LIFE EXPECTANCY PREDICTION USING MACHINE LEARNING

A project report submitted as a part of internship (smartinternz) under

SMARTBRIDGE

Submitted by

ALLU.JASWANTHI (17A31A1201)
B.TECH (INFORMATION TECHNOLOGY)
PRAGATI ENGINEERING COLLEGE

Internship Title	**:** Predicting Life Expectancy using Machine Learning - SB52176
Project ID	**:** SPS_PRO_215
Project Title	**:** Predicting Life Expectancy using Machine Learning
Duration	**:** 23.5 Days

ABSTRACT

The term “life expectancy” refers to the number of years a person can expect to live. By definition, life expectancy is based on an estimate of the average age that members of a particular population group will be when they die. The life expectancy varies widely across the globe. The measure differs considerably by sex, age, race, and geographic location etc. The aim of the project is to predict the average life expectancy rate of the people in a country based on several factors. At this point, a common question arises for everyone that is, how the life expectancy will be predicted. By applying machine learning techniques, one can easily predict the life expectancy. The model can be build using the existing historical data which consists of dependent and independent variables. In our case, life expectancy which will be predicted depends on several independent measures. This application is a perfect use case for regression, which determines the relationship between one dependent variable (life expectancy) and many independent variables (development indicators). In this internship, project is done using python and also without python on IBM cloud using Watson, machine learning, node red, autoAI services.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Life expectancy, an estimate of the number of remaining years of **life** a person has, is an important consideration for making clinical decisions in primary care. Predicting Life Expectancy helps analyze the average lifespan of the countrymen which helps in making crucial health decisions.

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. This problem statement provides a way 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 given.

This project is to build a model while considering historical data from a period of 2000 to 2015 for all the countries. The model trained in this project will be able to predict the average lifetime of a human being given some input factors .With the help of this project any country is able to predict the expected lifetime of their countrymen and then accordingly take preventive measures to improve on their healthcare measures. This will also help countries in improving a particular field such as GDP, alcohol intake, etc. which have a high impact on a country's life expectancy.

Good prognostication helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. So this problem statement is aimed at predicting Life Expectancy rate of a country given various features. It predicts the average lifetime of a human being and predicts on the basis of various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. So the end product will predict the future life expectancy of the person with the help of prior given appropriate matrix of features by the user like current 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.

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The data offers a timeframe from 2000 to 2015.

The data originates from here: https://www.kaggle.com/kumarajarshi/life-expectancy-who/data.

1.2. PURPOSE

Life expectancy is the most significant aspect for decision making. Good projection for example helps to decide the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning.

Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients. This project can be used in hospitals and the doctors can use it to predict the life expectancy of a patient with the underlying disease or a new born baby. It can be used by government predict the life expectancy of the economic backward people due to poverty.

With the help of this project it will be easy for governments of the countries with less life expectancies to improve their medical and healthcare services. This Project will give overall prediction about the life expectancy of people living in various countries who have various diseases like Diptheria, HIV, Hepatites, Polio, Measles and also people taking alcohol based on Body Mass Index(BMI), GDP, Population, Mortality Rate of a particular country.

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 percapita. Increase in life expectancy also leads to an increase in the “manpower” of 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.

CHAPTER 2

LITERATURE REVIEW

2.1. EXISTING PROBLEM

Past studies have revealed a lot of work in the field of predicting life expectancy of a human being. After reviewing existing works and techniques in the prediction of human Life Expectancy, and finally reached a conclusion that it is possible to predict a Average Life Expectancy for individuals using advancing technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearables and mobile health monitoring devices, IOT. It is noticed that the collection of data is a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The interworking of a heterogeneous health network is also a challenge for data collection. Despite these challenges, a possibility of predicting Life by proposing an approach of data collection and application by smartphone, in which users can enter their information to access the cloud server to obtain their own predicted Lifespan based on the given inputs.

To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.

