Predicting Life Expectancy Using Machine Learning

(Smart Internz)

Category : Machine Learning

Time Period : 30 Days

Data : WHO Data Sheet from kaggle expectancy with 22 features from 193 countries and 2938 records (2000-2015) years data.

Tools : Python, IBM Watson Studio, IBM Cloud, Node Red App.

Project Submitted By : Navneet Sharma

1. INTRODUCTION

Overview

This project is done under the guidance from Smart-internz Team. This helped me in getting started with Machine Learning. Including this I learned about using Cloud based services in this project because of using IBM Cloud service for the project.

I guess this project will surely going to help in setting up the foundation for my career in ML field.

Purpose

This project is designed to predict Life Expectancy of a country by employing some dependent measures.

This Project will surely help the Government and other non-Govt. organizations to improve life expectancy statistics of their country by taking adequate measures.

International Organizations like WHO, can also take this life expectancy stat into account for instructing the countries about the future consequences.

2. LITERATURE SURVEY

✦ Existing Problems on Life Expectancy

One of the most important causes of death in Western Countries are cardio, cerebrovascular diseases and malignancies.
For ex : In Germany in 2008, out of all the deaths, 68.6% women and 65.9% men died from these diseases.

Around 10% deaths each year are a result of result of respiratory diseases.

Climatic changes, natural disasters, human errors in industries, etc. also account for large number of deaths each year.
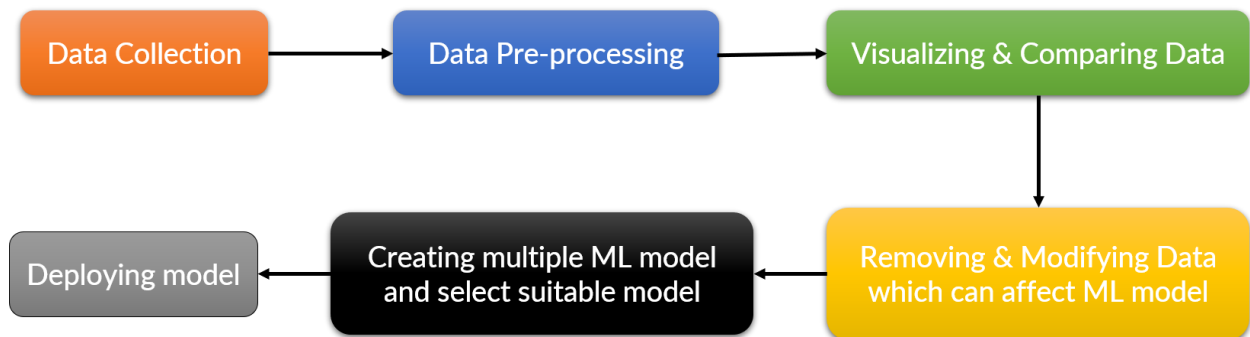
✦ Proposed Solution

