PREDICTING LIFE EXPECTANCY USING PYTHON

By: Ritika Rao

INDEX

1. Introduction

1.1. Overview

Life expectancy is a statistical measure of the average time a person is expected to live, based on the year of its birth, its current age, and other demographic factors including gender.

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. However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

To predict the life expectancy rate of a given country, we will be using Machine Learning algorithms to draw inferences from the given dataset and provide an output. For better usability by the customer, we are also going to create a UI for the user to interact with using Node-Red.

1.2. Purpose

- Personal Purpose

  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.

- Economic Purpose

  Predicting life expectancy would play a vital role in judging the growth and development of the economy. The most obvious explanation behind the connection between life expectancy and income is the effect of food supply on mortality. Higher-income also implies better access to housing, education, health services and other items which tend to lead to improved health, lower rates of mortality and higher life expectancy.

  Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to a rise in the workforce of a nation.

- Health Sector

  Based on the factors used to calculate the 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

  The insurance sector will be able to provide individualized services to people based on life expectancy outcomes and factors.

- 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 skill set acquired by people over many years.

2. Literature Survey

2.1. Existing Solution

As a result of the evolution of biotechnologies and related technologies such as the development of sophisticated medical equipment, humans can enjoy longer life expectancies than previously before. Predicting a human's life expectancy has been a long-term question to humankind. Many calculations and research have been done to create an equation despite it being impractical to simplify these variables into one equation.

Currently, there are various smart devices and applications such as smartphone apps and wearable devices that provide wellness and fitness tracking. Some apps offer health-related data such as sleep monitoring, heart rate measuring, and calorie expenditure collected and processed by the devices and servers in the cloud. However, no actual works provide the Personalized Life expectancy.

2.2. Proposed Solution

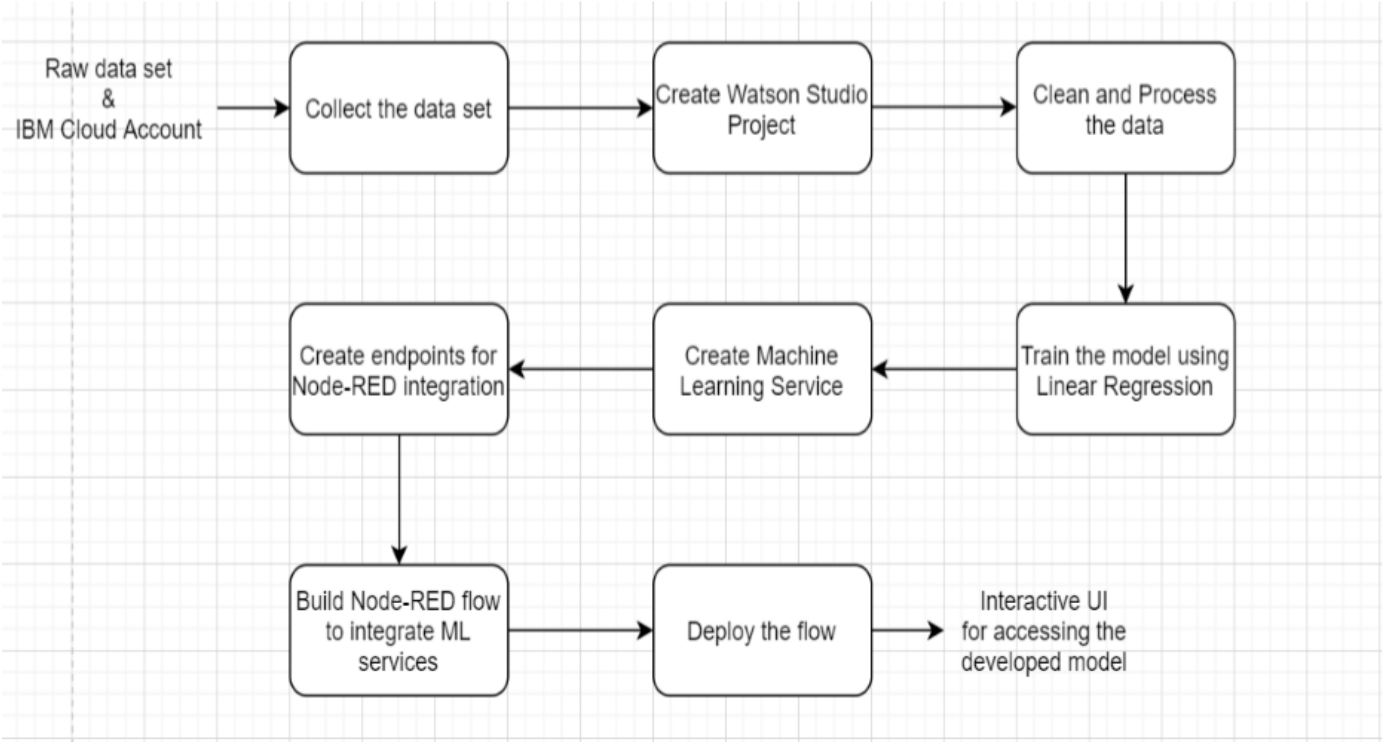
Recent advancements in machine learning industry have enabled many thoughts that were considered impossible to be achieved. Machine learning is capable of doing many wonders in the field of healthcare.

