Predicting Life Expectancy Using Machine Learning

1. INTRODUCTION

1.1 Overview

This project is based on Machine Learning in which the goal is to predict the Life Expectancy using datasets. Life Expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age and other demographic factors including gender. The most commonly used measure is life expectancy at birth (LEB). Life Expectancy is depends on various factors like 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 the 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 the healthcare system and some specific disease related deaths that happened in the country. This project aims to automate this task and provide life expectancy when values for different factors are given.

1.2 Purpose

An important point to note when interpreting life expectancy estimates is that very few people actually die at the age indicated by life expectancy, even if mortality patterns stay constant. Most will die much earlier or much later, since the risk of death is not uniform across the lifetime. Life expectancy is the average.

In societies with high infant mortality rates many people die in the first few years of life; but once they survive childhood, people often live much longer. Indeed, this is a common source of confusion in the interpretation of life expectancy figures: It is perfectly possible that a given population has a low life expectancy at birth, and yet has a large proportion of old people. So life expectancy gives a detail about the health care conditions of a country.

2. LITERATURE SURVEY

2.1 Existing problem

- The government doesn't get to know the reasons for the increasing death rate.
- Government doesn't get to know which particular area requires more medical and healthcare facilities.
- People are not aware of the cause of deaths in their surroundings.
- People don't know which disease is how dangerous and for which they should take more precautions.

2.2 Proposed solution

- Our proposed system makes this whole process of calculating Life Expectancy much easier so anyone can calculate the Life Expectancy without any domain knowledge.
- So for calculating Life Expectancy we consider some attributes like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors etc. But our proposed system makes this calculation automated and this system has a predicting tool which can predict the Life Expectancy from various attributes.
- So, using machine learning techniques we suppose to predict the value of Life Expectancy based on some common attributes like year, GDP, education, alcohol intake of people in the country, expenditure on the health care system and some specific disease related deaths that happened in the country etc. Anyone can find this data and get the Life Expectancy value based on their Country and Year.

3. THEORETICAL ANALYSIS

3.1 Block diagram

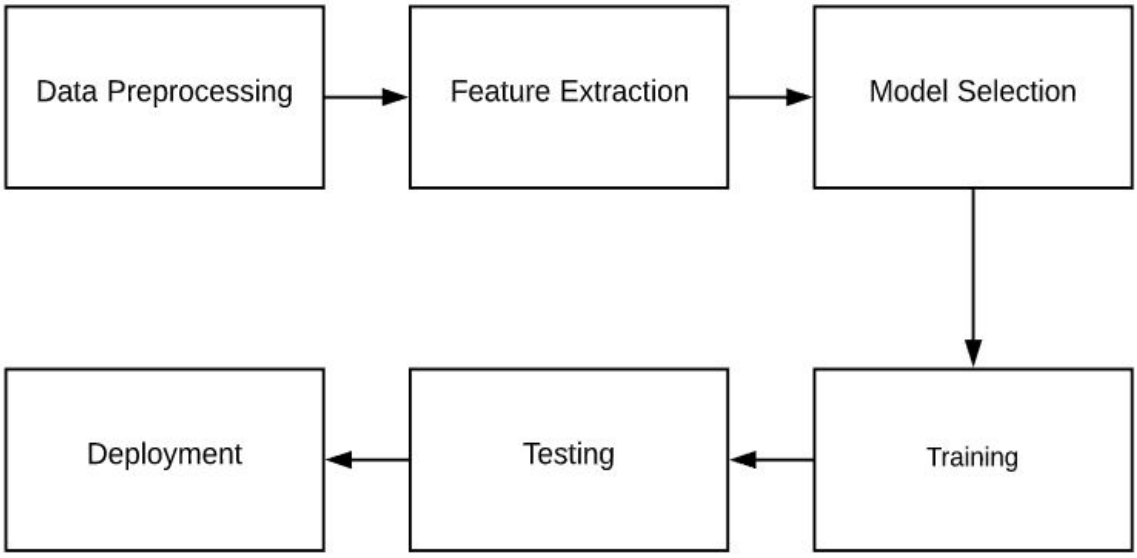

Figure: 3.1

3.2 Software designing

- IBM Cloud Account
- Watson Studio
- Node-RED
- Web Browser for running IBM Cloud services
- Any Environment to run Python code

