Importance of region and other socio-economic factors in the model of business efficiency

Mária Michňová[1] and Silvia Megyesiová[2]

[1,2] University of Economics in Bratislava, Faculty of Business Economics with seat in Košice, Department of Quantitative Methods, Tajovského 13, 041 30 Košice, Slovak Republic

maria.michnova@euba.sk

https://doi.org/10.53465/EDAMBA.2021.9788022549301.317-326

Abstract. The importance of analyzing and predicting business efficiency, both overall and economic, is proving necessary, especially in the current global situation, when the world has been hit by a pandemic, causing economic slowdown, closures and, in some cases, fatalities and of course negative effects on the overall economy of countries and the world. In this paper, we focused on the analysis of indicators as predictors of business efficiency. We used a sample of 150,000 companies accounting in the double-entry bookkeeping system in Slovakia. Using data on enterprises, we calculated a number of financial indicators to which we added socio-demographic indicators and focused on the indicator of return on assets, in which we determined the strength of the region as a predictor of other indicators of corporate efficiency. As part of the analysis, we used Kruskal-Wallis statistical tests to determine differences and machine learning techniques such as the XGBoost Tree model and the CHAID algorithm to determine the significance of the predictor.

Keywords: Business, Efficiency, Machine Learning

JEL classification: *C01, C38, C44, C53, M21*

1 Business Efficiency

We understand the concept of efficiency in the economy as a state when the economy allocates its resources efficiently and thus uses them efficiently. Economic efficiency is a state where each resource is allocated in a way where resource waste is minimized. If the economy is economically efficient, any changes made to help one subject would harm another subject [1].

Business efficiency of enterprises is one of the main problems undertaken by economics studies [2]. Its always very important, it provides entrepreneurs with the possibility to survive, what is the key condition to realize other objectives such as growth, development, maximizing owners benefits or building market value therefore it has a timeless character [2,3].

This issue is addressed by a number of authors, whose opinions differ in some cases but in many cases complement each other. In the literature one can find different definitions and interpretations for efficiency term, but generally it is considered from two perspectives: economic and organizational [4]. Lisý et al. (2007) speak of efficiency as the ability of the economy and its subjects to use resources as rationally as possible and to produce at the limit of production possibilities [5]. The output of the company is provided products and services, which arise from the consumption of production factors representing the inputs of the company. Thus, the ratio of output to input expresses the efficiency of the company [6,7].

The complexity of objective category for the organization and the variety of criteria to evaluate its efficiency the issue of its measurement and evaluation has multi-faceted character [2].

1.1 Measuring business efficiency

We can measure the efficiency of companies on the basis of several criteria. In its analysis, it is necessary to take into account both financial and socio - economic factors that affect it. The approach of different authors to evaluating the performance of organizations varies across studies, research, and articles, and the methods used to quantify them vary [8].

The most commonly used methods of measuring the economic efficiency of enterprises are considered to be:

- economic efficiency indicators,
- broader financial analysis, including the identification of economic standards.

In the analysis, we used several financial and non-financial socio-demographic indicators, focusing on the return on assets (ROA), because this indicators are considered indicators that characterize the efficiency of companies. The ROA indicator indicates how much profit an organization has made as a result of investing in its assets. The ROA indicator was calculated using the ratio:

$$ROA = \text{net profit} / \text{total assets};$$

it is the ratio between net income and total assets held by the entity [8]. The high level of this indicator highlights a high performance. The ROA indicator is often used to measure the efficiency and performance of companies, and is considered to be the most comprehensive indicator of measuring an organization's performance due to a combination of efficiency and effectiveness [9,10].

2 Data analysis

In the analysis, we focused on larger Slovak companies accounting in the system of double-entry bookkeeping, we worked with financial and non-financial indicators, which include the region in which the company is located, its size, ownership. In the initial phase of data analysis, we calculated financial and economic indicators, while after adjusting the data, we were left with a final sample containing 149,236 companies containing data between 2016 and 2019 (because we only had data available in this time interval), on the basis of which we created the entire analysis. It was an analysis of companies in the time interval just before the outbreak of the Covid-19 pandemic.

