PROJECT REPORT

TITLE: - Predicting Life Expectancy Using Machine Learning.

CATEGORY: - Machine Learning

SUBMITTED TO: - The SmartBridge

SUBMITTED BY: - Monika (monikajuneja16@gmail.com)

<u>Table of Contents</u>

1. Introduction

Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age and other demographic factors. Life Expectancy is one of the most important factors. It is often used to gauge the overall health of a community.

1.1 Overview

Life expectancy is a measure of premature death and it shows large differences in health across the world. Life expectancy is the key metric for assessing population health. Which focus solely at mortality at a young age, life expectancy captures the mortality along the entire life course. It tells us the average age of death in a population. Shifts in Life Expectancy are often used to describe trends in mortality.

1.2 Purpose

The aim of the project is to develop a model that will predict the life expectancy of the 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 in the dataset. After Successful Creation of the model, model will be deployed on IBM Watson and Node-red flow will be created.

2. Literature Review

2.1 Existing Problem

According to past Studies a lot of work has been done on Predicting Life Expectancy of the healthy Person and also of the people suffering from various Diseases. There

is a lot of shift in the Average Life Expectancy of the people of the Country over the Years. Various Factors affects the Life Expectancy of an individual. In previous studies not all factor has been taken into account such as HDI (Human Development Index) and also only data of one year per country was considered.
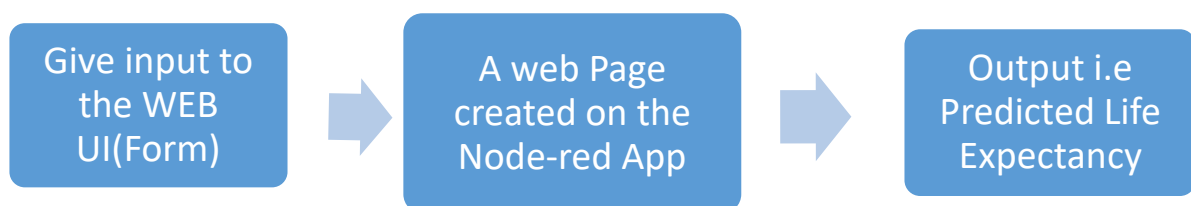
2.2 Required Solution

This Project will try to aim at accounting all the critical factors possible for efficient prediction. The project relies on the accuracy of data and for that the dataset has been taken from kaggle. The data-set related to life expectancy, health factors for 193 countries have been collected from the WHO data repository website and its corresponding economic data was collected from the United Nations website. Among all categories of health-related factors, only those critical factors were chosen which are more representative. The Dataset contain data of the years 2000 to 2015 of all the countries. Dataset covers all the factors possible and data will be preprocessed and various regression algorithms will be applied to train the model .Best model will be chosen and deployed.

IBM Cloud Platform is used to develop this project which provides a great environment and services.

3. <u>Theoretical Analysis</u>

3.1 Block Diagram

Give input to the WEB UI(Form)	→	A web Page created on the Node-red App	→	Output i.e Predicted Life Expectancy

3.2 Hardware/Software Designing

Software Requirements are:-

❖ Programming Language: Python

❖ Machine Learning libraries (NumPy, SciPy, matplotlib, scikit-learn, pandas)

