iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60136

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

An Application for Prediction of World Countries GDP

Dr. J. Sasi Bhanu[1], N. C. Tejaswini[2], P. Pranaya Yadav[3], T. Keerthi[4], M. Sai Nigam[5]
[1]*Professor, Department of CSE (Artificial Intelligence & Machine Learning), CMR College of Engineering & Technology, Hyderabad, Telangana State, India*
[2, 3, 4, 5]*B.Tech Students, Department of CSE (Artificial Intelligence & Machine Learning), CMR College of Engineering & Technology, Hyderabad, Telangana State, India*

Abstract: ***Gross Domestic Product (GDP) is an important metric for assessing the economic well-being of a nation. In this research paper, we will employ and compare multiple machine learning algorithms to predict the GDP across multiple countries and areas, including Linear Regression, SVM, Gradient boosting, and Random Forest. Ultimately created an API where users can enter the current values of parameters on which GDP depends, to get a more accurate prediction.***
Keywords: Gross Domestic Product (GDP), Linear Regression, SVM, Gradient Boosting, Random Forest, API, prediction.

I. INTRODUCTION

GDP is a major indicator of the health of the economy of every nation. Understanding Gross Domestic Product (GDP) dynamics is paramount in assessing a nation's economic health, serving as a fundamental measure of its overall economic activity. It denotes the total value of all goods and services produced within national boundaries, thus giving insights into how big or small an economy is and its future direction. Over time, estimates and predictions on GDP have had to undergo significant changes from traditional means because now these are supplemented by artificial intelligence.

This paper investigates the use of machine learning approaches in predicting GDP. We will conduct a comparative study on multiple countries and regions using different algorithms such as Linear Regression, Support Vector Machines (SVM), Gradient Boosting, Random Forest, etc. This study should provide an understanding of how various factors contribute to predicting GDP using these ML methods and evaluation of each model's performance is done through metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R2 score to enable a comprehensive assessment of predictive capabilities.

Additionally, the paper proposes the development of a specialized Application Programming Interface (API) tailored for Gross Domestic Product prediction which allows the user to enter the current parametric values that the GDP depends on, like population, infant mortality rate, population density, annual net migration, etc., in order to achieve the most accurate prediction.

II. RELATED WORK

A. "Data Science Project on GDP Analysis with Python."

It is possible to analyze the GDP (gross domestic product) of various nations using data science techniques. It does, however, make abundantly evident that the GDP serves as the main benchmark for evaluating the country's financial status. Furthermore, based on the country's economic growth, the GDP per capita is used as a measure to assess the prosperity of the country. Here, the focus is on data science techniques that may be applied to analyze big data sets critically in order to gain insight into the process of making decisions, generating predictions, and identifying patterns related to variations in GDP. Additionally, it provides essential insights from statistics, data pre-processing, forecasting models, and machine learning in order to answer the numerous challenges of various industries.

B. "Estimation, Analysis and Projection of India's GDP."

Analyzing the GDP growth of India through time series to predict the values of future GDP and to recognize key trends underlying policy-related reforms. For the assessment of the correlation between time and GDP, as well as the importance of trade liberalization dummy variables, the facilities of the Microsoft Excel spreadsheet program are used for regression analysis. The stationarity of the GDP is examined by means of unit root tests and exponential trend relationship procedure. The study concerns both main sectoral components of the GDP as well as a dummy variable that demonstrates the division between growth rates before and after the liberalization. Even with the removal of the dummy variable, the model fit is still very good.

©IJRASET: All Rights are Reserved

This indicates that India's GDP is a stationary process This analysis illustrates a significant link between the Trade, Transport, Storage, and Communication sector and GDP growth, thus concluding it is a good tool for forecasting and planning.

C. "Forecasting GDP: A Linear Regression Model."

Examining how GDP is calculated in India through statistical analysis by applying a linear regression model for predicting future GDP values based on the past. The aim is to enhance economic development by analyzing GDP with expenditure, production, and income approaches. Consequently, the utilization of regression techniques enables policymakers to spot trends as well as patterns from the data of GDP thus giving valuable knowledge to various decision-makers in government and researchers in economic planning.

