PROJECT REPORT

PROJECT NAME - PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

PROJECT ID - SPS_PRO_215

SUBMITTED BY - DHRUVILSINH JHALA

Index:

1. Introduction

1.1. Overview

Life expectancy refers to the number of years a person is expected to live. In mathematical terms, life expectancy refers to the expected number of years remaining for an individual at any given age.

The life expectancy for a particular person or population group depends on several variables such as their lifestyle, access to healthcare, diet, economic status and the relevant mortality and morbidity data. However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

In order to predict life expectancy rate of a given country, we will be using Machine Learning algorithms to draw inferences from the given dataset and give an output. For better usability by the customer, we are also going to be creating a UI for the user to interact with using Node-Red.

1.2. Purpose

- Economic growth

Predicting life expectancy would play a vital role in judging the growth and development of the economy. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country. The knowledge asset of a country increases with the number of individuals in a country.

- Population Growth

Helps the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skillset acquired by people over many years.

- Personal growth

This project would also help an individual assess his/her lifestyle choices and alter them accordingly to lead a longer and healthier life. It would make them more aware of their general health and its improvement or deterioration over time.

- Growth in Health Sector

Based on the factors used to calculate life expectancy of an individual and the outcome, health care will be able to fund and provide better services to those with greater need.

- Insurance Companies

Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

2. Literature Survey

2.1. Existing Problem

Predicting a human's life expectancy has been a long-term question to humankind. Many calculations and research have been done to create an equation despite it being impractical to simplify these variables into one equation. Currently there are various smart devices and applications such as smartphone apps and wearable devices that provide wellness and fitness tracking. Some apps provide health related data such as sleep monitoring, heart rate measuring, and calorie expenditure collected and processed by the devices and servers in the cloud. However no existing works provide the Personalized Life expectancy.

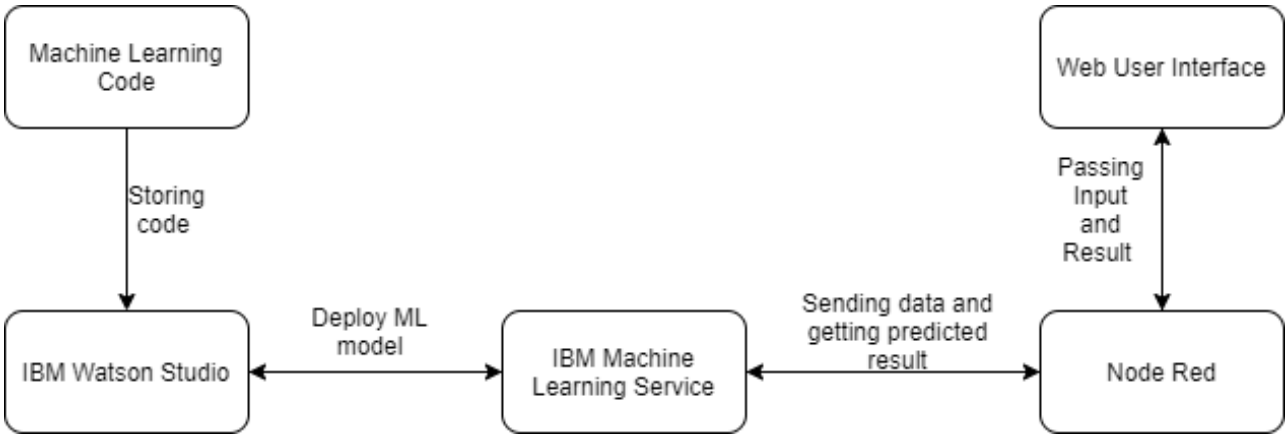
2.2. Proposed Solution

There has been an explosion of breakthroughs in the field of Machine Learning over the past few years. Machine Learning algorithms are capable of a lot and can-do wonders for the healthcare sector.

The proposed solution involves the use of Machine Learning algorithms specifically Regression models. Life expectancy is highly correlated over time among countries and its characteristics. These associations can be used to improve forecasts. Here we propose a method for forecasting life expectancy of an individual from a country taking into certain factors such as Adult Mortality rate, Infant deaths, Alcohol, Hepatitis B, Measles, BMI, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP of a country, Population, Income composition of resources, Schooling and status of the country in terms of Developing or Developed. This machine learning model will be made accessible to the users by integrating it with Node-Red to create an interactive and user-friendly User Interface.

