PROJECT REPORT

By: Pallavi Saumya

Project Topic: Predicting life expectancy of people in a country

Email: pallavisaumya25@gmail.com

WebPage Link: <https://node-red-sxeyq.eu-gb.mybluemix.net/ui/#!/0?socketid=eXO-egddVo0Jyt5RAABR>

1. INTRODUCTION

1.1. Overview

Life expectancy is an estimate of how long a person would live, on average.

Life expectancy is affected by many factors such as:

• Socioeconomic status, including employment, income, education and economic wellbeing. , The quality of the health system and the ability of people to access it; health behaviors such as tobacco and excessive alcohol consumption, poor nutrition and lack of exercise.

• Social factors; genetic factors; and environmental factors including overcrowded housing, lack of clean drinking water and adequate sanitation, etc..

It is very important to predict average life expectancy of a country to analyse further requirements to increase its rate of growth or stabilise the rate of growth in that country. So this is a typical Regression Machine Learning project that leverages historical data to predict insights into the future.

The end product will be a webpage where you need to give all the required inputs and then submit it . Afterwards it will predict the life expectancy value based on your regression technique.

**1.2. Purpose**

Through this project , I have tried to analyze the dataset i.e The Global Health Observatory (GHO) data repository under World Health Organization (WHO) to predict the life expectancy of people in different countries.

Since the observations of this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

Also since, life expectancy takes into account several individual-level as well as population-level factors to arrive at a figure, so, at individual level, people can take personal care like leave bat habits like alcohol consumption, life insurance planning,etc.

2. LITERATURE SURVEY

2.1. Proposed Problem

This problem statement is aimed at predicting Life Expectancy rate of a country given various features.

Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors.

This problem statement provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

2.2. Proposed Solution

The proposed solution specifically aims at getting a end product which will be a webpage where you need to give all the required inputs and then submit it . Afterwards it will predict the life expectancy value based on your regression technique.

**Steps:**

1. a) Create IBM cloud services
2. b) Configure Watson Studio
3. c) Create Node-Red Flow to connect all services together
4. d) Deploy and run Node-Red app

**2.2.1. Create IBM cloud Services**

• Watson Studio

• Machine Learning resource

• Node-Red

**2.2.2. Configure Watson Studio**

Get started with Watson Studio. Then create an empty project and add machine learning resource as associated services in settings.

Then add dataset and empty Jupyter notebook into Assets.After that go to notebook and write your code to build model and get the scoring endpoint URL.

**Steps for notebook:**

* **data collection -(**to import proper dataset**)**
* **Import necessary libraries**
* **feature selection -(**to select those features that will contribute most in predicting the life expectancy and thus it will help in increasing the accuracy of prediction)
* **Data Preprocessing-** Replacing nan values if any with their mean values.

**Train and Test** - The dataset was splitted into two parts i.e dependent variables and Dependent variables.

As Life Expectancy needs to be predicted so it is to be treated as output and all other columns are treated

as independent variables. Afterwards as we need regression technique to build our model .

* **Model Building and Deployment-**  At first the machine learning service credentials was stored in a variable Then the model was deployed and scoring\_endpoint url was generated .

**2.2.3. Create Node-Red Flow**

Go to Node-Red Editor from resource list. Install node-red Dashboard from manage pallette.

Now create the flow with the help of following node.

o Inject

* 1. o Debug
  2. o Function
  3. o Ui\_Form
  4. o Ui\_Text

Deploy and run Node Red app. Deploy the Node Red flow. Then copy the link url upto .net/ and paste at a new tab by ui at the end of the url .Deploy and run Node Red app.

Deploy the Node Red flow. Then copy the link url upto .net/ and paste at a new tab by ui at the end of the url like this.

