

# Economic Expansion and its Effects on Natural Resources

## A Clustering-Based Approach

Ion-Florin RĂDUCU\*

Bucharest University of Economic Studies, Bucharest, Romania

\*florin.raducu@csie.ase.ro

Stelian STANCU

Bucharest University of Economic Studies, Bucharest, Romania

stelian.stancu@csie.ase.ro

**Abstract.** *Over time, it has been constantly discussed and concluded that natural resources are essential elements for the socio-economic dimension of a country, having relevance both in terms of well-being and survival. The use of finite natural resources must follow the rationale that there is a need for future generations to benefit from these resources, because the depletion of non-renewable resources makes them indestructible in historical periods. Economic growth is a complex process of long-term evolution, which manifests itself as an increase in the capacity of a national economy to produce goods and services. This can only be analyzed over a long period of time. The economic balance represents the overall movement of an economy after eliminating seasonal or cyclical movements. The relationship between economic growth and natural resources must be viewed from a complex and sometimes contradictory perspective. Although natural resources are the foundation of all production, their abundance does not always imply long-term economic viability. Moreover, the relationship between these two elements is influenced by the limited availability of resources, which are not eternal in time. The purpose of this paper is to analyze economic growth in different countries of the world, using clustering algorithms. Cluster analysis is a useful tool for reducing sets of objects, or even variables, to a smaller number of information entities known as classes or clusters. The results showed that an optimal number of clusters to group the analyzed countries based on economic characteristics would be five.*

**Keywords:** Natural Resources, Cluster Analysis, Economic Growth, K-means.

## Introduction

The phenomenon of the development of national economies based on an intense industrialization process available in many countries has led to an increase in the demand for metals, which has however begun to be tempered in recent years by the economic crisis. Regardless of whether they have reserves of metal ores or not, most countries have made metallurgy a priority in the context of their economic development and industrialization processes, while ensuring the transformation of resources into primary processing raw materials needed by other industrial branches. Industrially developed countries that have reserves of metal resources (such as the United States, Australia, Canada and Sweden), rely on protecting the national heritage, while using these resources in an optimal way. For example, the United States of America, which imports bauxite, is carrying out extensive research for the extraction of alumina using various substitutes such as clay, kaolin and so on, in order to protect its reserves with increased quality.

Natural resources, especially natural gas, prove their usefulness in a variety of industries. They are widely used as fuel for automobiles, to generate electricity, or to heat homes. Natural gas is a popular resource for both consumers and businesses due to its benefits, which include lower greenhouse gas emissions than other energy sources. On the other hand, economic growth based

solely on the consumption of non-renewable resources is unsustainable, because once the resource is depleted, a country's economy risks collapsing due to a lack of diversity.

In this paper, we proposed the use of two clustering techniques (the hierarchical Ward method and the k-means algorithm) to analyze the economic growth and evolution of different countries around the world through several economic indicators.

Cluster analysis aims to identify classes, groups or clusters of objects or shapes that are as similar as possible within the same class (this means minimal variability within classes) and as distinct as possible from each other if these elements belong to different classes. Therefore, cluster analysis allows examining the similarities and differences between objects belonging to a given set, in order to group them into distinct and homogeneous classes. Each object in the analyzed set belongs to a single class, which is a discrete and unordered set. Clusters are classes or groups that structure a large number of objects (Ionescu, 2015).

## Literature review

Numerous researchers are concerned with the study of economic growth (and macroeconomic indicators), as well as the use of clustering techniques in research, which has enriched the specialized literature in this area.

A research study (Georgantopoulos, 2012) on tourism in Greece looked at the country's economic growth from 1988 to 2011, using both aggregate and disaggregated models. The study was based on processing full empirical data through ADF, PP, KPSS, Johansen cointegration, VAR models with error correction mechanism, impulse responses, variance decomposition and, finally, forecasts were made for real production and total tourism expenditure based on time series models such as VAR/VECM. The empirical results of the research highlighted that all variables are integrated to the second order and that there is still a long-term relationship between tourism expenditure and real GDP.

