PROJECT REPORT

on

PREDICTING LIFE EXPECTENCY

Using

Machine Learning

Project ID:SPS_PRO_215

Project Platform:SMART BRIDGE

BY: G.AMRUTHA

g.amruthа9812@gmail.com

CONTENTS

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. In mathematical terms, life expectancy refers to the expected number of years remaining for an individual at any given age. It is very important to predict average life expectancy of a country to analyze further requirements to increase its rate of growth or stabilize the rate of growth in that country.

So this is a typical Regression Machine Learning project that leverages historical data to predict insights into the future.To predict the life expectancy rate of a particular country, we will be using machine learning to draw inferences from the given data set and give a prediction. We will also be creating a UI using Node-RED for making the model accessible to general users.

1.2)PURPOSE:

Predicting life expectancy plays 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.

*The purpose of the project is to design a model for predicting Life Expectancy rate of a country given various features 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.

2)LITERATURE SURVEY

2.1)EXISTING PROBLEM:

In order to increase the life expectancy of any given country first we need to identify which factors are responsible in determining the life expectancy of a given country. By analyzing the results we can identify those features which play a crucial role, and also we can predict the life expectancy by giving certain inputs like GDP, BMI, Year, Alcohol intake and etc.

*In the context of life expectancy, we are often devoid of ideas of where to concentrate resources on to increase the average life expectancy of a community. this is because in the modern world there are so many other factors other than medical factors that our lifespan depends on like our Lifestyle, our choice of food, etc. Thus it is necessary to find the right areas to pour resources and spent time working on so that the results may improve in our favor.

2.2)PROPOSED SOLUTION:

Machine learning is an application of artificial intelligence

(AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

*Solution for this problem is that we can create a machine learning regression model which can predict the life expectancy of people based on different affecting factors which will help in knowing what to implement for betterment of humans. In order to do so we would take the help of IBM cloud services, Watson studio and Node – red for deploying the model.

