

Predicting Life Expectancy Using Machine Learning

Introduction

1.1 Overview :

The project is aimed at predicting the life expectancy rate of a country given various features. This project can be used to predict the average life expectancy of people living in a country considering various factors such as Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors where life expectancy is a statistical measure of the average time a human being is expected to live. For prediction of life expectancy, Machine Learning algorithms are used and the project is deployed with the help of IBM cloud.

The web interface for the project is deployed using Node RED application where the user can fill the required information and will get average Life Expectancy value as an output.

1.2 Purpose :

The purpose of this project is to predict average life expectancy of a country. In most countries, life expectancy calculation is derived from a national statistical agency based on large amounts of data. There are many uses for it in the financial world, including life insurance, pension planning, and U.S. Social Security benefits. Life expectancy is used in pricing and underwriting life insurance and insurance products like annuities, as well as in retirement and pension planning.

Literature Survey :

Existing Problem :

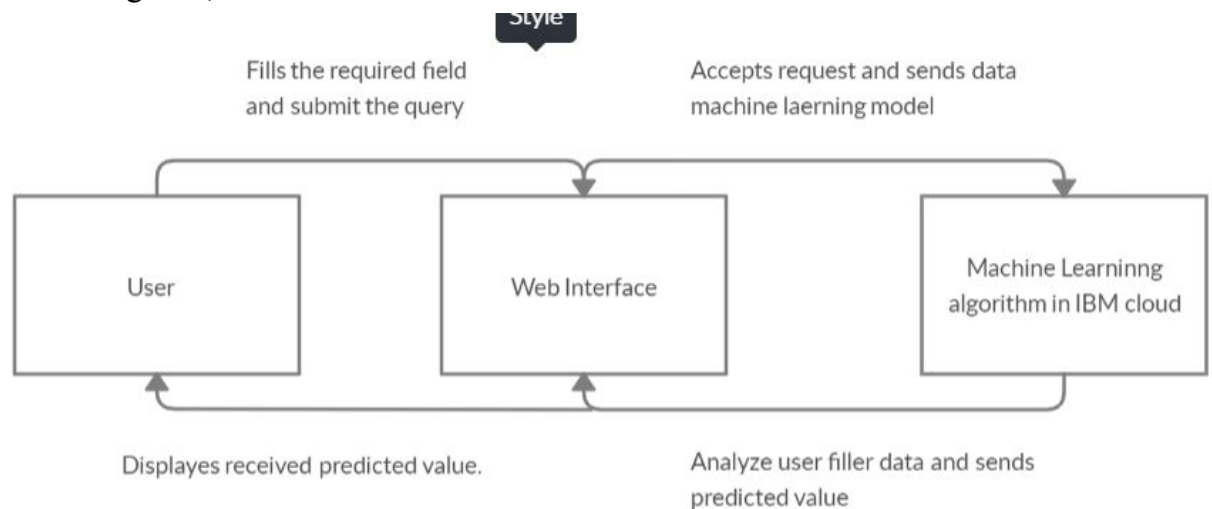
1. The government doesn't get to know the reasons for the increasing death rate.
2. People are not aware of the cause of deaths in their surroundings.
3. People don't know which disease is how much fatal and for which they should take more precautions.
4. Government doesn't get to know which particular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

Proposed Solution :

1. Proposed solution consists of machine learning algorithm which will analyze data and provide required results such as average life expectancy of a country, reasons for increasing death rate, etc.
2. People will get aware about threat of diseases in their surroundings and hence will take precautions accordingly.
3. People will also get to know about factors affecting their life duration and thus will take preventive measures in order to improve their life duration.
4. Government will also get to know about which states or cities need more healthcare and medical facilities.

Theoretical Analysis :

Block Diagram ;



User :

User fills the required information of all fields on the web-app form and sends a request to the web interface by submitting the data.

Web interface :

