SMART BRIDGE

Let's Bridge the gap

Summer Internship Report

llSPS-INT-2048-Predicting-Life-Expectancy-using-Machine-Learning

Submitted By

Maneesh D

Preface

This report is the work I have done during the online summer internship at the **Smart Bridge Educational Services**. I did the project "**Prediction of Life Expectancy Using Machine Learning"** with the help and the supervision of the mentors of Smart Bridge. This report will give the overview of the work done during the course of the internship.

Maneesh D

Index

1. Introduction

Problem Statement- Predicting Life Expectancy Using Machine Learning

Problem Description :

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.

Life expectancy is a statistical measure and predicting life expectancy helps to determine the course of treatment, managing health care services and facilities, help in planning, managing resources, care planning improves the quality of the final phase of life by simulating doctors to explore the preference for end of life.

2. LITERATURE REVIEW:

Predicting the **lifespan** of people, or their "Personal **Life Expectancy**" (PLE) would greatly alter our lives. On one hand, it may have benefits for policy making, and help optimize an individual's health, or the services they receive.

2.1 Existing Solution:

- In our regular system, there are some problem arise because whole concepts depends upon morbidity and mortality like smoking, alcohol, consumption, overweight and others health issues.
- From previous researches, we take a data-set from 2000-2015 and applied regression techniques.

2.2 Proposed Solution:

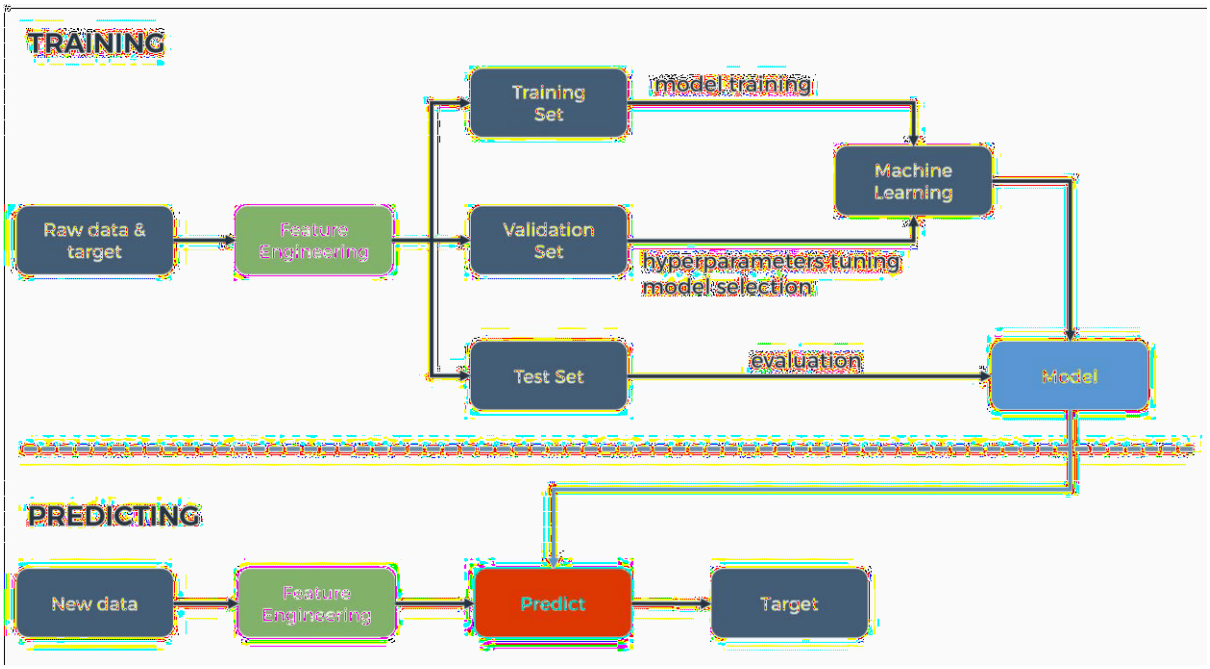
For our problem we have a dataset of different country consist of various factors, some important factors such as HIV, Hepatitis B, Polio, Diphtheria are considered.

The data set we considered related to health factor of 193 countries and has been collected from WHO data repository.

In our project we use some immunization factors, morality factors, economic and social factors to predict life expectancy using Machine learning model.

3. THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware and Software Requirement:

a) Project Requirement: Python, IBM cloud, IBM Watson

b) Functional Requirement: IBM cloud

c) Technical Requirement: ML Watson Studio, Node-Red

d) Software Requirement: Watson Studio, Node-Red

4. EXPERIMENTAL INVESTIGATIONS

A).Choose a project idea:

