}

¹ Merdeka University Malang, Terusan Dieng Street 62-64, Klojen, Pisang Candi, Sukun District, Malang City, East Java 65146, Indonesia

Article History

Submitted : 08-07-2026

Revised : 26-07-2026

Accepted : 05-08-2026

Published : 20-09-2026

DOI: 10.32664/smatika.v16i03.2474

Keyword

Clustering; Davies-Bouldin Index; Electronic Integrated Circuit (EIC); Export; K-Means; K-Medoids; Silhouette Score

*Corresponding Author:

rahmatina.hidayati@unmer.ac.id

Abstract

One of Indonesia's export commodities is Electronic Integrated Circuits (EICs). Understanding the characteristics of EIC export destination countries is important for identifying export patterns and to support strategic decision-making. Clustering is one approach that can be used to analyze export destination countries by grouping them based on their similar characteristics. The objective of this study is to evaluate the performance of K-Means and K-Medoids algorithms in analyzing Indonesia's EIC export destination countries. The data used in this study were obtained from Badan Pusat Statistik (BPS) and consist of export value and weight data from 2021 to 2025 containing 84 destination countries. This dataset is selected because limited studies have compared K-Means and K-Medoids on Indonesia's EIC export destination countries. The data from BPS were then preprocessed, then normalized by applying Min-Max normalization method. The Elbow Method was subsequently employed to find the optimal number of clusters, with the analysis indicating that four clusters were optimal. Following the Elbow Method, K-Means and K-Medoids algorithms were then applied and assessed based on the Davies-Bouldin Index and Silhouette Score. The findings of this study demonstrate that K-Means outperformed K-Medoids in both the DBI and Silhouette Score by achieving lower DBI value and higher Silhouette Score. The results of both algorithms revealed that most destination countries belonged to the very low export value and weight cluster, accounting for around 88% of countries in K-Means and 60% in K-Medoids. The findings suggest that K-Means is more suitable for clustering Indonesia's EIC export destination countries and can provide useful insight into export patterns that may support future trade analysis and decision-making.

1. Introduction

Export is one of the main components of a country's economic activity that has significant impact on the country's sustainability and its growth. A country's export activities, whether in form of goods or services, help increase its foreign exchange earnings, which can be used to support the country's economic growth [1]. One of Indonesia's exported goods is electronic integrated circuit (EIC). EIC is a semiconductor chip that integrates complex electronic component into a single chip. These chips are the foundation of modern electronic devices such as smartphones, laptops, and televisions. Their primary functions include digital signal processing, memory management, and communication interfaces [2]. Export data for Indonesia's electronic integrated

circuit (HS code 8542) from Badan Pusat Statistik (BPS) showed an upward trend during 2021 to 2025. The increasing volume and complexity of export data highlight the importance of applying data analysis technique to support decision-making and extract meaningful information from large dataset [3].

The official BPS website contains large amount of export and import data, including export value and weight the of EICs. The EIC export data contain information on export value (Free on Board) and export weight of numerous destination countries, each of which has different export characteristics. The data provided by BPS are currently available in the form of tables, which makes it difficult to identify export patterns, priority export destination countries, and classify countries with similar characteristics.

Therefore, this study implements the clustering method to analyze EIC export data obtained from BPS, using export value (FOB) and export weight as the clustering variables. Clustering is a method that helps to identify patterns in data and can be used to support decision-making processes as well as the identification of characteristics that may be overlooked by the conventional method [4]. This study uses the K-Means and K-Medoids clustering methods to analyze Indonesia's electronic integrated circuits export from 2021 to 2025. Previous studies have shown that K-Means is less time-consuming than K-Medoids [5]. However, another study found that K-Medoids is more effective in producing stable clusters than K-Means [6]. These contrasting findings motivate the comparison of both algorithms in the present study. Furthermore, the characteristics of Indonesia's EIC export data, which exhibit considerable variation and contain outliers, provide an opportunity to evaluate both algorithms for this specific application. In addition, computational efficiency and cluster stability are used as the comparison criteria.

A previous study [7] compared the performance of K-Means and K-Medoids for clustering Indonesian industrial export commodities. However, unlike the previous study, the present research focuses on Indonesia's EIC export destination countries rather than export commodities. This study analyzes export value (FOB) and export weight data from 2021 to 2025 to identify groups of destination countries with similar export characteristics. Therefore, this study extends the previous work by providing a destination-country perspective for a specific export commodity, which may support the identification of priority export markets and the development of export strategies.

This study is expected to contribute to a better understanding of Indonesia's electronic integrated circuit exports. The results of this study may be used to identify clusters of export destination countries, determine priority markets, and support the development of export strategies. Furthermore, this study is expected to provide a valuable insight for comparative analysis of the K-Means and K-Medoids methods for export data clustering.

