Golden Sun-Rise
International Journal of Multidisciplinary on Science and Management
ISSN: XXXX-XXXX / Volume 1 Issue 1 January 2024 / Page No: 15-20
Paper Id: IJMSM-V1I1P103 / Doi:

*Research Article*

# Advanced GDP Analysis Using Artificial Intelligence

**Naveen K[1], Santhosh R[2], Jayalakshman A[3]**

[1,2,3] *Department of Electronics and Communication Engineering, Chalapathi Institute of Engineering and Technology (Autonomous), Andhra Pradesh, India.*

Received: 29 December 2023 Revised: 16 January 2024 Accepted: 28 January 2024 Published: 02 February 2024

***Abstract -*** By using datasets unique to each nation, the project uses artificial intelligence (AI) to predict GDP per capita and automate GDP analysis. The system will perform a thorough analysis of the state of the economy, using algorithms to find trends and connections between different variables and GDP. Giving users access to an intuitive platform for displaying insights is the aim in order to support data-driven decision-making. Data science is essential for transparently managing and deriving insights from economic data in a world when technology is everywhere. This research is in line with the increasing significance of data science in helping people and governments comprehend and navigate economic environments.

***Keywords -*** *Jupyter, Colab, Artificial Intelligence, Gross Domestic Product, Auto Regressive Integrated Moving Average Model.*

## I. INTRODUCTION

A prominent technique that aids governments in boosting economic output, efficiency, and openness is data science. Through statistical analysis, data preparation, predictive modeling, and machine learning, it addresses a variety of issues in many businesses and the economy. Self-reliance is a key component of sustainability goals, and it has a high correlation with measures of a country's economic health such as GDP. This project's primary objective is to forecast Gujarat State's GDP figures using Random Forest Regressor and ARIMA machine learning techniques. Time-series analysis and regression models are developed for in-depth GDP analysis and visualization. The relationship between GDP and influential factors is estimated by regression models. Time-series forecasting increases projected accuracy by utilizing patterns and the ARIMA model.

## II. LITERATURE SURVEY

Davy Cielen, Manning Publications Co., M. A., a. D. B. M. (2016) Data scientists do idea research in order to fulfill the main duties assigned to them. Some of the key components of data science include using No SQL, graph databases, data visualization, and proposing an appropriate data science approach. Data-related languages like Rare and Python are commonly utilized. Learn how using Python can help with decision-making when working with enormous datasets that need to be stored on multiple machines or when data is moving too fast for one system to handle. With the use of Sci-kit-learn, Stats Models, and several Python data science packages, this research provides a hands-on experience.

M. Kumar Manoj (2014, vol. 14, Studies in Business and Economics).Through time series forecasting, models that project values into the future are developed using present (historical) data. A forecast was an estimation or assessment that projects future events based on data from past occurrences and present patterns. Conversely, a forecast is an anticipation of an event, whether or not it is known beforehand.

Data on GDP is not a linear format. A method for turning a non-stationary time series into a stationary one is called difference. This is a crucial phase in the ARIMA model's data preparation procedure. Plots of the partial auto correlation function (PACF) and auto correlation function (ACF) are used to provide the ARIMA parameters. Patel, M. (2020). The paper covers a variety of machine learning techniques that have improved understanding of regression models.

D. REID, Economical, 35 (1968), 140-Researchers have long used a variety of combination strategies to increase forecasting accuracies. Specifically, better outcomes are anticipated when different approaches are used to forecast time series data. The basic combination strategies, such mean and median, continue to be effective and widely used despite the many combination approaches that have been presented to date. To incorporate the benefits of both of these approaches, this research suggests a novel combination method that is based on the mean and median combination methods.

E. B. Valeria Fonti, Research Paper in Business Analytics, VU Amsterdam, p. 25 One of the most crucial and challenging parts of statistical modeling is feature selection. This is due to the possibility that developing a model that addresses every possible scenario would be challenging, and the performance standards for various data sets may differ. Economic growth theory commonly evaluates and explains steady-state or long-run growth using the human development index (HDI) or other standard-of-living metrics, or the percentage increase in national income (Sengupta, 2011).Two topological descriptors, the kappa shape index and connection index, have been used to predict the antagonistic activity of β-carboline and its thirteen derivatives. Cache Pro software was used for molecular modelling and geometry optimization of all the compounds, and MOPAC 2002 was used in the semi-empirical PM3 technique to assess the descriptor values. To forecast activity, multiple linear regression analysis (MLR) was applied.

