***Heart Disease Data Description:***

The dataset is from an on-going cardiovascular study on residents of the town of Framingham, Massachusetts. The classification goal is to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). The dataset provides the patients’ information on his health and lifestyle habits. It includes over 4,000 records and 15 attributes. Each variable attribute is a potential risk factor. There are demographic, behavioural and medical risk factors enlisted within.

***Variable description:***

***Demographic variables:***

 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)

***Behavioural variables:***

 Current Smoker: whether or not the patient is a current smoker (Nominal)

 Cigs Per Day: 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*** ***variables:***

 BP Meds: whether or not the patient was on blood pressure medication (Nominal)

 Prevalent Stroke: whether or not the patient had previously had a stroke (Nominal)

 Prevalent Hyp: whether or not the patient was hypertensive (Nominal)

 Diabetes: whether or not the patient had diabetes (Nominal)

***Current medical variables:***

 Tot Chol: total cholesterol level (Continuous)

 Sys BP: systolic blood pressure (Continuous)

 Dia BP: diastolic blood pressure (Continuous)

 BMI: Body Mass Index (Continuous)

 Heart Rate: 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)

***Target variable (TenYearCHD):***

* 10 year risk of coronary heart disease CHD (binary: “1”, means “Yes”, “0” means “No”)