

Predicting Life Expectancy Using Machine Learning

BY: Rashmi Praveen Mishra

INDEX

S. No	Title	Page No.
1	Introduction	3
2	Literature Survey	4
3	Theoretical Analysis	5-7
4	Experimental Investigations	8
5	Flowchart	9
6	Result	10-12
7	Advantages and Disadvantages	13
8	Applications	14
9	Conclusion	14
10	Future Scope	15
11	Bibliography	15
12	Source Code	16

1. Introduction

1.1. Overview

Life expectancy refers to the number of years a person is expected to live. In mathematical terms, life expectancy refers to the expected number of years remaining for an individual at any given age.

The life expectancy for a particular person or population group depends on several variables such as their lifestyle, access to healthcare, diet, economic status and the relevant mortality and morbidity data. However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

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

1.2. Purpose

Economic growth

Predicting life expectancy would play a vital role in judging the growth and development of the economy.

Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country.

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 skillset acquired by people over many years.

Personal growth

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.

Growth in Health Sector

Based on the factors used to calculate 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

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

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 are able to 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 provide 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 existing works provide the Personalized Life expectancy.

2.2. Proposed Solution

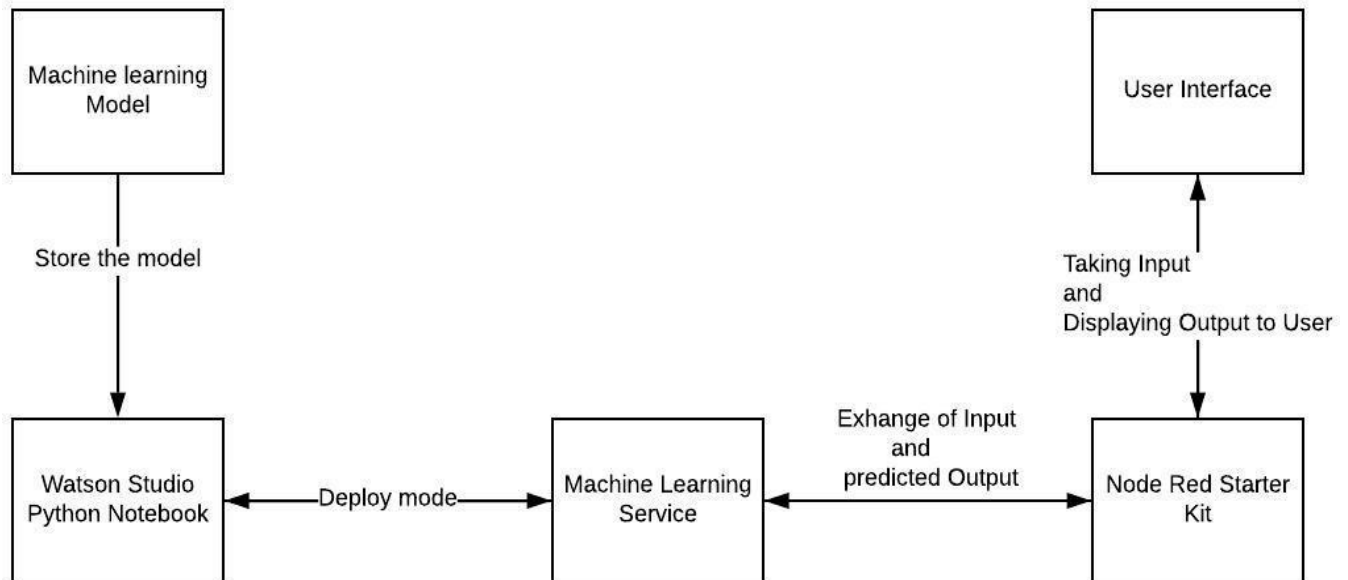
There has been an explosion of breakthroughs in the field of Machine Learning over the past few years. Machine Learning algorithms are capable of a lot and can-do wonders for the healthcare sector.

The proposed solution involves the use of Machine Learning algorithms specifically Regression models such as Linear Regression, Ridge regression, etc. Life expectancy is highly correlated over time among countries and between males and females. These associations can be used to improve forecasts. Here we propose a method for 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

