Predict Life Expectancy Using Machine Learning

Project Report

(By: Amarjeet Singh)

1. Introduction

1.1 Overview

Life expectancy refers to the number of years a person is expected to live. In mathematical terms, life expectancy refers to the expected number of years remaining for an individual at any given age. Based on actuarial science, the life expectancy for a particular person or a population group depends on several individual-level as well as population-level factors such as a person't lifestyle, historical mortality data of country, etc. However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

To predict the life expectancy rate of a particular country, we will be using machine learning to draw inferences from the given dataset and give a prediction. We will also be creating a UI using Node-RED for making the model accesible to general users.

1.2 Purpose

Economic growth

Predicting life expectancy plays a vital role in judging the growth and development of the economy. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country.

Population Growth

Helps the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skillset acquired by people over many years.

Personal growth

This project would also help an individual assess his/her lifestyle choices and alter them accordingly to lead a longer and healthier life. It would make them more aware of their general health and its improvement or deterioration over time.

Growth in Health Sector

Basedonthefactorsusedtocalculatelifeexpectancyofan individualandtheoutcome, health care will be able to fund and provide better services to those with greaterneed.

Insurance Companies

Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

2. Literature Survey

2.1 Existing Problem

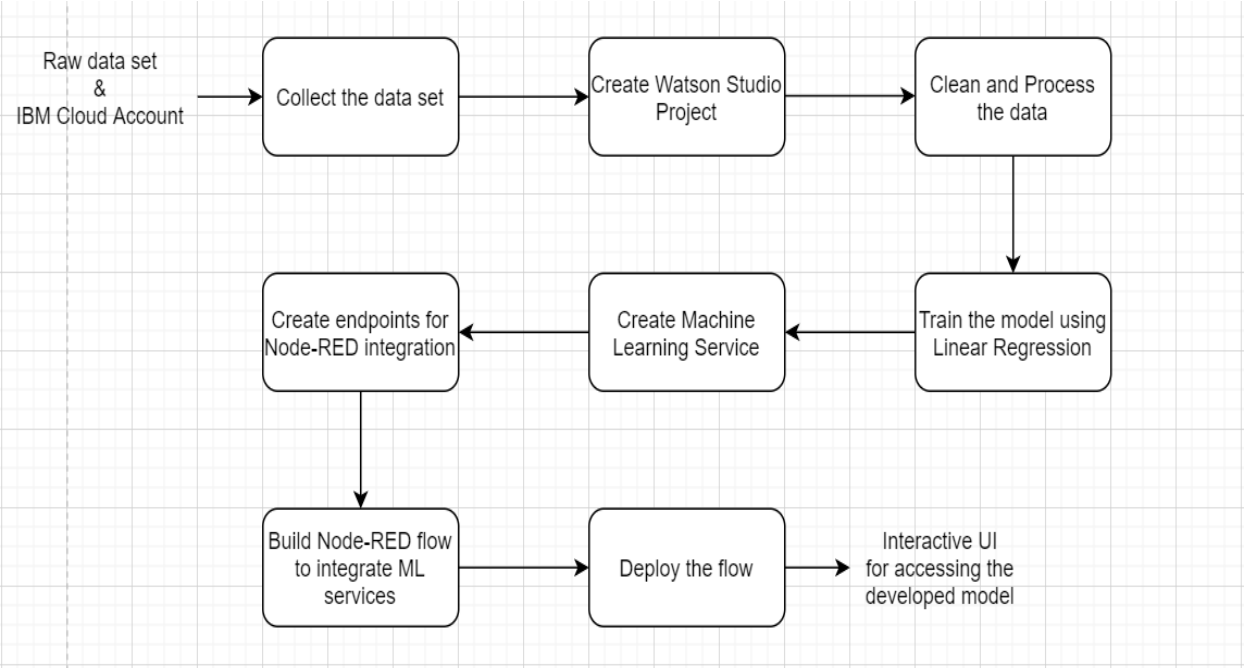
The World Health Organization (WHO) began producing annual life tables for all member states in 1999. These life tables are a basic input to all WHO estimates of global, regional and country-level patterns and trends in all-cause and cause-specific mortality. After the publication of life tables for years till 2009, in the 2011 edition of World Health Statistics, WHO has shifted to a two year cycle for the updating of life tables for all member states. Even still the model is not really updated in every fields. WHO applies standard methods for the analysis of member state data to ensure comparability of estimates across countries. This will inevitably result in differences for some member states with official estimates for quantities such as life expectancy, where a variety of different projection methods and other methods are used.

