PREDECTING LIFE EXPECTANCY OF A COUNTRY USING ML

Name : Anirudh Hosur
College : Vellore Institute of Technology, Chennai Campus

A) PROJECT PLANNING AND KICKOFF

Task 1 : Project scope, schedule, team and deliverables -

Part 1 - Project Kick-off template
Kick off meeting agenda
Hello this is Anirudh Hosur currently pursuing B.Tech from VIT Chennai campus.

Discuss project background
1. What we have today - The project is predicting life expectancy of a person living in that country using machine learning.
2. Why we need to change - We need to discipline ourselves and time management should be taken seriously to finish the project.
3. What are the key measures of success? - Time management and dedication

Identify stakeholders
1. Who is impacted? - Anyone who is intrested to know the life expectancy
2. Who are the key stakeholders - Sponsor and/or other decision makers who represent constituencies that will be impacted, and whose support is critical to the success of the project? -

Review project objectives
1. Objectives - To predcit the life expectancy of a person in their country
2. Deliverables - To deliver a strong model with good accuracy to predict the life expectancy
3. Assumption - All asumptions are based on the data and its values

Review team member roles & responsibilities
1. Project Team - An indivitual project

Review other potential issues, risks, questions and concerns
What might get in the way of success? - If the data dosen't fit properly to the model, low accuracy, etc.
How could we address those concerns? - By trying to train diffrent models and figure out the most appropriate amongst them.

Identify next steps and timing
- Team communications - The communication shall take place through slack
- Frequency of team meetings - Mondays and Thursdays shall be mentoring sessions every week. Saturday shall be project review day.

Part 2 - How to kick off a project
A project kick-off check list:

1) Contract - I am the authorised indivitual responsible for the sucess of this project
2) Client needs and priorites - The client requires a life expectancy of a person living in that country and the priority is always to deliver one with highest accuracy.
3) Project Plan - To train the dataset such that, the tested data responses with good accuracy.
4) Team member roles - An indivitual project
5) Scope of work - Not a phased approach, definetly a direct approach
6) Project Schedule - A duration of 29 days
7) Project Budget - Nil
8) Project Reviews - All reviews and milestones to be showcased on Saturdays
9) Project Management Process - Going to manage the data and derive meaningful insights from it
10) Project Manager Expectations - A strong communication with the mentors for their constant support and guidance & to provide a good model with better accuracy.

Task 2 : Set up the developement environemnt -

Part 1 - Create a github account : https://github.com/AnirudhHosur

Part 2 - What is Github?

A platform where people showcase their projects which is available for the world to learn from.

Software touches every corner of our lives and thus we can contribute and pull requests to update the software.

Part 3 - What is Slack ?

A very easy to use, interactive & friendly software to communicate with your project mates and mentors.

Part 4 - Install slack and create account - https://app.slack.com/client/T013KTY8SH3/D013J1RCYR4/user_profile/U013PB5DB1T

Part 5 - Working with document writer

A very friendly software to update your work and keep records manually.

B) Explore IBM Cloud Platform -

Task 1 : Create IBM cloud account -

Part 1 - Sign up for IBM accademic initative account - Done

Part 2 - Signup for IBM Cloud - https://cloud.ibm.com/user

Part 3 - Getting started with IBM Cloud - Build and deploy apps for the project

Task 2 : Create a Node-RED Starter Application -

Part 1 : Starter application created in the IBM cloud , including a cloudant database to store the application flow configuration.

Added the dashboard and modbus pallette features onto the Node-red app.

Name : Node-RED Life
https://cloud.ibm.com/devops/toolchains/6ebd6d9d-83f7-44d2-ac68-2f8d58fdca4c?env_id=ibm:yp:eu-gb

Part 2 : What is Node-RED?
Low code programmin gor event driven apps.

Part 3 : Learn more with Node-RED Labs
Learning how to implement , style and add more nodes to integrate your overall node-red flow.

Part 4 : what is an api ?
Application Programming Interface : Behind the scenes which makes possbile for underlying communications between users and machines.
Helps out in reuest-response mechanishms to fullfil communications.

Part 5 : Create a simple webpage
A basic website is to be built using CSS & basic HTML

C) Explore IBM Watson Services:

Task 1 : Explore IBM watson usecases -

Part 1 - Explore Watson products and services:

AI shall be infused with our application to make more accurate predcitions of life expectancy of a human in a country. Data from IBM cloud shall be infused with IBM watson to make accurate predictions.

Part 2 : Watson at work -

AI changes the way world works making buisness better. With watson personalized experiments, minimising risks and making strong AI-infsed predictions is possible.

