Predicting Life Expectancy

Work From:- SMART INTERNZE

Category : Machine Learning

Time Period : - 30 Days

Data :- From (WHO) Life Expectancy with 22 features from 193 countries and 2938 records (2000 - 2015)years data

Tools : - Python,IBM Cloud

Project Submitted by:
Jyothi Prakash Muddana
III year ,CSE
Tirumala Engineering College
Andhra pradesh

1. INTRODUCTION

1.1 Over View

This Project is done under the guidelines from SmartInterze Team.

This helped me to start learning Machine Learning.

We are going to use a huge data and some IBM Cloud Tools to complete our project .

I think it one of My grate Opportunities
In this project we are going to work on Huge data set with python to handle that data

1.2 Purpose

Purpose of this project is to predict life expectancy of a country by using some dependent features .

This will help Government and other Non-Governmental Organizations to take most effective steps to increase the life space of Humans

2. LITERATURE SURVEY

2.1 Existing Problems on Life Expectancy

The most important causes of death in Western industrialized countries are cardio and cerebrovascular diseases and malignancies. For instance, in Germany in 2008, 68.6% of all women and 65.9% of all men died from these diseases

The third most frequent cause of death are respiratory diseases which cause less than 10% of deaths each year Important risk factors for cardio and cerebrovascular

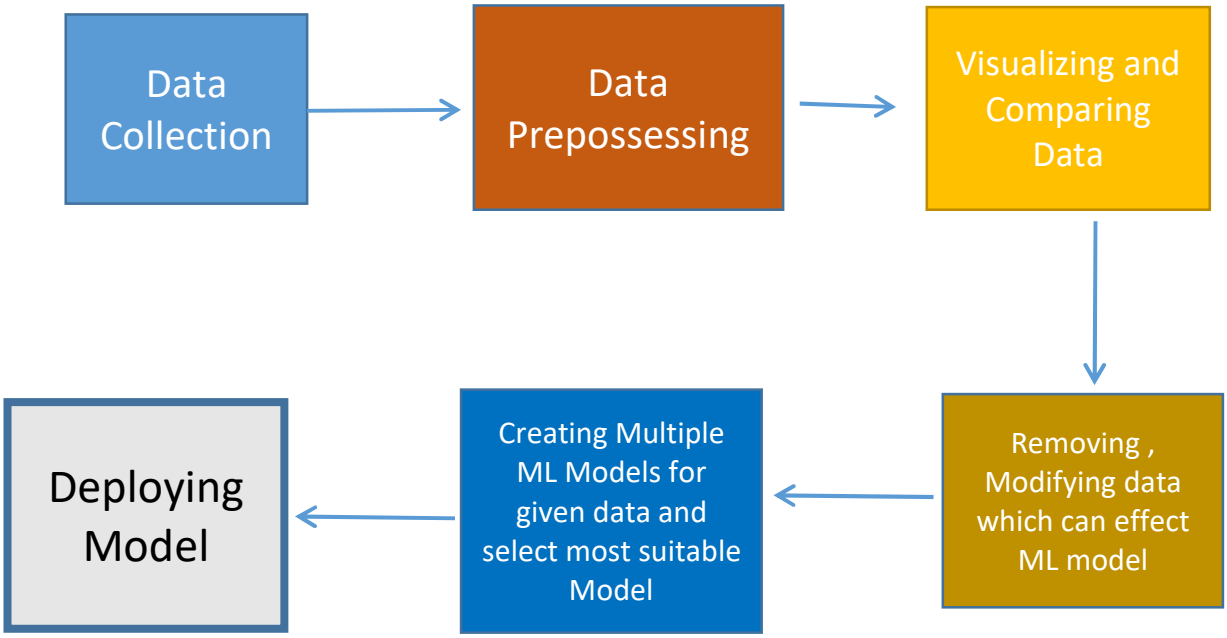
Not only because of diseases even climatic changes,Natural disaster , Human Mistakes in Huge Industries where more people work

