PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

PROBLEM DESCRIPTION:

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 SCOPE, SCHEDULE, TEAM & DELIVERABLES:

- **PROJECT SCOPE:**

  Good prognostication helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. So this problem statement is aimed at predicting Life Expectancy rate of a country given various features. It predicts the average lifetime of a human being and predicts on the basis of various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. So the end product will predict the future life expectancy of the person with the help of prior given appropriate matrix of features by the user like current 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 SUMMARY :**
  - This project is to build a model while considering historical data from a period of 2000 to 2015 for all the countries.
  - The model trained in this project will be able to predict the average lifetime of a human being given some input factors.
  - With the help of this project any country can able to predict the expected lifetime of their countrymen and then accordingly take preventive measures to improve their healthcare.
  - This will also help countries in improving a particular a particular field such as GDP, alcohol intake, etc. which have a high impact on country's life expectancy.

- **PROJECT REQUIREMENTS :**

  Predicting life expectancy with python:
  - Download the dataset of WHO.
  - Analyze it and clean the dataset.
  - Create IBM account.

- Create the appropriate cloud and node red services.
- Train the regression model on different algorithms.
- Check for the best one and finalize that algorithm to train our model.
- Create scoring end point for integrating our model to node red.
- Build Node red flow for GUI (web app).

Predicting life expectancy without python:

- Create autoAI service.
- Deploy model.
- Integrate it using node red.

➢ FUNCTIONAL REQUIREMENTS :

- Provide the model with the inputs fields.
- The model will return the output as the average predicted lifespan.

➢ TECHNICAL REQUIREMENTS:

- The GUI must be integrated with the backend trained model.
- The model before training must be given with clean dataset (done by preprocessing).

➢ SOFTWARE REQUIREMENTS:

- Python IDE
- Excel
- IBM Cloud Account
- IBM Watson

➢ REGARDING THE DATASET:

This dataset is available in the Global Health Observatory (GHO) data repository under World Health Organization (WHO) which keeps track of the health status as well as many other related factors for all countries. The data-sets are made available to public for the purpose of health data analysis. In this dataset, related to life expectancy, health factors for 193 countries have been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The dataset consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables are divided into several broad categories like Immunization related factors, Mortality factors, Economical factors and Social factors.

***Source*:** https://www.kaggle.com/kumarajarshi/life-expectancy-who/

➢ DELIVERABLES:

- The dataset aims to answer the following key questions:
- Does various predicting factors which has been chosen initially really affect the Life expectancy ?
- What are the predicting variables actually affecting the life expectancy?

- Should a country having a lower life expectancy value increase its healthcare expenditure in order to improve its average lifespan?
- How does Infant and Adult mortality rates affect life expectancy?
- Does Life Expectancy has positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.
- What is the impact of schooling on the lifespan of humans?
- Does Life Expectancy have positive or negative relationship with drinking alcohol?
- Do densely populated countries tend to have lower life expectancy?
- What is the impact of Immunization coverage on life Expectancy?
- Finally, a user interface will be implemented where you can get the predicted value of life.

➢ **PROJECT TEAM:** Individual

➢ **PROJECT DURATION:** 1 month