1)INTRODUCTION

1.1 Overview

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: 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.

1.2 Purpose

It is important to have a prediction of life expectancy for every country, it can help in death rate analysis and may also lead to find out the main cause of it, so that officials can plan accordingly to work on it and have better health facilities for citizens.Government Health bodies collect the data every year, so data is available in abundance and can be utilised for a good cause.

2)LITERATURE SURVEY

2.1 Existing problem

The major problem with merely increasing life expectancy is that it also increases morbidity simply because people live long enough to get more age-related disease, disability, dementia and dysfunction. Many serious diseases have increased prevalence with age, including cancer, heart disease, stroke, respiratory disease, kidney disease, dementia, arthritis and osteoporosis. As we can see, the life expectancy prediction can have a huge impact in deciding the yearly health care plan for the health ministry of the respective country.So life expectancy prediction is important and various features need to be taken into consideration for the same.Public and private investment in medical research is primarily focused on reducing death rates, rather than reducing ageing and age-related disease.

2.2 Proposed solution

To predict life expectancy we need to collect data first, as we know health care organizations already have past data available so we can use a dataset which is uploaded on kaggle.The data set has following columns:

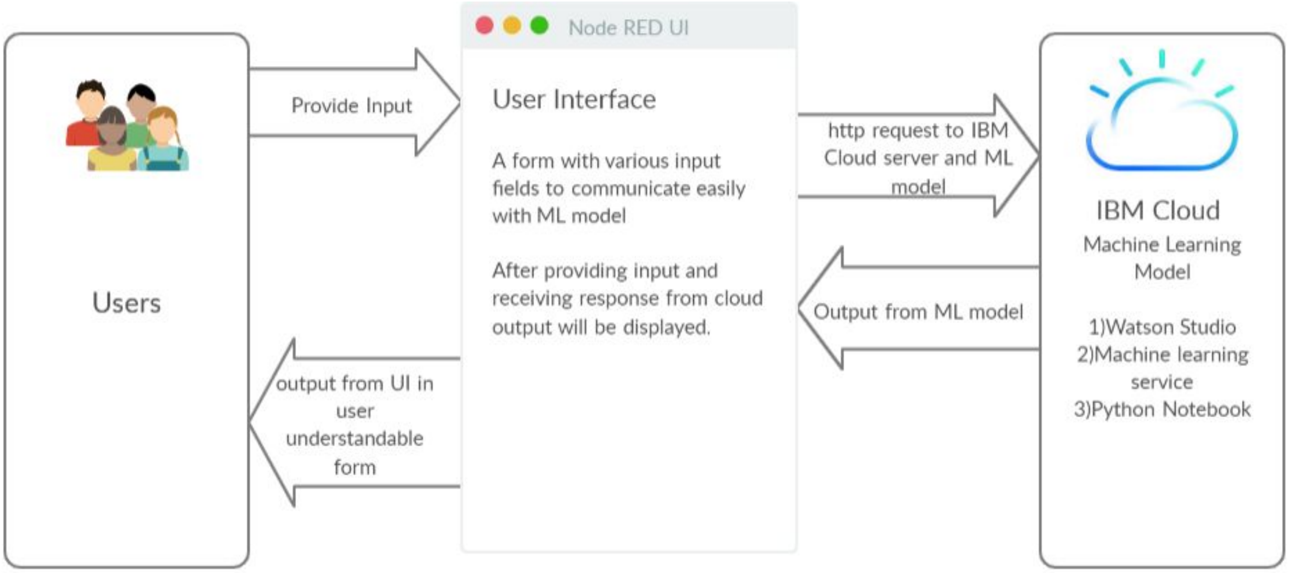
'Country','Year','Status', 'Adult Mortality', 'infant deaths', 'Alcohol', 'percentage expenditure', 'Hepatitis B', 'Measles ', ' BMI ', 'under-five deaths ', 'Polio', 'Total expenditure', 'Diphtheria ', ' HIV/AIDS', 'GDP', 'Population', ' thinness 1-19 years',' thinness 5-9 years', 'Income composition of resources', 'Schooling'

Based on the columns mentioned I tried to learn the importance of features in life expectancy prediction and how to go for data processing as a few of the columns are categorical and we need to handle these types of data well.Also converting string columns into numerical ones.

