

# Machine Learning

## **Problem statement:**

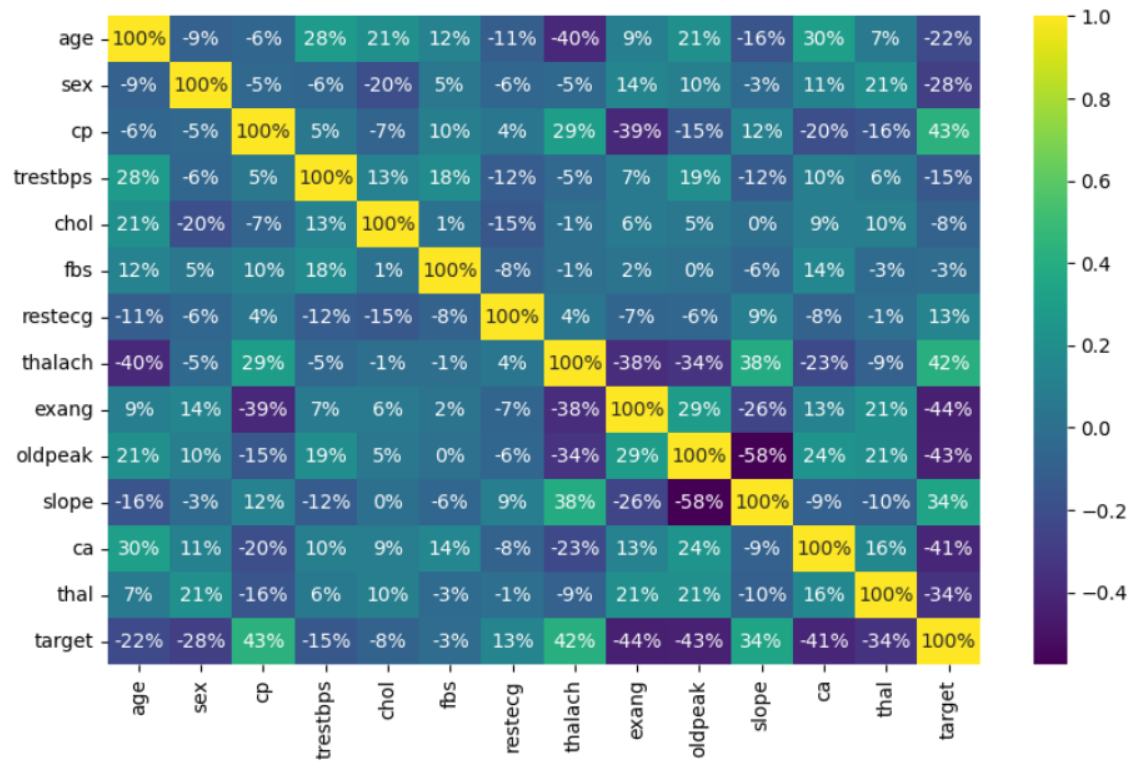
Cardiovascular diseases are the leading cause of death globally. It is therefore necessary to identify the causes and develop a system to predict heart attacks in an effective manner. The dataset given has information about the factors that might have an impact on cardiovascular health.

