

## Tarea 2

**ENTREGA:** Miércoles 15 de Mayo 2019 hasta las 23:55 hrs vía moodle.

### Instrucciones

- Puede realizar la tarea de forma individual o en grupos de hasta 3 personas.
- Debe adjuntar todo el código fuente utilizado.
- Puede programar en cualquier lenguaje, pero el uso de librerías está restringido (sección 4).
- Si el lenguaje lo permite, se acepta un Jupyter Notebook que incluya todo lo pedido.

### 1 Objetivo

El objetivo es diseñar un sistema de recomendación que funcione en base a alguno de los métodos vistos en clases (User-Based, Item-Based, Slope-One, Cluster-based) y sea capaz de realizar recomendaciones personalizadas para cualquier nuevo usuario.

### 2 Dataset

Para realizar las recomendaciones se trabajará con el dataset de películas MovieLens 10M Dataset (<https://grouplens.org/datasets/movielens/10m/>): 10 millones de ratings, 10.000 películas y 72.000 usuarios.

El archivo ratings.dat contiene ratings provistos por usuarios sobre películas. Cada fila puede ser vista como la preferencia de un usuario sobre una película:

UserID::MovieID::Rating::Timestamp.

	UserID	MovieID	Rating	Timestamp
	1	122	5	838985046
	1	185	5	838983525
	...	...	...	...

Utilice ratings.dat para crear una matriz de ratings.

El archivo movies.dat asocia el título de cada película a un identificador único:

MovieID::Title::Genres.

	MovieID	Title	Genres
	1	Toy Story (1995)	Adventure Animation ...
	2	Jumanji (1995)	Adventure Children ...
	...	...	...

Utilice `movies.dat` para obtener los títulos de las películas que recomiende.

### 3 Recomendador

El sistema debe recomendar películas a un usuario nuevo (que no está en el archivo `ratings.dat`). Por lo tanto, la primera etapa consiste en recolectar información sobre sus preferencias: El usuario debe calificar 10 películas con un rating entre 1 y 5. Las películas deben ser las más populares en el archivo `movies.dat`, es decir estar ordenadas de mayor a menor según cantidad de ratings. Considere que el usuario puede no haber visto alguna de esas películas.

En la segunda etapa el sistema debe recomendar las Top-5 películas que más le gustarían al usuario según sus preferencias (incluya el rating estimado). En ningún caso debe aparecer alguna película calificada en la etapa anterior. A continuación se muestra un ejemplo del sistema:

```
Star Wars: Episode IV - A New Hope (1977):5
The Godfather (1972):2
Titanic (1997):NA
Fargo (1996):3.2
...

RECOMENDACIONES
1 - 12 Monkeys (Twelve Monkeys) (1995) 4.8
2 - Stargate (1994) 4.6
...
5 - Highlander III: The Sorcerer (1994) 3.5
```

### 4 Implementación

La implementación de su algoritmo de recomendación debe ser programada por su cuenta basándose en la materia vista en clases. Solo se permitirán librerías para calcular distancias (similitudes) y realizar clustering. Utilice estructuras de datos adecuadas para el manejo de matrices grandes y dispersas. Está permitido almacenar información y trabajar en memoria secundaria. El sistema debe ser capaz de entregar recomendaciones en un tiempo prudente.