Evaluating Environmental Sustainability Position of Turkey Via Bric and Mint Countries With K-Nn Algorithm

Özge EREN[1]

ABSTRACT

Environmental sustainability is one of the biggest difficulties met by humanity. It is also one pillar of sustainable development purposes required to be executed. For evaluating and comparing all world countries with respect to this perspective, an index is derived from the United Nations. With this index, the countries of the world are divided into four categories, such as "Very High Developed (VHD)", "High Developed (HD)", "Medium Developed (MD)" and lastly, "Low Developed (LD)". According to the 2019 environmental sustainable index value, Turkey is located in the VHD category. Some of the experts put forward that Turkey will compete with economic giants in 2050. Our hope economically continued efforts will be achieved environmentally also. These thoughts became the starting point of this study. With this aim, the position of Turkey is predicted with the BRIC and MINT countries data regarding to environmental sustainability with the k-NN (Nearest Neighbor) algorithm technique.

Keywords: *k-NN Algorithm; Environmental Sustainability; BRIC and MINT Countries.*

Türkiye'nin Çevresel Sürdürülebilirlik Kategorisinin Brıc ve Mınt Ülkeleri Yardımıyla K-Nn Algoritması Üzerinden Değerlendirilmesi

ÖZ

Çevresel sürdürülebilirlik, insanlığın günümüz dünyasında yüzleştiği en büyük sorunlardan biridir. Ayrıca sürdürülebilir kalkınmanın tüm yönleriyle yürütülmesi açısından oldukça önemli bir parçasıdır. Tüm dünya ülkelerini çevresel sürdürülebilirlik açısından değerlendirmek ve karşılaştırmak için Birleşmiş Milletler bir endeks türetmiştir. Çevresel sürdürülebilirlik indeksi olarak oluşturulan bu indeks ile ülkeler "Çok Yüksek Gelişmiş", "Gelişmiş", "Orta Gelişmiş" ve son olarak da "Düşük Gelişmiş" olmak üzere dört farklı kategoride sınıflandırılmaktadır. 2019 çevresel sürdürülebilir endeks değerine göre, Türkiye son sıralarda da olsa Çok Yüksek Gelişme göstermiş ülke kategorisinde yer almaktadır. Türkiye›nin 2050'li yıllarda ise ekonomik devlerle rekabet edeceğini öne sürülmektedir. Umudumuz ekonomik olarak devam eden çabalarımızın çevresel olarak da sağlanması yönündedir. Bu düşünceler bu çalışmanın başlangıç noktası olmuştur. Bu amaçla, Türkiye›nin çevresel sürdürülebilirlik kategorisi, k-NN (k-En Yakın Komşu) algoritma tekniği ile BRIC (Brezilya, Rusya, Hindistan, Çin) ve MINT (Meksika, Endonezya, Nijerya, Türkiye) ülkeleri verileri yardımıyla tahmin edilmiştir.

Anahtar Kelimeler: *k-NN Algoritması, Çevresel Sürdürülebilirlik, BRIC ve MINT Ülkeleri.*

INTRODUCTION

The term of sustainable development was built on the three pillars of sustainability including social, economic and environmental perspective. It is certain that all the pillars will be achieved for the real achievement on sustainable development goals. In this study, it is focused on one pillar that is environmental sustainability. It is defined as "Current needs of us without anyway compromising the quality of environment/ecosystem thus it will be able to support future generations equally remains" (Ferranti et al., 2018). One another about the concept of environmental sustainability was developed from Goodland (Goodland, 1995). According to his explanations, environmental sustainability "Seeking to improve human welfare by preserving the all needed sources of raw materials used for

human needs and ensuring that the swamps for human wastes are not exceeded, in order to restrain harm to humans".

United Nations for the world countries created an index to measure and rank them. This index is called Environmental Sustainability Index. It contains twelve indicators that cover environmental sustainability and its threats. Indicators related to environmental sustainability are energy consumption of fossil fuel, fresh water withdrawals, carbon dioxide emissions, renewable energy consumption, forest area, and natural resource depletion with percentage calculation of GNI (Gross National Income). Indicators used for environmental threats are the rates of mortality attributed to hygiene services and sanitation, degraded land, domestic and environmental air pollution, and unsafe water, and lastly, Nature's Red Index, which counts up extinction risks of species.