❖ Cloud Service: IBM Cloud

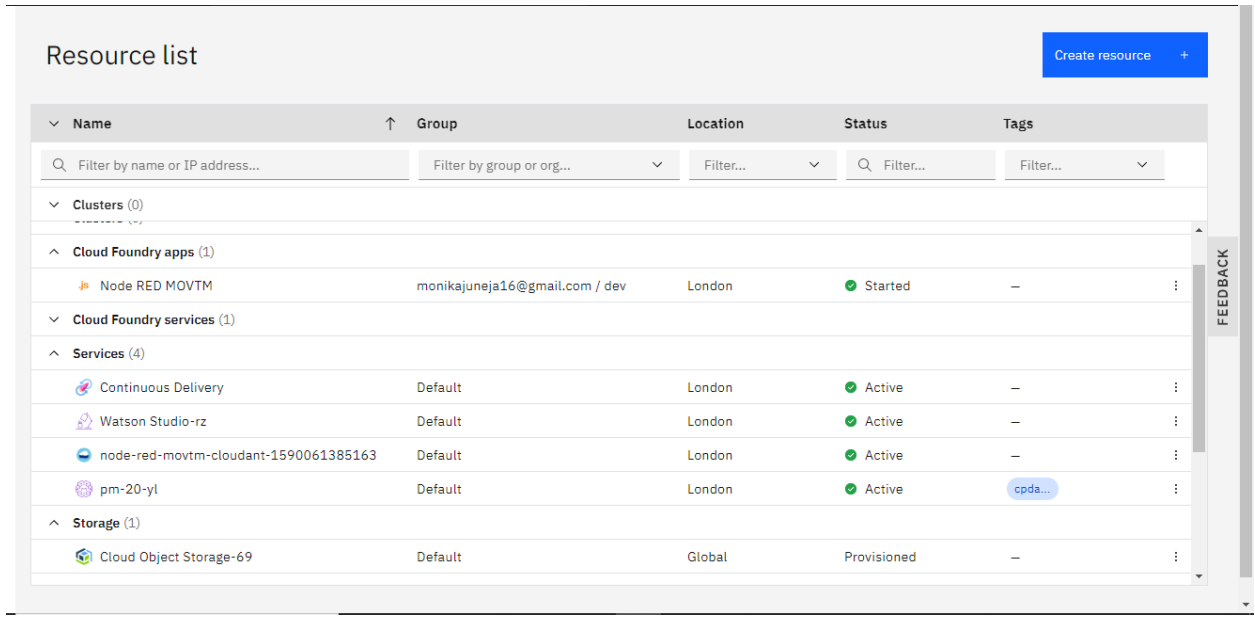❖ IDE like Jupyter

❖ IBM Watson

❖ Node-red App to create WEB UI

4. Experimental Investigations

4.1 Import and analyze the Data

❖ The Dataset is downloaded from the Link: -

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

❖ After importing and exploring the dataset.The dataset needs to preprocessed and cleaned which includes handling all the Null values.

❖ Dataset contains a lot of Null values, Null values are handled by: -

▪ Removing all the rows which have Null values in *GDP* and *Population* Column.

▪ Replacing rest of the Null values with the mean of the Column.

❖ Pipeline is used transform categorical columns to numerical columns and StandardScaler is applied on Numerical Columns.

❖ Then the dataset will be splitted into training and test set. Model is trained with train data and test set is used to test the model.

4.2 Creating an account and required services on IBM Cloud Platform
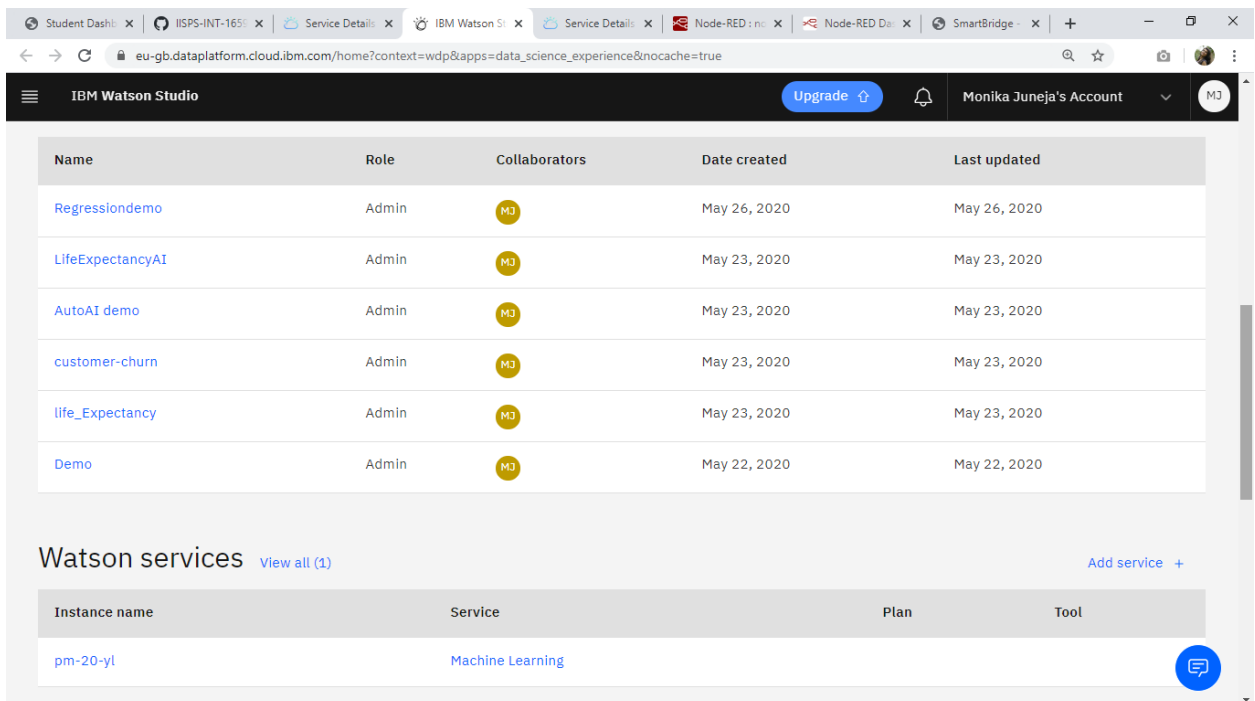
▪ Resource List:-

