PREDICTING LIFE EXPECTENCY USING MACHINE LEARNING

MACHINE LEARNING PROJECT REPORT

TEJAS K N

1. INTRODUCTION

1.1 Overview

With the emergence of technologies such as electronic health and mobile health,(cloud computing, big data, and the Internet of Things (IoT), health related data are increasing and many applications such as smartphone apps and wearable devices that provide wellness and fitness tracking are entering the market. Some apps provide health related data such as sleep monitoring, heart rate measuring, and calorie expenditure collected and processed by the devices and servers in the cloud. These requirements can be extended to provide a personalized life expectancy (PLE) for the purpose of wellbeing and encouraging lifestyle improvement. No existing works provide this PLE information that is developed and customized for the individual. This project is based on the concurrent models and methodologies to calculate and predict life expectancy (LE) and proposes an idea of using multi-phased approaches to the solution as the project requires an immense and broad range of work to accomplish. As a result, the current prediction of LE, which was found to be up to a maximum of five years could potentially be extended to a lifetime prediction by utilizing generic health data. In this project, the novel idea of the solution proposing a PLE on an individual basis, which can be extended to life time is presented in addition to the existing works.

1.2 Purpose

The main purpose of this project is Predicting Life Expectancy using Machine Learning deliver the outcome for the given dataset of any county. It will predict the life expectancy based on the – • 'Country', • 'Year', • 'Status' • 'Life expectancy ' • 'Adult Mortality' • 'infant deaths' • 'Alcohol' • 'percentage expenditure' • 'Hepatitis B' • 'Measles ' • ' BMI ' • 'under-five deaths ' • 'Polio' • 'Total expenditure' • 'Diphtheria ' • ' HIV/AIDS' • 'GDP' • 'Population' • ' thinness 1-19 years' • ' thinness 5-9 years' • 'Income composition of resources' • 'Schooling'.

2. LITERATURE SURVEY

2.1 Existing system

As a result of the evolution of biotechnologies and related technologies such as the development of sophisticated medical equipment, humans are able to enjoy longer life expectancies than previously before. For example, a clinical research center claims that in 10 to 12 years from now, for every year that humans live, science is extending the life for more than a year using health intelligence platform integrating genomics, advanced clinical imaging and robust machine learning in a spa-like setting. Predicting a human's life expectancy has been a long-term question to human kind , and there have been many attempts to make the prediction accurate and popular since the prevalence of smart phones and apps. However, the effectiveness of those apps is limited due to the constraints of developing a classification of meta-data, such as the complexity and variety of environmental, geographic, genetic, and living factors of humans. For example, a report showed that people living in a village called Yuzurihara in Japan, also known as "the village of long life", were ten times more likely to live beyond the age of 85 than anywhere in North America. These people also had similar traits such as smooth skin, flexible joints and thick hair . This implies that geographic and living environments affect the longevity of human life, and the use of statistics can make it possible to forecast a life expectancy of a person who lives in a similar environment village with a similar lifestyle.

2.1 Proposed system

Following are the features of the data: • 'Country', • 'Year', • 'Status' • 'Life expectancy ' • 'Adult Mortality' • 'infant deaths' • 'Alcohol' • 'percentage expenditure' • 'Hepatitis B' • 'Measles ' • ' BMI ' • 'under-five deaths ' • 'Polio' • 'Total expenditure' • 'Diphtheria ' • ' HIV/AIDS' • 'GDP' • 'Population' • ' thinness 1-19 years' • ' thinness 5-9 years' • 'Income composition of resources' • 'Schooling' Target is Life Expectancy, measured in number of years. The assumptions are:

1. These are country level average .

2. There is no distinction between male and female.

Proposed system is a machine learning trained model with 95% of accuracy.

3. THEORETICAL ANALYSIS

3.1 Hardware /software designing

Predicting Life Expectancy using Machine Learning is based on the supervised machine learning using linear regression algorithm. we need following software to develop the project – • IDE(Integrated development environment Software) for write the python code . • Database software to stone the dataset of the project . • Communication software to communicate with the team members and discuss the problem and future planes of the project.

