International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 01 | Jan 2024 www.irjet.net p-ISSN: 2395-0072

Forecasting of India's GDP using Various Regression Algorithms

Mitali Sarnaik[1], Renuka Dad[2], Priyanka Shahane[3]

[1]Student, Department of Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, India
[2]Student, Department of Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, India
[3]Assistant Professor, Department of Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, India

---

Abstract - *The only global metric used to evaluate the status of a nation's economy is its gross domestic product (GDP), which is a single figure that expresses the monetary value of all completed products and services produced inside its borders in a specific time period. With India's economy developing at one of the quickest rates in the world, forecasting the country's GDP with precision is crucial for scholars, investors, and politicians alike. By enabling policymakers to make well-informed decisions about fiscal and monetary policies, accurate GDP estimates have the ability to mitigate economic downturns and foster sustainable growth. This study collected, preprocessed, and employed historical economic indicators from various industries, such as agriculture, industry, and services, as features for training and assessing machine learning models. Many techniques, such as Support Vector Machine, Auto Regression, and Linear Regression, were used to forecast India's GDP. fared better than the others in terms of expected accuracy and generalization to new data, according to the trial findings. With any luck, we'll be able to use methods like Multiple Linear Regression, Decision Trees, Random Forests, and Polynomial Regression with more precision. The study's findings will show how machine learning techniques may be used to anticipate India's GDP with a fair degree of accuracy. The results highlight the importance of feature engineering and selection and the need to consider external and internal economic factors. The knowledge gained from this study can help economists and politicians create plans that will promote stability and economic progress.*

Key Words: Machine Learning, Gross Domestic Product (GDP), Mean Absolute Error, Root Mean Squared Error, Mean Squared Error, Score, Random Forest, Decision Tree, Polynomial Regression, Multiple Linear Regression.

1.INTRODUCTION

In a time where data-driven technologies are always evolving, there is going to be a change in the field of economic research and forecasting. The capacity to predict economic indicators with extraordinary accuracy becomes imperative as countries endeavor to manage the intricacies of a fast evolving global environment. The Gross Domestic Product (GDP) is a crucial indicator of a country's economic health that impacts investments, policies, and the general well-being of society. GDP projection has historically been based on econometric models that take historical data and certain economic variables into account. But the complex interactions between an increasing number of variables, along with the nonlinear dynamics of contemporary economies, have exposed shortcomings in traditional approaches. Within this framework, a novel approach to use Machine Learning (ML) appears to be a viable path towards improving GDP prediction accuracy and resilience. This study is a groundbreaking exploration of the merging of modern machine learning techniques with economic understanding, primarily focusing on India, a major global economy with a diverse range of industries, populations, and governmental actions. Our research not only emphasizes the necessity of a modernized method for estimating GDP, but it also provides guidelines for the use of ML models in the field of economics. Within this framework, a novel approach to use Machine Learning (ML) appears to be a viable path towards improving GDP prediction accuracy and resilience. This study is a groundbreaking exploration of the merging of modern machine learning techniques with economic understanding, primarily focusing on India, a major global economy with a diverse range of industries, populations, and governmental actions. Our research not only emphasizes the necessity of a modernized method for estimating GDP, but it also provides guidelines for the use of ML models in the field of economics. Additionally, our research clarifies the interpretability of these models by providing insights into the underlying causes of economic trends—a point that traditional approaches frequently miss. This paper is a call to action for economists, data scientists, and policymakers to create a new alliance between technology and economics, as well as a testament to the potential of machine learning to reshape conventional paradigms. Through this synergy, we set off on a path of transformation that will enable nations with the vision to navigate uncertainty and make wise decisions in a global economy that is becoming more and more complex. Come along as we explore the future of machine learning-based economic forecasting and establish a standard for innovation in GDP forecasting

2. LITERATURE SURVEY

2.1. Literature Survey Summary

Rudrani Bhattacharya et. al. [1] pioneered a study. Their research incorporated advanced methodologies such as Factor Augmented Time Varying Coefficient Regression and Autoregressive Integrated Moving Average. Data was sourced from the Central Statistical Organisation, Ministry of Statistics and Programme Implementation, yielding an impressive Root Mean Square Error (RMSE) of 0.35.

Mrs. Jyoti Bagate et. al. conducted a research study. Their investigation involved the use of Multiple Linear Regression and Random Forest techniques to forecast the Indian Gross Domestic Product (GDP). They utilized data from reputable sources such as the World Bank and the Government of India, achieving an impressive accuracy rate of 94%. Their work showcased the effectiveness of these methods in accurately predicting economic indicators.

Tanvi Gharte et. al. research, offered a comprehensive analysis. Their study involved Gradient Boosting, Random Forest, and Linear Regression techniques, with data sourced from Kaggle. They achieved varying accuracy rates: 82% with linear regression, 87% with random forest, and an impressive 89% with Gradient Boosting. This research underscored the importance of meticulous technique selection and data analysis in economic prediction models.

