**PROJECT REPORT**

**SMARTINTERNZ INTERNSHIP**

**By**: Shivam Yadav

**Predicting Life Expectancy using Machine Learning**

**Internship Title:** Predicting Life Expectancy using Machine Learning - SB30615

**Project Title:** Predicting Life Expectancy using Machine Learning

**1. INTRODUCTION:**

It is Machine Learning project that uses Python3, IBM Cloud services like Node Red and IBM Watson Studio to Predict Life Expectancy of a Country. In this project I am going to use Linear Regression Model to predict Life Expectancy using various features given in the dataset.

**2. LITERATURE SURVEY:**

**2.1 EXISTING PROBLEM**

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. Life expectancy is a measure that is often used to gauge the overall health of a community. Life expectancy at birth measures health status across all age groups. Shifts in life expectancy are often used to describe trends in mortality.

**2.2 PROPOSED SOLUTION**

To design a machine learning model using regression techniques that can predict the life expectancy of a country given the various factors that affect the health of its citizens. Using this model, we can assess the extent of effect these factors have on the average life expectancy. This can give us an idea about the changes that need to be brought about in order to improve the health conditions of people around the world.

**3. THEORETICAL ANALYSIS**

**3.1 BLOCK DIAGRAM**

Data Pre-Processing and Cleaning

Import Dataset

Import Libraries

END

Model Deployement

Model Training

START

Model Testing

Model Validation

Model Selection

Feature Extraction

**3.2 HARDWARE / SOFTWARE DESIGNING:**

3.2.1 HARDWARE REQUIREMENTS:

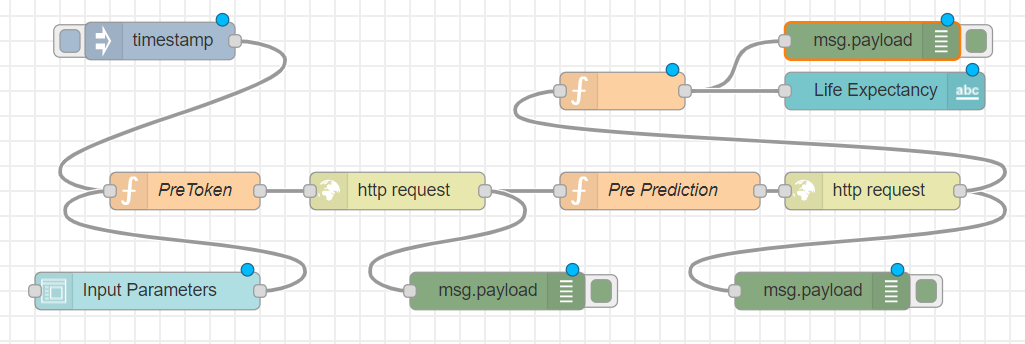
Any working Laptop/PC with good graphics card and stable internet connection.

