Project Documentation

Project Report on

Predicting Life Expectancy using Machine Learning

Under

Remote Summer Internship Program 2020 by SmartInternz

Project by :

H.Bharat Chandra

B.Tech 2nd year (IT) Vignan's Institute of Information Technology, Duvvada.

Email: bharat.chandra200@gmail.com

1. INTRODUCTION :

- **Overview :** Predicting Life Expectancy using Machine Learning is a typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features. Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors like [Country,Year,Status,Life expectancy ,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].
- **Purpose :** The purpose of this project is to predict the life expectancy of a person considering the various factors. The project will be helpful in improving the health condition of the society and give insights about some crucial factors such as [Alcohol intake, GDP growth, schooling, adult mortality, total and cost expenditure and etc]. The project uses a **Regression** which is a classification algorithm.It is a measure of the relation between the mean value of one variable (e.g. output) and corresponding values of other variables . The dataset used or the training of the model was downloaded from kaggle.com and Python is used to write the code for machine learning model.

2. LITERATURE SURVEY :

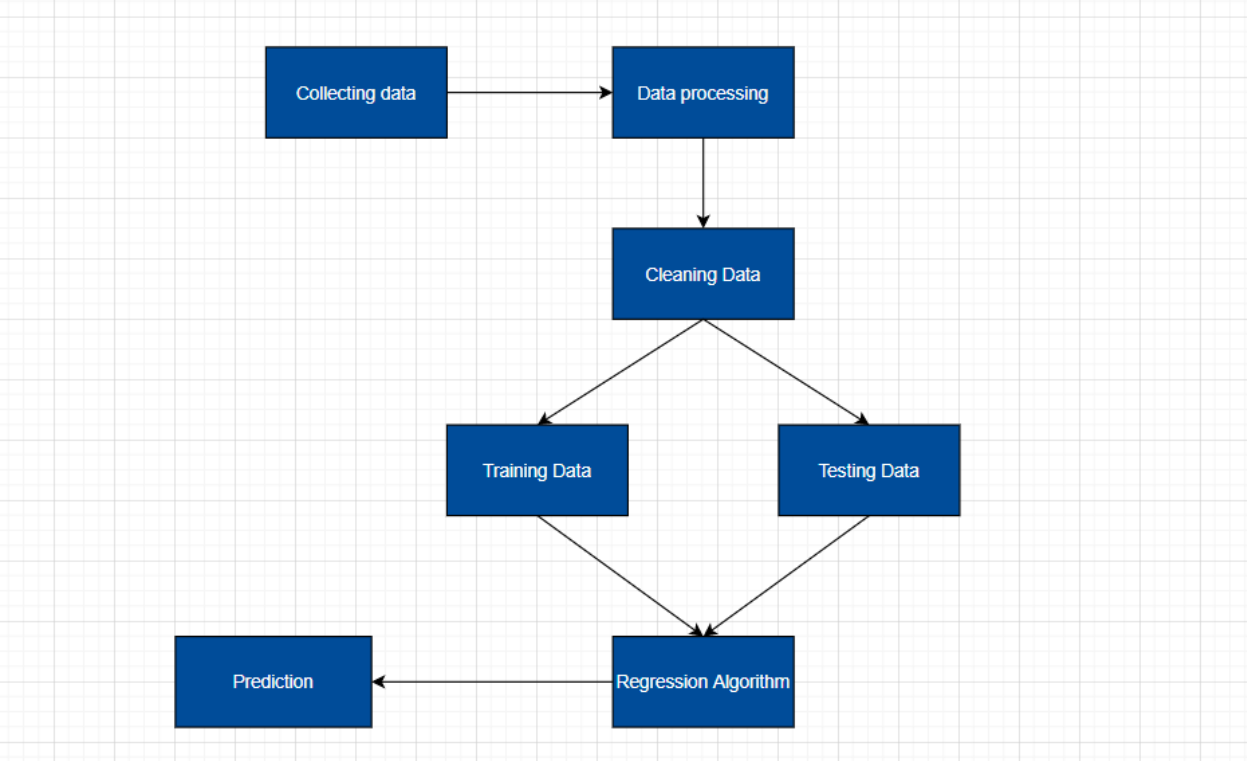
- **Existing problem :**
  It has been noted that data collection for predicting the life/health using the machine learning/big data is a big challenge due to considerations relating to privacy and government policy, which will require the collaboration of various health sector bodies. Despite these challenges, Life expectancy can be predicted by proposing a data collection and application approach. As Artificial intelligence and Machine Learning technologies are developing and quickly being implemented, the ease of gathering health data from the public as well as current government agencies such as centralized health servers could be increased
- **Proposed Solution :**
  I explored life expectancy and looked for dataset on the following aspects (features) : Birth Rate · Cancer Rate · Dengue Cases · Gross Domestic Product (GDP) · Expenditure · Heart Disease Rate · Population · Area · Population Density . life Expectancy, measured in number of years.

