**Predicting Life Expectancy Using Machine Learning**

**Project Manager: Chinkit Manchanda Date: 22/05/2020**

**PROJECT SCOPE DOCUMENT**

1. **PROJECT SUMMARY:**

The Regression analysis technique is used for forecasting time series data that make predictions of a target using several features. When the prediction is dependent on the historical data points, which are used to recognize the behaviour of the data and to make predictions for future time series. This problem statement is aimed at predicting **Life Expectancy rate** of a country given various features.

**Life expectancy** is based on an estimate of the average age that members of a particular population group will be when they die. In practical however things are far more complicated because we can categorize predicting of life expectancy based on factors and the relationship between these factors.

As mentioned in the context, there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates but the factors immunization and human development index were not taken into consideration for the predictions.

1. **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 separated datasets as per there categorization.
* And linear regression will be applied on them individually 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.

1. **PROJECT REQUIREMENTS:** 
   1. **FUNCTIONAL REQUIREMENTS:**

Predicting the life expectancy rate of a country.

* 1. **TECHNICAL REQUIREMENTS:**

Python, IBM Cloud, IBM Watson.

* 1. **HARDWARE REQUIREMENTS:**

*Processor:* i3 7th gen or higher

*Speed:* 2GHz or more

*Hard disk space:* 10 GB or more

*Ram Memory:* 4 GB or more

* 1. **SOFTWARE REQUIREMENTS:***Operating system:* Windows XP or higher

*Browser:* Google Chrome, Mozilla Firefox, etc.

Spyder, Jupyter, Anaconda Navigator

1. **PROJECT DELIVERABLES:**

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

1. **PROJECT TEAM:**

***Project Manager***: Chinkit Manchanda (Individual Work)