2.2. PROPOSED SOLUTION

Although there have been a lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on a data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations in this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy.

The model of ”Predicting Life Expectancy using Machine Learning” uses IBM Cloud services, which helps to avoid any storage issues. The UI Presented to the users is a website url i.e. on users fingertips.

CHAPTER 3

TEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM

The overall project is based on two steps building a machine learning model deploying and integrating it with node red.

The below figure shows how to build a model for predicting the output

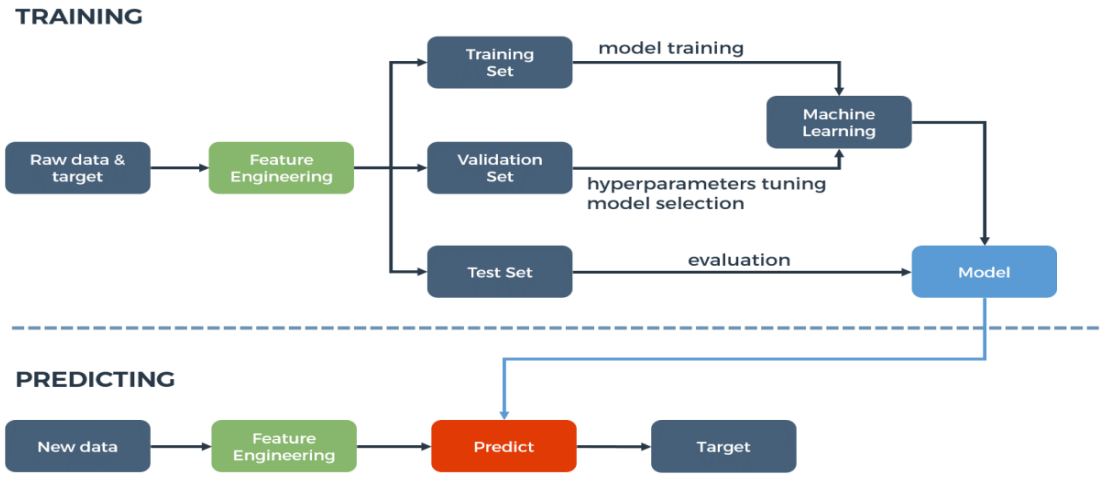

Figure 3.1.1: ML Model steps

The below figure illustrates the model deployment its storage and the process required to build UI.

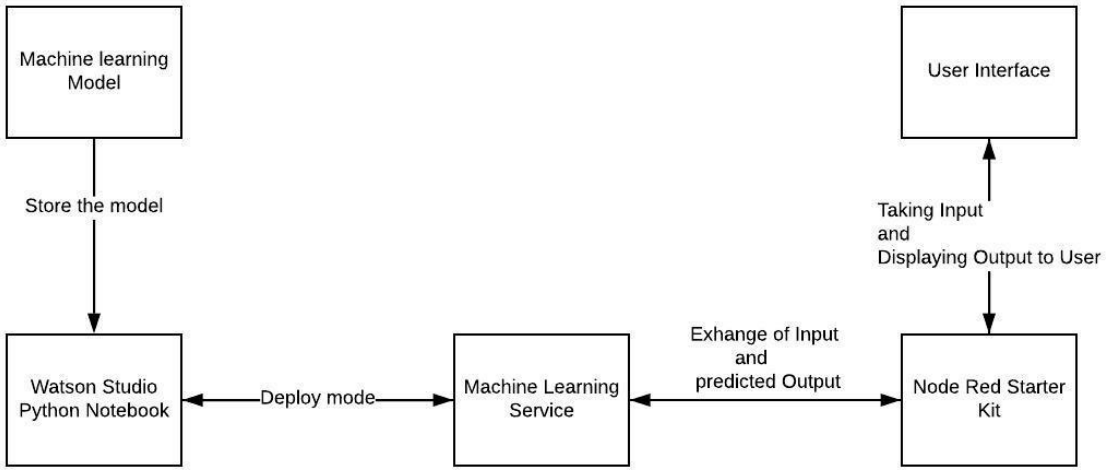

Figure 3.1.2: Deployment and integration

The entire process is shown below.

