Predicting Life Expectancy Using Machine Learning

Prepared by: Adithi Macherla

Project Name: Predicting Life Expectancy using Machine Learning

Kickoff Date: 17-06-2020

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 the 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 the 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 its 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 the 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 the healthcare system, and some specific disease-related deaths that happened in the country are given.

Project Name: Predicting Life Expectancy using Machine Learning

Kickoff Date: 17-06-2020

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 to the 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 of having a long life one should have a fit life as well.

2 LITERATURE SURVEY

2.1 Existing problem

Past studies have revealed a lot of work in the field of predicting the 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 an 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 the 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 the motivation to resolve both the factors stated previously by formulating a regression model based on the 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 a 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 the user's fingertips.

3 PROJECT REQUIREMENTS

This project mainly aims at predicting life expectancy. The basic requirement of the project is the availability of a suitable dataset which will aid the prediction. So in this project, I have used the standard WHO dataset on Kaggle. The machine learning model is trained on the basis of the data provided, such that it could predict the average lifespan of an individual in the coming years.

3.1 Functional Requirements

- Download the dataset.
- Analyze and clean the dataset.
- Create an IBM account.
- Create the required services i.e.WatsonStudio, Watson Machine Learning, and Node-red.
- Train the regression model on different algorithms.
- Check for the best one and finalize that algorithm to train our model.
- Build Node-red flow for GUI.
- Create a scoring endpoint for integrating our model to node-red flow.

3.2 Technical Requirements

- The GUI must be integrated with the backend model.
- The model before training should be provided with the clean dataset(which

is preprocessed).

3.3 Software Requirements

- Python IDE
- Microsoft Excel
- IBM Cloud Account
- IBM Watson
- Node-red

4 FLOWCHART

A flowchart is a diagram that depicts the flow of a process, system, or computer algorithm. They are widely used in multiple fields to document, study, 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.
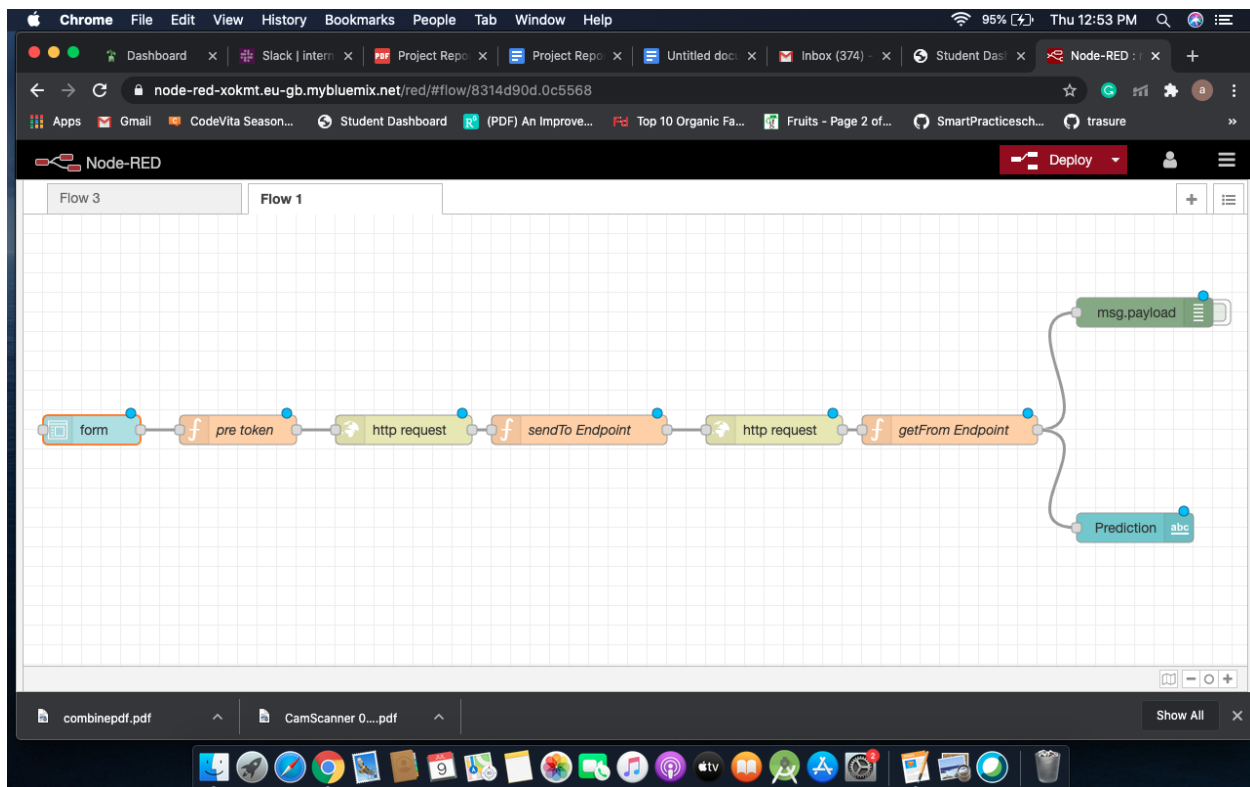
Project Name: Predicting Life Expectancy using Machine Learning

