

Curs 2021-22

Programació Avançada
Projecte de l'assignatura

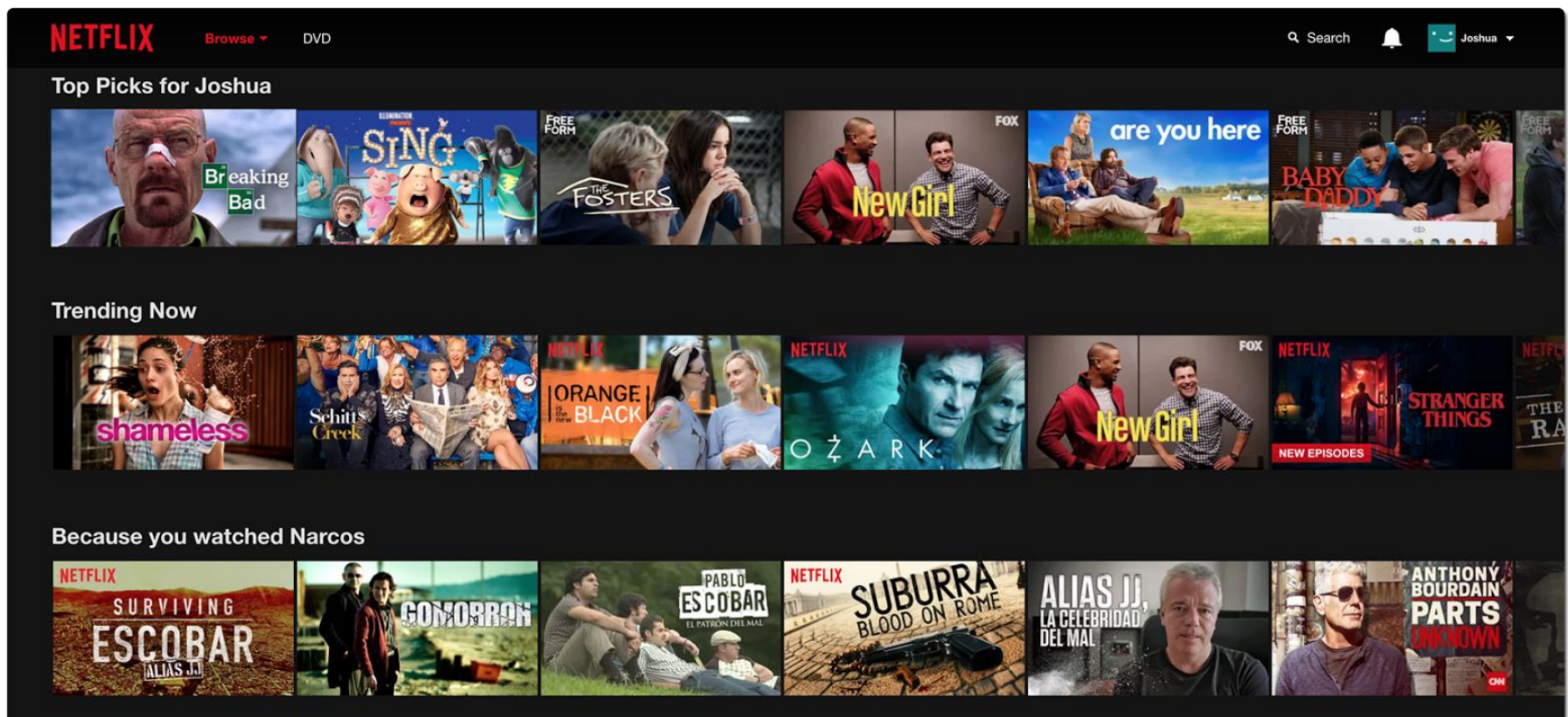
Organització del projecte

- **Grups de 2 persones.** Hem obert a Caronte una inscripció a grups perquè pugueu indicar amb qui fareu el projecte.
- No és obligatori que feu el projecte que us proposem aquí. Podeu fer propostes de **projecte alternatiu**: enviar mail al vostre professor abans del dia **1 d'abril**
- El projecte és part del **treball autònom** de l'assignatura. A les sessions de classe només farem seguiment i avaluació del treball.
- Recordeu: **avaluació del projecte**
$$\text{Nota Projecte} = (0.3 * \text{Entrega Parcial}) + (0.7 * \text{Entrega Final})$$
 - Un lliurament parcial + un lliurament final
 - Nota final projecte ≥ 5
 - Nota Entrega Final projecte ≥ 5

