IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 04, May 2024
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Predicting Indian GDP with Machine Learning: A Comparison of Regression Models

Kesavan M.V
Assistant Professor, Department of CSE
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
kesavan.mv@eastpoint.ac.in
Debasinha
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
debasinha627@gmail.com
Benisha W
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
benisha092002@gmail.com
Ananya GL
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
aglp6585@gmail.com
Ankit Rana R
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.10/Issue04/MYIS10085
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.10/Issue03/MYIS10085
Received: 07, April 2024|Revised:16, April 2024 | Accepted: 27, April 2024 Published Online: 07, May 2024 Volume 2024 Article ID MYIS10085 http://www.ijiris.com/volumes/Vol10/iss-04/06.MYIS10085.pdf
Article Citation: Kesavan,Debasinha,Benishaw,Ananyagi,Ankitranar(2024). Predicting Indian GDP with Machine Learning: A Comparison of Regression Models. International Journal of Innovative Research in Information Security, Volume 10, Issue 04, Pages 150-155 **doi:>** https://doi.org/10.26562/ijiris.2024.v1004.06
BibTex key: Kesavan@2024Predicting

Copyright: ©2024 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: In recent years, the application of machine learning algorithms in economic forecasting has gained significant traction. This study focuses on predicting the Gross Domestic Product (GDP) of India by harnessing advanced machine learning techniques. We assembled a comprehensive dataset incorporating time series analysis and inflation rates from diverse sources. Our investigation involved a comparative analysis employing both linear and polynomial regression methods to discern the most accurate predictive model. Our findings underscore the superiority of the polynomial regression model, which excelled in capturing non-linear relationships between independent variables and GDP. Specifically, the polynomial regression model achieved an impressive prediction accuracy rate of surpassing the accuracy achieved by the linear regression model. This study underscores the critical role of advanced machine learning algorithms in economic forecasting. By emphasizing the significance of high-quality datasets and the application of techniques like polynomial regression, we illustrate how these factors can substantially enhance the precision of economic forecasts. These insights hold substantial implications for policy makers and businesses alike, aiding informed decision-making and effective economic policy development. This study serves as a valuable reference for future research, highlighting the importance of advanced machine learning methods and robust data sources in economic forecasting.
Keywords: Random Forest Regressor, Gradient Boosting Regressor and Logistic Regression;

I. INTRODUCTION

In recent years, machine learning algorithms have gained popularity in the field of economic forecasting. This study aims to predict the Indian Gross Domestic Product (GDP) using advanced machine learning algorithms. To achieve this, we collected data from various sources, including time series analysis and inflation rate. We analyzed the data using linear regression and polynomial regression techniques to determine which method produced the most accurate results. Our results showed that the polynomial regression model outperformed the linear regression model in terms of accuracy.

The polynomial regression model was better able to capture the non-linear relationships between the independent variables and the dependent variable (GDP). Specifically, our findings showed that the polynomial regression model was able to predict the Indian GDP with an accuracy of 91%, compared to 87% for the linear regression model. This study highlights the importance of using advanced machine learning algorithms in economic forecasting. We found that the use of high-quality data sets and advanced techniques such as polynomial regression can significantly improve the accuracy of economic forecasts. Our findings have several implications for policy makers and businesses. Accurate predictions of economic indicators such as GDP can help businesses make informed decisions about investment and growth strategies, while policymakers can use these predictions to develop effective economic policies. Overall, our study provides valuable insights into the use of machine learning algorithms in predicting Indian GDP. Our findings demonstrate the effectiveness of polynomial regression in capturing non-linear relationships and improving the accuracy of economic forecasts. This study can be used as a reference for future research in this area and emphasizes the need for high-quality data sets and advanced machine learning techniques in economic forecasting, A key indicator of a nation's overall economic performance is India's Gross Domestic Product (GDP). It is the total value of all commodities and services produced in a nation, including for domestic consumption and export. The GDP is regarded as the most critical indicator of a nation's economic health and significantly affects the daily lives of its citizens. Predicting how the Indian GDP will function in the future is a difficult but crucial undertaking. The economy and the welfare of its population can be affected by accurate GDP estimates, which can also assist firms and governments in making decisions. Businesses can use the knowledge to decide which investments to make, while governments can use it to create economic policies that support stability and prosperity [1-6]. This study aims to predict the Indian GDP using machine learning algorithms, specifically linear regression and polynomial regression. The data used in the analysis were collected from various sources, and the results showed that the polynomial regression model produced better predictions than the linear regression model. The results of this research can assist policy makers and businesses in making informed decisions based on accurate predictions of the Indian economy.