3.2 Data Analysis

Following are some graphs generated using the refine data set .
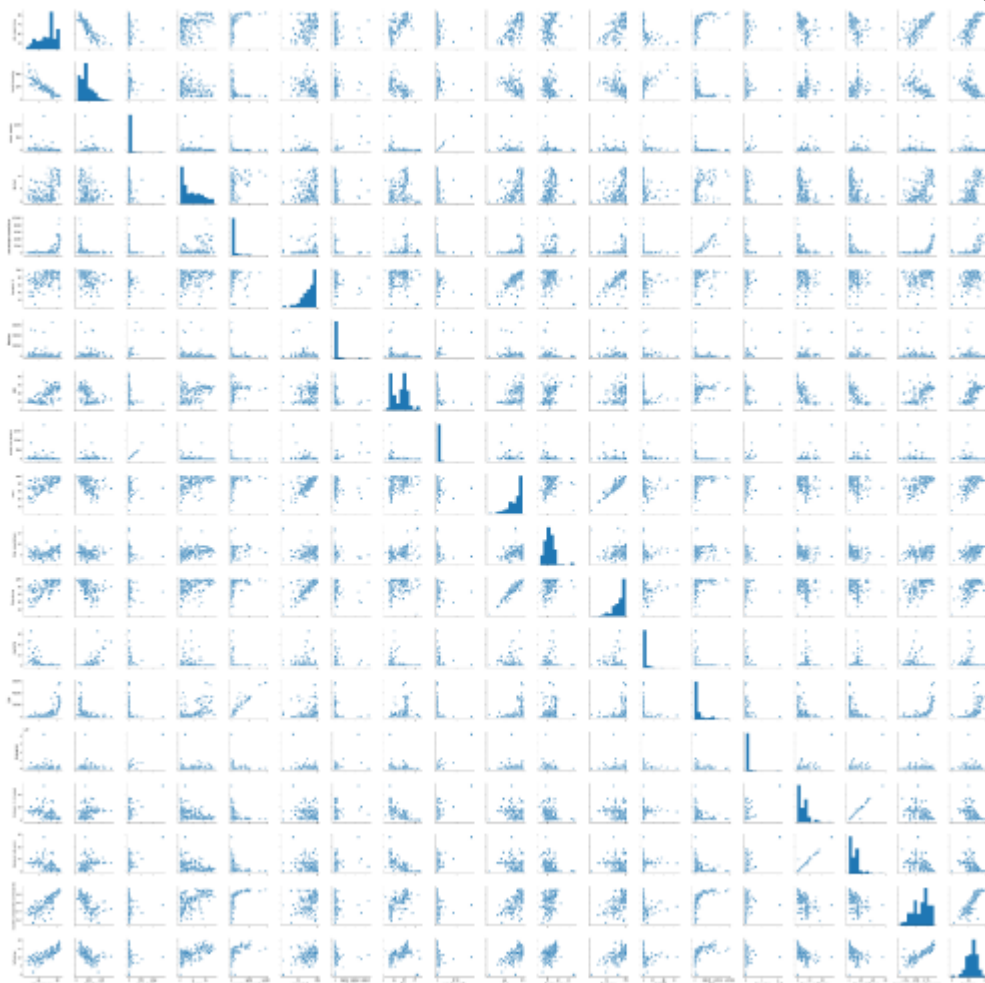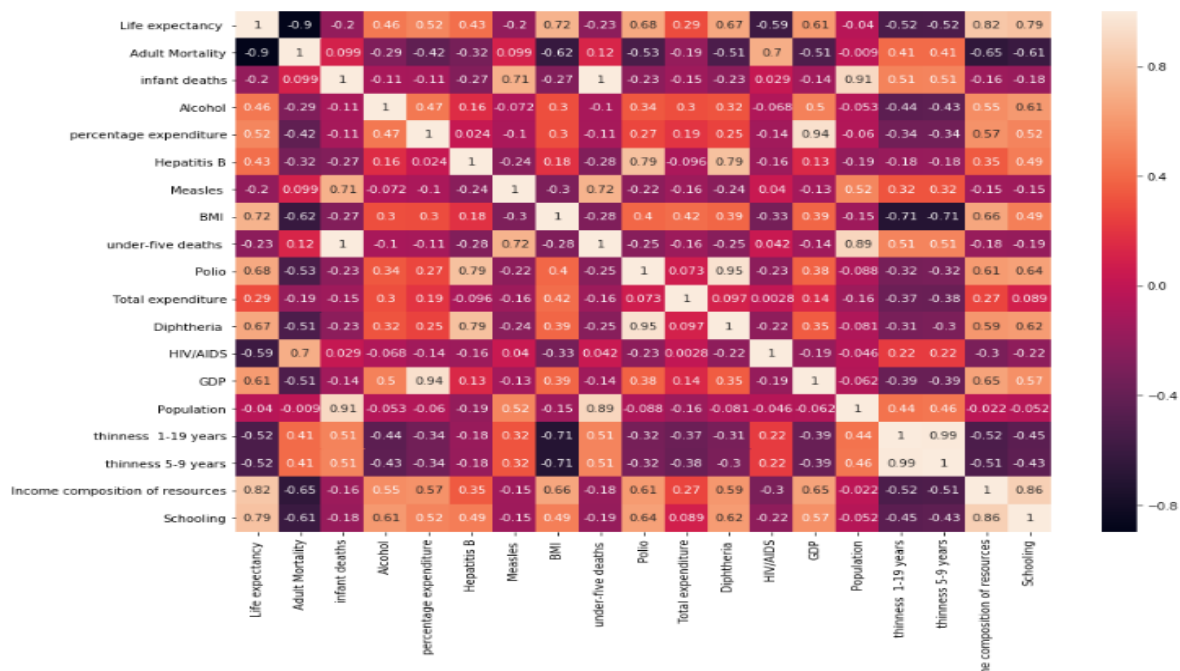
4.EXPERIMENTAL INVESTIGATIONS

