LIFE EXPECTANCY MODEL

PROJECT DOCUMENT

15-06-20

Nikitha Mary Thomas

Contents

1. INTRODUCTION

1.1 Overview

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more. Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age, and other demographic factors including BMI, gender etc. Life expectancy varies by geographical area and by era. In the Bronze Age, for example, life expectancy was 26 years, while in 2010, it was 67 years.

1.2 Purpose

This project aims at creating a new model based on the data provided is to evaluate the life expectancy of the country. The model takes in various factors including adult mortality etc. to calculate the result. The data set from WHO offers a time frame from 2000 to 2015.This is used to produce the model. The algorithm that produces the best result is ELASTIC NET regression algorithm and is used here.

2. LITERATURE SURVEY

a. Existing Problem

The World Health Organization (WHO) began producing annual life tables for all Member States in 1999. These life tables are a basic input to all WHO estimates of global, regional and country-level patterns and trends in all-cause and cause-specific mortality. After the publication of life tables for years to 2009 in the 2011 edition of World Health Statistics, WHO has shifted to a two year cycle for the updating of life tables for all Member States. Even still the model is not really updated in every fields. WHO applies standard methods to the analysis of Member State data to ensure comparability of estimates across countries. This will inevitably result in differences for some Member States with official estimates for quantities such as life expectancy, where a variety of different projection methods and other methods are used.

b. Proposed Solution

It is planned to create a model that will make a system train to do the job of prediction, rather than the use of complex mathematical processes. The behind scenes of training the system does involve mathematics but it is way easier than doing all equations on own. By using supervised machine learning techniques. We can extract a model that will be able to predict the life expectancy of future years. One method of approach is to use LSTM models to achieve this task.

3. THEORETICAL ANALYSIS

a. Algorithm

Machine Learning:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to learn automatically and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

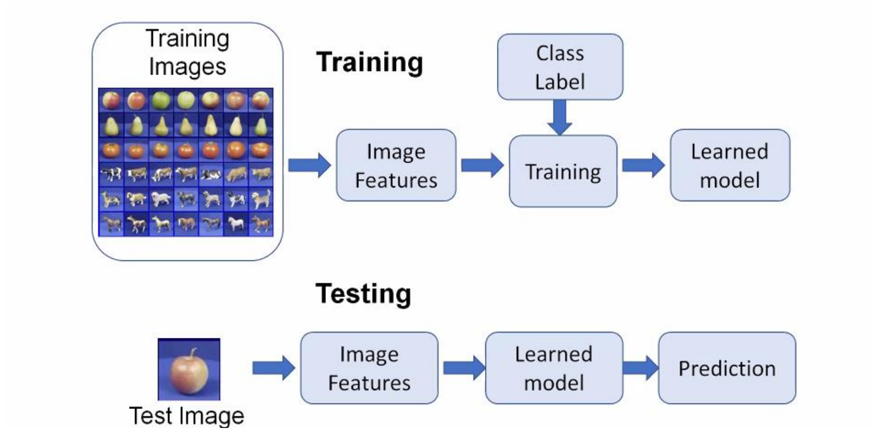
The machine learning algorithm is classified into Supervised, unsupervised and reinforced learning. For this project a type of supervised model is used.

Supervised Learning:

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training data set, the learning algorithm produces an inferred function to make predictions about the output values. The system can provide targets for any new input after sufficient training. Supervised learning problems can be further grouped into Regression and Classification problems

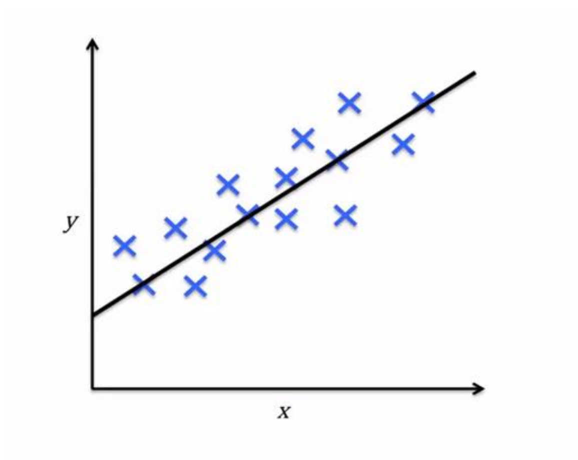
Classification:

A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes.