Objectiu del projecte

Sistema de recomanació de contingut

Dissenyar i implementar un sistema de recomanació de nou contingut (pel·lícules, sèries, llibres, cançons, videojocs, etc.) a un usuari a partir de les valoracions passades dels usuaris, del perfil de cada usuari i de les característiques de cada ítem.



Objectiu del projecte

Sistema de recomanació de contingut

Dissenyar i implementar un sistema de recomanació de nou contingut (pel·lícules, sèries, llibres, cançons, videojocs, etc.) a un usuari a partir de les valoracions passades dels usuaris, del perfil de cada usuari i de les característiques de cada ítem.

Taula de valoracions

	ítems				
	i_1	i_2	i_3	...	i_n
u_1	2	0	5	...	3
u_2	4	0	0	...	1
u_3	3	4	0	...	0
.
.
.
u_n	5	2	1	...	0

Característiques dels ítems

pel·lícula -> descripció, gènere, director, actors, ...

llibre -> gènere, autor, resum, idioma, ...

Perfil dels usuaris

usuari -> edat, població, professió, ...

Valoració per part de l'usuari i de l'ítem j. Un 0 indica que no l'ha valorat



Usuari -> Llista de ítems recomanats

Sistemes de recomanació

Sistemes de recomanació simples (*Top popular items*):

- Recomanen els ítems més populars en funció d'una puntuació global per cada ítem (per exemple, la mitja de les valoracions de tots els usuaris).
- Donen la mateixa recomanació per tots els usuaris. No tenen en compte ni el perfil dels usuaris ni les seves preferències passades.

Sistemes de recomanació col·laboratiu (*other users also like...*):

- Recomanen en funció de les preferències o valoracions d'altres usuaris que tenen preferències passades similars.
- Si un altre usuari ha valorat de forma similar els mateixos ítems que nosaltres, és probable que ens agradin altres ítems que aquest usuari ha valorat positivament.

Sistemes de recomanació basats en contingut (*because you liked...*):

- Recomanen en funció de les preferències passades de l'usuari i també de les característiques dels ítems o del perfil de l'usuari (per exemple, si sempre valorem positivament pel·lícules d'un determinat gènere, el sistema tendirà a recomanar-nos aquest tipus de pel·lícules)
- Necessiten informació sobre les característiques dels ítems i/o usuaris apart de la taula de valoracions passades.

Datasets

En el projecte haureu d'aplicar el vostre sistema de recomanació com a mínim a dos conjunts de dades diferents: *MovieLens* i un altre que vosaltres escolliu.

MovieLens data set

- Conté valoracions de pel·lícules: 100k valoracions de 9.000 pel·lícules fetes per 600 usuaris

<https://grouplens.org/datasets/movielens/latest/>

Altres conjunts de dades

<https://www.kaggle.com/>

The screenshot shows the Kaggle search results for 'recommender systems'. The page is filtered to show 21 results, sorted by date. The left sidebar contains filters for Date (Last 90 days: 3, Last week: 1), Viewed By You (Viewed: 1, Not Viewed: 20), and Dataset Size (medium: 10, small: 10). The main content area displays two datasets:

- Recommender Systems Data** by Naidu Suraj Vardhan, 5 days ago, 36 MB, 0 upvotes. It is a Dataset.
- 9000+ Movies Dataset** by Doula Isham Rashik Hasan, 9 days ago, 2 MB, 0 upvotes. It is a Dataset.

