FINAL REPORT

ON

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

BY

IRISH MEHTA

Prepared in partial fulfilment of the
Smartbridge internship procedure,
20th June, 2020

SMARTBRIDGE
Let's Bridge the Gap

IN PARTNERSHIP WITH

Table of Contents

1.)PROJECT INTRODUCTION & SUMMARY

Life expectancy is a measure of various factors that determine the expected number of years a person is predicted to live. There are mainly two kinds of life-expectancy evaluations to be done according W.H.O

The project is aimed at creating a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years, implemented on IBM cloud using NODE RED integration for the endpoints and using AUTO AI for training without the code. The data offers a timeframe from 2000 to 2015. The data being used to create the model is available on kaggle at the following link. There are many existing models based on the same dataset which utilize different supervised learning algorithms like regression and xgboost.

1.1 Overview

Regression Machine Learning project analyses historical data predict insights into the future. This problem statement is aimed at predicting Life Expectancy of an average person of a country on basis of given features.

1.2 Purpose

Life expectancy is a statistical measure of the average time a human being is expected to live, and 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 predicting 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.

Economic growth

Predicting life expectancy would allow a better flow of economy by allocation of correct resources and will help it grow

Population Control and Workforce

A lot of countries are facing either a population scarcity or overpopulation, and having an estimate of duration of each citizen allows a better planning in terns of infrastructure and

countries having a higher life expectancy tend to have a better ability to grow because as population increases, so does the pool of knowledge

Benefit for health and insurance sectors

Insurance companies can roll out tailored policies and health professionals can guide patients to live a healthy lifestyle, along with providing adequate support to those who require it the most. Governments may be able to use predictions to more efficiently allocate limited resources, such as social welfare assistance and health care funding, to individuals and areas of greater need

2.)LITERATURE SURVEY

2.1 Existing problem

Predicting life expectancy is not a new concept. Experts do this at a population level by classifying people into groups, often based on region or ethnicity.
Despite its importance and prominence in research and policy, it is surprisingly difficult to find a simple yet detailed solution for predicting life expectancy

The **cohort life expectancy** is the average life length of a particular cohort – a group of individuals born in a given year.

Period life expectancy is estimating the average length of life for a *hypothetical* cohort assumed to be exposed, from birth through death, to the mortality rates observed at one particular *period* – commonly a year. This is the common definition used by most of the worldwide organizations like WHO

An important point to bear in mind when interpreting life expectancy estimates is that very few people will die at precisely the age indicated by life expectancy, even if mortality patterns stay constant.

For example, very few of the infants born in South Africa in 2009 will die at 52.2 years of age, as per the figures in the map above. Most will die much earlier or much later, since the risk of death is not uniform across the lifetime. Life expectancy is the average.

In societies with high infant mortality rates many people die in the first few years of life; but once they survive childhood, people often live much longer. Indeed, this is a common source of

confusion in the interpretation of life expectancy figures: It is perfectly possible that a given population has a low life expectancy at birth, and yet has a large proportion of old people.

People may become distressed if their life expectancy is unexpectedly low, or at the thought of having one at all. This raises concerns about how such predictions could impact those who experience or are at risk of mental health problems.

Having people's detailed health data could also let insurance companies more accurately profile applicants, leading to discrimination against groups or individuals.

Also, pharmaceutical companies could coordinate targeted medical campaigns based on people's life expectancy. And governments could choose to tax individuals differently, or restrict services for certain people.

2.2 Proposed solution

With machine learning and artificial intelligence, it's becoming feasible to analyse larger quantities of data. The use of deep learning and cognitive computing, such as with IBM Watson, helps doctors make more accurate diagnoses than using human judgement alone.

