Project Report on :

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

-SmartInternz Project

By : TAMOJIT DAS

Introduction:

Overview:

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The data offers a timeframe from 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 haven't been trained.

Purpose:

Life expectancy plays an important role when decisions about the final phase of life need to be made. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. Advance Care Planning (ACP) is the process during which patients make decisions about the health care they wish to receive in the future, in case the patient loses the capacity of making decisions or communicating about them. Successful ACP enhances the quality of life and death for palliative patients, by providing timely palliative care and documenting preferences regarding resuscitation and euthanasia, among other things.

Accurate prognosis of life expectancy is essential for general practitioners (GPs) to decide when to introduce the topic of ACP to the patient, and it is a key determinant in end-of-life decisions. Increasing the accuracy of prognoses has the potential to benefit patients in various ways by enabling more consistent ACP, earlier and better anticipation on palliative needs, and preventing excessive treatment. This study focuses on automatic life expectancy prediction based on medical records.

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. Firstly, we will 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.

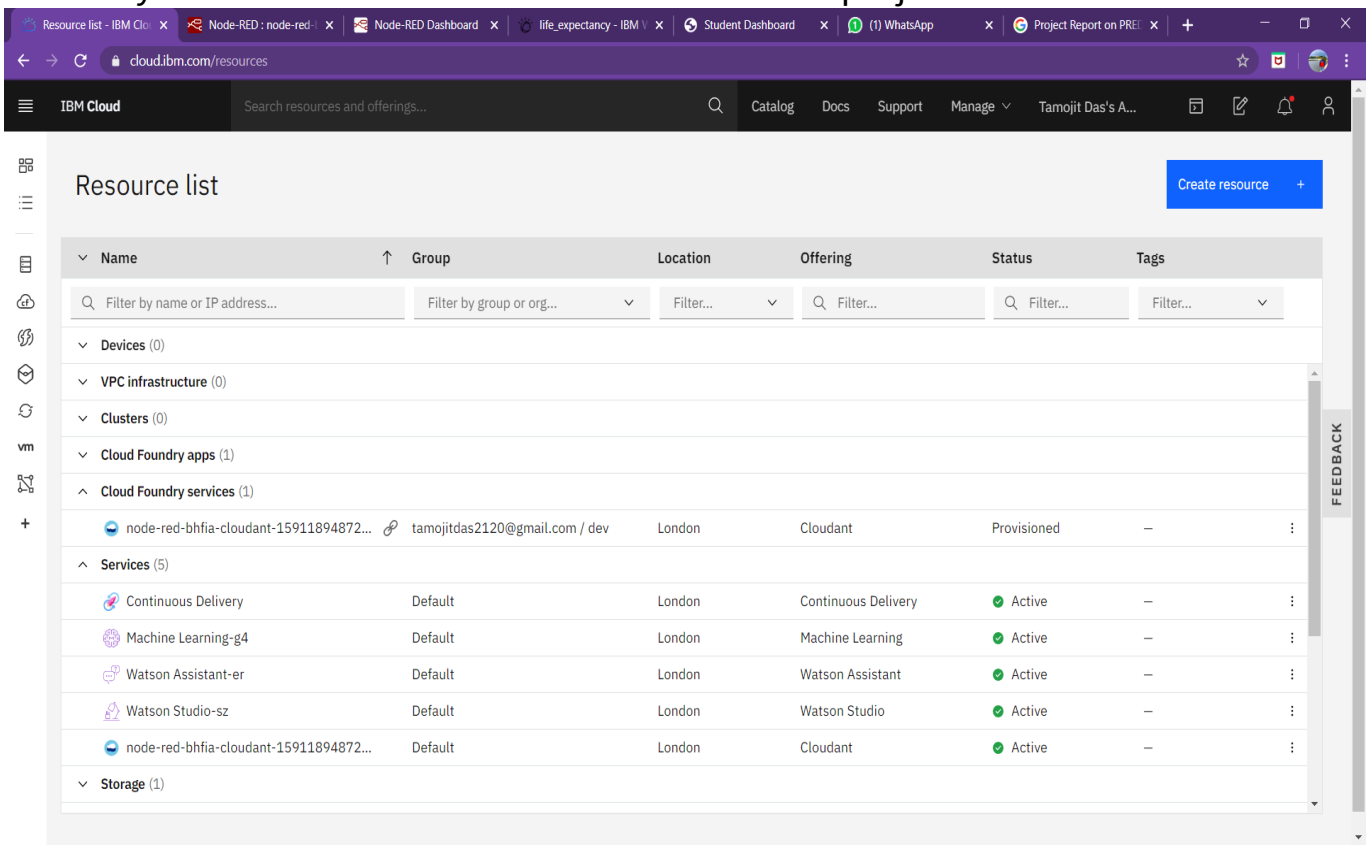
Project Schedule:

The project was scheduled to me from 19 May, 2020 to 17 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.

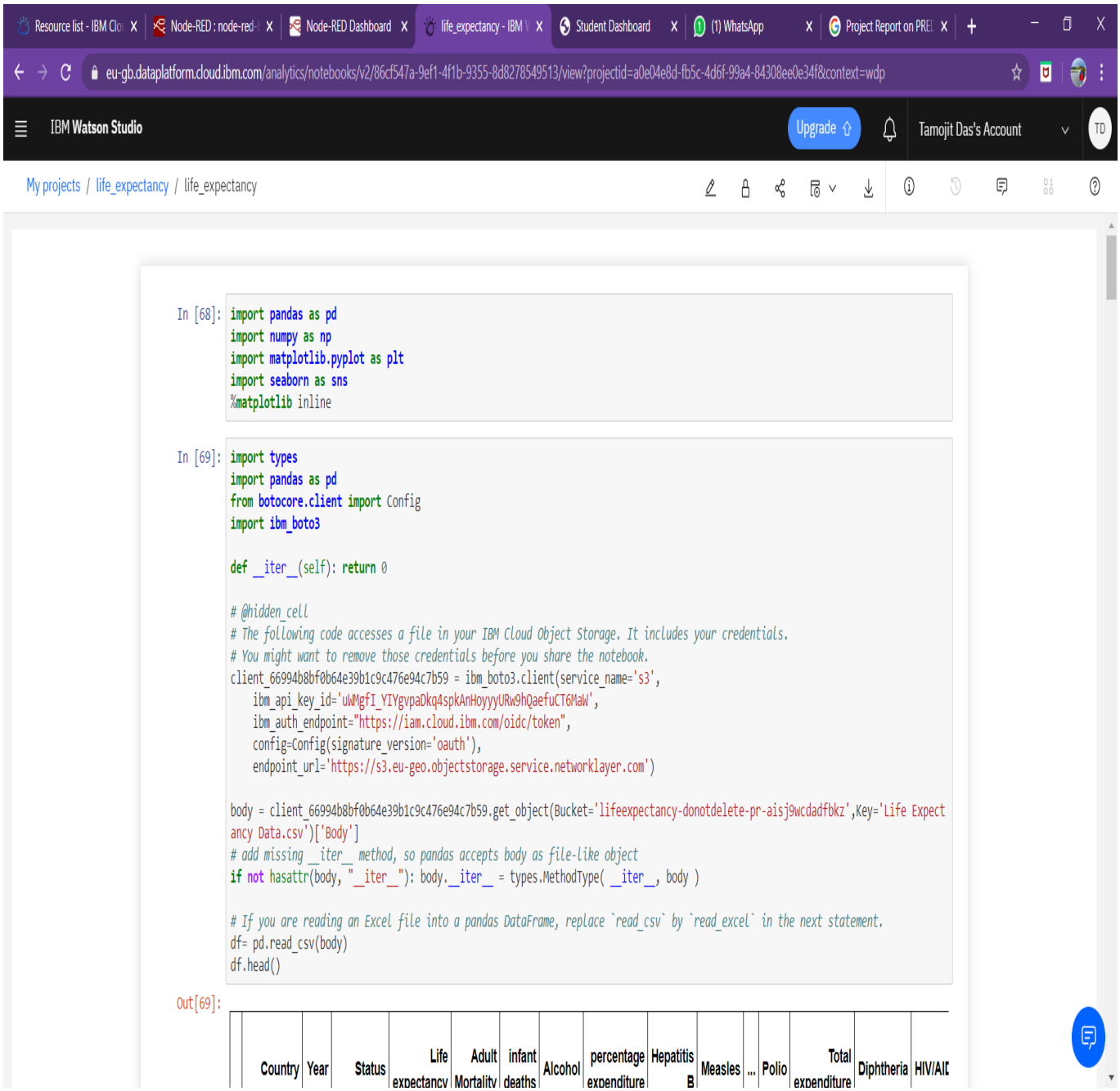


3) Creating a Watson Project: Once the services required for the project are enabled you can go with for the creation of the project. Watson Studio allows you to create various project using different tools like Jupyter Notebook, Auto AI, R Studio etc.