In the first phase, based on the ROA indicator, we analyzed whether there are differences between individual regions of Slovakia, considering the distribution by region. The ROA indicator shows whether a company can use its resources efficiently. For more accurate results of the analysis, we cleaned the sample with companies whose ROA values were outliers. We calculated the values of the ROA indicator and from them we calculated the median values of return on assets in the regions in individual years.

Table 10. Median of ROA indicators of companies by regions in Slovakia

	Year	**region_id**							
		62	**63**	**64**	**65**	**66**	**67**	**68**	**69**
Median of ROA in %	**2016**	2,45	2,65	3,18	2,71	3,02	2,40	2,59	2,36
	2017	2,45	2,48	3,05	2,57	2,92	2,34	2,37	2,31
	2018	1,85	1,92	2,49	1,99	2,52	1,99	1,69	1,86
	2019	1,44	1,41	1,85	1,31	2,10	1,36	1,36	1,38

** 62-Bratislavský kraj; 63-Trnavský kraj; 64-Trenčiansky kraj; 65-Nitriansky kraj; 66-Žilinský kraj; 67-Banskobystrický kraj; 68-Prešovský kraj; 69-Košický kraj*

The return on assets indicator tells us about the profit that the company earned from 1 euro of assets. Due to the large number of companies, we decided to point out the median values of return on assets. The median values are positive but low. We see that in none of the regions did the median value exceed 5 %. We can point out the fact that from 2016 to 2019 the values decreased in all regions of Slovakia which means that the return on assets gradually decreased, in some cases there could be a loss, ie the investment exceeded the profit. In 2016, the median return on assets was above 2 % in all regions and above 3 % in regions 64 (Trenčiansky kraj) and 66 (Žilinský kraj). In the following years, we observe a decline, while by 2019 these values were lower than 2 %, with the exception of region 66, where we record a median of 2.1 %. The highest decrease was recorded in region 65 (Nitriansky kraj) by 1.41%, the lowest decrease by 0.92% was in region 66.

We tested the values based on the Kruskal-Wallis test. The Kruskal-Wallis test is a nonparametric test, which we used for comparing all regions of Slovakia.

We have two basic hypotheses:

H0: There are no differences between regions.

H1: There are differences between regions.

We analyzed data for the entire period as well as individual years. There were differences between individual regions throughout the period and in individual years, which meant that we rejected the H0 hypothesis and at the 95% confidence level we leaned towards the H1 hypothesis that there are differences within the ROA indicator between Slovak regions.

Table 11. Kruskal-Wallis test about the differences between regions of Slovakia

Sample 1-Sample 2	Adj.Sig. 2016-2019	2016	2017	2018	2019
62-64	**0.000**	**0.012**	0.238	**0.006**	0.330
63-64	**0.000**	1.000	0.751	0.266	0.109
62-66	**0.000**	**0.000**	0.093	**0.000**	**0.002**
63-66	**0.000**	1.000	0.476	0.054	**0.001**
67-64	**0.000**	0.220	0.135	1.000	**0.007**
65-66	**0.000**	1.000	1.000	0.185	**0.005**
68-64	**0.000**	0.739	0.155	**0.002**	0.335
69-64	**0.000**	**0.001**	**0.004**	**0.009**	**0.026**
67-66	**0.000**	**0.030**	0.069	0.424	**0.000**
68-66	**0.000**	0.138	0.080	**0.000**	**0.007**
69-66	**0.000**	**0.000**	**0.001**	**0.001**	**0.000**
65-64	**0.003**	1.000	1.000	0.739	0.271
69-65	**0.004**	**0.048**	1.000	1.000	1.000
69-62	0.079	1.000	1.000	1.000	1.000
69-63	0.108	0.176	1.000	1.000	1.000
62-63	1.000	1.000	1.000	1.000	1.000
62-65	1.000	0.615	1.000	1.000	1.000
63-65	1.000	1.000	1.000	1.000	1.000
67-62	1.000	1.000	1.000	1.000	1.000
68-62	1.000	1.000	1.000	1.000	1.000
67-63	1.000	1.000	1.000	1.000	1.000
64-66	1.000	1.000	1.000	1.000	1.000
68-63	1.000	1.000	1.000	1.000	1.000
67-65	1.000	1.000	1.000	1.000	1.000
68-65	1.000	1.000	1.000	1.000	1.000
68-67	1.000	1.000	1.000	1.000	1.000
69-67	1.000	1.000	1.000	1.000	1.000
69-68	1.000	1.000	1.000	1.000	1.000

