Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

Clustering the Economic Conditions of Various Countries From 2010-2023 by Conducting A Comparative Analysis of the K-Means and K-Medoids Algorithms

Yunita Yunita [1,*], Hanifah Ekawati[2], Amelia Yusnita[2]

[1]Prodi Bisnis Digital, STMIK Widya Cipta Dharma, Samarinda, Indonesia
[2]Prodi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia
Email: [1,*]yunita@wicida.ac.id, [2]hanifah@wicida.ac.id, [3]amelia@wicida.ac.id
Correspondence Author Email: yunita@wicida.ac.id
Submitted: **11/05/2025**; Accepted: **30/06/2025**; Published**: 30/06/2025**

Abstrak−Understanding the similarities and differences in economic conditions across countries is crucial for various stakeholders. This research investigates the global economic landscape by clustering countries based on their economic indicators, including GDP, inflation rate, unemployment rate, and economic growth, spanning the period of 2010 to 2023. This timeframe encompasses significant global economic events, making it pertinent for analysis. The study employs and compares two prominent clustering algorithms: K-Means and K-Medoids, to identify groups of countries exhibiting similar economic patterns. Utilizing secondary data from Kaggle encompassing 19 countries, the research assesses the ability of each algorithm to delineate meaningful economic clusters. The K-Means algorithm, with a determined optimal number of four clusters, demonstrated a reasonably good cluster separation and moderate internal cohesion, evidenced by a Silhouette Coefficient of 0.58 and a Davies-Bouldin Index of 0.63. In contrast, the K-Medoids algorithm yielded a distinct clustering structure with a lower Silhouette Coefficient (0.26) and a higher Davies-Bouldin Index (1.16), suggesting less distinct cluster separation and potential sensitivity to data characteristics. This comparative analysis provides insights into the applicability and performance of K-Means and K-Medoids in discerning global economic structures, contributing to a deeper understanding of the world economic map and the utility of clustering techniques in economic data analysis.

Keywords: K-Means Method; K-Medoids Method; Data Mining, Comparative Algoritma; Clustering

1. INTRODUCTION

The state of a country's economy is the complex result of various interconnected factors. These include government financial and monetary policies, the rate of increase in the prices of goods (inflation) and unemployment, as well as the growth in the total value of goods and services produced (GDP) and the difference between the value of exports and imports (trade balance). Understanding the similarities and differences in economic conditions between countries is important for many parties, including policymakers, investors, and researchers. This information is useful for recognising opportunities and threats at the global level, developing more targeted economic development strategies, and understanding the dynamics of the world economy as a whole.

In recent decades, economic ties and interdependence between countries around the world have intensified due to globalisation. Economic events in one country can quickly spread and affect the economic conditions of other countries. Therefore, comparing economic conditions between countries has become increasingly important. By grouping countries based on the similarity of their economic conditions, we can gain a better understanding of how the global economy is structured and segmented.

One popular way to analyse data and find groups or clusters is through clustering. This technique aims to divide a set of data into groups in such a way that data within a group is highly similar, while data between groups is significantly different. Two clustering algorithms that are often used are K-Means and K-Medoids.

The K-Means algorithm[1][2][3] works by finding the centroid which is the average value of the group members. This method is known to be efficient for large data. However, it is susceptible to outliers and tends to assume a spherical group shape. On the other hand, the K-Medoids algorithm[4][5][6] uses actual data points (medoids) as group centres. This makes it more resistant to outliers and less dependent on a particular group shape.

This research aims to cluster the economic conditions of various countries in the time span of 2010 to 2023. The selection of this period is based on the existence of various important economic events at the world level that are expected to affect the economic conditions of many countries, such as the debt crisis in Europe, significant changes in commodity prices, and the COVID-19 pandemic and its recovery process. To achieve this goal, this research will apply and compare two different clustering algorithms, namely K-Means and K-Medoids. This comparison will assess the ability of both algorithms to identify groups of countries with similar economic conditions based on several key economic indicators. The results of this study are expected to provide a deeper understanding of the global economic map and contribute to the use of clustering algorithms to analyse economic data.

Some previous research has been done before, namely research conducted in 2023 by ÖMer N. Kenger and other research colleagues[7]. This paper tackled the limitation of the Smart City Index (SCI)'s rigid city clustering approach. The study employed three more flexible clustering algorithms: K-Means, Fuzzy C-Means, and K-Medoids to categorize cities based on SCI data, aiming for a more nuanced understanding of smart city dynamics. The final result indicated that these clustering techniques offered a different perspective compared to the SCI's original groupings, with K-Means generally performing well with a MSE of 0.033 for the “technology” category, and Fuzzy

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

C-Means showing strength in specific "structure" groups. The research highlighted that the choice of clustering algorithm impacts the resulting city classifications. The identified gap in this study lies in its limited scope, encompassing only 118 cities and relying on survey data from the SCI, suggesting a need for broader and potentially more objective data sources for future investigations.