Task 2 : Explore IBM watson Machine Learning

Part 1 - Introduction to ML:

Ibm watson shall behave like the backend and the ML model shall be stored in it. The Ml algorithms shall proceed in the following steps -

a) Importing the libraries
b) Importing the dataset
c) Removing the country column
d) Removing the rows which contain Nan or 0 values
e) One-hot encoding the categorical variables
f) Splitting the dataset into training and testing
g) Applying regression models namely, Linear regression, decision trees, Random forest
h) Assesing the accuracy

Part 2 - About IBM watson ML:

We first need to provsion the IBM cloud services with watson to integrate these two to create your desired Ml model. Watson studio tools can be used to train-test and deploy models with ease.
The watson service can be set-up through IBM cloud.
Object storage should also be provisioned to act as a bucket of data.

D) Introduction to watson studio:

Task 1 - Build your own ML model in IBM watson studio -

Part 1 - Build your own ML model in IBM watson studio using watson studio services -

Like the famous jupyter notebook, waston provides a notebook of its own in which our coding shall be implemented. From computation to visualization everything is taken care of. To implement it the pre-requisits include:

a) A working project in watson studio
b) The concerned dataset
c) Notebook
d) WML credentials
e) Access tokens

The model shall then be deployed as a web service. Test deployement endpoint shall acess your model and provide you the required output with regards to the accuracy you have achived.

Task 2 - Automate your ML model

This section is for those who have are using the AutoML service of watson studio to automate their model. I am hard coding the model using python.

E) Predicitng Life Expectancy with python:

Task 1 - Collect the dataset for the project

The WHO dataset is available on kaggle - https://www.kaggle.com/kumarajarshi/life-expectancy-who

Task 2 - Create nessassary ibm cloud services -

Sign up for IBM watson studio which is powered by ibm cloud itself. Now the watson studio and ibm cloud are integrated in one.

Task 3 - Create a watson studio project

A new watson studio project needs to be created. This will be integrated with cloud storage bucket.

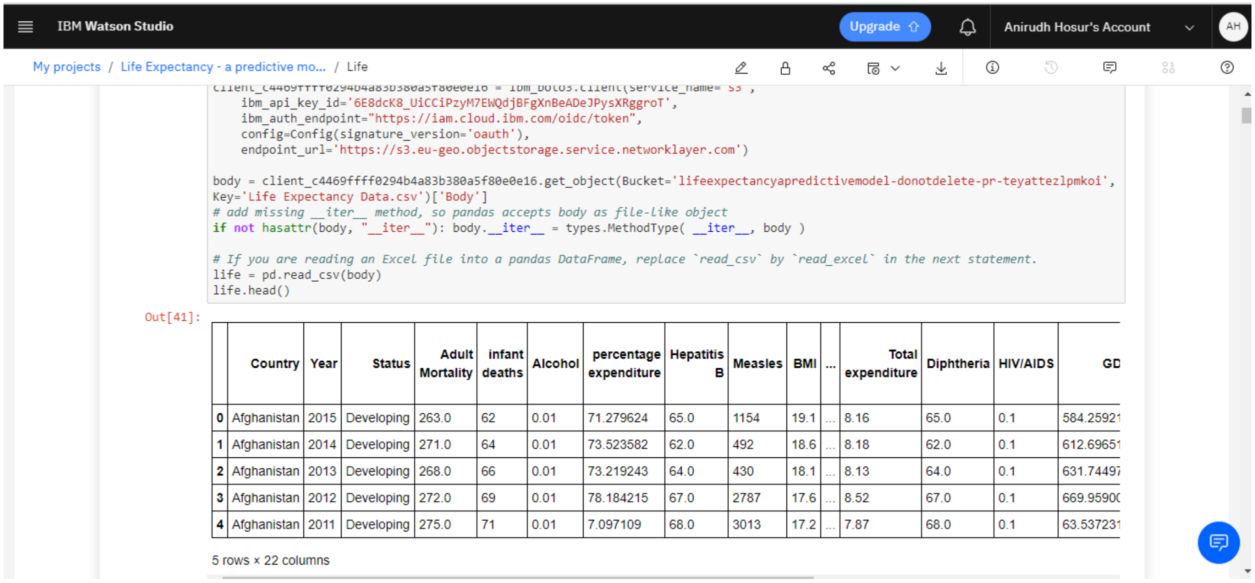
Task 4 - Configure Watson studio

The tokens and WML credentials need to be added, i.e 3 things need to be taken care of - Cloud Storage, Access Tokens and Associated services i.e. watson studio machine learning.

