PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

By : SOHAM R. SAHARE

SOHAM R. SAHARE

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

Project Summary :

A typical Regression Machine Learning project leverages historical data to predict insights into the future. 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.

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
- Random Forest Regression
- Decision Trees

And by Hyperparameter Tuning increasing the accuracy of predictive model.

Project Requirement :

SOHAM R. SAHARE

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 :

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

The merged data set by using the databases in the .csv formats from Kaggle. Datasets need to be integrated into the Python IDE.

Software Requirements :

- Python IDE
- IBM Cloud
- IBM Watson
- IBM Node-Red

Project Deliverables :

The project is scheduled for 29 days from 15th May 2020 to 14th June 2020.

Project Team :

The project is done individually by Soham R. Sahare.

Phases of the development

1) Project Planning & Kickoff :
   - Writing Project Scope, Schedule, Team & Deliverables for the project.
   - Setup the Development Environment – Creating GitHub account, Slack account, working with zoho writer and Understanding GitHub.

2) Explore IBM Cloud Platform :
   - Creating IBM account
   - Creating a Node-Red Starter Application

3) Explore IBM Watson Services :
   - Exploring IBM Watson usecases.
   - Exploring IBM Watson for Machine Learning.

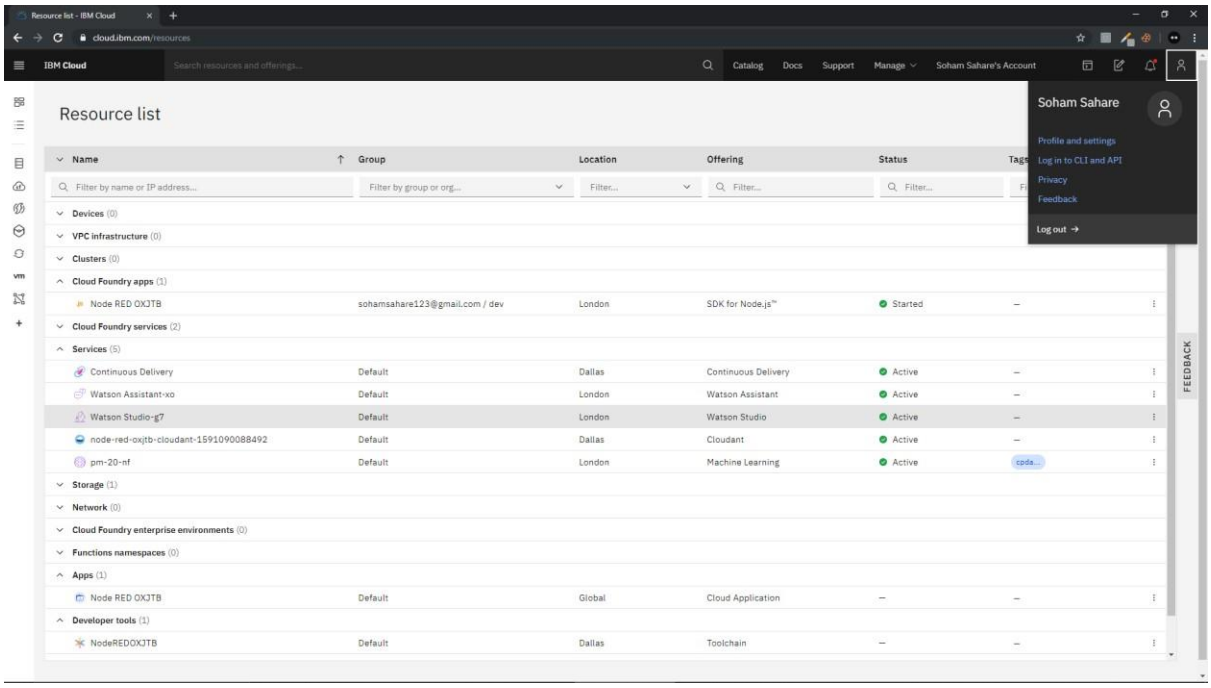
4) Introduction To Watson Studio :
   - Building a Machine Learning model in IBM Watson Studio.
   - Automating Machine Learning Model.