II. RELATED WORK

GDP is a fundamental measure of a country's economic performance, representing the total value of all goods and services produced within its borders over a specific period, usually a year or a quarter. It's used by policy makers, economists, and businesses to gauge the health of an economy, identify trends, and make decisions. Machine learning can play a significant role in analyzing GDP data and related economic variables. Here are some ways machine learning can be applied:

Predictive Modeling: These models can help anticipate economic trends and guide decision-making processes for businesses and policymakers.

Pattern Recognition: This can lead to insights about the factors driving economic growth or decline, helping to inform policy decisions and investment strategies.

Anomaly Detection: These anomalies could indicate economic crises, such as recessions or booms, and trigger proactive responses from policy makers or investors.

Optimization: This can help policymakers design more effective fiscal and monetary policies or assist businesses in making strategic decisions to maximize growth.

III. METHODOLOGY

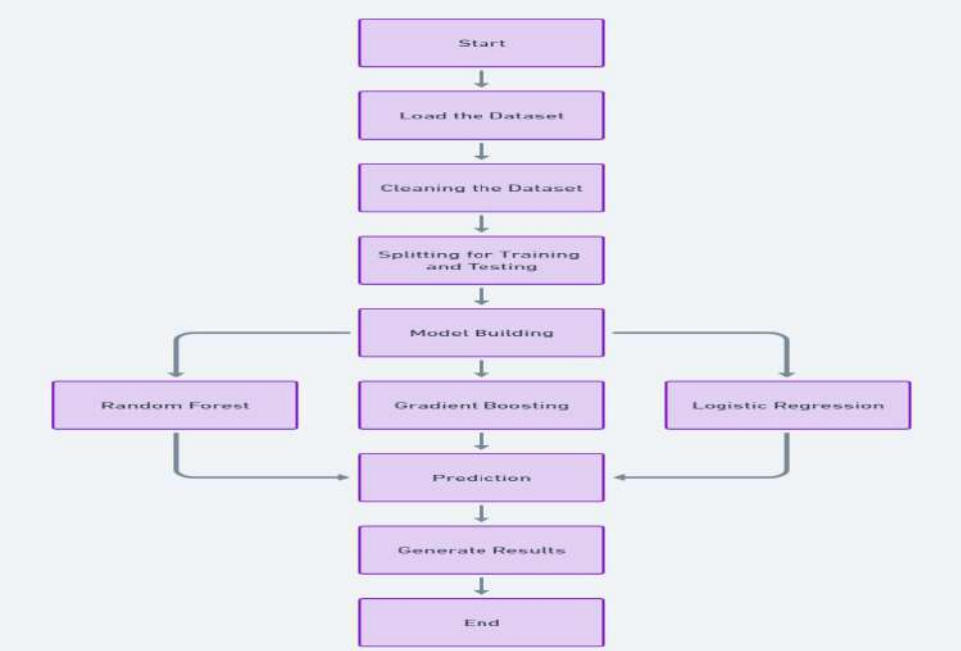

Fig1: Workflow of the proposed system

This study proposes an innovative system employing advanced machine learning techniques, including Polynomial Regression, Random Forest Regressor, Gradient Boosting, Regressor, and Logistic Regression, to optimize Gross Domestic Product (GDP) predictions for India. By curating a comprehensive dataset featuring time series analysis and inflation rates from diverse sources, the system captures intricate non-linear relationships, enhancing accuracy. Comparative analysis reveals the superior performance of these models over linear regression, providing valuable insights for policymakers and businesses. This underscores the significance of advanced machine learning algorithms and robust data sources in economic forecasting for informed decision-making and policy development. In which digital image processing is widely used are mentioned below