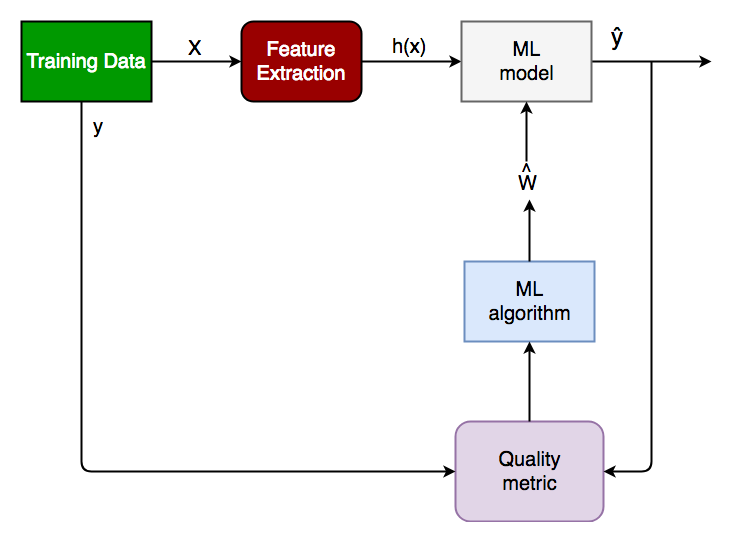
3. THEORETICAL ANALYSIS

3.1. BLOCK DIAGRAM

Input values to the fields such as 'country', 'BMI', 'Total expenditure' , 'measles', 'Status', HIV/AIDS', 'Alcohol' , 'percentage expenditure' and etc to the blank fields in webpage.

Deployed machine learning model with maximum accuracy score

Predicted Life Expectancy value



3.2. HARDWARE / SOFTWARE DESIGNING

o **Project Requirements**: Python, IBM Cloud, IBM Watson

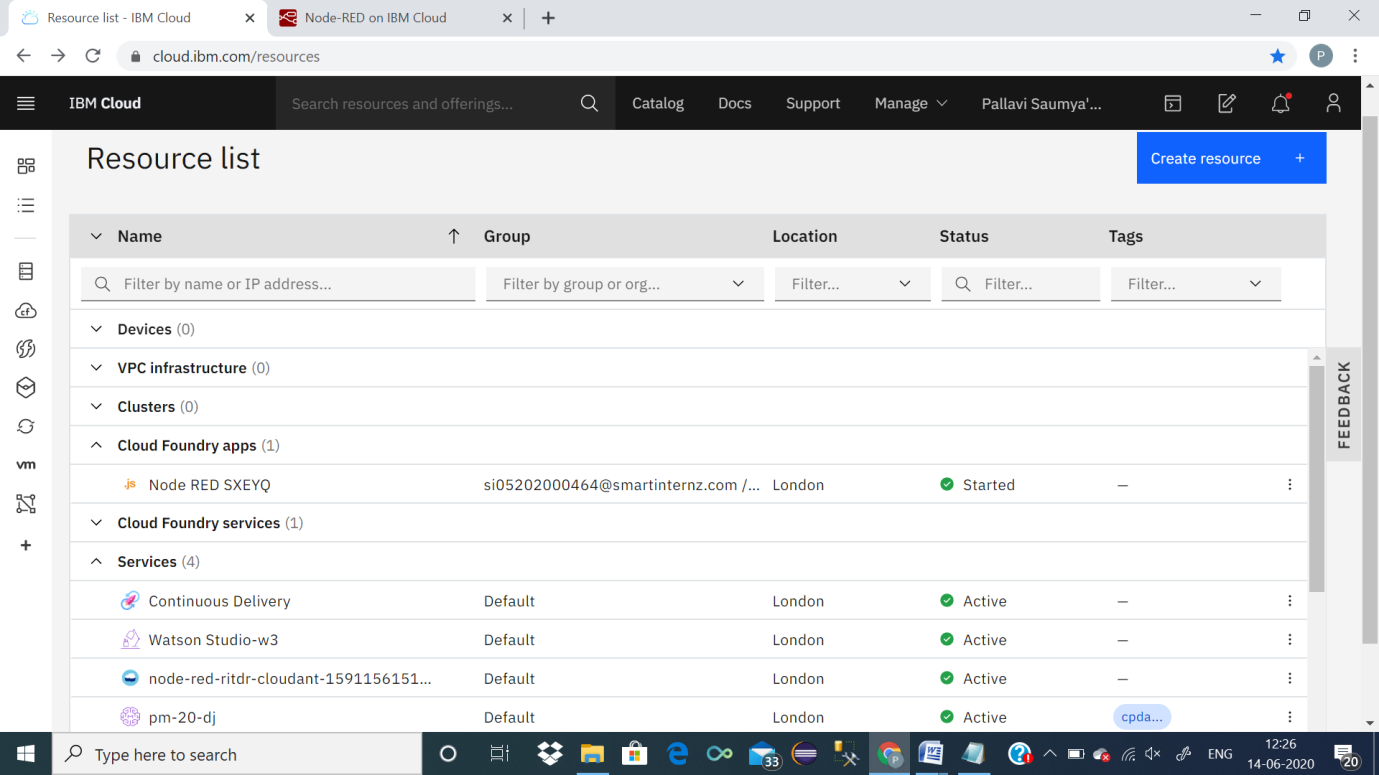
o **Functional Requirements**: IBM cloud