  Then I have performed data cleaning and data analysis using the statistical tools in python and selected the dependent and independent features and created a

machine learning model using regression (a classification algorithm). using that model when we give the inputs(features) the model will give prediction (life expectancy in years) as output. and finally that model is deployed to IBM cloud and made it useful for all the people.

3. THEORETICAL ANALYSIS :

- **Block Diagram :**



- **Hardware/Software designing :**

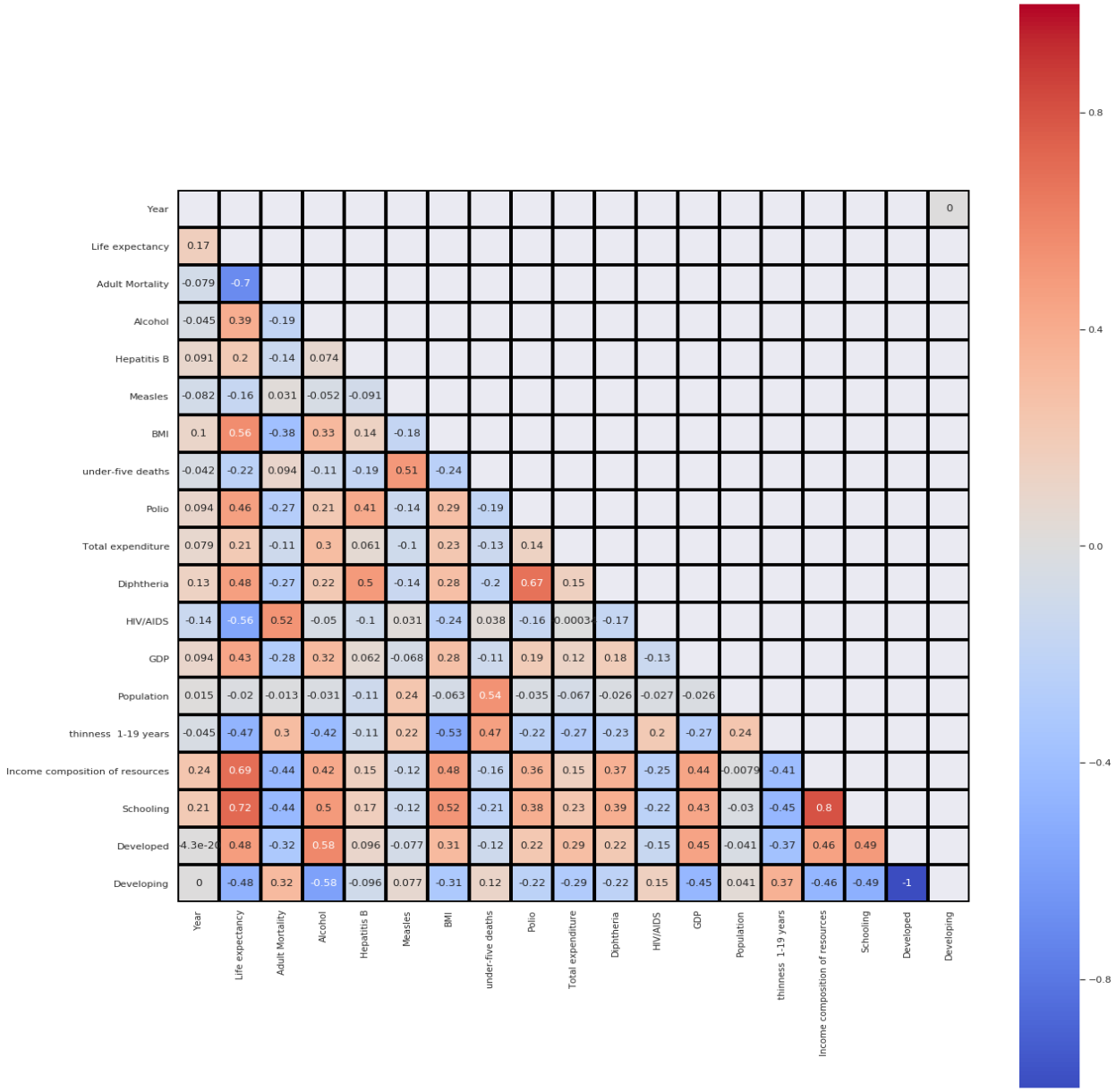
  I have designed a User interface that could help to calculate the life expectancy of an individual in years. It is provided with input values and output shows the life expectancy inn years.

