Original Paper

European Journal of Digital Economy Research
2020 Vol.1 Issue.2 pp.57-68

CLUSTERING LIVABLE COUNTRIES USING SELF-ORGANIZING MAPS

Meltem KARAATLI*, Merve ÇOLAK, Bülent KARAATLI*****

* Assoc. Prof. Dr., Süleyman Demirel University, Faculty of Economics and Administrative Sciences, Dept. of Business Administration, meltemkaraatli@sdu.edu.tr, https://orcid.org/0000-0002-7403-9587.

** Corresponding Author, Ph.D. Candidate, Süleyman Demirel University, Institute of Social Sciences, ie_mervecolak@yahoo.com, https://orcid.org/0000-0001-5327-0678.

*** Dr., bkaraatli@hotmail.com, https://orcid.org/0000-0003-4834-5828.

ABSTRACT

People living in developing and underdeveloped countries are looking for a way to live in countries that are developed in terms of democracy, have high economic welfare, and have high social life. In this study, considering the 2017 "Better Life Index" data, prepared by the OECD a cluster analysis has been conducted for the livable countries. Better Life Index consists of 22 variables in housing, income, employment, society, education, environment, civil participation, health, life satisfaction, security, work life balance groups. One-way ANOVA and Kruskal Wallis tests are also carried out in order to examine the presence of any statistically significant difference among the clusters.

Keywords: *Better Life Index, OECD, Clustering Analysis, Livable Countries, Data Mining, Self-Organizing Maps.*

1. INTRODUCTION

Data mining, which is also known as knowledge discovery, machine learning, knowledge mining, is the extraction process of valuable information from big data structures (Jaseena & David, 2014: 131). The aim of data mining is to find out hidden pattern, increase the value of data and convert data to knowledge (Burbidge & Buxton, 2001:3). Data mining also determines the rules for predicting the future through using available data (Parlar & Kakıllı Acaravcı, 2017: 693).

Information technologies, which enable long term storage of high-volume information, lead widespread usage of data mining applications (Fayyad et al. 1996: 28). Data mining models are mainly descriptive and predictive. Patterns that will be used in decision making constitue descriptive models. In descriptive models, the patterns that will be used in decision making are defined. In predictive models, however, results are are estimated through existing data. (Jain & Srivastava, 2013: 116).

“Clustering Analysis”, which is one of the data mining methods, used in the study and shown in the descriptive model, is a commonly used technique that allows to classify the examined units according to their similarities in certain groups, to reveal the common attibutes of the units and to make definitions about the created groups. (Chen & Wang, 2008: 4262).

The Organization for Economic Cooperation and Development (OECD) was created in 1961 with the Paris Peace Agreement. Defining itself as an international organization that aims to create better policies for better lives, OECD offers governments a platform where they can share their experiences with each other and produce mutual solutions to problems. OECD tries to understand the factors behind economic, social and environmental changes, and measures global trends and productivity in the fields of trade and investment (OECD, 2020).

The Better Life Index, published since 2011 by OECD is generated through housing, revenue, employment, society, environment, civic engagement, health, life satisfaction, security, and work-life balance criterion in order to compare prosperity levels of countries on different fields (Do Carvalhal Monteiro et al. 2019: 478).

In study, 38 countries were grouped based on 22 different variables through clustering analysis. Clustering analysis allows collecting the countries in groups in terms of the specified variables (attributes), and also allowing each group (cluster) to be identified and differentiate from each other so that livability can be assessed from different aspects. The aim of this study is to make clustering analysis using “Better Life Index” data belongs to 35 OECD member countries as well as South Africa, Russian Federation and Brazil. Although

South Africa, Russian Federation and Brazil are not members of OECD, these countries are possible members that OECD cooperate with through the "The Centre for Cooperation with Non-Members" established within the organization (MFA, 2020).

In the study, Artificial Neural Networks (ANN) based Self-Organizing Maps (SOM) used for clustering analysis. Furthermore, Dunn Index is used to determine the optimum cluster numbers. In the following sections, a brief literature review, methodology and the findings of the analysis will be given.

2. LITERATURE REVIEW

Some recent clustering studies are as follow:

Azar et al. (2013) applied clustering analysis using the data of the patients with thyroid, a hormone which plays an important role on the regulations of body functions. At this research all the cluster numbers from 2 to 11 are tested and the optimum number is concluded as three.

In the study of Hudson et al. (2016), Australian railway workers; were divided into four clusters according to asleep and awake hours by using SOM method.