o **Technical Requirements**: ML, WATSON Studio, Python, Node-Red

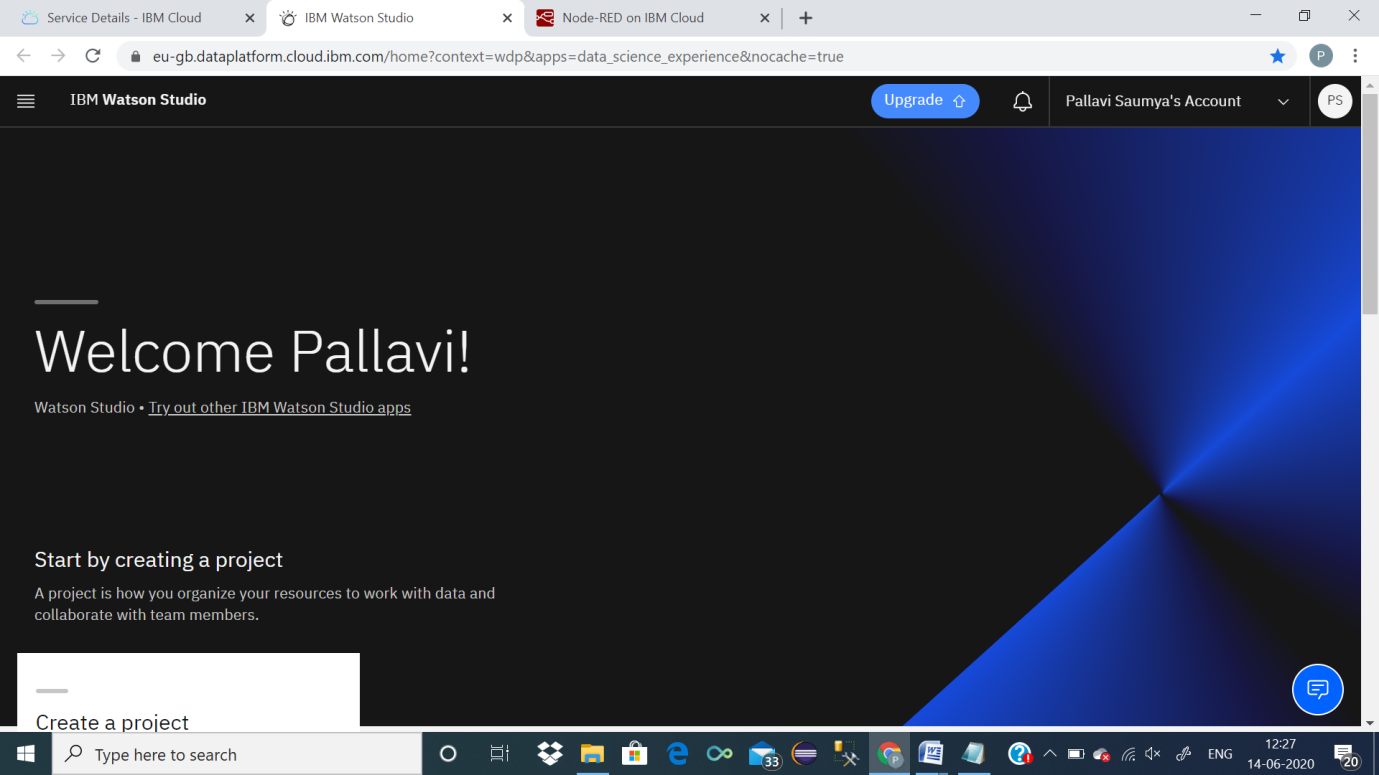
o **Software Requirements**: Watson Studio, Node-Red