a. Predicting life expectancy of a Country

B). Conducting a background research

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

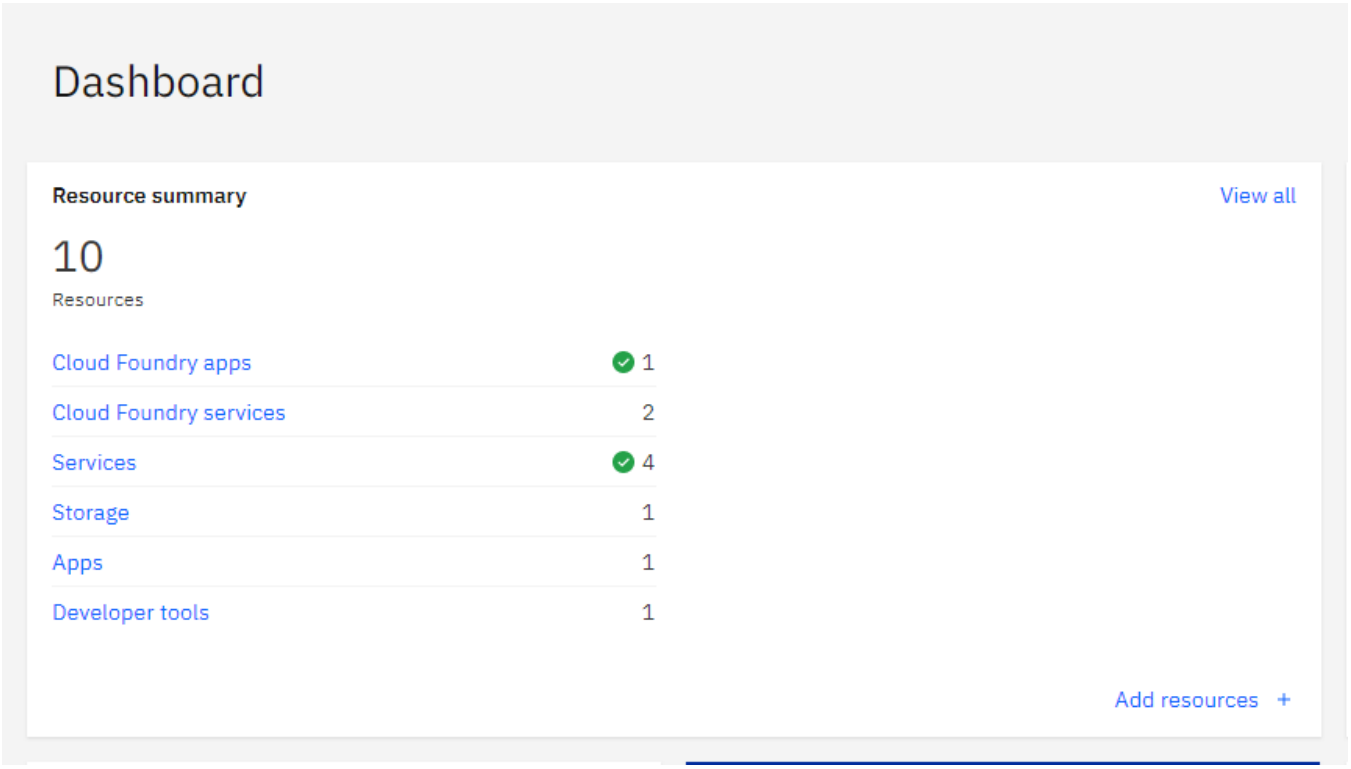
C). Some important Factors Which I have used are:

I. Adult Morality
II. Infant deaths
III. Alcohol
IV. Percentage Expenditure
V. Hepatitis B
VI. Measles
VII. BMI
VIII. Under five deaths
IX. Polio
X. Total expenditure
XI. Diphtheria
XII. HIV/AIDS
XIII. GDP
XIV. Populations
XV. Thinness 10-19 years
XVI. Thinness 5-9 years
XVII. Income Expenditure
XVIII. Schooling

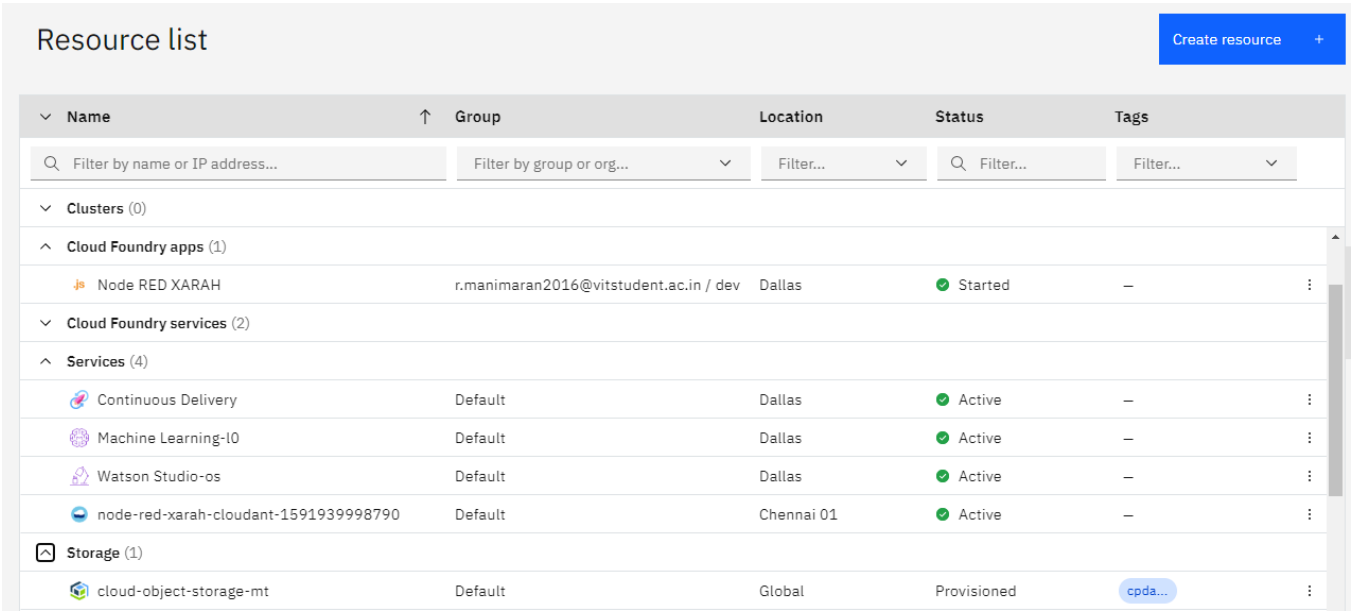
D). Finding most suitable algorithm: Linear Regression gives me the highest accuracy of 82% .

