Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia

Hazrul Anshari Ulvi[1)*], Muhammad Ikhsan[2)]
[1,2)] Computer Science Study Program, State Islamic University of North Sumatra, Indonesia
[1)]hazrul0701202100@uinsu.ac.id, [2)]mhd.ikhsan@uinsu.ac.id

Submitted : Jul 2, 2024 | **Accepted** : Jul 6 2024 | **Published** : Jul 9, 2024

Abstract: International relations affect the economic growth of each country, which can affect the economic growth of each country. As a result, global economic growth is necessary, which means that the global economy has a greater capacity to produce goods and services. Exports and imports are very important to drive economic growth. but if exports and imports are not balanced, it will have a bad impact if the value of imports is greater than exports, export prices abroad will definitely fall. An analysis comparing export and import categories is needed to determine which goods are most imported and exported in Indonesia in 2021-2023. This study uses a quantitative methodology and machine learning methods, namely k-means and k-medoids algorithms. These two methods will be compared to determine which is the most effective for export and import data of goods in Indonesia in 2021-2023. The results of the study were obtained by K-Means more effectively in handling data on the grouping of exports and imports of goods in Indonesia in 2021-2023. The dataset shows the results of the evaluation of K-Means using DBI of 0.59, while the results of the evaluation using K-Medoids show a result of 1.7868. Because the evaluation value of K-Means has low computing performance compared to K-Medoids. The largest amount of the value and weight of exports and imports of goods in Indonesia is in C1 where in the HS code [27], namely Mineral fuels with a total export value of goods in 2021 to 2023 of 134,999,470,522 US$ and a total import value of 113,714,568,740 US$. Meanwhile, the total export weight of goods from 2021 to 2023 in mineral fuel goods is 1,505,006,250,327 Kg or around 1,658,985,413 tons and the total import weight is 186,446,782,134 Kg or around 205,522,397 tons.

Keywords: Export; Import; K-Means; K-Medoids; DBI

INTRODUCTION

The development of a country's economy is inseparable from global economic conditions. Global economic relations affect the economic growth of each country. This condition causes competitiveness as one of the decisive factors in competition between countries in order to benefit from the openness of the world economy (Hanifah, 2022). Therefore, global economic growth is needed which is an increase in the ability of the economy around the world to produce goods and services. Increasing the export of goods and services allows a country to accelerate its economic growth.

In international trade export and import activities are one of the most important factors to increase economic growth. Export and import activities provide benefits for a country that participates in it. Exports are one of the sources of foreign exchange that are urgently needed by countries whose economies are open because exports can work widely in various countries and allow an increase in the amount of production that encourages economic growth, while through imports the country can meet its

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

domestic needs that cannot be produced domestically so that the costs incurred for a product of goods and services will be cheaper (Fauziah & Khoerulloh, 2020).

Indonesia exports foodstuffs and production materials thanks to its abundance of natural resources, which opens up opportunities in global trade. Meanwhile, citizens' needs for products that are not made domestically are met through imports. Export and import activities cannot be separated from a country, because basically no country can stand alone without relations and cooperation between countries. However, export and import activities have a bad impact if the imbalance between exports and imports is not balanced. The value of imports that are higher than exports will certainly affecBt the selling value of exports carried out to experience a decrease in the selling value of foreign countries. For this reason, it is necessary to have an analysis to compare the grouping of exports and imports in Indonesia (Iskandar et al., 2020). In 2021, exports and imports of Indonesian goods decreased, with a total export value of 231,522 million dollars and imports of 196,190 million dollars. This shows that Indonesia has a trade deficit, which means the value of imports is greater than the value of exports. Apart from that, product distribution also showed a significant decline in the number of goods imported from Indonesia. However, data from the Central Statistics Agency (BPS) shows that the value of Indonesia's exports and imports in 2022 will increase by 26.07% on an annual basis, reaching a record high in the last ten years. Indonesia's export value will reach US$291.97 billion in 2022, a jump from US$231.6 billion in the previous year. Indonesia's exports will decrease by 4.26% annually and imports by 6.18% in 2023. Exports of non-oil and gas goods, such as mineral fuels, fats, animal and vegetable oils, as well as electrical machinery and equipment, as well as oil and gas exports, such as natural gas, Oil yields, and crude oil, experienced this decline. Furthermore, prices of Indonesia's main commodities, such as coal and palm oil, fell drastically, which had an impact on exports and imports of goods. Export and import conditions in 2021-2023 in Indonesia experienced drastic ups and downs (BPS, 2023).

In Wahyudi research it shows that results of his research where the DBI K-Means Clustering value is 0.094 and K-Medoids is 0.072 so that the K-Medoids algorithm is better than the K-Means algorithm in classifying fresh milk production in Indonesia. This is because K-Medoids has lower computing performance compared to K-Means (Wahyudi & Pujiastuti, 2022).

Based on this, this research requires a data mining process. The way to group it is by clustering. The clustering algorithm used is k-means and k-medoids. The goal is that the resulting partitions are less sensitive because the dataset has extreme values so that the use of medoids is not based on the observed average of each cluster (Luchia et al., 2022). The technique used in this study is Euclidean distance. Euclidean distance describes the degree of similarity between two or more by calculating the value of the Euclidean distance, whereas the distance gets closer, it indicates objects with the same group (Nuraini, 2022). And the limitation in this study is to arrive at a performance where to calculate performance is by DBI (Davies Bouldin Index) technique. The dataset used is a dataset obtained from the Central Statistics Agency (BPS) data on export and import data from 2021 to 2023.