So for data handling and processing we need to understand meaning of every column and perform data analysis.Here in this project we are using ML model to predict life expectancy.After data processing we need to fit our model on training data and validate using testing data (data which model has never seen before).Calculated the accuracy after that.Based on accuracy we can decide model which can be finalised.Random forest regressor gives better accuracy then linear regression.After ipython notebook is ready we can upload it on IBM watson and design node red app for user Interface.Connect backend code with UI and verify the results.

3)THEORETICAL ANALYSIS

3.1 Block diagram


Users
Provide Input
Node RED UI
User Interface
A form with various input fields to communicate easily with ML model
After providing input and receiving response from cloud output will be displayed.
output from UI in user understandable form
http request to IBM Cloud server and ML model
Output from ML model
IBM Cloud
Machine Learning Model
1)Watson Studio
2)Machine learning service
3)Python Notebook

3.2 Software designing

1)Jupyter Notebook

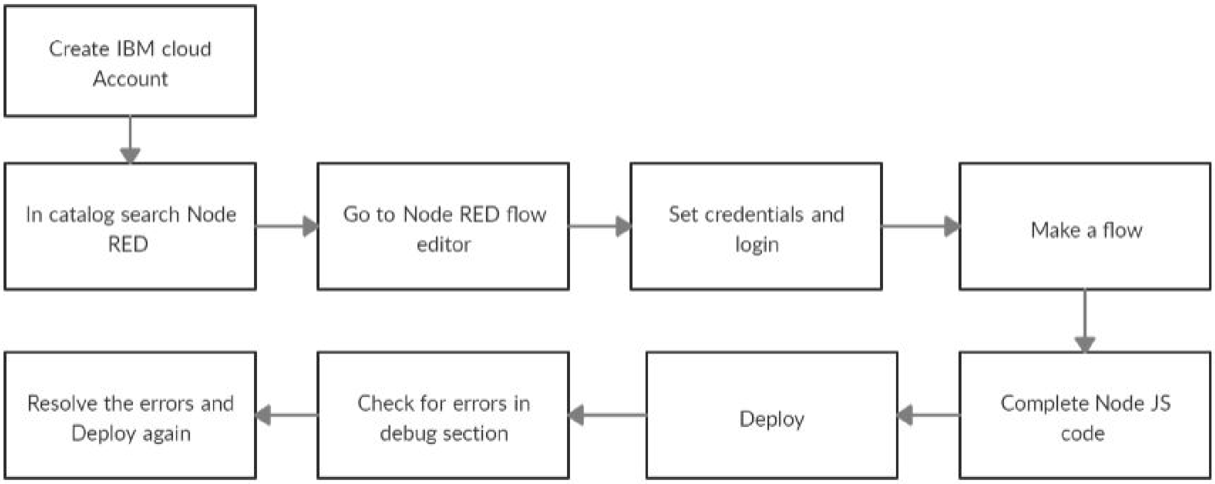
2)Watson Studio

3)Machine learning service

4)Node RED

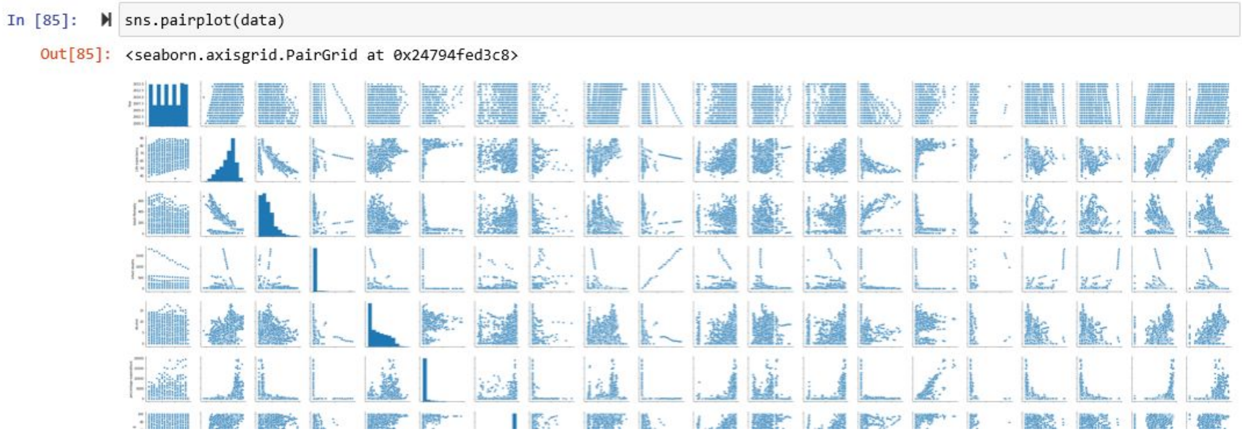
5)Python machine learning Libraries

Node RED steps:


Create IBM cloud Account
In catalog search Node RED
Go to Node RED flow editor
Set credentials and login
Make a flow
Complete Node JS code
Deploy
Check for errors in debug section
Resolve the errors and Deploy again

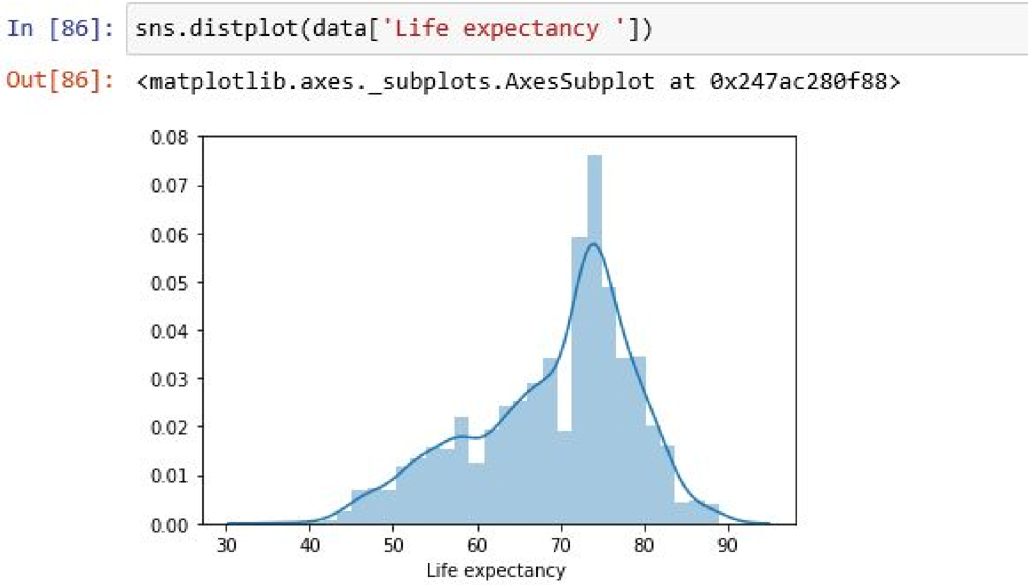
4)EXPERIMENTAL INVESTIGATIONS

i)sns pairplot:



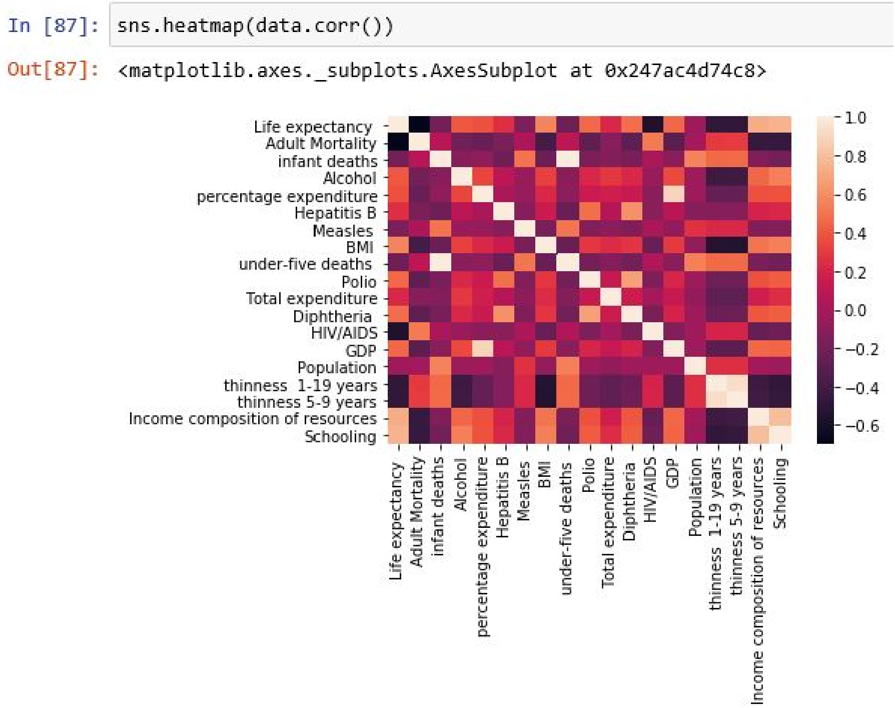
```
In [85]:  sns.pairplot(data)
Out[85]: <seaborn.axisgrid.PairGrid at 0x24794fed3c8>
```

ii)distplot



```
In [86]: sns.distplot(data['Life expectancy '])

Out[86]: <matplotlib.axes._subplots.AxesSubplot at 0x247ac280f88>
```