3.2.2 SOFTWARE REQUIREMENTS:

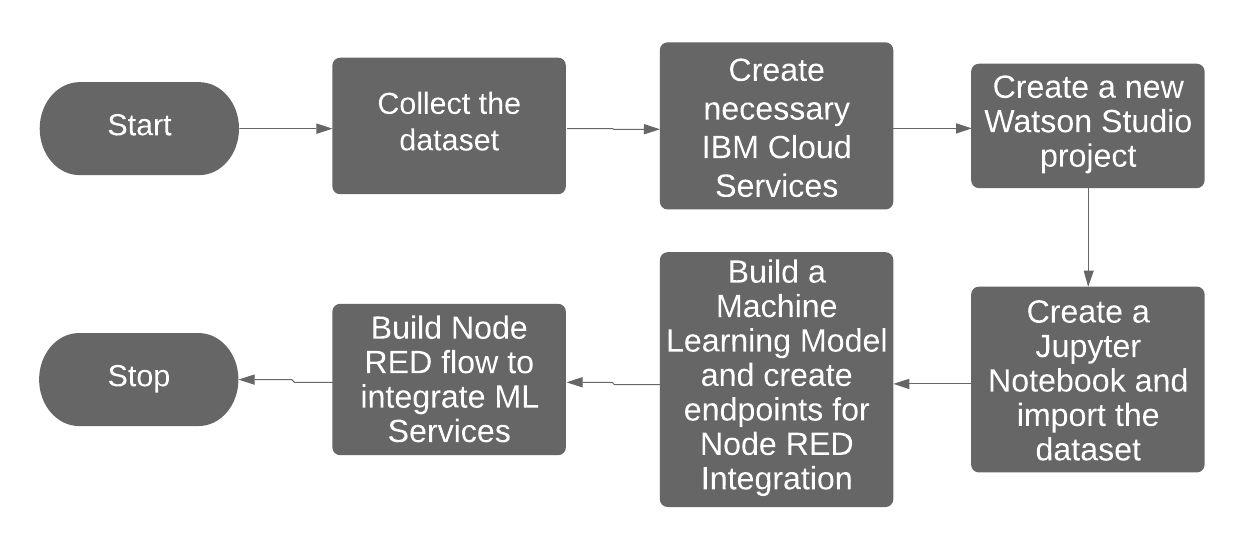
Python3, IBM Cloud, Zoho Writer, GitHub, IBM Watson, Node Red

**4. EXPERIMENTAL INVESTIGATIONS**

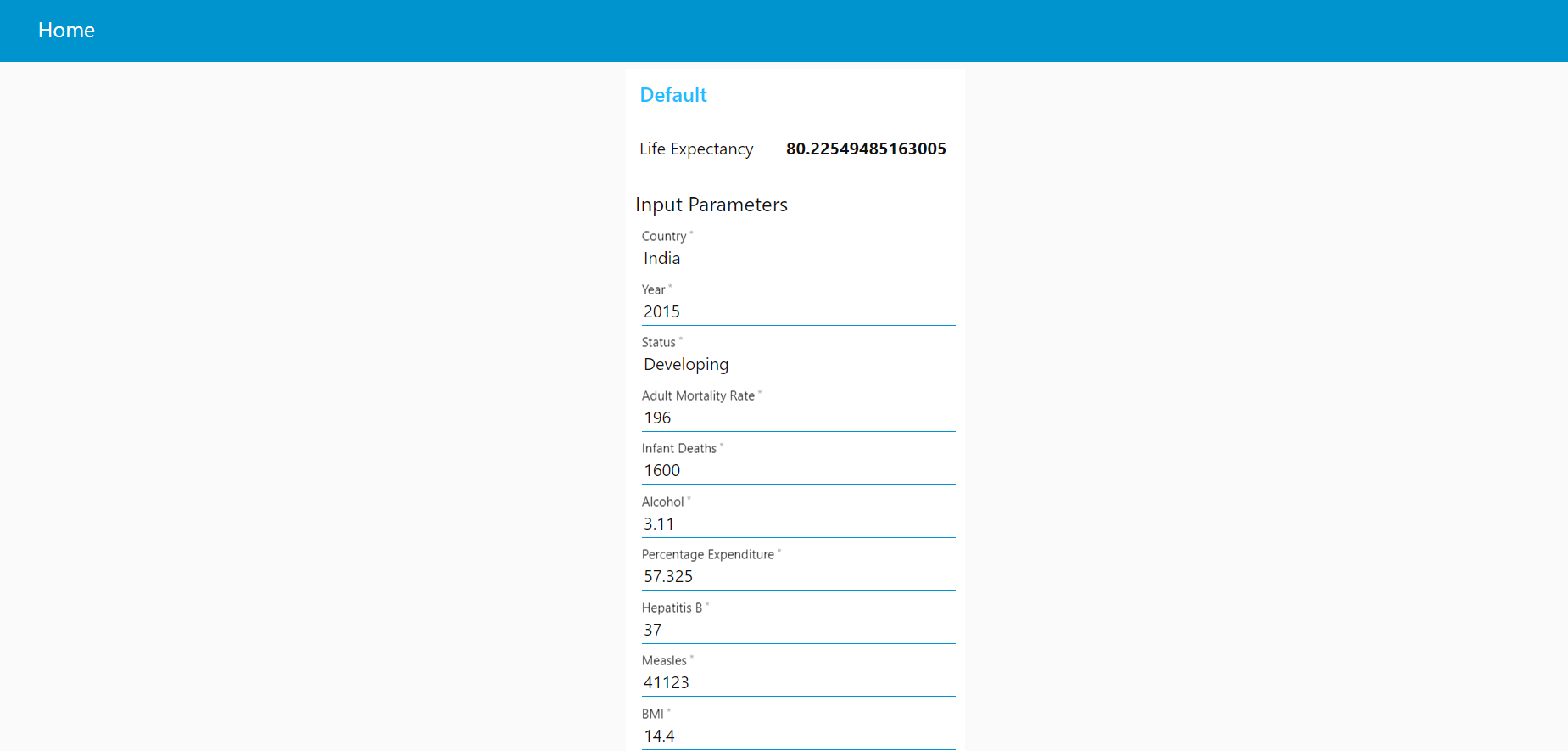
1. Collect Dataset from Kaggle.
2. Create necessary IBM Cloud Services.
3. Create new Watson Studio Project and then create new Jupyter Notebook and import the Dataset
4. Apply preprocessing and Data cleaning method on dataset and fill all the NAN values in the dataset for better performance of model.
5. Take Life Expectancy column as Output and all the other column as Input features.
6. Divide the Dataset into training and testing using sklearn.model\_selection library.
7. Use the Linear Regression model on training and testing dataset and create a model.
8. Using model predict the accuracy of the model.
9. Create the endpoints of the notebook to link it with the Node Red for deploying model.
10. Create the Node Red flow with form, function, HTTP, timestamp, text and Debug node.



