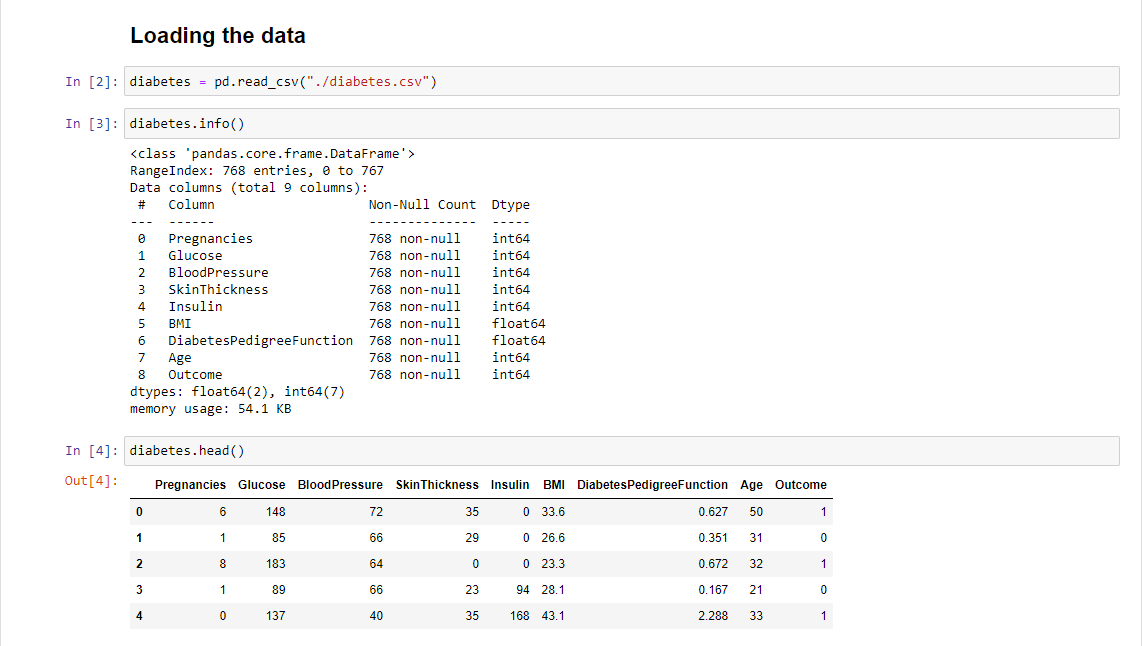
Case study 1: Xây dựng mô hình dự đoán 1 người có bị tiểu đường hay không?

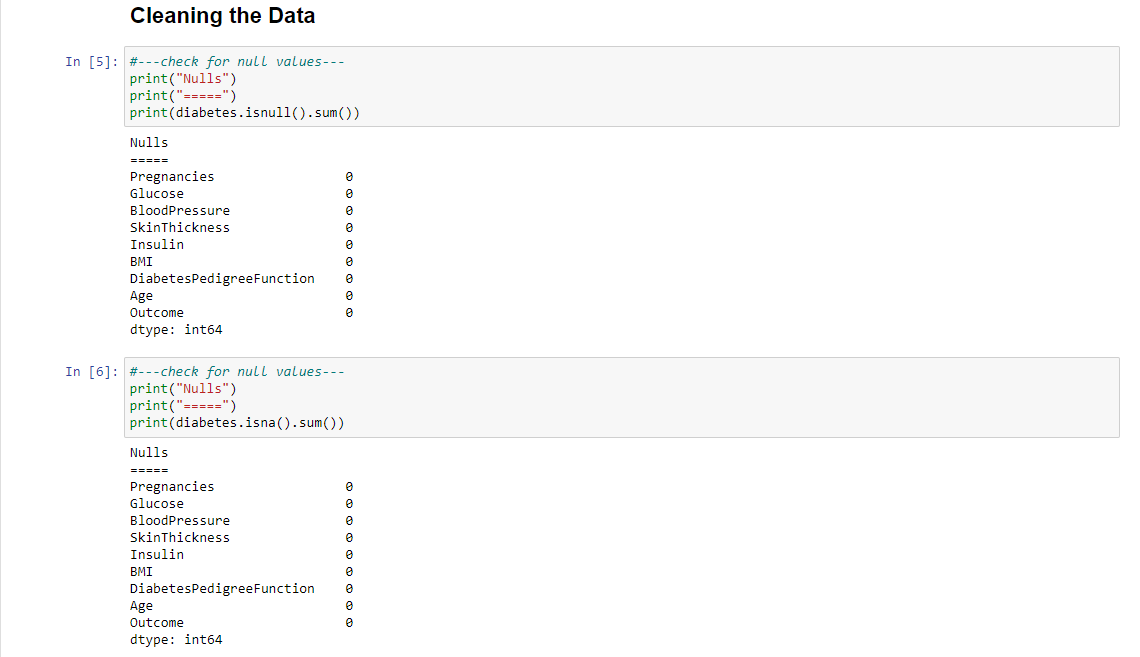
(Outcome: 1: Tiểu đường, 0: Không bị tiểu đường)

