Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 10, Issue 2, 2022

Received: September 29, 2021
Accepted: July 07, 2022
Published Online: December 31, 2022

AJ ID: 2022.10.01.MIS.01
DOI: 10.17093/alphanumeric.1002461
Research Article

Analysis of The Countries According to The Prosperity Level with Data Mining

Şebnem Koltan Yılmaz, Ph.D. *

Assist. Prof., Faculty of Economics and Administrative Sciences, Inonu University, Malatya, Turkiye, sebnem.yilmaz@inonu.edu.tr

Sibel Şener, Ph.D.

Assist. Prof., Faculty of Economics and Administrative Sciences, Sivas Cumhuriyet University, Sivas, Turkiye, sblsener@hotmail.com

* İnönü Üniversitesi, Iktisadi ve Idari Bilimler Fakültesi, Elazığ Yolu 15. km, 44280, Malatya, Türkiye

ABSTRACT

Data mining (DM) includes techniques for finding meaningful information hidden in these massive data stacks. The aim of this study is to divide the countries according to their prosperity levels with Cluster Analysis (CA), which is one of the DM techniques. In this context, the 2019 data of 167 countries within the updated 12 prosperity indicators in The Legatum Prosperity Index (LPI) were used. Countries were divided into clusters with the Ward's algorithm, and the Elbow method was used for verifying of the optimal cluster number. The similarities between the countries were determined with the K-Means, and Tukiye's place in the clusters was determined. The results show that countries are divided into three clusters. The most significant indicators in separating them into clusters are "market access and infrastructure, education, investment environment", and the least significant indicators are "social capital, natural environment, safety and security". It has been determined that Turkiye is located in the middle prosperity level cluster and its "health, living conditions, education" indicators are the highest, while its "natural environment, personal freedom, management" indicators are the lowest.

Keywords: Data Mining, Prosperity Level, Legatum Prosperity Index, Cluster Analysis

1. Introduction

As the prosperity of the countries increases, the economies of the countries develop and happy individuals and societies are formed (Akar, 2014). Prosperity measurement is important for the purposes of meeting expectations from the welfare state, determining the location of the countries and comparing prosperity levels (Akar, 2015). The most important criterion for this is welfare indicators. Gross Domestic Product (GDP) is an indicator that has been used for a long time and is known as the basic indicator of prosperity. A country with a higher GDP per capita is generally considered a better place to live. Nevertheless, there have been criticisms about the necessity of considering social indicators in determining prosperity (Markou et al., 2015). Since material well-being distracts people from other important values such as selflessness and justice, GDP does not reflect the way people perceive their welfare and happiness (Budsaratragoon & Jitmaneeroj, 2021). For this reason, the prosperity of a country is expressed in sustainable development goals aimed at meeting basic human needs, and the quality of life of its citizens, rather than material wealth represented by GDP. Sustainable and stable economies that prioritize social and environmental goals, can prevent the increase in prosperity demands generated by only growth-based economies (Büchs, 2021; Budsaratragoon & Jitmaneeroj, 2021).

In multidimensional welfare measures, there are factors that cannot be met at the same time, and the difficulty in balancing them complicates the concept of prosperity. In order to eliminate this complexity, more than one indicator should be included in the composite index, which is a tool for assessing the prosperity performance of a country (Budsaratragoon & Jitmaneeroj, 2021). In this context, attempts have been made around the world to measure prosperity; many indices have been developed that aim to measure welfare with financial, social, environmental and cultural indicators since the 1990s. The oldest of these is the Human Development Index (HDI), which has been published regularly every year since 1990 by the United Nations Development Program (UNDP) (UNDP, HDI, 2019). Global Competitiveness Index (GCI) (GCI, 2019) introduced by the World Economic Forum (2004), Better Life Index (BLI) (BLI, 2019) developed by the Organization for Economic Corporation and Development (OECD) (2011), the World Happiness Report (WHR, 2020) published by UNDP (2012) and the Social Progress Index (SPI) (SPI, 2020) developed by Michael Green and Luke Greeves are other important multidimensional well-being indexes.

The Legatum Prosperity Index (LPI), developed by the Legatum Institute, is another multidimensional welfare index. Unlike others, LPI gives a unique insight into both material wealth and other dimensions of prosperity, and it defends that a nation's prosperity can be achieved through inclusive societies, open economies and people empowerment. This index provides ranking based on the equally weighted sum of overall scores and sub-index scores, in addition to individual sub-index scores. In this direction, LPI, published annually since 2007, determines the prosperity level of countries and provides a comparable analysis. While the LPI covered 149 countries within 9 prosperity indicators in 2018, it covered 167 countries within 12 prosperity indicators in 2019 (LPI, 2018; LPI, 2019a).

The increasing interest in prosperity assessment at regional, national and international levels in recent years has brought along the use of various methods in this regard. One of these methods is Data Mining (DM). DM, which enables to reveal meaningful information hidden in large data stacks, combines many techniques from fields such as statistics, machine learning, pattern recognition, database and data warehouses, visualization, algorithms, high-performance computing (Han, Kamber & Pei, 2012). In DM, there is statistical induction instead of generalization of population results (Tüzüntürk, 2010) and basically three functions are performed as "classification and regression", "clustering", "association rules and sequential time patterns" (Albayrak & Koltan Yılmaz, 2009). Clustering, which is one of these functions, is a DM method used in many fields to find similar data groups from data and to create a model from the data (Ali & Kadhum, 2017). It produces insights from data analyzed and interpreted by humans (Maylawati et al., 2020). With Cluster Analysis (CA), the data is divided into subgroups (clusters) that are similar to each other but not known before, according to their basic characteristics. Algorithms using for this purpose are collected in two groups: Hierarchical Clustering (HC) and Non-Hierarchical Clustering (NHC). The aim of this study is to divide countries according to their prosperity levels with CA, and to show the applicability of the method. The research questions are based on three main themes:

What similarities (or differences) are there between the emerged clusters according to the prosperity levels of the countries?

What are the indicators that are significant in separating countries according to their prosperity levels and their levels of influence?

What is the place of Türkiye among these clusters?

A two-step method involving HC and NHC is used in this article to address the research questions. This method ensures that the prosperity levels of countries are determined and compared according to the 2019 data of LPI within the updated 12 indicators.

2. Literature Review

CA make important contributions to divide countries according to their prosperity level. Some of the studies using CA are as follows. The algorithms, indexes and indicators are summarized in Table 1.

Bambra (2007), Abu Sharkh & Gough (2010), Kowalski & Wałęga (2015) examined welfare state regime models. Abu Sharkh & Gough (2010) aimed to develop a model to determine the welfare regimes of 65 developing countries and to evaluate their stability for the years 1990-2000. Countries that are similar to each other have been revealed with CA. The number of clusters was determined with HC, and these clusters were explained with NHC. The results show that 4 important clusters were obtained from the countries in 1990 and 8 important clusters in 2000. These results provide a model to rank countries, starting with the cluster that is most similar to OECD welfare states. Unlike this study, Bambra (2007), Kowalski & Wałęga (2015) evaluated the issue within the framework of a gender approach. They used CA to propose a model for comparing states on the basis of women's economic independence and autonomy. Bambra (2007) used data from UNDP for 21 countries for 2003 and 2004. The result obtained presents an approach that divides the countries into 5 clusters. Kowalski & Wałęga (2015) used the same indicators as Bambra (2007) and clustered

18 countries with 2011 data. In the country selection, unlike previous studies, they included highly developed social democratic, liberal and conservative countries with post-communist countries that joined the European Union (EU) in 2004 or 2007 that have similar labor market conditions within the indicators used. They examined the similarities in the 4 clusters, and revealed that post-communist countries were divided into 2 different clusters, but these countries did not form a homogeneous group different from developed countries.

