Dataset Documentation

Sanjana Farial (14.02.04.100)
Sharmin Sultana (13.02.04.006)

1 Title: Heart Disease Dataset

2 Data-set Information:

This dataset contains various data to work on for both male and female patients of different ages to predict the existence of heart disease.

3 Sources:

1. Processed Cleveland data [1]
2. Heart Disease Male [2]

4 Instance Information:

There are 275 instances.

5 Attribute Information:

9 attribute:

1. age
2. gender
3. chest-pain
4. rest-bpress
5. blood-suger
6. rest-electro
7. max-heart-rate
8. exercise-angina
9. disease(target attribute)

6 Complete Attribute Documentation:

1. age: Minimum age is 28 and maximum age is 76
2. gender: Number of **male** patients is 177 and Number of **female** patients is 97
3. chest-pain: chest pain type
   - typ angina = typical angina
   - atyp angina = atypical angina
   - non anginal = non anginal pain
   - asympt = asymptomatic pain
4. rest bpress: resting blood pressure(in mm Hg)
5. blood suger: fasting blood sugar > 120 mg/dl(t = true, f = false)
6. rest electro: resting electrocardiographic results
   - normal
   - st-t-wave abnormality(T wave inversions and/or ST elevation or depression of > 0.05 mV)
   - left-vent-hyper(showing probable or definite left ventricular hypertrophy by Estes criteria)
7. max heart rate: maximum heart rate achieved
8. exercise angina: exercise induced angina(yes or no)
9. disease(target attribute): positive for presence of disease and negative for absence of disease

7 Class Distribution

Number of data belongs to "Positive" class = 117 and "Negative" Class = 157

References

[1] Aha, D., and Kibler, D. Instance-based prediction of heart-disease presence with the cleveland database. Tech. rep., University of California, Mar 1980.

[2] Ricco Rakotomalala. Data scientist.