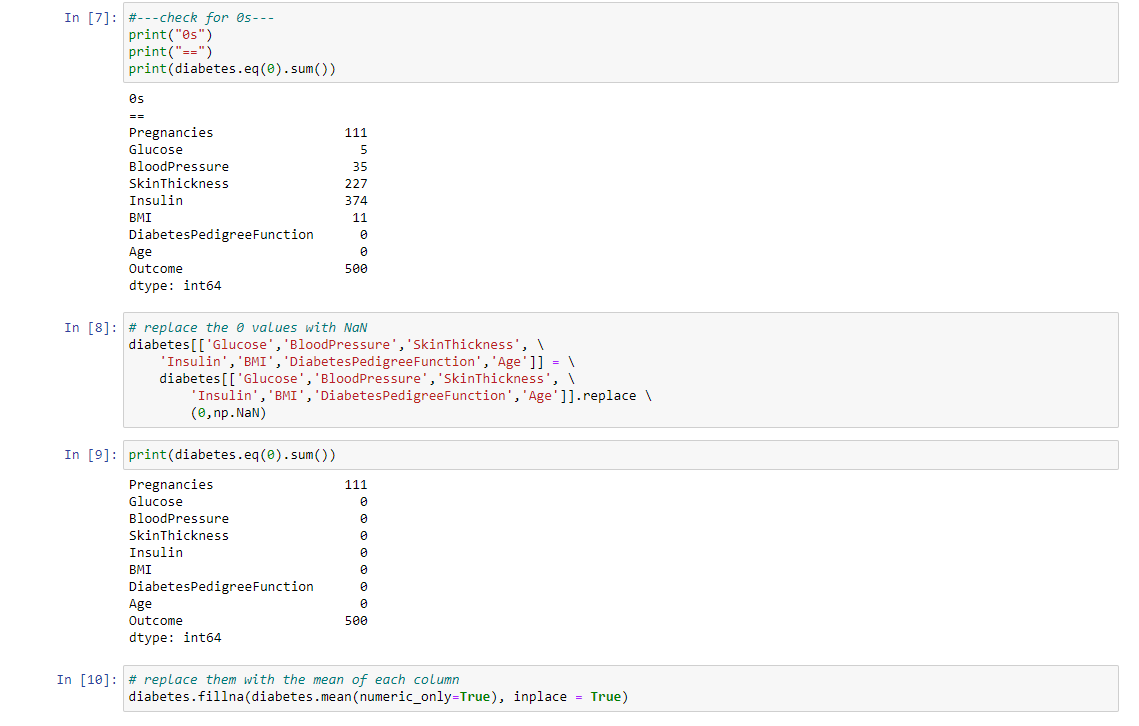
1. Xử lí dữ liệu

* Đọc dữ liệu từ file csv: với target=Outcome

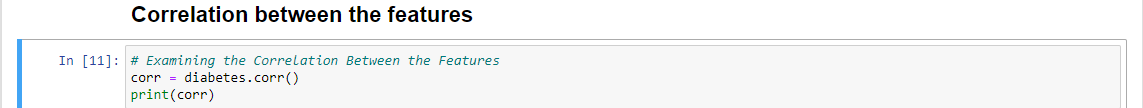


* Làm sạch dữ liệu: thay những giá trị null bằng giá trị trung bình của cột đó





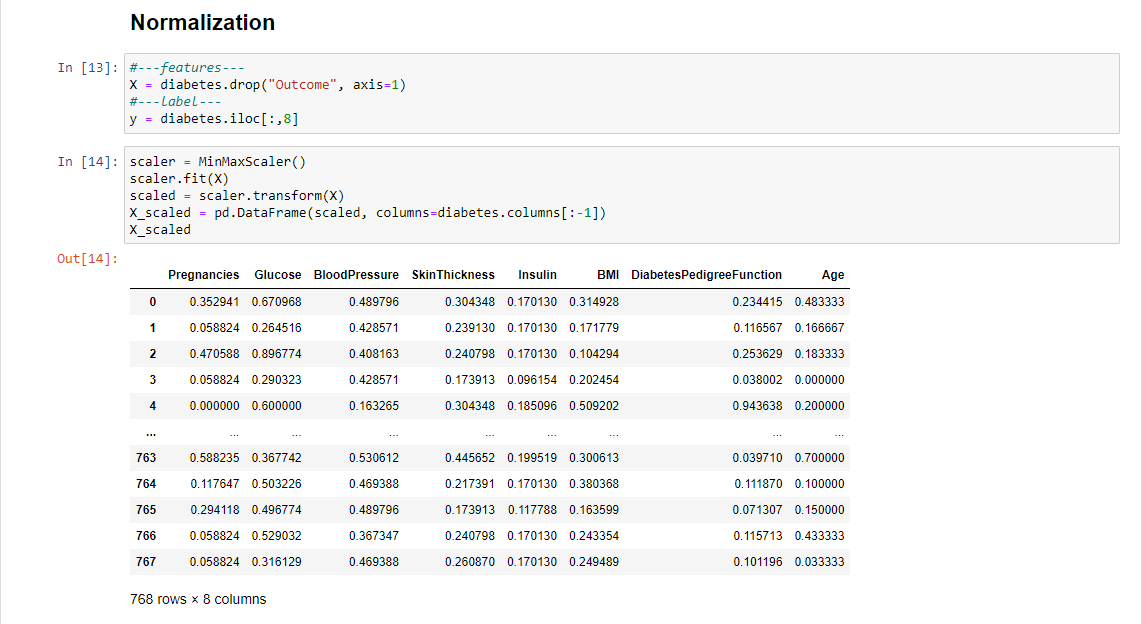
* Thể hiện độ tương quan giữa các features với seaborn





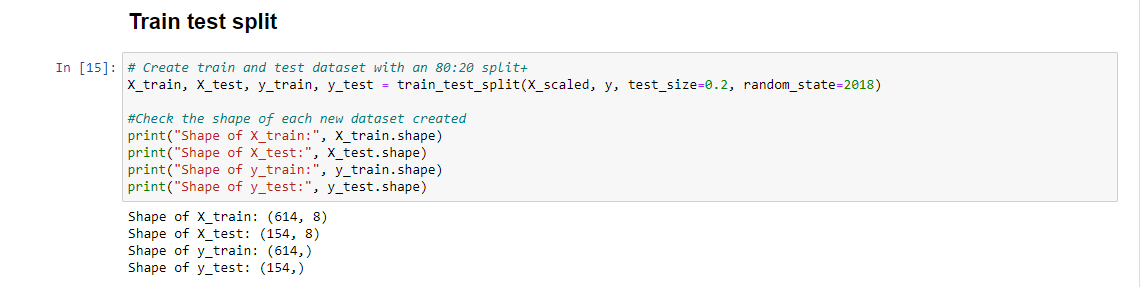


* Normalization: MinMaxScaler() -> dữ liệu sẽ nằm trong khoảng (0, 1)