Objectiu del projecte

Característiques del sistema de recomanació

- S'ha de per poder aplicar com a mínim a dos tipus de conjunts de dades diferents: *MovieLens* i un altre que escolliu vosaltres.
- S'hauran d'implementar diferents mètodes (com a mínim un de cada tipus) per obtenir la llista de recomanacions per un usuari.
- Haurà d'incloure un sistema per avaluar i comparar els diferents mètodes.
- Podreu afegir altres funcionalitats extra a la versió bàsica. Totes les extensions del sistema bàsic es tindran en compte positivament a l'avaluació final del projecte.
- La implementació del projecte s'haurà de fer utilitzant classes i herència i fent servir les tècniques i principis de disseny orientat a objecte que anirem explicant al llarg del curs.
- Per implementar algunes de les tècniques de recomanació, farem servir algunes funcionalitats de llibreries en Python, que ja us anirem indicant a mesura que avancem amb el projecte.

Planificació del projecte

Primera versió

L'objectiu és implementar una versió bàsica d'un sistema de recomanació que pugui funcionar amb diferents conjunts de dades i implementar diferents tècniques de recomanació:

- Haurà de permetre treballar amb dos conjunts de dades diferents.
- Haurà d'incloure dos mètodes de recomanació:
 - Un sistema de recomanació simple
 - Un sistema col·laboratiu a partir de les interaccions entre usuaris
- Podrem escollir quan executem el sistema amb quin conjunt de dades volem treballar i quin tipus de sistema volem utilitzar.

Segona versió

L'objectiu serà ampliar el sistema per:

- Incloure altres mètodes de recomanació
- Incloure un mòdul per avaluar i comparar els resultats dels diferents mètodes
- Visualitzar els resultats

Organització del projecte

Properes tasques a realitzar

- Diagrama UML amb el disseny de classes necessari per poder fer la primera versió del projecte:
 - Per cada classe, atributs (indicant el seu tipus) i mètodes (indicant tipus dels paràmetres i del valor de retorn).
 - Relacions d'herència i de composició entre les diferents classes.
- **Data de lliurament:** dimecres **20 d'abril**.
- **Avaluació:** comptarà un **25%** de la nota del **lliurament parcial**

Lliurament de la primera versió del projecte: dimecres **4 de maig**.

Properes tasques a realitzar

Diagrama de classes. Alguns consells de disseny

- Eina per fer diagrames UML: <http://www.draw.io/>
- Penseu com podeu utilitzar l'herència i el polimorfisme per guardar la informació dels diferents conjunts de dades a partir d'una mateixa classe base.
 - Penseu quines dades (atributs) o mètodes poden ser comuns a tots els conjunts de dades i poseu tot el que sigui comú a una classe base.
 - Poseu tot el que és específic de cada tipus de dades a les classes derivades.
- Penseu com podeu utilitzar l'herència i el polimorfisme per implementar els diferents mètodes de recomanació.
 - Penseu quina pot ser la interfície pública comuna a tots els sistemes de recomanació que implementeu.
 - Penseu quina ha de ser la implementació específica per cada sistema de recomanació.
- Penseu bé els atributs i mètodes de cada classe, separant entre el que és privat i la interfície pública. Procureu que cada classe i cada mètode tinguin responsabilitats i funcions clarament definides. Penseu com es relacionen els mètodes d'una classe amb els de les altres.
- Si us ajuda a completar el disseny, comenceu a implementar el codi d'alguna de les classes o dels seus mètodes.

