

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH



**UIT**  
Trường Đại học  
Công nghệ Thông tin

**Khoa Khoa học  
và Kỹ thuật Thông tin**

## **BÁO CÁO BÀI TẬP THỰC HÀNH 6**

### **XÂY DỰNG MÔ HÌNH MÁY HỌC TRÊN DỮ LIỆU ĐÃ THU THẬP**

Môn học: Thu thập và tiền xử lý dữ liệu - DS103.N21

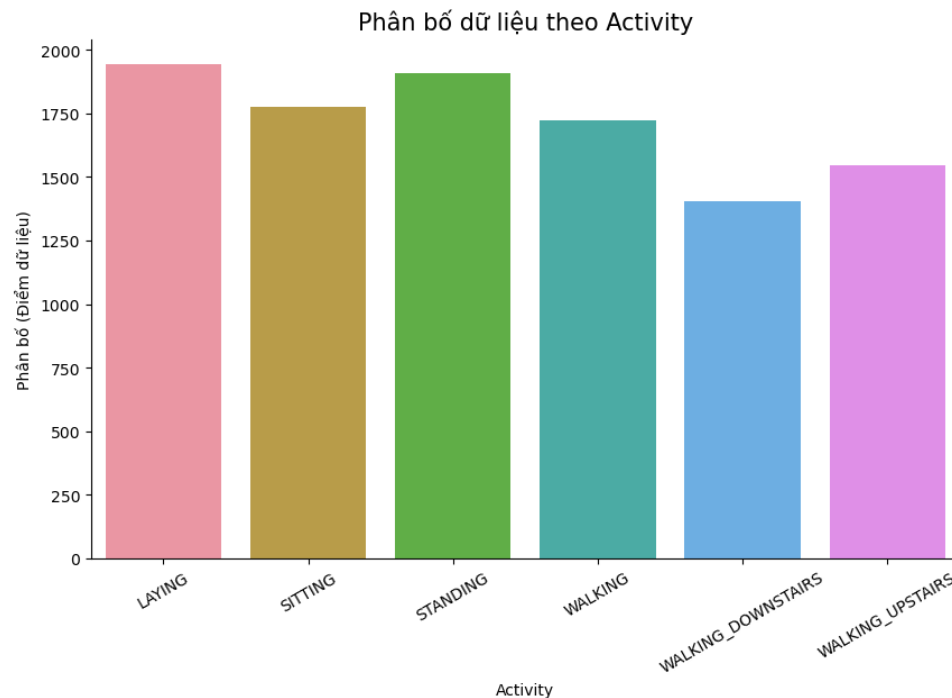
Tên: Nguyễn Thị Huyền Trang

MSSV: 21520488

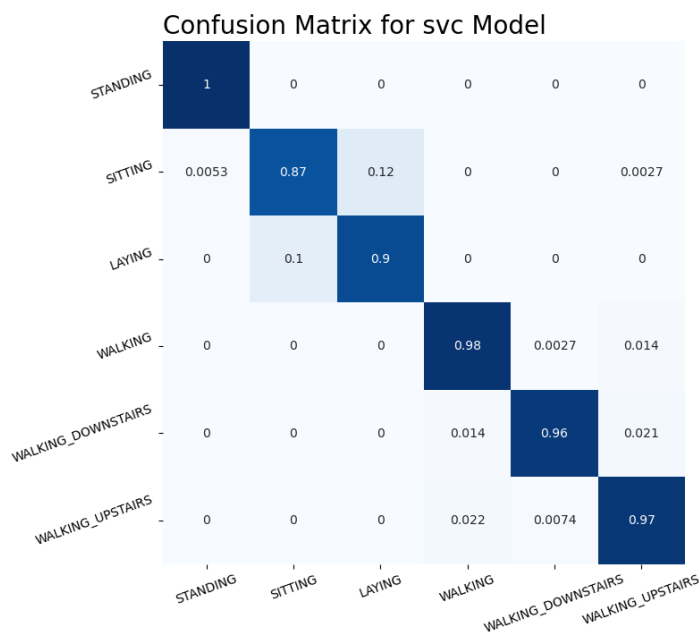
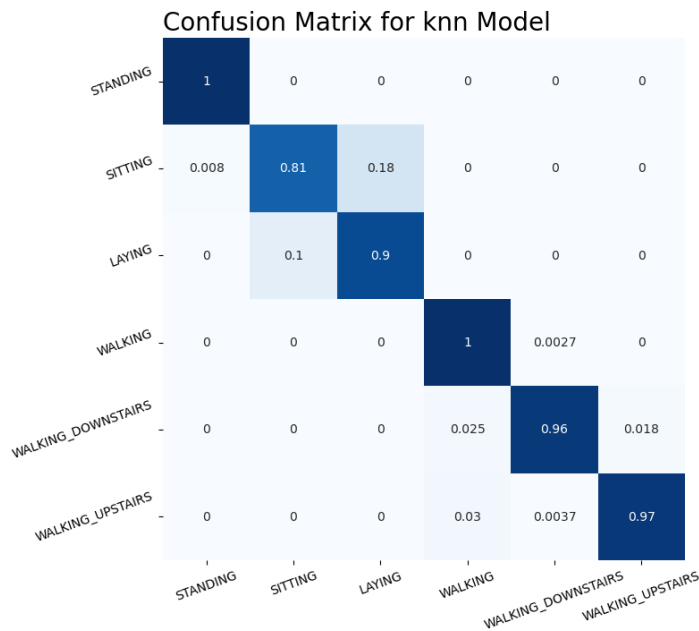
## Bài 1: Sử dụng lại bộ dữ liệu Human Activity Recognition Using Smartphones Data Set

1. Đọc dữ liệu đã được tiền xử lý trước đó. Bỏ đi thuộc tính Subject.
2. Thống kê số lượng nhãn, số lượng thuộc tính và phân bố dữ liệu theo nhãn.

- Số lượng nhãn: 6
- Số lượng thuộc tính: 80
- Biểu đồ phân bố điểm dữ liệu:



3. Chuẩn bị dữ liệu: phân chia dữ liệu ban đầu thành 2 tập: huấn luyện (train) và kiểm thử (test) với tỉ lệ 8-2.
4. Cài đặt mô hình: kNN và SVM. Sử dụng kỹ thuật cross validation.
5. Đánh giá mô hình: Độ chính xác (Accuracy) Ma trận nhầm lẫn (Confusion matrix).