2. Research Method

In general, this study is divided into several stages, as follows:

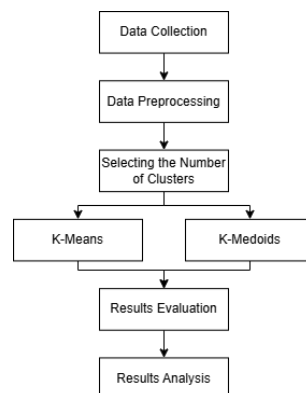


Figure 1. Research Flow

Figure 1 illustrates the overall research workflow. The study begins with collecting Indonesia's EIC export value (FOB) and export weight data from BPS. The collected data then preprocessed through data aggregation, data cleaning, and Min-Max normalization before being analyzed using the K-Means and K-Medoids clustering algorithms. Finally, the clustering results are evaluated using the Davies-Bouldin Index and Silhouette Score to determine the algorithm that provides better clustering performance.

2.1 Data Collection

This study utilizes data that consist of Indonesia's electronic integrated circuit exports destination countries from 2021 to 2025, sourced from the official website of Badan Pusat Statistik (BPS). The dataset consists of two main variables, namely export value (FOB) measured in U.S Dollars (US\$) and export weight measured in kilograms (kg). The data used in this study consist of the combined value of five main categories of electronic integrated circuits (EICs), classified under HS Code 85423100, 85423200, 85423300, 85423900, and 85429000.

2.2 Data Preprocessing

Data preprocessing procedure consisted of 3 main stages: data cleaning, data selection and data transformation. During the data cleaning stage, record with missing value were replaced with 0 while records that already have a value of 0 remained unchanged. Blank values were replaced with 0 because they represent no recorded export activities rather than missing observations [8]. Furthermore, outlier records were not removed or modified because they reflect the actual conditions of export activities.

In the data selection stage, two primary variables are selected, namely export value (FOB) measured in U.S Dollars (US\$) and export weight measured in kilograms (kg). These two variables are selected because they are expected to improve the representativeness of the resulting clusters and to capture export characteristics from both value and volume perspectives. By using both variables, the analysis is expected to distinguish countries with high export values but low export volumes, as well as countries with high export volumes and low export values.

During the data transformation stage, the data will be normalized by using the Min-Max method. Normalization is a preprocessing technique used to transform data from its original range into a common scale, thereby facilitating subsequent data analysis and prediction process [9]. Min-Max normalization is a linear transformation technique that scales attribute values into a specified range, typically between 0 and 1, while preserving the relationships among the original data values [10]. In this study, normalization process is necessary because the FOB values are considerably larger in scale compared than the export weight values, which may bias the clustering results if the data are left unnormalized.

Min-Max method is calculated using (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where:

x : original values;

x' : normalized values;

x_{min} : lowest value of the variable; and

x_{max} : highest value of the variable.

2.3 Clustering

Clustering is a data analysis technique that groups objects in a dataset based on their similarities. The clusters produced from the method exhibit distinct characteristics from one another, while the objects within each cluster share a high degree of similarities in their characteristics [11]. Clustering is a form of data mining, which is a data analysis process used to identify relevant patterns, relationships and trends within a dataset [12].

The clustering method employed in this study are K-Means and K-Medoids. K-Means is a clustering algorithm that groups data objects into clusters based on their similarities, with the objective of producing a cluster whose

member are highly similar to one another. In K-Means, clusters assignment is determined based on the distance between each data point and the centroid of a clusters [13]. K-Medoids is a clustering algorithm that partition data into multiple clusters. The primary distinction between the two methods is that instead of representing each cluster with centroid, K-Medoids uses an actual data point selected as the representative center of a cluster, known as medoid [14].

The clustering process starts with determining the optimal number of clusters using the Elbow Method. This method determines the optimal number of clusters by using inertia value, also known as Within-Cluster Sum of Squares (WCSS). WCSS is used to measure the total error represented by the distances between data points and their corresponding cluster centroids [15]. A lower WCSS value indicates that resulting clusters are more compact. However, WCSS value naturally decreases as the number of clusters increases because each cluster contains fewer data points, resulting in smaller distances between data points and their corresponding centroids. Therefore, Elbow Method identifies the point at which the rate of decrease in WCSS begins to slow down significantly, this point is referred as the elbow point [16]. Once the optimal number of clusters has been determined, the K-Means and K-Medoids algorithm are then applied.