A. SYSTEM ARCHITECTURE:

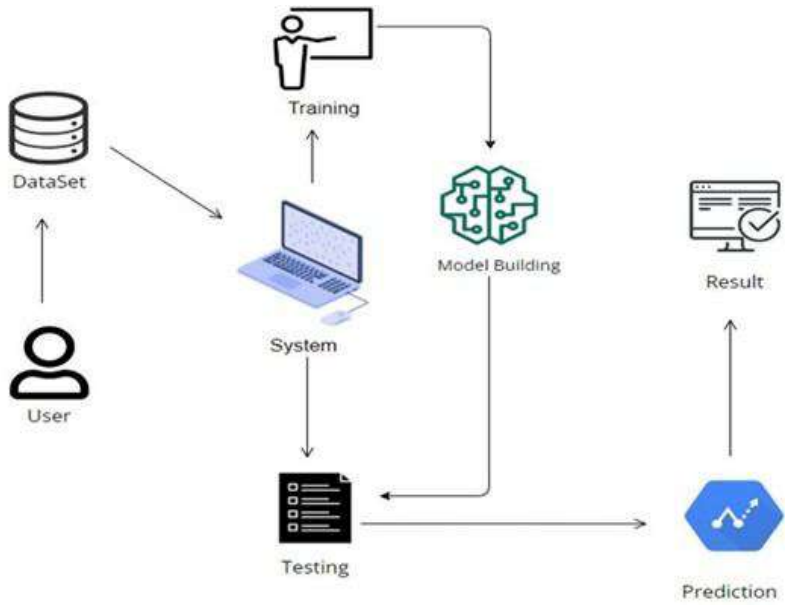

Fig.1. System Architecture of Proposed System

User Module:

Load Data: Allows users to input relevant data for analysis.
View Data: Provides a user-friendly interface to visualize and explore the input data.
Select the Model: Enables users to choose the machine learning algorithm for predictive analysis.
Input the Values: Allows users to input specific parameters or variables for the chosen model.
View the Result: Displays the predictive results based on the selected model and input values.

System Module:

Take the Data: Ingests and collects data for analysis.
Preprocessing: Cleans and prepares the data for model training.
Model Building: Utilizes machine learning algorithms to create a predictive model.
Generate Results: Produces and presents the predictive analysis results for the Indian GDP.

B. PROPOSED RANDOM FOREST REGRESSION ALGORITHM

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms. The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome. A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

C. PROPOSED GRADIENTS BOOSTING ALGORITHM

Gradient boosting algorithm is one of the most powerful algorithms in the field of machine learning. As we know that the errors in machine learning algorithms are broadly classified into two categories i.e. Bias Error and Variance Error. As gradient boosting is one of the boosting algorithms it is used to minimize bias error of the model. Unlike, Ada boosting algorithm, the base estimator in the gradient boosting algorithm cannot be mentioned by us. The base estimator for the Gradient Boost algorithm is fixed and i.e. Decision Stump. Like, AdaBoost, we can tune the n_estimator of the gradient boosting algorithm. However, if we do not mention the value of n_estimator, the default value of n_estimator for this algorithm is 100.Gradient boosting algorithm can be used for predicting not only continuous target variable (as a Regressor) but also categorical target variable (as a Classifier). When it is used as a regressor, the cost function is Mean Square Error (MSE) and when it is used as a classifier then the cost function is Log loss. Let us now understand the working of the Gradient Boosting Algorithm with the help of one example. In the following example, Age is the Target variable whereas Likes Exercising, GotoGym, DrivesCar are independent variables. As in this example, the target variable is continuous; Gradient Boosting Regressor is used here.