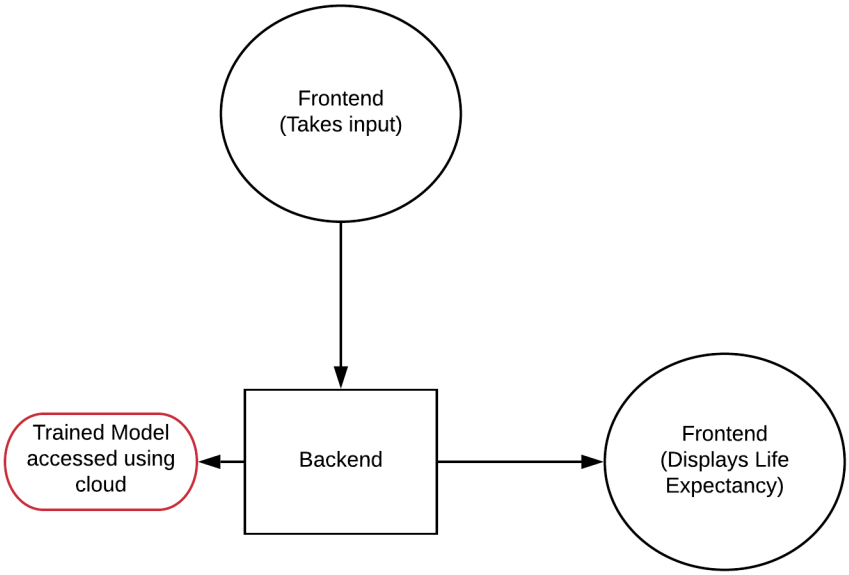
Human behaviour and activities are so unpredictable, it's almost impossible to measure, classify and predict lifespan. A personal life expectancy, even a carefully calculated one, would only provide a "natural life expectancy" based on generic data optimised with personal data. The key to accuracy would be the quality and quantity of data available. Much of this would be taken directly from the user, including gender, age, weight, height and ethnicity.

Access to real-time sensor data through fitness trackers and smart watches could also monitor activity levels, heart rate and blood pressure. This could then be coupled with lifestyle information such as occupation, socioeconomic status, exercise, diet and family medical history.

The proposed solution involves the use of Machine Learning algorithms specifically Regression models such as Linear Regression, Ridge regression, etc. Life expectancy is highly correlated over time among countries and between males and females. These associations can be used to improve forecasts. Here we propose a method for forecasting life expectancy of an individual from a country taking into certain factors such as Adult Mortality rate, Infant deaths, Alcohol, Hepatitis B, Measles, BMI, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP of a country, Population, Income composition of resources, Schooling and status of the country in terms of Developing or Developed

3.)THEORITICAL ANALYSIS

3.1 Block diagram



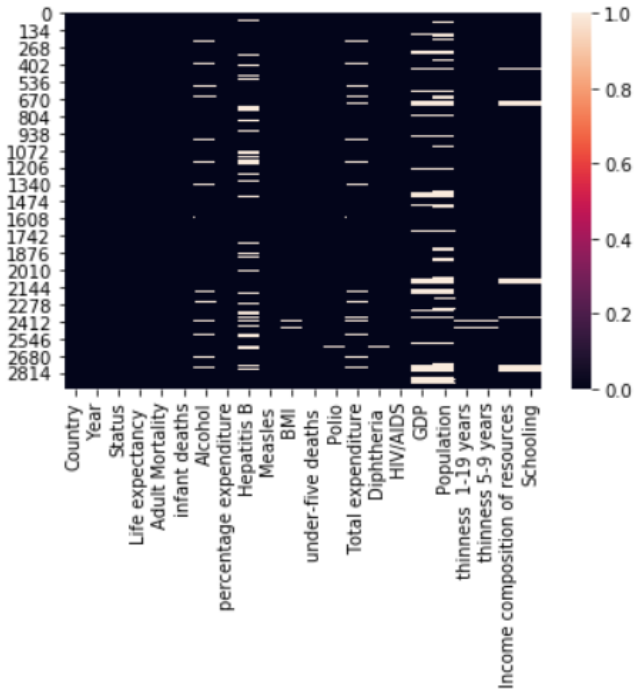
3.2 Hardware / Software designing

- Working internet connection with access to IBM cloud
- A working knowledge of the entire ML pipeline from data cleaning to deploying the model on NODE RED