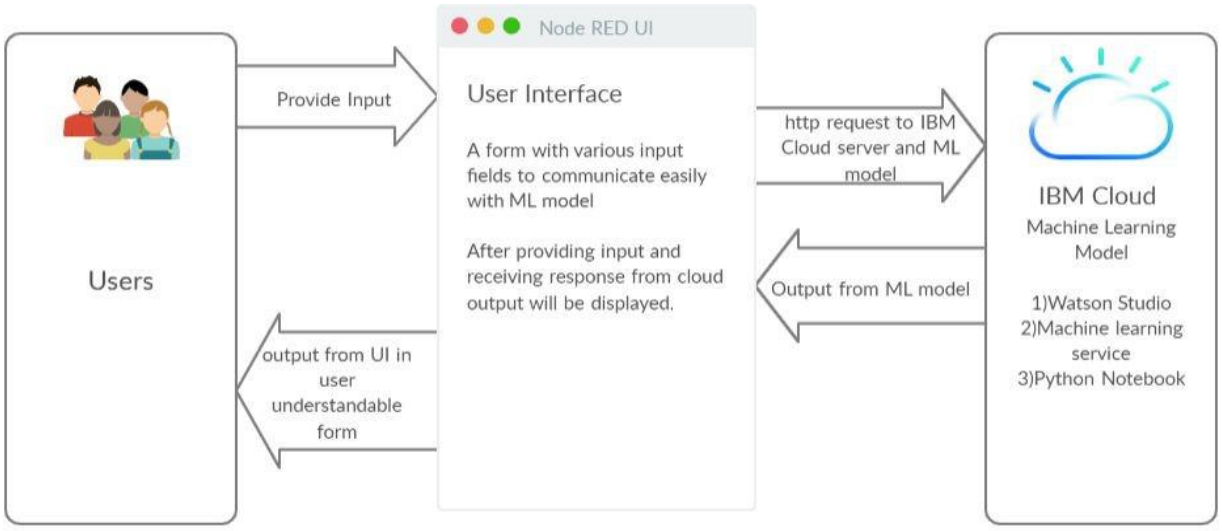

Figure 3.1.3: Entire block diagram

3.2. SOFTWARE DESIGNING

- Create necessary IBM Cloud services
- Create Watson studio project
- Configure Watson Studio
- Create IBM Machine Learning instance
- Create machine learning model in Jupyter notebook
- Deploy the machine learning model
- Create flow and configure node
- Integrate node red with machine learning model
- Deploy and run Node Red app.

Input is taken from the user using a "Form" element in Node-Red. Then, an HTTP request is made to the IBM cloud that further makes an HTTP request to the deployed model using model's instance id. After verification of id, the model sends an HTTP response which is finally parsed by the Node-Red application.

Linear regression is one of the simplest and easiest algorithm to understand and apply. It is a predictive machine learning algorithm which belongs to both, statistics, as well as machine learning.

Steps to build the model:

- Define the problem.
- Gather the data.
- Clean & Explore the data.
- Model the data.
- Evaluate the model.

CHAPTER 4

FLOWCHART

A flowchart is a diagram that depicts a flow of process, system or computer algorithm. They are widely used in multiple fields to document, study, plan, improve and communicate complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence.