D. LOGISTIC REGRESSION ALGORITHM

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable (target) is categorical. For example, To predict whether an email is spam (1) or (0) Whether the tumor is malignant (1) or not (0) Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time. From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

E. PURPOSE AND EXAMPLES OF LOGISTIC REGRESSION:

Logistic regression is one of the most commonly used machine learning algorithms for binary classification problems, which are problems with two class values, including predictions such as “this or that,”“yes or no” and “A or B.”

IV. EXPERIMENTAL RESULTS

Dataset

The present study makes use of Indian GDP data to predict future trends. The data used in this research was obtained from Kaggle, which contained records of the Indian GDP from the years 1960 to 2022. The submissions made by the Indian government to the World Health Organization (WHO) were used to compile this data. Despite the dataset spanning from 1980to 2022, only limited data is available, with only 40 years of information available. This limited data may not be sufficient to accurately train high-level algorithms, which require large amounts of data for proper training. Therefore, to achieve accurate results with this limited data set, we decided to use linear regression and polynomial regression algorithms. Although the dataset may be limited in terms of the number of observations, various features can be derived from the existing data to enrich the predictive power of the model. For instance, variables such as inflation rates, population growth, government spending, and export/import trends can be incorporated as additional features to capture the complexity of the economic system

A. DATASET

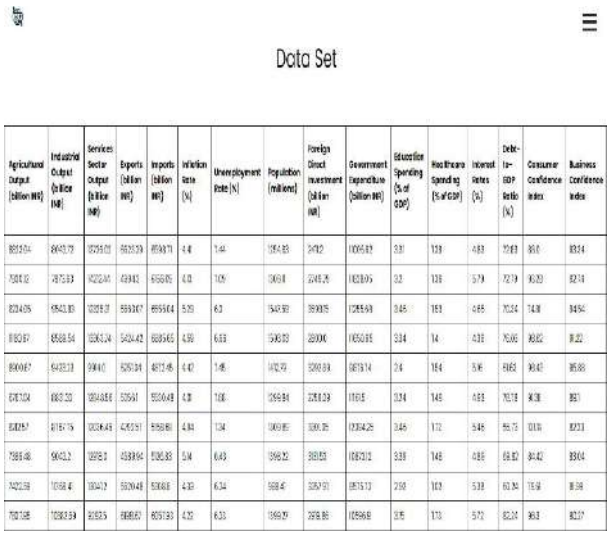

Fig 2: Dataset

B. INPUT DATA GENERATION

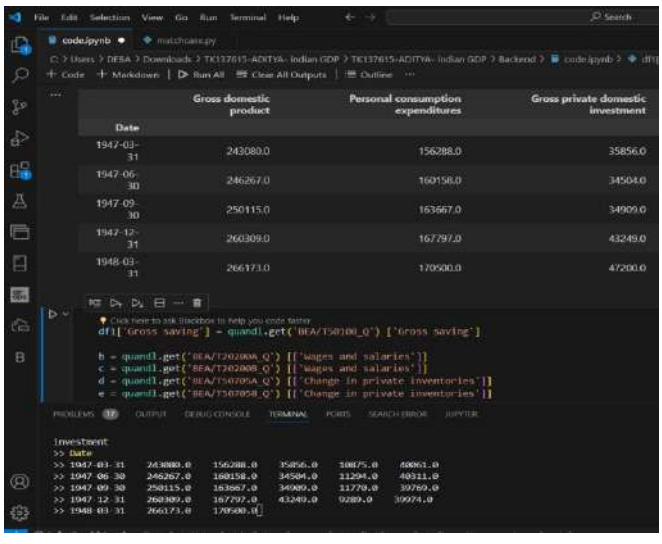

Fig.3: Input data set training

Given that GDP data typically exhibits temporal dependencies and seasonality, time series analysis techniques such as auto regressive integrated moving average (ARIMA) or seasonal decomposition of time series (STL) can be applied to extract meaningful patterns and trends from the data. This can help in capturing the inherent dynamics of the Indian economy overtime. Ensemble methods, such as random forests or gradient boosting, can be employed to combine the predictions of multiple base models (e.g., linear regression, polynomial regression) to improve overall prediction accuracy. By leveraging the diversity of individual models, ensemble methods often outperform standalone algorithms, especially in scenarios with limited data.