Wu et al. (2026) provide a fast filtering method for retired power batteries using the k-means++ algorithm, with the goal of improving classification efficiency and accuracy during the battery reuse process. The optimal battery classification number  $k$  is determined for the two application scenarios, and the advantages and disadvantages of the FCM algorithm, OPTICS algorithm, k-means++, and k-means++ mixed with Mahalanobis distance (M-k-means++) algorithms in terms of clustering effect are compared on the data of two types of batteries: NCR18650B and LISHEN 18650, respectively. The findings show that, without increasing the complexity of computation, the improved M-k-means++ algorithm significantly improves the battery clustering effect in two application scenarios. In the energy storage application and low-speed electric vehicle scenarios, the M-k-means++ algorithm outperforms the other algorithms, improving the profile coefficient (SC) by 42.6% and 41.4%, respectively, and lowering the Davidson-Bourdain index.

Brida et al. (2020) investigated the evolving relationship between income inequality and economic growth using a non-parametric approach and a numerical taxonomy, which relies on data coding and clustering methods. The researchers concluded, among other things, that income inequality is the most pressing problem. Income inequality is an important factor for poverty, growth, and equity.

Li et al. (2019) examined the relationship between carbon dioxide emissions and economic growth in the context of global warming using clusters. In this study, clusters are determined by the differences between the fitting models of the countries in each cluster, in contrast to the traditional

clustering technique, which uses geometric distances. This paper attempts to address the issue by proposing a novel clustering approach based on symbolic regression and the apriori algorithm.

The Ward hierarchical clustering method has been successfully used in a wide range of research studies, spanning diverse fields of knowledge such as economics, medicine, chemistry, and seismology: Ogasawara & Kon (2021), Varin et al. (2009), Chadwick et al. (2019), Hale & Dougherty (1988), Ilieva et al. (2025), Hafeezallah et al. (2024).

## Methodology

Ward's clustering method is defined as a hierarchical agglomerative method that is based on the following methodological approach: assuming there are  $N$  elements to group, one starts with  $N$  groups, each consisting of a single entity, then searches the similarity matrix for the most similar pair of groups and combines them to reduce the number of groups by one (Hervada-Sala & Jarauta-Bragulat, 2004). The hierarchical clustering technique is an unsupervised learning technique in which similar data points are assigned to groups based on distance or similarity. Instead of predetermining the number of groups, it creates a tree structure known as a dendrogram, which describes how the groups merge or split at different levels.

The objective of this clustering technique is to identify, at each stage, the two clusters whose fusion results in the minimum increase in the sum of squares of the total error within the group (or in the distances between the centroids of the merged clusters).

The total within group error sum of squares,  $V_T(K)$ , in a single stage with  $K$  groups,  $J$  variables, and  $N_i$  elements in each group is expressed as:

$$V_T(K) = \sum_{k=1}^K (\sum_{j=1}^J (\sum_{i=1}^{N_i} (x_{ijk} - \bar{x}_{jk}(i))^2)) \quad (1)$$

$$\bar{x}_{jk}(i) = \frac{1}{N_i} \sum_{i=1}^{N_i} x_{ijk}, \quad (2)$$

$$\sum_{i=1}^K N_i = N \quad (3)$$

K-means clustering is an unsupervised learning algorithm that classifies unlabeled data points into groups or clusters. It is one of the most widely used clustering methods in machine learning. Unlike supervised learning, the training dataset for this algorithm contains unlabeled data, which means that the data points do not follow a defined classification structure. K-means is an iterative clustering algorithm that divides a data set into groups based on the distance between their centroids. The centroid, also known as the cluster center, is the average or median of all the points in the cluster, based on the characteristics of the data.

For a data point  $x = (x_1, x_2, \dots, x_n)$  and a centroid  $c = (c_1, c_2, \dots, c_n)$ , the Euclidean distance (the most common type of distance, frequently used) is computed as follow:

$$d(x, c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2} \quad (4)$$

Formally, the k-means algorithm can be described as follows:

1. Determine the number of clusters ( $k$ );
2. Choose  $k$  data points at random to serve as the cluster center;
3. Use the Euclidean distance formula to calculate the distance between each data point and each cluster center;
4. Assign each data point to the cluster whose center is closest to it;
5. Recalculate the centers of newly formed clusters. The center of a cluster is calculated by taking the average of all the data points in that cluster;

6. Continue with the procedure from Step 3 to Step 5 until one of the following stopping criteria is met: if data points integrate into the same cluster, reached the maximum of iteration or the newly formed cluster's center points remain constant.