  Software Requirements:

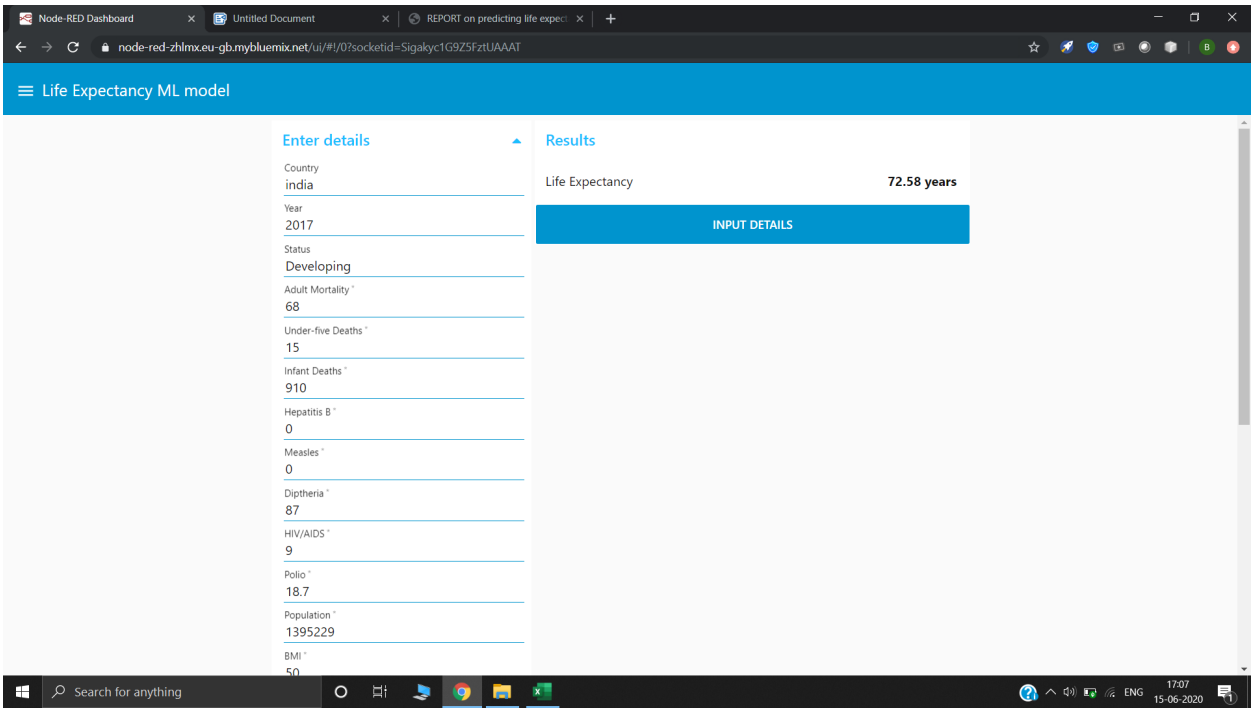
  IBM cloud services
  IBM Watson Studio
  IBM AutoAI experiment
  IBM Node-Red application 7
  Jupyter Notebook
  Github
  Zoho document writer