4. EXPERIMENTAL INVESTIGATIONS

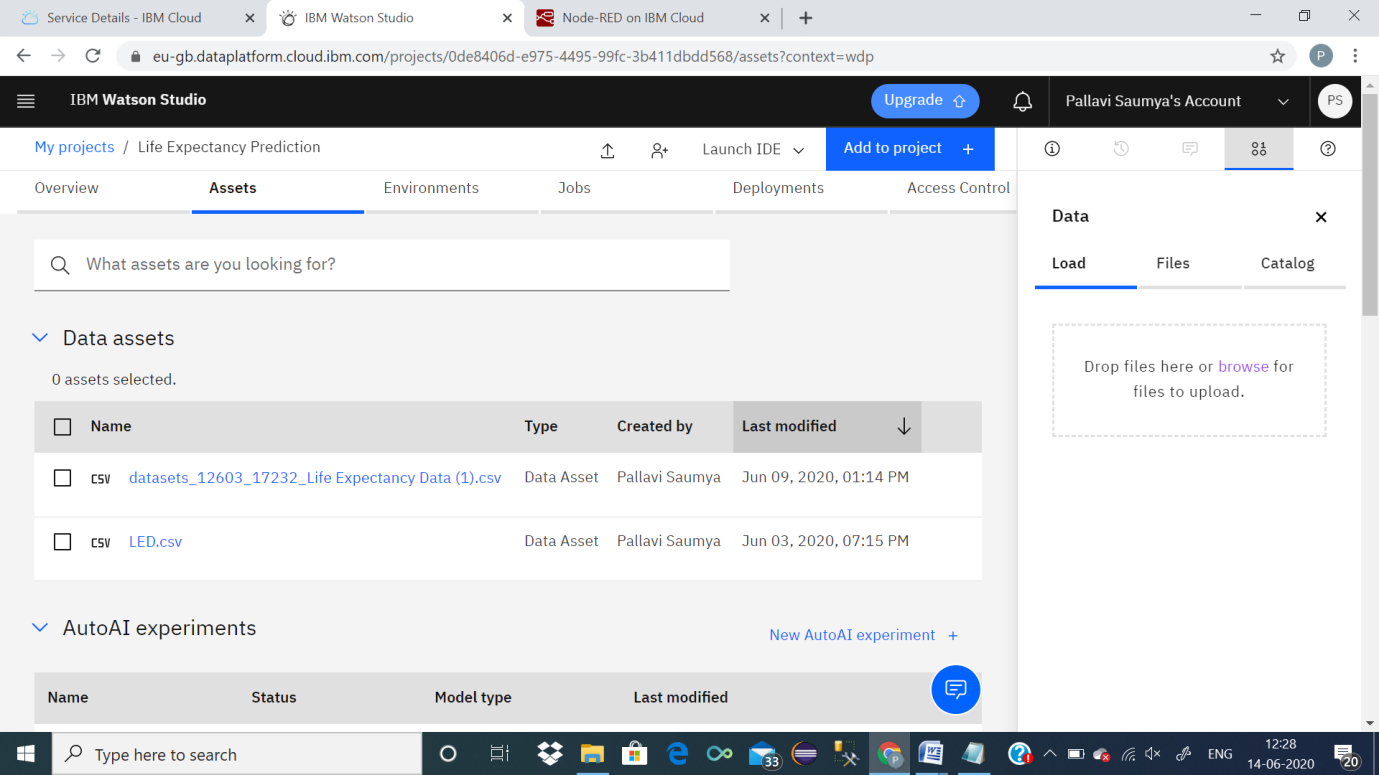
**A) IBM Cloud Resource List**



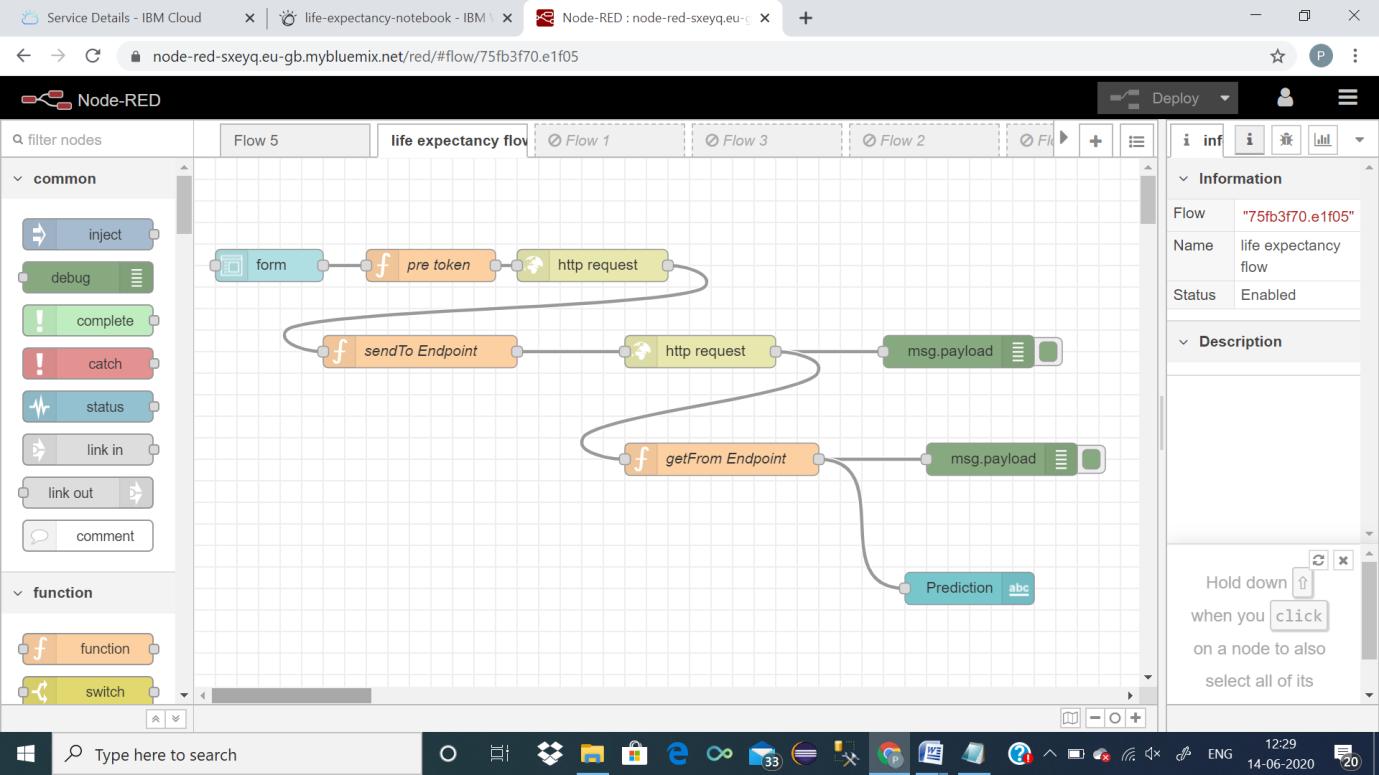
**B) IBM Watson Studio**

****

