Project Name: Predicting Life Expectancy using Machine Learning - SB45648

PROJECT PLANNING & KICKOFF

➢ Project Scope Document

- **Project Summary :**

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. This problem statement provides a way 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.

- **Project Requirements:**

To analyze the insights, the data set is required to build the model. IBM Watson Studio is use to work on Python Jupyter Notebook. Variables are analyze to check the dependencies of on life expectancy. Build a suitable algorithm with best accuracy and predict life expectancy.

- **Functional Requirement:**

Predicting Life Expectancy rate of a country on the basis of various features.

- **Technical Requirements:**

Python, IBM Cloud, IBM Watson

- **Software Requirements: NA**

- **Project Deliverables:**

Machine Learning Model to predict Life Expectancy rate of country data insights.

- **Project Team:**

Divyanshi Garg

- **Project Schedule:**

Assigned on 6 June 2020 and complete it in 1 month i.e. 5 July 2020.

SET UP THE DEVELOPMENT ENVIRONMENT

1. **Where do you develop the code ?**

IBM Watson & IBM Cloud

2. **Where do you save it ?**

IBM Cloud

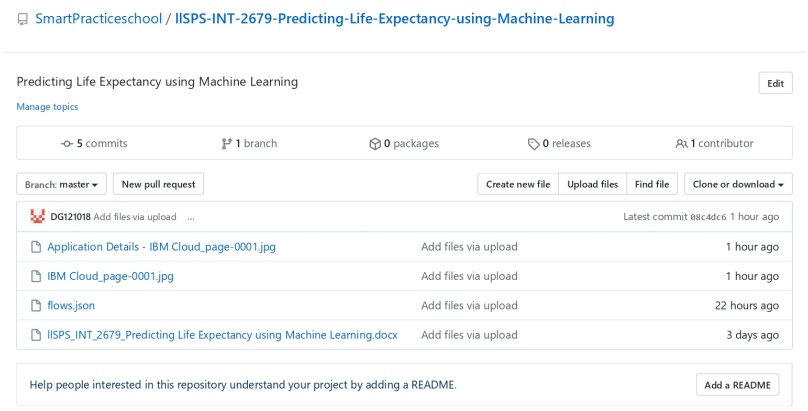
3. **How do you communicate with your team ?** Slack Channel
4. **Where do you document your work ?**

ZOHO Writer

5. **How do you share deliverables with your client ?**

GITHUB Repositry

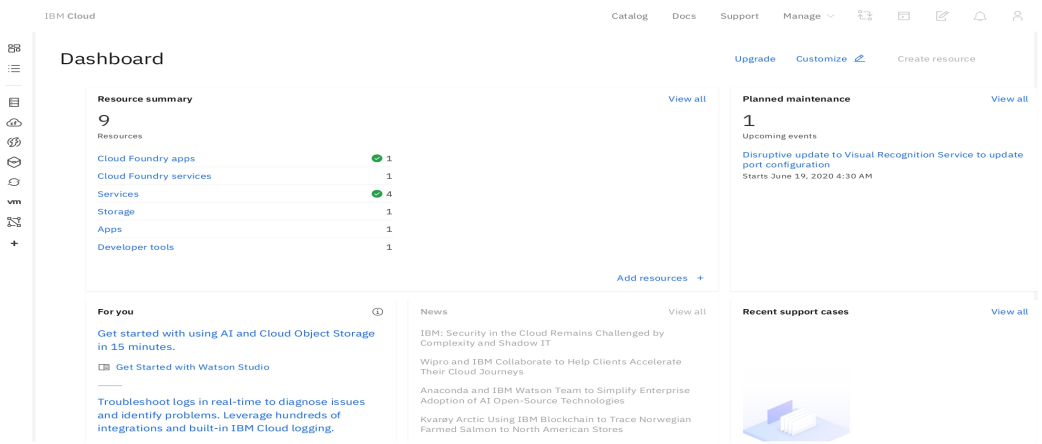
- **GITHUB Account:**



EXPLORE IBM WATSON CLOUD SERVICES

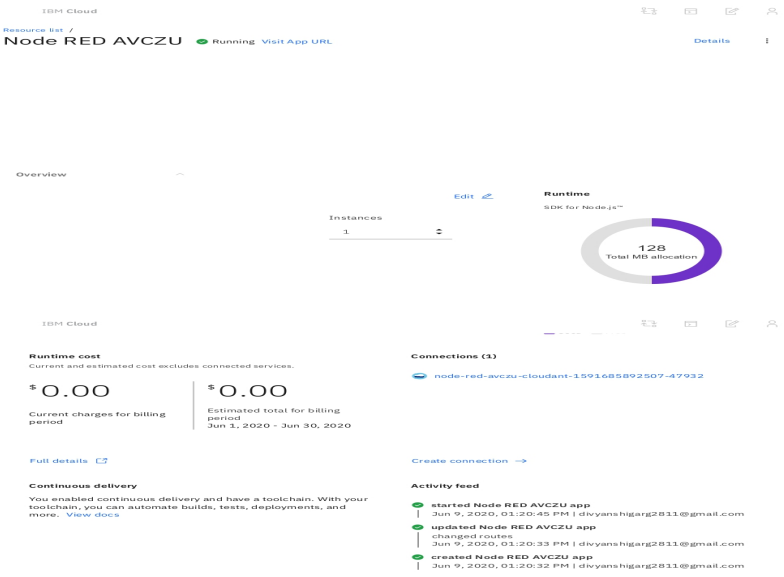
- **IBM Cloud Account Link:**

https://cloud.ibm.com/



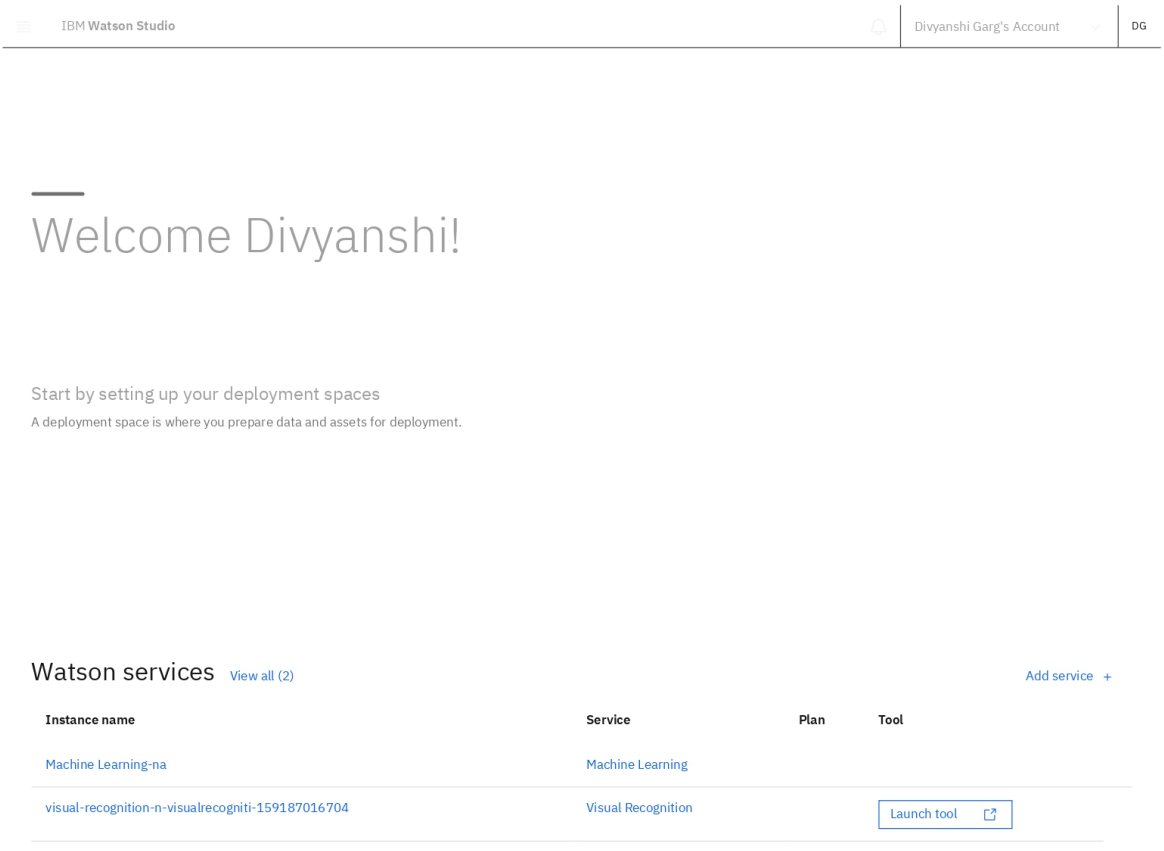
- **Node-Red Application Link :**

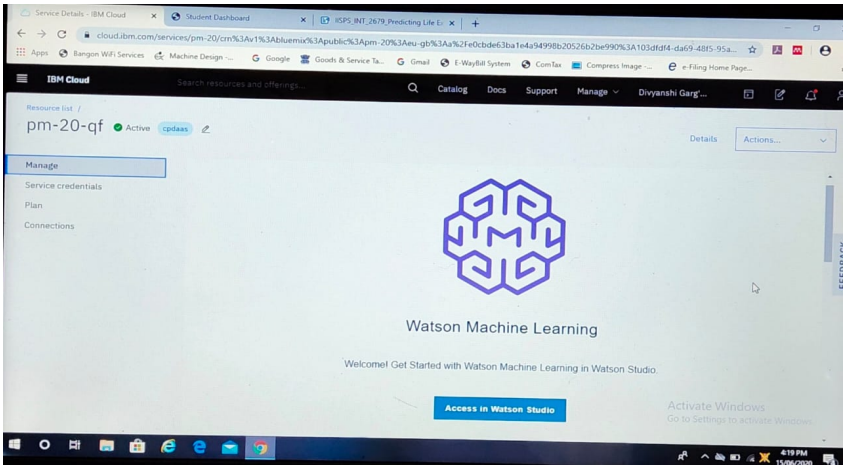
https://node-red-avczu.eu-gb.mybluemix.net/red/

