LESSON 6 : ENSEMBLE LEARNING

**- Deadline : *Giảng viên / Mentor set deadline và điền vào mục này trước khi gửi cho học viên***

**- Phương thức nộp bài : *Nêu rõ học viên sẽ gửi bài tập trực tiếp cho mentors hay gửi vào 1 folder đã được tạo sẵn. Nếu là folder tạo sẵn, vui lòng đính kèm link nộp bài vào đây !***

# Mục đích

*Bài tập này sẽ giúp các bạn củng cố kiến thức đã học, sau khi hoàn thiện các bài tập, các bạn có có thể đạt được :*

**Kỹ năng :**

* Thành thạo việc xây dựng mô hình thuộc nhóm thuật toán Ensemble learning như Random Forest và Adaboost

**Kiến thức :**

Các định nghĩa, khái niệm sẽ được củng cố thông qua phần bài tập này, bao gồm :

* Imbalance Data
* Exploratory Data Analysis
* Ensemble Learning
* Bagging
* Boosting (Adaboosting)

# Bài tập

## **Bài tập** :

#### **Problem Statement:**

Các ngân hàng đều muốn giữ chân khách hàng của mình để duy trì hoạt động kinh doanh và ngân hàng Đa quốc gia ABC cũng muốn điều đó.

Dưới đây là dữ liệu khách hàng của các khách hàng tại Ngân hàng Đa quốc gia ABC có phát sinh giao dịch và mục đích của dữ liệu sẽ là dự đoán Tỷ lệ khách hàng rời bỏ.

Giả sử bạn là Data Analyst cho ngân hàng ABC. BOD đang cố gắng tìm hiểu xem tại sao lại xảy ra vấn đề trên và liệu người dùng các dịch vụ có rời bỏ ABC hay không (hủy sử dụng dịch vụ) trong vài ngày tới.

#### **Objective:**

Bạn được yêu cầu xây dựng một mô hình dự đoán khách hàng sẽ rời bỏ hay tiếp tục sử dụng dịch vụ. Nó sẽ được team Strategy sử dụng để ước lượng số lượng khách hàng rời bỏ và lên các phương án cải thiện. Với kiến thức đã học, bạn hãy xây dựng mô hình **Random Forest** kết hợp với **GridSearchcv** để tìm ra best parameter cho mô hình và **Adaboost** để giải quyết bài toán trên.

#### **Dataset Reference & Description:**

[Lin](https://drive.google.com/file/d/1fFqLRizk_W9WCtRVKusQclwDU_5KaUg7/view?usp=drive_link)k

**HẾT !**