The RStudio software was used to obtain the case study and results. This is a software that offers numerous benefits and opportunities for modeling, data manipulation, statistical analysis, and programming

## Results and discussions

The World Bank provided the dataset used in the analysis, which included 32 countries and ten indicators.

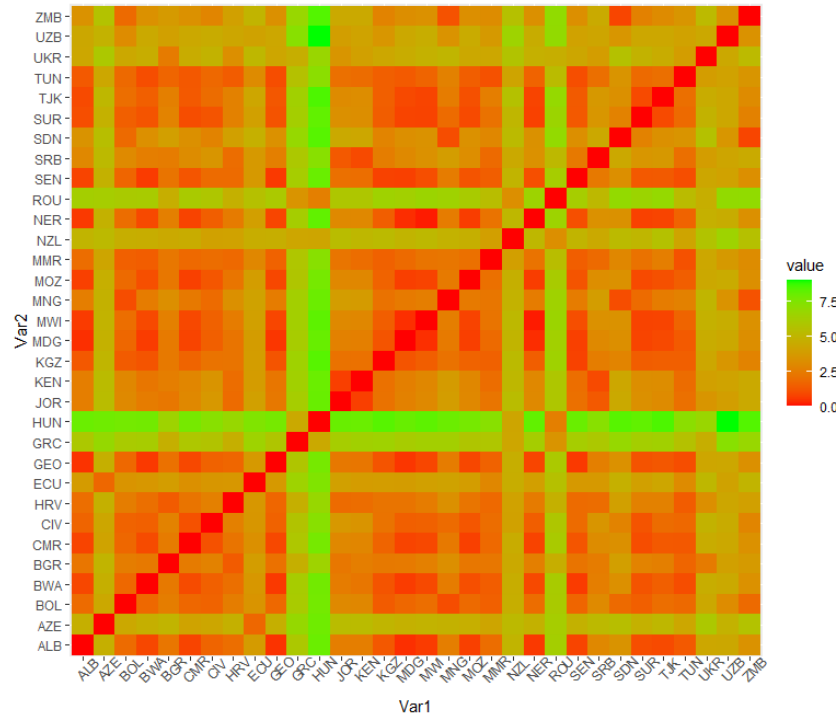
Table 1 outlines the indicators used in the analysis.

**Table 1. Indicators description**

| Indicators                                                     | Description                                                                                                                                                        |
|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Adjusted savings: fixed capital consumption (ASFCC)            | The consumption of fixed capital represents the replacement value of capital consumed during the production process.                                               |
| Adjusted savings: education expenditure (ASEE)                 | Education expenditures include the education system's current operating expenses, such as salaries and bonuses.                                                    |
| Adjusted savings: energy depletion (ASED)                      | Energy depletion is defined as the ratio between the stock of energy resources and their remaining lifespan.                                                       |
| Adjusted savings: mineral depletion (ASMD)                     | They are the ratio between the value of the mineral resource stock and its remaining lifespan.                                                                     |
| Services import (SI)                                           | Services are defined as the production of intangible goods that can be produced, transferred, and consumed simultaneously.                                         |
| Goods import (GI)                                              | Imports of goods include all movable goods (including non-monetary gold) involved in a transfer of possession from non-residents to residents.                     |
| Services export (SE)                                           | -                                                                                                                                                                  |
| Goods export (GE)                                              | -                                                                                                                                                                  |
| Net primary income (net income from abroad – NPI)              | Includes net employment income, as well as the SNA's net property and entrepreneurial income components.                                                           |
| Net secondary income (net current transfers from abroad - NSI) | Current transfers are income transfers made between residents of the reporting country and the rest of the world that do not include provisions for reimbursement. |

Source: WorldBank.

Before beginning to apply the two methods, a distance matrix between objects was created to identify similarities and differences.