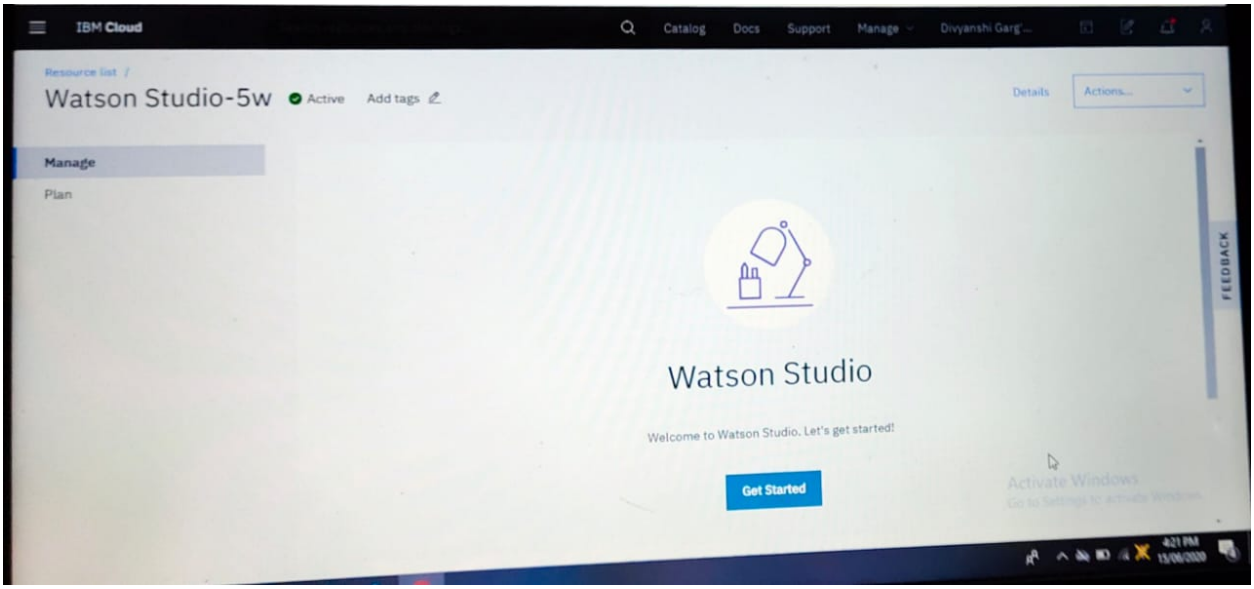


EXPLORE IBM WATSON SERVICES

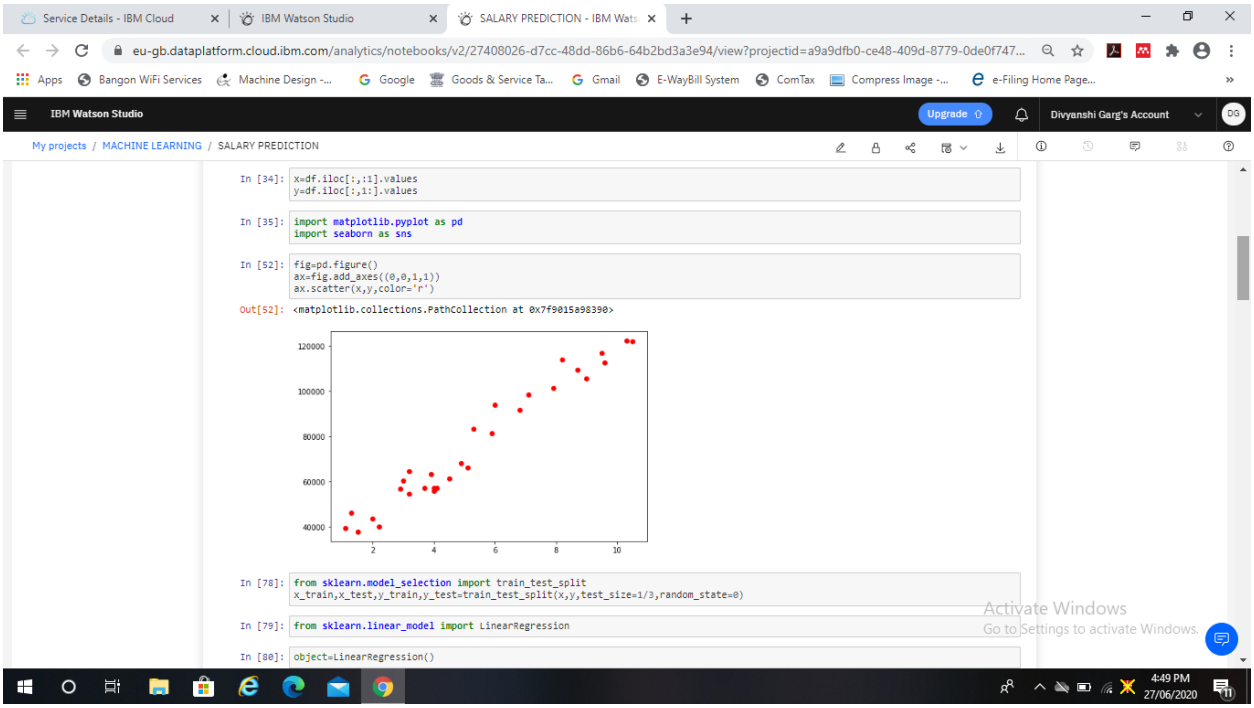
- **IBM Watson Use Cases & Explore IBM Watson ML :**

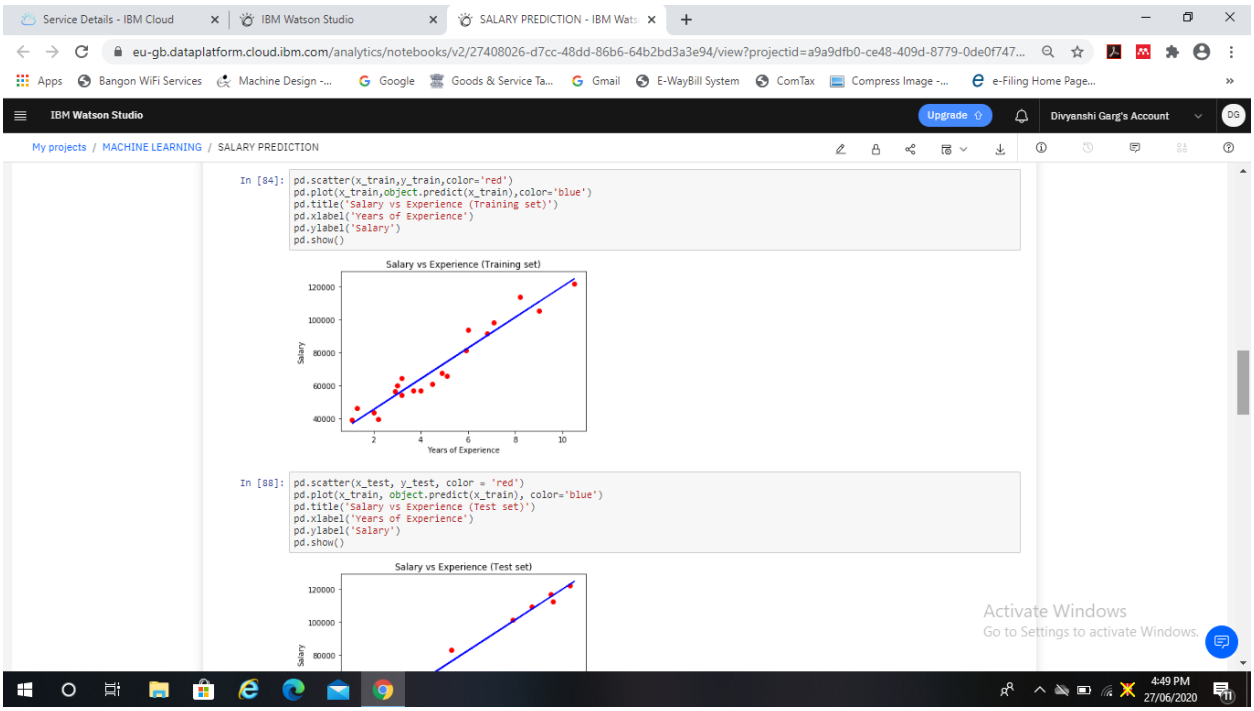



IBM Cloud
Catalog
Docs
Support
Manage
Divyanshi Garg'...
pm-20-qf
Active
cpdaas
Details
Actions...
Manage
Service credentials
Plan
Connections
Watson Machine Learning
Welcome! Get Started with Watson Machine Learning in Watson Studio.
Access in Watson Studio
Activate Windows


IBM Cloud
Catalog
Docs
Support
Manage
Divyanshi Garg'...
Watson Studio-5w
Active
Add tags
Details
Actions...
Manage
Plan
Watson Studio
Welcome to Watson Studio. Let's get started!
Get Started
FEEDBACK
Activate Windows

INTRODUCTION TO WATSON STUDIO





Service Details - IBM Cloud | IBM Watson Studio | SALARY PREDICTION - IBM Wats

eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/27408026-d7cc-48dd-86b6-64b2bd3a3e94/view?projectid=a9a9dfb0-ce48-409d-8779-0de0f747...