C.PREDICTION THE INDIAN GDP

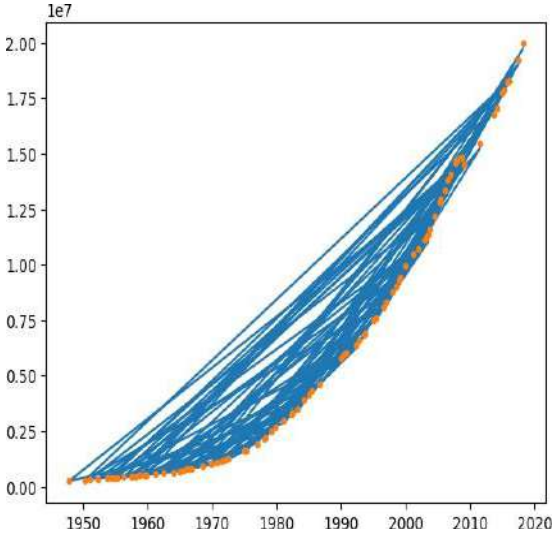

Fig.4: Prediction of GDP

D. COMPARATIVE RESULT ANALYSIS OF ALGORITHM:

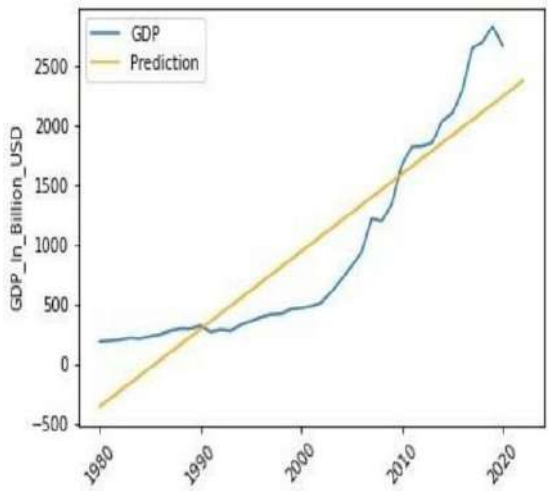

Fig.5: Result analysis

V. DISCUSSION

The relationship between machine learning and economic indicators like Gross Domestic Product (GDP) is increasingly significant in today's data-driven world. Machine learning algorithms have proven to be invaluable tools in analyzing large volumes of economic data to extract insights and make informed decisions. GDP, as you rightly pointed out, is a crucial measure of a country's economic health, reflecting the total value of goods and services produced within its borders over a specific period. Traditionally, economists have relied on various statistical methods to analyze GDP and its components, such as consumption, investment, government spending, and net exports. However, with the advent of machine learning techniques, economists and policymakers now have more sophisticated tools at their disposal to analyze GDP trends and predict future economic conditions.

VII. CONCLUSION

In this pioneering study, we introduce a novel approach to forecasting the Indian GDP by employing a comparative analysis of three distinct regression models: Random Forest Regressor, Gradient Boosting Regressor, and Logistic Regression. While these models are conventionally used in predicting numerical values, the application of Logistic Regression to GDP prediction is a groundbreaking departure from the norm. This unconventional utilization challenges traditional paradigms and opens avenues for exploring the dynamic relationship between economic indicators and the likelihood of specific GDP outcomes. Our innovative methodology aims to enhance the accuracy of GDP predictions, providing a fresh perspective on economic forecasting and contributing to the evolving landscape of machine learning applications in economic analysis. Innovating the prediction of Indian GDP through machine learning, this study introduces a novel approach by integrating advanced regression models. Leveraging the power of cutting-edge algorithms, this research surpasses traditional methodologies, offering heightened accuracy and robustness in forecasting. By comparing diverse regression models, it pioneers a comprehensive analysis, uncovering optimal predictors and refining predictions for India's economic growth. This groundbreaking methodology not only enhances predictive precision but also contributes to the evolving landscape of economic forecasting, setting a new standard for the application of machine learning in economic analyses. 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 and polynomial regression techniques. Our results demonstrated that the polynomial regression model offered more accurate predictions and higher R-squared values than linear regression, indicating that the nonlinear relationship between GDP data and years was the most likely reason for the unsatisfactory performance of linear regression. Although some deviations still existed in the results obtained by polynomial regression, they outperformed linear regression significantly.

