VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANASANGAMA" BELAGAVI - 590 018

KARNATAKA

REPORT OF INTERNSHIP/PROFESSIONAL PRACTICE

Carried out in

TEQUED LABS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

MAHESH A C
[1CG17CS049]

INTERNAL GUIDE

Mr. Suhas K C, M.Tech
Assistant Professor
Dept. of, CSE
C.I.T, Gubbi, Tumkur.

EXTERNAL GUIDE

Mr. Adithya S K
Director
Tequed Labs Pvt Ltd
Banashankari, Banglore

HOD

Dr. Shantala C P, Ph.D
Professor & Head,
Dept. of CSE
C.I.T, Gubbi

Channabasaveshwara Institute of Technology
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

Channabasaveshwara Institute of Technolog
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

UNDERTAKING

I, **MAHESH A C** bearing **1CG17CS049,** student of VIII Semester B.E. in CSE C.I.T, GUBBI, TUMKUR hereby declare that the Internship carried out in **Tequed Labs (P) Ltd, Banashankari, Banglore** and submitted in partial fulfillment of the requirements for the award of the degree Bachelor of Engineering in Computer Science & Engineering of the Visvesvaraya Technological University, Belagavi during the academic year 2020-2021.

Place: GUBBI **MAHESH A C**

Date: **[1CG17CS049]**

Channabasaveshwara Institute of Technology
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-21

BONAFIDE CERTIFICATE

This is to certify that the Internship carried out in **Tequed Labs (P) Ltd, Banashankari, Banglore** is a bonafide work of **MAHESH A C - [1CG17CS049],** student of **VIII** semester **B.E.-** CSE from **Channabasaveshwara Institute of Technology, Gubbi, Tumkur,** in partial fulfillment of the requirements for the award of degree **B.E.,** in **COMPUTER SCIENCE & ENGINEERING** of **Visvesvaraya Technological University**, **Belgaum** during the academic year 2020-2021. It is certified that the Internship work carried out was under my supervision and guidance.

Guide

Mr. Suhas K C, M.Tech
Assistant Professor
Dept., of CSE
C.I.T, Gubbi.

Channabasaveshwara Institute of Technology

(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

CERTIFICATE

This is to certify that the internship entitled "Population Prediction using Machine Learning" has been carried out by **MAHESH A C – [1CG17CS049]** bonafide student of **CHANNABASAVESHWARA INSTITUTE OF TECHNOLOGY, GUBBI, TUMKUR**, in partial fulfillment of the requirement for the award of the degree **Bachelor of Engineering** in **COMPUTER SCIENCE & ENGINEERING** from the **Visvesvaraya Technological University, Belagavi** during the year **2020-2021.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report. The Internship report has been approved as it satisfies the academic requirements in respect of Internship/Professional practice prescribed for the said degree.

Signature of Guide

Mr. Suhas K C, M.Tech
Assistant Professor
Dept., of CSE
C.I.T, Gubbi.

Signature of HOD

Dr. Shantala C P, Ph.D
Professor & Head,
Dept., of CSE
C.I.T, Gubbi.

Signature of Principal

Dr. SURESH D S, Ph.D
Director & Principal
C.I.T, Gubbi.

External Viva

Examiners Name	Signature with Date
1. _____	_____
2. _____	_____

ACKNOWLEDGEMENT

Several special people have contributed significantly to this effort. First of all, I am grateful to my institution, **Channabasaveshwara Institute of Technology, Gubbi**, which provides me an opportunity in fulfilling my most cherished desire of reaching my goal.

I, acknowledge and express my sincere thanks to our beloved Director & Principal, **Dr. Suresh D S** Ph.D., for his many valuable suggestion and continued encouragement by supporting me in mt academic endeavors.

I, express my sincere gratitude to **Dr. Shantala C P,** Ph.D **Professor and Head, Department of CSE** for providing her constructive criticisms and suggestions.

I, extend my gratitude to my Internship guide **Mr. Suhas K C** M.Tech., **Assistant Professor, Department of CSE**, for her guidance, support and suggestions throughout the period of this Internship.

I express my deep sense of gratitude to **Tequed Labs (P) Ltd, Banashankari, Banglore** for giving such an opportunity to carry out the internship in their esteemed industry/organization.

I sincerely thank **Mr. Adithya S K, Director, Tequed Labs (P) Ltd, Banashankari, Banglore** for exemplary guidance and supervision.