Apps | Bangon WiFi Services | Machine Design -... | Google | Goods & Service Ta... | Gmail | E-WayBill System | ComTax | Compress Image -... | e-Filing Home Page...

IBM Watson Studio | Upgrade | Divyanshi Garg's Account

My projects / MACHINE LEARNING / SALARY PREDICTION

Years of Experience

In [88]:
```
pd.scatter(x_test, y_test, color = 'red')
pd.plot(x_train, object.predict(x_train), color='blue')
pd.title('Salary vs Experience (Test set)')
pd.xlabel('Years of Experience')
pd.ylabel('Salary')
pd.show()
```

In [90]:
```
new_salary_pred = object.predict([[15]])
new_salary_pred
```

Out[90]:
```
array([[167005.32889087]])
```

In [92]:
```
from watson_machine_learning_client import WatsonMachineLearningAPIClient
```

In [94]:
```
wml_credentials = {
  "apikey": "QfNQhxIJpTr3TXJyU-Cctk4mqV4pk62bmueieJDfCA_S",
  "iam_apikey_description": "Auto-generated for key 19df97fa-8591-4f09-b790-e8a3c873bd40",
  "iam_apikey_name": "wdp-writer",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/e0cbde63ba1e4a94998b2052bb2be990::serviceid:ServiceId-f7b8cdf6-b3c
```

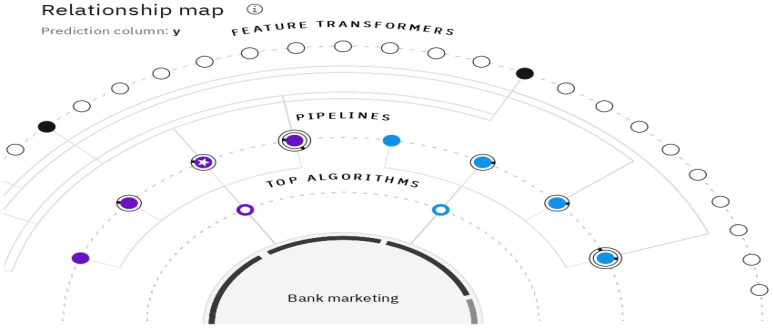

4:49 PM
27/06/2020

IBM Watson Studio | Upgrade | Divyanshi Garg's Account | DG

Experiment summary | Pipeline comparison | Rank by: ROC AUC (Optimized) | Cross validation | Holdout

Relationship map

Prediction column: y



Progress map

Swap view

Experiment completed

8 PIPELINES GENERATED

8 pipelines generated from alg
pipeline leaderboard below for

Time elapsed: 19 minutes

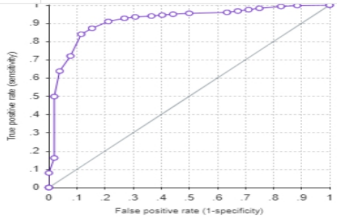
View full log

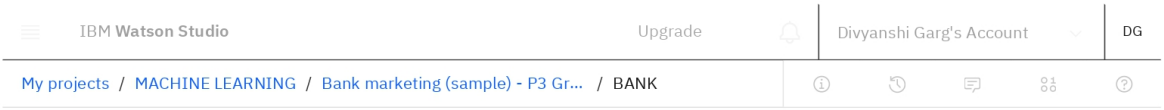
Pipeline leaderboard

	Rank↑	Name	Algorithm	ROC AUC (O...	Enhancements	Build time
⌄	★ 1	Pipeline 3	Gradient Boosting Classifier	**0.899**	HPO-1 FE	00:02:05

My projects / MACHINE LEARNING / Bank marketing (sample)

	Cross validation score	Holdout score
Accuracy	0.898	0.896
Average precision	0.511	0.619
F1	0.449	0.390
Log loss	0.242	0.216
Normalized Gini Coefficient	0.359	0.194
Precision	0.597	0.600





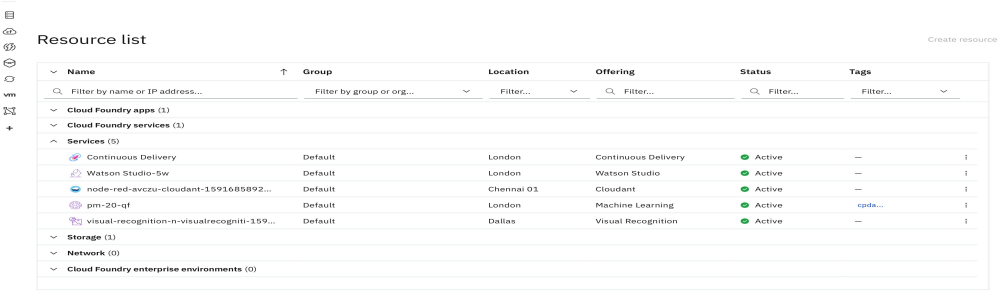
BANK

Overview | Implementation | Test

Deployment

Name	BANK
Type	Web Service
Deployment ID	f7fb2fb6-cdf9-4941-99e8-9737e5aa72d2
Status	Ready
Asset type	Model
Asset name	Bank marketing (sample) - P3 GradientBoostingClassifierEstimator
Machine learning service	pm-20-qf
Created	Jun 15, 2020 6:18 PM
Last modified	Jun 15, 2020 6:18 PM

PREDICTING LIFE EXPECTANCY WITHOUT PYTHON



PREDICTING LIFE EXPECTANCY DEPLOY

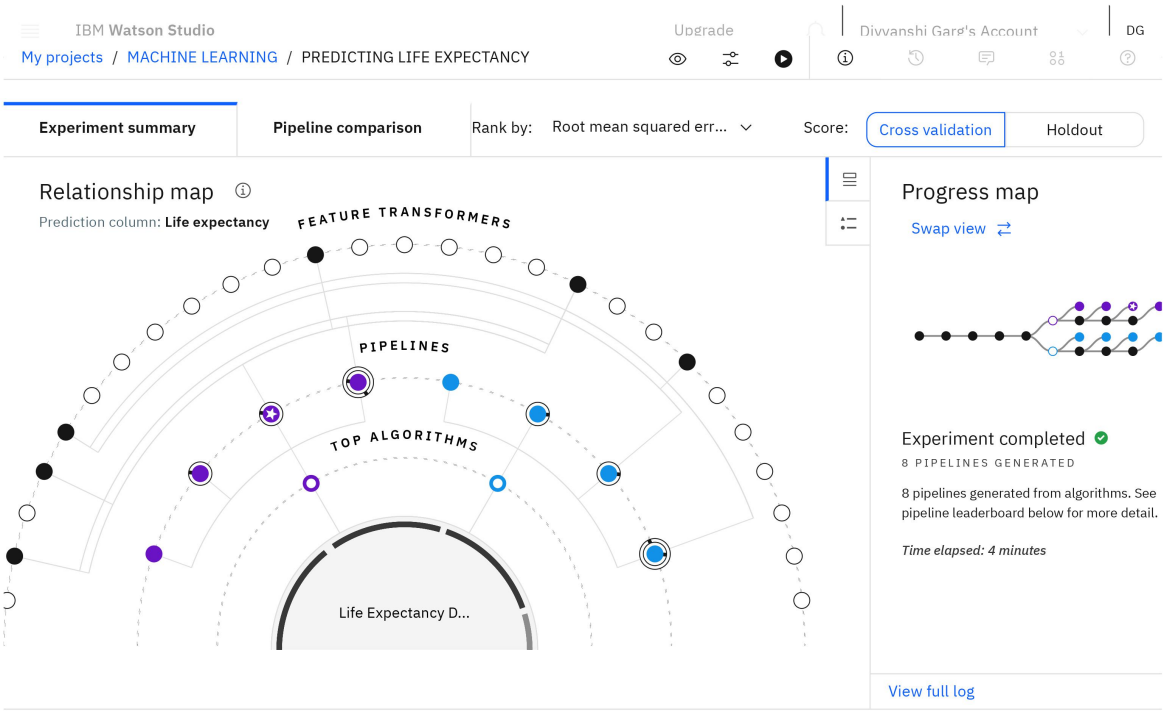
Overview

Deployment

Name	PREDICTING LIFE EXPECTANCY DEPLOY
Type	Web Service
Deployment ID	e909127e-0cc9-4c0e-854c-dffa6c64a14a
Status	Ready
Asset type	Model
Asset name	PREDICTING LIFE EXPECTANCY - P3 ExtraTreesRegressorEstimator
Machine learning service	pm-20-qf
Created	Jun 16, 2020 4:39 PM
Last modified	Jun 16, 2020 4:39 PM

Model

Name	PREDICTING LIFE EXPECTANCY - P3 ExtraTreesRegressorEstimator
ID	1b4cabce-e1de-414f-b893-86391bafedae
Version ID	22fb6178-5c90-40af-86c7-5634774f5e2e



Pipeline leaderboard

Rank	Name	Algorithm	RMSE (Optimiz...	Enhancements	Build time

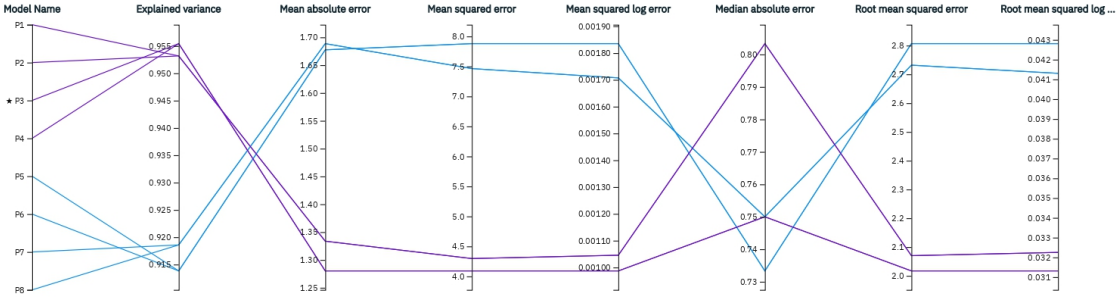
IBM **Watson Studio** Upgrade Divyanshi Garg's Account DG