2.2 Proposed Solution

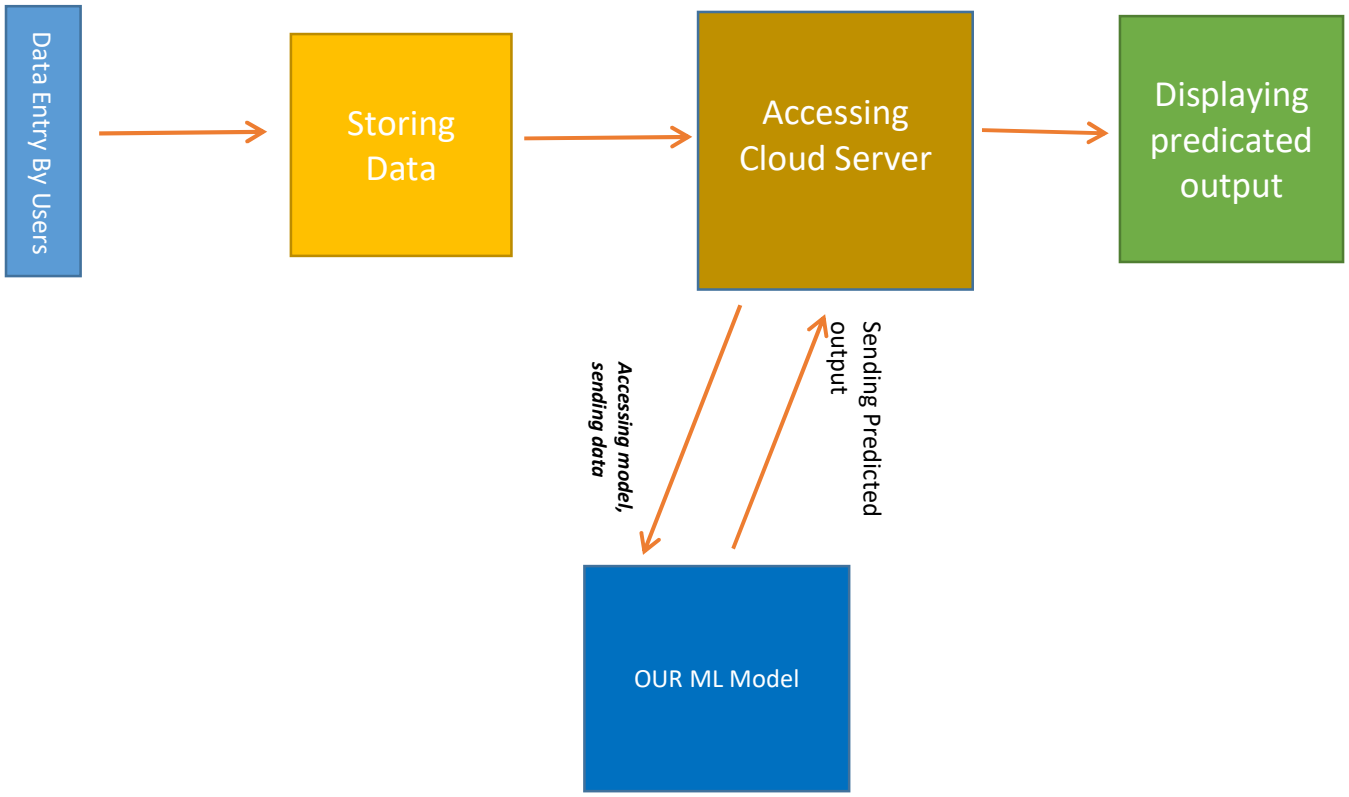
Regular physical activity reduces the risk of and/or improves many diseases and conditions including arterial hypertension, diabetes , obesity, coronary heart disease, chronic heart failure, and chronic obstructive pulmonary disease . In addition, the risk of colon , breast , lung diseases, and pancreatic cancer is reduced, Even there physical Fitness will help them to withstand the natural Disaster and climatic changes

3. Theoretical Analysis

3.1.1 Machine Learning Model



3.1.2 Block Diagram



3.2 Hardware And Software

3.2.1 Hardware

- IBM Cloud Storage

Features

- ✓ 1 COS Service Instance
- ✓ Storage up to 25 GB/mo.
- ✓ Up to 20,000 GET requests/mo.
- ✓ Up to 2,000 PUT requests/mo.
- ✓ Up to Data Retrieval 10 GB/mo.
- ✓ Up to 5GB Public Outbound