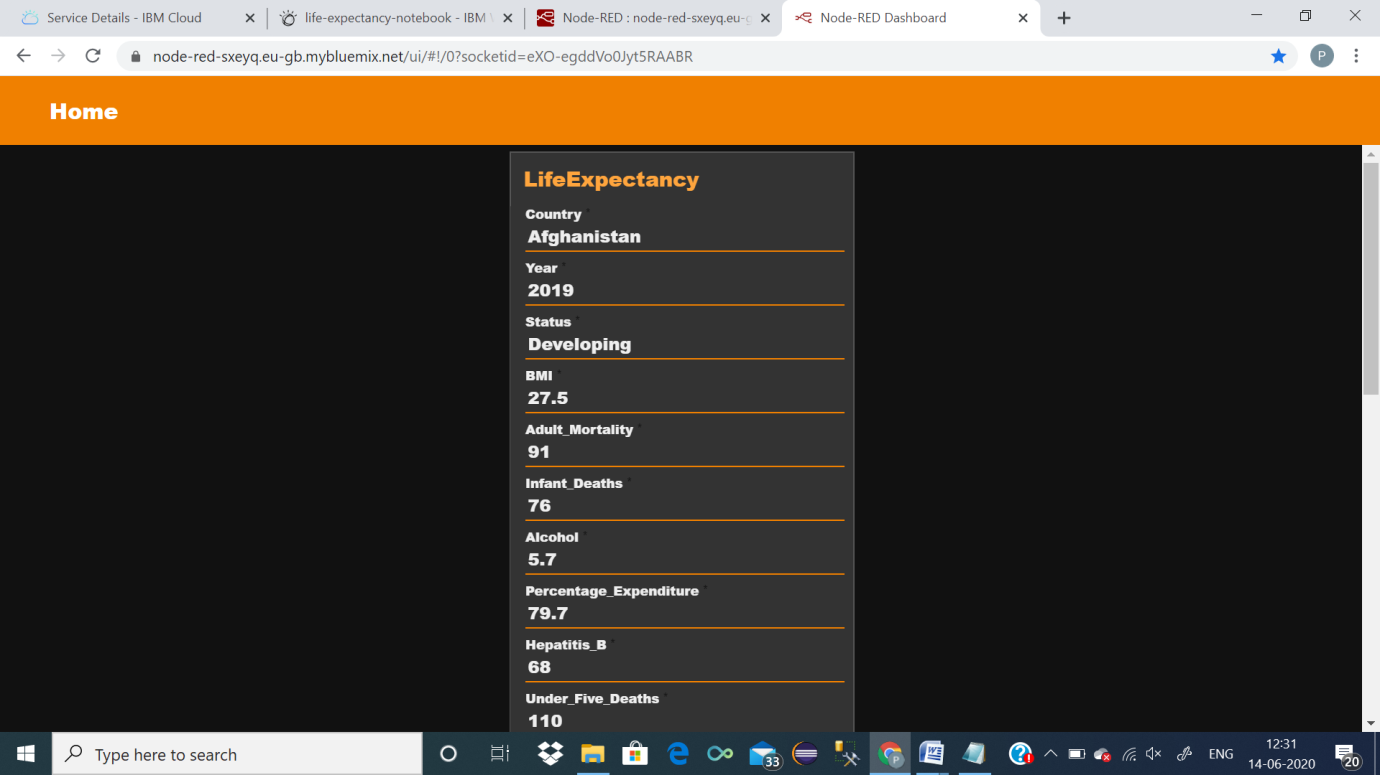
**C) IBM Cloud Project Details**



**D) Node-Red Flow**



**E) Life Expectancy Prediction UI**



**FLOWCHART**



The above flow chart