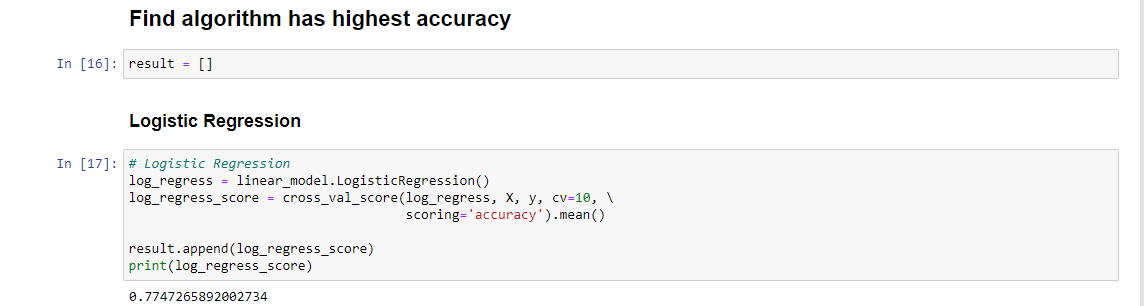


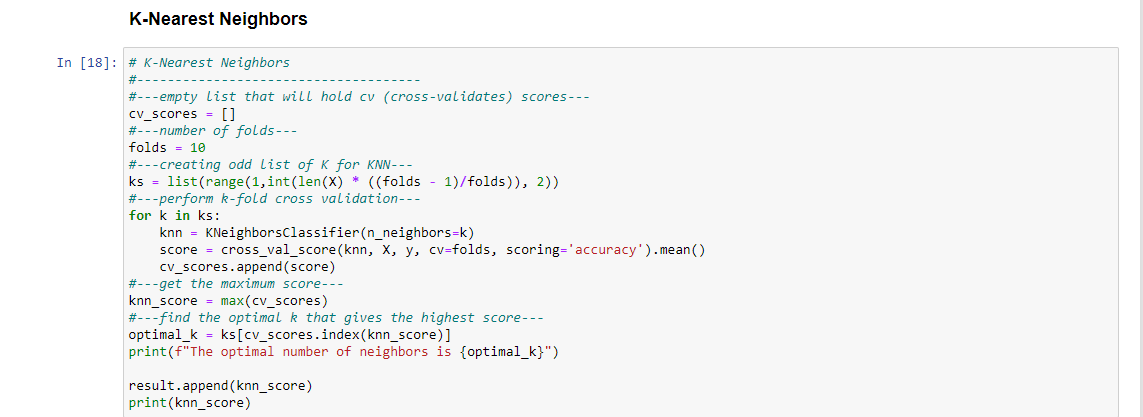
1. Mô hình machine learning

* Chia dữ liệu thành 2 tập train & test với tỉ lệ 80:20

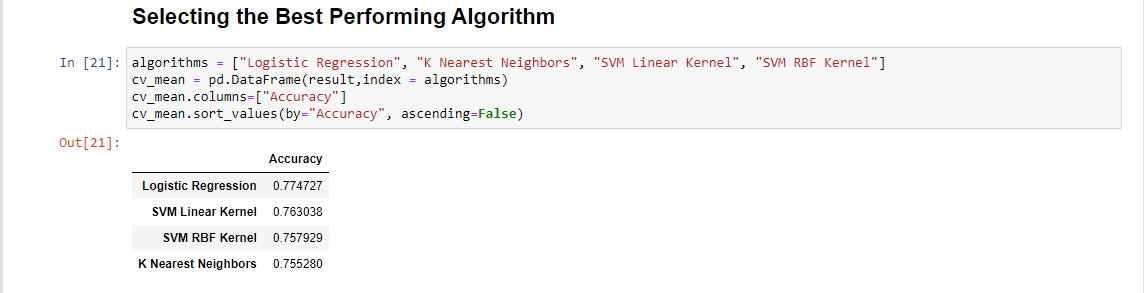


* Tiến hành lựa chọn thuật toán machine learning cho độ chính xác cao nhất trong 4 thuật toán: Logistic Regression, KNN, Support Vector Machines, RBF Kernel => kết quả chọn thuật toán Logistic Regression