LITERATURE REVIEW

International trade is an activity of exchange or even the activity of buying and selling that occurs between countries as one of the efforts to obtain benefits and profits from these activities (Trilarasati et al., 2023). International trade or international business is carried out through sale and purchase agreements, called "exports and imports". Selling activities are called "exports", and buying activities are called "imports" in this sale and purchase agreement, export is an activity or activity of exporting products and goods from the country to abroad by following regulatory standards and applicable provisions. Export activities are generally carried out by a country that is able to produce large quantities of goods and this amount has been fulfilled domestically (Tri Sugiarti Ramadhan, Nanik Wahyuningtiyas, 2023). Import activities are the process of importing goods or services from other countries into the domestic territory of a country. This process is carried out by individuals, companies, or governments who buy products from abroad for various purposes, from personal consumption to industrial use.

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases (Tinendung & Zufria, 2023). Data mining is also called an analytical process designed to

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

examine large amounts of data in search of valuable and consistent hidden knowledge. The purpose of data mining is to look for trends or patterns that are desired in a large database to help in future decision-making. The branch of artificial intelligence (AI) known as machine learning (ML) concentrates on creating and developing statistical algorithms that can learn from data and apply this knowledge to perform tasks without explicit instructions. This allows computer systems to improve their performance based on the data they consume. There are several machine learning techniques, namely Supervised Learning and Unsupervised Learning. In supervised learning, directed learning can be in the form of classification or regression (Septhiani & Hendry, 2023). Unsupervised learning analyzes and groups unlabeled datasets using machine learning algorithms.

Data preprocessing in data mining is a critical process that involves cleaning, transforming, and integrating data to prepare the data for analysis. The purpose of data preprocessing is to improve the quality of data and make it more suitable for specific data mining tasks. Where this study uses the Standard Scaler technique which is a preprocessing method where the method will standardize features by removing the average and scaling the variant units. The process of calculating pre-processing on data such as tables uses the standard scaler formula as follows:

$$Z_i = \frac{x_i - \mu}{\sigma} \qquad (1)$$

Where:
Zi : The standard score or z-score of the ith data
Xi : The original data value of the ith data
μ : The mean of all data
σ : Standard deviation from the entire data

A clustering algorithm is a collection of data processing techniques used to group a dataset or object into groups (or clusters) based on their similarity in characteristics or features. Clustering, or partitioning data into similar groups with different items, is a core technique for data analysis (Ghadiri et al., 2021).

The elbow method is one of the commonly used methods to determine the best number of clusters in clustering. Analysis of the elbow method in determining the best number of clusters by looking at the shape of the elbow in the resulting curve, which is the comparison between the number of clusters that will form an elbow at a certain point in the curve (Orisa & Faisol, 2021).

K-Means clustering is a cluster analysis method that aims to break down objects into k clusters and then observed where each cluster object is obtained through the nearest average. The K-Means algorithm is also a very basic and widely used approach, which has sparked great research interest in overcoming various challenges, improving limitations, and exploring new adaptations or developments of the algorithm. The problems studied include, but are not limited to, handling high-dimensional datasets, clustering data with complex cluster shapes, determining the optimal number of clusters, and sensitivity to early centroid selection (Xia et al., 2022). To determine the center point of the cluster on the data is done using equation (2)

$$\text{Ckj} = \frac{x_1 j + x_2 j + \cdots + x_n j}{n} \qquad (2)$$

Where :
Ckj : Center Cluster k on variable j (j = 1, 2,...,p)
n : A lot of data on cluster k
And to update the centroid point value in equation (3)

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \qquad (3)$$

Where:
μ_k : Centroid point of the Kth cluster
Nk : The amount of data in the K cluster
xi : Data I in cluster K

The K-Medoids algorithm or Partitioning Around Medoids (PAM) is a clustering method that uses representations (medoids) to divide the set (n) of several objects into (k) clusters. The advantage of K-

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

Medoid is that it overcomes the weakness of K-Means and the results of the clustering process do not depend on the order in which the dataset is entered. The weakness of this algorithm is that it determines the medoid is random, so I don't know if the similarity is high or not, and it has a great effect on the cluster and this algorithm can run well for a small number of datasets because the computation takes a long time and results in complicated calculations. To allocate data into the closest cluster, the Euclidean Distance measure with equation (4) is used.

$$D = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

Where:
D : Euclidean Distance
Xi : First data attribute
Yi : Second data attribute

DBI (Davies Bouldin index) is an internal assessment framework, at the validation or performance stage assessing the advantages of clustering that has been carried out using the number and features in the dataset. The DBI approach has the goal of maximizing the distance between clusters and different clusters and trying to minimize the distance between objects in the cluster. The lower the DBI value (non-negative ≥ 0), the better the cluster obtained from the clustering grouping that has been used. Formula for DBI in equations (6)

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(Ri, j, \dots k) \quad (6)$$

Where:
DBI: Scalar value of DBI
Ri,j : Ratio of the number of clusters
K : Number of clusters used

Python is a programming language that uses interpreters to execute its program code. The interpreter can translate the code directly. Python adopts the programming paradigm of several other languages. This combination of paradigms makes it easier for programmers to develop various projects using Python (Rahman et al., 2023). The use of Python is mainly aimed at data analysis, data visualization, and AI development and creation (Ikhsan et al., 2022). In this study, the tools used are jupyter notebooks. Many libraries are used such as numpy, pandas, matplotlib.pyplot and so on.

METHOD

This study uses a quantitative research methodology which is an approach in research that emphasizes the collection and analysis of numerical data to find causal and statistical relationships between variables. It involves measuring phenomena using numbers and statistics, as well as using statistical techniques for data analysis. The stages of the research are shown in the following figure.

