Predicting Life Expectancy Using Machine Learning

Prepared by: Anjali Singh

1. INTRODUCTION

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.

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 2000to 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.

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.

3. PROJECT REQUIREMENTS:

This project mainly aims at predicting life expectancy. The basic requirement of the project is the availability of the 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 of WHO

◊ Analyze it and clean the dataset

◊ Create IBM account

◊ Create the appropriate cloud and node red services

◊ Train the regression model on different algorithms

◊ Check for the best one and finalize that algorithm to train our mode

◊ Build Node red flow for GUI (web app)

◊ Create scoring end point for integrating our model to node red

◊ Provide the model with the input's fields

◊ The model will return the output as the average predicted lifespan

3.2. Technical Requirements:

◊ The GUI must be integrated with the backend trained model.

◊ The model before training must be given with clean dataset (done by preprocessing)

3.3. Software Requirements:

◊ Python IDE

◊ Excel

◊ IBM Cloud Account

◊ IBM Watson

◊ Node Red

4. FLOWCHART:

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, 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.

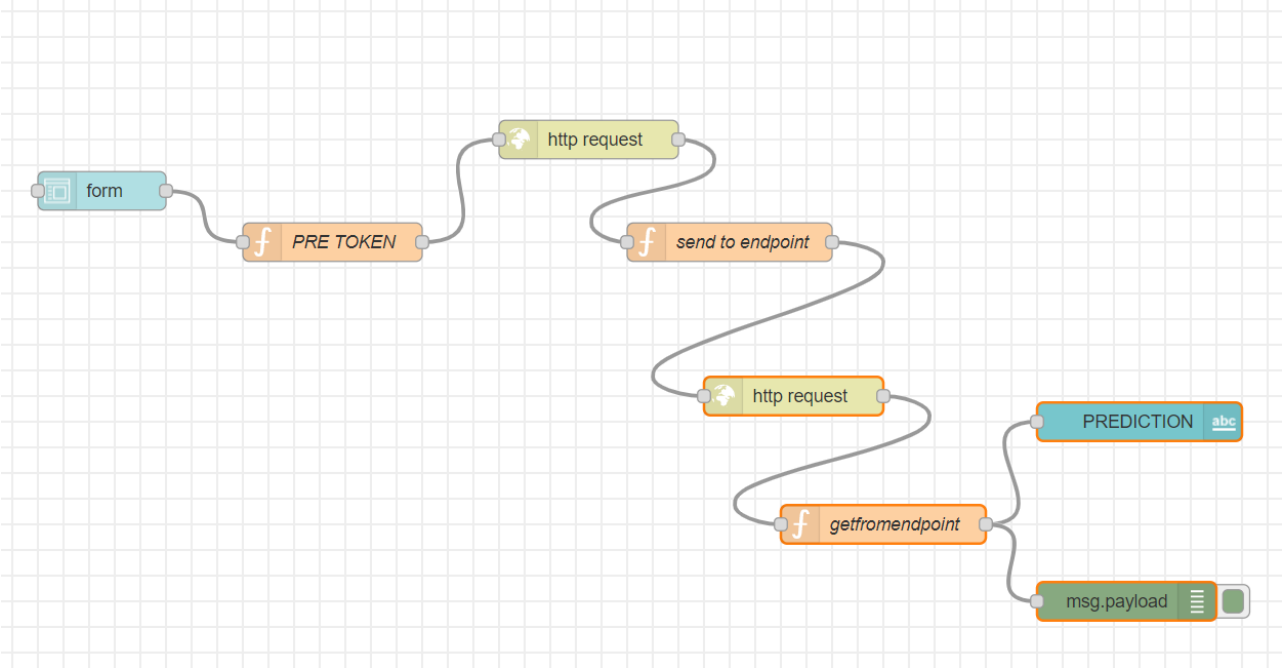

Figure A: Node-red flow

5. RESULT:

The user-friendly Graphical User interface is shown in Figure 2. 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 B.