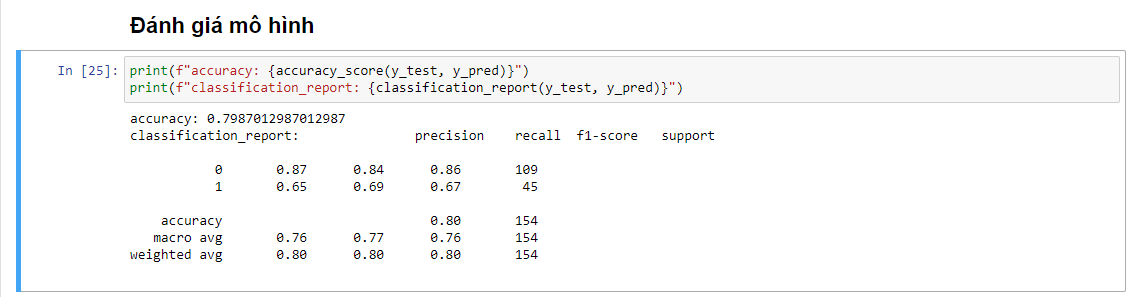




* Xây dựng mô hình

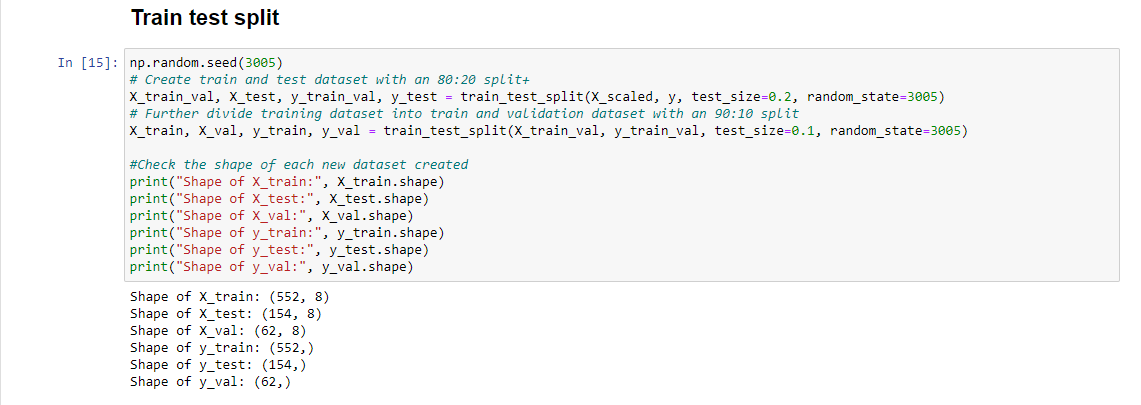


* Đánh giá mô hình:

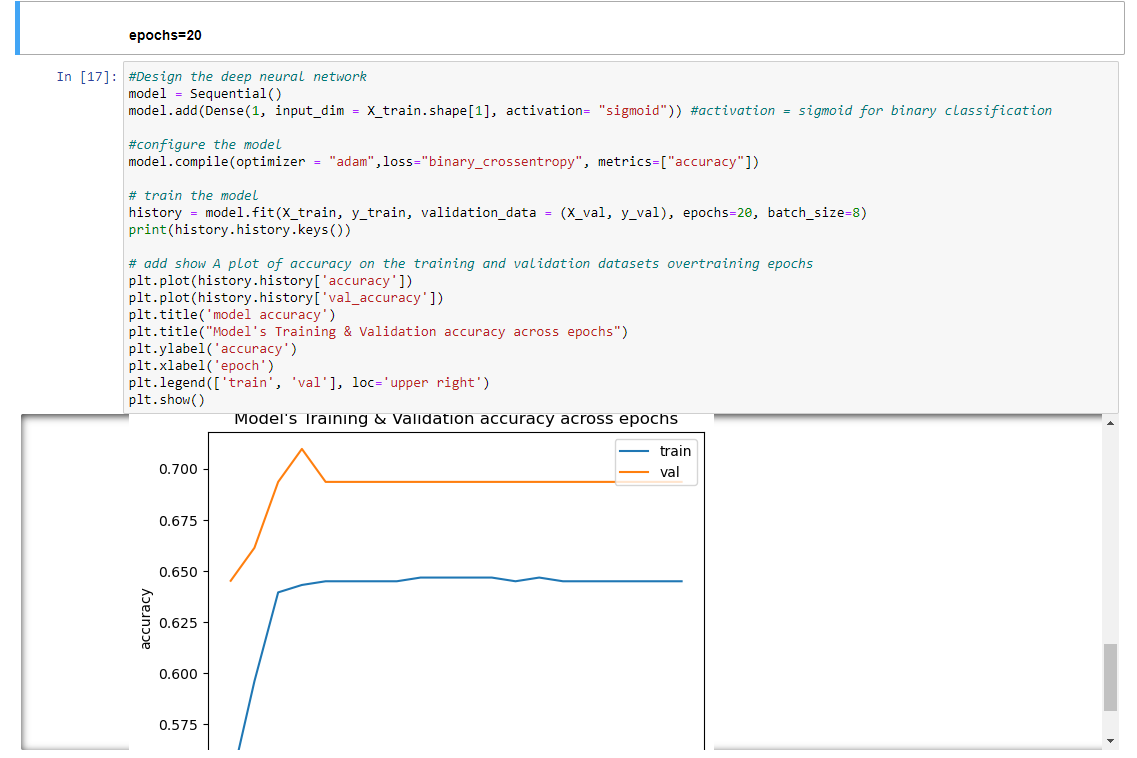


1. Mô hình deep learning

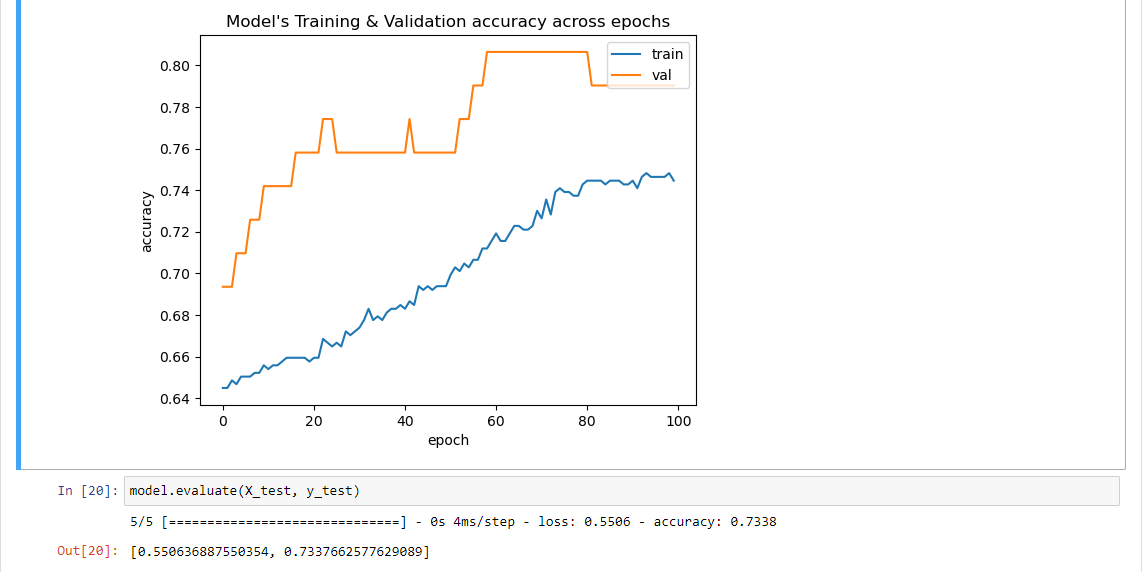
* Chia dữ liệu thành 3 tập train, val, test