• Configure the Watson studio: Once you are done 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:

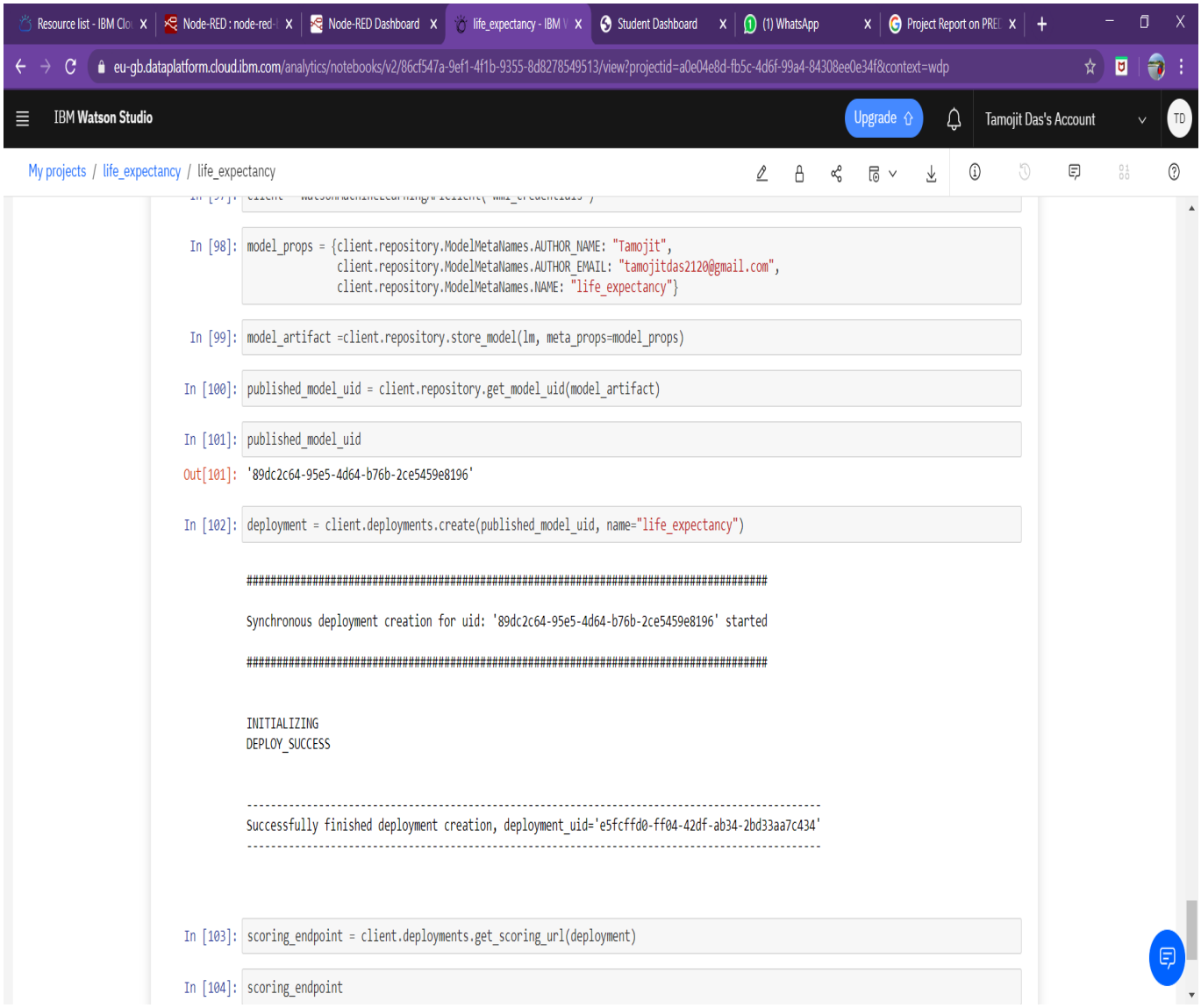
As we are creating the Machine Learning Model for the prediction of the Life Expectancy, we must create the Machine learning services in IBM cloud which will help in building up the model.