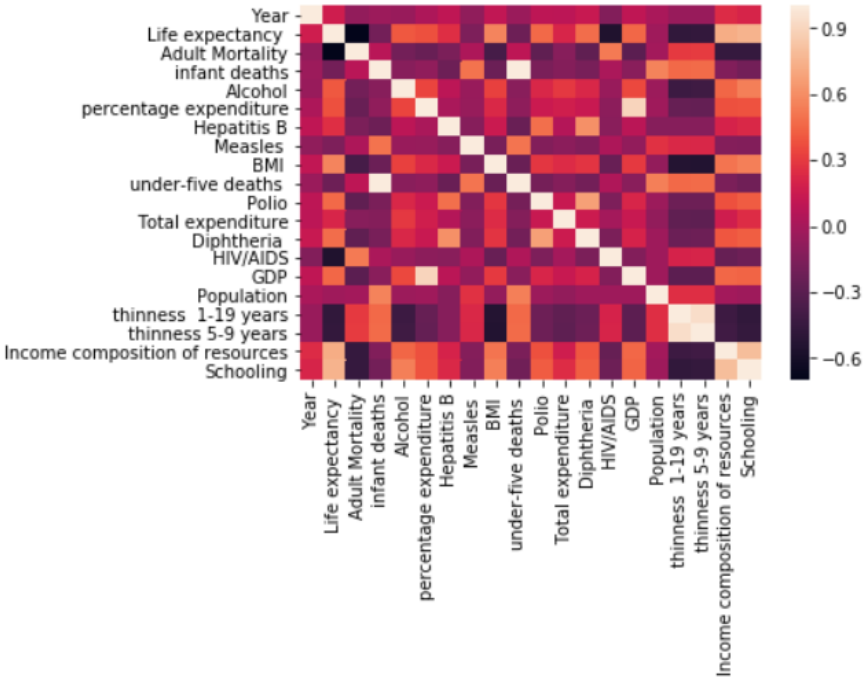
- Jupyter Notebook (with Python 3.6+)
- IBM Cloud services
- Slack (for communication)

4.)EXPERIMENTAL INVESTIGATIONS

While performing some cleaning steps on the data, these were he null values present in the dataset

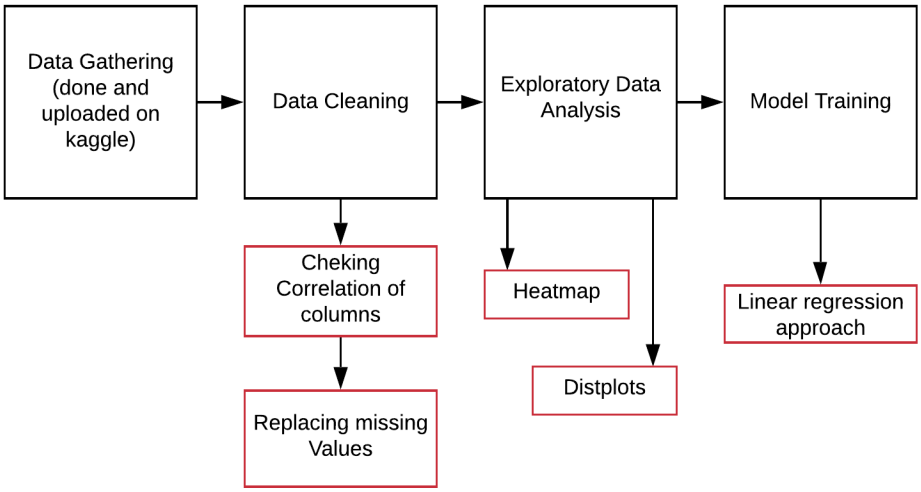


This was the heatmap between the existing columns, helps to figure out the dependencies of each column