In this study, it was aimed to perform classification efforts of Turkey's position by using BRIC (Brazil, China, India, and Russia) and MINT (Mexico, Turkey, Nigeria, Indonesia) countries' data which covers environmental sustainability and environmental threats. BRIC is an acronym used to represent four countries that have similar economic development, also known as the BRIC's or the Big Four, including Brazil, Russia, India, and China. MINT is also another acronym, first used in 2001 by Jim O'Neill, including Mexico, Indonesia, Nigeria, and Turkey. They have been predicted to be gigantic economies by 2050. The BRIC and MINT countries changed direction the economic weight from developed counties to emerging ones. It is obvious that they will continue to grow with their young population and gross national income, which could also refer to environmental changes (Hawksworth and Cookson, 2006). Population with in a larger and richer form may initiate environmental collapse (Bradshaw et al., 2010; Scherer et al., 2018; Ward et al., 2016; Scherer, 2019). To make a prediction for Turkey's position with the k-NN (Nearest Neighbor) algorithm by using BRIC and MINT countries' data is that reason.

LITERATURE REVIEW
In the literature review, it can be seen that lots of different application areas including the attempts to progress the methodology going on related to the k-NN algorithm. Especially, it is widely used in text classification area.

Text classification is the process of determining the class to which any document belongs. Soucy and Mineauin at 2001 propose to use a smooth non-weighted features k-NN algorithm for text classification efforts. Using feature interaction, it has been found the relevant features for the learning task (Soucy &Mineau, 2001).

The conventional k-NN has a very big defect that computing the steps causes huge time, as the k-NN algorithm is applied to text classification with the high size and big samples. The paper from Wang and Wang at 2007 is progressed to search the exact k nearest neighbors quickly which is called TFK-NN (Tree-Fast-K-Nearest-Neighbor) (Wang and Wang, 2007). According to some studies, k-NN text categorization can be effective, but less efficient method. Jiang and others at 2021 proposed an improved k-NN algorithm for text categorization, which constructs the new classification model by combining constrained one pass clustering algorithm and k-NN text categorization (Jiang et al.,2012).

Another important application area is related to the choice of k parameter. Lee and others in 2003 prepared a study about the choice of the parameter k. It is very likely that a constant k value will be ended in a bias on large categories. Nearest neighbors can be used for assigning whether a test document should be classified into a category in the training set, which has more samples (Li et al., 2003). Zhang et al. studied an example-driven k-parameter calculation, which assigns different k values for different test samples in k-NN prediction practices including classification, regression. The proposed approach shows that this algorithm is better than the primitive *k-NN* algorithms in terms of data mining tasks (Zhang et al., 2018). Deng et al. firstly proposed k-means clustering technique to split the whole dataset into small parts, and then conducted k-NN classification. It has been conducted lots of experiments on medical big data. According to experimental efforts, the proposed technique works better in terms of accuracy and efficiency (Deng et al.,2016).

The paper progressed from Alkhatib et al. in 2013 related to the financial application of the k-NN algorithm; they applied non-linear regression k-NN algorithm in order to estimate stock prices for on a sample of companies listed on the Jordanian stock market to help investors, decision makers. One important finding is depending on the actual stock prices data

the estimated results were so close, and almost parallel to actual stock prices (Alkhatib et al., 2013). In 2019, Li and others achieved a research in power energy load prediction method, based on the k-NN algorithm. The results verify the validity and accuracy of the prediction method (Li and Jin,2019).

Another study is to evaluate the position of Lebanon compared to other Middle East countries in terms of sustainable development. For this purpose, it has been adopted an integrated of data mining techniques like a tool to study the link between these indicators. By using K-means clustering in this study defined the different levels of living. In order to extract the most relevant indicators, feature selection technique was applied. Finally, k-NN classification technique was used to estimate the model (Katat et al., 2019). The study of Katat et al., is base study for us because of some similarities like trying to make classification of the any country with k-NN algorithm.

METHODOLOGY
In pattern recognition area k-NN is one of the most important statistical methods. The original k-NN algorithm was introduced in 1967 by T. M. Cover and P. E. Hart. It is called also very simple nonparametric classification algorithm in which you choose the k closest neighbors to a point, and for its label each of the neighbors constitutes a "vote". Then, it can be assigned the point the label with the most votes. k-NN depends to a case-approach, because this is essentially predicting an input's label based on similar samples. In the below, it can be seen clearly all the steps:

1- It must be loaded with the training as well as test data.

2- Determine "k" (must be any integer number) number, which represents the number of nearest neighbors

3- Calculate distance (which distance it depends to you) one by one between the query-instance, and all the training samples.

It is certain that the result of k-NN algorithm depending on the distance type can be changed. Usually, in the literature, Euclid distance can be preferred (Kurniadi et al.,Ali et al., Todeschini, R.).

Euclid Distance
$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Minkowski Distance
$$d(x, y) = \sqrt[p]{\sum_{i=1}^{n} (x_i - y_i)^p}$$

Manhattan Distance
$$d(x, y) = \sum_{\square=1}^{n} |x_i - y_i|$$

4- Sort in ascending order by distance value

5- It will choose the best "k" rows from the sorted array.

