Predicting Life Expectancy using Machine Learning

Introduction

Life expectancy can be stated as a statistical measure of average time a human being is
expected to live. This measure depends upon various features such as Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. Life expectancy is an average for all people in the population including those who die shortly after birth, those who die in early adulthood (e.g. childbirth, war), and those who live unimpeded until old age whereas maximum lifespan is an individual-specific concept, it is an upper bound rather than an average.

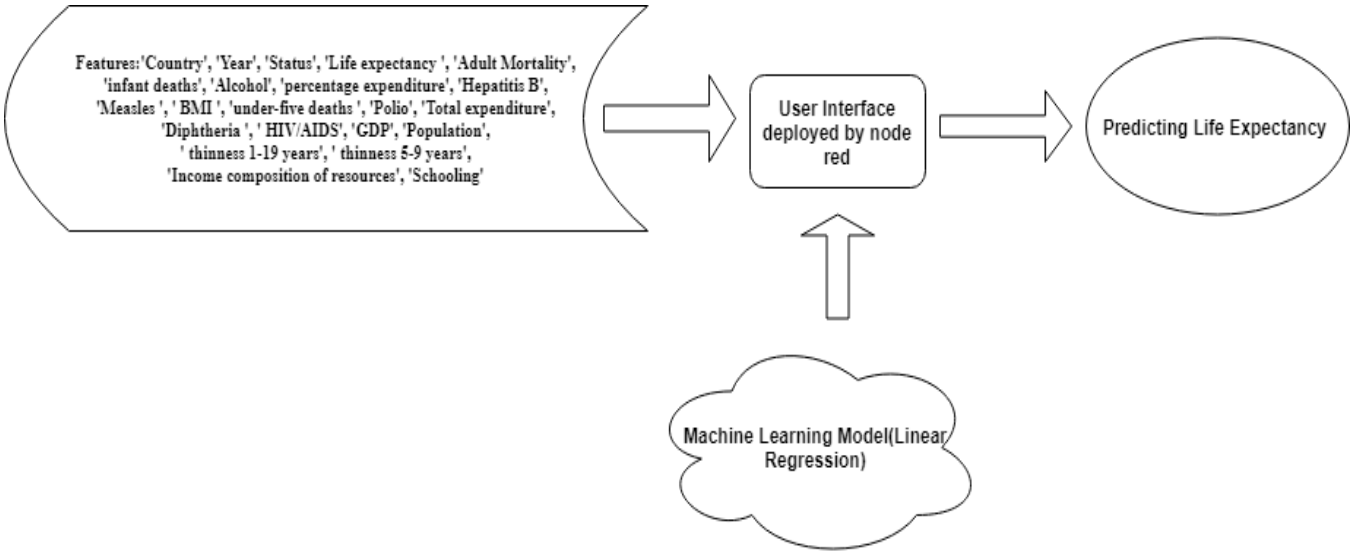
Literature Survey

Many models go in the globe which are accustomed to estimate the life expectancy were characterized by the utilization of life years instead of risk of death, implausible estimates, questionable recommendations and poor clinical feasibility.

Due to this there are some appreciable errors in the prediction so in order to overcome this error I have used features such as Year, Education, Adult Mortality rate, alcohol intake of people in the country, expenditure on healthcare system ,some specific disease related deaths and other demographic factors.

I tested three models on this data and model which gave best accuracy was chosen by me. The model used by me gave the best life-expectancy prediction given all the specified data.

Block Diagram



Requirements

Programming Language:Python 3.7

Notebook: Jupyter notebook and Watson Studio

Machine Learning libraries: pandas,numpy,scikit-learn and scipy Visualization libraries: seaborn,matplotlib

Algorithm:Linear Regression(Partition Trainingset=60% and testing set=40%)

Node-Red to display the flow of project and for UI.