Some random inputs are given to the deployed machine learning model. We got the following output.

Default

Prediction   55.367250915750915

Adult Mortality *
12

infant deaths *
33

Alcohol *
0

percentage expenditure *
43

Hepatitis B *
23

Measles *
1234

BMI *
1

under-five deaths *
4

Polio *
1

5.RESULT

Based on the given data, the machine learning model understands the data and cross reference the data to watch what are the factors that are affecting the results we require i.e life expectancy. Then when we give any input, it has already run algorithm to get the output based on previously given data.

So the results we get are approximations, they are not definitely true, but it works in maximum number of cases, except for some exceptional ones.

6.ADVANTAGES AND DISADVANTAGES

● Since we can predict the life span, we can know what factors are influencing the expectancy on life span in what ways.

● So, therefore by trying to change those factors in the real world we can increase the life span.

7.APPLICATIONS

I could possibly collect more data by expanding the scope to cities instead of countries, and to explore other features (factors) affecting life expectancy. Also, I could split the data to male and female categories for such life expectancy regression analysis. To conclude, here are some interesting insights:

1. Japan has the highest life expectancy (83.7 years). Central African Republic (49.5 years) and many countries in the African continent are at the bottom of scale. Singapore is ranked #5 (82.7 years).

2. Take good care of the environment. It has the largest coefficient (impact) on the country's life expectancy.

9. SCOPE

The problem of processing datasets such as electronic medical records(EMR) and their integration with genomics, environmental factors, socioeconomic factor and patient behavior variations have posed a problem for researchers the health industry. Due to rapid innovations in machine learning field such as big data, analytics, visualization, deep learning, health workers now have improved way of processing, and developing meaningful information from huge datasets that have been accumulated over many years . Big data and machine learning can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy. We can use demographics of selected regional areas and multiple behavioral health disorders across regions to find correlation between individual behavior indicators and behavioral health outcomes.

10.APPENDIX

o WEBPAGE

- https://node-red-repcv.eugb.mybluemix.net/ui/#!/0?socketid=iIUkpZfOA-UchtZyAAAF
- WHO data for life expectancy of different country.

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