5) Predicting Life Expectancy with Python :

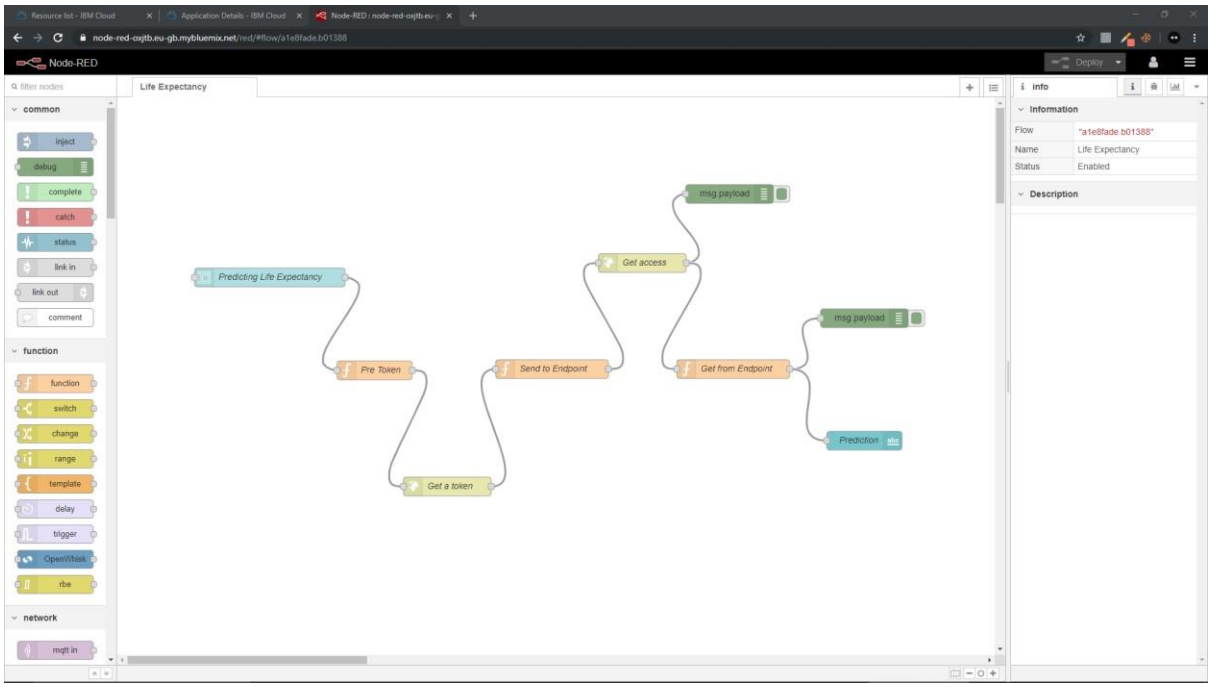
- Collecting Dataset from Kaggle.
- Creating Necessary IBM Cloud Services.
- Watson Studio Project
- Creating Machine Learning Service
- Creating a Jupyter Notebook in IBM Watson Studio and Import the dataset.
- Building a Machine Learning Model And Create Endpoints For Node-RED Integration.
- Build Node-RED Flow To Integrate ML Services

Some Screenshots while Project Develpoment :

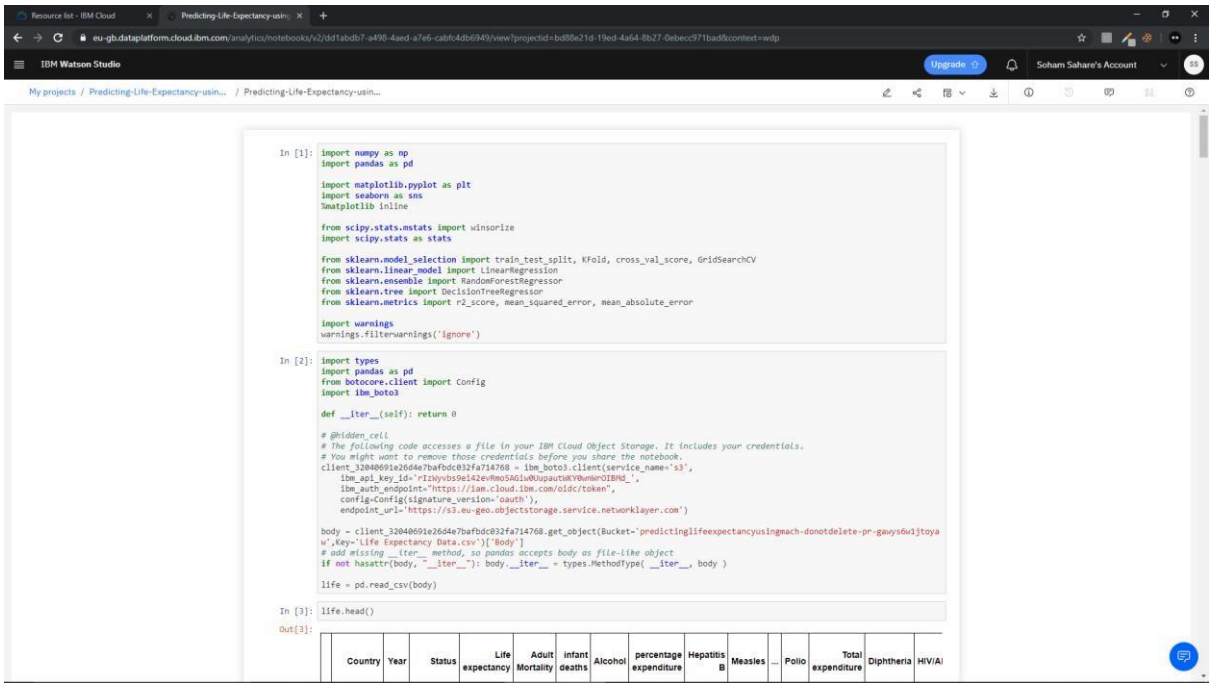
1) Resource list of all the apps and services developed in IBM Cloud.