can be explained in reference to my project in the following way:

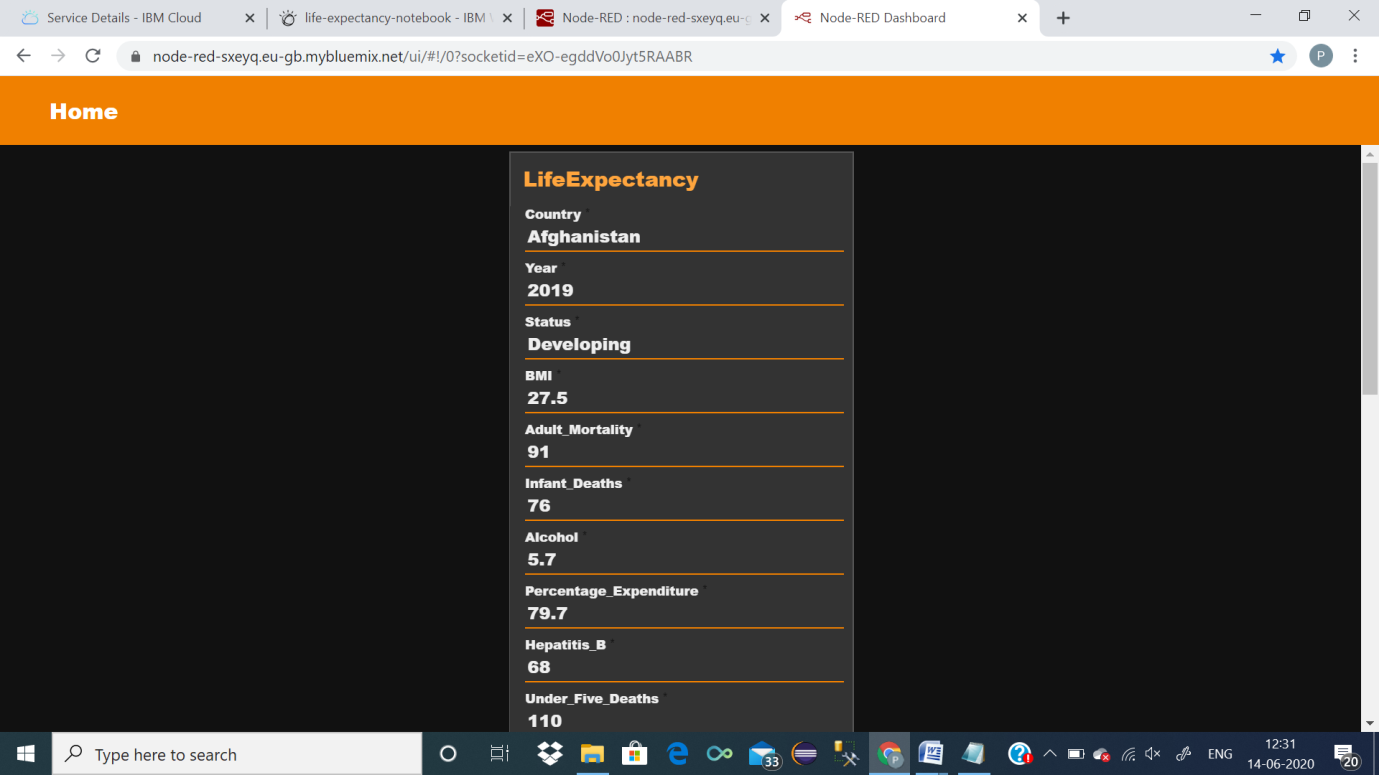
1. a) Firstly, the user will provide input for all the required values .