My projects / MACHINE LEARNING / PREDICTING LIFE EXPECTANCY

Experiment summary | **Pipeline comparison** | Rank by: Root mean squared err... | Cross validation | Holdout

Metric chart

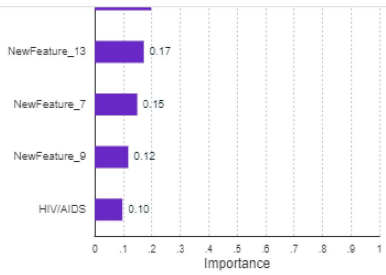
Prediction column: **Life expectancy**



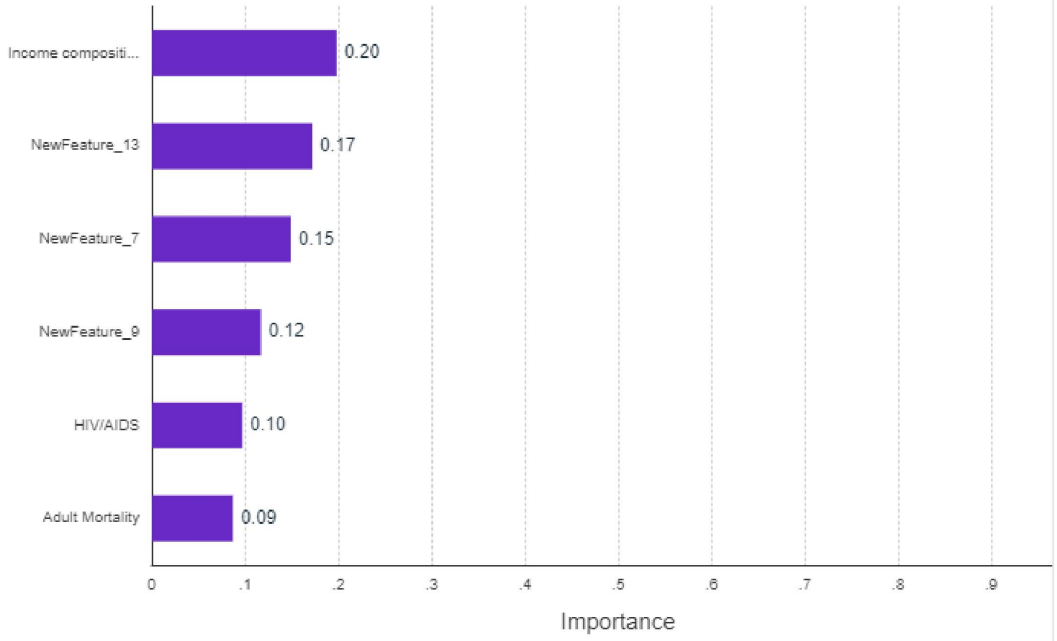
Pipeline leaderboard

Experiment summary | Pipeline comparison | Rank by: Root mean squared err... | Score: Cross validation | Holdout

Explained variance	0.955	0.952
MAE	1.280	1.282
MSE	4.091	4.101
MSLE	0.001	0.001
MedAE	0.750	0.700
RMSE	2.017	2.025
RMSLE	0.031	0.031
R^2	0.955	0.952



>	2	Pipeline 4	Extra Trees Regressor	**2.017**	HPO-1	FE	HPO-2	00:00:38
>	3	Pipeline 1	Extra Trees Regressor	**2.070**	*None*			00:00:01
>	4	Pipeline 2	Extra Trees Regressor	**2.070**	HPO-1			00:00:12
>	5	Pipeline 7	Decision Tree Regressor	**2.732**	HPO-1	FE		00:00:41
>	6	Pipeline 8	Decision Tree Regressor	**2.732**	HPO-1	FE	HPO-2	00:00:09
>	7	Pipeline 5	Decision Tree Regressor	**2.807**	*None*			00:00:01


Income compositi...
0.20
NewFeature_13
0.17
NewFeature_7
0.15
NewFeature_9
0.12
HIV/AIDS
0.10
Adult Mortality
0.09
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
Importance

	Holdout Score	Cross Validation Score
Root Mean Squared Error (RMSE)	2.025	2.017
R^2	0.952	0.955
Explained Variance	0.952	0.955
Mean Squared Error (MSE)	4.101	4.091
Mean Squared Log Error (MSLE)	0.001	0.001
Mean Absolute Error (MAE)	1.282	1.280
Median Absolute Error (MedAE)	0.700	0.750
Root Mean Squared Log Error (RMSLE)	0.031	0.031

PREDICTING LIFE EXPECTANCY DEPLOY

Test

Enter input data

64

thinness 5-9 years

64

Income composition of resources

64

Schooling

64

Predict


```
{
 "predictions": [
  {
   "fields": [
    "prediction"
   ],
   "values": [
    [
     65.78000030517578
    ]
   ]
  }
 ]
}
```

SMARTBRIDGE

Let’s Bridge the gap

Summer Internship report

Predicting Life Expectancy using Machine Learning

06/06/2020-4/07/2020

Submitted by: Divyanshi Garg

INDEX

INTRODUCTION

Overview

The life expectancy dataset for this project is collected from Kaggle. It is a platform for predictive modelling and analytics competitions. Statisticians and data scientist from all over the world post the data and compete to produce the solve dataset problem with best of model. The Machine Learning regression algorithm is applied to the dataset to get best results.

Problem Statement:

Predicting Life Expectancy Using Machine Learning

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. This problem statement provides a way 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.

Purpose:

Built a machine learning model for the prediction of life expectancy.

LITERATURE SURVEY

Existing Problem:

Predicting life expectancy of people around the world on the basis of different set of attributes such as BMI, Country, Adult Mortality etc.

Proposed Solution:

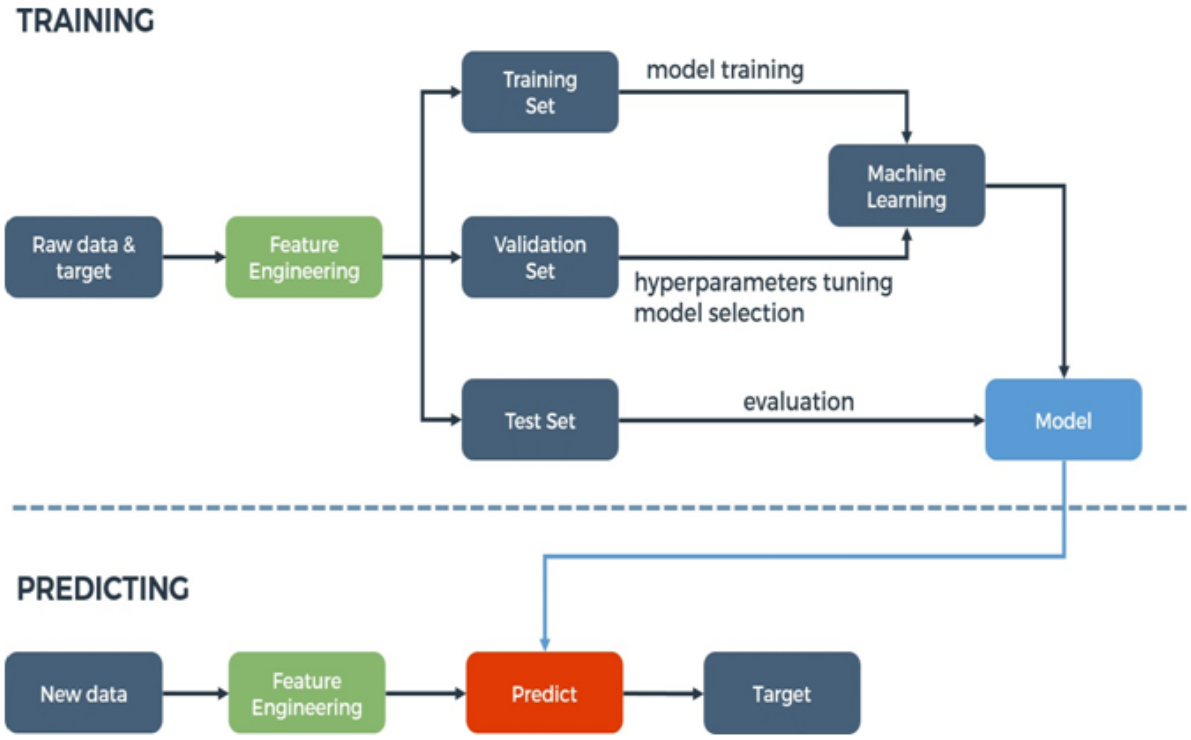
To study at world level, it is required to classify population into groups on the basis of age, gender, region etc. For this and other complex statistics, the new technologies such as Artificial Intelligence, Machine Learning, Neural Networks are available.

To calculate the life span of an individual a system is required that consider environmental, geographic, health and lifestyle attributes. With the help of Machine Learning, regression algorithm it's feasible to analyze the dataset and predict the life expectancy on individual. IBM Cloud platform made the process a much more easier to predict accurate and efficient results.

