PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

MACHINE LEARNING PROJECT REPORT

Thushar.V.Nikam

1. INTRODUCTION

1.1 Overview:

With the emergence of technologies such as electronic health and mobile health,(cloud computing, big data, and the Internet of Things (IoT), health related data are increasing and many applications such as smartphone apps and wearable devices that provide wellness and fitness tracking are entering the market. 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. These requirements can be extended to provide a personalized life expectancy (PLE) for the purpose of wellbeing and encouraging lifestyle improvement. No existing works provide this PLE information that is developed and customized for the individual. This project is based on the concurrent models and methodologies to calculate and predict life expectancy (LE) and proposes an idea of using multi-phased approaches to the solution as the project requires an immense and broad range of work to accomplish. As a result, the current prediction of LE, which was found to be up to a maximum of five years could potentially be extended to a lifetime prediction by utilizing generic health data. In this project, the novel idea of the solution proposing a PLE on an individual basis, which can be extended to lifetime is presented in addion to the existing works

1.2 Purpose

The main purpose of this project is Predicting Life Expectancy using Machine Learning deliver the outcome for the given dataset of any county. It will predict the life expectancy based on the 'Country','Year', 'Status' 'Life expectancy ' 'Adult Mortality' 'infant deaths' 'Alcohol' 'percentage expenditure' 'Hepatits B' 'Measles ' ' BMI ' 'under-five deaths ' 'Polio' 'Total expenditure' 'Diphtheria ' ' HIV/AIDS' 'GDP' 'Populaton' ' thinness 1-19 years' ' thinness 5-9 years' 'Income composition of resources' 'Schooling'

2. LITERATURE SURVEY:

2.1 Existing System:

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. For example, a clinical research center claims that in 10 to 12 years from now, for every year that humans live, science is extending the life for more than a year using health intelligence platform integrating genomics, advanced clinical imaging and robust machine learning in a spa-like setting. Predicting a human's life expectancy has been a long-term question to human kind , and there have been many attempts to make the prediction accurate and popular since the prevalence of smartphones and apps. However, the effectiveness of those apps is limited due to the constraint of developing a classification of meta-data, such as the complexity and variety of environmental, geographic, genetic, and living factors of humans. For example, a report showed that people living in a village called Yuzurihara in Japan, also known as "the village of long life", were ten times more likely to live beyond the age of 85 than anywhere in North America. These people also had similar traits such as smooth skin, flexible joints and thick hair . This implies that geographic and living environments affect the longevity of human life, and the use of statistics can make it possible to forecast a life expectancy of a person who lives in a similar environment village with a similarlifestyle.

2.2 Proposed Solutions:

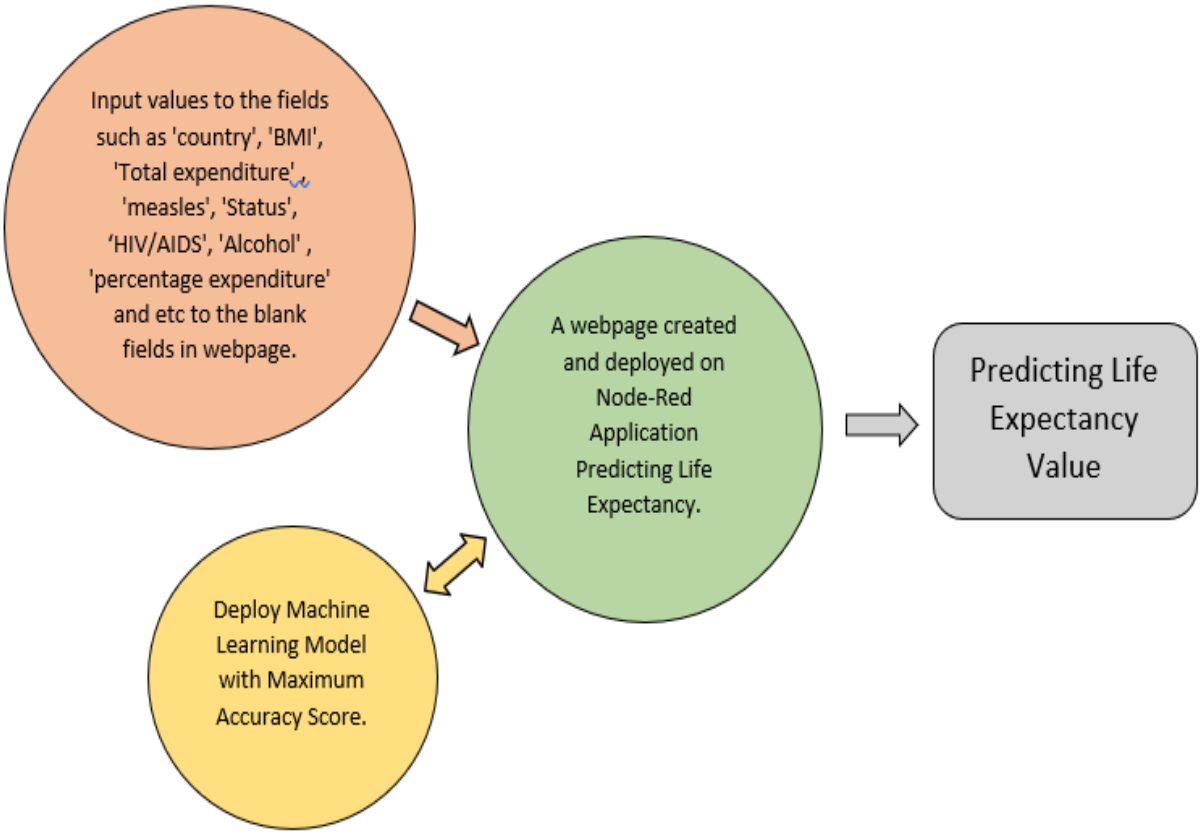
Following are the features of the data :'Country','Year', 'Status' 'Life expectancy ' 'Adult Mortality' 'infant deaths' 'Alcohol' 'percentage expenditure' 'Hepatitis B' 'Measles ' ' BMI ' 'under-five deaths ' 'Polio' 'Total expenditure' 'Diphtheria ' 'HIV/AIDS' 'GDP' 'Population' ' thinness 1-19 years'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

