Project Report

on

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

By: TIYASHA BISWAS

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

Project Summary:

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This project is aimed at predicting Life Expectancy rate of a country given various features. The Global Health Observatory (GHO) data repository under World Health Organisation (WHO) keeps track of the health status as well as many other related factors for all countries The data-sets are made available to public for the purpose of health data analysis. The data-set related to life expectancy, health factors for 193 countries has been collected of a time frame 2000 to 2015. The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they have not been trained. Some of the algorithms that can be possibly used are:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Elastic Net Regression
- Linear Regression with Polynomic features
- Decision Tree Regression
- Random Forest Regression

So, we will be developing an application using ML algorithm for the prediction of the life expectancy of such patients.

Project Requirement:

This project fundamentally aims in predicting the life expectancy. The primary requirement of the project is the suitable dataset which will aid the prediction.
The dataset will provide various details like kind of diseases leading to the death.
By using supervised machine learning techniques, we can extract a model that will be able to predict the life expectancy of future years.

Functional Requirement:

1. Create a data model present on the database.
2. The data set are made available to the public to the purpose of health data analysis.

3. It is related to the different countries depending on the different countries while finding the data set in different countries might be difficult and hence some countries are excluded from the final data set.

Technical Requirements:

1. The merged data set by using the databases in the .csv formats from Kaggle

2. Datasets need to be integrated into the Python IDE.

Software Requirements:

1. Python IDE

2. Excel

3. IBM Cloud

4. IBM Watson

5. IBM Node-Red Service

Project Deliverables:

1. Collect the data
2. Prepare a model for predicting life expectancy based on the collected data.
3. Prepare a module for prediction.

At the end, we will be able to predict the life expectancy of an individual.

Project Schedule:

The project is scheduled for 15 May, 2020 to 14 June, 2020.

Project Team:

The project is done individually with the help of IBM Cloud, the project can be written in Watson Studio and deployed using Node-Red Apps.

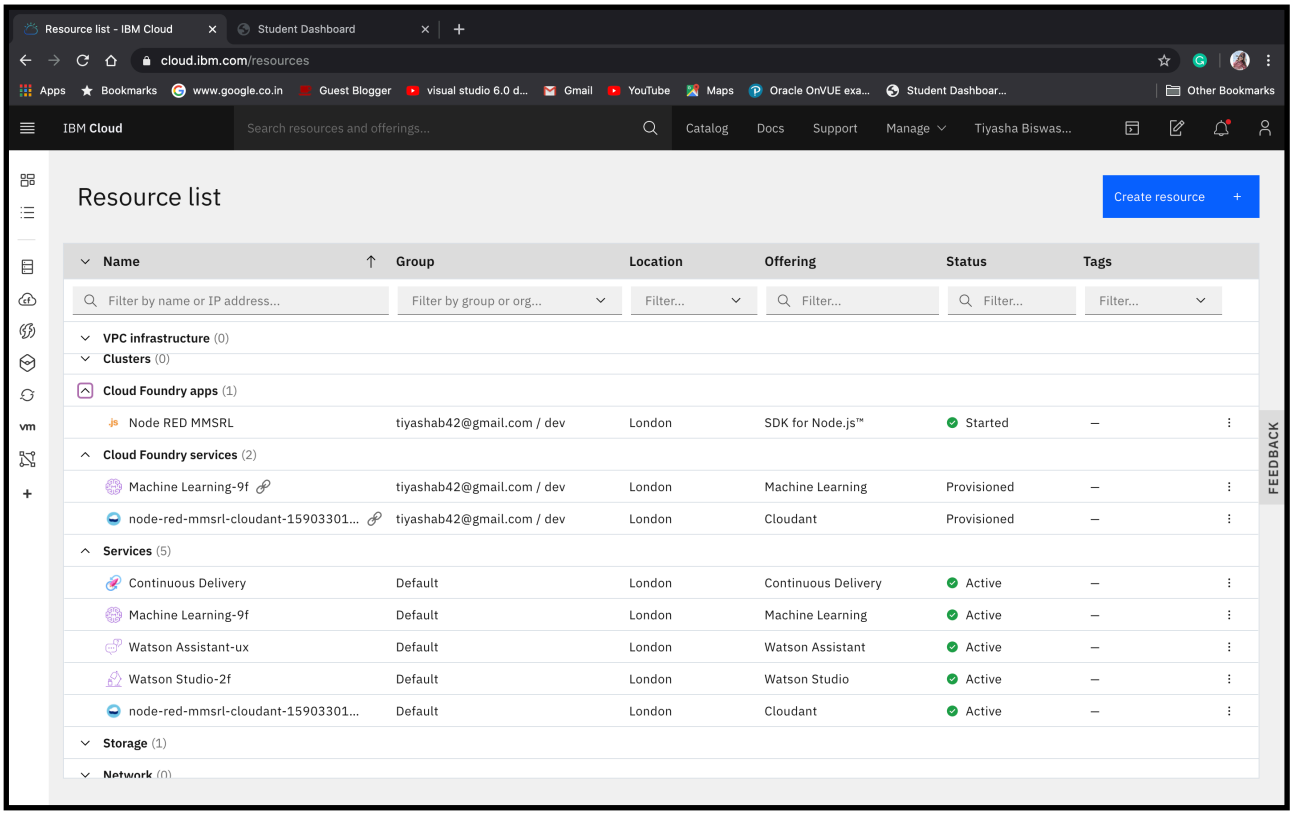
Phrases of the development:

