Final Project

CUNY621 – MS in Data Science

Group5, April 21, 2019

Presence of Heart Disease prediction

# Description

Data Science is used across many spheres around the world. The healthcare industry is no exception. It can play an essential role in predicting presence Heart diseases. Such information, if predicted well in advance, can provide important insights to doctors who can then adept their diagnosis and treatment per patient basis.

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

See if you can find any other trends in heart data to predict certain cardiovascular events or find any clear indications of heart health.

# DATA

The 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.

There are a total of 13 features and 1 target variable.

Attribution Information:

* Age
* Sex
* Chest pain types (4 values)
* Resting blood pressure
* Serum cholesterol in mg/dl
* Fasting food 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 fluoroscopy
* thal: 3 = normal; 6 = fixed defect; 7 = reversible defect

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.

# Hypothesis and Discussion

According to Centers for Disease Control and Prevention there are several health conditions, lifestyle, and you age and family history can increase your risk for heart disease. About half of all Americans (47%) have at least one of three key risk factors for heart disease: high blood pressure, high cholesterol, and smoking.

# Challenges

* Only 303 cases available.
* There is no single feature that has a very high correlation with our target value.

REFERENCES

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

[2] <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>

[3] <https://www.cdc.gov/heartdisease/risk_factors.htm>