Regular physical activities reduces the risk and improves many diseases including arterial hypertension, diabetes , obesity, coronary heart disease, chronic heart failure, and chronic obstructive pulmonary disease .
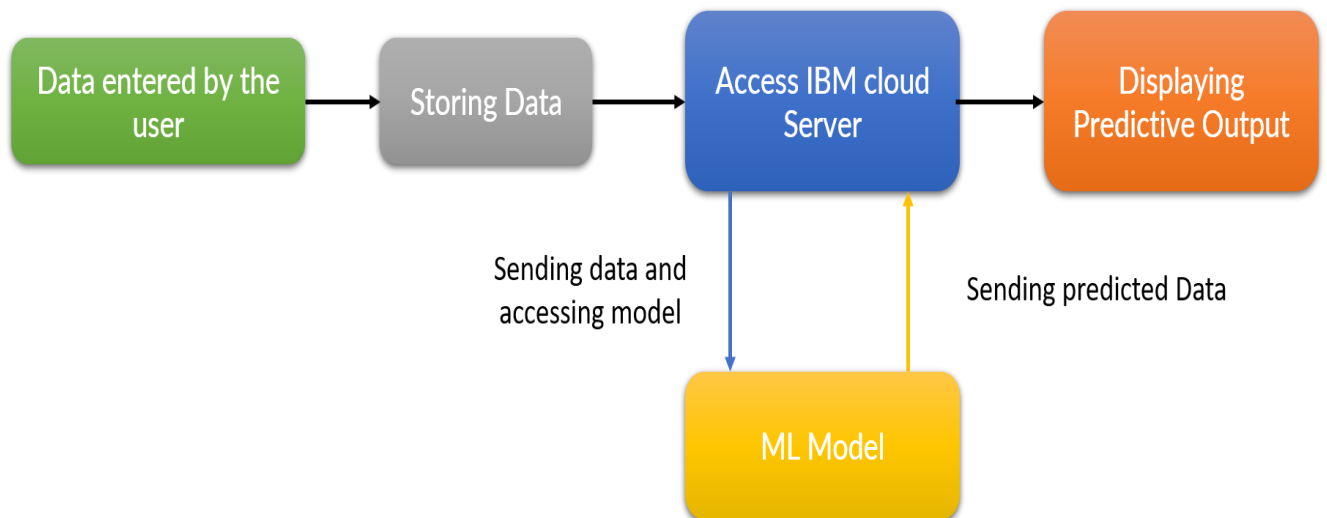
In addition the risk of colon, breast, lung diseases, and pancreatic cancer is reduced. Even there some physical fitness programmes which will help them to withstand natural disaster and climatic changes.

3. THEORETICAL ANALYSIS

Machine Learning Model

Data Collection → Data Pre-processing → Visualizing & Comparing Data

Deploying model ← Creating multiple ML model and select suitable model ← Removing & Modifying Data which can affect ML model

Block Diagram

Data entered by the user → Storing Data → Access IBM cloud Server → Displaying Predictive Output

Sending data and accessing model

Sending predicted Data

ML Model

Software & Hardware

➢ Software from IBM :

- ❖ IBM Cloud Service
- ❖ **Watson Studio** : To create model using Jupyter notebook.
- ❖ **Machine Learning Service** : Deploying the created model.
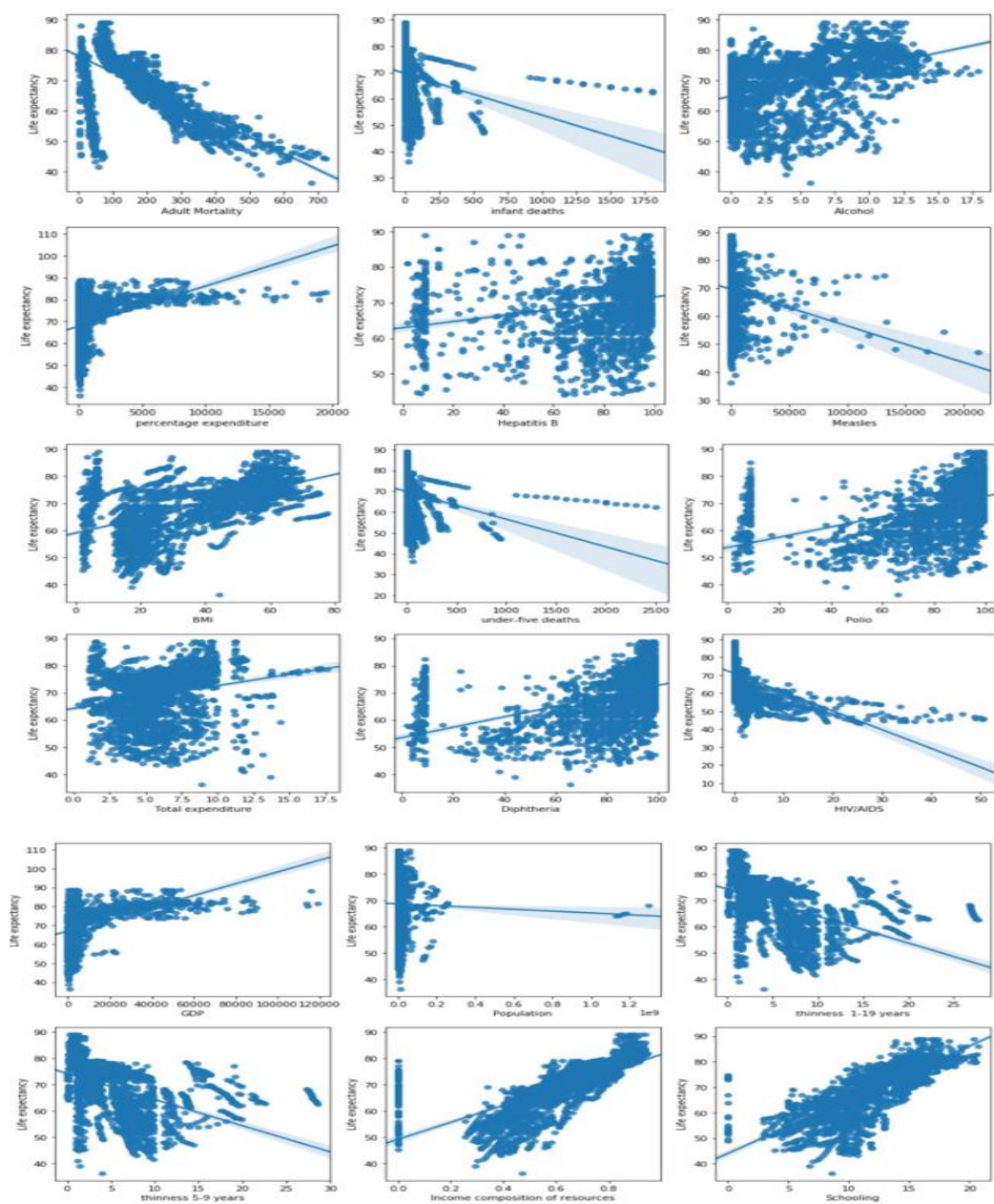- ❖ **Node Red App** : Creating Dashboard for our ML model.

