

Đồ án cuối kì Khoa học dữ liệu và Ứng dụng

Hoàng Xuân Trường

Ngày 22 tháng 11 năm 2020

1 Mục tiêu

Ôn tập và vận dụng các kiến thức đã học thực hiện một Quy trình Khoa học dữ liệu cho một vấn đề có tính dự đoán.

Cần phải hiểu kỹ và áp dụng từng giai đoạn của một quy trình một cách rõ ràng với câu hỏi được đặt ra.

2 Nội dung đồ án

- Ứng dụng Quy trình Khoa học dữ liệu để trả lời cho loại câu hỏi dự đoán. Các bạn sẽ tự lựa chọn câu hỏi sao cho phù hợp. Ví dụ như dự đoán thời tiết, dự đoán giá của laptop,...
- Phải dùng Máy Học để mô hình hóa dữ liệu.
- Tự thu nhập dữ liệu cho câu hỏi của bạn. Không được lấy dữ liệu có sẵn. Thu nhập bằng cách sử dụng parse HTML, dùng API.
- Các thành viên trong nhóm phải cùng thực hiện chung và hiểu rõ về đồ án của mình. Trình bày rõ ràng công việc của từng người.

2.1 Quy định

Các bạn làm trên một hoặc nhiều file Jupyter notebook sao cho phù hợp với đồ án của bạn. Nên trình bày các file notebook một cách dễ hiểu cho người đọc, như có tiêu đề từng phần, ghi chú phần nào quan trọng, visualize dữ liệu,... Như vậy sẽ dễ trong việc hiểu các bạn đang làm gì và để chấm bài.

Cần có phải có 1 file (README.md) trình bày thông tin của đồ án bao gồm:

1. Câu hỏi được đặt ra là gì?
2. Nếu trả lời được câu hỏi thì có ý nghĩa gì?
3. Cách thức thu nhập dữ liệu như thế nào? (từ đâu, parse HTML hay API, tham khảo từ đâu,...)

4. Tổng quan dữ liệu thu nhập được có bao nhiêu dòng, cột và cột cần dự đoán là gì.
5. Với mỗi cột dữ liệu thu nhập có ý nghĩa gì, kiểu dữ liệu và ví dụ.
6. Tự đánh giá đề án (kết quả, thiếu sót, cần làm gì phát triển thêm,...).
7. Phân công công việc.
8. Hướng dẫn chạy các file notebook (tất cả các quy trình, cả code thu nhập dữ liệu)

Dữ liệu thu nhập được sẽ lưu trong (các) file *.csv*.

2.2 Nộp đề án

Để tất cả các file trong một thư mục đặt tên `<MSSV1>_<MSSV2>` và gửi file `<MSSV1>_<MSSV2>.zip`

Nếu sau khi nén có kích thước lớn không thể nộp trên moodle thì có thể upload drive và gửi link (nhớ mở share link nếu không đề án không được chấm).

Điểm cộng cho các nhóm có câu hỏi hay và không phổ biến (như dự đoán thời tiết,...)

Chú ý: Bài làm giống nhau 0 điểm. Một thành viên nào trong nhóm không hiểu hết code của đề án mình, cả nhóm 0 điểm

3 Liên hệ

Mọi thắc mắc liên hệ qua mail: hxtruong6@gmail.com