E). Steps:

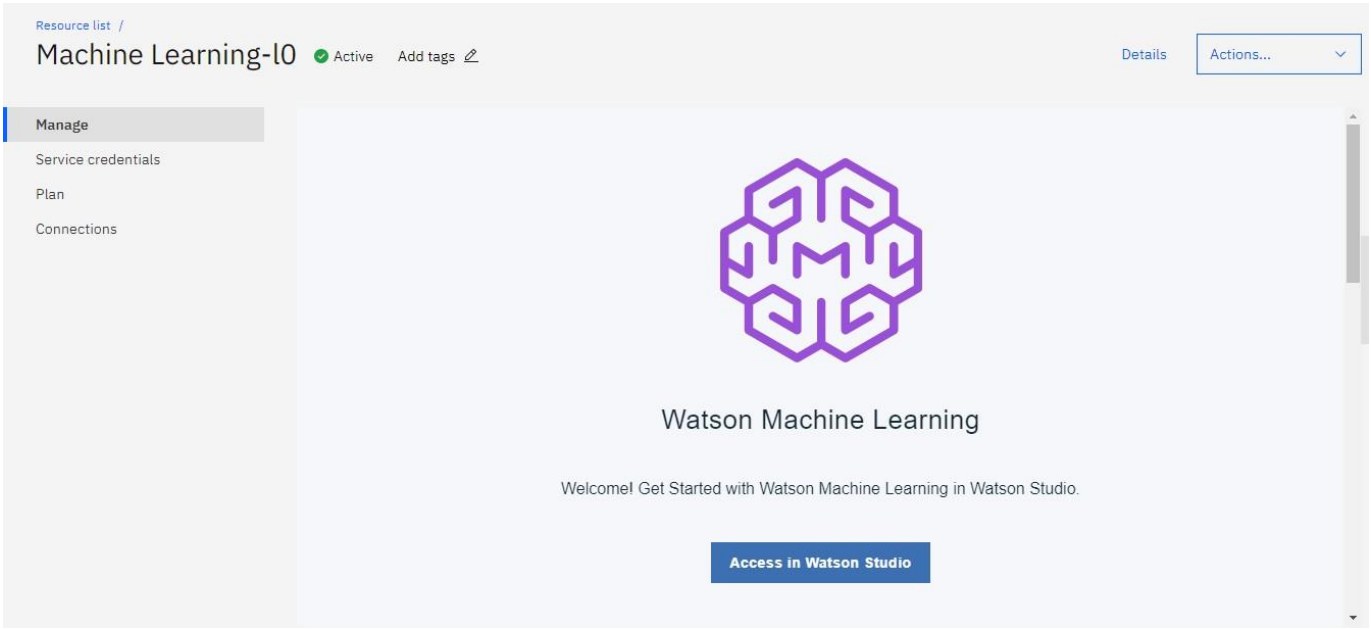
1). Create IBM Cloud Services



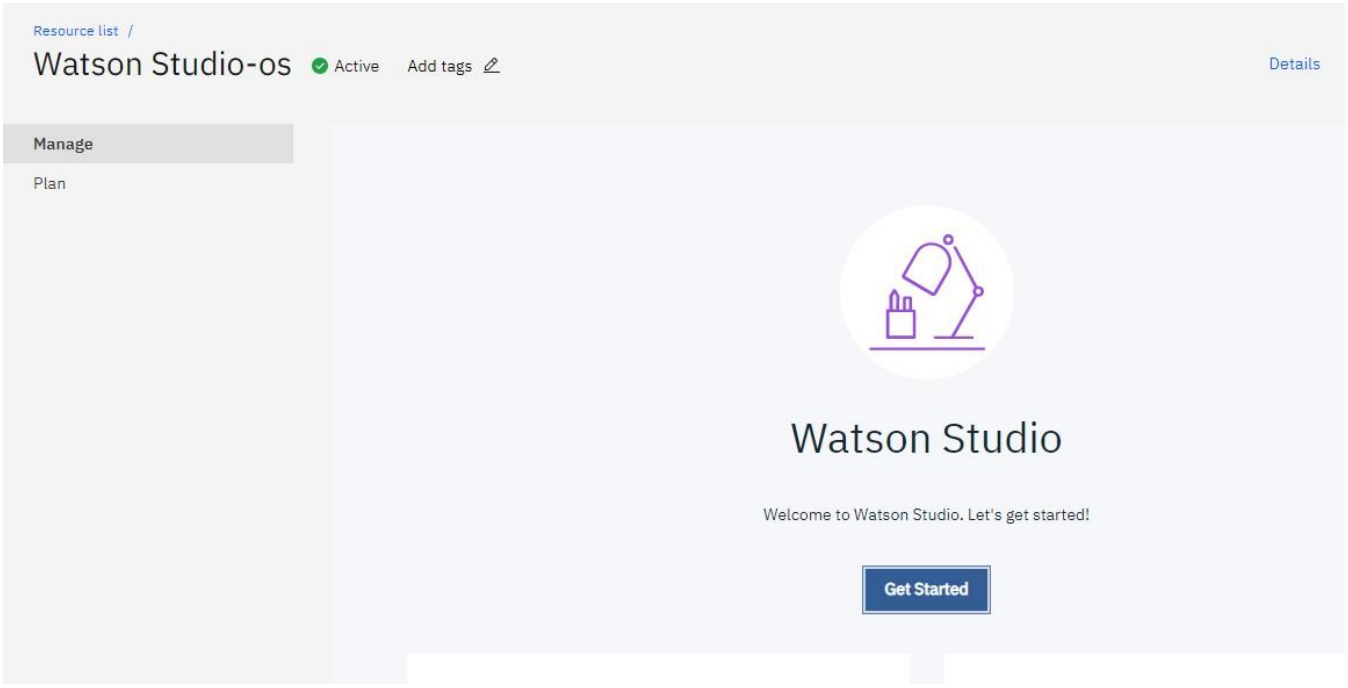
2). Resources List