Wu et al. (2016) also used SOM along with Kernel Principal Component Analysis (KPCA) and Support Vector Machine (SVM) methods in order to classify the quality of a solar cell patents in the solar industry.

Voutilainen and Arvola (2017) using 27 meteorological, biologic, physical and chemical variables for a duration of 21 years of a lake found in Finland, obtained three clusters by SOM Methodology at their study.

Pal et al. (2018), using World Health Organization (WHO) accident database, evaluated 176 countries in terms of 44 attributes. SOM technique is used in the study divided and countries into three clusters. There are 89 countries in the first cluster of middle-income countries, 51 countries in the second cluster of low-income countries, and 36 countries in the third cluster of high-income countries.

Özdemir and Kaya (2018) in their study, clustered OECD member countries according to carbon dioxide emission indicators derived from fossil fuel consumption by using K-Medoids and Fuzzy C-Means algorithms. Countries have divided into two clusters according to K-Medoids method and four clusters according to Fuzzy C-Means method.

Chang and Chen (2018) in their study, clustered OECD member countries according to PISA 2015 data. Countries are divided into three clusters in the study using hierarchical clustering.

In the study of Do Carvalhal Monteiro et al. (2019) OECD member countries clustered according to better life index data using a proposed new clustering analysis approach. In addition, Silhouette coefficient was used to determine the optimal number of clusters.

3. METHODOLOGY

3.1. Artificial Neural Networks (ANN)

ANN is a method inspired from the human brain that can learn and produce new information. Learning is achieved by training ANN with examples. A trained ANN can carry out mathematically, the procedures like, data classification, identification, data association, optimisation and prediction for the future (Maind & Wankar, 2014: 96).

An ANN is composed of neural cells called neurons which are interconnected in different forms and generally organized as layers. Three layers (input layer, hidden layer and output layer) are generally found in ANN. Input Layer receives information from the exterior while the output layer extracts the results. While input and output are composed of single layers, many hidden layers can be found between them. (Park et al. 1991: 444). Hidden layer is the part where, the data received from the input layer are transferred to the output layer. (Sarle, 1994: 4).

3.2. Self Organizing Map

SOM, also known as the Kohonen Network, is a ANN-based unsupervised learning algorithm used to perform clustering in data analysis (Gassen et al. 2015; 636; 680: Jin et al., 2015: 84). Nodes found at the input layer of SOM algorithm signifies variables, and the nodes at the output layer (Kohonen Layer) construct clusters (He & Deng, 2005: 252).

In SOM-based clustering analysis, assuming that there are m variables and n clusters, the distance of each input from each set is examined. In order to form sets, the distance between the weight vector (w) of each variable and the input vector (x) is calculated. Equality (1) is used when calculating the distance.

$$d_i = \sqrt{\sum_{j=1}^{m}\left(x_j - w_{ij}\right)^2} \qquad (1)$$

The steps of SOM-based clustering analysis are as follows (Kohonen, 1990: 1465; Gan & Wu, 2007: 56-59):

Step 1: An initial random value between 0 and 1 is attributed to weight vector.

Step 2: The distance between each input and weight vector is calculated through equation 1.

Step 3: Input having the shortest distance between the distances calculated at the previous step is chosen and is entitled as the winner node.

Step 4: Through the winner node and learning parameter, weights are updated by means of Equation 2. It is assumed that α has a value between 0 and 1, generally close to 0.

$$w_j(new) = w_j(current) + \alpha\left[x - w_j(current)\right] \text{ (2)}$$

Step 5: Until the ending conditions are met, step 3 and step 4 are reiterated.

The biggest advantage of SOM-based clustering analysis is that it can reduce a multidimensional data set to two dimensions.

3.3. Dunn Index

Several methods are used to measure the cluster quality. The Dunn Index is one of those methods. Minimizing the distance among clusters and maximizing the in-cluster distance is the main objective of the Dunn Index. Higher values of the Dunn Index indicate the quality of the cluster. (Azar, 2013: 8). The disadvantage of the index is the extension of the analytical need necessary fort he calculation of increasing c and n numbers and calculation difficulties. Dunn index is calculated as at the equation 3.