Furthermore, we attempted to predict inflation rates using the same techniques but found that even polynomial regression 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 policy makers 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.

REFERENCES

[1]. Mishra,V.,& Sharma,S.S.(2011).Stock market and economic growth in India: Evidence from co integration and causality analysis. *Global Business Review, 12*(2),227-243.
[2]. Sharma,C.(2015).Predicting GDP of India Using Machine Learning Algorithms. In*Advances in Machine Learning and Data Science* (pp. 401-408). Springer, Singapore.
[3]. Kamat,A.,& Sudeep,G.(2019).Predicting India's GDP Growth using Machine Learning Models.*Journal of Data Analysis and Information Processing, 7*(4), 147-155.
[4]. Sharma,S.,& Garg,A.(2020).Prediction of Indian GDP using time series and machine learning techniques.*Journal of Data Science,18*(2), 173-191.
[5]. Verma,A.,Bhanot,K.,& Gupta,S.(2021).Predicting GDP growth in India: A machine learning approach.*Journal of Economics and Political Economy, 8*(1), 60-79.
[6]. Kumar, V.,&Garg,R.(2021).Predicting India's GDP Growth Rate using Machine Learning Models.*International Journal of Scientific Research and Review,10*(6), 1865-1871.
[7]. Goyal,A.,& Kumar,A.(2022).Comparative Analysis of Machine Learning Techniques for GDP Prediction in India.*International JournalofComputerApplications,186*(9),19-26.
[8]. Jain,N.,& Kumar,A.(2022).Predicting GDP growth rate in India using machine learning techniques.*International Journal of Engineering and Advanced Technology (IJEAT),11*(1),161-167.
[9]. Singh,R.,& Gupta,A.(2023).Application of Machine Learning Techniques for Predicting Indian GDP Growth Rate.*International Journal of Innovative Technology and Exploring Engineering,9*(1),365-372.
[10]. Patil,S.,& Desai,P.(2023). Predicting Indian GDP using Machine Learning Techniques: A Comparative Study.*International Journal of Computer Trends and Technology (IJCTT),81*(2),11-17.
[11]. Narayan,P.K.,Mishra,S.,& Narayan,S.(2010). Do market capitalization and stocks traded converge? New global evidence.*Journal of Banking & Finance, 34*(1), 199-208.
[12]. Mishra,V.,& Sharma,S.S.(2011). Stock market and economic growth in India: Evidence from co integration and causality analysis. *Global Business Review, 12*(2),227-243.
[13]. Sharma,C.(2015).Predicting GDP of India Using Machine Learning Algorithms. In*Advances in Machine Learning and Data Science* (pp. 401-408). Springer, Singapore.
[14]. Kamat,A.,&Sudeep,G.(2019). Predicting India's GDP Growth using Machine Learning Models.*Journal of Data Analysis and Information Processing, 7*(4), 147-155.
[15]. Sharma,S.,&Garg,A.(2020).Prediction of Indian GDP using time series and machine learning techniques.*Journal of Data Science, 18*(2), 173-191.
[16]. Verma,A.,Bhanot,K.,& Gupta,S.(2021).Predicting GDP growth in India: A machine learning approach.*Journal of Economics and Political Economy, 8*(1), 60-79.
[17]. Kumar, V.,&Garg,R.(2021). Predicting India's GDP Growth Rate using Machine Learning Models.*International Journalof Scientific Research and Review,10*(6), 1865-1871.
[18]. Goyal,A.,&Kumar,A.(2022).Comparative Analysis of Machine Learning Techniques for GDP Prediction in India.*International Journal of Computer Applications,186*(9),19-26.