The proposed solution involves the use of machine learning algorithms, specifically Regression models, to predict life expectancy. Here we recommend a method for the forecasting life expectancy of an individual from a country taking into certain factors such as Adult Mortality rate, Infant deaths, Alcohol, Hepatitis B, Measles, BMI, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP of a country, Population, Income composition of resources, Schooling and status of the country in terms of Developing or Developed.

This machine learning model will be made accessible to the users by integrating it with Node-Red to create an interactive and user-friendly User Interface.

3. Theoretical Analysis

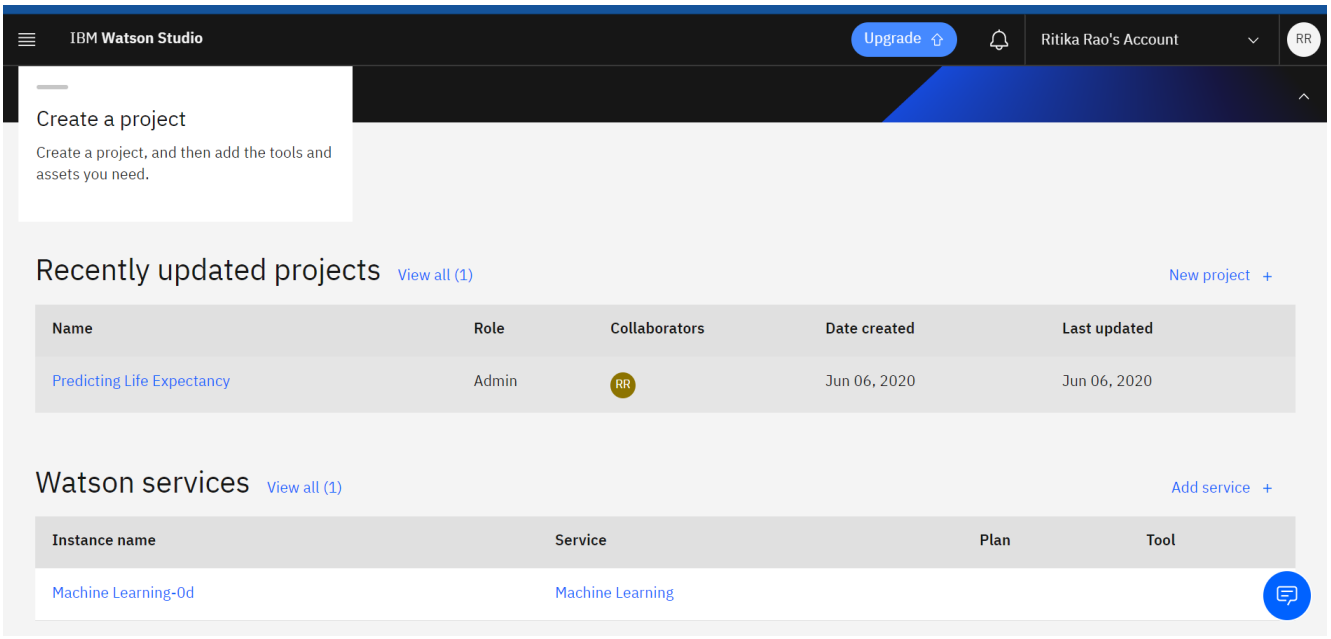