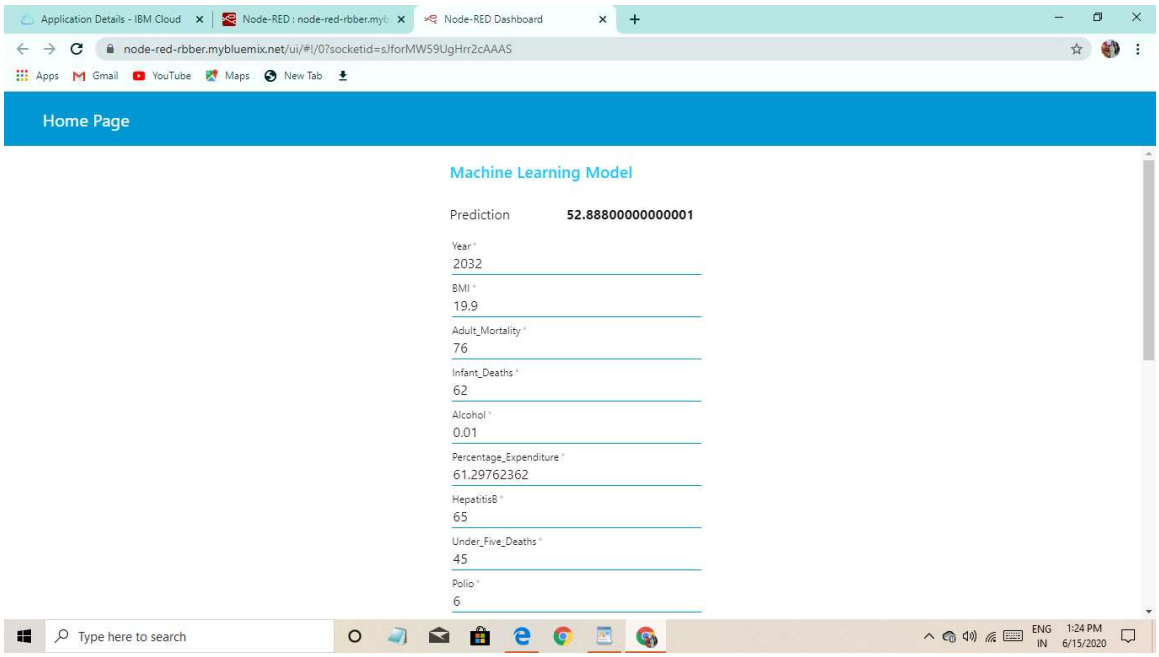
1)Create IBM cloud Services

- Watson Studio
- Machine Learning resource
- Node-Red

2)Configure Watson Studio

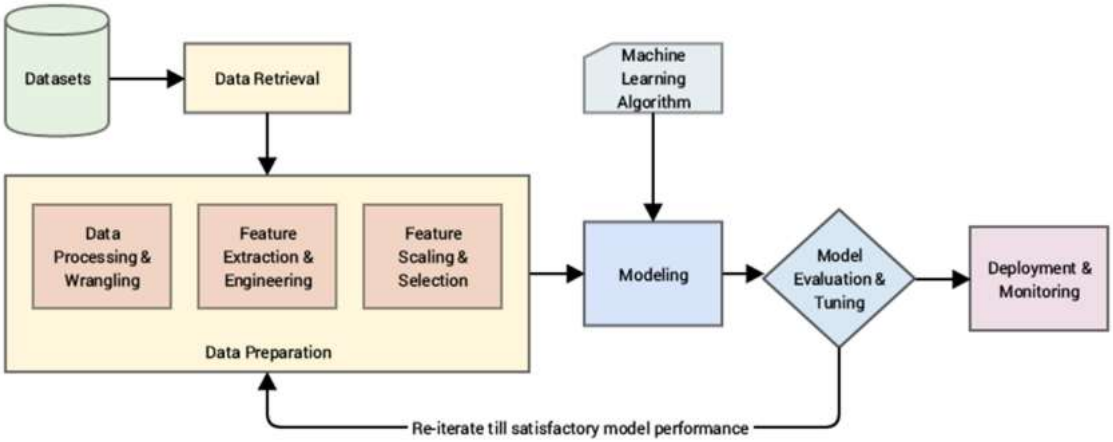
3)Create Node-Red Flow to connect all services together.

Deploy and run Node Red app. Deploy the Node Red flow. Then copy the link url upto .net/ and paste at a new tab by ui at the end of the url like this . An example screenshot is attached below:



3)THEORETICAL ANALYSIS:

3.1)BLOCK DIAGRAM:



The above block diagram summarizes the entire process in a simple manner. The blocks include:

Datasets:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

*The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

Data Preparation:

Data preparation is the act of manipulating (or pre-processing) raw data (which may come from disparate data sources) into a form that can readily and accurately be analysed, e.g. for business purposes.

*Data preparation is the first step in data analytics projects and can include many discrete tasks such as loading data or data ingestion, data fusion, data cleaning, data augmentation, and data delivery.

Modelling:

The process of modeling means training a machine learning algorithm to predict the labels from the features, tuning it for the business need, and validating it on holdout data.

Machine Learning Algorithms:

Machine Learning algorithm is an evolution of the regular algorithm. It makes your programs “smarter”, by allowing them to automatically learn from the data you provide. The

algorithm is mainly divided into:

1)Training Phase

2)Testing phase

Model Evaluation and Tuning:

Methods for evaluating a model's performance are divided into 2 categories: namely, holdout and Cross-validation. Both methods use a test set (i.e data not seen by the model) to evaluate model performance. It's not recommended to use the data we used to build the model to evaluate it. This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set.

Deployement and Monitoring:

Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data. It is one of the last stages in the machine learning life cycle and can be one of the most cumbersome.