Dwarakanath G V et. al. conducted a study. Their research involved Support Vector Machine and Linear Regression techniques, utilizing data from data.gov.in. Despite the absence of specific accuracy, their research emphasized the utilization of machine learning for global economic predictions.

Giovanni Maccarone et. al. explored GDP Forecasting using Machine Learning. Their study employed K-Nearest Neighbors and Linear Regression techniques, utilizing the Global Data Hub. Remarkably, their approach resulted in a Mean Absolute Error of 1.73e-03, indicating a highly precise prediction model.

Saurabh Ghosh et. al. delved into GDP Nowcasting in their study. Their research incorporated Prophet forecasting and artificial neural network methodologies, utilizing RBI's Database for Indian Economy. They achieved a notable accuracy rate of 90%, showcasing the effectiveness of their approach in capturing the nuances of real-time economic data.

Vaishnavi Padmawar et. al. study focused on Random Forest and Linear Regression techniques. Their research, based on data from Kaggle, achieved a Root Mean Square Error of 3097.194, highlighting the accuracy and reliability of their GDP prediction model.

Parag Kar utilized Python to predict Indian Nominal GDP in USD, employing Gradient Descent and Linear Regression techniques. The results were shared on the platform Medium, showcasing a high accuracy score of 0.9403951277253917. Kar's work demonstrated the power of Python in economic forecasting, providing valuable insights into India's economic landscape.

Navyasri S et. al. delved into predicting the GDP of India utilizing machine learning. While the specific accuracy was not mentioned, the authors experimented with K-means clustering and Linear Regression techniques, employing data from data.gov.in. Their research laid the foundation for further exploration into economic forecasting using advanced algorithms.

Ronan Flannery study utilized Artificial Neural Networks and Random Forests, analyzing raw data for their predictions. The model achieved an RMSE of 964.11, highlighting Flannery's expertise in machine learning techniques for economic forecasting.

2.2. Previous Algorithms Applied

Time Series Analysis: Time series models have been widely used for GDP forecasting. Examples include Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing. To increase these models' capacity for prediction, researchers have integrated machine learning techniques.

Regression Models: To incorporate a variety of economic indicators, including industrial production, inflation rates, and fiscal policies, as features in GDP prediction models, multiple linear and linear regression models have been used.

Neural Networks: To better predict economic data and identify complex patterns, deep learning methods such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs) have been used.

2.3. Challenges

Data Quality and Availability: One of the main issues is India's economic data quality and availability, which can be erratic or lacking, especially at the regional levels.

Model Complexity: When working with sparse historical data, complex machine learning models are susceptible to overfitting, which calls for rigorous regularization and validation techniques.

Non-Stationarity: Conventional time series models are challenged by the non-stationarity of economic time series data, necessitating further preprocessing or sophisticated machine learning methods.

Interconnected Variables: It can be difficult to pinpoint the most important characteristics and develop an accurate

model of them due to the complex relationships that exist between economic variables.

