Comprehensive Examination of Gross Domestic Product

Dr. D. Anusha[1], A.Nanditha[2], N. Manoj[3], S. Teja Shree Ram[4], V. Sravani[5]
[1]*Associate Professor, Department of CSE (Data Science), SRK Institute of Technology, Vijayawada, A.P., India*
[2,3,4,5]*Department of CSE (Data Science), SRK Institute of Technology, Vijayawada, A.P., India*

Abstract: Gross Domestic Product (GDP) is a measure of the value of goods and services produced in a country over a period of time. It's a broad measure of a country's economic health. A country's GDP depends on various factors, from micro industries like cafes and saloons to macro and major industries like IT, Banking and Tourism, all of them contribute in the growth of the economy of a country. Countries around the world collect data on consumption, investment, government spending and net exports to measure the GDP of that country. This makes GDP a universal measurement. If economy is healthy GDP growth expands, if economy is in bad shape GDP contracts. In the recent years 'recession' has become a trending topic, two consecutive quarters of negative GDP growth can lead to recession, it can largely affect the country's economy in all the sectors leading to issues such as Unemployment, Business Closures, Decreased sales and profits, Housing prices, etc. To overcome these issues governments design various ways, goals and set up policies for a better economic growth. This project's main aim is to identify the factors that are affecting the GDP by reviewing the previous economic movements and predicting how economic changes can amend patterns of previous trends. A more accurate prediction can help the country to build a strong economy. This research-based project also calculates the importance of features affected for the calculation of the GDP. It is very useful for the viewer to view the GDP growth of the different countries and also, they can see the best and worst predict performance in form of data visualizations such as graphs and plots by using python Matplotlib and Seaborn libraries. We used mathematical models and simple ML algorithms like Linear Regression, Random Forest Regression and Decision Tree Regression to predict future developments. To evaluate the model's performance, Mean Squared Error (MSE) and Mean Squared Logarithmic Error (MSLE) are used.

Keywords: Economy, Gross Domestic Product, Linear Regression, Decision Tree Regression.

1. INTRODUCTION

Gross Domestic Product (GDP) is an important gauge of the overall health of the economy, it is the total monetary or the market value of all the finished goods and services produced within a country in a specific period of time. It also includes the income earned from production, or the total amount spent on final goods and services. If we buy a cup of coffee from a random coffee shop and pay 5-10 rupees those 5 10 rupees are factored into the India's GDP. From very minute sectors to the major factories everything is a part of GDP. As for the reports United States is the world's largest economy, with a GDP of more than $20 trillion in 2024. GDP plays an important role in the development of the country, if the GDP of the country is high the economy of the country is said to be healthy, if the GDP is low the country's economy plummets and leads to lower living standards, less infrastructure, and industries. less advanced GDP was first introduced to the world in the late 1930's and in India GDP was introduced in the year 1992. In the first year to be estimated India's GDP was at 5.5% which was comparatively higher than other countries. A country's national statistical agency usually calculates GDP. It can be calculated in three ways, first way is by the Income Approach, adds up what everyone earned, this includes wages, rent, interest and corporate profits. Second is the Expenditure method, summation of what everyone spent in the year, which also includes private and government consumption expenditures and the Third approach is the Production Output approach, Measures based on the value of the goods and services produced. This includes all money spent on products and services, investments, infrastructure, and government spending. The most commonly used equation to measure GDP is: GDP=Consumption + Investment + Government Spending + NetExports

GDP = C + I + G + NX

Statisticians often claim that the formula is not whole and it only looks at production and manufacture missing out services and digital economy. GDP doesn't measure economic equality and well-being. Even if the country is rich economically, wealth maybe spread unevenly. In the project the Data for predicting factors influencing the growth of GDP is taken from Kaggle. The dataset consists of 227 samples of data with 20 different factors namely literacy, net migration, population etc. The main aim of the project is to consider all the factors that are actively affecting the GDP of different countries, visualize the results using python libraries such as Matplotlib and Seaborn. By the help of ML algorithms like Linear Regression, Random Forest Regression and Decision Tree Regression we are predicting the future developments of the economy, analyse all the factors and predict which factor contributes the most in the growth of the economy such that governments can actively plan policies which can benefit in upbringing those factors in future. Performance of the model is evaluated by the Mean Squared Error (MSE) and Mean Squared Logarithmic Error (MSLE). MSE and MSLE are the metrics used to measure the performance of regression models in machine learning. Both metrics are sensitive to outliers in the data. Each models performance is evaluated for a decent prediction performance. The main focus of traditional economics models is mainly on explanations of relationships whereas machine learning classifiers target predictions.

