AN EXAMINATION ON THE EFFICIENCY LEVEL OF G20 ECONOMIES TO NEW INVESTMENT AND TECHNOLOGICAL DEVELOPMENT: DATA ENVELOPMENT ANALYSIS AND MACHINE LEARNING

[1]TALHA KARADAYI, [2]EMRAH ONDER

[1]Research Assistant, Istanbul Ticaret University, İstanbul, Türkiye
[2]Professor, Istanbul University, İstanbul, Türkiye
E-mail: [1]tkaradayi@ticaret.edu.tr, [2]emrah@istanbul.edu.tr

Abstract -

Problem Statement: All countries race to increase the level of development and welfare."Which of the countries rapid move?" is among the most frequently spoken table talks. However, not everyone competes on equal terms. While analyzing countries, each country should be evaluated according to its resources. Advantages and disadvantages should be considered in subjects such as economic conjuncture, business environment, resources, geography, technology, etc. The issues of comforting the business environment and adapting to technology constitute the infrastructure of economic progress of countries. It will be useful to discover questions about which countries are more efficient in these actions according to their dynamics, and which action is more important.

Methods: G20 countries were divided into two subgroups developed and developing to further increase homogeneity, and examined. As comparable data, GDP per capita was chosen as the input, and country efficiencies between 2016 and 2020 were examined in reaching Doing Business and Global Connectivity Index scores as outputs. Data Envelopment Analysis was used to explore efficiency. Then, Machine Learning methods (Decision Tree and Random Forest) were used to determine which output is more important.

Results: South Korea (in 2016 and 2020) and India (in 2019) shine out as the most efficient variables. Doing Business scores were more important than Global Connectivity Index for both groups (developed and developing).

Conclusion: It may be beneficial for developed countries to focus on South Korea (especially in 2016 and 2020) while developing countries concentrate on India (especially in 2019). Additionally, an examination of their policies can be beneficial to the roadmap of entrepreneurship and policymakers. Regardless of the development level of the countries, providing a facilitating environment for entrepreneurs and investors and increasing their comfort can make significant contributions. In any case, there are opportunities, however, the matter is whether they are supported or not.

Abbreviations: GDP (Gross Domestic Product), ICT (Information and Communication Technology), GCI (Global Connectivity Index), DB (Doing Business), DEA (Data Envelopment Analysis), ML (Machine Learning), DMUs (Decision Making Units)

Keywords - G20, Doing Business, Global Connectivity Index, Data Envelopment Analysis, Machine Learning, Decision Tree, Random Forest.

I. INTRODUCTION

From the 1990s to the present, the number of Internet users has grown from a few million to almost five billion. This trend has brought about a digital transformation that is transforming our societies and economies. The Internet, which is at the forefront of ICT, facilitates access to resources to develop and acquire skills, offers alternatives to ways of doing business such as remote working methods, and provides businesses with access to a wider talent pool. On the other hand, it enables innovation by giving rise to new business models. In addition, it contributes to job creation and economic development by increasing productivity and generating innovation (ITU, 2022). In addition to technology and innovation, the establishment of new businesses and the increase in enterprises are also strongly related to economic growth. Countries are driven to compete and often produce incentive reforms to increase the level of entrepreneurship (Nave & Rodrigues, 2022). On the other hand, it can be said that the acceleration of innovation and technological developments act in connection with the entrepreneurial culture. While innovation-driven economies provide more competitive and more favorable environments for entrepreneurs (Nave & Rodrigues, 2022), entrepreneurial activities play a positive role in technological change and economic growth (Ács & Varga, 2005). Across economic research, entrepreneurship is particularly associated with rapid growth and rapid job creation. It is a generally accepted view that entrepreneurship is needed to achieve better economic performance (Wennekers and Thurik, 1999; Ács and Varga, 2005; Nave & Rodrigues, 2022). The differences in

entrepreneurship rates among countries are affected not only by the individual qualities of entrepreneurs but also by institutional frameworks and the macroeconomic environment (Baumol and Strom, 2007; Dreher and Gassebner, 2008; Bosma and Schutjens, 2011). For this reason, it would be appropriate to make an assessment based on the data of the Doing Business (DB) research, which emphasizes the aspects that facilitate entrepreneurship. This study will be carried out on the basis of "GDP Per Capita, ICT connectivity level, and activities to facilitate entrepreneurship" and will focus on the efficiency of G20 countries in increasing national welfare.