Despite the fact that we found that there are differences between regions, we cannot consider this as a sufficient result to claim that there are differences between companies in Slovakia depending on their operation in the regions. For regions with a p-value lower than 0.05, it is assumed that there are statistically significant differences and thus we accept the hypothesis H1, with the strongest values being equal to zero. These differences were not found in all years in the same regions, with the exception of regions 69-64 (Košický and Trenčiansky kraj) and 69-66 (Košický and Žilinský kraj), which means that the region is not a strong predictor. The values of the Kruskal-Wallis test (K-W test) when testing the diversity of regions were in many cases equal to 1.00. If the p value of the K-W test is equal to 1.00, we can say that there are no statistically significant differences between these regions [11].

To find out that the same differences in the same regions did not come out every year, we decided to test the region through classification and prediction algorithms to find out what role the region plays in classification and prediction for different financial indicators such as e.g., ROA indicator.

2.1 XGBoost Tree algorithm

One of the newer machine learning techniques is the XGBoost Tree algorithm, which we used to analyze whether the region is a suitable and reliable predictor in evaluating the effectiveness of companies. Tree highlighting is a very effective and very commonly used method of machine learning [12]. It is the method of highlighting / amplifying the tree that is found in various successful current applications. Due to its frequent use, the tree amplification method has been shown to provide the best results within many classification criteria, in a wide range of issues [13]. The most important factor in the success of XGBoost is its scalability in all scenarios. The system is ten times faster than other current solutions and is scalable to billions of examples in distributed or limited memory settings thanks to several important systems and algorithmic optimizations [14].

Within this model, we used the default settings, number boost round was 10, maximum depth 6, minimum child weight 1.0, maximum delta step 0.0, sub sample 1.0, Eta 0.3; Gamma 0.0; Colsample by tree 1.0; Colsample by level 1.0; Lambda 1.0 and Alpha 0.0.

To reduce the negative effect of overfitting in classification and prediction methods, we applied the model on both the training and test set according to the standard 80:20 distribution so that we could trust the model in terms of stability.

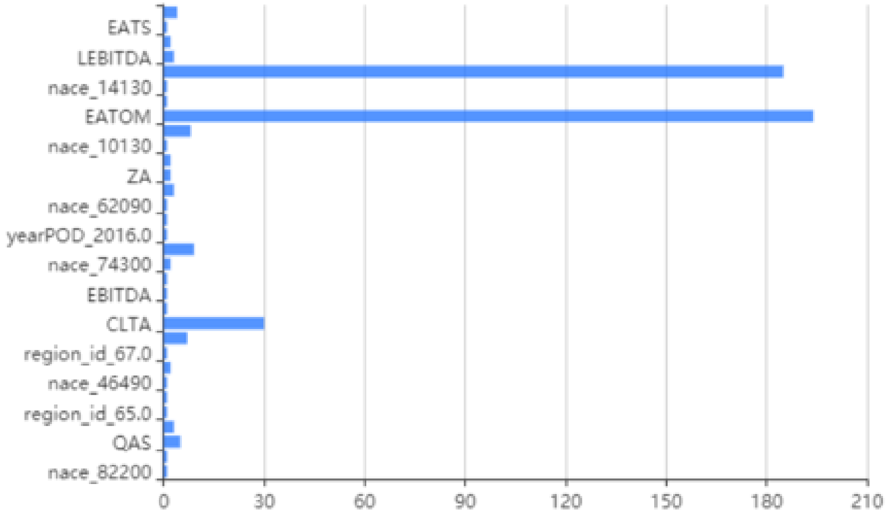

Fig. 2. Predictor importance with financial and social-demographic indicators

Within predictor importance, the region emerged as a significant predictor, but financial indicators were much stronger predictors. Among strong predictors of business efficiency, we include factors and indicators such as the industry, the region in which the company is located, EAT / S (net profit / sales), L / EBITDA (liabilities / EBITDA), EAT / OM (EAT / current assets), Z / A (liabilities / total assets), and QA / S (current assets / sales).

