**Project Report**

**INTRODUCTION:**

1.1  Overview:

**Life expectancy** is a statistical measure of the average (see below) time an organism is expected to live, based on the year of its birth, its current age, and other [demographic](https://en.wikipedia.org/wiki/Demographic) factors including gender.

                       To demonstrate how to build a regression model in Python, I used the ‘Life Expectancy (WHO) dataset on Kaggle [here](https://www.kaggle.com/kumarajarshi/life-expectancy-who). My goal was to create a model that could predict the average life expectancy of a person in a given country on a given year based on a number of variables.

                    The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they haven’t been trained. Four algorithms have been used:

Linear Regression  
Ridge Regression  
Lasso Regression  
ElasticNet Regression  
Linear Regression with Polynomic features  
Decision Tree Regression  
Random Forest Regression

1.2  Purpose:

**Life expectancy** is a measure that is often used to gauge the overall health of a community. **Life expectancy** at birth measures health status across all **age** groups. Small increases in **life expectancy** translate into large increases in the population.

**Life expectancy** is a statistical prediction for how long a person will live. Based on actuarial science, **life expectancy** takes into account several individual-level as well as population-level factors to arrive at a figure.

**LITERATURE SURVEY**

2.1 Existing Problem:

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.

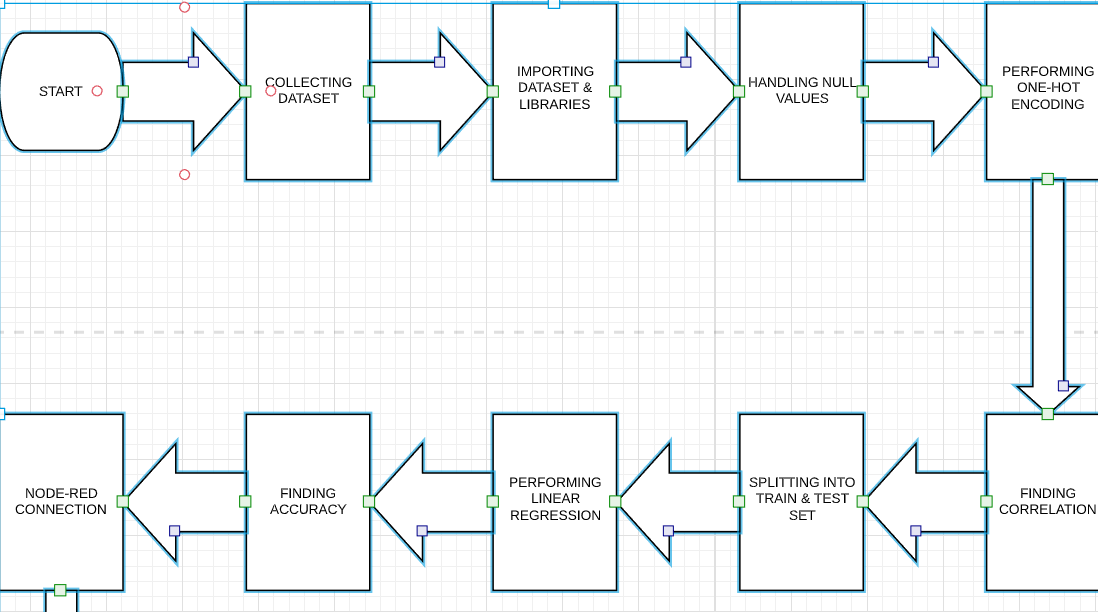
2.2 Proposed Solution:

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The data offers a timeframe from 2000 to 2015. The data originates from here: https://www.kaggle.com/kumarajarshi/life-expectancy-who/data The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they haven’t been trained. Four algorithms have been used:

Linear Regression  
Ridge Regression  
Lasso Regression  
ElasticNet Regression  
Linear Regression with Polynomic features  
Decision Tree Regression  
Random Forest Regression