6- It will assign a class to the test point according to the most frequently class of these rows.

7- End

FINDINGS
This index normally includes 12 indicators, but two indicators (Degraded land and Fresh water withdrawals) are missing in the lots of countries (UNDP/Dashbord/Environmental Sustainability), so these are ignored. It is also one limitation of study.

Table 1: Indicators of Environmentally Sustainability Index

Notation	Definition
K_1	Fossil fuel energy consumption (% of total energy consumption)
K_2	Renewable energy consumption (% of total final energy consumption)
K_3	Carbon dioxide emissions (Per capita (tons))
K_4	Carbon dioxide emissions (kg per 2010 US$ of GDP)
K_5	Forest area (% of total land area)
K_6	Forest area (Change (%))
K_7	Natural resource depletion (% of GNI)
K_8	Household and ambient air pollution (per 100,000 population)
K_9	Unsafe water, sanitation and hygiene services (per 100,000 population)
K_{10}	Red List Index

The indicators formed environmental sustainability index can be seen at Table 1. It contains a selection of ten indicators that cover environmental sustainability and its threats. Table 2 and Table 3 also show the MINT and BRIC countries also raw indicators' values. According to Table 2, Mexico

and Indonesia are located in HD Segment. Turkey's position is located VHD Segment for the years 2019. In the Table 3, only Russia is only in VHD segment, Brazil and China in HD segment, and lastly India attends MD segment also.

Table 2: MINT Countries

Country	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}	Category
Mexico	90.4	9.2	3.6	0.21	33.9	-5.5	2.2	37	1.1	0.677	HD
Indonesia	66.1	36.9	1.7	0.17	49.9	-23.8	1.9	112	7.1	0.754	HD
Nigeria	18.9	86.6	0.5	0.09	7.2	-61.8	4.4	307	68.6	0.874	LD
Turkey	86.8	13.4	4.2	0.18	15.4	22.8	0.2	47	0.3	0.875	VHD

Table 3: BRIC Countries

Country	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}	Category
Brazil	59.1	43.8	2.0	0.15	58.9	-9.9	1.9	30	1.0	0.902	HD
India	73.6	36.0	1.6	0.26	23.8	10.8	1.0	184	18.6	0.678	MD
Russian Federation	92.1	3.3	9.9	0.45	49.8	0.8	5.8	49	0.1	0.955	VHD
China	87.7	12.4	6.4	0.47	22.4	33.6	0.9	113	0.6	0.744	HD

When we check Turkey's position by using MINT countries data as illustrated in Table 4, first and second neighbor categories are HD, and the third one is LD, which means that if the value of k is equal to one, two and three, the prediction for Turkey is High HD. Since none of the MINT countries other than Turkey is in High HD category, it is obvious that we could not make the right prediction for Turkey with the knowledge of MINT Countries. Maybe, economically they can be seen in the same segment, but from the side of environmental sustainability, not.

Table 4: k-NN Results for MINT Countries

Country	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}	Distance	Neighbor	Category
Mexico	90.4	9.2	3.6	0.21	33.9	-5.5	2.2	37	1.1	0.677	35.78	First	High HD
Indonesia	66.1	36.9	1.7	0.17	49.9	-23.8	1.9	112	7.1	0.754	92.89	Second	High HD
Nigeria	18.9	86.6	0.5	0.09	7.2	-61.8	4.4	307	68.6	0.874	299.18	Third	Low HD
Turkey	86.8	13.4	4.2	0.18	15.4	22.8	0.2	47	0.3	0.875			

When we check Turkey's position by using BRIC's countries data as illustrated in Table 5, first and second neighbor categories are VHD, and third and fourth categories are HD, which means that if the value of k is equal to one, two and three, the prediction for Turkey is Very High HD. If the value of k is four, no prediction is obtained from the analysis.

Table 5: k-NN Results for BRIC Countries

Country	K_1	K_2	...	K_8	K_9	K_{10}	Distance	Neighbor	Category
Brazil	59.1	43.8	...	30	1.0	0.902	70.44	43.19	Very High HD
India	73.6	36.0	...	184	18.6	0.678	141.48	43.19	Very High HD
Russian Federation	92.1	3.3	...	49	0.1	0.955	43.19	67.28	High HD
China	87.7	12.4	...	113	0.6	0.744	67.28	70.44	High HD
Turkey	86,8	13,4	...	47	0,3	0,875			

CONCLUSION

According to some studies for the future prediction with the year of 2050, Turkey will be the giant economy in the world. Even one idea asserted that, Russia and Turkey are both expected to be larger than present-day China's status (Dadush and Stancil., 2010). In this study, BRIC and MINT countries, including Turkey, that are considered to be in the same segment in terms of some economic indicators. With this study, it has been wondered whether the economic similarity can be achieved in terms of environmental sustainability indicators. The main idea of this study depends this hypothesis. With k-NN algorithm by using separately BRIC and MINT countries data checked again the possibility of Turkey's real position to make prediction. With the help of k-NN algorithm the results are not meaningful. One reason should be Turkey compared with the others really better (environmentally sustainability). For further studies,

this effort should be progress with different countries and with different techniques.