The disadvantage of the index is that as the number of c and n increases calculation becomes more difficult. The Dunn Index is calculated as in Equation 3

$$\mathrm{DI}(c) = min_{i\in c}\left\{min_{j\in c, i\neq j}\left\{\frac{min_{x\in c_i, y\in c_j} d(x,y)}{max_{x,y\in c} d(x,y)}\right\}\right\} \text{ (3)}$$

4. ANALYSES

Better Life Index is prepared by the OECD and taken from the Internet address of Turkey Statistical Institute (TUIK, 2017). In this study, housing, revenue, jobs, community, education, environment, civic engagement, health, life satisfaction, safety, work-life balance are examined by considering 22 different variables from 38 countries. The variables are dwellings without basic facilities (%), housing expenditures (%), room per person (%), household net adjusted disposable income ($), household net wealth ($), labor market insecurity (%), employment rate (%), long-term unemployment rate (%), personal earnings ($), educational attainment (%), years in education, air pollution (mcg per m3), water quality (%), stakeholder engagement for the developed regulations (average score), voter turnout (%), life expectancy (years), health declaration (%), life satisfaction (average score), feeling safe walking alone at night (%), murder rate, employees working very long hours (%), time devoted to leisure and personal care (hours).

Clustering analysis will enable countries to be divided into groups in terms of specified variables (attributes), enabling each group (cluster) to be identified, differences in groups emerging, and an assessment of livability in the context of the specified variables.

The set of variables of 38 countries are seen at Table 1. Main purpose of this study is to group these 38 countries in terms of livable countries using SOM algorithm that is an ANN based clustering technique and interpret obtained results. Moreover, One-way ANOVA and Kruskal Wallis tests are also carried out in order to examine the presence of any statistically significant difference among the clusters.

Table 1. Variables Used in the Study

Kod	Variable Name	Kod	Variable Name
X_1	Dwellings without basic facilities (%)	X_{12}	Air pollution (mcg per m3)
X_2	Housing expenditures (%)	X_{13}	Water quality (%)
X_3	Room per person (%)	X_{14}	Stakeholder engagement for the developed regulations (average score)
X_4	Household net adjusted disposable income ($)	X_{15}	Voter turnout (%)
X_5	Household net wealth ($)	X_{16}	Life expectancy (years)
X_6	Labor market insecurity (%)	X_{17}	Health declaration (%)
X_7	Employment rate (%)	X_{18}	Life satisfaction (average score)
X_8	Long-term unemployment rate (%)	X_{19}	Feeling safe walking alone at night (%)
X_9	Personal earnings ($)	X_{20}	Murder rate
X_{10}	Educational attainment (%)	X_{21}	Employees working very long hours (%)
X_{11}	Years in education	X_{22}	Time devoted to leisure and personal care (hours).

The number of clusters in SOM algorithm are determined by the user. At this study, the Dunn Index was used to limit the number of clusters. The Dunn Index values are found in Table 2.

Table 2. The Dunn Index Values with Respect to Different Number of Clusters

Number of Clusters	The Dunn Index Values
2	0,26
3	0,27*
4	0,27*
5	0,25

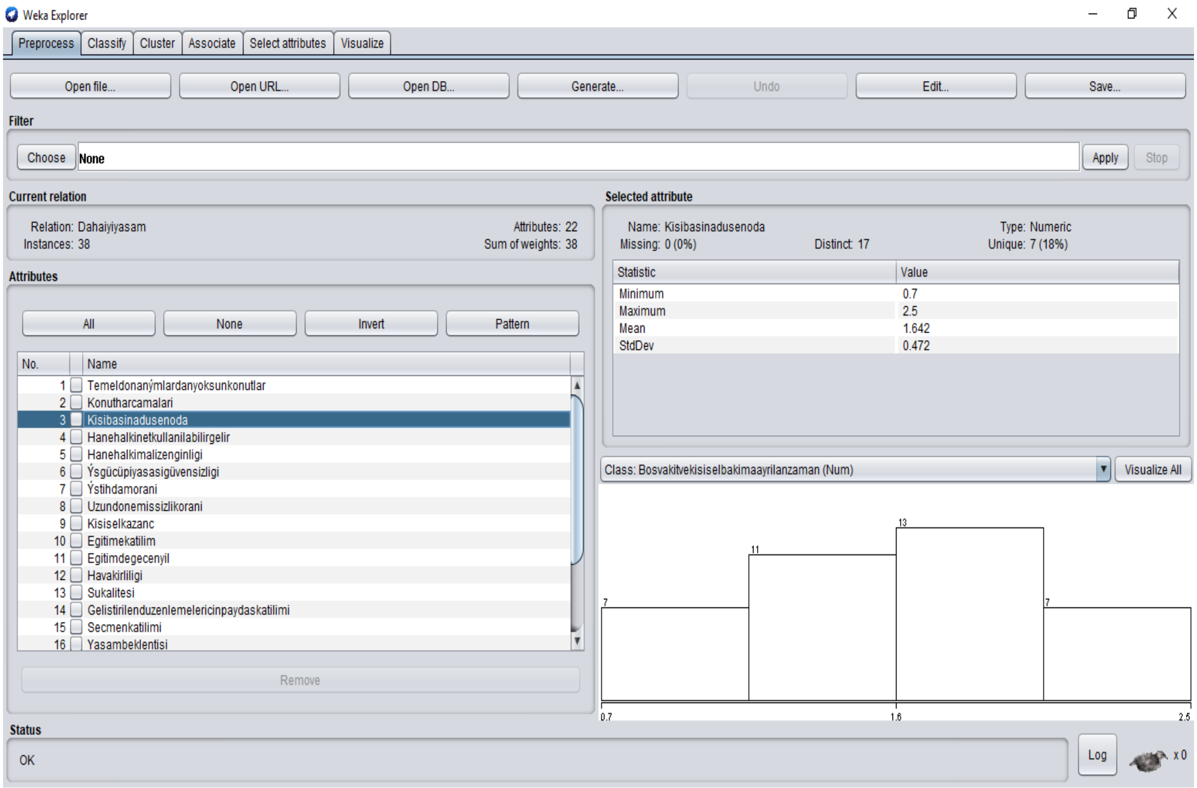
As seen in Table 2, the highest Dunn value is obtained if the countries are clustered in three or four groups.