II. INSTRUMENTS TO ANALYSIS

Global Connectivity Index (GCI)

The GCI was created to analyze a broad spectrum of indicators for Information and Communication Technology (ICT) Infrastructure and digital transformation. It has provided a comprehensive map of the global digital economy as an annual index since 2014. The GCI comprehensively and objectively evaluates connectivity from both a national and industrial perspective. In addition, it provides a detailed assessment of the current state of connectivity and how close the world is to achieving universal and meaningful connectivity. The research framework of the GCI comprises a combination of advanced and fundamental technologies, enabling analysis of how past, current, and future intersect to help exhibit the global digital economy (Huawei, 2023).

The index benchmark countries according to their performance in 40 indicators which are given. These various indicators have been determined to explore the impact of ICT on a country, its level of digital competitiveness and its future growth (Huawei, 2023). In summary, the GCI monitors the relationship between ICT infrastructure investment and economic growth to provide policymakers and investors with the trends and information they need to make sound decisions (Huawei, 2022).The widespread use of ICT in a country or industry increases the efficiency (task efficiency, functional efficiency, system efficiency, organizational efficiency and agility, and ecosystem efficiency and resilience) in that country or sector (Huawei, 2022). Therefore, ICTs, which are directly related to efficiency, were chosen as one of the focuses of the research part of the study.

Doing Business (DB)

The Doing Business data focuses on the role of government policy in the daily operations of small and medium-sized domestic firms. The goal is to encourage efficient, transparent, and simple regulation so that businesses can thrive. DB data focuses on regulations that affect small and medium-sized domestic companies in an economy's largest commercial city (World Bank, 2020). In another word, World Bank constituted the DB reports to provide objective measures of business regulations and their enforcement per country, and involve 190 economies (World Bank, 2023a). While doing this examination, DB focuses on 12 areas of business regulation.

These business regulations are respectively: starting a business, employing workers, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across borders, contracting with the government, enforcing contracts, and resolving insolvency. However, "employing workers" and "contracting with the government" are not included in the ease of DB score and ranking. In conclusion, the DB Index includes 10 of these 12 criteria (World Bank, 2020).Clear and facilitative rules are far more effective in encouraging development and growth if they are efficient, transparent, and easily accessible to those who are supposed to benefit from them. Such rules foster an environment in which entrepreneurs with drive and innovative ideas can begin their businesses and productive firms can invest, expand, and create new jobs (World Bank, 2020). As a result, the DB Index focuses on the extent to which initiatives to be carried out in a country are encouraged and paved the way.

When the literature is examined, there are findings indicating that the criteria scored by the DB research give effective results on economic growth (Ani, 2015). In addition, DB already provides data on how governments regulate businesses, contributing to research into how regulation affects development (World Bank, 2020). Thus, these scores show indicators that all countries want to maximize in terms of their national and economic interests. Therefore, it is inevitable that the scores obtained from DB Index should be included in the study while examining the effectiveness of countries in stimulating the economy.

Gross Domestic Product (GDP) Per Capita

Gross Domestic Product (GDP) is the most extensively used single indicator of a country's total economic activity. It indicates the total value at current prices of final goods and services produced within a country over a given time period divided by the average population for the same year (IMF, 2022). GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products (World Bank, 2023b). GDP per capita attempts to basic measure the value of output per person, which is an indirect indicator of per capita income. GDP growth and GDP per capita are broad metrics of growth in the economy (World Bank,

2023c). In another word, national wealth has critical importance for a country to increase its investments and keep up with technological developments. In addition, the high economic wealth makes possible advanced infrastructure. As the Internet, ICT adoption will be influenced by a nation's overall economic health because wealthier nations are more likely to have the resources and capital needed to advance the technology (Hargittai, 1999). Research highlights that the difference in the adoption of ICT, known as the Digital Divide, continues to increase day by day between developed and developing countries (Ayanso et al., 2014). The economic situation of the country acts as a resource in the progress of nations in ICT and in the creation of necessary regulations, subsidies, and supports to encourage entrepreneurship. As an indicator of the current economic situation, the GDP Per Capita will be used as an input in Data Envelopment Analysis (DEA) to be comparable data when examining the G20 group, which consists of countries with different levels of development and diverse populations.

