

## **1. INTRODUCTION**

Predicting Life Expectancy using Machine Learning A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features.

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.

## **2. LITERATURE SURVEY**

I explored life expectancy and looked for data on the following aspects (features):

- Birth Rate
- Cancer Rate
- Dengue Cases
- Environmental Performance Index (EPI)
- Gross Domestic Product (GDP)
- Health Expenditure
- Heart Disease Rate
- Population
- Area
- Population Density
- Stroke Rate

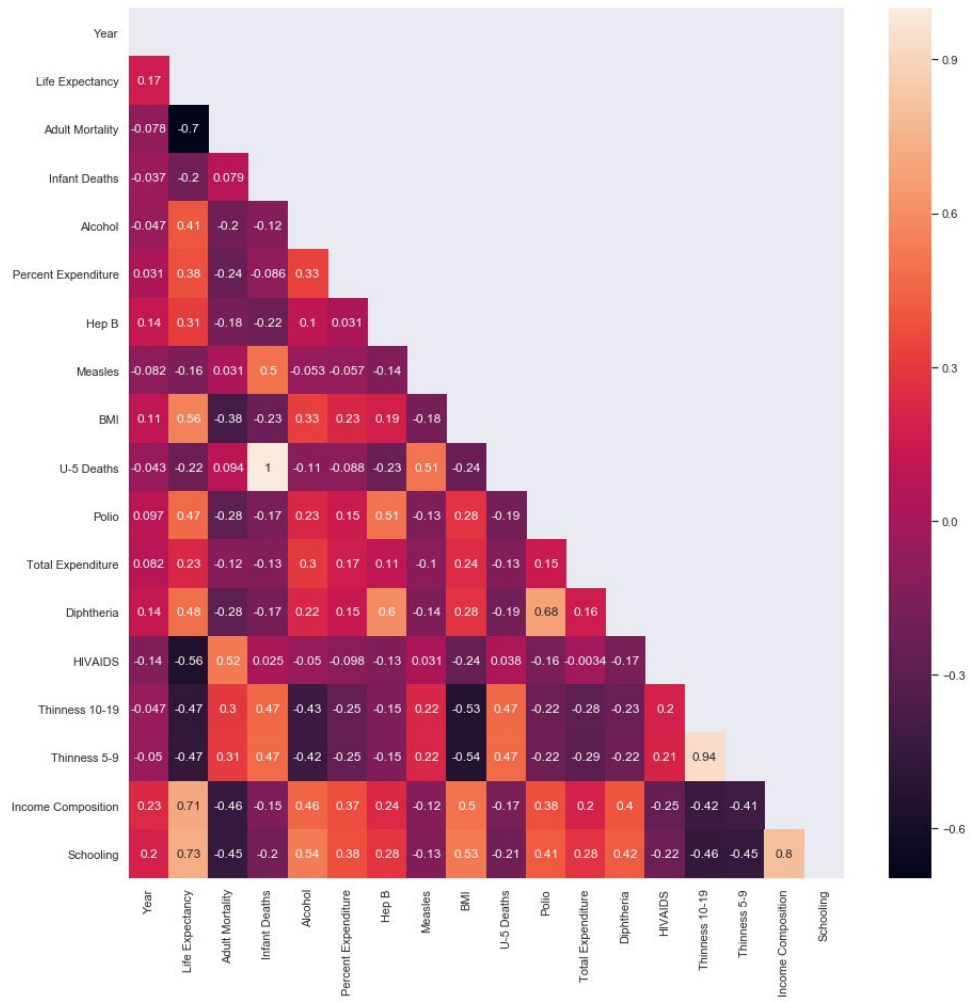
Target is Life Expectancy, measured in number of years.