2. LITERATURE SURVEY

1. TITLE -"World GDP prediction using Machine Learning"

YEAR - 2024

AUTHOR - G. Dwarakanath, T. Shivakumar, Ms. Shraddha

This project endeavours to harness the power of machine learning algorithms to predict global GDP trends and calculate year-to-year growth rates. By leveraging advanced predictive analytics, the authors aimed to provide stakeholders with accurate forecasts of GDP performance across various countries. Such forecasts serve as invaluable tools for anticipating economic trends, enabling informed decision-making and strategic planning. In addition to predicting GDP trends, this project includes an analysis of the factors driving GDP fluctuations. By assessing the relative importance of various features in GDP calculation, the authors aimed to elucidate the underlying drivers of economic growth. This feature importance analysis offers valuable insights for policymakers and economists, guiding efforts to promote sustainable economic development and mitigate potential risks.

2. TITLE - "Predicting GDP using Autoregressive Models"

YEAR - 2017

AUTHOR - John Roush, Keith Siopes, Gongzhu Hu

National economies are multifaceted systems influenced by numerous interconnected factors, making accurate GDP forecasting inherently uncertain and short-term. Previous research efforts in GDP prediction have highlighted the difficulty of capturing the intricacies of economic dynamics effectively. While various modelling approaches have been explored, such Autoregression as (VAR), Vector their predictive capabilities often fall short of providing actionable insights. As such, there is a recognized need for more sophisticated models to improve the accuracy and reliability of GDP forecasts. In this study, the authors aimed to employ a VAR (4) model to analyse a limited set of economic indicators for GDP forecasting. However, the results obtained are deemed consistent yet lacking in substantive insights. Despite the efforts, the predictive power of the model remains limited, underscoring the complexity of economic systems and the challenges inherent in forecasting GDP accurately. As a result, we conclude that more advanced modelling techniques and a broader array of economic indicators are necessary to enhance the effectiveness of GDP forecasting efforts.

3. TITLE – "Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach"

YEAR - 2020

AUTHOR - Jaehyun Yoon This paper introduces a novel approach utilizing machine learning techniques, specifically gradient boosting and random forest models, to forecast real GDP growth. Focusing on Japan, the study spans from 2001 to 2018 and aims to produce accurate forecasts. Benchmark forecasts from the International Monetary Fund and Bank of Japan serve as points of

comparison. To enhance out-of sample prediction accuracy, the study employs cross- validation, a process aimed at selecting optimal hyperparameters. By fine- tuning model settings through cross-validation, the researchers aimed to improve the reliability and robustness of their forecasts. The accuracy of the forecasts is evaluated using two metrics: mean absolute percentage error (MAPE) and root squared mean error (RMSE). These metrics provide quantitative measures of the forecast performance, allowing for comparisons between different models and benchmark forecasts. The results of the study indicate that both the gradient boosting and random forest models outperform the benchmark forecasts for the 2001– 2018 period. Moreover, the gradient boosting model demonstrates superior accuracy compared to the random forest model. These findings highlight the efficacy of machine learning approaches in macroeconomic forecasting and underscore the potential for their increased utilization in future research and practice.

3. EXISTING SYSTEM

The existing system that was proposed in the year 2021 used simple ML algorithms like Linear Regression and Random Forest Regression for predicting the future developments of the country's GDP. The authors used world dataset from Kaggle. The dataset consists of 227 samples of data with 20 different factors namely literacy, net migration, population etc. The machine learning algorithm, namely, linear regression, is utilized to model and analyze the information provided for Gross Domestic Product prediction. Random Forest is one of the well-known machine learning algorithms that belongs to supervised learning technique. It is used both for regression and classification problems in machine learning. Random Forest relies on the concept of ensemble learning. The authors also evaluated the Performance of the model using evaluation metrics like Mean Absolute Error (MAE) and Root Mean Square regression used in the project is to predict the individual attribute of the dataset. Feature selection and Feature scaling was done on the linear regression model to attain a decent prediction performance, for Random Forest Regression Scaling was not tested since it should not affect the algorithm's performance. The authors used grid search in order to obtain good parameters for Random Forest. For this some of the parameters were optimized and the resultant accuracy of the optimized Random Forest Regression without feature selection and feature scaling is relatively higher than the performance of the Linear Regression.