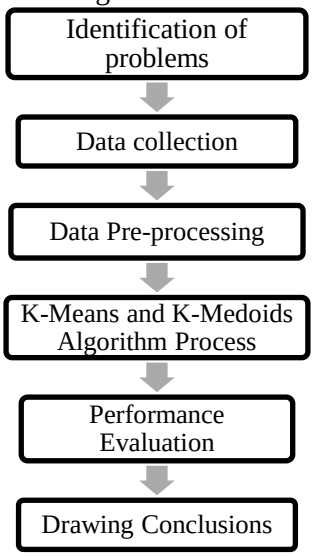

Figure 3.1 Research Stages

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

Identify the problem, namely if there is an imbalance between exports and imports. The value of imports that are higher than exports will certainly affect the selling value of exports carried out to experience a decrease in the selling value of foreign countries. The imbalance in the value of exports and imports has a great impact on entrepreneurs in Indonesia. To address these issues, clustering can help in optimizing export and import strategies, such as determining which goods are most popular and being able to understand specific market dynamics.

The data used is a dataset obtained from the Central Statistics Agency (BPS) with the file name "import and export data" and the data is taken from the internet and accessed from https://www.bps.go.id/id/exim links consisting of export and import information on goods in Indonesia. The goods used in this study have a two-digit HS code, where the HS Code in export and import data in Indonesia stands for Harmonized System (HS). HS is an international standard classification system used to categorize and identify internationally traded products.

The process of data preparation is known as data pre-processing. In this process, 98 items with the period 2021 - 2023 were implemented using Jupyter Notebook and the programming language used was python to pre-process data. The data pre-processing process is carried out by transforming data, namely by normalizing data. The data normalization technique is used on export and import value and weight labels in 2021-2023 using StandardScaler.

At this stage, data that has gone through the data pre-processing stage will be tested using the K-Means and K-Medoids algorithms with the help of Jupyter Notebook. The data was first normalized using python tools before processing and the algorithm used for data grouping in this study was the K-Means and K-Medoids algorithms.

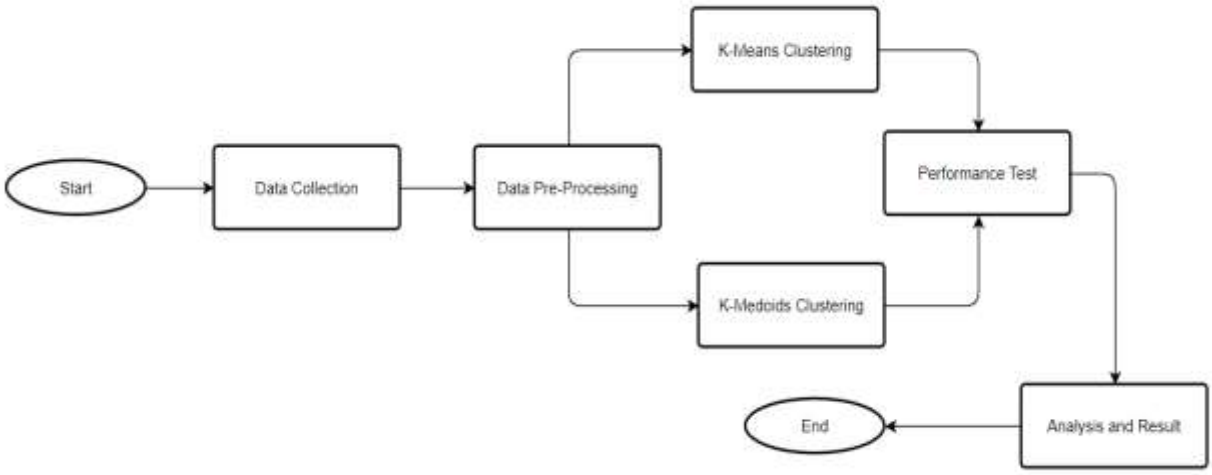

Figure 3.2 Research Flowchart

The performance evaluation process of clustering results with the K-Means and K-Medoids algorithms uses the Davies-Bouldin Index (DBI) method. And after getting the performance evaluation of the clustering results with the K-Means and K-Medoids algorithms, the next stage is to draw conclusions from the research conducted.

RESULT

This study aims to determine the quality of the two clustering algorithms, namely K-Means and K-Medoids in the grouping of exports and imports of goods in Indonesia, which algorithm is suitable in the dataset of grouping exports and imports of goods in Indonesia in 2021-2023. The dataset was obtained from BPS (Central Statistics Agency) which consists of information on the export and import of goods in Indonesia on a scale of 2021-2023 and a dataset of 1176 data with categories of 98 goods.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

Table 1 . Dataset of Export and Import of Goods 2021-2023

HS Code and Description	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	Import Weight (KG) 2021	Export Value (USD) 2022	…	Import Weight (KG) 2023
[01] Live Animals	63934122.92	25226699.14	572021619	146767208	61863070.85	…	127637381
[02] Animal Meat	20022587.48	2414899.53	965343727	276762560	15692101.92	…	312988066
[03] Fish, crustaceans and mollusks	3707404752	787473197.31	333392815	129149962	4009098550.74	…	218261127
[04] Buttermilk and eggs	581420780.27	38637079.56	1394289968	496672929	662276440.76	…	470674762
[05] Animal products	31218065.58	11756620.36	141119915	60103032	37522917.33	…	6118386 5
…	…	…	…	…	…	…	…
[99] Software, digital goods and consignments	2935607412.31	635945923.73	8168584665.02	1744777005.11	25050181.65	…	171124

In the Exports and Imports of Goods in Indonesia dataset, the next is the pre-processing process. The technique used is a standard scaler which is used in export and import data for goods from 2021 to 2023.