➢ Hardware :

- ❖ **IBM Cloud Storage.**
- ❖ **IBM Cloud Server.**

4. EXPERIMENTAL INVESTIGATION

➕ Comparing Life Expectancy with other features

In above graphs 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 intake
3. Schooling
4. Percentage Expenditure
5. HIV/AIDS
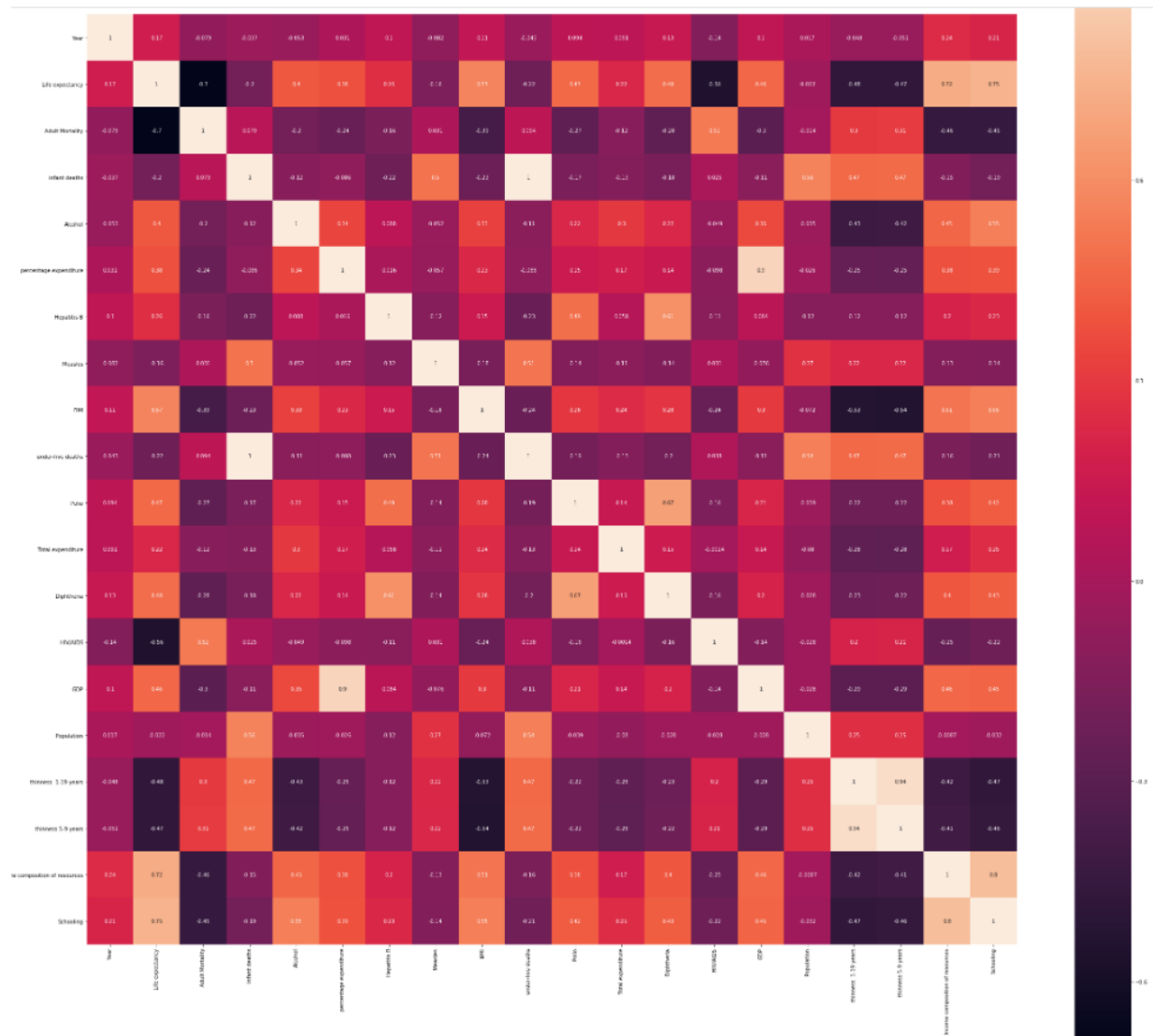6. Income composition of resources ...