D. "Analysis of GDP using Linear Regression."

The analysis of GDP by applying a linear regression tool is done to understand changes in the economy and their relation to the stock market, aiming to create a predictive model for policymakers, businessmen, and investors. Using data from Yahoo Finance, BEA.gov, and Bloomberg Terminal the thoroughness increased while calculating everything from the GDP model became simple. The regression analysis is conducted on the S&P 500 and GDP indexes to have the possibility of knowing what must be followed to any change in an economy. Focusing on the GDP impact on the S&P industry and particular areas including stock market performance and found, acting in different sectors, but very few of them show a positive relation versus the rest.

III. SYSTEM REQUIREMENTS

A. Hardware Requirements

1) Processor -Intel i7 core or above

2) RAM- 16GB or above

B. Software Requirements

1) Windows

2) Visual Studio Code

IV. METHODOLOGY

A. Importing Libraries and Data

The dataset of "countries_of_world.csv" from Kaggle contains 227 rows representing different countries and 20 columns representing various features that contribute to the GDP. These features include Country name, Population, Population density, Area, Literacy rate, Birth rate, Death rate, Industry, Service, GDP per capita, etc. To perform steps such as uploading the dataset, preprocessing, and forecasting we import different libraries such as numpy, pandas, seaborn, etc.

B. Data Preprocessing

The dataset consists of many missing data so normalization is much required. To do so, firstly, we rename the columns to short forms and assign float/string types to them as they are all objects. Then we fill in the missing data values. We do this by observing the same region's data, some can be filled by calculating the mean values of that region and assigning it to them. Then, EDA is implemented in order to understand the correlation between all the features.

C. Data Split for Training and Testing

Data is split into four types i.e., selecting all features without scaling, all features with scaling, selecting a few features with scaling, and without scaling.

D. Model Training

Though most features have no linear relation we apply linear regression on the data and then SVM is implemented, along with Gradient boosting and Random Forest algorithms. After implementing, we compare the results of all models by using model evaluation metrics. The algorithm that provides the most accurate score and least error is chosen as the final algorithm which will be used in the API to predict the GDP. Optimized random forest is also implemented to get better results.

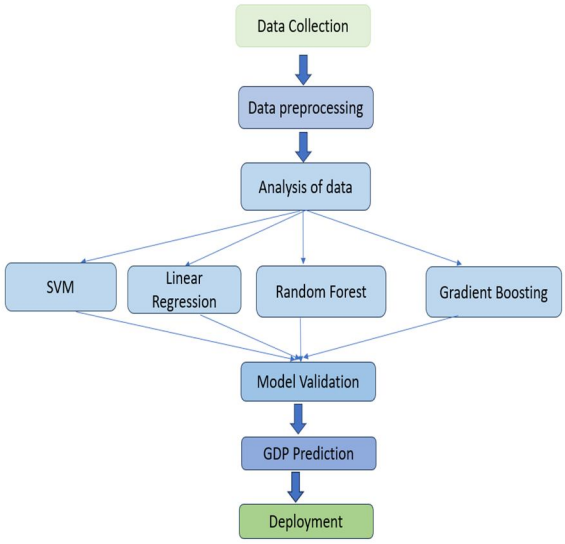
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. Model Evaluation

To calculate which algorithm gives the most accurate score, MAE, RMSE, and R2 scores are found and compared. The model with the least error is deployed for the estimation of GDP.

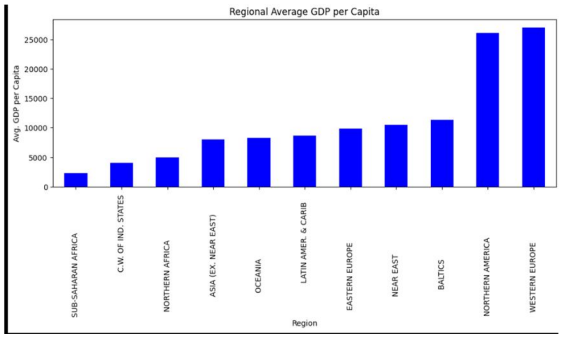
F. Block Diagram

The figure below depicts the flow for execution of the trained model.