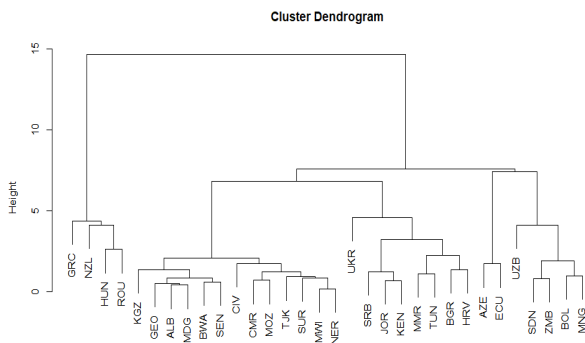
**Figure 1. Distances between objects**

Source: Authors' own research results.

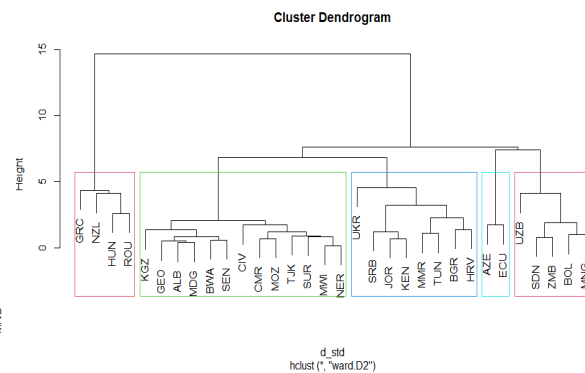
In Figure 1, the green boxes indicate a large Euclidean distance between countries, while the red boxes indicate a small distance. Small distances result in a high level of similarity. Large distances indicate a high level of dissimilarity. For example, the shapes from BOL (Bolivia )to HRV (Croatia) are similar because they use dark colors.

### ***Ward hierarchical method***

This clustering method creates clusters of countries in successive stages, at varying heights. This process can be represented graphically in the form of a dendrogram.



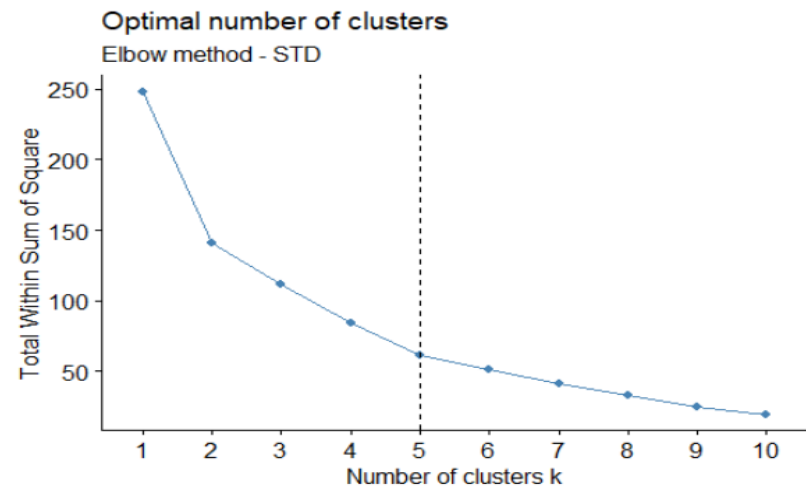
**Figure 2. Dendrogram**



**Figure 3. Dendrogram with clusters**

Source: Authors' own research results

Figure 2 illustrates a clustering tree. The height is represented on the 0–15 axis. Clusters form at varying heights, as seen from bottom to top.



