

## Phân tích dữ liệu thông minh.

### Đề án thực hành

# Predict Student Performance from Game Play

## 1. Giới thiệu

Đề án này sẽ dựa trên cuộc thi [Predict Student Performance from Game Play](#) trên Kaggle.

## 2. Yêu cầu của đề án

Các nhóm sẽ tham dự cuộc thi trên. Download bộ dữ liệu, xây dựng mô hình dự đoán, submit lên Kaggle, xem kết quả.

1. Tổ chức thư mục cho đề án: các nhóm có thể dùng một hay nhiều file notebooks nhưng phải được tách biệt rõ ràng cho từng giai đoạn: khám phá dữ liệu, tiền xử lý dữ liệu rút trích đặc trưng, xây dựng mô hình, đánh giá và phân tích kết quả.
2. Nên có nhiều hình vẽ biểu đồ trực quan để giải thích và trình bày trong quá trình làm.
3. Phải có giải thích rõ ràng cho mọi cell code trong file jupyter notebook. Tức là, mỗi cell code nên có một cell markdown kèm theo để giải thích.

## 3. Quy định nộp bài:

Mỗi nhóm cần nộp:

- File jupyter notebook. File này sẽ bao gồm source code và báo cáo. Nội dung cần trình bày bao gồm
  - o Giới thiệu thành viên và phân công công việc
  - o Giới thiệu chung về đề án

- o Quá trình khám phá dữ liệu
  - o Quá trình tiền xử lý dữ liệu
  - o Quá trình tạo đặc trưng (feature engineering)
  - o Xây dựng mô hình (Có thể dùng nhiều mô hình)
  - o Thử nghiệm:
    - Training
    - Testing
    - Kết quả đạt được, nhận xét và phân tích
  - o Kết luận và hướng phát triển. (Cần đưa kết quả trên leaderboard của Kaggle)
- Các mã nguồn python khác (nếu có)
  - Một đoạn video ngắn (15-20p) trong đó nhóm trình bày đồ án của mình.

#### 4. Thang điểm:

- a. Kết quả trên Leaderboard (2đ) – Không tính Perfect score. (Dự kiến):
  - Top 10%: 2đ
  - Top 20%: 1.5đ
  - Top 30%: 1đ
  - Top 40%: 0.5 đ
- b. Nội dung file notebook + code (5đ)
 

Theo yêu cầu như phần 3. Các nhóm cần trình bày rõ ràng, chi tiết về mô hình mình đã dùng để đạt được kết quả như phần 4a.
- c. Nội dung trình bày trong video (3đ)