Shahbaz, Iftikhar & Mahmood (2013) analyzed data from the World Bank and Yale Center for Environmental Law and Policy of 58 countries for the years 2000-2009 in order to measure economic welfare and environmental empathy, and determined the empathy levels of the countries. Countries were clustered with NHC; then the effect of individual empathy indicators was visualized with Decision Trees (DT). The results show the 5 clusters formed according to the empathy levels of the countries and the change in empathic attitudes depending on the indicators.

Kangallı, Uyar & Buyrukoğlu (2014) and Gülden & Karakış (2019) evaluated the economic freedoms of OECD countries with HC and NHC. Kangallı et al. (2014) analyzed 34 countries for 2011. They revealed that the countries were divided into 3 clusters and that the similarity between the CA methods was parallel to each other. Gülden & Karakış (2019) examined 36 countries for the years 2018 and 2019. They revealed that the countries were divided into 5 clusters in 2018 and 6 clusters in 2019, and that the most similarity among the methods was obtained with HC.

Alptekin & Yeşilaydın (2015), Mut & Akyürek (2017), Değirmenci & Yakıcı Ayan (2020) aimed to divide OECD countries, including Türkiye, with CA according to their health indicators. They used data obtained from the World Bank and OECD database. Alptekin & Yeşilaydın (2015) analyzed the 2012 data of 34 countries. 5 clusters were obtained as a result of the analysis performed with Fuzzy C-Means and the similarities of these clusters were evaluated. When the cluster which Türkiye belongs to is examined, it is stated that a common feature of the countries in this cluster, excluding Türkiye, is that they were not the founding countries and joined the OECD later. In the study of Mut & Akyürek (2017), 35 countries for 2013 were analyzed. It was determined that countries should be divided into 3 clusters with HC, and the differences of countries were evaluated with the NHC. It has been determined that all of the health indicators are significantly effective, and Türkiye is below the OECD average. Değirmenci & Yakıcı Ayan (2020), it is aimed to divide with Fuzzy CA and rank countries with Multi Criteria Decision Making (MCDM) and evaluate the position of Türkiye. They analyzed 32 countries with 2015 data. The countries forming 4 clusters were ranked according to their scores. It has been determined that Türkiye is in the cluster with the lowest average score along with 3 countries (Korea, Mexico and Poland).

Peiro-Palomino & Picazo-Tadeo (2018) present a combined prosperity indicator covering the period 2013-2016 for 38 OECD and non-OECD member countries. In the first stage, composite indicators strengthened by Data Envelopment Analysis (DEA) were then sequenced with MCDM. In the second stage, prosperity groups were determined with different HC. Among the 5 clusters obtained, the cluster with the highest number of countries is the cluster with the lowest prosperity level.

Akkuş & Zontul (2019) aimed to divide countries according to development criteria with NHC and Artificial Neural Networks (ANN). In this context, 2015 data from 214 countries were used. The obtained values by the countries were compared according to 12 predetermined parameters; then the position of Türkiye in these clusters were evaluated. The results showed that 16 clusters were formed by both methods. Among the clusters obtained by NHC, Türkiye is in the cluster in which Middle Eastern countries predominate, while it is in the cluster in which developing countries in the Americas predominate among the clusters obtained with ANN.

Dinç Cavlak (2019) examined 154 countries for 3 dimensions (human-environmental-economic welfare) with HC. It is aimed to reveal the sustainability levels of the countries in the current and sub-dimensions with the data of 2016 obtained from different sources (from World Health Organization Healthy Life Expectancy, to Research Institute of Organic Agriculture, International Monetary Fund) and to evaluate the sustainability performances. The countries were grouped into 4 clusters and the clusters differed significantly in the determined dimensions and sub-dimensions.

Levent & Özarı (2019) and Taşçı & Özarı (2019) used MCDM and NHC to rank the countries within the freedom criteria and divided them according to their similarities. The main aim of the study of Levent & Özarı (2019) is to examine the similarities of the ranking obtained with MCDM and the clusters obtained with NHC. 11 G-10 member countries were analyzed for 2017 by using MCDM, and their performances were determined by ranking the economic freedom criteria of the countries. Since it was aimed to compare countries with the results obtained, NHC was used after the cluster number was determined as 2. The results show that countries have less similarity compared to each other in the clusters obtained by NHC. Taşçı & Özarı (2019), on the other hand, ranked 35 OECD member countries with MCDM for the years 2015-2019, using same indicators. These countries are divided into 2 clusters and similar countries are divided with the NHC. The results of the study show that most of the countries that are ranked close to each other are in the same cluster.

In the studies of Levy-Carciente, Phélan & Perdomo (2020), a 12-year comparative analysis of prosperity is presented for the years 2007-2018 using 9 indicators of the LPI for 18 Latin American countries and Spain. An exploratory and descriptive statistical analysis was performed along with trend analysis over time, rates of variation, distributions and means, estimates. They analyzed countries in 3 clusters with HC and revealed that there is a positive trend of convergence between countries in different ways (with the exception of Venezuela), both in general and in specific areas. Budsaratragoon & Jitmaneeroj (2021) created a 4-stage model to identify causal relationships between sub-indices for multidimensional welfare. In the model they created using the 8 indicators of LPI (annual 2015 LPI), CA constitutes the first stage to divide 142 countries. The 7 clusters they obtained in this way formed the categorical variables for DM in the second stage and the accuracy of the model was evaluated. In the third stage, the impact of the sub-indices was determined, and these impact levels were compared with the importance performance analysis in the last stage. The results reveal that education and the student/teacher ratio are the main drivers of welfare. In this study, unlike previous studies, 12 updated prosperity indicators of LPI were utilized. It is thought that this study will contribute to the literature in terms of index and indicators.

In the study of Timor & Yüzbaşı Künç (2021), 35 countries (34 EU members and Türkiye) were divided according to their general economic levels with NHC.