DISADVANTAGES:

• Linear Regression might not capture Non-Linear Relations leading to under fitting.
• Sensitive to outliers.
• Random Forest Ensemble Learning can be less interpretable and cause over fitting.
• Random Forest, if not fine-tuned can suffer suboptimal perform.

4. PROPOSEDSYSTEM

Developed a predictive model to review the previous economic movements and predict how economic changes can amend patterns of previous trends. At each stage visualized the outcomes to provide more user viewable interface, such that user will be able to understand the action taken at each stage. Dataset – Kaggle, 227 samples, 20 factors Algorithms - Linear Regression, Random Forest Regression, Decision Tree Regression 8 Evaluation Metrics – RMSE, MSLE. The performance of each regression algorithm is evaluated to identify the best working algorithm. Every regression model is based on dependent and independent variables and resultant is considered as the final prediction. At different phases the given data set is trained, tested and the model is built or proposed and the final outcome of the project is expected to be a valuable insight for a country to plan a healthier economy by considering all the available factors.

ADVANTAGES:

• Most of the existing models failed to provide prediction of future trends.
• Here we are accurately predicting the factors influence on trends and suggesting the methods to improve the economy by improving the factors.
• Decision Trees are inherently more interpretable than linear regression or UGC CARE Group-1,
• Ensemble Learning, ensemble approaches can lead to more robust predictions.
• The proposed system is expected to achieve high prediction accuracy compared to the existing system.

