

# Project Name: Predicting Life Expectancy using machine learning Kick off date: 06-06-2020

•	Predicting Life Expectancy using Machine Learning - SB48483
Project ID	SPS_PRO_215
Project Title	Predicting Life Expectancy using Machine Learning

# <u>Project Scope</u>

# 1. 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.

#### 2.Project Requirements:

#### The project requirements are:

- 1. IBM Cloud
- 2. IBM Watson Services
- 3. Node red
- 4. Python

#### 3. Functional Requirement:

Predicting life expectancy rate of a country.

#### <u>4. Technical Requirements:</u>

Knowledge In the field of Python, IBM cloud, IBM watson

#### 5. Software Requirements:

- 1. IBM Cloud
- 2. IBM Watson Services
- 3. Node red
- 4. Python platform

#### 6. Algorithm:

- 1. Loading the packages which are required.
- 2. Reading the data.
- 3. Exploring the data and analyzing it.
- 4. Preprocessing the data.
- 5. Create a training and Test sets.
- 6. Create a model and train it

#### 7.Project Deliverables:

A Trained Model which predicts the life expectancy of people living in a country with the given data.

#### 8.Details:

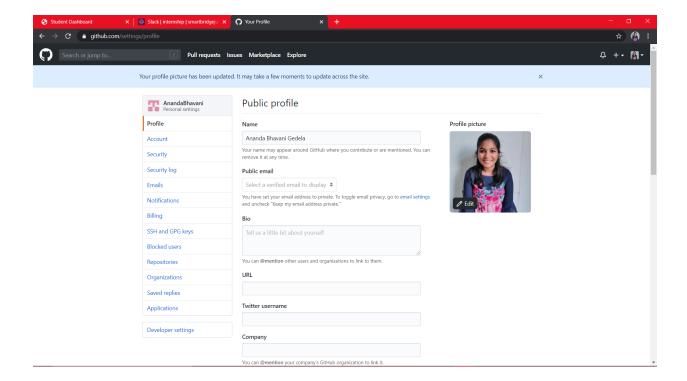
Name: Ananda Bhavani Gedela

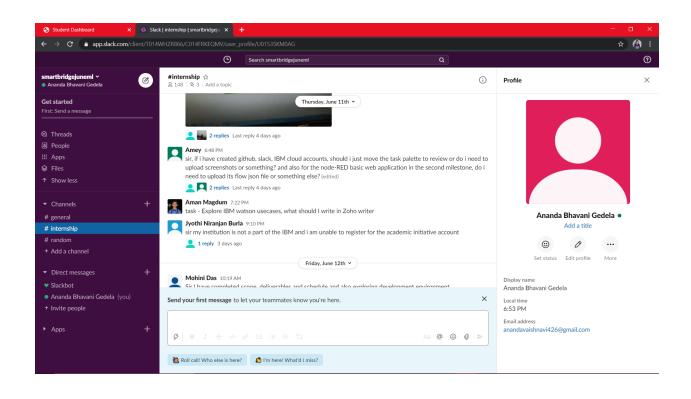
College: Vignan's institute of engineering for women.