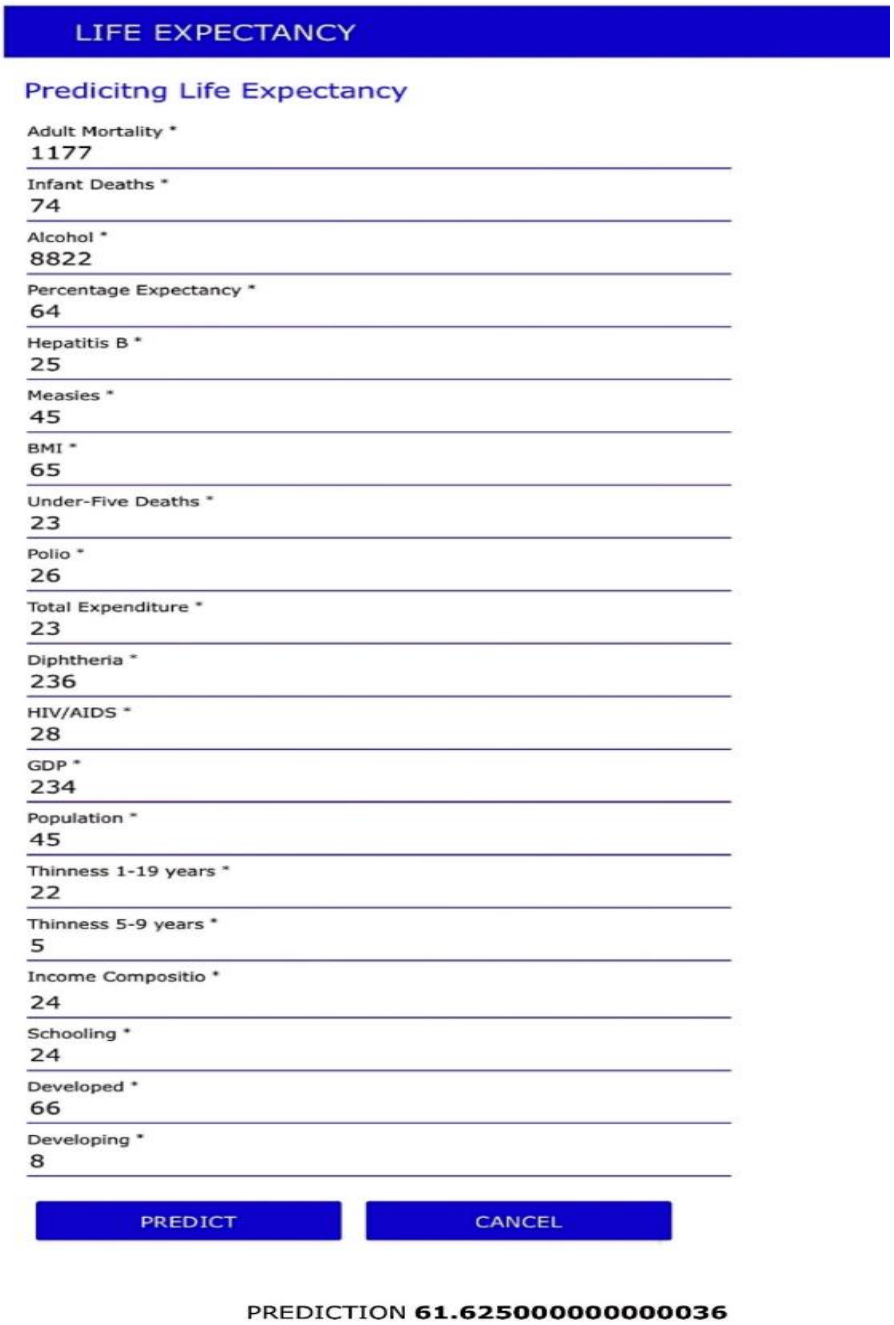

Figure B: Output

6. ADVANTAGES AND DISADVANTAGES:

6.1. 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.

6.2. 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

7. APPLICATIONS

We are also to distinguish different risk factors for life expectancy, such as smoking-status, occupation, socio-economic class, and others. More complex analyses for assessing cancer survival, that involves comparisons between two populations or a population in two points in time can also be undertaken.

In addition to public health domains, life tables are also used by insurance companies and actuary departments.

When used in biology, age specific fertility rates are also included in the calculations.

When data has not been available, such as in low income countries, life tables have been modelled using what data are available, usually childhood mortality data.

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 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.

APPENDIX:

A. **Dataset Reference: https://www.kaggle.com/kumarajarshi/life-expectancy-who**
B. **Drive link: https://drive.google.com/file/d/1q5Tbyyp4OonZ78JtFOkq1FfK0Jsd5AqX/view?usp=sharing**
C. **GUI URL link: https://node-red-dcpai-2020-07-30.eu-gb.mybluemix.net/ui/#!/0?socketid=MpLdFlcmKRJ5TBEMAAAK**

D. **Source Code (In my GitHub link): https://github.com/SmartPracticeschool/llSPS-INT-3508-Predicting-Life-Expectancy-using-Machine-Learning/blob/master/Life%20Expectancy.ipynb**