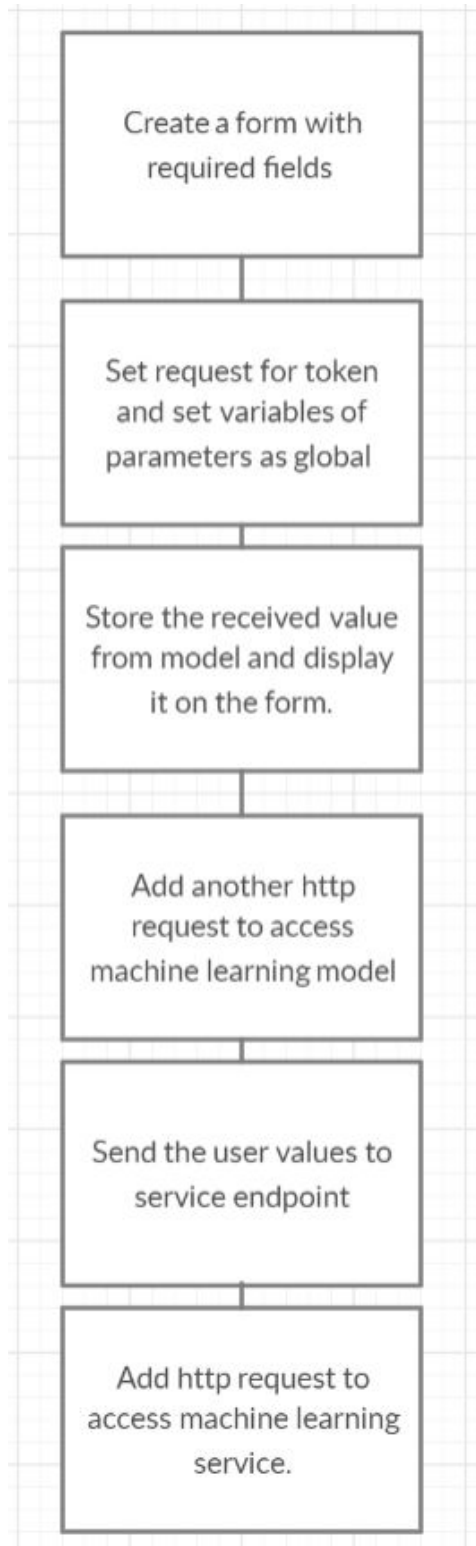
Web interface verifies the request, processes the data and passes it to the machine learning algorithm. When data received from the model, same displays it on we-app to the user.

Machine learning algorithm :

When a machine learning algorithm receives the data, it analyzes the data and performs required operations and sends the result back to the web-app form.

Software designing :

Node-Red App Designing :

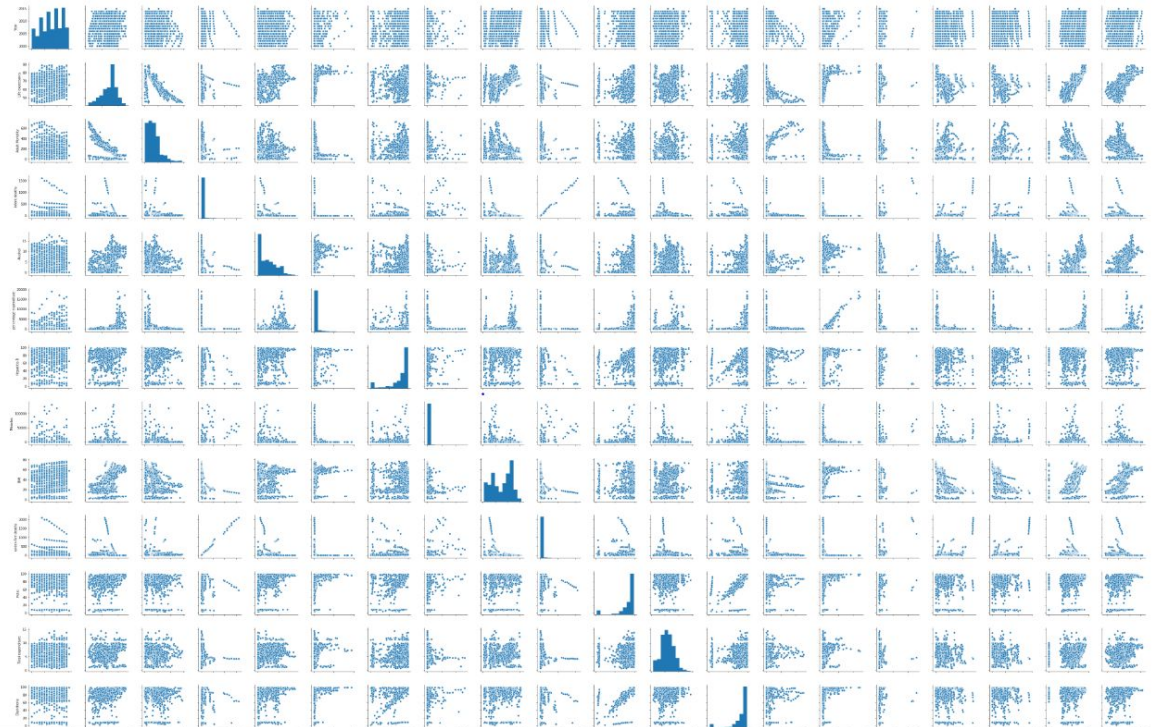


Experimental Investigations :