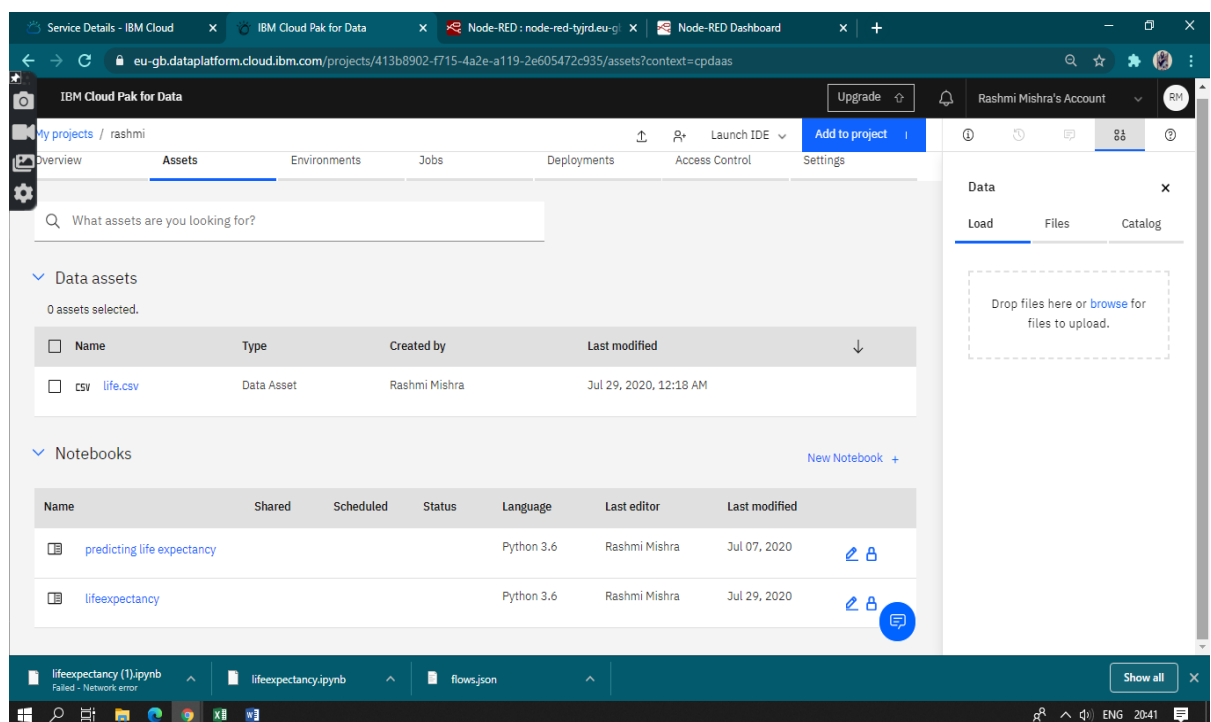
3.1. Block Diagram



3.2. Hardware/ Software Designing

Model Designing (Watson Studio) :

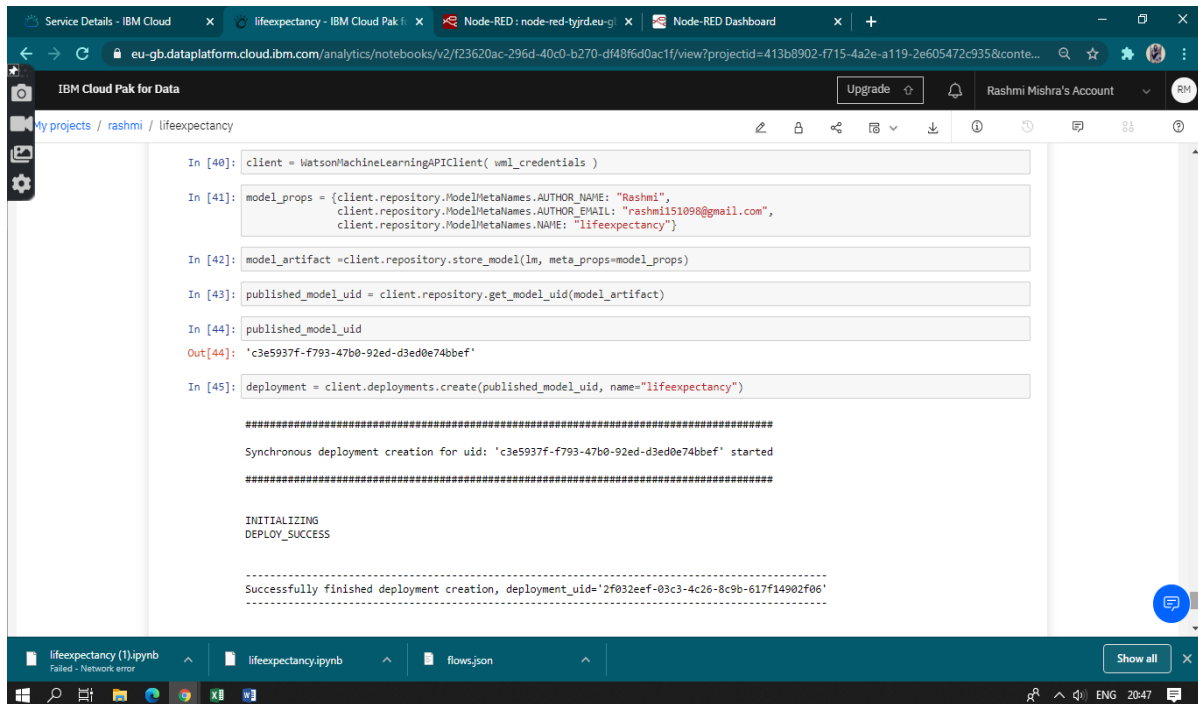
Steps: New Project => Create an empty Project => Give project name => Click Create => Add to Project => Notebook



