**Project-1**

**Heart Attack Analysis & Prediction Dataset**

#### **Objective: To predict the chance of Heart attack**

## **About Data**

age - Age of the patient

sex - Sex of the patient

cp - Chest pain type ~ 0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic

trtbps - Resting blood pressure (in mm Hg)

chol - Cholestoral in mg/dl fetched via BMI sensor

fbs - (fasting blood sugar > 120 mg/dl) ~ 1 = True, 0 = False

restecg - Resting electrocardiographic results ~ 0 = Normal, 1 = ST-T wave normality, 2 = Left ventricular hypertrophy

thalachh - Maximum heart rate achieved

oldpeak - Previous peak

slp - Slope

caa - Number of major vessels

thall - Thalium Stress Test result ~ (0,3)

exng - Exercise induced angina ~ 1 = Yes, 0 = No

output - Target variable

# Data Preprocessing

Before providing the dataset to any model, it is essential to check outliers and transform it so that its distribution will have a mean of 0 and a standard deviation of 1.

### **Normalisation**

### Since most of the machine learning algorithms use Euclidean distance between two data points in their computations, this is a problem. To suppress this effect, we need to bring all features to the same level of magnitudes. This can be achieved by a method called feature scaling.

**Modelling**

In this notebook 5 different machine learning algorithms will be evaluated on the dataset for prediction analysis:

1. Logistic Regression (Logistic)
2. Naive Bayes (NaiveBayes)
3. Classification and Regression Trees or CART (REPTree)
4. k-Nearest Neighbors or KNN (IBk)
5. Support Vector Machines or SVM (SMO)
6. Random Forest and Desion Trees
7. XGBoost

Each algorithm will be evaluated using classification accuracy, to measure the performance of each model. First step in the data modelling is to label the dataset with X (matrix of independent variables) and y (vector of the dependent variable). Then create an instance of the model to train and fit the model, then calculate predictions of test set in order to get the classification report.

**Project-2**

**World Happiness Report Analysis**

### **About**

The World Happiness Report is a landmark survey of the state of global happiness . The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

### **Content**

The happiness scores and rankings use data from the Gallup World Poll . The columns following the happiness score estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world’s lowest national averages for each of the six factors. They have no impact on the total score reported for each country, but they do explain why some countries rank higher than others.

# Data Preprocessing

* There is no null value in our data (each column has 149 non-null count).
* There are 18 float value columns.
* Only two columns are having object values which are 'Country name' and 'Regional indicator'.

## **Data Visualization**

## **Dataset information :**

**1. Country**

Name of the country

**2. Happiness Rank**

Rank of the country based on the Happiness Score.

**3. Happiness Score**

A metric measured in 2016 by asking the sampled people the question: "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest"

**4. Whisker High**

Lower Confidence Interval of the Happiness Score

**5. Whisker Low**

Upper Confidence Interval of the Happiness Score

**6. Economy (GDP per Capita)**

The extent to which GDP contributes to the calculation of the Happiness Score.

**7. Family**

The extent to which Family contributes to the calculation of the Happiness Score

**8. Health (Life Expectancy)**

The extent to which Life expectancy contributed to the calculation of the Happiness Score

**9. Freedom**

The extent to which Freedom contributed to the calculation of the Happiness Score

**10. Trust (Government Corruption)**

The extent to which Perception of Corruption contributes to Happiness Score

**11. Generosity**

The extent to which Generosity contributed to the calculation of the Happiness Score

**12. Dystopia Residual**

The extent to which Dystopia Residual contributed to the calculation of the Happiness Score.