1) Collecting the Dataset:

First most important thing for any project is collecting the data as per requirement of the model. Thus, we collect the data from the given source. For the project the dataset was "Life Expectancy". Thus dataset was provided by the WHO in order for the analysis purpose. We have used this dataset for the prediction purpose.

2) *Setting up IBM Cloud Services:*

For using the various Cloud services for the project development. One must first create an IBM Cloud account. Once the account is created you can access various services used for ML projects.



3) Creating a Watson Project:

Once the services required for the project are enabled you can go with for the creation of the project. Watson Studio allows you to create various project using different tools like Jupyter Notebook, Auto AI, R Studio etc.

- *Configure the Watson studio:*

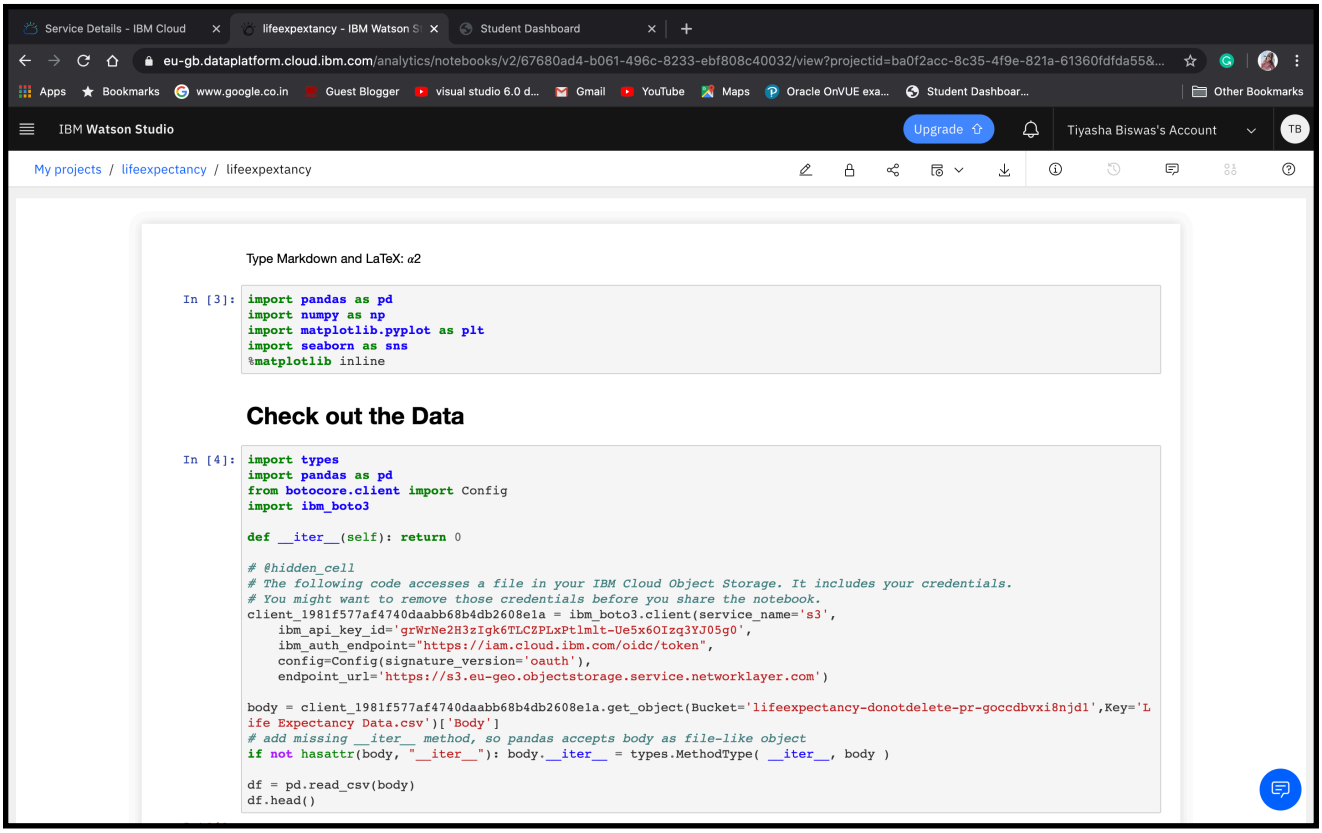
  Once you are done with the creation of the Watson project you can configure the various services associated with it. Also you can look for the various tools associated with it.

