Predictive Analysis Of Indian GDP Using Machine Learning Algorithms

C Siva Kumar[1], P Lakshmi Sagar[2] , Samala Pavan Kumar[3*] , Shaik Mohammad Abrar[4] , Renati Venkata Sai Susanth[5] , Sangaraju Sai Yashwanth Varma[6]

[1]Assistant Professor, Department of DS, Mohan Babu University(Erstwhile Sree Vidyanikethan Engineering College), Tirupati, A.P., India
[2]Assisatnt Professor, Department of CSE, SV College Of Engineering, Tirupati, A.P., India
[3,4,5,6] UG Scholar, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, A.P., India

[1]Sivakumar.c@vidyanikethan.edu,[2]lakshmisagar.p@svec.edu.in
[3*]samalpavan999@gmail.com,[4]smohammadabrar18@gmail.com,
[5]Susanthrenati5@gmail.com, [6]naniyashwanth793@gmail.com

Abstract. In this research endeavor, machine learning algorithms—specifically Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor—are harnessed to anticipate the future trajectory of the Indian Gross Domestic Product. Employing an extensive dataset that incorporates historical GDP, per capita income, imports, exports, and GDP growth rate, the study seeks to evaluate the predictive precision of each model. Post data preprocessing and model training, assessment metrics will be employed to juxtapose the efficacy of these models. The research yields insightful perspectives into the adeptness of these algorithms in predicting Indian GDP, providing policymakers and economists with valuable information to make well-informed decisions. The identification of the most accurate predictive model and critical economic indicators is paramount in this context.

Keywords: Indian GDP Prediction, Economic Indicators,GradientBoosting ,Regressor, Linear Regression, Random Forest Regressor,Performance Evaluation

1 INTRODUCTION

Understanding and predicting India's Gross Domestic Product is paramount for navigating the complexities of its economic landscape. GDP, a vital indicator of economic features, reflects the total products produced within a specific time frame. In our project, we employ advanced Machine Learning algorithms to analyze historical GDP data comprehensively. This endeavor goes beyond traditional analysis, aiming to uncover intricate patterns and trends that can offer deeper insights

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_56

into India's economic dynamics. By harnessing the power of predictive analysis, we seek to enhance our ability to anticipate changes, providing valuable foresight for policymakers, businesses, and researchers alike.Our initiative emphasizes the proactive use of predictive modeling to strengthen India's economic forecasting capabilities. Machine Learning algorithms play a pivotal role in this endeavor, allowing us to delve into historical GDP data and extract meaningful patterns. Through meticulous analysis, we aim to uncover hidden correlations and dynamics that contribute to a more accurate understanding of India's economic trends. This predictive insight is invaluable for stakeholders in making timely and informed decisions, ensuring adaptability in response to evolving economic scenarios. The project's overarching goal is to provide a robust and proactive tool for anticipating shifts in India's economic trajectory, fostering a deeper comprehension of the factors influencing economic growth.By integrating predictive analysis into the study of India's GDP, our project underscores the importance of staying ahead of economic trends. The utilization of advanced Machine Learning algorithms enables us to not only analyze historical data but also to extrapolate meaningful patterns that can aid in forecasting future economic growth. In addition to fostering a proactive understanding of India's economic landscape, our project emphasizes the dynamic nature of predictive modeling in adapting to emerging challenges. The incorporation of Machine Learning algorithms enables us to not only identify historical trends but also to respond dynamically to evolving economic scenarios. By continuously refining our models based on real-time data, we aim to create a robust forecasting framework that can capture the nuanced shifts in India's GDP trajectory.