Scoring Endpoint:

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

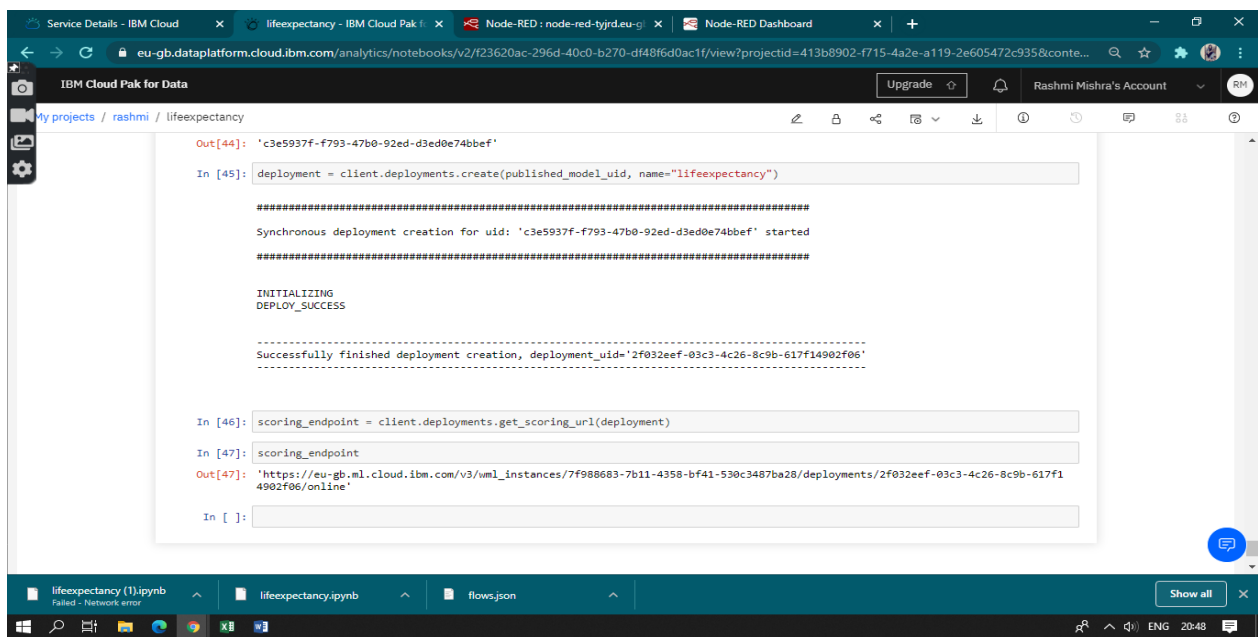
Services => Machine Learning Service => Service Credentials => Copy the credentials



The screenshot shows a Jupyter Notebook interface within the IBM Cloud Pak for Data environment. The notebook is titled 'lifeexpectancy' and is located in the 'rashmi' project. The code in the notebook performs the following steps:

- Imports the `WatsonMachineLearningAPIClient` using `wml_credentials`.
- Creates `model_props` with repository metadata including author name, email, and model name.
- Stores the model artifact using `client.repository.store_model`.
- Retrieves the `published_model_uid` using `client.repository.get_model_uid`.
- Creates a deployment using `client.deployments.create` with the `published_model_uid` and name 'lifeexpectancy'.
- Prints a confirmation message: 'Successfully finished deployment creation, deployment_uid='2f032eef-03c3-4c26-8c9b-617f14902f06''.

The output of the notebook shows the deployment creation process, including a 'Synchronous deployment creation' message and a 'Successfully finished deployment creation' message.



This screenshot shows the continuation of the Jupyter Notebook from the previous image. The code in the notebook performs the following steps:

- Retrieves the `scoring_endpoint` using `client.deployments.get_scoring_url` with the `deployment` object.
- Prints the `scoring_endpoint` value: 'https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/7f988683-7b11-4358-bf41-530c3487ba28/deployments/2f032eef-03c3-4c26-8c9b-617f14902f06/online'.

The output of the notebook shows the `scoring_endpoint` value.