1. These are country level average.
2. There is no distinction between male and female

Proposed system is a machine learning training model with 95% accuracy

3. THEORITICAL ANALYSIS:

3.1 Block Diagram:



3.2 Hardware/Software Design:

a. Functional Requirements

To be able to predict the life expectancy accurately using Machine Learning models.

b. Hardware Requirements:

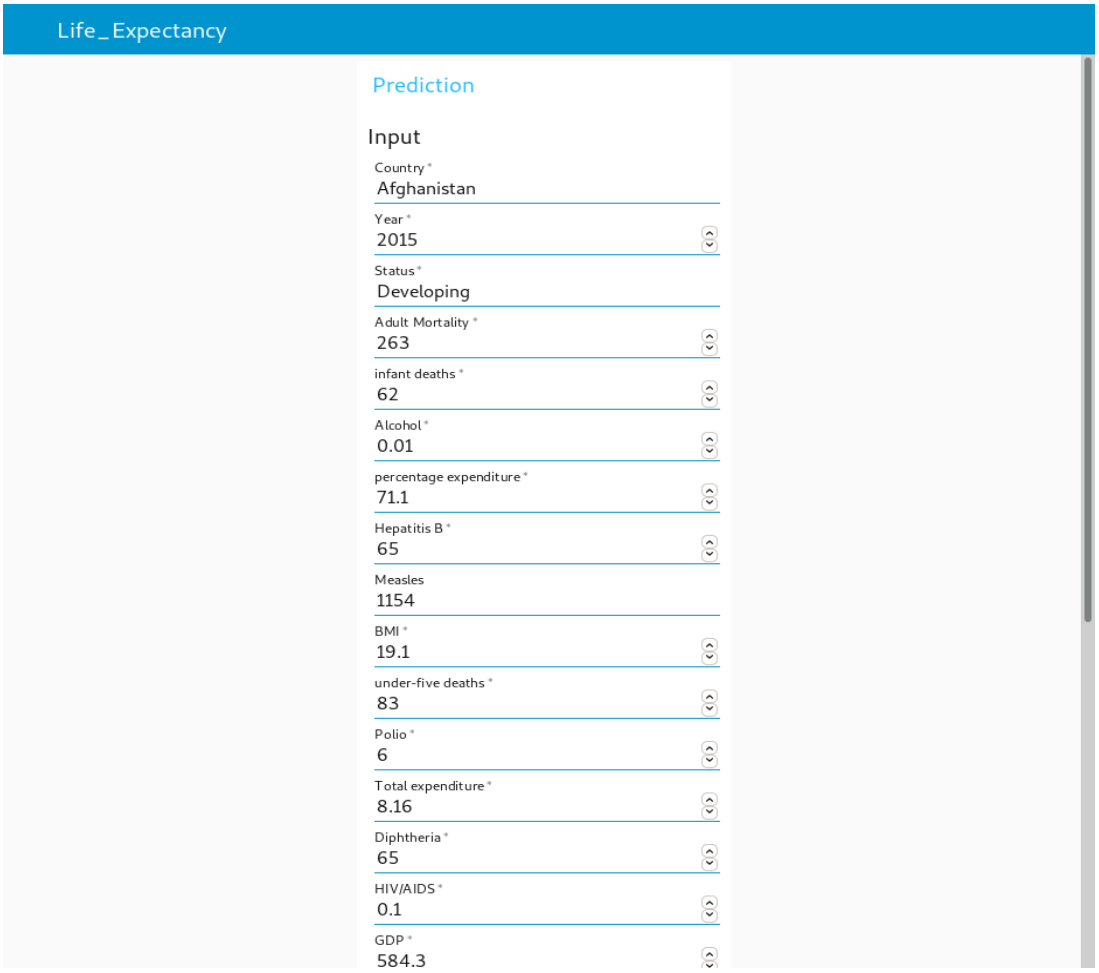
i. Processor - i3 or higher
ii. Hard disk-Min 20GB or more