✚ Life Expectancy Plot
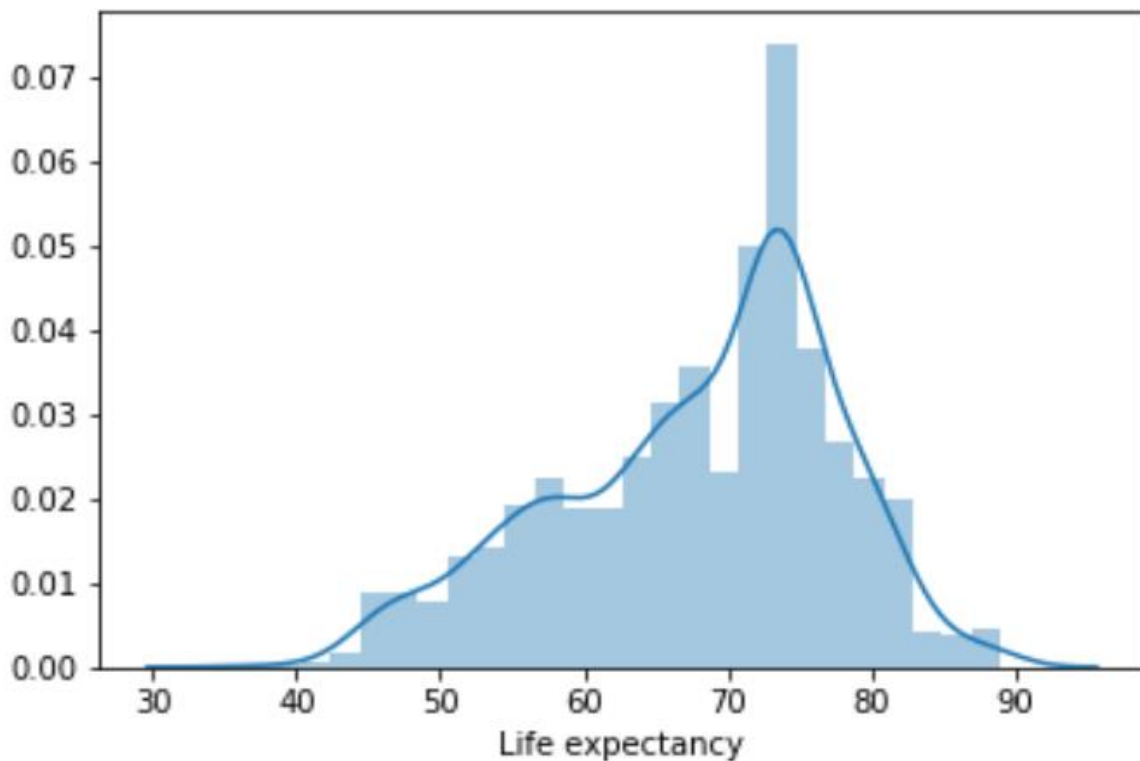
Above graph shows that most of the countries have life expectancy in between 70-80 years.