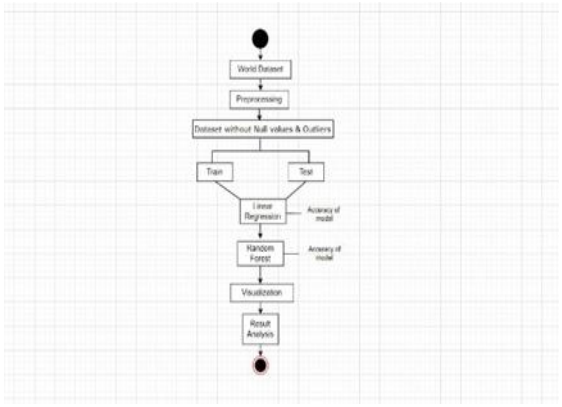

Fig 1 : Activity Diagram

5. RESULTS

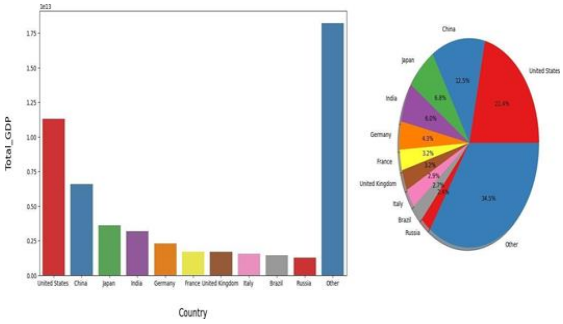

Fig 2: Top 10 countries with highest total GDP

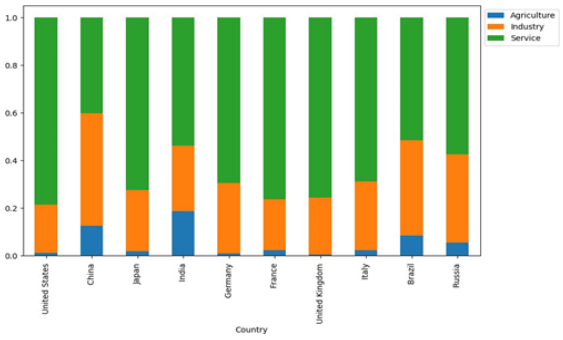

Fig 3: Comparison of the Economy structure for Top 10

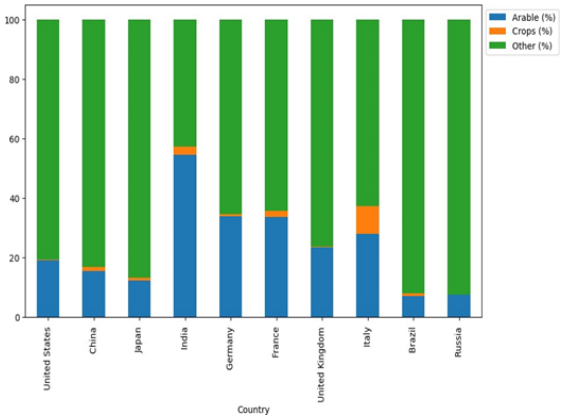

Fig 4 :Comparison of land usage of top 10 countries

CONCLUSION

In conclusion, this project has shed light on the multifaceted nature of Gross Domestic Product (GDP) and its pivotal role as a barometer of a country's economic health. By delving into the diverse factors influencing GDP, ranging from micro-level industries to major sectors such as IT, Banking, and Tourism, we've underscored the intricate web of contributor's growth. Through shaping our economic research-based approach, we've endeavoured to identify key factors influencing GDP and offer insights into how economic changes can potentially alter established patterns. By harnessing mathematical models and machine learning algorithms, such as Linear Regression, Random Forest Regression, and Decision Tree Regression, we've strived to provide accurate predictions of future economic developments. The utilization of performance evaluation metrics such as Mean Squared Error (MSE) and Mean Squared Logarithmic Error (MSLE) has enabled us to gauge the efficacy of our predictive models. Moreover, the incorporation of data visualizations, including graphs and plots generated using Python libraries like Matplotlib and Seaborn, has facilitated a comprehensive understanding of GDP growth trends across different countries. In essence, this project serves as a testament to the potential of interdisciplinary approaches in macroeconomic analysis. By leveraging advanced analytical tools and techniques, we aim to empower stakeholders with actionable insights for fostering robust economic growth and resilience in the face of uncertainty.

FUTURE SCOPE

• In this research-based project "Comprehensive Examination of Gross Domestic Product (GDP) "we are actively predicting the factors that can help a country improve their economy.

• A healthy economy can improve the lifestyle of the population. This project can provide a base for the Govt to understand the factors that are needed to develop in order to increase the GDP of their country. • In future, this model can be improved by using a better machine learning algorithm which may result in even better performance.

REFERENCE

[1] " Long Gang, "GDP Prediction by Support Vector Machine Trained with Genetic Algorithm," in

2nd International Conference on Signal Processing Systems (ICSPS), 2010.
[2] ” Martin Schneider, Martin Spitzer, "Forecasting Austrian GDP using the generalized dynamic factor model," 17 September 2004.
[3] ” Gary L. Shelley, Frederick H. Wallace, "Inflation, money, and real GDP in Mexico: a causality analysis," Applied Economics Letters, vol. 11, no. 4, p. 223–225, 2004.
[4] ” Anwar Ali Shah G.Syed, Faiz Muhammad Shaikh, "Effects of Macroeconomic Variables on Gross Domestic Product (GDP) in Pakistan," in International Conference on Applied Economics (ICOAE), 2013.
[5] Maciej Aleksandra Woźniak, Gąsior, Joanna Tomasz Duda, Bernat, "Relations of GDP growth and development of SMEs in Poland," in 23rd International Conference on Knowledge-Based and Intelligent Information Systems, 2019. & Engineering
[6] Carlos Encinas-Ferrer, Eddie Villegas-Zermeño, "Foreign direct investment and gross domestic product growth," in International Conference on Applied Economics, ICOAE, Kazan, Russia, 2-4 July 2015.
[7] Luca Coscieme, Lars F. Mortensen, Sharolyn Anderson, James Ward, Ian Donohue, Paul C. Sutton, "Going beyond Gross Domestic Product as an indicator to bring coherence to the Sustainable Development Goals," Journal of Cleaner Production, vol. 248, 2019.
[8] Gourav Kalbalia, Vivek Tambi, "Forecasting GDP: A Linear Regression Model," DU Journal of Undergraduate Research and Innovation, vol. 2, no. 2, pp. 41-46, 2016.