IBM cloud platform has an in-build AUTO AI that makes it easier for the user to get the results only with loading the dataset. Auto AI split the dataset into training and testing and apply regression to it. After applying regression, RMSE value is automatically calculated which reveal the accuracy of model.

THEORITICAL ANALYSIS

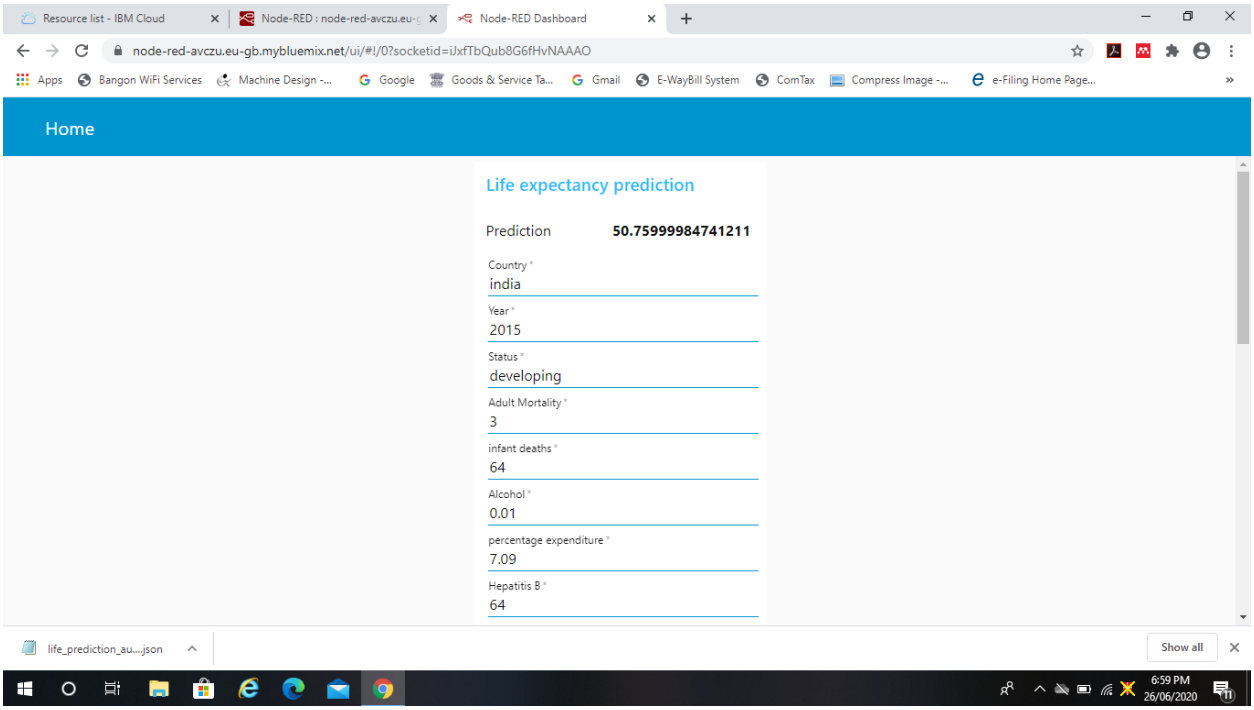
Block diagram



Hardware and Software Designing

An Auto AI model is integrated with node-red flow to predict life expectancy. The node-red flow URL is shared below

https://node-red-avczu.eu-gb.mybluemix.net/ui/#!/0?socketid=0N8o0NQ1Ar8bb0cTAAAR

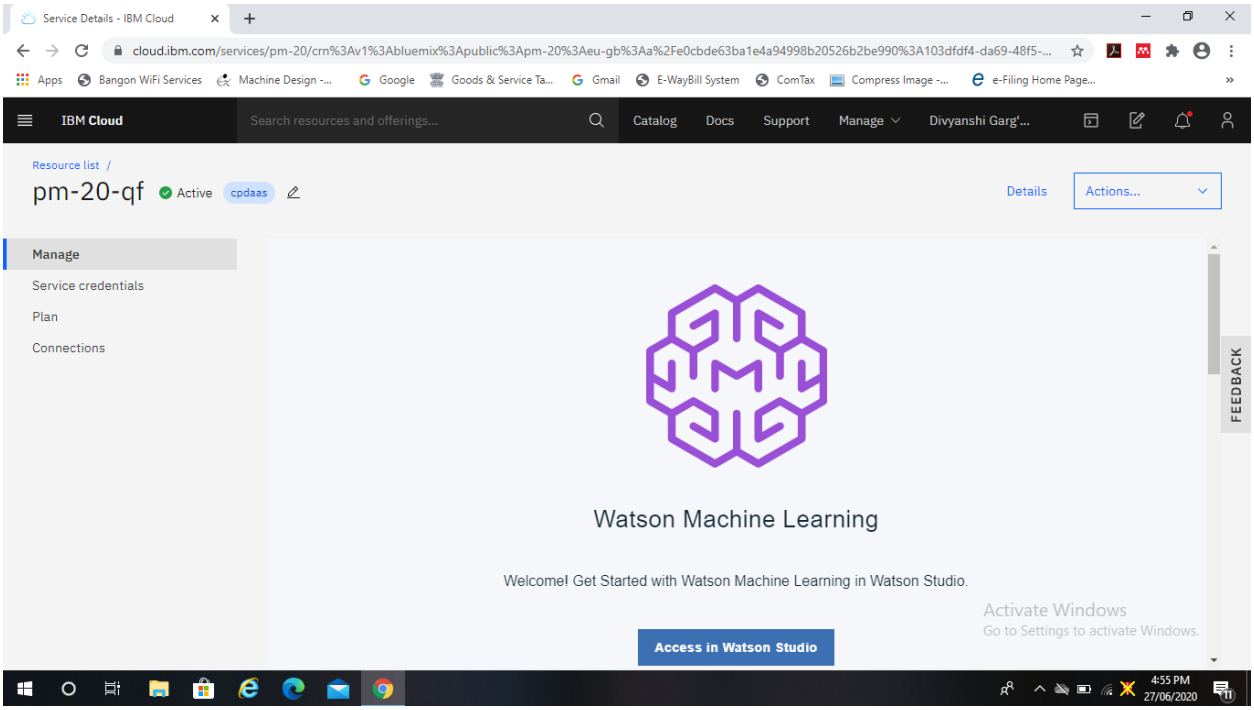


Experimental Investigations

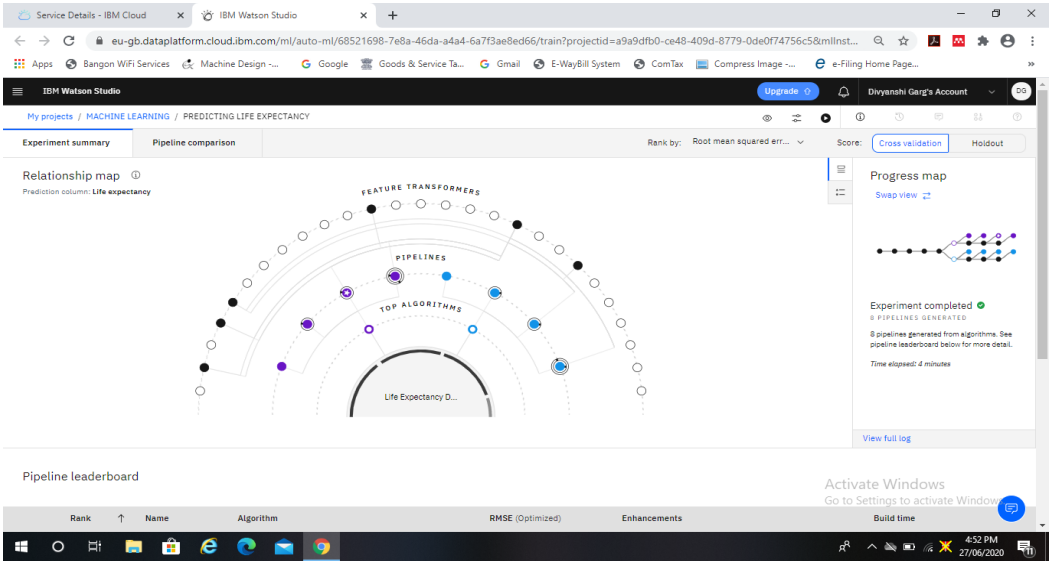
- Collect the dataset from Kaggle.

