

Đồ án nhóm cuối kỳ

Môn: Khoa Học Dữ Liệu, Lớp: CQ2016

Nội dung đồ án

- Vận dụng qui trình Khoa Học Dữ Liệu để trả lời cho loại câu hỏi dự đoán (nhóm tự chọn câu hỏi); nghĩa là phải dùng Máy Học để mô hình hóa dữ liệu. Chẳng hạn, trong ví dụ dự đoán chất lượng rượu Bordeaux trên lớp, câu hỏi là: **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)**? Một cách nói khác: cho **các thông tin của rượu (tuổi, các thông tin về thời tiết của vụ nho dùng để sản xuất rượu)**, hỏi **giá rượu** tương ứng bằng bao nhiêu?
- Phải tự thu thập dữ liệu (parse HTML, dùng web API); không được lấy dữ liệu có sẵn.
- Làm việc nhóm: tất cả các thành viên trong nhóm đều phải hiểu về đồ án của nhóm. Dùng Git & GitHub để quản lý phiên bản và hợp tác 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 lần 1 (02/12/2019)

Đến buổi học ngày 02/12/2019, 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 file 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? Nguồn gốc của câu hỏi: nhóm tự nghĩ ra hay lấy cảm hứng từ đâu?
- 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 hay có vấn đề gì khác không?

Báo cáo cuối kỳ (sẽ thông báo sau)