In 2023, Ermawati, et al. conducted research.[8]. This study addressed the problem of categorizing Indonesian state universities (PTNs) based on their scientific publication productivity to gauge their performance. The method employed was cluster analysis, specifically comparing the k-means and k-medoids algorithms with Silhouette index validation. The final result indicated that the k-means method yielded a better clustering outcome (Silhouette score of 0.8018) compared to k-medoids (0.7281). The k-means clustering identified four PTN-BH in Java as having high publication productivity, sixteen PTN-BH and PTN-BLU predominantly in Java as having medium productivity, and one hundred and two PTN-Satker mostly located outside Java as having low productivity. The gap in this paragraph lies in the lack of discussion regarding the reasons behind these clustering results. While the study successfully categorized universities, it does not delve into the underlying factors contributing to the observed differences in scientific publication productivity across the identified clusters and university types.

Lia Puspita Sari, et al also conducted research in 2023[9]. The research addresses the problem of grouping cities in Indonesia based on their inflation indicators. The method employed to tackle this was a comparative analysis of two clustering techniques: Ward's method and K-Means, utilizing 11 expenditure group variables from the Consumer Price Index (IHK). The final result of the analysis indicated that K-Means clustering yielded a lower standard deviation ratio (1.43) compared to Ward's method (1.77), suggesting that the K-Means clustering results are more suitable for grouping cities based on inflation indicators. The gap in this paragraph lies in the lack of specific details about the characteristics of the resulting clusters. While it identifies K-Means as the preferred method, it doesn't describe the distinct inflation patterns or expenditure profiles that define each of the three identified clusters.

Fathimatuz Zahra, et al also conducted research in 2024[10]. This research addressed the problem of varying poverty levels across Indonesian provinces. To tackle this, the study employed the K-Medoids clustering algorithm to group the 34 provinces based on their poverty data from 2015 to 2022, sourced from the Central Bureau of Statistics (BPS). The analysis successfully identified three distinct poverty clusters: very poor, poor, and vulnerable poor, with the majority of the population falling into the extreme poverty (50%) and poverty (45%) categories. The Silhouette Score of 0.4 indicated a very good clustering result. However, a gap in the paragraph is the lack of specific details about the characteristics or distinguishing factors of each identified cluster, such as which provinces fall into each category and the key poverty indicators that define them.

2. RESEARCH METHODOLOGY

This research methodology is designed to cluster the economic conditions of various countries during the period of 2010-2023 using the K-Means and K-Medoids algorithms, and to conduct a comparative analysis of the clustering results generated by both algorithms. The main stages in this research can be seen in Figure 1 below:

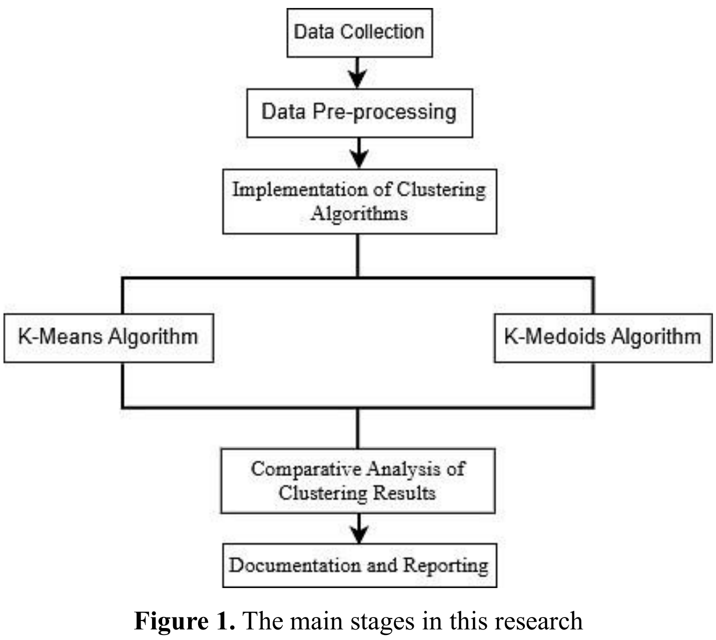

Figure 1. The main stages in this research

2.1 Data Collection:

Secondary data collected from credible and reliable sources such as the World Bank, the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), and 1 relevant national statistical

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

agencies of the selected countries. Data collected annually for the period from 2010 to 2023. This time frame is chosen to encompass various significant global economic dynamics of recent years. A set of key economic indicators have been selected to represent the economic conditions of a country. These indicators may include: Annual Gross Domestic Product (GDP) Growth (%), Inflation Rate (Consumer Price Index - CPI, % annual change), Unemployment Rate (%) and Economic Growth (%).There are 19 countries that will be analysed to serve as data samples in this study.