https://www.kaggle.com/kumarajarshi/life-expectancy-who

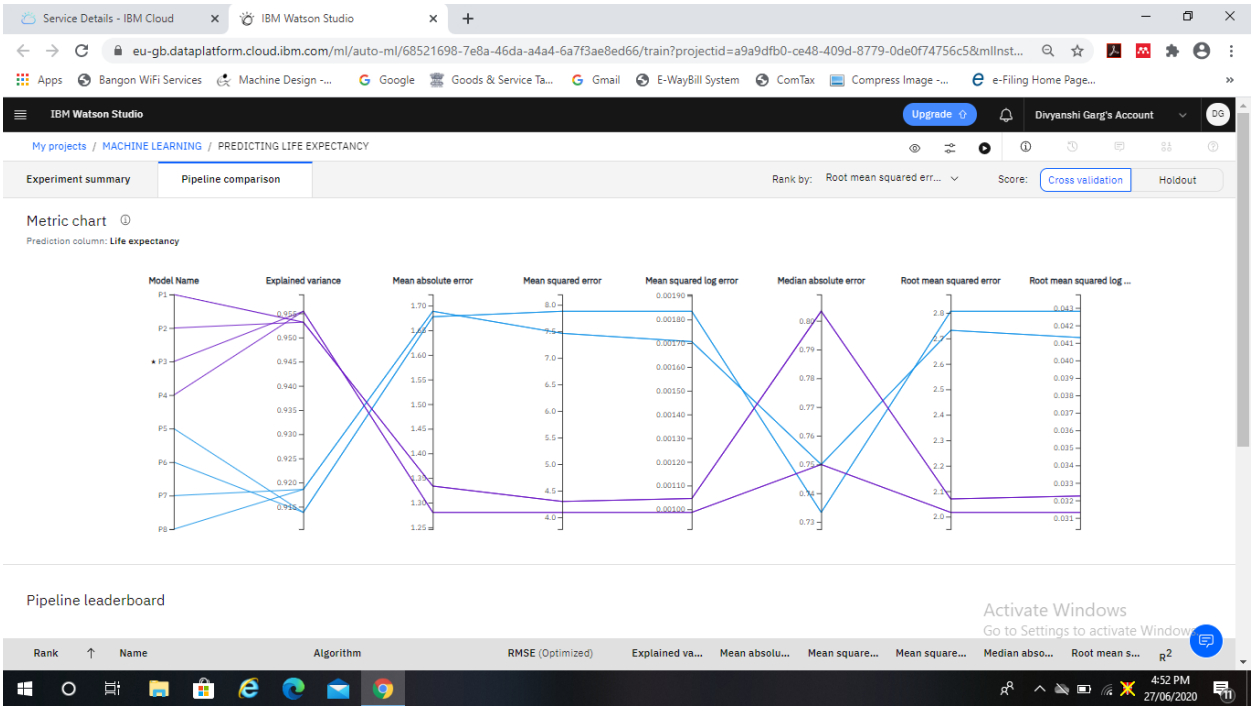
- Create a Machine Learning Service in IBM Cloud to build a model.

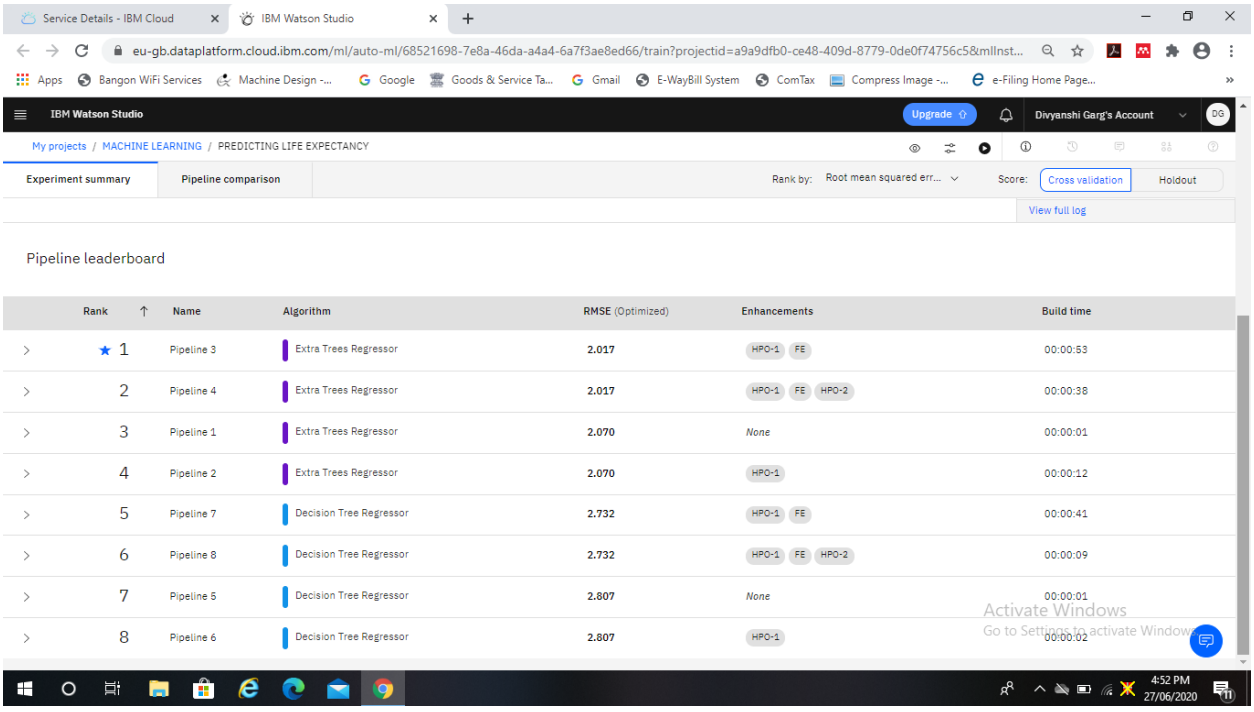


- Build an Auto AI model to predict life expectancy.

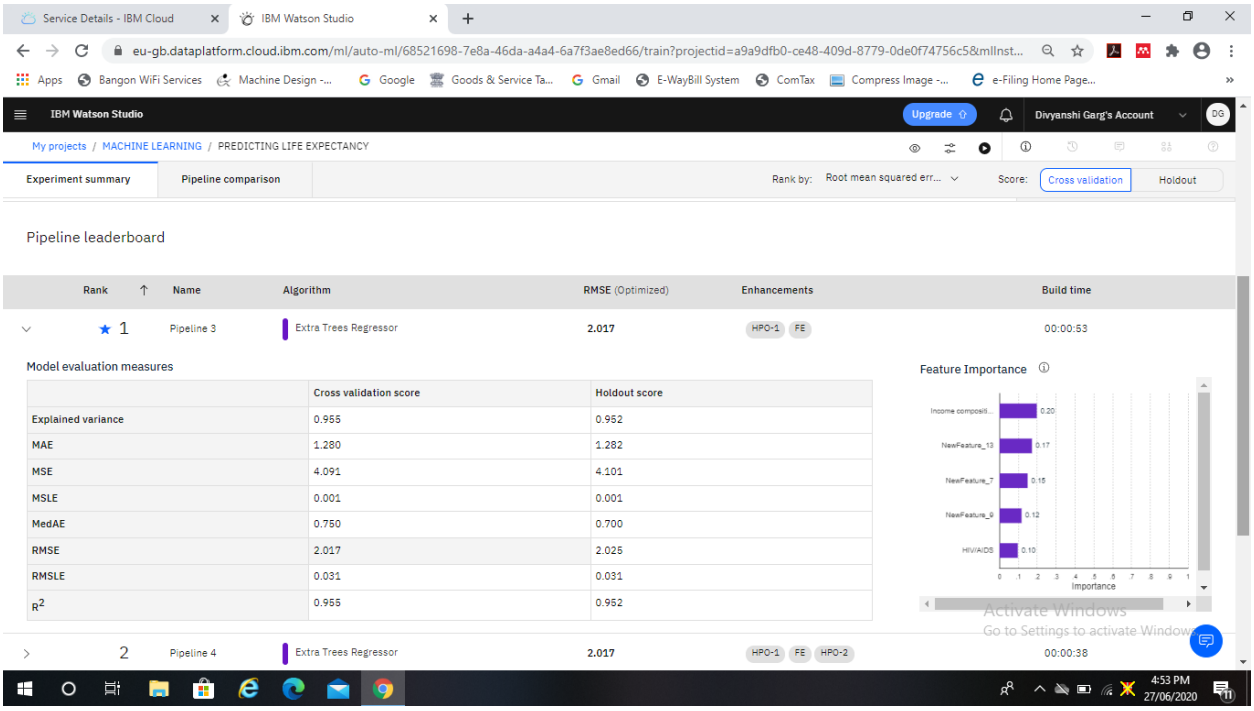


- Compare the pipelines created by model and choose the one that best fits the model.