▪ Project and ML service on IBM Watson:-

- Assets in the Project:-

- Node-red Flow

- WEB UI

5. Flowchart

5.1 ML Flow:- Flow for creating and deploying the model.

Import the Dataset

⬇

Read and Explore the Data

⬇

Preprocess The Dataset

⬇

Split the dataset into training and test set

⬇

Train the Model

⬇

Make predictions for test data

⬇

Compare Predicted and Actual Results

⬇

Calculate MSE and R^2

⬇

Choose the best Algorithm and Deploy the Model
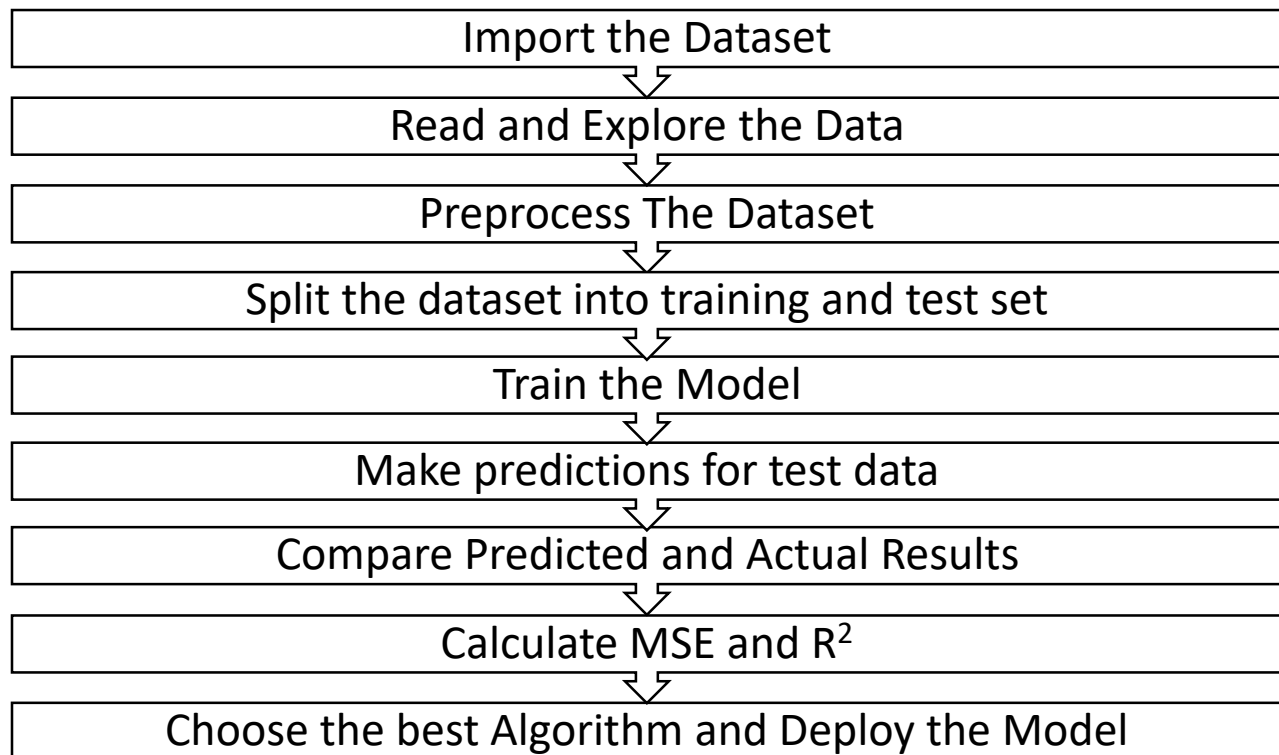
5.2 Node-red Flow
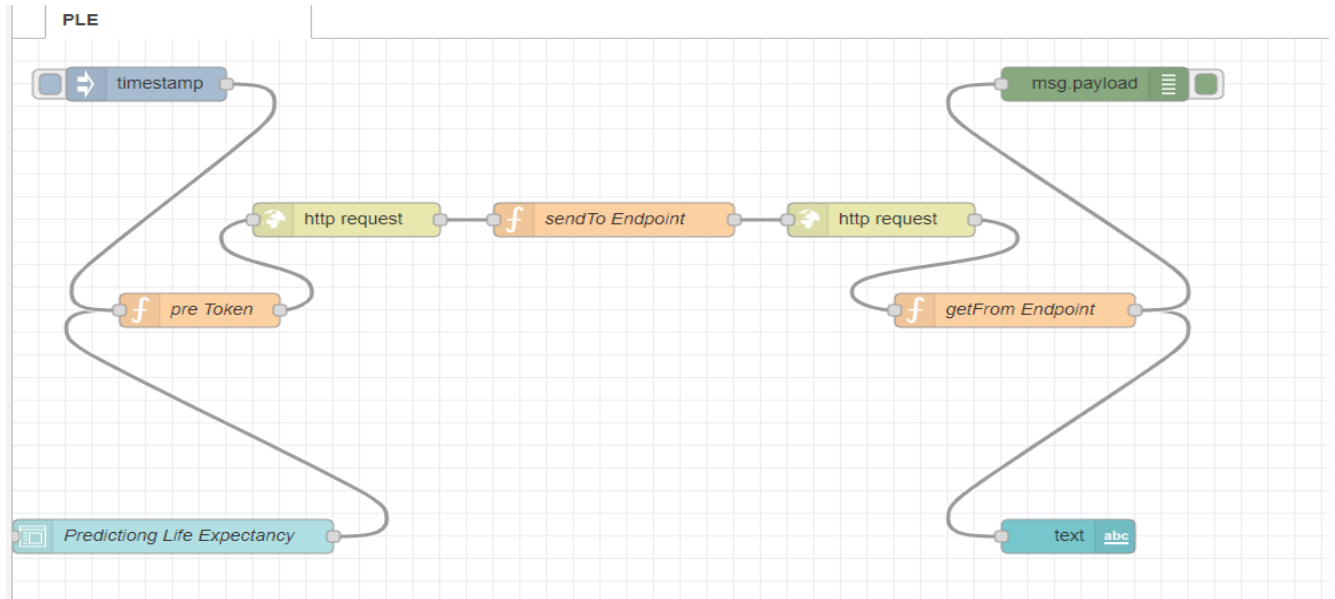
A Node-red flow is integrated with the Machine Learning Service.

6. RESULT

7. Advantages and Disadvantages

7.1 Advantages:-

❖ Predicting Life Expectancy can help Government and other Organizations to target the variables that need improvement. Where should Government Focus and what policies can be made.

❖ It helps to evaluate how different factors impact Life Expectancy. It helps in Understanding Health Inequalities.

❖ Anyone can predict the Life Expectancy of the people of their country or any other country with the required data.

❖ It's free and no cost is included.

❖ User Friendly Interface, It has a simple and easy to use interface where user needs to provide required Input and Output i.e Life Expectancy will be generated.

7.2 Disadvantages:-

❖ The Model has an accuracy of 97% and it does not guarantee to produce accurate value.

❖ The Model is highly dependent on the Input provided by the user. So, it's necessary the input is valid and true.

8. Applications

❖ It can be used by Government of the Country to make different policies and laws.

❖ It can act as an important statistics to deal with Health Inequalities in the country and the world.

❖ It can be used by Insurance Companies.

❖ It can play important role in Country development Process.

