

# Báo cáo

Nhóm gồm 1 thành viên : Phạm Gia Hy | MSSV: 21110311 | Lớp: 21TTH

## Nội dung báo cáo:

1. Nội dung cần thiết đều nằm trong tất cả các file notebook, có thể không cần đọc file này cũng được
2. Trong repo này, ta tiến hành thí nghiệm với 5 model trên tập dữ liệu **icml face data.csv**. Ta apply cả 5 model vào dữ liệu đã qua xử lý PCA lẫn chưa qua xử lý PCA.  
Các model:
  - a. K-nearest Neighbor
  - b. Gaussian Naïve Bayes
  - c. Random Forest
  - d. Multi-layer Perceptron
  - e. Convolutional Neural Network
3. Các kỹ thuật đã được áp dụng trong thí nghiệm này :
  - a. PCA
  - b. Data Augmentation
4. Model (Machine Learning) đạt được hiệu suất cao nhất trên tập dữ liệu gốc :  
**RandomForestClassifier()** : 45.4%  
Điểm chung giữa các model khi áp dụng vào data này: **Class 5 lúc nào cũng đạt hiệu suất tốt nhất**
  - Class đạt hiệu suất tốt nhất : Class 5
  - Class 0 và Class 4 luôn đạt hiệu suất thấp nhất
  - Class 1 tuy có rất ít dữ liệu so với các Class còn lại, nhưng hiệu suất vẫn ổn định khi cho train đủ lâu
  - Precision cao nhất cũng là của class 5 (dùng **RandomForestClassifier()**)  
(với tập dữ liệu đã qua xử lý PCA) : 79%
5. Model (Deep Learning) **CNN** đạt hiệu suất 52% (metric: accuracy)