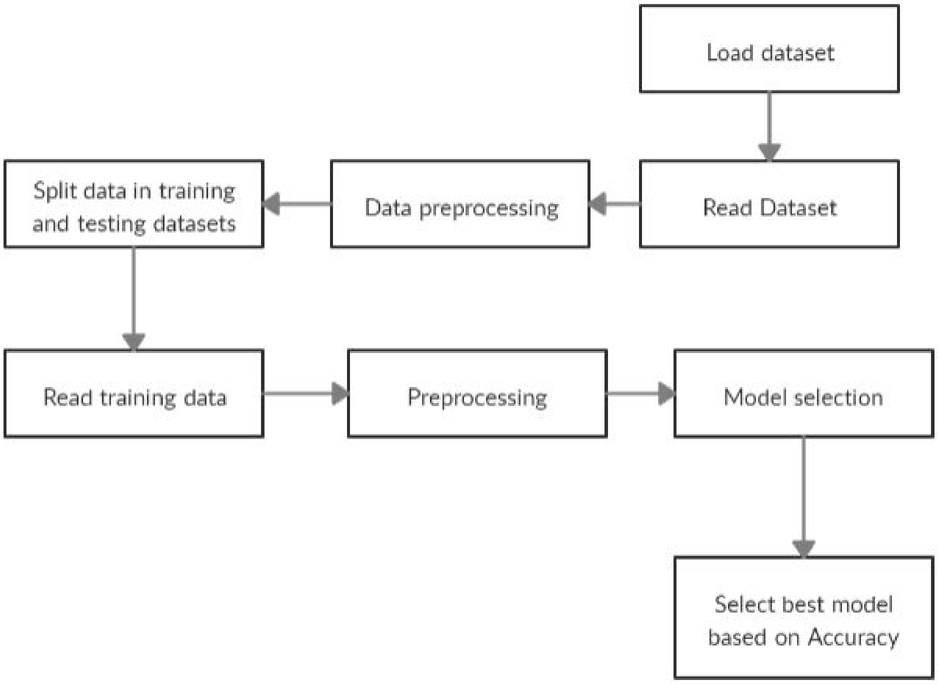
iii)heatmap



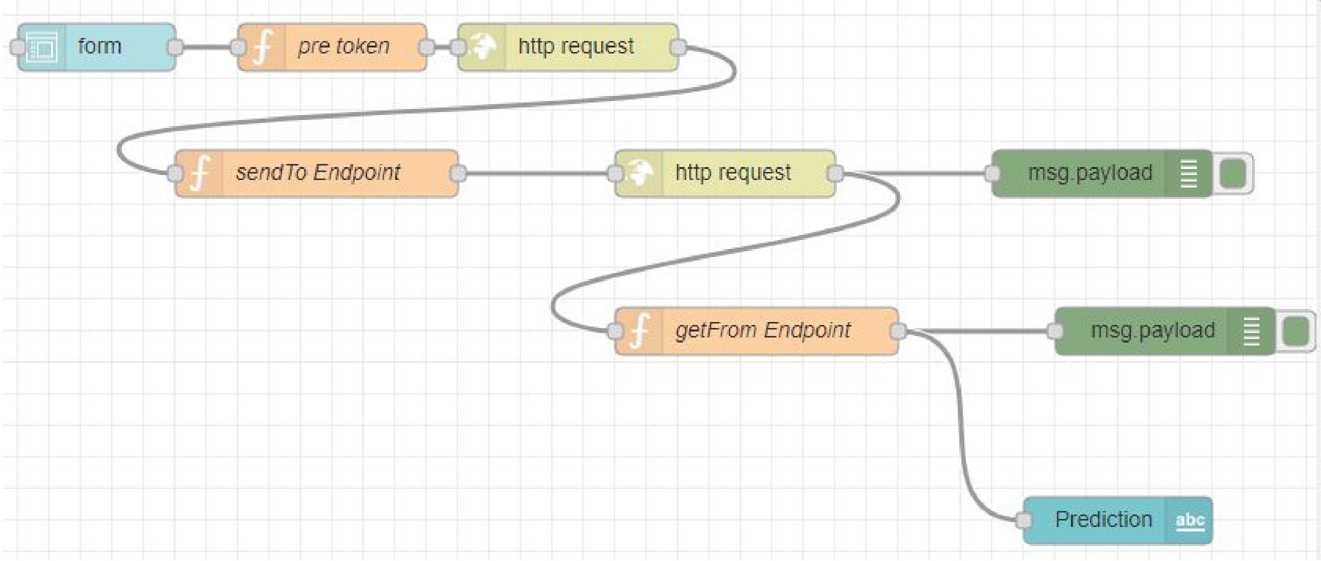
```
In [87]: sns.heatmap(data.corr())

Out[87]: <matplotlib.axes._subplots.AxesSubplot at 0x247ac4d74c8>
```