Tipus de sistemes de recomanació

Sistema de recomanació simple

- Es basa en calcular per cada ítem una valoració global basada en les puntuacions de tots els usuaris i ordenar els ítems encara no puntuats per l'usuari en funció d'aquesta valoració.
- Una manera simple d'obtenir la valoració d'un ítem seria calcular la mitja de totes les valoracions fetes per tots els usuaris. Però aquest mètode simple té algun inconvenient ja que no considera el nº de puntuacions que ha tingut cada ítem.
 - Per exemple una pel·lícula que hagi estat valorada només per 5 usuaris amb una valoració mitja de 4,5 tindria una valoració més alta que una altra pel·lícula que hagi estat valorada per 200 usuaris amb una valoració mitja de 4,4
 - Per tant, introduïrem un factor corrector per tenir en compte el nº de puntuacions que ha rebut cada ítem

Tipus de sistemes de recomanació

Sistema de recomanació simple

1. Fixarem un n^o mínim de puntuacions *min_vots* que ha d'haver rebut un ítem per considerar-lo. Descartarem tot els ítems que hagin rebut menys puntuacions que *min_vots*.
2. Per calcular la valoració de cada ítem, apart del n^o mínim de puntuacions *min_vots*, tindrem en compte 3 valors més:
 - *avg_item*: valoració mitja que li han donat els usuaris a l'ítem (considerant només els vots rebuts, descartem valoracions amb un 0).
 - *num_vots*: n^o d'usuaris que han puntuat aquet ítem.
 - *avg_global*: valoració mitja de tots els ítems considerats.
3. A partir d'aquesta valors obtindrem la valoració de cada ítem aplicant la fórmula següent:

$$score = \left(\frac{num_vots}{num_vots + min_vots} \cdot avg_item \right) + \left(\frac{min_vots}{num_vots + min_vots} \cdot avg_global \right)$$

D'aquesta forma, els ítems que tinguin un n^o de vots més petit veuran disminuït el pes de la seva pròpia valoració que es compensarà agafant com a referència la valoració global de tots els ítems.

Tipus de sistemes de recomanació

Sistema de recomanació simple

Exemple

Volem predir les valoracions pel nou usuari de les pel·lícules que no ha valorat (P1 i P6) a partir de la taula de valoracions usuaris – pel·lícules.

1. Fixem $min_vots = 3$. Descartem P3
2. Calculem els valors necessaris

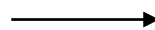
	P1	P2	P4	P5	P6
avg_item	3,75	5	3,25	3,5	4
num_vots	4	3	4	4	3

$$avg_global = 3,83$$

3. Calculem *score* per les pel·lícules a valorar (P1 i P6)

$$Score_{p1} = \left(\frac{4}{7} \cdot 3,75\right) + \left(\frac{3}{7} \cdot 3,83\right) = 3,78$$

$$Score_{p6} = \left(\frac{3}{6} \cdot 4\right) + \left(\frac{3}{6} \cdot 3,83\right) = 3,91$$



Recomanació: P6, P1

		Pel·lícules					
Usuaris		P1	P2	P3	P4	P5	P6
	U1	7	6	7	4	5	4
	U2	6	7	0	4	3	4
	U3	1	2	0	3	3	4
	U4	1	0	1	2	3	0

Nou Usuari

	P1	P2	P3	P4	P5	P6
Usuari	0	3	3	1	1	0

Tipus de sistemes de recomanació

Sistema de recomanació col·laboratiu

- Es basa en trobar usuaris que hagin donat puntuacions similars als mateixos ítems que l'usuari pel que volem recomanar nous ítems.
- Les recomanacions a l'usuari es basaran en les puntuacions que hagin donat els usuaris més similars als ítems que el nou usuari encara no hagi valorat.
- Seguirem aquests passos:
 1. Calculem una mesura de similitud entre l'usuari pel que volem recomanar i tots els altres usuaris a partir de les valoracions que han fet de cada ítem.
 2. Ordenem tots els usuaris a partir d'aquesta mesura de similitud i seleccionem els k usuaris més similars.
 3. Assignem a cada nou ítem una puntuació basada en la mitja ponderada de les valoracions que li han donat els k usuaris més similars

Tipus de sistemes de recomanació

Sistema de recomanació col·laboratiu

1. Calculem una mesura de similitud entre l'usuari pel que volem recomanar i tots els altres usuaris a partir de les valoracions que han fet de cada ítem.