M. R. Segal, Random Forest Regression and Machine Learning Benchmarks, 2003-Recently, Breiman (2001a, b) developed an ensemble strategy for classification and regression on a suite of benchmark datasets, demonstrating exceptional performance in terms of prediction error. The technique is known as "random forests" since the basic components of the ensemble are tree-structured predictors, and each of them is created using a random injection. It is even more amazing that the extraordinary performance may be achieved with what appears to be a single tuning paddle, to which very little sensitivity is required. Every single tree that makes up the forest has reached its full potential in depth. Although this reduces regard bias, variance comes with the known trade-off.

## III. EXISTING WORK

To understand a nation's overall economic performance, a variety of aspects of the Gross Domestic Product are examined in GDP analysis. These are a few typical techniques for GDP analysis.

### *A. GDP Growth Rate*

The GDP growth rate, which shows the rate at which the economy is growing or decreasing, is one of the main metrics in GDP analysis. It is computed by quantifying the change as a percentage and comparing the GDP of two periods (often year over year or quarter over quarter). Positive growth rates point to an expanding economy, whereas negative growth rates point to a contracting one.

### *B. Sectoral Analysis*

Analysing GDP involves looking at the contributions made by the various economic sectors. To do this, the GDP is divided into many sectors, including manufacturing, services, construction, and agriculture. Policymakers and analysts can determine areas of strength or weakness and comprehend the causes of economic growth by evaluating the growth rates, relative sizes, and trends of various sectors.

### *C. International Trade Analysis*

Trends in international commerce are commonly examined in GDP analyses. This means calculating the GDP contribution that imports and exports make as well as looking at the trade balance, or the difference between imports and exports. A positive trade balance (exports exceeding imports) has a positive influence on GDP, whereas a negative trade balance (imports exceeding exports) has a negative impact on GDP.

### *D. Data Processing*

Following collection, the data is subjected to thorough processing and validation. To ensure correctness, this entails cleansing the data, examining it for errors or inconsistencies, and making the necessary corrections. When data is not readily available, data processing may also entail approximating missing numbers or estimating certain GDP components.

### *E. Methodological Framework*

Using the selected technique (production, income, or spending), workers calculate GDP according to accepted methodological standards and recommendations. These frameworks offer a set of uniform guidelines for data aggregation, sector and tackling measurement issues.

## IV. PROPOSED SOLUTION

### *A. Google Collabs*

Python running in the browser can be made simpler without requiring complicated configurations with Google Collab, a free Jupyter notebook. Introduced by Google in 2017, it offers smooth collaboration like Google Docs and comes with built-in Python and necessary libraries. It's a useful tool for data scientists working on ML and Deep Learning projects with cloud storage capabilities because it has integrated GPU support. Tensor Flow and Co-laboratory (formerly called Collab)—a development tool—showcase Google's dedication to AI research. Having offered free GPU support since 2017, Collab hopes to establish itself as a benchmark in academic ML and data science education, perhaps increasing the number of users of Google Cloud APIs. The construction of machine learning applications has been made easier since its launch, particularly for people who are already familiar with Jupyter notebook.

### *B. Jupyter Notebook*

A programming language is called Python. You may be able to speak with the computer using it. You'll need the assistance of a particular program or application to accomplish that. More precisely, Jupyter, the Jupyter Notebook App, could assist us in doing that. You are able to create and share documents with text, equations, live code, and visualizations using an open-source web tool called Jupyter Notebook. The Jupyter Notebook project is run by the Jupyter Notebook team and is a spin-off of the I Python project, which once had its own I Python Notebook project.