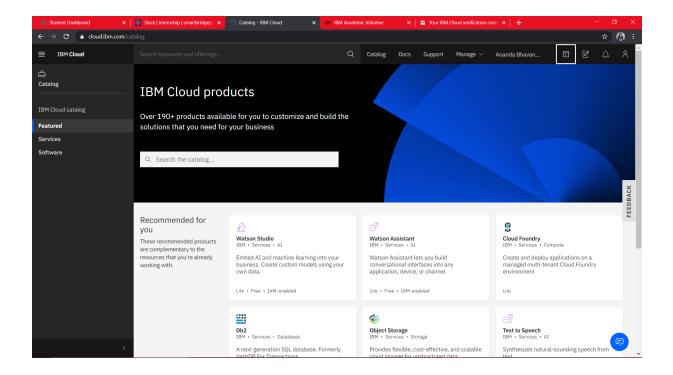
Email: anandavaishnavi426@gmail.com

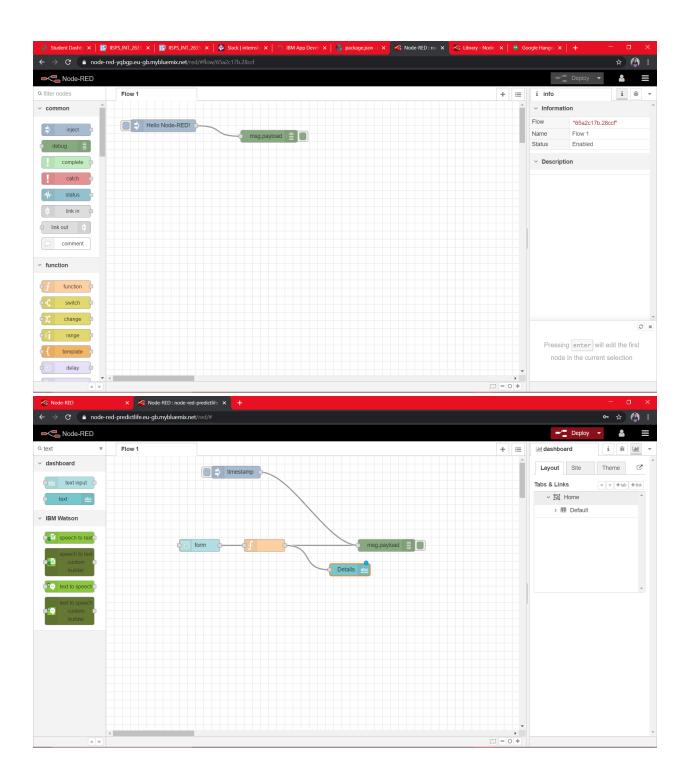
#### 9.Project Schedule:

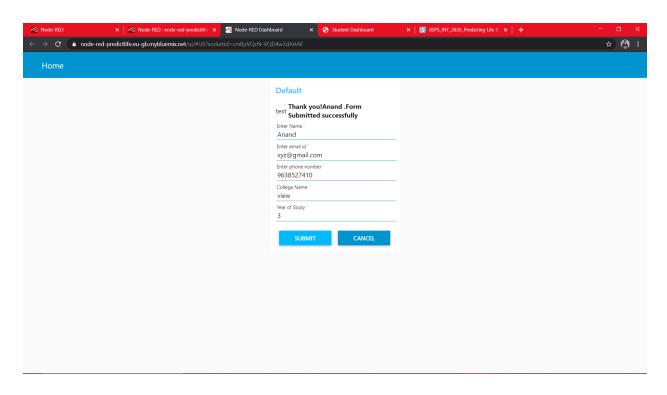
Starting from 06-06-2020 the project has to completed and submitted within one month.