III. METHODOLOGY

This research aims to examine how efficient G20 countries are in accessing and adopting ICT (based on the GCI) and in creating facilitator conditions for entrepreneurship (based on DB Scores). Accordingly, "GDP per capita", a comparable unit of wealth, was taken as input. On this occasion, it was desired to explore how far each country can go within the framework of its economic power. To measure the efficiency level of G20 economies in the catch-up new investment and technological development, the first part of this study will employ the Data Envelopment Analysis (DEA) technique. As a second step, Machine Learning techniques will be used, and it will be determined which of the outputs has more critical importance by using the Decision Tree and Random Forest methods.

Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a non-parametric method of linear programming in which the efficiency of each Decision-Making Units (DMUs) included in the study is estimated according to the most efficient units forming the production frontier (Mihaylova-Borisova & Nenkova, 2021). Farrell (1957), who focused on the productivity of firms and industries, first introduced the DEA to literature. The model was further developed by Charnes et al. (1978) and has been taken advantage in the literature as part of productivity analysis until now.

Machine Learning: Decision Tree & Random Forest

Machine learning (ML) is a subset of artificial intelligence and is a collection of algorithms that extract insights from data. This insight can be used by humans or other machines to make decisions (Lewis, 2017). ML is the process of automatically discovering patterns and trends in data (Ghatak, 2017). ML methods are divided into three subgroups supervised, unsupervised, and reinforcement. Decision trees and random forest are examples of supervised ML techniques. A decision tree contains a series of logical decisions like a flowchart. Decision trees are a very useful and widely used technique to facilitate decision-making and can be applied to model almost any type of data (Lantz, 2015). The Random Forest algorithm is based on the wisdom of the crowd concept and consists of many trees. In this way, different training datasets are made up of subsets of the dataset. In classification problems, the class with the most votes gives the result, while in regression problems, the average of the outputs of the trees can make the final decision (Lewis, 2017).

Data

In this study, three different databases were used. GDP Per capita (current price), which constitutes the input of the study, is taken from the October 2022 issue of the World Economic Outlook published by the IMF. One of the outputs, the GCI, consists of data that Huawei has published annually since 2014. Another output is DB data which is published by the World Bank. Due to the lack of missing data, the study covers the period from the beginning of 2016 to the end of 2020 (5 years).

IV. ANALYSIS

The data obtained from the relevant databases were listed and brought together in a five-year period. Since the data (USA $, percent, etc.) should be comparable in DEA, all the different types of data were normalized and compressed between 0-1.On the other hand, the data groups to be examined should have a homogeneous structure. To provide it, G20 countries were examined and analyzed as two separate groups Developed (9) and Developing (10) Countries. Since data for 5 years is collected for each country, there are 45 variables for developed countries (9 countries x 5 years = 45) and 50 variables for developing countries (10 countries x 5 years = 50). In this way, data and years have been converted into a comparable form. The efficiency values between inputs and outputs were examined by inserting the normalized data into DEA through R software. Then, again through R software, Decision Tree and Random Forest analyses for DB and GCI.

Data Envelopment Analysis (DEA)

As can be seen in Figures 2 and 4, 2 out of 45 variables from developed countries were determined as efficient and 43 variables as inefficient due to DEA. On the other hand, 1 out of 50 variables from developing countries is evaluated as efficient and 49 of them as inefficient.

