**Predicting Life Expectancy using Machine Learning**

**By**

**Kedarisetti Vallabh Ashish**

**Under**

**SMARTBRIDGE**

**1.Introduction**

**1.1 Project Overview**

Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age and other demographic factors including gender. 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.

The project aims at comparing such factors and explore the relationship between them by using various machine learning algorithms multiple linear regression, random forest etc. and choose the best or optimum model for prediction. Based on the input of various factors of a country, the target of the model would be to predict an age.

**1.2 Purpose**

This project aims at predicting life expectancy rate of any country given the necessary factors like adult mortality, infant deaths, status etc. If life expectancy of a country is higher it implies that the country is doing well in health and public sectors. So, with the help of machine learning we can predict life expectancy and know what the country has to do to achieve high life expectancy.

**2. Literature Survey**

**2.1 Existing Problem**

Few works have been done to provide an individually customized life expectancy prediction. We have reviewed existing works and techniques in the prediction of human LE, and reached a conclusion that it is feasible to predict a PLE for individuals using evolving technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearable’s and mobile health monitoring devices. We also identified that the collection of data will be 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 a PLE prediction by proposing an approach of data collection and application by smartphone, with which users can enter their information to access the cloud server to obtain their own PLE, was shown.

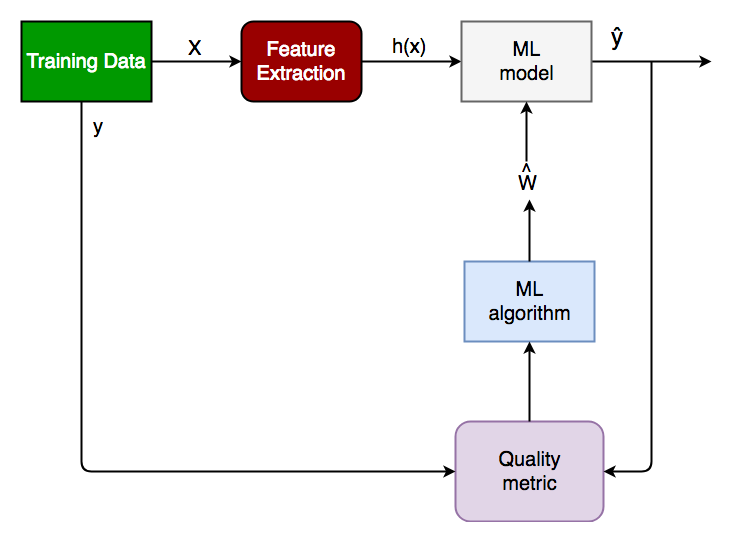
To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing 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 lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that 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 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 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 and hence they need not download any application to predict the results, which saves the storage space as that is the need of the hour.

A comparative study has been performed by using MLR through Jupyter Notebook in IBM Watson Studio and also by using IBM’s Auto AI. Auto AI chooses the best algorithm possible for the dataset and problem statement.

