

CSC10108 – TRỰC QUAN HÓA DỮ LIỆU

LAB-03: ỨNG DỤNG VÀ CÔNG CỤ

TRỰC QUAN HÓA DỮ LIỆU

(Cập nhật lần cuối: 07/03/2023)

1 THÔNG TIN CHUNG

Loại bài tập	Bài tập thực hành
Thời gian thực hiện	
Hạn nộp bài	
Hình thức thực hiện	Làm việc nhóm
Hình thức nộp bài	Nộp thông qua Moodle FIT
Biên soạn	Nguyễn Thị Thu Hằng (NTT Hằng) Nguyễn Bảo Long (NB Long) Lê Nhựt Nam (LN Nam)
Giáo viên hướng dẫn thực hành	Nguyễn Thị Thu Hằng (NTT Hằng) Nguyễn Bảo Long (NB Long) Lê Nhựt Nam (LN Nam)
Thông tin liên lạc	nam.lnhut@gmail.com (LN Nam) baolongnguyen.mac@gmail.com (NB Long) ntthuhang0131@gmail.com (NTT Hằng)

2 NỘI DUNG

Đây là bài tiếp nối bài lab 02 mà phần trước chúng ta đã thực hiện việc lấy dữ liệu, tìm hiểu một số quan hệ trên đó. Trong lab này, chúng ta sử dụng lại tập dữ liệu và vận dụng một số công cụ như Tableau, Streamlit, Charted, Google Charts, Grafana Labs, Dygraphs, D3.js, RapidMiner, Chart.js, RAW Graphs, Datawrapper, hay Redash để việc trực quan hóa dữ liệu trở nên tốt hơn.

Các yêu chính của lab này:

- Nhóm sinh viên thực hiện viết báo cáo tìm hiểu về công cụ thực hiện:
 - o Giới thiệu về công cụ
 - o Các tính năng hỗ trợ, có minh họa và giới thiệu ngắn gọn. Có thể kèm những minh họa trên dữ liệu ngoài tập dữ liệu thực hiện
- Nhóm sinh viên thực hiện áp dụng công cụ trên tập dữ liệu đã chọn.
- Nhóm sinh viên thực hiện làm video trình bày dashboard, sản phẩm đạt được.

3 NHỮNG GIỚI HẠN

- Bài tập thực hành này được giới hạn trong môi trường công cụ mà nhóm sinh viên lựa chọn thực hiện.
- Một số thư viện như NumPy, Pandas, Seaborn, Matplotlib có thể được sử dụng. Các thư viện khác muốn sử dụng cần phải hỏi ý kiến của giáo viên thực hành.
- **Có thể chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu nhưng không bắt buộc.**

4 HÌNH THỨC NỘP BÀI

Bài tập thực hành được thực hiện theo nhóm. Thời gian và cách thức nộp, xem trên Moodle. Sinh viên tổ chức thư mục nộp bài như sau:

- **Thư mục docs:** chứa báo cáo trình bày trong file .doc/.docx/pdf (Khuyến khích sử dụng định dạng .pdf). Sinh viên trình bày cáo đảm bảo các nội dung như: Thông tin nhóm: tên nhóm, mssv...; Mức độ hoàn thành tổng thể của mỗi yêu cầu; Mức độ hoàn thành của từng thành viên; Chi tiết thuật toán, chạy ví dụ, nhận xét; Khuyến khích trình bày đơn giản, có hình minh họa.

- **Thư mục products:** chứa mã nguồn bao gồm Jupyter notebooks, Python script của nhóm, có kèm hướng dẫn sử dụng (đối với mã nguồn khác Python); dashboard, sản phẩm đã hoàn thành.
- **Thư mục videos:** chứa đường dẫn đến video trình bày của nhóm. Nhóm có thể kênh Youtube để chia sẻ.

5 HÌNH THỨC ĐÁNH GIÁ

Tiêu chí	Tỷ lệ
Tìm hiểu về công cụ	20%
Áp dụng trực quan hóa trên tập dữ liệu được chọn	50%
Rút ra ý nghĩa hợp lý sau mỗi dữ liệu được trực quan.	10%
Báo cáo trình bày bố cục và định dạng hợp lý, rõ ràng.	10%
Có những phân tích, trực quan hóa bằng những biểu đồ mới lạ và rút ra những thông tin hữu ích. Sử dụng mô hình học máy cơ bản.	5%
Tổng	105%

Lưu ý: nếu số biểu diễn trực quan sử dụng công cụ quá ít thì sẽ xem xét giảm tỉ lệ ở mức 2.

6 CÁC QUY ĐỊNH

- Bài không có báo cáo sẽ không chấm.
- Thành viên không tham gia sẽ không có điểm.
- Các nguồn tài liệu tham khảo (nếu có) cần ghi đầy đủ trong báo cáo ở mục Tài liệu tham khảo. **Lưu ý cần phân biệt giữa tham khảo và đạo văn.**
- Đặt tên thư mục bài làm là MSSV1_MSSV2_MSSV03_Lab01, với MSSV là mã số sinh viên, nén toàn bộ bài nộp thành 1 tập tin trước khi nộp. Nếu kích thước >20MB thì upload lên server ngoài như Google Drive, ..., nộp link và giữ link public ít nhất trong 2 năm.
- **Bài giống nhau sẽ 0 điểm môn học.**