

1 Objectifs du TP

Cette séance de travaux pratiques propose un court tutoriel sur une application possible du classifieur bayésien naïf. Il s'agit de construire un système de recommandations de films similaire à celui proposé par Netflix.

Le TP peut s'effectuer sur Google Colab, sous la forme d'un notebook Python.

2 Recommandations de films avec MovieLens

Le dataset utilisé est « MovieLens latest-small », fourni par GroupLens. Il contient 100 836 notes, 3 683 tags, 9 742 films pour 610 utilisateurs ; les données couvrent des évaluations créées entre 1996 et 2018.

On souhaite construire un petit système de recommandation de films de type Netflix. À partir des notes d'utilisateurs MovieLens, on transforme le problème en tâche de classification binaire :

- 1°. Charger le dataset MovieLens.
- 2°. Explorer les données : nombre de notes, distribution des ratings, genres les plus fréquents.
- 3°. Construire une variable cible binaire Y définie par $Y = 1 \iff$ l'utilisateur aime le film :

$$Y = \mathbb{1}_{[\text{rating} \geq 4]}. \quad (1)$$

- 4°. Transformer les genres en variables binaires.
- 5°. Entraîner un classifieur binaire de type Naive Bayes.
- 6°. Évaluer le modèle à partir de courbes donnant le taux d'erreur, la matrice de confusion, des courbes ROC, éventuellement des courbes de type précision-rappel.
- 7°. Utiliser le modèle pour recommander des films à un utilisateur donné.