Kickoff Date: 17-06-2020

5 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(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 predicted label, based on the inputs provided as shown in the following Figure.
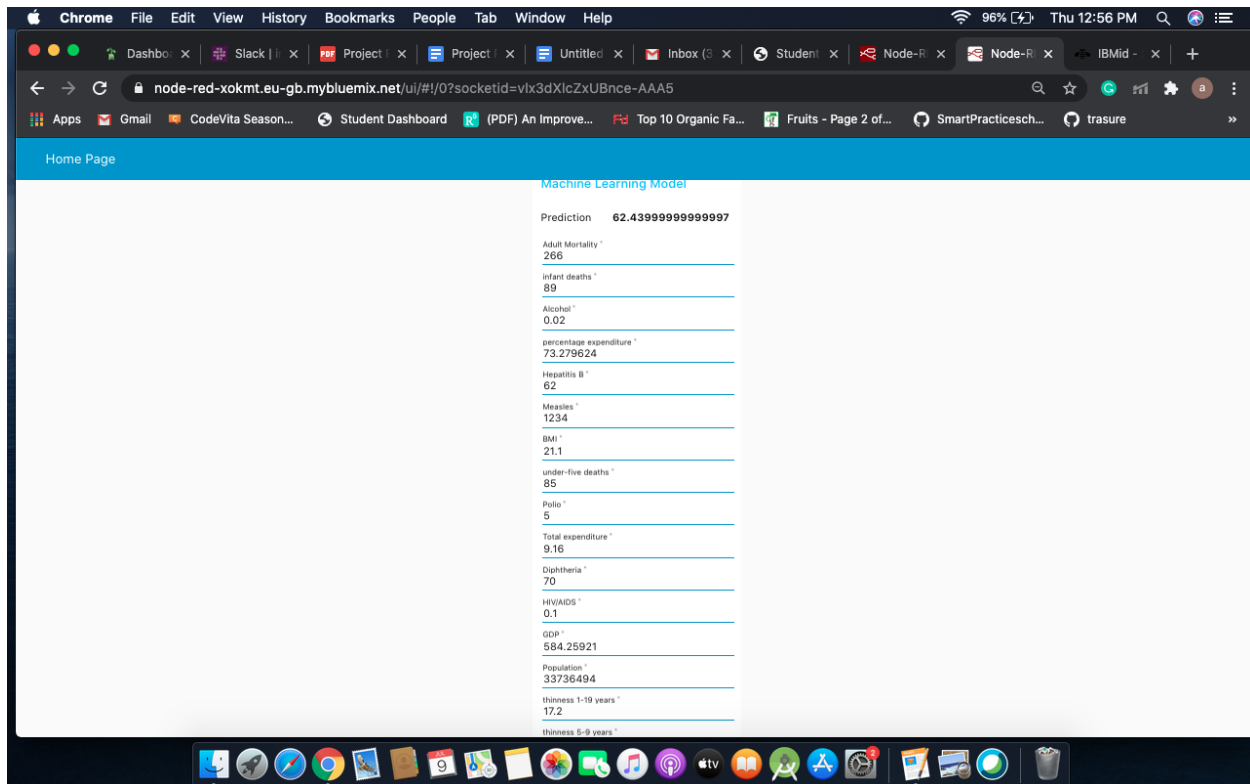
6 ADVANTAGES & DISADVANTAGES

6.1 Advantages

Advantages of using IBM services:

- Processes unstructured data.
- Acts as a decision support system.
- Improves performance.
- Handles very large datasets.
- Easy for users to interact with the model via UI.

- User-friendly.
- Easy to build and deploy.
- Doesn't require much storage page.

6.2 Disadvantages

- Language restrictions
- Seen as a disruptive technology
- High maintenance and requires an internet connection.
- Doesn't process structured data directly.
- The increasing rate of data with limited resources.

7 APPLICATIONS

Life expectancy is the primary factor in determining an individual's risk factor and the likelihood they will make a claim. This project/idea is useful for Insurance companies as they consider age, lifestyle choices, family medical history, and several other factors when determining premium rates for individual life insurance policies. The principle of life expectancy suggests that you should purchase a life insurance policy for yourself and your spouse sooner rather than later. Not only will you save money through lower premium costs, but you will also have longer for your policy to accumulate value and become a potentially significant financial resource as you age.

It can be used by researchers to make meaningful research out of it and thus, bring something that will help increase the expectancy considering the impact of a specific factor on the average lifespan of people in a specific country.

8 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.

9 FUTURE SCOPE

As the 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

Project Name: Predicting Life Expectancy using Machine Learning

Kickoff Date: 17-06-2020

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.

APPENDIX
A. Dataset Reference: https://www.kaggle.com/kumarajarshi/life-expectancy-who
B. GUI URL link:
https://node-red-xokmt.eu-gb.mybluemix.net/ui/#!/0?socketid=AGHr_F6pxA-uiFO_AAA6
C. Source Code(Github link):
https://github.com/SmartPracticeschool/llSPS-INT-2973-Predicting-Life-Expectancy-using-Machine-Learning