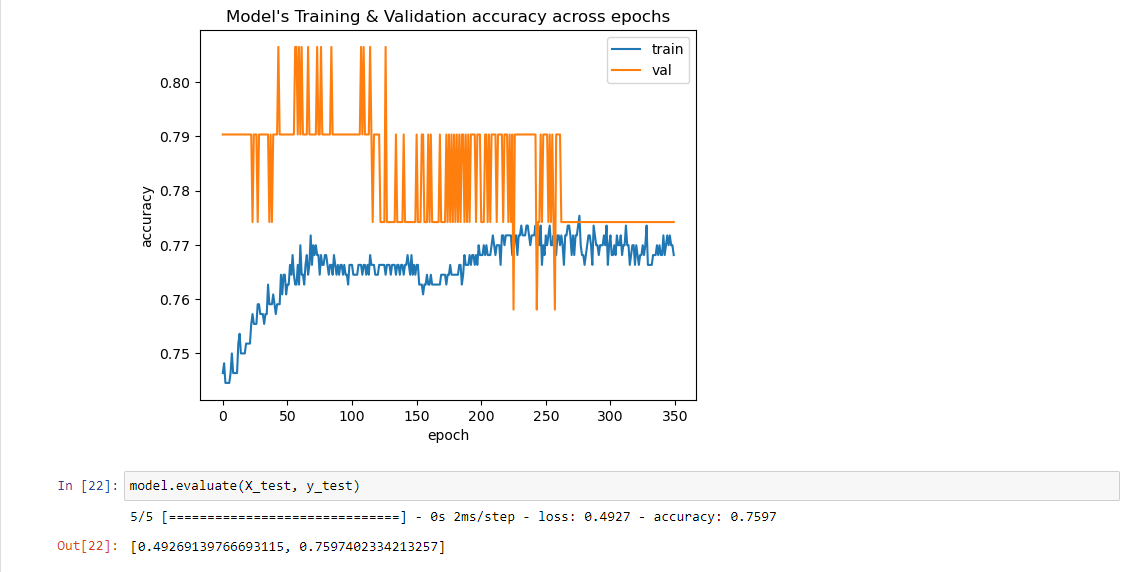
* Xây dựng mô hình Deep Learning phân loại nhị phân
* 0 hidden layer - 1 neurons, batch\_size=8
  + epochs=20



