SRI SIDDHARTHA ACADEMY OF HIGHER EDUCATION
(Declared as Deemed to be University Under Section 3 of the UGC Act, 1956)

AGALAKOTE, TUMAKURU-572107

Report On:

GDP Prediction and Forecasting using Linear Regression Algorithm

Submitted By:
Bhumika R(20CS018)

Under the Guidance of:

SINDHU T N.
MADHUMALA G.

In partial fulfilment of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SRI SIDDHARTHA INSTITUTE OF TECHNOLOGY
(A Constituent College of Sri Siddhartha Academy of Higher Education)

Maralur, Tumkur-572 105

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
SSIT, TUMKUR

LINEAR REGRESSION ALGORITHM:

The implementation of this can be done by using Linear Regression. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

It is a supervised machine learning algorithm that performs a regression task. Basically, it is the mathematical model that analyses the linear relationship between a dependent variable with a given set of independent variables. In the project we will use simple linear regression to predict the individual attribute of the dataset. For this 80% of the dataset was the training dataset i.e., used for training the model and the remaining 20% was used to test the dataset.

WORKING:

Gross domestic product is the monetary value of all finished goods and services made within a country during a specific period. We are going to calculate the Gross domestic product for the upcoming years. For doing this we are using technologies like pandas, matplotlib, Scikit-learn and The Jupyter Notebook. Pandas is a software library written for the Python programming language for data manipulation and analysis. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.Scikit learn is a free software machine learning library for the Python programming language. The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. We are going to predict the Gross domestic product for the future years by using the technologies like pandas, matplotlib, Scikit-learn and The Jupyter Notebook. The implementation of this can be done by using Linear Regression. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

The Gross Domestic Product is calculated for the future years by using regression coefficient. The regression coefficients are a statically measure which is used to measure the average functional relationship between variables. In regression analysis, one variable is dependent and other is independent. . Also we can calculate the Gross Domestic Product for upcoming years using other algorithms also. We explored all the supervised regression models in order to get the best fitting models. We have trained the model using Linear Regression, Random Forest and Gradient Boosting machine learning algorithms and also estimated the performance of these models.

CODE

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

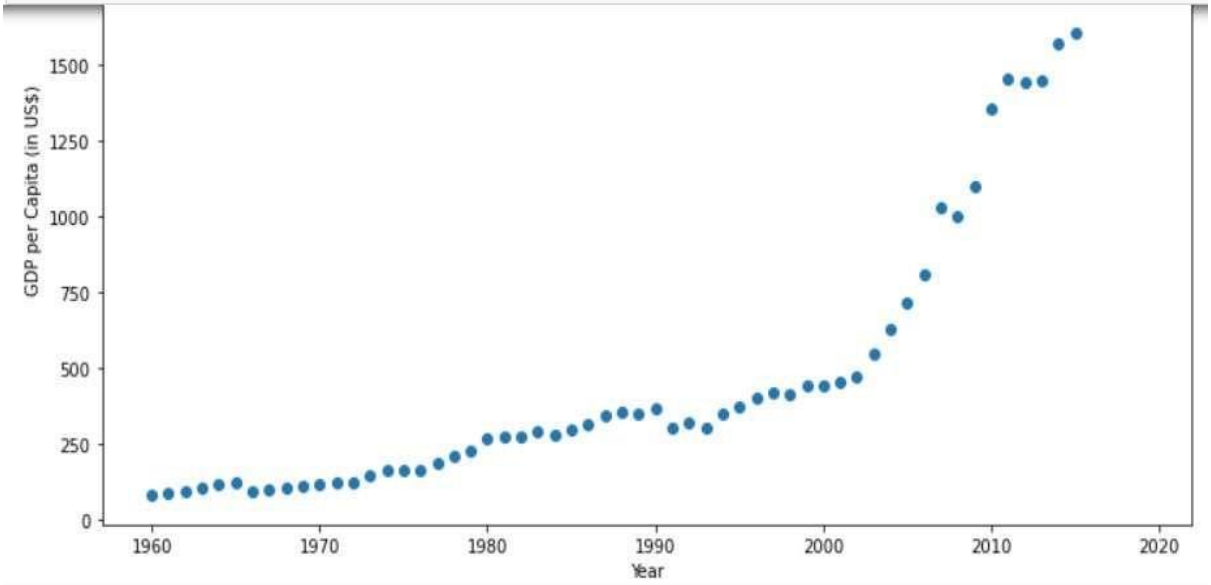
```
Matplotlib is building the font cache; this may
```

```
df = pd.read_csv("GDP_per_Capita_India.csv")
df
```