Correlation between attribute visualizing using Heat-map

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

✛ Visualizing which Model is most Efficient
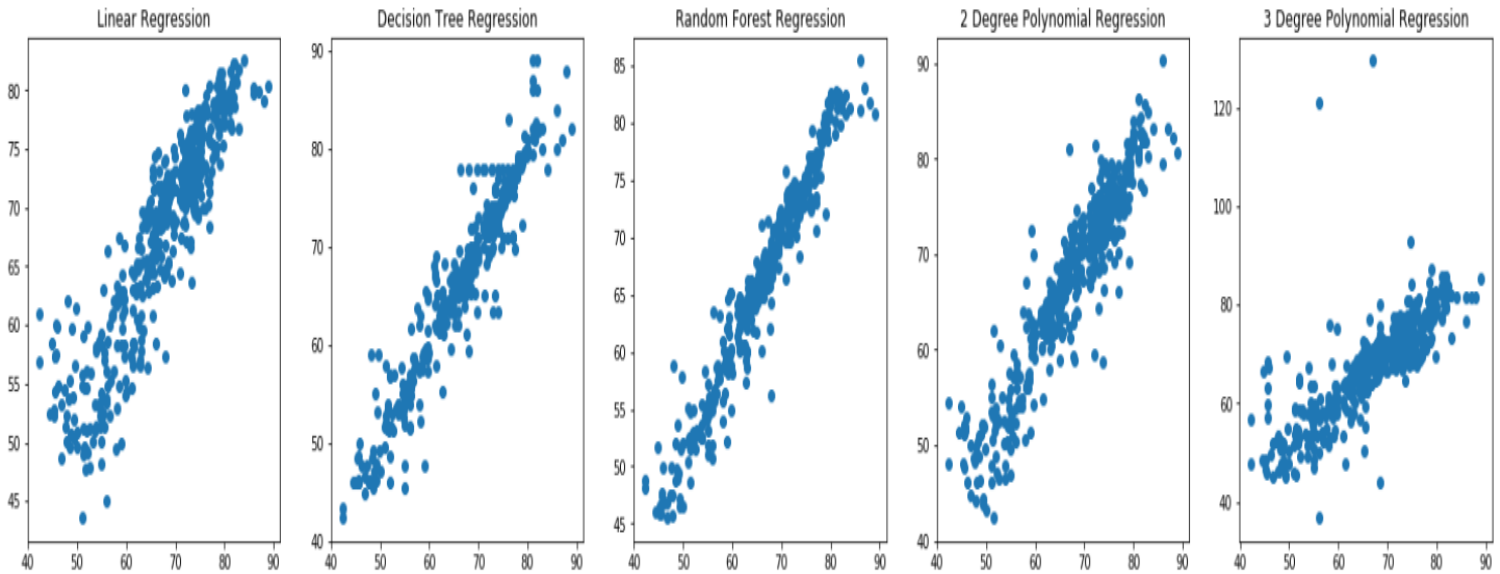
ML Models Used :

1. Linear regression model
2. Decision tree model
3. Random forest regression model
4. 2-degree polynomial regression
5. 3-degree polynomial regression

Random forest regression is the most suitable model for our data, as we can infer from the above graph.