4. EXPERIMENTAL INVESTIGATION

Our Data of dataset is analyzed using different plots:

1. Pair Plot:

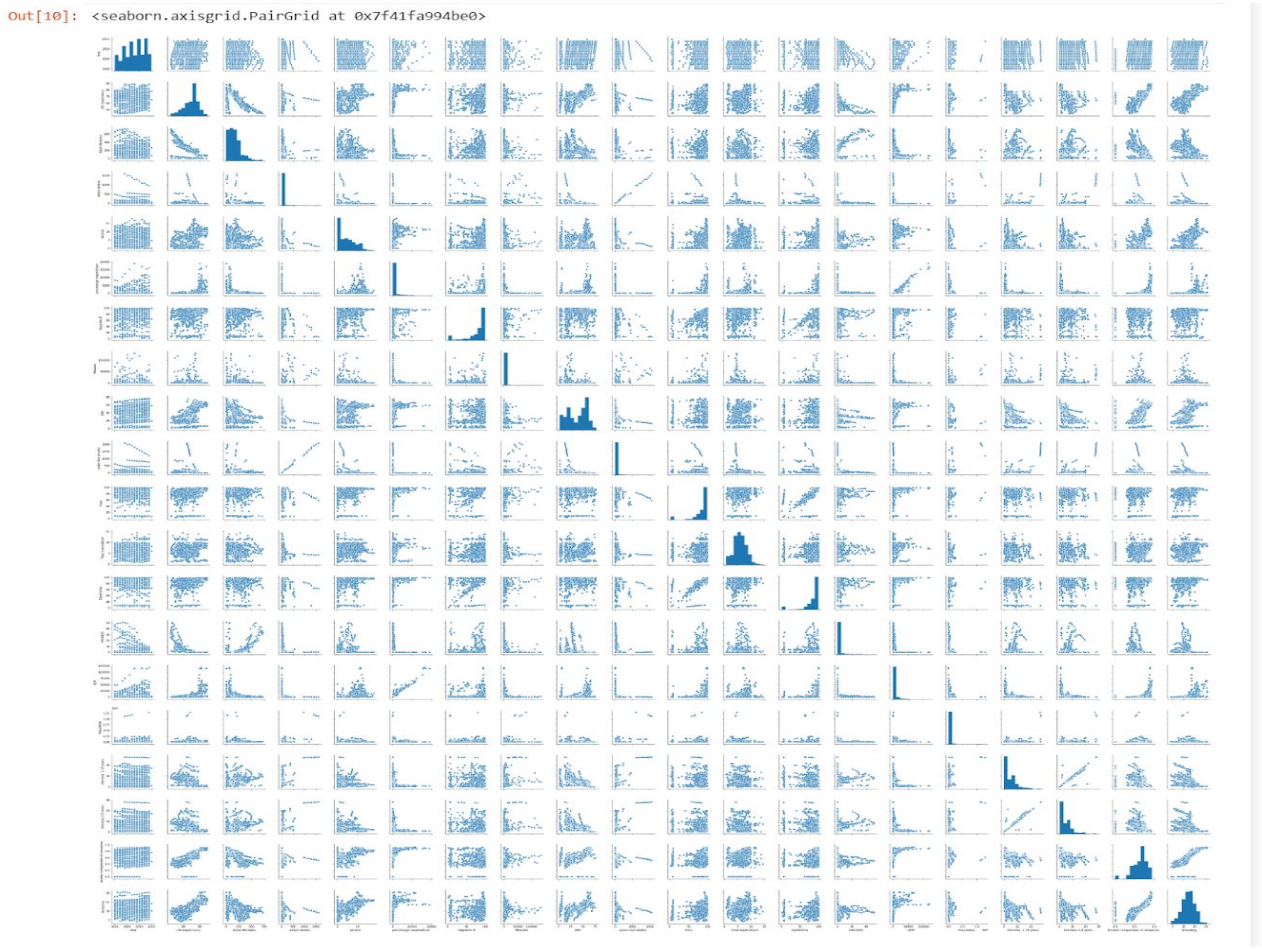

Figure: Pairplot

Pair Plots are a really simple way to visualize relationships between each variable. It produces a matrix of relationships between each variable in your data for an instant examination of our data. Hence using Pairplot we can see the relationships between two variables and can be understood.

2. DistPlot:

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f41ed42ab70>
```

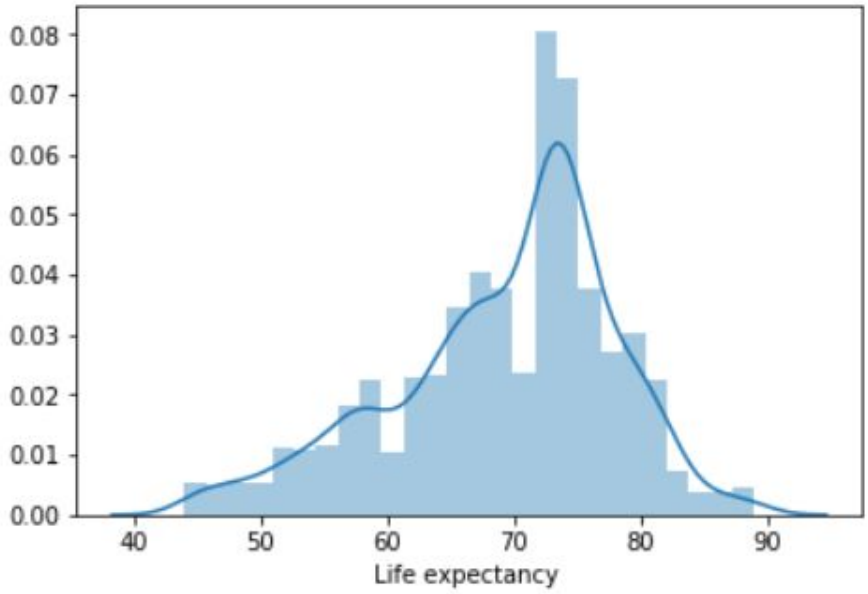

Figure: Distplot

A Distplot is like a histogram which gives how a particular value falls in a particular column.

3. HeatMap:

```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7f41ecefc668>
```

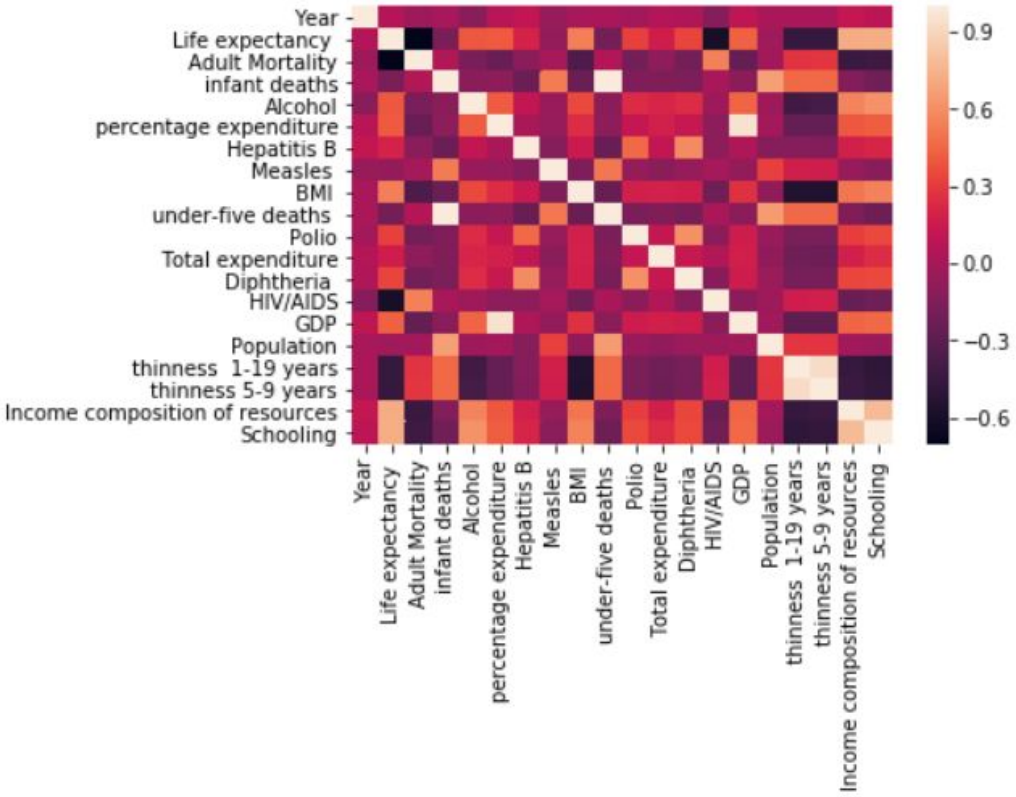

Figure: HeatMap

A heatmap is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors. It is used to identify which column contributes