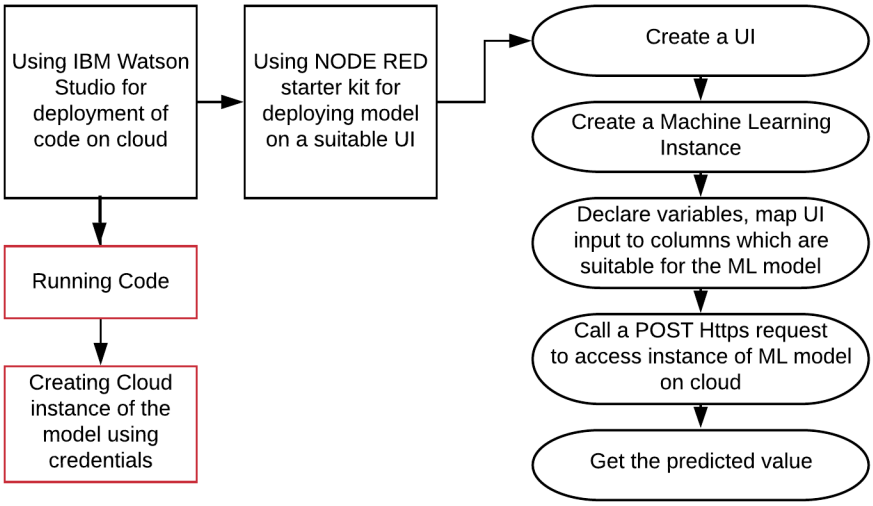


5.)FLOWCHART

Stage 1



Stage 2

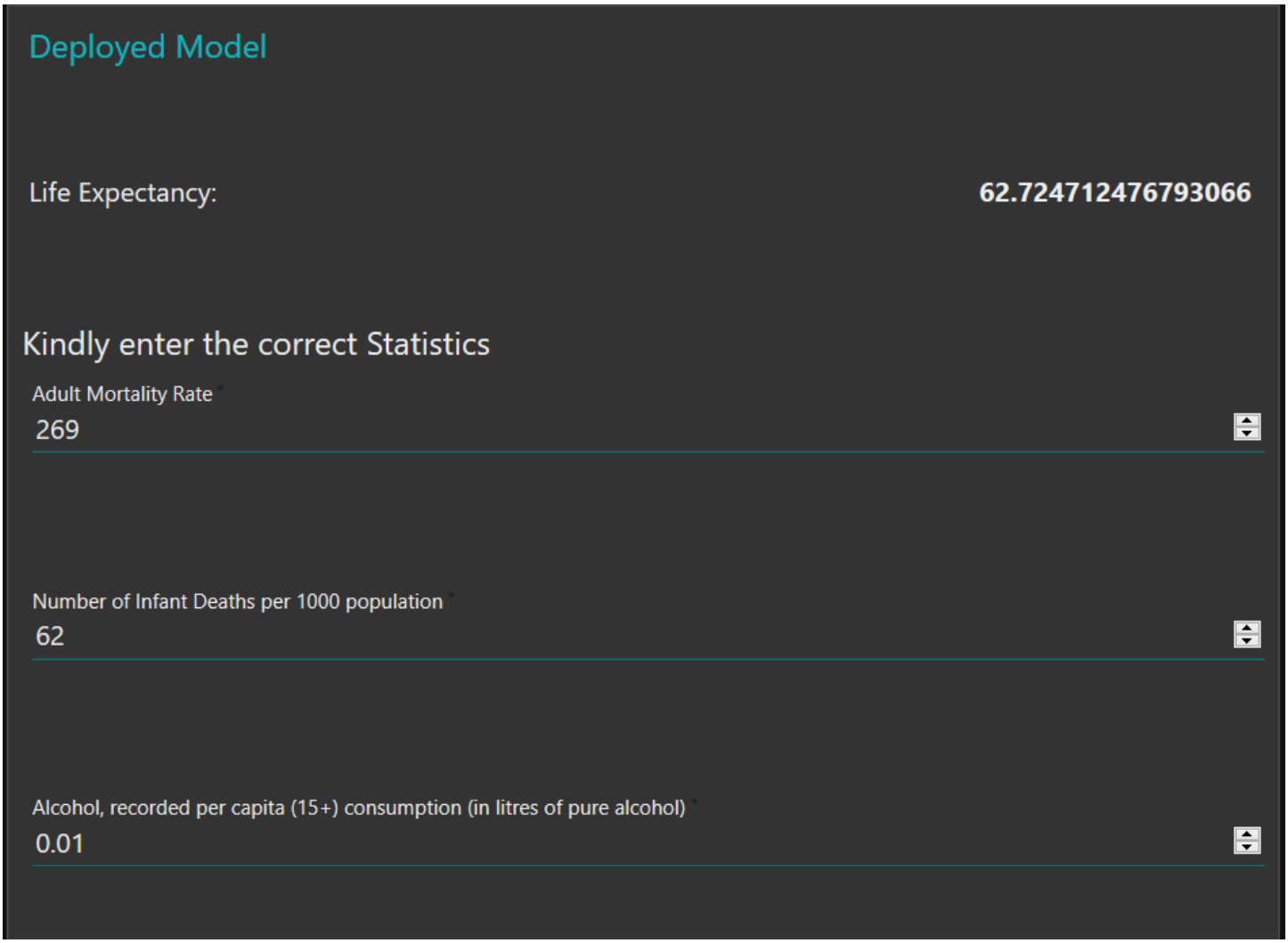


6.)RESULT

The final deployment can be viewed on this link

https://node-red-xskfb.eu-gb.mybluemix.net/ui/#!/0?socketid=OrQbhOPGTo_993iMAAAA

Briefly, it looks like this:



7.)ADVANTAGES & DISADVANTAGES

Advantages

One of the biggest advantages of embedding machine learning algorithms is their ability to improve over time. Machine learning technology typically improves efficiency and accuracy thanks to the ever-increasing amounts of data that are processed. The application learns the patterns and trends hidden within the data without human intervention which makes predicting much simpler and easier. The more data is fed to the algorithm, the higher the accuracy of the algorithm is. It is also the key component in technologies for automation. Using Node-Red also simplifies the effort put into a creating the front-end. The programmer doesn't need extensive knowledge on HTML and JavaScript. It also makes the integration between Machine learning model and the UI much easier.

Disadvantages

- People may become distressed if their life expectancy is unexpectedly low, or at the thought of having one at all. This raises concerns about how such predictions could impact those who experience or are at risk of mental health problems.
- Having people's detailed health data could also let insurance companies more accurately profile applicants, leading to discrimination against groups or individuals.
- Also, pharmaceutical companies could coordinate targeted medical campaigns based on people's life expectancy. And governments could choose to tax individuals differently, or restrict services for certain people.

8.)APPLICATIONS

1) Personalized Life Expectancy: It would make people more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.

2) Government: Governments may be able to use predictions to more efficiently allocate limited resources, such as social welfare assistance and health care funding, to individuals and areas of greater need

3) 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.

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

9.)CONCLUSION