Article	**Method**	**Index**	**Indicators**
Bambra (2007), Kowalski & Wałęga (2015); Abu Sharkh & Gough (2010)	HC (Ward), NHC (K-Means)	Human Development Index	Relative female economic activity rate, maternity leave compensation, compensated maternity leave duration; Life expectancy, literacy and poverty.
Shahbaz et al. (2013)	NHC (K-Means), DT	Economic and Environmental Prosperity Indicators	GDP, Gini, unemployment, population growth, public spending on education, inflation, public health expenditure, Gross National Income, air and water pollution, quantity of CO2, literacy-interest-tax rates, food -crop productions, forest area environmental performance, environmental burden of disease, ecosystem vitality, agriculture, biodiversity & habitat, climate change.
Kangallı et al. (2014)	HC (Ward), NHC (K-Means)	Economic Freedom Index	Business-trade-fiscal-monetary-financial-investment-labor freedoms, public spending, property rights, anti-corruption.
Alptekin & Yeşilaydın (2015); Mut & Akyürek (2017)	NHC (Fuzzy C- Means); HC (Ward), NHC (K-Means)	Global Competitiveness Index (Health)	Life expectancy at birth and maternal mortality rate, percentage of children vaccinated against measles, hospital beds and physicians numbers, years of schooling, health expenditures per capita, percentage of adults who smoke, fruit consumption, carbon monoxide emissions, Gini.
Peiro-Palomino & Picazo-Tadeo (2018)	DEA, HC (Ward, Single Linkage, Average Linkage, Complete Linkage, Centroid)	Better Life Index	Health, work-life balance, safety, housing, community, environment, income, jobs, civic engagement, education.
Akkuş & Zontul (2019)	NHC (K-Means), ANN (Self Organizing Map)	World Data Bank Indicators	Agriculture, death rate, GDP, foreign debt, numbers of bank branches-internet users and female parliamentarians, workability, human rights, technology exports, clean water.
Dinç Cavlak (2019)	HC (Ward)	The Sustainable Society Index	employment, population growth, adequate food and drink, education, healthy living, gender equality, income distribution, good governance, forestry & protected area, consumption, biodiversity, energy use and saving, greenhouse gases, renewable energy, organic agriculture, real savings, GDP, public debt, health protection.
Gülden & Karakış (2019); Levent & Özarı (2019); Taşçı & Özarı (2019)	HC (Ward, Linkage, Neighborhood, Centroid, Median), NHC (K-Means); MCDM (EDAS), NHC (K-Means); MCDM (Grey Relational Analysis), NHC (K-Means)	Economic Freedom Index	Business-labor-monetary-trade-investment-financial freedoms, property rights, state integrity, judicial efficiency, public expenditure, tax burden, financial soundness.
Değirmenci & Yakıcı Ayan (2020)	MCDM (TOPSIS); NHC (Fuzzy C-Means)	Global Competitiveness Index (Health)	Health and pharmaceutical expenditures, Numbers of physicians, nurses and hospital beds.
Levy-Carciente et al. (2020); Budsaratragoon & Jitmaneeroj, (2021)	Exploratory and Descriptive Statistical Analyses, HC (Ward); Expectation Maximization CA, ANN (Bayesian Network-Naive Bayes), Partial Least Square Path Modeling, Importance-Performance Analysis	Legatum Prosperity Index	Open economies (enterprise conditions, economic quality), inclusive societies (safety and security, personal freedom, management, social capital), people empowerment (health, education, natural environment)
Timor and Yüzbaşı Künç (2021)	NHC (K-Means), DT (C5.0)	Economic Prosperity Indicators	GDP, export import rate, unemployment

Table 1. Methods, Indices and Indicators Used in Previous Studies

In the study, the 2016-2019 data of the countries were used and the information economy variables that distinguish the countries from each other in the 3 obtained clusters according to these data were determined. With the results obtained, it has been revealed that it is necessary to give importance to education, R&D and innovation activities in order for countries to progress in terms of welfare and economic level. According to Table 2, it is seen that in previous studies, CA was used alone or, as in this study, supported by different CA algorithms. However, CA has also been used with different DM or statistical methods. Studies mostly cover economy, health and environment-oriented welfare indicators in terms of indices. Furthermore, the LPI is included only in the studies of Levy-Carciente et al. (2020) and Budsaratragoon & Jitmaneeroj (2021) with a limited number of indicators.

3. Material and Methods

Clustering is finding groups of similar data in a data set. Regions where unit density is higher than other regions are called “clusters”. A certain number of groups and accordingly, a general model can be formed from the data by defining these clusters (Shahbaz et al., 2013). In the study, a two-stage CA was applied to the 2019 data of 167 countries within the updated 12 prosperity indicators in the LPI. Countries were firstly divided into groups with Ward’s, one of the HC, and then the similarities between the groups and the position of Türkiye in these groups were determined with K-Means, one of the NHC.

3.1. Legatum Prosperity Index

LPI is a transformation tool that offers original ideas on the formation and change of welfare (LPI, 2019a). It provides a rich and holistic policy-driven dataset that represents more than 99% of the world's population (LPI, 2019a). The index is defined as the combination of two main elements, economic, and social welfare. Economic welfare refers to more than GDP per capita, including quantitative and qualitative aspects not covered by monetary value. Social welfare refers to all aspects of life, including subjective ones, such as happiness and life satisfaction. The definition of welfare is holistic and organized into three broad areas: inclusive societies, open economies, and people empowerment. Welfare may have different levels at different times in each country within the scope of these areas; however, a combination of indicators shown in Table 2 is needed for each country to achieve welfare (Levy-Carciente et al., 2020).

Inclusive Societies	**Open Economies**	**People Empowerment**
Safety and Security	Investment Environment	Living Conditions
Personal Freedom	Enterprise Conditions	Health
Management	Market Access and Infrastructure	Education
Social Capital	Economic Quality	Natural Environment

Table 2. Legatum Prosperity Index Indicators (LPI, 2019b)

According to Table 2, the section of inclusive societies shows the structure of relations between individuals and institutions, and the extent to which these relations provide or hinder social cohesion and development. Social and legal institutions are necessary to protect the fundamental freedoms and developmental abilities of individuals. The section of open economies shows to what extent the economy is open to competition. Economy encourages innovation, investment, business and trade, and facilitates growth. In order for a society to live in prosperity and wealth, the economy must accommodate these ideals. The people empowerment section, on the other hand, shows the quality of people's experiences and the aspects that allow individuals to reach their autonomy and full potential (LPI, 2019b).

3.2. Cluster Analysis

CA divides the units, variables or units and variables in the ungrouped data matrix with unknown natural groupings into subsets that are similar to each other according to their basic characteristics (Akkuş & Zontul, 2019; Alptekin & Yeşilaydın, 2015; Özdamar, 2004). While the variance between the clustered units is minimized, the variance between the groups is maximized (Ketchen & Shook, 1996). As a result of the analysis, the units in the same cluster are close to each other, and different clusters are noticeably far from each other. Thus, while the units forming the cluster are similar to each other, they are dissimilar to the units of other clusters. While similarity is explained as the strength of the relationship between two features, clusters are determined according to dissimilarity, which brings up the issue of how to measure dissimilarity. For this, distance measurements are used, and distance measurements change according to the different measurement units and measurement techniques of the variables (Kangallı et al., 2011; Turan, Özarı & Demir, 2016). In the variables, “distance and correlation” are selected as distance measurements when ratio or interval scales are used, “chi-square or phi-square” is selected in case of counting, "Euclidean, square Euclidean, measure difference, pattern difference, Lance and Williams difference" are selected when binary observations are used (Özdamar, 2004).

After the selection of the distance criterion, the appropriate algorithm for CA is determined. Algorithms determine the rules for measuring distances between units in order to assign cluster membership (Dinç Cavlak, 2019). There are HC and NHC algorithms for dividing units into appropriate number of clusters. The difference between HC and NHC stems from the formation of clusters, their representation and determination of the appropriate number of clusters. HC lays out hierarchical tree diagrams (dendrogram) and can decide how many clusters should be formed, whereas, in NHC, the cluster number is determined by the researchers (Kangallı et al., 2011; Koltan Yılmaz & Patır, 2011; Turan et al., 2016; Dinç Cavlak, 2019; Gülden & Karakış, 2019). HC consists of Additive (Linkage: Single, Average, Complete; Variance: Ward′s; Centralization: Median, Centroid) and Divisive (Monothetic, Polythetic) clustering algorithms, and NHC consists of K-Means algorithms (Koltan Yılmaz & Patır, 2011).

