PROJECT REPORT

Predicting Life Expectancy using Machine Learning

## Internship Title : Predicting Life Expectancy using Machine Learning - SB22623

## Project ID : SPS\_PRO\_215

## Project Title : Predicting Life Expectancy using Machine Learning

## Duration : 23.5 Days

**Vaibhav Chaudhary**

B.Tech.

2nd Year

Computer Science And Engineering

### 

### Abstract

Life expectancy is a statistical measure of the average time a human being is expected to live and this depends on various factors.Predicting life expectancy is one of the most used summary indicators for the overall health of a population.By predicting life expectancy we determine risk factors and future developments. Life expectancy is based on several factors like adult morality , alcohol etc. This project is based on the models and methodologies to calculate and predict life expectancy. As a result, the current prediction of life expectancy could potentially be extended to a lifetime prediction by utilizing generic health data.In this project i am going to develop a machine learning model using IBM services (watson studio and node-red) and WHO dataset.Methods i approached the task of predicting life expectancy as a supervised machine learning task. I trained a machine learning model to predict life expectancy with the help of Random Forest Regression algorithm.Outcome of this project is a dashboard where we can predict life expectancy.

# INTRODUCTION

1.1 Overview - A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features. Life expectancy is a statistical measure of the average time a human being is expected to live, 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.

1.2 Purpose - In this project i am going to predict average life expectancy of people living in a country when various factors 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.In this project we also learn what is the impact of alcohol etc. in our life expectancy.The aim of this project is provide a simple interface by this we can easily predict the life expectancy.

# LITERATURE SURVEY

2.1 Existing problem - Life expectancy is a statistical measure of the average time a human being is expected to live and this depends on various factors. There is a problem if we don’t know the life expectancy and it depends on which factors determine how we increase the life expectancy rate and there is a problem of accuracy of life prediction and there is no simple user interface by which anyone can predict life expectancy.

2.2 Proposed solution - In this Project i developed a model using IBM services (watson studio and node-red) and WHO dataset.Methods i approached the task of predicting life expectancy as a supervised machine learning task. I trained a machine learning model to predict life expectancy with the help of Random Forest Regression algorithm and provide a simple user interface(dashboard) by which anyone can easily predict the life expectancy.

# THEORETICAL ANALYSIS

Life expectancy is a regression problem for predicting life expectancy Random Forest Regression model is used.

Predictive Model-Making use of past data and other attributes,Predicting the future using this data.

Stages of predictive modeling

* Problem Definition
* Hypothesis Generation
* Data Extraction / Collection
* Data Exploration and Transformation
* Predictive Modeling
* Model Deployment / Implementation

3.1 Block diagram - This is the block diagram of workflow for building and deploying the model.



3.2 Hardware / Software designing - For designing model i collected the dataset from the given reference link after this import the libraries and read the dataset using pandas after this apply data cleaning technique and make dataset clean.Then for model building split dataset into two parts train and test using tain\_test\_split after this fit and train the data using random forest regression method and build a model for testing and prediction. After building the model , deploy the model using Watson machine learning service. After deploying the model , connecting this model to Node-red for the user interface and deploying the flow , I built a dashboard for easy life expectancy prediction.