Predicting lifespan of human beings can greatly alter our lives. Human behavior and activities are so unpredictable, it may almost be impossible to correctly predict lifespan. However, with the help of Machine learning algorithms such as Regression models, we can get close to predicting a roundabout value. This breakthrough can widely impact health sectors and economic sectors by improving the resources, funds and services provided to the common people. It can also increase the ease of access to the individuals. With the help of Machine Learning algorithms, one can ease the process of automating the application and predicting the expectancy with an admirable accuracy. It also reduces the effort and time put into deploying the application and making it more accessible to the users.

10.)FUTURE SCOPE

With machine learning and artificial intelligence, it’s becoming feasible to analyse larger quantities of data. The use of deep learning and cognitive computing, such as with IBM Watson, helps doctors make more accurate diagnoses than using human judgement alone.

This, coupled with predictive analytics and increasing computational power, means we may soon have systems, or even apps, that can calculate life expectancy.

While the collection of enough data will be challenging, we can likely expect to see advances in this area in the coming years.

If so, issues related to data compliance, as well and collaboration with government and state agencies will need to be carefully managed. Any system predicting life expectancy would handle highly sensitive data, raising ethical and privacy concerns.

It would also attract cybercriminals, and various other security threats.

BIBILOGRAPHY

https://www.kaggle.com/kumarajarshi/life-expectancy-who

Kent M, Vickers AJ. A systematic literature review of life expectancy prediction tools for patients with localized prostate cancer. Journal of Urology 2014: epub.

https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted

https://www.kaggle.com/najeedosmani/96-r2-score-using-linear-regression

APPENDIX

The source code can be found in the GitHub Repository, link to which is:

https://github.com/SmartPracticeschool/llSPS-INT-2347-Predicting-Life-Expectancy-using-Machine-Learning