

# Đồ án nhóm cuối kỳ

Môn: Khoa Học Dữ Liệu, Lớp: CQ2015/2

## Nội dung đồ án

Vận dụng quy trình Khoa Học Dữ Liệu để tìm câu trả lời cho một câu hỏi (nhóm tự chọn câu hỏi).

Yêu cầu:

- Phải tự thu thập dữ liệu (ví dụ, dùng web API); không được lấy dữ liệu có sẵn.
- Phải dùng Máy Học để mô hình hóa dữ liệu. Như vậy, câu hỏi cần trả lời sẽ có dạng: **output** được tính như nào từ **input**? Chẳng hạn, trong ví dụ dự đoán chất lượng rượu Bordeaux trên lớp, câu hỏi là: **chất lượng rượu Bordeaux (giá rượu)** được tính như thế nào từ **các thông tin của rượu (tuổi, các thông tin về thời tiết của vụ nho dùng để sản xuất rượu)**?
- Làm việc nhóm: tất cả các thành viên trong nhóm đều phải hiểu về đồ án của nhóm.
- Phải ghi nhận rõ ràng và đầy đủ các bước đã thực hiện (bằng jupyter notebook) để người khác có thể kiểm chứng được; sau này, khi mình chạy lại notebook của nhóm bạn thì phải ra được kết quả giống như nhóm bạn đã báo cáo.

## Báo cáo giữa kỳ (30/11/2018)

Đến 30/11/2018, nhóm bạn nên đã thu thập xong dữ liệu và đưa ra câu hỏi cần trả lời (lúc đó có thể đưa ra một vài câu hỏi cũng được, nhưng sau cùng thì phải chốt lại một câu hỏi).

Nhóm bạn chuẩn bị slide (dùng jupyter notebook cũng được) để trình bày. Mỗi nhóm sẽ có khoảng 5 phút để trình bày. Nội dung cần trình bày:

- Câu hỏi đặt ra là gì? Nếu trả lời được câu hỏi thì có lợi ích gì?
- Thu thập dữ liệu
  - Dữ liệu được thu thập như thế nào? Có hợp pháp không?
  - Dữ liệu gồm bao nhiêu dòng, bao nhiêu cột, ý nghĩa của mỗi cột?
  - Dữ liệu có giá trị thiếu không? Nếu có thì nhóm bạn định xử lý như thế nào?

## Báo cáo cuối kỳ (sẽ thông báo sau)