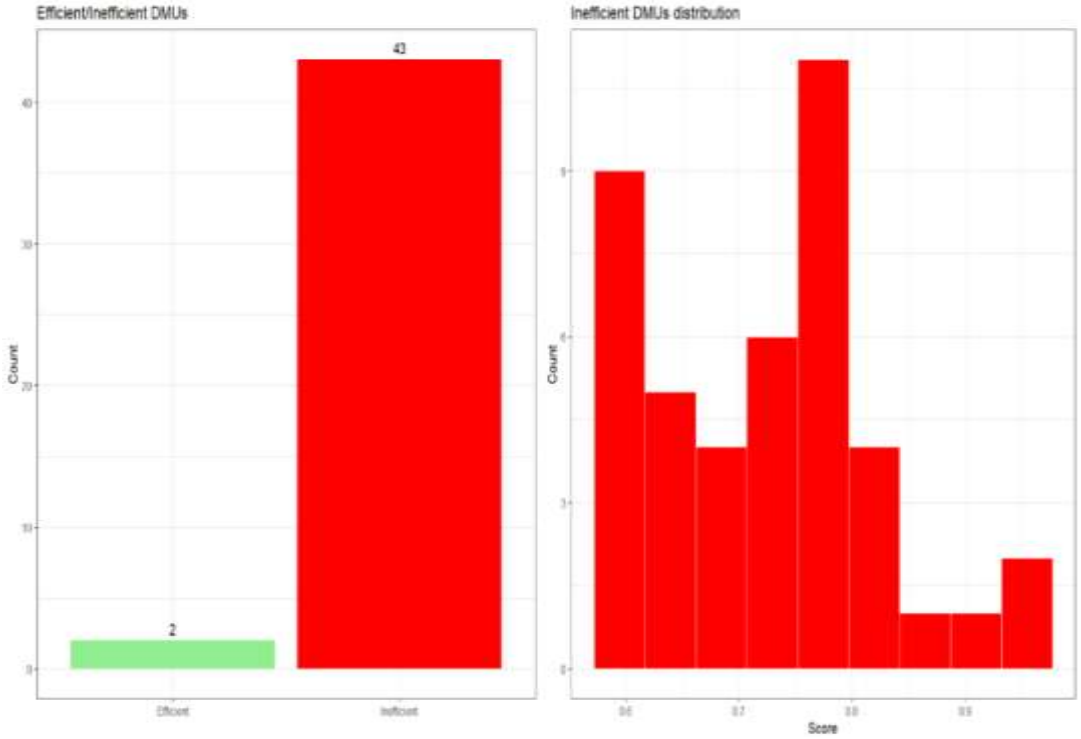


Figure 2:Efficient/Inefficient DMUs of Developed Country Variables
Figure 3: Inefficient DMUs Distribution of Developed Country Variables

However, the situation is slightly different in the developing country group. It is seen that the majority of DMUs are between 0.00-0.10. This shows that compared to developed countries, the productivity levels of developing countries in their own groups are much lower.

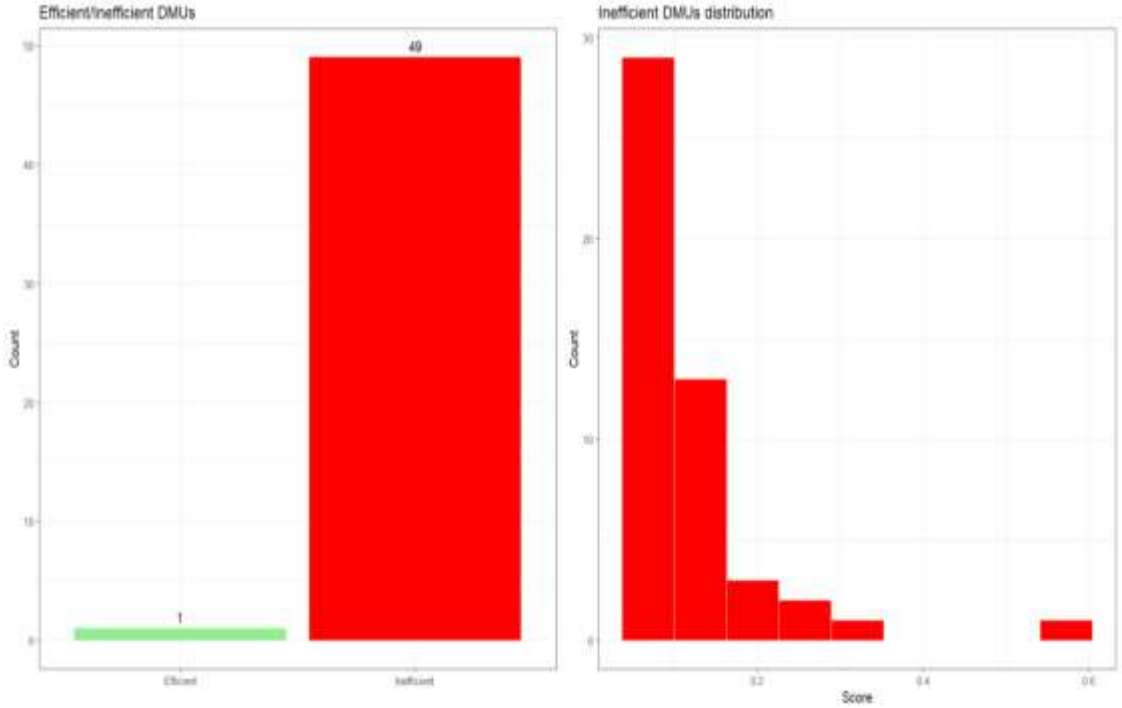


Figure 4:Efficient/Inefficient DMUs of Developing Country Variables
Figure 5: Inefficient DMUs Distribution of Developing Country Variables

