

<https://www.kaggle.com/ronitf/heart-disease-uci>

#### Data Set 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. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this directory.

To see Test Costs (donated by Peter Turney), please see the folder "Costs"

#### Attribute Information:

Only 14 attributes used: 1. #3 (age) 2. #4 (sex) 3. #9 (cp) 4. #10 (trestbps) 5. #12 (chol) 6. #16 (fbs) 7. #19 (restecg) 8. #32 (thalach) 9. #38 (exang) 10. #40 (oldpeak) 11. #41 (slope) 12. #44 (ca) 13. #51 (thal) 14. #58 (num) (the predicted attribute)

1)age

age in years

2)sex

(1 = male; 0 = female)

3)cp

chest pain type

4)trestbps

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

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

12)ca

number of major vessels (0-3) colored by flourosopy

13)thal

3 = normal; 6 = fixed defect; 7 = reversable defect

14)target

1 or 0