more when predicting a particular column value.

5. FLOWCHART

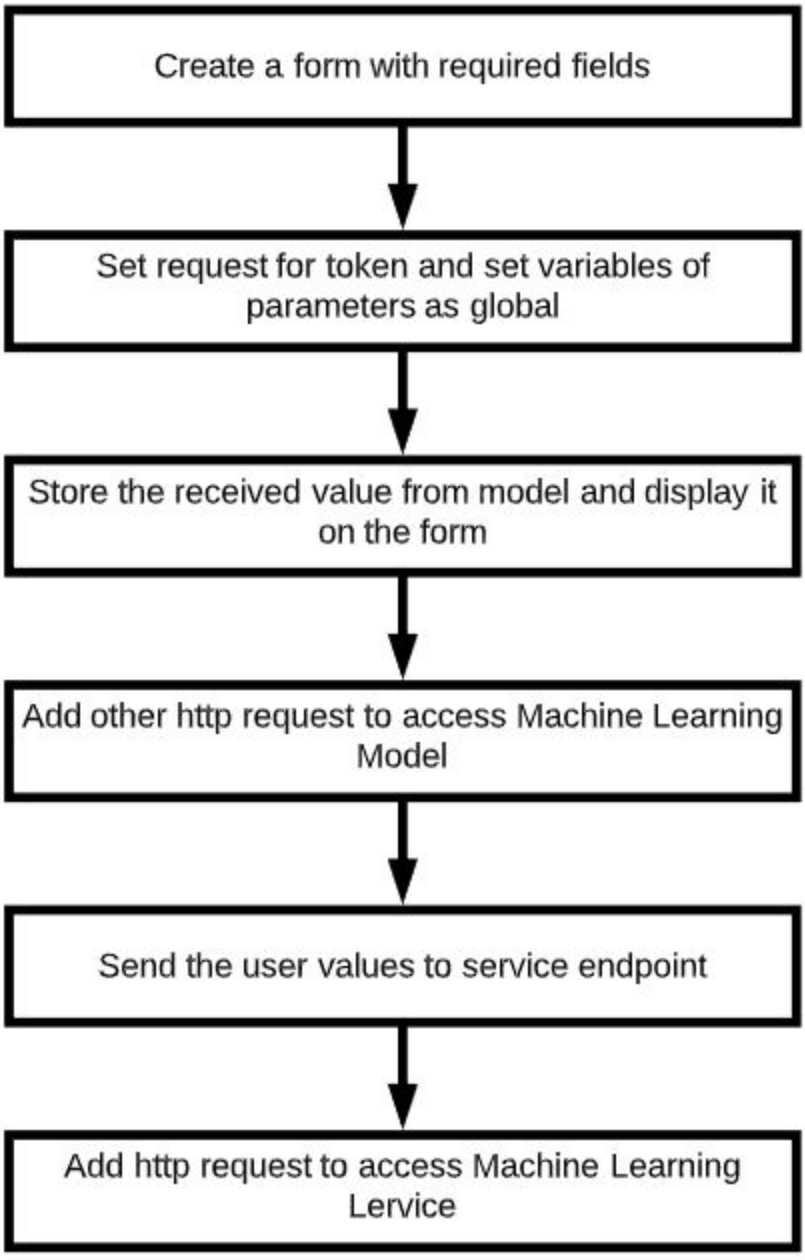

Figure: Flowchart

6. RESULT

Algorithm used	Forest Regression
Accuracy	95.6%
MAE	1.1583535108958865
MSE	3.4548060968523013
RMSE	1.8587108696223578

