**Heart Disease prediction using Machine Learning.**

Introduction

Heart diseases is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases have increased significantly over the past few decades in India, in fact it has become the leading cause of death in India.

A study shows that from 1990 to 2016 the death rate due to heart diseases have increased around 34 per cent from 155.7 to 209.1 deaths per one lakh population in India.

Thus preventing Heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives. This is where Machine Learning comes into play. Machine Learning helps in predicting the Heart diseases, and the predictions made are quite accurate.

Problem Description:

A dataset is formed by taking into consideration some of the information of the individuals. The problem is: based on the given information about each individual we have to calculate that whether that individual will suffer from heart disease.

# Dataset:

The dataset consists of 920 individual’s data. There are 15 columns in the dataset, however the first column name is not a good parameter as far as machine learning is considered so, there are effectively 14 columns.

1. Age : displays the age of the individual.
2. Sex : displays the gender of the individual using the following format : 1 = male 0 = female.
3. Chest-pain type : displays the type of chest-pain experienced by the individual using the following format : 1 = typical angina 2 = atypical angina 3 = non - anginal pain 4 = asymptotic
4. Resting Blood Pressure : displays the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholestrol : displays the serum cholestrol in mg/dl (unit)
6. Fasting Blood Sugar : compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)
7. Resting ECG : 0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hyperthrophy
8. Max heart rate achieved : displays the max heart rate achieved by an individual.
9. Exercise induced angina : 1 = yes 0 = no
10. ST depression induced by exercise relative to rest : displays the value which is integer or float.
11. Peak exercise ST segment : 1 = upsloping 2 = flat 3 = downsloping
12. Number of major vessels (0-3) colored by flourosopy : displays the value as integer or float.
13. Thal : displays the thalassemia : 3 = normal 6 = fixed defect 7 = reversable defect
14. Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not : 0 = absence 1,2,3,4 = present.

# Model Training and Prediction:

We can train our prediction model by analysing existing data because we already know whether each patient has heart disease. This process is also known as supervision and learning. The trained model is then used to predict if users suffer from heart disease. The training and prediction process is described as follows:

# Splitting:

First, data is divided into two parts using component splitting. In this experiment, data is split based on a ratio of 67:33 for the training set and the prediction set. The training set data is used in some machine learning algorithms: Artificial Neural Networks, Gaussian Naïve Bayes, KNN, Random Forest, Support Vector Classifier component for model training, while the prediction set data is used in the prediction component.

# Performance Metrics

Accuracy, Precision, Recall, F1-Score, ROC-AUC, Log loss used to compare the algorithms.