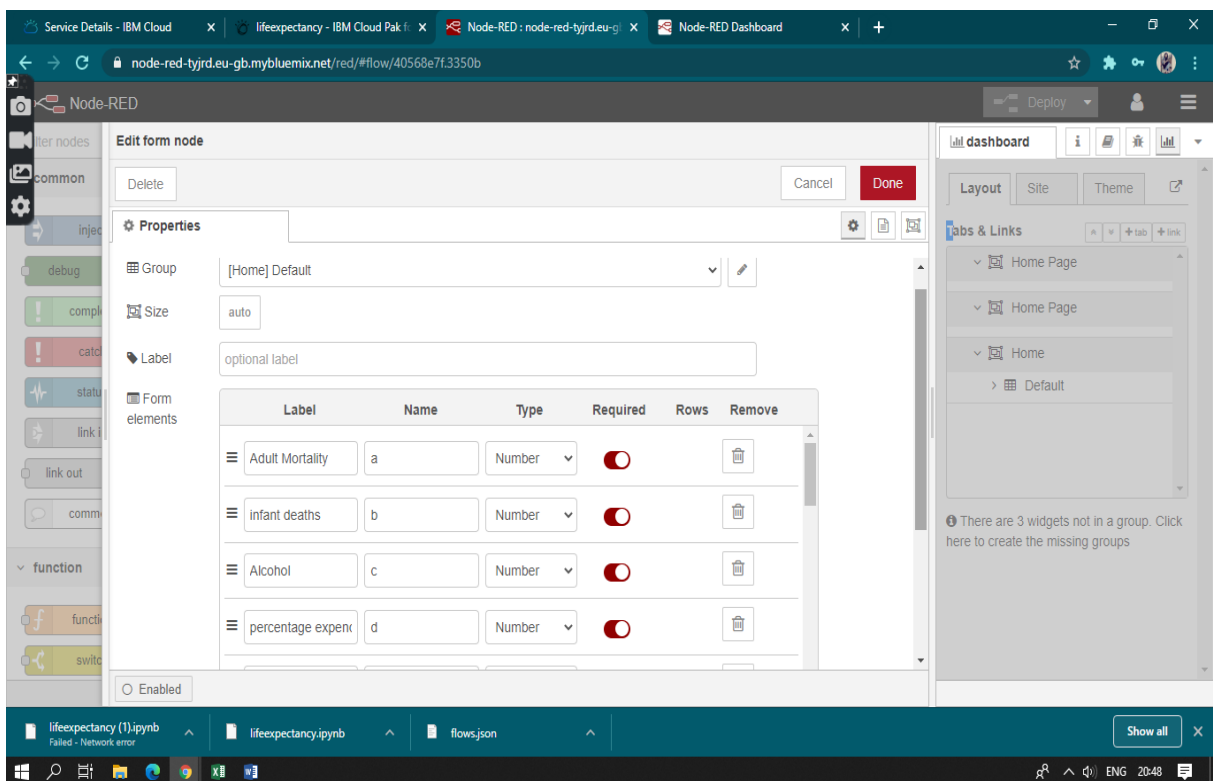
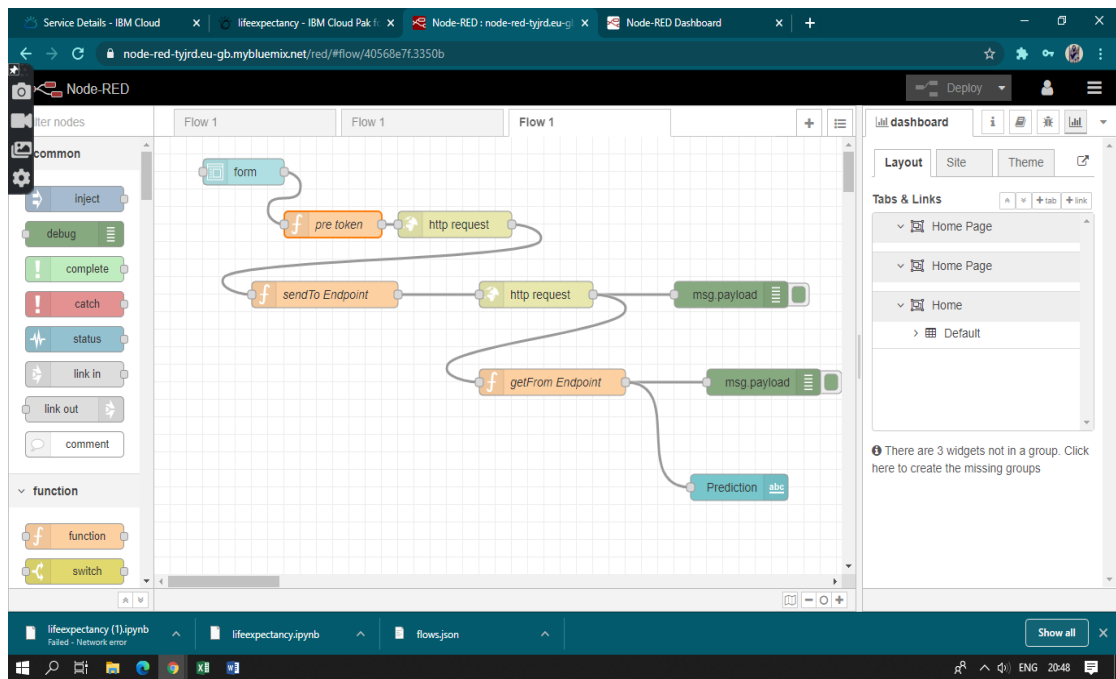
User Interface Integration with ML Model (Node-