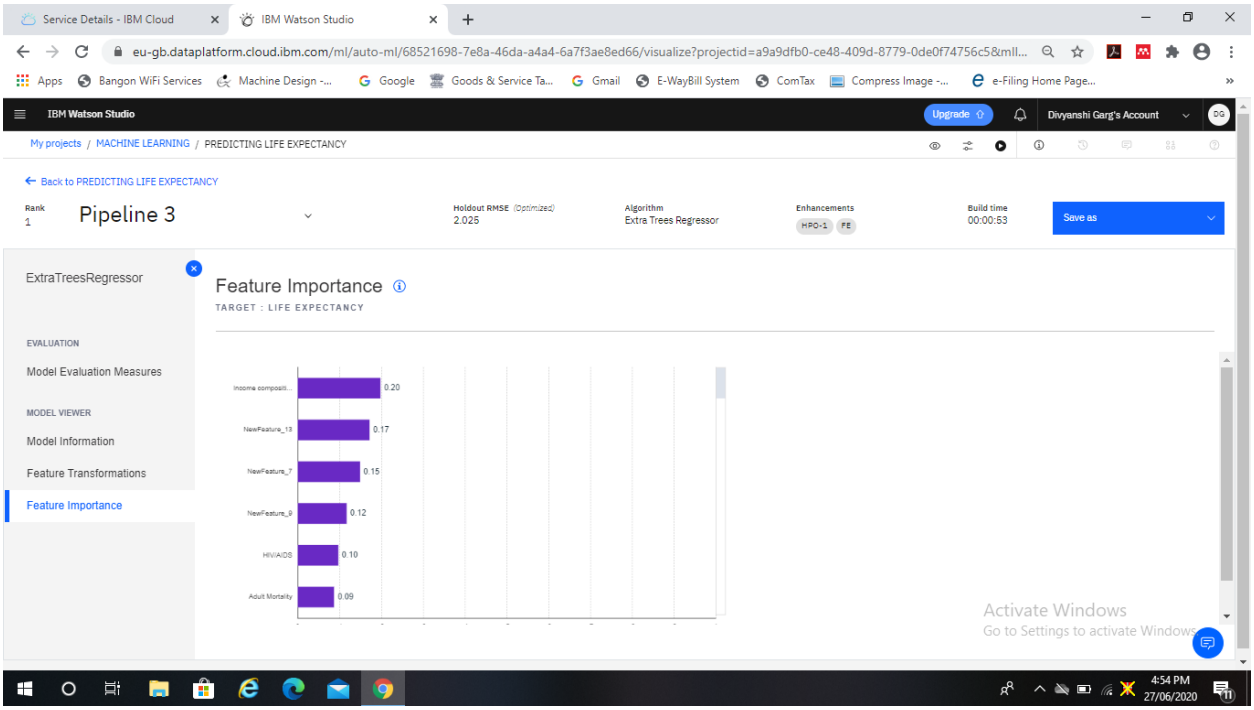




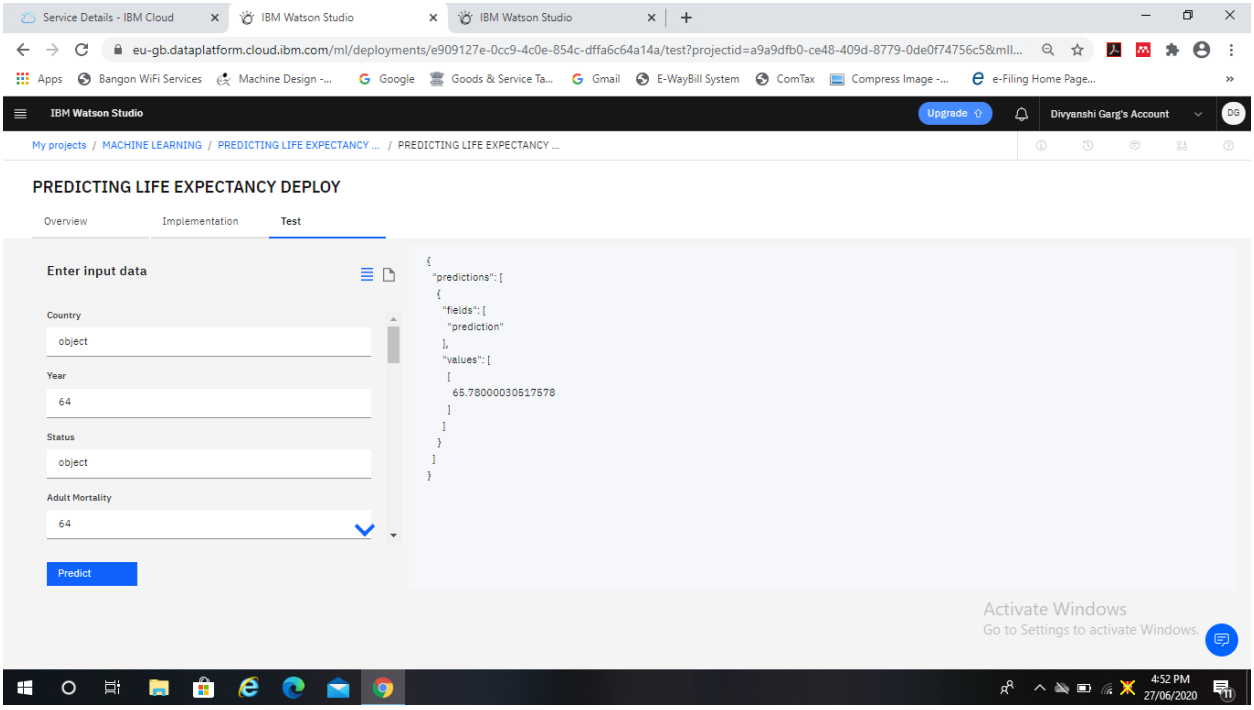
- Look for the RMSE and other values to know the accuracy of model.



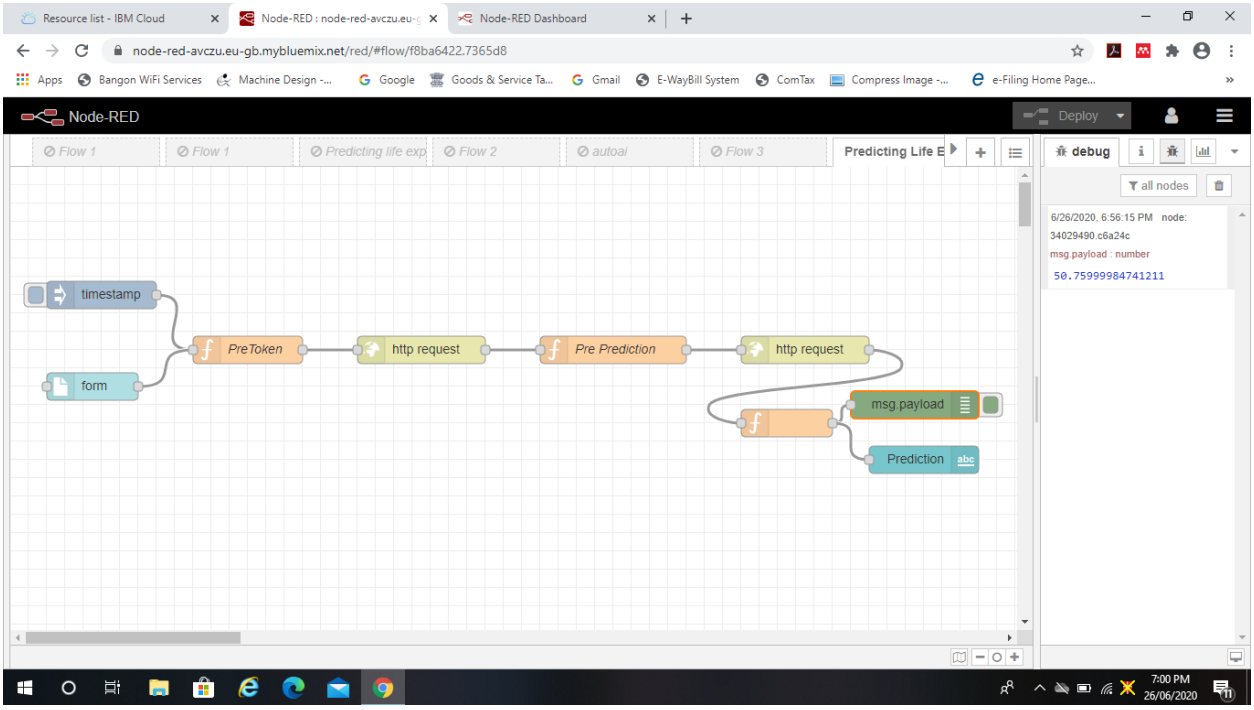
- Study the importance of attributes in prediction of life expectancy to know the influence of them on predicited results.



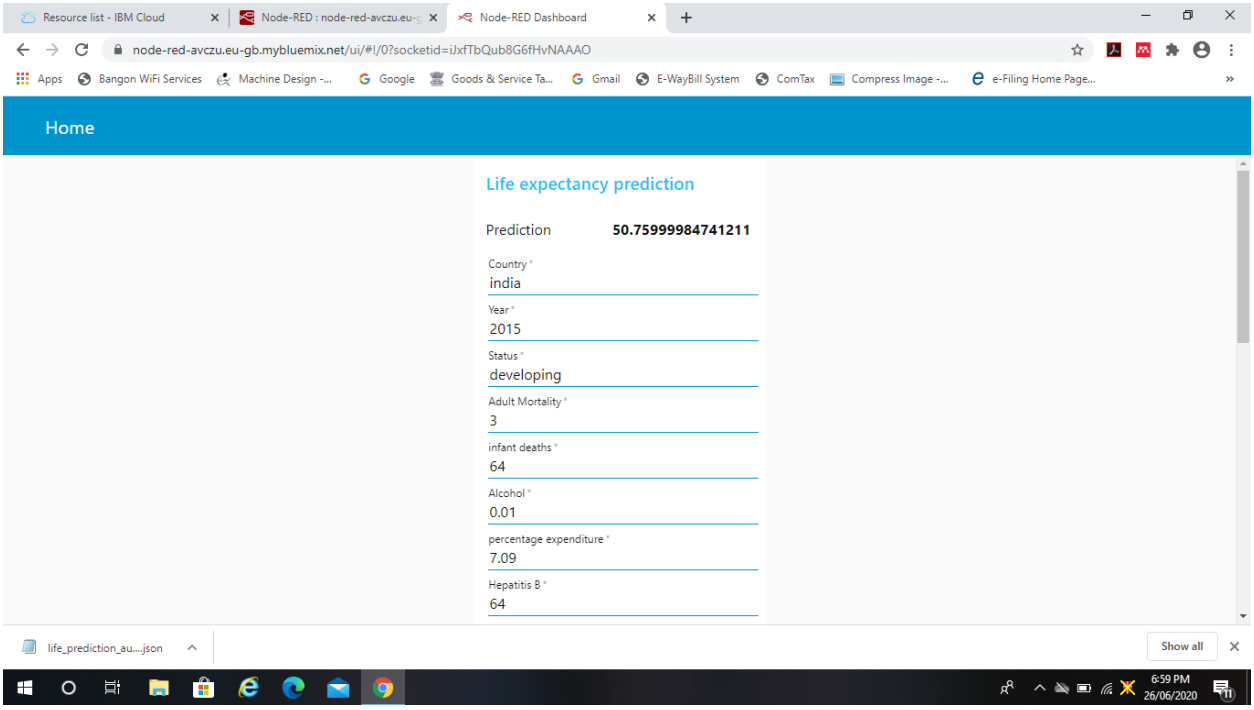
- Test the model by assigning values to the attributes and look for prediction.