Node-Red Flow -

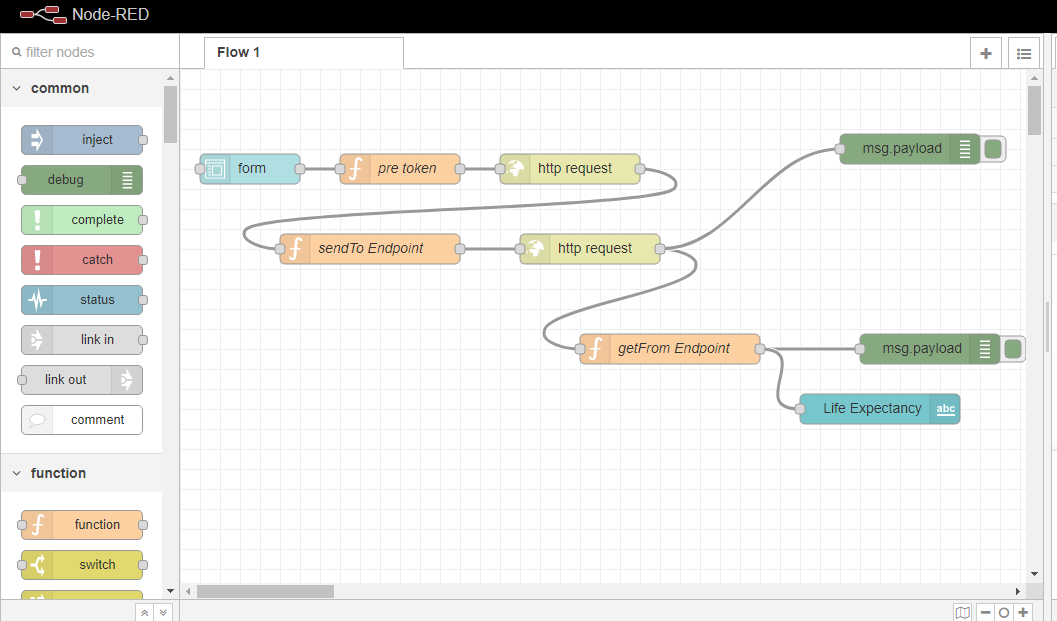
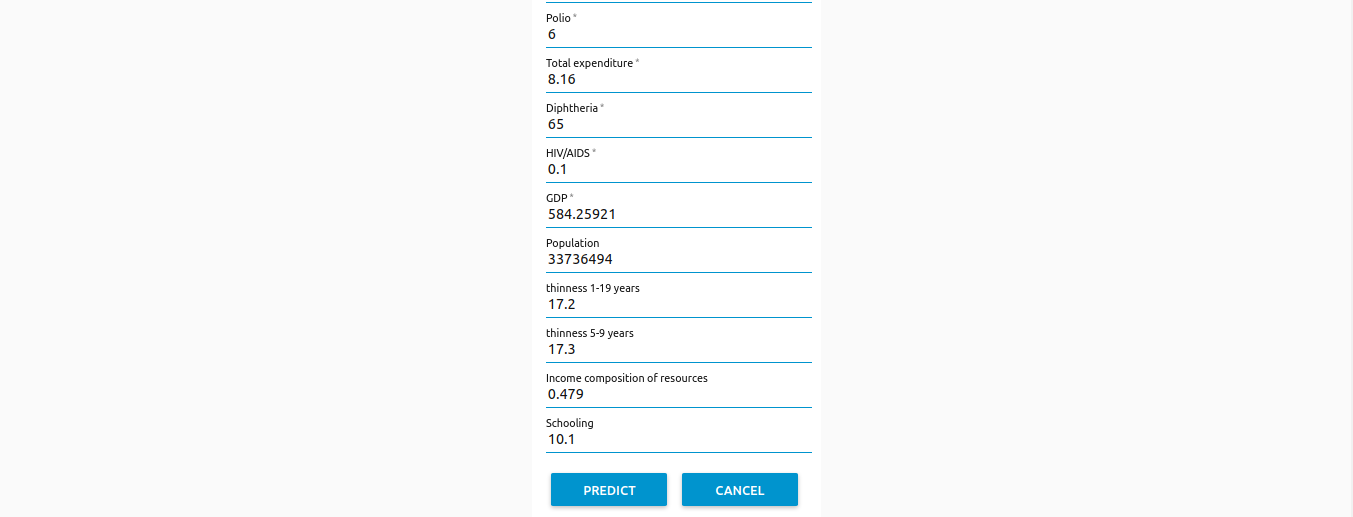
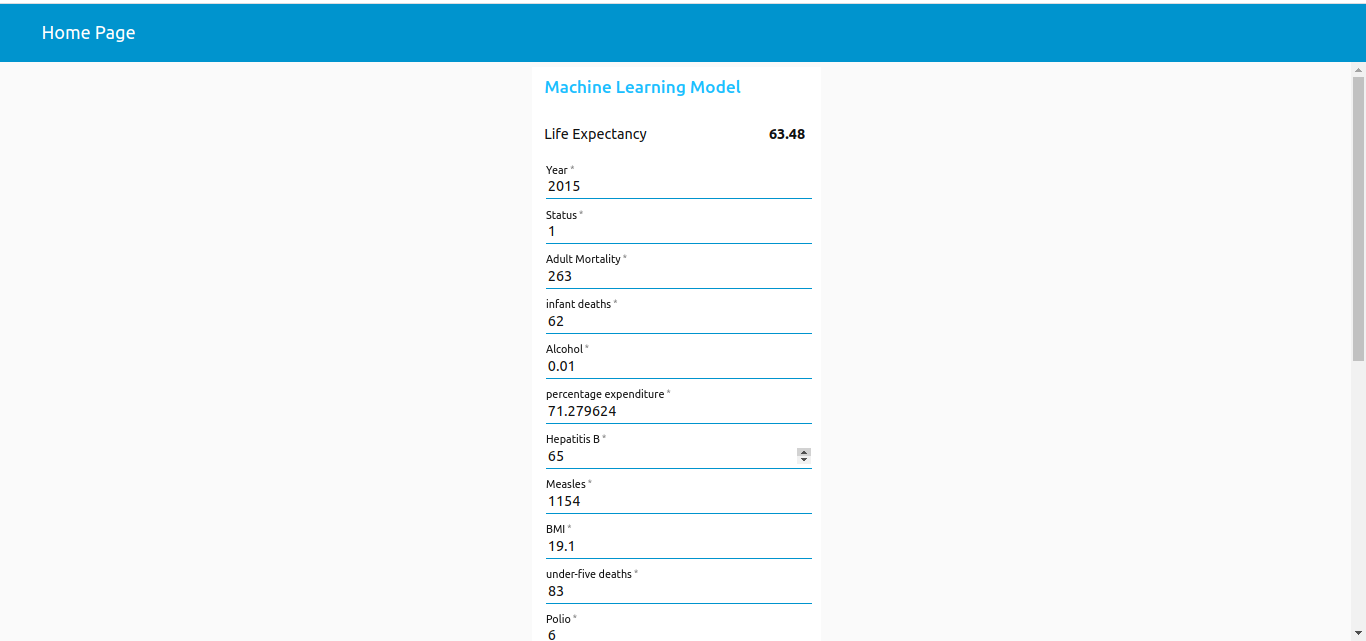


