iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: I Month of publication: January 2024

DOI: https://doi.org/10.22214/ijraset.2024.57958

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Role of Gross Domestic Expenditure on Research and Development on Economic GDP A Time Series Forecasting Approach Using ARIMA Model

Zeba[1], M. Afshar Alam[2], Harleen Kaur[3], Bhavya Alankar[4], Ihteram Raza Khan[5]
[1, 2, 3, 4, 5]*Department of Computer Science & Engineering, School of Engineering Sciences & Technology (SEST), Jamia Hamdard, New Delhi, India*

Abstract: Research and development is considered as one of the important factor to contribute strongly for sustainable development goals (SDG -9). These further may help in building resilient infrastructure, promoting industrialization, and fostering innovation. Contributions to research and development (GERD) and the number of researchers employed in R&D activities have a significant impact on research and development. The Innovation Index delves deeper than just total GERD figures. It analyzes how effectively countries allocate their R&D resources. This particularly provides reward to the countries that prioritize research in key areas like renewable energy, healthcare, and digital technologies while also considering the efficiency and impact of their R&D spending.

The current availability of limited data for index-based studies is not conducive to the design of policy scenarios and technology deployment models. In this paper, we have studied various machine learning models and have employed the ARIMA model to study the impact of data variables to forecast time series forecasting. In this study, a comprehensive R&D spending estimate and its correlation with other variables is analyzed to reveal the global GERD shuffle to escalate the studies on technology impact.
Keywords: Global expenditure on R&D (GERD), Autoregressive Integrated Moving Average (ARIMA), R&D intensity, S-curve forecasting, economic development, patent accumulation, and homoscedasticity.

I. INTRODUCTION

Technology and innovation diffusion show the realistic pathways of development using advanced machine learning and data screening techniques. Statistical learning advancement has resulted in the creation of machine learning methods, such as supervised and unsupervised methods.

With swift technology success in foremost areas, it has become necessary for policy analysts to design strategically the relevant decision-making for the commercialization of their inputs. Such inputs will help in generating sustainable concepts of global expenditure on R&D i.e. (GERD) and global innovation index (GII). Any emerging economy lays down its direct contribution to the total factor of production, where R&D and innovation play a pivotal role in scaling and production of the economy.

In this paper, the author has considered indicators of the global innovation index (GII) to analyze the policy purview of technology development, analysis, and its indiscriminate effect on technology translation. The fourth industrial revolution is an amalgamation of automation and a new exchange of networks using artificial intelligence and machine learning practices. Major changes occur in potential activity localization to alter the growth development dynamics and product & service value addition (Yangdol & Singh, 2022).

Supply chain management, inventory management, consumer analytics, construction, and shipping have now entered under the aegis of efficiently automated management, ever since the start of a new phase of productivity transformation. India's rank in terms of the *manufacturing value added index* has improved from 14th in 2000 to 5th rank in 2019. However competitive industrial performance rank is lowered to 38th rank in 2019, which is due to lower manufacturing sector share in GDP during the marked year. It has contributed only 17.1% of Gross value addition (GVA) citing only 11.2% of wholly owned employment (RBI Bulletin, June 2022). According to the OECD classification of economic activities on research and development, intensity and manufacturing lead economies broadly can be classified into five categories: High R&D intensity, medium-high R&D intensity, medium R&D intensity, medium-low R&D intensity, and low R&D intensity. Whereas R&D intensity is measured as expenditure on R&D as a percentage of GDP. As per the latest R&D intensity of India, it is about 0.7% in 2018 in contrast to other nations, which is on the lower side.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

In (Fig.1) the Technology Intensity of India's manufacturing sector, is represented, showing the R&D intensity of 31.8% in the medium and low industrial manufacturing sectors, with the lowest percentage of 10.8% in high R&D intensity industries.