Everitt (1974) asserts that if the number of clusters cannot be determined, then Equation 4 can be used.

$$K=\sqrt{\frac{n}{2}} \tag{4}$$

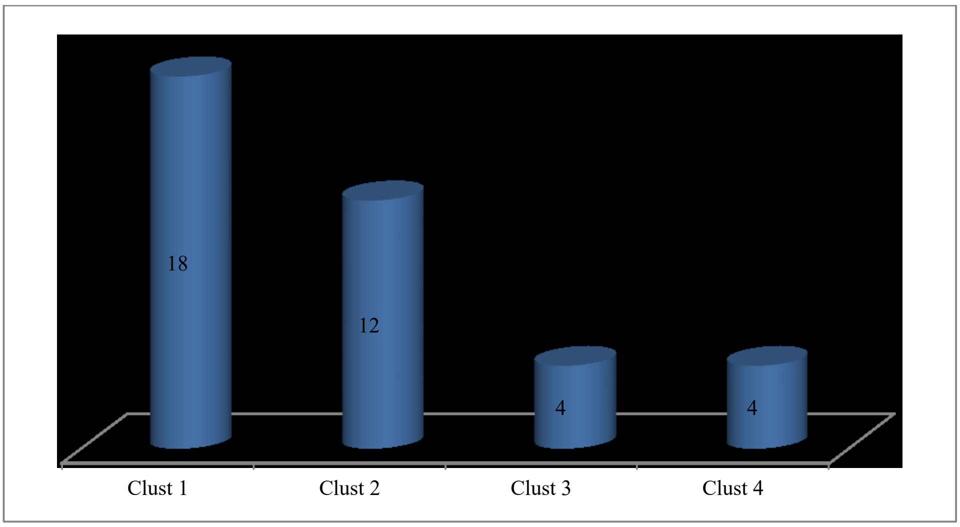
Equation 4 suggests that the optimum number of clusters for this study is four. After determining the number of clusters, basic statistical analyzes are conducted through the WEKA program. In Figure 1, the ranges, average and standard deviation values and histogram graph of the "room per person" variable are seen as an example.

Figure 1. Screenshot from the WEKA Window



In Figure 2, the number of countries in each cluster is shown on a bar graph.