Task 5 - Create machine learning service

In IBM cloud, from the catalog ibm watson macine learning service can pe provisioned.

Task 6 - Create a jupyter notebook in waston studio and import the dataset



```
client_c4469ffff0294b4a83b380a5f80e0e16 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='6E8dcK8_UiCCiPzyM7EWQdjBFgXnBeADeJPysXRggroT',
    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_c4469ffff0294b4a83b380a5f80e0e16.get_object(Bucket='lifeexpectancyapredictivemodel-donotdelete-pr-teyattezlpmkoi',
Key='Life Expectancy Data.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 )

# If you are reading an Excel file into a pandas DataFrame, replace `read_csv` by `read_excel` in the next statement.
life = pd.read_csv(body)
life.head()
```

Out[41]:

	Country	Year	Status	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	...	Total expenditure	Diphtheria	HIV/AIDS	GD
0	Afghanistan	2015	Developing	263.0	62	0.01	71.279624	65.0	1154	19.1	...	8.16	65.0	0.1	584.25921
1	Afghanistan	2014	Developing	271.0	64	0.01	73.523582	62.0	492	18.6	...	8.18	62.0	0.1	612.69651
2	Afghanistan	2013	Developing	268.0	66	0.01	73.219243	64.0	430	18.1	...	8.13	64.0	0.1	631.74497
3	Afghanistan	2012	Developing	272.0	69	0.01	78.184215	67.0	2787	17.6	...	8.52	67.0	0.1	669.95900
4	Afghanistan	2011	Developing	275.0	71	0.01	7.097109	68.0	3013	17.2	...	7.87	68.0	0.1	63.537231

5 rows × 22 columns

Task 7 - Build a machine learning model and create end-points for node-red integration

The Ml models being implemented are Random forest, Linear Regression, Desicion trees. WML credentials are integrated using machine learning service credentials.

The accuracy(r2_score) obtained successfully after regression are -

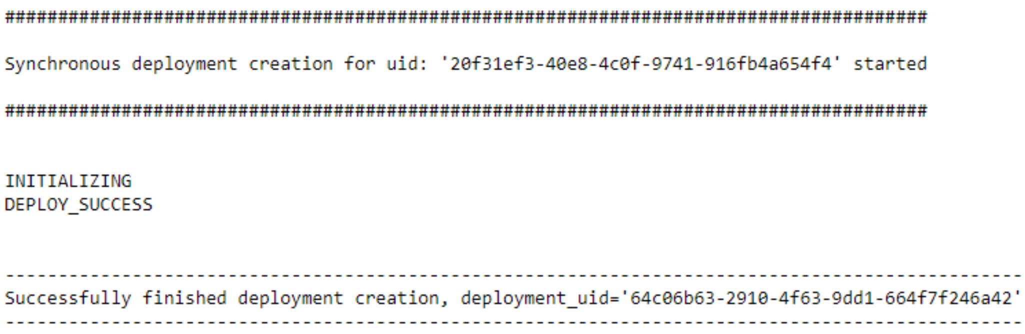
a) **Linear Regression - 0.8557**

b) **Decision Trees - 0.91996**

c) **Random Forests - 0.96128**

Clearly, Random Forest was the most sucessful one.