Step 1: Installing and loading all necessary python packages

Step 2:Importing the Life-Expectancy csv file

Step 3:Perform certain visualization and determine the usable features

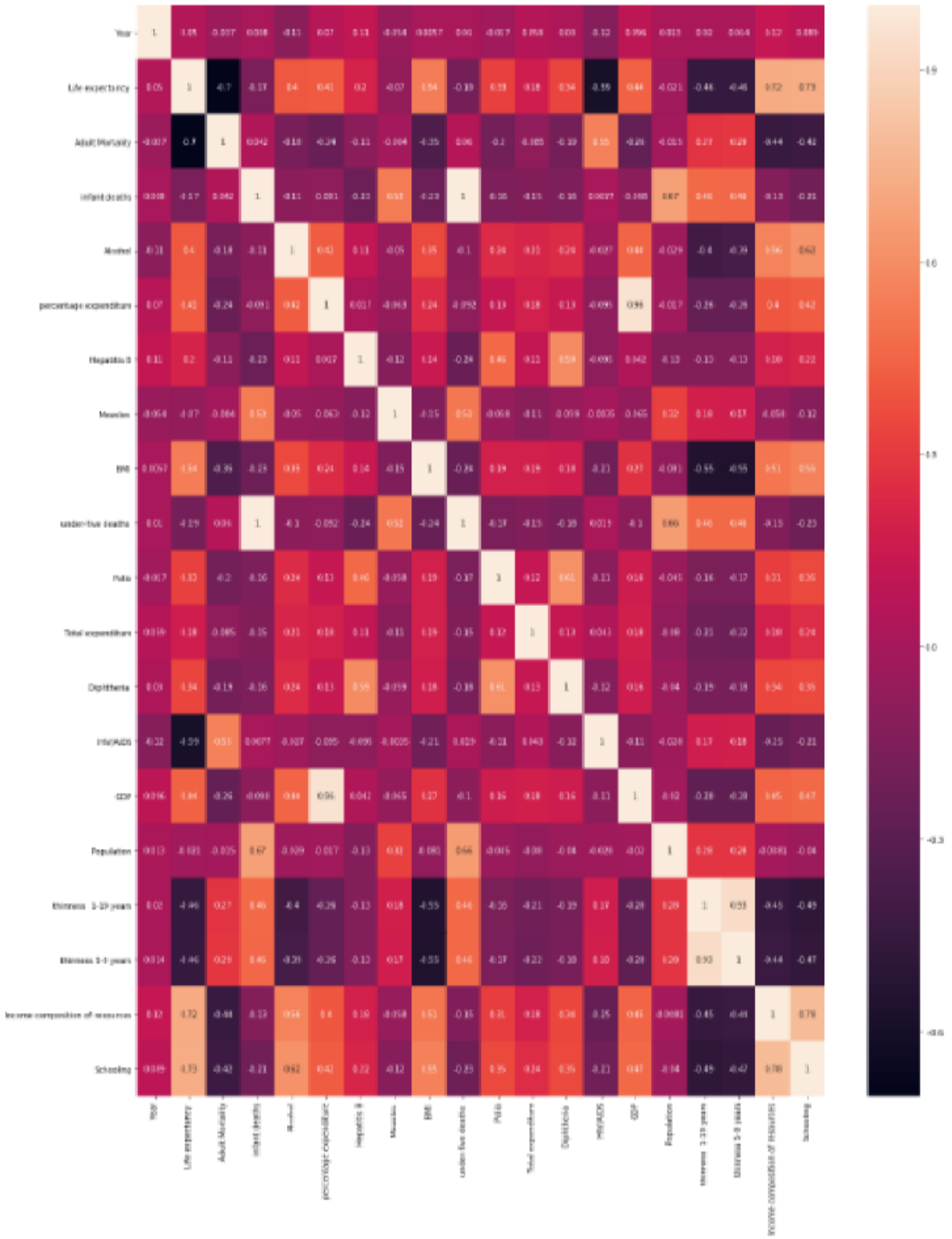
Step 4:Perform preprocessing of data

Step 5: Apply suitable algorithm in order to get a satisfactory prediction of Life-Expectancy.

EXPERIMENTAL INVESTIGATIONS

Analyzing every feature in our dataset is very important which helps us to build a model which gives more accurate result

027b8>



.