2.4 Results Evaluation

Clustering evaluation metrics are used to evaluate the quality of clustering solutions and compare the performance of different clustering algorithm [17]. In this study, the Davies-Bouldin Index and Silhouette Score were used as the clustering evaluation metrics to compare the performance of K-Means and K-Medoids algorithms. Davies-Bouldin Index (DBI) is a metric used to evaluate the quality of cluster structures through the assessment of cluster cohesion and separation [18]. Cohesion refers to the compactness of a cluster and commonly measured by the distances between data points and their corresponding centroid. Separation represents the degree of distinction between clusters and is measured by the distance between centroids of different clusters. Davies-Bouldin Index evaluates cluster quality by analyzing intra-cluster similarity and inter-cluster separation. Clustering method with lower DBI values are considered to have better performance [19]. Davies-Bouldin Index is calculated using (2), where R_i is obtained from (3) and R_{ij} is calculated using (4).

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i \quad (2)$$

$$R_i = \max R_{ij} \quad (3)$$

$$R_{ij} = \frac{\text{var}(c_i) + \text{var}(c_j)}{\|c_i - c_j\|} \quad (4)$$

Where:

k: number of clusters;

R_i : maximum similarity ratio for cluster i;

R_{ij} : similarity ration between clusters i and j;

$\text{var}(c_i)$: dispersion of data within cluster i;

c_i : centroid of cluster i;

$\text{var}(c_j)$: dispersion of data within cluster j;

c_j : centroid of cluster j; and

$\|c_i - c_j\|$: distance between the centroids of clusters i and j.

Another metric that used to assess clustering quality is Silhouette Score, which evaluate clustering results by how well each data object fits within its assigned cluster relative to other clusters [7]. Its values range from -1 to 1, with larger scores indicate better clustering performance and smaller scores indicates poorer clustering performance [20]. Silhouette score is calculated using (5).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

where:

$a(i)$: average distance between data point i and all other data within the same cluster;
 $b(i)$: average distance between data point i and the nearest neighboring cluster; and
 $s(i)$: silhouette score for data point i .

3. Result and Discussions

3.1 Data Collection and Preprocessing

The data on Indonesia's electronic integrated circuit exports destination countries from 2021 to 2025 is obtained from the official website of Badan Pusat Statistik (BPS) and downloaded in Microsoft Excel Spreadsheet file (.xlsx) format. The dataset contains a total of 84 destination countries. The data were obtained in two separate files: one containing export value (FOB) and the other containing export weight. In addition to export value and weight, both files also contain information on the departure port or airport for each destination country. Furthermore, the data for each year is also divided into monthly periods.

The preprocessing stage begins with data aggregation, in which the EICs exports data were reorganized. Information on individual departure ports was removed, and the corresponding data is combined so that each destination country was represented by a single record. In addition, the temporal data were aggregated from monthly data to annual data and the EIC subcategories were aggregated into single category representing EIC as a whole. Subsequently, the data were further aggregated to obtain the cumulative export value and export weight for the 2021-2025. The aggregation process was conducted to provide overall representation of export patterns during the selected period. Furthermore, aggregating the data is expected to reduce the influence of year-to-year fluctuations, which may lead to less stable clustering results. Afterward, the export value and weight datasets were merged into single table. Subsequently, data cleaning is performed to handle missing value.

Table 1. Sample of Merged Export Value and Weight Data

No.	Country	FOB (US\$)	Weight (kg)
1	Algeria	\$4.242.469,82	28992,61
2	Anguilla	\$14.951,43	40
3	Argentina	\$13.389.750,15	153558,2
4	Australia	\$363.676,06	662,35
5	Austria	\$935.646,93	1447,46
6	Bangladesh	\$569.999,33	3950,35
7	Belgium	\$174,28	2,41
8	Brazil	\$126.702.245,02	401701,2
9	Brunei Darussalam	\$88,00	10,62
10	Bulgaria	\$258.867,37	199,11

Following data cleaning, the data were transformed through Min-Max Normalization method to eliminate the influence of differences in variables. Due to the large number of normalized values, all data presented in this study are displayed with six decimal places, while the original value is still retained for all calculations.

Table 2. Sample of Normalized Export Data

No.	Country	FOB (US\$)	Weight (kg)
1	Algeria	0,002583	0,006702
2	Anguilla	0,000009	0,000009
3	Argentina	0,008153	0,035494
4	Australia	0,000221	0,000153

No.	Country	FOB (US\$)	Weight (kg)
5	Austria	0,000570	0,000335
6	Bangladesh	0,000347	0,000913
7	Belgium	0,000000	0,000001
8	Brazil	0,077149	0,092852
9	Brunei Darussalam	0,000000	0,000002
10	Bulgaria	0,000158	0,000046

3.2 Clustering

Prior to the clustering process, optimal number of clusters were identified by applying the Elbow Method. The Elbow Method itself starts by applying K-Means algorithm to calculate the Within-Cluster Sum of Squares (WCSS) for each number of clusters. In this study, the number of clusters considered in the Elbow Method range between $k = 2$ and $k = 6$, where k is the number of clusters. According to the Elbow Method results, $k = 4$ was the most appropriate number of clusters for this study.

