Báo cáo IOT

Họ và tên:

Dương Văn Giang – B19DCCN204

Dataset: Dữ liệu về bệnh tim mạch

<https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset?fbclid=IwAR1KSxZnoOmIZADAF2rLB5kTNjZSJfxLLseiUWVG59zBS8aXAjvvq4VhjRk>

1. Về bộ dữ liệu

Đây là bộ dữ liệu thu nhập về thông tin của bệnh tim mạch

1 Tập dữ liệu gồm 12 thông tin bao gồm:

2 Age: tuổi

3 Height: chiều cao bệnh nhân(cm)

4 Weight: cân nặng của bệnh nhân(kg)

5 Gender: giới tính

6 Systolic blood pressure: Huyết áp tâm thu

7 Diastolic blood pressure: Huyết áp tâm trương

8 cholesterol,

9 Glucose,

10 Smoking: bệnh nhân có hút thuốc lá không

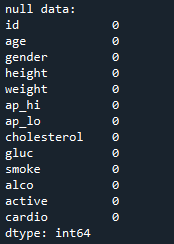
11 Alcohol intake: Bệnh nhân có thường uống rượu

12 Physical activity: Bệnh nhân có thường hoạt động



1. Xử lý dữ liệu

* Mising data
* 



Ta thấy không có dataset nào tồn tại 1 giá trị bên trong là rỗng

, Duplicate data



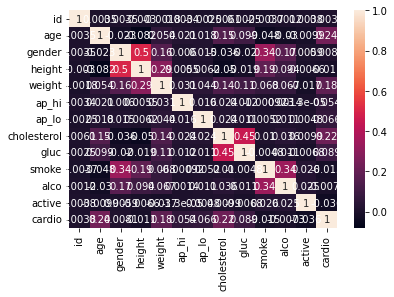


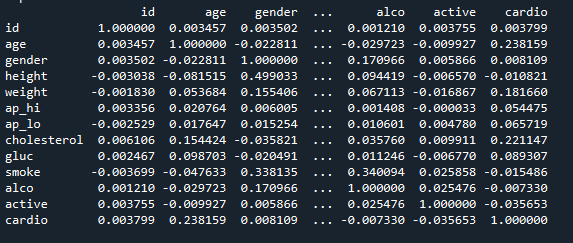
Ta thấy không có hàng nào trùng lặp trong tập dữ liệu

1. Chọn feature

Ta có tận 12 feature để lựa chọn, để biểu thị sự tương quan giữa feature và kết quả ta dùng heatmap





Ta cũng có bảng tương quan sau 

Ta lựa chọn các feature sau: age, height, weight, ap\_hi, ap\_lo, cholesterol, gluc, smoke, active.

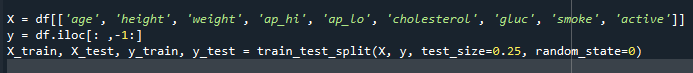
1. Tách tập dữ liệu.

Để thuận lợi cho việc đào tạo và đánh giá hiệu xuaart của mô hình ta chia dataset thành 2 tập dữ liệu con là: train dataset và test dataset với tỉ lệ 70:30

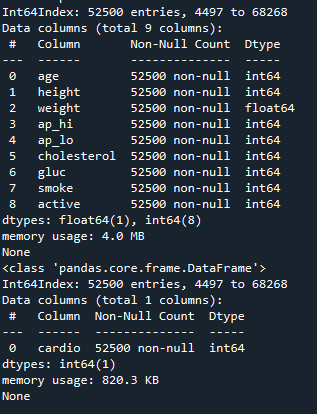
Train dataset dùng để huẩn luyện mô hình

Test dataset dùng để đánh giá mô hình

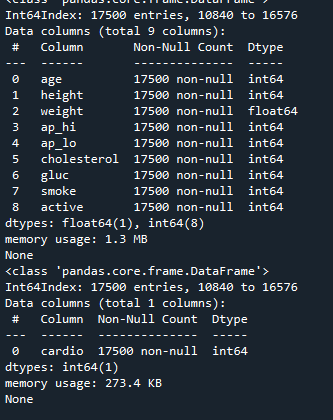
Ta sử sụng sklearn để tách ra làm 2 dataset



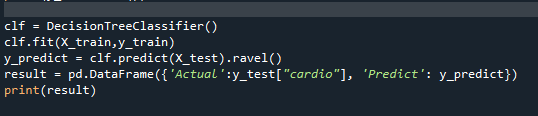
Tập train:



Tập test:

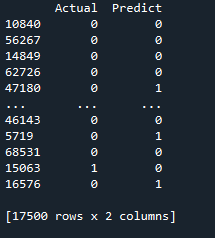


1. Đào tạo model

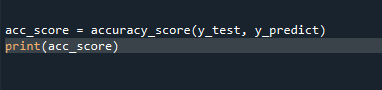
* Mục đích là ta có thể dự đoán được người bệnh có khả năng bị bệnh tim mạch hay không
* Đây là bài toán classification do vậy ta sử dụng DecitionTreeClassifier để sử lí bài toán này
* 

Sau khi train model ta sử dụng tập test để ước lượng độ chính xác của model

Ta có kết quả



Để dễ dàng tính toán được độ chính xác của model ta sử dụng accuracy\_score



Và có kết quả như sau