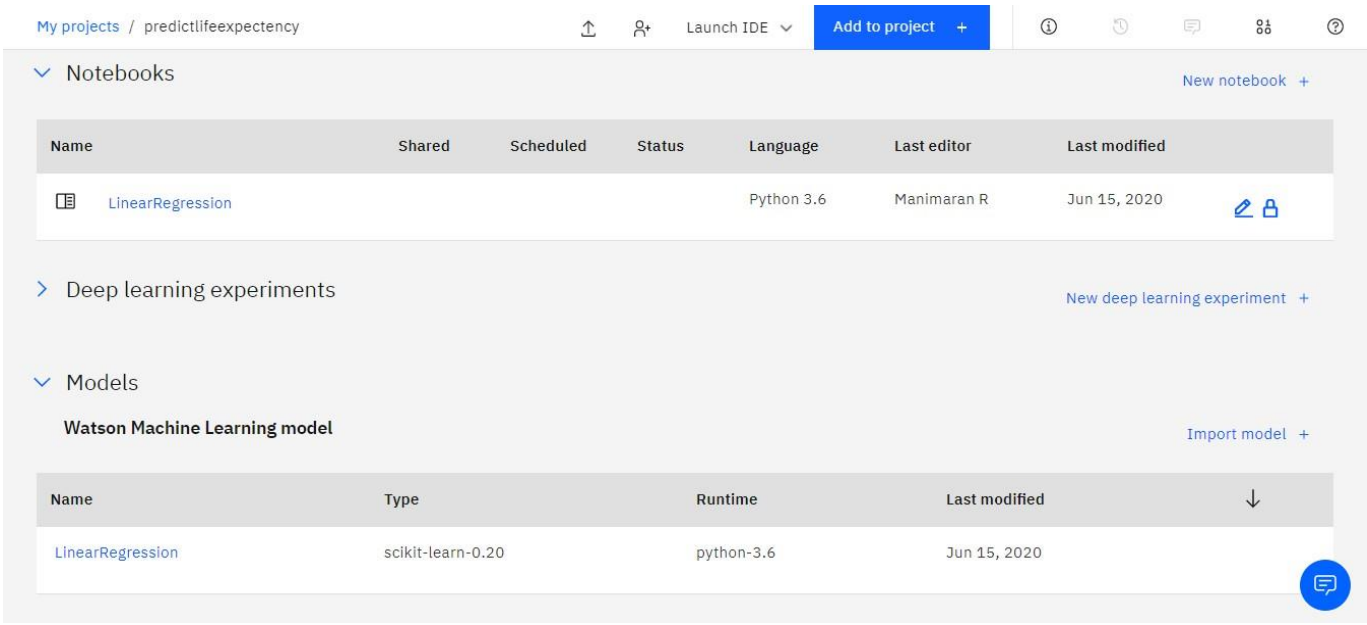
3). Watson Machine Learning Service:



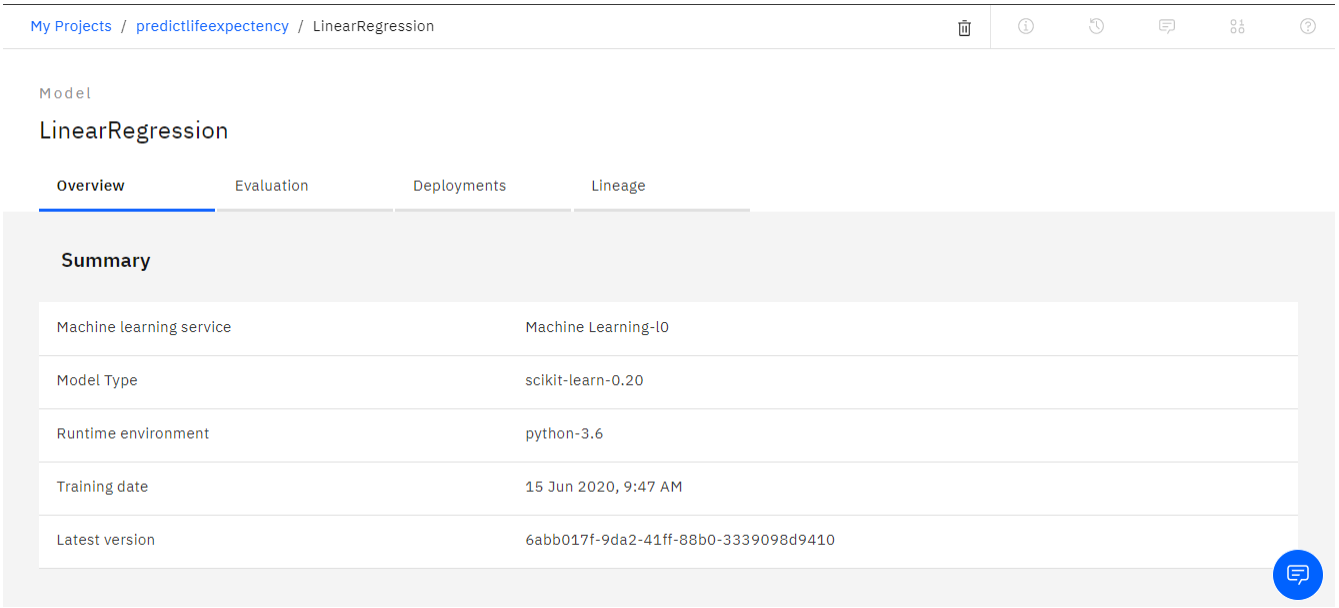
4). Watson Studio:



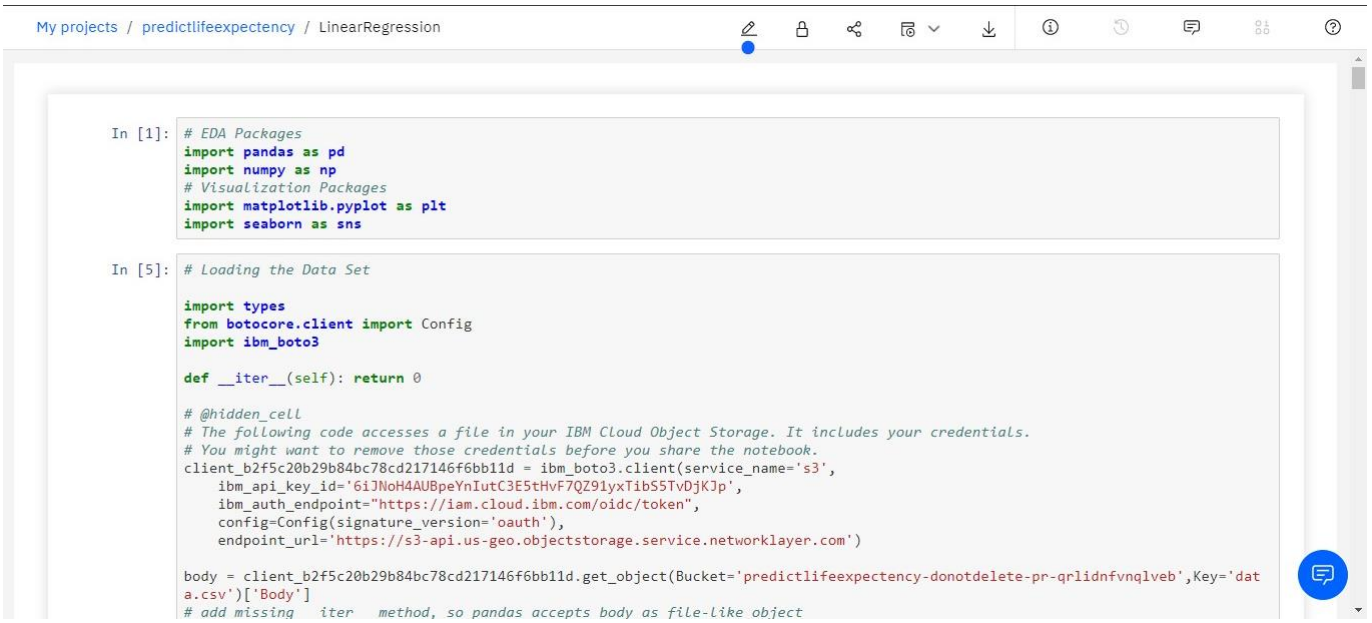
5). Watson Notebook



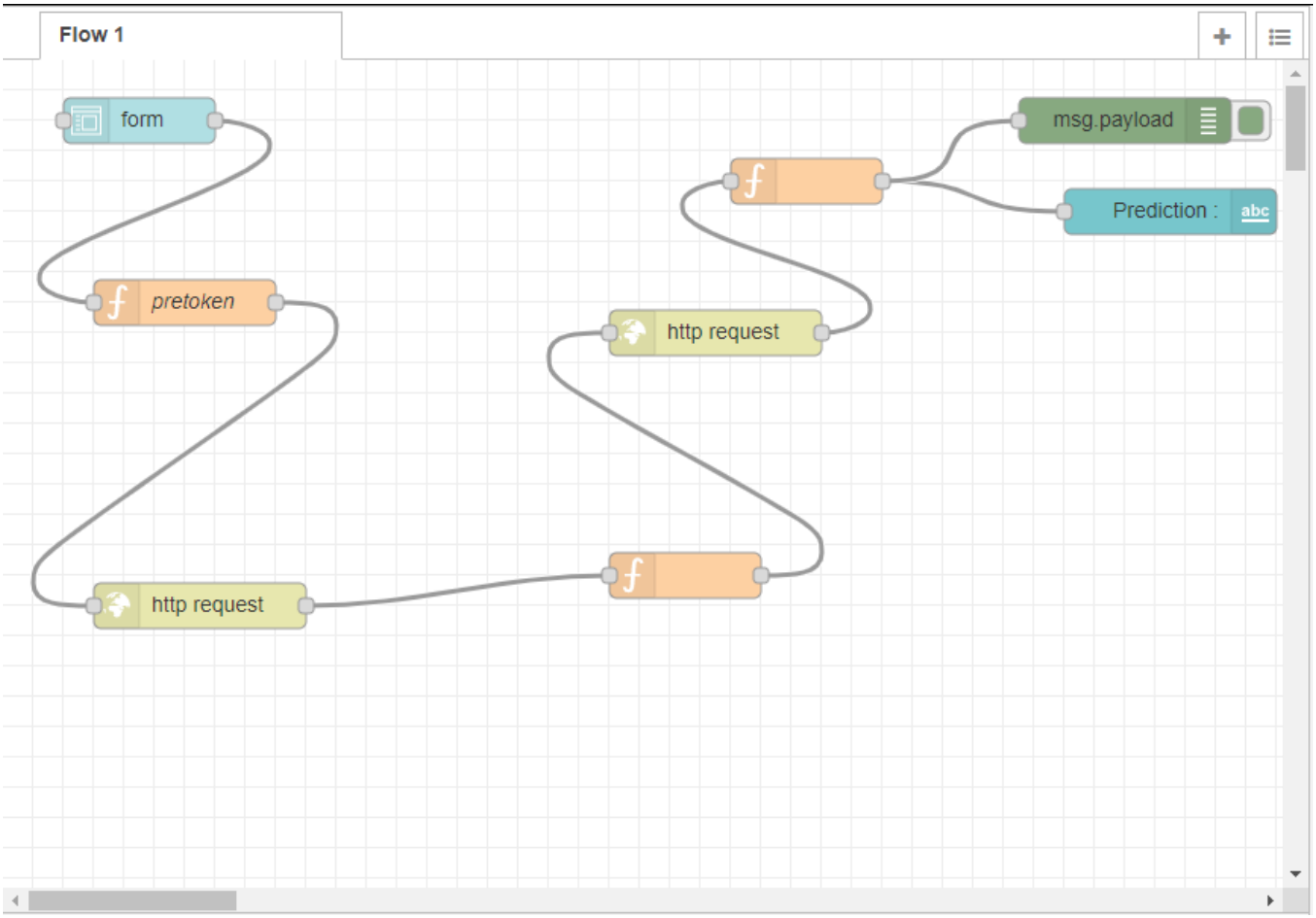