In the study, Ward′s algorithm from HC and K-Means algorithm from NHC were used. The Ward′s, known as the minimum variance method, is based on the average distance of the observation falling in the middle of a cluster from the observations in the same cluster, and utilizes the total deviation squares (Koltan Yılmaz & Patır, 2011). Clusters are determined by the level of closeness or distance of the observations to each other (Dinç Cavlak, 2019). The most similar clusters are combined until all observations are in a single cluster. The optimal cluster number is chosen out of all cluster solutions (Cornish, 2007).

The K-Means algorithm takes k input parameters and divides a series of n objects into k clusters (Ali & Kadhum, 2017; Morissette & Chartier, 2013). After the number of clusters is determined, for each cluster $(\mu_1, \mu_2, \ldots, R)$ the central value is randomly initialized and the calculation continues until formulas (1) and (2) approach each other (Cornish, 2007; Ogbuabor & Ugwoke, 2018; Maylawati et al., 2020).

It is calculated for each *i* value: $c^i \coloneqq argmin_j \left\| x^i - \mu_i \right\|^2 \ldots$ (1)

It is calculated for each j value: $\mu_j \coloneqq \frac{\sum_{i=1}^{m} 1\{c^{(i)}=j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)}=j}\ldots$ (2)

The appropriate number of k clusters can be determined by methods such as different CA/statistical methods, trial and error, the researcher's experience/preliminary knowledge, the number of repetitions of the procedures, the examination of convergence criteria and the cluster validity indexes (Demiralay & Çamurcu, 2005; Kangallı et al., 2011; Shahapure & Nicholas, 2020). In the study, Ward's algorithm was applied to determine the number of clusters in separating countries into clusters. Ward's shares the total Error Sum of Squares (ESS) with K-means partitioning. The solution is generated using random starts of the algorithm and maintaining a solution that minimizes the total ESS in the K-means. Applying the Ward's algorithm to the data, identifying the partition of the objects in the dendrogram and using this partition as the starting approximation are an appropriate solution. The solution can be advanced by iterations of the K-means (Murtagh & Legendre, 2014). However, the interpreting of the obtained dendrogram vary to researchers. It is only used as a preliminary information and different cluster number determination methods are needed (Sökut Açar & Ayman Öz, 2020). Therefore, optimal cluster number obtained from dendrogram in this study is verified by cluster validity index that is measure used to evaluate the clustering (Mamat et al., 2018; Nidheesh et al., 2020). Cluster validity indexes are classified as internal and external. The Elbow method which is internal index (Saputra et al., 2019) was used to determine the validity of a clustering. Internal indexes use the similarity measurements and the clustering results to compute the index (Nidheesh et al., 2020).

The Elbow Method is one of the most popular methods (Chatzopoulos & Derri, 2004; Jeon et al., 2016; Humaira & Rasyidah, 2018; Syakur et al., 2018; Umargono et al.,2019; Yuan & Yang, 2019) to measure the cohesion a cluster (Saputra et al., 2019). Basic idea is identifying the initial optimal cluster number (k) as 2, and increasing this to the maximum for the estimated cluster number (Shi et al., 2021). This method produces graphics that will give ideas of the optimal cluster number. The optimal value was determined by looking at the point that give an angle known as "elbow criterion" in the graph. Performance indicators use number of squared errors (Bholowalia & Kumar, 2014; Wulandari, 2020).

In CA, the normality of distance values is sufficient (Dinç Cavlak, 2019). However, when mean and variance are very different from each other, variables with large mean and variance suppress other variables to a certain extent and reduce their effectiveness relatively. For this reason, it is appropriate to standardize the data or convert them to observed values at certain intervals (Özdamar, 2004).

4. Analysis of Data

The CA methods allows to divide the countries into clusters based on their prosperity levels. The data were analyzed using the IBM SPSS Statistics for Windows, version 20 (IBM Corp., Armonk, N.Y., USA).

The predictor variable(s) and the objects to be clustered are the two main components of CA (Crowther et al., 2021). In this research, countries have been the objects of clustering and LPI indicators have been predictor variables. Researchers also examine the scales of predictor variables, because the extreme values of the variables in the data matrix have negative effects on clustering. For this reasons, each of the variables was converted to standard values known as "Z values" while preparing the data for analysis. Squared Euclidean Distance was chosen as the distance criterion. In the first stage, Ward's algorithm was used. The Ward's algorithm was preferred since the results of the analysis showed clustering more clearly, and it was decided how many clusters countries should be divided into. Also, the Elbow Method was used for verifying of the optimum cluster number. In the second stage, the countries in the clusters, the indicators that are significant in separating the countries into clusters, and the similarities (differences) between the clusters were determined by using the K-Means in line with the number of clusters decided to be suitable. The Elbow Method was repeated for K-Means to verify the optimum cluster number.

5. Results

In the study, countries were divided into clusters according to their prosperity levels and the following findings were obtained.

5.1. Findings Obtained by Ward's Algorithm

With the Ward's algorithm, it is possible to determine how many different clusters countries should be divided into, with a dendrogram. The clusters formed by prosperity levels within the LPI indicators of 167 countries are shown in Figure 1. The dendrogram is scaled from bottom to top as 0-25 units (evenly spaced). Vertical lines show the distance. As distance increases, new countries are added to the cluster formed due to similarity. When the distance is 25 units, a single cluster is formed. Horizontal lines indicate agglomerative clustering. The junction points of the clusters indicate which groups are formed.