Finally, I would like to thank all the individuals who supported me directly and indirectly for the successful completion of this internship work.

MAHESH A C
[1CG17CS049]

TEQUED LABS

INVENT - INNOVATE - ITERATE

RESEARCH AND INNOVATION HUB

No 10 Anjaneya Nagar Banashankari 3rd Stage Bangalore 85

CERTIFICATE ID: TQL2020AIS101

CERTIFICATE
OF
COMPLETION

IS PROUDLY PRESENTED TO

MAHESH AC

FOR SUCCESSFULLY COMPLETING THE
ONLINE SKILL DEVELOPMENT PROGRAM ON
ARTIFICIAL INTELLIGENCE & MACHINE LEARNING
CONDUCTED BY TEQUED LABS
FROM 20-09-2020 TO 20-10-2020

Aditya. S. K.

DIRECTOR
TEQUED LABS

CEO
TEQUED LABS

ABSTRACT

Population may be considered positive hindrance in the way of economic development of the country. But too much population is not good for economic growth. In India, the current population is about 1.36 billion which is second most populated country in the world.

In this project, by using previous population data we can represent it the form of graphs which is easy for analysis and we can make use of that graphs in different business statistics. By improvising this with the help of machine learning algorithms [Linear Regression], and also with python libraries we can predict future population using current datasets. It is also possible to represent the future predicted values in the form of graphs so that it will be easy to understand and these predicted values will be used in different business operations.

The previous population data is in the form of .csv [comma separated values] file. By using the csv file, we can import it using python libraries i.e., pandas, matplotlib etc. By using scikit learn library importing the linear regression algorithm and develop a model for predictive analysis. Then model will give the predictive result in both numerical as well as graphical representation.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Population may be considered positive hindrance in the way of economic development of a country. But too much population is not good for economic development. In this evolving world, Population growth is the primary source of environmental damage in different aspects. Now a days use of human resource has been reduced in many fields. So increase in population leads to population explosion.

Current Population of India - According to latest United Nations data, India, with 1,366,417,754(1.36billion) people is the second most populous country in the world, while China is on the top with over 1,381,433,578 (1.38billion) people. The India represents almost 17.7% of the world's population, which means one out of six people on this planet live in India. Although, the crown of the world's most populous country is on China's head for decades, India is all set to take the numerous uno position by 2027. India is predicted to have more than (1.5022) billion people by the end of 2027.

To overcome this, by using machine learning algorithms and supported libraries we can predict the future population and make use that data for future analysis to reduce population growth.

1.1 Objectives:

The main objective of this project is to predict the future population and make use of it in any other business intelligence. With the help of this prediction we can easily find out the change in population for each year.

This data can be used in different business statistical analysis and using graphs for that data gives pictorial understanding of requirement analysis.

1.2 Problem Statement:

In this project the user can analyse the data with help of graphs and gets predicted population strength, using some previous years data and some machine learning algorithms, libraries, graphs, python.

1.3 Scope of the Project:

The aim of the project is to help in business statistics where data analysis plays a major role, based on the requirement user can make use of it. This project gives graphical analysis of the data which is easy to understand and get algorithm predicted output [population] with graphical analysis using machine learning algorithm and python.

CHAPTER 2

LITERATURE SURVEY

The India represents almost 17.7% of the world's population, which means one out of six people on this planet live in India. Although, the crown of the world's most populous country is on China's head for decades, India is all set to take the numerous uno position by 2027. India is predicted to have more than (1.5022) billion people by the end of 2027. During 1975–2010, the population doubled to 1.2 billion. The Indian population reached the billion mark in 1998. India is projected to be the world's most populous country by 2024, surpassing the population of China. It is expected INDIAN POPULATION to become the first country to be home to more than 1.5 billion people by 2030, and its population is set to reach 1.7 billion by 2050. Its population growth rate is 1.13%, ranking 112th in the world in 2017.

- **Some of the major population problems of India are as follows:**
  - Rapid Growth of Population
  - Disproportionate Gender Composition
  - Poor Standard of Living and Malnutrition
  - Unemployment

After getting information of population we need to analyse using the data of population and make use of it in different fields. Data will help us to give actual picture of population for analysis.

CHAPTER 3

TRAINING

In 30 days of internship, we learnt the need of Artificial Intelligence and machine learning, basics of python and its libraries, machine learning algorithm to develop the project i.e., population prediction.