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.

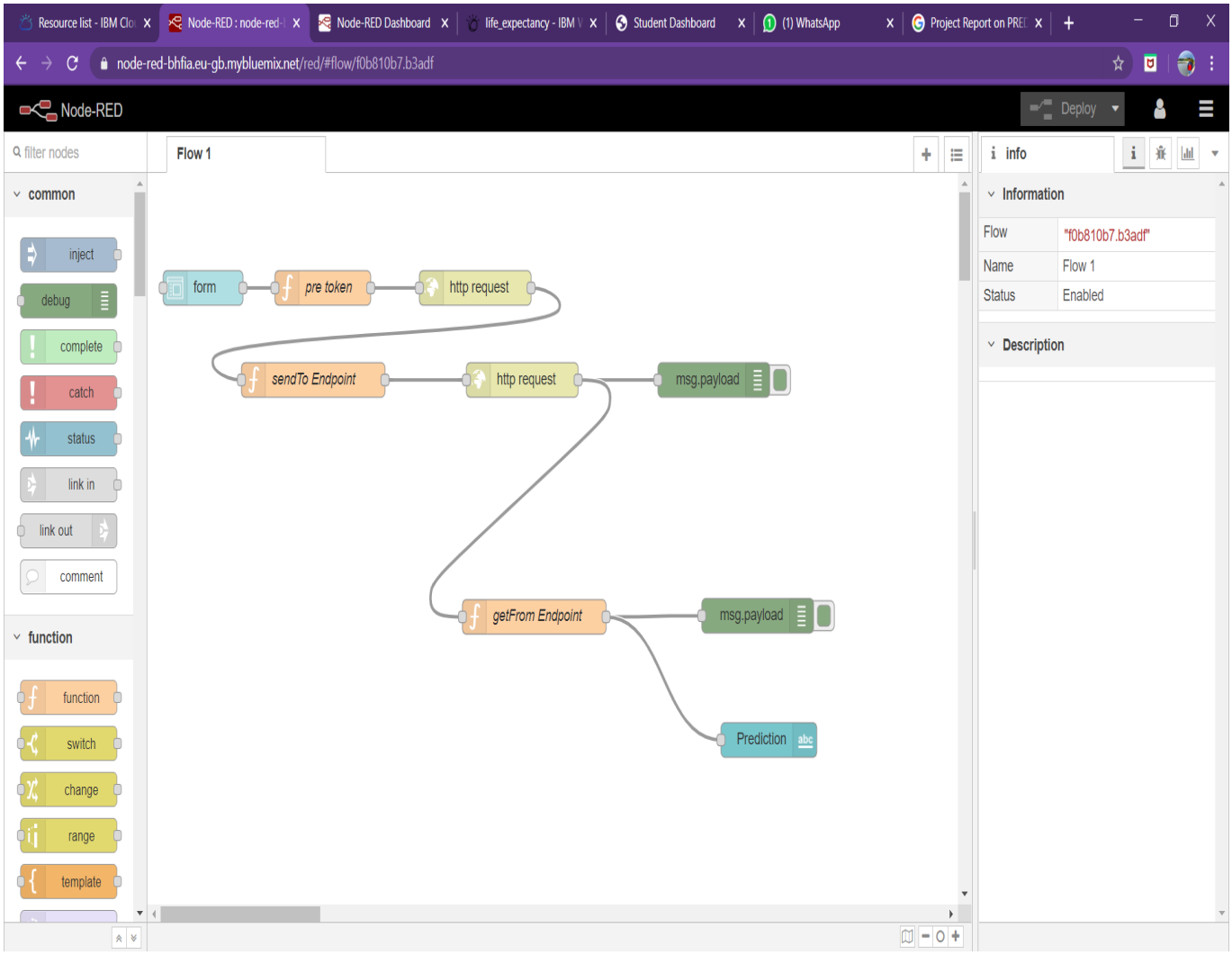


b) Choose the appropriate model for prediction: We can use any model for the the prediction person and with the help of it you can train and test the dataset. For the project I have