4. EXPERIMENTAL INVESTIGATIONS :

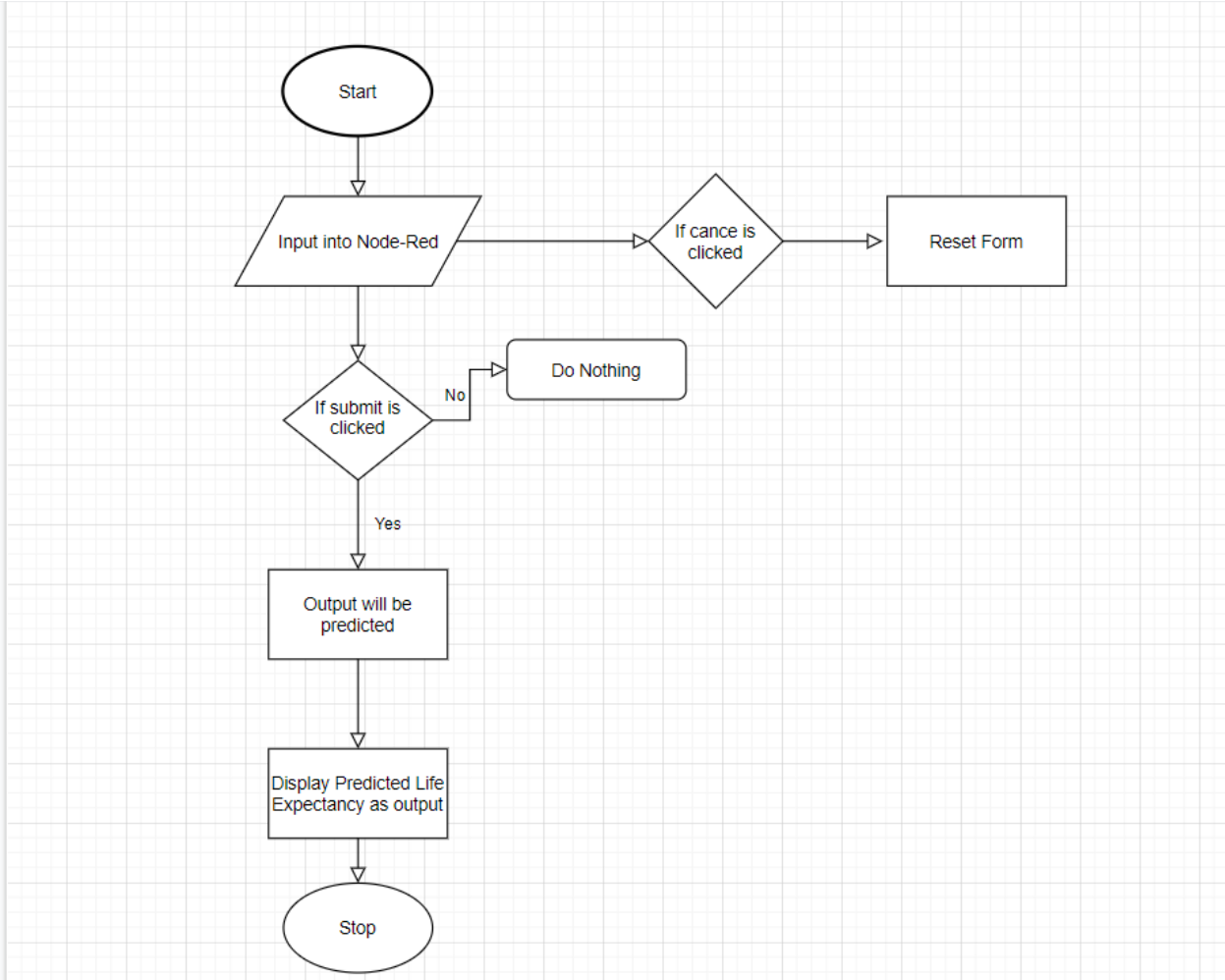
These are some graphs from the refined data analysis, that makes us understand collinearity.