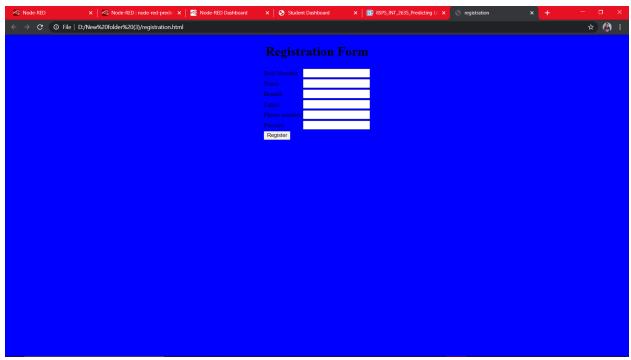


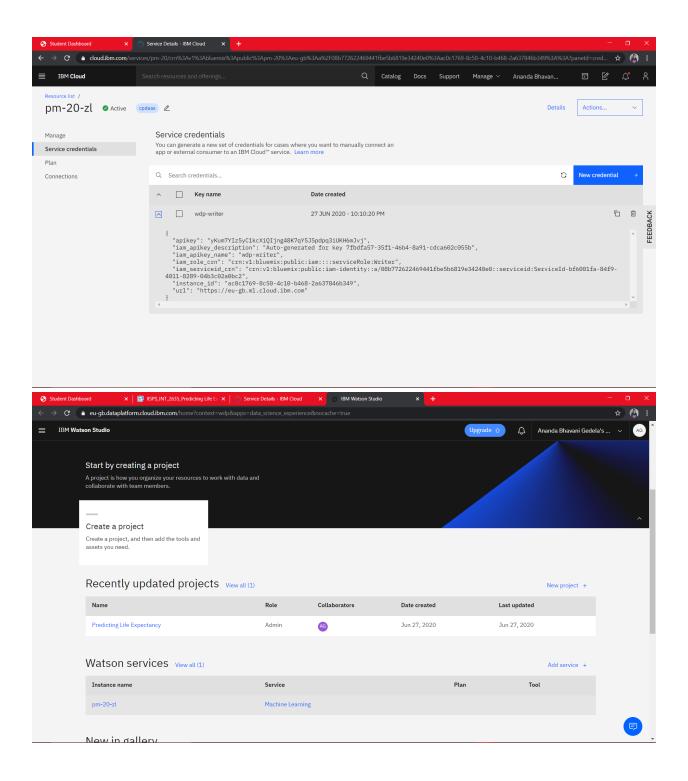


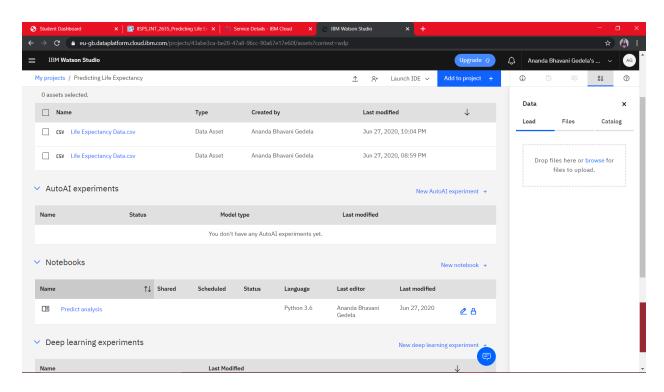


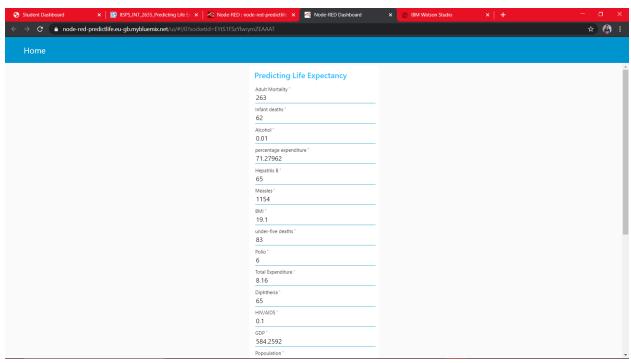


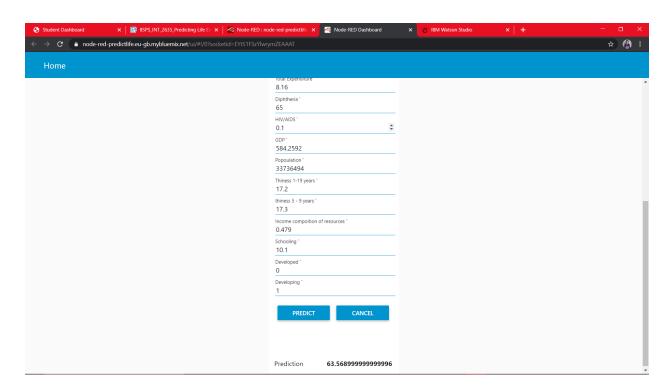


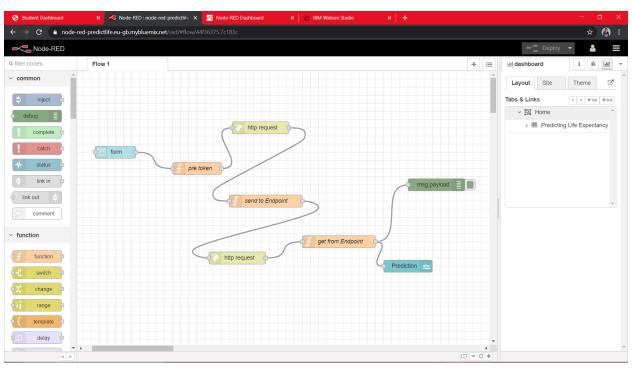












# INTERNSHIP PROJECT REPORT ON

**Predicting Life Expectancy Using Machine Learning** 

In Partial Fulfillment of

For the Award of the Degree of

IN

**COMPUTER SCIENCE AND ENGINEERING** 

BY

**ANANDA BHAVANI GEDELA** 

# **Under**

# **SMARTINTERNZ**

- 1 INTRODUCTION
  - 1.1 Overview
  - 1. 2 Purpose
- 2 LITERATURE SURVEY
  - 2.1 Existing problem
  - 2.2 Proposed solution
- 3 THEORITICAL ANALYSIS
  - 3.1 Block diagram
  - 3.2 Software /Hardware designing
- 4 EXPERIMENTAL INVESTIGATIONS
- 5 FLOWCHART

- 6 RESULT
- 7 ADVANTAGES AND DISADVANTAGES
- 8 APPLICATIONS
- 9 CONCLUSION
- 10 FUTURESCOPE
- 11 REFERENCES

# INTRODUCTION

The term "life expectancy" refers to the number of years a person can expect to live. By definition, life expectancy is based on an estimate of the average age that members of a particular population group will be when they die.

For example, colorectal cancer screening guidelines state that clinicians should only screen patients with an estimated life expectancy of at least 10 years because otherwise benefits of cancer detection are unlikely to outweigh the harms and costs. Referral to hospice care is often based on a life expectancy of less than 6 months. Implantable cardiac defibrillators are not indicated if the patient is not expected to live longer than one year.

#### 1.1 Overview

In practical terms, estimating life expectancy entails predicting the probability of surviving successive years of life, based on observed age-specific mortality rates, 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.

Good prognostication 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. So this problem statement is aimed at predicting Life Expectancy rate of a country given various features. It predicts the average lifetime of a human being and predicts on the basis of various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. So the end product will predict the future life expectancy of the person with the help of prior given appropriate matrix of features by the user like current 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.

# 1.2 PURPOSE

Although there have been a lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of