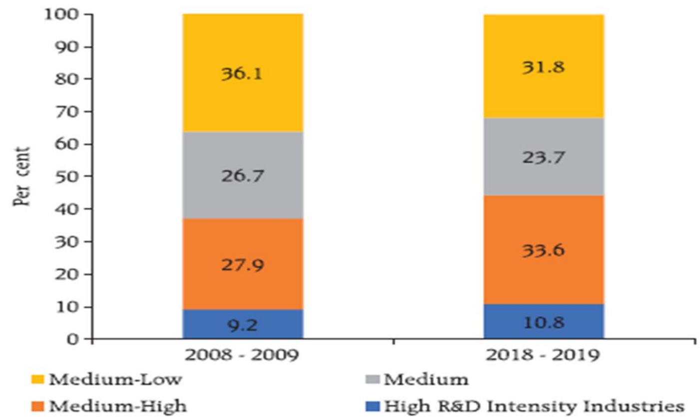


Figure 1: Technology Intensity of India's manufacturing sector

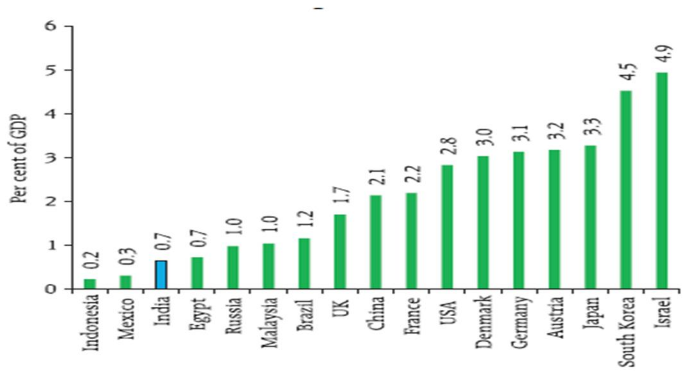


Figure 2. R&D percentage of GDP:2018
(Source: RBI-June 2022 bulletin, based on an annual survey of industries)

Globally, technology has led to an uplift and boost towards economic development in terms of the global average GDP structure. The trend shown in (Fig.2) has put focus on GERD i.e., Global expenditure on Research and Development, which directly hampers the cost-effectiveness of current research and development. Growth factors of the national innovation system are based on revisited characteristics of the drivers of economic growth and its relative indicators.

Here, GERD is one of the key factorial indicators, which helps in the diffusion of national research development, to empower innovation management and its further deliverables. This study deals with GERD at national and global levels, with numerous studies. (Siedlecki et al., 2020) analyzed innovation level studies at regional and based on geographical origin. Policymakers have driven emerging economies by closely monitoring the indicators of the national innovation system and its ranking solutions. This gives an analytical approach to drive the critical indicators of scientific policymaking.

II. LITERATURE REVIEW

Numerous studies have been reported on the importance of studying GERD as a key factor for the growth of the national innovation system, to identify research and development indicators for any emerging economy. According to (Siedlecki et al., 2020), the GDP of the economy stimulates other factors including research and development spending and ICT spending. Their study reflects the idea of higher spending on R&D in transition economies, which indicates productivity improvement, represented by the application of S-curve forecasting. Many pieces of literature have reported the rate of R&D spillovers in developing and industrially developed economies. (Almeida & Teixeira, 2007) have shared the negative effect of patenting, over a study of 88 countries from 1996-2003. They have found diverse interactions to support the patenting activity for R&D investment. Also, their penal data study shows an impactful scenario for less developed economies on their R&D intensity, from the accumulated patents, whereas it has given patent accumulation has given a slightly lower impact on highly developed economies.

(Otomo, 2017) has examined the impact of R&D expenditure via patent applications in the USA and Europe and found that despite an increase in the number of patents, there remains a diminishing and marginal increase in the country's R&D expenditure. (Guellec & Pottelsberghe de la Potterie, 2001) have quoted that the higher R&D intensity of the country gets benefits from foreign R&D spillovers. They have also mentioned that there is a positive impact of domestic R&D intensities between the two countries. It was analyzed that 0.1 % of the difference in intensity of research, between the two countries, generates a spread of about 0.002% between their development elasticity. In the same study, it was also analyzed that, if a firm of an individual country wants to take full advantage of international innovation spillovers, then expenditure of R&D plays a crucial role, by investing in adaptive research activities.

