Problem Statement

Classification of a specific heart disease using machine learning techniques.

Problem Description:

In this bootcamp, you are expected to build a machine learning model(s), that can detect between a subject afflicted with heart disease and someone who is normal. Problems such as this are common in the healthcare field where such medical diagnoses can be made with the aid of machine learning and AI techniques, usually with much better accuracy. Hospitals and medical enterprises often employ specialists such as machine learning engineers and data scientists to carry out these tasks.

Attribute Information:

Using the 13 attributes which are already extracted, in the heart disease dataset, you are expected to detect either the presence of or the absence of the heart disease in human subjects.

There are 13 attributes:

1. age: age in years
2. sex: sex (1 = male; 0 = female)
3. cp: chest pain type
   -- Value 0: typical angina
   -- Value 1: atypical angina
   -- Value 2: non-anginal pain
   -- Value 3: asymptomatic
4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5. chol: serum cholesterol in mg/dl
6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg: resting electrocardiographic results
   -- Value 0: normal
   -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
   -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina (1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment
    -- Value 0: upsloping
    -- Value 1: flat
    -- Value 2: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy

13. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect
    and the label
14. condition: 0 = no disease, 1 = disease

Objectives:

In this bootcamp, you are expected to perform the following:

1. Perform exploratory data analysis
2. Compare performance with at least 2 classification models
3. Compute appropriate performance metrics.