**Mini Project Report on**



**Prediction of GDP Growth**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

***Parth Sarthi 2017520***

***Under the Mentorship of***

**Mr. Vikas Tomar**

**Asst. Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January 2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Prediction of GDP Growth”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Mr. Vikas Tomar, Asst. Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Parth Sarthi University Roll no: 2017520

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1** |
| Chapter 2 | Literature Survey | **3** |
| Chapter 3 | Methodology | **4** |
| Chapter 4 | Result and Discussion | **6** |
| Chapter 5 | Conclusion and Future Work | **8** |
| Chapter 6 | References | **9** |

**Chapter 1**

**Introduction**

**1.1 Introduction**

The Gross Domestic Product (GDP) is a unified, all-encompassing measure of economic activity that considers the total value of goods and services produced in the economy. Academics, investors, and regulators regard it as a proxy for the economy's wealth and an informative indicator that drives decision-making processes. As a result, the GDP forecast is an important issue. Indeed, it is worthwhile to target national economic policies as well as other areas ranging from non-performing loans to natural disasters.

GDP is a key metric for assessing a country's health and condition in comparison to other countries. As a result, knowing GDP ahead of time can help determine whether a country's economic health is improving or deteriorating. Gross domestic product (GDP) is a measure used to assess a country's overall economic performance; it includes all products and services produced by the economy, as well as personal consumption, government expenditures, and so on.

The social, economic, and cultural environments all have an impact on the country's economic growth. We used parameters such as population, area, population density, coastline, net migration, literacy, phones, stillbirths, arable, crop land, and other land, birth rate, death rate, region, and climate to calculate the country's GDP per capita. As a result, we created a prediction model with all these parameters as predictor factors and global GDP as the dependent variable.

Our goal is to use machine learning algorithms to predict and forecast per capita GDP for a country. GDP forecasting entails using applied mathematics and mathematical models to forecast future economic developments. It enables us to examine previous economic movements and forecast how current economic changes may alter the correlations of previous trends. As a result, more accurate forecasting would greatly assist the government in setting economic development goals.

As a result, a correct GDP prediction provides a number one insight which affiliates an understanding of future economic trends.

GDP is considered one of the most essential economic indicators, helping specialists to build a more realistic picture of a country's financial situation. When the economy is doing well, wages rise, and the unemployment rate declines as businesses hire more people to meet increasing economic needs. Increased GDP reflects rising profits inside the country, as does consumer purchasing power, and vice versa. A significant change in GDP, whether positive or negative, has an influence on the stock market.

Diagram

Description automatically generated

**1.2 GDP Calculation**

Personal consumption expenditures, government spending, corporate investment, and the trade balance are all common components of GDP.

There are three ways to calculate GDP. The revenue approach, the expenditure approach, and the production approach are the three.

The income method assumes that all expenditures in an economy should equal the total revenue generated by the production of all economic products and services. GDP is calculated as the total revenue obtained by all economic entities in the country in the form of factor income such as profit, wages, rental income, dividend income, and interest income.

The expenditure technique is predicated on the premise that all merchandise and services must be purchased by someone. It indicates that the overall value of manufacturing output must equal the total value of consumer spending on goods. GDP is calculated as the total amount of personal consumer expenditure, gross domestic private capital, government expenditures, and the net of exports minus imports within the economy over a particular period.

The most direct technique is the production or value-added strategy. It is calculated by adding the gross value added of all industries in the nation. In other words, GDP is calculated as the sum of the value added by all services and items produced inside the economy during a certain period.

**Chapter 2**

**Literature Survey**

GDP not only aids in the diagnosis of economic problems, but also in their resolution. Keeping all these factors in mind, we picked a topic of estimating GDP that can be utilized by any ordinary citizen who wants to know their country's GDP.

There are co-integration links between inflation and money. According to one study, slowing money expansion may result in lower inflation. Inflationary fluctuations were separated into two categories: predictable and unexpected. In the case of differentiated inflation, foreseeable increases have a negative impact on GDP growth. Unpredictable increases have a favorable impact on real GDP growth. However, this study solely analyses money as an economic element and ignores other crucial indices of economic health.

In another study published in 2010, a support vector machine trained with a genetic approach is used to anticipate GDP. The researchers determined that the evolutionary algorithm could provide optimum solution in a relatively short amount of time, making it a good tool for selecting support vector machine settings. The SVM parameters are then simultaneously optimized using a genetic algorithm. GDP statistics from 1989 to 2002 were utilized for training, while data from 2003 to 2007 was used for testing. However, the project's drawback is that various Machine Learning algorithms potentially outperform SVM.

Another research, titled "**Predicting GDP Using Autoregressive Models**" was released in 2017. To forecast GDP, they employed autoregressive models and subsequently built a vector autoregressive model. The expected outcome corresponds to past GDP data and forecasts consistent future growth. The limitation of this technique is that it fails to surmount the historic economic downturn. This study also did not take into consideration other variables such as trade, economic, and geographical factors in predicting GDP growth.