**Figure 4. Elbow graphic – optimal number of clusters**

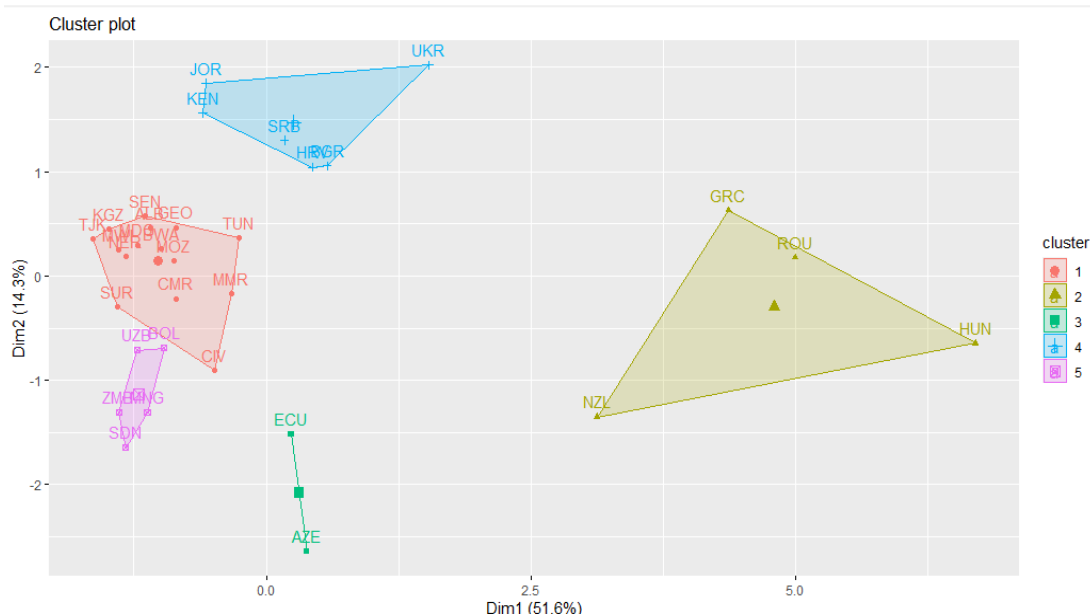
Source: Authors' own research results.

Figure 4, the so-called "turning point" (*elbow*) graph, depicts the optimal number of clusters. These clusters are also visible in Figure 3's dendrogram, which includes the countries that combine to make up each cluster. The graph depicts the relationship between the number of clusters ( $k$ ), represented on the abscissa axis, and the Total Within Sum of Squares (WSS) of Errors within the clusters (WSS), represented on the coordinate axis. The goal is to strike a balance between model accuracy (as low as possible WSS) and simplicity (a smaller number of clusters). Up to  $k = 5$ , the formation of a new cluster significantly improves the explanation for the data's variance. Increasing this number of clusters results in a fairly small information surplus, which implies that considering more clusters in the analysis may overfit the model or artificially segment some natural groupings.

The centroids were also calculated and are included in Table 2 of the appendix. The centroids are the arithmetic average of the variables in a cluster. The first cluster, on average, has the smallest adjusted economies, imports, and exports. So economic growth is quite slow. Cluster 4 has the highest imports and exports, but the lowest net income. This cluster exhibits higher economic growth.

### ***K-means algorithm***

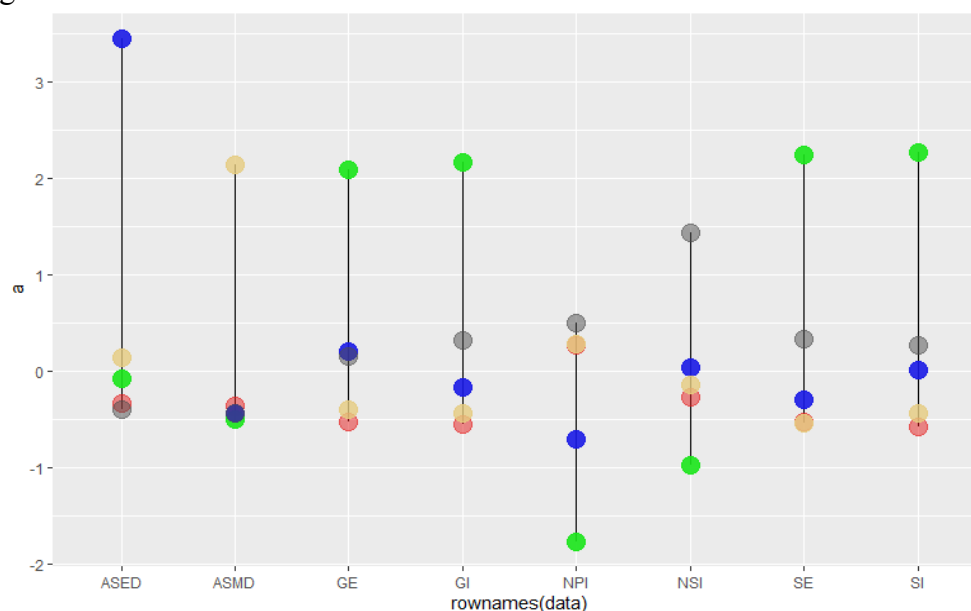
Table 3 in the appendix displays the output of the k-means algorithm, which was executed with the RStudio program.