Table 3. WCSS Values for Various of Cluster Numbers

Number of Clusters	WCSS
$k = 2$	0,286847
$k = 3$	0,102382
$k = 4$	0,051881
$k = 5$	0,048954
$k = 6$	0,043586

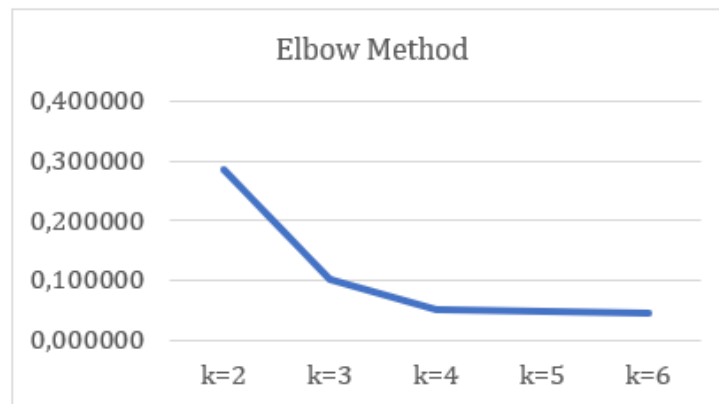


Figure 2. Elbow Method Results

Once optimal number of clusters had been determined, data clustering process was carried out using the K-Means and K-Medoids algorithm. Since K-Means was previously applied during the Elbow Method to determine the optimal number of clusters, the algorithm was performed again using different initial centroids. This process was carried to verify the clustering results from the previous stage. The randomly selected initial centroids used in the K-Means algorithm are shown in Table 4.

Table 4. K-Means Initial Centroids

No.	Country	FOB (US\$)	Weight (kg)
28	Germany	0,154264	0,183625

No.	Country	FOB (US\$)	Weight (kg)
39	Japan	0,008791	0,010124
76	Thailand	0,006067	0,001360
80	UAE	0,000017	0,000598

Based on the initial centroid, initial cluster assignment resulted in 4 members in Cluster 1, 15 members in Cluster 2, 6 members in Cluster 3, and 59 members in Cluster 4. Cluster membership is established by computing the distance from each data point to every centroid. Next, the average of export value and export weight for each cluster were calculated to obtain the new centroids. In this study, the process is repeated eighth times. After the eighth iteration the clusters stabilized and no further changes in cluster membership were observed. The final centroids coordinates obtained from the K-Means algorithm were C1 (1.000000, 1.000000), C2 (0.117143, 0.238351), C3 (0.064149, 0.068651), and C4 (0.002056, 0.002566), where X represent the normalized export value and Y represent the normalized export weight. In addition, the final clustering results consisted of 1 member in Cluster 1, 3 members in Cluster 2, 6 members in Cluster 3, and 74 members in Cluster 4. After repeating this process ten times using different initial centroids, the same clustering results were obtained in every run.

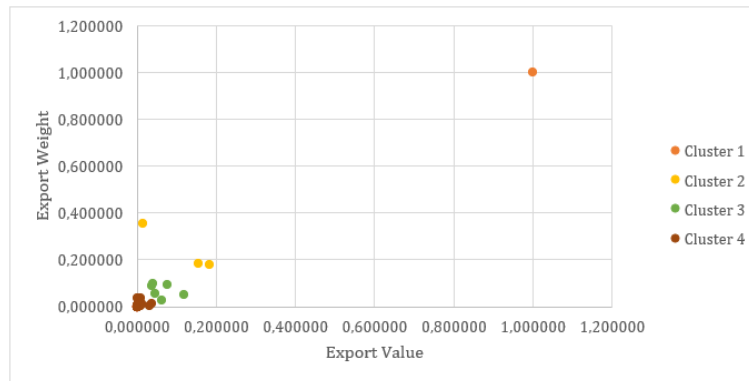


Figure 3. Scatter Plot of K-Means Clustering Results

Following the K-Means clustering process, K-Medoids clustering was applied using the same number of clusters, $k = 4$. The process began with the selection of initial medoids. The randomly selected initial medoids are shown in Table 5.

Table 5. K-Medoids Initial Medoids

No.	Country	FOB (US\$)	Weight (kg)
53	Nigeria	0,000306	0,002110
72	Sweden	0,000135	0,000030
73	Switzerland	0,000251	0,000407
76	Thailand	0,006067	0,001360

The initial cluster assignment resulted in 8 members in Cluster 1, 48 members in Cluster 2, 8 members in Cluster 3, and 20 members in Cluster 4. Similar to K-Means, in K-Medoids the cluster assignment is performed by calculating the distance from each data point to their every medoid. For K-Medoids, the next step is to calculate the total cost within each cluster and identify the medoid that produces the lowest total cost. The example of the total cost calculation is presented in Figure 4.