Table 2 . Data Pre-Processing

No.	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	Import Weight (KG) 2021	Export Value (USD) 2022	…	Import Weight (KG) 2023
0	-0.38187	-0135631	-0.30711	-0.26340	-0.34748	…	-0.254951
1	-0.38906	-0.13612	-0.22264	-0.24286	-0.35289	…	-0.229982
2	0.21541	-0.11928	-0.35937	-0.26618	0.11432	…	-0.242743
3	-0.29703	-0.13534	-0.13051	-0.20811	-0.27724	…	-0.208740
No.	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	Import Weight (KG) 2021	Export Value (USD) 2022	..	Import Weight (KG) 2023
4	-0.38723	-0.13592	-0.39966	-0.27709	-0.35033	…	-0.263904
…	…	…	…	…	…	…	…
97	0.08888	-0.12253	-0.41866	-0.28658	0.60096	…	-0.272163

The process of calculating pre-processing on data such as tables uses the standard scaler formula as follows:

$$Z_i = \frac{x_i - \mu}{\sigma}$$

Where initially it is to calculate the average of each column with an example column "Export Value (USD) 2021", namely:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

$$\mu = \frac{63.934.123 + 20.022.587 + 3.707.404.752 + 581.420.780 + \cdots + 2.935.607.412}{98}$$

μ = 2.393.382.332

Next is to calculate the variance (S2):

$$S^2 = \frac{(63.934.123 - 2.393.382.332) + (20.022.587 - 2.393.382.332) + \cdots + (2.935.607.412 - 2.393.382.332)}{98}$$

S2 = 6.100.104.037

The last step is to calculate the Standard Scaler

Xi =63.934.123

$$Z_i = \frac{63.934.123 - 2.393.382.332}{6.100.104.037} = -0.381870$$

Xi = 20.022.587

$$Z_i = \frac{20.022.587 - 2.393.382.332}{6.100.104.037} = -0.38906$$

Xi = 3.707.404.752

$$Z_i = \frac{3.707.404.752 - 2.393.382.332}{6.100.104.037} = 0.21541$$

The calculation process is carried out continuously until "Import Weight (KG) 2023" and the results are in table 4.2 above. Before the K-Means testing process, first to determine the best k, namely by analyzing the elbow method which uses WCSS (Within-Cluster sum of Squares) from K=1 to K=10.

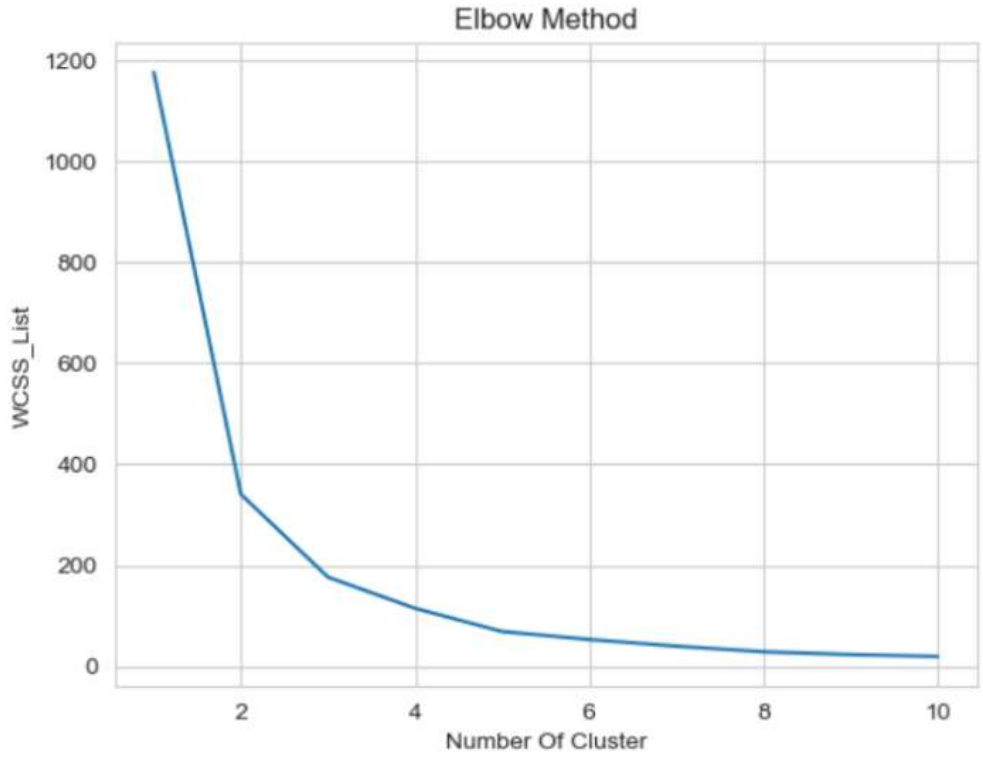

Figure 4.1 Elbow Method To Determine The Best K

Figure 4.1 above shows that a significant decrease in WCSS occurred at K=1 to K=3. In the figure, it can be concluded that the optimal K is at K=3, because the decrease in WCSS becomes more sluggish

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

which shows that the addition of further clusters does not result in a significant decrease in WCSS. Next is the test on K-Means, the visualization for the clustering model is shown in figure 4.2.