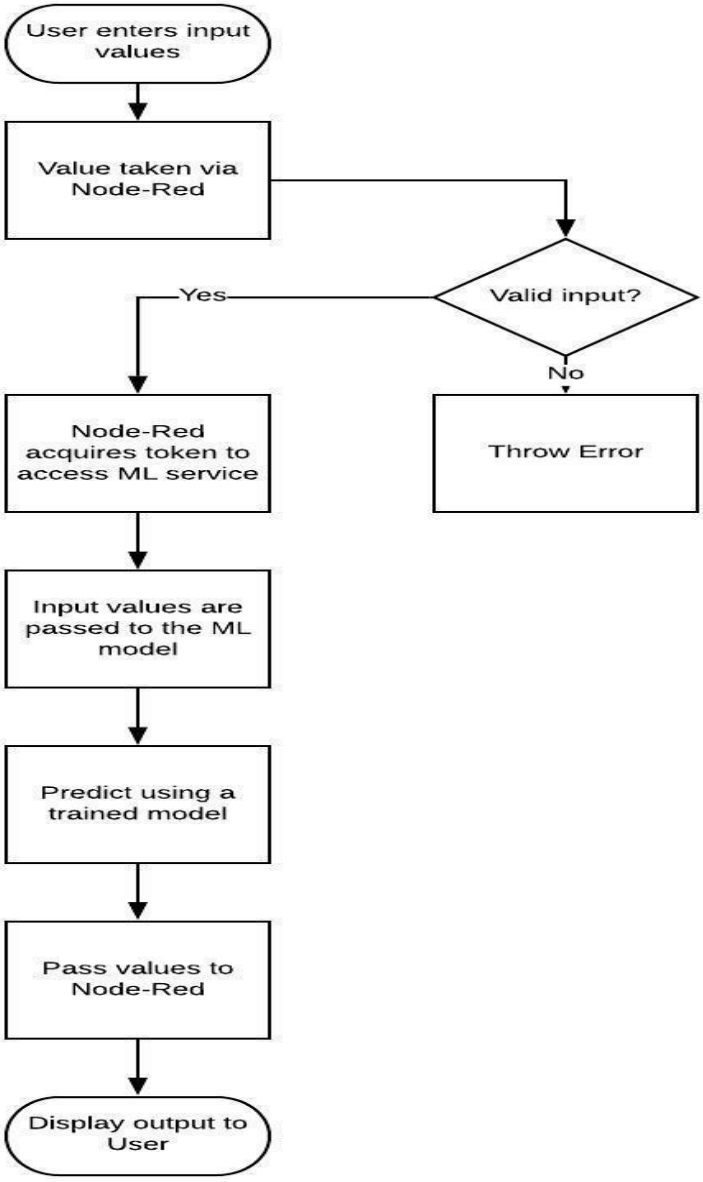

Figure 4.1: Overall flowchart

4.1. FLOW CHART OF ML MODEL

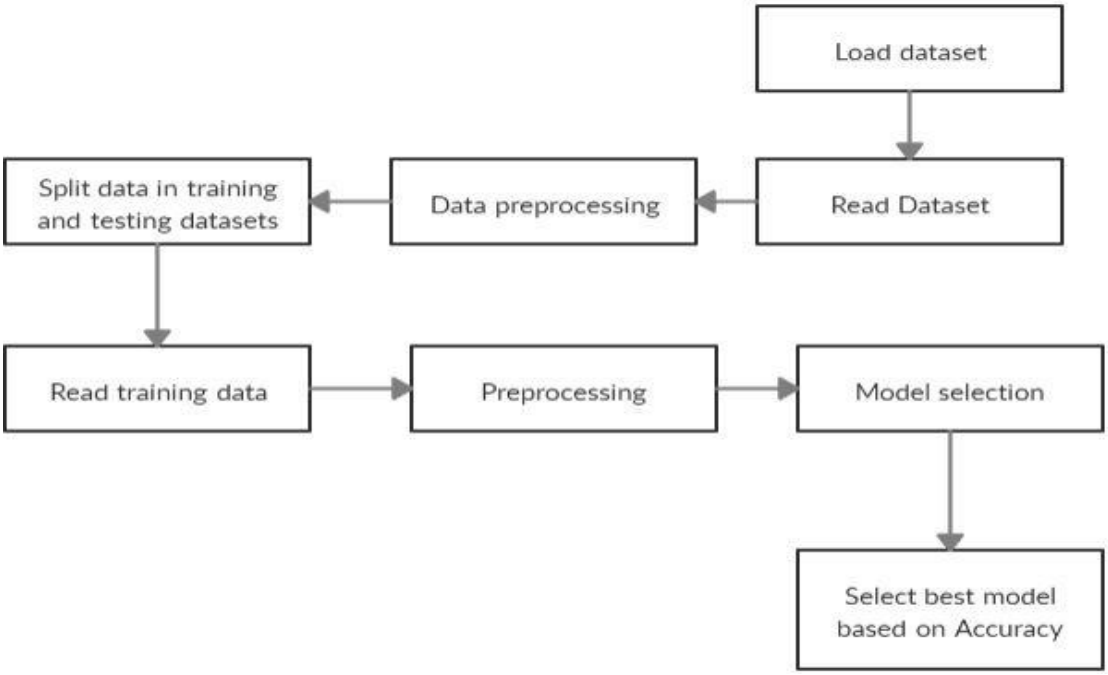

Figure 4.1.1: ML Model Flowchart

4.2. NODE RED FLOW

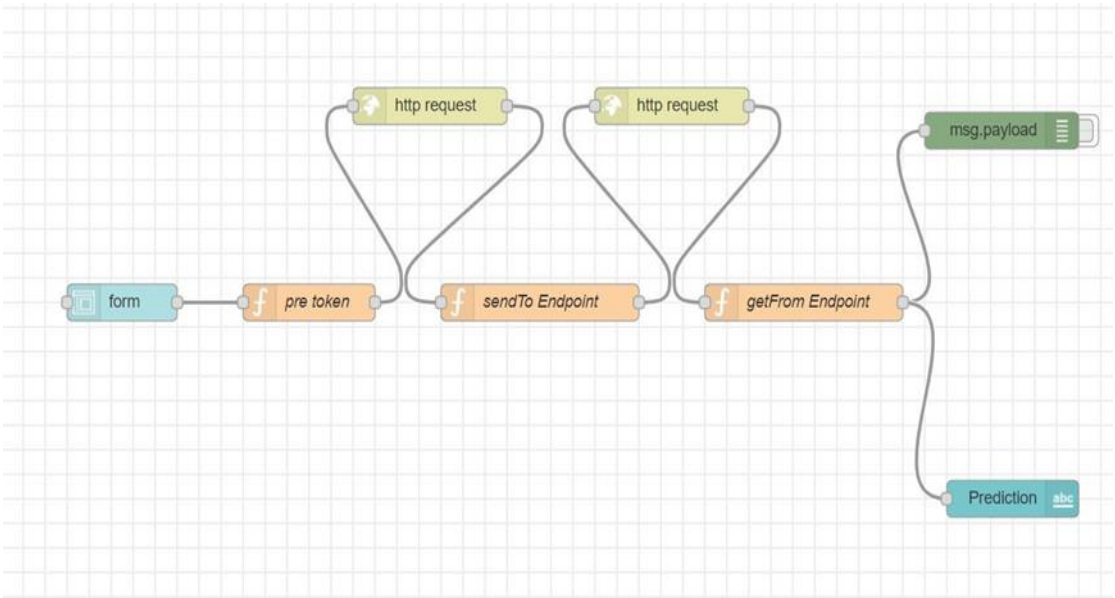

Figure 4.2.1: Node red flow

CHAPTER 5

RESULTS

The user friendly Graphical User interface is shown in Figure 5.1. This GUI is connected to the trained machine learning model present in the backend(IBM watson notebook). The user has to fill in the inputs accordingly and click on the “Predict” button present at the end of the form. On clicking the “Predict” button, the user will be displayed the predicted life expectancy at the predict label, based on the inputs provided as shown in Figure 5.1.