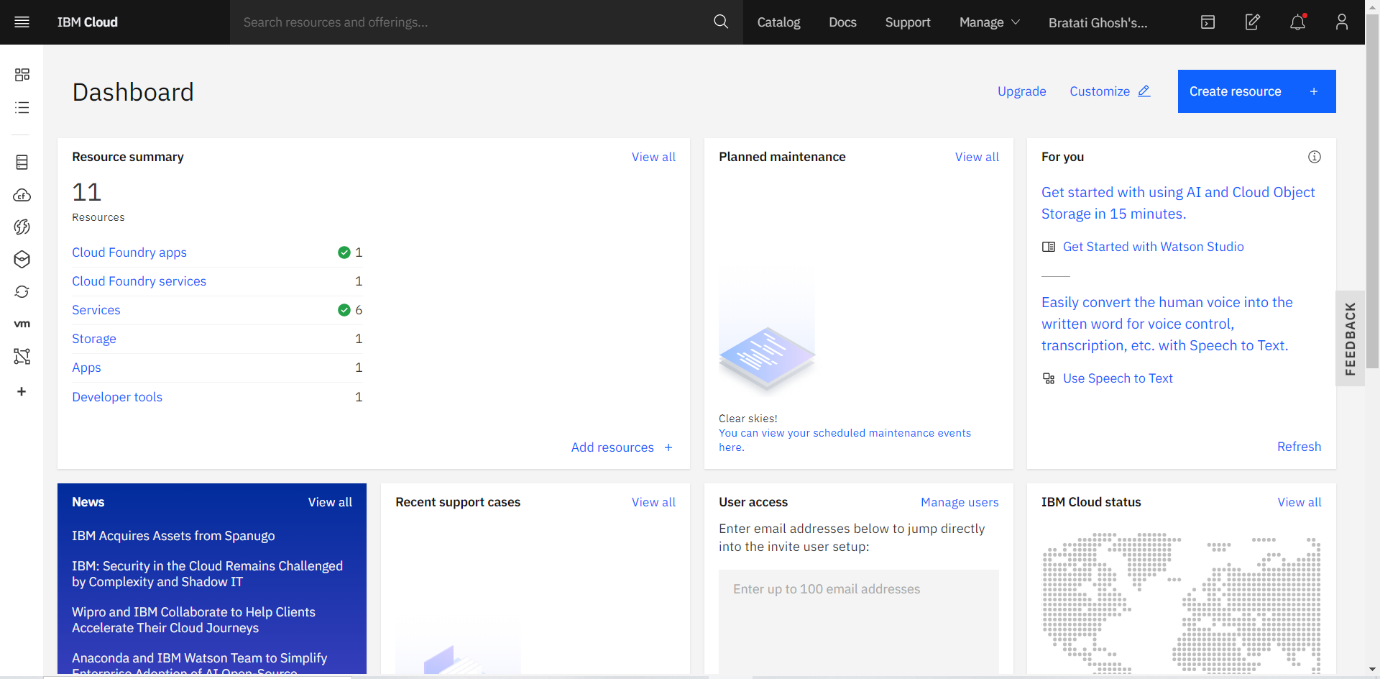
**Theoritical Analysis:**

3.1 Block Diagram:

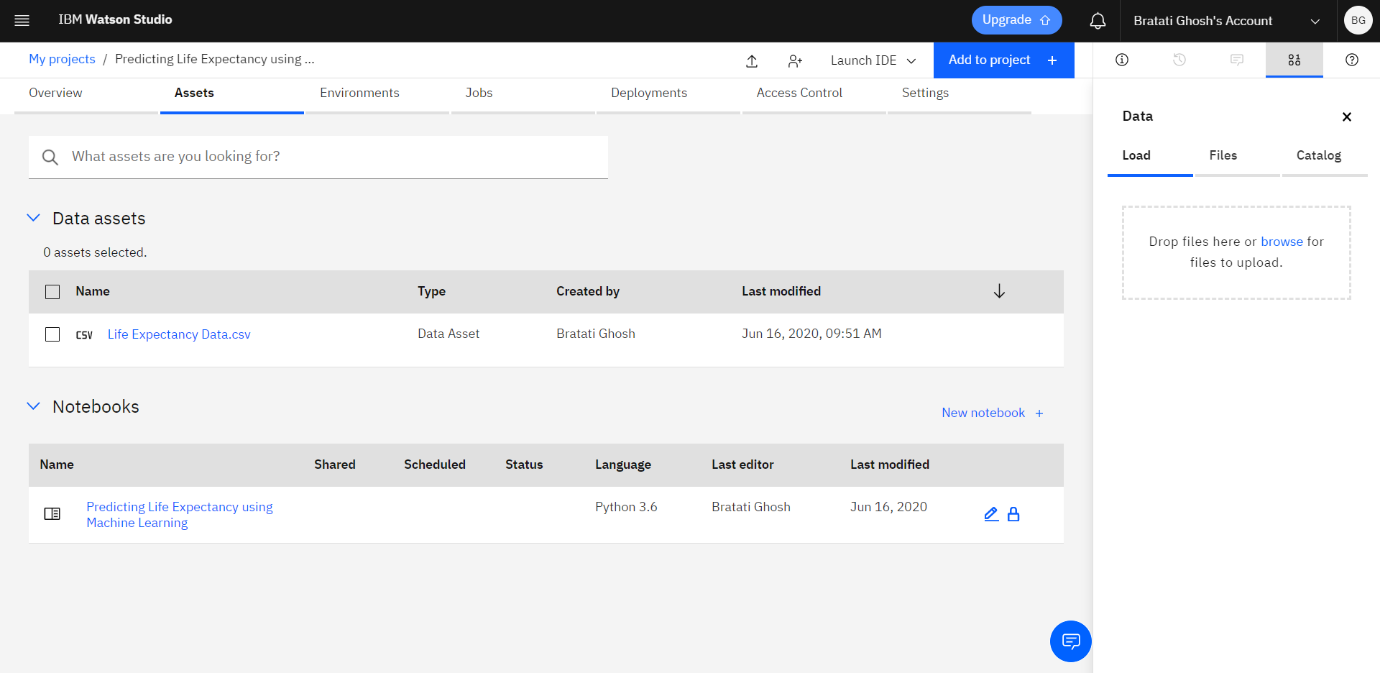


3.2 Hardware And Software Desing:

1) Creating IBM cloud account



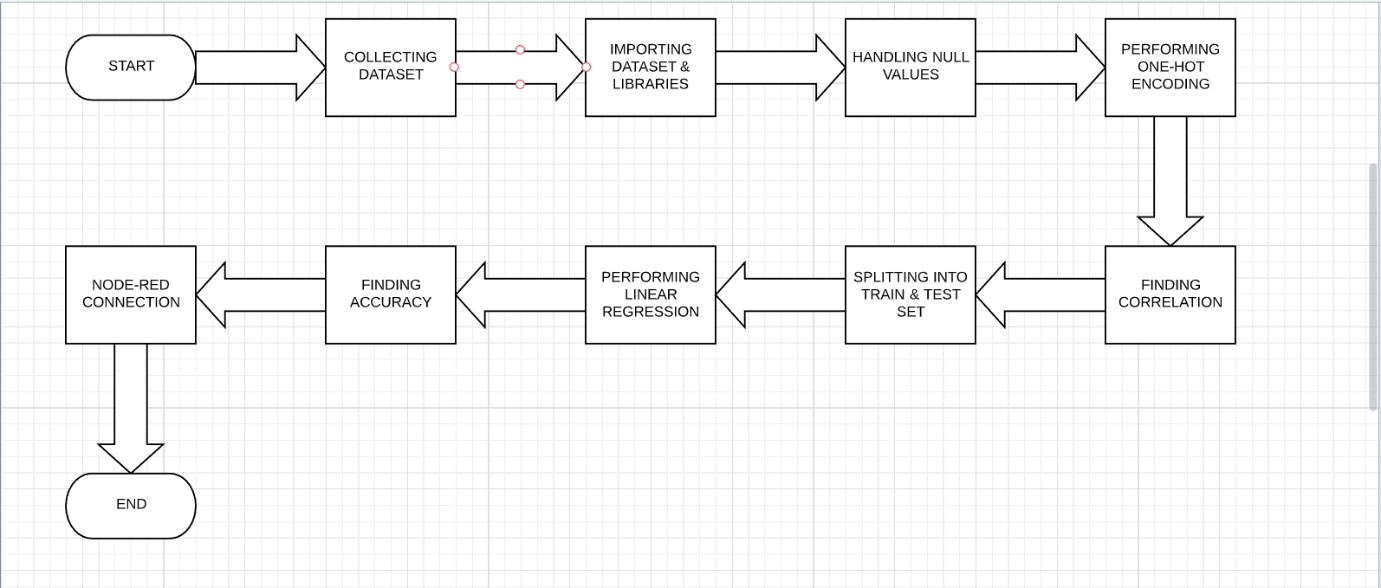
2)Creating Watson Studio



**Experimental Investigation:**

 Here, we use simple linear regression technique to predict which gives a good r-square value and a less MSE value.

**Flowchart:**



**Result:**

 1. R square on the test data of 92%

2. MAE of 2.32

       3.   MSE of 9.80

**Advantage And Disadvantage:**

**Advantages**:

**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.

**Disadvantages**:

At smaller geographies may be influenced by nursing homes in the area.

**conclusion:**

After comparing all the algorithms we can conclude the Linear Regression offer which are the same:

1. Best Parameters: {‘alpha’: 0, ‘max\_iter’: 10}

2. R square on the test data of 92%

3. MAE of 2.32

4. MSE of 9.80

**Bibliography:**

The following sources have been used:

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

2. Introduction to Machine Learning with Python by Andreas C. Müller & Sarah Guido

3. Labs of the couse

4. stack overflow

5. Lectures of the course

**Appendix:**

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