2.2 Data Pre-processing:

2.2.1 Handling Missing Values

Appropriate methods will be applied to address missing data, such as imputation using the mean, median, or more advanced imputation techniques, depending on the pattern and extent of missing data[11][12][13].

2.2.2 Data Normalization/Standardization:

Given that the economic variables have different scales, normalization techniques (e.g., min-max scaling) or standardization techniques (e.g., z-score standardization) will be applied to ensure that all variables contribute equally to the clustering process and to avoid bias due to scale differences[14][15][16].

2.3 Implementation of Clustering Algorithms:

2.3.1 K-Means Algorithm:

The K-Means algorithm[3][4][17][18][19] is a popular partitional clustering method that aims to divide a dataset into a pre-specified number of K groups (clusters). It operates by iteratively assigning each data point to the nearest cluster based on a distance metric, typically the Euclidean distance. The process begins with the initialization of K cluster centers (centroids) either randomly or using a specific heuristic. Subsequently, each data point is assigned to the cluster whose centroid is closest. Once all data points have been assigned to their respective clusters, new centroids for each cluster are calculated as the arithmetic mean of all the data points belonging to that cluster. This step of assigning data points and updating centroids is repeated until convergence is achieved, which is indicated by no significant change in the centroid positions or cluster memberships. The primary objective of the K-Means algorithm is to minimize the within-cluster variance, which is the sum of the squared distances between each data point within a cluster and its cluster's centroid.

a. The optimal number of clusters (K) will be determined using appropriate methods, such as the Elbow Method, Silhouette Score, or Gap Statistic.
b. The K-Means algorithm will be run on the pre-processed data with multiple initializations of cluster centroids to avoid suboptimal solutions.
c. The clustering results will be evaluated using internal metrics (e.g., Silhouette Score, Davies-Bouldin Index) and interpretation of cluster profiles.

Mathematically, the K-Means algorithm seeks to minimize the objective function (often referred to as the distortion function or inertia) defined as:

$$J(V) = \sum_{i=1}^{K} \sum_{x_j \in S_i} \left\| x_j - v_i \right\|^2 \quad (1)$$

Where K is the desired number of clusters, x_j is the j-th data point, S_i is the set of data points belonging to the i-th cluster and v_i is the centroid vector (mean) of the i-th cluster, calculated as:

$$v_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j \quad (2)$$

$||x_j - v_i||^2$ is the squared Euclidean distance between the data point x_j and the centroid v_i

The algorithm is known for its efficiency in handling large datasets but is sensitive to the initial initialization of centroids and the presence of outliers in the data. Furthermore, K-Means tends to produce clusters that are convex in shape and relatively of equal size.

2.3.2 K-Medoids Algorithm (PAM - Partitioning Around Medoids)

The K-Medoids algorithm[20][21][22], also known as PAM (Partitioning Around Medoids)[23], is a partitional clustering algorithm that aims to divide a dataset into a pre-specified number of K clusters. Unlike K-Means, which uses the mean of data points (centroid) as the cluster center, K-Medoids selects actual data points from within the cluster itself to serve as the cluster centers, called medoids. A medoid is defined as the point within a cluster that has the smallest average dissimilarity to all other points in the same cluster. The K-Medoids algorithm starts by randomly selecting K initial data points as medoids. Then, each remaining data point is assigned to the cluster of the nearest medoid based on a distance metric (often Euclidean distance or other dissimilarity measures). After all data points have been assigned, new medoids for each cluster are selected by evaluating all points within the cluster and choosing the one that has the minimum total distance to all other points in that cluster. This process of assigning data points and updating medoids is repeated until there is no significant change in the medoids or cluster memberships. By using actual data points as cluster centers, K-Medoids is more robust to the presence of outliers compared to K-Means[24].

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

a. The same number of clusters (K) as used in K-Means will be applied for easier comparison. The same methods for determining K will be used.
a. The PAM algorithm will be run on the pre-processed data.
b. The clustering results will be evaluated using internal metrics (e.g., Silhouette Score, Davies-Bouldin Index) and interpretation of cluster profiles.

Mathematically, the K-Medoids algorithm seeks to minimize a cost function E which is the sum of the dissimilarities between each data point in a cluster and the medoid m_i of that cluster:

$$E = \sum_{i=1}^{K} \sum_{x_j \in S_i} d(x_j, m_i) \quad (3)$$

Where K is the desired number of clusters, x_j is the j-th data point, S_i is the set of data points belonging to the i-th cluster, m_i is the medoid of the i-th cluster, which is one of the data points in S_i and $d(x_j, m_i)$ is the distance (dissimilarity) function between the data point x_j and the medoid m_i (e.g., Euclidean distance). The selection of the medoid m_i is done by finding the point $x_k \in S_i$ that minimizes:

$$\sum_{x_j \in S_i} d(x_j, x_k) \quad (4)$$

The K-Medoids algorithm is more computationally expensive than K-Means, especially for large datasets, because the medoid selection process involves calculating distances between all pairs of data points within each cluster at each iteration. However, its advantage in handling outliers and its ability to work with various types of distance metrics make it an attractive choice in certain situations.

