## **Heart Disease Prediction**

## **Goal:**

Aim of the project is to determine if the patient has the heart disease or not based on the culture blood test report .

## **Data**

This database contains 14 physical attributes based on physical testing of a patient. Blood samples are taken and the patient also conducts a brief exercise test. The "goal" field refers to the presence of heart disease in the patient. It is integer (0 for no presence, 1 for presence). In general, to confirm 100% if a patient has heart disease can be quite an invasive process, so if we can create a model that accurately predicts the likelihood of heart disease, we can help avoid expensive and invasive procedures.

Content

Attribute Information:

* age
* sex
* chest pain type (4 values)
* resting blood pressure
* serum cholestoral in mg/dl
* fasting blood sugar > 120 mg/dl
* resting electrocardiographic results (values 0,1,2)
* maximum heart rate achieved
* exercise induced angina
* oldpeak = ST depression induced by exercise relative to rest
* the slope of the peak exercise ST segment
* number of major vessels (0-3) colored by flourosopy
* thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
* target:0 for no presence of heart disease, 1 for presence of heart disease

Original Source: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

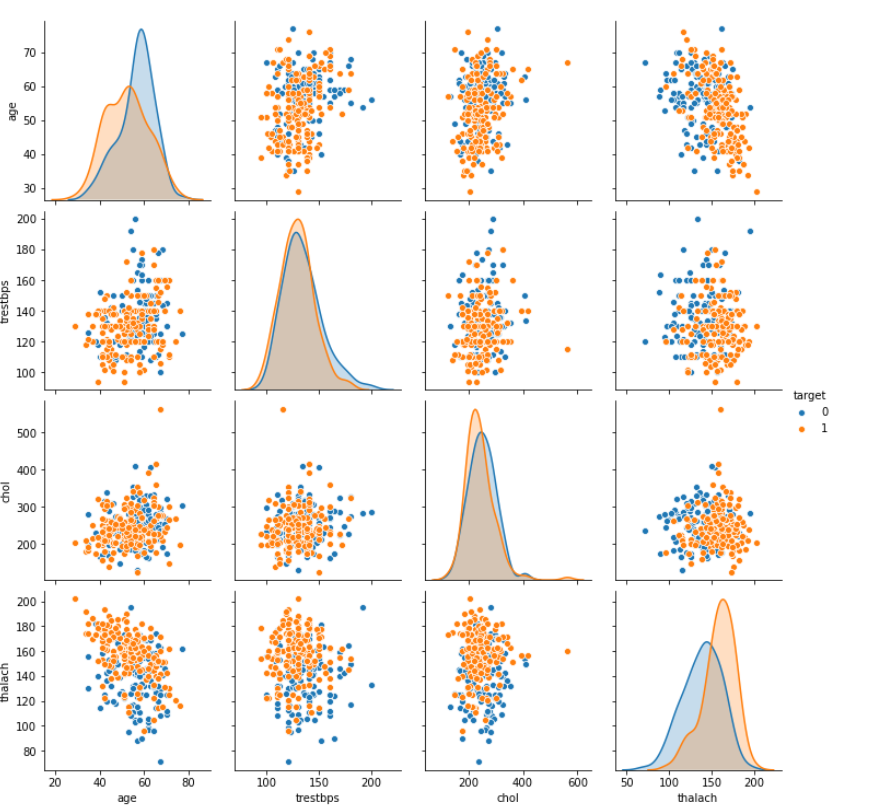
Creators:

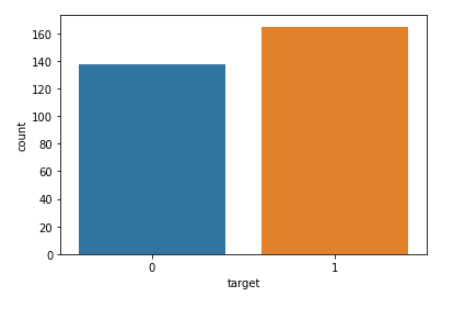
Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