5. FLOW CHART

```
                    ┌─────────┐
                    │  Start  │
                    └────┬────┘
                         │
                         ▼
              ┌──────────────────────┐
              │  Pre-processing Data │
              └──────────┬───────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │    Processed Data    │
              └──────────┬───────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │   Feature Selection  │
              └──────────┬───────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │    Dividing Data     │
              │   using ML method    │
              └──────────┬───────────┘
                   ┌─────┴─────┐
                   ▼           ▼
            ┌───────────┐ ┌───────────┐
            │ Test Data │ │ Test Data │
            └─────┬─────┘ └─────┬─────┘
                  └─────┬───────┘
                        ▼
              ┌──────────────────────┐
              │    Creating Models   │
              └──────────┬───────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │    Model Selection   │
              └──────────┬───────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │  Calculating ML Score│
              └──────────┬───────────┘
                         │
                         ▼
                    ┌─────────┐
                    │  Stop   │
                    └─────────┘
```
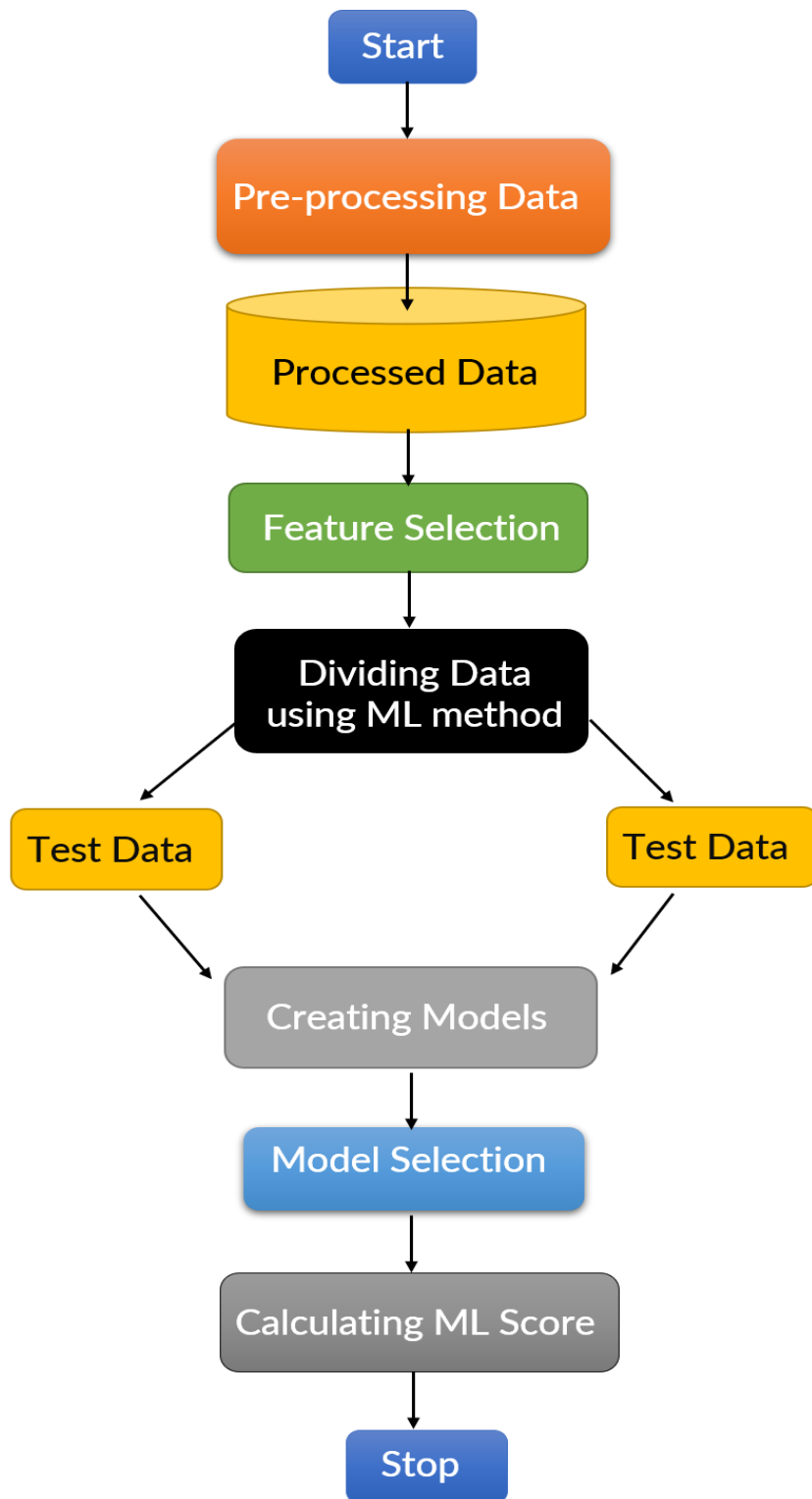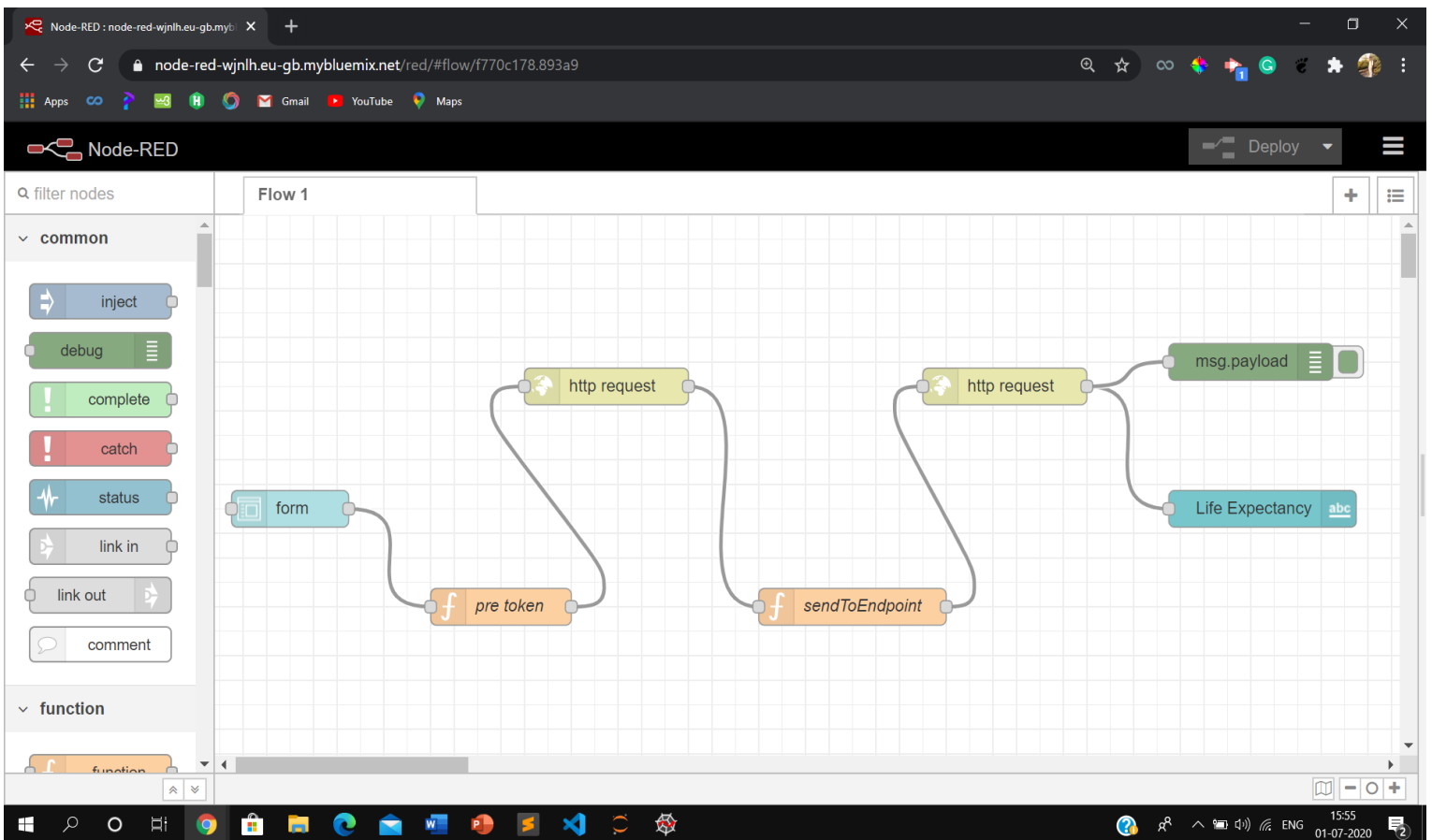
Node Red Flow