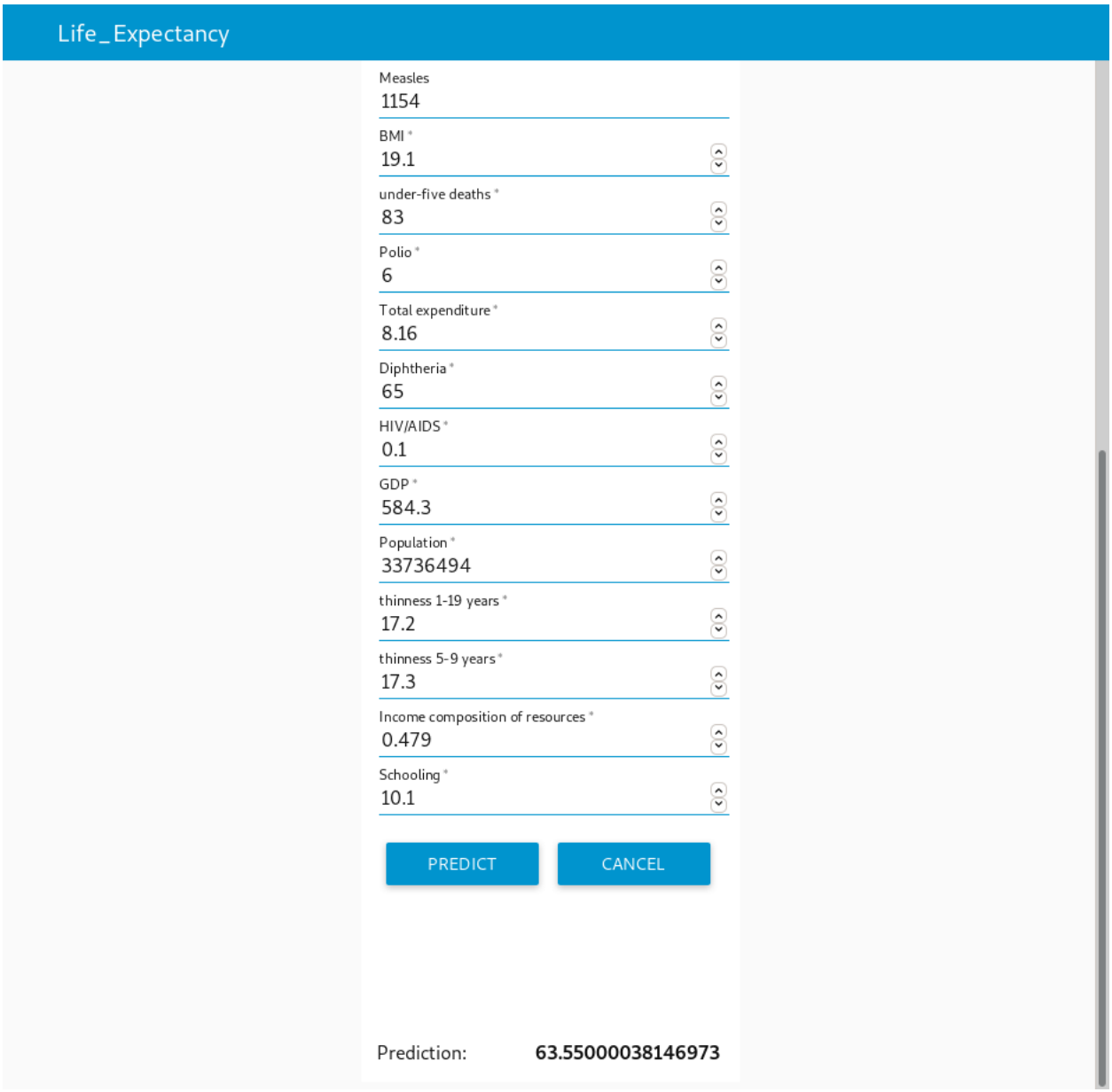
c. Software Requirements:

i. Python(pandas,numpy,matplotlib,seaborn,sklearn)
ii. Machine Learning Algorithms
iii. Jupyter Notebook Enviroment
iv. watson studio
v. Html
vi. Node-Red

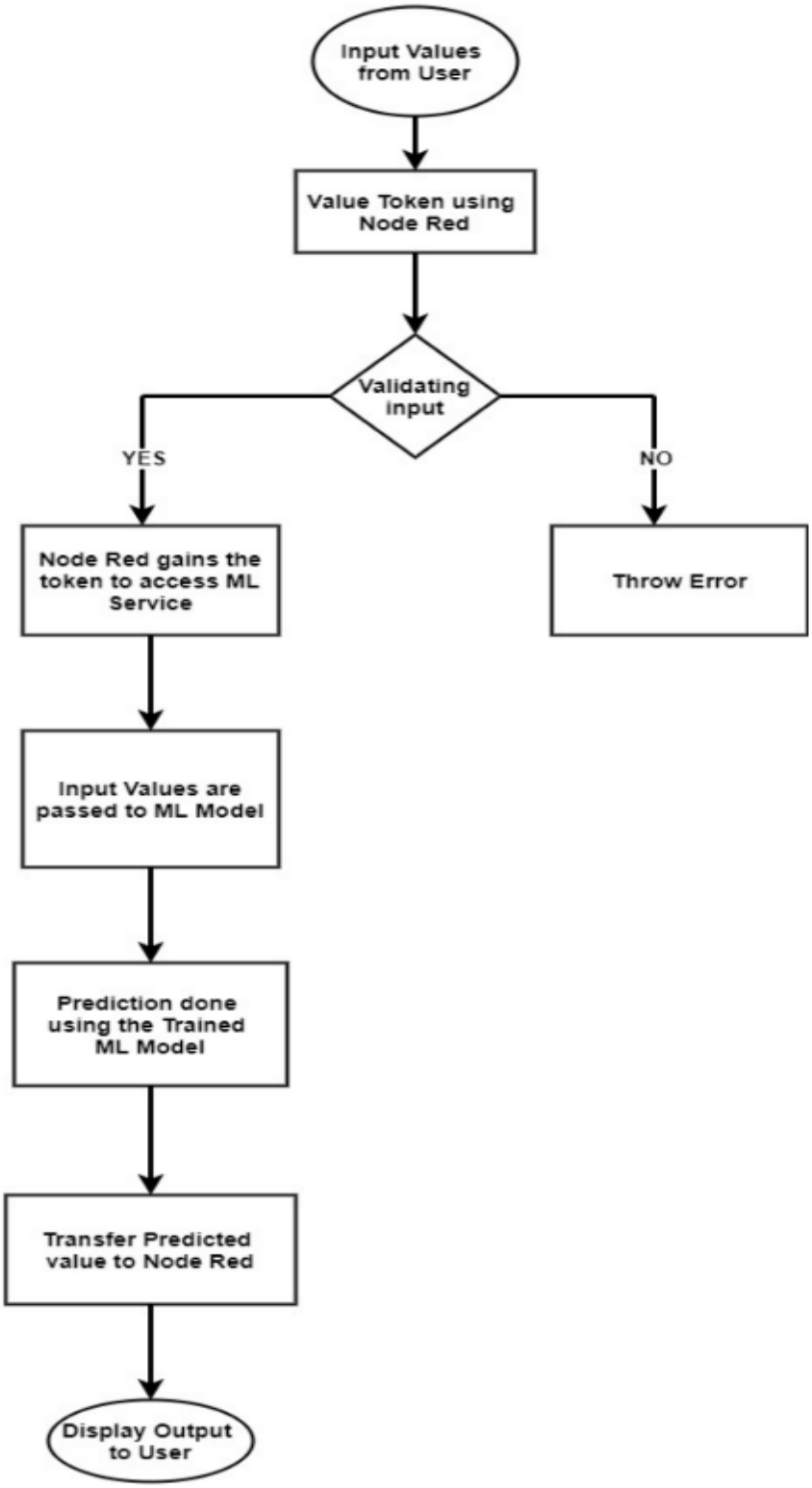
4. EXPERIMENTAL INVESTIGATIONS:

Some random inputs are given to the machine learning model.We got the following Output




Life_Expectancy
Measles
1154
BMI *
19.1
under-five deaths *
83
Polio *
6
Total expenditure *
8.16
Diphtheria *
65
HIV/AIDS *
0.1
GDP *
584.3
Population *
33736494
thinness 1-19 years *
17.2
thinness 5-9 years *
17.3
Income composition of resources *
0.479
Schooling *
10.1
PREDICT
CANCEL
Prediction:
63.55000038146973

5. FLOWCHART:


Input Values from User
Value Token using Node Red
Validating input
YES
NO
Node Red gains the token to access ML Service
Throw Error
Input Values are passed to ML Model
Prediction done using the Trained ML Model
Transfer Predicted value to Node Red
Display Output to User

6. RESULT:

Based on the given data, the machine learning model understands the data and cross reference the data to watch what are the factors that are affecting the results we require i.e life expectancy. Then when we give any input, it has already run algorithm to get the output based on previously given data.So the results we get are approximations, they are not definitely true, but it works in maximum number of cases, except for some exceptional ones.

7. ADVANTAGES AND DISADVANTAGES:

7.1 ADVANTAGES:

1. Health Inequalities: Life ex pectancy has been 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.

7.2 DISADVANTAGES:

1. Wrong Prediction: As it depends completely on user, so if user provides some wrong values then it will predict wrong value.
2. Average Prediction: The model predicts average or approximate value with 97.07% accuracy but not accurate value.

8. APPLICATIONS:

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

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

9. CONCLUSION:

In this project we develop a model that predicts a life expectancy by using some factors on which life expectancy depends on.To develop this project we use varius IBM cloud services such as IBM Watson,IBM Machine Learning service ,Node Red application

10. FUTURE SCOPE:

The problem of processing datasets such as electronic medical records(EMR) and their integration with genomics, environmental factors, socioeconomic factor and patient behavior variations have posed a problem for researchers the health industry. Due to rapid innovations in machine learning field such as big data, analytics, visualization, deep learning, health workers now have improved way of processing, and developing meaningful information from huge datasets that have been accumulated over many years . Big data and machine learning can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy. We can use demographics of selected regional areas and multiple behavioral health disorders across regions to find correlation between individual behavior indicators and behavioral health outcomes.

11. BIBILOGRAPHY:

- https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://nodered.org/
- https://www.kaggle.com/kumarajarshi/life-expectancy-who

12. APPENDIX

12.1 Souce Code

- Node Red Application:

[{"id":"70e49228.93aec4","type":"tab","label":"Flow 1","disabled":false,"info":""},{"id":"e61a1070.c29888","type":"ui_form","z":"70e49228.93aec4","name":"","label":"Input","group":"197538ef.773357","order":1,"width":0,"height":0,"options":[{"label":"Country","value":"a","type":"text","required":true,"rows":null},{"label":"Year","value":"b","type":"number","required":true,"rows":null},{"label":"Status","value":"c","type":"text","required":true,"rows":null},{"label":"Adult Mortality","value":"d","type":"number","required":true,"rows":null},{"label":"infant deaths","value":"e","type":"number","required":true,"rows":null},{"label":"Alcohol","value":"f","type":"number","required":true,"rows":null},{"label":"percentage expenditure","value":"g","type":"number","required":true,"rows":null},{"label":"Hepatitis B","value":"h","type":"number","required":true,"rows":null},{"label":"Measles ","value":"i","type":"text","required":false,"rows":null},{"label":"BMI","value":"j","type":"number","required":true,"rows":null},{"label":"under-five deaths ","value":"k","type":"number","required":true,"rows":null},{"label":"Polio","value":"l","type":"number","required":true,"rows":null},{"label":"Total