Table 12. Reliability of XGBoost Tree model

Partition	1_Training	2_Testing
Minimum Error	-17,647	-82,538
Maximum Error	12,295	15,59
Mean Error	-0,017	-0,022
Mean Absolute Error	0,046	0,054
Standard Deviation	0,173	0,544
Linear Correlation	0,998	0,977
Occurrences	149 236	37 563

In the training group, the accuracy of the created model was 0.998, but within the test group it was 0.977, which represents a high accuracy of the model, which means that sufficiently correct and important indicators were used in the model.

2.2 CHAID algorithm

Since the region also emerged as a significant predictor of financial indicators, we decided to use the CHAID decision tree algorithm to find out how strong a predictor is among socio-demographic indicators. The CHAID algorithm is based on chi-square statistics. The result of the test is a probability that is between 0 and 1. If the chi-square value approaches zero, there is a significant difference between the two classes being compared. If the value approaches to the one, it means that there is no significant difference between the two classes. CHAID is a segmentation method that can identify the relationship between a dependent variable and independent variables or predictors [15,16,17]. In addition to the region, we used the type of business ownership, the size of the business and the year as inputs. We used the default CHAID setting with an alpha of 0.05.

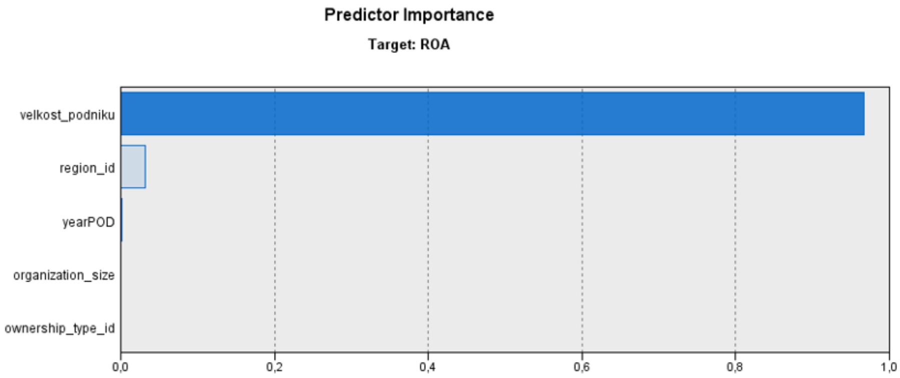

Fig. 3. Predictor importance of social-demographic indicators

Among socio-demographic indicators, the region is the second most important predictor at 0.03. In the CHAID decision tree (visualized below), other socio-demographic factors also play the most important role in micro-enterprises; in other types of enterprises, other socio-demographic factors also play an important role.

The created model of the decision tree CHAID in which the ROA indicator containing a sample of 149236 companies was divided in the first basic division according to the factor related to the size of the company into 3 branches with a p-value of 0.000. The Node 1 contained 30,203 companies, which was divided into two further branches Node 4 (27896 enterprises) and Node 5 (2307 enterprises) according to the type of ownership of the companies. For branch Node 5, the decision tree algorithm did not find a significant factor according to which the statistical significance should continue the division. However, the algorithm found a significant factor that can divide the Node 4 branch, depending on the year.

The Node 2 branch contained 70,614 data relating to micro-enterprises. This is the most numerous node, which further branches into 3 branches Node 6, 7 and 8 according to the organization size. For the Node 6 and Node 8 branches, the algorithm did not find a significant factor according to which they could be divided further, which means that for them this division was final. However, for the Node 7 branch, a factor was found according to which it was possible to divide the dressage into the other three Node 13, 14 and 15 branches according to the type of region in which the company is located.