According to the Organization for Economic Cooperation and Development (OECD) research and working papers, it was identified that the elasticity of public research is higher when the business R&D concentration of an economy is higher. It helps in developing a stronger link between public and private research activities. OECD studies have shown that in 2019, total growth in real expenditure on R&D in most OECD countries was driven by business R&D growth, supporting 71% of research performance in these nations. The said expenditure rises by 4.6% in 2019, while the research expenditures in the government sector have given a steady rise of only 3.4%, and thus a decline in government institutions as relative R&D performers. Research and development (R&D) spending was forecasted to reach over 2.47 trillion U.S. dollars globally in 2022 (once local currencies are converted for purchasing power parity). This compares to around one trillion U.S. dollars in 2005 and around 555 billion U.S. dollars in 1996. Spending decreased in 2020 following the outbreak of COVID-19 but increased again in 2021 and was forecasted to do so in 2022 too. Given the increase in higher dimensional information to stratify the existing technology trends to emancipate policy scenarios and methods to optimize model selection. Although model selection is more challenging in terms of theoretical and empirical analysis of high dimensional data insights, here machine learning emphasizes algorithms that show transformative intelligence. There are many instances where machine learning has played a crucial role in analyzing the data forecast using time series analysis of high dimensional and multivariate data. (Liu & Chia Liang, n.d.) Have adopted a principal component analysis (PCA) to classify a national innovation system and to identify the economic growth in emerging countries. In the same study, LASSO clustering was also followed, which was used to calculate the coefficient on R&D expenditure as a percentage of GDP, which came up to 0.003. In this study, PCA and LASSO approach was undertaken to describe the intensity of external and internal constraints on the growth of developing countries.

In research (Tudor & Sova, 2022), it was identified that the driving factors for R&D intensity are key model data characteristics for estimation of the impact of high technology exports, the number of impactful researchers, and trade openness. Using exploratory data analysis and a dynamic system Generalized Method of Moments (GMM) panel model were the studied indicators by (Tudor & Sova, 2022) to share research and development expenditure in gross GDP. In the said study, results were scripted for the year 2007-2015, in which high technology exports have been statistically created, providing a significant effect on R&D expenses for middle- and low-income panel economies. Developing countries have made progress in the structural transformation of manufactured exports and trade openness. At the same, patents are the indicator of high R&D intensity in high-income panel economies. Many other studies have taken OLS (Ordinary Least Square) regression which used cluster analysis, as its first step in choosing the innovation indicators. The given data is then reduced to two main factors by running the principal component analysis (PCA). Another step in the same approach includes clustering analysis to classify the innovation indicators into classes, to categorize each country's innovation cluster.

Following previous research, the empirical model to introduce research and development-based innovation index indicators, the author has taken innovation indicators to compare the responsible research productivity among various groups of countries. A group of OECD countries is taken in this paper, including central Asia and South Asian countries, with their data on GERD value as a percentage of global GDP expenditure.

The author has also considered the number of researchers and their impact on bringing up the GDP and capital productivity of the concerned economies. Although, there are no such previous studies with the GERD data analysis and its impact on economic growth, the author has tried to bring ahead the relationship between the GERD and number of researchers employed in R&D, in a full-time equivalent ratio.

In this study, machine learning applications, including automated machine learning is employed. This will help in the identification of dependent variable factors available in the database. It will further help in the analysis of GERD as a performing indicator and its impact on economic development. Here ARIMA time series analysis was performed to forecast the next five years, and to study the data changes over a given period, using the historical data provided.

III. DATA

This study includes three variables from over twenty-four countries from the year 2015 – 2020. The variables are chosen based on the availability of the data. The data was drawn from the web portal of the world bank statistics domain of SDG indicators. Here, research and development indicators were chosen as one of the standard nine SDG indicators among seventeen sustainable development goals.

