**PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING**

INTRODUCTION:

**Overview:**

* Life expectancy refers to the number of years a person is expected to live based on the statistical average. Life expectancy varies by geographical area and by era.
* The life expectancy for a particular person or population group depends on several variables such as their lifestyle, access to healthcare, diet, economical status and the relevant mortality and morbidity data.
* However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

**Purpose:**

* 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.

LITERATURE SURVEY:

**Existing Problem:**

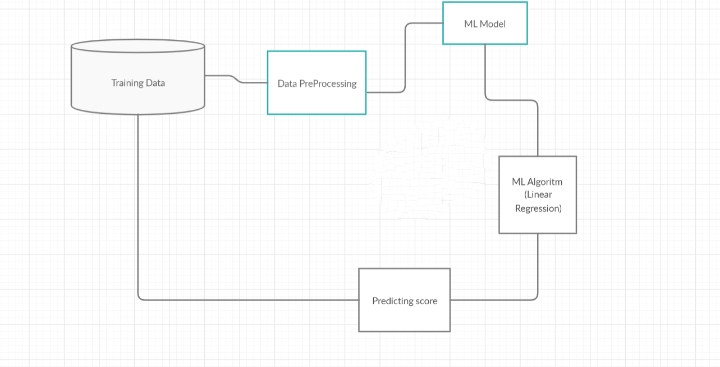
* Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors.
* 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.

**Proposed Solution:**

* Design a Regression model to predict life expectancy ratio of a given country based on some features provided such as year, GDP(gross domestic product), education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country.

THEORITICAL ANALYSIS:

**Block Diagram:**

****

**PROJECT REQUIREMENTS:-**

* A Supervised Machine learning Regression algorithm with maximum accuracy to be trained and tested on the dataset.
* The Dataset consists of 21 columns excluding the predicting column i.e. Life expectancy.

**SOFTWARE REQUIREMENTS:-**

* IBM Cloud
* IBM Watson Studio
* Node-red App

**Experimental Investigation:**

1. Choose a Project Idea:

Predicting Life Expectancy of a person.

1. Conduct Background Research:

<https://www.kaggle.com/kumarajarshi/life-expectancy-who>

[https://ourworldindata.org/life-expectancy-how-is-it- calculated-and-how-should-it-be-interpreted](https://ourworldindata.org/life-expectancy-how-is-it-%20%20%20%20%20calculated-and-how-should-it-be-interpreted)

1. Compose a Hypothesis:

Based on our study and information gathered we can predict the average age of a person.

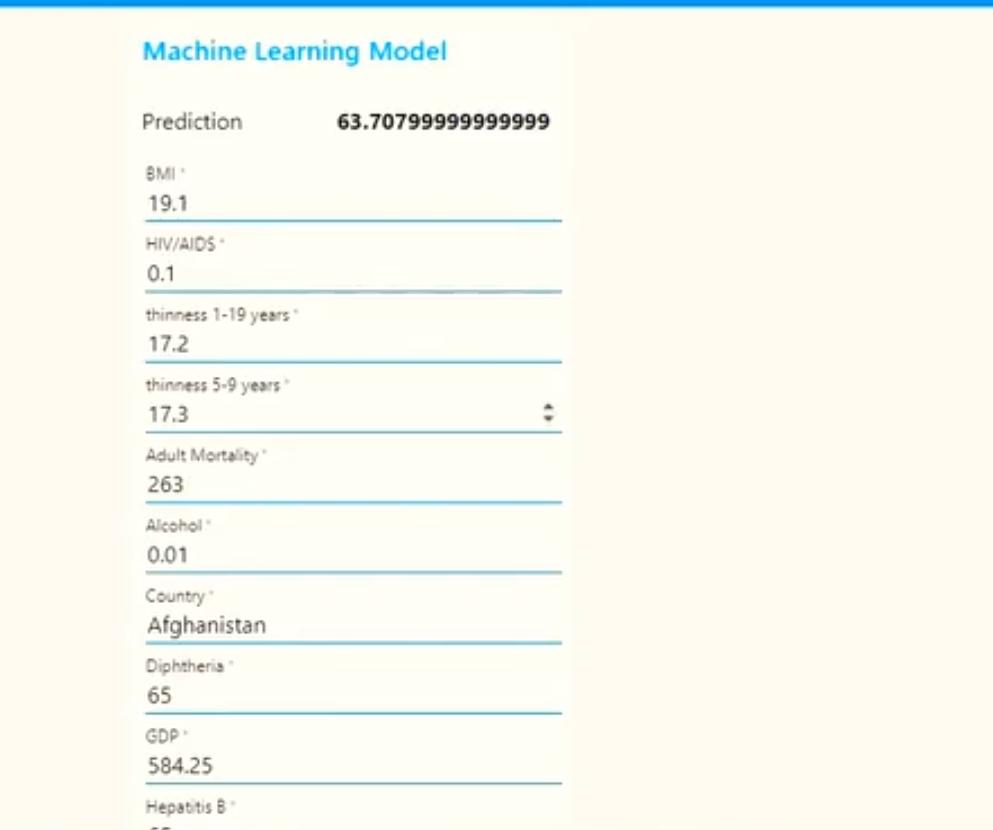
1. Design your Experiment:

First we need to collect the suitable data for our problem statement. Next we need to construct the model for this problem we use regression model.

1. Draw Conclusions:

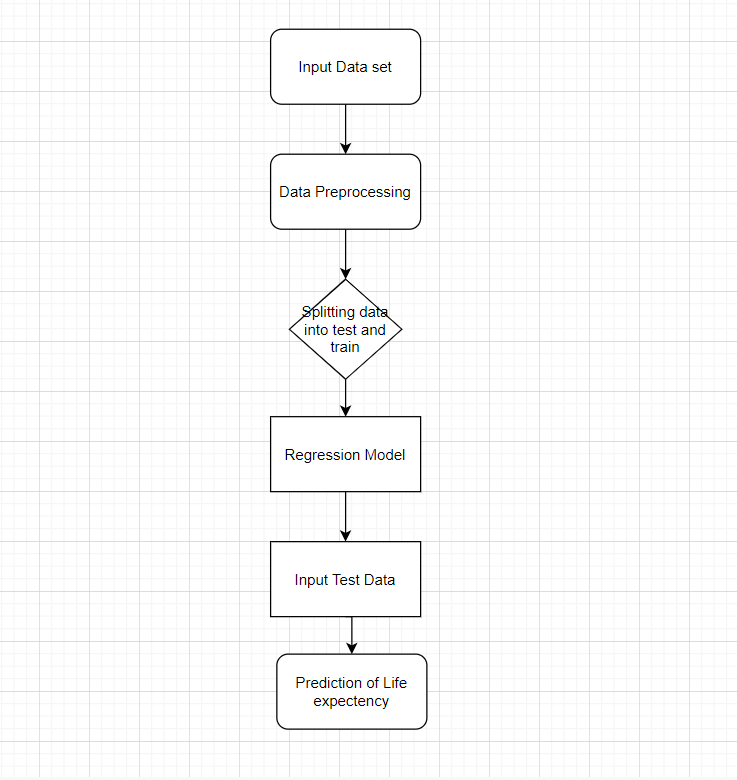
After construction of our model, we can predict the average life expectancy of a person.

**Result:**

****

By training the model using regression, we can predict the average life of a person based on various factors that effect his health. So, hence based on the various factors in the data set we can predict the average life span of a person.

**FLOWCHART:**

****

**CONCLUSION:**

The end product is an webpage created and deployed on node-red app of IBM cloud. The backend of webpage is an Extra Tree Regressor Model with 97.07% R2 score created and deployed on watson studio using machine learning service.

The web-page has input fields similar to dataset columns such as Country, BMI, percentage expenditure, Alcohol etc and an output field named as prediction i.e similar to dataset column Life expectancy which gives the life expectancy prediction based on the inputted values.

**SOURCE CODE:**

****