Mesura de similitud entre 2 usuaris: distància del cosinus

$$s(u, v) = \frac{\sum_{i \in I} p_{ui} \cdot p_{vi}}{\sqrt{\sum_{i \in I} p_{ui}^2} \sqrt{\sum_{i \in I} p_{vi}^2}}$$

u, v : usuaris que volem comparar

I : conjunt de tots els ítems

p_{ui} : puntuació que l'usuari u dona a l'ítem i

Només tenim en compte aquells ítems que els dos usuaris han valorat (puntuació diferent de 0)

		Pel·lícules					
Usuaris		P1	P2	P3	P4	P5	P6
	U1	7	6	7	4	5	4
	U2	6	7	0	4	3	4
	U3	1	2	2	3	3	4
	U4	1	0	1	2	3	3

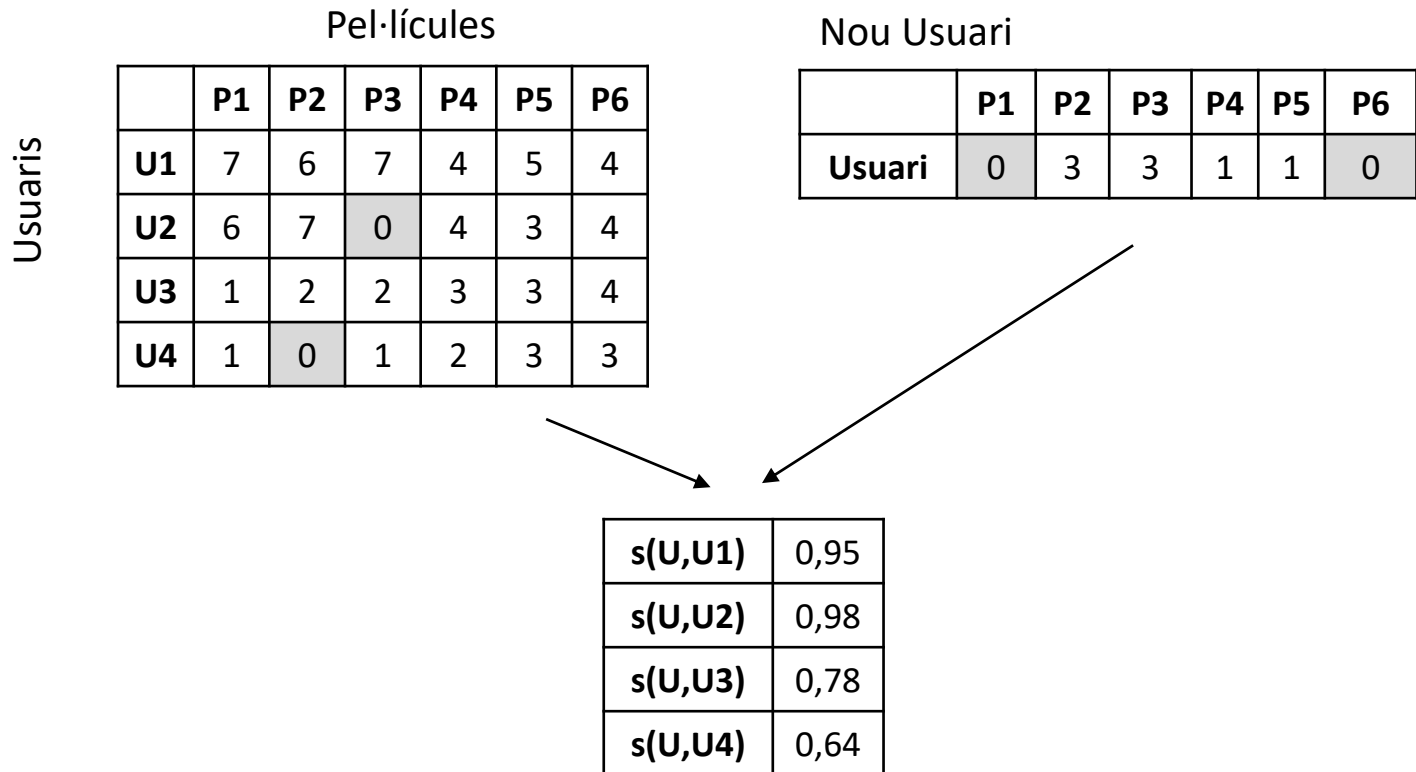
Nou Usuari						
	P1	P2	P3	P4	P5	P6
Usuari	0	3	3	1	1	0