```
Out[22]: <matplotlib.collections.PathCollection at 0x7f41ec984710>
```

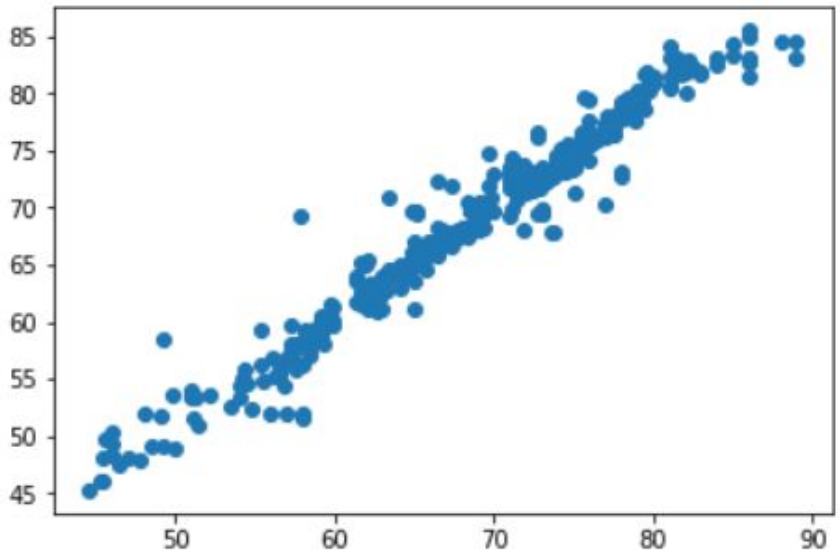

Figure: Scatter Plot

7. ADVANTAGES

- People will be aware about life expectancy and reasons behind increasing death rate and hence will take preventive measures accordingly.
- Government will be aware of the reason behind the death.
- Using this model all the rigorous calculations are eliminated.

8. DISADVANTAGES

- This model fails to work for multiple sets of data.
- It requires all the parameters which are mentioned in the prediction form.
- It requires Internet connection.

9. APPLICATIONS

- Healthcare facility management.
- It can be used in the financial world, including insurance, pension planning.
- It has social benefits.

10. CONCLUSION

Hence the project has automated the entire task of back-breaking calculations and gives a better experience to our users. The information obtained can be useful to the society and this method is also much cheaper than hiring people to do the calculations.

11. FUTURE SCOPE

- A system to update our model parameters when there is a change in consistency of data like when a sudden epidemic occurs or during a recession, then all the attributes used in our model need to be updated to provide the most precise life expectancy.
- Accuracy of the model can be improved.

12. BIBLIOGRAPHY

- https://www.kaggle.com/kumarajarshi/life-expectancy-who
- https://cloud.ibm.com/login
- https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0775-2
- https://www.youtube.com/watch?v=Jtej3Y6uUng
- https://www.youtube.com/watch?v=NmdjtezQMSM

13. APPENDIX

Source code Link:

https://drive.google.com/drive/folders/15cKfnMHBFMRtRG3-4HU05KRA0lBfhb57?usp=sharing