IBM Watson Studio | Upgrade | Dheeraj Poojary's Account

My projects / Life-Expectancy / Untitled14

In [8]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [9]:
```
#df=pd.read_csv("C:\\Users\\Dell\\Downloads\\Life Expectancy Data.csv")
```

In [10]:
```
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_67373f8b4e9e4844b2e9563758a9a7d7 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='JkkHr-PJdh8SkKn3_1i19zIH1ri3A9gre04m_5fXcs_8',
    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com')

body = client_67373f8b4e9e4844b2e9563758a9a7d7.get_object(Bucket='lifeexpectancy-donotdelete-pr-zliammnq6xsuym',Key='Life Expectancy Data.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

# If you are reading an Excel file into a pandas DataFrame, replace `read_csv` by `read_excel` in the next statement.
df = pd.read_csv(body)
df.head()
```

Out[10]:

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria	HIV/AIDS

IBM Watson Studio | Upgrade | Dheeraj Poojary's Account

My projects / Life-Expectancy / Untitled14

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2938 entries, 0 to 2937
Data columns (total 22 columns):
Country                          2938 non-null object
Year                             2938 non-null int64
Status                           2938 non-null object
Life expectancy                  2928 non-null float64
Adult Mortality                  2928 non-null float64
infant deaths                    2938 non-null int64
Alcohol                          2744 non-null float64
percentage expenditure           2938 non-null float64
Hepatitis B                      2385 non-null float64
Measles                          2938 non-null int64
 BMI                             2904 non-null float64
under-five deaths                2938 non-null int64
Polio                            2919 non-null float64
Total expenditure                2712 non-null float64
Diphtheria                       2919 non-null float64
 HIV/AIDS                        2938 non-null float64
GDP                              2490 non-null float64
Population                       2286 non-null float64
 thinness  1-19 years            2904 non-null float64
 thinness 5-9 years              2904 non-null float64
