PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING
BY - SHREYA TIWARI

Project Scope:

Project Title:
Predicting Life Expectancy using Machine Learning

Purpose:
- We can predict life expectancy of peoples in a country using machine learning with the help of this. In any country there are many factors that decrease the life of a person and also decreases the average life of peoples in any country like many people die from starvation, diseases etc. with machine learning we can predict their country's life expectancy.
- When any country knew the factors that are decreasing the life expectancy of their people, they can try to reduce those factors and increase life expectancy.

Schedule:
1st week:
1. Project planning and kick off
2. Explore IBM cloud platform
3. Explore IBM Watson services

2nd week: Introduction to Watson Studio
3rd week: Predicting Life Expectancy with Python
4th week: Predicting Life Expectancy with Python

Team:
Name: Shreya Tiwari

1. ## Introduction
   ### 1.1 Overview
   In this project, I have created a Machine Learning Mode using the data provided by World Health Organization(WHO) to evaluate the life expectancy for different countries in years, The data offers time-frame from 2000 to 2015.

   ### 1.2 Purpose
   Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication helps to anticipate the procurement of health care services and facilities. Thus, a life expectancy predicting model will be a great help to the health-care

sector as this will help in finding the loop-holes in the health-care sector.

2. <u>Literature Survey</u>

2.1 Existing Problem

Predicting the lifespan of people, or their Personal Life Expectancy(PLE) would greatly alter our lives. On the other hand, it will definitely help in optimizing an individual's health or the health services that they receive.

2.2 Proposed Solution

In this projected, I have tested two machine learning algorithms and I have deployed the machine learning model which had the best result. The machine learning algorithms that I tested are :

- Linear Regression
- Random Forest Regression

After deploying the model, I have created a Node-Red Application as User Interface(UI) which the user uses to predict the life expectancy.
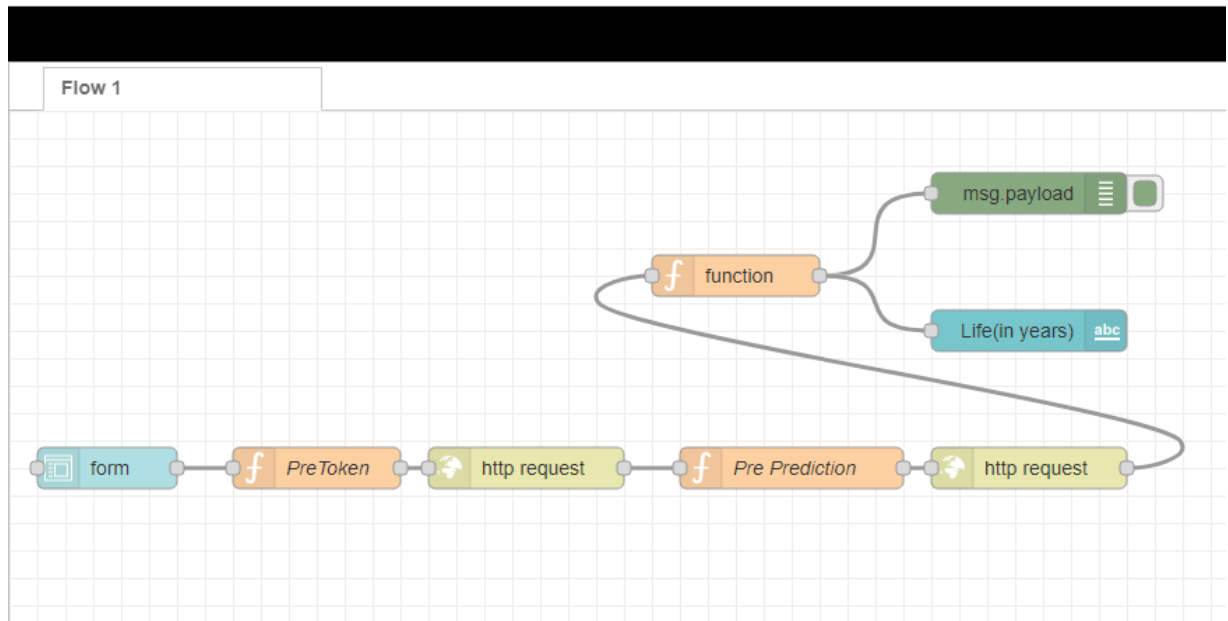
3. <u>Theoretical Analysis</u>