So which countries and years do these DMUs belong to? In the developed countries group (Figure 6) South Korea is presented as efficient with its performance in 2016 and 2020. On the other hand, in the group of developing countries, the only productive variable (Figure 7) is India's 2019

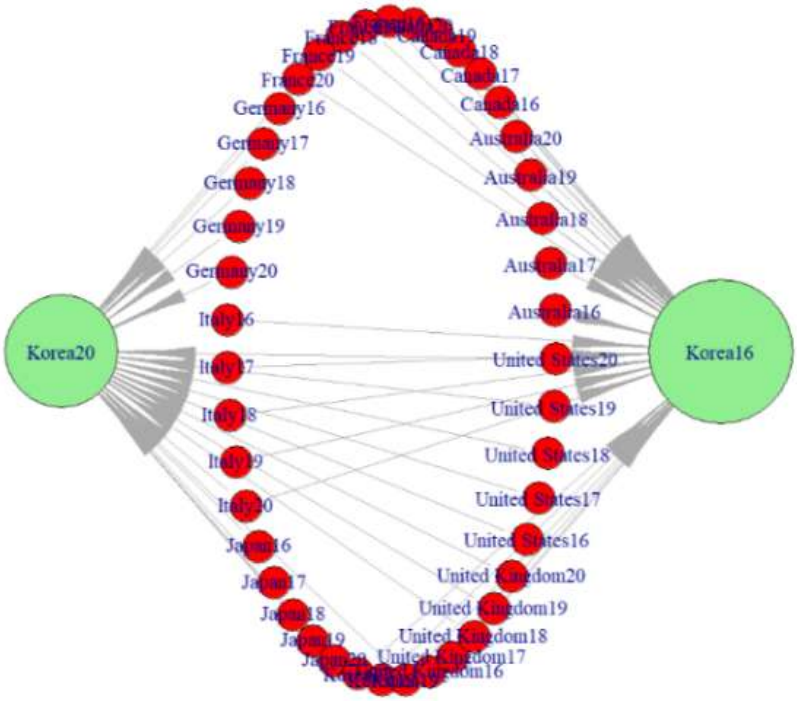


Figure 6:Reference Graph of Efficient DMUs of Developed Country Variables

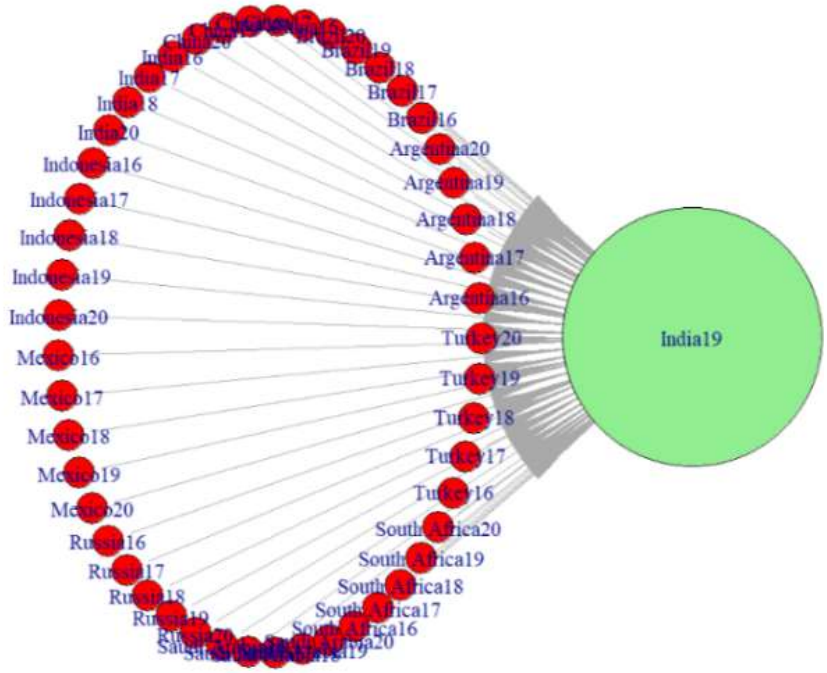


Figure 7: Reference Graph of Efficient DMUs of Developing Country Variables

Decision Tree & Random Forest

After analyzing the efficiency of countries in getting the maximum score from the DB and the GCI, the Decision Tree and Random Forest analysis to discover which output (DB &GCI) is more important. First, the productivity scores for all variables as a result of DEA were organized from the largest to the smallest, and the median value (mean value) was found. Values of more than the median value are given "1" labels, and the low values are given "0" labels. These labels were used as inputs for random forest and decision tree analysis. When creating decision trees, the top root contains all the classifications that represent the original data set. The tree then continues growing up to the first node where a particular feature variable is used to categorize the classifications. Since classifications can be divided into many patterns, the one with the greatest purity is taken care of (Lantz, 2015).



Figure 8: Decision Tree to Developed Country Variables

The results of the Random Forest analysis, which is a collection of decision trees, also support the results of the Decision tree analysis. DB scores were found to have a higher significance level for both developed and developing countries (also see Figures 8, 9, 10, and 11).

