Classification

“Heart Disease”

# DATASET from UCI Machine learning Repositories website:

# “Stat log (Heart Disease)”

**Dataset Information**

* This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them.
* In particular, the Cleveland database is the only one that has been used by ML researchers to this date.
* The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.
* The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

|  |  |
| --- | --- |
| **Stat log (Heart) Data Set**  This dataset is a heart disease database similar to a database already present in the repository (Heart Disease databases) but in a slightly different form |  |

* This database contains 13 attributes which have been extracted from a larger set of 75
* Dataset is available at **UCI** website
* No missing values.
* 270 observations
* Variable to be predicted: Absence (1) or presence (2) of heart disease.

**Attribute Information**

1. age: age in years
2. sex: sex (1 = male; 0 = female)
3. chest pain type (4 values)
   * + Value 1: typical angina
     + Value 2: atypical angina
     + Value 3: non-anginal pain
     + Value 4: asymptomatic
4. resting blood pressure (in mm Hg on admission to the hospital)
5. chol: serum cholestoral in mg/dl
6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg: resting electrocardiographic results (values 0,1,2)
   * 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
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina (1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment
    * + Value 1 : upsloping
      + Value 2: flat
      + Value 3: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy
13. thal:
    * + value 3 = normal
      + value 6 = fixed defect
      + value 7 = reversable defect

# Code section:

# Importing

# Screenshot_2020-03-03 classification assignment - Jupyter Notebook.png

# Data read

# 2.png

# Dataset info

# 3.png

# Balance factor

# 4.png

# Presence 150 and absence 120 its balanced

# Creating model class

# 5.png

# Decision tree classifier: Accuracy

# 6.png

Accuracy of Decision tree classification: **0.7407407407407407**

# Tree plotting

# 7.png

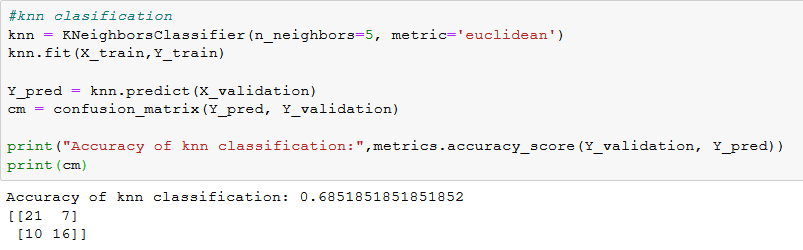
# index.png

# Random forest classifier Accuracy:

# 8.png

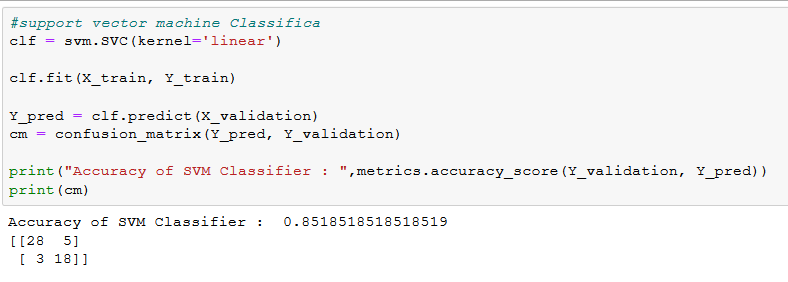
Accuracy random forest classification: **0.7592592592592593**

1. KNN Classification Accuracy :



Accuracy of KNN classification: **0.6851851851851852**

1. Support vector machine Classification(SVM)



Accuracy of SVM Classifier: **0.8518518518518519**

1. MultilayerPercetron classification

# 11.png

Accuracy of MLP Classifier: **0.5740740740740741**

…………………………………………………………