



ĐỒ ÁN THỰC HÀNH #1 – NHÓM

CSC12102 - CHUYÊN ĐỀ CHỌN LỌC TRONG HỆ THỐNG THÔNG TIN

1 Thông tin chung

| | |
|----------------------|--|
| Mã số đồ án: | 2122_CDCLHTTT_DA#1 |
| Thời gian thực hiện: | 7 tuần (chi tiết trên moodle) |
| Hình thức thực hiện: | NHÓM |
| Cách nộp bài: | Qua hệ thống Moodle |
| GV phụ trách: | Hồ Thị Hoàng Vy |
| Thông tin liên lạc: | hthvy@fit.hcmus.edu.vn |

2 YÊU CẦU ĐỒ ÁN

- Dataset: <https://grouplens.org/datasets/movielens/100k/>
- Hãy cài đặt hệ tư vấn phim theo kỹ thuật items to items Collaborative Filtering (memory based) sử dụng ngôn ngữ python.
 - Mô tả dataset cách chi tiết.
 - Mô tả bài toán item to item CF với movilen dataset (mô tả ý tưởng thực hiện)
 - Đọc và chuyển dữ liệu vào ma trận
 - Quan sát mô tả dữ liệu
 - Biến đổi, xử lý dữ liệu
 - Giai đoạn offline:
 - tính toán độ tương đồng sử dụng KNN
 - tính toán độ tương đồng sử dụng Jaccard
 - Giai đoạn online:
 - Phát sinh các gợi ý
 - Đánh giá thuật toán:
 - Lựa chọn độ đo đánh giá độ chính xác của dự đoán

3 YÊU CẦU NỘP

- .doc: báo cáo giải thích về dataset, ý nghĩa thuộc tính, ý tưởng thực hiện thuật toán (mã giả), kết quả thực hiện, đánh giá...
- Video: quay lại quá trình trao đổi và thực hiện của mỗi thành viên có mở camera
- Source code: .pybn và link google colab
- Làm bài trên google colab
- Lần 1:
 - Nộp source + báo cáo
 - Source: .iynb (có hiển thị kết quả thực thi của mỗi đoạn lệnh)
 - File data trước và sau khi xử lý làm sạch dữ liệu



- Báo cáo: giải thích, mô tả dataset, các bước làm sạch, biến đổi dữ liệu, ...
- Tính toán độ tương đồng sử dụng Jaccard
- Lần 2: nộp hoàn chỉnh
 - Source + báo cáo
 - Source: gồm đầy đủ các bước cả giai đoạn offline + online, file dữ liệu
 - Báo cáo: phân tích, cài đặt, đánh giá
 - Video giải thích bài làm.