None of the study described above evaluated the holistic perspective of GDP dependency on Social, Economic, Geographical, and Environmental aspects to estimate global GDP. Most of the effort included developing mathematical or statistical models. As a result, we are attempting to integrate all such comprehensive parameters in our suggested model. We will investigate which parameters have an influence on GDP and which parameters are connected.

**Chapter 3**

**Methodology**

**3.1 Design and Framework**

The goal of our proposed system is to create a GDP Estimation Tool with a higher accuracy ML model than the existing one. Furthermore, our system is open to everyone and is not limited to a certain set of users.

**Diagram

Description automatically generated**

**A. Data Collection:**

We will use macrotrends to acquire data to anticipate factors driving GDP growth. The collection includes GDP of last 62 years.

**B. Data Pre-Processing:**

Data pre-processing is required to clean the data and prepare it for use with a machine-learning model. It contributes to a model's efficiency and accuracy. The primary goal of data cleaning is to detect and eliminate inconsistencies and duplicate data to generate a reliable dataset. Our dataset might consist of some missing data points.

**Text

Description automatically generated**

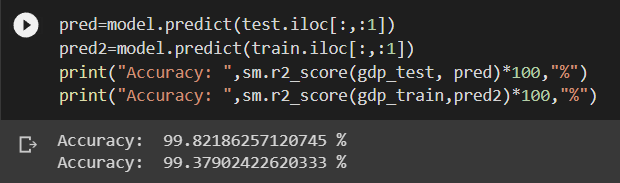
**C. Model Training:**

The different algorithms can be used to train and test the split data. We have applied MARS (**Multivariate Adaptive Regression Splines**). It is an algorithm designed for multivariate non-linear regression problems. **Regression Problems** require a model to forecast a numerical value. **Multivariate** refers to the presence of more than one input variable, whereas **Non-Linear** refers to the fact that the connection between the input variables and the target variable is not linear, and so cannot be expressed using a straight line.

The MARS method works by identifying a series of basic piecewise linear functions that characterize the data and combining them to create a forecast. In other ways, the model is a collection of linear functions. A piecewise linear function is one that is formed up of smaller functions. It is a function in this scenario which either outputs 0 or the input value directly.

**D. Model Evaluation:**

After completion of model training, we have checked the accuracy of our machine learning algorithm. Performance of these algorithms is evaluated using r2 score. The accuracy of our model was found to be above **99%.**

****

**E. GDP Prediction:**

We will be able to anticipate GDP per capita for any year once the model has been evaluated.

**Text

Description automatically generated**

**Chapter 4**

**Result and Discussion**

To evaluate the model, true GDP per capita was plotted against the prediction.

1. **Dataset**

Figure shows the depiction of Dataset for true GDP per capita. True GDP per capita was plotted against the year to be used for model creation.

**Chart, histogram

Description automatically generated**

1. **Regression**

**Chart, line chart, scatter chart

Description automatically generated**

1. **Training Set**

The Dataset was divided into Training Set and Test Set in the ratio of 80-20. Figure shows the Training Set used to train our GDP Prediction Model.

Chart

Description automatically generated

1. **Test Set**

The following graph shows the prediction of the test set used to test the accuracy of the model.

Chart, scatter chart

Description automatically generated

**Chapter 5­**

**Conclusion and Future Work**

We investigated all the supervised regression models to find the best-fitting models. We used the MARS technique to train the model and to assess its performance. To provide a detailed overview of performance, MAE and RMSE approaches are used. MARS predicts the genuine GDP per capita with an accuracy of more than 99%.

To summarize, the project I created on GDP growth projection is around 99% correct. Because GDP growth was moderate before 2001 but expanded at a faster pace afterward, fitting the model with a linear regression model was not possible, hence Multivariate Adaptive Regression Spline was utilized. It tackles the difficulties of linear regression by providing a continuous, more uniform curve to fit the data onto, and it outperforms random forest techniques in terms of predicting values beyond the dataset's extremes.

The Hinge function in it made it easier to fit the curve and make predictions. Because the data accessible online only spanned 62 years, from 1960 to 2021, it is impractical to make projections for the following 50 years or so because we do not know what the pace of growth will be.

As a result, our study provides near-accurate forecasts for the next 20-30 years and may be improved with a much bigger dataset.

**Chapter 6**

**References**

1. PyEarth Documentation

<https://contrib.scikit-learn.org/py-earth/content.html>

1. MARS algorithm

<https://medium.com/@aishwaryapatange/pyearth-multivariate-adaptive-regression-splines-on-python-86e8dd647030>

[MARS: Multivariate Adaptive Regression Splines — How to Improve on Linear Regression? | by Saul Dobilas | Towards Data Science](https://towardsdatascience.com/mars-multivariate-adaptive-regression-splines-how-to-improve-on-linear-regression-e1e7a63c5eae)

1. Dataset

<https://www.macrotrends.net/countries/IND/india/gdp-gross-domestic-product>

<https://www.kaggle.com/datasets/josebinjose/indias-gdp-and-per-capita-from-1961?resource=download>

1. PyEarth Installation

<https://stackoverflow.com/questions/66503039/how-to-install-pyearth-in-google-colab>

1. Regression types

<https://www.geeksforgeeks.org/types-of-regression-techniques/>

1. MARS Information

<https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_spline>