

1. INTRODUCTION

Prediction of life expectancy is difficult for humans. Our research shows that machine learning and natural language processing techniques offer a feasible and promising approach to predicting life expectancy.

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

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'*

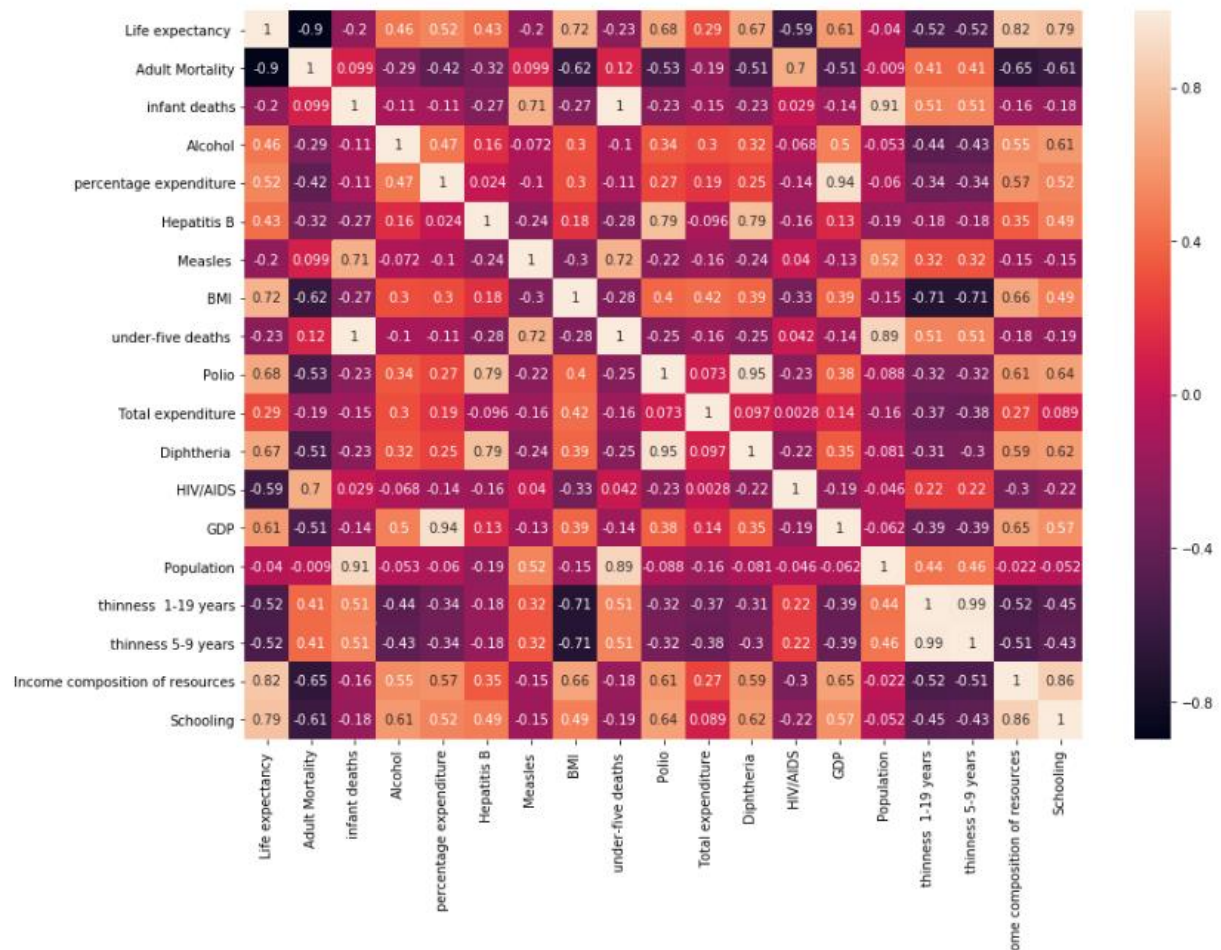
Target is Life Expectancy, measured in number of years.