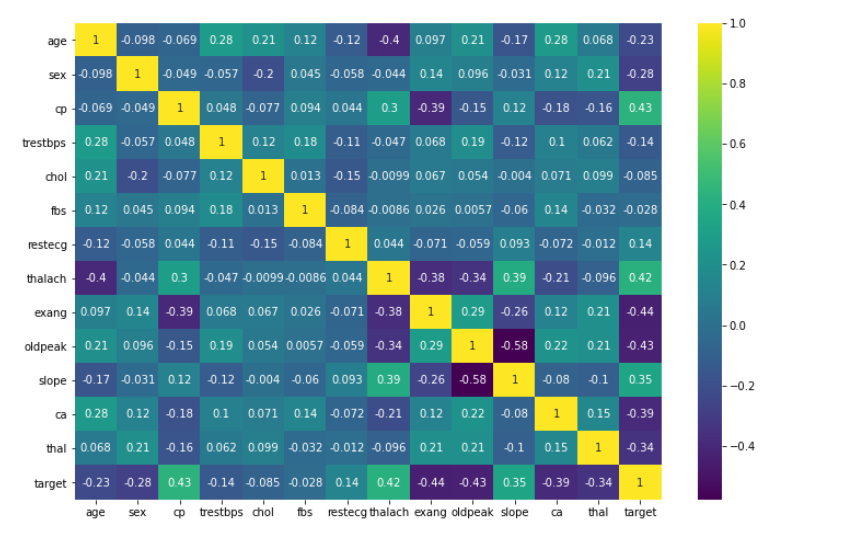
#1 Perfor Exploratory analysis

#2 Check Missing data ,Duplicates

#3 Data Visualizations





 ,

#4 **Train | Test Split and Scaling**

**#5 Perform a train test split on the data, with the test size of 10% and a random\_state of 101**

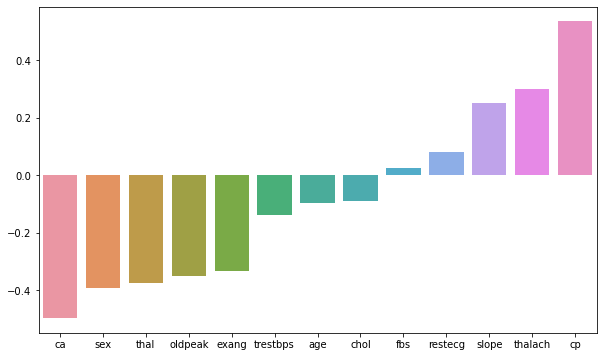
#6 **Create a StandardScaler object and normalize the X train and test set feature data. Make sure you only fit to the training data to avoid data leakage (data knowledge leaking from the test set).**

**#7 Create a Logistic Regression model and use Cross-Validation to find a well-performing C value for the hyper-parameter search. You have two options here, use *LogisticRegressionCV* OR use a combination of *LogisticRegression* and *GridSearchCV*. The choice is up to you.**

**#8 Report back your search's optimal parameters, specifically the C value.**

**#9 Report back the model's coefficients**

**#10 Create a visualization of the coefficients by using a barplot of their values**

****

**#11 Model Performance Evaluation**

**Create the following evaluations**

**\* Confusion Matrix Array**

**\* Confusion Matrix Plot**

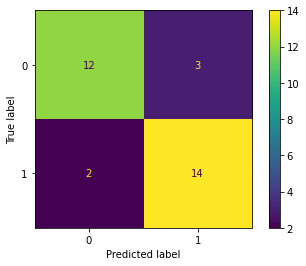
**\* Classification Report**

**\* Confusion Matrix Array**

**array([[12, 3],**

**[ 2, 14]], dtype=int64)**

**\* Confusion Matrix Plot**

****

**\* Classification Report**

precision recall f1-score support

0 0.86 0.80 0.83 15

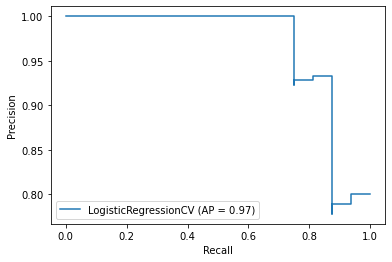
1 0.82 0.88 0.85 16

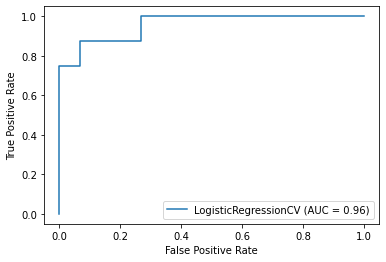
accuracy 0.84 31

macro avg 0.84 0.84 0.84 31

weighted avg 0.84 0.84 0.84 31

Performance Curves





Finally do the Prediction for the following patient

**A patient with the following features has come into the medical office:**

age 48.0

sex 0.0

cp 2.0

trestbps 130.0

chol 275.0

fbs 0.0

restecg 1.0

thalach 139.0

exang 0.0

oldpeak 0.2

slope 2.0

ca 0.0

thal 2.0