Analytical results presented in the diagram (Figure 8) showed that the DB is more important than the GCI for the group of developed countries in the G20. Additionally, it is seen that developed countries with a threshold value of more than 0.81 are efficient countries.

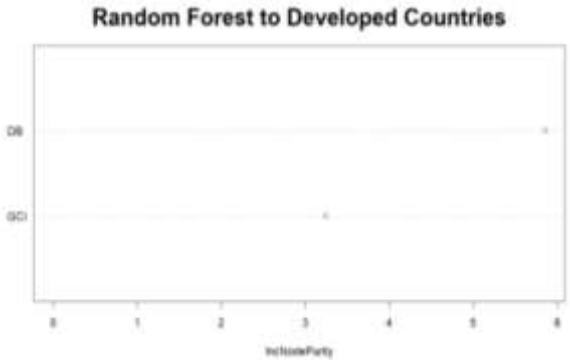


Figure 9: Random Forest to Developed Country Variables

Random Forest presents the importance level of the variables in the "IncNodePurity" graph. In Figure 9, "IncNodePurity" scores the purity levels of the nodes of the DB and GCI variables. In other words, the variables are scored according to whether the "node variance" is high or low. Accordingly, it is seen that the importance level of DB is calculated as 5.78 while the importance level of GCI is 3.39.



Figure 10: Decision Tree to Developing Country Variables

On the other hand, when the decision tree is carried out to the developing countries group (Figure 10), the DB data comes into prominence like the result of the developed countries group. Another remarkable point here is that countries with a threshold value between 0.29 and 0.52 are more effective countries.

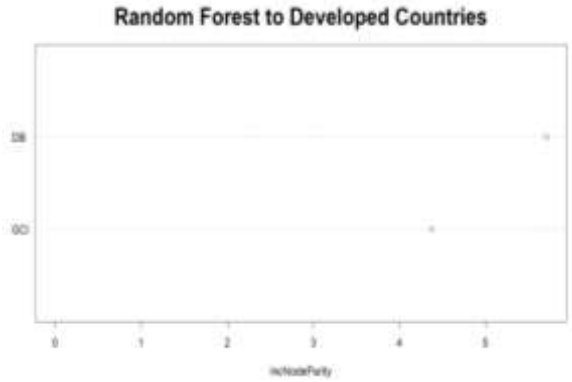


Figure 11: Random Forest to Developing Country Variables

In Random Forest analysis (Figure 11), DB appears to be slightly more efficient, and the nodes are a little purer than GCI. GCI scored 4.54, while DB scored 5.57. The difference in purity level is only 0.98 for the developing country group while it is 2.39 for developed countries. Although this is at different levels, it shows that DB scores are more important in the decision tree and random forest analysis methods performed for all G20 countries.

