

Les Simpsons

Prédire le score IMDB d'un épisode

IMDB est un site internet qui sauvegarde diverses informations sur des films, des séries, etc. Entre autres, vous pourrez y retrouver des résumés, des avis d'utilisateurs et, surtout, la note que ce site décerne à chacun des titres qui s'y trouvent ([cliquez ici pour un exemple](#)).

Le jeu de données se situe sur la plateforme Connected¹ ainsi qu'un notebook vierge.

L'objectif de cet exercice sera de prédire cette note en fonction de deux scénarios et de trouver lequel d'entre les deux est le plus pertinent. Chaque scénario montrera quelles seront les « **features** » à employer. Lors de l'évaluation de chaque modèle, vous calculerez leur MSE afin de pouvoir les comparer.

Cet entraînement vous permet de construire des modèles et de les comparer.

Ligne de conduite à appliquer

- Vous utiliserez la technique du « KFold cross validation » en effectuant 5 « splits ». Les données devront être mélangées en faisant appel à un « random_state » dont la valeur sera fixée à 777.
- Vous utiliserez la méthode « cross_val_score » pour calculer les différents MSE rapidement.
- Vous afficherez la moyenne des MSE des deux scénarios.
- Vous sélectionnerez le meilleur scénario en expliquant la raison de votre choix.

Scénarios

Premier scénario :

- Nombre de voisins : 10
- Variables explicatives : « us_viewers_in_millions », « views », « imdb_votes ».

Deuxième scénario :

- Nombre de voisins : 15
- Variables explicatives : « us_viewers_in_millions », « imdb_votes ».

¹ Source : <https://www.kaggle.com/wcukierski/the-simpsons-by-the-data>