WEEK 1:

In starting day of our internship, we are addressed by Mr. Aditya S K (director of the company). Later our guide has given some introduction about Artificial Intelligence, then basics of python.

WEEK 2:

On the second week, we got usage of essential python libraries like numpy, pandas, matplotlib, etc. and different forms of data. Later we learnt some machine learning algorithm.

WEEK 3:

On the third week, our guide as assigned one project with the problem statement that, make use population data of any country and try to analyse the data later create a model using ‘Linear Regression’ and get some predictive outcome out of it.

WEEK 4:

By the end of week, we are successfully completed the assigned project as per given problem statement. We got more accurate predicted values from the model with graphical representation which easy to understand for various business activities.

CHAPTER 4

SYSTEM ANALYSIS AND DESIGN

System design is the process of defining the architecture, components, modules, interfaces and data for a system to satisfy specified requirements. One could see it as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering. If the broader topic of product development blends the perspective of marketing, design, and manufacturing into a single approach to product development, then design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user.

4.1 EXISTING SYSTEM:

In existing system, by using simple python libraries like NumPy, pandas, matplotlib we can represent the data in the form of pictorial representation. Otherwise, we can't able to get the future data to access the data for future references.

Advantages:

- It gives pictorial representation of data.

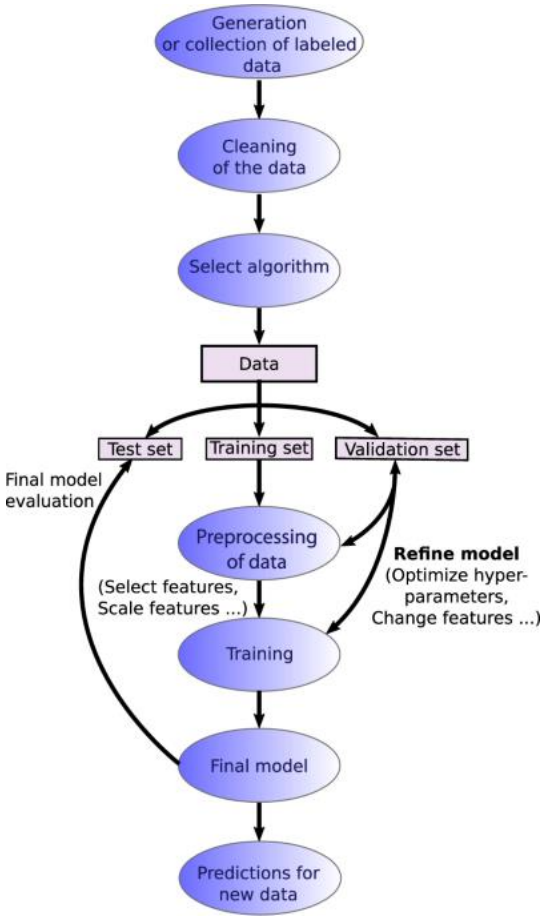
Disadvantages:

- It doesn't give any predictive data or future prediction.

4.2 PROPOSED SYSTEM:

It is an enhanced version of existing system, i.e., by using machine learning algorithms under scikit-learn library of python we can able to create a machine learning model which is capable of predicting the future values using given dataset with very good accuracy. We can also represent the predictive data using graphs for easy analysis.

4.3 ARCHITECTURE:



4.3 Architecture of system design

4.4 LIBRARIES USED:

- NumPy (for Numerical Analysis).
- Pandas (for handling data files).
- Matplotlib (for visualizations inline & figure settings).
- Scikit Learn (for model building & data pre-processing).

4.5 DATASET:

For this project, we are using the dataset of Indian population since 1960 to 2019, available on git repository.

You can find data on the link as follows:

https://github.com/acmahesh10/population-dataset

CHAPTER 5

IMPLEMENTATION

To be used efficiently, all computer software needs certain hardware components or other software resources to be present on a computer. These prerequisites are known as (computer) system requirements and are often used as a guideline as opposed to an absolute rule. Most software defines two sets of system requirements: minimum and recommended. With increasing demand for higher processing power and resources in newer versions of software, system requirements tend to increase over time. Industry analysts suggest that this trend plays a bigger part in driving upgrades to existing computer systems than technological advancements.