### *C. Python Software*

Python is freely used and disseminated, even for commercial reasons, thanks to the Python Software Foundation's open-source license, which has been authorized by OSI. Python is a widely used tool for data analysis, visualization, process automation, and software development. Because Python is so simple to learn, even non-programmers like biologists and accountants are starting to use it. The late 1980s saw Guido van Rossum design Python, which was published in 1991 and is well known for its extensive library and "batteries included" philosophy. Significant enhancements were made to Python 3.0 when it was launched in 2008, and Python 2 was phased out with the release of Python 2.7.18 in 2020.

### *D. Proposed Model*

*i). Load Libraries*

I am concentrating my investigation for my artificial intelligence (AI) computerized GDP analysis study on the "Countries of the World" dataset. Using the information from the 227 countries in the dataset, I will be concentrating on the variables that influence a nation's GDP per capita and attempting to develop a model. I'll also talk a little bit about the GDP.

*ii). Overview of the Table*

**Table 1. Overview of the Table**

| Index | Country | Region | Population | Area (sq.mi.) | Pop. Density (per sq.mi.) | Coastline (coast /area ratio) | Net Migration | Infant Mortality (per 1000 births) | GDP ($ per capita) | Literacy (%) | Phones |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Count** | 227 | 227 | 227.0 | 227.0 | 227.0 | 227.0 | 224.0 | 224.0 | 225.0 | 209.0 | |
| **Unique** | 227 | 11 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | |
| **Top** | Afghanistan | Sub-Saharan Africa | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | |
| **Freq** | 1 | 51 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | |
| **Mean** | NaN | NaN | 28740284.365638766 | 598226.95594713 65 | 379.04713656387 66 | 21.16533009647577 | 00381249999999985 | 35.506964285714 28 | 8230088495 58 | 3827751196 175 | 0614348775 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Std** | NaN | NaN | 117891326.54347652 | 20 | 185824610408 | 7228686315156448 | 4.888269211168732 | 35.38989877264037 | 10049.13851319723 | 19.7221729206261225 | 4.888269211168732 |
| **Min** | NaN | NaN | 70260 | 4647.5 | 00 | 700 | 0.9275 | 185824610408 | 4647.5 | 70260 | |
| **25%** | NaN | NaN | 4376240 | 86800.0 | 29.15 | 01 | 0.0 | 00 | 86800.0 | 4376240 | |
| **50%** | NaN | NaN | 47869940 | 441811.0 | 788 | 0.73 | 0.9975 | 29.15 | 441811.0 | 47869940 | |
| **75%** | NaN | NaN | 174977725 | 86800.0 | 190.15 | 1034499999999999999 | 190.15 | 210 | 86800.0 | 174977725 | |
| **Max** | NaN | NaN | 1313973713.0 | 17075200.0 | 162715 | 870.66 | 23.06 | 505.75 | 23.06 | 0.73 | |

*iii). Feature and Benefits of Python*

It has an English-like syntax and is compatible with a variety of operating systems, including Windows, Mac, Linux, Raspberry Pi, and others, it requires fewer lines of code than other programming languages.