Figure 2. Number of Countries in Each Cluster

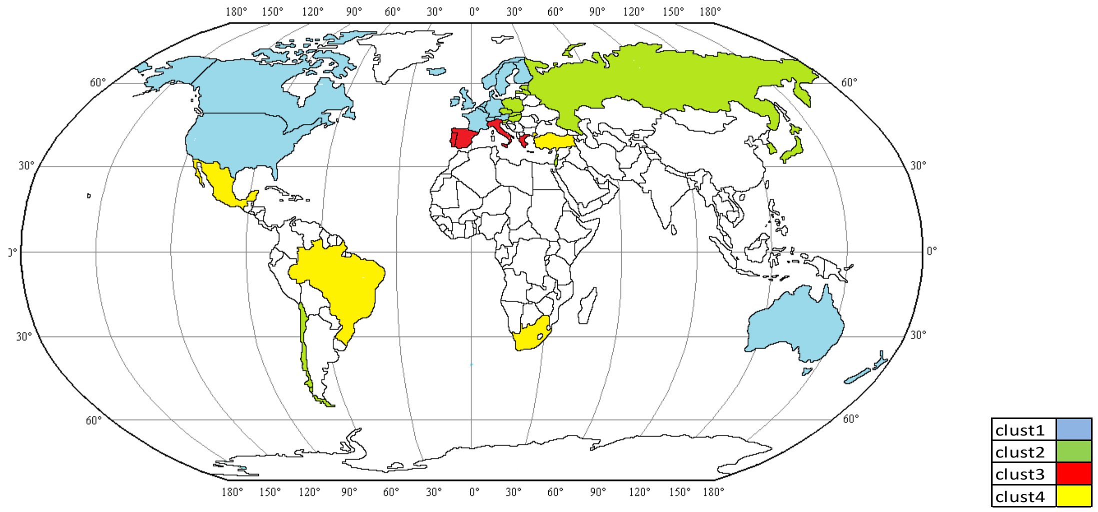


According to the findings obtained from the SOM Algorithm, there are 18 countries in "Cluster 1", 12 in "Cluster 2", 4 in "Cluster 3" and 4 in "Cluster 4". The clusters with the countries are presented in Table 3, while visualisation of the clusters are given in a world map in Figure 3.

Table 3. Clusters

Clusters	Countries
Cluster 1	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Netherlands, Luxembourg, Norway, New Zealand, Sweden, Switzerland, United States of America
Cluster 2	Czech Republic, Estonia, Israel, Japan, Korea, Latvia, Hungary, Poland, Slovakia, Slovenia, Chile, Russian Federation
Cluster 3	Spain, Italy, Portugal, Greece
Cluster 4	Mexico, Turkey, Brazil, South Africa

Figure 3. Visualisation of the Clusters on the Map



Following the clustering procedure, an examination in detail were made to through the One-Way ANOVA and Kruskal Wallis test analysis which determine whether or not there exist a statistically significant difference among clusters and normal distribution of variables. Normality tests through Kolmogorov-Smirnov were performed for each variable of clusters firstly. Then secondly, One-Way ANOVA test for the variables meeting the normality assumption, Kruskal Wallis test for the variables that are not meeting the assumption are used to determine the existence of a statistically significant difference. According to Kolmogorov-Smirnov test, variables X_2, X_3, X_4, X_5, X_7, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{18}, X_{19}, X_{22} provide assumption of normality. However, because of Levene test statistic value of X_{18} and X_{22} variables is less than 0.05, Kruskal Wallis test was applied to these variables. The results obtained from these tests are shown in Table 4.

Table 4. Test Results

Variables	Normality Test	Test Results of the Variables with Normal Distribution			Test Results of the Other Variables	
	Kolmogorov-Smirnov (p)	Anova Test (p)	Tukey	Tukey Test (p)	Kruskal Wallis Test (p)	Mean Rank
X_1	0.000	-	-	-	0.000	Clust1:11.36 Clust2:28.42 Clust3:16.25 Clust4:32,63
X_2	0.196	0.244	-	-	-	-
X_3	0.099	0.000	Clust1: 2, 3, 4 Clust2: 1, 4 Clust3: 1, 4 Clust4: 1, 2, 3	0.000-0.022-0.000 0.000-0.017 0.022-0.007 0.000.007	-	-
X_4	0.200	0.000	Clust1: 2, 3, 4 Clust2: 1 Clust3: 1 Clust4: 1	0.000-0.001-0.000 0.000 0.001 0.000	-	-
X_5	0.057	0.000	Clust1: 2, 4 Clust2: 1 Clust4: 1	0.001-0.002 0.001 0.002	-	-
X_6	0.000	-	-	-	0.002	Clust1:14.61 Clust2:18.25 Clust3:34.13 Clust4:30.63
X_7	0.090	0.000	Clust1: 3, 4 Clust2:3, 4 Clust3:1, 2 Clust4:1, 2	0.000-0.000 0.028-0.001 0.000-0.028 0.000-0.001	-	-
X_8	0.000	-	-	-	0.013	Clust1:15.83 Clust2:18.67