Income composition of resources  2771 non-null float64
Schooling                        2775 non-null float64
dtypes: float64(16), int64(4), object(2)
memory usage: 505.0+ KB
```

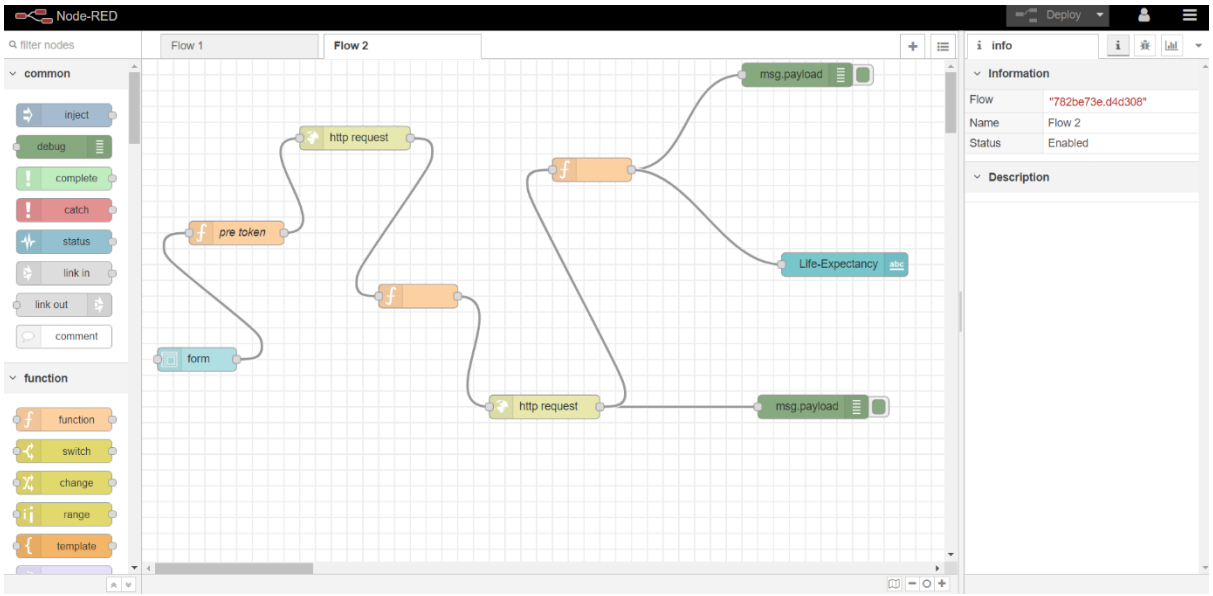
In [12]:
```
df.dropna(inplace=True)
df['Life expectancy ']=df['Life expectancy '].astype(int,copy=True)
```

In [13]:
```
df.describe()
```

Out[13]:

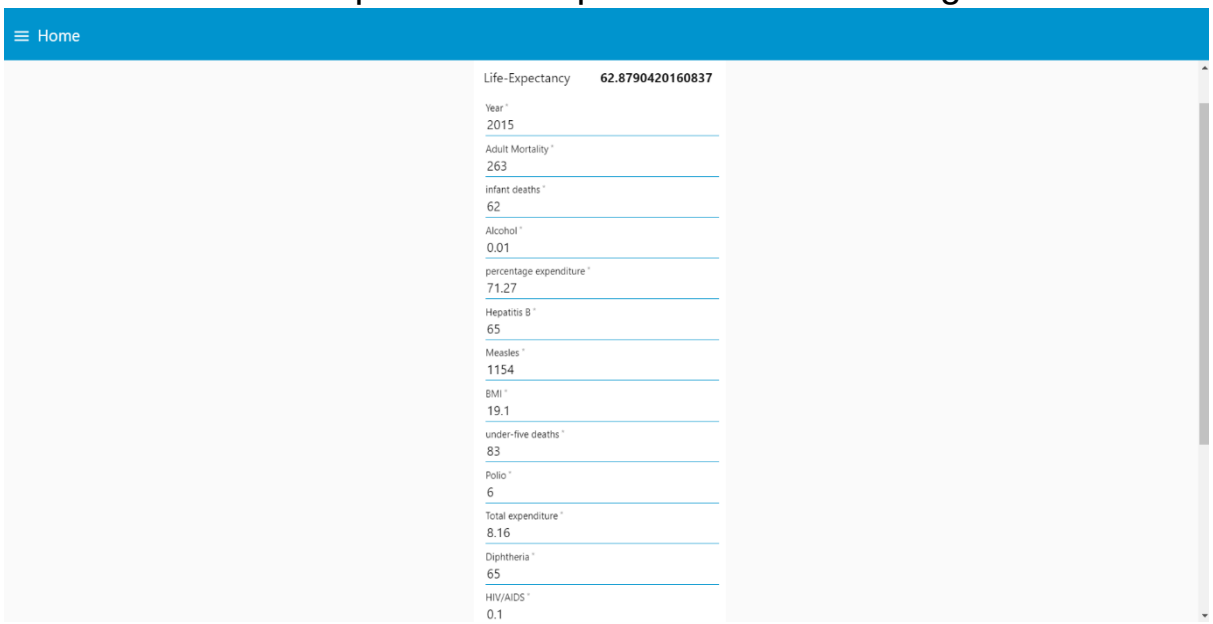
	Year	Life	Adult	infant	Alcohol	percentage	Hepatitis B	Measles	BMI	under-five

Flow



RESULT

The end product or the required result will be a webpage where you need to give all the required inputs and then submit it . Afterwards it will predict the life expectancy value based on your regression technique. Scatter Plot and Hist plot between predicted data and original data



Scatter and hist plot between predicted and original data

ADVANTAGES:

1)Helpful for a country for analysis: It will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

Life expectancy can be used nationally to monitor health inequalities of a country.

2) Reduced Costs: This is a simple webpage and can be accessed by any citizen of a country to calculate life expectancy of their country and does not required any kind of payment neither for designing nor for using.

3) User Friendly Interface: This interface requires no background knowledge of how to use it. It's a simple interface and only ask for required values and predict the output.

DISADVANTAGES:

a)Prediction may not be accurate: As it depends completely on user, so if user provides some wrong values then it will predict wrong value.

b)User input is not saved in any database.

APPLICATION

a) It can be used to monitor health inequalities of a country.

b) It can be used to develop statistics for country development process.

c) It can be used to analyse the factors for high life expectancy.

CONCLUSION

We have prepared a model that will predict the life expectancy of a person in a country.

The end product which is a user interface which will be useful for the user to predict life expectancy value of their own country or any other country based on some required details such as GDP, BMI, Year, Alcohol Intake, Total expenditure and etc.