immunization and human development index was not taken into account in the past. By predicting life expectancy and having good prognostication can help in making valuable decision like the course of treatment and helps to anticipate the procurement of health care services and facilities.

The average life Expectancy of a certain country says many things about that particular country. It ultimately helps in predicting the health conditions and the development of the health sector in that particular country. This ultimately helps the nation to find the area which needs attention in an urge to improve its contribution in average lifespan of a human being. The expectancy obviously depends upon the country's population, GDP, the economy of the country and many more factors. It is not enough to have a long life; instead with having a long life one should have a healthy and fit life as well.

# 2 LITERATURE SURVEY

# 2.1 EXISTING PROBLEM

Predicting Life Expectancy has been a long-term question to humankind. Many calculations and Research have been done to create an equation despite it being impractical to simplify these variables into one equation. Past studies have revealed a lot of work in the field of predicting life expectancy of a human being. After reviewing existing works and techniques in the prediction of human Life Expectancy, and finally reached a conclusion that it is possible to predict a Average Life Expectancy for individuals using advancing technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearable's and mobile health monitoring devices, IOT. It is noticed that the collection of data is a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The interworking of a heterogeneous health network is also a challenge for data collection.

Despite these challenges, a possibility of predicting Life by proposing an approach of data collection and application by smart phone, in which users can enter their information to access the cloud server to obtain their own predicted Lifespan based on the given inputs. To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.

#### 2.2 PROPOSED SOLUTION

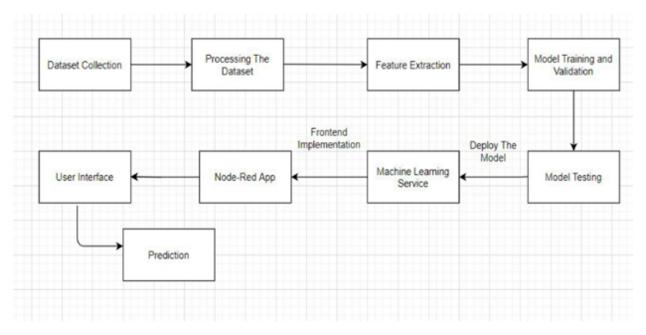
Although there have been a lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on a data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regressions while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations in this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. The model of "Predicting Life Expectancy using Machine Learning" uses IBM Cloud services, which helps to avoid any storage issues. The UI Presented to the users is a website URL i.e. on user's fingertips.