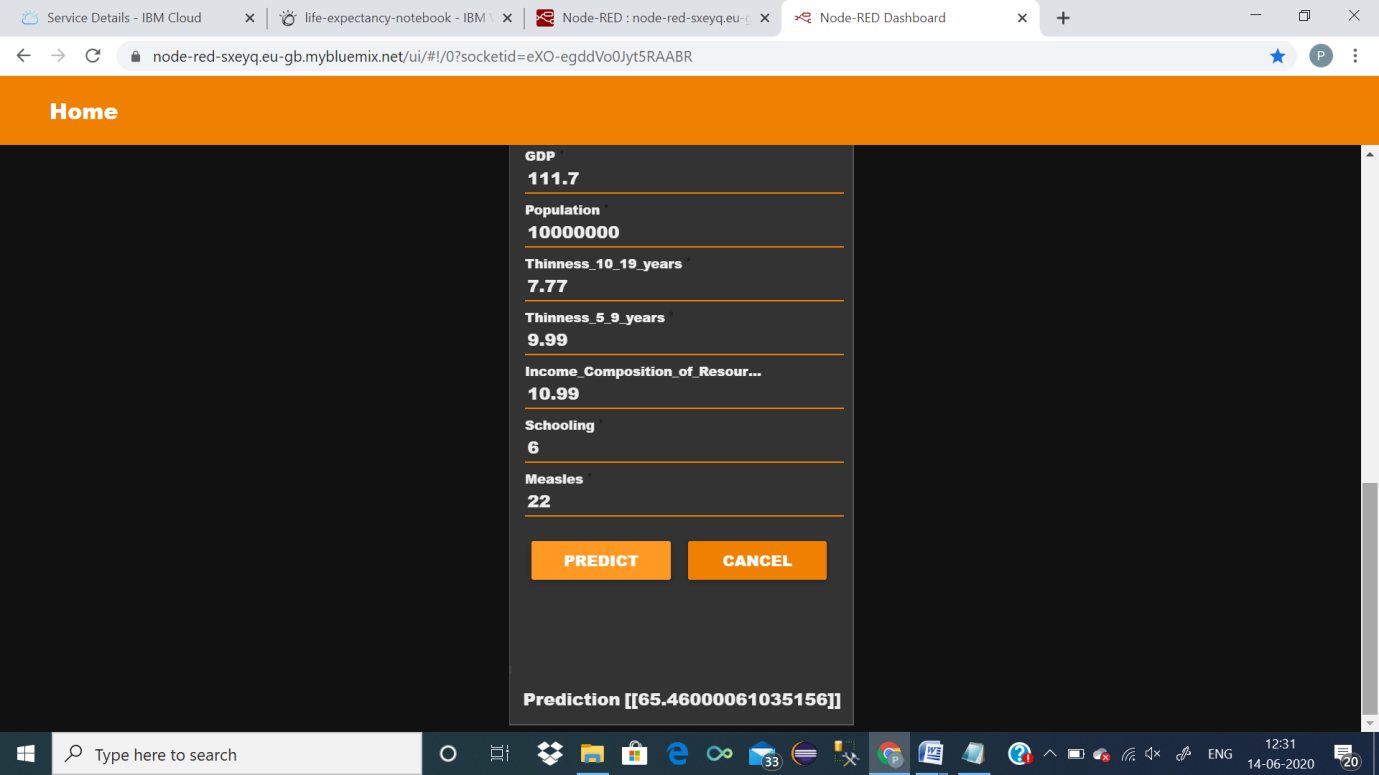
b) The data then entered into watson and the scoring endpoint url matches with the