Red) : Nodes: 1) Form Node: Edit => Add

New UI Tab

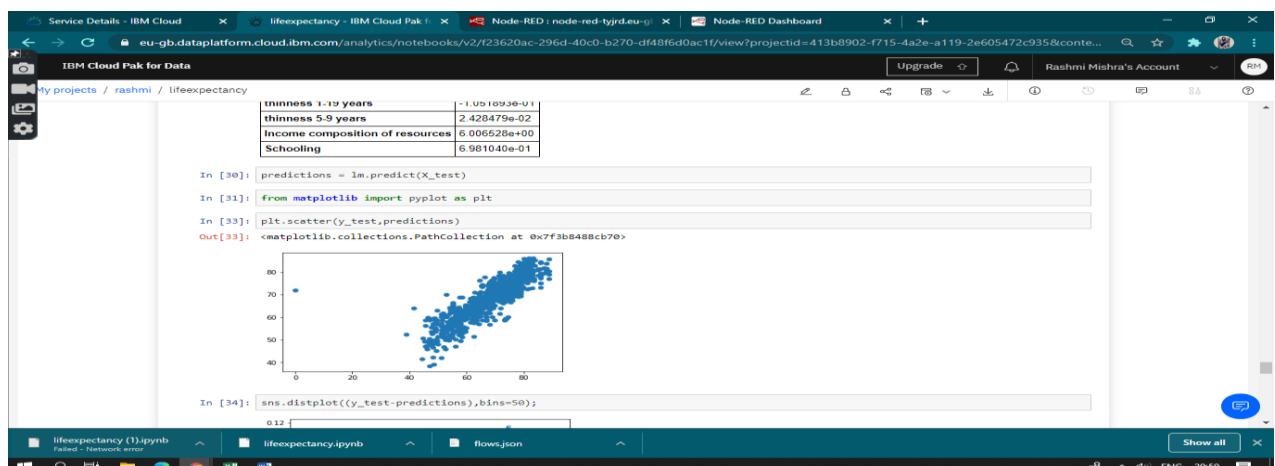
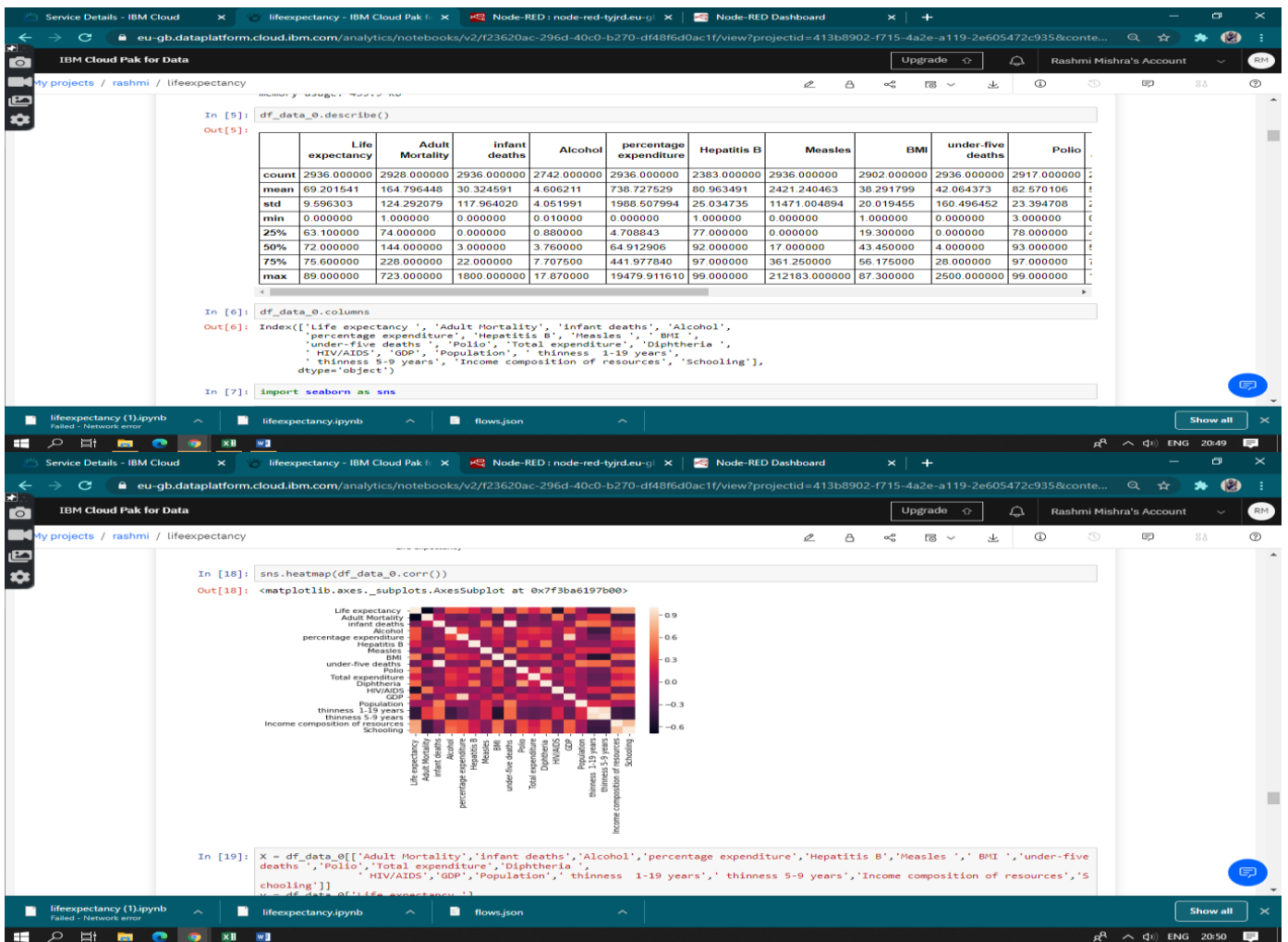
2) Function Node: To obtain access to Machine Learning Services. Requires API Key