9. Conclusion

After importing, exploring and preprocessing the data. Dataset is splitted in to training and test set and different Machine learning algorithms has been applied. **The best algorithm which produce minimum Mean Squared Error (3.12) and maximum R^2 (97%) is Extra Trees Regressor**. So, this model has been chosen and deployed. Then simple User Interface is created through Node-red App. Machine Learning Model is integrated with the Node-red WEB User Interface.

The User Interface is very friendly and simple which enables any user with no technical background knowledge to predict Life expectancy efficiently.

10. Future Scope

In Future, several work can be done such as

❖ Feature Engineering to select only those features which are required and removing all the unnecessary features.

❖ Visualizing Dashboard can also be created to view the results and to see which feature has more impact on the result.

❖ Interactive and Attractive UI.

11. Biblography

■ https://www.ibm.com/cloud/get-started

■ https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/

■ https://github.com/watson-developer-cloud/node-red-labs

■ https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be

- https://www.w3schools.com/howto/howto_make_a_website.asp

- https://www.ibm.com/watson/products-services

- https://www.youtube.com/watch?v=W3iPbFTAAds&feature=youtu.be

- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/

- https://www.youtube.com/watch?v=NmdjtezQMSM

- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html

- https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L

- https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2

- https://www.youtube.com/watch?v=Jtej3Y6uUng

- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service

12. Appendix

12.1 Python Code

github link to my code: - https://github.com/SmartPracticeschool/llSPS-INT-1905-Predicting-Life-Expectancy-using-Machine-Learning/tree/master/Notebook%20File

12.2 Node-red flow

[{"id":"2dc227ac.c86478","type":"ui_form","z":"1a97385.daeeac8","name":"Predict iong Life Expectancy","label":"Predicting Life Expectancy of people of the country","group":"2d99c54c.4dfb7a","order":1,"width":0,"height":0,"options":[{"label":"Country","value":"c","type":"text","required":true,"rows":null},{"label":"Year","value":"year","type":"number","required":true,"rows":null},{"label":"Status","value":"s","type":"text","required":true,"rows":null},{"label":"Adult Mortality","value":"a","type":"number","required":true,"rows":null},{"label":"infant deaths","value":"i","type":"number","required":true,"rows":null},{"label":"Alcohol"

,"value":"alchl","type":"number","required":true,"rows":null},{"label":"percentage expenditure","value":"pe","type":"number","required":true,"rows":null},{"label":"Hepatitis B","value":"hb","type":"number","required":true,"rows":null},{"label":"Measles ","value":"m","type":"number","required":true,"rows":null},{"label":"   BMI ","value":"bmi","type":"number","required":true,"rows":null},{"label":"under-five deaths ","value":"ufd","type":"number","required":true,"rows":null},{"label":"Polio","value":"p","type":"number","required":true,"rows":null},{"label":"Total expenditure","value":"te","type":"number","required":true,"rows":null},{"label":"Diphtheria ","value":"dpthra","type":"number","required":true,"rows":null},{"label":"HIV/AIDS","value":"hiv","type":"number","required":true,"rows":null},{"label":"GDP","value":"gdp","type":"number","required":true,"rows":null},{"label":"Population ","value":"popltn","type":"number","required":true,"rows":null},{"label":" thinness 119years","value":"thns1_19","type":"number","required":true,"rows":null},{"label":"thinness5years","value":"thns5_9","type":"number","required":true,"rows":null},{"label":"Incomecompositionofresources","value":"ic","type":"number","required":true,"rows":null},{"label":"Schooling","value":"schlng","type":"number","required":true,"rows":null}],"formValue":{"c":"","year":"","s":"","a":"","i":"","alchl":"","pe":"","hb":"","m":"","bmi":"","ufd":"","p":"","te":"","dpthra":"","hiv":"","gdp":"","popltn":"","thns1_19":"","thns5_9":"","ic":"","schlng":""},"payload":"","submit":"submit","cancel":"cancel","topic":"TestingofModel","x":120,"y":440,"wires":[["4c55aa1e.b321c4"]]},{"id":"2d99c54c.4dfb7a","type":"ui_group","z":"","name":"","tab":"924aedde.d1b57","order":1,"disp":true,"width":"6","collapse":true},{"id":"924aedde.d1b57","type":"ui_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]