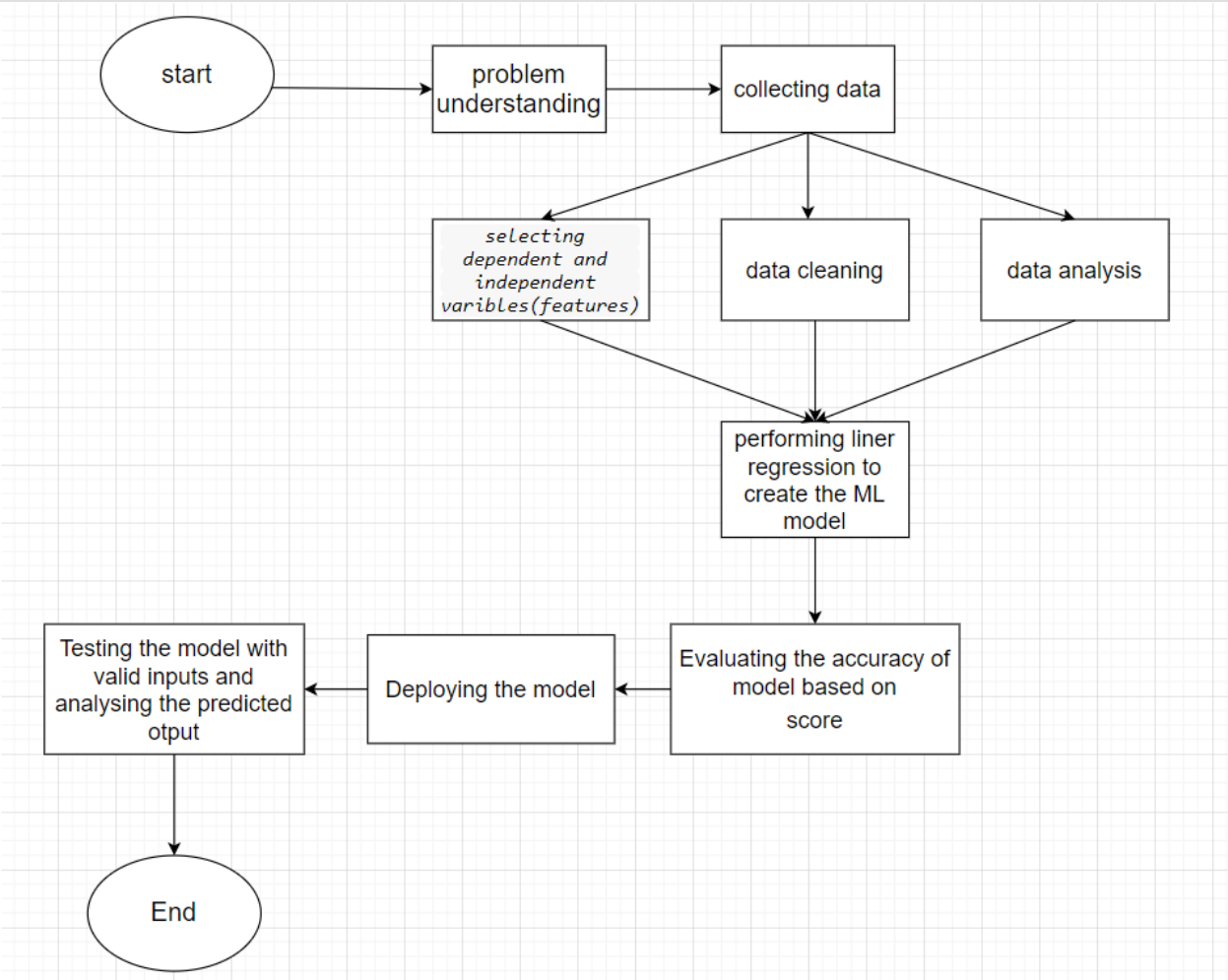


Some random inputs are given to the deployed machine learning model and output is here :