2.4 Comparative Analysis of Clustering Results:

2.4.1 Comparison of Evaluation Metrics:

The values of internal evaluation metrics (such as Silhouette Score and Davies-Bouldin Index) generated by both algorithms will be compared to assess the quality of clustering for each algorithm.

2.5 Cluster Profile Analysis:

The economic characteristics of each cluster generated by K-Means and K-Medoids will be analyzed. This involves examining the mean values (for K-Means) or the medoid values (for K-Medoids) of each economic variable within each cluster.

2.6 Cluster Stability Comparison:

The stability of the clusters produced by both algorithms can be tested using techniques such as subsampling or by running the algorithms multiple times with different initializations and measuring the consistency of the groupings.

2.7 Documentation and Reporting:

The entire research process, from data collection and pre-processing to algorithm implementation, comparative analysis, and result interpretation, will be documented in detail. The research report will present the main findings, a comparison of the performance of the K-Means and K-Medoids algorithms in the context of economic data, and the implications of the resulting country groupings.

3. RESULTS AND DISCUSSION

3.1 Results

The ‘Economic Indicators And Inflation’ dataset sourced from the Kaggle platform is a collection of secondary data used in this study, the dataset contains various economic indicators and inflation data from 19 countries in the time period from 2010 to 2025. The indicators used as an assessment include GDP (in billion USD), Inflation Rate (%), Unemployment Rate (%) and Economic Growth (%). This dataset is highly relevant for clustering as it allows researchers to identify groups of countries or time periods that have similar patterns or characteristics based on their economic indicators and inflation rates. There are two clustering methods that will be analysed including the K-Means method and the K-Medoids method. The following Table 1 contains the sample.

Table 1. Sample Data

Country	Year	GDP (in billion USD)	Inflation Rate (%)	Unemployment Rate (%)	Economic Growth (%)
USA	2010	15000.0	1.64	9.63	2.55
…	…	…	…	…	…
USA	2025	22500.0	3.5	4.2	2.6
China	2010	6700.0	3.3	4.2	10.3
China	2011	7000.0	5.4	4.1	9.3

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

Country	Year	GDP (in billion USD)	Inflation Rate (%)	Unemployment Rate (%)	Economic Growth (%)
China	2012	7300.0	2.6	4.1	7.7
…	…	…	…	…	…
China	2025	12000.0	2.8	5.4	3.8
Japan	2010	5500.0	-0.72	5.1	4.2
Japan	2011	5600.0	0.0	4.7	-0.1
Japan	2012	5700.0	0.02	4.6	1.5
…	…	…	…	…	…
Japan	2019	6400.0	0.4	2.4	1.1
Japan	2020	6500.0	0.2	2.8	-4.8
…	…	…	…	…	…
Pakistan	2025	500.0	20.0	5.0	3.2

Based on the sample data in Table 1 above, GDP is one of the features in the dataset that has a significantly different value scale, it can be seen that the feature has a much larger value than the inflation rate (in percentage). If the clustering algorithm is directly applied to data with these different scales, features with larger scales can dominate the calculation of the distance between data points, so that features with smaller scales become less influential. Standardisation is done so that all features have similar scales (mean around 0 and standard deviation around 1), so that each feature contributes equally to the clustering process. The following Figure 2 displays the results of the normalisation that has been done.

```
Hasil Normalisasi (NumPy Array:)
[[ 2.713 -0.258  0.423 -0.232]
 [ 2.828 -0.246  0.32  -0.514]
 [ 2.943 -0.254  0.194 -0.307]
 ...
 [-0.636 -0.118 -0.27   0.168]
 [-0.625 -0.08  -0.27  -0.108]
 [-0.613 -0.118 -0.27  -0.053]]
```

Figure 2. Normalisation Results

After normalisation in Figure 2 above, the next process is to implement the K-Means algorithm, in this study it is done iteratively by testing various number of clusters (k). For each value of k, the K-Means model is trained on the scaled data, and the inertia value (WCSS) of the clustering result is recorded. The aim is to find the optimal number of clusters using the elbow method, which is to observe the decrease in the inertia value as the number of clusters increases and look for the point where the decrease starts to slow down significantly, which is believed to be an indication of a balance between cluster compactness and the number of clusters that is not excessive. The results of this process are then visualised in the form of a plot to facilitate the identification of the number of clusters that best fit the data structure which can be seen in Figure 3 below.