REFERENCES

[1]Ali, Z., Abbas, A. W., Thasleema, T. M., Uddin, B., Raaz, T., & Abid, S. A. R. (2015). Database development and automatic speech recognition of isolated Pashto spoken digits using MFCC and K-NN. International Journal of Speech Technology, 18(2), 271-275.

[2]Alkhatib, K., Najadat, H., Hmeidi, I., &Shatnawi, M. K. A. (2013). Stock price prediction using k-nearest neighbor (k-NN) algorithm. International Journal of Business, Humanities and Technology, 3(3), 32-44.

[3]Bradshaw, C. J., Giam, X., & Sodhi, N. S. (2010). Evaluating the relative environmental impact of countries. PloS one, 5(5).

[4]Cover, T.M., Hart, P.E.(1967), Nearest neighbor pattern classification. IEEE Trans. Inf. Theory IT-13, 21–27

[5]Dadush, U. B.,& Stancil, B. (2010). The world order in 2050. Carnegie Endowment for International Peace.

[6]Deng, Z., Zhu, X., Cheng, D., Zong, M., & Zhang, S. (2016). Efficient k-NN classification algorithm for big data. Neurocomputing, 195, 143-148.

[7]El Katat, S., Kalakech, A., Kalakech, M., & Hamad, D. (2019). Financial Development Indicators: A Comparative Study between Lebanon and Middle East Countries Based on Data Mining Techniques. Internatıonal Arab Journal Of Informatıon Technology, 16(3 A), 499-505.

[8]Ferranti, P., Berry, E., & Jock, A. (2018). Encyclopedia of Food Security and Sustainability. Elsevier.

[9]Goodland, R. (1995). The concept of environmental sustainability. Annual review of ecology and systematics, 26(1), 1-24.

[10]Hawksworth, J.,& Cookson, G. (2006). The world in 2050. How big will the major emerging market economies get and how can the OECD compete.

[11]Jiang, S., Pang, G., Wu, M., &Kuang, L. (2012). An improved K-nearest-neighbor algorithm for text categorization. Expert Systems with Applications, 39(1), 1503-1509.

[12]Kurniadi, D., Abdurachman, E., Warnars, H. L. H. S., &Suparta, W. (2018, November). The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm. In IOP Conference Series: Materials Science and Engineering (Vol. 434, No. 1, p. 012039). IOP Publishing.

[13]Li, B., Yu, S., & Lu, Q. (2003). An improved k-nearest neighbor algorithm for text categorization. arXiv preprint cs/0306099.

[14]Li, F.,& Jin, G. (2019). Research on power energy load forecasting method based on K-NN. International Journal of Ambient Energy, 1-6.

[15]O'neill, J. (2001). Building better global economic BRICs.in Global Economics Paper No: 66, GS Global Economics Website, available online at https://www.goldmansachs.com/insights/archive/index.html

[16]Scherer, L., Behrens, P., de Koning, A., Heijungs, R., Sprecher, B., &Tukker, A. (2018). Trade-offs between social and environmental Sustainable Development Goals. Environmental Science & Policy, 90, 65-72.

[17]Scherer, L., de Koning, A., &Tukker, A. (2019). BRIC and MINT countries' environmental impacts rising despite alleviative consumption patterns. Science of the Total Environment, 665, 52-60.

[18]P. Soucy and G. W. Mineau, "A simple K-NN algorithm for text categorization," Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 2001, pp. 647-648.

[19]Todeschini, R. (1990). Weighted k-nearest neighbour method for the calculation of missing values. Chemometrics and Intelligent Laboratory Systems, 9(2), 201-205.

[20]Ward, J. D., Sutton, P. C., Werner, A. D., Costanza, R., Mohr, S. H., & Simmons, C. T. (2016). Is decoupling GDP growth from environmental impact possible?. PloS one, 11(10).

[21]Y. Wang and Z. Wang, "A Fast K-NN Algorithm for Text Categorization," 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, 2007, pp. 3436-3441.

[22]Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018). A novel k-NN algorithm with data-driven k parameter computation. Pattern Recognition Letters, 109, 44-54.

URL 1: http://hdr.undp.org/en/content/dashboard-4-environmental-sustainability-0