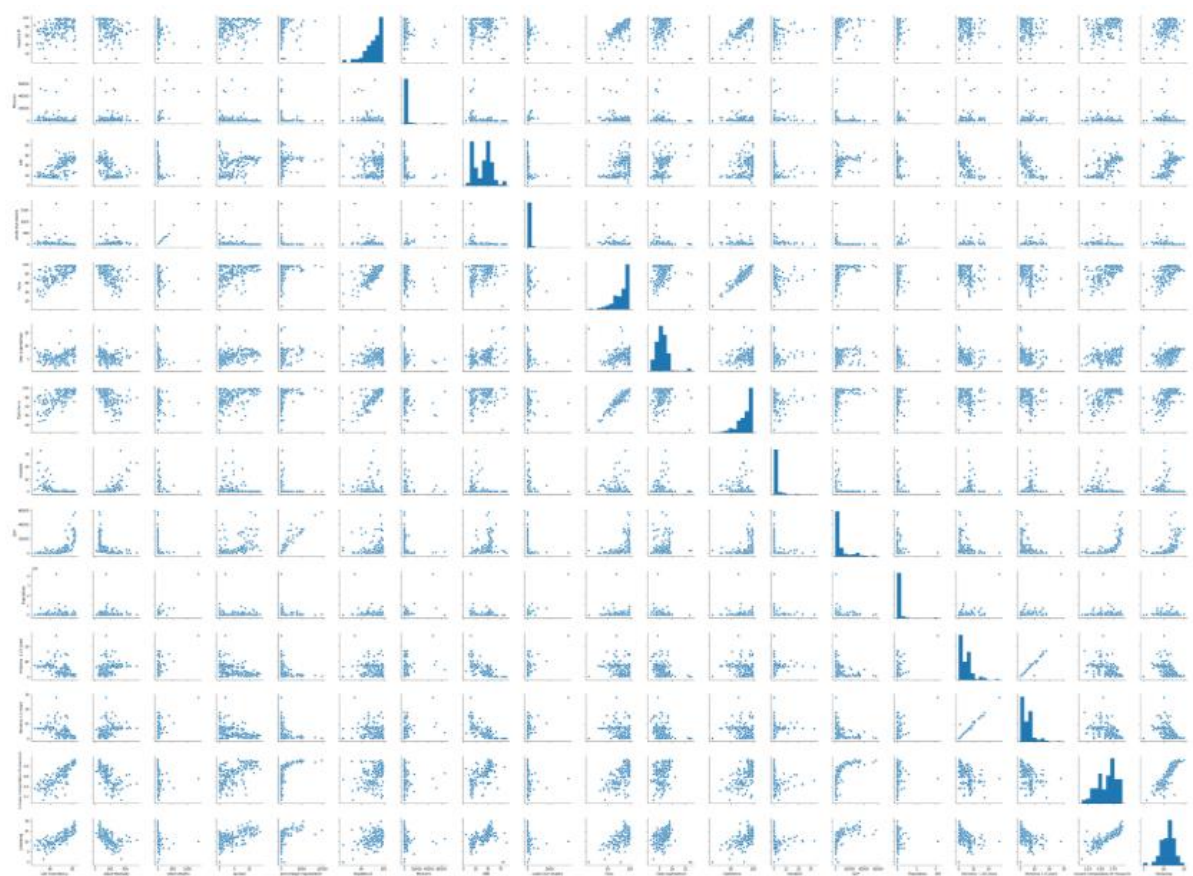
The assumptions are:

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

3.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.

Home

Default

Prediction **58.61537499999999**

Adult Mortality *
263

infant deaths *
62

Alcohol *
0.01

percentage expenditure *
71.27

Hepatitis B *
65

Measles *
1154

BMI *
19.1

under-five deaths *
83

Polio *
6

Total expenditure *
8.16

Diphtheria *
65

HIV/AIDS *
0.1

GDP *
584.2592

Population *
33736494

thinness 1-19 years *
17.2

thinness 5-9 years *
17.3

Income composition of resources *
0.479

Schooling *
10.1

SUBMIT

CANCEL

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.

APPENDIX

○ *WEBPAGE*

- <https://node-red-repcv.eu-gb.mybluemix.net/ui/#!/0?socketid=iUkpZfOA-UchtZyAAAF>
- *WHO data for life expectancy of different country.*
 - <https://www.kaggle.com/kumarajarshi/life-expectancy-who/data>