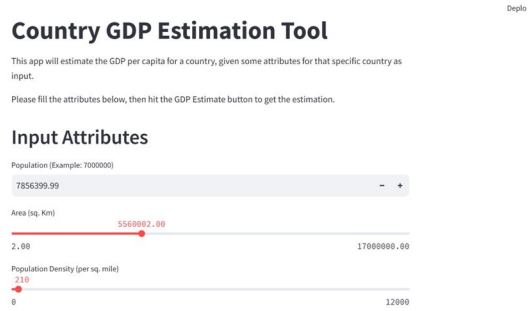


V. RESULTS AND DISCUSSIONS

The below image depicts the regional average GDP per capita.



Fig(1)Regional Average GDP per capita



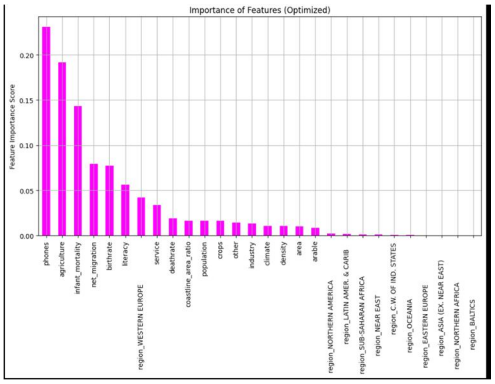
Fig(2) API of GDP estimation tool

The result of the research project is an API that allows the user to enter the key attribute values that contribute most to the GDP, they are processed and used to predict the GDP.



Fig(3)Some parameters determining GDP

Parameters like Area Ratio, Infant mortality rate, Population literacy rate, Land %, Cropland %, etc. are some of the indicators used for analysis. The sum of percentages of Arable, Crops, and other land should be 100%. Agricultural, Industrial, and Services Economy should sum up to 1.



Fig(4)Importance of features

By choosing which area one wants to calculate the GDP in, and by quickly searching the exact parametric values on Google, the GDP prediction varies to give the most accurate prediction.



Fig(5)Result estimation of GDP

The resultant predicted GDP value is displayed as shown above.