5.1 TECHNOLOGY USED:

MACHINE LEARNING:

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data.[1] It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

LINEAR REGRESSION:

Linear Regression attempt to model the relationship between two variables by fitting a linear equation to observed data. The other is considered to be dependent variable. For Example: A modeller might want to relate weights of individuals to their heights using a linear regression model.

Linear regression is useful for finding relationship between multiple continuous variables. There are multiple independent variables and single independent variable.

$y = m1X1+m2X2+\ldots\ldots+b$

Where,

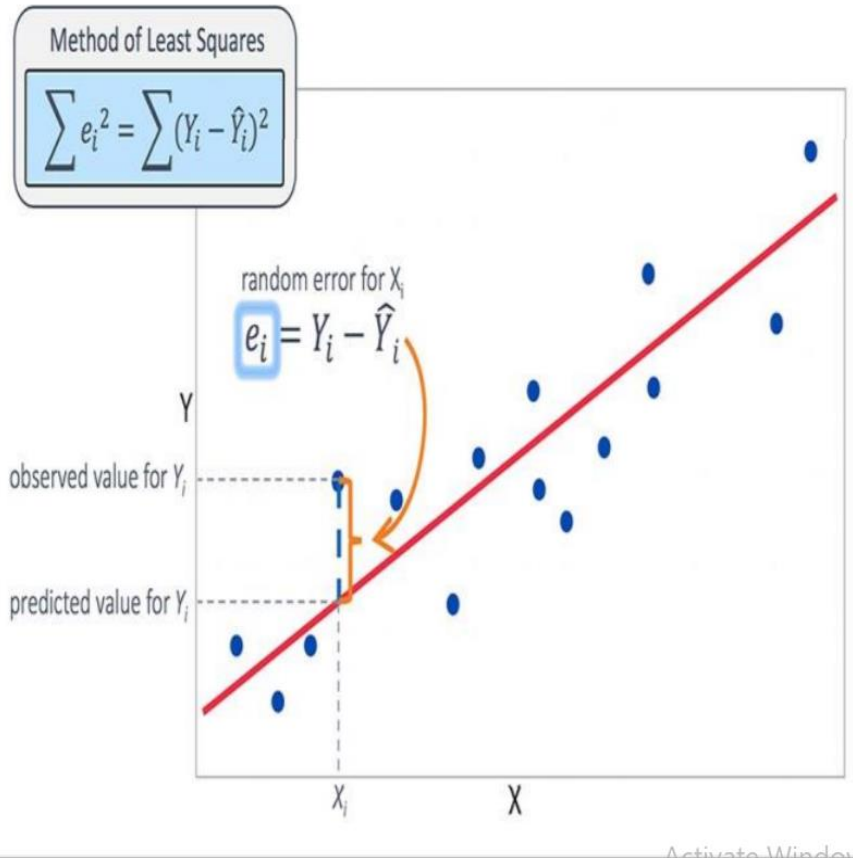
m1, m2, m3 …. →slope.

b → y intercept.

X1, X2, X3 …… →independent variables.

y → dependent variables.

As the function is a regression model, score function will help us find the accuracy of our model. Our model will be much accurate when the score is nearer to 1.0



Linear Regression

5.2 MODEL BUILDING:

First, import all the libraries/packages which are necessary to analyse the given dataset about Indian population.

```
1   # Required python packages
2   import pandas as pd
3   import numpy as np
4   import matplotlib.pyplot as plt
```

Next step is to import the given dataset about the Indian population, as the dataset is too large to analyse, here only the head part of the dataset is about the Indian population, as the dataset is too large to analyse, here only the head part of the dataset is extracted for better analysation of data.

```
: dataset.head()
```

:

	years	pop_india
0	1960	450547679
1	1961	459642165
2	1962	469077190
3	1963	478825608
4	1964	488848135