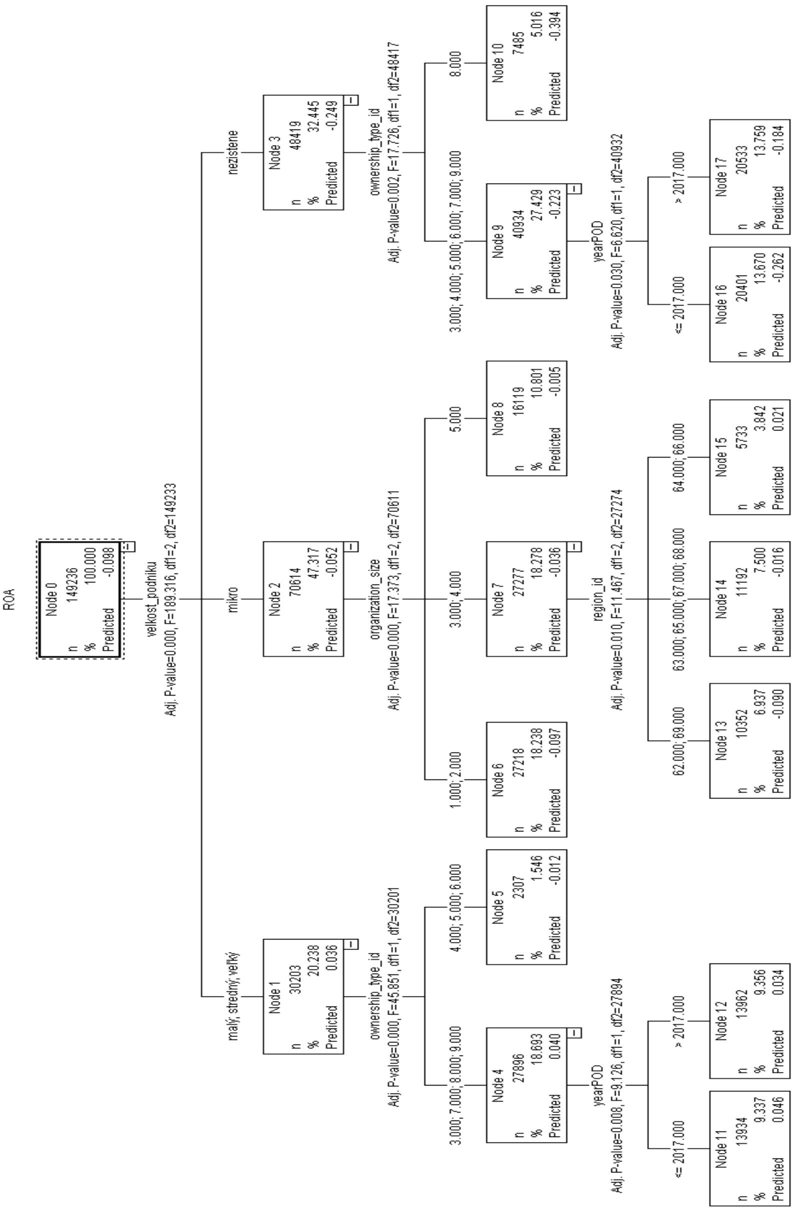

Fig. 4. CHAID Decision Tree of ROA indicators with social-demographic factors

The Node 3 branch, containing 48419 enterprises, contained enterprises whose size was not determined. This branch could be divided into two more according to the ownership type factor on Node 9 and 10. The Node 9 branch contained 40934 data, Node 10 only 7495, while the Node 9 branch was divided by the algorithm into two branches by year, according to the years on Node 16 (2016 and 2017 years) and on the Node 17 branch, containing data from 2018 and 2019.

The CHAID algorithm showed us the importance of the region as a factor that affects the efficiency of companies. The regional factor was also significant in addition to financial indicators, but among the socio-demographic factors we can consider it the second most important factor in the efficiency of individual companies.

Conclusion

We can define, measure and analyze the efficiency of a company through several ways and methods. This is due to the way we look at efficiency and what kind of company we consider effective, what values the financial indicators must acquire in order to be able to call a company efficient. At present, however, there are a number of opinions from different authors and a number of different studies on the latter we can rely on and create our own analyzes based on them. As part of the analysis, we decided to focus on the ROA indicator as a financial indicator and based on it, we analyzed the strength of the regional factor. Based on the performed analysis, we state that the region is a factor that has an impact on the efficiency of companies. Taking into account both financial and non-financial indicators and factors, the region came out as a significant factor, as its strength was not clear in addition to financial indicators, we decided to use CHAID decision tree to determine its strength in socio-demographic indicators, in this case as the second strongest predictor of business efficiency. The analysis was created on the basis of data on organizations in the time period 2016-2019, which represents the period before the outbreak of the COVID-19 pandemic. In further research, we will analyze organizations during a pandemic, the change of various financial and non-financial indicators, how the crisis affected their development, the organizations in which sectors of the economy were hit the crisis the most, which less and which not at all.