						Clust3:35.75 Clust4:22.25
X_9	0.090	0.000	Clust1: 2, 3, 4 Clust2:1, 4 Clust3:1, 4 Clust4:1, 2, 3	0.000-0.000-.0.000 0.000-0.016 0.000-0.013 0.000-0.016-0.0013	-	-
X_{10}	0.000	-	-	-	0.000	Clust1:19.69 Clust2:29.13 Clust3: 6.50 Clust4: 2.75
X_{11}	0.200	0,008	Clust1: 2, 4 Clust2: 1 Clust4: 1	0.047-0.017 0.047 0.017	-	-
X_{12}	0.200	0.000	Clust1: 2, 4 Clust2: 1 Clust4: 1	0.000-0.038 0.000 0.038	-	-
X_{13}	0.166	0.000	Clust1: 2, 3, 4 Clust2: 1 Clust3: 1 Clust4: 1	0.000-0.003-0.000 0.000 0.003 0.000	-	-
X_{14}	0.200	0.405	-	-	-	-
X_{15}	0.200	0.001	Clust1: 2 Clust2: 1	0.001 0.001	-	-
X_{16}	0.000	-	-	-	0.001	Clust1:24.61 Clust2:14.00 Clust3:27.50 Clust4: 5.00
X_{17}	0.009	-	-	-	0.000	Clust1:27.44 Clust2: 9.21 Clust3:17.38

						Clust4:16.75
X_{18}	0.098	-	-		0.000	Clust1:28.78 Clust2:12.58 Clust3:7.50 Clust4:10.50
X_{19}	0.200	0.000	Clust1: 2, 4 Clust2: 1, 4 Clust3: 4 Clust4:1, 2, 3	0.002-0.000 0.002-0.004 0.003 0.000-0.004-0.003	-	-
X_{20}	0.000	-	-	-	0.007	Clust1:14.78 Clust2:22.42 Clust3:16.63 Clust4:34.88
X_{21}	0.000	-	-	-	0.083	-
X_{22}	0.109	-	-	-	0.030	Clust1:23.83 Clust2:16.58 Clust3:21.50 Clust4:6.75

The test statistics at Table 4 assert that there is no significant difference among the clusters, as the p values of X_2 and X_{14} that meet the condition of normal distribution, are higher than 5%. As the p value of X_3, X_4, X_5, X_7, X_9, X_{11}, X_{12}, X_{13}, X_{15}, and X_{19} variables are lower than 5% it can be said that a significant diference among the clusters exist. Considering the variables meeting the normality condition and having a significant difference among the clusters, the differences between the clusters are indicated in the Table 4 in details. The findings suggest that the Cluster 1 is different from the others.

Kruskal Wallis was applied to X_1, X_6, X_8, X_{10}, X_{16}, X_{17}, X_{20}, X_{21} variables that do not meet the normal distribution condition as well as X_{18}, X_{22} variables in which Anova test could not be applied as a result of the Levene test. According to Kruskal Wallis test results, since the p value of X_{21} variable is greater than 5%, it can be interpreted that there is no significant difference between the clusters. As p values belonging to X_1, X_6, X_8, X_{10}, X_{16}, X_{17} X_{18} X_{20} and X_{22} are lower than 5%, it can be said that a significant difference among variables exist.

Main indicators among variables; dwellings without basic facilities (X_1), labor market insecurity (X_6), long-term unemployment rate (X_8), air pollution (X_{12}), murder rate (X_{20}) and employees working very long hours (X_{21}) are negative indicators and it is preferred that they take lower values. Higher values are expected for rest of the positive variables.

Finally, the variables in Table 4. that do not meet the condition of normal distribution and having a significant relation among clusters are X_1, X_6, X_8, X_{20} variables. When the clusters mean ranks of Kruskal Wallis Test results for these variables are examined, Cluster 3 and cluster 4 are ranked first, whereas Cluster 1 and Cluster 2 are ranked in the last rows.

5. CONCLUSION

In this study; using “Better Life Index” data provided by the OECD, has been done clustering with WEKA software by taking the SOM algorithm into account. Optimum cluster number is determined by considering the Dunn Index.

As seen at Figure 3. Cluster 1 consists of Australia, Austria, Belgium, Germany, Denmark, Finland, France, Germany, Netherlands, Ireland, Sweden, Switzerland, Iceland, Canada, Luxembourg, Norway and New Zealand. This group is composed of the countries, having a higher democracy culture and economic prosperity along with modern social lifestyle. In this context, the considered variables of the countries in this cluster within the framework of economic, social and political dimensions have taken close values.