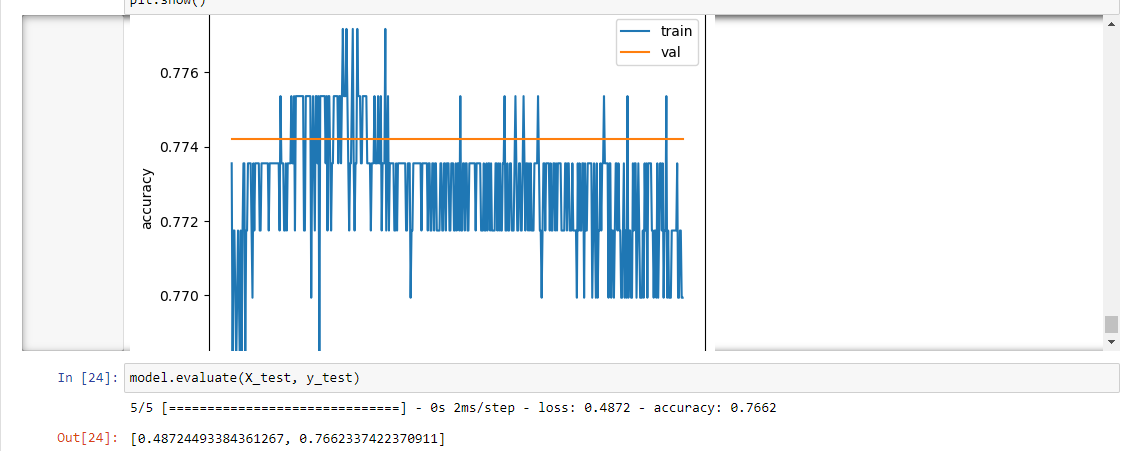
* + epochs=100



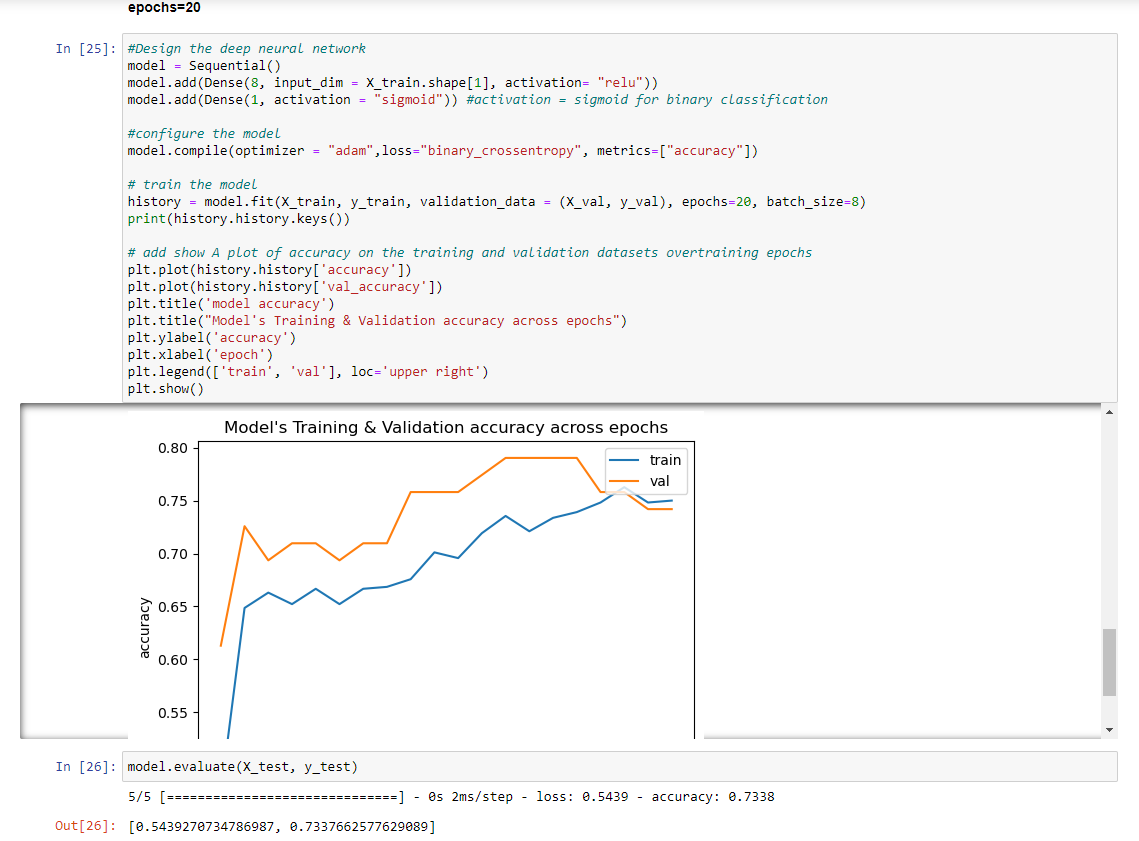
* + epochs=350



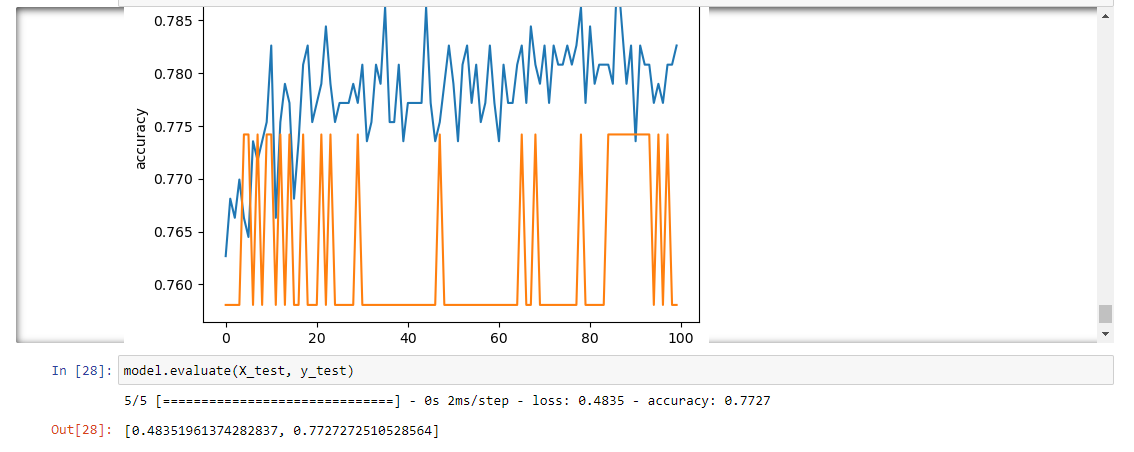
* + epochs=500



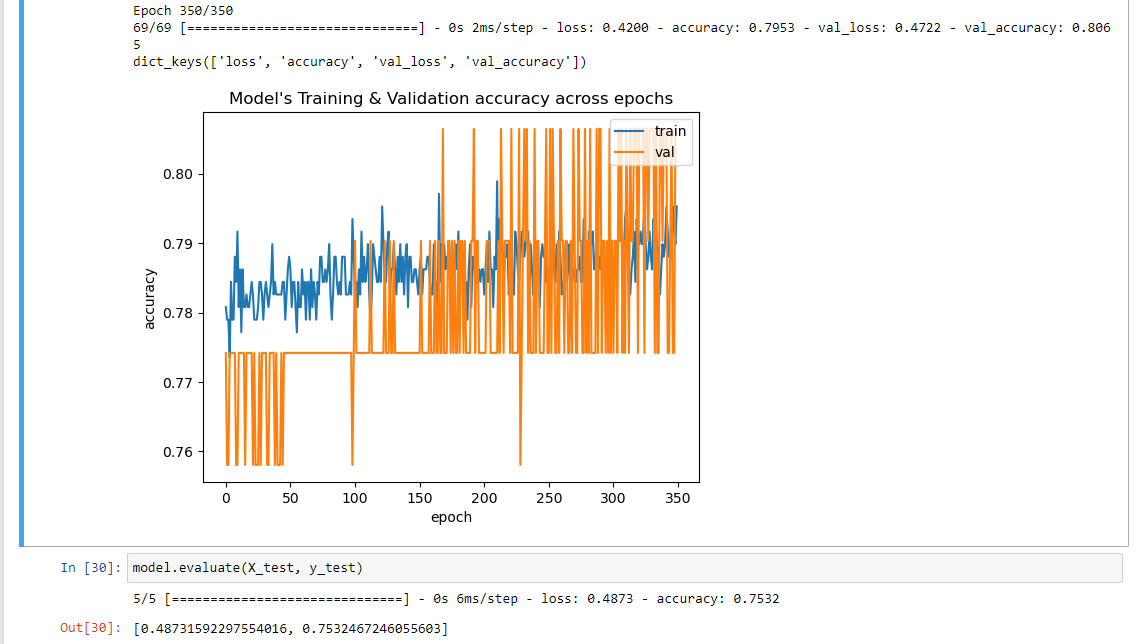
* 1 hidden layer - 8 neurons, batch\_size=8
  + epochs=20