## **To Identify the cause and predict heart attacks:**

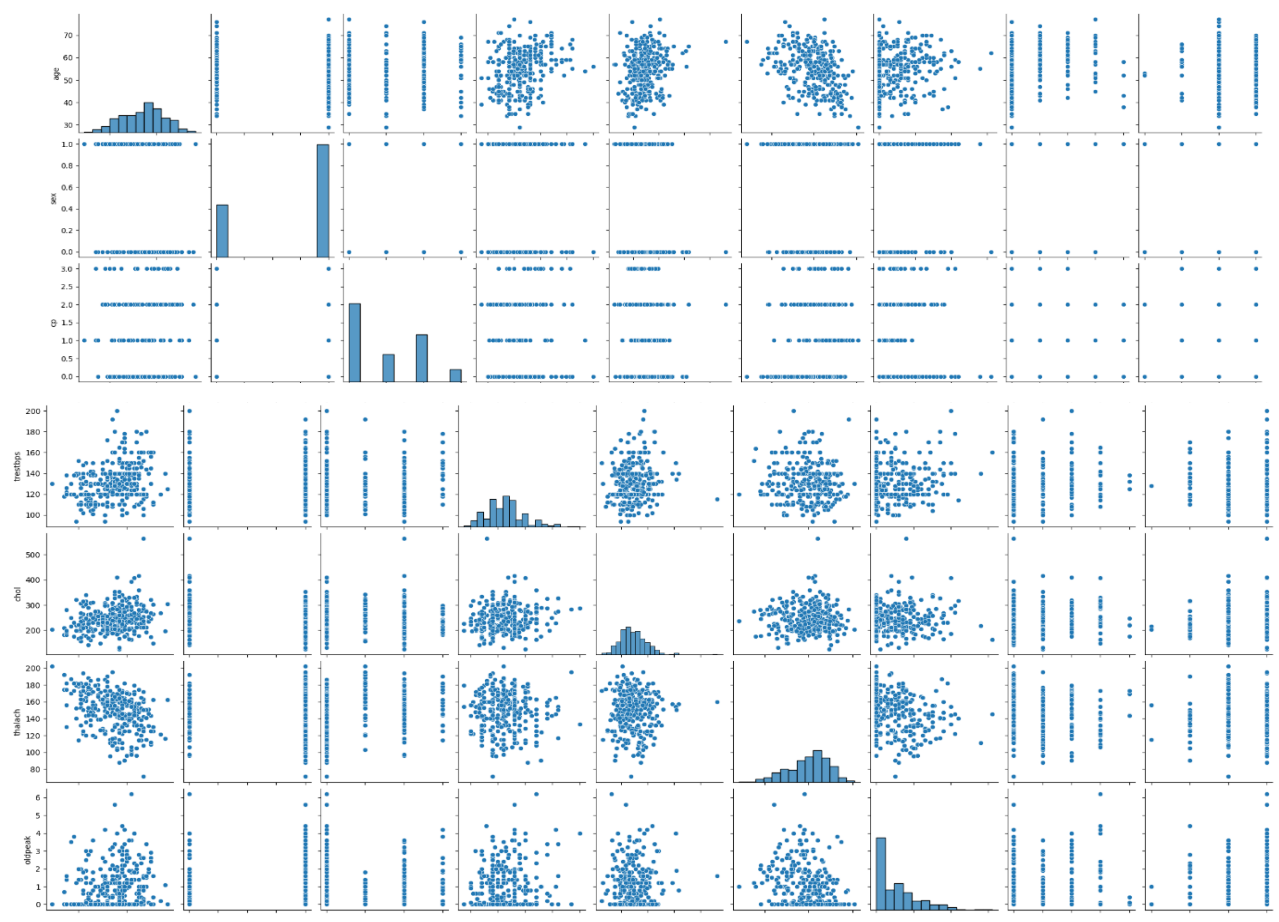
## **Preliminary analysis:**

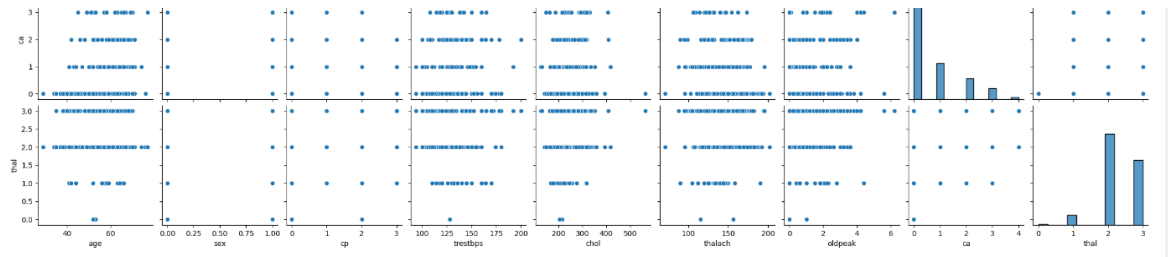
The dataset contains 303 rows and 14 columns, there is no missing value in the dataset, I drop the duplicates if it is there and check for the statistical summary of the data. There are 206 males and 96 females in the sex column, and check the target count with 164 people with heart disease and 138 people without heart disease. Hence, there are more people suffering from heart problems.

Then I perform the visual interpretation of the data, where I check the correlation of the data with the help of the heat map and find the slope and old peak, target and exang, target and old peak to have a negative correlation, chest pain and target, and target and thalach to have a positive correlation. And from the count plot analysis, I observe that low resting blood pressure (trestbps), people with 210-280 mg/dl cholesterol, thal-2 results, and resting electrocardiographic '1', slope-2, results are associated with more heart diseases.



And from the pair plot I can analyze the relations between variables and found Age and cp, age and sex, trestbps and sex have negative relation and ca and age, trestbps and age, age and trestbps have a positive relation.





Now we split our data into an independent variable and a dependent variable (X and Y), apply a logistic regression algorithm to our model, and find an accuracy of 86% in the training data and a 78% accuracy in the testing data.

Applying an random forest algorithm to the model, find an accuracy of 78%.

The accuracy of the random forest classifier and logistic regression is fairly good and the cross-validation scores obtain is quite good.