			Algeria	Cote Divoire	East Timor	Malaysia	Nigeria	Russia	Spain	Sudan	
		x	0,002583	0,000652	0,000433	0,014560	0,000306	0,002747	0,000035	0,000244	
Cluster 1	x	y	0,006702	0,003533	0,036873	0,354836	0,002110	0,024825	0,000040	0,003229	Total Cost
Algeria	0,002583	0,006702	0,000000	0,003711	0,030248	0,348340	0,005126	0,018124	0,007133	0,004187	0,416868
Cote Divoire	0,000652	0,003533	0,003711	0,000000	0,033341	0,351578	0,001465	0,021394	0,003548	0,000508	0,415545
East Timor	0,000433	0,036873	0,030248	0,033341	0,000000	0,318276	0,034764	0,012269	0,036836	0,033644	0,499378
Malaysia	0,014560	0,354836	0,348340	0,351578	0,318276	0,000000	0,353014	0,330223	0,355094	0,351898	2,408423
Nigeria	0,000306	0,002110	0,005126	0,001465	0,034764	0,353014	0,000000	0,022846	0,002088	0,001121	0,420424
Russia	0,002747	0,024825	0,018124	0,021394	0,012269	0,330223	0,022846	0,000000	0,024933	0,021740	0,451528
Spain	0,000035	0,000040	0,007133	0,003548	0,036836	0,355094	0,002088	0,024933	0,000000	0,003197	0,432827
Sudan	0,000244	0,003229	0,004187	0,000508	0,033644	0,351898	0,001121	0,021740	0,003197	0,000000	0,416295

Figure 4. Sample of Total Cost Calculation for Cluster 1

After determining the data points with the lowest total cost, that data point was selected as the new medoid for its cluster. Based on the example shown Figure 4, Cote d'Ivoire was selected as the new medoid. This process is then repeated until the cluster stabilized, where no changes in cluster membership were observed. In this study, the K-Medoids is repeated until fifth iteration, after which no changes in cluster membership were observed from the sixth iteration onward. The final medoids obtained from the K-Medoids algorithm were Japan (0.008791, 0.010124), Anguilla (0.000009, 0.000009), United Kingdom (0.000543, 0.000931), and Brazil (0.077149, 0.092852). In addition, the final clustering results consisted of 12 members in Cluster 1, 50 members in Cluster 2, 13 members in Cluster 3, and 9 members in Cluster 4.

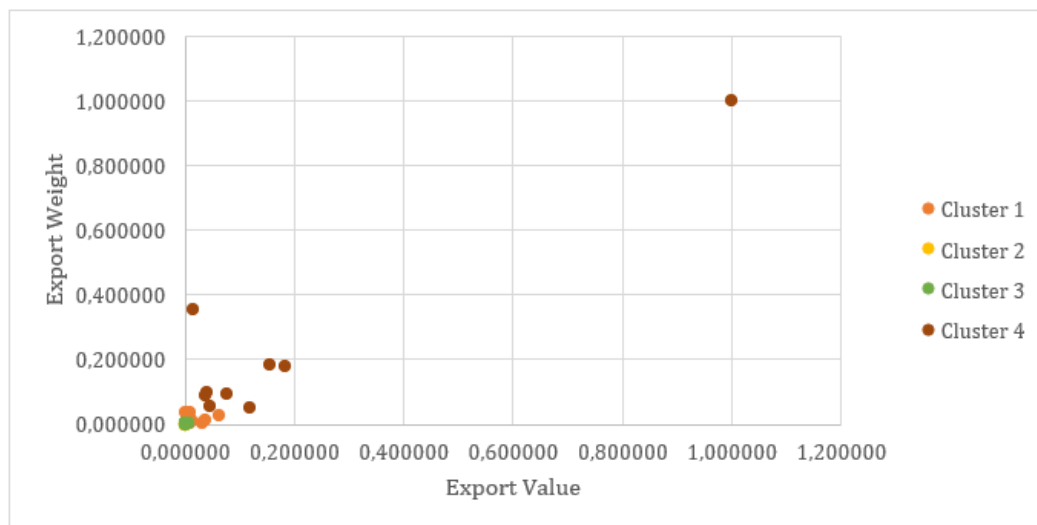


Figure 5. Scatter Plot of K-Medoids Clustering Results

3.3 Clustering Results Evaluation

After the clustering process, the results of both clustering algorithms were evaluated using the Davies-Bouldin Index and Silhouette Score. In this study, both evaluation methods were also calculated in Microsoft Excel. The evaluation results indicates that K-Means showed superior performance compared with K-Medoids based on both Davies-Bouldin Index and Silhouette Score. The results of both clustering algorithms are presented in Table 6.