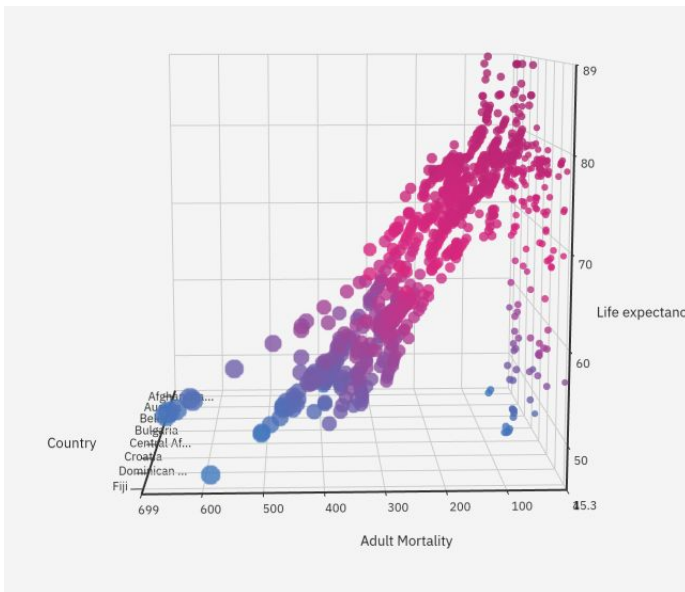
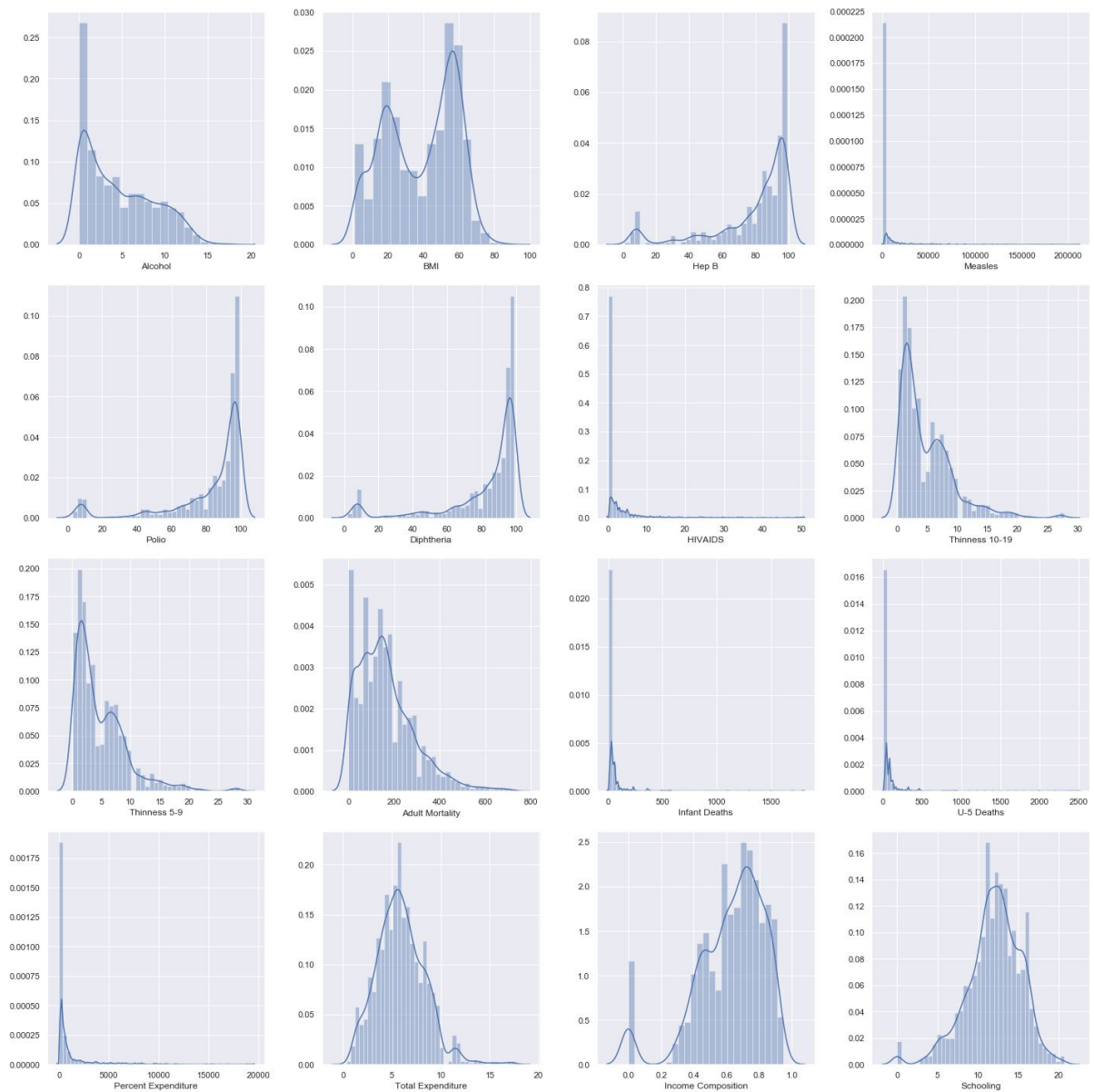
The assumptions are:

1. These are country level average
2. There is no distinction between male and female

## **3. DATA ANALYSIS**

These are some graphs from the refined data analysis, that makes us understand collinearity.





## 4. EXPERIMENTAL INVESTIGATIONS

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

Life expectancy

ML experiment

EXPECTED LIFE SPAN  
75.31082014620361 years

Enter details to Predict life span expectancy

Country

India

Year

2005

Status

Developing

Adult mortality

56

Infant deaths

0

Alcohol

0.05

Percentage expenditure

5628.53

Hepatitis B

25

Measles

54

BMI

25

under-five deaths

2

Polio

35

Total expenditure

9.56

Diphtheria

6

HIV/AIDS

0.5

GDP

55258685

Population

575155328862

thinness 1-19 years

5.6

thinness 5-9 years

5.8

Income composition of resources

0.856

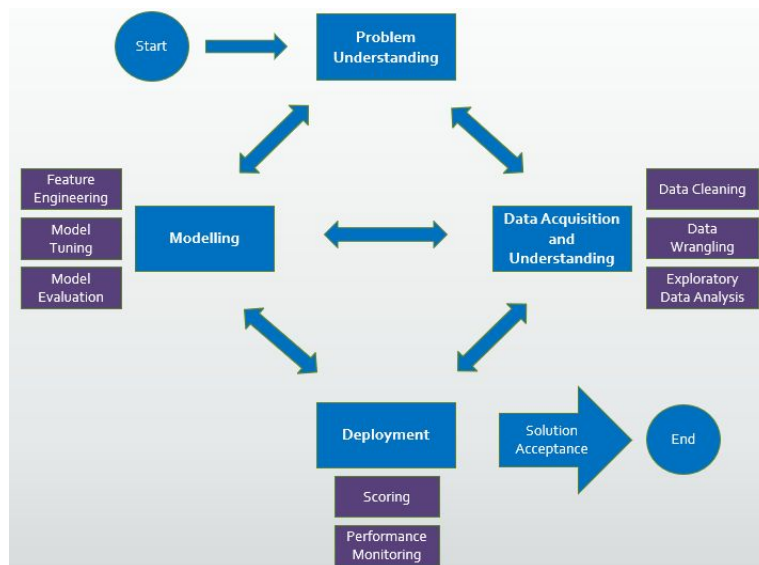
Schooling

20

SUBMIT

CANCEL

## 5. FLOW CHART



## 6. RESULT

Based on the given data, the auto ai 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.

## 7. 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.

## 8. 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.

## **APPENDIX**

- A. WEBPAGE <https://node-red-ml.eu-gb.mybluemix.net/ui>
- B. YouTube link for demonstration vedio<https://youtu.be/yUe2juE-AMl>