The end result will be a web page where it will predict the life expectancy based upon various inputs provided by the user by the technique of regression algorithm. It depend on factors such as Status of the country, Adult Mortality Rate, Infant deaths, Alcohol, Hepatitis B, Measles, BMI, Polio, Total Expenditure, Diphtheria, HIV/AIDS, GDP of a country, Population, Income Composition of Resources, Schooling status of the country.

To access the trained model, we will use Node-Red App from IBM Cloud.

# 3. THEORITICAL ANALYSIS

#### 3.1 BLOCK DIAGRAM



# 3.2 HARDWARE /SOFTWARE DESIGNING

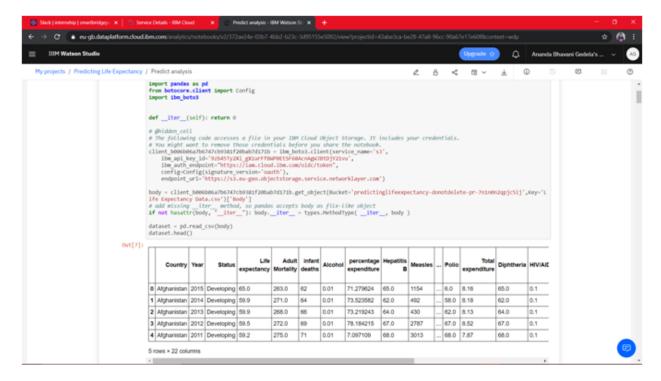
# 1. Collecting the Dataset

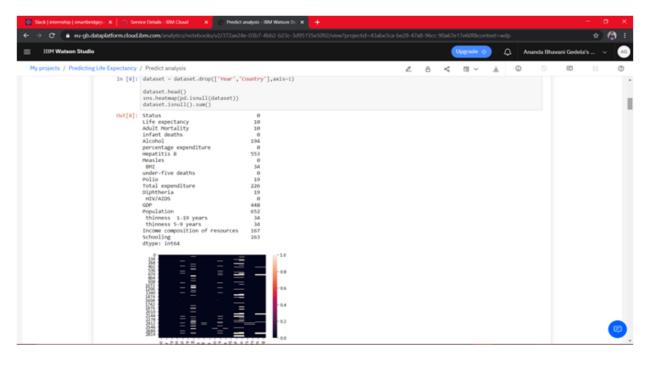
- 2. Creating Necessary IBM Cloud Service
- 3. Creating and Configuring Watson Studio
- 4. Create Machine Learning Service
- 5. Adding Jupyter Notebook
- 6. Build ML model and create Scoring Endpoint for Node-Red Integration
  - 7. Build Node-Red Flow and integrate ML services and deploy.

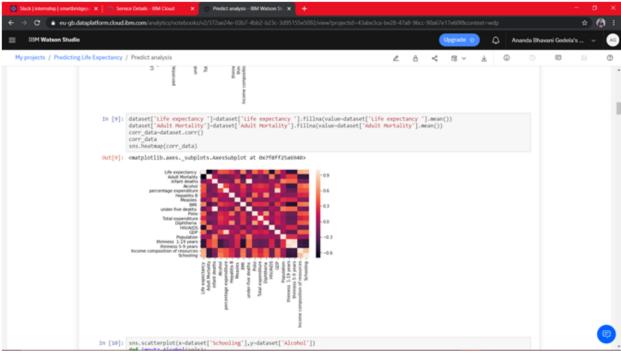
Software Requirements: ➤ Python IDE ➤ IBM Cloud ➤ IBM Watson ➤ IBM Watson Studio ➤ Node-Red

# 4. THEORITICAL ANALYSIS

Analyzing each and every feature in our dataset is very important as it helps us to build a model which gives more accurate result.

