

Project Report

On

**Predicting Life Expectancy Using Machine
Learning**

Under

Remote Summer Internship Program 2020

By SmartInternz

Presented by

SOMYA DAS

Predicting Life Expectancy using Machine Learning - SB47315

1.INTRODUCTION:

1.1.Overview:

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. The aim of this project is 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 provided.

1.2.Purpose:

The purpose of the project is to predict the Life Expectancy based on user's input. The dataset used for the training for the model was downloaded from kaggle.com. In this project an UI is designed with the help of Node-RED application provided by IBM Cloud where user can enter input values like country, year, adult mortality, deaths, alcohol, expenditure, BMI, diseases, GDP, population, thinness, income, schooling, etc. and output will be generated in the form of predicted life expectancy. Random Forest Regression technique is used to predict the life expectancy which is deployed on IBM Cloud using services like IBM Watson Studio and IBM Watson Machine Learning Service. IBM Cloud also provide IBM AutoAI experiment which allow to create a machine learning model without any knowledge of coding and can deploy the model and integrate with Node-RED UI .

2.LITERATURE SURVEY:

2.1.Existing Problem:

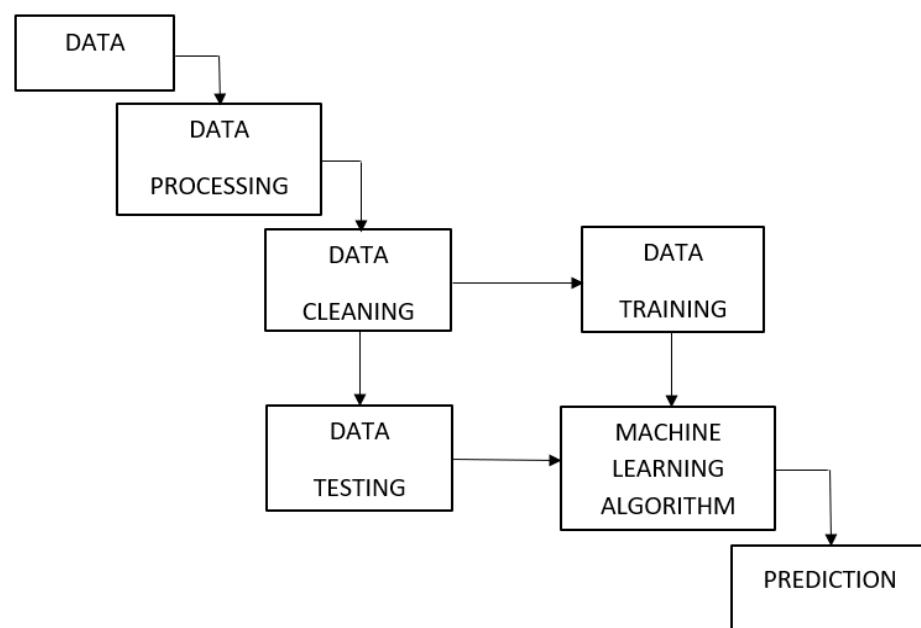
Currently the life expectancy of a person is calculated on the basis of already collected data and the life expectancy in future cannot be predicted exactly due to lack of data and technology. Hence, many problem may arise because people are not prepared to tackle the issues due to the change in factors that may affect the life expectancy.

2.2.Proposed Solution:

If the life expectancy of a person can be predicted for the next coming years then we can have an idea about the factors that may affect the life expectancy either in positive or negative way and accordingly can take necessary actions and precautions.

3.THEORETICAL ANALYSIS:

3.1.Block Diagram:



3.2.Hardware/Software Design:

Technical Requirements:

- The GUI must be integrated with the backend trained model.
- The clean data set must be given to model before training.

Software Requirements:

- SmartInternz Project Workspace
- Github
- Zoho Writer
- Slack Channel
- Microsoft Excel
- IBM Cloud Services
- IBM Watson Studio
- IBM Watson Machine Learning
- IBM Node-RED application
- IBM AutoAI Experiment
- Python IDE

4.EXPERIMENTAL INVESTIGATIONS:

The dataset was available in SmartInternz Project Workspace under the milestones. After downloading it I analysed the data and find out the relation among the columns and which column is necessary for project and then clean it accordingly. Then found the best model to fit my data to give accurate output. After completing the backend it was time to design front end UI which was done with the help of Node-RED application. All the tasks were completed with the help of IBM Cloud Services.

5.FLOWCHART:

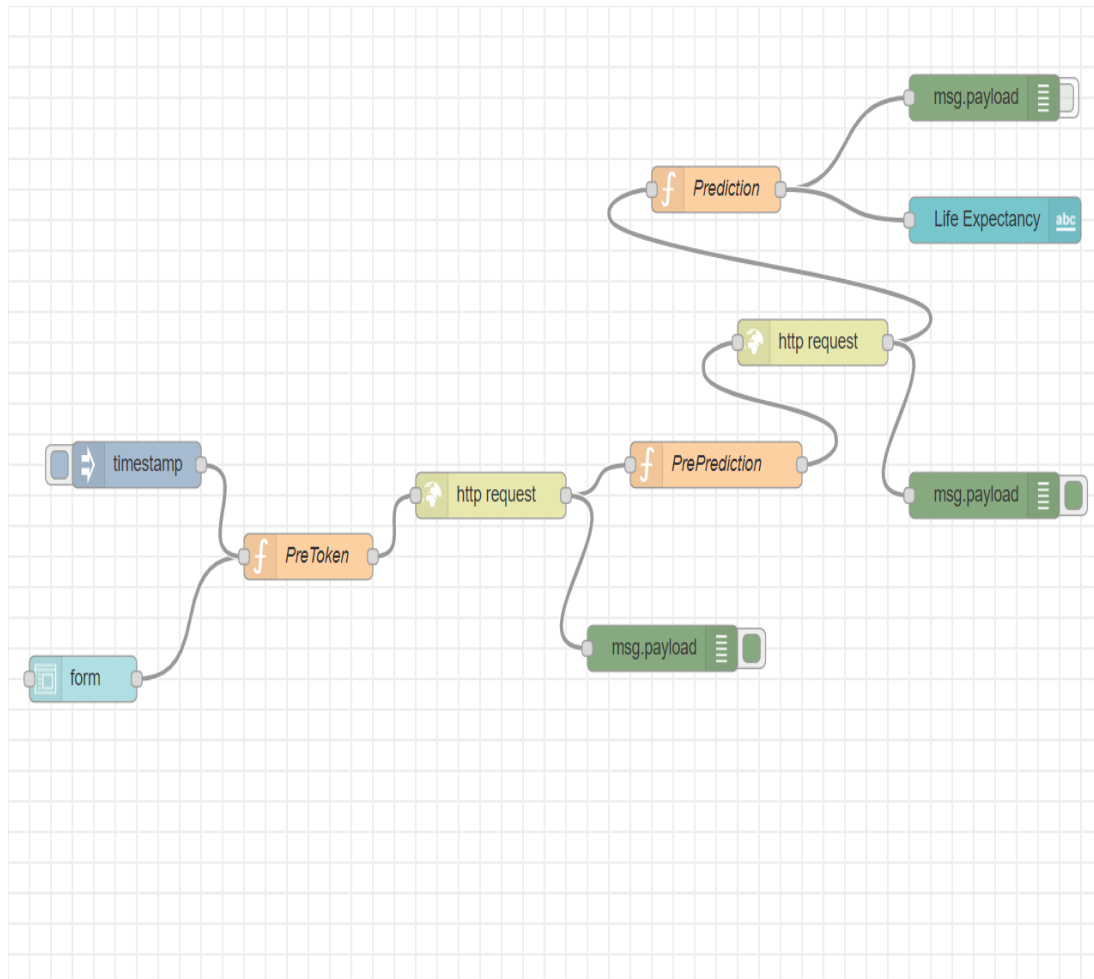


Figure. Node-RED Flow

6.RESULTS:

The UI is used to take input from user, it is integrated with trained Machine Learning model present at the backend, IBM Watson Notebook. When user fill provide input to the form and click on Predict button then predicted life expectancy will be displayed at the Life Expectancy label.

Life Expectancy Prediction using Python

Form

Life Expectancy **66.52599999999994**

Adult Mortality **196**

Infant Deaths **1200**

Alcohol **2.77**

Percentage Expenditure **57.7336**

Hepatitis B **38**

Measles **31458**

BMI **15.9**

Under-Five Deaths **1600**

Polio **76**

Total Expenditure **4.28**

Diphtheria **79**

HIV/AIDS **0.2**

Figure. Life Expectancy Prediction using Python

Life Expectancy Prediction using AutoAI

Form

Life Expectancy **66.37000045776367**

Country **India**

Year **2010**

Status **Developing**

Adult Mortality **196**

Infant Deaths **1200**

Alcohol **2.77**

Percentage Expenditure **57.7336**

Hepatitis B **38**

Measles **31458**

BMI **15.9**

Under-Five Deaths **1600**

Polio **76**

Figure. Life Expectancy Prediction using AutoAI

7.ADVANTAGES AND DISADVANTAGES:

7.1.Advantages:

- Easy process of unstructured data
- Fills human limitations
- Handle enormous quantities of data
- Easy for users to interact with the model via UI
- User-friendly
- Easy to build and deploy
- Requires less storage space

7.2.Disadvantages:

- IBM Cloud is only available in English
- Requires Internet Connection
- Error in data can result in wrong prediction
- Accuracy is not 100%

8.APPLICATIONS:

This project can be used by Researchers to make meaningful research out of it and thus, bring some improvement that will help in increasing the expectancy considering the impact of a specific factor on the average lifespan of a people in a specific country.

9.CONCLUSION:

We can conclude that this project will be able to predict the life expectancy of a person using machine learning with the help of various factors that have significant impact on lifespan such as country, year, adult mortality, deaths, alcohol, expenditure, BMI, diseases, GDP, population, thinness, income, schooling, and many more.

10.FUTURE SCOPE:

The machine learning model used for the project can be trained well using more accurate data and trained model can be used to increase the accuracy. This model can be further used to predict the lifespan of plant, trees and animals using dataset for it.

11.BIBLIOGRAPHY:

- SmartInternz Dashboard:-
<https://smartinternz.com/Student/dashboard>
- Dataset Reference:-
<https://www.kaggle.com/kumarajarshi/life-expectancy-who>
- IBM Cloud:-
<https://www.ibm.com/cloud/get-started>
- Node-RED application:-
<https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>
- Watson Studio:-
<https://bookdown.org/caoying4work/watsonstudio-workshop/>
- IBM Watson Machine Learning:-
<https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/>
- AutoAI:-
<https://developer.ibm.com/tutorials/watson-studio-auto-ai/>
- YouTube:-
https://www.youtube.com/watch?v=KdSqvpVvmgs&list=PLjIJJFiCdXMIInlWHEsgsnY3P55kGdSDh_

APPENDIX

Node-RED GUI link:-

<https://node-red-dmavo.mybluemix.net/ui/>

Source code link:-

<https://github.com/SmartPracticeschool/IIIPS-INT-2028-Predicting-Life-Expectancy-using-Machine-Learning>

Project Demonstration Video link:-

<https://drive.google.com/file/d/19TRo0mcITT0OgZk-IhY2xk7uYMTqJ8Dx/view?usp=sharing>