The SDG dataset provides the depth insight into the national innovation system which includes annual GDP growth percentage, number of researchers in R&D per millions of people in a particular economy, and R&D expenditure in terms of percentage of gross domestic product (GDP), calculated over the years. We have applied research intensity ratio projections and the number of researchers to GDP forecasts to derive global expenditure on research and development (GERD) projections. The variable selected in this study is mentioned with the keywords used during ARIMA modeling:

1) *AGRP:* Annual gross domestic product (GDP) growth percentage
2) *RDGDP:* R&D expenditure as a percentage of GDP
3) *RRDPM:* Researchers in R&D per million full-time equivalents (FTE)

IV. METHODOLOGY

With the change in global expenditure on research and development (GERD), the practice of innovating and its historical assessment changes as per the availability of data. We have first drawn the dataset series from the World Bank repository and cross-verify the same using other sources like the UNESCO cross-country dataset, to harmonize the authenticity of available data, for each country's GERD intensity. GERD estimates along with the other two variables were obtained for twenty-four countries for the year 2015-2020.

The gaps for several years of a few countries, for which data was not available in a dataset, were extrapolated using the earlier year dataset, and in some cases were set constant, where the continuous yearly data was not available for any selected variables. This has been done in the countries for which no reported GERD, GDP, and number of full-time equivalent (FTE) of researchers was not available.

Notably this dataset is also made available for several countries which lower- and middle-income group countries, as well as European and Central Asian countries, countries that are presently the part of International Development Association (IDA) and International Bank for Reconstruction and Development (IBRD) which lends in a group of poorest countries. As the future is uncertain, but based on the assumptions that the empirical regularities may persist over the coming years, we have developed the newly generated series of GERD for the period of the next 5 years to form projections of regional and global GERD intensities.

To enhance the results, a machine learning approach was carried out, to analyze the impact of GERD and researchers as per million full-time equivalents over the GDP of the country.

In time series analysis, ARIMA models are flexible and are used in a wide range of analyses. The time series model can provide short-run forecasts, in a very precise manner for the variable data on the concerned variable. ARIMA where stands for (p, d, q), which represents Autoregressive-Integrated-Moving Average with the three parameters where, p, represents the order of autoregression, d, and q represent the degree of differencing, and the order of moving average, respectively. As Box and Jenkins have worked for the forecasting of a large variety of time series data, ARIMA is also referred to as a Box-Jenkins model. ARIMA is the process of three steps, based on random disturbances that arise in empirical time series observation with the following three steps: (i) (AR) Auto-regressive process, (ii) Differencing process, (iii) Moving Average.

As the GDP, GERD variables are going to show an exponential rise in their values therefore we have used the simple time series method of the ARIMA model to forecast the GERD and percentage GDP forecast for upcoming years. ARIMA shows higher fitting with more accuracy than the exponential smoothening, hence it sufficiently captures both seasonal and non-seasonal forecasting trends in the dataset.

Here, we have focused on non-seasonal forecasting trends to describe the growth of variables over the period of years taken, analyzing that the GERD pattern will increase over the period and is in the upgrowing non-seasonal pattern. We believe that the ARIMA model will fit appropriately within the current dataset, to reflect the best models. To predict the model fit, (p, d, q) is identified by the autocorrelation function (ACF) and partial autocorrelation function (PACF) whereas p is the autoregressive term, d being different ordered, and q is the moving average term. ARIMA results are derived out of the Akaike information criterion, which is the goodness of fit test, where a minimum of AIC is considered the best-fit outcome. All statistical analysis was conducted using R library "forecast, series and zoo."

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

V. RESULTS AND DISCUSSIONS

In this section, we first tested the stationarity of the data based on ACF (autocorrelation, and PACF (Partial autocorrelation function) and plotted the autocorrelation time series. Autocorrelation is a correlation between the observation of time series separated by the k time units. While (fig. 6) below for ACF shows a correlation between points, up to and including its lag unit, its correlation coefficient is at the x-axis and the number of legs is represented at the y-axis. Similarly, the partial autocorrelation function measures relationship strength with other terms that are accounted for, i.e., the intervening lags in the model. PACF is the combination of observations excluded during the initiation phase and other intervening observations.