Table 6. Results of DBI and Silhouette Score

Algorithm	Davies-Bouldin Index	Silhouette Score
K-Means	0,535698	0,845739
K-Medoids	2,186415	0,462290

3.4 Comparison with Orange Software

To validate the clustering results, the result of Min-Max Normalization was also analyzed using Orange Data Mining with identical clustering parameters, including the predefined number of clusters. The Davies-Bouldin Index and Silhouette Score obtained using Orange are presented in Table 7.

Table 7. Orange Data Mining Evaluation Results

Algorithm	Davies-Bouldin Index	Silhouette Score
K-Means	0,343404	0,857940
K-Medoids	1,166796	0,607108

The differences in the results can be attributed to the different cluster member distribution. K-Means produced 76 members in Cluster 1, 1 member in Cluster 2, 6 members in Cluster 3, and 1 member in Cluster 4, whereas K-Medoids produced 11 members in Cluster 1, 5 members in Cluster 2, 9 members in Cluster 3, and 59 members in Cluster 4. The observed differences are likely due to the random selection of initial centroids. In the Excel no outlier was selected as the initial centroids during the random initialization process.

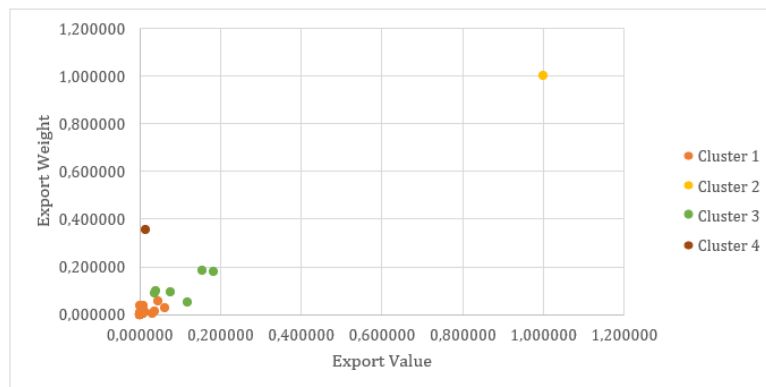


Figure 6. Scatter Plot of Orange K-Means Clustering Results

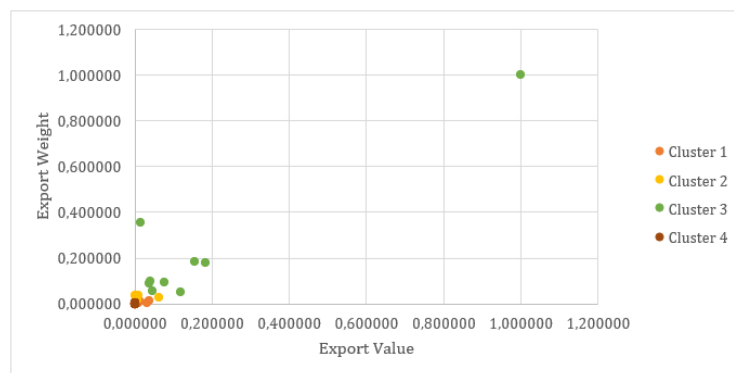


Figure 7. Scatter Plot of Orange K-Medoids Clustering Results

3.5 Results Analysis

Table 8 summarizes the cluster characteristics obtained using Excel and Orange Data Mining for both the K-Means and K-Medoids algorithms.

Table 8. Comparison of Cluster Categories Between Excel and Orange Data Mining

Algorithm	Platform	Very Low	Low	High	Very High
K-Means	Excel	C4	C3	C2	C1
K-Means	Orange	C1	C4	C3	C2
K-Medoids	Excel	C2	C3	C1	C4
K-Medoids	Orange	C4	CC1	C2	C3

Although the cluster label differed between the two implementations, the clusters were consistently categorized according to their export value and weight characteristics.

From the results obtained from Excel, there were a total of 33 different cluster assignments. In K-Medoids, more countries were assigned to the very high export and value weight cluster, while in K-Means, only Singapore was assigned in this cluster. This is suspected to be the cause of the poorer evaluation results obtained by the K-Medoids algorithm, as Singapore's and Malaysia's export data can be considered an outlier due to its substantially higher export value and weight compared to other countries. However, the other three clusters obtained from K-Medoids showed lower average distances than their corresponding clusters in K-Means.

In addition, the other suspected cause of the poorer K-Medoids evaluation results is the smaller distance between its medoids compared to the distance between centroids in K-Means, resulting in poorer Davies-Bouldin Index and Silhouette Score values.