References

1. Horváthová, J., Mokrišová, M. Výkonnosť verzus efektívnosť podniku. Journal of Global Science. ISSN: 2453-756X. Online: http://www.jogsc.com (2017).
2. Nawrocki, T.L. The Use of Fuzzy Logic in the Enterprises Business Efficiency Assessment. ESSENCE AND MEASUREMENT OF ORGANIZATIONAL EFFICIENCY. Book SeriesSpringer Proceedings in Business and Economics, 229-248 (2016).
3. Jonek, K. I. Racjonalizacja kosztów jako sposób poprawy efektywności działania w spółce restrukturyzacji kopalń. In: Dudycz T, Osbert-Pociecha G, Brycz B (eds) Efektywność—rozważania nad istotą i pomiarem, vol 261. Wydaw. UE we Wrocławiu, Wrocław, pp 103–115 (2012).

4. Kozuń, C.G. Efektywność—rozważania nad istotą i topologią. Studia i Prace Kwartalnik KES SGH, No 4. Wydaw. SGH, Warszawa, pp 13–42 (2013).
5. Lisý, J. et all. Ekonómia v novej ekonomike. Wolters Kluwer (lura Edition). ISBN 80-8078-164-4 (2007).
6. Synek, M., Kopkáňe, H., Kubálková, M. Manažérske výpočty a ekonomická analýza. Vyd. 1. V praze: C.H. Beck, xvii, 301s. Beckova edice ekonomie. ISBN 978-80-7400-154-3 (2009).
7. Kotulič, R., Király, P., Rajčániová, M. Finančná analýza podniku. Tretie vydanie. 232 s. ISBN 978-80-8168-888-1 (2018).
8. Fiala, R.; Hedija, V.; Dvořák, J.; Jánský, J. Are profitable firms also financially healthy? Empirical evidence for pigbreeding sector. CUSTOS E AGRONEGOCIO ON LINE. Vol. 16, Issue: 1, 173-201 (2020).
9. Bumbescu, S.S. ANALYSIS OF ECONOMIC PERFORMANCE IN AGRICULTURE USING ECONOMETRIC MODELING. STUDIA UNIVERSITATIS VASILE GOLDIS ARAD SERIA STIINTE ECONOMICE. Vol. 30, Issue 3, 118-128 (2020).
10. Courtis, P. DuPont Ratio: A comprehensive measure of business performance, European Research Studies, Volume VI, Issue (1-2), 18-31 (2003).
11. Nasser Veiga, L.G, Tortato, U. Sustainable Market Indexes Behavior Analysis: a Study on the Brazilian Stock Market. POMS 23rd Annual Conference, Chicago, Illinois, U.S.A (2011).
12. Friedman, J. Greedy function approximation: a gradient boosting machine. Annals of Statistics,29(5):11891232.Online:https://scholar.google.sk/scholar?q=Greedy+function+approximation:+a+gradient+boosting+machine.+Annals+of+Statistics,&hl=en&as_sdt=0&as_vis=1&oi=scholar (2001).
13. Lin, J., Qi, CH., Wan, H., Min, J., Chen, J., Zhang, K., Zhang, L. 2021. Prediction of Cross-Tension Strength of Self-Piercing Riveted Joints Using Finite Element Simulation and XGBoost Algorithm. Chinese Journal of Mechanical Engineering. Online: https://link.springer.com/article/10.1186/s10033-021-00551-w (2021).
14. Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. DOI: 10.1145/2939672.2939785. Online: https://arxiv.org/abs/1603.02754 (2016).
15. Zounemat-Kermani, M., Stephan, D., Barjenbruch, M., Hinkelmann, R. Ensemble data mining modeling in corrosion of concrete sewer: A comparative study of network-based (MLPNN & RBFNN) and tree-based (RF, CHAID, & CART) models. Advanced Engineering Informatics, 43, 101030. DOI: 10.1016/j.aei.2019.101030.Online: https://www.sciencedirect.com/science/article/abs/pii/S1474034619306032 (2020).
16. Ramzai, J. Simple guide for Top 2 types of Decision Trees: Chaid & Cart. Machine Learning Fungamentals. Towards data science. Online: Simple guide for Top 2 Types of Decision Trees- CHAID & CART | Towards Data Science (2020).
17. Jijo, T.B., Abdulazeez, M.A. Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technnology Trend (2021).