IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Prediction of Indian GDP using Multiple Linear Regression and Random Forest Regression

[1]Mrs. Jyoti Bagate, [2]Malavika Anoop, [3]Rudrakshi Deshpande, [4]Payal Sadhwani, [5]Om Bharwani,

[1]Assistant Professor, VESIT, [2-5]Student, VESIT,

[1-5]Department of Electronics and Telecommunication,

[1-5]Vivekanand Education Society's Institute of Technology, University of Mumbai, India

Abstract: Machine learning is still a relatively new field in economics. Furthermore, there is still a shortage of knowledge about machine learning. What machine learning is, how it differs from conventional econometrics, and how analysts and organizations may best use it. Our research aims to evaluate prediction models employing modern ML methods. We test if ML algorithms can enhance performance and identify characteristics that promote economic recovery or forecast economic recessions. Some indicators can provide insight into the status of the economy while others cannot. In this paper, we will study India's GDP and make predictions with the help of Machine Learning.

Keywords: Economics, GDP, Machine Learning, Forecasting, Linear Regression, Random Forest Regressor

I. INTRODUCTION

Machine learning can be used to evaluate a vast amount of data to make the best financial decisions. In a flourishing country like India, economic measures play an essential role in commercial and governmental sectors when making judgments. Economic forecasting helps business executives, officials, and investors make better decisions. As a result, making accurate projections about these measures is increasingly critical.

In its most fundamental form, econometrics is nothing more than economic statistics. Machine learning is used in economics to accomplish a similar goal, but with massive amounts of data. Keeping this in mind, we examine the conceptual framework and apply it to the forecasting of India's gross domestic product (GDP).

Gross Domestic Products (GDP): Gross domestic product (GDP) is a measure of the total worth of goods and services generated inside a country's borders over a year. Indicators such as GDP growth rate are critical to assessing a country's economic health. It's a crucial theory for policymakers and government planners. The Gross Domestic Product (GDP) can tell us if the economy is in a recession, depression, or boom. The GDP is a comprehensive indicator of the country's total economic output.

The formula is given as:

$$\mathbf{GDP = C + I + G + (X- M)}$$

Where,

- C = Annual Consumption (Personal Consumer Expenditure)
- I = Gross Private Domestic Investment
- G = Government Spending
- X = Total Amount of Exports
- M = Total Amount of Imports
- (X-M) = Total Net exports

India's economy is the world's sixth-biggest, with a GDP of $2.94 trillion (U.S.). With PPP (purchasing power parity), India's GDP ranks third in the world, with a value of $10.51 trillion. [1]

India's economy is made up of agriculture, handicrafts, industrial plants, and a multitude of other stuff. Even though two-thirds of Indians work in agriculture, the service sector is the primary driver of economic development today. Due to a large number of educated individuals who are fluent in English, India has emerged as a global leader in information technology.

There were severe government regulations in several areas of India's independent history, including banking, foreign direct investment (FDI), and telecommunications. Since the early 1990s, India has gradually opened its markets to international commerce and investment by diminishing government control. It was around this time that India's economy took off.

II. LITERATURE REVIEW

The development of the country's gross domestic product (GDP) is used as a gauge of economic stability. Economic growth may be predicted in a variety of ways, as evidenced by several research projects.

The paper, "Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach" by Jaehyun Yoon [2], presents a method for creating machine learning models, specifically a gradient boosting model and a random forest model, to forecast real GDP growth. This study focuses on the real GDP growth of Japan and produces forecasts for the years from 2001 to 2018. The forecasts by the International Monetary Fund and Bank of Japan are used as benchmarks. To improve out-of-sample prediction, the cross-validation process, which is designed to choose the optimal hyperparameters, is used. The accuracy of the forecast is measured by mean absolute percentage error and root squared mean error. Random Forest uses bootstrapped data, the regression trees trained independently, and the output of trees is averaged to produce predictions.

The paper, Gross Domestic Product Prediction using Machine Learning" by Vaishnavi Padmawar, Pradnya Pawar, Akshit Karande [3], shows that the effect of varying inflation on real gross domestic product growth is negative. It indicates that the volatility of the rate of inflation can induce a negative impact on the real GDP. The study encourages the use of machine learning classifiers such as Random Forest and linear regression in macroeconomic data forecasting. The main thrust of traditional economics models is usually on explaining relationships and their effects on a dependent variable. However, machine learning models can also produce high-quality predictions.The methodology used was Linear Regression Model followed by Random Forest to achieve the best possible accuracy.

The paper, "Predicting economic growth using machine learning" by James Thomas Bang, Atin Basuchoudhary, Tinni Sen [4], shows that while more traditional models will still be used to predict the effects of various variables on the margin, Machine Learning models can provide better predictions than those derived from traditional models. It is found that these models do a better job than some traditional models for the purpose of predicting as they solve the problem of missing data more constructively. Also, studies of growth through traditional models can be considered severely biased because traditional models drop any observation that has a missing data point for even one of the variables specified in the model.

The findings of our research promote the use of machine learning classifiers, namely Linear Regression and Random Forest regression, in the forecasting of macroeconomic indicators.

III. METHODOLOGY

Our proposed work is divided into two parts:

1. Establishing a model to reliably forecast economic measures using historical data from the World Bank [5] and the Government of India website [6] as well as Kaggle datasets [7].

2. Streamlining the analytical process by combining all forecasts into a single webpage and implementing the most accurate one.

The steps we took to implement this are as follows:

1. Data Collection

The first step is to determine what sort of data is needed to address a specific problem and whether or not there are any privacy problems associated with the data. It is the initial stage in the process of developing a machine learning model. To conduct our research, we used World Bank [5] open source data.

2. Data Preparation

Data preprocessing is the process of extracting raw data, which is data that has been acquired in the actual world and has been turned into a clean data collection before it is used. A series of procedures were carried out to reduce the data to a manageable amount. Null values have been dropped and filled up throughout our work.

3. Data Visualization

In data visualization, we can observe how the data appears and what type of association exists between its many characteristics. It's the quickest approach to verify if the attributes match the output. We've utilized Python packages like Matplotlib and Seaborn to show data in an appealing way.

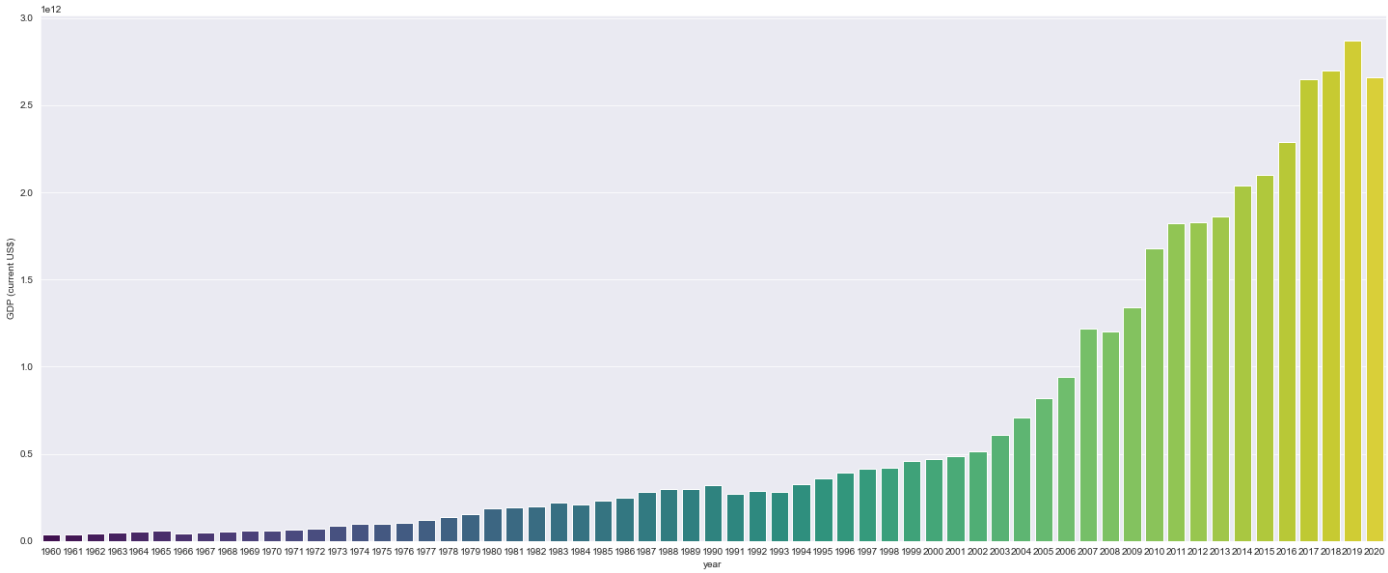

Fig 3.1: Visualization of GDP per Capita from 1960-2020

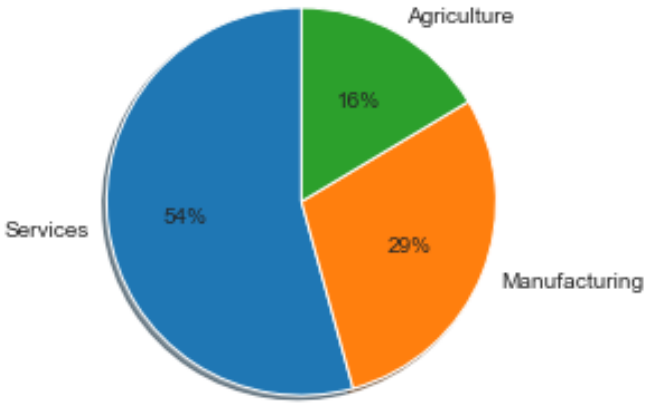

Fig 3.2: Sector-wise contribution

4. Data Analysis

For GDP prediction, we have made use of Linear Regression algorithm and Random Forest Regression algorithm.

A. Linear Regression

Linear Regression is a supervised learning-based machine learning technique. Variable correlation and forecasting are two of its most common applications. There are a variety of different regression models, each of which has a different number of independent variables and a different correlation between the dependent and independent variables.

B. Random Forest Regression

A random forest is a meta estimator that employs aggregating to increase predicted accuracy and mitigate over-fitting by fitting a number of classification decision trees to various subsets of the dataset.

5. Model Training

When developing algorithms, model training is the stage of the data science development lifecycle in which researchers attempt to find the optimal mix of weights and biases for the methodology to limit a loss function throughout the predicted range. There are 70% training and 30% testing datasets in our research.

6. Model Evaluation

Model evaluation is the method of assessing a machine learning model's progress and identify any flaws. Our model compares predicted data to actual data. Among the metrics we utilized for our assessments were R-squared and Root Mean Squared Error on Prediction (RMSE/RMSEP).

IV. RESULTS

This study presents a method for predicting the annual GDP of India from 1960 to 2020. The prediction includes features that encompass correlation with GDP such as services, manufacturing and agriculture having a significant contribution. We performed regression using Multiple Linear Regression with manufacturing, value added; services, value added; industry (including construction), value added; population, total as independent variables while GDP (in current USD) as dependent variable. The table below shows the difference between actual and predicted value and figure 3.2 shows the plot of true values vs predicted values using Multiple Linear Regression.

	Actual Value	Predicted	Difference
26	2.490000e+11	2.500229e+11	-1.022875e+09
35	3.600000e+11	3.560548e+11	3.945150e+09
60	2.660000e+12	2.158813e+12	5.011869e+11
28	2.970000e+11	3.004441e+11	-3.444148e+09
11	6.735099e+10	6.685568e+10	4.953077e+08

Fig 4.1: Difference in Actual vs Predicted Values

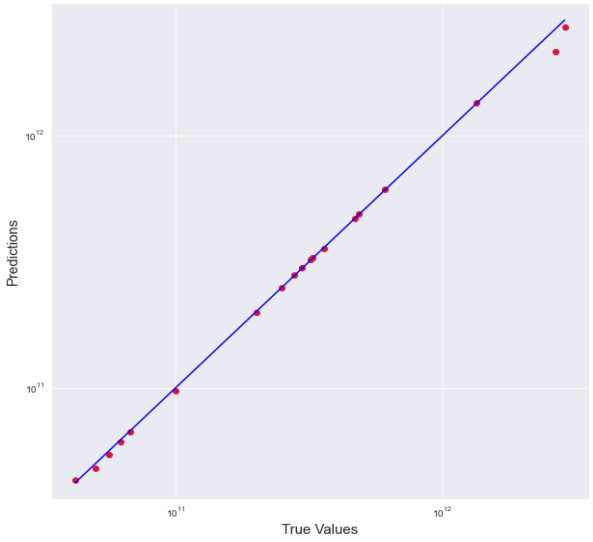

Fig 4.2: Plot of True Values vs Predicted Values using Multiple Linear Regression

We furthermore performed Random Forest Regression and the table below shows the comparison of the two algorithms.

	Multiple Linear Regression	Random Forest Regressor Algorithm
R Squared Value	0.9769236583724383	0.9441972244945368
RMSE	90.00550551537863	283.9706006869975

Table 4.1: Comparison Table of the ML Algorithms

V. CONCLUSION

In this study, we provide a comparison of two machine learning algorithms, Multiple Linear Regression algorithm and Random Forest Regressor for the prediction of GDP of India. The results of our analysis show that the Multiple Linear Regression algorithm outperformed the Random Forest Regressor algorithm with a R-square value of 0.97 as opposed to that of 0.94 by the latter algorithm. In a developing country like India, evaluating economic metrics such as GDP plays an important part in taking decisions, both in private sectors as well as public sectors. In this study, we have used machine learning algorithms to predict one of the key economic metric GDP (in current USD) with the data obtained from the World Bank.

VI. REFERENCES

[1] Economy of India, Wikipedia - https://en.wikipedia.org/wiki/Economy_of_India#cite_note-53
[2] Yoon, J. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. Comput Econ 57, 247–265 (2021)
[3] Padmawar, Vaishnavi & Pawar, Pradnya & Karande, Akshit. (2021). Gross Domestic Product Prediction using Machine Learning. 08. 2395-0056.
[4] Bang, James & Sen, Tinni & Basuchoudhary, Atin. (2015). New Tools for Predicting Economic Growth Using Machine Learning: A Guide for Theory and Policy.
[5] World Bank Dataset - https://data.worldbank.org/country/IN
[6] Government of India Dataset - https://data.gov.in/resources/indian-economy-selected-indicators-2010-11-2016-17
[7] Kaggle Dataset - https://www.kaggle.com/paree24/india-gdp-growth-world-bank-1961-to-2017