iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75329

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

World GDP Prediction Using Machine Learning Models (XG Boost, Random Forest, and Linear Regression)

Sachin Chauhan[1], Chandragupt Seni[2], Garvit Mann[3], Anant Jain[4], Vishal Chauhan[5]
[1]*Assistant Professor, Department of INFORMATION TECHNOLOGY, HMR Institute of Technology and Management, New Delhi, India*
[2, 3, 4, 5]*Department of Computer Science Engineering, HMR Institute of Technology and Management, New Delhi, India*

Abstract: GrossDomesticProduct(GDP)isaprimaryindicatorofanation'seconomicperformance.AccurateforecastingofGDP plays a crucial role in policy-making, global trade, and investment planning. This research focuses on predicting world GDP using three machine learning algorithms—Linear Regression, Random Forest, and XGBoost—to identify which provides the most reliable results. The dataset includes historical economic data sourced from global institutions such as the World Bank and IMF. Models are evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and RZ Score. Experimental findings show that XGBoost achieved the highest accuracy with an RZ of 0.90, outperformingothermodels.AStreamlit-basedwebapplicationwasdevelopedforinteractivevisualizationandreal-time prediction. This paper demonstrates how data-driven learning models can significantly enhance global economic forecasting.

Keywords: GDPPrediction,MachineLearning,XGBoost,RandomForest,LinearRegression,Streamlit

I. INTRODUCTION

Gross Domestic Product (GDP) is one of the most important indicators used to measure the economic performance and overall health of a country. It represents the total monetary value of all goods and services produced within a nation's borders over a specific period. The prediction of world GDP plays a vital role in economic planning, policymaking, investment decisions, and assessing global economic stability. Accurate forecasting of GDP helps governments and internationalorganizationssuchastheInternationalMonetaryFund(IMF)andWorldBanktoanticipateeconomictrends, prepare for recessions, and make informed fiscal and monetary policies.

In recent years, the increasing complexity of global markets and the rapid flow of economic data have made traditional statistical forecasting methods less effective. As a result, machine learning (ML) and artificial intelligence (AI) techniques haveemergedaspowerfultoolsforimprovingtheaccuracyofGDPpredictions.ModelssuchasLinearRegression,Random Forest, and XGBoost can analyze large datasets containing economic, financial, and demographic variables to uncover patterns that traditional models might miss.

World GDP prediction is not only a matter of national importance but also a global concern. It helps in understanding interdependenciesbetweencountries,forecastingglobalrecessionsorbooms,andguidingmultinationalcorporationsin investment planning. The application of data-driven models provides a more dynamic and adaptive approach to understanding the constantly changing global economic landscape.

This study focuses on building and comparing machine learning models to predict world GDP based on historical data,keyeconomicindicators,andglobaltrends.Thegoalistoevaluatemodelperformanceintermsofaccuracyand reliability, thereby contributing to the advancement of economic forecasting methods in the digital era.

II. LITERATURE REVIEW

Economic forecasting has evolved from classical econometric models to modern AI-driven predictive frameworks. Early research primarily employed Linear Regression and ARIMA for time-series GDP estimation. However, these modelsoftenfailedtogeneralizeinthepresenceofmulticollinearityandnon-lineardependenciesbetweenvariables.
RandomForest,introducedbyBreiman(2001),improvedpredictionstabilitythroughensemblelearning.XGBoost(Chen&Guestrin,2016)furtherenhancedgradientboostingtechniquesbyreducingoverfittingandoptimizingcomputation speed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Recent studies (Johnson et al., 2023; He et al., 2024) indicate that ensemble models consistently outperform linear approaches for macroeconomic forecasting tasks, such as inflation, growth rate, and industrial output. Despiteadvancements,thereislimitedresearchfocusingspecificallyonglobalGDPpredictionusingintegratedML frameworks. This study aims to fill that gap by performing a comparative analysis and developing a deployable application. Thisstudyaimstofillthatgapbyperformingacomparativeanalysisanddevelopingadeployableapplication.