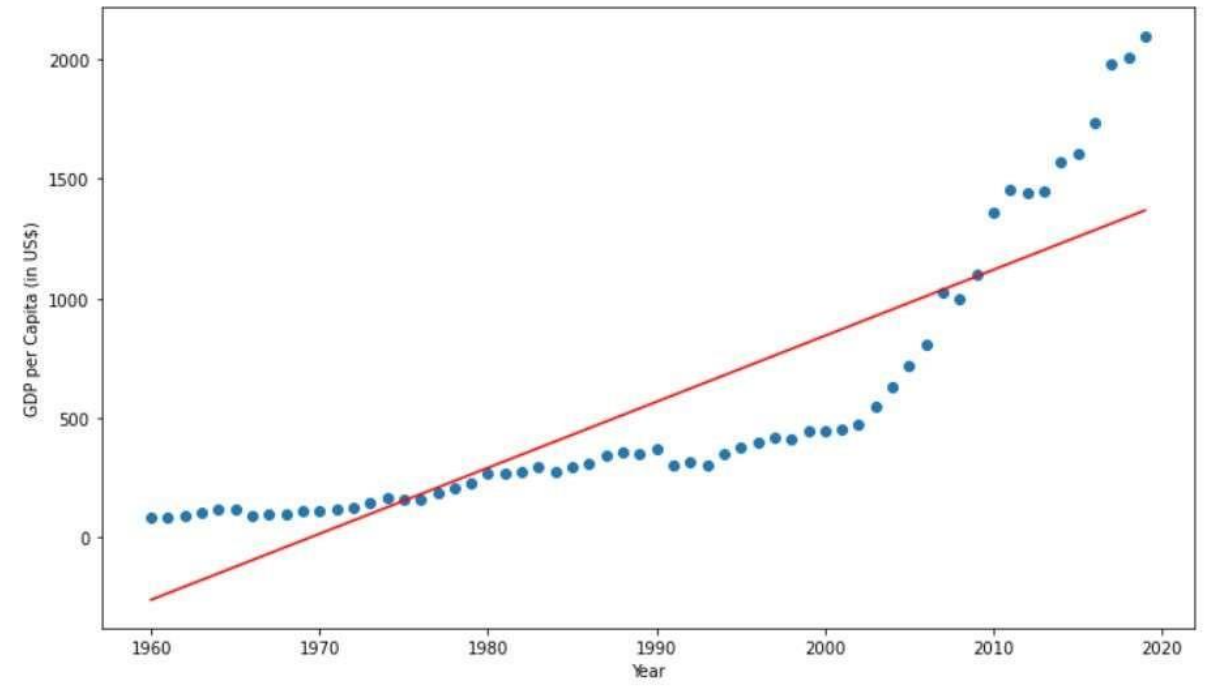
```
df.describe()
```

	Year	GDP_per_Capita
count	60.000000	60.000000
mean	1989.500000	553.144349
std	17.464249	561.886853
min	1960.000000	82.188602
25%	1974.750000	160.328112
50%	1989.500000	328.685380
75%	2004.250000	649.545939
max	2019.000000	2099.599048

```
plt.figure(figsize=(12,7))
plt.scatter(df.Year, df.GDP_per_Capita)
plt.xlabel("Year")
plt.ylabel("GDP per Capita (in US$)")
plt.show()
```




```
plt.figure(figsize=(12,7))
plt.scatter(df.Year, df.GDP_per_Capita)
plt.xlabel("Year")
plt.ylabel("GDP per Capita (in US$)")
plt.plot(df["Year"], reg.predict(df[["Year"]]), color='red')
plt.show()
```



```
pd.DataFrame(data={"Years" : [2020,2021,2022,2023,2024], "Predict_GDP": Predict_GDP})
```

	Years	Predict_GDP
0	2020	1395.675300
1	2021	1423.299266
2	2022	1450.923232
3	2023	1478.547197
4	2024	1506.171163

```
df = pd.read_csv("GDP_per_Capita_India.csv")
df
```

	Year	GDP_per_Capita
0	1960	82.188602
1	1961	85.354301
2	1962	89.881757
3	1963	101.126428
4	1964	115.537497
5	1965	119.318917

Results

True GDP per capita was plotted against the prediction in order to evaluate the model using linear regression. Fig. shows the depiction of linear regression model for true GDP per capita prediction. True GDP per capita was plotted against the prediction in order to evaluate the model using linear regression.