4) Creating Machine Learning Services:

As we are creating the Machine Learning Model for the prediction of the Life Expectancy we must create the Machine learning services in IBM cloud which will help in building up the model.

a) Create Jupyter Notebook and Import Dataset:

Firstly in the project we need to add the Jupyter Notebook (it is the platform for developing the model and actual implementation). Once the Jupyter Notebook is created we must import the data. The data set is Inserted to code In pandas data frame.



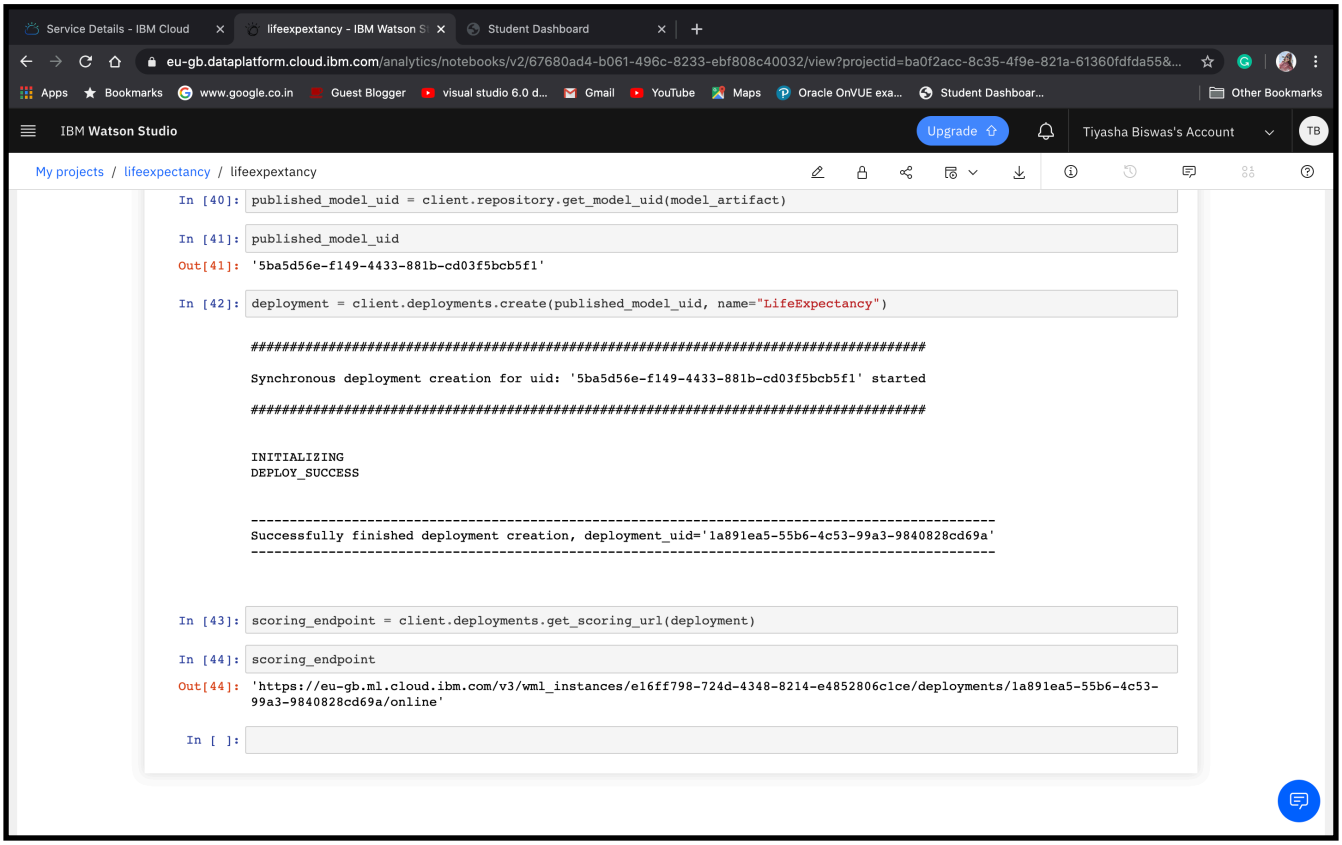
b) Choose the appropriate model for prediction:

We can use any model for the the prediction person and with the help of it you can train and test the dataset.

For the project I have been chosen the Linear Regression and Lasso and Ridge Regression for the development purpose.

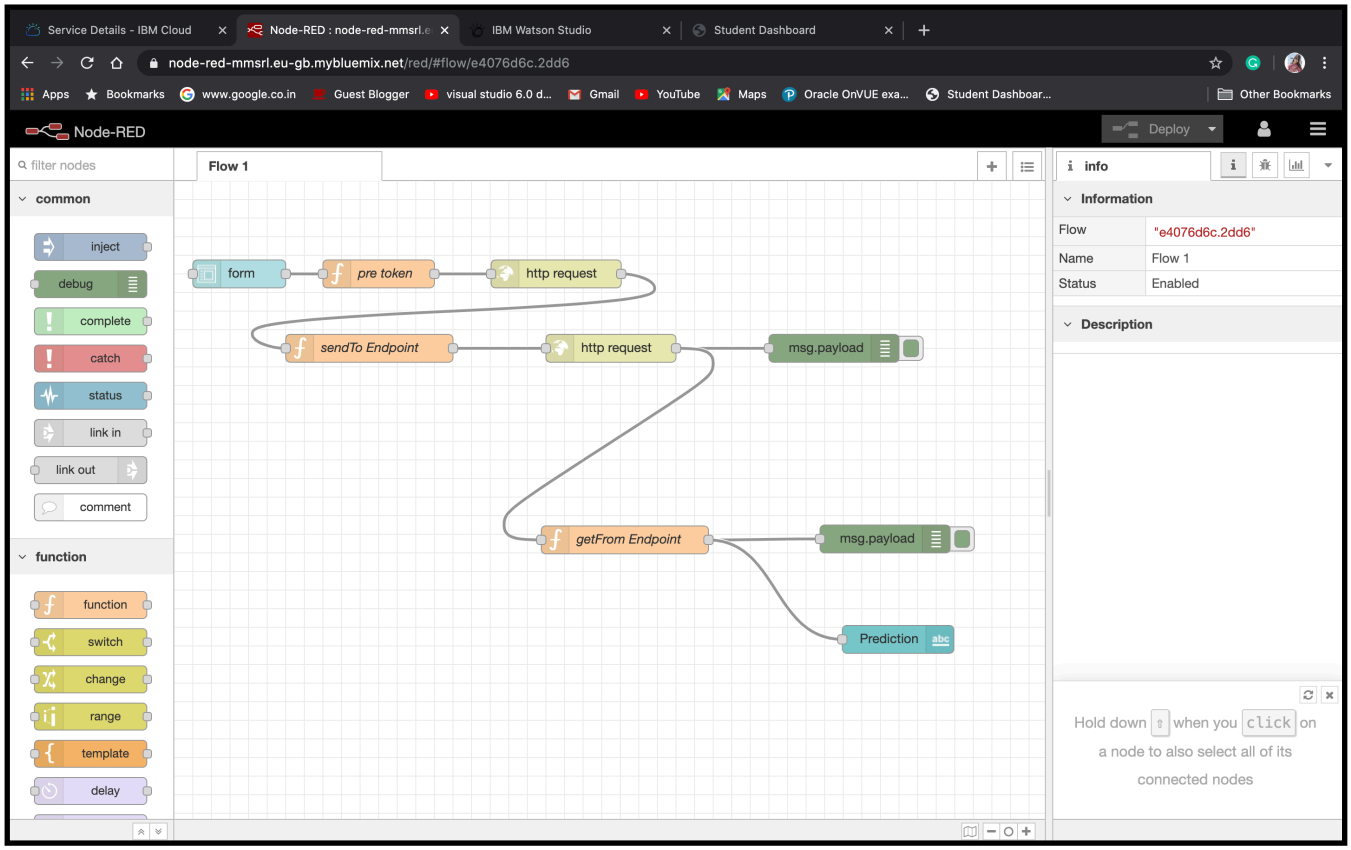
c) Deployment of Model:

Once we are done with building the model. We must deploy the model. The deployed model will be stored in IBM Cloud Storage.

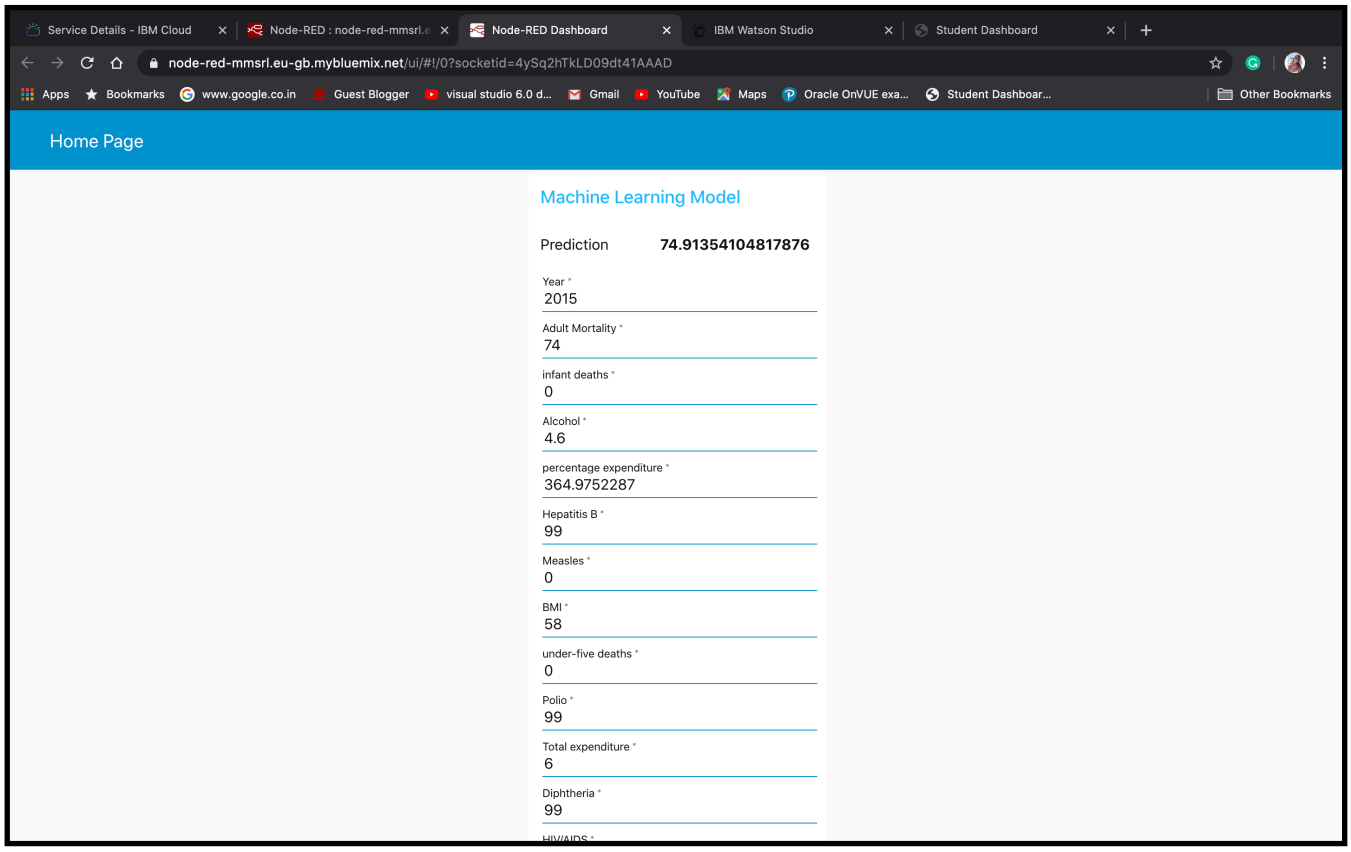


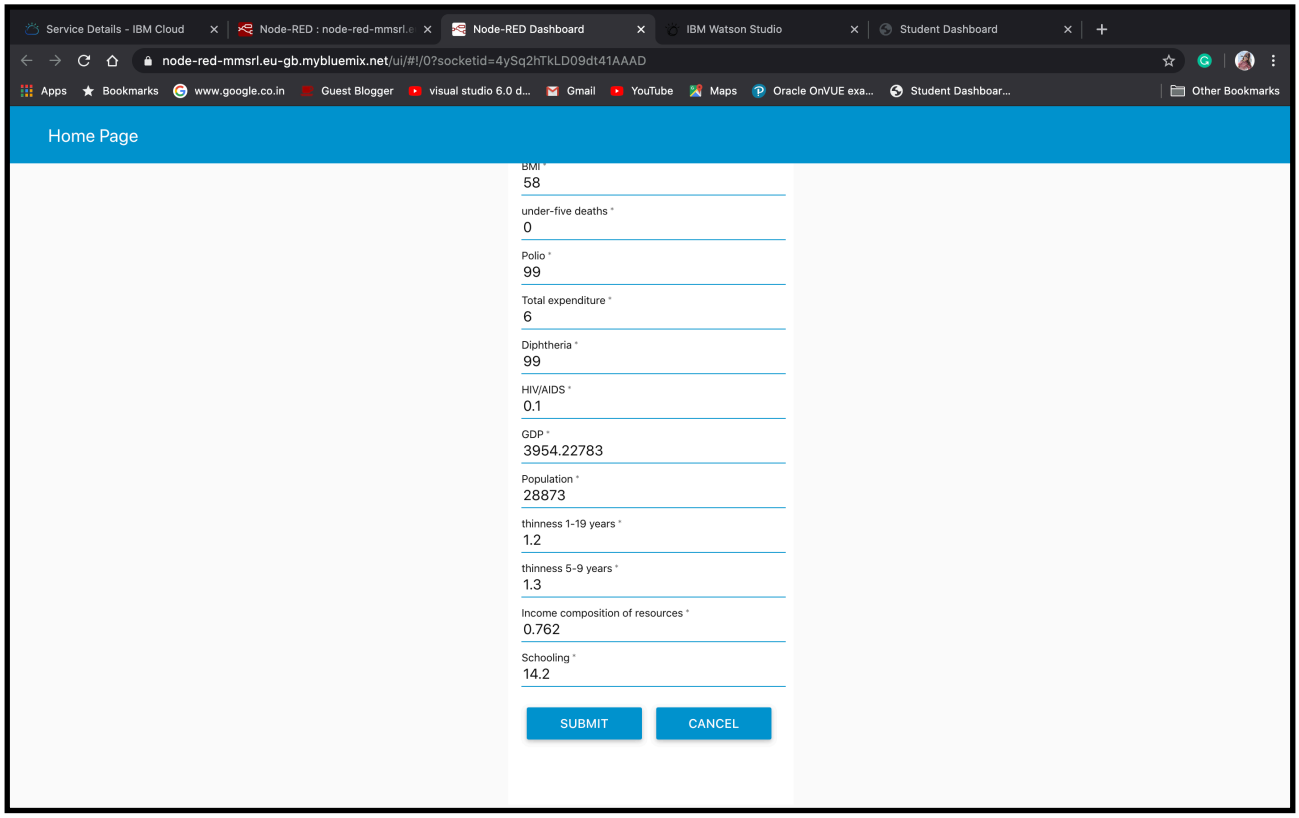
5) Create a Node-red Flow:

Once the model is deployed you can create the node red flow to create an API for the model . thus API will act as an front end to the model.



Output of the project:





Conclusion:

From the project we can predict the life expectancy from a particular set of values given in dataset.