3. Theoretical Analysis

3.1. Block Diagram



3.2. Hardware/ Software Designing

1) Model Designing (Watson Studio) :

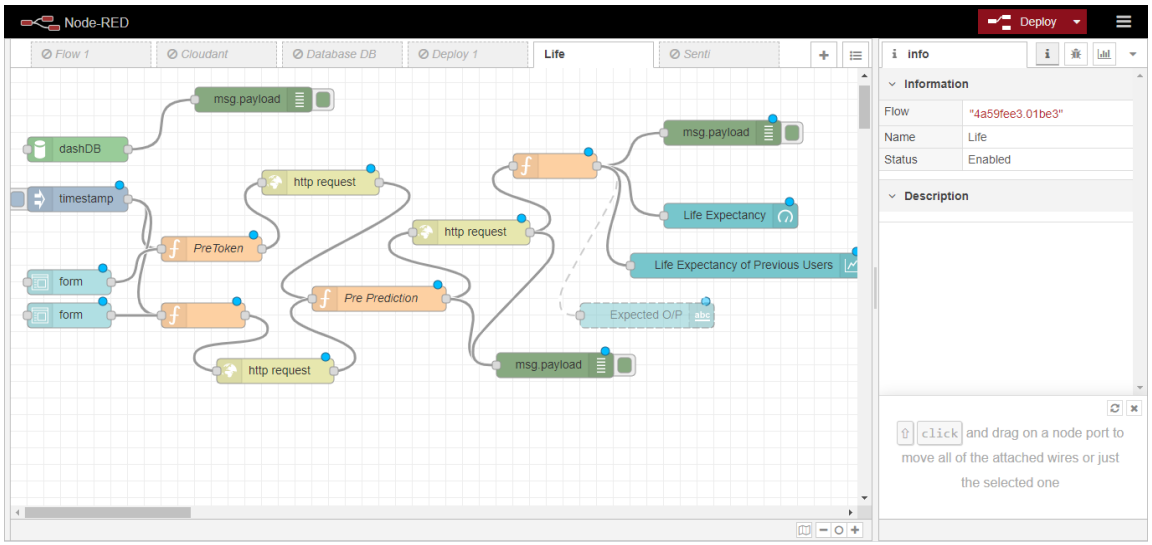
Steps: New Project, Create an empty Project, Give project name, Click Create, Add to Project, Notebook

2) Model Deployment (Machine Learning Service):

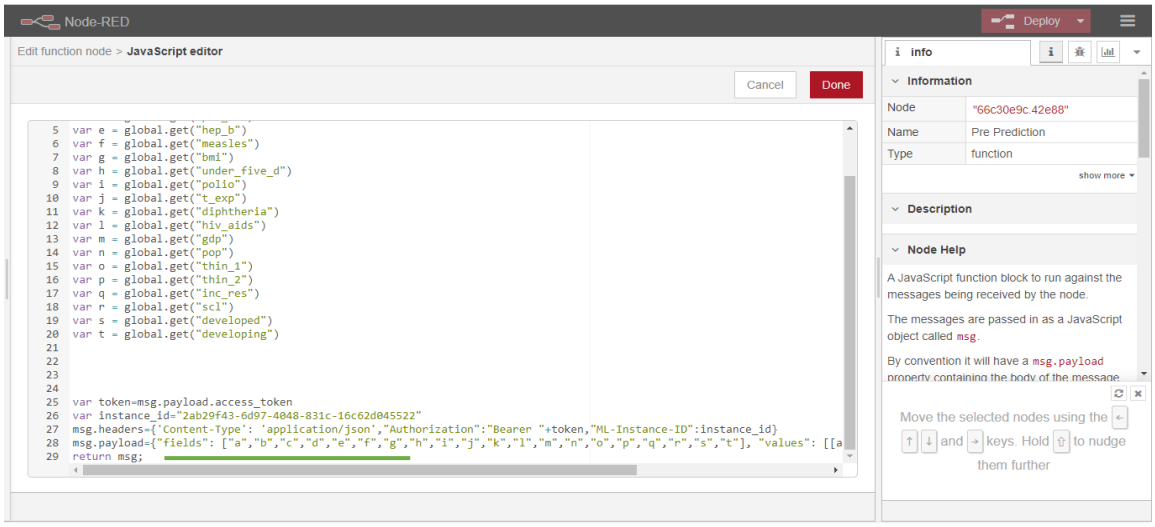
Services, Machine Learning Service, Click Create, Service Credentials, Copy the credentials

3) User Interface Integration with ML Model (Node Red) :