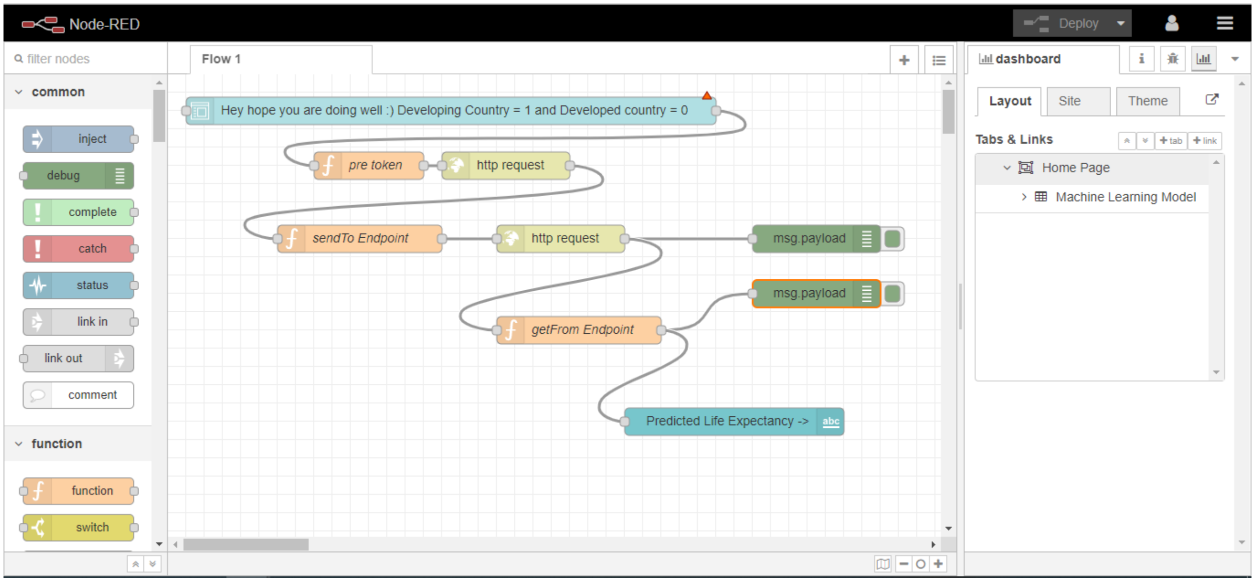
End points for node-red creation have also been taken care of, giving the following message -



```
#######################################################################################

Synchronous deployment creation for uid: '20f31ef3-40e8-4c0f-9741-916fb4a654f4' started

#######################################################################################


INITIALIZING
DEPLOY_SUCCESS


------------------------------------------------------------------------------------------------
Successfully finished deployment creation, deployment_uid='64c06b63-2910-4f63-9dd1-664f7f246a42'
------------------------------------------------------------------------------------------------
```

Task 8 - Build Node-red FLow to integrate node-red services

With regards to .json coding, the node-red has been successfully deployed.



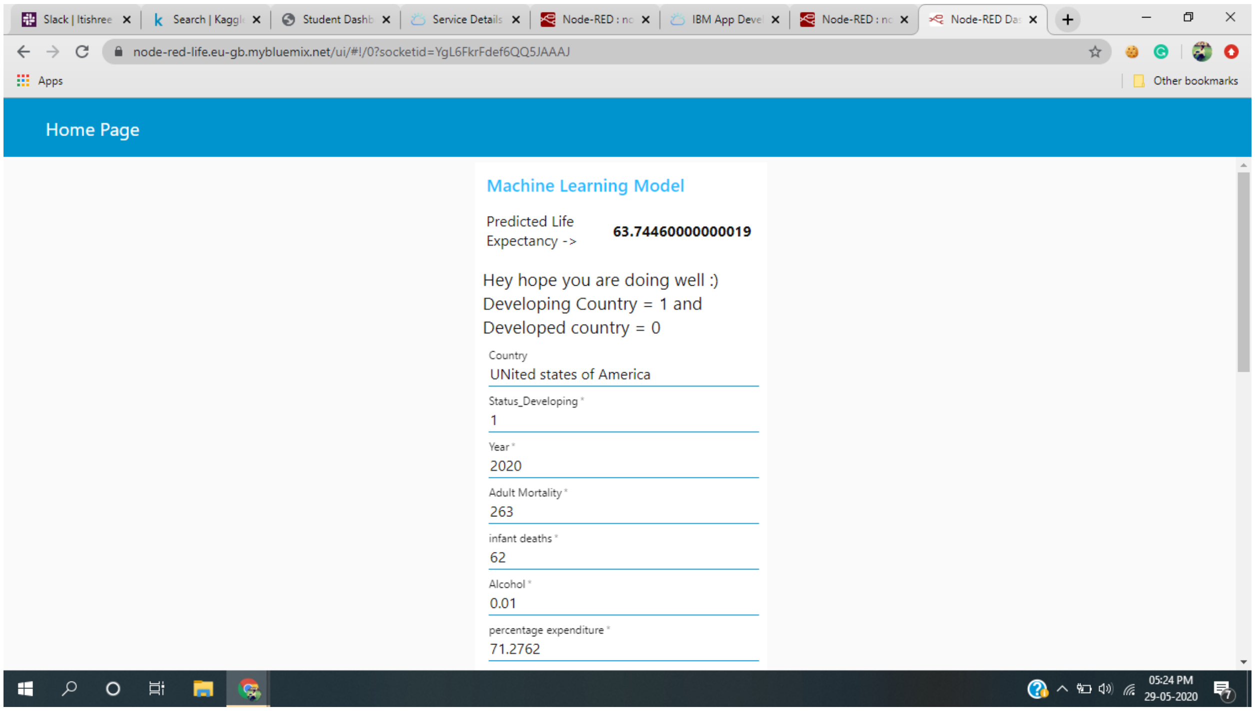
Hey hope you are doing well :) Developing Country = 1 and Developed country = 0

CONCLUSION :

IBM Watson behaves as the back-end which contains the ML model and the database. Whereas, node-red acts like the UI through which the user can speculate.
We clearly observe from below that the predcition from the ML watson studio model and the prediction from the node-red UI are the same.
i.e. 63.7446

This Prediction is done with an accuracy of 96.128%. Obtained by training "Random Forests" on the model.