Fig - Node-Red FlowFig- Node-Red Dashboard

# EXPERIMENTAL INVESTIGATIONS

# In this part we are going to investigate our model and look at our model performance. For any country having values as following ‘Year’: 2015, ‘Status’: 1, ‘Adult\_Mortality’: 263, ‘infant\_deaths’: 62, ‘Alcohol’: 0.01, ‘percentage\_expenditure’: 71.27963362, ‘Hepatitis\_B’: 65, ‘Measles’: 1154, ‘BMI’: 19.1, ‘under\_five\_deaths’: 83, ‘Polio’: 6, ‘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’ gives us the result of Life Expectancy value as **63.48**.

# The original value of Life Expectancy was 65 in 2015.

# FLOWCHART

This is the workflow of the process and services used for building and deploying the model.



# RESULT

# Final output of this project is a machine learning model by which we can predict life expectancy.

Model score : 0.8916400509418365

Accuracy : 89.16%

# MSE(mean\_squared\_error) : 8.096138983050842

# RMSE (root\_mean\_squared\_error) : 2.8453715017640215

# MAE(mean\_absolute\_error) : 2.051138014527844.

# ADVANTAGES & DISADVANTAGES

# 7.1. Advantage

# 7.1.1 By this model we can predict life expectancy.

# 7.1.2 By this we can analyse how Infant and Adult mortality rate affected life expectancy.

# 7.1.3 By this we can find Life Expectancy has positive or negative correlation with alcohol.

# 7.1.4 The impact of schooling on life expectancy.

# 2. Disadvantage

Life expectancy measures do not include non-fatal health/disability.

# APPLICATIONS

The applications of life expectancy are

8.1 Health Care analysis

8.2 Economy analysis

8.3 immunization factors analysis

8.4 social factors analysis

# CONCLUSION

Finally we are able to predict life expectancy using machine learning and check what factors affect life expectancy. 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.Outcome of this project is a dashboard which is easy to use and by this anyone can predict life expectancy.

# FUTURE SCOPE

This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

# BIBLIOGRAPHY

References

11.1 IBM services

11.1.1 DashBoard - <https://cloud.ibm.com/>

11.1.2 Resources - <https://cloud.ibm.com/resources>

11.1.3 Watson Studio - <https://cloud.ibm.com/services/data-science-experience/crn%3Av1%3Abluemix%3Apublic%3Adata-science-experience%3Aeu-gb%3Aa%2F122c5402e42945b4851be75e266abe3a%3A48e73ece-1405-414d-bee3-120421f9ba99%3A%3A?paneId=manage>

11.1.4 Node-Red- <https://node-red-dqwgr.eu-gb.mybluemix.net/red/#flow/343e3cb4.555874>

11.1.5 Node-Red Dashboard - <https://node-red-dqwgr.eu-gb.mybluemix.net/ui/#!/0?socketid=_7zSab04BpIdCOX5AAAj>

11.1.6 Watson Machine Learning - <https://cloud.ibm.com/services/pm-20/crn%3Av1%3Abluemix%3Apublic%3Apm-20%3Aeu-gb%3Aa%2F122c5402e42945b4851be75e266abe3a%3A9a65746c-b0ae-47e9-9856-b1a8764d5bf1%3A%3A?paneId=manage>