**Figure 5. Graphic representation of the 5 clusters**

Source: Authors' own research results.

Figure 5 depicts the five clusters and the countries that are members of them. The two dimensions represent the two primary components. The first has 51.6% of the information, while the second has 14.3%. Countries in the first cluster have low net exports but high mineral revenues. Cluster 4 countries have higher revenues, while cluster 3 countries have lower revenues and net exports. Countries in the second cluster have high values for both dimensions, resulting in increased economic growth.



**Figure 6. Discriminating power of variables**

Source: Authors' own research results.

To assess the external validity and discriminatory power of the variables, we examined the centroid profiles of the five clustering solutions (Figure 6). In the context of cluster analysis, the discriminatory power of a variable is defined as the ability of that variable to clearly separate and distinguish between different clusters. A variable has high discriminatory power if its mean values (centroids) are significantly different between the groups.

Discriminant power is commonly defined as the ratio of between-group variance to within-group variance. A variable is useful in the analysis and should be kept if it reduces intra-cluster distances while increasing the distance between centroids.

The distribution of standardized values presents a significant degree of heterogeneity, indicating that the K-means algorithm achieved a favorable segmentation of the population into distinct groups. Variables such as ASED and ASMD have the highest discrimination power and are responsible for defining the extreme clusters (blue and yellow). The GE and GI variables, on the other hand, show greater convergence between the red and gray groups, which denotes their lower relevance in distinguishing these specific segments.

## Conclusion

To summarize the course of this article, the purpose of the analysis was determined by capturing the economies of various countries around the world and their evolution using cluster analysis, which is useful for capturing groups of countries that have similar or identical economic situations.

Using both the Ward hierarchical method and the k-means algorithm, we obtained the same number of clusters, namely five. Based on the results of the analysis and the graphical representations of the clusters, we can conclude that Romania is no longer classified as a resource-dependent economy or in the process of primary development. It is part of a group of economies that have reached institutional maturity and market integration, with growth driven by stability and services. Romania (ROU) is part of the red cluster, which also includes New Zealand (NZL), Greece (GRC) and Hungary.

Countries such as Niger (NER), Mozambique (MOZ) and Bolivia (BOL) are examples of economies where growth is likely hindered by dependence on natural resources or structural instability.

## Acknowledgement(s)

We are grateful to participate of this conference. This work was supported by the Bucharest University of Economic Studies during the phd studies of Ion-Florin Răducu.

## References

- Brida, J.G., Carrera, E.J.S., & Segarra, V. (2020). Clustering and regime dynamics for economic growth and income inequality. *Structural Change and Economic Dynamics*, 52, 99-108
- Chadwick, P., Peel, A., Thirkell, G., & Pichon, B. (2019). Use of whole genome sequencing to resolve a surgical ward cluster of meticillin-resistant *Staphylococcus aureus*. *Infection, Disease & Health*, 24 (1), S8-S9
- Ilieva, R., Kalaydzhiev, P., Slavchev, B., Spasova, N., Kinova, E., Goudev, A. (2025). Clinical phenotypes of severe atrial cardiomyopathy and their outcome: A cluster analysis. *IJC Heart & Vasculture*, 58, 101679.
- Ionescu, S.A. (2015). Hierarchical clustering with applications in financial analysis. *The Romanian Journal of Statistics*, 8, 3-14