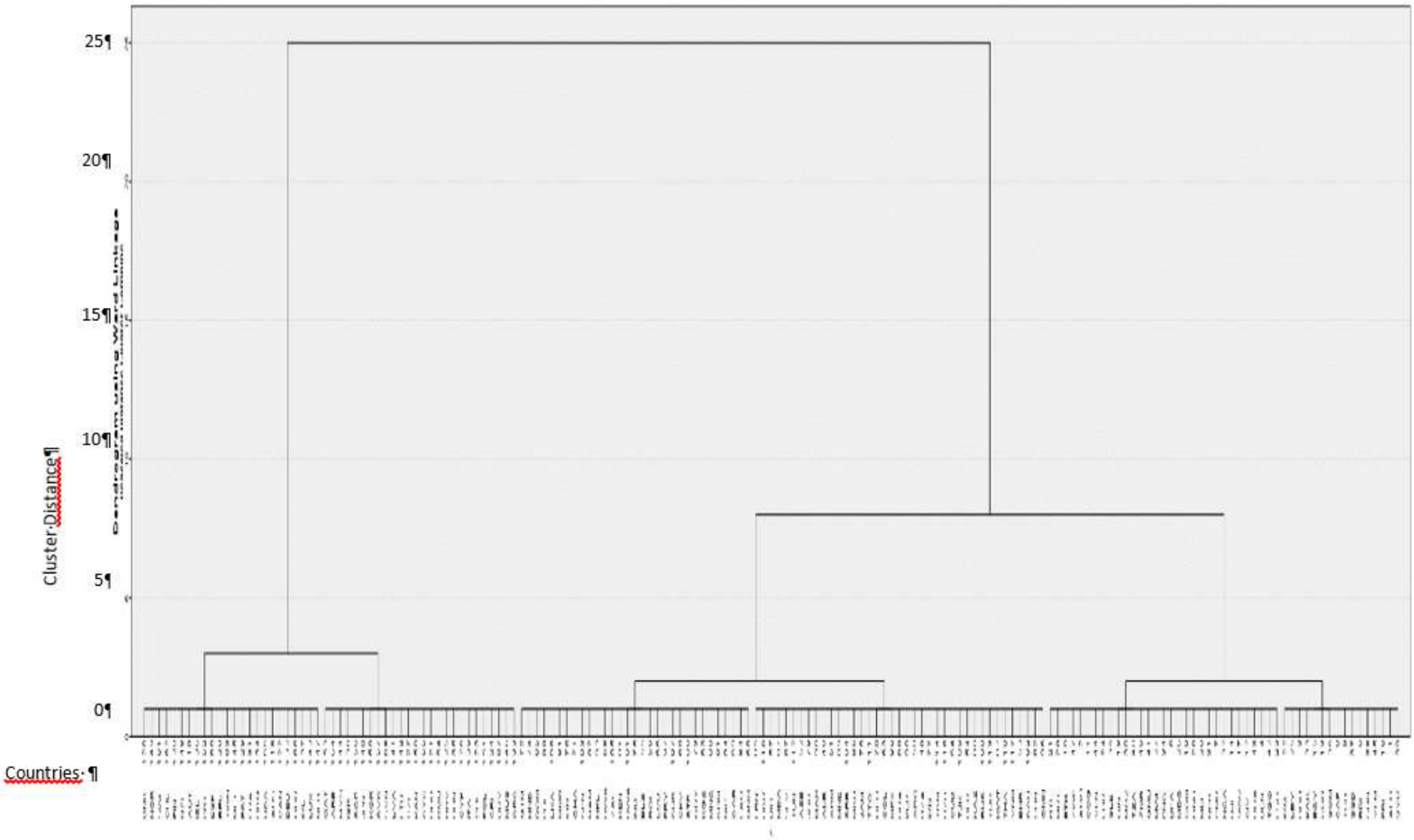

Figure 1. Clusters Formed by Ward's Algorithm

When Figure 1 is examined, it is seen that the countries that are most similar to each other in the dendrogram come together at the closest distance. Countries with very strong similarity in terms of indicators form a group at a distance of 1 unit. In this case, although it can be thought that 5 clusters are formed at a distance of 1 unit, it is seen that the countries recombine at close distances. Therefore, country groups, which are similar to each other, form the 1st cluster at a distance of 3 units, and the 2nd and 3rd clusters at a distance of 2 units. As the similarity of the clusters decreased, at a distance of about 8 units, the 2nd and 3rd clusters merged, and these merged with the 1st cluster at a distance of 25 units and became a single cluster. Accordingly, it was decided that the optimal number of clusters was three in the range of approximately 3-8 units. This can be verified by using Elbow Method. Figure 2 shows the change in agglomeration coefficients (within-cluster sums of squares) (Egloff, et al., 2003) according to the different cluster numbers for Ward's algorithm.

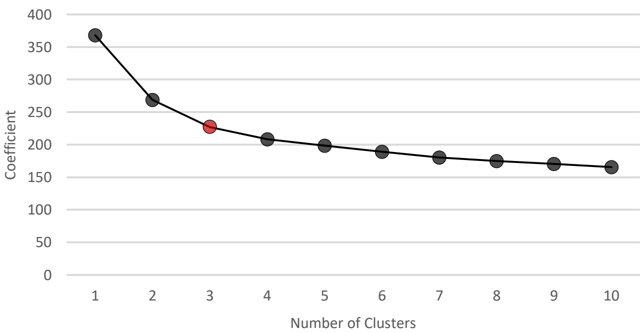

Figure 2. Elbow-criterion by Ward's algorithm

According to Figure 2, coefficients indicate the distance among the countries. A significant increase in the coefficients (Elbow-criterion, the elbow in the curve) indicates an optimal cluster number (Ketchen and Shook, 1996; Chatzopoulos & Derri, 2004). It is seen that the number of the best cluster can be “2” or “3”. Combining the results of the dendrogram and the Elbow method, it was concluded that using three clusters was an appropriate choice.

5.2. Findings Obtained by K-Means Algorithm

In the study, the K-Means was implemented to determine the indicators that are significant in separating the countries into clusters and the similarities that emerged in this direction. While applying the K-Means, the cluster number must be determined by the researcher. Since the countries are divided into 3 clusters with Ward’s algorithm and the Elbow method, the number of clusters has been designated as "3". Also, the Elbow Method was repeated. Figure 3 shows the mean average distances according to the different cluster numbers for K-Means and it is indicated that the number of the best cluster was “3”.

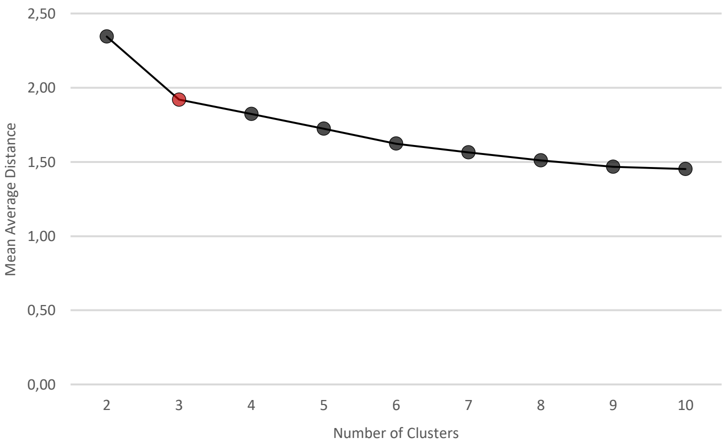

Figure 3. Elbow-criterion by K-Means algorithm

Thus, the 3 clusters obtained can be distinguished as countries with “low, medium and high” prosperity levels within the scope of indicators. As a result of the analysis, tables showing which cluster the countries belong to and the distance between clusters were obtained, and these values are shown respectively in Table 3 and Table 4. The clusters are named C1, C2 and C3 in the tables.

Clusters	Countries	n
C1	Afghanistan, Angola, Bangladesh, Benin, Burkina Faso, Burundi, Cambodia, Cameroon, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Eswatini, Ethiopia, Gambia, Guinea, Guinea-Bissau, Haiti, Iraq, Laos, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mozambique, Myanmar, Nepal, Niger, Nigeria, Pakistan, Papua New Guinea, Senegal, Sierra Leone, Somalia, South Sudan, Sudan, Syria, Tanzania, Togo, Uganda, Venezuela, Yemen, Zambia, Zimbabwe.	53
C2	Albania, Algeria, Argentina, Armenia, Azerbaijan, Bahrain, Belarus, Belize, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Cabo Verde, China, Colombia, Croatia, Cuba, Dominican Republic, Ecuador, El Salvador, Gabon, Georgia, Ghana, Greece, Guatemala, Guyana, Honduras, Hungary, India, Indonesia, Iran, Jamaica, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Kuwait, Lebanon, Mauritius, Mexico, Moldova, Mongolia, Montenegro, Morocco, Namibia, Nicaragua, North Macedonia, Oman, Panama, Paraguay, Peru, Philippines, Qatar, Romania, Russia, Rwanda, Saudi Arabia, São Tomé and Príncipe, Serbia, Seychelles, South Africa, Sri Lanka, Suriname, Tajikistan, Thailand, Trinidad and Tobago, Tunisia, Türkiye, Turkmenistan, Ukraine, Uzbekistan, Vietnam.	73
C3	Australia, Austria, Belgium, Canada, Chile, Costa Rica, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Hong Kong, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Malaysia, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, Singapore, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan China, United Arab Emirates, United Kingdom, United States, Uruguay.	41
Total		167

Table 3. Countries in the Clusters Formed by the K Means Algorithm

According to Table 3, it is seen that among the clusters obtained by the K-Means, there are 53 countries in C1, 73 countries in C2, and 41 countries in C3.