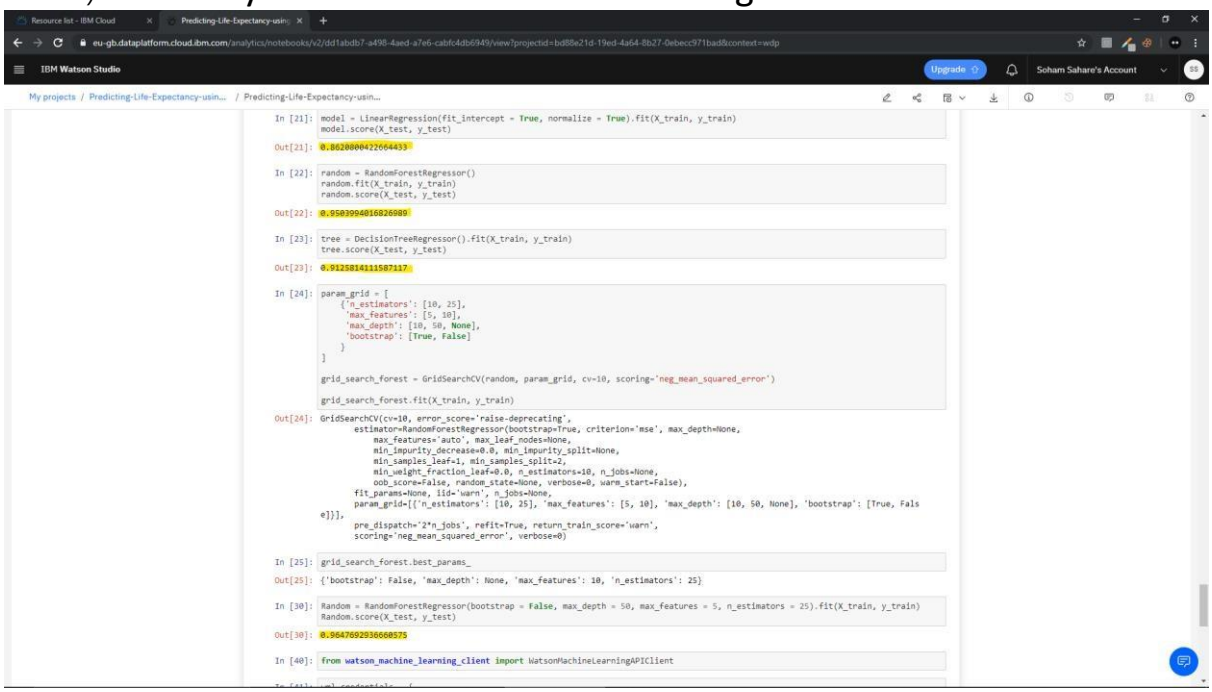
2) Node-Red Flow Editor.