3.1. Block Diagram

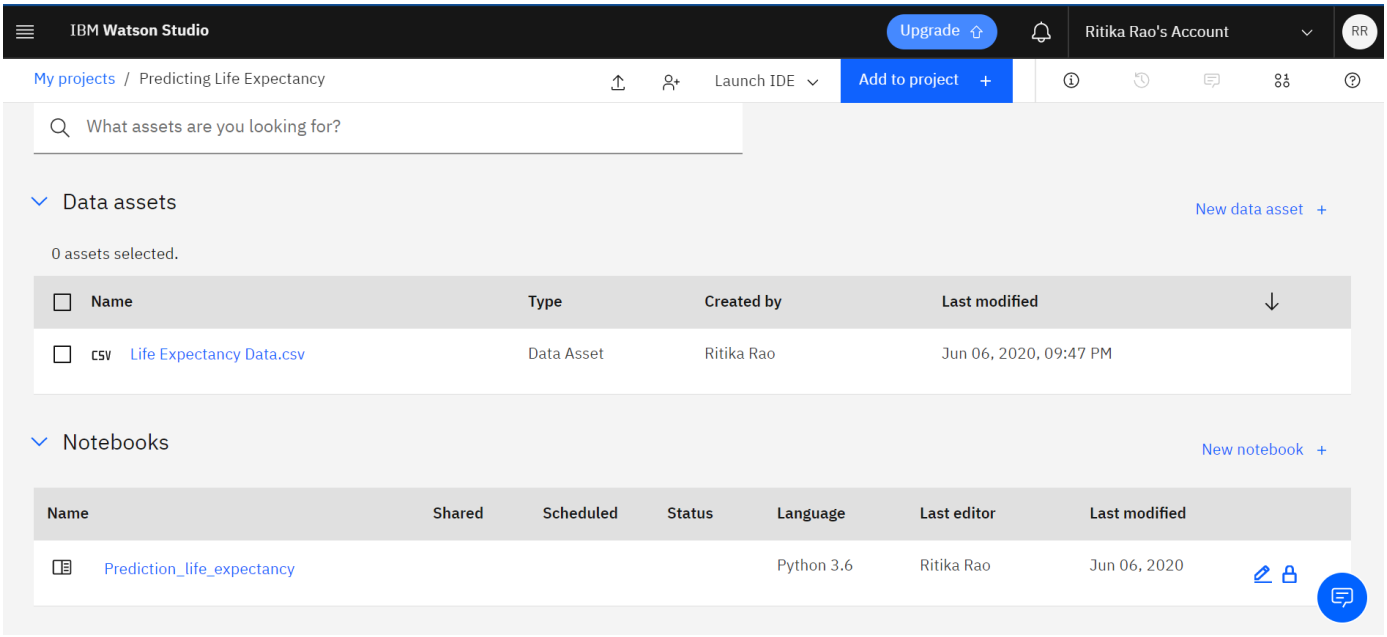


3.2. Hardware/ Software Designing

Model Designing **(Watson Studio):**

Steps: New Project => Create an empty Project => Give project name => Click Create => Add to Project => Notebook


IBM Watson Studio
Upgrade
Ritika Rao's Account
RR
Create a project
Create a project, and then add the tools and assets you need.
Recently updated projects
View all (1)
New project +
Name
Role
Collaborators
Date created
Last updated
Predicting Life Expectancy
Admin
RR
Jun 06, 2020
Jun 06, 2020
Watson services
View all (1)
Add service +
Instance name
Service
Plan
Tool
Machine Learning-0d
Machine Learning



Scoring Endpoint:

For wml credentials, replace with your own credentials of the service.