Cluster	C1	C2	C3
C1		3.529	7.454
C2	3.529		4.179
C3	7.454	4.179	

Table 4. Distances Between Final Cluster Centers

When Table 4 is examined, it is seen that C1 and C2 are the closest clusters to each other (3,529), while C1 and C3 are the farthest clusters (7,454). This means that the countries in C1 and C2 are more similar to each other in terms of indicators, while the countries in C1 and C3 are the least similar to each other. The ANOVA results showing the separation of the indicators among the clusters are given in Table 5.

Areas	**Indicators**	**Cluster**		**Error**		**F**	**Sig.**
		Mean Square	**df**	**Mean Square**	**df**		
Inclusive Societies	Safety and Security	40.481	2	.519	164	78.069	.000
	Personal Freedom	42.066	2	.499	164	84.267	.000
	Management	65.803	2	.210	164	313.762	.000
	Social Capital	30.268	2	.643	164	47.067	.000
Open Economies	Investment Environment	67.387	2	.190	164	353.919	.000
	Enterprise Conditions	61.414	2	.263	164	233.301	.000
	Market Access and Infrastructure	68.846	2	.173	164	398.844	.000
	Economic Quality	59.510	2	.286	164	207.734	.000
People Empowerment	Living Conditions	65.112	2	.218	164	298.477	.000
	Health	61.419	2	.263	164	233.370	.000
	Education	67.847	2	.185	164	367.161	.000
	Natural Environment	32.069	2	.621	164	51.630	.000

Table 5. ANOVA Results Showing the Significant of Indicators

When Table 5 is examined, it is seen that all of the indicators had significance in the separating of the countries in 3 clusters ($p<0.05$). Moreover, variables with large F values provide the greatest separation among the clusters. It is seen that the most significant indicators are, respectively, “market access and infrastructure, education, investment environment, management”, and the least significant indicators are “social capital, natural environment”. The difference of the indicators by clusters is given in Table 6.

Areas	Indicators	Clusters		
		C1	C2	C3
Inclusive Societies	Safety and Security	-.81333	-.00376	1.05807
	Personal Freedom	-.68386	-.16746	1.18217
	Management	-.89829	-.16524	1.45540
	Social Capital	-.67965	-.03267	.93673
Open Economies	Investment Environment	-1.08163	.03759	1.33128
	Enterprise Conditions	-.91763	-.10541	1.37388
	Market Access and Infrastructure	-1.13259	.09332	1.29793
	Economic Quality	-.95685	-.04111	1.31010
People Empowerment	Living Conditions	-1.18772	.23284	1.12076
	Health	-1.18907	.29912	1.00451
	Education	-1.19731	.20997	1.17389
	Natural Environment	-.53468	-.20599	1.05793

Table 6. Final Cluster Centers

When Table 6 is examined, the averages of the indicators in the 3 clusters are seen. According to this, C1 has the lowest averages for all indicators. In this regard, it can be said that countries with "low" prosperity levels are included in this cluster. In C1, values closest to the average are observed in the “natural environment, social capital, personal freedom” indicators, while the farthest values are observed in the “living conditions, education, health” indicators. When the averages of C2, which includes Turkiye, are examined, in this cluster, it is observed that “health, living conditions, education” indicators have values above the average, and “natural environment, personal freedom, management” indicators have the lowest values below the average. C3 is the cluster with the highest above-average values of all indicators. In light of this, it can be said that countries with "high" prosperity levels are included in this cluster. This is also in line with the fact that C1, which is seen to be least similar to C3 in Table 4, has a "low" prosperity level. It is observed that the "management, enterprise conditions, investment environment, economic quality" indicators of the countries in C3 are higher than the others. The indicator that can be evaluated as the lowest among the indicators is “social capital”.

6. Conclusion

DM includes methods that enable users to reveal meaningful information from data stacks to make predictions. CA is a DM method that divides data into clusters that are similar to each other based on their fundamental characteristics, but which are not known before. The aim of the study is to divide the countries into clusters according to their prosperity level with CA. Two-stage method consisting of HC and NHC was used to analyze the 2019 data of 167 countries by utilizing updated 12 indicators in the LPI. The countries in the clusters, the indicators that are significant in separating the countries into clusters, and the position of Turkiye in the clusters have been determined.

The results reveal that CA exhibits effective results in order to divide countries according to prosperity levels. As a result of the analysis, countries were divided into 3 clusters. It has been found that all 12 indicators reflecting all areas of LPI are significant in separating countries into clusters, and the most significant indicators are “market access and infrastructure, education, investment environment, management” and the least significant indicators are “social capital, natural environment, safety and security”.

On the basis of clusters, it is possible to evaluate C3, which has the highest averages in terms of indicators, at the "high" prosperity level; C1 with the lowest averages at the "low" prosperity level, and C2 at the "medium" prosperity level. In the C3 countries with high prosperity level, it has been determined that the indicators of “management” representing the field of inclusive societies and “enterprise conditions, investment environment, economic quality” representing the field of open economies are higher than the others in all fields where prosperity is defined according to LPI. This shows that in countries with high welfare, growth is facilitated by innovation, investment, business and trade incentives, and that the relationship and harmony between individuals and institutions is strong. In the C1 countries with the lowest welfare level, it has been determined that the “living conditions, education, health” indicators in the people empowerment field have the lowest values in all fields where welfare is defined according to LPI. This shows that aspects that represent social welfare and enable individuals to reach their autonomy and full potential are weak in low-income countries.

When the prosperity levels are evaluated for Türkiye, it is seen that the cluster (C2) that Türkiye is in is at the "medium" prosperity level. It was determined that the “health, living conditions, education” indicators representing the people empowerment field were above the average and higher than the other indicators of the cluster and C1 in low welfare level in all fields where prosperity is defined according to LPI. On the other hand, “natural environment” representing the people empowerment field and “personal freedom, management” representing the inclusive societies field had the lowest values. This shows that countries in C2 are better than C1 in matters that represent social welfare, and that the relationship structures between institutions and individuals, which are necessary for individuals′ fundamental freedoms and self-development, are weak.

Levy-Carciente et al. (2020) compared the 12-year development of Latin American countries with Spain using HC. According to the LPI, they determined that the countries closest to Spain and showing the most positive developments are Costa Rica, Panama, Chile and Uruguay. In addition, it has been determined that Venezuela is different from all regions and is the most negative country in terms of all indicators. When the placement of the same countries is examined in this study, it is seen that Spain is in the cluster with the highest prosperity levels, along with Costa Rica, Chile and Uruguay. Venezuela, on the other hand, is similarly in the cluster with the lowest prosperity levels. In another study that grouped countries using CA, Budsaratragoon & Jitmaneeroj (2021) revealed that education is the most important indicator of increasing prosperity. Even in this study, education was determined as the second most significant indicator in determining the level of prosperity in general.

According to the above results, in the study, it has been revealed that CA is effective in dividing countries into clusters according to their prosperity levels. A general model can be formed from the data with the help of obtained clusters. Insights about the data can be produced by interpreting the model, and new information can be obtained by using these data again and supporting it with different methods. This is also one of the main purposes of the DM. In future studies, it will be possible to reach more detailed information with new and different patterns by using different DM methods or increasing the amount of data.