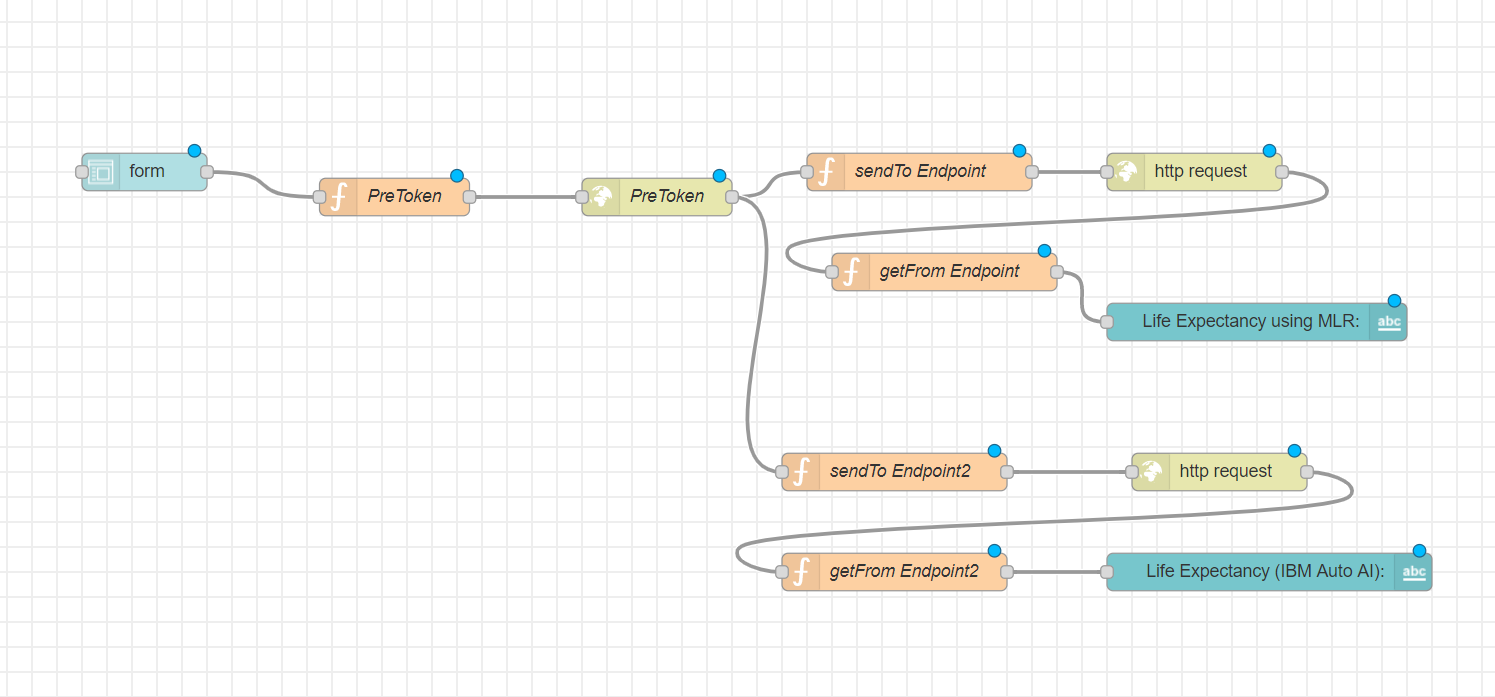
**3.Theoretical Analysis**



*Figure 1: block diagram*

**4. Node-red flow**

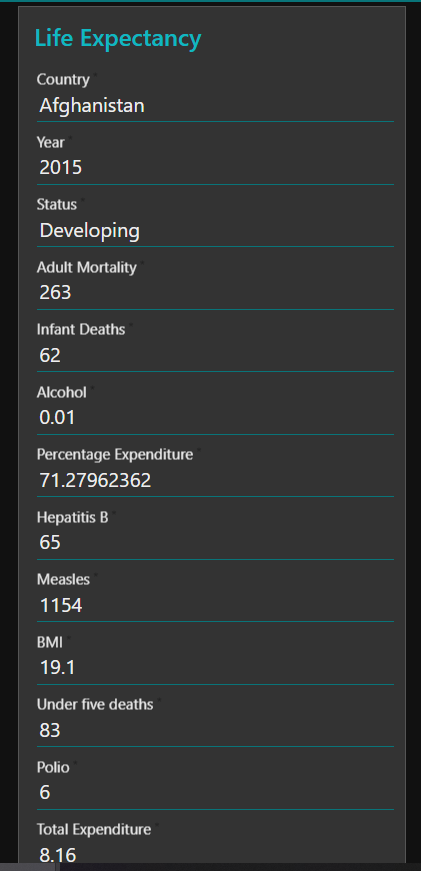
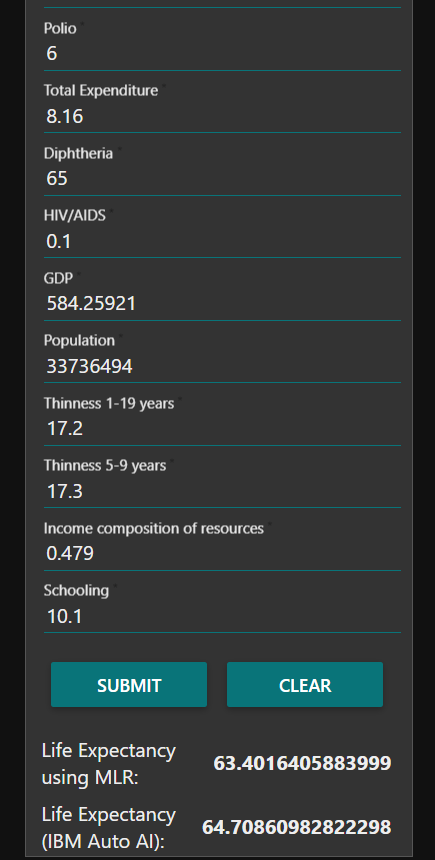
A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to document, study, plan, improve and communicate often complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals,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



*Figure 2: Node-red flow*

**5.Result**

The model appears to the user in the form of an interface as shown in the Figure 2. The user has to fill in the inputs and click on “Submit” button at the end of the form. On clicking the “Submit” button, the user will be displayed the predicted life expectancy, based on the inputs provided, at the top of the page as shown in Figure 3.



*Figure 3: Result*

**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 user 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

• Doesn't process structured data directly

• Increasing rate of data, with limited resources

2. Not connected to database, hence no record of input.

3. Requires internet connection.

**7. APPLICATIONS**

One of the biggest question is “How long can a person live?”. It has given rise to a plethora of religions and spiritual paths over thousands of years, and more recently, some highly amusing apps. This system will be used for people wondering with such questions.

Life expectancy is the primary factor in determining an individual's risk factor and the likelihood they will make a claim. Insurance companies 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 researches out of it and thus, bring about something that will help increase the expectancy consider 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, and Alcohol, HIV, Hepatitis B, GDP, Percentage Expenditure and many more. User can interact with the system via a simple user interface which is in the form of a form with input spaces which the user needs to fill the inputs into.

**9. FUTURE SCOPE**

As future scope, we can connect the model to the database to have the record of predictions. This will help us analyse the trends in the life span.

A model with country wise bifurcation can be made, which will help to segregate the data demographically.