1. Deploy the Node Red flow and give the inputs and see the Predicted output.

**5. FLOWCHART**

**6. RESULT**



**7. ADVANTAGES & DISADVANTAGES**

**7.1 ADVANTAGES**

* It provides a powerful statistical method that allows a business to examine the relationship between two or more variables of interest.
* It helps determine which factors matter most, which it can ignore, and how those factors interact with each other.

**7.2 DISADVANTAGES**

* It cannot be used in case of qualitative phenomenon.
* The functional relationship that is established between any two or more variables on the basis of some limited data may not hold good if more and more data are taken into consideration
* It is assumed that the cause and effect relationship between the variables remains unchanged. This assumption may not always hold good and hence estimation of the values of a variable made on the basis of the regression equation may lead to erroneous and misleading results.

**8. APPLICATIONS**

* It shows how patterns of change in life expectancy are described by trajectories of mortality improvements over age and time.
* Life expectancy gives an idea of the health issues prevalent in particular country and the healthcare system available to combat these issues.
* Life expectancy provides a useful measure of average life spans and gives insights into uncertainty about age at death.

**9. CONCLUSION**

We implemented a machine learning model using regression analysis to predict the life expectancy of a particular country. We used IBM Cloud, IBM Watson and Node RED to work on this project. The dataset we used related the life expectancy and health factors for 193 countries from the WHO data repository website and its corresponding economic data was collected from United Nation website. The following factors were applied to predict the life expectancy: Country, Year, Status, Adult Mortality, Infant Deaths, Alcohol, Percentage Expenditure, Hepatitis B, Measles, BMI, Under-Five Deaths, Polio, Total Expenditure, Diphtheria, HIV/AIDS, GDP, Population, Thinness 1-19 years, Thinness 5-9 years, Income Composition of Resources, Schooling.

**10. FUTURE SCOPE**

In order to extend this work, we can increase the size of the dataset and observe the effect on the results. We can also apply other Machine Learning models in order to obtain higher accuracy.

**11. BIBLIOGRAPHY**

**APPENDIX**

1. Source Code : <https://github.com/Shiv1O1/Predicting-Life-Expectancy-using-ML/blob/master/SmartBridge_ML.ipynb>
2. Dataset : <https://github.com/Shiv1O1/Predicting-Life-Expectancy-using-ML/blob/master/Life_Expectancy_Data.csv>