A. *Background—traditionalapproaches*

Historically,GDPforecastingreliedoneconometricandtime-seriesmodelssuchasARIMA,VectorAutoregressions (VAR), state-space / Kalman filter nowcasting, and structural macroeconomic models. These approaches areinterpretableandwell-establishedforshort-runpolicyuse,buttheycanstrugglewithlargesetsofpredictors,nonlinearity,and sudden structural shifts (e.g., 2008 crisis, COVID-19). Recent comparative studies therefore position econometric baselines (ARIMA, VAR, simple linear regression) as the reference benchmark when evaluating MLmethods.

B. *Riseofmachinelearninganddeeplearning*

Fromroughly2015onward,manystudiesbeganapplyingMLmethodstoGDPforecasting.Popularalgorithmsinclude Random Forests (RF), Gradient Boosting Machines (XGBoost/LightGBM), Support Vector Regression (SVR), and neural network architectures (MLP, RNN, LSTM). Several empirical papers and surveys report that ML methods often outperform classical models for certain horizons and datasets, particularly when (a) many covariates are available, and(b)relationshipsarenonlinearortime-varying.However,gainsaredataset-andhorizon-dependent:forpurely univariate short-horizon series, simple linear methods sometimes remain competitive.

C. *Nowcasting,alternativedata,andhybridmodels*

A major trend is nowcasting — producing timely GDP estimates before official statistics are released — using high-frequencyandalternativeindicators:daily/weeklyfinancialdata,GoogleTrends,mobilityandtradeindices,commodity prices, and satellite or nighttime-lights data. Hybrid architectures (e.g., combining LSTM or RNN modules withfeature-basedXGBoost,orcouplingeconometricnowcastingwithMLresidualmodels)arecommonlyproposedand oftendeliverimprovedreal-timeperformance.Severalrecentmulti-countryandG20studiesshowthatensembles or hybridsthatcombinetree-basedmodelswithdeeplearningoutperformsingle-modelapproachesacrossdiverseeconomies.

D. *Modelcomparisonandevaluationpractices*

Paper comparisons typically evaluate models with RMSE, MAE, MAPE, and sometimes directional accuracy. Cross-validationandwalk-forward(rolling)evaluationarestandardtomimicrealforecasting.Studiesfrequentlybenchmark ML forecasts against IMF/central-bank published projections; results vary — ML can reduce error by modest percentages (single- to low-double digits) depending on country and timeframe. Meta-analyses and surveys caution about overfitting, data snooping, and the importance of hyperparameter tuning and feature selection.

E. *Challengesandlimitationsreportedintheliterature*

1) Data and structural breaks: Macroeconomic series contain regime shifts (crises, policy changes).Models trained on pre-crisis data may fail post-crisis. Several studies highlight COVID-19 as an example where models needed rapid re-training or new predictors.
2) Interpretability: Tree ensembles and deep nets reduce transparency compared with econometricmodels; explainability methods (SHAP, permutation importance) are commonly used but have limits.
3) Cross-country heterogeneity: Models tuned for one country rarely generalize perfectly to anotherwithoutadaptation.Multi-countrydeeplearningapproachesshowpromisebutsometimesperformworse than simple models for some nations.