11.1.7 Notebook - https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/f6344e28-1fdc-4011-9d56-c021d48d2fbd/view?access\_token=22dbc4704ffbabcce895767f65eb82a2d0ba66cb58e2dcf3878e53d47294ef4d

11.2 Dataset - <https://www.kaggle.com/kumarajarshi/life-expectancy-who/data>

# APPENDIX

#### A. Source code

Notebook

In [1]:**Import pandas as** **as** **pd**

**import** **numpy** **as** **np**

In [2]:**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\_c69e348d48d348aea86cb49f41a4924f = ibm\_boto3.client(service\_name='s3',

ibm\_api\_key\_id='SPYRkLOvFuNf49Qh3j3UcjZRhlFneKsTwz5Sv4e5dn2J',

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\_c69e348d48d348aea86cb49f41a4924f.get\_object(Bucket='smartinternz-donotdelete-pr-svjqfwiwmb99pv',Key='Life Expectancy Data.csv')['Body']

**if** **not** hasattr(body, "\_\_iter\_\_"): body.\_\_iter\_\_ = types.MethodType( \_\_iter\_\_, body )

data = pd.read\_csv(body)

data.head()

In [3]:data.head()

In [4]:data.tail()

In [5]:data.isnull()

In [6]:data.isnull().sum()

In [7]:data=data.dropna()

In [8]:data.isnull().sum()

In [9]:data.types

In [10]:data=data.drop(['Country'],axis=1)

In [11]:a={ 'Developing':1,

'Developed':0}

data['Status']=data['Status'].map(a)

In [12]:data.info()

In [13]:data.columns

In [14]:x = data.drop(['Life expectancy '], axis=1)

y = data['Life expectancy ']

In [15]:**from** **sklearn.model\_selection** **import** train\_test\_split

train\_x,valid\_x,train\_y,valid\_y = train\_test\_split(x,y, random\_state = 101, shuffle=**False**)

In [16]:**from** **sklearn.ensemble** **import** RandomForestRegressor

In [17]:model = RandomForestRegressor(random\_state=45)

In [18]:model.fit(train\_x,train\_y)

In [19]:pred=model.predict(valid\_x)

In [20]:model.score(valid\_x, valid\_y)

Out[20]:0.8916400509418365

In [21]:**from** **sklearn** **import** metrics

In [22]:print('MAE:', metrics.mean\_absolute\_error(valid\_y, pred))

print('MSE:', metrics.mean\_squared\_error(valid\_y, pred))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(valid\_y, pred)))

MAE: 2.051138014527844

MSE: 8.096138983050842

RMSE: 2.8453715017640215

In [23]:!pip install watson-machine-learning-client

In [24]:**from** **watson\_machine\_learning\_client** **import** WatsonMachineLearningAPIClient

In [25]:wml\_credentials={

"apikey": "y\_e6f\_g1rVLrELPIDogIe1TZzwCNjnqlANNDSzKDzwXO",

"instance\_id": "9a65746c-b0ae-47e9-9856-b1a8764d5bf1",

"url": "https://eu-gb.ml.cloud.ibm.com"

}

In [26]:client = WatsonMachineLearningAPIClient( wml\_credentials )

In [27]:model\_props = {client.repository.ModelMetaNames.AUTHOR\_NAME: "Vaibhav",

client.repository.ModelMetaNames.AUTHOR\_EMAIL: "vaibhavchaudhary14107@gmail.com",

client.repository.ModelMetaNames.NAME: "Life\_Expectancy"}

In [28]:model\_artifact =client.repository.store\_model(model, meta\_props=model\_props)

In [29]:published\_model\_uid = client.repository.get\_model\_uid(model\_artifact)

In [30]:published\_model\_uid

Out[30]:'362d8605-4335-4bfe-b73f-cb2dedef74e1'

In [32]:deployment = client.deployments.create(published\_model\_uid, name="Life\_Expectancy")

In [33]:scoring\_endpoint = client.deployments.get\_scoring\_url(deployment)

In [34]:scoring\_endpoint

Out[34]:'<https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/9a65746c-b0ae-47e9-9856-b1a8764d5bf1/deployments/23d0ba7e-3e07-47ee-a5ca-9dd6a6c96e3a/online>'