3.2) <u>HARDWARE / SOFTWARE DESIGNING :</u>

*IBM Cloud

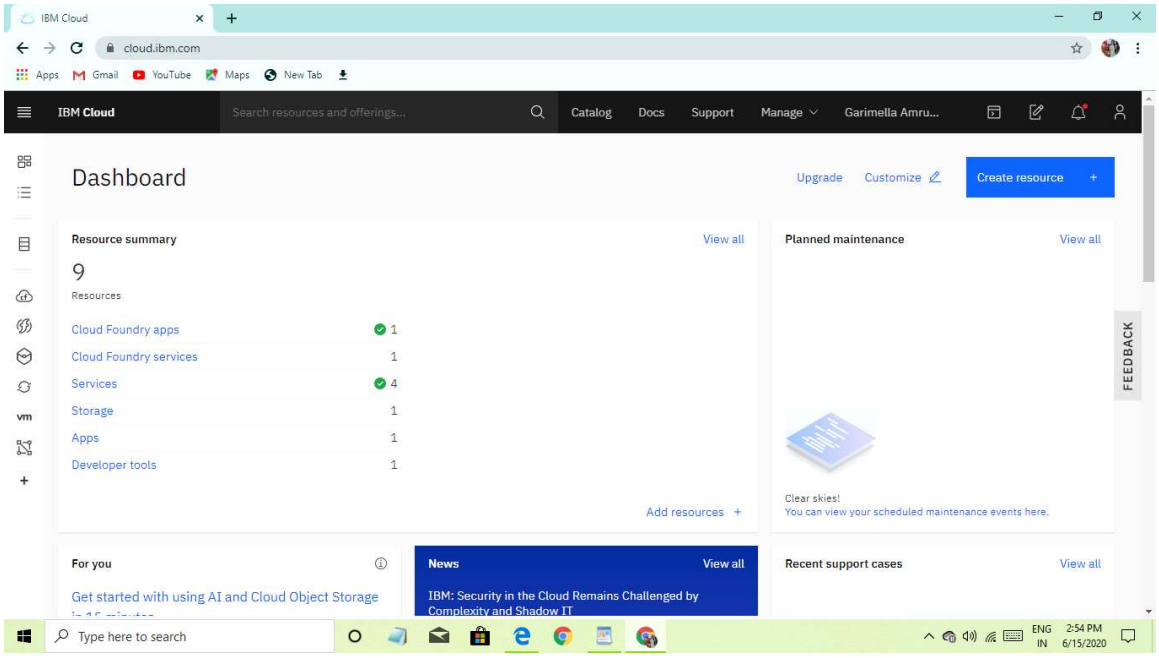
*Node-RED

*Watson Studio

*Watson Machine Learning

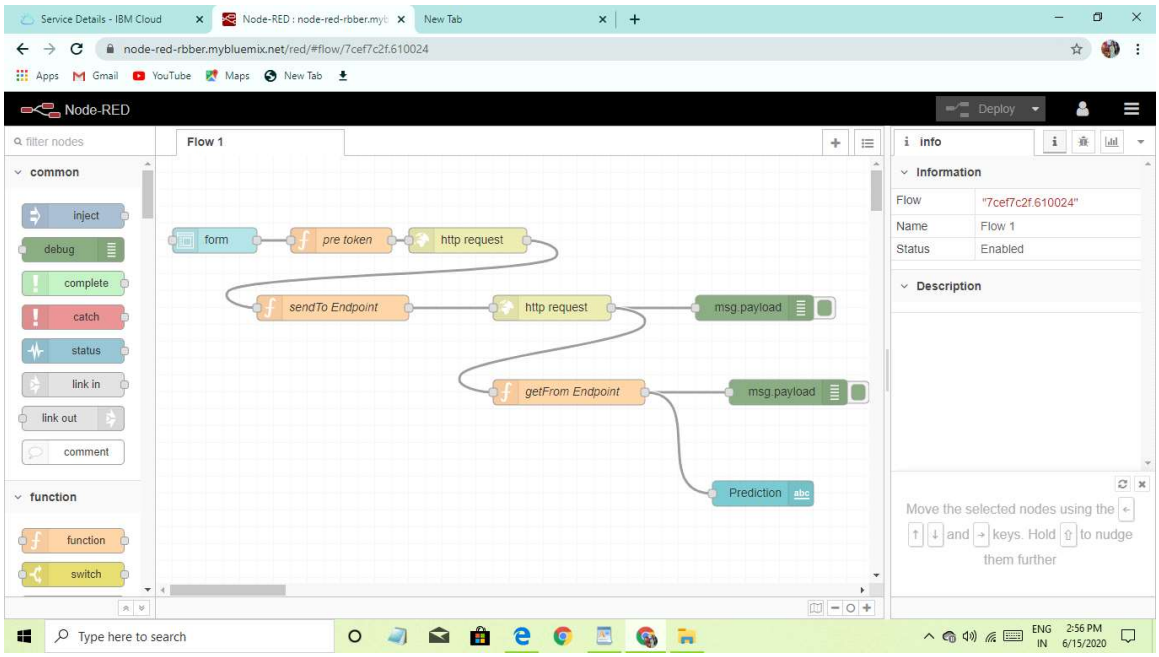
IBM Cloud:

IBM Cloud is a set of cloud computing services for business offered by the information technology company IBM. It combines platform as a service (PaaS) with infrastructure as a service (IaaS). The platform scales and supports both small development teams and organizations, and large enterprise businesses. It is globally deployed across data centers around the world.IBM's main competitors in the cloud computing market include Amazon Web Services, Microsoft Azure and Google Cloud Platform.