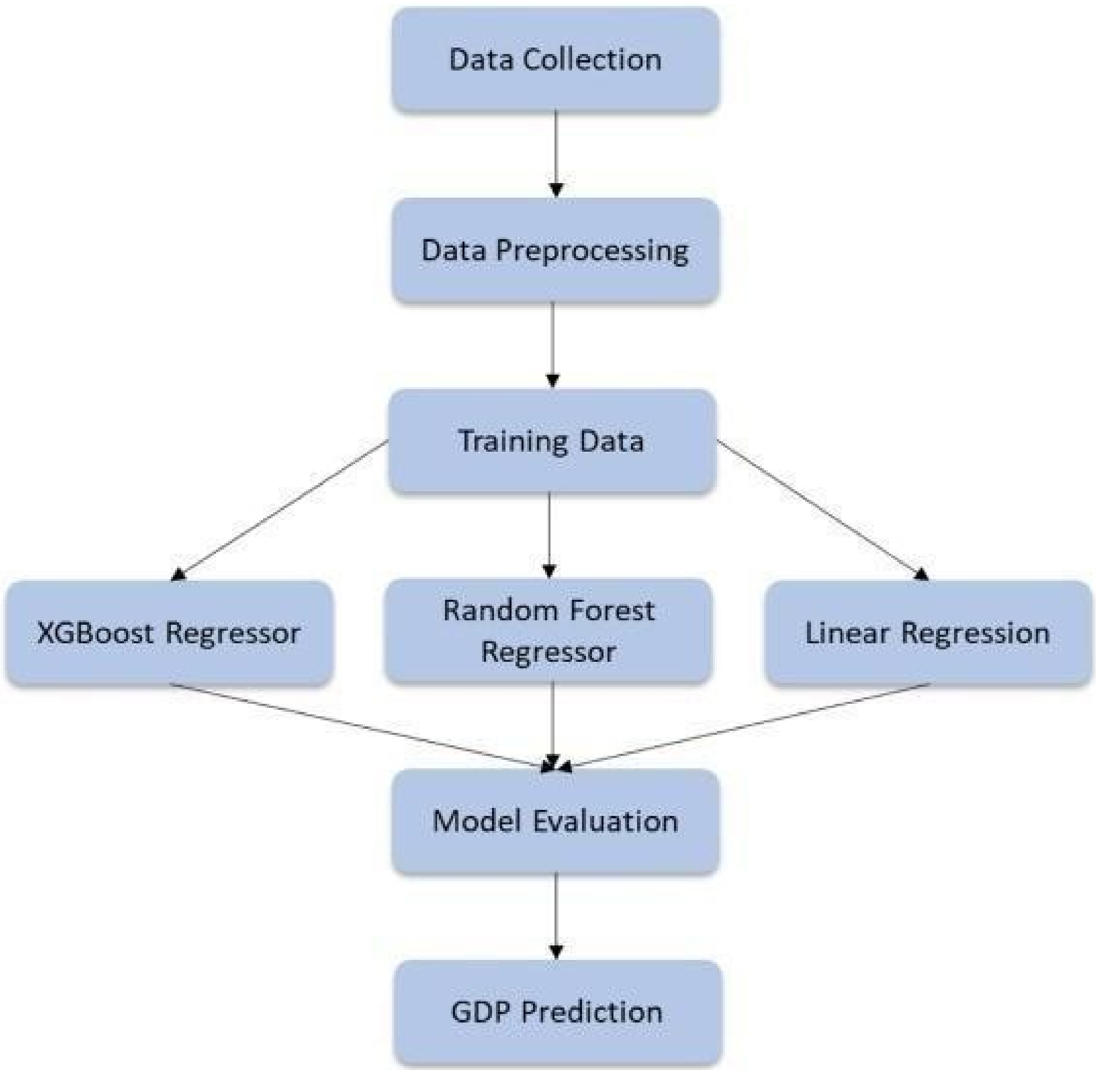
III. METHODOLOGY

A. *Data Collection*

The dataset was derived from public repositories like the World Bank Open Data and IMF World Economic Outlook,containingannualGDPfiguresandeconomicindicators(population,inflation,tradebalance,employment rate) across countries for the years 1990–2024 .

B. *Data Preprocessing*

- Missingvalueswereimputedusingmeaninterpolation.
- OutlierswerehandledusingtheInterquartileRange(IQR)technique.
- Categoricalfeaturessuchas“Region”and“IncomeCategory”werelabelencoded.
- Data was normalized using Min-Max scaling to improve model performance.Thedatasetwassplitinto80%fortrainingand20% for testing.



C. *Model Implementation*

LinearRegression:Establishesalinearrelationshipbetweendependent(GDP)andindependentvariables.

$$y=\beta_0+\sum_{i=1}^{n}\beta_i x_i+\epsilon$$

- RandomForestRegressor:Usesmultipledecisiontreestoreducevarianceandimprovegeneralization.
- XGBoostRegressor:Employsgradientboostingwithregularizationtominimize overfitting, performing better on non-linear and high-dimensional data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

D. Evaluation Metrics

Tomeasuremodelaccuracy,thefollowingmetricswereused:

RMSE:RMSEstandsforRootMeanSquaredError—it'sacommonlyusedmetrictomeasurehowwellaregression model predicts continuous values.