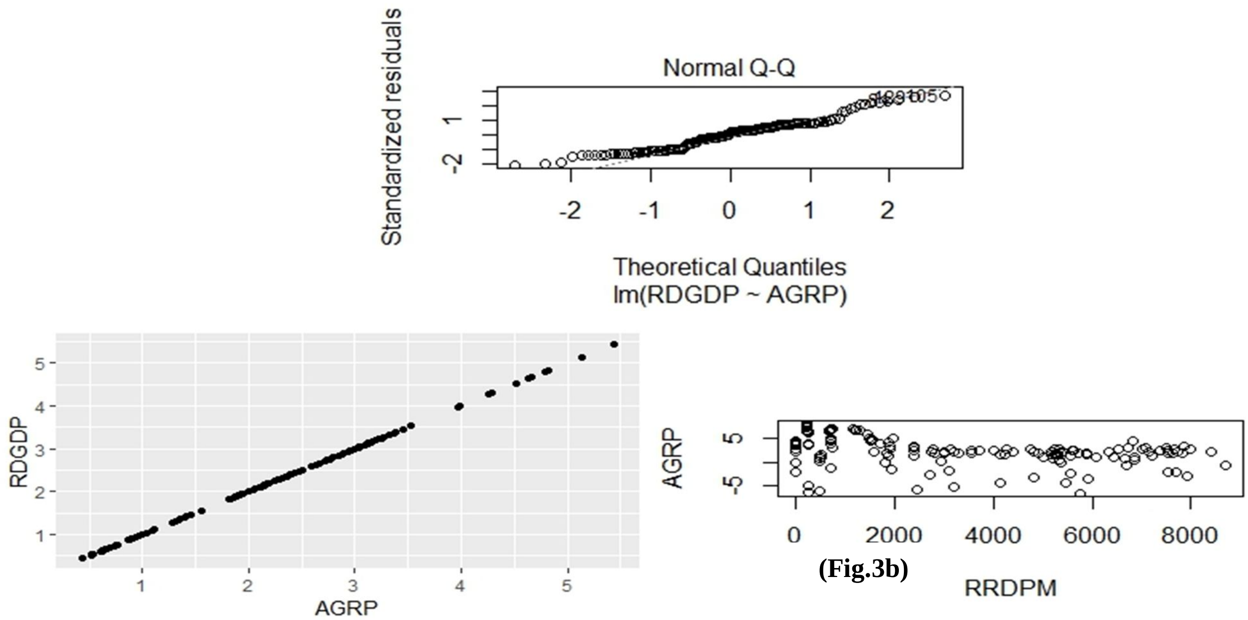


Figure 3. Full time Researchers in R&D in per million (RRDPM) Vs Annual GDP growth percentage (AGRP)

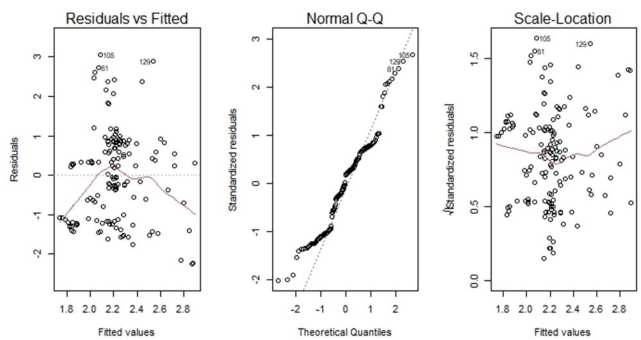


Figure 4. Regression values for Annual GDP growth percentage (AGRP) and R&D expenditure as a percentage of GDP (RDGDP)

(Fig.6) below represents the partial autocorrelation (PACF)and autocorrelation (ACF) plots as observed. In the data we have performed multiple regression, to check whether our data meets our model assumptions and obtained the residual plots, where residuals are the unexplained variance.

The red line in (Fig.5) representing the mean of residuals is horizontal and centered around zero, showing the outliers and biases in data. In a normal Q-Q plot placed at the center, we can see the real residuals from an almost perfect one-to-one line with the theoretical residues except for a few of the outliers at the end of the data, thus forming a perfect model. Based on residuals, we can say that our model meets the assumption of homoscedasticity.