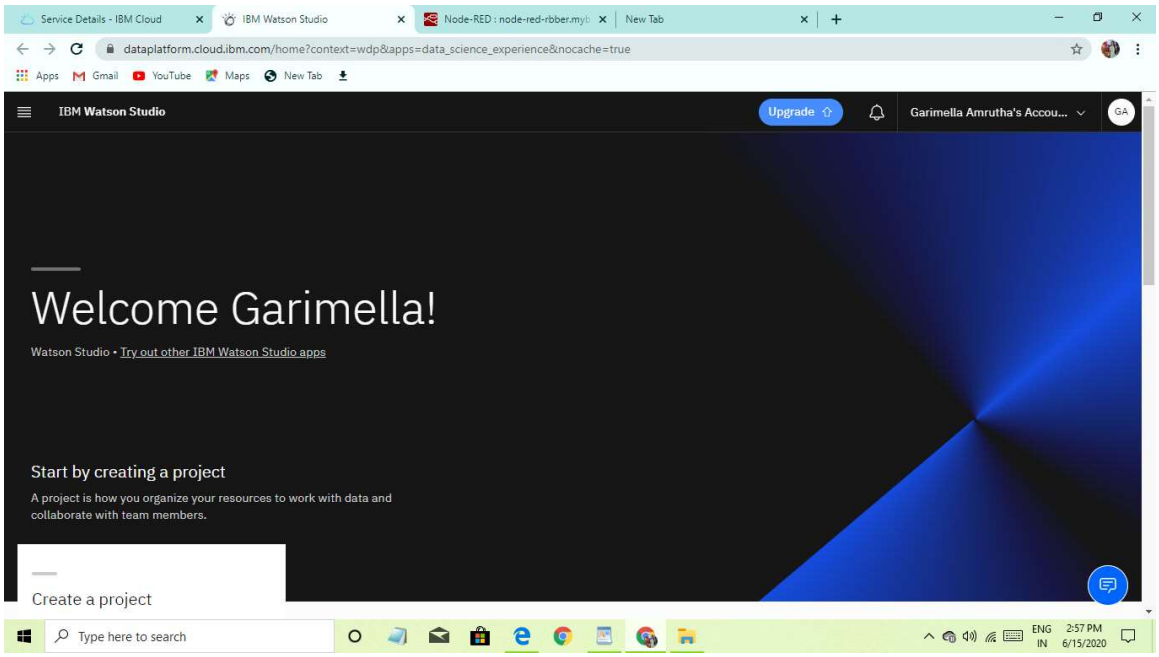
Node-RED:

Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things. Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions.



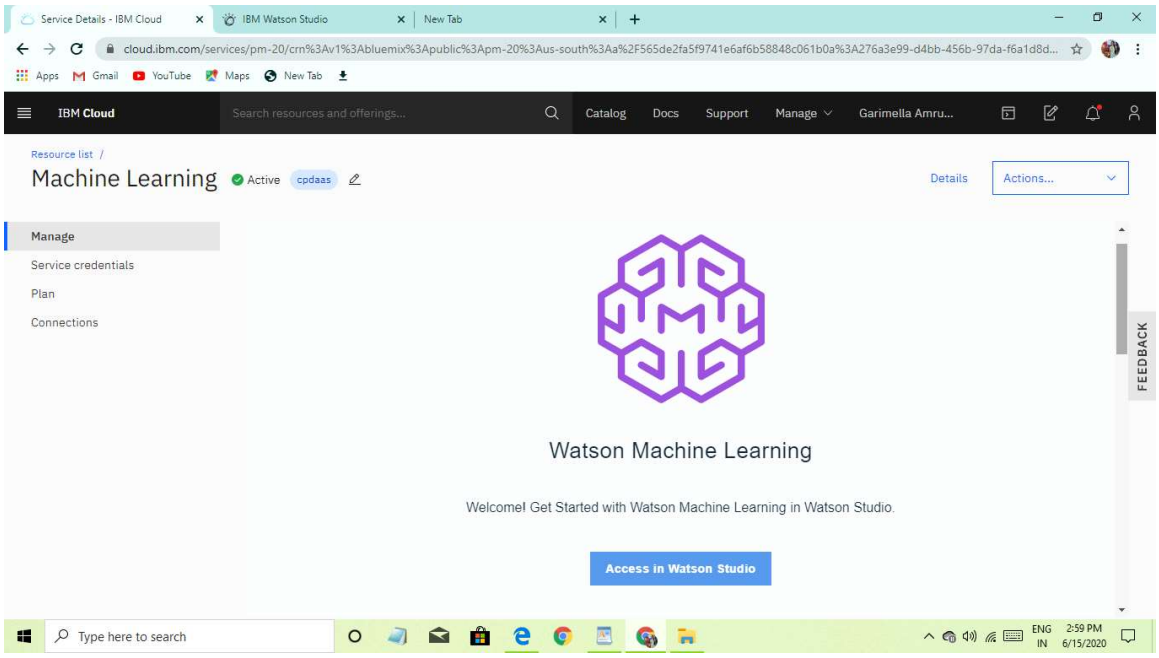
Watson Studio:

IBM Watson® Studio helps data scientists and analysts prepare data and build models at scale across any cloud. With its open, flexible multicloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI.