*iv). Top Countries with Highest GDP Per Capita*

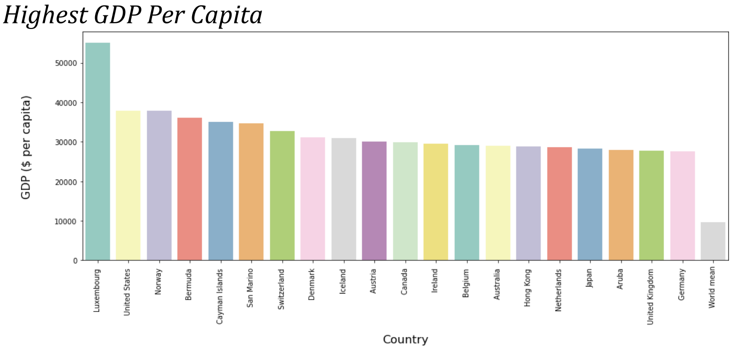


***Figure 1. Top Countries with Highest GDP Per Capita***

*v). Correlation Between Variables*

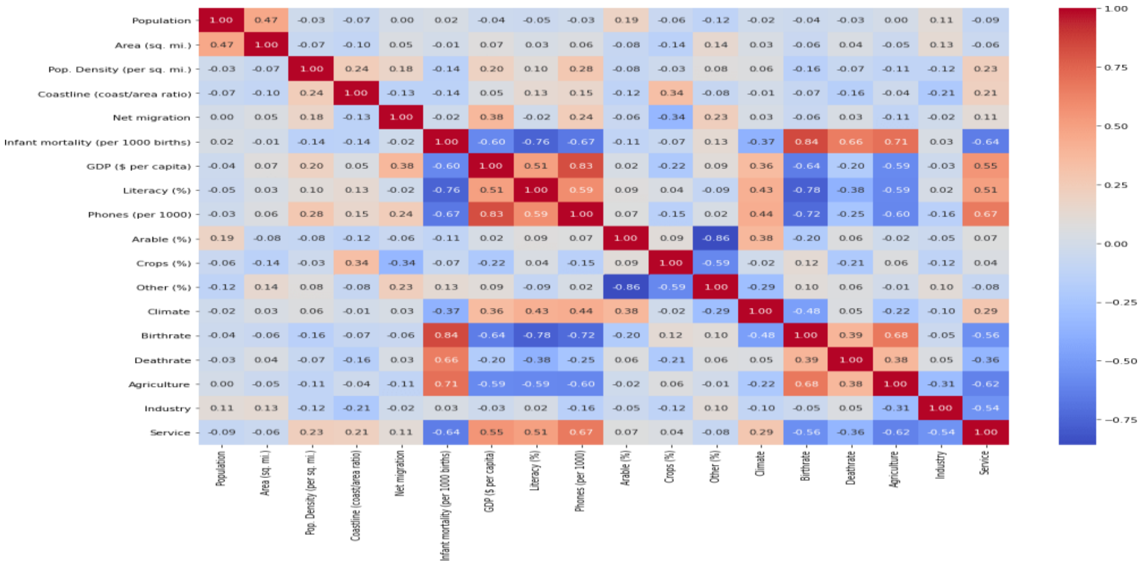


***Figure 2. Correlation between Variables***

## V. FACTORS AFFECTING TOTAL GDP

The relationship between the overall GDP and the other columns can also be examined. Following numerous characteristics that have also been determined to be largely connected with GDP per capita, the top two factors are area and population. A comparison of the top ten Let's compares the economic structures of the top ten nations based on their total GDP now.