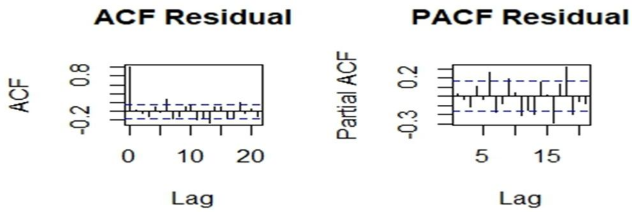


Figure 5. ACF and PACF residual plots. Results of forecasts and Growth Rate of GERD over a period (2021-2029)

Standardized residual plot of R&D expenditure as percentage of GDP (RDGDP) Vs Annual GDP growth percentage (AGRP)

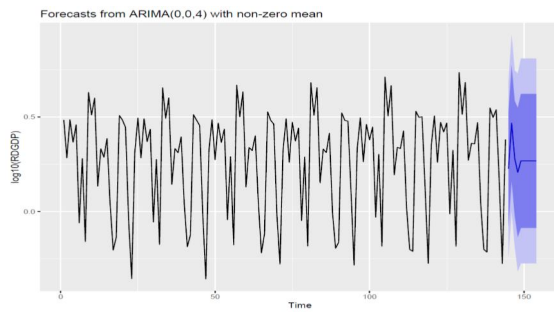


Figure 6. ARIMA forecast map for GERD.
(2021-2029)

According to the results obtained via the ARIMA forecasting, it was scrutinized that the forecasted value of GERD i.e., global research and development expenditure increases over the period (2021-2029). However, the GERD growth rate suggests that there is a steady and constant rate also available after a certain period, and a lower trend in all categories from lower limit growth to higher limit growth rate. Hence it is recommended that it is essential to present the lower and upper growth chart with the point forecasted value to attain a clear picture of the innovation index and GERD.

The lower and higher projections imply that the global R&D expenditure could vary over the upcoming years. This shows that a range of prospective global research in the mentioned countries taken in the database, reflects the prospective global research future, largely in terms of increased R&D expenditure along with the increase in the number of researchers in FTE which is the major supplicant in GERD upliftment and hence the GDP of any economy.

VI. CONCLUSION

In this study, an ARIMA model has been estimated to forecast GERD and its growth rates for years ahead by utilizing time series data over the period 2015-2020. Instead of using the ordinary method of least square estimation algorithm ARIMA has been used for more accurate forecast prediction. This aspect confirms that for any country GERD is a positive contributor to the rise in GDP and its further economic development in terms of technology and infrastructure build-up.

This study has also been checked by using the Ljung-Box statistic, the study shows that only two variables have shown significant coefficient values with AR and MA, showing their effectiveness in forecasting the GERD forecast to show its impact on real GDP. The findings of this study are more important for the policymakers in implicating the assessment using the macroeconomic values for economic development. Policymakers can use such methods to identify other macro indicators for the innovation index, like foreign R&D expenditure and R&D investment globally, and their further impact on the current GDP scenario. However, it would be interesting to expand this research by including such important contributing factors that influence the GDP and growth rates, in the future.

VII. ACKNOWLEDGMENTS

We are thankful to the Visvesvaraya scheme of the Ministry of Electronics, Information Technology (MEITy), for their support in carrying out this research of capability building and policy experimentation.

REFERENCES

[1] Yangdol, V. D., Rigzen, & Singh, S. (2022, June) RBI Bulletin: https://www.rbi.org.in/Scripts/BS_ViewBulletin.aspx?Id=21072.

[2] Almeida, A., & Teixeira, A. A. C. (2007, December). Does Patenting negatively impact on R&D investment: An international panel data assessment.