Although Orange Data Mining produced better results from those obtained in Excel, both approaches consistently indicate that K-Means achieved better clustering performance than K-Medoids according to the Davies-Bouldin Index and the Silhouette Score Values in analyzing Indonesia's EIC export data. Furthermore, the members of each cluster were assigned according to their respective export value and weight characteristics.

The findings in this study are consistent with the results reported in [5], which showed that K-Means require less computational effort than K-Medoids. In this study, K-Means algorithm is also faster and easier to calculate manually in Microsoft Excel, requiring approximately 1 hour obtain the stable results, while K-Medoids required approximately 2-3 hours to complete. However, the findings differ from those reported in [6], which suggested that K-Medoids produces more stable clusters. In this study, K-Means algorithm achieved better performance than K-Medoids in analyzing Indonesia's EIC exports, as evaluated using the Davies-Bouldin Index and Silhouette Score

4. Conclusions and Future Works

The result of this study indicates that Indonesia's EIC exports are generally characterized by very low export value and export weight. This can be seen from the clustering result, where the majority of destination countries were classified into the very low export value and weight category by both algorithms. Specifically, 88% of countries assigned by K-Means and 60% of countries assigned by K-Medoids. The clustering results obtained using Orange also shows that majority of countries are assigned to very low export value and weight category. These cluster characteristics can assist policymakers in identifying priority export markets and designing market-specific export promotion strategies. In addition, exporters may use the clustering results to distinguish high-potential destination countries from the market with relatively low export performance.

This study has limitations that should be considered when interpreting the results. First, although the clustering process was repeated multiple times using different initial centroids and produced consistent results, the stability evaluation was limited to a small number of random initializations. Future studies may perform more extensive stability testing using a larger number of random initialization or different initialization strategies. Second, although Microsoft Excel was used as the primary tools for clustering

calculations, the computational process was performed using the original data values without rounding. The rounded values were only displayed for visualization purposes to improve readability. However, the use of spreadsheet-based computation may still present limitations in terms of scalability and automation compared to specialized data mining software. Third, this study relies on single national data source, namely Badan Pusat Statistik (BPS), which may limit the scope of analysis.

From this study, future studies may investigate the impact of outliers on clustering performances and clustering quality, particularly in cases where small number of countries exhibit substantially higher export values and weights than the rest of the dataset. In addition, other clustering method may be used to improve the identification clusters of export destination countries, determine priority markets, and support the development of export strategies.

5. Declarations

Author Contributions

Christian Andrew: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft; Rahmatina Hidayati: Methodology, Supervision, Validation, Writing – Review & Editing.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data used in this study are secondary data obtained from the official website of Badan Pusat Statistik (BPS). The data supporting the findings of this study are publicly available from BPS website.

Use of Artificial Intelligence (AI)

Generative artificial intelligence tools were used to assist with language translation and manuscript preparation. The author review and verified all AI-assisted content and remain full responsible for the final manuscript.

Ethical Clearance

This study used secondary statistical data and did not involve human participants, animals, or the collection of personal data. Therefore, ethical clearance from institutional ethic committee was not required. The study was conducted in accordance with applicable research and publication ethic.

6. References

- [1] S. Agustina *dkk.*, “PENGARUH EKSPOR TERHADAP PERTUMBUHAN EKONOMI DI INDONESIA,” *Jurnal Manajemen dan Ekonomi Kreatif*, vol. 1, no. 1, hlm. 113–126, 2023, doi: 10.59024/jumek.v1i1.31.
- [2] K. M dan D. R, “INTEGRATED CIRCUITS AND THEIR APPLICATIONS IN ELECTRONICS,” *American Journal of Applied Science and Technology*, vol. 4, no. 04, hlm. 24–27, Apr 2024, doi: 10.37547/ajast/Volume04Issue04-05.
- [3] E. Subiyantoro, A. R. Muslikh, M. Andarwati, G. Swalaganata, dan F. Y. Pamuji, “Analisis Pemilihan Media Promosi UMKM untuk Meningkatkan Volume Penjualan Menggunakan Metode Analytical Hierarchy Process (AHP),” *Jurnal Teknologi dan Manajemen Informatika*, vol. 8, no. 1, hlm. 1–8, Jul 2022, doi: 10.26905/jtmi.v8i1.6760.
- [4] A. Khalif, A. N. Hasanah, M. H. Ridwan, dan B. N. Sari, “Klasterisasi Tingkat Kemiskinan di Indonesia menggunakan Algoritma K-Means,” *Generation Journal*, vol. 8, no. 1, hlm. 54–62, Mar 2024, doi: 10.29407/gj.v8i1.21470.

