Project Report

on

"Predicting Life Expectancy using Machine Learning"

Category: Machine Learning

Submitted By : Muhammed Jawad

Ashfaq Shiningwala

Email : jawadshiningwala@gmail.com

Submitted To :

SmartInternz.com

Index

1. INTRODUCTION

Life expectancy, an estimate of the number of remaining years of life a person has, is an important consideration for making clinical decisions in primary care. Predicting Life Expectancy helps analyze the average lifespan of the countrymen which helps in making crucial health decisions

1.1. Overview

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.

This project is to build a model while considering historical data from a period of 2000 to 2015 for all the countries. The model trained in this project will be able to predict the average lifetime of a human being given some input factors .With the help of this project any country is able to predict the expected lifetime of their countrymen and then accordingly take preventive measures to improve on their healthcare measures. This will also help countries in improving a particular field such as GDP ,alcohol intake,etc which have a high impact on a country's life expectancy.

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

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 fit life as well.

Muhammed Jawad

2. LITERATURE REVIEW

2.1. Existing Problem

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, wearables 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 smartphone, 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 regression 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 users fingertips.

Muhammed Jawad

3. THEORITICAL ANALYSIS

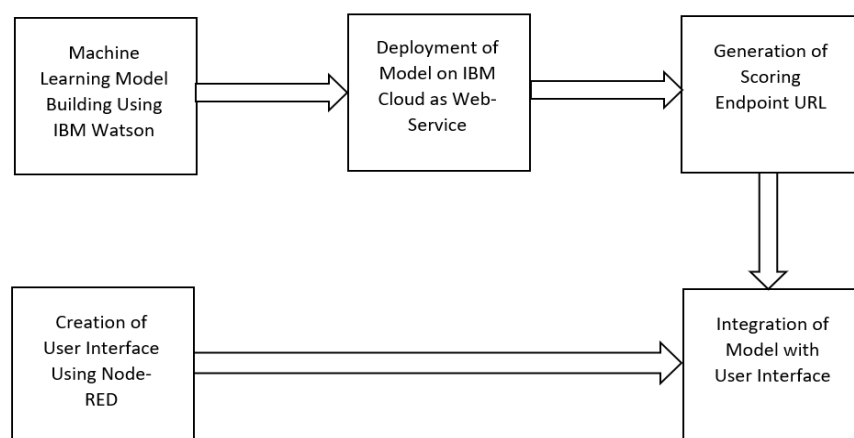## 3.1 Block diagram

TRAINING
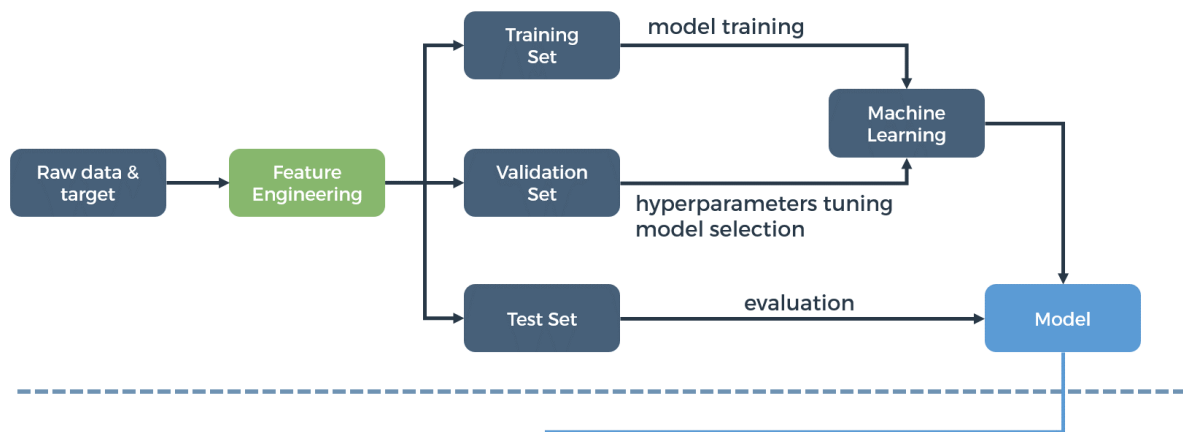
PREDICTING

3.2 Hardware / Software designing

There is no Hardware involved as such. The web-app has IBM Watson Machine Learning as its backend service and Node-RED for the User Interface development. The web-app has been deployed on IBM Cloud as a Web Service. Basic Flow is as follows:
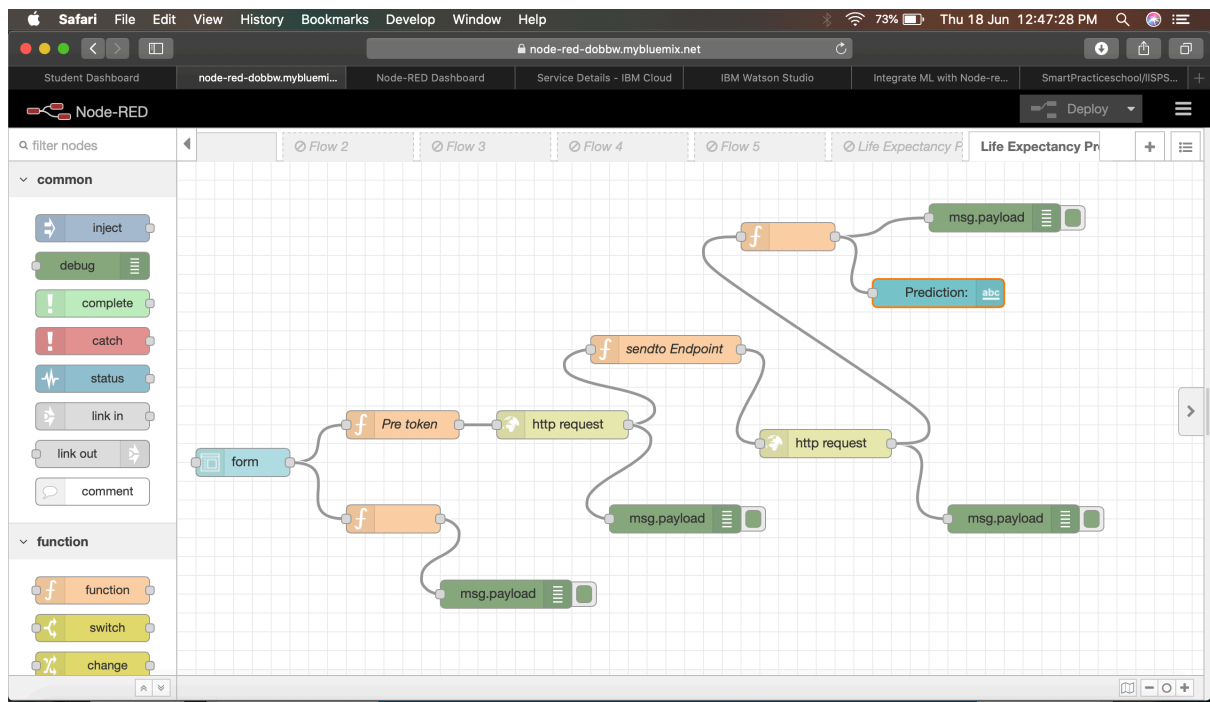
Muhammed Jawad