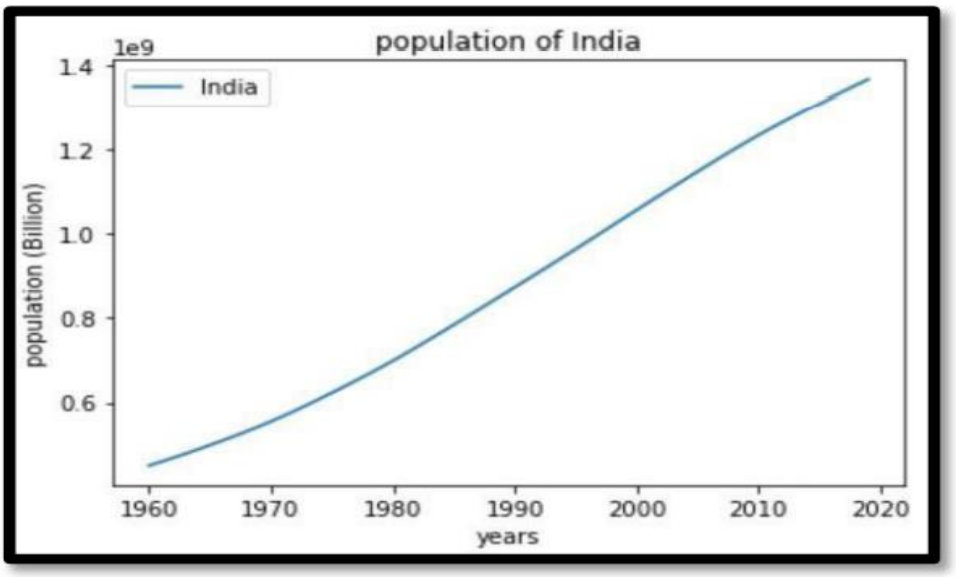
Now to we can also know, what type of data is being used

```
: dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   years       60 non-null     int64
 1   pop_india   60 non-null     int64
dtypes: int64(2)
memory usage: 1.0 KB
```

By using matplotlib library lets plot the graph against year and population, here is the code

```
: # Select the particular rows & column
X = dataset.iloc[:, :-1].values
Y = dataset.iloc[:, 1].values

# Data visualisation by ploting graph
plt.plot(X, Y, label="India")
plt.title('population of India')
plt.xlabel('years')
plt.ylabel('population (Billion)')
plt.legend()
plt.show()
```

The graph is plotted against year and population.



Splitting dataset into training and testing data (training the model), 80% of dataset is going for training and remaining 20% of dataset is going for testing, here we using linear regression algorithm for predicting the population.

```
# splitting dataset into training and testing data (training the model)
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test= train_test_split(X,Y, test_size=0.20,random_state=0)

# choose the prediction model - linear model
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(X_train,Y_train)

y_pred= regressor.predict(X_test)
```

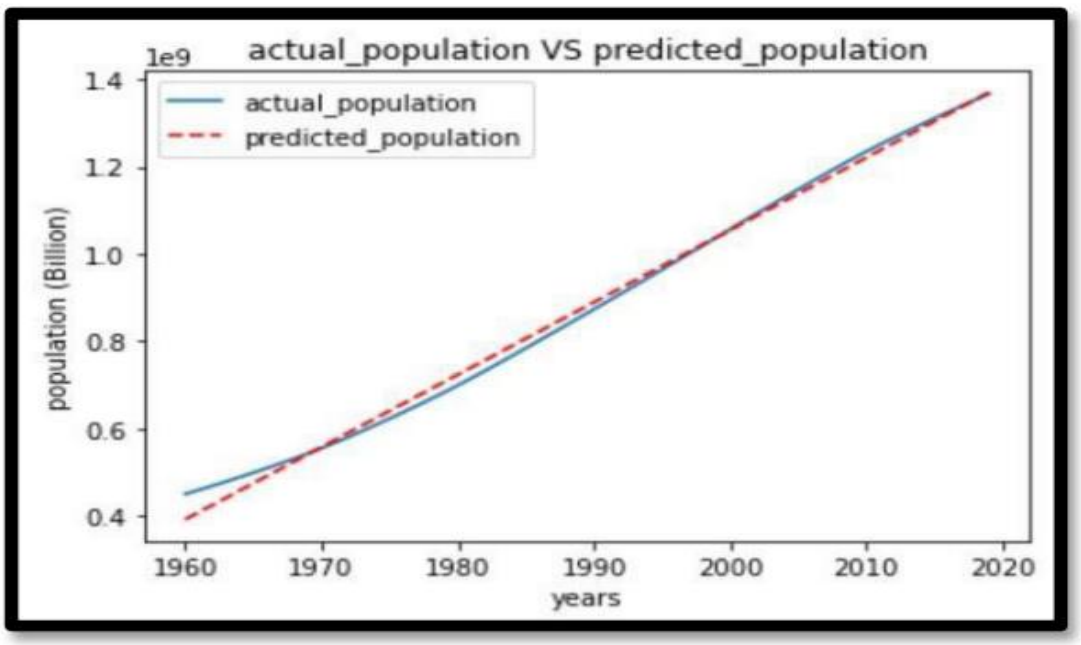
Now the model is ready for predictive analysis.