- IBM Server

1 authorized user

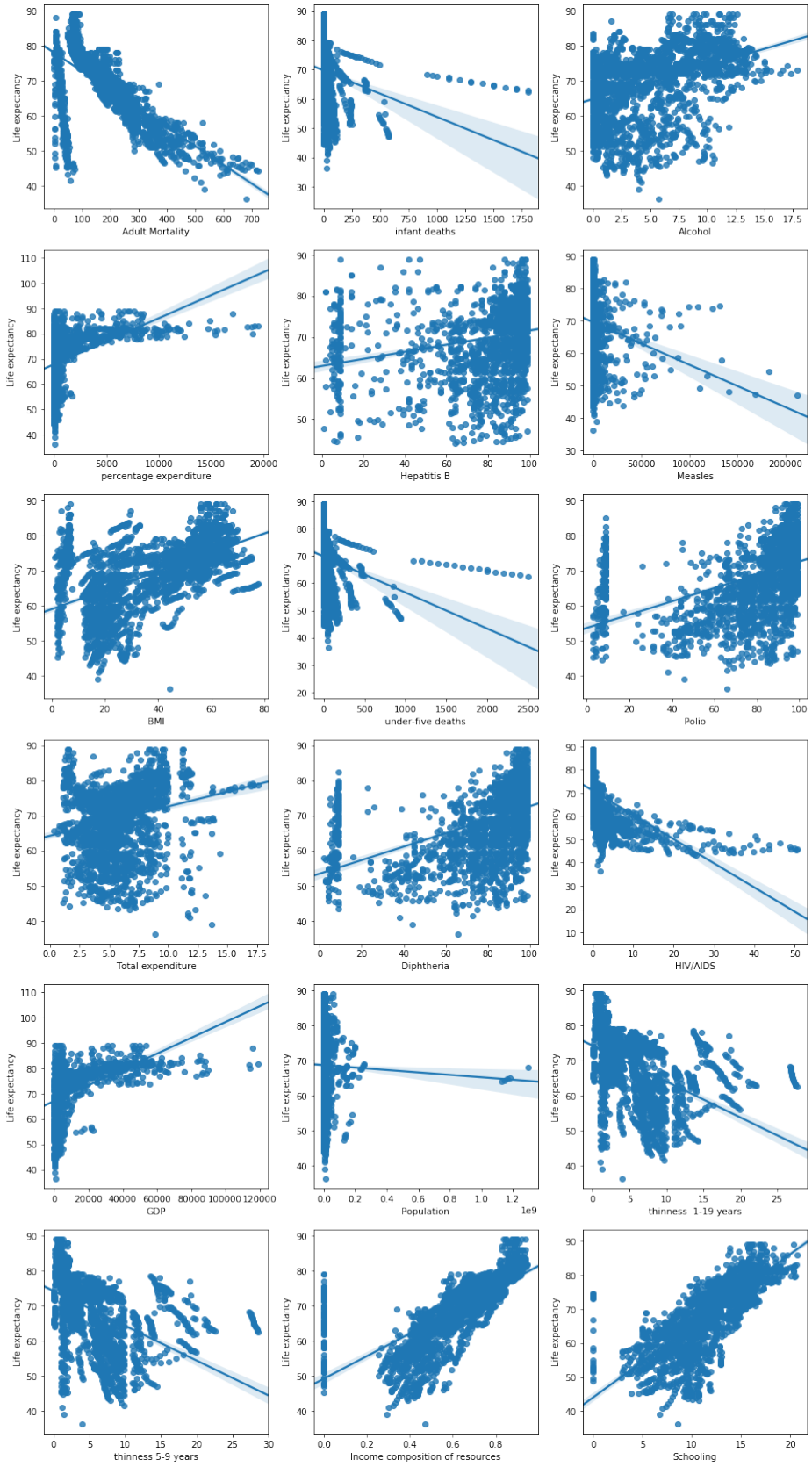
- ✓ 50 capacity unit-hours monthly limit
- ✓ Environment = # of capacity units required per hour
  - 1 vCPU + 4 GB RAM = 0.5
  - 2 vCPU + 8 GB RAM = 1
  - 4 vCPU + 16 GB RAM = 2
- ✓Decision Optimization = Environment + 5

3.2.2 Software from IBM

- IBM Cloud services
- Watson Studio
  To Create our Model using Jupyter Notebook
- Machine Learning service
  Deploying the created Mode
- Node red App
  Creating Dashboard for our model (flow)

4. Experimental Investigations

4.1 Comparing Life Expectancy with other features

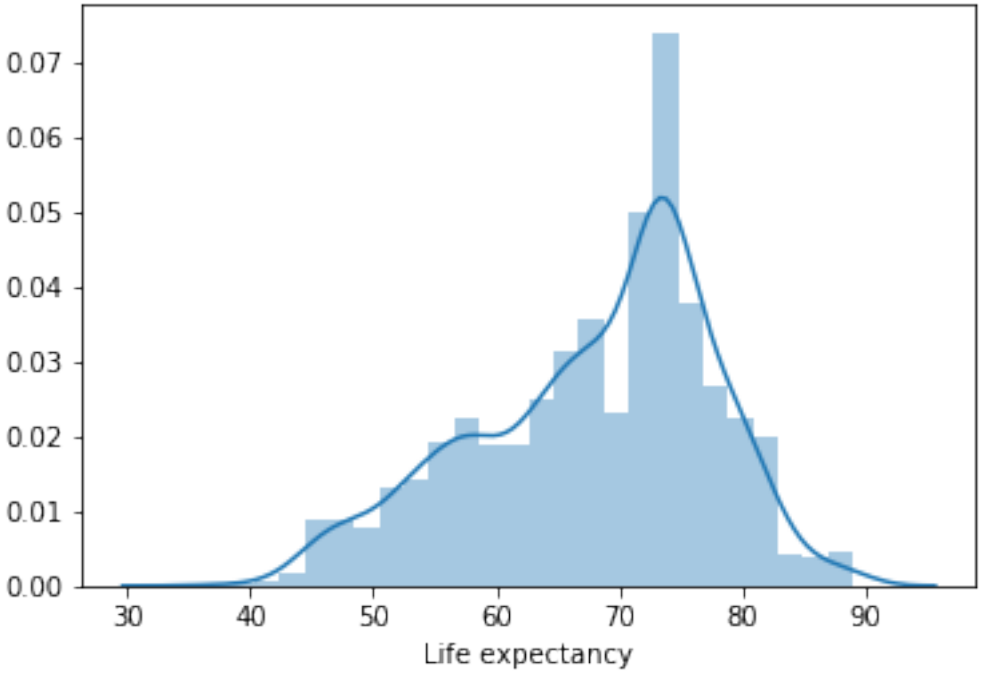


In above gave we can see some regular pattern in between Life Expectancy and other features