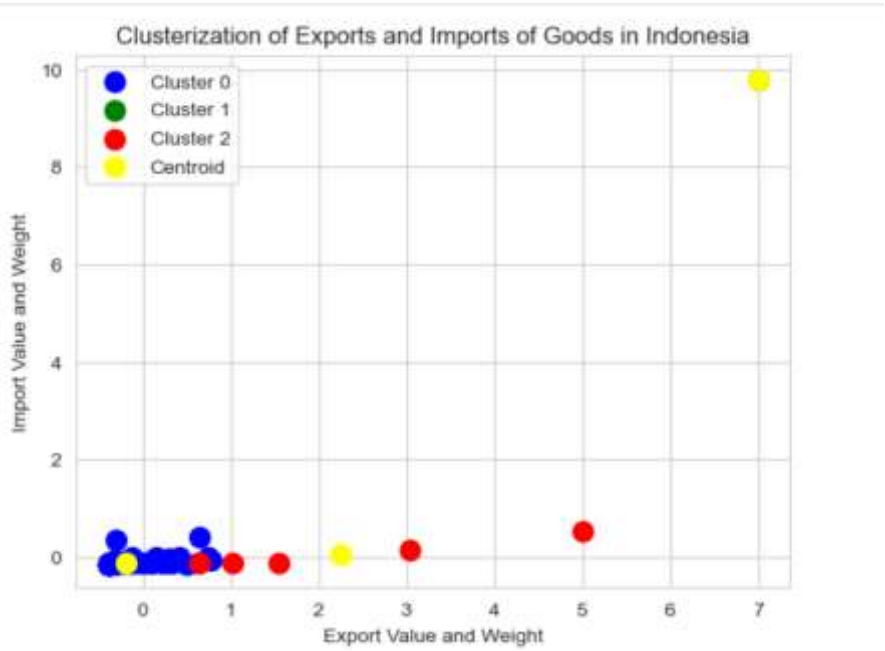

Figure 4.2 Clustering Visualization

Figure 4.2 shows that there are 3 clusters, namely cluster 0 in blue, cluster 1 in green with the highest value and weight, cluster 2 in red and centroid point in yellow. Here's the calculation to find K-Means:

a. The number of clusters used is 3, namely low (C0), high (C1) and medium (C2) clusters.
b. Centroid points are randomly selected from the dataset. The center of the cluster can be known from the following table 3:

Table 3
Centroid Early Data

Cluster	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	…	Import Weight (KG) 2023
[03] Fish, crustaceans, and mollusks (C0)	3707404752	787473197	333392815	…	218261127
[27] Mineral fuels (C1)	45078570715	463131109581	28840088962	…	68524768595
[54] Artificial filament (C2)	687285600	242282216	1638860854	…	545374317

Calculating the distance of each data using Euclidean distance

$$Data_{(0,1)} = \sqrt{(63934123 - 370740452)^2 + (25226699 - 787473197)^2 + (61863071 - 333392815)^2 + \cdots + (127637381 - 218261127)^2}$$
$$= 6593508594$$

$$Data_{(0,2)} = \sqrt{(20022587 - 45078570715)^2 + (787473197 - 787473197)^2 + (15692102 - 333392815)^2 + \cdots + (312988066 - 218261127)^2}$$
$$= 88625E+11$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

$$Data_{(2,99)} = \sqrt{(2935607412 - 687285600)^2 + (635945924 - 242282216)^2 + (8168584665 - 1638860854)^2 + \cdots + (171124 - 545374317)^2}$$

= 2166201220

The results of the first iteration can be seen in table 4 below.

Table 4 . Results of the Calculation of the Center Distance of the First Iteration Cluster

HS Code and Description	Distance to Centroid			Nearby
	C0	C1	C2	
[01] Live animals	6593508594	88625E+11	2498033304	C2
HS Code and Description	Distance to Centroid			Nearby
	C0	C1	C2	
[02] Animal meat	885114E+11	886183E+11	1772239105	C2
[03] Fish, crustaceans, and mollusks	2419008748	88428E+11	5850878512	C0
[04] Milk, butter, and eggs	3252934678	885889E+11	533677555,5	C2
[05] Animal products	6685481919	8863117E+11	3091882893	C2
…	…	…	…	…
[99] Software, digital goods and consignments	5627902160	886257E+11	2166201220	C2

Next, determine the new centroid for the next iteration by calculating the mean for each cluster.

Table 5 . Centroid New Data

Cluster	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	…	Import Weight (KG) 2023
[03] Fish, crustaceans, and mollusks (C0)	4241010082	4839461304	3138642435	…	3526448662
[27] Mineral fuels (C1)	45078570715	463131109581	28840088962	…	68524768595
[54] Artificial filament (C2)	1158995080	5303358184,5	1234497628	…	573504112,3

Repeat steps 3-4 until the data position has not changed. So that the final result of the cluster position is shown in table 6.

Table 6 .Results of K-Means Clustering

HS Code and Description	Cluster
[01] Live animals	0
[02] Animal meat	0
[03] Fish, crustaceans, and mollusks	0
[04] Milk, butter, and eggs	0
[05] Animal products	0

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

[06] Live trees and cut flowers	0
[07] Vegetables	0
HS Code and Description	Cluster
[08] Fruits	0
[09] Coffee, tea, and spices	0
…	…
[99] Software, digital goods and consignments	0

Next, calculate the performance of K-Means using DBI (Davies Bouldin Index).