Cluster 2 is composed of Czech Republic, Estonia, Israel, Japan, Korea, Latvia, Hungary, Poland, Slovakia, Slovenia, Chile and Russia. The cluster basicly consists of Eastern Europe countries as well as some emerging countries in the Middle East, South America and Far East. Czech Republic, Estonia, Latvia, Hungary, Poland, Slovakia and Slovenia taking place in this group are the countries that are recent EU member states after the disintegration of the Soviet Union. They have lower economical power than the developed EU countries.

It is seen that Russia, another country in this cluster, after taking some steps on democratization and eliminating the old communist system, was still far from fulfilling the requirements of a true democratic government, even though it structured the previous 1977 Constitution between 1991-1993. It can be stated that liberal democracy does not rule in Russia today as the political and economic system in the country is still in transition (Baharçiçek & Ağır, 2014: 17-19).

Other countries of this cluster like Israel, Japan, Korea and Chile can be given examples to the countries, having prominent local economic powers. Nevertheless, Japan and Korea, the countries with developped economic dinamics does not taking place in cluster 1 can be explained due to their variable values of “life satisfaction” and “very long working rate” being low within the framework of social dimension of this study.

However, it can be stated that especially countries with developed economic dynamics such as Japan and Korea are not included in "Cluster 1" because of the low values of variables such as "life satisfaction" and "working very long hours" within the framework of the social dimension. Therefore, it can be thought that the economic dimension alone is not effective in determining clusters.

Cluster 3 can be defined as a cluster in which the EU countries like Spain, İtaly, Greece and Portugal having closer geographical positions, cultural habits, economical situations and lifestyles. Although these countries are in EU not in cluster 1 but in cluster 3, the economic problems that these countries have recently faced are “labor market insecurity”, “employment”, “long-term unemployment”, etc., so it can be interpreted that negatively affects the values of the variables and causes the “life satisfaction” rate to decrease within the social dimension. As the fact that those EU member states are in Cluster 3 instead of Cluster 1, it is possible to state that the economic problems that these countries have been experiencing recently negatively affects the variables such as “labor market insecurity”, “employment”, “long-term unemployment”, causing a decrease in the “life satisfaction” in those countries.

Cluster 4; is a group composed of countries like Turkey, Mexico, South Africa and Brazil which are located at different continents. While their economical and political situation, population and areas sizes are similar, the economical variables: “employment rate”, “long-term unemployment rate”, “personal earning rate”, “household net adjusted disposable income”, the social variables: “life satisfaction”, “feeling safe walking alone at night”, and the political varibable “political participation rates” are relatively lower from the other countries.

The study of clustering countries according to the Better Life Index is a multi variable study taking many economical, political etc. variables into account that present significant findings. Even though the most important factor on grouping the countries in different clusters seems the economic dimension, the social and political dimensions such as political stability, geographical position, socio-cultural conditions, habits are also seen very important.

REFERENCES

1. AZAR, A. T., SHAIMAA, A. E. S. & ABOUL, E. H. (2013). "Fuzzy and Hard Clustering Analysis for Thyroid Disease", Computer Methods and Programs in Biomedicine, 111 (1), pp. 1-16.

2. BAHARÇİÇEK, A. & AĞIR, O. (2014). "Rusya'nın Başarısız Demokratikleşme Tarihi", Birey ve Toplum, 4 (8), pp. 5-27.

3. BURBIDGE, R. & BUXTON, B. (2001). "An Introduction to Support Vector Machines for Data Mining", Keynote Papers, young OR12, pp. 3-15.

4. CHANG, D. F. & CHEN, C. C. (2018). "Cluster Analysis for Student Performance in PISA 2015 among OECD Economies", ICIC Express Letters, Part B, Applications: An International Journal of Research and Surveys, 9 (11), pp. 1139-1146.

5. CHEN, C. & WANG, L. (2008). "Product Platform Design through Clustering Analysis and Information Theoretical Approach", International Journal of Production Research, 46 (15), pp. 4259-4284.

6. DO CARVALHAL MONTEIRO, R. L., PEREIRA, V. & COSTA, H. G. (2019). "Analysis of the Better Life Index trough A Cluster Algorithm", Social Indicators Research, 142 (2), pp. 477-506.

