Heart Disease Prediction

Milestone: Data collection and Processing

Group 3

Mugdha Sanjay Parbat - 002142372

Pranav Chandrakant Pulkundwar - 002121679

Telephone

+1 (617) 901-8417

+1 (617) 901-8418

Email ID

parbat.m@northeastern.edu

pulkundwar.p@northeastern.edu

Percentage of contribution by Student 1: 50%

Percentage of contribution by Student 2: 50%

Submission Date: 28th February 2022

Data Collection and Processing

Understanding what the variables are:-

The dataset that we have selected from UCI repository consists of 303 individuals data. There are 14 columns in the dataset as follows:-

1. Age: This tells us an individual's age
2. Sex: This tells us an individual's gender :
   1 = male
   0 = female
3. Chest-pain type: This describes 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: This tells us the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholesterol: This tells us the serum cholesterol in mg/dl (unit)
6. Fasting Blood Sugar: This compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar > 120mg/dl then : 1 (T) else : 0 (F)
7. Resting ECG : This tells us the resting electrocardiographic results
   0 = normal
   1 = having ST-T wave abnormality
   2 = left ventricular hypertrophy
8. Max heart rate achieved : This tells us the maximum heart rate achieved by an individual.
9. Exercise induced angina :- This tells us if there is an exercise induced angina or not.
   1 = yes
   0 = no
10. ST depression induced by exercise relative to rest: This shows a value which is an integer or float.
11. Peak exercise ST segment :
    1 = upsloping
    2 = flat
    3 = down sloping
12. Number of major vessels (0–3) coloured by fluoroscopy : This shows the value as integer or float.
13. Thala : This describes the type of thalassemia :-
    3 = normal
    6 = fixed defect
    7 = reversible defect
14. Diagnosis of heart disease : This tells us whether the individual is suffering from heart disease or not :-
    0 = absence
    1, 2, 3, 4 = present.

Reason behind choosing these parameters:-

In the dataset, there are 76 parameters but we decided to choose only 14 variables among them and here's why:-

1. **Age** : One of the crucial factor in being targeted by heart diseases is age. According to research, it roughly triples the risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. According to estimations, people with age 65 and older are the ones who die of coronary heart disease. And at the same time, the risk of stroke increases two fold every decade after age 55.
2. **Sex**: Men are at greater risk of heart disease than pre-menopausal women. Once past menopause, it has been a matter of argument that a woman's risk is similar to a man's although recent data from the WHO

and UN disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.

3. **Angina (Chest Pain)**: Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood. It might feel like pressure or squeezing in your chest. The discomfort is not limited to chest but also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.
4. **Resting Blood Pressure**: Over a period of time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.
5. **Serum Cholesterol**: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of a heart attack.
6. **Fasting Blood Sugar**: High blood sugar is one of the main reasons contributing to the risk of heart attack. Low levels of insulins secreted by the pancreas or not maintaining the safe level of insulin can lead to increased sugar level in body which in response increases the risk of heart attack. The blood sugar levels are measured on an empty stomach since food can cause the blood sugar to rise significantly which gives false results.
7. **Resting ECG**: According to reports of U.S. Preventive Services Task Force, the probable harms of screening resting, or exercise ECG may be high for people with low risk of cardiovascular disease. The current data is insufficient to analyze the benefits and harms of the screening ECG for people with intermediate or high risk of cardiovascular diseases.
8. **Max heart rate achieved**: The risk of heat attack because of acceleration of heart rate is comparable to the risk due to high blood pressure. According to recent studies it is shown that the increase of 10 beats per minute of heart rate increases the risk of cardiac arrest by at least 20%. This is similar to the increased risk due to the increase in systolic blood pressure by 10 mm Hg.
9. **Peak exercise ST segment**: When the ST-segment depression is less than 1 mm at 60–80 ms following the J point on a treadmill ECG stress test, it is considered abnormal. Exercise ECGs with up-sloping ST-segment depressions are usually classified as "equivocal." In general, horizontal or down-sloping ST-segment depression at a lesser workload (measured in METs) or heart rate predicts a poor prognosis and a greater risk of multi-vessel disease. The duration of ST-segment depression is particularly crucial, as a positive treadmill ECG stress test is associated with a protracted recovery after peak stress. The presence of ST-segment elevation > 1 mm (typically indicating transmural ischemia) is another indicator of severe CAD, and these patients are frequently referred for coronary angiography right away.
10. **Exercise induced angina**: The tight, gripping or squeezing pain because of exercise or physical work is associated with the angina which may vary from mild to severe pain. This pain is generally felt in the chest center and slowly spreads to the shoulders, or the back, jaw, neck, arms. Types of Anginas are generally divided as:-
    a. Stable Angina / Angina Pectoris
    b. Unstable Angina
    c. Variant (Prinzmetal) Angina
    d. Microvascular Angina.