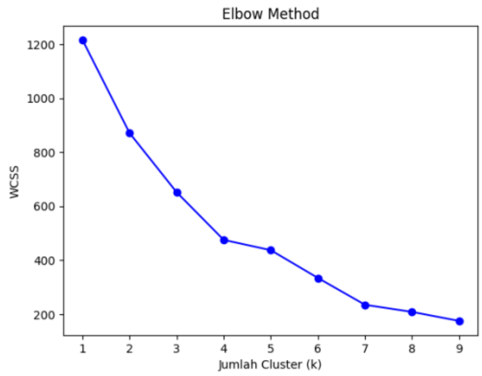

Figure 3: Determination of the Number of K (Cluster)

From the graph in Figure 3 above, it can be seen that the decrease in WCSS is very steep from k=1 to k=3 or k=4, and then the rate of decrease slows down significantly after that point. The point where this change in the rate of decrease occurs (forming an 'elbow') is considered to be the optimal number of clusters, as adding more clusters after this point does not provide a substantial reduction in WCSS, which means that adding clusters no longer significantly improves the homogeneity within the cluster. Based on the visual of this graph, the optimal number of clusters is around 3 or 4, where there is a visible 'elbow' formation so this study will form 4 clusters. Here is Figure 4 of the implementation results using the K-Means Clustering method.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

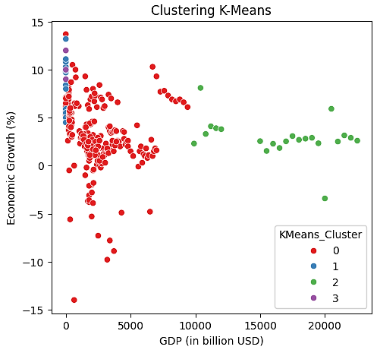

Figure 4. Application of K-Means Clustering

Figure 4 is the visualisation result of data clustering using the K-Means algorithm, where each point represents a data point mapped based on two features: GDP (Gross Domestic Product) in billion USD on the X-axis and Economic Growth (%) on the Y-axis. The different colours at each point indicate the different clusters that have been identified by the K-Means algorithm. The right-hand side indicates that red represents cluster 0, blue represents cluster 1, green represents cluster 2, and purple represents cluster 3. From this visualisation, we can observe how the K-Means algorithm has divided the data points into four different clusters based on their proximity in the two-dimensional space of GDP and Economic Growth features. For example, most of the data points with low GDP and variable economic growth are grouped in the red cluster (0), while the green cluster (2) tends to have higher GDP with moderate economic growth. Blue (1) and purple (3) clusters appear to have relatively low GDP and economic growth characteristics and are concentrated in smaller GDP areas. This visual analysis helps us understand how K-Means identifies groups of similar data based on the considered features.

Based on the three indicators, researchers visualised using only two indicators because GDP is often considered a leading indicator of the size and scale of a country or region's economy. The economic growth rate shows the dynamics and changes in the size of the economy. The combination of the two provides a fundamental picture of economic conditions and developments. The researcher's decision to only visualise GDP and economic growth rate does not mean that other variables in the dataset are not important to the clustering process itself. The performance results of K-Means clustering can be seen in Figure 5 below.


Silhouette Coefficient: 0.5836641281834816
Davies-Bouldin Index: 0.6265039382613496

Figure 5. K-Means performance evaluation

Figure 5 presents two commonly used evaluation metrics to measure the quality of K-Means clustering results, namely Silhouette Coefficient and Davies-Bouldin Index. The Silhouette Coefficient has a value between -1 and 1, where higher values indicate that objects are more similar to their own cluster than to the nearest neighbouring cluster. A value of 0.5836641281834816 indicates that on average, data points within a cluster are fairly well separated and have moderate cohesion within their cluster. On the other hand, the Davies-Bouldin Index measures the ratio of within-cluster similarity to between-cluster separation, with lower values indicating better clustering. A value of 0.6265039382613496 indicates that there is fairly good separation between clusters compared to how dispersed the data is within each cluster. Overall, these two metrics give a positive indication of the quality of the resulting clustering, with the Silhouette Coefficient indicating good separation and the Davies-Bouldin Index confirming that the clusters are relatively dense and well separated from each other.

Figure 6 below shows the visualisation results of clustering the data using the K-Medoids algorithm, which is similar to K-Means but uses actual data points (medoids) as cluster centres instead of centroids.

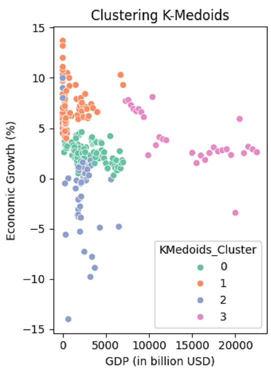

Figure 6. Application of K-Medoids Clustering

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

In Figure 6, each point on the graph represents a data point positioned by GDP value (in billion USD) on the X-axis and Economic Growth (%) on the Y-axis. The different colours at each point indicate the cluster membership assigned by the K-Medoids algorithm, with the legend on the right-hand side identifying each colour with a cluster number (0 to 3). From this visualisation, we can observe how the data has been clustered into four different groups based on the similarity of GDP values and economic growth. For example, the turquoise green coloured cluster (0) tends to have lower to medium GDP values with economic growth varying around positive values. Orange clusters (1) are concentrated at very low GDP values but with a wide range of economic growth, including high values. Purple clusters (2) indicate groups with relatively low to medium GDP values and economic growth that tends to be negative or low positive. Finally, the pink clusters (3) appear to have higher GDP values with varying economic growth. This visualisation helps us understand how K-Medoids groups the data based on the medoid representation of each cluster. The performance results of K-Medoids clustering can be seen in Figure 7 below.

```
Silhouette Coefficient: 0.26113900034086496
Davies-Bouldin Index: 1.1638253156038378
```