$$s(\text{usuari}, u_2) = \frac{7 \cdot 3 + 4 \cdot 1 + 3 \cdot 1}{\sqrt{7^2 + 4^2 + 3^2} \sqrt{3^2 + 1^2 + 1^2}} = 0,98$$

Tipus de sistemes de recomanació

Sistema de recomanació col·laboratiu

2. Ordenem tots els usuaris a partir d'aquesta mesura de similitud i seleccionem els k usuaris més similars.



Si ordenem per similitud i fixem $k=2$, ens quedarem amb U2 i U1

Tipus de sistemes de recomanació

Sistema de recomanació col·laboratiu

3. Assignem a cada nou ítem una puntuació basada en la mitja ponderada de les valoracions que li han donat els k usuaris més similars

Càlcul de la puntuació final de cada ítem

$$p_{ui} = \mu_u + \frac{\sum_{v \in U} s(u, v) \cdot (p_{vi} - \mu_v)}{\sum_{v \in U} |s(u, v)|}$$

u : usuari pel que volem recomanar nous ítems

i : cadascun dels nous ítems pels que volem predir una valoració

p_{ui} : puntuació que l'usuari u dona a l'ítem i .

$s(u, v)$: similitud entre dos usuaris calculada segons s'ha explicat abans

μ_u : mitja de totes les puntuacions que ha donat un usuari. S'utilitza per normalitzar els valors de les puntuacions entre usuaris diferents (un usuari pot puntuar de forma global més alt o més baix que els altres).

	P1	P2	P3	P4	P5	P6	μ_v	$s(u, v)$
U2	6	7	0	4	3	4	4,8	0,98
U1	7	6	7	4	5	4	5,5	0,95

	P1	P2	P3	P4	P5	P6	μ_u
Usuari	0	3	3	1	1	0	2.0

$$score_{p1} = 2,0 + \frac{0,98 \cdot (6 - 4,8) + 0,95 \cdot (7 - 5,5)}{0,98 + 0,95} = 3,56$$

$$score_{p6} = 2,0 + \frac{0,98 \cdot (4 - 4,8) + 0,95 \cdot (4 - 5,5)}{0,98 + 0,95} = 0,98$$

→ Recomanació: P1, P6

Lectura dels fitxers de dades

Normalment les dades estaran en format **csv (comma separated values)**

Exemple: MovieLens dataset

ratings.csv

```
userId,movieId,rating,timestamp
1,1,4.0,964982703
1,3,4.0,964981247
1,6,4.0,964982224
1,47,5.0,964983815
```

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
```

Lectura de fitxers en format csv

```
import csv
```

```
with open('ratings.csv', 'r', encoding='utf8') as csv_file:
```

```
    csvreader = csv.reader(csv_file) → Crea un objecte per llegir un fitxer csv
```

```
    fields = next(csvreader) → Llegeix una línia del fitxer csv i guarda els valors en una
```

```
    for row in csvreader:
```

```
        ...
```

llista. En aquest cas,
fields = ['userId', 'movieId', 'rating',
'timestamp']

Itera per totes les línies del fitxer csv. Per cada línia, guarda els valors en una llista.

Per exemple, a la primera iteració:

```
row = ['1', '1', '4.0', '964982703']
```