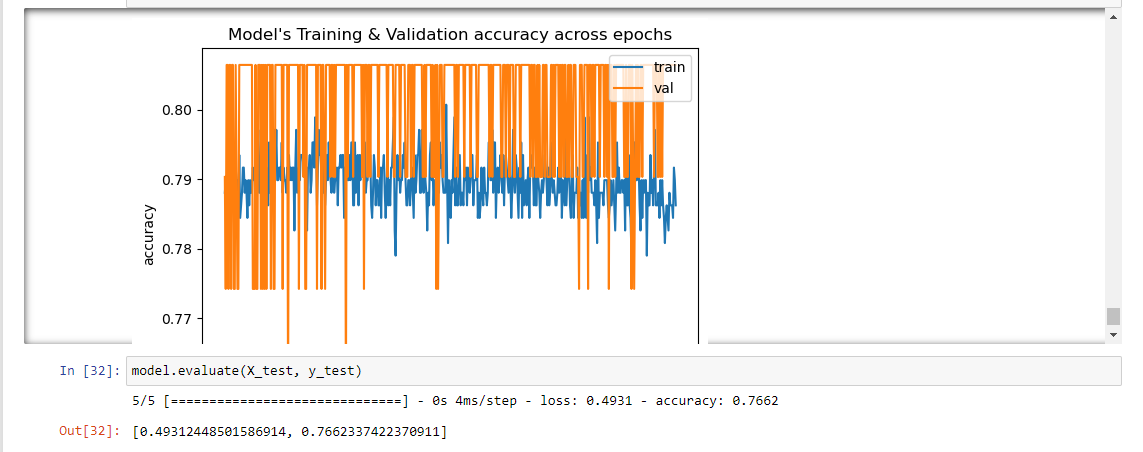
* + epochs=100



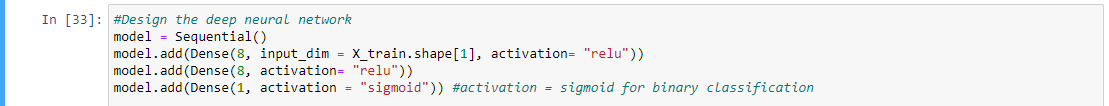
* + epochs=350

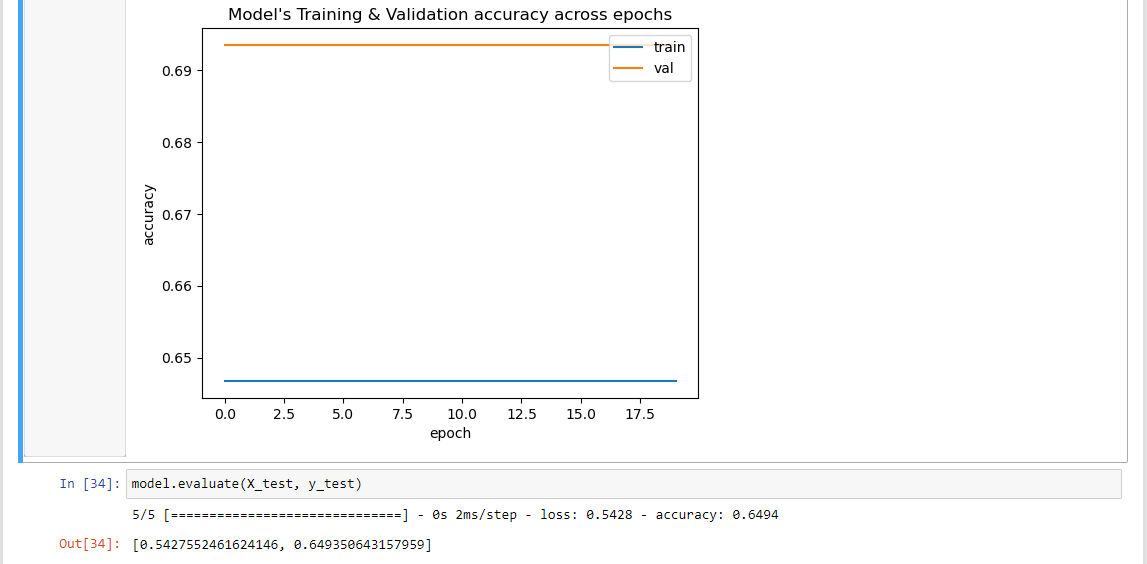


* + epochs=500

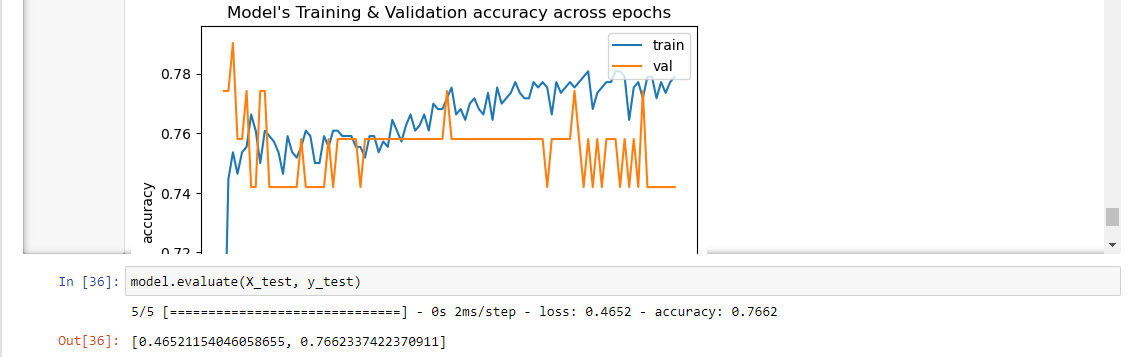


* 2 hidden layer - 8 neurons, batch\_size=8
  + epochs=20

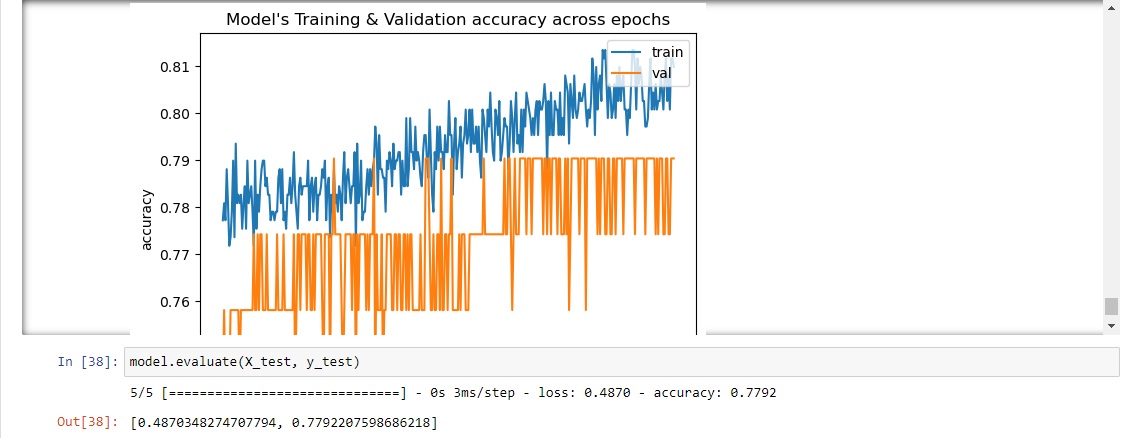