Some of the important features based on the above graph are

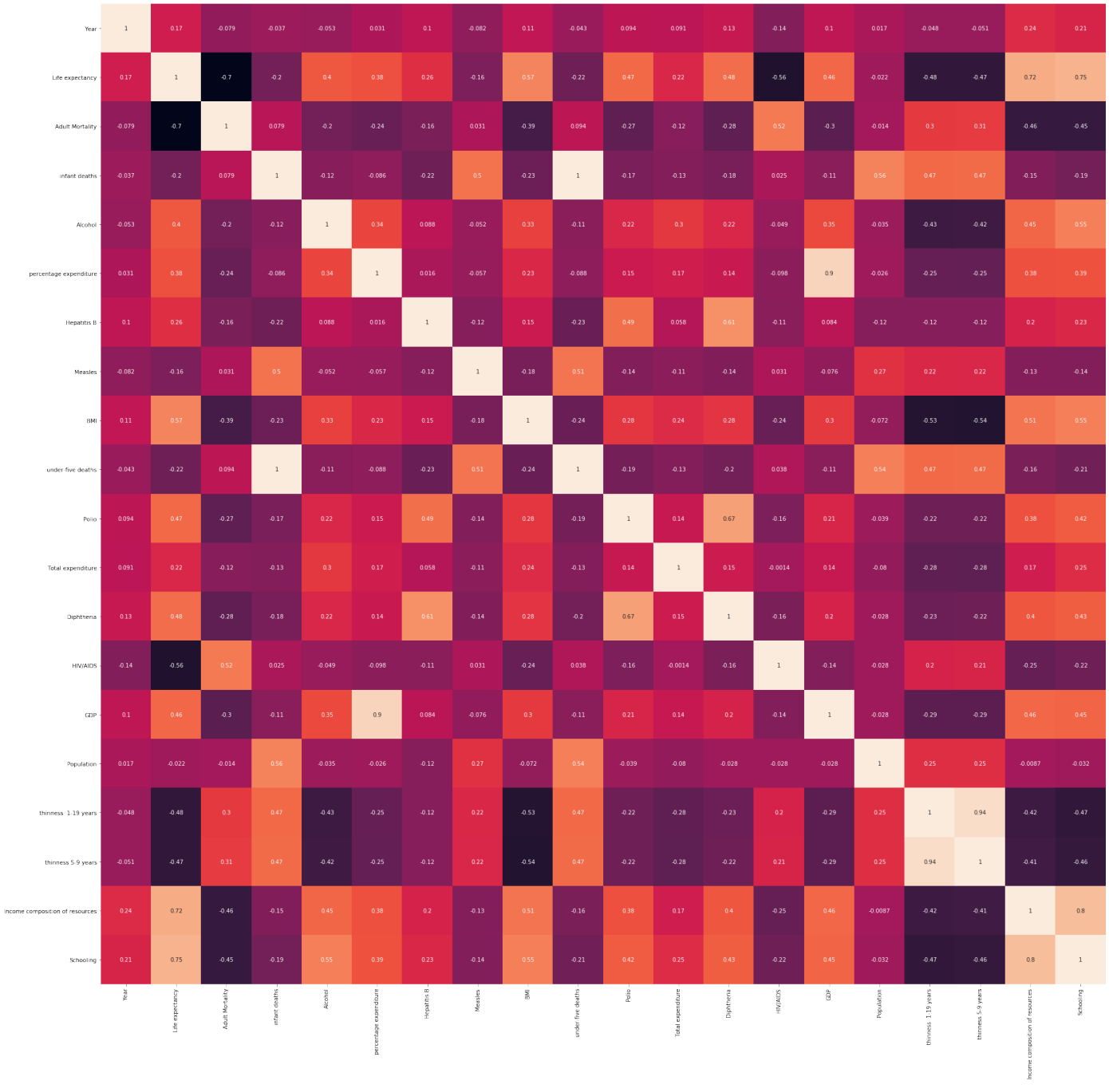
1. Adult Mortality
2. Alcohol in take
3. Schooling
4. Percentage Expenditure
5. HIV/AIDS
6. Income composition of resources …

4.2 Life Expectancy dist Plot



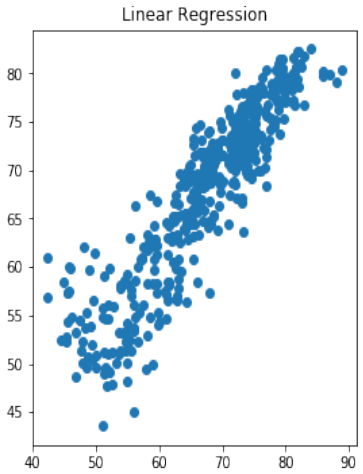
In above graph we can view that most of the countries having Life Expectancy in between 70 -80 years

