

BI - TP1 - Prise en main de RapidMiner et Frequent Item Sets

Ludovic Denoyer

Pentaho

Prise en main de RapidMiner

Nous allons travailler sur le dataset movieLens 1m disponible sur le Web.

0.1 Importation et nettoyage des données

- Importez les données issues des fichiers ratings.dat
- Sélectionner les lignes dont le rating est égal à 5
- Calculer l'ensemble des films qui ont au moins 1000 ratings
- En déduire un nouveau jeu de données qui ne conserve que les lignes dont le rating est égal à 5 pour les films fréquents (opérateur 'Intersect')

Outils utiles : Filter Examples, Select Attributes, Aggregate, Set Role, Intersect, Write CSV

0.2 Extraction de règles d'association

- Extraire des règles d'association pertinentes

Outils utiles : Join, Select Attributes, Pivot, Replace Missing Values, Numerical to Binomial, Set Role, FP-Growth, Create Association Rules

Qui aime Star-wars 1 ?

Répondre à cette question en utilisant un classifieur linéaire.....