Figure 7. K-Medoids performance evaluation

Figure 7 presents two performance evaluation metrics of the K-Medoids algorithm, namely Silhouette Coefficient and Davies-Bouldin Index. Silhouette Coefficient has a value between -1 and 1, where a higher value indicates that each data point is more similar to its own cluster member than to other cluster members. A Silhouette Coefficient value of 0.26113900034086496 indicates that the separation between clusters is not very clear and the cohesion within clusters is also not very strong. On the other hand, the Davies-Bouldin Index measures the average similarity of each cluster to the cluster it is most similar to, where lower values signify better clustering (dense and well-separated clusters). A Davies-Bouldin Index value of 1.1638253156038378 indicates that there is significant similarity between clusters, and the separation between clusters may not be optimal. Overall, the low Silhouette Coefficient value and relatively high Davies-Bouldin Index suggest that the quality of the resulting clustering may be poor, with indications of overlap between clusters and weak internal cluster cohesion. The following comparison graph of the K-means and K-Medoids methods can be seen in Figure 8.

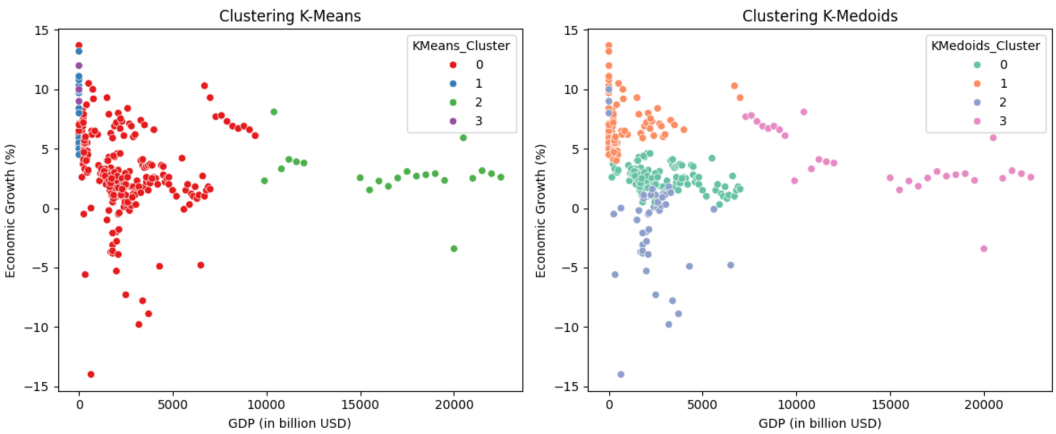

Figure 8. Visualisation results of scatter plot comparison of K-Means and K-Medoids methods

Based on the clustering analysis results visualised in two scatter plots in Figure 8 above, the comparison between K-Means and K-Medoids algorithms in clustering data based on GDP and Economic Growth features shows differences in cluster formation. In the K-Means plot, it can be seen that the algorithm tends to produce clusters with a more rounded shape and relies on the centroid as the cluster centre, which results in some data points with extreme values of low GDP and variable economic growth being grouped together in the dominant red (0) cluster. Meanwhile, the K-Medoids plot displays clusters with a more organic shape and is defined by medoids (actual data points), resulting in clusters that may be more representative of the actual density of the data, as seen in the separation of turquoise green clusters (0) with low to medium GDP values and positive economic growth, and orange clusters (1) that separate data with very low GDP but a wide range of economic growth. This difference indicates that the choice of clustering algorithm may result in different interpretations of the cluster structure in the data, where K-Medoids may be more sensitive to the presence of outliers or data with non-linear distributions compared to K-Means.

3.2 Discussion

This study analyses and interprets the clustering results for the time period (2010-2025) of various countries using two different clustering algorithms, namely K-Means and K-Medoids. In general, the clustering results using the K-Means algorithm tend to show higher consistency in assigning countries to a single cluster throughout the 2010-2025 time period. This can be seen in countries such as Australia, Bangladesh, Brazil, Canada, France, Germany, India,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

Italy, Japan, Malaysia, and Saudi Arabia, where most or all of the time period is grouped into Cluster 0. This consistency indicates that based on the Euclidean distance metric used by K-Means, the data characteristics of these countries are relatively stable throughout the observed period. In contrast, the K-Medoids algorithm, which uses actual data points (medoids) as cluster centres and measures dissimilarity based on Manhattan distance, tends to produce more variable groupings over time for many countries. This difference could be due to the sensitivity of K-Medoids to the presence of outliers or significant changes in one or a few data points (specific years), which could affect the selection of medoids and the overall cluster configuration.