Watson Machine Learning:

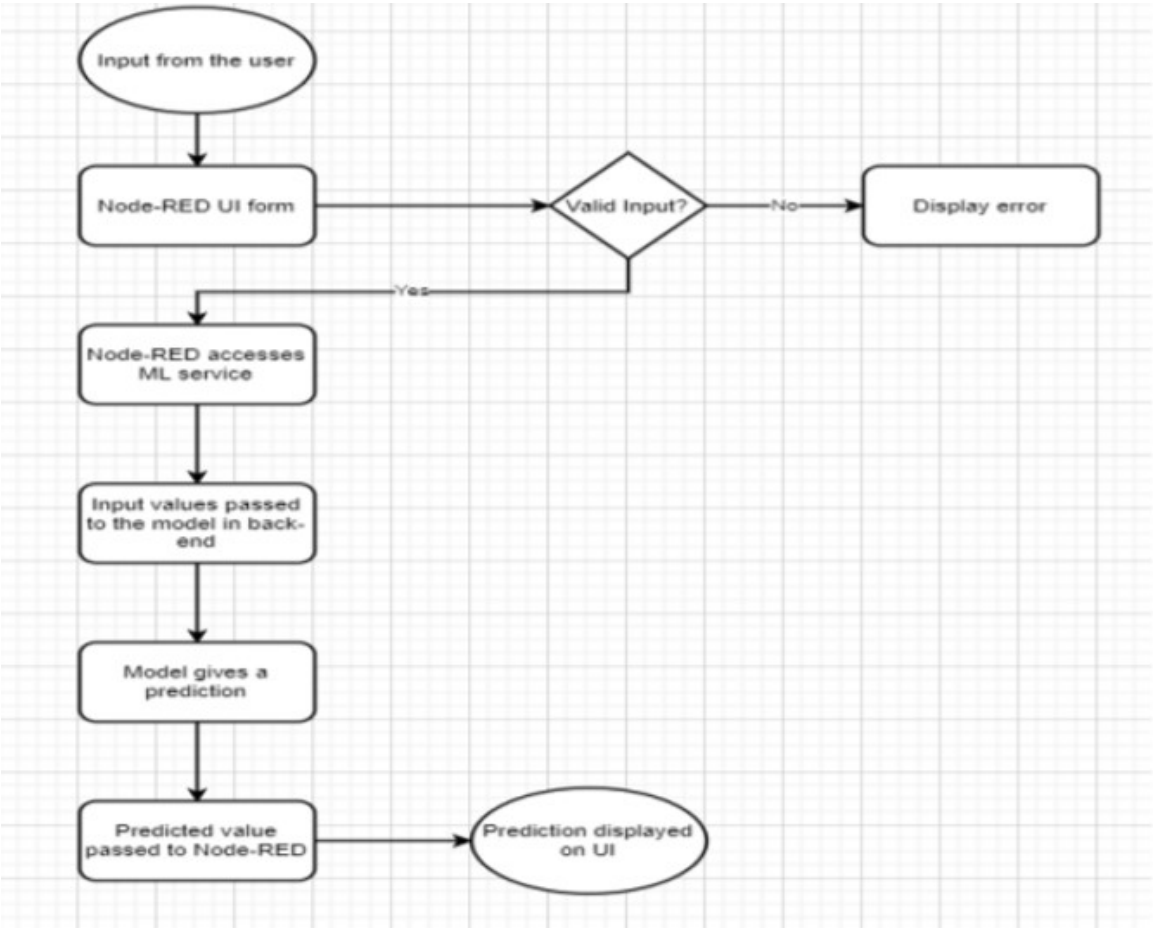
IBM Watson® Machine Learning helps data scientists and developers accelerate AI and machine-learning deployment. With its open, extensible model operation, Watson Machine Learning helps businesses simplify and harness AI at scale across any cloud.



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 2015. There are numerous rows with null values for some of the columns and outliers as well. These null values will be replaced with the mean for the respective columns and the outliers will be adjusted. All the columns except the "Country" and "Status" are of integer type. The country column is not considered into 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.