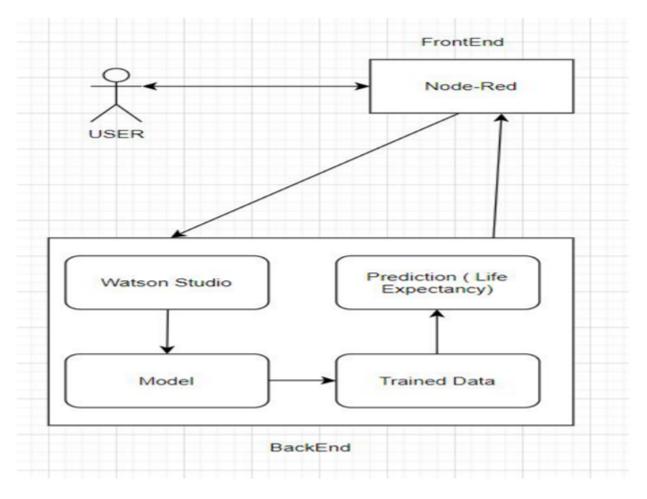


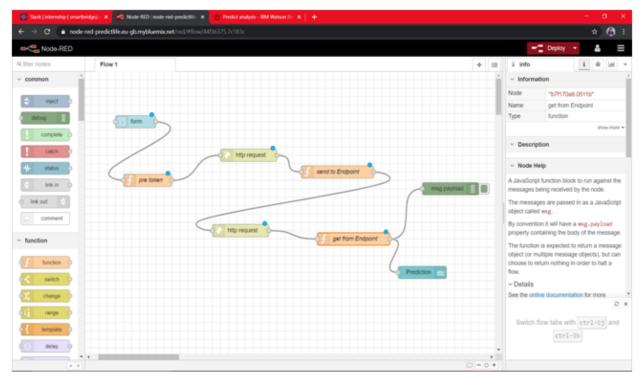




# 5. FLOW CHART

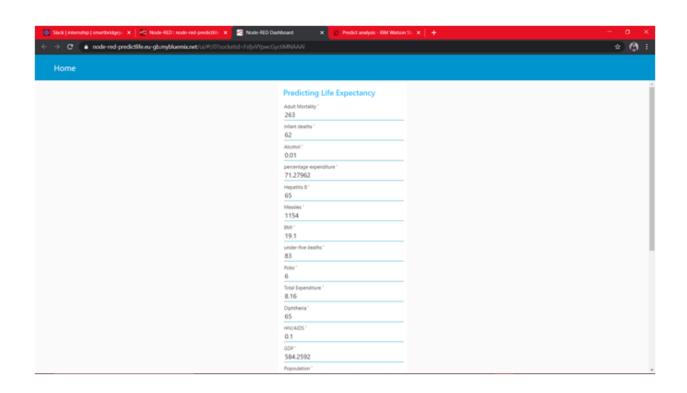
A flowchart is a diagram that depicts a flow of process, system or computer algorithm. They are widely used in multiple fields to document, study, and plan, improve and communicate complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence.

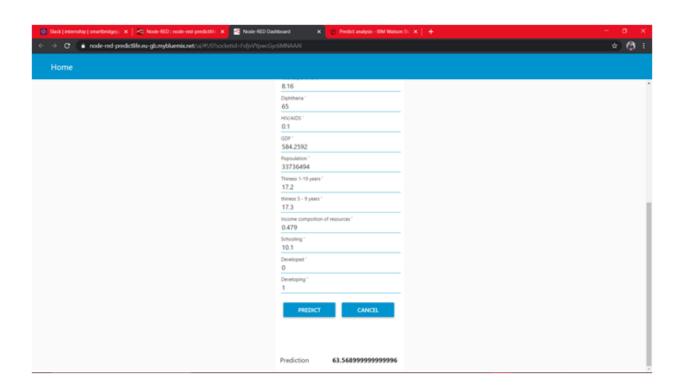




# 6. RESULT

The user friendly Graphical User interface is shown in Figure below. This GUI is connected to the trained machine learning model present in the backend (IBM Watson notebook). The user has to fill in the inputs accordingly and click on the "Predict" button present at the end of the form. On clicking the "Predict" button, the user will be displayed the predicted life expectancy at the predict label, based on the inputs provided as shown in Figure below.