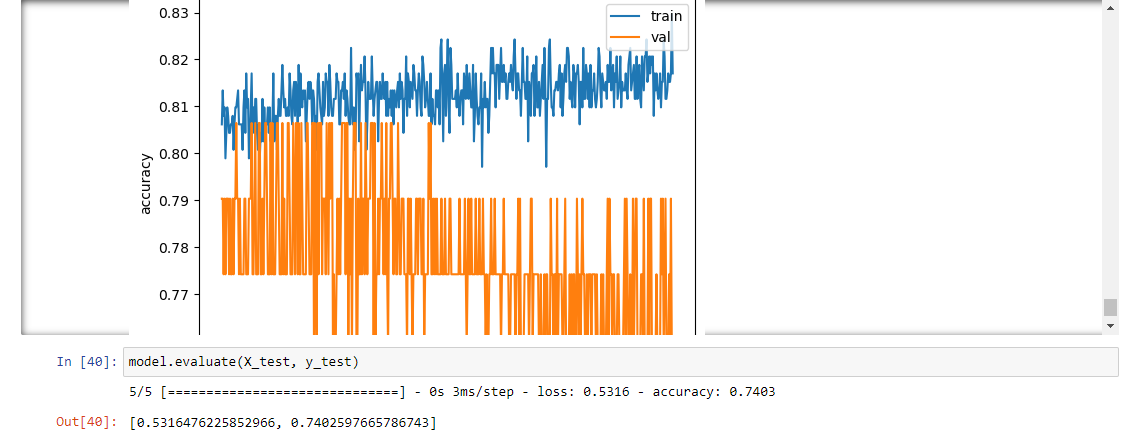
* + epochs=100



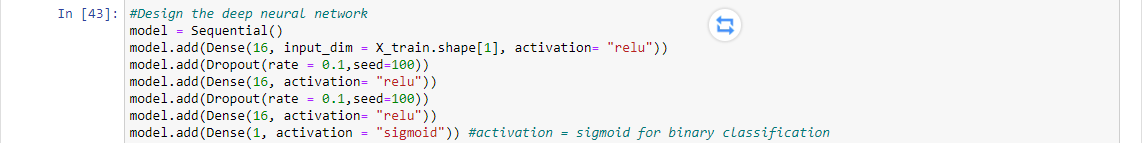
* + epochs=350

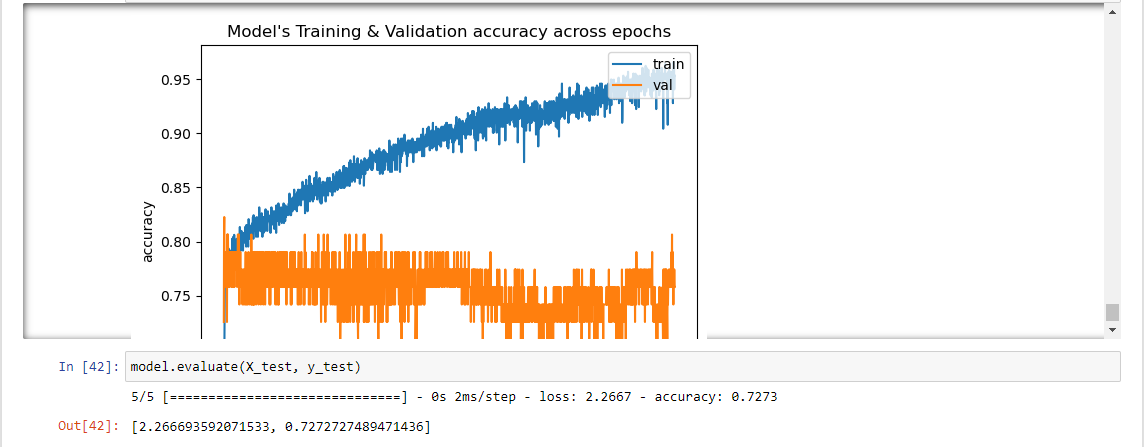


* + epochs=500



* Deepest - 3 hidden layer - 16 neurons, batch\_size=8, epochs=2000
  + Không dùng Dropout





* + Dùng Dropout