3. METHODOLOGY

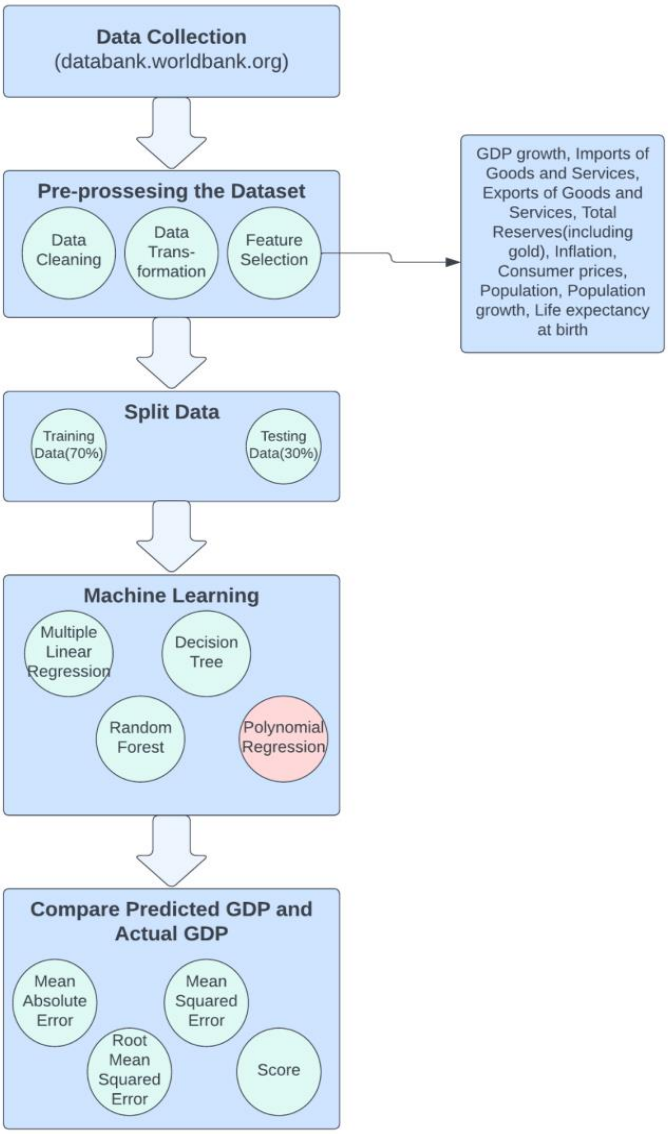

Fig -1: Architecture Diagram

3.1. Data Source

Originally from the World Bank's database, we downloaded the dataset from Kaggle.

3.2. Features

The dataset includes a number of features pertaining to population statistics and economic indicators. GDP growth, imports and exports of goods and services, total reserves (including gold), consumer prices, inflation, population growth, and life expectancy at birth are some of these characteristics.

3.3. Data Cleaning

Using the pandas library, we eliminated rows and columns that contained null or empty values in order to perform data cleaning. This is necessary to guarantee that there is no missing data in the dataset, which could affect the analysis's quality.

3.4. Data Transformation

We used mean values in place of empty values. One common method for handling missing values and making sure the dataset is still suitable for analysis is to impute the mean for the missing data.

3.5. Train-Test Split

To divide the dataset into training and testing subsets, we employed the train_test_split function from the sklearn_model_selection module. For this purpose, a typical split of 70:30 was selected.

3.6. Regression Models

To forecast India's GDP, we used a variety of regression techniques. Among the regression techniques are:

1. A linear regression model with more than one predictor variable is known as multiple linear regression.

$$y = \beta 0 + \beta 1x + \varepsilon$$

2. Polynomial Regression: This kind of regression uses an nth-degree polynomial to represent the relationship between the independent variable or variables and the dependent variable.

$$y = \beta 0 + \beta 1x + \beta 2x2 + \beta 3x3 + \ldots\ldots\ldots + \beta nxn + \varepsilon$$

The polynomial terms' coefficients are 0, 1, 2,..., n, and y is the polynomial's dependent variable.

3. Decision Tree Regression: This type of regression models the relationship between variables by utilizing decision trees.

4. Random Forest Regression: To increase predictive accuracy, this ensemble learning technique combines several decision trees.

3.7. Evaluation Metrics

We used a range of evaluation metrics to compare the actual and predicted GDP values, such as:

The average absolute difference between the expected and actual values is measured by the Mean Absolute Error (MAE).

The average squared difference between the expected and actual values is measured by the Mean Squared Error (MSE).

Score (Accuracy): This metric compares the actual and predicted values to determine how accurate the regression models were.

Root Mean Squared Error (RMSE): The average magnitude of error is measured by taking the square root of the mean squared error (MSE).

4. IMPLEMENTATION RESULTS

Table -1: Result Table

Model Name	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error	Score
Multiple Linear Regression	1476156 0293.969 036	3.88472 7146040 704e+20	1970971 1175.05 4554	99.949%
Polynomial Regression	2997889 304.2289 51	8.63753 4807592 975e+19	9293833 873.914 991	99.986%
Decision Tree	7328759 4628.210 53	1.49648 1719924 0363e+2 2	1021350 43942.4 8788	98.641%
Random Forest	1495037 89728.05 554	4.39355 3451419 084e+22	2096080 49736.1 4645	92.700%

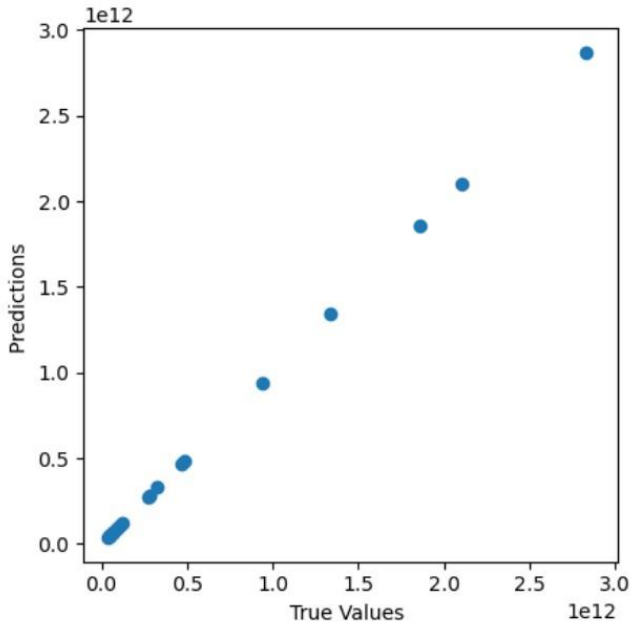

Chart -1: Polynomial Regression(Best Accuracy)