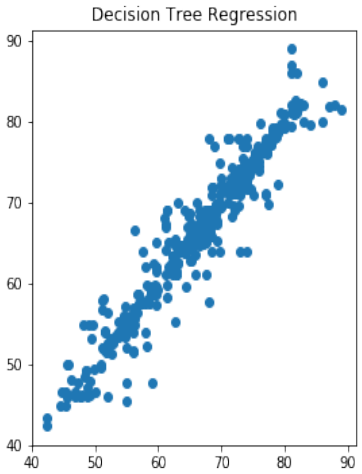
4.3 correlation between attribute visualizing using Heat-map

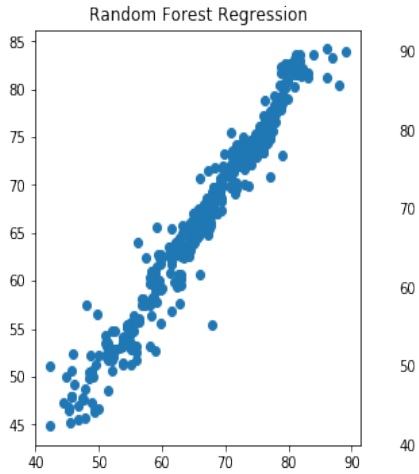


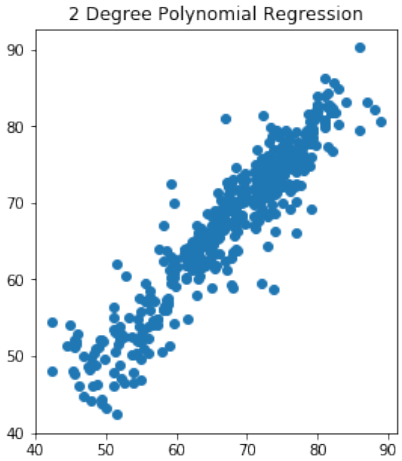
- Correlation to check how strong the attributes are dependent on each other
- In the above graph , light color are show highly dependent attributes and dark are show less dependent attributes

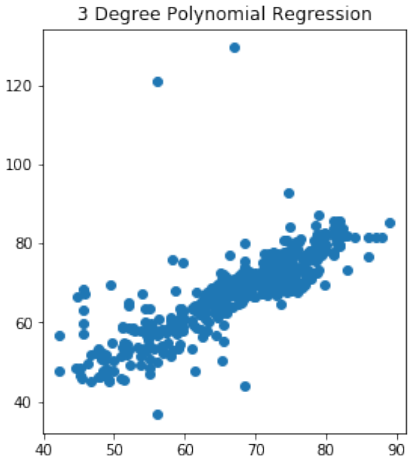
4.4 Visualizing which model is most efficient









