

# Predicting Life Expectancy Using Machine Learning

Project Manager: Anuj Kumar Pradhan

Date: 24/05/2020

Email : anujpradhan208@gmail.com

## PROJECT SCOPE DOCUMENT

### 1. **PROJECT SUMMARY:**

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting the 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 the 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. **PROJECT GOALS:**

A comparison is required to show how our model is better than the previous approach. And this follows a few steps:

- Multiple linear regression is applied on all the 20 predicting variables. The data of these 20 variables will be used in predicting the life expectancy.
- And linear regression will be applied on them to answer the mentioned questions of which factors affect the most and by how much.
- P-value analysis will be performed to check the statistical significance of columns/variables (the p-value is the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. The p-value is used as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected. A smaller p-value means that there is stronger evidence in favour of the alternative hypothesis).
- The resulting columns will be taken for final implementation of regression to obtain the results.
- Data visualization through seaborn and matplotlib will be done to show difference between the accuracy and dependency of variables.

### 3. PROJECT REQUIREMENTS:

#### a. FUNCTIONAL REQUIREMENTS:

Predicting the life expectancy rate of a country.

#### b. TECHNICAL REQUIREMENTS:

Python, IBM Cloud, IBM Watson.

#### c. HARDWARE REQUIREMENTS:

*Processor:* i3 7<sup>th</sup> gen or higher

*Speed:* 2GHz or more

*Hard disk space:* 10 GB or more

*Ram Memory:* 4 GB or more

#### d. SOFTWARE REQUIREMENTS:

*Operating system:* Windows XP or higher

*Browser :* Google Chrome

*IDE :* Spyder, Jupyter, Anaconda Navigator

### 4. PROJECT DELIVERABLES:

- Project Documentation
- Machine Learning Prediction Model
- Node red flow diagram + Watson Auto ML Mode

### 5. PROJECT TEAM:

**Project Manager:** Anuj Kumar Pradhan (Individual Work)