

## TME 9 - Systèmes de recommandation

### Recommandation et le module Surprise

Le module surprise est un module spécifiquement dédié aux systèmes de recommandation qui suit les conventions de `scikit-learn`.

L'objectif de ce TME est de prendre en main et de comparer différents algorithmes basiques de recommandation en utilisant ce module. Vous utiliserez le jeu de données MovieLens (le small, le jeu de données nettoyé et restreint aux films/utilisateurs ayant plus de 10 votes est fourni en complément du sujet).

Pour charger les données, utilisez `surprise.Dataset.load_from_file("movielenssmall.csv", surprise.Reader())`. L'autre fichier contient les titres des films.

Dans un premier temps, faites une analyse sommaire des données en traçant les histogramme des notes, du nombre d'avis par film/par utilisateurs, des notes moyennes des films/utilisateurs. Visualisez également la matrice de distance pair à pair entre utilisateurs avec une distance euclidienne et une distance cosinus. Calculez la sparsité du jeu de données.

Dans un deuxième temps, divisez votre jeu de données en apprentissage/test (en retirant au hasard un pourcentage de ratings de la matrice de notes). Testez et comparez au moins les algorithmes suivants (en lisant la documentation pour comprendre la différence, et sur plusieurs paramétrages) :

- BaselineOnly
- KNNBasic
- KNNWithMeans
- SVD
- NMF

Pour les méthodes issues de factorisation de matrices, vous pouvez également tracer en 2D les représentations latentes des films (en utilisant un algorithme de réduction de dimension type t-SNE - par exemple par le biais de `tensorboard`).

Il peut être intéressant également pour ces algorithmes de faire un clustering des représentations latentes des films et faire une interprétation des clusters ainsi trouvés.