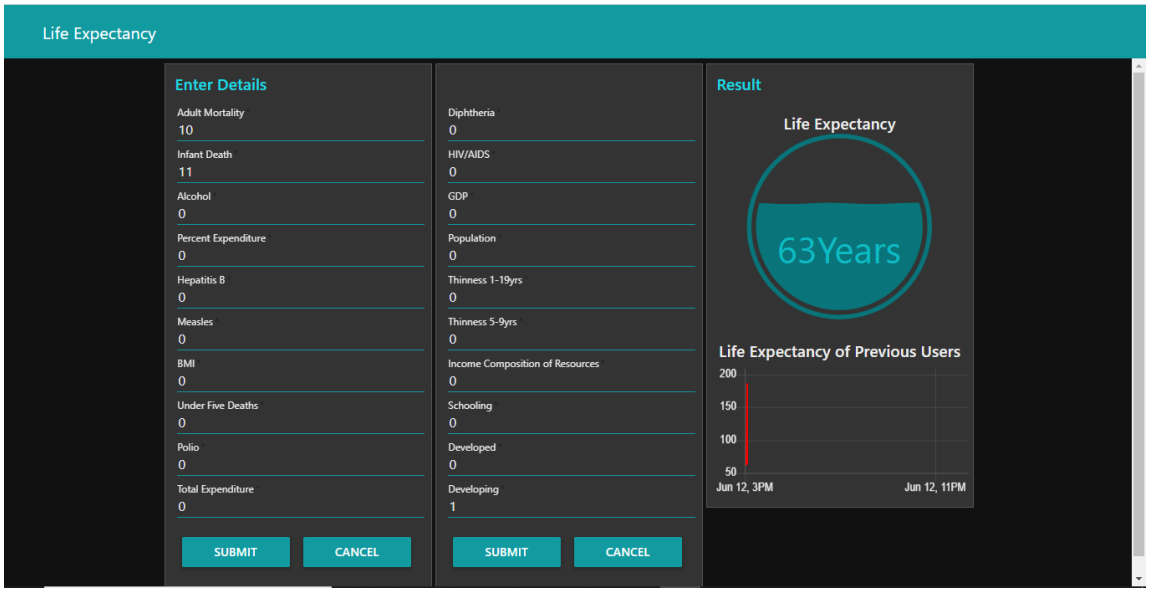
Search Node Red and open, Click Create , Go to URL, Design Flow