3.1 Block Diagram
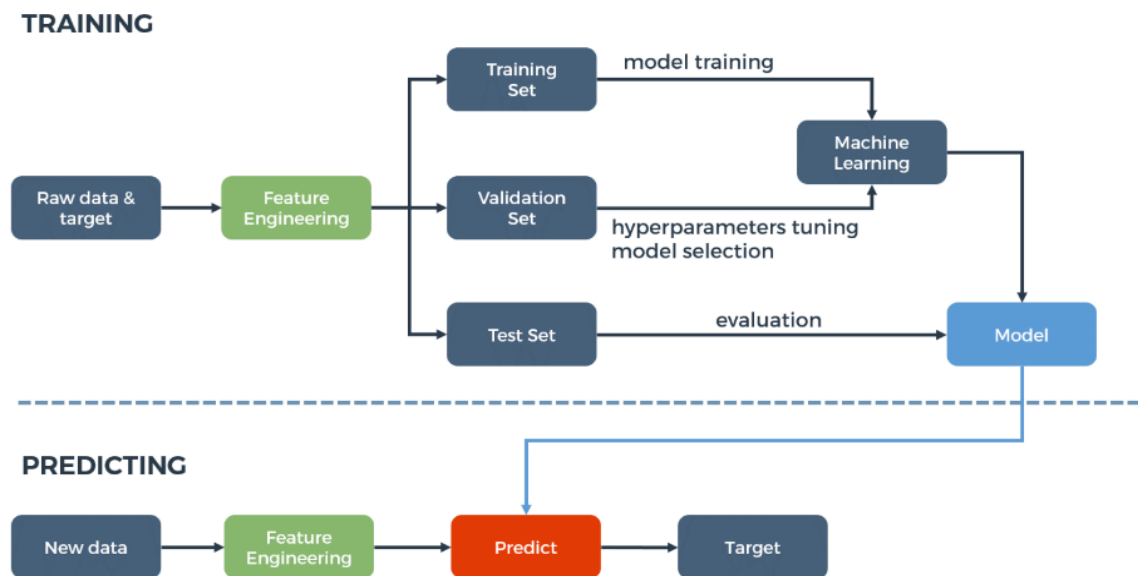
3.2 Hardware / Software Designing

This project is implemented using IBM Node-Red Application. The Node-Red Flow for this project is given below.

The link for the Node-Red Application:
https://node-red-peexu-comeon.eu-gb.mybluemix.net/ui/#!/0?socketid=TyXYeBO
XxsQa8IOEAAAI

4. **Experimental Investigations**

- The data for this project is collected from Kaggle, the kaggle link for the data is:
  https://www.kaggle.com/kumarajarshi/life-expectancy-who
- Following are the steps taken in order to implement this project:

1. Created an IBM Cloud Account.
2. Created a Watson Machine Learning Service.
3. Created a new project on Watson Studio named - 'Predicting Life Expectancy'.
4. Added the data that is downloaded from Kaggle to the project.
5. Create a blank Jupyter Notebook.
6. Loading the data in notebook.
7. Importing all the required python libraries that are required for the project.
8. Preprocessing the data for the missing / null values and converting the string data to the integer.
9. Observing the relation between data using a heat map.
10. Train and test different machine learning algorithms.
11. Choose the model with best accuracy and deploy the mode.
12. Get the Scoring Endpoint.
13. Create a node red application.
14. Integrate the node-red application with the machine learning model.