expenditure","value":"m","type":"number","required":true,"rows":null},{"label":"
Diphtheria","value":"n","type":"number","required":true,"rows":null},{"label":"
HIV/AIDS","value":"o","type":"number","required":true,"rows":null},{"label":"GD
P","value":"p","type":"number","required":true,"rows":null},{"label":"Population"
,"value":"q","type":"number","required":true,"rows":null},{"label":" thinness 1-19
years","value":"r","type":"number","required":true,"rows":null},{"label":"
thinness 5-9
years","value":"s","type":"number","required":true,"rows":null},{"label":"Income
composition of
resources","value":"t","type":"number","required":true,"rows":null},{"label":"Sch
ooling","value":"u","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":"","u":""},"payload":"","submit":"Predict","cancel":"
cancel","topic":"","x":390,"y":440,"wires":[["47941fe7.798688"]]},{"id":"47941fe
7.798688","type":"function","z":"70e49228.93aec4","name":"pre
token","func":"//make user given values as global
variables\nglobal.set(\"a\",msg.payload.a);\nglobal.set(\"b\",msg.payload.b);\n
global.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.payl
oad.g);\nglobal.set(\"h\",msg.payload.h);\nglobal.set(\"i\",msg.payload.i);\nglob
al.set(\"j\",msg.payload.j);\nglobal.set(\"k\",msg.payload.k);\nglobal.set(\"l\",ms
g.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);\nglobal.set(\"u\",msg.payload.u);\
n\n//following are required to receive a token\nvar
apikey=\"sVNE7O7NxkaFOG2DA2M1FzTo16-oOFsYkkT8jH9rkvH0\";\nmsg.h
eaders={\"content-type\":\"application/x-www-form-urlencoded\"};\nmsg.paylo
ad={\"grant_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apike
y};\nreturn
msg;\n","outputs":1,"noerr":0,"x":605,"y":460,"wires":[["7c7eac43.e1041c"]]},{"i
d":"99894733.2a986","type":"http
request","z":"70e49228.93aec4","name":"","method":"POST","ret":"obj","payto
qs":false,"url":"https://us-south.ml.cloud.ibm.com/v4/deployments/8ad04919-7
22e-4571-9d07-cf380bf50740/predictions","tls":"","persist":false,"proxy":"","aut
hType":"basic","x":850,"y":540,"wires":[["c8dafcd9.a86ff","d3668e2a.b21988"]]
},{"id":"4fe34572.075a5c","type":"debug","z":"70e49228.93aec4","name":"","ac

tive":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload"
,"targetType":"msg","x":1130,"y":640,"wires":[]},{"id":"d3668e2a.b21988","type"
:"function","z":"70e49228.93aec4","name":"","func":"msg.payload=msg.payloa
d.predictions[0].values[0][0]\nreturn
msg;","outputs":1,"noerr":0,"x":830,"y":640,"wires":[["4fe34572.075a5c","ce04
9dd1.d02658"]]},{"id":"c8dafcd9.a86ff","type":"debug","z":"70e49228.93aec4","
name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complet
e":"payload","targetType":"msg","x":1090,"y":540,"wires":[]},{"id":"ce957fc6.ec
d72","type":"function","z":"70e49228.93aec4","name":"PrePrediction","func":"//
get token and make headers\nvar token=msg.payload.access_token;\nvar
instance_id=\"903d3708-097a-40b8-b43e-33a239065561\"\nmsg.headers={'C
ontent-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\");\nvar u =
global.get(\"u\");\n\n//send the user values to service endpoint\nmsg.payload
={\"input_data\":[{\"fields\":[\"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\"],\"values\":[[a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u]]}]};\
nreturn
msg;\n","outputs":1,"noerr":0,"x":580,"y":540,"wires":[["99894733.2a986"]]},{"id
":"7c7eac43.e1041c","type":"http
request","z":"70e49228.93aec4","name":"","method":"POST","ret":"obj","payto
qs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"
proxy":"","authType":"basic","x":750,"y":460,"wires":[["ce957fc6.ecd72"]]},{"id":
"ce049dd1.d02658","type":"ui_text","z":"70e49228.93aec4","group":"197538ef
.773357","order":2,"width":0,"height":0,"name":"","label":"Prediction:","format":
"{{msg.payload}}","layout":"row-spread","x":1010,"y":720,"wires":[]},{"id":"1975
38ef.773357","type":"ui_group","z":"","name":"Prediction","tab":"793f73c6.e62

58c","order":1,"disp":true,"width":"6","collapse":false},{"id":"793f73c6.e6258c",
"type":"ui_tab","z":"","name":"Life_Expectancy","icon":"","disabled":false,"hidd
en":false}]