[3] Cantner, U., & Vannuccini, S. (2018). Elements of a Schumpeterian catalytic research and innovation policy. Industrial and Corporate Change, 27(5), 833–850. https://doi.org/10.1093/icc/dty028

[4] Guellec, D., & Pottelsberghe de la Potterie, B. van. (2001). R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries. OECD Science, Technology, and Industry Working Papers, 3. https://doi.org/10.1787/652870318341

[5] Liu, Y., & Chia Liang, C.-. (n.d.). Exploring Innovation Factors Affecting Economic Performance in Emerging Asian Countries: A Machine Learning Approach. Singapore Economic Review. https://doi.org/10.1142/S0217590820450058

[6] Otomo, P. (2017). The Impact of P The Impact of Patents on Research patents on Research & De ch & Development Expenditure development Expenditure as a Percentage of Gross Domestic Product: A Case in the U.S. and EU Economies: A Case in the U.S. and EU Economies. Economics Student Theses and Capstone Project.

[7] https://creativematter.skidmore.edu/econ_studt_schol?utm_source=creativematter.skidmore.edu%2Fecon_studt_schol%2F55&utm_medium=PDF&utm_campaign=PDFCoverPages

[8] Siedlecki, R., Papla, D., Prędkiewicz, P., & Bem, A. (2020, April 20). Forecasting of GDP and R&D Spending in Transition Economies.: Estimation Using S-curve and Modified S-curve. International Business Information Management Association (IBIMA)

[9] Tudor, C., & Sova, R. (2022). Driving Factors for R& D Intensity: Evidence from Global and Income-Level Panels. Sustainability, 14(3), 1854. https://doi.org/10.3390/su14031854

[10] Wee, C. K., & Nayak, R. (2019a). Adaptive load forecasting using reinforcement learning with database technology. Journal of Information and Telecommunication, 3(3), 381–399. https://doi.org/10.1080/24751839.2019.1596470

[11] Faloutsos, C., Gasthaus, J., Januschowski, T., & Wang, Y. (2019). Classical and Contemporary Approaches to Big Time Series Forecasting. Proceedings of the 2019 International Conference on Management of Data. https://doi.org/10.1145/3299869.3314033.

[12] Byungun Yoon, M. (2012, January). Articles – Manupatra. https://articles.manupatra.com/#. https://articles.manupatra.com/article-details/Applicability-of-Patent-Information-in-Technological-Forecasting-A-Sector-Specific-Approach

[13] Choi, Y., Park, S., & Lee, S. (2021). Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data. Scientometrics. https://doi.org/10.1007/s11192-021-04001-1

[14] Firat, Woon, Madnick, A. K., Wei Lee, Stuart & Composite Information Systems Laboratory (CISL) Sloan School of Management, Room E53-320 Massachusetts Institute of Technology Cambridge, MA 02142. (2008). Technological Forecasting – A Review. https://skat.ihmc.us/rid=1NMM9Z943-1197XJF-5CGB/2008-15.pdf

[15] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009

[16] Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

[17] Jang, H. J., Woo, H. G., & Lee, C. (2017). Hawkes process-based technology impact analysis. Journal of Informetrics, 11(2), 511–529. https://doi.org/10.1016/j.joi.2017.03.007

[18] Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. World Patent Information, 65, 102035. https://doi.org/10.1016/j.wpi.2021.102035

[19] Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. Technological Forecasting and Social Change, 127, 291–303. https://doi.org/10.1016/j.techfore.2017.10.002

[20] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020a). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18.https://doi.org/10.3390/e23010018

[21] Martino, J. P. (2003). A review of selected recent advances in technological forecasting. Technological Forecasting and Social Change, 70(8), 719–733. https://doi.org/10.1016/s0040-1625(02)00375-x.