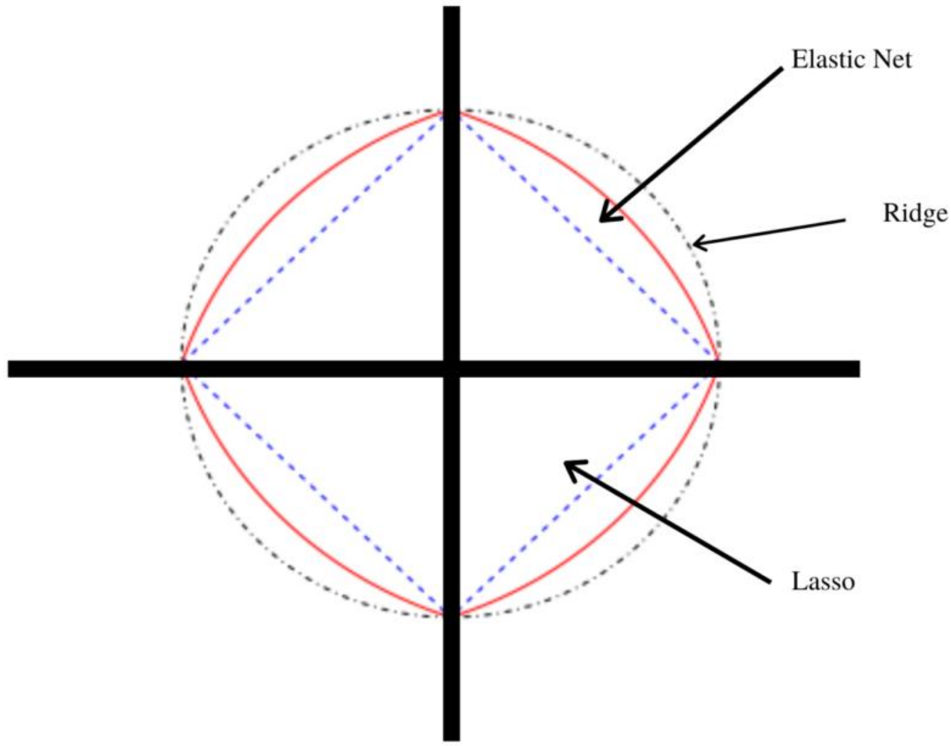
Regression:

A regression problem is when the output variable is a real or continuous value. Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points.



In this project we are using Elastic Nest Regression as it produces the least error rate among other algorithms available.

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve on the regularization of statistical models.



The elastic net method improves on lasso's limitations, i.e., where lasso takes a few samples for high dimensional data, the elastic net procedure provides the inclusion of "n" number of variables until saturation. In a case where the variables are correlated groups, lasso tends to choose one variable from such groups and ignore the rest entirely.

To eliminate the limitations found in lasso, the elastic net includes a quadratic expression in the penalty, which, when used in isolation, becomes ridge regression. The quadratic expression in the penalty elevates the loss function toward being convex. The elastic net draws on the best of both worlds – i.e., lasso and ridge regression.