# 7. ADVANTAGES & DISADVANTAGES

#### 7.1. Advantages:

- 1. Advantages of using IBM Watson:
- Processes unstructured data
- Fills human limitations
- Acts as a decision support system, doesn't replace humans
- Improves performance + abilities by giving best available data
- Improve and transform customer service
- Handle enormous quantities of data
- Sustainable Competitive Advantage
- 2. Easy for users to interact with the model via the UI.
- 3. User-friendly.
- 4. Easy to build and deploy.
- 5. Doesn't require much storage space.

# 7.2. Disadvantages:

Disadvantages of using IBM Watson:

• Computation cost and training the model will increase the cost if dataset is

too large in size

- Node-Red doesn't give much flexibility to design own templates, although it's a great service.
- Required Internet connection
- Increasing rate of data, with limited resources
- Wrong prediction if the user provides wrong information as it is user dependent.

# 8. APPLICATIONS

- To analyze all the factors and plan out measures to increase the life expectancy of the country
- To help government prepare life insurance policies for people. This will benefit the people.
- To analyze country's growth statistics in future years.
- It can be used to monitor health inequalities of a country.
- This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

# 9. CONCLUSION

Thus, we have developed a model that will predict the life expectancy of a specific demographic region based on the inputs provided. Various factors have a significant impact on the life span such as Adult Mortality, Population, Under 5 Deaths, Thinness 1-5 Years, alcohol, HIV, Hepatitis B, GDP, Percentage Expenditure and many more. Users can interact with the system via a simple Graphical user interface which is in the form of a form with input spaces which the user needs to fill the inputs into and then press the "predict" button.

Predicting Life Expectancy can lead to the development of the country. It can widely impact Health Sectors, Public Sectors and Economic Sectors by improving the resources, funds and services provided to people.

#### 10. FUTURE SCOPE

As future scope, we can connect the model to the database which can predict the life Expectancy of not only human beings but also of the plants and different animals present on the earth. This will help us analyze the trends in the life span. A model with country wise bifurcation can be made, which will help to segregate the data demographically.

Other factors such as sentiment analysis and mental health can be added to predict life expectancy. Happiness index is also one such feature which can be proved vital determining life expectancy.

# 11. Bibliography:

- 1. Project planning and kick off:
- a. https://www.allbusinesstemplates.com/download/?file code=2KBA4&lang=en&iuid=9f9faa69-9fab-40ee-8457- ea0e5df8c8de
- b. https://www.youtube.com/watch?v=LOCkVmENq8&feature=youtu.be
- c. https://github.com/
- d. https://slack.com/intl/en-in/
- e. https://www.zoho.com/writer/help/working-withtext.html
- 2. IBM Cloud:
- a. https://cloud.ibm.com/login

- b. https://developer.ibm.com/tutorials/how-to-create-anode-red-starter-application/
- c. https://www.w3schools.com/howto/howto\_make\_a\_w ebsite.asp
- 3. IBM Watson services:
- a. https://www.ibm.com/watson/products-services

b.https://developer.ibm.com/technologies/machinelearning/series/learning-path-machinelearning-fordevelopers/

- c. https://www.youtube.com/watch?v=NmdjtezQMSM
- d. https://bookdown.org/caoying4work/watsonstudioworkshop/jn.html
- e. https://developer.ibm.com/tutorials/watson-studioauto-ai/

**f**.https://bookdown.org/caoying4work/watsonstudioworkshop/jn.html#deploy-model-a s-web-service

- 4. Dataset:
- a. https://www.kaggle.com/kumarajarshi/life-expectancywho

# 12. Appendix:

GUI url link: <a href="https://node-red-predictlife.eu-gb.mybluemix.net/ui/">https://node-red-predictlife.eu-gb.mybluemix.net/ui/</a>

#### Source code:

https://github.com/SmartPracticeschool/IISPS-INT-2635-Predicting-Life-Expectancy-using-Machine-Learning/blob/master/Predict%20analysis%20(1).ipynb

