**Chuẩn đoán bệnh tim**

1. **Tập dữ liệu**

- Sử dụng tập dữ liệu Cleveland đây là tập dữ liệu duy nhất được các nhà nghiên cứu ML sử dụng ngày nay.

- Tập dữ liệu bao gồm 14 thuộc tính và 303 mẫu tin được sử dụng cho mục đích phân loại bệnh tim của một bệnh nhân.

- Sử dụng 13 thuộc tính làm đặc trưng, bao gồm: age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca. thal

1. age: age in years (tuổi)

2. sex: sex (giới tính) (1 = male; 0 = female)

3. cp: chest pain type (loại đau ngực)

-- Value 1: typical angina (đau thất ngực điển hình)

-- Value 2: atypical angina (đau thất ngực không điển hình)

-- Value 3: non-anginal pain (không đau)

-- Value 4: asymptomatic (không triệu chứng)

4. trestbps: resting blood pressure (in mm Hg on admission to the

hospital)

5. chol: serum cholestoral in mg/dl (cholestoral trong máu) (đơn vị mg/dl)

6. fbs: fasting blood sugar > 120 mg/dl (lượng đường trong máu > 120mg/dl) (1 = true, 0 = false)

7. restecg: resting electrocardiographic results (kết quả điện tâm đồ)

-- Value 0: normal (bình thường)

-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) (có bất thường sóng ST-T (nghịch đảo sóng T và / hoặc độ cao ST hoặc trầm cảm> 0,05 mV))

-- Value 2: showing probable or definite left ventricular hypertrophy

by Estes' criteria (cho thấy phì đại thất trái hoặc xác định

                     theo tiêu chí của Estes)

8. thalach: maximum heart rate achieved (nhịp tim tối đa đạt được)

9. exang: exercise induced angina (tập thể dục gây đau thắt ngực) (1 = true; 0 = false)

10. oldpeak = ST depression induced by exercise relative to rest (ST trầm cảm gây ra bởi tập thể dục liên quan đến phần còn lại)

11. slope: the slope of the peak exercise ST segment (độ dốc của đoạn ST tập thể dục đỉnh)

-- Value 1: upsloping (lên dốc)

-- Value 2: flat (bằng phẳng)

-- Value 3: downsloping (xuống dốc)

12. ca: number of major vessels (0-3) colored by flourosopy (số lượng tàu lớn (0-3) được tô màu bằng phương pháp fluoroscopy)

13. thal: 3 = normal; 6 = fixed defect (khuyết tật cố đình); 7 = reversable defect (khuyến tật đảo ngược)

- Sử dụng thuộc tính pred\_attribute làm nhãn chuẩn đoán bệnh. Để đơn giản cho việc chuẩn đoán qui định nhãn có giá trị 0 thì có giá trị 0, các nhãn có giá trị 1, 2, 3, 4 có giá trị là 1

pred\_attribute: diagnosis of heart disease (angiographic disease status) (tình trạng tim mạch)

-- Value 0: < 50% diameter narrowing (xác xuất bị bệnh nhỏ hơn 50%)

-- Value 1: > 50% diameter narrowing (xác xuất bị bệnh lớn hơn 50%)

Database: 0 1 2 3 4 Total

Cleveland: 164 55 36 35 13 303

Review Data Source

The authors of the databases have requested:

...that any publications resulting from the use of the data include the names of the principal investigator responsible for the data collection at each institution. They would be:

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.