4. EXPERIMENTAL ANALYSIS

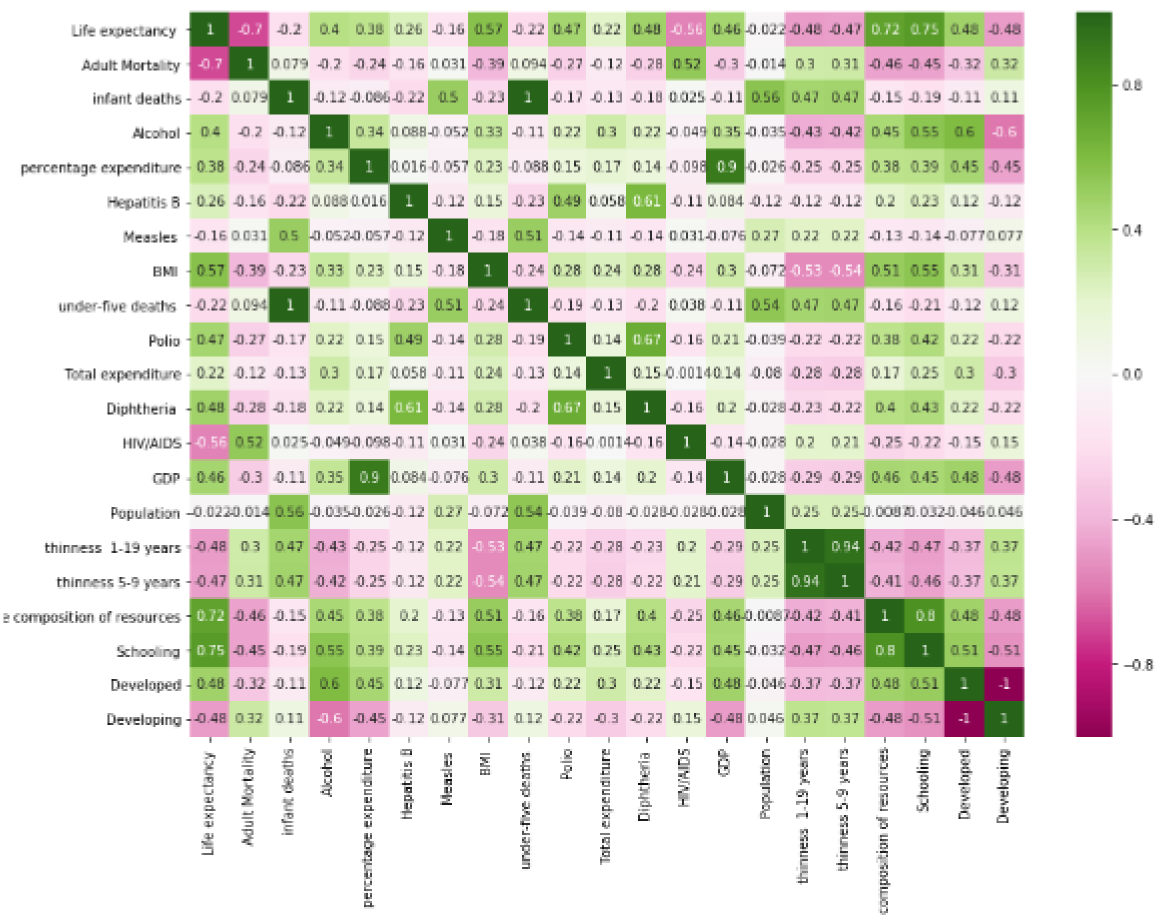
Various regression algorithms were tried out to find out the best model with least error rates. The algorithms that can be used here include Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Linear Regression with Polynomic features, Decision Tree Regression, Random Forest Regression