2 LITERATURE REVIEW

In their inquiry, Adam Richardson [1] delves into the precision of predicting New Zealand's real GDP growth through the common ML algorithms on real-time datasets. The resultant forecasts are then compared with the predictive accuracy of a basic auto regressive benchmark and other data-intensive methodologies. The results unveiled that the majority of machine learning models surpassed both auto regressive and traditional statistical approaches in terms of forecasting reliability.As a consequence, the findings advocate for the adoption of Machine Learning algorithms as a compelling alternative to GDP now casting models, enhancing the predictive toolkit available to forecasters. [2] Pirasant Premraj, in his research , conducts a thorough comparative analysis for Gross Domestic Product forecasting across various economies. Among the machine learning algorithms utilized. In contrast, traditional time series regression approaches, including Auto regressive models, Auto regressive Integrated Moving Average models, and Vector Auto regressive models, are also incorporated. The findings underscore the effectiveness of multivariate models, highlighting their superior performance in predicting GDP growth based on selected variables. This emphasizes the model's suitability for forecasting economic trends in the specified economies. [3]Liliana and Togar Alam Napitupulu, researchers from Universitas Surabaya, employ Artificial Neural Networks (ANN) in their study focusing on Gross Domestic Product (GDP) forecasting in Indonesia. Acknowledging the pivotal role of accurate GDP projections for business planning and government fiscal policies, the study compares the efficacy of ANN with traditional government methodologies. The findings reveal that ANN, with its ability to handle non-linear and non-parametric data, outperforms government-issued GDP forecasts, the study suggests the broader applicability of ANN in predicting various economic indicators, urging further exploration and refinement of forecasting methodologies in the Indonesian economic landscape. In the exploration of Gross Domestic Product (GDP) prediction[4] for Indonesia, Siti Saadah and Muhammad Satrio Wibowo delve into the use of deep learning algorithms, specifically LSTM and RNN. Acknowledging GDP as a crucial indicator for assessing a country's financial stability, the researchers emphasize the significance of understanding economic conditions, particularly in the context of sudden fluctuations,such as those experienced during the COVID-19 pandemic. Leveraging deep learning methods, the study demonstrates that LSTM and RNN, with specific architectural configurations, exhibit high prediction accuracy.

3 METHODOLOGY

Data Collection :Collected comprehensive data containing relevant economic indicators and historical GDP data for India from kaggle website.The data contains economic indicators like GDP growth, GDP Per capita, Imports of goods, Exports of goods, inflation and consumer prices, Population total, Population Growth.

Data Preprocessing:Cleansed the collected data by addressing missing values outliers, and inconsistencies.Standardize or normalize numerical features and encode categorical variables appropriately. Ensure the dataset is in a suitable format for analysis.

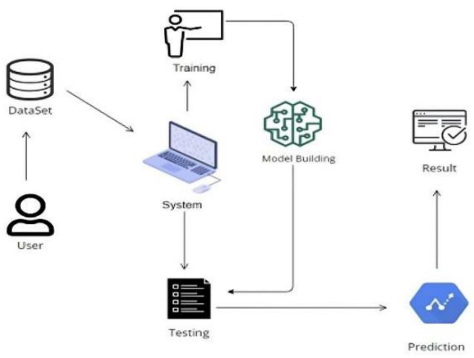

Fig1:Architecture

I. Feature Selection

Conducted a thorough analysis to identify and select pertinent features for GDP prediction. Considered economic indicators which includes that have historically demonstrated a significant impact on GDP variations. Use statistical methods or domain expertise to guide the selection process.

IV. Model Selection

We have choosen Random Forest Regressor, Gradient Boosting Regressor and Linear Regression algoritms to build models and we pick the model that gives best performance. Model Training & Testing:In order to accurately evaluate the model's performance, the data was split into two sets: one for testing and the other for training.In order to help the models understand patterns, we used 80% of the data for training. The remaining 20% of the data was set aside for testing, which assessed the models' prediction power.

Model Description

Dividing the data into sets for testing and training.Creating predictions on the testing data using each trained multiple regression model that has been developed.

Random Forest Regressor : Multiple decision trees are used by the Random Forest Regressor, a potent ensemble learning method designed for regression applications, to improve prediction accuracy and stability. This approach, which comes from bagging, builds a lot of decision trees during training and averages their predictions to get the result. A random portion of the training data is used to form each tree in the Random Forest framework, and a random subset of features is taken into consideration for splitting at each node. This injection of randomness helps reduce model variance, mitigating the risk of overfitting to the training data when compared to a single decision tree. Significantly, Random Forest excels in managing large datasets with higher dimensionality and adeptly handles missing values. Moreover, it provides valuable insights into feature importance, identifying variables that play a substantial role in the prediction process. Owing to its adaptability and robustness, Random Forest Regression finds extensive applications across diverse domains, effectively addressing complex regression problems.