1. deployed model.
2. c) Then it enters into trained data and predicts the life expetancy value .

**6. RESULT**

The end product or the required result will be a webpage where you need to give all the required inputs and then submit it . Afterwards it will predict the life expectancy value based on your regression technique.





7. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

1)Helpful for a country for analysis: It will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

Life expectancy can be used nationally to monitor health inequalities of a country.

2) Reduced Costs: This is a simple webpage and can be accessed by any citizen of a country to calculate life expectancy of their country and does not required any kind of payment neither for designing nor for using.

3) User Friendly Interface: This interface requires no background knowledge of how to use it. It’s a simple interface and only ask for required values and predict the output.

DISADVANTAGES:

a)Prediction may not be accurate: As it depends completely on user, so if user provides some wrong values then it will predict wrong value.

b) Average Prediction: The model predicts average or approximate value with 94.07% accuracy but not accurate value.

c)User input is not saved in any database.

8. APPLICATION

a) It can be used to monitor health inequalities of a country.

b) It can be used to develop statistics for country development process.

c) It can be used to analyse the factors for high life expectancy.

d) It will help government prepare life insurance policies for people .

9. CONCLUSION

We have prepared a model that will predict the life expectancy of a person in a country.

The end product which is a user interface which will be useful for the user to predict life expectancy value of their own country or any other country based on some required details such as GDP, BMI, Year, Alcohol Intake, Total expenditure and etc.

11. BIBLIOGRAPHY

* <https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform>
* <https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>
* <https://nodered.org/>
* <https://github.com/watson-developer-cloud/node-red-labs>
* <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

Appendix

Source Code

**import** **types**

**import** **pandas** **as** **pd**

**from** **botocore.client** **import** Config

**import** **ibm\_boto3**

**def** \_\_iter\_\_(self): **return** 0

df\_data\_1 = pd.read\_csv(body)

df\_data\_1.head()

**import** **numpy** **as** **np**

**import** **seaborn** **as** **sns**

**import** **matplotlib.pyplot** **as** **plt**

df\_data\_1=df\_data\_1.dropna()

df\_data\_1=df\_data\_1.fillna(df\_data\_1.mean())

x=df\_data\_1.iloc[:,[0,1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21]].values

y=df\_data\_1.iloc[:,3].values

**from** **sklearn.preprocessing** **import** LabelEncoder, OneHotEncoder

labelencoder\_x=LabelEncoder()

x[:,0]=labelencoder\_x.fit\_transform(x[:,0])

x[:,2]=labelencoder\_x.fit\_transform(x[:,2])

onehotencoder=OneHotEncoder(categorical\_features=[0,2])

x=onehotencoder.fit\_transform(x).toarray()

**from** **sklearn.model\_selection** **import** train\_test\_split

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.5,random\_state=77)

**from** **sklearn.linear\_model** **import** LinearRegression

lr=LinearRegression()

lr.fit(x\_train,y\_train)

y\_pred=lr.predict(x\_test)

plt.scatter(y\_test,y\_pred ,color='black')

plt.title('life expectancy Prediction(Graph)')

plt.xlabel('Actual Life Expectancy')

plt.ylabel('Predicted Life Expectancy')

plt.show()