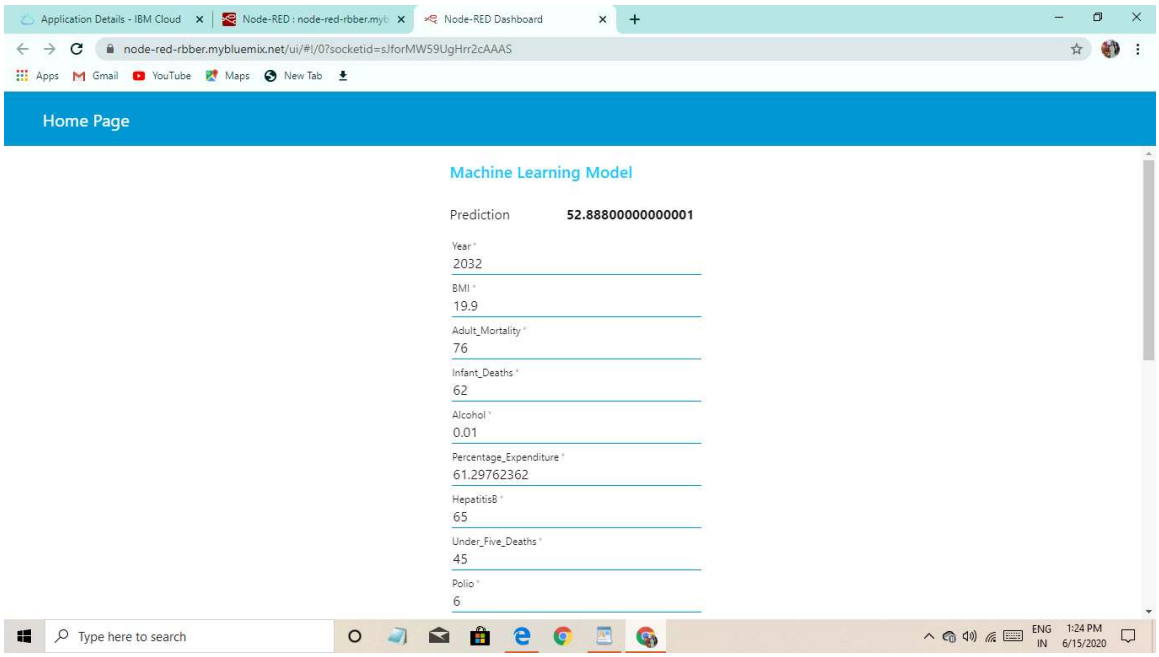
5)FLOW CHART:



6)RESULT:

A machine learning model developed using Linear Regression and IBM Cloud and its services. The model has a high accuracy.

An interactive UI deployed using the Node-RED service for the user to interact with the model in the back-end and get predictions for a given set of input values. This is my User Interface.



7)ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

a) This is a straightforward site page and can be accessed by any resident of a nation to figure life expectancy of their nation and doesn't required any sort of installment. One can directly enter the values into the dashboard.

b) This interface requires no foundation information on the most proficient method to utilize it. It's a straightforward interface and just request required qualities and anticipate the yield.

c) Using PCA we can also determine which factors affect life expectancy most.

d) By utilizing this crucial information the governments of respective countries can make changes in the policies in order

to increase their life expectancy.

DISADVANTAGES:

a) Wrong Prediction: As it depends completely on user, so if user provides some wrong values then it will predict wrong value.

b) Still, the accuracy of the model is not that great. So, sometimes the result fetched may nit be accurate.

8.APPLICATIONS:

1)Useful in the medical field for studying and isolating causes of deviation in human life expectancy over an observed period of time.

2) Supporting timely recognition of the right moment to start Advance Care Planning.

3)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.

4)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 an increase in the “manpower”

of a country. The knowledge asset of a country increases with the number of individuals in a country.

5)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.

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

9)CONCLUSION:

Prognostication of life expectancy is difficult for humans. Our research shows that machine learning and natural language processing techniques offer a feasible and promising approach to predicting life expectancy. The research has potential for real-life applications, such as supporting timely recognition of the right moment to start Advance Care Planning.

10)FUTURE SCOPE:

This research should be considered to be exploratory. In order to replicate and extend this research, we can expand the data set substantially, by collecting additional data of both deceased and active patients. This will allow us to zoom in on specific illness trajectories, and to rephrase the task in such a way that it will match clinical settings more closely, for example by aiming to make predictions about patients

while they are still active.

*We will be able to compare a range of predictive models, alternative patient representations, and (interpretations of) output variables in future work. To provide a better comparison between automatic and human prognostication, we can investigate the prediction accuracy of both the system and general practitioners by presenting them with the same task and test data.