- Integrate Auto AI model with node-red flow.

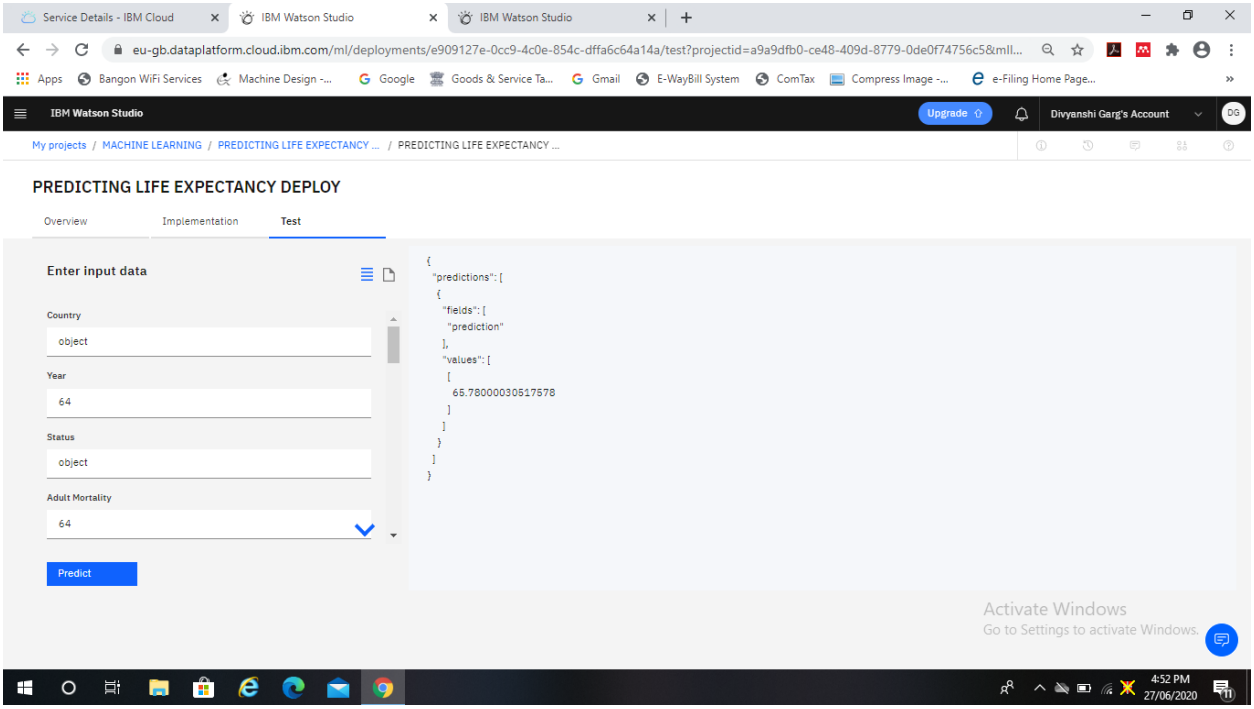


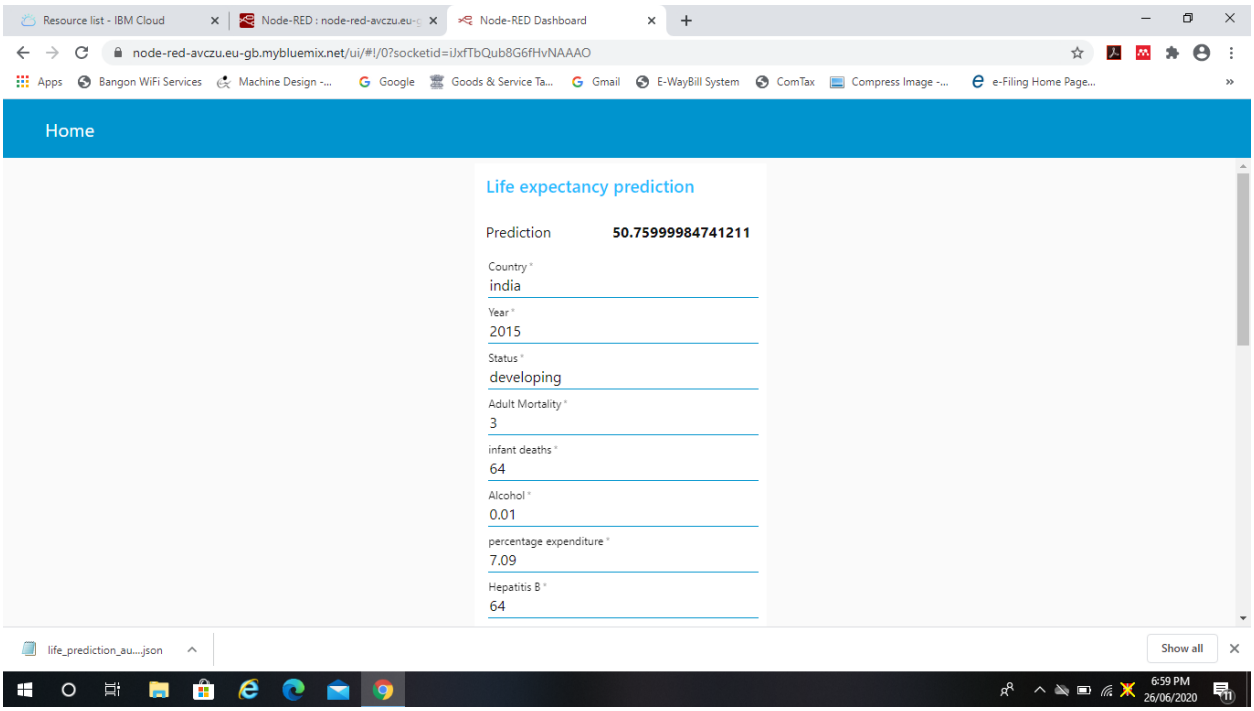
- Get the output in User Interface.



RESULTS

- Prediction of life expectancy on the basis of different distinct features.





- Income composition is the most important feature that plays a vital role in predicting the life expectancy. It's importance is the highest among other features. It depicts that an individual with higher income tends to have a great lifesytle and so the life span.
- A positive correlation between‘Schooling’ and ‘LifeExpectancy’ is observed.
  Education system of rich countries is well established and have high life expectancy rate.
- A positive correlation between ‘GDP’ and ‘Life Expectancy’ is observed.
  As the countries with good healthcare, education, infrastructure facilities are more likely to have high life expectancy rate.

Advantages

- Life expectancy predictions have the potential to be beneficial to individuals, health service providers and governments. It would make people more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.
- Life expectancy can be estimated at any age, e.g. life expectancy at 65 years. Gives more weight to deaths at younger ages.
- Life expectancy has been used nationally to monitor health inequalities. the problems caused can be tackled with the joint efforts made by people and it is urgent for human beings to do the best to save the environment and find out solution to the difficulties caused by the larger and larger population on the planet.

Disadvantages

- At smaller geographies may be influenced by nursing homes in the area.
- A younger death does not necessarily have any more impact on the rate than an older death. It is quite possible that the death of a person aged 70 years may increase the rate to the same extent as an infant death.

Application

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 broadly: facilitates Advance Care Planning. Evidence based studies indicate the factors based on longetivity of an individual.

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.

Income Composition has most influenced the life span predictions. The accuracy of model found to be 95% which is a great percentage and from the results it can be concluded that the model best fits the dataset. On the other hand, the benefits both people and the world can get from the phenomenon of people living longer are irreplaceable and undeniable. It is the truth that longevity is a symbol of civilization and better life.

FUTURE SCOPE

Life expectancy inequality is forecast to continue to rise across districts, however, with present and future inequalities partly related to district deprivation and partly associated with variation within deprivation quintiles, especially within the deprived quintiles. Furthermore, I found that life expectancy varied more in the more deprived quintiles, perhaps because deprived communities are more vulnerable to factors that affect health and longevity, but vary independently of deprivation.

Most of the gains in longevity will be in those older than 65 years of age, and are, hence, highly relevant for planning pensions and health and social services. We also forecast that the closing of the female–male life expectancy gap will continue steadily. This narrowing will occur because death rates in middle-aged and old age are estimated to decrease more slowly in the future than they did in the past, perhaps partly due to accumulation of risks from smoking in middle-aged and old age individuals.

BIBILOGRAPHY

- Create a Kick Off Meeting Agenda

https://www.allbusinesstemplates.com/download/?filecode=2KBA4&lang=en&iuid=9f9faa69-9fab-40ee-8457-ea0e5df8c8de

- GitHub account

https://github.com/

https://www.youtube.com/watch?v=7YUTc4Cigc8&feature

=youtu.be

- Work with Zoho Writer

https://www.zoho.com/writer/help/working-with-text.html

- IBM Academic Initiative account

https://my15.digitalexperience.ibm.com/b73a5759-c6a6-4033-ab6b-d9d4f9a6d65b/dxsites/151914d1-03d2-48fe-97d9-d21166848e65/

- IBM Service

https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L

- Watson Studio

https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2

- Build Auto AI model

https://bookdown.org/caoying4work/watsonstudio-workshop/auto.html#add-asset-as-auto-ai

- Create Node-red application

https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/

https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be

https://www.w3schools.com/howto/howto_make_a_website.asp

https://www.youtube.com/

APPENDIX

Source Code

https://node-red-avczu.eu-gb.mybluemix.net/ui/#!/0?socketid=0N8o0NQ1Ar8bb0cTAAAR