- [5] R. Hidayati, L. Indana, M. Y. S. Putra, dan R. Z. Sasongko, "Analisis Cluster dengan K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Pembangunan TIK Tahun 2021-2022," *jtik*, vol. 8, no. 2, hlm. 405–411, Apr 2024, doi: 10.35870/jtik.v8i2.1815.
- [6] T. A. Wahyudi, A. Arifia, dan A. Adi Suryanto, "Application of Text Mining and K-Medoids for Job Clustering Based on Job Description Analysis," *Jurnal Teknologi dan Manajemen Informatika*, vol. 11, no. 1, hlm. 100–109, Jul 2025, doi: 10.26905/jtmi.v11i1.15863.
- [7] P. L. Arifa, H. D. Rahmasari, C. A. Aimandiga, A. Fitrianto, dan R. B. Yudhianto, "PERBANDINGAN K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKKAN KOMODITAS EKSPOR INDUSTRI DI INDONESIA," *MUST: Journal of Mathematics Education, Science and Technology*, vol. 10, no. 2, hlm. 37–56, Des 2025, doi: 10.30651/must.v10i2.28661.
- [8] R. W. Saputra, Nursobah, dan A. A. Khair, "Segmentation Analysis of Tourist Visits Using K-Means Algorithm at the Kutai Kartanegara Tourism Office," *SMATIKA JURNAL*, vol. 16, no. 02, hlm. 459–469, Jun 2026, doi: 10.32664/smatika.v16i02.2324.
- [9] S. G. K. Patro dan K. K. Sahu, "Normalization: A Preprocessing Stage," 19 Maret 2015, *arXiv*: arXiv:1503.06462. doi: 10.48550/arXiv.1503.06462.
- [10] S. García, J. Luengo, dan F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. dalam Intelligent Systems Reference Library, vol. 72. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-10247-4.
- [11] E. R. Arini, "Penerapan K-Means Cluster di Provinsi Jawa Timur Berdasarkan Ketahanan Pangan," *Journal of Science Nusantara*, vol. 3, no. 1, hlm. 32–36, Apr 2023, doi: 10.28926/jsnu.v3i1.866.
- [12] F. Anggraeny, A. Eviyanti, dan Sumarno, "Penerapan K-Means dengan Evaluasi Davies-Bouldin Index untuk Pengelompokan Kelas Unggulan SMP Wijaya Sukodono," *SMATIKA JURNAL*, vol. 15, no. 02, hlm. 372–382, Des 2025, doi: 10.32664/smatika.v15i02.1689.
- [13] M. A. Septianto, A. Faqih, dan A. R. Rinaldi, "KLAUSTERISASI DATA PRODUKSI PERTANIAN DI KABUPATEN CIREBON DENGAN ALGORITMA K-MEANS," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Art. no. 2, Apr 2025, doi: 10.23960/jitet.v13i2.6174.
- [14] D. Anggraini dan M. S. Hasibuan, "Initial Centroid Pada Algoritma K-Means Dan K-Medoids," *Jurnal Multimedia dan Teknologi Informasi (Jatilima)*, vol. 7, no. 03, hlm. 487–500, Agu 2025, doi: 10.54209/jatilima.v7i03.1615.
- [15] N. G. Syah, A. S. Fitriani, I. A. Kautsar, dan C. Taurusta, "Klasifikasi Kualitas Redaman Kabel Fiber Optik Berdasarkan Terminasi Menggunakan Algoritma K-Means," *SMATIKA JURNAL*, vol. 15, no. 02, hlm. 360–371, Des 2025, doi: 10.32664/smatika.v15i02.1899.
- [16] A. P. Putra, J. Tshivana, dan E. Rilvani, "PERBANDINGAN TEORITIS DAN EKSPERIMEN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM KLAUSTERISASI DATA," *Kohesi: Jurnal Sains dan Teknologi*, vol. 10, no. 2, hlm. 61–70, Agu 2025, doi: 10.2238/0b2z5035.
- [17] A. M. Ikotun, F. Habyarimana, dan A. E. Ezugwu, "Cluster validity indices for automatic clustering: A comprehensive review," *Heliyon*, vol. 11, no. 2, Jan 2025, doi: 10.1016/j.heliyon.2025.e41953.
- [18] P.-N. Tan, M. Steinbach, dan V. Kumar, *Introduction to Data Mining*. Pearson India, 2016.
- [19] N. K. Zuhal, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," *Seminar Nasional Teknologi & Sains*, vol. 1, no. 1, hlm. 200–205, Feb 2022, doi: 10.29407/stains.v1i1.1495.

- [20] A. Firdausiyah dan S. Widoyoningrum, "Implementation of K-Means Clustering with Attribute Adjustment and Cluster Validation in a Web-Based TPQ Student Assessment Information System," *SMATIKA JURNAL*, vol. 16, no. 02, hlm. 481–500, Jun 2026, doi: 10.32664/smatika.v16i02.2386.