11)BIBLIOGRAPHY:

APPENDIX:

https://cloud.ibm.com

https://www.ibm.com/in-en/cloud/watson-studio

https://nodered.org

NODE-RED Flow:


```
[
 {
"id":"7cef7c2f.610024",
"type":"tab",
"label":"Flow1",
```

"disabled":false,

"info":""

},

{

"id":"a5167841.bce1b8",

"type":"ui_form",

"z":"7cef7c2f.610024",

"name":"",

"label":"",

"group":"bc2b0978.0ce728",

"order":0,

"width":0,

"height":0,

"options":[{"label":"Year","value":"a","type":"number","required":true,"rows":null

},

{

"label":"BMI",

"value":"b",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Adult_Mortality",

"value":"c",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Infant_Deaths",

"value":"d",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Alcohol",

"value":"e",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Percentage_Expenditure",

"value":"f",

"type":"number",

"required":true,

"rows":null

},

{"label":"HepatitisB",

"value":"g",

"type":"number",

"required":true,

"rows":null

},

{"label":"Under_Five_Deaths",

"value":"h",

"type":"number",

"required":true,

"rows":null

},

{"label":"Polio",

"value":"i",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Total_Expenditure",

"value":"j",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Diphtheria",

"value":"k",

"type":"number",

"required":true,

"rows":null

},

{

"label":"HIV",

"value":"l",

"type":"number",

"required":true,

"rows":null

}, {

"label":"GDP",

"value":"m",

"type":"number",

"required":true,

"rows":null

},

{

"label":"Population",

"value":"n",

"type":"number",

"required":true,

"rows":null

},

{

"label":"thinness_1to19_years",

"value":"o",

"type":"number",

"required":true,

"rows":null

},

{

"label":"thinness_5to9_years",

"value":"p",

"type":"number",

"required":true,

"rows":null

```
},
{
"label":"Income_Comp_Of_Resources",
"value":"q",
"type":"number",
"required":true,
"rows":null
},
{
"label":"Schooling",
"value":"r",
"type":"number",
"required":true,
"rows":null
},
{
"label":"Measles",
"value":"s",
"type":"number",
"required":true,
"rows":null
},
{
```

"label":"Developing(1 or 0)",

"value":"t",

"type":"number",

"required":true,

"rows":null

}

],

"formValue":

{

"a":"",

"b":"",

"c":"",

"d":"",

"e":"",

"f":"",

"g":"",

"h":"",

"i":"",

"j":"",

"k":"",

"l":"",

"m":"",

"n":"",

"o":"",

"p":"",

"q":"",

"r":"",

"s":"",

"t":""

},

"payload":"",

"submit":"submit",

"cancel":"cancel",

"topic":"",

"x":70,

"y":100,

"wires":[["6581e5f2.ceb65c"]

]

},

{

"id":"6581e5f2.ceb65c",

"type":"function",

"z":"7cef7c2f.610024",

"name":"pre token",

"func":"//make user given values as global variables\nglobal.set(\"a\",

msg.payload.a);

\nglobal.set(\"b\",msg.payload.b);

\nglobal.set(\"c\",msg.payload.c);

\nglobal.set(\"d\",msg.payload.d);

\nglobal.set(\"e\",msg.payload.e);

\nglobal.set(\"f\",msg.payload.f);

\nglobal.set(\"g\",msg.payload.g);

\nglobal.set(\"h\",msg.payload.h);

\nglobal.set(\"i\",msg.payload.i);

\nglobal.set(\"j\",msg.payload.j);

\nglobal.set(\"k\",msg.payload.k);

\nglobal.set(\"l\",msg.payload.l);

\nglobal.set(\"m\",msg.payload.m);

\nglobal.set(\"n\",msg.payload.n);

\nglobal.set(\"o\",msg.payload.o);

\nglobal.set(\"p\",msg.payload.p);

\nglobal.set(\"q\",msg.payload.q);

\nglobal.set(\"r\",msg.payload.r);

\nglobal.set(\"s\",msg.payload.s);

\nglobal.set(\"t\",msg.payload.t);

\n\n//following are required to receive a token\nvar apikey=\"KlEwVVpWG93A5
_fQHTy4CMWhPYwS26B_CAxdcc9RANUN\";

\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"};
\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey
\":apikey};

\nreturn msg;\n","outputs":1,

"noerr":0,

"x":220,

"y":100,

"wires":[["5c1f2137.b21cd"]

]

},

{

"id":"480a9911.f2bf48",

"type":"http request",

"z":"7cef7c2f.610024",

"name":"",

"method":"POST",

"ret":"obj",

"paytoqs":false,

"url":"https://us-south.ml.cloud.ibm.com/v3/wml_instances/276a3e99-d4bb-456b-97da-f6a1d8d1207d/deployments/a18a759f-f2be-47a6-a73b-839c81332a55/online",