5. **Flowchart**
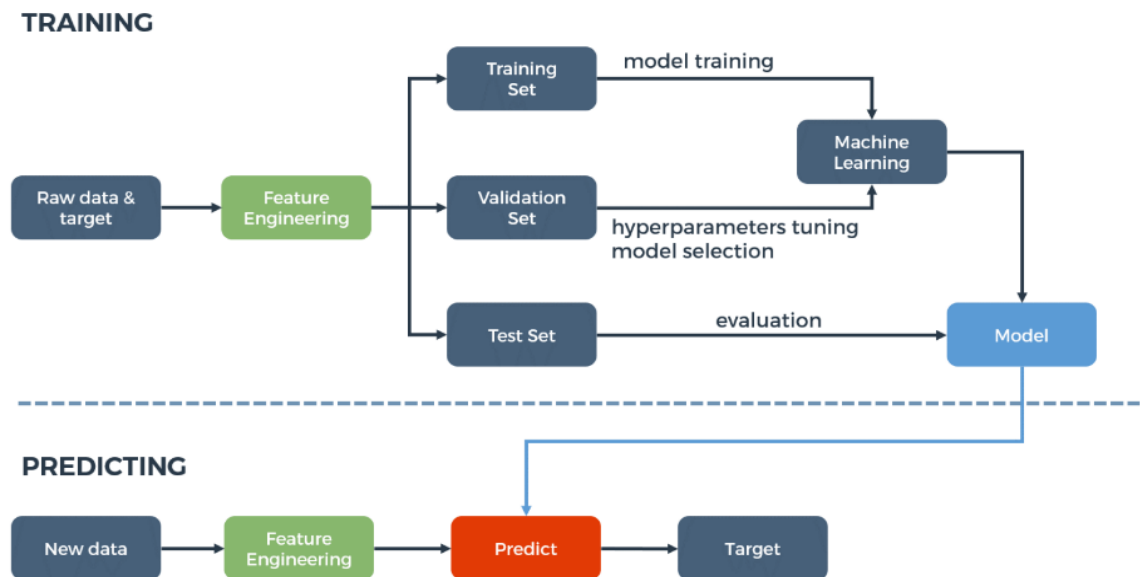
6. **Result**

The R^2 score and the root-mean-squared error for the Linear Regression Model is :

```
In [166]: reg_all = LinearRegression()
          reg_all.fit(X_trainset, y_trainset)
          y_pred = reg_all.predict(X_testset)
          print('R^2: {}'.format(reg_all.score(X_testset, y_testset)))
          print('rmse: {}'.format(np.sqrt(mean_squared_error(y_pred, y_testset))))

          R^2: 0.8427037778089941
          rmse: 3.6216271382784853
```

The R^2 score and the root-mean-squared error for the Random Forest Regression
Model is :

```
In [190]: random_forest_reg2 = RandomForestRegressor(max_depth=5, random_state=0)
          random_forest_reg2.fit(X_trainset, y_trainset)
          y_pred = random_forest_reg2.predict(X_testset)
          print('R^2: {}'.format(random_forest_reg2.score(X_testset, y_testset)))
          print('rmse: {}'.format(np.sqrt(mean_squared_error(y_pred, y_testset))))

          R^2: 0.9324113282844745
          rmse: 2.3740045419314146
```

Hence, I deployed the Random Forest Regression Model as it had better result.

7. Advantages And Disadvantages

ADVANTAGES:
- This project will be really helpful for the government to make changes in the health care unit of the country.
- A life prediction model would really help the researchers who are researching on the affect of a disease on an entire country.
- A life predicting model will also help the health care unit of the country to identify which disease is more threatning and causing more human deaths.

DISADVANTAGES:
- The life pexpectancy redicting algorithm may sometimes give wrong information as the current accuracy(93%) of the algorithm is still low.
- This program may fail if the input data is from a very different distribution from the training data.

8. Applications

The project has great application in:
- Health - Care Industry.
- Government making new health policies.
- Some researchers reseaching about the life expectancy in a country.

9. Conclusion

Predicting life expectancy with good accuracy is not an easy thing to achieve, but the research that I did in my project suggests me that with the help of Machine Learning we can do this task. This project has potential for real-life applications, such as supporting timely recognition of the right moment to start advance care planning.

10. Future Scope

I tried to get the best accuracy in results, but with more deep research and after testing more different Machine Learning Algorithms, I am sure that one can get better results. We can further add more data or we can apply neural networks and deep learning to get more accurate results.

11. Bibliography

https://www.youtube.com/watch?v=w3jLJU7DT5E&feature=youtu.be
https://www.zoho.com/writer/help/working-with-text.html
https://www.ibm.com/cloud/get-started
https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
https://nodered.org/
https://github.com/watson-developer-cloud/node-red-labs
https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be
https://www.w3schools.com/howto/howto_make_a_website.asp
https://www.ibm.com/watson/products-services
https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/
https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html
https://www.kaggle.com/kumarajarshi/life-expectancy-who