6). Watson Model



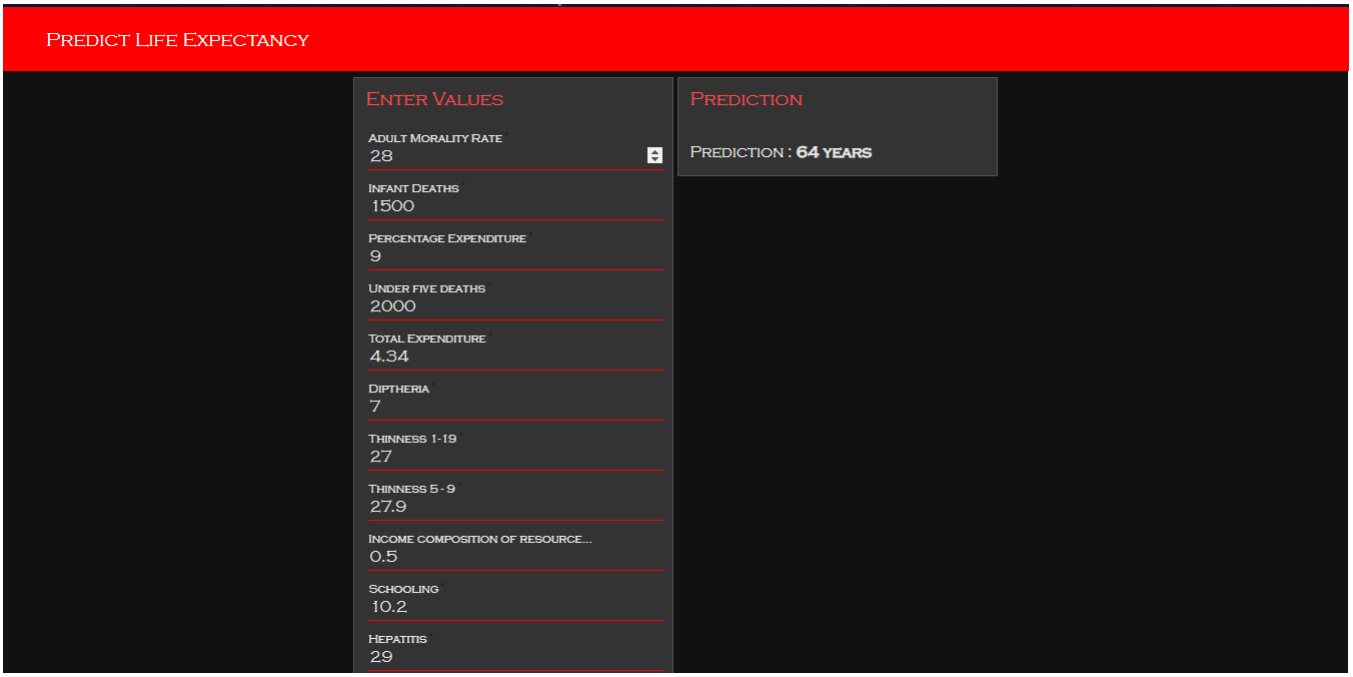
7). Notebook:

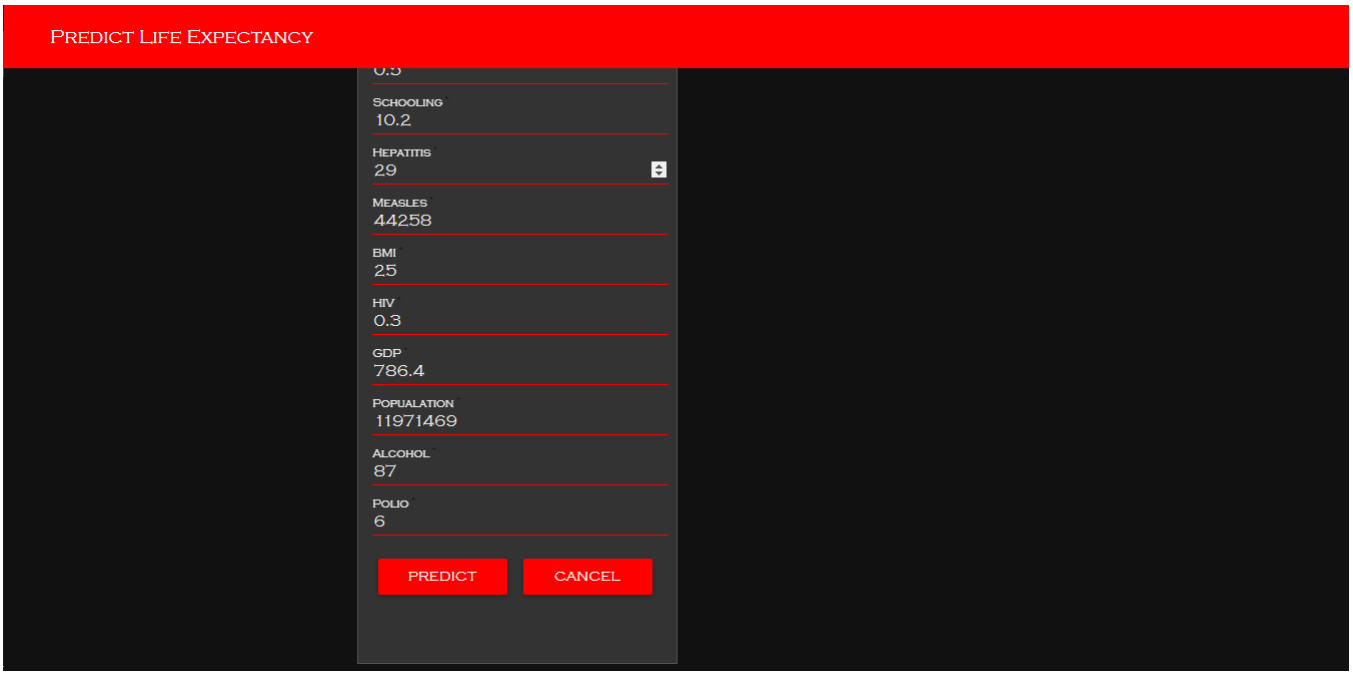


8). Node Red Flow:

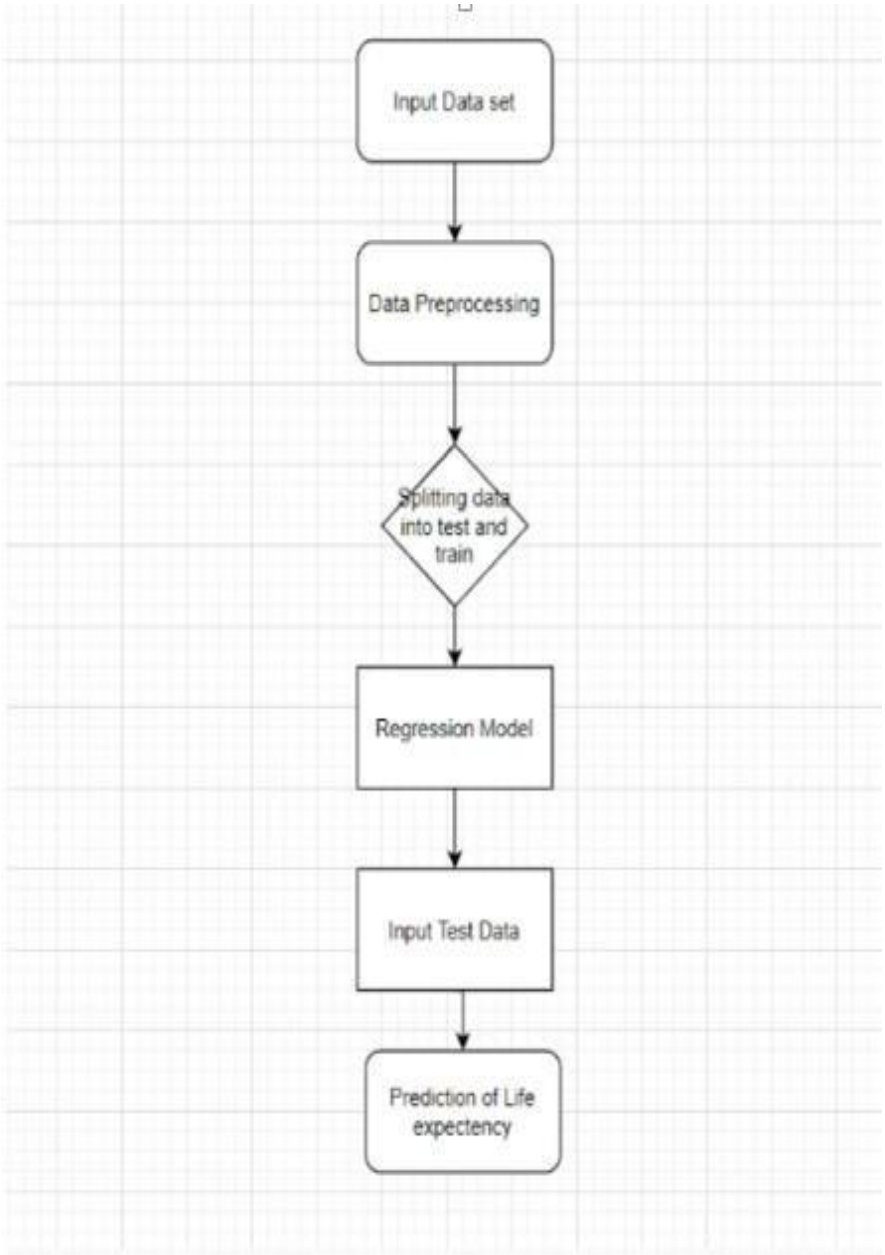


9). Node Red UI:


PREDICT LIFE EXPECTANCY
ENTER VALUES
ADULT MORALITY RATE
28
INFANT DEATHS
1500
PERCENTAGE EXPENDITURE
9
UNDER FIVE DEATHS
2000
TOTAL EXPENDITURE
4.34
DIPTHERIA
7
THINNESS 1-19
27
THINNESS 5-9
27.9
INCOME COMPOSITION OF RESOURCE...
0.5
SCHOOLING
10.2
HEPATITIS
29
PREDICTION
PREDICTION : 64 YEARS

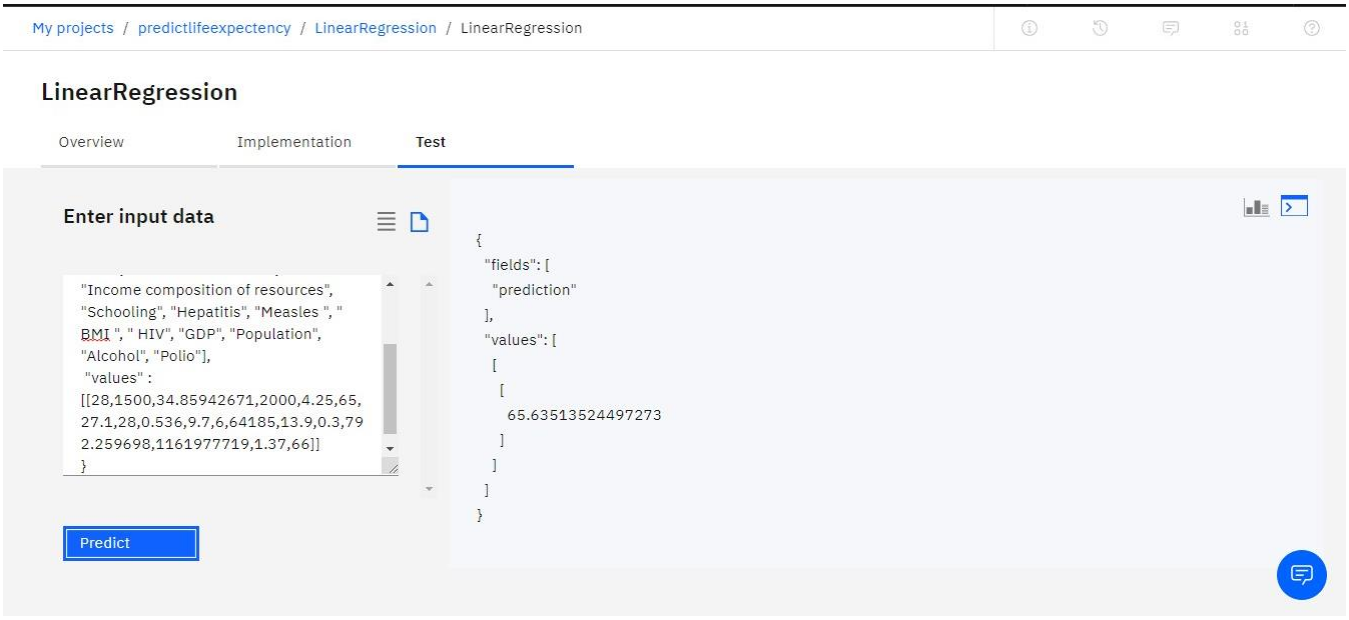

PREDICT LIFE EXPECTANCY
SCHOOLING
10.2
HEPATITIS
29
MEASLES
44258
BMI
25
HIV
0.3
GDP
786.4
POPULATION
11971469
ALCOHOL
87
POLIO
6
PREDICT
CANCEL

5. FLOWCHART


Input Data set
Data Preprocessing
Splitting data into test and train
Regression Model
Input Test Data
Prediction of Life expectency

6. RESULT

Prediction of life expectancy based on adult molarity, GDP and population etc.



7. ADVANTAGES & DISADVANTAGES

7.1 Advantages:

- User friendly Interface: This interface is very attractive, no background knowledge requires, it is simple web UI and ask for required input and predict output.
- Reduced costs: This simply a web page and does not required any kind of payment neither for designing nor for using.
- Can be used in any organization to analyze the data
- Regression technique is comparatively less impacted by noise.
- The dataset are available to public for the purpose of health data analysis.

7.2 Disadvantages:

- Can be only used by the people having the knowledge of data analysis.
- As the model is deployed on Cloud, so one requires good internet connections to use the applications.
- The model predicts averages or approximates value with 82 % accuracy.

8. APPLICATIONS

- It can be used to monitor health inequalities of acountry.
- It can be used to develop statistics for country development process.
- It can be used to analyse the factors for high life expectancy.
- It is user friendly and can be used by anyone.

9. Conclusion

Some interesting correlations here:

- There is a strong positive correlation between ‘Schooling’ and ‘Life Expectancy’. This may be because education is more established and prevalent in wealthier countries. This means countries with less corruption, infrastructure, healthcare, welfare, and so forth.

- Similarly to the point above, there is a moderate positive correlation between ‘GDP’ and ‘Life Expectancy’, most likely due to the same reason.

- Surprisingly there’s a moderate positive correlation between ‘Alcohol’ and ‘Life Expectancy’. I’m guessing that this is due to the fact that only wealthier countries can afford alcohol or the consumption of alcohol is more prevalent among wealthier populations.

10. BIBLIOGRAPHY

- https://cloud.ibm.com/login
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://nodered.org/
- https://www.ibm.com/watson/products-services
- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/
- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/
- https://developer.ibm.com/tutorials/watson-studio-auto-ai/
- https://www.kaggle.com/kumarajarshi/life-expectancy-who
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service

11. APPENDIX

Demonstration Video Link: https://youtu.be/NF-arWMUvMQ

GitHub Link: https://github.com/SmartPracticeschool/llSPS-INT-2048-Predicting-Life-Expectancy-using-Machine-Learning