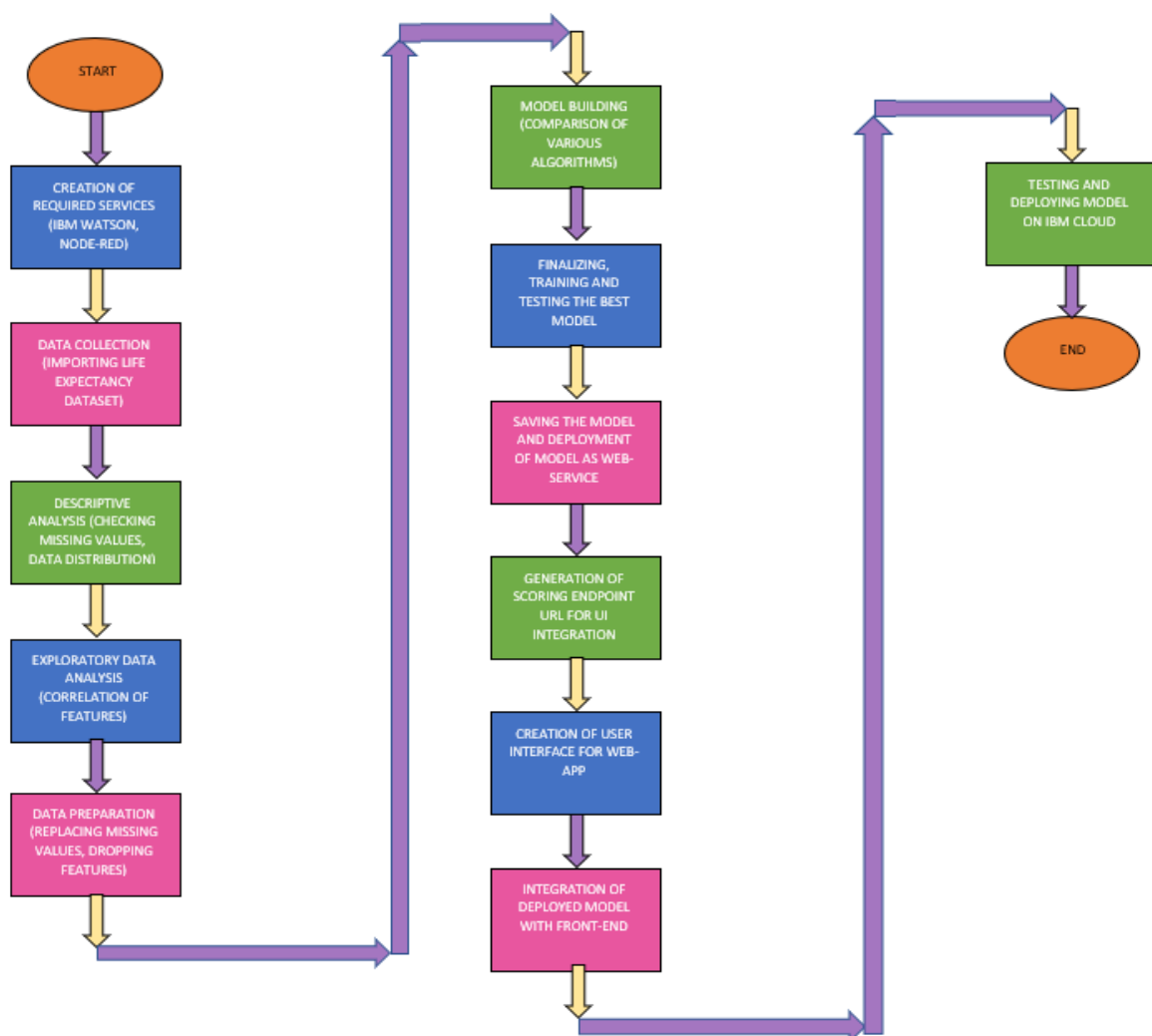A node red flow is made for the prediction of life expectancy. Which is given below.
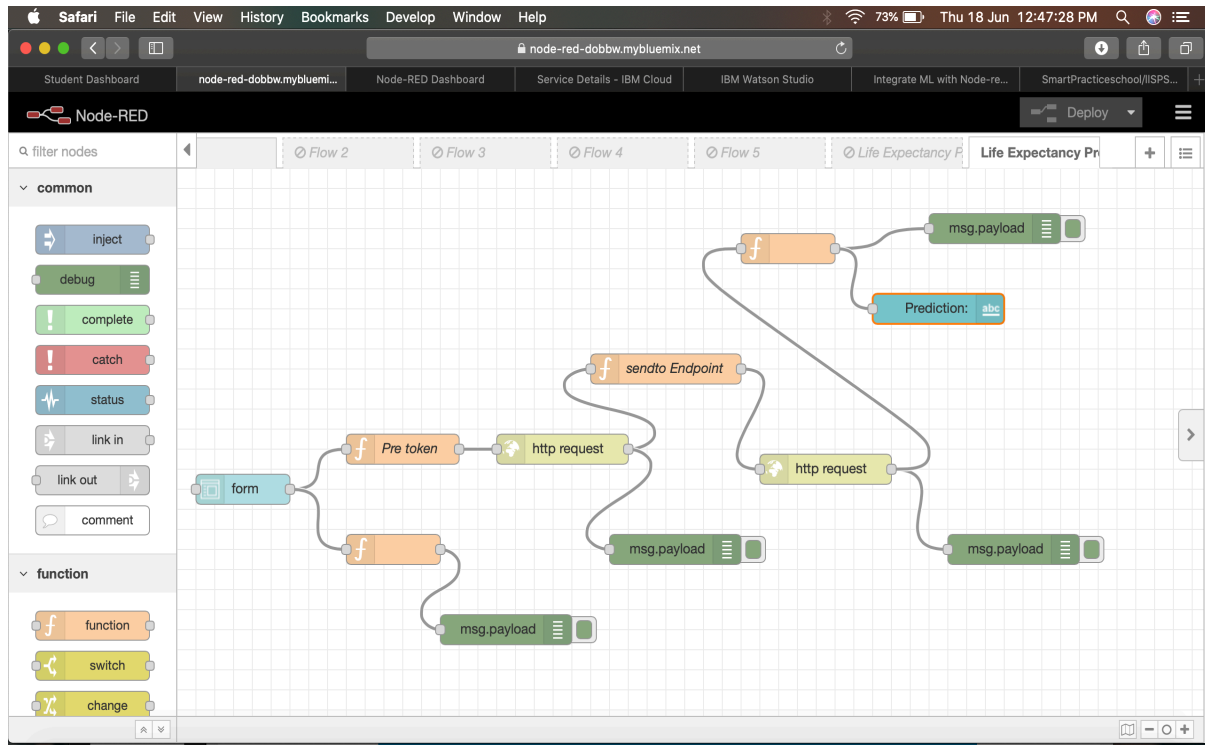
4. EXPERIMENTAL INVESTIGATIONS

