HỆ KHUYẾN NGHỊ

BÀI TẬP THỰC HÀNH TUẦN 2

THUẬT TOÁN LỌC CỘNG TÁC

1. Quy định về việc nộp bài

- Thời gian: Được giảng viên thiết lập trên hệ thống Moodle.
- Hình thức nộp: Trên Moodle.
- Bài nộp bao gồm các file .**ipynb** trong một folder và nén lại thành một tập tin (**.zip**).
- Cách đặt tên: **BTTH2_MSSV.zip**
- Công cụ thực hành. **Google colab**
- Lưu ý: Sai quy định thì sẽ nhận 0 điểm.

2. Nội dung thực hành

2.1. Theo dõi giảng viên hướng dẫn

Xây dựng mô hình khuyến nghị phim dựa trên dữ liệu rating từ người dùng.

- Xây dựng thuật toán lọc cộng tác dựa trên User
  - Hệ số tương quan Pearson
- Xây dựng thuật toán lọc cộng tác dựa trên Item
  - Độ đo tương tự cosine
- Bộ dữ liệu: Movielens ml-latest-small.zip (size: 1 MB)

+ Link download: https://grouplens.org/datasets/movielens/

+ Rating scale: 1 → 5

+ Ratings.csv

	userId	movieId	rating	timestamp
2	1	1	4.0	964982703
3	1	3	4.0	964981247
4	1	6	4.0	964982224
5	1	47	5.0	964983815
6	1	50	5.0	964982931
7	1	70	3.0	964982400
8	1	101	5.0	964980868
9	1	110	4.0	964982176
10	1	151	5.0	964984041
11	1	157	5.0	964984100
12	1	163	5.0	964983650
13	1	216	5.0	964981208

+ Movies.csv

movieId	title	genres
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action|Crime|Thriller
7	Sabrina (1995)	Comedy|Romance
8	Tom and Huck (1995)	Adventure|Children
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action|Adventure|Thriller
11	American President, The (1995)	Comedy|Drama|Romance
12	Dracula: Dead and Loving It (1995)	Comedy|Horror
13	Balto (1995)	Adventure|Animation|Children
14	Nixon (1995)	Drama
15	Cutthroat Island (1995)	Action|Adventure|Romance
16	Casino (1995)	Crime|Drama
17	Sense and Sensibility (1995)	Drama|Romance
18	Four Rooms (1995)	Comedy
19	Ace Ventura: When Nature Calls (1995)	Comedy

+ Users.csv

	movieId	tag	timestamp
2	60756	funny	1445714994
2	60756	Highly quotable	1445714996
2	60756	will ferrell	1445714992
2	89774	Boxing story	1445715207
2	89774	MMA	1445715200
2	89774	Tom Hardy	1445715205
2	106782	drugs	1445715054
2	106782	Leonardo DiCaprio	1445715051
2	106782	Martin Scorsese	1445715056
7	48516	way too long	1169687325
18	431	Al Pacino	1462138765
18	431	gangster	1462138749
18	431	mafia	1462138755
18	1221	Al Pacino	1461699306
18	1221	Mafia	1461699303
18	5995	holocaust	1455735472

2.2.Yêu cầu về nhà.

Xây dựng lại các hệ khuyến nghị phim cho bộ dữ liệu trên

- Yêu cầu 1: Xây dựng thuật toán lọc cộng tác dựa trên User với độ tương tự cosine
  - ➔ Xuất ra top 10 bộ phim (chưa được người dùng xem) có dự đoán rating cao nhất cho UserID =20
- Yêu cầu 2: Xây dựng thuật toán lọc cộng tác dựa trên Item với hệ số tương quan pearson
  - ➔ Xuất ra 10 người sẽ rating cao nhất (nghĩa là những người này chưa rating) cho bộ phim có movieId = 50
- Thử nghiệm với K lân cận 5 và 10 cho cả hai yêu cầu trên và đánh giá so sánh kết quả trả về trong hai trường hợp này: có bao nhiêu kết quả trả về giống nhau, khác nhau, thứ tự trả về của các kết quả giống nhau của hai trường hợp K này như thế nào.
- Sử dụng **Tổng hợp đánh giá dựa trên khoảng cách đánh giá** để dự đoán giá trị rating bị khuyết.