"tls":"",

"persist":false,

"proxy":"",

"authType":"",

"x":470,

"y":180,

"wires":[["5ed77fb.b13e88",

"a39c1161.d47e2"]

]

},

{

"id":"5b2a8ced.e8a324",

"type":"debug",

"z":"7cef7c2f.610024",

"name":"",

"active":true,

"tosidebar":true,

"console":false,

"tostatus":false,

"complete":"payload",

"targetType":"msg",

"x":750,

"y":280,

"wires":[]

},

{

"id":"a39c1161.d47e2",

"type":"function",

"z":"7cef7c2f.610024",

"name":"getFrom Endpoint",

"func":"msg.payload=msg.payload.values[0][0];

\nreturn msg;

","outputs":1,

"noerr":0,

"x":490,

"y":280,

"wires":[["5b2a8ced.e8a324","e555521.6fd29b"]

]

},

{

"id":"5ed77fb.b13e88",

"type":"debug",

"z":"7cef7c2f.610024",

"name":"",

"active":true,

"tosidebar":true,

"console":false,

"tostatus":false,

"complete":"payload",

"targetType":"msg",

"x":710,

"y":180,

"wires":[]

},

{

"id":"c27528c8.45c5a8",

"type":"function",

"z":"7cef7c2f.610024",

"name":"sendTo Endpoint",

"func":"//get token and make headers\nvar token=msg.payload.access_token;

\nvar instance_id=\"276a3e99-d4bb-456b-97da-f6a1d8d1207d\"\nmsg.headers={'Co

ntent-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID
\":instance_id}\n\n//get variables that are set earlier\nvar a = global.get(\"a\");

\nvar b = global.get(\"b\");

\nvar c = global.get(\"c\");

\nvar d = global.get(\"d\");\nvar e = global.get(\"e\");

\nvar f = global.get(\"f\");

\nvar g = global.get(\"g\");

\nvar h = global.get(\"h\");

\nvar i = global.get(\"i\");

\nvar j = global.get(\"j\");

\nvar k = global.get(\"k\");

\nvar l = global.get(\"l\");

\nvar m = global.get(\"m\");

\nvar n = global.get(\"n\");

\nvar o = global.get(\"o\");

\nvar p = global.get(\"p\");

\nvar q = global.get(\"q\");

\nvar r = global.get(\"r\");

\nvar s = global.get(\"s\");

\nvar t = global.get(\"t\");

\n\n\n//send the user values to service endpoint\nmsg.payload = \n{\"fields\":[\"Year \",\"BMI\",

\n\"Adult Mortality\",

\n\"infant deaths\",

\n\"Alcohol\",

\"percentage expenditure\",

\"Hepatitis B\",

\"under-five deaths \",

\"Polio\",

\"Total expenditure\",

\"Diphtheria \",

\" HIV/AIDS\",

\"GDP\",

\"Population\",

\" thinness 1-19 years\",

\" thinness 5-9 years\",

\"Income composition of resources\",

\"Schooling\",

\"Measles\",

\"Developing\"],

\n\"values\":[[a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t]]};\n\nreturn msg;\n",

"outputs":1,"noerr":0,"x":210,"y":180,"wires":[["480a9911.f2bf48"]

]

},

{

"id":"5c1f2137.b21cd",

"type":"http request",

"z":"7cef7c2f.610024",

"name":"",

"method":"POST",

"ret":"obj",

"paytoqs":false,

"url":"https://iam.cloud.ibm.com/identity/token",

"tls":"",

"persist":false,

"proxy":"",

"authType":"basic",

"x":370,

"y":100,

"wires":[["c27528c8.45c5a8"]

```
]

},

{

"id":"e555521.6fd29b",

"type":"ui_text",

"z":"7cef7c2f.610024",

"group":"bc2b0978.0ce728",

"order":1,

"width":0,

"height":0,

"name":"",

"label":"Prediction",

"format":"{{msg.payload}}",

"layout":"row-spread",

"x":720,

"y":400,

"wires":[]

},

{

"id":"bc2b0978.0ce728",

"type":"ui_group",

"z":"",

"name":"Machine Learning Model",
```

"tab":"398dcff3.5a27f",

"order":1,

"disp":true,

"width":"6",

"collapse":false

},

{

"id":"398dcff3.5a27f",

"type":"ui_tab",

"z":"",

"name":"Home Page",

"icon":"dashboard",

"disabled":false,

"hidden":false

}

]

My Github Repository Link:

https://github.com/SmartPracticeschool/llSPS-INT-2033-Predicting-Life-Expectancy-using-Machine-Learning