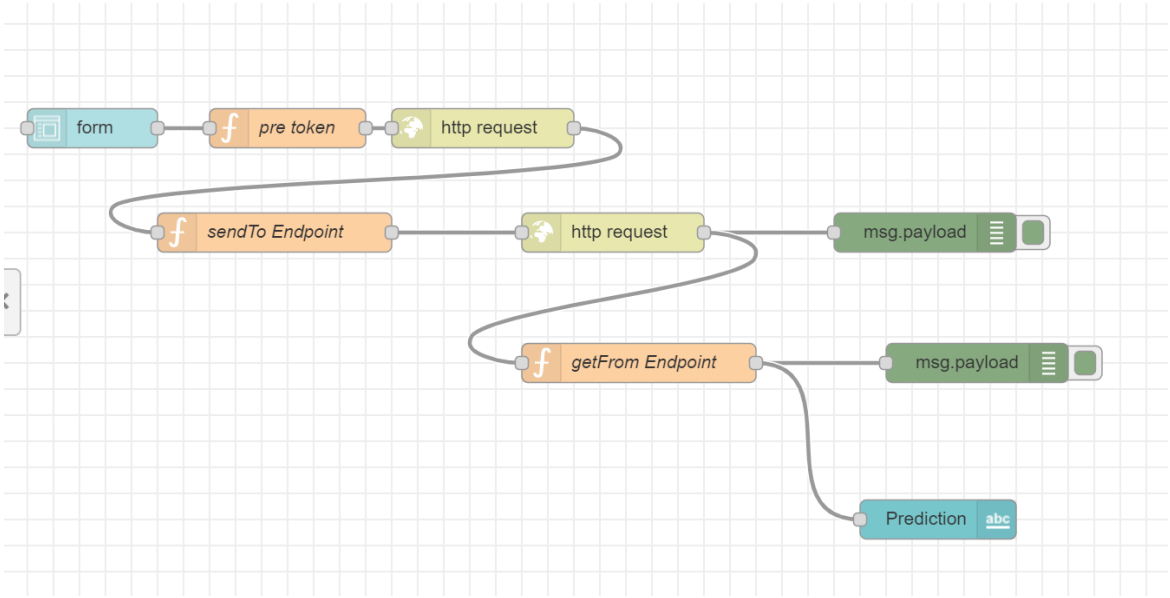
Services => Machine Learning Service => Service Credentials => Copy the credentials

```
In [90]: deployment = client.deployments.create(published_model_uid, name="Prediction_life_expectancy")

#######################################################################################################
####

Synchronous deployment creation for uid: '4f013ef8-e272-4b5c-9540-1ac1d003da6f' started

#######################################################################################################
####


INITIALIZING
DEPLOY_SUCCESS


------------------------------------------------------------------------------------------------
Successfully finished deployment creation, deployment_uid='b2e8d020-98db-44cd-9f0d-9334cf2bb23b'
------------------------------------------------------------------------------------------------


In [91]: scoring_endpoint = client.deployments.get_scoring_url(deployment)

In [92]: scoring_endpoint
Out[92]: 'https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/f6bcad98-e36e-40d2-bc1c-f41034f7dc4c/deployments/b2e8d020-98db-44cd-9f0d-9334cf2b
b23b/online'
```

User Interface Integration with Ml Model(Node-Red):

Nodes used:

1. Form Node: Edit => Add New UI Tab
2. Function Node: To obtain access to Machine Learning Services. Requires API Key
3. HTTP Request Node: POST method and returns a parsed JSON object. Gains access to Machine Learning services.



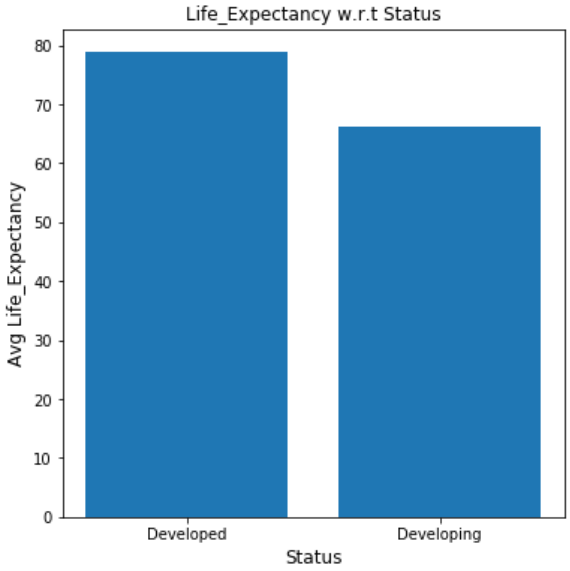
4. Experimental Investigations

The WHO data set for life expectancy contains 22 columns in total. The data set has historical data of life expectancy for the period between 2000 to 2016. There are numerous rows with null values for some of the columns and outliers. These null values are replaced with the mean for the respective columns. After which the outliers are adjusted. All the columns except the "Country" and "Status" are of integer type. The country column is not included in the model training because of its less relation to the "Life Expectancy" column. The "Status" column is changed to integer type such that "Developing" is mapped to 1 and "Developed" is mapped to 0. After these steps, the data is ready for the model to be trained.

Analyzing the relations between various features can help us improve the performance of the model as well as decide which model would be more suitable. We decided using many visualizations.