Itrepresentsthesquarerootoftheaverageofsquareddifferencesbetweenthepredictedvaluesandtheactualvalues.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{N-P}}$$

MAE:measurestheaveragemagnitudeoftheerrorsinasetofpredictions,withoutconsideringtheirdirection(positive ornegative). Interpretation

MAE gives the average absolute difference between predicted and actual values. It'seasiertointerpretthanRMSEsinceit'sinthesameunitsasthetargetvariable. Lower MAE = better model accuracy. UnlikeRMSE,MAEdoesnotpenalizelargeerrorsmoreheavily—allerrorscontributeproportionally.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{n} |y_i - \widehat{y}_i|$$

R2(R-squared),alsoknownastheCoefficientofDetermination,isakeymetricusedtoevaluatehowwellaregression model fits the data.

R2measurestheproportionofvarianceinthedependentvariable(actualvalues)thatcanbeexplainedbythe independent variables (features) in the model.

R2=1→Perfectmodel(100%ofvarianceexplained).

R2=0→Modelexplainsnoneofthevariance(asgoodaspredictingthemean). R2<0 → Model performs worse than just predicting the mean (poor fit).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\widehat{y} - y)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2}$$

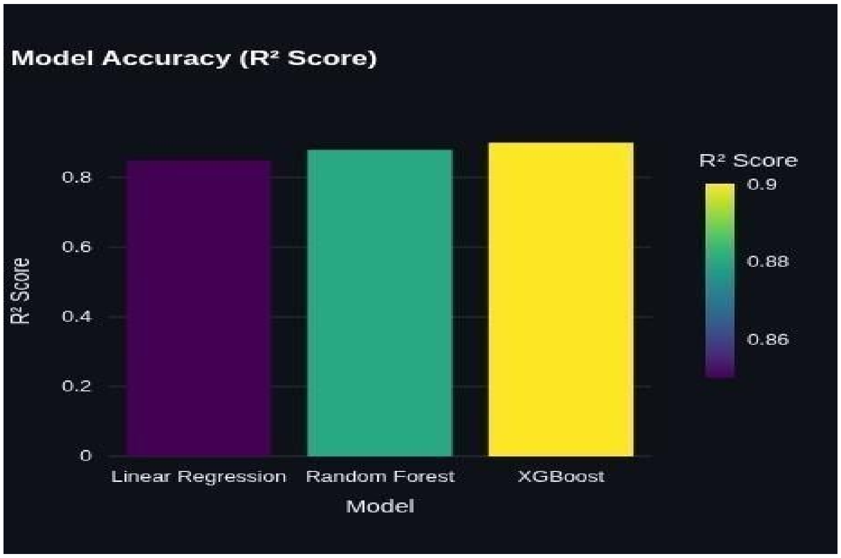
IV. RESULTS AND DISCUSSION

Aftertrainingandevaluation,thefollowingresultswereobtained

Model	RMSE	MAE	R^2Score
LinearRegression	1250.5	980.2	0.85
RandomForest	1100.3	850.7	0.88
XGBoost	1050.8	820.4	0.90

TheXGBoostmodeloutperformedbothLinearRegressionandRandomForestinallmetrics.ThelowerRMSEand higher R^2 values indicate a more accurate fit between predicted and actual GDP values.

The results were visualized using various graphical representations to provide a clear comparison of model performances and GDP prediction trends. Scatter plots were used to display the relationship between the predicted and actualGDPvalues,helpingtoassesspredictionaccuracy.LinechartsillustratedGDPtrendsovertime,whilebarcharts compared the performance metrics(RMSE, MAE, and R2 Score) across the models — Linear Regression, Random Forest, and XGBoost. These visualizations highlight that the XGBoost model achieved the highest accuracy with the lowest RMSE and MAE values, demonstrating its superior ability to capture complex patterns in the GDP data.