Gradient Boosting Regressor: Gradient Boosting Regression stands as an advanced machine learning technique that extends the boosting concept, where multiple weak predictive models collaborate to establish a robust predictor. In this approach, predictors are sequentially incorporated into an ensemble, with each one rectifying the errors of its predecessor. A pivotal feature of Gradient Boosting is its utilization of a gradient descent algorithm to minimize errors within a predictive model. In the progression, each new model concentrates on precisely predicting the residuals or errors of the preceding models. Gradient Boosting Regression constructs a series of decision trees in a step-wise manner. Following the addition of each tree, the algorithm adjusts instance weights depends on outliers of previous tree, emphasizing more challenging-to-predict instances. This iterative refinement persists until a specified number of trees are generated or no further enhancements can be achieved. This technique is renowned for its efficacy in handling diverse data types, including non-linear relationships, and its flexibility in model tuning.

Linear Regression: The relationship between a dependent variable and one or more independent variables can be modeled using linear regression, a basic statistical and machine learning tool. By using observed data to create a linear equation, this is accomplished. When an analysis is done in its most basic version, known as simple linear regression, it only takes into account one independent variable.

Formulae: $\hat{Y} = \boldsymbol{b_0} + \boldsymbol{b_1 x_1} + \boldsymbol{b_2 x_2} + \ldots\ldots + \boldsymbol{b_n x_n}$. eq(1)

$\hat{y}$ represents the predicted value, b_0 is the y-intercept, and $b_1, b_2, \ldots, b_n$ are the coefficients corresponding to the independent variables $X_1, X_2, \ldots, X_n$, respectively.

Model Evaluation

Assessed the performance of each model using the testing data. Utilize evaluation metrics such as MAPE, and R-squared (R^2) to quantify the accuracy of GDP predictions. Compare the results across the three models.

$$\text{MAPE} = 1/n \sum_{i-1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \times \mathbf{100} \quad \ldots \quad \text{eq(2)}$$

Where, A_i is the variable's actual value that is being predicted for observation
F_i is the estimated value for the first observation
n total observations .

$$\mathbf{R^2} = 1 - \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \quad \ldots \quad \textbf{eq(3)}$$

Where ,$\mathbf{y_i}$ is the dependent variable's actual value for the ith observation.

$\hat{\mathbf{y}}_{\mathbf{i}}$ is the dependent variable's expected value for the ith observation.
$\bar{\mathbf{y}}$ is the dependent variable's values' mean.

n total number of observations

4 RESULTS

Three algorithms were utilised in this effort to create forecast models for Indian GDP utilising data from the Indian economy. The MAPE and R square score of the models varied. Gradient Boosting Regressor performed well, having the highest R square score and lowest MAPE, according to metrics measuring MAPE and R square score

.

Table1: Comparision Table

Algorithms	**MAPE**	**R square Score**
Linear Regression	2.1635167675172	0.6138532968201
Random Forest Regressor	0.0352450881336	0.9988494186115
Gradient Boosting Regressor	0.0124051247933	0.9999862153992

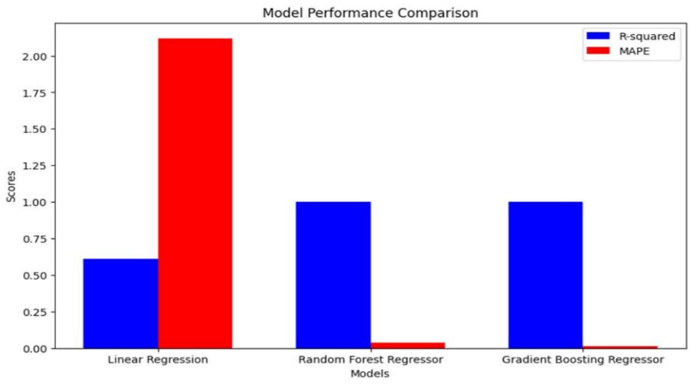

Fig2: R-squared and MAPE performance comparision

5 CONCLUSION

In this groundbreaking study, we introduce an innovative approach to predicting the Indian GDP through a comparative analysis of three distinct regression models : Random Forest , Gradient Boosting and Linear Regression. Our inventive methodology seeks to improve the accuracy of GDP predictions, offering a fresh perspective on economic forecasting and contributing to the evolving landscape of machine learning applications in economic analysis. In conclusion, our project reveals that, when compared, the three algorithms exhibit varying performance. Notably, Linear Regression displayed low performance in comparison to the remaining models, Gradient Boosting Regressor performed well with highest R square score and least MAPE. This insight underscores the potential for advanced algorithms to outperform traditional linear approaches in economic forecasting, enhancing our understanding of predictive modeling in economic analysis.