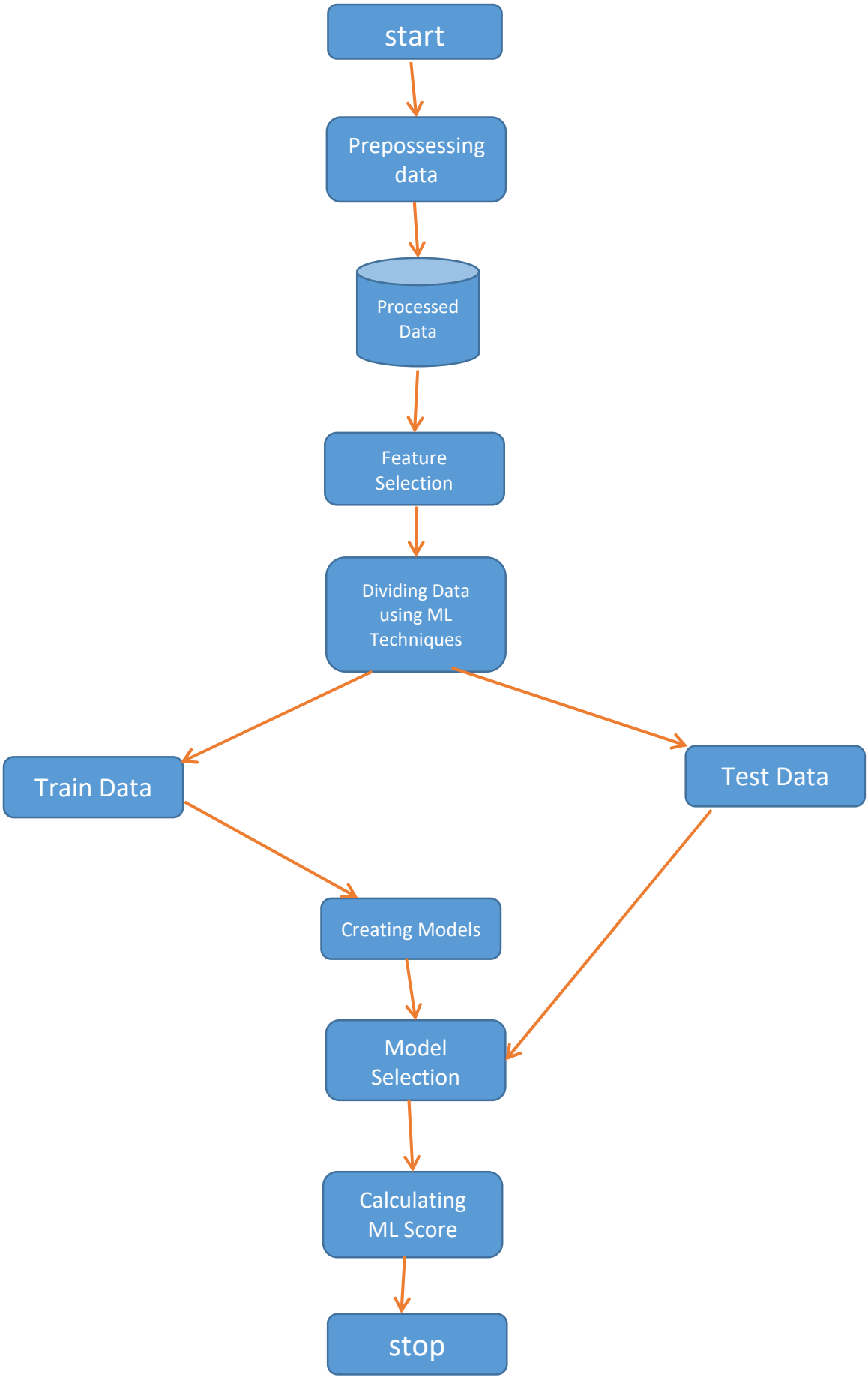


Here I use 5 models

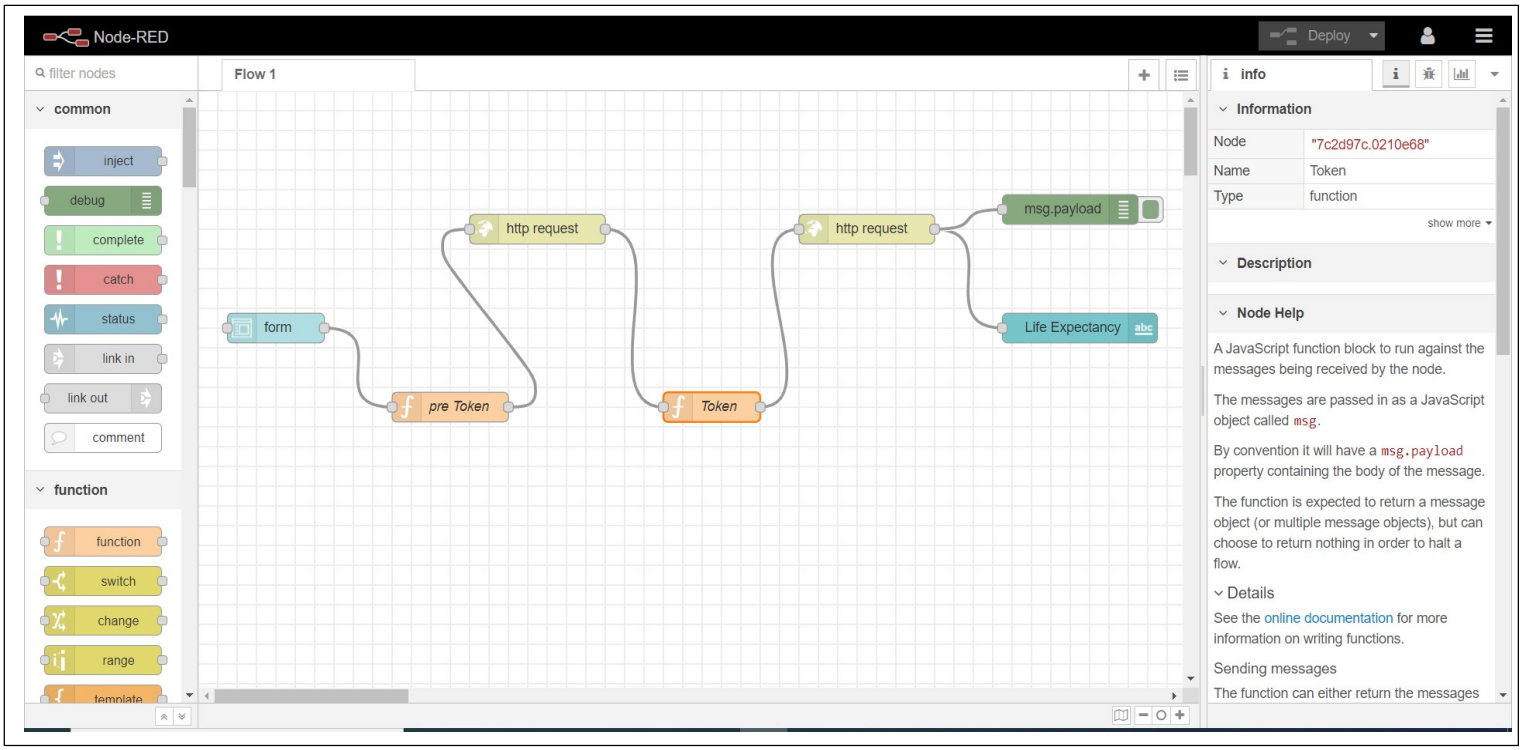
- Linear Regression Model
- Decision Tree Model
- Random Forest Regression Model
- 2-degree Polynomial Regression
- 3-degree Polynomial Regression