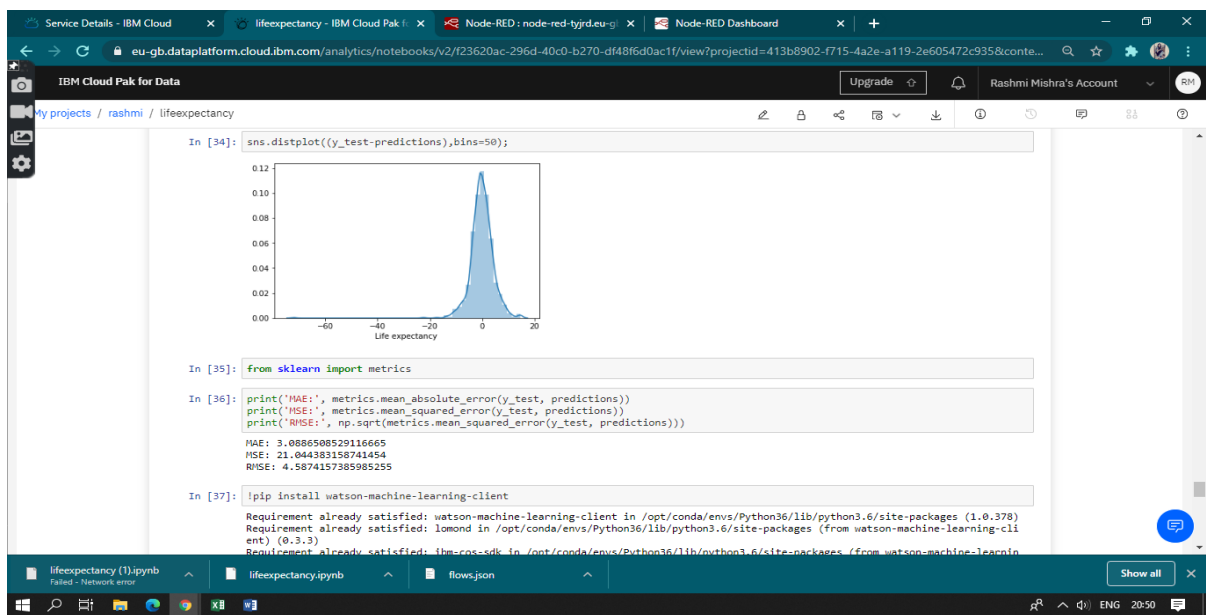
3) HTTP Request Node: POST method and returns a parsed JSON object.
Gains access to Machine Learning services.



