

# 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](https://grouplens.org/datasets/movielens/) (size: 1 MB)

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

+ Rating scale: 1 → 5

+ Ratings.csv

1	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

## DS300 – Hệ khuyến nghị

+ Movies.csv

	A	B	C
1	movieId	title	genres
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
3	2	Jumanji (1995)	Adventure Children Fantasy
4	3	Grumpier Old Men (1995)	Comedy Romance
5	4	Waiting to Exhale (1995)	Comedy Drama Romance
6	5	Father of the Bride Part II (1995)	Comedy
7	6	Heat (1995)	Action Crime Thriller
8	7	Sabrina (1995)	Comedy Romance
9	8	Tom and Huck (1995)	Adventure Children
10	9	Sudden Death (1995)	Action
11	10	GoldenEye (1995)	Action Adventure Thriller
12	11	American President, The (1995)	Comedy Drama Romance
13	12	Dracula: Dead and Loving It (1995)	Comedy Horror
14	13	Balto (1995)	Adventure Animation Children
15	14	Nixon (1995)	Drama
16	15	Cutthroat Island (1995)	Action Adventure Romance
17	16	Casino (1995)	Crime Drama
18	17	Sense and Sensibility (1995)	Drama Romance
19	18	Four Rooms (1995)	Comedy
20	19	Ace Ventura: When Nature Calls (1995)	Comedy

+ Users.csv

	A	B	C	D
	Name Box	movieId	tag	timestamp
2	2	60756	funny	1445714994
3	2	60756	Highly quotable	1445714996
4	2	60756	will ferrell	1445714992
5	2	89774	Boxing story	1445715207
6	2	89774	MMA	1445715200
7	2	89774	Tom Hardy	1445715205
8	2	106782	drugs	1445715054
9	2	106782	Leonardo DiCaprio	1445715051
10	2	106782	Martin Scorsese	1445715056
11	7	48516	way too long	1169687325
12	18	431	Al Pacino	1462138765
13	18	431	gangster	1462138749
14	18	431	mafia	1462138755
15	18	1221	Al Pacino	1461699306
16	18	1221	Mafia	1461699303
17	18	5995	holocaust	1455735472

### 2.2. Yêu cầu về nhà.

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

- 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
- 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