✓ From above graph we can see that Random Forest Regression is most effective Model for our data

5. Flow Chart


start
Prepossessing data
Processed Data
Feature Selection
Dividing Data using ML Techniques
Train Data
Test Data
Creating Models
Model Selection
Calculating ML Score
stop

Node - Red (Flow)

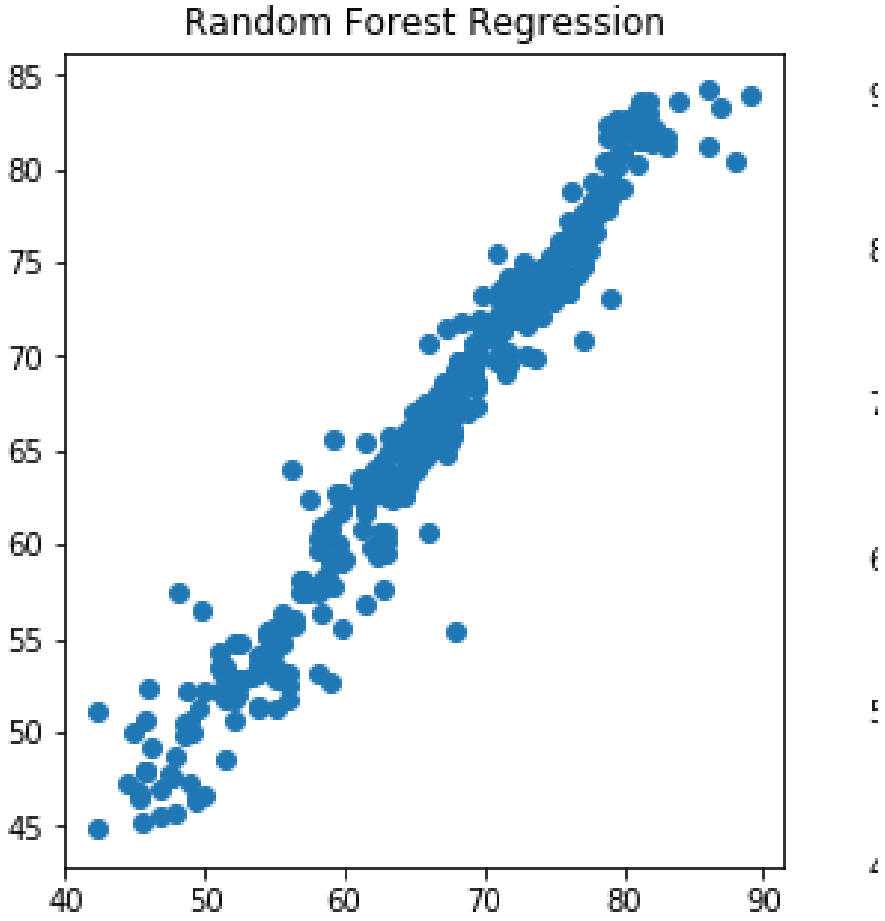


Explanation

- We are using form Node for taking input from user in the form Format
- Pre Token is a function which is used convert local variables in to global and creating a MSG for HTTP request to get access from IBM cloud
- Token function is used to create msg.header from token,instance id and data in the format of json object
- In Second HTTP request we are sending the json object to the model and getting predicted Life Expectancy from it
- Edit text node is used to print the predicted values on dashboard