Use of DBI for Cluster Evaluation

```
[8]: dbi = davies_bouldin_score(data_skalered,kmeans.labels_)

print(f"Indeks Davies-Bouldin: {dbi:.2f}")

Indeks Davies-Bouldin: 0.59
```

Figure 4.3 K-Means Clustering Performance

In the K-Means Algorithm, the performance or evaluation obtained from the DBI results is 0.59. Next is the test on K-Medoids, just like the K-Means process, the optimal K value used is K=3 and the K-Medoids calculation process. First, initialize the cluster center as many as 3 clusters from the dataset. For the selection of each medoid is randomly selected. And normalize the data.

Table 7 . Random Selection of Early Medoids and Data Normalization

Cluster	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	…	Import Weight (KG) 2023
[02] Animal meat (C0)	0.00044	0.000	0.033	…	0.005
[27] Mineral fuels (C1)	1	1	1	…	1
[52] Cotton (C2)	0.01924	0.0001	0.065	…	0.007

Table 8 . Normalization of Export and Import of Goods Medoids Dataset

HS Code and Description	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	Import Weight (KG) 2021	Export Value (USD) 2022	…	Import Weight (KG) 2023
[01] Live Animals	0.001	0.000	0.020	0.003	0.001	…	0.002
[2] Animal Meat	0.00044	0.000	0.033	0.005	0.00021	…	0.005
[03] Fish, crustaceans, and mollusks	0.082	0.002	0.012	0.002	0.056	…	0.003
[04] Milk, butter, and eggs	0.013	0.00083	0.048	0.009	0.009	…	0.007

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

HS Code and Description	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	Import Weight (KG) 2021	Export Value (USD) 2022	…	Import Weight (KG) 2023
[05] Animal products	0.001	0.00025	0.005	0.001	0.001	…	0.001
…	…	…	…	…	…	…	…
[98] Software, digital goods and consignment	0.065	0.001	0.002	0.000	0.115	…	0.000

The formula for finding normalization is:

$$Normalisasi\ data = \frac{data\ ke\ n - min}{max - min}$$

$$Normalisasi\ data_1 = \frac{63934123 - 38272.91}{45078570715 - 38272.91} = 0.001$$

$$Normalisasi\ data_2 = \frac{20022587 - 38272.91}{45078570715 - 38272.91} = 0.00044$$

$$Normalisasi\ data_3 = \frac{3707404752 - 38272.91}{45078570715 - 38272.91} = 0.082$$

Then calculate the distance value with the Euclidean Distance equation shown in the equation on the K-Means algorithm. The following is the result of the distance calculation in the 1st iteration of the K-Medoids algorithm.

Table 9 . Results of the 1st Iteration of K-Medoids Calculation

HS Code	Cost 1	Cost 2	Cost 3	Closeness	Cluster
[1]	0.0245	3.4498	0.0701	0.0245	0
[2]	0	3.4363	0.0489	0	0
[3]	0.12	3.4302	0.1195	0.1195	2
[4]	0.0319	3.4187	0.0208	0.0208	2
[5]	0.0428	3.4602	0.0882	0.0428	0
[…]	…	…	…	…	…
[98]	0.138	3.4077	0.1429	0.138	0
Total Cost Simpangan				**14.014374**	

After obtaining the results of the distance of each medoids object in the 1st iteration, then proceed to the 2nd iteration. The new medoids candidates in the 2nd iteration can be seen in the following table 10:

Table 10 . New Medoids Candidates

Cluster	Export Value (USD) 2021	Export Weight (KG) 2021	Import Value (USD) 2021	…	Import Weight (KG) 2023
[11] Milling Results (C0)	0.005	0.001	0.007	…	0.004
[21] Various processed foods (C1)	0.032	0.001	0.035	…	0.004
[52] Musical instruments and parts (C2)	0.017	0.00013	0.007	…	0.00015

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

The results of the calculation of the distance of the new medoids points to the overall data are as shown in table 11.

Table 11 . Results of Distance Calculation to Each New Non-Medoids

HS Code	Cost 1	Cost 2	Cost 3	Closeness	Cluster
[1]	0.0179	0.0534	0.027	0.0179	0
[2]	0.0378	0.0474	0.0462	0.0378	0
[3]	0.112	0.0761	0.0947	0.0761	1
[4]	0.0631	0.0366	0.0657	0.0366	2
[5]	0.0144	0.0649	0.0221	0.0144	0
[…]	…	…	…	…	…
[98]	0.0286	0.0521	0.0367	0.0287	1
Total Cost Simpangan				**14.435253**	

Then the calculation of the total deviation (S) is carried out which is calculated from the calculation of the total cost (shortest distance) of the 1st iteration and the 2nd iteration.
S = total new cost – total cost old

Total Deviation (S) = Total Distance of the 2nd Iteration - Total Distance of the 1st Iteration
= 14.435253 - 14.014374
= 0.420879

Because the S value > 0, the clustering process was stopped. The following is the final result of the division of members of each cluster using the K-Medoids algorithm.