Data of dataset is analyzed using different plots :

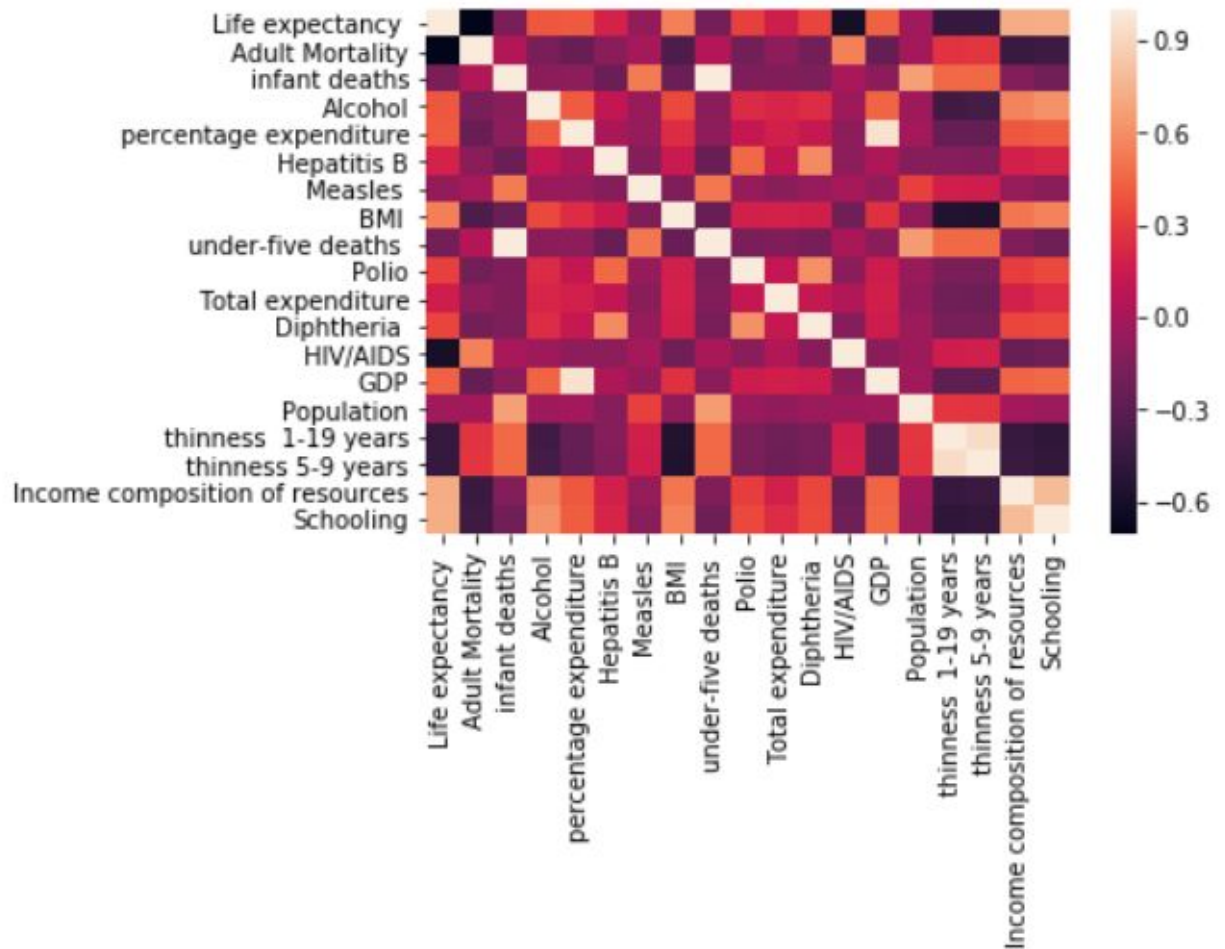
1. Pair Plot :

Out[16]: <seaborn.axisgrid.PairGrid at 0x7fe1183c4ac8>



A pairplot allows us to see both distribution of single variables and relationships between two variables. Hence using pair plot relationships between two variables of the dataset can be understood.