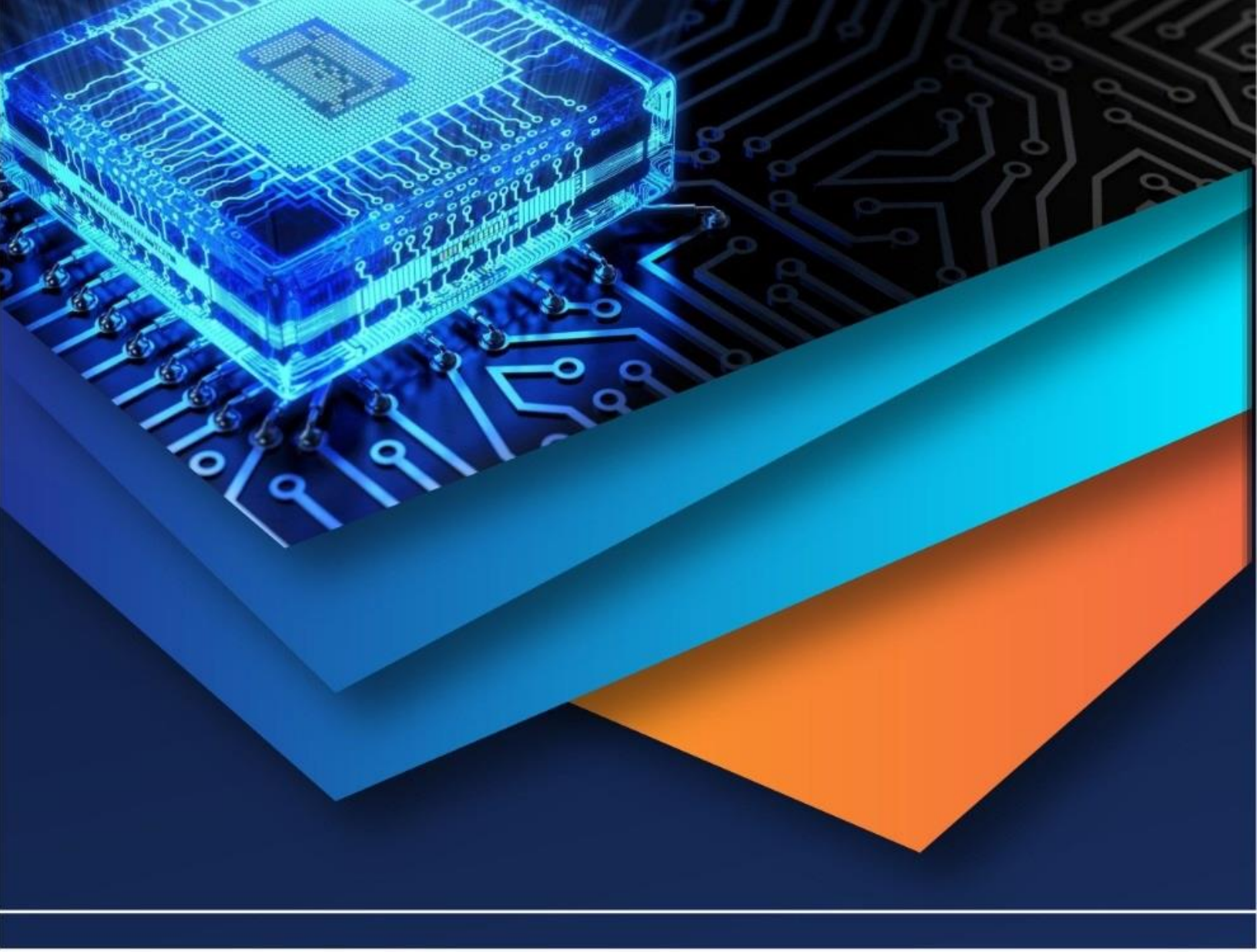
VI. CONCLUSION

After implementation of four algorithms (Linear regression, SVM(Support Vector Machine), Gradient boosting, and Random Forest) we get the best performance by random forest when all the features are considered without scaling with MAE: 2142.13, RMSE: 3097.19, R2_Score:0.88. Linear regression and SVM have given the least performance and Gradient boosting performance is almost similar to the Random forest score. Deployed the model of the "GDP Estimation Tool" which can estimate the GDP of a country by giving required attributes as input.

REFERENCES

[1] Data science project on GDP analysis with python by Srilakshmi Madadi UG Student, Department of Computer Science and Engineering, Kakatiya Institute of Technology and Science, Warangal Urban, Telangana, India
[2] Estimation, Analysis and Projection of India's GDP by Ugam Raj Daga and Rituparna Das and Bhishma Maheshwari
[3] Exploratory Data Analysis on Indian economy using Python by Obli Karthi M, Kamalesh S, Mohammed Riswan U, Student, Mechanical Engineering Department, Kumaraguru College of Technology
[4] Data Mining Methods and Models for Social and Economic Processes Forecasting by Yuliia Dehtiarova, Kyiv School of Economics and Yuvi Yevdokimov, University of New Brunswick
[5] Analysis of GDP using Linear Regression by Ryan Mayolo
[6] Forecasting GDP: A Linear Regression Model by Gourav Kalbalia, Vivek Tambi Cluster Innovation Centre, University of Delhi
[7] Random forest and support vector machine on features selection for regression analysis by Christine Dewi and Rung-Ching Chen

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)