- Collection of data set from Kaggle.
  https://www.kaggle.com/kumarajarshi/life-expectancy-who

- On IBM Watson studio machine learning using auto ai build a model to predict life expectancy.
  o To do so first create account on IBM Watson studio.
  o Using Add to project choose auto ai.
  o Then upload data set
  o Choose best way to predict.
  o Save as a model which is on the top
  o Deploy the model.
  o Test the model.
  o Create service credential
  o Make node red flow

Muhammed Jawad

5. FLOWCHART

Muhammed Jawad

6. RESULT

The user friendly Graphical User interface is shown in Figure. This GUI is connected to the trained machine learning model present in the backend. 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.

7. ADVANTAGES AND DISADVANTAGES

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.

Muhammed Jawad

5. Doesn't require much storage space.

Disadvantages:

1. Disadvantages of using IBM Watson:
   - Only in English (Limits areas of use)
   - Seen as disruptive technology
   - Maintenance and even requires internet connection.
   - Doesn't process structured data directly
   - Increasing rate of data, with limited resources
2. The web-app is not handy to use for users as it is not a mobile app.
3. The user has to input all the fields and only then the prediction will be given to the user. However, inputting all these values is a tiresome job, and the user cannot input these values through speech.

8. APPLICATIONS

- 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.
  - Similarly more result can be abstract.

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.

10. FUTURE SCOPE

1. We can make the model more accurate and do more relevant feature extraction.
2. We can make a mobile app as now-a-days it is more preferred.

Muhammed Jawad

3. 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 analyse the trends in the life span.
4. A model with country wise bifurcation can be made, which will help to segregate the data demographically.

11. BIBILOGRAPHY

- Learn how to add, edit, delete text using Writer.
  https://www.zoho.com/writer/help/working-with-text.html

- GitHub account creation
  https://github.com
  https://www.youtube.com/watch?v=7YUTc4Cigc8&feature=youtu.be

- How to create a Kick Off Meeting Agenda
  https://www.youtube.com/watch?v=LOCkV-mENq8&feature=youtu.be

- Data visualization, preparation, and transformation using IBM Watson Studio
  https://developer.ibm.com/tutorials/watson-studio-data-visualization-preparation-transformation/

- Automate model building in IBM Watson Studio
  https://developer.ibm.com/tutorials/watson-studio-auto-ai/

- Create a Node-RED starter application
  https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
  https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be

APPENDIX

A. Source code

- The source code for IBM Watson Project and Node-RED Flow can be found at:
  https://github.com/SmartPracticeschool/llSPS-INT-2124-Predicting-Life-Expectancy-using-Machine-Learning

- The web-app can be used at:
  https://node-red-dobbw.mybluemix.net/ui/#!/0?socketid=hixPMJbxeSndVKfAAAAU