

## BÁO CÁO THỰC HÀNH KHAI THÁC DỮ LIỆU – TUẦN 9

Họ tên : Nguyễn Tiến Phong  
MSSV : 20280071

Phân lớp dữ liệu thông qua thuật toán Support Vector Machines (SVM) :

1. Đọc và xử lý tập dữ liệu banknote authentication từ UCI Machine Learning Repository :
  - Chuyển dữ liệu thành 2 mảng X (variance, skewness, curtosis, entropy) và y (class)
  - Xác định tính tuyến tính của dữ liệu thông qua ma trận tương quan và scatterplot :  
Nhận thấy rằng tập dữ liệu có tính tuyến tính mặc dù mức độ tương quan giữa các feature khác yếu.
  - Tách dữ liệu thành tập train với 80% và tập test với 20%
2. Xây dựng mô hình với kernel='linear' và C='100'.

In ra classification\_report thì nhận thấy rằng mô hình có vẻ bị overfitting

	precision	recall	f1-score	support
0	0.99	0.99	0.99	148
1	0.98	0.98	0.98	127
accuracy			0.99	275
macro avg	0.99	0.99	0.99	275
weighted avg	0.99	0.99	0.99	275

3. Vẽ các biên quyết định  
Phần này không biết vẽ cho 4 features thế nào nên em vẽ cho từng cặp feature nhưng thấy khá là rối do model áp dụng cho 4 features nhưng lại sử dụng cho từng cặp (em chỉ vẽ siêu phẳng lề lớn nhất).