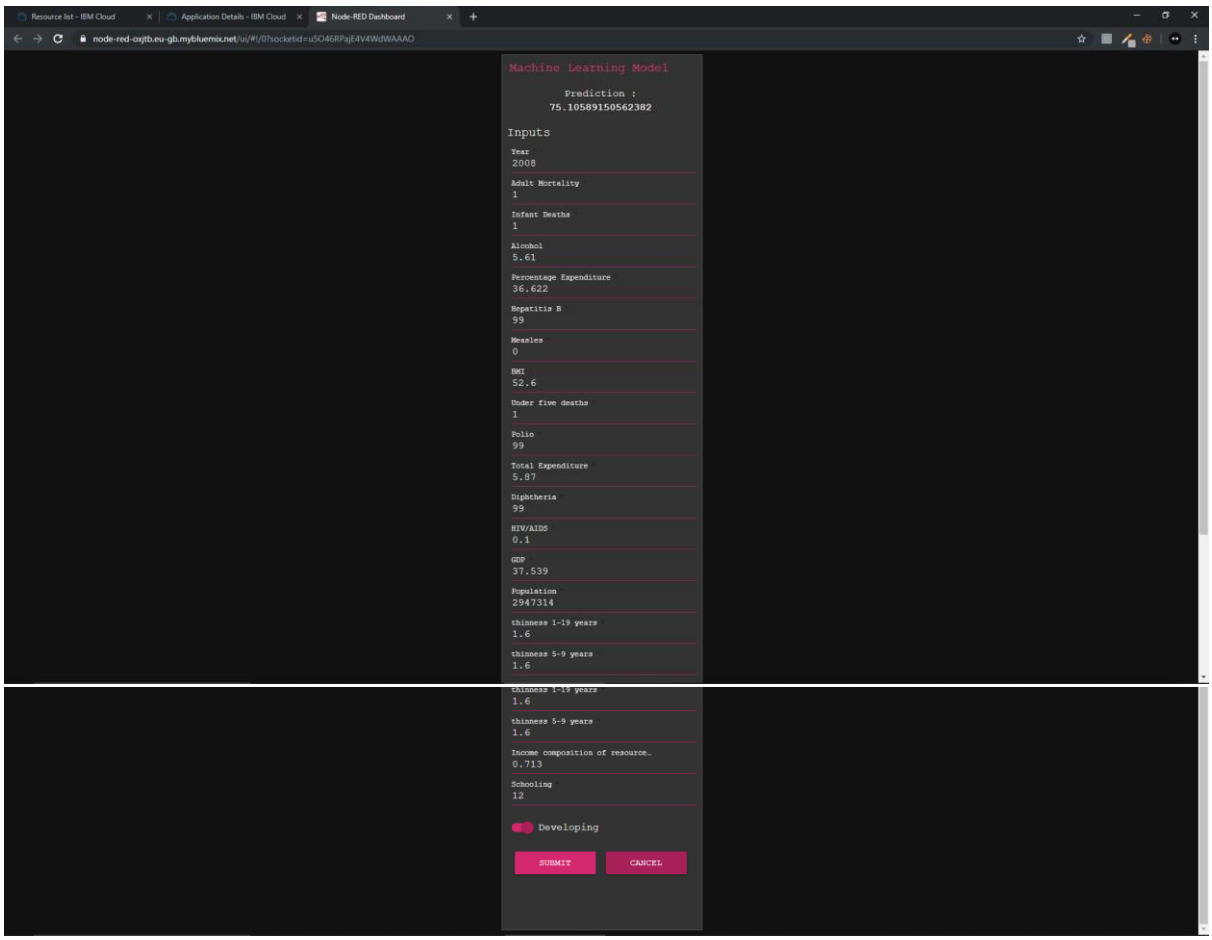
3) Jupyter Notebook in IBM Watson Studio for Development of Machine Learning Model.



4) Accuracy of different Machine Learning Models.



5) Node-Red Deployed.



Conclusion :

From this project we can predict Life Expectancy with 96.47 % accuracy. Using Hyper parameterized Random Forest Regressor.

Important Links :

Node-Red URL : https://node-red-oxjtb.eu-gb.mybluemix.net/ui/#!/0?socketid=YdWb57ES7FMjDH76AAAH

Node-Red Editor URL : https://node-red-oxjtb.eu-gb.mybluemix.net/red/#flow/a1e8fade.b01388

Deployed video : https://drive.google.com/file/d/1zj4NXGgM2DkH_eSawpeNB6SQiyef26p-/view?usp=sharing

Dataset : https://www.kaggle.com/kumarajarshi/life-expectancy-who