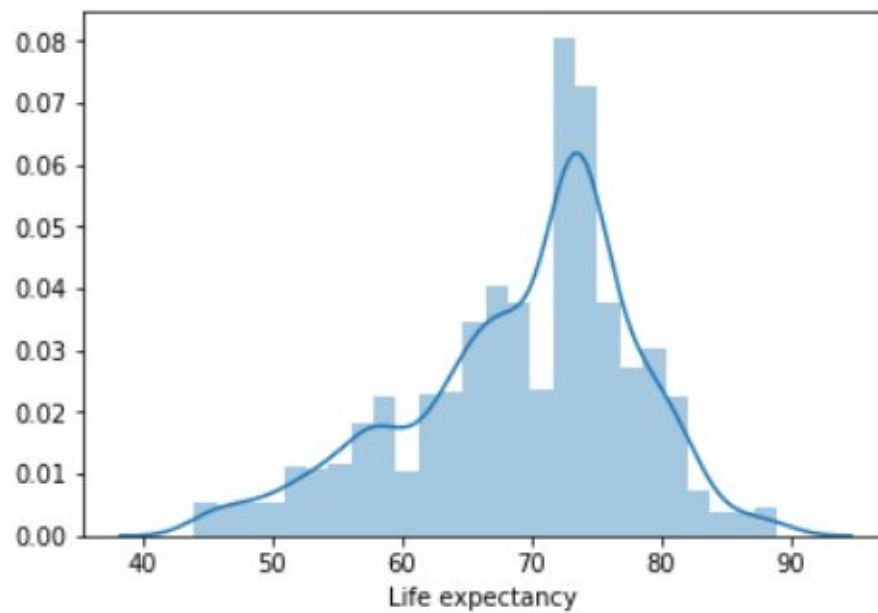
2. HeatMap :



Heatmap is useful in identifying which columns contribute more when predicting a particular column value. For example, how much adult mortality will contribute towards life expectancy can be known from this heatmap.

3.DistPlot :

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe0e6d81f60>
```



Distplot is like a histogram which gives how a particular value falls in a particular column . For eg here around 72-75 values has occurred most times in the dataset.

Flowchart :

Import required libraries

Collect raw data by
importing dataset

Preprocess the data and
remove null values

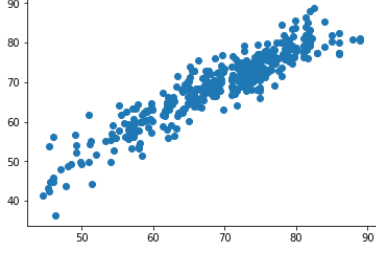
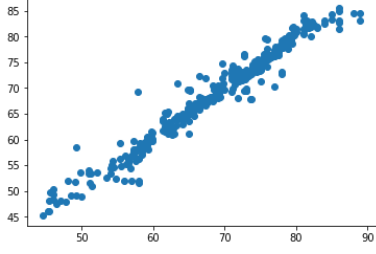
Analyze the data. Plot
data of dataset and
extract required
information.

Split data into training and
testing sets.

Train the model and
predicts the value for test
data.

Calculate evaluation
metrics parameters.

Result :

| Parameters | Linear Regression | Forest Regression |
|--------------|---|---|
| Accuracy | 85.2% | 95.6% |
| Scatter Plot |  |  |
| MAE | 2.7182713907025304 | 1.1583535108958865 |
| MSE | 12.277105932124066 | 3.4548060968523013 |
| RMSE | 3.5038701363098586 | 1.8587108696223578 |

Advantages :

1. Government will be aware of the reasons behind the death rate.
2. Government will get to know which area has a higher death rate, so can increase medical and healthcare facilities in that area.
3. People will be aware about life expectancy and reasons behind increasing death rate and hence will take preventive measures accordingly.
4. People will be aware of any increasing spread of fatal diseases in their surroundings.

Disadvantages :

1. It requires internet connection.
2. Application is unable to predict value for multiple sets of data at the same time.

Applications :

1. One of the applications of life expectancy is in the financial world, including life insurance, pension planning.
2. Social benefits and healthcare facility management.
3. Life expectancy is used in pricing and underwriting life insurance and insurance products like annuities, as well as in retirement and pension planning.

4. Also life expectancy can be used to increase healthcare and medical facilities where it is needed most.

Conclusion :

Hence considering various parameters life expectancy value can be calculated easily by the user with the help of Node-Red app. Thus reducing the complexity and also can be used by a person with less technical knowledge.

Future Scope :

Accuracy of the model can be improved further.

And the system can be improved in a way such that the user will be able to predict life expectancy values for multiple sets of data at the same time.

Bibliography :

- <https://www.who.int/whosis/whostat2006DefinitionsAndMetadata.pdf>
- [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)60296-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)60296-3/fulltext)
- <https://www.news-medical.net/health/What-is-Life-Expectancy.aspx>
- <https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4>

Appendix :

Source Code Link :

https://drive.google.com/open?id=1Z-5EA9yoRtAwD4neOL4CgtWdqboq7-j_