6 FUTURE SCOPE

Following the completion of the first machine learning-based prediction analysis of the Indian GDP, the models' performance can be improved by adjusting their hyperparameters. Furthermore, investigating other ensemble techniques or machine learning algorithms might yield insightful information and raise the prediction accuracy. The prediction potential of the model could also be improved by adding additional features or economic indicators, such as political stability or conditions in the global economy, that may affect GDP. Additionally, carrying out a more thorough examination of feature importance using methods like SHAP values or permutation importance could provide insightful information about the variables influencing GDP and improve the prediction model.

REFERENCES

[1]S. G, S. K. K, J. S. A. and H. V. V, "Forecasting GDP using ARIMA and Artificial Neural Networks Models under Indian Environment", International Journal of Mathematics Trends and Technology, vol. 56, no. 1, pp. 60-70, 2018

[2] M. Jahn, "Artificial neural network regression models: Predicting GDP growth," Hamburg Institute of International Economics

[3] S. Sa'adah and M. S. Wibowo, "Prediction of Gross Domestic Product (GDP) in Indonesia Using Deep Learning Algorithm," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),

[4]V. Bhardwaj, P. Bhavsar and D. Patnaik, "Forecasting GDP per capita of OECD countries using machine learning and deep learning models," 2022 Interdisciplinary Research in Technology and Management (IRTM), Kolkata, India, 2022,

[5] I. -S. Wong, K. -H. Ng and K. -C. Khor, "The Influence of Data Preprocessing in Predicting Malaysia's GDP Growth Rate Using Multilayer Perceptron," 2021 International Conference on Decision Aid Sciences and Application (DASA),

[6]C. N. Wang and V. T. Phan, "Enhancing the Accurate of Grey Prediction for GDP Growth Rate in Vietnam," 2014 International Symposium on Computer, Consumer and Control, Taichung, Taiwan, 2014,

[7]U. Salma, M. M. Hasan and M. I. Sultana, "A Regression Model-Based Approach to Identifying Determining Factors for GDP Growth in Bangladesh," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020,

[8]L. Shao and W. Zhou, "The forecasting for GDP development of Liaoning province based on Grey System Theory," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China, 2013, pp. 300-303

[9]L. Na and L. Xinpeng, "Forecast and scenario simulation analysis of power demand considering a variety of external environmental factors," 2016 China International Conference on Electricity Distribution (CICED), Xi'an, China, 2016, pp. 1-4

[10] Avanija, J., G. Sunitha, and K. Reddy Madhavi. "Semantic Similarity based Web Document Clustering Using Hybrid Swarm Intelligence and FuzzyC-Means." Helix 7, no. 5 (2017): 2007-2012.

[11] Madhavi, K. Reddy, A. Vinaya Babu, A. Anand Rao, and S. V. N. Raju. "Identification of optimal cluster centroid of multi-variable functions for clustering concept-drift categorical data." In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, pp. 124-128. 2012.

[12] Q. Gang, Cui Shengnan and T. Jiafu, "Time series forecasting of medicare fund expenditures based on historical data—Taking Dalian as an example," The 26th Chinese Control and Decision Conference (2014 CCDC), Changsha, China, 2014,

[13] S. Kumari and S. K. Singh, "Deep Learning-based Time Series Models for GDP and ICT Growth Prediction in India," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 250-256,

[14] X. Liu, D. Xiong, Z. Latif and Y. Lu, "A Comprehensive Research of Models Between Internet Development and China's Economic Growth," 2020 International Signal Processing, Communications and Engineering Management Conference (ISPCEM), Montreal, QC, Canada, 2020, pp. 58-62,

[15]V. I. Antipov, F. F. Pashchenko, Y. I. Kudinov and E. S. Duvanov, "Forecast of Economic Development in Russia 2022-2035," 2022 2nd International Conference on Technology Enhanced Learning in Higher Education (TELE), Lipetsk, Russian Federation, 2022, pp. 220-224,