  - Ta sử dụng độ \*chính xác (Accuracy)\* để đánh giá cho mô hình
  - Độ chính xác của mô hình kNN: 93.6408



## 6. Có nhận xét gì về hiệu quả giữa 2 mô hình?

Độ chênh lệch về hiệu quả của 2 mô hình (Accuracy) tương đối giống nhau, mô hình SVC có kết quả Accuracy lớn hơn mô hình kNN 1 khoảng rất rất nhỏ. (không đáng kể).

## 7. Thực hiện lại các bước từ 1 tới 7 khoảng 5 lần, ứng với mỗi lần thực hiện ghi lại giá trị Accuracy và thay đổi giá trị seed. Sau đó tính và ghi ra kết quả Accuracy trung bình của 5 lần thực hiện

Sau khi thực hiện 5 lần lặp và thay đổi giá trị seed của 2 mô hình kNN và SVC, thu được các giá trị Accuracy của mỗi lần lặp như sau:

Lần lặp	seed	Knn	SVC
1	643	95.194175	95.000000
2	694	93.689320	94.029126
3	424	94.757282	94.708738
4	113	94.417476	94.660194
5	19	94.368932	94.077670

Avg.Accuracy sau 5 lần lặp của kNN: 94.4854

Avg.Accuracy sau 5 lần lặp của SVC: 94.4951

*Nhận xét:*

Khi thực hiện nhiều lần thí nghiệm hơn, kết quả trung bình thu được cho thấy hiệu quả của mô hình gần như không có sự khác biệt. Với 5 lần, hiệu quả của 2 mô hình chỉ khác nhau gần 0.2%.