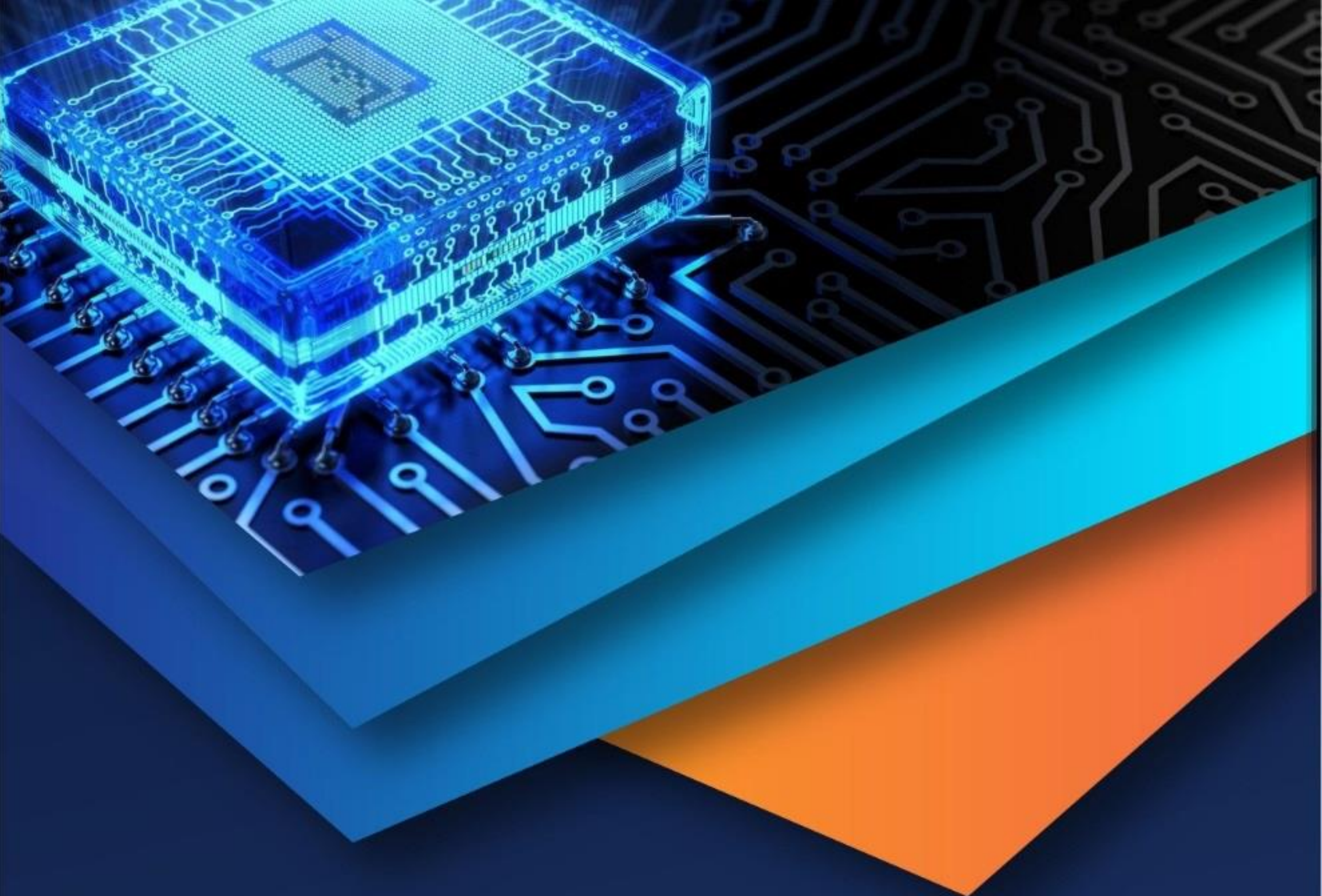
[22] OECD.OECD, Patent-Statistics-Manual. https://www.oecd.org/sti/inno/oecdpatentstatisticsmanual.html

[23] Shobana, G., & Umamaheswari, K. (2021). Forecasting by Machine Learning Techniques and Econometrics: A Review. 2021 6th International Conference on Inventive Computation Technologies (ICICT). https://doi.org/10.1109/icict50816.2021.9358514

[24] Smith, G. W. (2005). Intellectual Property Rights, Developing Countries, and Trips. The Journal of World Intellectual Property, 2(6), 969–975. https://doi.org/10.1111/j.1747-1796.1999.tb00102.x

[25] Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. Technological Forecasting and Social Change, 75(9), 1396–1405. https://doi.org/10.1016/j.techfore.2008.04.004

[26] Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Information Fusion, 77, 29–52. https://doi.org/10.1016/j.inffus.2021.07.016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY