IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204252 |

Indian GDP Prediction using Different Machine Learning Algorithms

G. Rekha [1] , K. Thapaswi [2] , G. Naresh Reddy [3] , K. Sai Kiran Reddy [4] , A. Manasa [5]

Assistant Professor, Dept. of CSE., JNTUA University, Kuppam Engineering College, Chittoor, India[1]

UG Students, Dept. of CSE., JNTUA University, Kuppam Engineering College, Chittoor, India[2-5]

ABSTRACT: GDP is significant for measuring the strength of national economies. This study aims to predict the economy and growth analysis of a country. It will be helpful for the prediction of Gross Domestic Product (GDP) by overcoming the existing system. By using Advanced machine learning algorithms such as the Gradient boosting algorithm, linear regression, and Random Forest algorithm we can predict the most accurate values. GDP will depend on various factors such as Agriculture, manufacturing, exports, imports, and Land, etc. So we want to collect a huge amount of data from various sources for Accurate results. The existing system using Arima models makes it difficult to capture non-linear relationships. GDP will be predicted annually there is always a change in GDP based on the above factors. It will be also helpful for different businessmen, policymakers, and investors in making informed decisions and sustainable economic growth. The results are produced 91% accurate compared to 87% of existing output.

KEYWORDS: GDP, Random Forest algorithm, Gradient Boosting Algorithm, Linear Regression

I. INTRODUCTION

The indication of the country's overall economic performance is India's Gross Domestic Product (GDP). It is the final value of overall goods and services produced in a country, including for import and export. The GDP is denoted as the most critical indicator of a country's economic health and significantly affects the lives of citizens. Predicting how the GDP will function in the future is a difficult but crucial undertaking. Accurate prediction of output will be helpful for firms and governments in making decisions. Investors can use the knowledge to decide which investments to make, while governments can use it to create economic policies that support stability in the economy. This paper predicts the Indian GDPusing machine learning algorithms, specifically linear regression, gradient boosting, and Random Forest Algorithms. The data used in the analysis were collected from various sources, and the results were produced accurately. The results of this research can assist policymakers and businesses in making informed decisions based on accurate predictions of the Indian economy.

II. PROPOSED APPROACH

The objective of this study was to use machine learning algorithms to predict the Indian Gross Domestic Product (GDP) based on information that is readily available in publicly available datasets. To be representative of the practical use of such predictions in real-world scenarios, the study excluded information that is not easily obtained or would require costly or impractical methods of assessment. All the data analysis and modeling were performed using Google Collab andthe datasets used in the study were obtained from Kaggle.

THE DATASET

The present study makes use of Indian GDP data to predict future trends. The data used in this research was gained from Kaggle, which contains the records of theIndian GDP from the years 1960 to 2020. The limited data cannot be sufficient to accurately train Advanced machine learning algorithms, so it requires large amounts of data for training. Then it will produce accurate results with this limited data set, we decided to use linear regression, Gradient Boosting, and Random Forest Algorithms.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204252 |

Data Set

Agricultural Output (billion INR)	Industrial Output (billion INR)	Services Sector Output (billion INR)	Exports (billion INR)	Imports (billion INR)	Inflation Rate (%)	Unemployment Rate (%)	Population (millions)	Foreign Direct Investment (billion INR)	Government Expenditure (billion INR)	Education Spending (% of GDP)	Healthcare Spending (% of GDP)	Interest Rates (%)	Debt-to-GDP Ratio (%)	Consumer Confidence Index	Business Confidence Index
8823.04	8043.72	12736.02	6623.39	6598.71	4.41	7.44	1254.63	2411.2	11006.62	3.31	1.38	4.83	72.83	86.0	93.34
7800.12	7873.63	14212.44	4994.3	6156.05	4.13	7.09	1309.11	2749.79	11828.05	3.2	1.36	5.79	72.79	93.28	92.74
8234.05	9543.83	12325.31	5563.07	6555.04	5.28	6.0	1547.58	3598.15	11255.68	3.46	1.53	4.65	70.34	74.81	94.54
9180.67	8589.54	13363.74	5424.42	6885.65	4.59	6.66	1593.03	2800.0	11650.65	3.34	1.4	4.36	76.03	98.62	91.22
8900.67	9423.23	9914.0	6251.04	4872.45	4.42	7.45	1412.79	3292.89	9876.74	2.4	1.54	5.16	61.63	98.43	95.88
6767.04	8831.33	12848.56	5356.1	5530.48	4.01	7.88	1299.84	2759.09	11161.5	3.24	1.49	4.93	78.78	91.38	89.1
8212.57	8767.75	12036.46	4792.57	5158.68	4.94	7.34	1309.89	3301.05	12094.25	3.46	1.72	5.46	55.73	101.18	82.23
7386.48	9043.2	12918.0	4589.94	5126.83	5.14	6.43	1398.22	3131.53	10873.12	3.39	1.48	4.89	69.82	84.42	83.04
7422.59	10168.41	13041.2	5520.48	5308.6	4.39	6.34	988.41	3357.97	8575.73	2.92	1.02	5.38	60.24	79.61	81.99
7807.95	10382.69	9292.5	6198.67	6057.93	4.22	6.33	1399.27	2919.86	10596.9	3.15	1.73	5.72	82.24	96.3	90.27

III. REGRESSION

Regression is a technique for modeling the relationship between one or more independent variables and a dependent variable. It is commonly used to predict future values based on past data. In particular, forecasting future values of a variable based on its historical values is a common application of regression. The accuracy of these predictions can be helpful in decision-making such as economics, policymakers, and business. Determining the relationship between the independent variables and a dependent variable that best fits the observed data is the basis of the technique. In this dataset, there are multiple independent variables such as Inflation rates, imports, exports, etc. There are two types of regression techniques linear regression and polynomial regression. But here we are using only one type of regression for accurate prediction.

A. Linear Regression

Linear regression looks for a linear relationship between the GDP and other factors, like inflation, and population growth, exports, and imports. It assumes that there is a linear relationship between the independent variables and the dependent variable. This represents the relationship with a linear equation, where the coefficients of the independent variables map their relationship with the dependent variable and the intercept value maps the point on the Y-axis where the regression line crosses. Thismethod is mostly used for simple regression for which one independent variable is used to predict the numerical value of the one dependent variable. However, in the case of multiple regression, consists of multiple independent variables can beused to predict the dependent variable.

The formula of linear regression is

$$Y = a + bX \quad (1)$$

where Y is the dependent variable, a is the intercept and b is the coefficient for the independent variable X

B. Gradient boosting Algorithm

The main purpose of this algorithm is to produce accurate predictions. It combines multiple weak models called decision trees to create a strong predictive model. It keeps improving itself by learning from past experiences. By utilizing Gradient boosting we can take advantage of its ability to handle complex relationships and capture non-linear patterns in the data. This can help us make more accurate predictions about the Indian GDP and gain insight into the factors that influence its growth. It's worth mentioning that the accuracy of the predictions depends on the Quality of the data used for training and the features considered. Additionally, it is important to regularly update the model with new data to ensure its predictions stay relevant.

Formula for Gradient boosting

At each iteration n, the model prediction y^n(a) is updated using the gradient of the loss function L for the previous predictions

$$Y^n(a)=y^n-1(a)+r.hn(a) \quad (2)$$

Y^n-1(x) is a prediction of the ensemble at iteration n-1

r is the learning rate controlling the step size of the update

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204252 |

hn(a) is the output of a new weak learner trained to correct the errors

C. Random forest algorithm

Random Forest is a supervised learning algorithm that is used for both classification and regression. This technique can be used to predict the GDP of India by creating multiple decision trees, each looking at different factors such as inflation rate, population, and government spending. These decision trees make individual predictions, and then their prediction is combined to get a more accurate estimate of GDP. Based on the majority of decisions the output is predicted. Hence it will help in improving the performance of the model.

IV. RESULTS

A. Linear Regression

The linear regression model, which is widely used in machine learning, that the connection between the independent and dependent variables is linear. It attempted to predict the Indian GDP values using linear regression. However, the resultsshowed that the GDP values and the years were not in a linear relationship, making linear regression an approach for the prediction. The predictions generated by the algorithm had low accuracy, as the algorithm failed to capture the complex non-linear relationship between the years and GDP values. The difference between theactual and predicted values was observed to be different. To obtain more accurate predictions, it needs to use a more advanced algorithm that can capture the non-linear relationship between the variables. In summary, while linear regression is a useful tool for predicting dependent variables based on independentvariables; it may not be suitable for all datasets. For non-linear relationships, it is necessary to use more complex algorithms, such as gradient boosting, and random forest to obtain more accurate predictions.

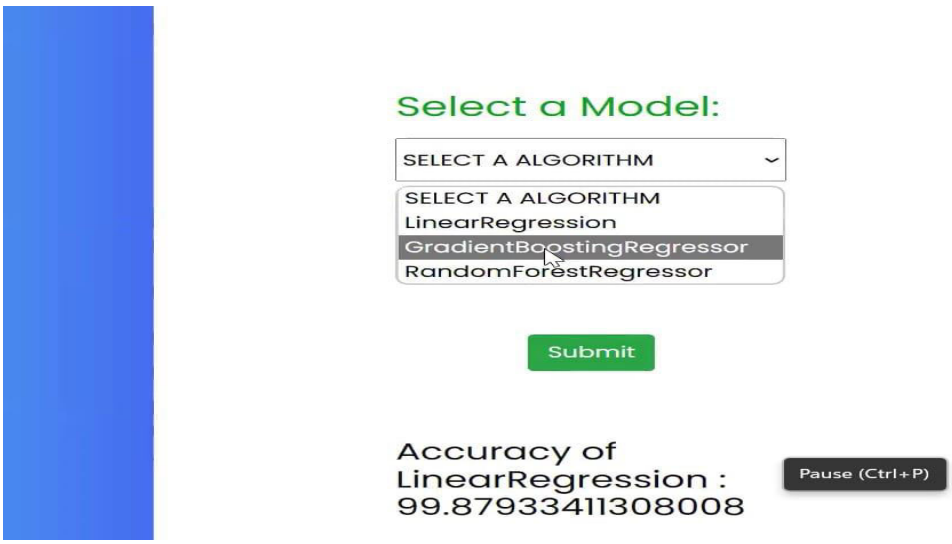


Fig. 1. This image shows the predicted value of Indian gross domestic value based on the past data .It is more accuracy compared to the existing model.

B. Gradient Boosting Algorithm

Gradient Boosting algorithm is a method of standing out for its prediction speed and accuracy, particularly with large complex datasets. From Kaggle competitions to machine learning solutions for business, this algorithm has produced the best result in the Indian dataset. The Gradient Boosting algorithm is used to combines weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean square or cross entropy of the previous model using gradient descent. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall problem prediction error. This algorithm builds an additive model in a forward stage-wise function; it allow for the optimization of arbitrary differentiate loss function and then main aim of gradient boosting algorithm is used to shown an accuracy and performance are unmatched for tabular Supervised learning tasks.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204252 |

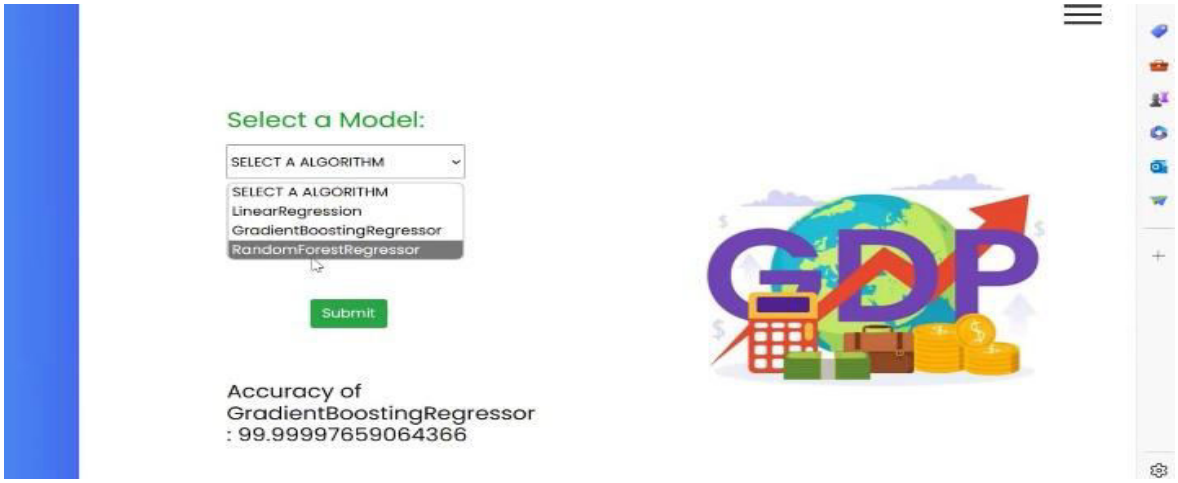
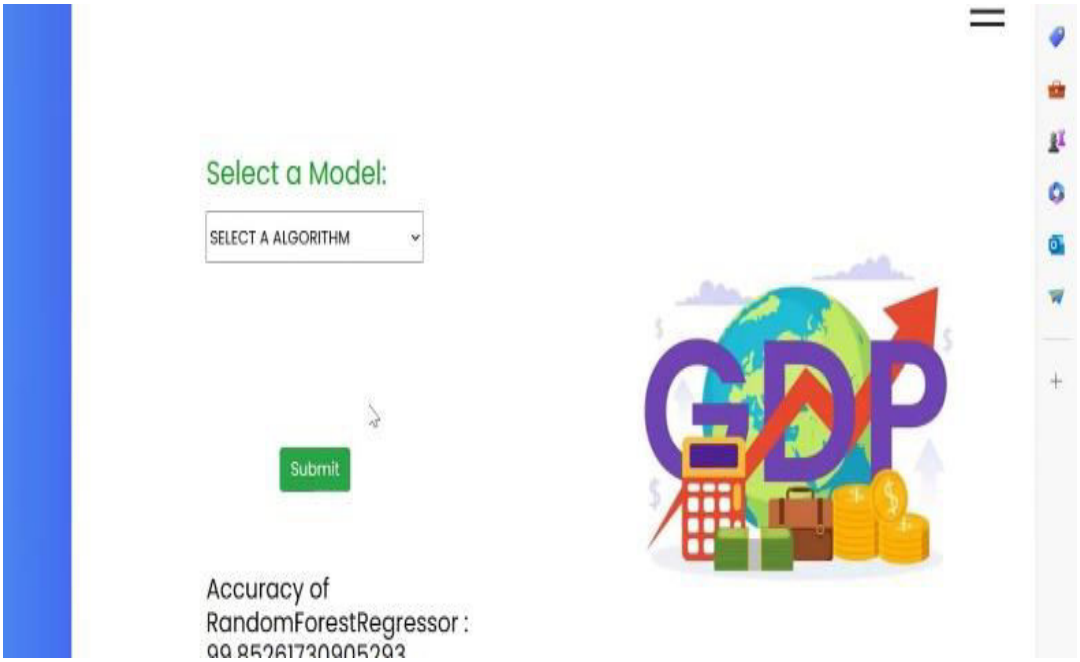


Fig. 2. The image shows a comparison between the actual values of the Indian Gross Domestic Product (GDP) over time and the predicted values using Gradient boosting algorithm.

C. Random Forest Algorithm

The random forest algorithm its works by creating multiple decisions trees, where each tree is trained on a different subset of the data. Each tree independently predicts the GDP based on different features lite inflation, interest rates, government spending and more. when its time to make a prediction, all the trees in the forest vote on the outcome & the final prediction is determined by the majority vote. This approach helps to reduce overfitting & improve the accuracy of the predictions. When a model is outfit, it performs really well on the training data but doesn't generalize well to new unseen data. Random forest algorithm combines multiple decisions tree to reach a single result. The random forest tree is belonging to a machine learning which does ensemble classification. The term ensemble implies a method which does ensemble implies a method which makes predictions by averaging over the predictions of several independent base models. The main aim of random forest algorithm to works by creating many decisions trees in one training session.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204252 |

V. DISCUSSION

In this study, we attempted to predict the Indian GDP values for the years 1960-2020 using regression analysis. However, after applying linear regression to the dataset, it was evident that the relationship between GDP values andthe years was not linear. Therefore, it was futile to use linear regression to predict the GDP values, and the algorithm had negligible accuracy.

To obtain better results, we turned to Gradient Boosting, which is a more flexible algorithm. By using this method, we could model the relationship between the years and GDP values more accurately. The result showed that the Gradient boosting much better suited to the dataset, as it could capture the nonlinear relationship between the variables. By selecting the optimal degree for the polynomial, we were able to improve the accuracy of the model and obtain a better prediction of the GDP values.

Other model is ARIMA not used due to the limited amount of data available. These models require a large amount of data for training and may not be suitable for smaller datasets. The use of regression algorithms allowed us to make accurate predictions, but we recognize that these models have their own limitations. Despite this, we believe that the resultsof this study can be useful for policymakers and economists to make informed decisions about the Indian economy. Even with the Random Forest algorithm, we were unable to makeaccurate predictions. Hence, accurately predicting the inflation rate continues to be a significant challenge, and we suggest additional research to enhance the quality of the data and increase the quantity of available data to enhance the precision of the forecasts. Despite these constraints, our study offers useful perspectives on the correlation between GDP values and the year.

VI. CONCLUSION

The prediction of the Indian Gross Domestic Product (GDP) is crucial for understanding and interpreting the economy's functioning, and machine learning algorithms have shown immense promise in this regard. Our investigation aimed to predict Indian GDP using machine learning algorithms and compare the performance of linear ,Gradient Boosting and Random Forest algorithm. Our results demonstrated that the Gradient Boosting offered more accurate predictions than linear regression, indicating that the non- linear relationship between GDP data and years was the most likely reason for the unsatisfactory performance of linear regression. Furthermore, we attempted to predict inflation rates using the sametechniques but found that even Random Forest could not provide accurate predictions due to irregularities in the data.

Therefore, the findings highlight the need for high- quality and more extensive datasets for Precision Nutrition research. Our study underscores the potential of machine learning algorithms in predicting economic indicators and provides insights for policymakers and businesses to make informed decisions based on accurate predictions of the Indian GDP. The application of machine learning algorithms can offer useful information to decision-makers in the realm of economics, leading to successful policies and decision-making. These insights may further enable the government to implement policies that can lead to economic growth and stability. We recommend the use of time-series algorithms in futureresearch, which require more data for training but could improve the results further. As the quality of the dataimproves, it may be possible to develop even more accurate models, leading to better policy-making and improved economic growth.

In conclusion, our study highlights the significance of employing machine learning algorithms in predicting Indian GDP. The findings of this study could be beneficial for policymakers, economists, and other stakeholders involved in economic decision-making, leading to better policy-making and improved economic growth in the longrun.

REFERENCES

1. "Jovic, s., Miladinovic, j., Micic, R., &Rakic, G.(2019).Analyzing exchanging rate and GDP
2. "Katrina, W. (2021).A Guide to the types of machine learning algorithms and their applications.
3. "The Linear Regression Model (LRM)," Linear Regression: A Mathematical Introduction, pp. 1–21, 2019
4. "Educational data mining and learning analytics "Ryan baker 2014
5. "School Students Dropout Prediction by using Linear Regression and Logistics Regression," Adalya Journal, vol. 9, no. 10, Oct. 2020
6. "Amith K Shukla, Sandeep Kumar," Interval type2 Fuzzy Weighted Exterme learning machine for GDP Prediction", IEEE-2018

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details