Node-Red Flow

[{"id":"343e3cb4.555874","type":"tab","label":"Flow 1","disabled":false,"info":""},{"id":"c3dc5ec5.ce742","type":"function","z":"343e3cb4.555874","name":"pre token","func":"//make user given values as global variables\global.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//following are required to receive a token\nvar apikey=\"y\_e6f\_g1rVLrELPIDogIe1TZzwCNjnqlANNDSzKDzwXO\";\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":[["46d24213.b71fdc"]]},{"id":"12c8cf11.8e29f1","type":"http request","z":"343e3cb4.555874","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v3/wml\_instances/9a65746c-b0ae-47e9-9856-b1a8764d5bf1/deployments/9ebf6a06-52d9-4345-93dc-3ef17baedd5c/online","tls":"","persist":false,"proxy":"","authType":"","x":410,"y":180,"wires":[["45a95a6a.306bb4","45d8d8c8.1094e8"]]},{"id":"63daa766.38e1c8","type":"debug","z":"343e3cb4.555874","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":750,"y":280,"wires":[]},{"id":"45a95a6a.306bb4","type":"function","z":"343e3cb4.555874","name":"getFrom Endpoint","func":"msg.payload=msg.payload.values[0][0];\nreturn msg;","outputs":1,"noerr":0,"x":490,"y":280,"wires":[["63daa766.38e1c8","6870695.3bc6998"]]},{"id":"45d8d8c8.1094e8","type":"debug","z":"343e3cb4.555874","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":730,"y":80,"wires":[]},{"id":"a744415.0118dc","type":"function","z":"343e3cb4.555874","name":"sendTo Endpoint","func":"//get token and make headers\nvar token=msg.payload.access\_token;\nvar instance\_id=\"9a65746c-b0ae-47e9-9856-b1a8764d5bf1\"\nmsg.headers={'Content-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//send the user values to service endpoint\nmsg.payload=\n{\"fields\":['Year', 'Status', 'Adult Mortality', 'infant deaths', 'Alcohol',\n 'percentage expenditure', 'Hepatitis B', 'Measles ', ' BMI ',\n 'under-five deaths ', 'Polio', 'Total expenditure', 'Diphtheria ',\n ' HIV/AIDS', 'GDP', 'Population', ' thinness 1-19 years',\n ' thinness 5-9 years', 'Income composition of resources', 'Schooling'],\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":190,"y":180,"wires":[["12c8cf11.8e29f1"]]},{"id":"46d24213.b71fdc","type":"http request","z":"343e3cb4.555874","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"basic","x":390,"y":100,"wires":[["a744415.0118dc"]]},{"id":"6870695.3bc6998","type":"ui\_text","z":"343e3cb4.555874","group":"b393bf5e.dd2cb","order":1,"width":0,"height":0,"name":"","label":"Life Expectancy","format":"{{msg.payload}}","layout":"row-spread","x":700,"y":340,"wires":[]},{"id":"28fa70d2.352e8","type":"ui\_form","z":"343e3cb4.555874","name":"","label":"","group":"b393bf5e.dd2cb","order":0,"width":0,"height":0,"options":[{"label":"Year\t","value":"a","type":"number","required":true,"rows":null},{"label":"Status","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":"Hepatitis B","value":"g","type":"number","required":true,"rows":null},{"label":"Measles","value":"h","type":"number","required":true,"rows":null},{"label":"BMI","value":"i","type":"number","required":true,"rows":null},{"label":"under-five deaths","value":"j","type":"number","required":true,"rows":null},{"label":"Polio","value":"k","type":"number","required":true,"rows":null},{"label":"Total expenditure","value":"l","type":"number","required":true,"rows":null},{"label":"Diphtheria","value":"m","type":"number","required":true,"rows":null},{"label":"HIV/AIDS","value":"n","type":"number","required":true,"rows":null},{"label":"GDP","value":"o","type":"number","required":true,"rows":null},{"label":"Population","value":"p","type":"text","required":false,"rows":null},{"label":"thinness 1-19 years","value":"q","type":"text","required":false,"rows":null},{"label":"thinness 5-9 years","value":"r","type":"text","required":false,"rows":null},{"label":"Income composition of resources","value":"s","type":"text","required":false,"rows":null},{"label":"Schooling","value":"t","type":"text","required":false,"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":"Predict","cancel":"cancel","topic":"","x":70,"y":100,"wires":[["c3dc5ec5.ce742"]]},{"id":"b393bf5e.dd2cb","type":"ui\_group","z":"","name":"Machine Learning Model","tab":"d1cb38a4.f7dc78","order":1,"disp":true,"width":"6","collapse":false},{"id":"d1cb38a4.f7dc78","type":"ui\_tab","z":"","name":"Home Page","icon":"dashboard","disabled":false,"hidden":false}]