2.2 Proposed Solution

I will be building a solution using Machine Learning to predict the life expectancy given a set of inputs. Specifically, linear regression model will be used. We will train the model on a given WHO dataset with real values from the past. It will contain some information like the status of the country, adult mortality rate, infant deaths, alcohol, hepatitis B, measles, BMI, etc. This trained model will then be able to give a prediction for given input data points. Finally, this model will be deployed and made accessible to users using an interactive UI generated using the Node-RED service.

3. Theoretical Analysis

3.1 Block Diagram



3.4 Hardware/Software Designing

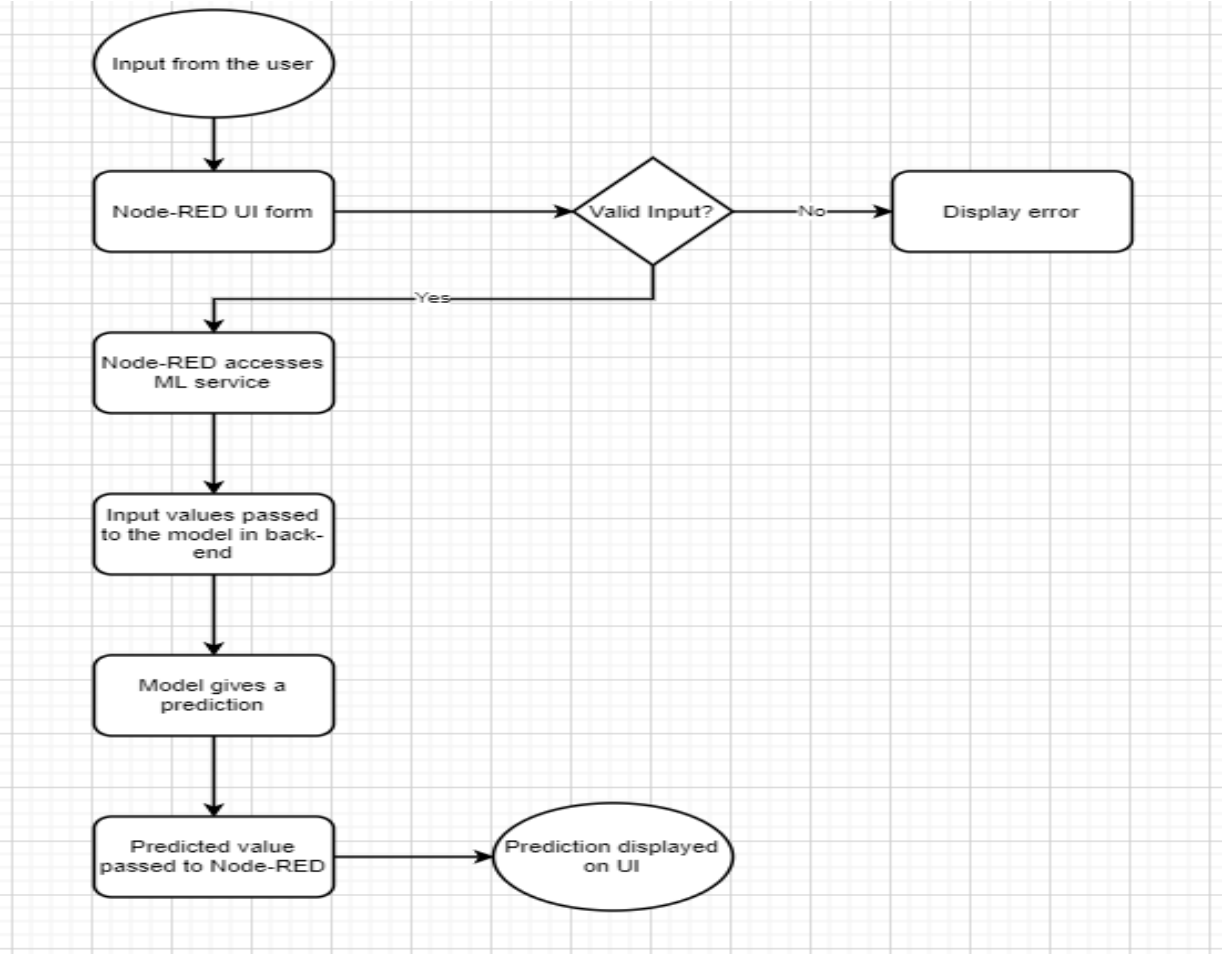
The project doesn't explicitly have any special hardware requirements apart from a basic PC with a browser and good internet connection.
The following are the software requirements:

- IBM Watson Studio
- Jupyter Notebook
- Machine Learning Service
- IBM Node-RED service

4. Experimental Investigations

The WHO data set for life expectancy contains 22 columns in total. The data set has historical data of life expectancy for the period between 2000 to 2015.
There are numerous rows with null values for some of the columns and outliers as well. These null values will be replaced with the mean for the respective columns and the outliers will be adjusted. Also, the "Year" column is not relevant to our model.
All the columns except the "Country" and "Status" are of integer type. The country column is not considered into the model training because of its less relation to the "Life Expectancy" column. The "Status" column is changed to integer type such that "Developing" is mapped to 1 and "Developed" is mapped to 0. After these steps, the data is ready for the model to be trained.

5. Flowchart



6. Result

A machine learning model developed using Linear Regression and IBM Cloud and its services. The model has a high accuracy.
An interactive UI deployed using the Node-RED service for the user to interact with the model in the back-end and get predictions for a given set of input values.

7. Advantages & Disadvantages

7.1 Advantages

The biggest advantage of using Machine Learning for predictions is that it learns by itslef. It could evolve over time as more data becomes available. The algorithm analyzes the trends in the data and adjusts itself to increease the acuracy.

The services provided by IBM cloud reduces the over head of learning different skill sets for the development of the application and its deployment as well. Node-RED makes it really easy to design a flow and deploy the model with an interactive & user-friendly UI.

7.2 Disadvantages

It is a well-know fact that machine learning is a boon as well as a bane for a developer. Due to minimal human intervention, it becomes difficult to track down the causes of erros and resolve them. Furthermore, there can be cases when the model takes a lot of time for training due to the preence of humongous amount to data limited computing power. Though Node-RED allows for easy deployment of UI for a model, there are limited number of customizations available for a developer.

8. Applications

Personalized Life Expectancy

Individuals can predict their own life expectancy by inputting values in the corresponding fields. This could help make people more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.

Government

It could help the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skillset acquired by people over many years. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country. The knowledge asset of a country increases with the number of individuals in a country.

Health Sector

Based on the factors used to calculate life expectancy of an individual and the outcome, health care will be able to fund and provide better services to those with greater need.

Insurance Companies

Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

9. Conclusion

Life Expectancy prediction can play a vital role in numerous sectors of the industry and machine learning can be used to achieve this with a high level of accuracy.

10. Future Scope

The application can be further improved by integrating it with the functionality to provide suggestions and medications to the individuals using the application. This will help predict as well as increase the individual's life expectancy.
The scalability and flexibility of the application can also be improved with advancement in technology and availability of new and improved resources.
With the growth in Artificial Intelligence and Computer vision, we can also try to take into account the physical health and appearance of a person for predicting the life expectancy. Mental health can also be considered as a major factor while predicting life expectancy with the help of sentiment analysis systems.

11. Bibliography

- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application
- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html
- https://www.kaggle.com/kumarajarshi/life-expectancy-who
- https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2
- https://www.youtube.com/watch?v=Jtej3Y6uUng
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service

12. Appendix

Source Code: *Here*