

# **CAPSTONE PROJECT PROPOSAL**

UDACITY MACHINE LEARNING NANODEGREE 2020

## **Predicting Life Expectancy rate of a Country**

Peyala Samarasimha Reddy

June, 2020

### **1. Domain Background :**

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

My personal motivation for this project is this project will definetily helps the human kind inorder reduce the deaths due to hungry and other crises on today using most trending technical solutions. Best part of this project is easy to use, user friendly comes as an end product.

### **2. Problem Statement :**

The main objective of this project will be to use Machine Learning techniques to predict the life expectancy rate of a country based on the inputs taken from the user like year, country, status, Mortality rate etc..about 22 required fields to get the almost accurate life expectancy

rate. Instead remaining the solution to the python code, the python code is integrated with node red application. So, that it will become user friendly.

### **3.1 Datasets and inputs :**

For this project we will use a dataset called Life Expectancy Data.csv which can be downloaded from kaggle <https://www.kaggle.com/kumarajarshi/life-expectancy-who> Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries. The dataset related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories: Immunization related factors, Mortality factors, Economical factors and Social factors.

- Country: 193 unique countries names
- Year
- Status : Developed or Developing status
- Life expectancy
- Adult Mortality
- infant deaths
- Alcohol
- percentage expenditure
- Hepatitis B
- Measles
- BMI
- HIV/AIDS
- thinness 1-9 years
- thinness 5-9 years
- Diptheria
- GDP
- Income composition of resources
- Polio
- Population
- Schooling

- total expenditure
- under-five deaths

## 3.2 Project Requirements:

### 2.1 Functional Requirements:

Predicting the life expectancy rate of a country

### 2.2 Technical Requirements:

Python, IBM CCloud, IBM Watson

#### Hardware Requirements

PC/Laptop/desktop with

Processor: i3 or above

Speed: 2 GHz or more

Hard disk space: 10 GB or more

RAM: 4GB or more

#### Software Requirements

Operating System: Linux/Windows

IBM Cloud

IBM Watson

Python compiler

## 4. Solution Statement

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. Project output comes by application of machine learning techniques and node red application