been chosen the Linear Regression and Lasso and Ridge Regression for the development purpose.

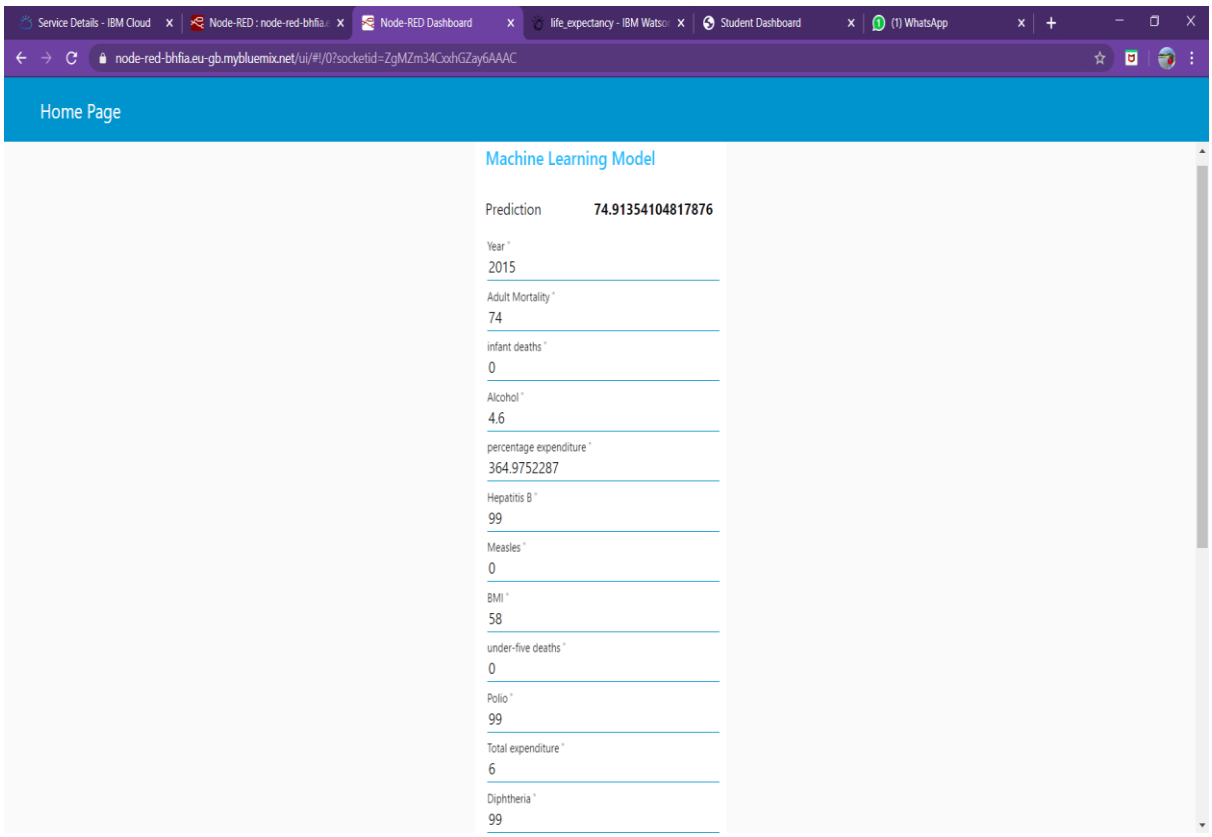
c) Deployment of Model: Once we are done with building the model. We must deploy the model. The deployed model will be stored in IBM Cloud Storage.

