* Tên bộ dữ liệu: HeartDisease
* Đặc điểm của tập dữ liệu
  + Tổng số mẫu: 21061
  + Số thuộc tính: 13 cột đầu tiên là các giá trị dạng số nguyên hoặc số thập phân
  + Mục tiêu: Cột 14 là biến nhị phân xác định có hay không mắc bệnh tim
* Giải thích các thuộc tính:
  + Age: tuổi tính theo năm
  + Sex: Giới tính, 1 – Nam, 0 – Nữ
  + Cp: Kiểu đau ngực 1, 2, 3, 4
  + Trestbps: Huyết áp khi nghỉ ngơi (mmHg)
  + Chol: cholesterol trong huyết thanh (mgDl)
  + Fbs: đường huyết lúc đói có lớn hơn 120 mg/dl không? 1 – Có, 0 - Không
  + Restecg: kết quả điện tâm đồ lúc nghỉ (1, 2, 3)
  + Thalach: Nhịp tim tối đa đạt được
  + Exang: đau ngực khi tập thể dục. 1 – Có, 0 - Không
  + Oldpeak: ST gây ra do tập thể dục liên quan đến nghỉ ngơi
  + Slope: độ dốc của đoạn ST tập thể dục đỉnh cao
  + Ca: số lượng các mạch chính (0-3) được tô màu bằng bột màu
  + Thal: Bệnh Thalassemia. 3 – Không bị, 6 – Mắc bệnh không hồi phục, 7 - Mắc bệnh có hồi phục