

## Enunciat del projecte de PROP Quadrimestre de tardor, curs 21/22

### Sistema Recomanador

Un sistema de recomanació és una eina que estableix un conjunt de criteris sobre les dades dels usuaris per a realitzar prediccions sobre recomanacions d'ítems que puguin ser d'utilitat o valor per a un usuari concret. Els ítems poden ser qualsevol cosa, com ara pel·lícules, cançons, llibres, restaurants, o qualsevol tipus de producte. Aquestes recomanacions normalment varien en funció de cada usuari, i existeixen diferents tipus de sistemes de recomanació en funció de les necessitats del sistema.

La idea és implementar un sistema recomanador d'ítems per a un usuari, anomenat l'usuari actiu, donats un conjunt d'usuaris i un conjunt d'ítems del mateix tipus que han sigut comprats/usats/seleccionats/valorats, etc. pels usuaris. Els ítems s'han de poder identificar de forma genèrica per un identificador més una sèrie d'atributs que poden variar segons el tipus d'ítem concret. Com a mínim han de poder haver-hi atributs numèrics, booleans o categòrics (ex: blau, verd, vermell).

Per a la recomanació s'hauran d'implementar, como a mínim, els mètodes més utilitzats, que són els següents:

- *Collaborative filtering*: són estratègies que es basen en recomanar a l'usuari actiu els ítems que han agradat a usuaris semblants a ell. Hi ha moltes variants i algorismes. En el projecte haureu d'implementar l'estratègia basada en els algorismes "*k-means*" + "*Slope one*" que us detallarem més endavant.
- *Content-based filtering*: són estratègies que es basen en recomanar a l'usuari actiu els ítems semblants als que li han agradat a ell. També hi ha moltes variants i tècniques. En el projecte haureu d'implementar una estratègia basada en l'algorisme "*k-nearest neighbours (k-NN)*", que us detallarem més endavant.
- *Hybrid approaches*: són estratègies que combinen tècniques basades en les dues aproximacions anteriors. La combinació pot ser de diferent maneres.

Tots 3 mètodes es basen en el concepte de similitud (concepte clau utilitzat en *case-based reasoning*). Els conjunts de dades seran molt simples (estructurats), i vindran en fitxers de text. Cada línia del fitxer s'anomenarà registre, i contindrà els valors dels atributs i/o les valoracions dels ítems per part dels usuaris. Us proporcionarem exemples de *data sets*.

El programa ha d'oferir un entorn tant còmode com sigui possible per al necessari preprocés de les dades. Obligatòriament, el programa ha de:

- Permetre la definició de tipus d'ítems a partir dels seus atributs
- Permetre la gestió d'usuaris
- Permetre la gestió d'ítems
- Permetre la valoració d'ítems per part d'un usuari
- Definir funcions de distància o similitud depenent dels atributs utilitzats. Depenent d'aquestes funcions, es podria requerir discretitzar/binaritzar els valors dels atributs.
- S'ha de donar l'opció de que l'usuari pugui guardar dades preprocessades per fer més eficient el procés en treballar de nou amb el mateix conjunt de dades.
- Recomana un conjunt d'ítems a un usuari a partir d'un conjunt de dades
- Avaluació d'un conjunt de recomanacions (és a dir, calcular la seva qualitat)
- Guardar i recuperar el resultat de l'algorisme principal (recomanacions)

I opcionalment:

- Es podria incloure qualsevol selecció manual de les dades que s'usaran per calcular les recomanacions, com ara eliminació de registres escollits (p.ex., *outliers* o redundants), eliminació d'atributs (que podrien ser, p.ex., poc rellevants), afegir atributs nous (p.ex., calculats o derivats dels altres), etc.

**Funcionalitats principals a entregar al primer lliurament:**

1. Recomanació d'ítems per un usuari utilitzant *collaborative filtering* i *content-based filtering*
2. Avaluació d'un conjunt de recomanacions

**Reducció de funcionalitat als equips de 3 persones:**

1. No és imprescindible implementar el mètode híbrid
2. A la gestió d'ítems, només és obligatòria la càrrega d'arxius
3. Al primer lliurament només cal presentar un dels 2 mètodes

**Dates dels lliuraments:**

- Primer: divendres 19 de novembre de 2021
- Segon: dimecres 15 de desembre de 2021
- Tercer: dimecres 22 de desembre de 2021 (lliuraments interactius: a partir del 10 de gener de 2022)