

Projet Statistique L3 SID

Description: Movielens est un système de recommandations personnalisées, non commercial, développé par l'université du Minnesota (<https://movielens.org>). Ce site web met à disposition un certain nombre de leurs données. Le jeu de données de ce projet en fait partie. L'objectif de ce projet est de tirer un maximum d'informations de ces données.

Pistes:

- Profil utilisateur :
 - Les utilisateurs notent-ils toujours les films qu'ils aiment? Ou prennent-ils le temps de noter les autres également?
 - Comment évoluent les notes dans le temps? Les premières notes d'un utilisateur sont-elles toujours bonnes puis ensuite plus diverses?
 - Les utilisateurs affichent-ils une préférence particulière pour un type de films (comédie romantique, biopic, science fiction, action, guerre, ...)
- Profil utilisateur-utilisateur :
 - Peut-on regrouper les utilisateurs par préférence?
 - Au sein d'un même groupe, les utilisateurs sont-ils toujours en accord ?
 - Quelles sont les données démographiques au sein d'un groupe ? (ex: Est ce que les étudiants entre 18 et 25 ans aiment des films similaires?)
- Films :
 - Des films ressortent-ils du lot ? Dans le bon et le mauvais sens?
 - Au sein d'un même type de film, peut-on détecter des groupes de films ?
 - Pour les films sortis au moment de la récolte de données, un effet vieillissement est-il détectable?

Caractéristiques du jeu de données:

- 100 000 notes de 0.5 à 5
 - 1682 films
 - 943 utilisateurs
 - Chaque utilisateur a au moins noté 20 films différents.
- Données récoltées sur la période du 19 septembre 1997 au 22 avril 1998
- Données de référence dans la communauté système de recommandation (recherche)

Chaine de traitement:

- Insertion en base des données (Compte Oracle) ? Pertinent/ Non pertinent?
- Traitement statistique (R)

Organisation/ Evaluation:

- Groupe de 2-3 personnes
- 4 séances de TP/Travail personnel
- Choix d'un aspect d'exploration (Utilisateur / Utilisateur-Utilisateur / Films)
- Les points ne sont que des pistes, n'hésitez pas à avoir vos propres idées
- Si question, n'hésitez pas à me joindre louedec@irit.fr
- Rapport (taille libre)
- Notation
 - Pertinence des choix effectués
 - Clarté du rendu (Rédaction du rapport, explications claires, graphiques correctement mis en forme, ...)