The UI :



Machine Learning Model

Predicted Life Expectancy -> 63.74460000000019

Hey hope you are doing well :)
Developing Country = 1 and
Developed country = 0

Country
UNited states of America

Status_Developing *
1

Year *
2020

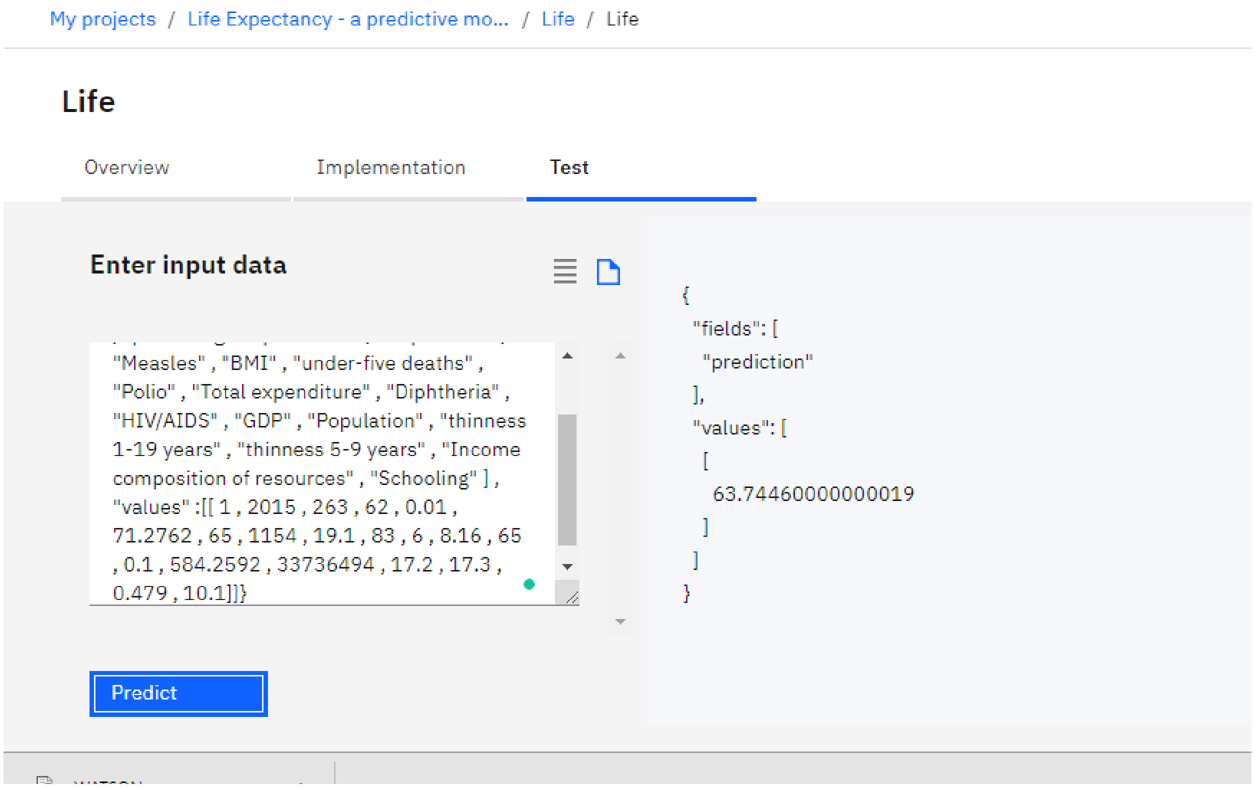
Adult Mortality *
263

infant deaths *
62

Alcohol *
0.01

percentage expenditure *
71.2762

The Watson prediction :


My projects / Life Expectancy - a predictive mo... / Life / Life
Life
Overview
Implementation
Test
Enter input data
Predict

```
"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" :[[ 1 , 2015 , 263 , 62 , 0.01 ,
71.2762 , 65 , 1154 , 19.1 , 83 , 6 , 8.16 , 65
, 0.1 , 584.2592 , 33736494 , 17.2 , 17.3 ,
0.479 , 10.1]]}
```

```
{
 "fields": [
  "prediction"
 ],
 "values": [
  [
   63.74460000000019
  ]
 ]
}
```