CHAPTER 6

RESULTS

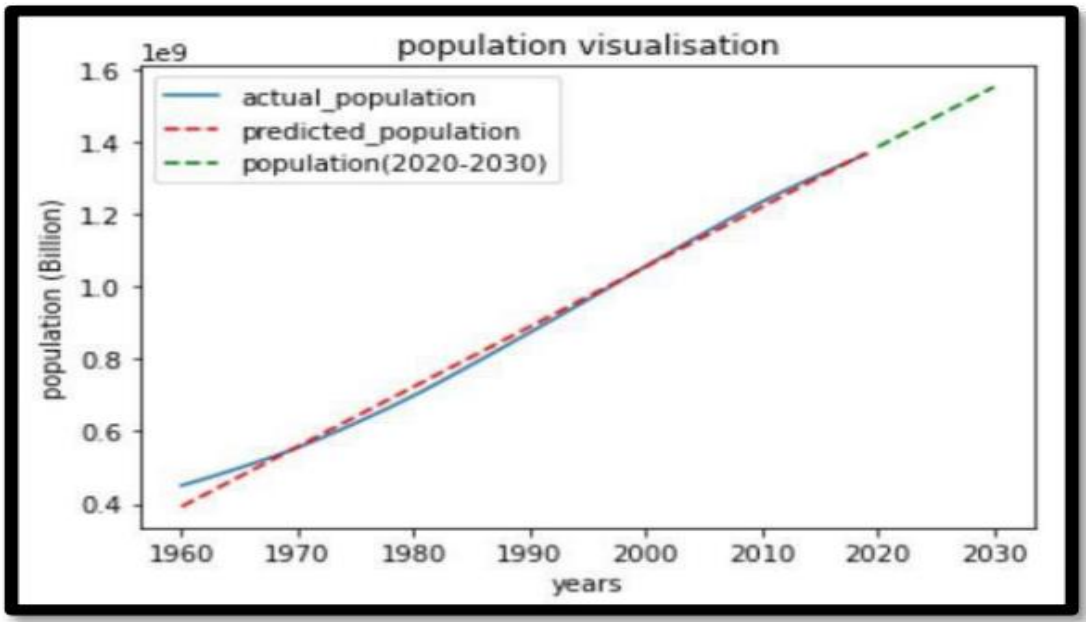
The graphical representation of actual population and predicted population, where population values present in dataset is called actual population and algorithm predicted values is called predicted population.

```
plt.plot(X, Y, label="actual_population")
plt.plot(year, predict, '--r', label="predicted_population")
plt.title('actual_population VS predicted_population')
plt.xlabel('years')
plt.ylabel('population (Billion)')
plt.legend()
plt.show()
```



Here we predicting the next ten years of data using algorithm and representing with the help of graph

```
plt.plot(X, Y, label="actual_population")
plt.plot(year, predict, '--r', label="predicted_population")
plt.plot(year1, predict1, '--g', label="population(2020-2030)")
plt.title('population visualisation')
plt.xlabel('years')
plt.ylabel('population (Billion)')
plt.legend()
plt.show()
```



Here user can get the particular year population by entering the year.

```
#user input : asking user to enter the year
#-------------------------------------------------------------------#
y_pred1 = regressor.predict([[int(input("enter the year : "))]])
print("population = ", y_pred1)
#-------------------------------------------------------------------#
```

```
enter the year : 2030
population =  [1.55196616e+09]
```

CHAPTER 7

CONCLUSION

As we observed in the above result our machine learning model is performing well with more accuracy when comparative actual data. By the help of this model, we can predict the future values. ML code using python we can analyse any data set given and as per the topic chosen about Indian population, the above-mentioned code using python helped a lot to predict the population of a particular year. By using linear regression, we can predict the future population and its accuracy. With the help of analysis, it helps to overcome the population problem in future.

REFERENCES

- https://worldpopulationreview.com/countries
- https://population.un.org/wpp/
- https://numpy.org/doc/stable/
- https://scikit-learn.org/stable/
- https://www.geeksforgeeks.org/machine-learning/
- https://github.com/acmahesh10/population-dataset
- https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86
- https://matplotlib.org/stable/contents.html