While K-Means often puts most countries into Cluster 0, K-Medoids shows a clearer distinction between countries in terms of their cluster profiles. For example Bangladesh, the entire time period is grouped into Cluster 0 by K-Means, but entirely into Cluster 1 by K-Medoids. This difference suggests that the two algorithms capture different data structures for Bangladesh. Brazil, China, Indonesia and Saudi Arabia also show differences.

Interestingly, the K-Medoids algorithm consistently identified 2020 as a different year for some countries (Australia, Canada, Germany, India, Malaysia, South Korea and the UK), placing them in Cluster 2. This reflects the significant global impact of events in that year (such as the COVID-19 pandemic) on these countries' economic data, which K-Medoids is more sensitive to detect.

The fundamental difference between K-Means (using the mean as the cluster centre and Euclidean distance) and K-Medoids (using actual data points as the cluster centre and Manhattan distance) is a key factor in the difference in clustering results. K-Means tends to be more sensitive to the variance and shape of the data distribution, while K-Medoids is more robust to outliers because it uses actual data points as cluster centres.

Changes in clusters over time, especially those identified by K-Means (as in China and Indonesia) and K-Medoids (as in Brazil, France, Saudi Arabia, and Turkey), indicate significant changes in the data patterns of these countries over the study period. Further analysis of the variables used in the clustering is required to understand the economic or other factors underlying these changes.

4. CONCLUSIONS

This research aims to cluster the economic conditions of various countries from 2010 to 2023 based on secondary data from Kaggle which includes economic indicators (GDP, inflation rate, unemployment rate, economic growth) of 19 countries. Through comparative analysis of K-Means and K-Medoids algorithms, this research identifies similar patterns and characteristics across countries or time periods. The results of the K-Means implementation with 4 clusters show a fairly good separation and moderate cohesion between clusters based on the visualisation of GDP and economic growth, which is supported by the Silhouette Coefficient value of 0.58 and Davies-Bouldin Index of 0.63. Meanwhile, K-Medoids produced a different clustering with a lower Silhouette Coefficient value (0.26) and a higher Davies-Bouldin Index (1.16), indicating a less clear separation between clusters and potential overlap, as well as a higher sensitivity to data with non-linear distribution or outliers than K-Means.

REFERENCES

[1] H.-L. Minh, T. Sang-To, M. A. Wahab, and T. Cuong-Le, “A new metaheuristic optimization based on K-means clustering algorithm and its application to structural damage identification,” *Knowledge-Based Systems*, vol. 251, p. 109189, 2022, doi: https://doi.org/10.1016/j.knosys.2022.109189.

[2] M. I. Faizi and S. M. Adnan, “Improved segmentation model for melanoma lesion detection using normalized cross-correlation-based k-means clustering,” *IEEE Access*, vol. 12, pp. 20753–20766, 2024, doi: 10.1109/ACCESS.2024.3360223.

[3] Z. Salman and A. Alomary, “Performance of the K-means and fuzzy C-means algorithms in big data analytics,” *International Journal of Information Technology*, vol. 16, no. 1, pp. 465–470, 2024, doi: https://doi.org/10.1007/s41870-023-01436-y.

[4] J. Heidari, N. Daneshpour, and A. Zangeneh, “A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers,” *Pattern Recognition*, vol. 155, pp. 1–12, 2024, doi: https://doi.org/10.1016/j.patcog.2024.110639.

[5] N. Hasdyna and R. K. Dinata, “Comparative Analysis of K-Medoids and Purity K-Medoids Methods for Identifying Accident-Prone Areas in North Aceh Regency,” *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 263–272, 2024, doi: https://doi.org/10.15294/sji.v11i2.3433.

[6] A. de Mathelin, N. E. Cecchi, F. Deheeger, M. Mougeot, and N. Vayatis, “OneBatchPAM: A Fast and Frugal K-Medoids Algorithm,” *arXiv preprint arXiv:2501.19285*, 2025, doi: https://doi.org/10.48550/arXiv.2501.19285.

[7] Ö. N. Kenger, Z. D. Kenger, E. Özceylan, and B. Mrugalska, “Clustering of Cities Based on Their Smart Performances: A Comparative Approach of Fuzzy C-Means, K-Means, and K-Medoids,” *IEEE Access*, vol. 11, pp. 134446–134459, 2023, doi: 10.1109/ACCESS.2023.3333753.