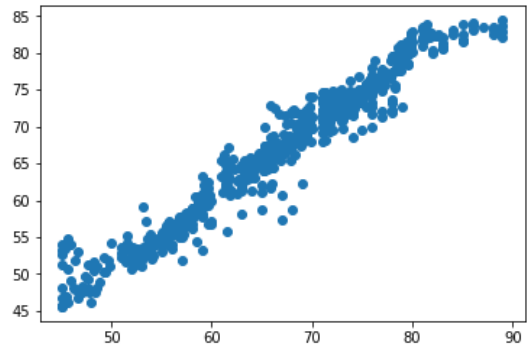
	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria	HIV/A
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	8.16	65.0	0.1
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.18	62.0	0.1
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	64.0	0.1
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	67.0	0.1
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	68.0	0.1

5 rows × 22 columns

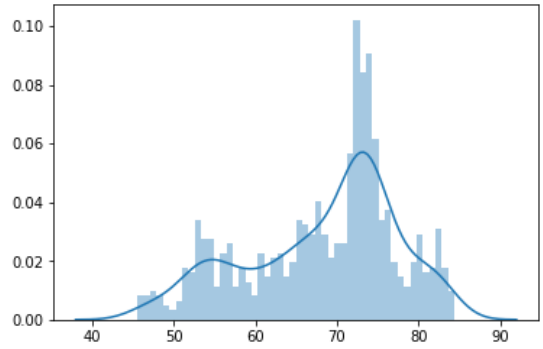


Results after applying the machine learning model

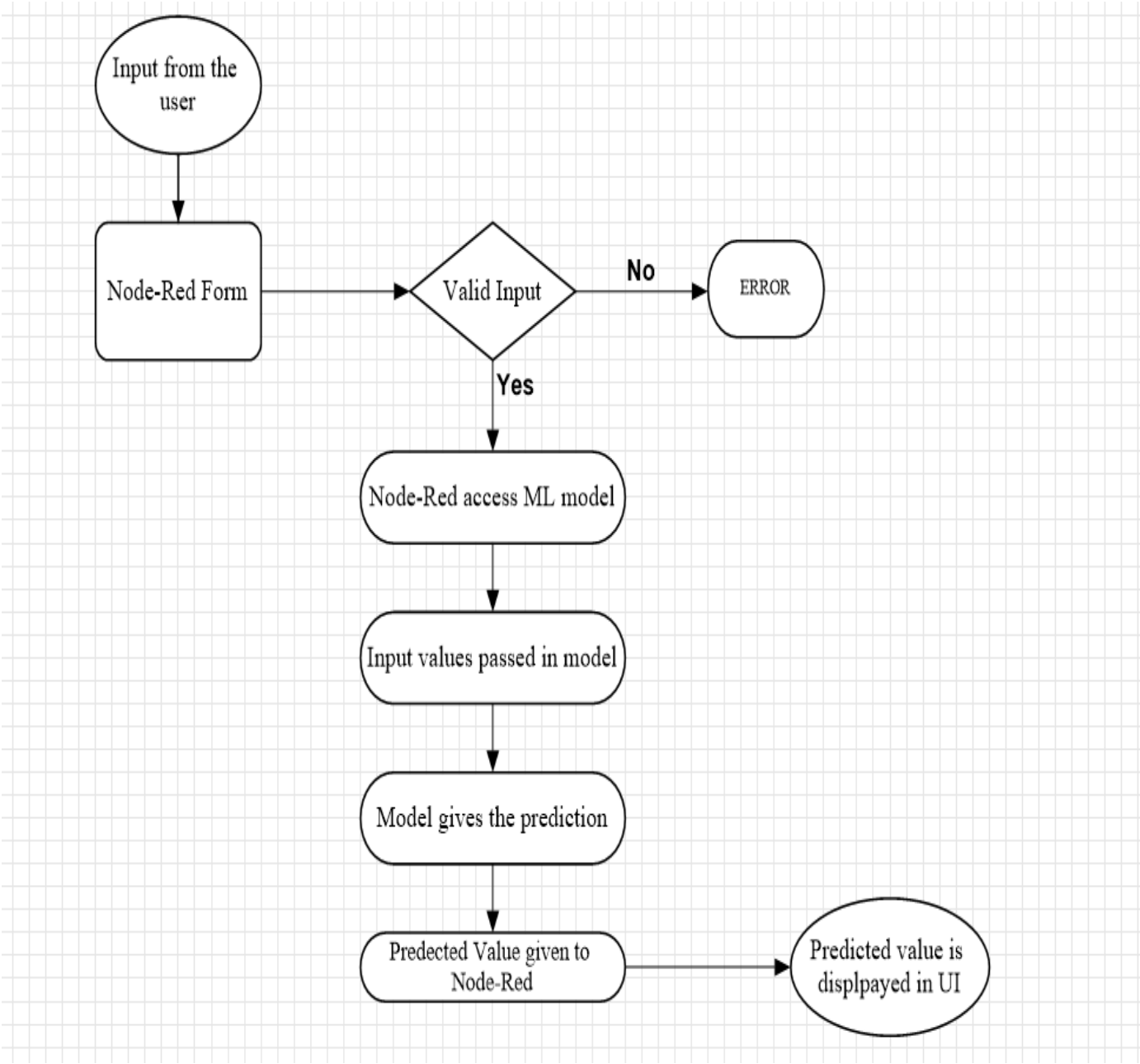
<matplotlib.collections.PathCollection at 0x7f749dda1cf8>



