

Mon système de recommandation

Introduction

Créé en 1998, “Netflix est un service de diffusion en streaming qui permet à ses membres de regarder une grande variété de séries TV, films, documentaires, etc. sur des milliers d'appareils connectés à Internet.” => [**Wikipédia Netflix**](https://fr.wikipedia.org/wiki/Netflix)

Aujourd’hui Netflix pèse plus de **20 milliards de dollars de chiffre d'affaires** et consomme **12,6% de la bande passante internet mondial.**

Lorsqu’on accède au service Netflix, le système de recommandations aide l’utilisateur à trouver aussi facilement que possible les séries TV ou films qu’il pourrait apprécier, grâce à un système de recommandation. Netflix calcule ainsi la probabilité que l’utilisateur regarde un titre donné du catalogue de Netflix, et peut ainsi optimiser ces partenariats ou plus globalement sa stratégie marketing. Netflix est l'archétype de la société data driven.

# Objectif

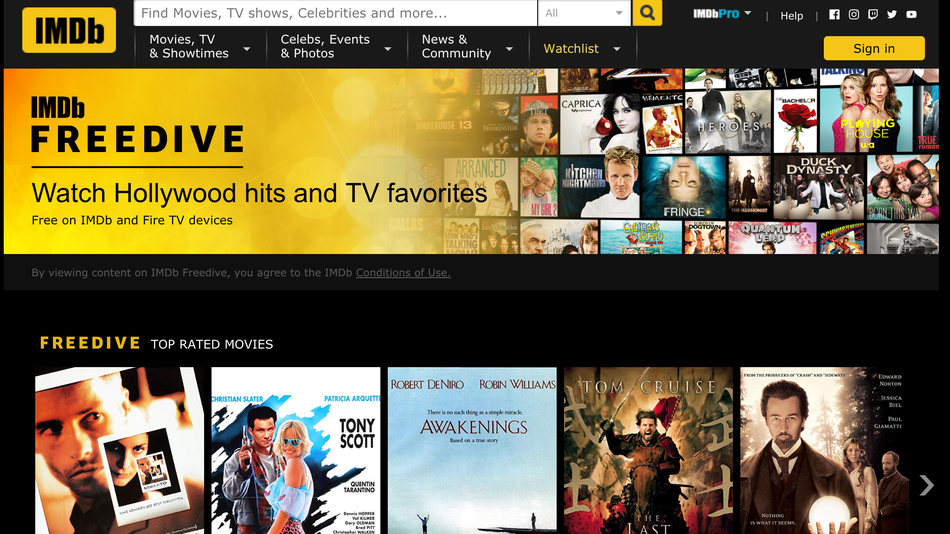
Vous êtes un Data analyst freelance. Un cinéma en perte de vitesse situé dans la Creuse vous contacte. Il a décidé de passer le cap du digital en créant un site internet taillé pour les locaux.



Pour aller encore plus loin, il vous demande de **créer un moteur de recommandations de films** qui à terme, enverra des notifications aux clients via SMS ou mails.

Pour l’instant, aucun client n’a renseigné ses préférences, **vous êtes dans une situation de cold start**. Mais heureusement, le client vous donne une base de données de films basée sur la plateforme IMDb.

Voici le lien pour télécharger le jeu de données : [Données projet Mon système de recommandation](https://grouplens.org/datasets/movielens/latest/). Ce lien propose 2 datasets de tailles différentes, libre à vous de choisir celui qui vous convient. Il est préférable de commencer par le petit dataset, afin de le prendre en main.



Apparemment il y a eu des erreurs lors de l’extraction des données. Le nettoyage prenant du temps, le client vous propose de travailler en collaboration avec d’autres Data analysts pour avoir un dataset bien propre. Après avoir exploré ce jeu de données, **vous concevrez un système de recommandations capable de retourner 5 films susceptible de plaire à un utilisateur.** Vous devrez donc choisir des métriques pertinentes pour trouver le système de recommandation.

Le client aurait souhaité l’intégrer à son site pour pouvoir le tester, mais le timing est trop serré. Force de proposition, vous lui proposer de **le rendre testable au moyen d’un invite de commande (input).** Il aura accès à un menu qui lui permettra :

* D’obtenir quelques statistiques sur les films (type, durée), acteurs (nombre de films, type de films) et d’autres. Vous le ferez notamment à l’aide de visualisations.
* De retourner une liste de films recommandés en fonction d'IDs de films qui ont été visionnés

Vous êtes assez libre sur ce projet. À vous d’allouer un temps suffisant pour le nettoyage, l’exploration et la création de l’invite de commande. L’objectif n’est pas d’arriver à un travail parfait, mais que le système fonctionne et que vous arrivez à déceler les points à améliorer.

Vous expliquerez votre démarche et les difficultés rencontrés lors d’une **présentation orale de 10 à 15 minutes (à préciser avec le client)**, **et** ferez une **démonstration** devant des représentants des clients. Votre client vous donnera quelques noms de films pendant la présentation, vous effectuerez des recommandations de films basés sur ces films. Les autres spectateurs pourront aussi proposer un film.

# Missions

* Créer un système de recommandation de film en utilisant des algorithmes de machine learning
* Faire une présentation pour présenter votre travail, expliquer votre démarche et proposer des pistes d’amélioration
* Faire une démonstration de l’invite de commande

# 

# Livrables attendus

* Un notebook contenant le nettoyage et exploration, accompagné de visualisations. Vous expliquerez vos choix de nettoyage et vos conclusions d’exploration dans un document de votre choix.
* Un notebook pour l’étape **Système de recommandation** avec le code source et des commentaires

Ressources

* <https://blogs.gartner.com/martin-kihn/how-to-build-a-recommender-system-in-python/>
* <https://linuxfr.org/news/predire-la-note-d-un-journal-sur-linuxfr-org>
* <https://www.mediego.com/fr/blog/netflix-success-story-basee-sur-algorithmes-de-recommandation/>
* <https://www.podcastscience.fm/dossiers/2012/04/25/les-algorithmes-de-recommandation/>
* Petits tips sur tu as lu jusqu'ici: des informations complétementaires sur chaque films -> <https://datasets.imdbws.com/>