[8] E. Ermawati, I. Sriliana, and R. Sriningsih, “Clustering of state universities in indonesia based on productivity of scientific publications using K-Means and K-Medoids,” *BAREKENG: Journal of Mathematics and Its Applications*, vol. 17, no. 3, pp. 1617–1630, 2023, doi: https://doi.org/10.30598/barekengvol17iss3pp1617-1630.

[9] L. P. Sari, A. Fanani, and A. H. Asyhar, “Analisis Perbandingan Pengelompokan Kota di Indonesia Berdasarkan Indikator Inflasi Tahun 2021 dengan Metode Ward dan K-Means,” *Jurnal Sains Matematika dan Statistika (JSMS)*, vol. 9, no. 2, pp. 108–118, 2021, doi: http://dx.doi.org/10.24014/jsms.v9i2.21100.

[10] F. Zahra, A. Khalif, and B. N. Sari, “Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids,” *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 2, pp. 1243–1249, 2024, doi: http://dx.doi.org/10.23960/jitet.v12i2.4199.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 793−801
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7330

[11] Y. Liu, B. Li, S. Yang, and Z. Li, "Handling missing values and imbalanced classes in machine learning to predict consumer preference: Demonstrations and comparisons to prominent methods," *Expert systems with applications*, vol. 237, p. 121694, 2024, doi: https://doi.org/10.1016/j.eswa.2023.121694.

[12] A. Tharwat and W. Schenck, "Active learning for handling missing data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 3273–3287, 2024, doi: 10.1109/TNNLS.2024.3352279.

[13] N. Kalpourtzi, J. R. Carpenter, and G. Touloumi, "Handling missing values in surveys with complex study design: a simulation study," *Journal of Survey Statistics and Methodology*, vol. 12, no. 1, pp. 105–129, 2024, doi: https://doi.org/10.1093/jssam/smac039.

[14] S. Kim, Y. Noh, Y.-J. Kang, S. Park, J.-W. Lee, and S.-W. Chin, "Hybrid data-scaling method for fault classification of compressors," *Measurement*, vol. 201, p. 111619, 2022, doi: https://doi.org/10.1016/j.measurement.2022.111619.

[15] T. T. Khoei and A. Singh, "Data reduction in big data: a survey of methods, challenges and future directions," *International Journal of Data Science and Analytics*, pp. 1–40, 2024, doi: https://doi.org/10.1007/s41060-024-00603-z.

[16] R. Shetty, M. Geetha, U. D. Acharya, and G. Shyamala, "Enhancing ovarian tumor dataset analysis through data mining preprocessing techniques," *IEEE Access*, vol. 12, pp. 122300–122312, 2024, doi: 10.1109/ACCESS.2024.3450520.

[17] R. Wu, "Behavioral analysis of electricity consumption characteristics for customer groups using the k-means algorithm," *Systems and Soft Computing*, vol. 6, p. 200143, 2024, doi: https://doi.org/10.1016/j.sasc.2024.200143.

[18] B. Wan, W. Huang, B. Pierre, Y. Cheng, and S. Zhou, "K-Means algorithm based on multi-feature-induced order," *Granular Computing*, vol. 9, no. 2, p. 45, 2024, doi: https://doi.org/10.1007/s41066-024-00470-w.

[19] G. Gwak, U. Hwang, and J. Kim, "Clustering of shoulder movement patterns using K-means algorithm based on the shoulder range of motion," *Journal of Bodywork and Movement Therapies*, vol. 41, pp. 164–170, 2025, doi: https://doi.org/10.1016/j.jbmt.2024.11.034.

[20] M. Chaudhry, "Optimized K-Medoid: A Comprehensive Approach to Medoid Discovery and Finding Optimal K," College of Electrical & Mechanical Engineering (CEME), NUST, 2024.

[21] L. Lenssen and E. Schubert, "Medoid Silhouette clustering with automatic cluster number selection," *Information Systems*, vol. 120, p. 102290, 2024, doi: https://doi.org/10.1016/j.is.2023.102290.

[22] Y. Mo, R. Xing, and H. Hou, "A Distributed Higher-Order $ k $-Medoids Clustering Algorithm for Network Partition," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 5, pp. 4181–4193, 2024, doi: 10.1109/TNSE.2024.3402383.

[23] M. Sahoo and S. Rai, "A partitioning around medoids (PAM) based sequential clustering approach for model order estimation of low-frequency oscillations in wide area measurement system," *Sādhanā*, vol. 49, no. 1, p. 94, 2024, doi: https://doi.org/10.1007/s12046-023-02408-5.

[24] N. Nurdin, F. Fajriana, R. Meiyanti, A. Adelia, and M. Maulita, "Clustering and Mapping of Agricultural Production Based on Geographic Information System Using K-Medoids Algorithm," *Journal of Artificial Intelligence and Technology*, vol. 5, pp. 116–124, 2025, doi: https://doi.org/10.37965/jait.2025.633.