References

Abu Sharkh, M. & Gough, I. (2010). Global welfare regimes a cluster analysis. Global Social Policy, 10(1), 27-58. DOI: 10.1177/1468018109355035

Akar, S. (2014). Türkiye'de daha iyi yaşam endeksi: OECD ülkeleri ile karşılaştırma [The better life index in Turkey: Comparison with OECD countries]. Journal of Life Economics, 1, 1-12. DOI: 10.15637/jlecon.201416987

Akar, H. (2015). Farklılaşan refah ölçüm yöntemleri ve eğitim açısından Türkiye'nin değerlendirilmesi [The diversifying measures of welfare and evaluation of Turkey in the context of education]. Finance, Politics & Economic Reviews, 52(606), 21-38. Retrieved from http://www.ekonomikyorumlar.com.tr/files/articles/152820006117_2.pdf

Akkuş, B. & Zontul, M. (2019). Veri madenciliği yöntemleri ile ülkeleri gelişmişlik ölçütlerine göre kümeleme üzerine bir uygulama [An application on clustering countries with data mining methods based on development criteria]. AURUM Journal of Engineering Systems and Architecture, 3(1), 51-64. Retrieved from https://dergipark.org.tr/tr/pub/ajesa/issue/47400/598179

Albayrak, A. S. & Koltan Yılmaz, Ş. (2009). Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama [Data mining: Decision tree algorithms and an application on ISE data]. Süleyman Demirel University Journal of Faculty of Economics Administ. Sciences, 14(1), 31-52. Retrieved from https://dergipark.org.tr/tr/pub/sduiibfd/issue/20831/223135

Ali, H. H. & Kadhum, L. E. (2017). K-Means clustering algorithm applications in data mining and pattern recognition. International Journal of Science and Research (IJSR), 6(8), 1577-1584.

Alptekin, N. & Yeşilaydın, G. (2015). OECD ülkelerinin sağlık göstergelerine göre bulanık kümeleme analizi ile sınıflandırılması [Dividing OECD countries according to health indicators using fuzzy clustering analysis]. Journal of Business Research Turk, 7(4), 137-155. Retrieved from https://isarder.org/index.php/isarder/article/view/274

Bambra, C. (2007). Defamilisation and welfare state regimes: A cluster analysis. International Journal of Social Welfare, 16, 326-338. DOI: 10.1111/j.1468-2397.2007.00486.x

Better Life Index (BLI) (2019). "What's the Better Life Index?", http://www.oecdbetterlifeindex.org/about/better-life-initiative/, (Accessed Date: 28.12. 2020).

Bholowalia, P. & Kumar, A. (2014). EBK-Means: A Clustering technique based on Elbow method and K-Means in WSN. International Journal of Computer Applications, 105(9), 17-24. DOI: 10.5120/18405-9674

Budsaratragoon, P. & Jitmaneeroj, B. (2021). Reform priorities for prosperity of nations: The Legatum Index. Journal of Policy Modeling, 43, 657-672. DOI: 10.1016/j.jpolmod.2020.09.004

Büchs, M. (2021). Sustainable welfare: Independence between growth and welfare has to go both ways. Global Social Policy, 21(2), 323-327. DOI: 10.1177/14680181211019153

Chatzopoulos, D. & Derri, V. (2004). Grading profiles of high school physical educators: A cluster analysis. Journal of Human Movement Studies, 47, 061-073.

Cornish, R. (2007). Statistics: Cluster Analysis. Mathematics Learning Support Centre, http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf, (12.05.2022).

Crowther, D., Kim, S., Lee, J., Lim, J. & Loewen, S. (2021). Methodological synthesis of cluster analysis in second language research. Language Learning, 71(1) 99-130. DOI: 10.1111/lang.12428

Demiralay, M. & Çamurcu, A. Y. (2005). CURE, AGNES ve K-Means algoritmalarındaki kümeleme yeteneklerinin karşılaştırılması [Comparison of clustering characteristics of CURE, AGNES and K-Means algorithms]. İstanbul Commerce University Journal of Science, 4(8), 1-18. Retrieved from https://dergipark.org.tr/tr/pub/ticaretfbd/issue/21348/229000

Değirmenci, N. & Yakıcı Ayan, T. (2020). OECD ülkelerinin sağlık göstergeleri açısından bulanık kümeleme analizi ve TOPSIS yöntemine göre değerlendirilmesi [Evaluation of OECD countries according to fuzzy clustering analysis and TOPSIS method in terms of health indicators]. Hacettepe University Journal of Economics and Administrative Sciences, (38)2, 229-241. DOI: 10.17065/huniibf.592991

Dinç Cavlak, Ö. (2019). Sürdürülebilir toplum göstergelerinin hiyerarşik kümeleme analizi yöntemiyle incelenmesi [Hierarchical clustering analysis of sustainable society indicators]. Third Sector

Social Economic Review, 54(4), 2053-2073. DOI: 10.15659/3.sektor-sosyal-ekonomi.19.12.1125

Egloff, B., Schmukle, S. C., Burns, L. R., Kohlmann, C. W. & Hock, M. (2003). Facets of dynamic positive affect: Differentiating joy, interest, and activation in the positive and negative affect schedule (PANAS). Journal of Personality and Social Psychology, 85(3), 528-540. DOI: 10.1037/0022-3514.85.3.528

Global Competitiveness Index (GCI) (2019). "Global Competitiveness Report", https://www.weforum.org/ reports/how-to-end-a-decade-of-lost-productivity-growth, (Accessed Date: 11. 01. 2021).

Gülden, T. & Karakış, E. (2019). OECD ülkelerinin ekonomik özgürlüklerine göre kümeleme analizi ile sınıflandırılması [Classification of OECD countries according to economic freedom with cluster analysis]. Journal of Economics and Administrative Sciences, 20(2), 297-316. Retrieved from http://esjournal.cumhuriyet.edu.tr/tr/pub/issue/50375/614708

Han, J., Kamber, M. & Pei, J. (2012). Data mining concepts and techniques. Waltham: Morgan Kaufmann Publishers is An Imprint of Elsevier.

Humaira, H. & Rasyidah, R. (2018). Determining the appropiate cluster number using Elbow method for K-Means algorithm, WMA-2 2018, DOI: 10.4108/eai.24-1-2018.2292388

Jeon, J. Y., Choi, J. S. & Byun, H. G. (2016). Implementation of Elbow method to improve the gases classification performance based on the RBFN-NSG Algorithm. Journal of Sensor Science and Technology, 25(6), 431-434. DOI: 10.5369/JSST.2016.25.6.431

Kangallı, S. G., Uyar, U. & Buyrukoğlu, S. (2014). OECD ülkelerinde ekonomik özgürlük: bir kümeleme analizi [Economic freedom in OECD countries: A cluster analysis]. International Journal of Alanya Faculty of Business, 6(3), 95-109. Retrieved from https://dergipark.org.tr/tr/pub/uaifd/issue/21601/231994

Ketchen, D. J. Jr. & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. Strategic Management Journal, 17, 441-458. Retrieved from http://www.jstor.org/stable/2486927

Koltan Yılmaz, Ş. & Patır, S. (2011). Kümeleme analizi ve pazarlamada kullanımı [Cluster analysis and its usage in marketing]. Journal of Academic Approaches, 2(1), 91-113. Retrieved from https://dergipark.org.tr/tr/pub/ayd/issue/3325/46150

Kowalski, R. & Wałęga, G. (2015). Defamilisation in Central and Eastern Europe: A Cluster Analysis. The 9th International Days of Statistics and Economics, September 10-12, Prague, 855-863.