6. Result

6.1 Model Result



graph between predicted,actual values

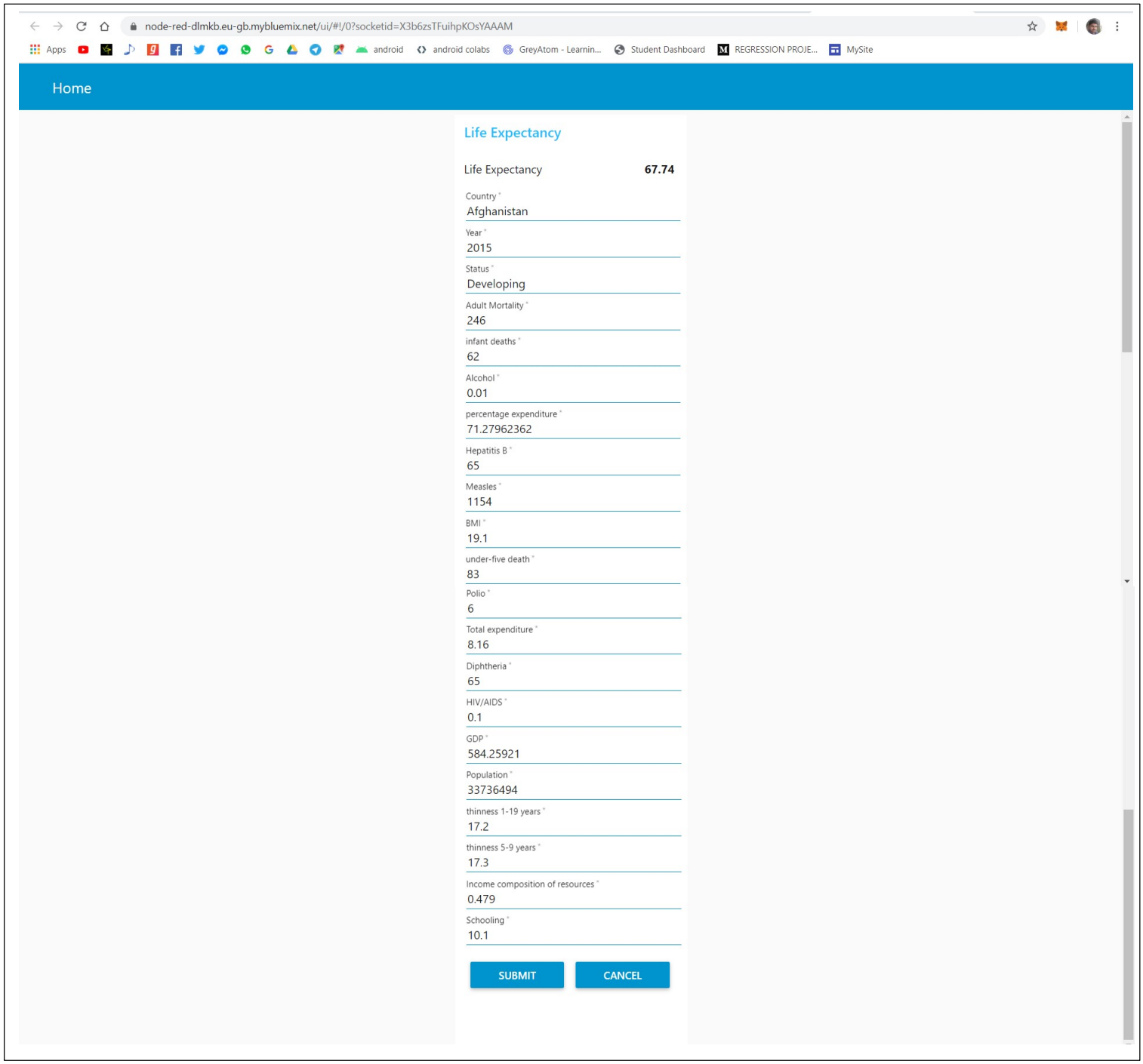
Mean Absolute Error: 1.2867203219315895
Mean Square Error: 3.8028237424547293
Root Mean Square Error: 1.9500830091190295

For This Record

[Afghanistan,2015,Developing,65,263,62,0.01,71.27962362,65,1154,19.1,83,6,8.16,65,0.1,584.25921,33736494,17.2,17.3,0.479,10.1]

We got 66.8 as output but actual is 65 , it is near to the result

6.2 Dashboard



✓ In the above example is the same input but change adult mortality value , by decreasing it the Life Expectancy value in increasing

7. Advantage and Disadvantage

7.1 Advantages

- ✓ Predicting life Expectancy will help us to monitor the health of the people
- ✓ By predicting it we can improve our health condition and public health care center
- ✓ We can compare Things and make decision for the future to increase life span of Humans

7.2 Disadvantages

- ✓ If our predictions are wrong that may lead to wrong judgement for future
- ✓ People may get upset by looking at the result ,If the result are published
- ✓ Some people make decisions by looking the age of the person ,If the person age's exceeds over Life Expectancy then there is a change of not taking care or him or they may take care than required

Both or problematic to that person

8. Applications

- By using this model we can create App
- Life-tables and Demographic applications

This type of thing may increase awareness on there lives

We cam create applications like sharing which type of activities will increase the life Expectancy

9. Conclusion

- Predicting life expectancy will help to improve our living conditions and life style

- Mostly it have Advantages than Disadvantage

- These type of applications will increase awareness on there life

- It is Very clear that some of our habits will effects our life(alcohol)

- Education is also required to increase life span of Humans

10. BIBILOGRAPHY

https://youtu.be/V0u6bxQOUJ8

https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/

https://www.coursera.org/learn/data-analysis-with-python/home/welcome

And Mentors of SmartInterze They really helped me a lot in deploying my ML model,Creating Dashboard for my project

Sincere Thank all the Mentor of SmartInterze