7. DUNN, J. C. (1974). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics, 3 (3), pp. 32-57.

8. EVERITT, B. (1974). Cluster Analysis, Heinmann, London, 122.

9. FAYYAD, U., PIATETSKY-SHAPIRO, G. & SMYTH, P. (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data", Communications of the ACM, 39 (11), pp. 27-34.

10. GAN, G., MA, C., WU, J. (2007). "Data Clustering Theory, Algorithms, and Applications", American Statistical Association.

11. HE, Z., XU, X. & DENG, S. (2005). "TCSOM: Clustering Transactions Using Self-Organizing Map", Neural Processing Letters, 22 (3), pp. 249-262.

12. HUDSON, I. L., LEEMAQZ, S. Y., KIM, S. W.,DARWENT, D., ROACH, G. & DAWSON, D. (2016). "SOM Clustering and Modelling of Australian Railway Drivers' Sleep, Wake, Duty Profiles", In Artificial Neural Network Modelling, pp. 235-279.

13. JAIN, N. & SRIVASTAVA, V. (2013). "Data Mining Techniques: A Survey Paper", IJRET: International Journal of Research in Engineering and Technology, 2 (11), pp. 2319-1163.

14. JASEENA, K. U. & DAVID, J. M. (2014). "Issues, Challenges, and Solutions: Big Data Mining", NeTCoM, CSIT, GRAPH-HOC, SPTM–2014, pp. 131-140.

15. JIN, C. H., POK, G., LEE, Y., PARK, H. W., KIM, K. D., YUN, U. & RYU, K. H. (2015). "A SOM Clustering Pattern Sequence-Based Next Symbol Prediction Method for Day-Ahead Direct Electricity Load and Price Forecasting", Energy Conversion and Management, 90, pp. 84-92.

16. KOHONEN, T. (1990). The Self-Organizing Map, Proceedings of the IEEE, 78 (9), pp. 1464-1480.

17. MAIND, S. B. & WANKAR, P. (2014). "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication, 2 (1), pp. 96-100.

18. MFA, (2020). http://www.mfa.gov.tr/ekonomik-isbirligi-ve-kalkinma-teskilati.tr.mfa/, Available online: 27.12.2020.

19. OECD, (2020). https://www.oecd.org/about/, Available online: 27.12.2020.

20. ÖZDEMIR, O. & KAYA, A. (2018). "K-Medoıds and Fuzzy C-Means Algorithms for Clustering Co2 Emissions of Turkey and Other OECD Countries", Applied Ecology and Environmental Research, 16 (3), pp. 2513-2526.

21. PAL, C., HIRAYAMA, S., NARAHARI, S., JEYABHARATH, M., PRAKASH, G. & KULOTHUNGAN, V. (2018). "An Insight of World Health Organization (WHO) Accident Database by Cluster Analysis with Self-Organizing Map (SOM)", Traffic Injury Prevention, 19 (sup1), S15-S20.

22. PARK, D. C., EL-SHARKAWI, M. A., MARKS, R. J., ATLAS, L. E. & DAMBORG, M. J. (1991). “Electric Load Forecasting Using An Artificial Neural Network”, IEEE Transactions on Power Systems, 6 (2), pp. 442-449.

23. PARLAR, T. & KAKILLI ACARAVCI, S. (2017). “Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data”, International Journal of Economics and Financial Issues, 7 (2), pp. 692-696.

24. SARLE, W. S. (1994). “Neural Networks and Statistical Models”, Proceedings of the Nineteenth Annual SAS Users Group International Conference, SAS Institute, pp. 1538–1550.

25. TUIK, (2017). http://www.tuik.gov.tr/PreIstatistikTablo.do%3Bjsessionid%3Dnp0ZVyJGtTSnGZhQlQrMvZGnlChLt14L4yBQ7dLfQ0rGlkpMGGLh!1863195282%3Fistab_id%3D9008&sa=U&ved=2ahUKEwixj5375pXjAhUH3aQKHfkBBEIQFjAAegQIBBAB&usg=AOvVaw3XjOABXWzM5HnPbmVQjEG7, Available Online: 05.03.2019.

26. VOUTILAINEN, A. & ARVOLA, L. (2017). “SOM Clustering of 21-Year Data of A Small Pristine Boreal Lake”, Knowledge & Management of Aquatic Ecosystems, 418 (36).

27. WU, J. L., CHANG, P. C., TSAO, C. C. & FAN, C. Y. (2016). “A Patent Quality Analysis and Classification System Using Self-Organizing Maps with Support Vector Machine”, Applied Soft Computing, 41, pp. 305-316.