- Georgantopoulos, A. (2012). Forecasting Tourism Expenditure and Growth: A VAR/VECM Analysis for Greece at both Aggregated and Disaggregated Levels. *International Research Journal of Finance and Economics*, 96, 155-167
- Hafeezallah, A., Al-Dhamari, A., & Abu-Bakar, S.A.R. (2024). Motion segmentation using Ward's hierarchical agglomerative clustering for crowd disaster risk mitigation. *International Journal of Disaster Risk Reduction*, 102, 104262
- Hale, R.L., & Dougherty, D. (1988). Differences between ward's and UPGMA methods of cluster analysis: Implications for school psychology. *Journal of School Psychology*, 26 (2), 121-131
- Hervada-Sala, C., & Jarauta-Bragulat, E. (2004). A program to perform Ward's clustering method on several regionalized variables. *Computers & Geosciences*, 30 (8), 881-886
- Li, W., Yang, G., Li, X., Sun, T., & Wang, J. (2021). Cluster analysis of the relationship between carbon dioxide emissions and economic growth. *Journal of Cleaner Production*, 225, 459-471
- Ogasawara, Y., & Kon, M. (2021). Two clustering methods based on the Ward's method and dendrograms with interval-value dissimilarities for interval-valued data. *International Journal of Approximate Reasoning*, 129, 103-121
- Varin, T., Bureau, R., Mueller, C., & Willett, P. (2009). Clustering files of chemical structures using the Székely–Rizzo generalization of Ward's method. *Journal of Molecular Graphics and Modelling*, 28 (2), 187-195
- Wu, Z., Xu, X., Li, H. (2026). Cluster sorting of retired power batteries based on improved K-means++ algorithm. *Journal of Energy Storage*, 144, 119822

## Appendix

**Table 2. Centroids**

|   | ASED   | ASMD   | SI     | SE     | GI     | GE     | NPI    | NSI    |
|---|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | -0.380 | -0.376 | -0.606 | -0.554 | -0.615 | -0.574 | 0.369  | -0.383 |
| 2 | 3.443  | -0.427 | 0.009  | -0.297 | -0.161 | 0.209  | -0.700 | 0.041  |
| 3 | 0.147  | 2.140  | -0.436 | -0.532 | -0.428 | -0.391 | 0.289  | -0.136 |
| 4 | -0.300 | -0.371 | 0.120  | 0.188  | 0.222  | 0.083  | 0.274  | 1.184  |
| 5 | -0.071 | -0.499 | 2.268  | 2.236  | 2.169  | 2.085  | -1.759 | -0.975 |

Source: Authors' own research results.

**Table 3. The result of the k-means algorithm**

K-means clustering with 5 clusters of sizes 15, 4, 2, 6, 5

Cluster means:

|   | ASED        | ASMD       | SI           | SE         | GI         | GE         | NPI        | NSI         |
|---|-------------|------------|--------------|------------|------------|------------|------------|-------------|
| 1 | -0.33099766 | -0.3518471 | -0.569472545 | -0.5155169 | -0.5451885 | -0.5176531 | 0.2671848  | -0.27234715 |
| 2 | -0.07134387 | -0.4992048 | 2.268358118  | 2.2363722  | 2.1692984  | 2.0848953  | -1.7587326 | -0.97534426 |
| 3 | 3.44282118  | -0.4274683 | 0.008761776  | -0.2967227 | -0.1609698 | 0.2086630  | -0.7002401 | 0.04084945  |
| 4 | -0.39495682 | -0.4285938 | 0.271477808  | 0.3400853  | 0.3271042  | 0.1601801  | 0.4973423  | 1.43092291  |
| 5 | 0.14688780  | 2.1402050  | -0.435546941 | -0.5319604 | -0.4280102 | -0.3906382 | 0.2887168  | -0.13613040 |

Clustering vector:

| ALB | AZE | BOL | BWA | BGR | CMR | CIV | HRV | ECU | GEO | GRC | HUN | JOR | KEN | KGZ | MDG | MWI | MNG | MOZ | MMR | NZL | NER | ROU | SEN | SRB | SDN | SUR |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| TJK | TUN | UKR | UZB | ZMB |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 1   | 3   | 5   | 1   | 4   | 1   | 1   | 4   | 3   | 1   | 2   | 2   | 4   | 4   | 1   | 1   | 1   | 5   | 1   | 1   | 2   | 1   | 2   | 1   | 4   | 5   | 1   |
| 1   | 1   | 4   | 5   | 5   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Within cluster sum of squares by cluster:

[1] 11.057362 21.243212 1.461416 15.514090 11.046404  
(between\_SS / total\_SS = 75.7 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size"  
"iter" "ifault"

Source: Authors' own research results.