# **Prediction of Heart Disease**

**Project Description**

The dataset is publically available.

In this dataset we are given the **patient’s information.**

It contains total **4240 records** and **15 features**.

Features includes Gender, Age , Education ,Cigs\_Per\_day, Current\_Smoker, Sys\_BP, Dia\_BP, Heart\_rate, Glucose…etc.

People have various problems regarding heath.

The goal of the project is to predict whether that person will have coronary heart disease in upcoming 10 years.

The Target variable is **TenYearCHD.**

We have to clean the data and apply all the algorithms to our model and check which one gives maximum accuracy.

Perform the following using “**framingham.csv**”

1. Read the csv file and find df.shape, df.describe().
2. Preprocess the data:-;

1] Check the null values from the dataset

2] Fill the null values using median or mode.

3] Remove the insignificant columns from the dataset if necessary.

4] Convert the categorical columns to numerical columns.

5] Find the statistical information of the dataset.

1. Perform EDA on the dataset using visualization such as bar plot, Box plot , count plot etc.
2. Apply standard scaling to standardize the data if necessary.
3. Split the data into training and testing.
4. Apply Logistic regression, Decision tree , Random forest, KNN & SVM to the same dataset and find :-

1] Accuracy of each model

2] Confusion matrix

3] Classification report

4] Run the model by giving the same inputs to all the models and check the accuracy and output of each model.

1. Observe the accuracy of each model and conclude which model gives the maximum accuracy.
2. So finally the model which gives max accuracy will be best model.

**Execute your project. Remove errors in the program if any.**

**Prepare the Project Report and Submit it to the Google Classroom.**