**Heart Disease Prediction using Logistic Regression**

**Dataset**: [Dataset Link](https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset)

**Dataset length:** Rows: It includes 4241 records and 15 attributes.

**Description:** The classification goal is to predict whether the patient has 10-year risk of future coronary heart disease (CHD). The dataset provides the patients’ information. Each attribute is a potential risk factor. There are both demographic, behavioral and medical risk factors.

* **Demographic:**sex: male or female;(Nominal)
* age: age of the patient;(Continuous - Although the recorded ages have been truncated to whole numbers, the concept of age is continuous)
* **Behavioral**
* currentSmoker: whether or not the patient is a current smoker (Nominal)
* cigsPerDay: the number of cigarettes that the person smoked on average in one day. (can be considered continuous as one can have any number of cigarettes, even half a cigarette.)
* **Medical(history):**
* BPMeds: whether or not the patient was on blood pressure medication (Nominal)
* prevalentStroke: whether or not the patient had previously had a stroke (Nominal)
* prevalentHyp: whether or not the patient was hypertensive (Nominal)
* diabetes: whether or not the patient had diabetes (Nominal)
* **Medical(current):**
* totChol: total cholesterol level (Continuous)
* sysBP: systolic blood pressure (Continuous)
* diaBP: diastolic blood pressure (Continuous)
* BMI: Body Mass Index (Continuous)
* heartRate: heart rate (Continuous - In medical research, variables such as heart rate though in fact discrete, yet are considered continuous because of large number of possible values.)
* glucose: glucose level (Continuous)
* **Predict variable (desired target):**
* 10-year risk of coronary heart disease CHD (binary: “1”, means “Yes”, “0” means “No”)