Legatum Prosperity Index (LPI) (2018). "The Legatum Prosperity Index 2018 Report", https://li.com/reports/2018-legatum-prosperity-index/, (Accessed Date: 09.04.2020).

Legatum Prosperity Index (LPI) (2019a). "The Legatum Prosperity Index", https://www.prosperity.com/about/summary/, (Accessed Date: 09.04.2020).

Legatum Prosperity Index (LPI) (2019b). "The Legatum Prosperity Index Methodology Report", https://prosperitysite.s3-accelerate.amazonaws.com/7515/8634/9002/Methodology_for_Legatum_Prosperity_Index_2019.pdf, (Accessed Date: 09.04.2020).

Levent, M. & Özarı, Ç. (2019). EDAS yöntemi ve kümeleme analizi ile G-10 ülkelerinin ekonomik özgürlük kriterleri ile değerlendirilmesi [Evaluating economic freedom′ criterias of G-10 countries with EDAS method and cluster analysis]. The Journal of Turk-Islam World Social Studies, 6(22), 219-235. DOI: 10.29228/TIDSAD.30876

Levy-Carciente, S., Phélan, C. M. & Perdomo, J. (2020). Prosperity in Spain and Latin America: myths and facts. International Journal of Advance Study and Research Work, 3(7), 2581-5997. DOI: 10.5281/zenodo.3958006

Mamat, A. R., Mohamed, F. S., Mohamed, M. A., Rawi, N. M. & Awang, M. I. (2018). Silhouette index for determining optimal k-means clustering on images in different color models. International Journal of Engineering & Technology, 7(2.14), 105-109. DOI: 10.14419/ijet.v7i2.14.11464

Markou, G., Palaiolouga, E., Kokkinakos, P., Markaki, O., Koussouris, S. & Askounis, D. (2015). "Prosperity Indicators: A Landscape Analysis", http://ceur-ws.org/Vol-1553/paper6.pdf, (Accessed Date: 08.02.2021).

Maylawati, D.S., Priatna,T. Sugilar, H. & Ramdhani, M.A. (2020). Data science for digital culture improvement in higher education using K-means clustering and text analytics. International

Journal of Electrical and Computer Engineering (IJECE), 10(5), 4569-4580. DOI: 10.11591/ijece.v10i5.pp4569-4580

Morissette, L. & Chartier, S. (2013). The K-Means clustering technique: General considerations and implementation in Mathematica, Tutorials in Quantitative Methods for Psychology, 9(1), 15-24. DOI:10.20982/tqmp.09.1.p015

Murtagh, F. & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?. Journal of Classification, 31, 274-295. DOI: 10.1007/s00357-014-9161-z

Mut, S. & Akyürek, Ç. E. (2017). OECD ülkelerinin sağlık göstergelerine göre kümeleme analizi ile sınıflandırılması [Classifying OECD countries according to health indicators using clustering analysis]. International Journal of Academic Value Studies (Javstudies), 3(12), 411-422. DOI: 10.23929/javs.283

Nidheesh, N., Abdul Nazeer, K. A. & Ameer, P. M. (2020). A Hierarchical Clustering algorithm based on Silhouette Index for cancer subtype discovery from genomic data. Neural Computing and Applications, 32, 11459-11476. DOI: 10.1007/s00521-019-04636-5

Ogbuabor, G. & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using Silhouette score value. International Journal of Computer Science & Information Technology (IJCSIT), 10(2), 27-37. DOI: 10.5121/ijcsit.2018.10203

Özdamar, K. (2004). Paket Programlar İle İstatistiksel Veri Analizi 2 [Statistical Data Analysis with Package Programs 2]. Eskişehir, Turkey: Kaan Kitabevi.

Peiro-Palomino J. & Picazo-Tadeo, A.J (2018). OECD: one or many? Ranking countries with a composite well-being indicator. Soc Indic Res, 139, 847-869. DOI: 10.1007/s11205-017-1747-5

Saputra, D. M., Saputra, D. & Oswari, L. D. (2019). Effect of distance metrics in determining k-value in Kmeans clustering using Elbow and Silhouette method. Advances in Intelligent Systems Research, 172, 341-346. DOI: 10.2991/aisr.k.200424.051

Shahapure, K. R. & Nicholas, C. (2020). "Cluster quality analysis using Silhouette score", 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 747-748.

Shahbaz, M., Iftikhar, M. & Mahmood, R. (2013). Classification based on Empathy level by Mining Economic Prosperity and Environmental Indicators. International Journal for e-Learning Security (IJeLS), 3(2), 340-349. DOI: 10.20533/ijels.2046.4568.2013.0043

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H. & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. J Wireless Com Network, 31. DOI: 10.1186/s13638-021-01910-w

Social Progress Index (SPI) (2020). "Global Index: Overview", https://www.socialprogress.org/index/global, (Accessed Date: 24.12.2020).

Söküt Açar, T. & Ayman Öz, N. (2020). The determination of optimal cluster number by Silhouette index at clustering of the European Union member countries and candidate Turkey by waste indicators. Pamukkale Univ Muh Bilim Derg, 26(3), 481-487. DOI: 10.5505/pajes.2019.49932

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S. & Satoto, B. D. (2018). Integration K-Means clustering method and Elbow method for identification of the best customer profile, IOP Conf. Series: Materials Science and Engineering, 336, 012017.

Taşçı, M. & Özarı, Ç. (2019), OECD ülkelerinin ekonomik özgürlük göstergelerinin K-Ortalamalar kümeleme yöntemi ve Gri İlişkisel Yöntemi ile analizi [Evaluating economic freedom criterias of OECD countries with grey relational analysis method and cluster analysis]. The Journal of Academic Social Science, 7(96), 464-488. DOI : 10.29228/ASOS.36738

Timor, M. & Yüzbaşı Künç, G. (2021). Ekonomik gelişmişliği etkileyen bilgi ekonomisi değişkenlerinin veri madenciliği ile belirlenmesi [Determination of knowledge economy variables that affect economic development by using data mining]. Optimum Journal of Economics and Management Sciences, 8(1), 1-18. DOI: 10.17541/optimum.748237

Turan, K. K., Özarı, Ç. & Demir, E. (2016). Kümeleme analizi ile Türkiye ve Ortadoğu ülkelerinin ekonomik göstergeler açısından karşılaştırılması [Comparing Turkey and The Middle East countries with cluster analysis: Economic perspective]. Istanbul Aydin University Journal, 29, 143-165. DOI: 10.17932/IAU.IAUD.m.13091352.2016.8/29.143-165

Tüzüntürk, S. (2010). Veri madenciliği ve istatistik [Data mining and statistics]. Bursa Uludağ Journal of Economy and Society, 29(1), 65-90.

Umargono, E., Suseno, J. & Vincensius Gunawan, S. K. (2019).K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median, In Proceedings of the International Conferences on Information System and Technology (CONRIST 2019), 234-240.

United Nations Development Program (UNDP) Human Development Index (HDI) (2019). "About Human Development", http://hdr.undp.org/en/humandev, (Accessed Date: 13. 07. 2020).

World Happiness Report (WHR) (2020). "World Happiness Report", https://worldhappiness.report/, (Accessed Date: 28.12.2020).

Wulandari, S. (2020). Analyze k-value selected method of k-means clustering algorithm to clustering province based on disease case. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(3), 121-124.

Yuan, C. & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. J, 2(2), 226-235. DOI: 10.3390/j2020016