5. RESULT

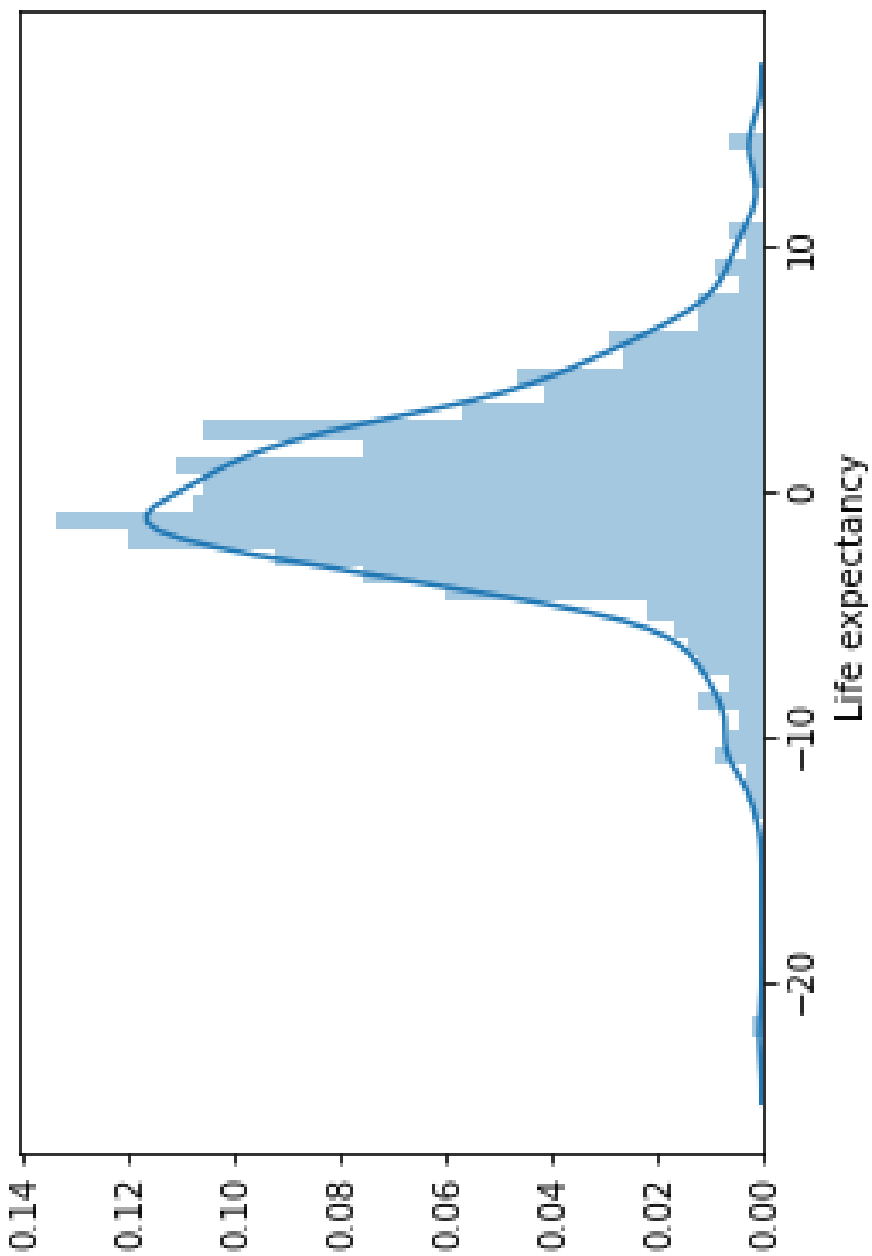
5.1 Project Model

The model is created in Watson studio and Jupyter notebook. The data set contains 22 features including countries, adult mortality etc. Since regression can only use number values the data set is grouped using countries, to avoid conflict.

The null values and infinity were removed and the data was then split into training and testing data. The following heat map was generated from the data.