V. CONCLUSION

Regardless of their economic or political systems, all countries struggle to increase their welfare and become economically stronger. The wealth and economic strength of each country are different from each other. While some countries have advantages in various fields such as technology, national wealth, geography, etc., some countries are in a disadvantageous position. Therefore, these differences should be considered when making comparisons and all countries should be evaluated according to their resources.In this study, GDP per capita was chosen as a comparable indicator of national resources and their efficiency in maximizing DB and GCI scores was analyzed by the DEA method.In addition, it has been tried to determine which of the DB and GCI is more important for G20 countries by Decision Tree and Random Forest analyses. The DB scores the facilitation and simplification of conditions to create new jobs and stimulate the economy per country. GCI scores the level of adoption and accessibility of information and communication technologies. In addition, the DB criterion is more important for both groups. From this point of view, within the framework of this analysis, it may be useful for developed countries to keep an eye on South Korea and for developing countries to focus on India in 2019 and examine their policies. On

the other hand, regardless of the development level of the countries, providing a facilitating environment for entrepreneurs and investors and preparing the necessary infrastructure to increase their comfort can make significant contributions.

REFERENCE

[1] Acs, Z. J., & Varga, A. (2005). Entrepreneurship, agglomeration and technological change. Small business economics, 24, 323-334.
[2] Ani, T. G. (2015). Effect of ease of doing business to economic growth among selected countries in Asia. Asia Pacific journal of multidisciplinary research, 3(5), 139-145.
[3] Ayanso, A., Cho, D. I., & Lertwachara, K. (2014). Information and communications technology development and the digital divide: A global and regional assessment. Information Technology for Development, 20(1), 60-77, DOI: 10.1080/02681102.2013.797378
[4] Baumol, W. J., & Strom, R. J. (2007). Entrepreneurship and economic growth. Strategic entrepreneurship journal, 1(3-4), 233-237.
[5] Bosma, N., & Schutjens, V. (2011). Understanding regional variation in entrepreneurial activity and entrepreneurial attitude in Europe. The Annals of regional science, 47, 711-742.
[6] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.
[7] Dreher, A., & Gassebner, M. (2008). Greasing the wheels of entrepreneurship? The impact of regulations and corruption on firm entry. CESifo Working Paper Series, 1-28.
[8] Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society: Series A (General), 120(3), 253-281.
[9] Ghatak, A. (2017). Machine Learning with R. Springer, doi: 10.1007/978-981-10-6808-9
[10] Hargittai, E. (1999). Weaving the Western Web: Explaining differences in Internet connectivity among OECD countries. Telecommunications policy, 23(10-11), 701-718.
[11] Huawei, (2020). Shaping the New Normal with Intelligent Connectivity. Global Connectivity Index 2020.
[12] Huawei, (2023). Methodology. Gci – Global Connectivity Index. Huawei. Reached at https://www.huawei.com/minisite/gci/en/methodology.html 31.01.2023
[13] International Monetary Fund (IMF), (2022). World Economic Outlook – October 2022 Edition. International Monetary Fund.
[14] International Telecommunication Union (ITU), (2022). Global Connectivity Report 2022. ITU Publications. ISBN: 978-92-61-33551-9
[15] Mihaylova-Borisova, G., Nenkova, P. (2021). DEA Efficiency Approach in Comparing Macroeconomic Performance of EU and Balkan Countries. Economic Studies, 30(6).
[16] Nave, E., & Rodrigues, R. G. (2022). Economic development levels and ease of doing business: is there a relationship?. Review of International Business and Strategy, (ahead-of-print).
[17] Lantz B, editor. Machine learning with R. 2nd ed. Birmingham: Packt Publishing, 2015.
[18] Lewis, N. D. C. (2017). Machine learning made easy with R: an inuitive step by step blueprint for beginners. ISBN-10: 1546483756
[19] Wennekers, S., & Thurik, R. (1999). Linking entrepreneurship and economic growth. Small business economics, 13, 27-56.
[20] World Bank, (2020). Doing Business 2020. The World Bank Group. DOI: 10.1596/978-1-4648-1440-2
[21] World Bank, (2023a). About Us - Doing Business Archive. The World Bank. Reached at https://archive.doingbusiness.org/en/about-us01.02.2023
[22] World Bank, (2023b). GDP per capita (current US$) – Data. The World Bank. Reached at https://data.worldbank.org/indicator/NY.GDP.PCAP.CD 02.02.2023
[23] World Bank, (2023c). Metadata Glossary – Data. The World Bank. Reached at https://databank.worldbank.org/metadataglossary/statistical-capacity-indicators/series/5.51.01.10.gdp 02.02.2023

★★★