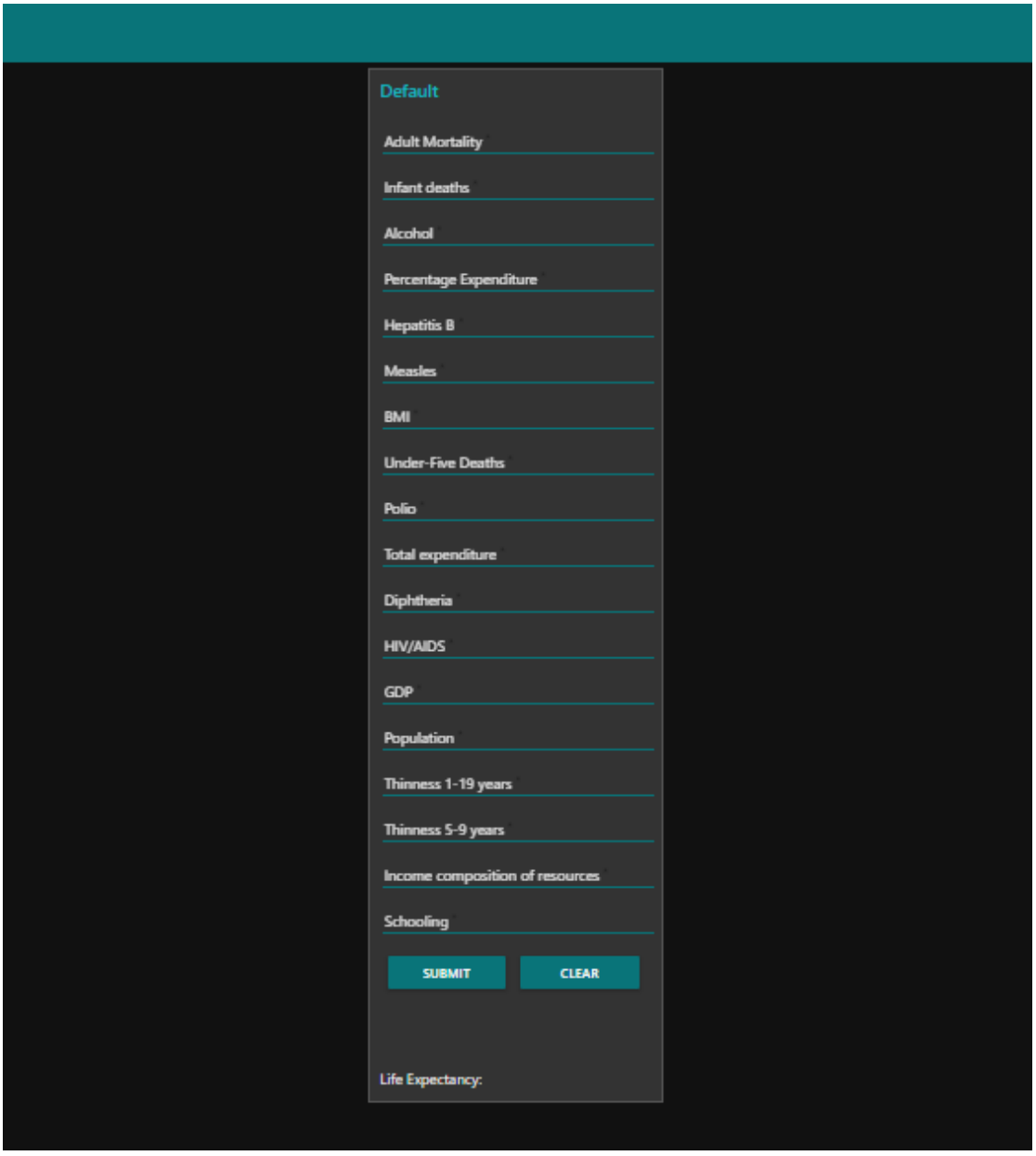

Figure 5.1: Result

CHAPTER 6

ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- Various health organizations around the world collect the data of spread of various diseases and health related issues so based on the dataset we can predict life expectancy of a particular region.
- With node red UI makes it easy for user to interact Life expectancy value can be used by various health care bodies to plan for future initiatives in the health care sector.
- One of the biggest advantages of embedding machine learning algorithms is their ability to improve over time. Machine learning technology typically improves efficiency and accuracy thanks to the ever-increasing amounts of data that are processed.