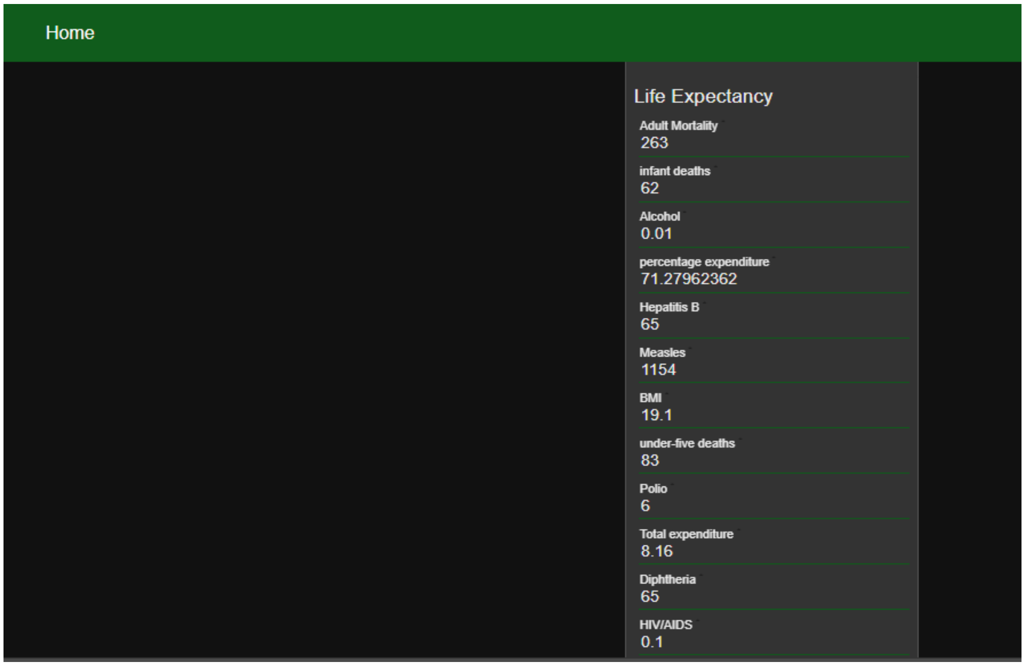


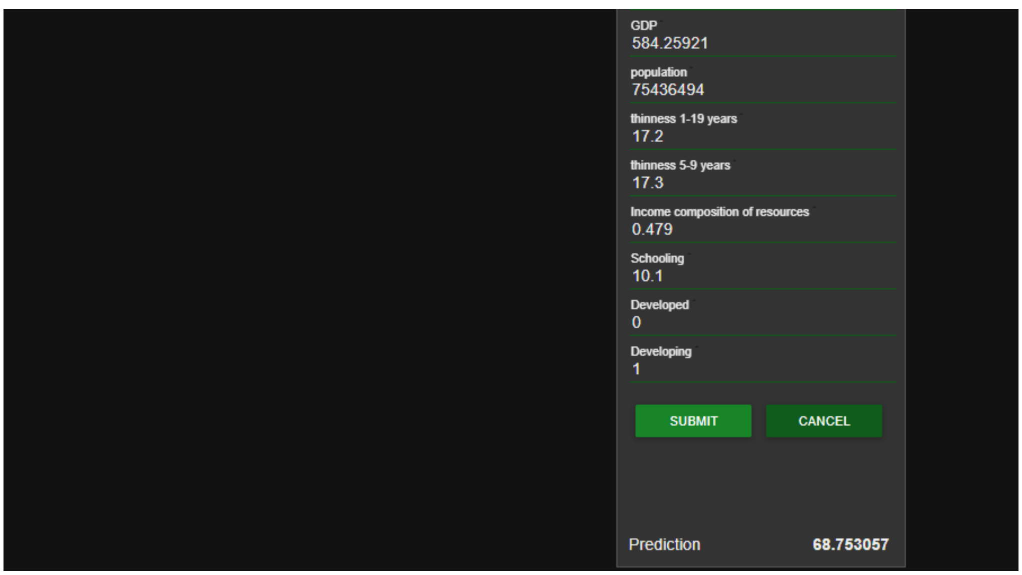
After training the data using Elastic net regression the prediction models were obtained,



5.2 User Interface

The UI output of the project is shown below with sample values. This was created with the help of Node Red Application provided by IBM Cloud.





6. ADVANTAGES AND DISADVANTAGES

a. Advantage

- Predicts Life Expectancy in very little time.
- Simple to use UI

b. Disadvantage

- Dataset is from 2000 – 2015
- Not error free

7. CONCLUSION

Life expectancy, estimate of the average number of additional years that a person of a given age can expect to live. The most common measure of life expectancy is life expectancy at birth. Life expectancy is a hypothetical measure. It assumes that the age-specific death rates for the year in question will apply throughout the lifetime of individuals born in that year. The estimate, in effect, projects the age-specific mortality (death) rates for a given period over the entire lifetime of the population born (or alive) during that time. The measure differs considerably by pandemics, geographic location and much more.

8. BIBILOGRAPHY

- Chen, Jonathan H., and Steven M. Asch. "Machine learning and prediction in medicine—beyond the peak of inflated expectations." The New England journal of medicine 376.26 (2017): 2507.
- https://www.britannica.com/science/life-expectancy
- Sebastian Raschka ,Python Machine Learning, PACKT Publishing, UK, 2015
- https://www.datasciencesociety.net
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering 160 (2007): 3-24.
- https://in.mathworks.com/