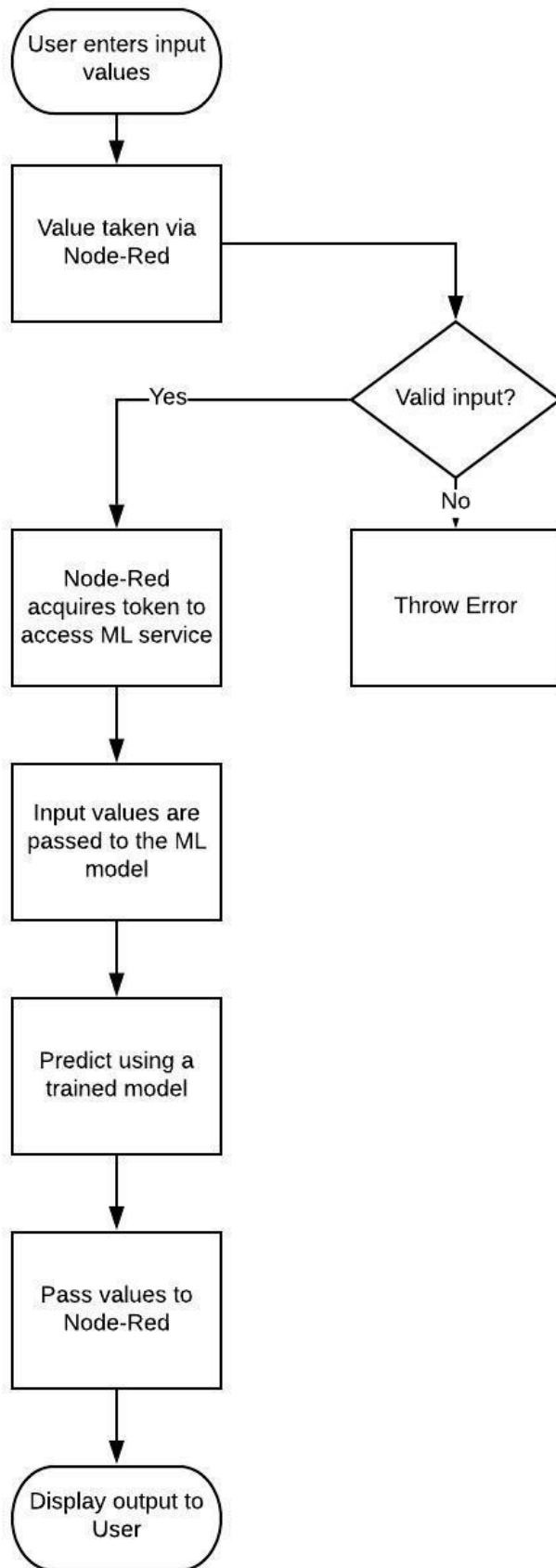
4. Experimental Investigations

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.

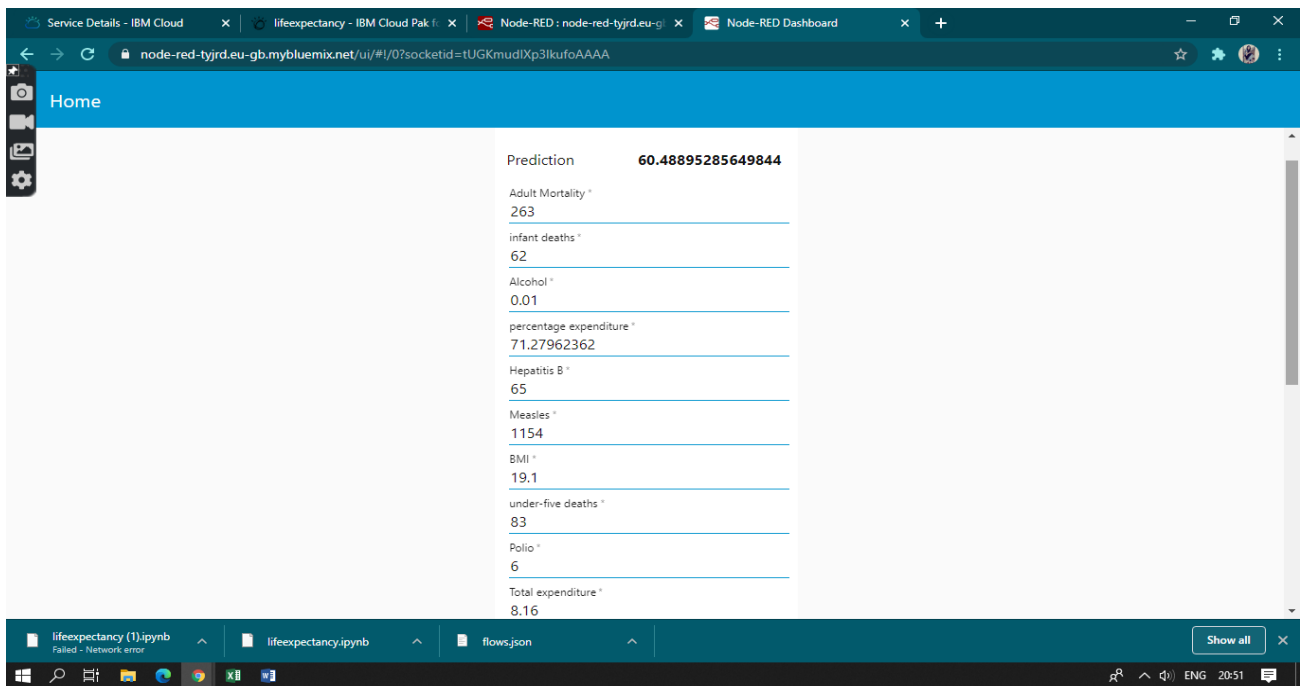




5. Flowchart



6. Result



Service Details - IBM Cloud x lifeexpectancy - IBM Cloud Pak f x Node-RED: node-red-tyjrd.eu-5 x Node-RED Dashboard x

node-red-tyjrd.eu-gb.mybluemix.net/ui/#/0?socketid=tUGK mudIXp3lkufoAAAA

Home

Total expenditure *
8.16

Diphtheria *
65

HIV/AIDS *
0.1

GDP *
584.25921

Population *
33736494

thinness 1-19 years *
17.2

thinness 5-9 years *
17.3

Income composition of resources *
0.479

Schooling *
10.1

SUBMIT CANCEL

lifeexpectancy (1).ipynb Failed - Network error lifeexpectancy.ipynb flows.json Show all x

Windows taskbar: 20:51 ENG

7. Advantages and Disadvantages

Advantages:

One of the biggest advantages of embedding machine learning algorithms is their ability to improve over time. Machine learning technology typically improves 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 easier. The more data is fed to the algorithm, the higher the accuracy of the algorithm is. It is also the key component in technologies for automation.