Explanation :

Form Node is been used for taking input from user in form format

Set 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
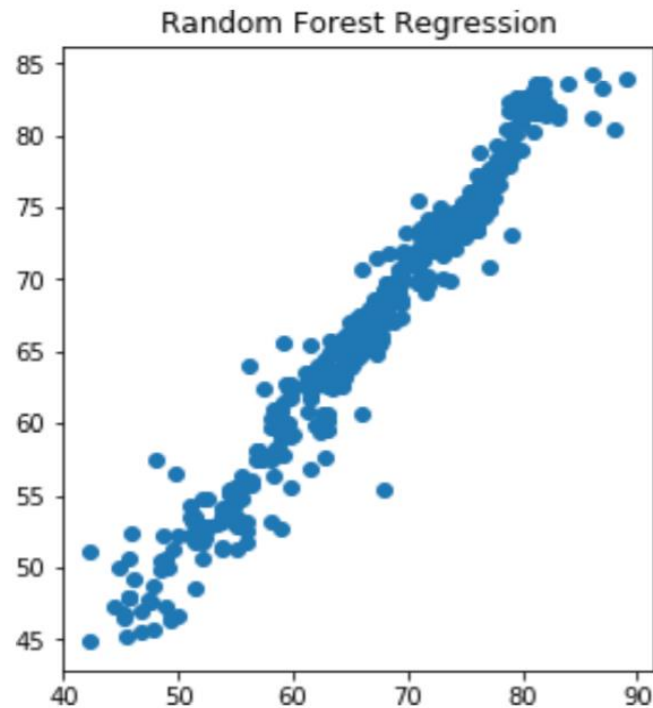
In Second HTTP request we are sending the json object to the model and getting predicted Life Expectancy from it

Edit text node is used to print the predicted values on dashboard

6. RESULT

🞧 Model Result

Random Forest Regression

graph between predicted and actual value

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, 71.2796236, 65, 1154 , 19.1 ,83 ,6 ,8.16 ,65 ,0.1 ,584.25921 ,33736494 ,17.2,17.3,0.479 ,10.1}

We got 67.64 as output but actual is 65, it is near to the result

🔳 Dashboard

Life Expectancy

Life Expectancy   **67.63999999999999**

Country
Afghanistan

Year
2015

Status
Developing

Adult Mortality
246

infant deaths
62

Alcohol
0.01

percentage expenditure
71.27962362

Hepatitis B
65

Measles
1154

BMI
19.1

under-five death
83

Polio
6

Total expenditure
8.16

Diphtheria
65

HIV/AIDS
0.1

GDP
584.25921

Population
33736494

thinness 1-19 years
17.2

thinness 5-9 years
17.3

Income composition of resources
0.479

Schooling
15.1

SUBMIT   CANCEL

Output as shown on the dashboard

7. ADVANTAGES & DISADVANTAGES

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 centre.
- We can compare things and make decision for the future to increase life span of humans.

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 results are published.
- Some people make decisions by looking the age of the person, If the age of the person exceeds over Life Expectancy then there is a chance of under caring of him/her.

8. APPLICATIONS

- By using this model we can create Apps.
- Life-tables and Demographic applications.
- These kind of software may increase awareness about life
- We can 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
- ➤ These type of applications will increase the awareness among people.
- ➤ It would be clear that which of our habits will affects our life.

10. BIBILOGRAPHY

Dashboard link : node-red-wjnlh.eu-gb.mybluemix.net/ui

Mentors of Smart-Internz really helped me a lot in deploying my ML model, and Creating Dashboard for my project.

THANK YOU