Cardiovascular Disease Prediction

1. <u>Problem Statement</u>:

   To predict the risk of cardiovascular disease in men and women depending on the various attributes obtained from the blood samples of the corresponding patients.

   Cardiovascular diseases have been a major concern because of the vastly changing lifestyle and the treatment of the same comes at a high expense especially in countries like United States. Thus, we want to put in efforts to analyse the data and based on various factors, determine the risk of the cardiovascular disease in a person so as to take let them be aware of the same.

2. <u>Description of the data set</u>:

   The data set that we are planning to use is from UCI Machine Learning Repository that contains 14 physical attributes based on the physical testing of a patient. In addition to having blood drawn, the patient also undergoes a quick exercise test. The links for the datasets we plan to use are from the provided link are as follows: Cleveland, Hungarian, Switzerland, VA.

   The 14 attributes are:

   ❖ age: age in years
   ❖ sex: sex (1 = male; 0 = female)
   ❖ cp: chest pain type
      → Value 1: typical angina
      → Value 2: atypical angina
      → Value 3: non-anginal pain
      → Value 4: asymptomatic
   ❖ trestbps: resting blood pressure (in mm Hg on admission to the hospital)
   ❖ chol: serum cholesterol in mg/dl
   ❖ fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
   ❖ 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
   ❖ thalach: maximum heart rate achieved
   ❖ exang: exercise induced angina (1 = yes; 0 = no)
   ❖ oldpeak = ST depression induced by exercise relative to rest
   ❖ slope: the slope of the peak exercise ST segment
      → Value 1: upsloping
      → Value 2: flat
      → Value 3: downsloping
   ❖ ca: number of major vessels (0-3) colored by flourosopy
   ❖ thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
   ❖ num: diagnosis of heart disease (angiographic disease status)
      → Value 0: < 50% diameter narrowing
      → Value 1: > 50% diameter narrowing
      (in any major vessel: attributes 59 through 68 are vessels)

The website has data obtained from Hungarian Institute of Cardiology, University Hospital (Zurich), University Hospital (Basel) and V.A. Medical Centre, Long Beach and Cleveland Clinic Foundation.

3. <u>Implementation Plan</u>:

Understanding and analysing the aspects of the data set	: Oct 20 - Oct 24
Exploratory Data Analysis (EDA)	: Oct 25 - Oct 31
Data Processing	: Nov 1 - Nov 9
ML model: Algorithm 1	: Nov 10 - Nov 17
ML model: Algorithm 2	: Nov 18 - Nov 25
ML model: Algorithm 3 and 4	: Nov 26 - Dec 2
Model Performance and its Evaluation	: Dec 3 - Dec 10
Checking the accuracy of the model and conclusions made	: Dec 11 - Dec 15
Final Project Report and Presentation work	: Dec 16 - Dec 21

4. <u>Team members and task allocation</u>:

1. Jaykumar Dineshkumar Patel (CWID: 2000 8512)

2. Juilee Sameer Thakur (CWID: 2000 8598)

3. Samyak Pradeep Upare (CWID: 2001 6643)

All the tasks mentioned above will be equally split and then collectively presented in the final report.