Table 12. Results of K-Medoids Clustering

HS Code and Description	Cluster
[01] Live animals	0
[02] Animal meat	0
[03] Fish, crustaceans, and mollusks	1
[04] Milk, butter, and eggs	2
[05] Animal products	0
[06] Live trees and cut flowers	0
[07] Vegetables	0
[08] Fruits	2
[09] Coffee, tea, and spices	0
…	…
[99] Software, digital goods and consignments	1

The DBI (Davies Bouldin Index) results from the clustering results of K-Medoids are shown in figure 4.4 below with a value of 1.7868.

```
CLuster Evaluation Using DBI on K-Medoids

[7]: dbi = davies_bouldin_score(data, kmedoids.labels_)
     print('Davies-Bouldin Index:', dbi)

Davies-Bouldin Index: 1.7868366567285963
```

Figure 4.4 DBI Performance Using K-Medoids Algorithm

From the two methods, the results of cluster evaluation on K-Means were 0.59 and K-Medoids were 1.7868. Therefore, it can be concluded that the best algorithm for grouping exports and imports of goods in Indonesia in 2021 – 2023 is to use the K-Means algorithm because the results of DBI or the results of the Cluster evaluation on K-Means are closer to 0 compared to K-Medoids.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13815
e-ISSN : 2541-2019
p-ISSN : 2541-044X

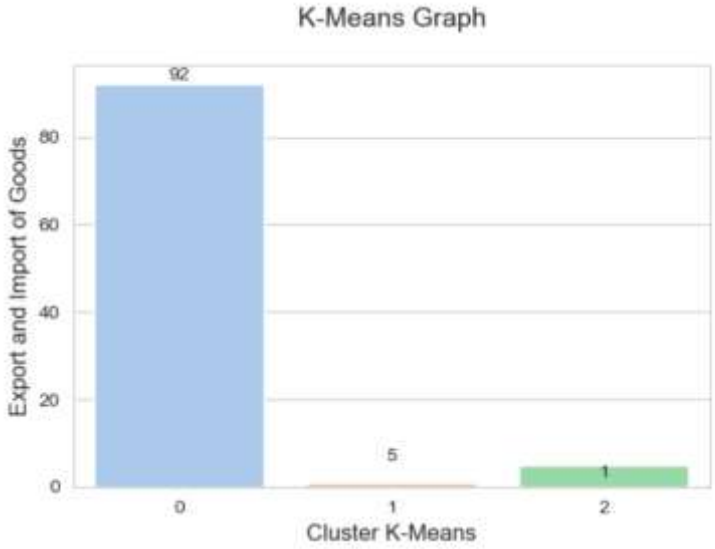

Figure 4.5 K-Means Clustering Graph

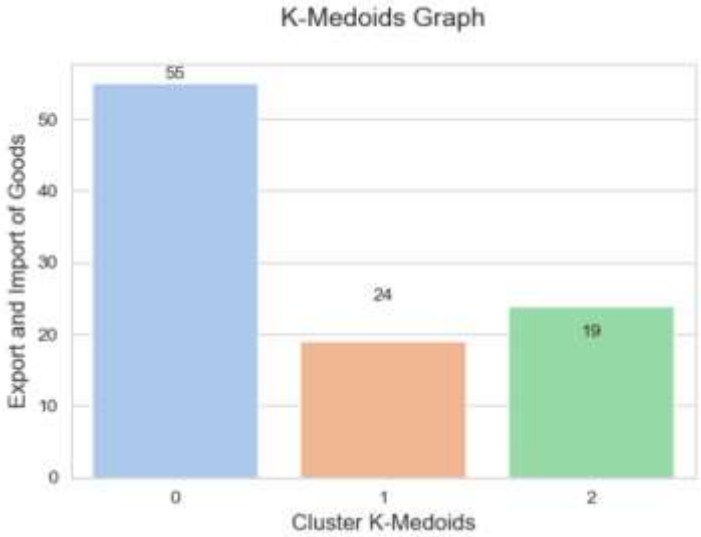

Figure 4.6 K-Medoids Clustering Graph

Figure 4.5 shows data clustering based on the K-Means algorithm where the first class or Cluster 0 has 92 data, Cluster 1 has 1 data and Cluster 2 has 5 data where the total of all data is 98 data.

Table 13.Number of K-Means Clusters

Cluster	Amount of Data
Cluster 0	92
Cluster 1	1
Cluster 2	5

Figure 4.6 shows data clustering based on the K-Medoids algorithm where the first class or Cluster 0 has 55 data, Cluster 1 has 19 data and Cluster 2 has 24 data where the total data is 98 data.

Table 14 .Number of K-Medoids Clusters

Cluster	Amount of Data
Cluster 0	55
Cluster 1	19
Cluster 2	24

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13815 p-ISSN : 2541-044X

DISCUSSIONS

From Wahyudi's research with a dataset of fresh milk production in Indonesia, it was found that k-medoids was more effective than k-means where the DBI value from K-Means was 0.094 and the DBI from K-medoids was 0.072. The fresh milk production dataset was taken from data from the Indonesian Central Statistics Agency regarding fresh milk production data in Indonesia from 2018 to 2020 with a total of 34 datasets according to provinces in Indonesia. Meanwhile, in this research, using the Export and Import of goods dataset obtained from BPS, it was found that the K-Means algorithm was more effective compared to K-Medoids because the computational performance value was lower than K-Medoids, where the DBI value of K-Means was 0.59 and DBI from K-medoids is 1.7868 with a total of 98 items and 1176 data on the value and weight of exports and imports of goods in Indonesia in 2021-2023.

By clustering it is found that the value and weight of exports and imports of goods in Indonesia is in the 1st cluster in the K-Means algorithm, namely mineral fuel with HS code = 27 with a total value of exports of goods in 2021 to 2023 of 134,999,470,522 US$ and the total import value was 113,714,568,740 US$. Meanwhile, the total weight of exported goods from 2021 to 2023 for mineral fuel goods is 1,505,006,250,327 Kg or around 1,658,985,413 tons and the total weight for imports is 186,446,782,134 Kg or around 205,522,397 tons.