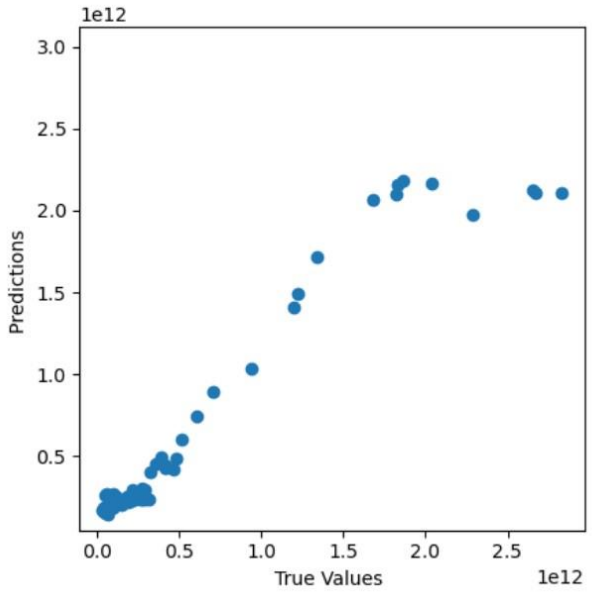

Chart -2: Random Forest(Worst Accuracy)

3. ADVANTAGES

• Economic Planning: Precise GDP forecasts assist policymakers and government officials in making plans for infrastructure development, resource distribution, and fiscal policies, among other areas of the economy. It provides information for choices about public spending, budgeting, and taxes.

• Investment Decisions: To make well-informed choices about expansion and investment in the Indian market, businesses and investors rely on GDP forecasts. Both domestic and foreign direct investment (FDI) can be stimulated by a robust GDP growth forecast.

• Employment and Labour Market: Job creation and unemployment rates are closely correlated with the GDP growth rate. Precise forecasts aid job searchers and labor market analysts in anticipating employment patterns and arranging their professional paths accordingly.

• Financial Markets: GDP projections are used by investors and financial institutions to evaluate market risks and make investment decisions. Foreign exchange rates, interest rates, and stock markets can all be impacted by predictions.

• Risk management: To evaluate and reduce economic risks, both individuals and businesses use GDP projections. Businesses might modify their supply chain strategies in response to anticipated changes in GDP growth or recession, for instance.

3. CONCLUSIONS

This work tackled the difficult task of forecasting India's GDP by utilizing a wide range of regression approaches,

including multiple linear regression, polynomial regression, random forest, and decision tree algorithms. We observed that polynomial regression had not been used to predict India's GDP in earlier research. So, after much testing and careful analysis, we found that Polynomial Regression was the best option, predicting India's GDP with 99% accuracy. This outcome shows how reliable and strong the approach is at capturing the complex relationships between the numerous factors affecting the nation's economic growth.

However, it's critical to recognise the challenges this endeavor will inevitably encounter. The quality of data, the interpretability of complex models, and the selection of hyper-parameters remain challenges in both research and practice. Future research should examine hybrid models, which integrate the benefits of multiple approaches to provide a more thorough understanding of the factors influencing GDP.

REFERENCES

[1] Bhattacharya, R., Bhandari, B. & Mundle, S. Nowcasting India's Quarterly GDP Growth: AFactor-Augmented Time-Varying Coefficient Regression Model (FA-TVCRM). J. Quant. Econ. 21,213–234 (2023). doi:10.1007/s40953-022-00335-6

[2] N. Srinivasan, K. M, N. V, K. S M, S. Kumar and S. R, "Predicting Indian GDP with Machine Learning: A Comparison of Regression Models," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp.1855-1858, doi: 10.1109/ICACCS57279.2023.10113035.

[3] N. Malik and B. Agarwal, "Time Series Nowcasting of India's GDP with Machine Learning," 2022 International Conference on Artificial Intelligence of Things (ICAIoT), Istanbul, Turkey, 2022, pp. 1-6, doi: 10.1109/ICAIoT57170.2022.10121883.

[4] Rahmani, A.M., Hosseini Mirmahaleh, S.Y. An Intelligent Algorithm to Predict GDP Rate and Find a Relationship Between COVID-19 Outbreak and Economic Downturn. Comput Econ (2022). doi:10.1007/s10614-022-10332-9

[5] S.C. Agu, F.U. Onu, U.K. Ezemagu, D. Oden,Predicting gross domestic product to macroeconomic indicators,Intelligent Systems with Applications,Volume 14,2022,200082,ISSN 2667-3053,doi:10.1016/j.iswa.2022.200082.