9. APPENDIX

A. Node- Red Flow

[{"id":"e31b199.a199ee8","type":"tab","label":"Flow
1","disabled":true,"info":""},{"id":"6adc9c91.631c14","type":"ui_form","z":"e31b199.a199ee8",
"name":"Life Expectancy
Model","label":"Life_Expectancy","group":"da40dc3.6fdee2","order":1,"width":0,"height":0,"op
tions":[{"label":"Adult
Mortality","value":"a","type":"number","required":true,"rows":null},{"label":"infant
deaths","value":"b","type":"number","required":true,"rows":null},{"label":"Alcohol
","value":"c","type":"number","required":true,"rows":null},{"label":"percentage
expenditure","value":"d","type":"number","required":true,"rows":null},{"label":"Hepatitis
B","value":"e","type":"number","required":true,"rows":null},{"label":"Measles","value":"f","typ
e":"number","required":true,"rows":null},{"label":"
BMI","value":"g","type":"number","required":true,"rows":null},{"label":"under-five
deaths","value":"h","type":"number","required":true,"rows":null},{"label":"Polio","value":"i","ty
pe":"number","required":true,"rows":null},{"label":"Total
expenditure","value":"j","type":"number","required":true,"rows":null},{"label":"Diphtheria","val
ue":"k","type":"number","required":true,"rows":null},{"label":"
HIV/AIDS","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 1-19
years","value":"o","type":"number","required":true,"rows":null},{"label":" thinness 5-9
years","value":" p","type":"number","required":true,"rows":null},{"label":"Income composition
of
resources","value":"q","type":"number","required":true,"rows":null},{"label":"Schooling","value
":"r","type":"number","required":true,"rows":null},{"label":"Developed","value":"s","type":"nu
mber","required":true,"rows":null},{"label":"Developing","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":"Reset","topic":"","x":1
60,"y":60,"wires":[["1fd1a0b5.06f35f"]]},{"id":"1fd1a0b5.06f35f","type":"function","z":"e31b19
9.a199ee8","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\n//following are required to receive a token\nvar apikey=\"fdQYOL7rQa9Lb_2ohzGk3JXqWa31P4WKr4p0IcQM2FVt\";\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":440,"y":60,"wires":[["5da3a5b7.682a3c"]]},{"id":"9f3d26e5.012e78","type":"debug","z":"e31b199.a199ee8","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":710,"y":260,"wires":[]},{"id":"26e28497.9fc0ac","type":"function","z":"e31b199.a199ee8","name":"getFrom Endpoint","func":"msg.payload=msg.payload.values[0][0];\nreturn msg;","outputs":1,"noerr":0,"x":430,"y":340,"wires":[["9f3d26e5.012e78","c6ef180a.2f78a8"]]},{"id":"9fb7ef40.a77ec","type":"function","z":"e31b199.a199ee8","name":"sendTo Endpoint","func":"//get token and make headers\nvar token=msg.payload.access_token\nvar instance_id=\"486da02e-390e-47aa-a823-141d46082113\"\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\");//0;\nvar t = global.get(\"t\");//1;\n\n//send the user values to service endpoint\nmsg.payload = \n{\"fields\":\n[\"Adult Mortality\",\n \"infant deaths\",\"Alcohol\", \"percentage expenditure\",\"Hepatitis B\",\n \"Measles \",\" BMI\", \"under-five deaths \", \"Polio\",\"Total expenditure\",\n \"Diphtheria \",\" HIV/AIDS\",\"GDP\",\"Population\",\" thinness 1-19 years\",\n \" thinness 5-9 years\",\"Income composition of resources\",\"Schooling\",\n \"Developed\",

\"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":150,"y":280,"wires":[["98a3780a.fdb758","f2bcc50f.70ca88"]]},{"id":"5da3a5b7.682a3c","type":"http request","z":"e31b199.a199ee8","name":"","method":"POST","ret":"obj","paytoqs":true,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":150,"y":160,"wires":[["9fb7ef40.a77ec"]]},{"id":"c6ef180a.2f78a8","type":"ui_text","z":"e31b199.a199ee8","group":"da40dc3.6fdee2","order":3,"width":0,"height":0,"name":"","label":"Prediction","format":"{{msg.payload}}","layout":"row-spread","x":700,"y":420,"wires":[]},{"id":"98a3780a.fdb758","type":"debug","z":"e31b199.a199ee8","d":true,"name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":470,"y":260,"wires":[]},{"id":"f2bcc50f.70ca88","type":"http request","z":"e31b199.a199ee8","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/486da02e-390e-47aa-a823-141d46082113/deployments/a43745a0-c9bc-44aa-9a87-f85c66a0a9c0/online","tls":"","persist":false,"proxy":"","authType":"","x":150,"y":400,"wires":[["26e28497.9fc0ac","1976ef42.38acb1"]]},{"id":"1976ef42.38acb1","type":"debug","z":"e31b199.a199ee8","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":410,"y":460,"wires":[]},{"id":"da40dc3.6fdee2","type":"ui_group","z":"","name":"Machine Learning Model","tab":"5c4e8246.da5d3c","order":2,"disp":true,"width":"6","collapse":false},{"id":"5c4e8246.da5d3c","type":"ui_tab","z":"","name":"Home Page","icon":"dashboard","disabled":false,"hidden":false}]