The clustering results also show significant differences in the grouping of the two algorithms, where the K-Means algorithm has cluster 0 with 92 data, cluster 1 with 1 data and cluster 2 with 5 data. And the K-Medoids algorithm has cluster 0 with 55 data, cluster 1 has 19 data and cluster 2 has 24 data.

CONCLUSION

Based on the results of the research conducted, it can be seen that the two algorithms, namely K-Means and K-Medoids, are more effective in handling data on grouping exports and imports of goods in Indonesia in 2021-2023. The dataset shows that the K-Means evaluation results using DBI are 0.59, while the evaluation results using K-Medoids show results of 1.7868. This is because the evaluation value of K-Means has lower computing performance than K-Medoids. The lower the DBI value (non-negative ≥ 0), the better the cluster obtained from the clustering that has been used.

This dataset shows the results of data clustering based on the K-Means algorithm where the first class or Cluster 0 has 92 data, Cluster 1 has 1 data and Cluster 2 has 5 data. Meanwhile, the data clustering results are based on the K-Medoids algorithm where the first class or Cluster 0 has 55 data, Cluster 1 has 19 data and Cluster 2 has 24 data, where the total data is 98 data.

And it can be concluded that the largest value and weight of export and import goods in Indonesia are in Cluster -1 (K-Means algorithm) where mineral fuels with HS code = 27. So it is hoped that this research can be a reference for the government so that it can improve the Indonesian economy. become better by making exports to other countries higher than imports or can continue to increase from year to year and never decline.

REFERENCES

BPS. (2023). Perkembangan Ekspor dan Impor Indonesia. *Berita Resmi Statistik No. 30/04/Th. XXII, 15 April 2019*, *64*, 1–8.

Fauziah, E. S., & Khoerulloh, A. K. (2020). Pengaruh Ekspor dan Impor Terhadap Pertumbuhan Ekonomi dengan Kurs Sebagai Variabel Intervening. *Khazanah Sosial*, *2*(1), 15–24. https://doi.org/10.15575/ks.v2i1.8191

Ghadiri, M., Samadi, S., & Vempala, S. (2021). Socially fair k-means clustering. *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 438–448. https://doi.org/10.1145/3442188.3445906

Hanifah, U. (2022). Pengaruh Ekspor Dan Impor Terhadap Pertumbuhan Ekonomi Di Indonesia. *Transekonomika: Akuntansi, Bisnis Dan Keuangan*, *2*(6), 107–126. https://doi.org/10.55047/transekonomika.v2i6.275

*name of corresponding author

Ikhsan, M., Armansyah, A., & Tamba, A. A. (2022). Implementasi Jaringan Syaraf Tiruan Backpropagation Pada Klasifikasi Grade Teh Hitam. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 387. https://doi.org/10.30865/json.v4i2.5312

Iskandar, B., Saifullah, S., Irawan, E., & ... (2020). Analisa Perbandingan Pengelompokkan Ekspor dan Impor di Indonesia Berdasarkan Bulan menggunakan K-Means. *Prosiding Seminar …*, *2*, 457–467. http://tunasbangsa.ac.id/seminar/index.php/senaris/article/view/195

Luchia, N. T., Handayani, H., Hamdi, F. S., Erlangga, D., & Octavia, S. F. (2022). Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *2*(2), 35–41. https://doi.org/10.57152/malcom.v2i2.422

Nuraini, R. (2022). Implementasi Euclidean Distance dan Segmentasi K-Means Clustering Pada Identifikasi Citra Jenis Ikan Nila. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *3*(1), 1–8.

Orisa, M., & Faisol, A. (2021). Analisis Algoritma Partitioning Around Medoid untuk Penentuan Klasterisasi. *Jurnal Teknologi Informasi Dan Terapan*, *8*(2), 86–90. https://doi.org/10.25047/jtit.v8i2.258

Rahman, S., Sembiring, A., Siregar, D., Khair, H., Gusti Prahmana, I., Puspadini, R., & Zen, M. (2023). Python : Dasar Dan Pemrograman Berorientasi Objek. In *Penerbit Tahta Media*.

Septhiani, A., & Hendry, H. (2023). Analisis Perbandingan Algoritma Supervised Learning untuk Prediksi Kasus Covid-19 di Jakarta. *Jurnal Sains Komputer Dan Informatika (J-SAKTI)*, *7*(2), 583–594. https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/668/643

Tinendung, I. S., & Zufria, I. (2023). *Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering*. *7*, 2014–2023. https://doi.org/10.30865/mib.v7i4.6908

Tri Sugiarti Ramadhan, Nanik Wahyuningtiyas, M. M. (2023). Understanding The Export And Import Process. In *Angewandte Chemie International Edition*, *6(11)*, *951–952.* (Vol. 3, Issue 1). https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf

Trilarasati, L. S., Khafiya, N. N., Adriananta, M., Fitriana, A. N., & Velma, M. G. (2023). Komparasi Daya Saing Ekspor Lada Indonesia Terhadap. *Jurnal Economina*, *2*(1), 231–240.

Wahyudi, M., & Pujiastuti, L. (2022). Komparasi K-Means Clustering dan K-Medoids dalam Mengelompokkan Produksi Susu Segar di Indonesia. *Jurnal Bumigora Information Technology (BITe)*, *4*(2), 243–254. https://doi.org/10.30812/bite.v4i2.2104

Xia, S., Peng, D., Meng, D., Zhang, C., Wang, G., Giem, E., Wei, W., & Chen, Z. (2022). Ball k-Means: Fast Adaptive Clustering With No Bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(1), 87–99. https://doi.org/10.1109/TPAMI.2020.3008694

*name of corresponding author