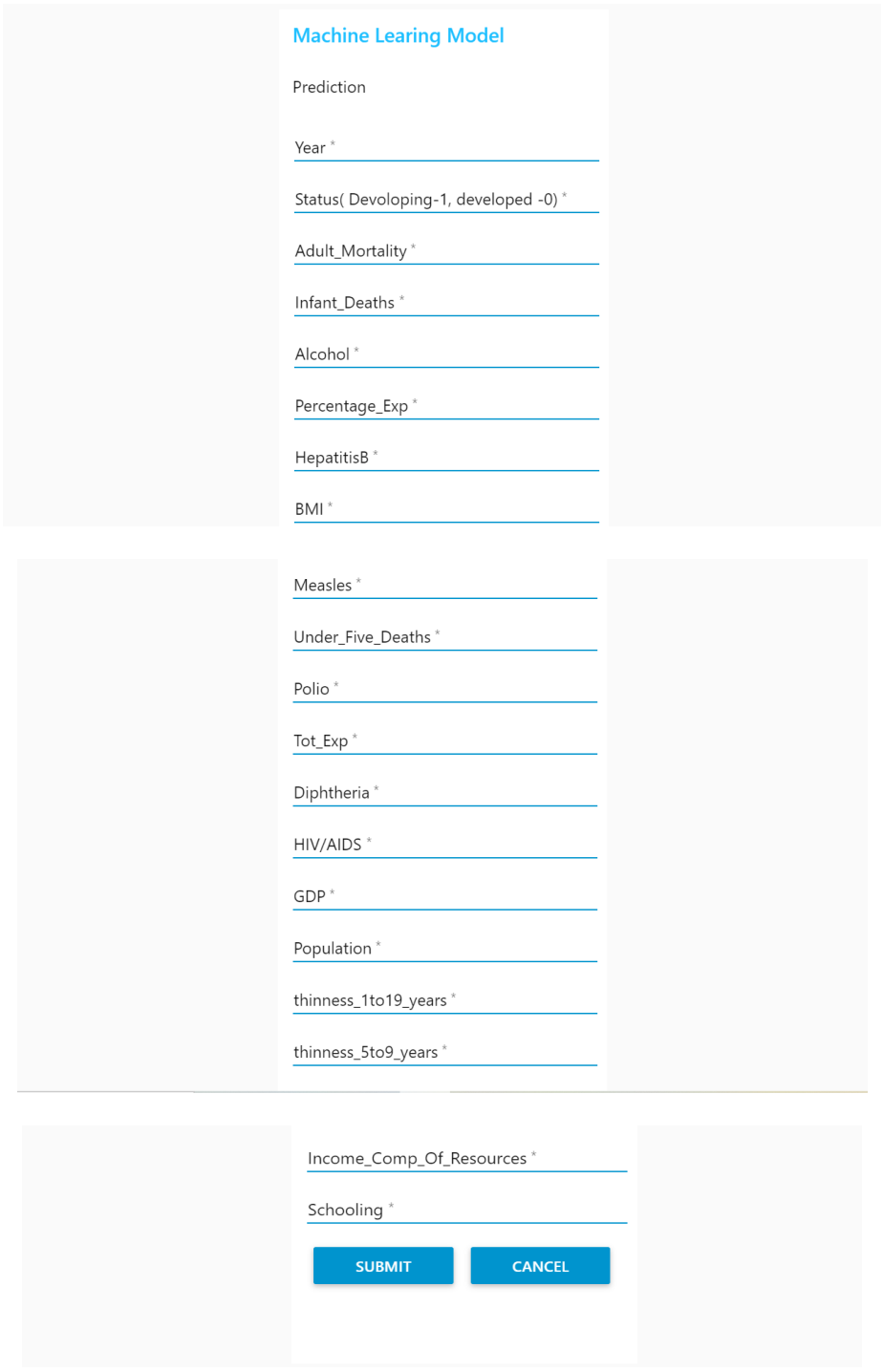
<matplotlib.axes._subplots.AxesSubplot at 0x7f749dd995f8>



