Problem Statement

Cardiovascular diseases are one of the leading causes of deaths globally. To identify the causes and develop a system to predict potential heart attacks in a effective manner is necessary. The data presented has all the information about relevant factors that might have an impact on cardiovascular health. The data to be studied in detail for further analysis.

There is one dataset data that has 14 attributes with more than 4000 data points.

You are required to determine and examine the factors that play a significant role in increasing the rate of heart attacks. Also, use the findings to create and predict a model.A table with numbers and letters

Description automatically generated

A screenshot of a data table

Description automatically generated

A screenshot of a medical report

Description automatically generatedA screenshot of a data

Description automatically generated

**Project Task: Week 1**

**Importing, Understanding, and Inspecting Data:**

1. Perform preliminary data inspection and report the findings as the structure of the data, missing values, duplicates, etc.

2. Based on the findings from the previous question, remove duplicates (if any) and treat missing values using an appropriate strategy.

3. Get a preliminary statistical summary of the data. Explore the measures of central tendencies and the spread of the data overall.

**Project Task: Week 2:**

**Performing EDA:**

1. Identify the data variable which might be categorical in nature. Describe and explore these variables using appropriate tools. For example: count plot.
2. Study the occurrence of CVD across different ages.
3. Can we detect heart attack based on anomalies in **resting blood pressure** of the patient?
4. Study the composition of overall patient w.r.t **gender.**

**Project task week 3:**

**Performing EDA and modelling:**

1. Describe the relationship between cholesterol levels and our target variable.
2. What can be concluded about the relationship between peak exercising and occurrence of CVD?
3. Is thalassemia a major cause of CVD? How are the other factors determing the occurrence of CVD?
4. Use a pair plot to understand the relationship between all the given variables.
5. Perform logistic regression, predict the outcome for test data, and validate the results by using the confusion matrix

**Project Task Week 4:**

**Dashboarding:**

1. Visualize the variables using tableau to create an understanding for attributes of diseased vs a healthy person.

2. Demonstrate the variables associated with each other and factors to build a dashboard