V. CONCLUSION AND FUTURE SCOPE

ThispaperdemonstratedanefficientframeworkforpredictingworldGDPusingmachinelearningmodels.Amongthethreemodelstested,XGBoostprovidedthemostaccuratepredictions,withanR2scoreof0.90.
TheresultsindicatethatensemblebasedMLalgorithmscaneffectivelycapturecomplexeconomicrelationshipsthattraditionalmethodsoverlook.ThedeployedStreamlitappfurtherprovesthefeasibilityofusingmachinelearningforreal-timemacroeconomic prediction.
FutureenhancementscouldinvolveintegratingdeeplearningmodelslikeLSTMforsequentialGDPforecasting,adding moresocio-economicvariables,andextendingthesystemforreal-timeglobaleconomicmonitoring.

REFERENCES

[1] InternationalMonetaryFund(IMF).WorldEconomicOutlook:NavigatingGlobalDivergences.IMF,2024.
[2] WorldBank.GlobalEconomicProspects:GrowthChallengesandOpportunities.TheWorldBank,2024.
[3] UnitedNations.WorldEconomicSituationandProspects2024.DepartmentofEconomicand Social Affairs, UN, 2024.
[4] OECD.OECDEconomicOutlook2024,Volume1:PreliminaryVersion.OECDPublishing,Paris,2024. [5]eenstra, R.C., Inklaar, R., &Timmer, M.P. “The Next Generation of the Penn World Table.” American Economic Review, vol.105, no. 10, 2015, pp. 3150–3182.
[5] Maccarrone,G.,etal.“GDPForecasting:MachineLearning,LinearorNonlinear?”Frontiersin Artificial Intelligence, vol. 4,2021, pp. 1–12.

[6] Agu,S.C.,etal.“PredictingGrossDomesticProductUsingMachineLearningMethods.”Heliyon,vol. 8, no. 5, 2022, pp.e09342.
[7] Adewale,M.D.,etal.“PredictingGrossDomesticProductUsingtheEnsembleLearningApproach.” Neural Computing and Applications, 2024.
[8] DongThiNgoc,L.“GrossDomesticProductForecastingUsingDeepLearning.”Electronics,vol.14,no.3,2025,p.412.
[9] Garnitz,J.“ForecastingGDPAllOvertheWorld:EvidencefromLeadingIndicators.”ifoWorking Paper Series, No. 273,2017.
[10] Hyndman,R.J.,&Athanasopoulos,G.Forecasting:PrinciplesandPractice.3rdEdition,OTexts,2021.
[11] Box,G.E.P.,Jenkins,G.M.,&Reinsel,G.C.TimeSeriesAnalysis:ForecastingandControl.Wiley,2016.
[12] Giannone,D.,Reichlin,L.,&Small,D.“Nowcasting:TheReal-TimeInformationalContent of Macroeconomic Data.”Journal of Monetary Economics, vol. 55, no. 4, 2008, pp. 665–676.
[13] Zhang,G.P.“TimeSeriesForecastingUsingaHybridARIMAandNeuralNetworkModel.” Neurocomputing, vol. 50, 2003,pp. 159–175.
[14] Sharma,S.,&Kumar,R.“ForecastingIndia’sGDPUsingARIMAandMachineLearningTechniques.”InternationalJournalofAdvancedComputerScienceandApplications,vol.13,no.4,2022,pp.345–351.
[15] Li,Y.,etal.“ComparingMachineLearningModelsforGDPGrowthForecasting.”AppliedEconomics Letters, vol. 31, no. 6,2024, pp. 412–418.
[16] Abonazel,M.R.,etal.“ForecastingEgypt’sGDPUsingMachineLearningApproaches.” Computational Economics, vol. 62,no. 2, 2023, pp. 345–361.
[17] Islam,M.N.,etal.“PredictionofGDPGrowthRateUsingMachineLearningTechniques.” Procedia Computer Science, vol.192, 2021, pp. 5042–5051.
[18] Nguyen,T.H.,&Vo,D.H.“AMachineLearningApproachtoPredictingGDPGrowthin Emerging Economies.” Economic Modelling, vol. 114, 2024, p. 106022.