5)FLOWCHART

Machine learning model

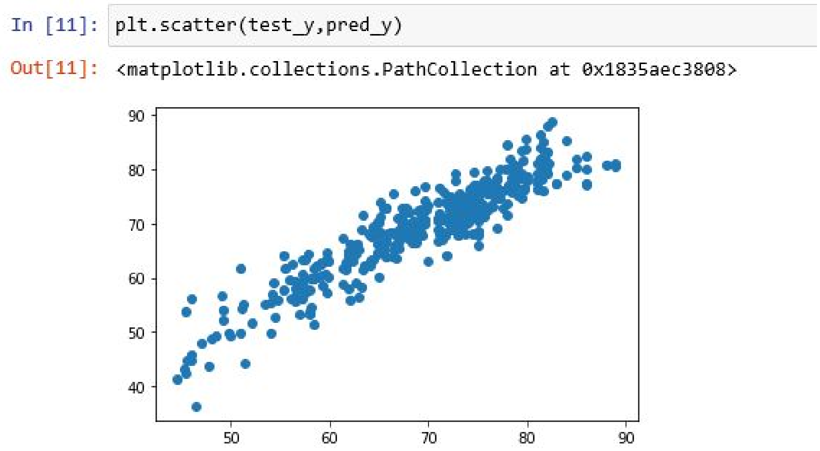


Node RED



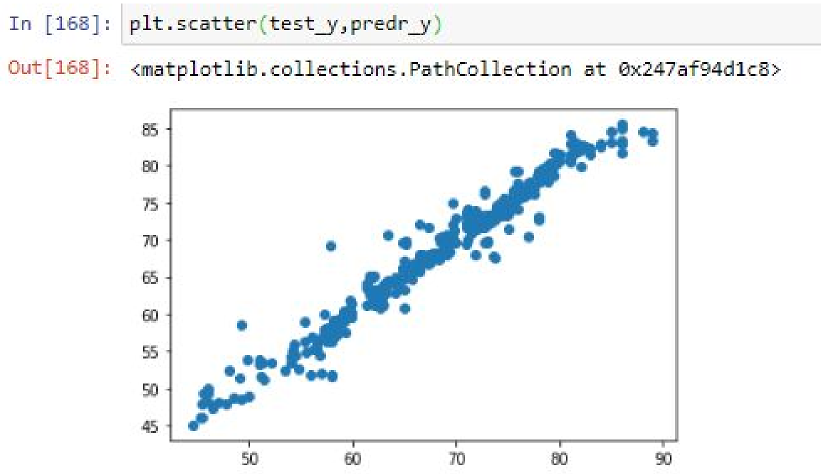
6)RESULT

i)Linear Regression:



```
In [11]: plt.scatter(test_y,pred_y)

Out[11]: <matplotlib.collections.PathCollection at 0x1835aec3808>
```

ii)Random Forest Regressor:



```
In [168]: plt.scatter(test_y,predr_y)

Out[168]: <matplotlib.collections.PathCollection at 0x247af94d1c8>
```

As we can observe that Random Forest Regressor provides better results.

7)ADVANTAGES & DISADVANTAGES

Advantages:

1)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.

2)with node red UI makes it easy for user to interact

3)Life expectancy value can be used by various health care bodies to plan for future initiatives in the health care sector.

Disadvantages:
1)The dataset which we used in here is not customised 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.
2)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.

8)APPLICATIONS

1)Can be used by various health organisations to predict life expectancy based on dataset provided.
2)Easy user interaction using node red app.
3)Deployed on IBM cloud.

9)CONCLUSION

1)While preparing for the above project I learned the basics of machine learning and statistics concepts.
2)I learned how to do Machine learning using scikit learn and data preprocessing.
3)How to handle different datasets and divide them into various data types, Identifying categorical columns.
4)Label Encoding for categorical data and Splitting data in training and testing data.
5)How to check accuracy of a model after fitting and prediction estimation.
6)I understood steps to deploy our model on IBM cloud and how to link node red UI with model

10)FUTURE SCOPE

1)We can have a file upload feature where multiple data sets and predict values for the same.

2)We can customize the dataset based on region and diseases commonly found in that region.
3)The life expectancy prediction can be given based on a particular disease, considering a single disease at a time.
4)We can study various other factors which affect life expectancy and add those in our dataset for better prediction.

11)BIBLIOGRAPHY

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

APPENDIX

A. Source code
https://drive.google.com/file/d/1aXo15Trtxpd_Fpjnhkj9e1rhTRjIS40G/view?usp=sharing