DISADVANTAGES:

- The dataset which we used in here is not customized for every region since every region has different climate the disease frequency and impact on public health in that region differs from other regions and based on that importance or weightage of those disease related columns increases in life expectancy prediction. So we can group together countries with similar disease pattern and predict life expectancy accordingly.
- UI can be used to predict only one set of values and returns only one output but what If we want to test on a large data set then we need to have an option to upload a csv file.
- Machine learning can also be very time-consuming. When the size of the data fed to the machine learning is very large, the computational cost and the time taken to train the model on the data increases drastically. This can increase the cost of resources required to implement the application on a large scale.

CHAPTER 7

APPLICATIONS

- Personalized Life Expectancy: Individuals can predict their own life expectancy by inputting values in the corresponding fields. This could help make people more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.
- 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.
- 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.

CHAPTER 8

CONCLUSION

- While preparing for the above project I learned the basics of machine learning and statistics concepts.
- I learned how to do Machine learning using scikit learn and data preprocessing.
- How to handle different datasets and divide them into various data types, Identifying categorical columns.
- Label Encoding for categorical data and Splitting data in training and testing data.
- How to check accuracy of a model after fitting and prediction estimation. I understood steps to deploy our model on IBM cloud and how to link node red UI with model.
- Predicting lifespan of human beings can greatly alter our lives. Human behavior and activities are so unpredictable, it may almost be impossible to correctly predict lifespan. However, with the help of Machine learning algorithms such as Regression models, we can get close to predicting a roundabout value.
- This breakthrough can widely impact health sectors and economic sectors by improving the resources, funds and services provided to the common people. It can also increase the ease of access to the individuals.

CHAPTER 9

FUTURE SCOPE

- For future use, one can integrate the life expectancy prediction with providing suggestions and medications to the individual using the application. This will help predict as well as increase the individual's life expectancy.
- The scalability and flexibility of the application can also be improved with advancement in technology and availability of new and improved resources. Also, with the growth in Artificial Neural networks and Deep learning, one can integrate that with our existing application. With the help of Convolutional Neural networks and Computer vision, we can also try to take into account the physical health and appearance of a person.
- Mental health can also be taken into account while predicting life expectancy with the help of sentiment analysis systems as well.

BIBLIOGRAPHY

- Dataset:
  https://www.kaggle.com/kumarajarshi/life-expectancy-who
- Understanding data:
  https://towardsdatascience.com/time-left-to-live-modeling-life-expectancy-and-prototyping-it-on-the-web-with-flask-and-68e3a8fa0fe4
- Understanding ML models:
  https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4
- IBM cloud Introduction:
  https://www.ibm.com/cloud/get-started
- Node-RED Tutorial:
  https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- Machine learning:
  https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/

APPENDIX

- **Git url:** https://github.com/SmartPracticeschool/llSPS-INT-2159-Predicting-Life-Expectancy-using-Machine-Learning

- **Video link:** https://drive.google.com/file/d/1aoI1_o8imPA_IheymLlOofwz_Au-XsVi/view

- **GUI links:**
  - Life expectancy prediction with python: https://node-red-myappml.eu-gb.mybluemix.net/ui/#!/0?socketid=o10AUJR8mOzH8rbwAABo
  - Life expectancy prediction without python: https://node-red-myappml.eu-gb.mybluemix.net/ui/#!/0?socketid=t1crLumB0-nnYpH3AABp