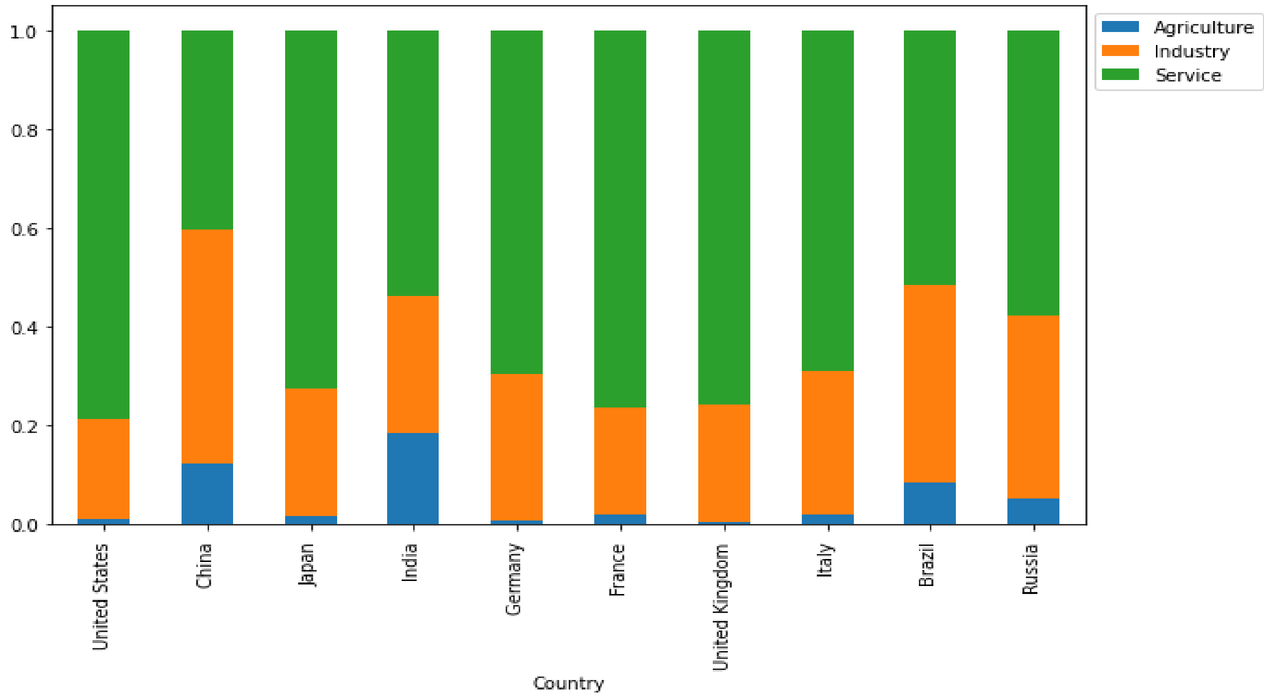


***Figure 3. Factors Affecting Total GDP***

### *A. Total Rank of GDP Per Capita*

**Table 2. Total Rank of GDP Per Capita**

| **Rank of total GDP** | **Rank of GDP per capita** |
|---|---|
| Germany | 15 |
| France | 15 |
| United Kingdom | 12 |
| Italy | 17 |
| Brazil | 84 |
| Russia | 75 |

### *B. Factors Affecting Total GDP*

**Table 3. Factors Affecting Total GDP**

| | |
|---|---|
| Population | 0.639528 |
| Area (sq. mi.) | 0.556396 |
| Phones (per 1000) | 0.233484 |
| Birthrate | 0.166889 |
| Agriculture | 0.139516 |

## VI. CONCLUSION

Our lives are infused with technology, and the nation's economy is no different. Better decision-making, predictive analysis, and pattern recognition are made possible by data science, which works with enormous volumes of data utilizing contemporary tools and methodologies. By applying data science to GDP analysis, we can determine the variables influencing GDP per capita in different nations. This aids in concentrating attention on the regions that support economic growth. This model can be used to save paperwork and make the process of calculating GDP less laborious. The likelihood of error is greatly reduced since a combination of many algorithms produces very high precision. Apart from events like worldwide pandemics, international wars, and economic crises, the model.

## VII. REFERENCES

1. Abas Omar Mohamed, "Modeling and Forecasting Somali Economic Growth Using ARIMA Models," *Forecasting*, vol. 4, no. 4, pp. 1-13, 2022. Google Scholar | Publisher Link
2. Chris Chatfield, and Haipeng Xing, *The Analysis of Time Series: An Introduction with R*, CRC Press, pp. 1-414, 2019. Google Scholar | Publisher Link
3. Girish Chandrashekar, and Ferat Sahin, "A survey on Feature Selection Methods," International Journal on Computers and Electrical Engineering, vol. 40, no. 1, pp. 16-28, 2014. Google Scholar | Publisher Link
4. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media, pp. 1-856, 2019. Google Scholar | Publisher Link
5. David J. Reid, "Combining Three Estimates of Gross Domestic Product," *Economics*, vol. 35, no. 140, pp. 431-444, 1968. Google Scholar | Publisher Link
6. Kumar Manoj, and Anand Madhu, "An Application of Time Series Arima Forecasting Model for Predicting Sugarcane Production In India," *Studies in Business and Economics*, pp. 1-14, 2014. Google Scholar | Publisher Link
7. Yin-Wong Cheung, and Kon S. Lai, "Lag Order and Critical Values of the Augmented Dickey– Fuller Test," *Journal of Business & Economic Statistics*, pp. 277-280, 1995. Google Scholar | Publisher Link
8. Polina Lemenkova, "Processing Oceanographic Data by Python Libraries NumPy, SciPy and Pandas," *Aquatic Research*, no. 2, vol. 2, pp. 73-91, 2019. Google Scholar | Publisher Link
9. Andy Liaw, and Matthew Wiener, "Classification and Regression by Random Forest," *R News*, pp. 18-22, 2002. Google Scholar | Publisher Link