Creating Node Red Flow



Passing Data to Machine Learning Model



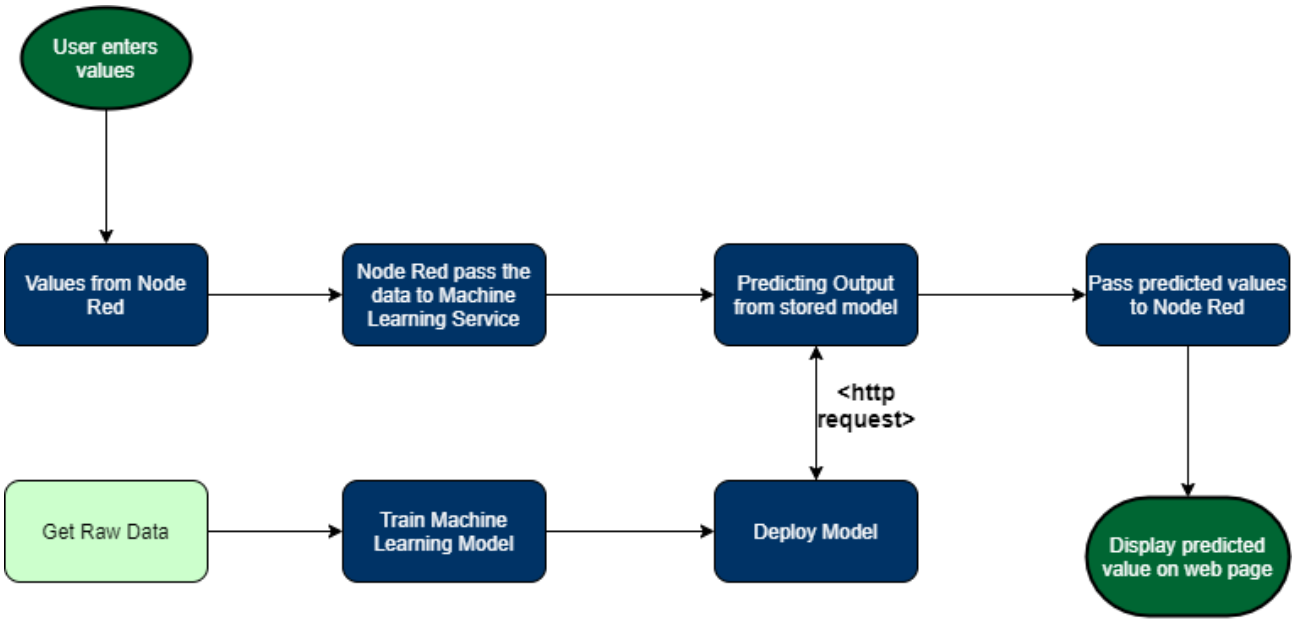
Deploy Flow

4. Experimental Investigation

Using Jupyter Notebook in Watson studio for model formation. Comparing multiple linear regression, polynomial regression, Random Forest Regression, Decision Tree Regression models using Cross Validation Score and R square scores.

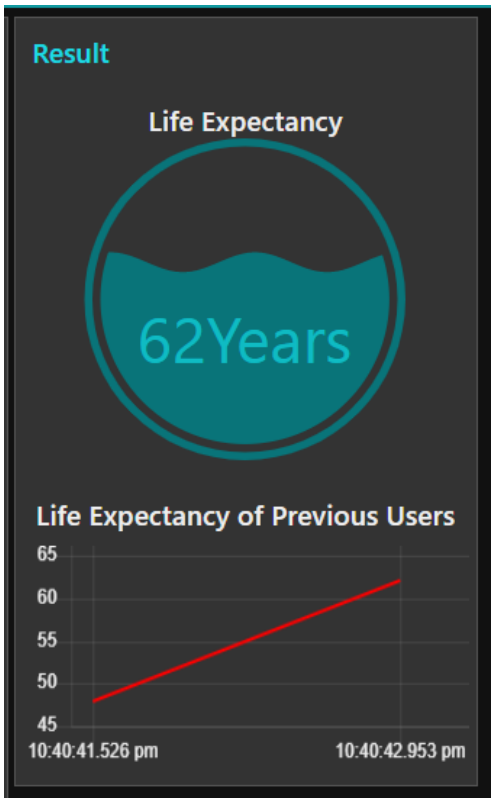
	R^2 score on training data	**R^2 score on test data**	**Mean squared error**
Linear Regression	0.92	0.91	9.8
Polynomial Regression	1	-18.45	1620.72
Random Forest Regression	0.98	0.88	6.71
Decision Tree Regression	0.98	0.80	16.59

5. Flowchart



6. Result

Enter the values in the two forms regarding Adult Mortality rate, Infant deaths, Alcohol, Hepatitis B, Measles, BMI, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP of a country, Population, Income composition of resources, Schooling and status of the country in terms of Developing or Developed. Then after pressing submit within a few seconds the result will be displayed. Result is 89% accurate.



7. 1. Advantages:

Machine learning model in this project is accurate up to 89%, hence good prediction can be done.

The application learns the patterns and trends hidden within the data without human intervention which makes predicting much simpler and easier. The more data is fed to the algorithm, the higher the accuracy of the algorithm is. It is also the key component in technologies for automation.

Using IBM Cloud, Node Red helps provide a user interface in a webpage to visualize the real-time process.

7. 2. Disadvantages:

In some exceptional cases of input data, the prediction may have a big amount of error.

The dataset used in this project will need to be updated from time to time to get new information on factors like GDP, schooling, etc as these factors change overtime.

8. Applications:

1) Personal Life Expectancy:
   Individuals can predict their own life expectancy by filling data in the form and get to know their life expectancy. They can take predictive measures accordingly.

2) Government:
   It could help the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skillset acquired by people over many years. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country. The knowledge asset of a country increases with the number of individuals in a country.

3) Health Sector:
   Based on the factors used to calculate life expectancy of an individual and the outcome, health care will be able to fund and provide better services to those with greater need.

4) Insurance Companies:
   Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

9. Conclusion:

This project is helpful in many ways in health care sector as predicting life expectancy can help alter our life. The machine learning regression model is highly efficient in predicting the output and node red provides great user interface.

10. Bibliography:

- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://www.kaggle.com/kumarajarshi/life-expectancy-who
- https://medium.com/@fraserjames06/life-expectancy-trends-and-disparities-my-first-data-science-project-b77bf28336d4
- https://www.ibm.com/watson/products-services