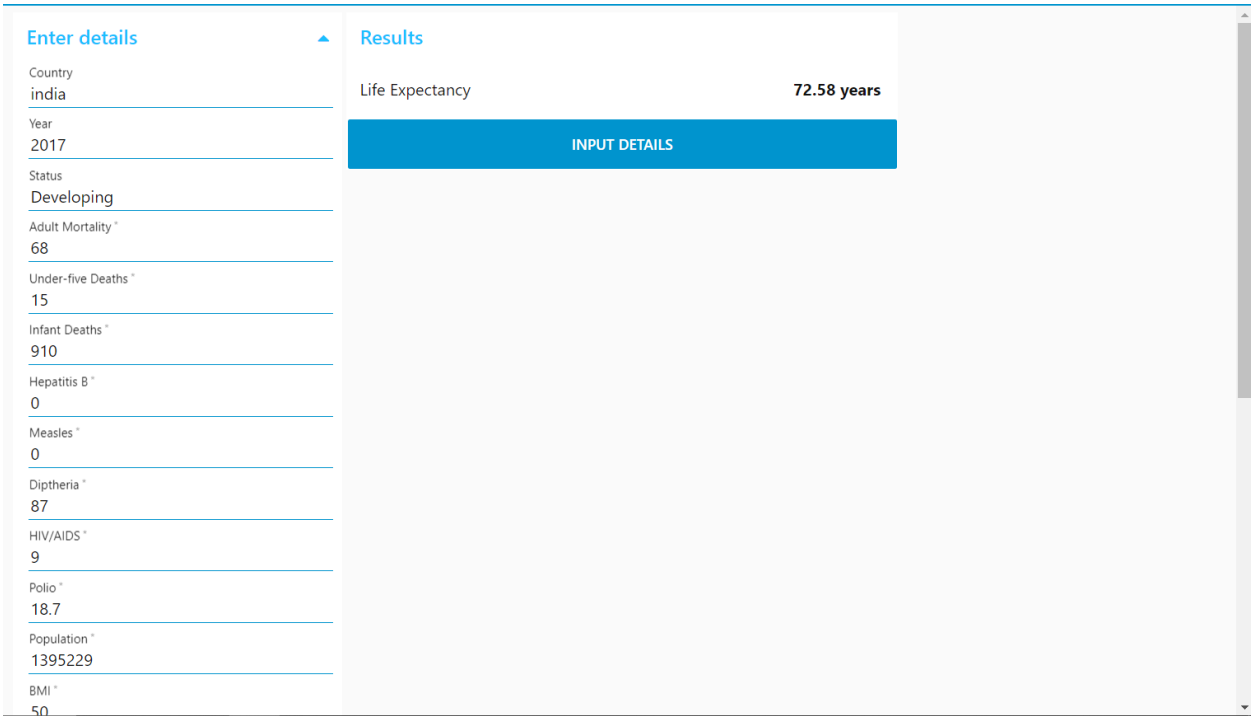


5. FLOW CHART :


Start
Input into Node-Red
If cance is clicked
Reset Form
If submit is clicked
No
Do Nothing
Yes
Output will be predicted
Display Predicted Life Expectancy as output
Stop



6. RESULT : Based on the given data, the autoAI model or ML model understands the data , what are the factors that are affecting the results we require i.e life expectancy. It will predict the output based on the features that we trained . Then when we give any input, based on the trained model, it will validate the features and give the accurate value as predicted output . So the results we get are approximations, they are not definitely true, but it works in maximum number of cases.
The output is the prediction of average life span of a person in years .



7. ADVANTAGES AND DISADVANTAGES :

- **Advantages :**

  1. Since we can predict the life span, we can know what factors are influencing the expectancy on life span in what ways.
  2. So, by improving the factors in real life , we can increase the life span.

- **Disadvantages :**

  1. Error in data can result in wrong prediction
  2. Accuracy is not 100%
  3. Error may occur due to inappropriate analysis of data .

8. APPLICATIONS :

- This project/idea is useful for Insurance companies as they consider age, lifestyle choices, family medical history, and several other factors when determining premium rates for individual life insurance policies. The principle of life expectancy suggests that you should purchase a life insurance policy for an individual.
- This will also help increase the expectancy considering the impact of a specific factor on the average lifespan of people in a specific country.

9. CONCLUSION :

Thus, we have developed a model that will predict the life expectancy of a specific demographic region based on the inputs provided. Various factors have a significant

impact on the life span such as Adult Mortality, Population, Under 5 Deaths, Thinness 1-5 Years, Alcohol, HIV, Hepatitis B, GDP, Percentage Expenditure and many more. Users can interact with the system via a simple Graphical user interface which is in the form of a form with input spaces which the user needs to fill the inputs into and then press the “Submit” button.

10. FUTURE SCOPE :

As future scope, we can connect the model to the database which can predict the life Expectancy of not only human beings but also of the plants and different animals present on the earth. This will help us analyze the trends in the life span. A model with country wise bifurcation can be made, which will help to segregate the data demographically

Big data and machine learning can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy. We can use demographics of selected regional areas and multiple behavioral health disorders across regions to find correlation between individual behavior indicators and behavioral health outcomes.

APPENDIX :

A. App/UI WEBPAGE : https://node-red-ml.eu-gb.mybluemix.net/ui
B. Dataset link : https://www.kaggle.com/kumarajarshi/life-expectancy-who
C. Source code : https://github.com/SmartPracticeschool/llSPS-INT-2166-Predicting-Life-Expectancy-using-Machine-Learning/blob/master/LifeExpectancy%20ml%20model%20.ipynb
D. YouTube : link for demonstration video - https://drive.google.com/file/d/1jtQ_oohO3K1BA5PVUsi_dtT8u3W_6PA6/view?usp=sharing