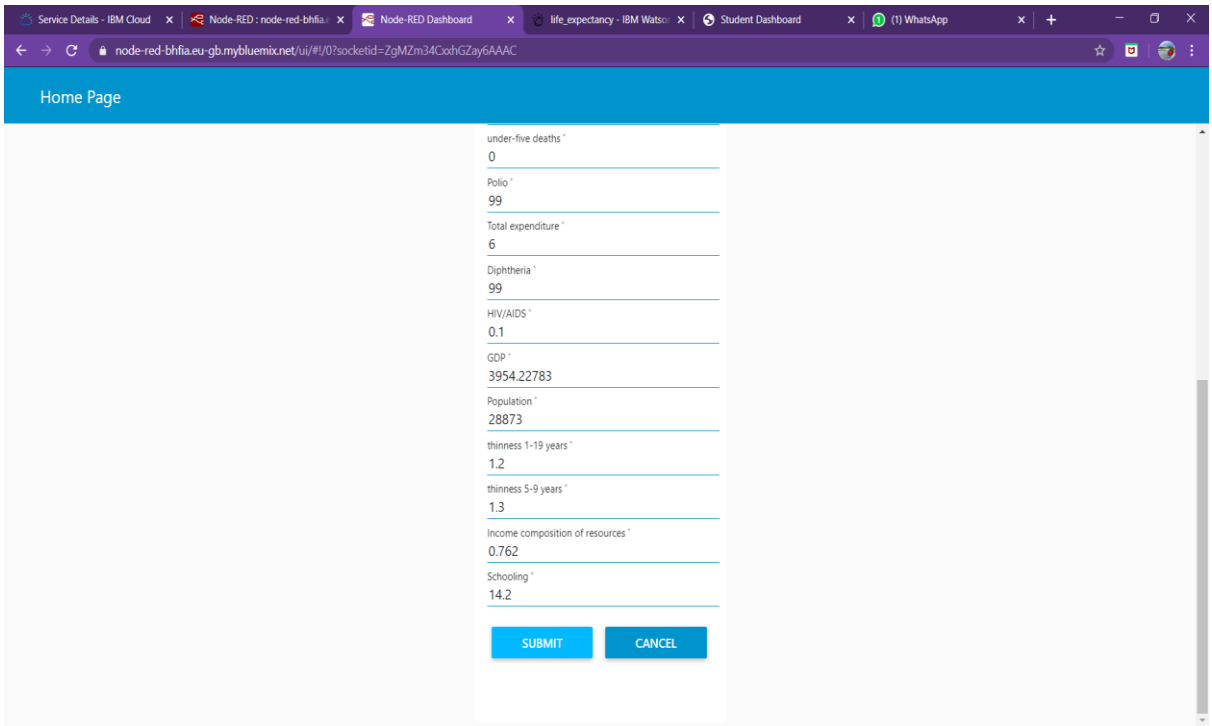


5) Create a Node-red Flow: Once the model is deployed you can create the node red flow to create an API for the model. Thus, API will act as a front end to the model.



Result of Machine Learning Model:
Output of the project:





Conclusion:

Prognostication of life expectancy is difficult for humans. Our research shows that machine learning and natural language processing techniques offer a feasible and promising approach to predicting life expectancy. The research has potential for real-life applications, such as supporting timely recognition of the right moment to start Advance Care Planning.

Future Scope:

Successful prediction model enhances the quality of life and death for palliative patients, by providing timely palliative care and documenting preferences regarding resuscitation and euthanasia, among other things.
Accurate prognosis of life expectancy is essential for general practitioners (GPs) to decide when to introduce the topic of ACP to the patient, and it is a key determinant in end-of-life decisions. Increasing the accuracy of prognoses has the potential to benefit patients in various ways by enabling more consistent ACP, earlier and better anticipation on palliative needs, and preventing excessive treatment.