5. Flowchart


Input from the user
Node-Red Form
Valid Input
No
ERROR
Yes
Node-Red access ML model
Input values passed in model
Model gives the prediction
Predected Value given to Node-Red
Predicted value is displpayed in UI

6. Result


Machine Learing Model
Prediction
Year *
Status(Devoloping-1, developed -0) *
Adult_Mortality *
Infant_Deaths *
Alcohol *
Percentage_Exp *
HepatitisB *
BMI *
Measles *
Under_Five_Deaths *
Polio *
Tot_Exp *
Diphtheria *
HIV/AIDS *
GDP *
Population *
thinness_1to19_years *
thinness_5to9_years *
Income_Comp_Of_Resources *
Schooling *
SUBMIT
CANCEL

7. Advantages and Disadvantages

Advantages:

One of the most significant advantages of embedding machine learning algorithms is their ability to improve over time. Machine learning technology typically enhances efficiency and accuracy thanks to the ever-increasing amounts of data that are processed.

The application learns the patterns and trends hidden within the data without human intervention which makes predicting much simpler and more manageable. As the amount of data used increases, accuracy also enhances with it. It is also the critical component in technologies for automation.

Using Node-Red also simplifies the effort put into a creating the front-end. The programmer doesn't need extensive knowledge of HTML and JavaScript. It also makes the integration between Machine learning model and the UI much more comfortable.

Disadvantages:

Using machine learning interface comes with its problems. Since the whole point of it is minimize human involvement, it also makes error detection and fixing much more problematic. It takes much time to identify the root cause of the problem.

Machine learning can also be very time-consuming. When the size of the data fed to the machine learning is enormous, the computational cost and the time taken to train the model on the data increases drastically. It can increase the value of resources required to implement the application on a large scale.

At the same time, Node-Red does not give many features to customize our UI.

8. Applications

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

2) 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 skill set 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 a rise in the "manpower" of a nation. The knowledge asset of a country increases with the number of individuals in a country.

3) Health Sector: The outcome will help the healthcare sector to fund and provide better services to those in need.

4) Insurance Companies: Insurance sector will be able to provide individualized services to people based on life expectancy outcomes and factors.

9. Conclusion

Predicting the lifespan of human beings can significantly alter our lives. Human behaviour and activities are unpredictable, and it may almost be impossible to predict lifespan correctly. 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 familiar people. It can also increase the ease of access to individuals.

With the help of Machine Learning algorithms, one can ease the process of automating the application and predicting the expectancy with reasonable accuracy. It also reduces the effort and time put into deploying the application and making it more accessible to the users.

10. Future Scope

For future use, one can integrate the life expectancy prediction by providing suggestions and medications to the individual using the application. It will help predict as well as increase the individual's life expectancy.

The scalability and flexibility of the application will also improve with the advancement in technology.
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.

11. Bibliography

- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as- web-service
- https://www.ibm.com/watson/products-services
- https://www.allbusinesstemplates.com/download/?filecode=2KBA4&lang=en&iuid=9f9faa 69-9fab-40ee-8457-ea0e5df8c8de
- https://www.datasciencesociety.net/using-machine-learning-to-explain-and-predict-the-life-expectancy-of-different-countries/
- https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4

12. Appendix

12.1. Source Code

https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/a54175c2-a9f5-421d-8028-805f385d6458/view?access_token=8040add831f9406d89812bd0705d7d0b1a5e578c03aec6a653d3e9ea9a291fdd

Services Used:

- **Watson Assistant**
- **Watson Studio**
- **IBM Cloud Function**
- **Node-Red**