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

Disadvantages:

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

Machine learning can also be very time-consuming. When the size of the data fed to the machine learning is very large, the computational cost and the time taken to train the model on the data increases drastically. This can increase the cost 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 own life expectancy by inputting values in the corresponding fields. This could help make people more aware of their general health, and its improvement or deterioration over time. This 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 skillset 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 an increase in the “manpower” of a country. The knowledge asset of a country increases with the number of individuals in a country.
- 3) Health Sector: Based on the factors used to calculate life expectancy of an individual and the outcome, health care will be able to fund and provide better services to those with greater need.
- 4) Insurance Companies: Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

9. Conclusion

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

With the help of Machine Learning algorithms, one can ease the process of automating the application and predicting the expectancy with an admirable 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 with providing suggestions and medications to the individual using the application. This will help predict as well as increase the individual's life expectancy.

The scalability and flexibility of the application can also be improved with advancement in technology and availability of new and improved resources. 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=9f9faa69-9fab-40ee-8457-ea0e5df8c8de>

12.

Appendix

12.1. Source Code

Services Used:

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

Python Notebook:

```
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_774c9cb3cf354dd5a555fd276d5e2f19 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='-QAUDh9lOr5_MMgSXgrSg-Af8eMA5kuEIXWBaldle5al',
    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_774c9cb3cf354dd5a555fd276d5e2f19.get_object(Bucket='rashmi-
donotdelete-pr-ittr6rztoojdfn',Key='life.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
)

df_data_0 = pd.read_csv(body)
df_data_0.head()

df_data_0.head()

df_data_0.info()

df_data_0.describe()

df_data_0.columns

import seaborn as sns

sns.pairplot(df_data_0)

df_data_0.isnull()

df_data_0.fillna(0)

df_data_0['Life expectancy '].fillna(df_data_0['Life expectancy '].mean())
```

```

df_data_0['Adult Mortality'].fillna(df_data_0['Adult Mortality'].mean())

df_data_0['infant deaths'].fillna(df_data_0['infant deaths'].mean())

import numpy as np

df_data_0['Adult Mortality'] = df_data_0['Adult
Mortality'].replace(np.nan,df_data_0['Adult Mortality'].mean())
df_data_0['Alcohol'] =
df_data_0['Alcohol'].replace(np.nan,df_data_0['Alcohol'].mean())
df_data_0['Hepatitis B'] = df_data_0['Hepatitis
B'].replace(np.nan,df_data_0['Hepatitis B'].mean())
df_data_0[' BMI '] = df_data_0[' BMI '].replace(np.nan,df_data_0[' BMI '].mean())
df_data_0['Polio'] = df_data_0['Polio'].replace(np.nan,df_data_0['Polio'].mean())
df_data_0['Total expenditure'] = df_data_0['Total
expenditure'].replace(np.nan,df_data_0['Total expenditure'].mean())
df_data_0['Diphtheria '] = df_data_0['Diphtheria
'].replace(np.nan,df_data_0['Diphtheria '].mean())
df_data_0['GDP'] = df_data_0['GDP'].replace(np.nan,df_data_0['GDP'].mean())
df_data_0['Population'] =
df_data_0['Population'].replace(np.nan,df_data_0['Population'].mean())
df_data_0[' thinness 1-19 years'] = df_data_0[' thinness 1-19
years'].replace(np.nan,df_data_0[' thinness 1-19 years'].mean())
df_data_0[' thinness 5-9 years'] = df_data_0[' thinness 5-9
years'].replace(np.nan,df_data_0[' thinness 5-9 years'].mean())
df_data_0['Income composition of resources'] = df_data_0['Income composition of
resources'].replace(np.nan,df_data_0['Income composition of resources'].mean())
df_data_0['Schooling'] =
df_data_0['Schooling'].replace(np.nan,df_data_0['Schooling'].mean())

df_data_0.isnull().sum()

sns.distplot(df_data_0['Life expectancy '])

sns.heatmap(df_data_0.corr())

X = df_data_0[['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']]
y = df_data_0['Life expectancy ']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4,
random_state=101)

from sklearn.linear_model import LinearRegression
lm = LinearRegression()
lm.fit(X_train,y_train)
np.where(df_data_0.values >= np.finfo(np.float64).max)
(array([], dtype=int64), array([], dtype=int64))
df_data_0 = df_data_0.replace([np.inf, -np.inf], np.nan)
import pandas as pd
print(lm.intercept_)
coeff_df = pd.DataFrame(lm.coef_,X.columns,columns=['Coefficient'])

```

```
coeff_df
predictions = lm.predict(X_test)

from matplotlib import pyplot as plt
plt.scatter(y_test, predictions)
sns.distplot((y_test-predictions), bins=50);

from sklearn import metrics
print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))
```