



# Application de recommandation de films

Kamel, AlexisG, Indira, Tristan



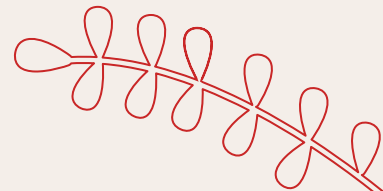
**METHFLIX**



01

Contexte

01010101  
01010101



# Objectifs

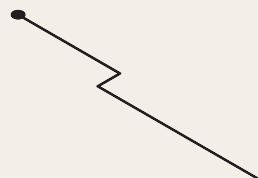


## Concurrencer Netflix

Proposer une recommandation de film adaptée à l'utilisateur

## Application

Proposer à l'utilisateur une interface ergonomique



# Fonctionnement de recommandation



## Content-based

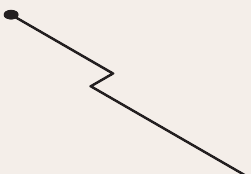

Données portant uniquement sur le contenu.  
Recommandation basée sur la proximité entre les films

## Collaborative filtering

Données portant sur la consommation des utilisateurs.  
On recommande à un utilisateur des films aimés par un utilisateur aux mêmes goûts

## Hybrid

Données portant sur le contenu ET les habitudes de consommation des utilisateurs.  
Mix des deux systèmes précédents



# Planning

## Jour 1

Veilles, découverte des algorithmes de recommandation, début de l'EDA.

## Jour 2

Suite et fin de l'EDA, premier modèle.

## Jour 3

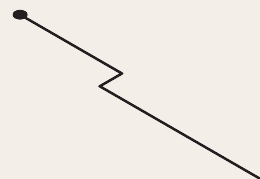
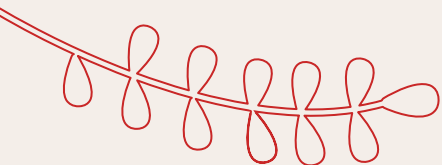
Modèle TF-IDF Vectorizer, Dernier test modèle.

## Jour 4

Création de l'API et déploiement

## Jour 5

Création du streamlit et déploiement

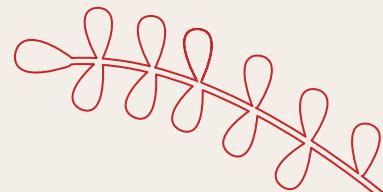




02

Nettoyage

01010101  
01010101



# Les données

## 28 features

Diverses informations sur le film  
(réalisateur, acteurs, genres, mots-clés  
...)

Données quantitatives **et** qualitatives.

Donc divers traitements

## ~5000 observations

Une observation par film

Valeurs nulles et doublons

# Nettoyage



## ● Doublons

*100 doublons retirés de la base de données*

## ● Valeurs nulles

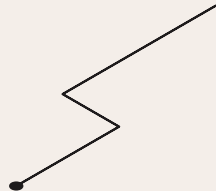
*Beaucoup de NaN dans plusieurs colonnes, suppression de ces colonnes  
Suppression des lignes avec des NaN pour le reste*

## ● Données textuelles

*Propres, probablement pas des données de saisie*

## ● Données pécuniaires

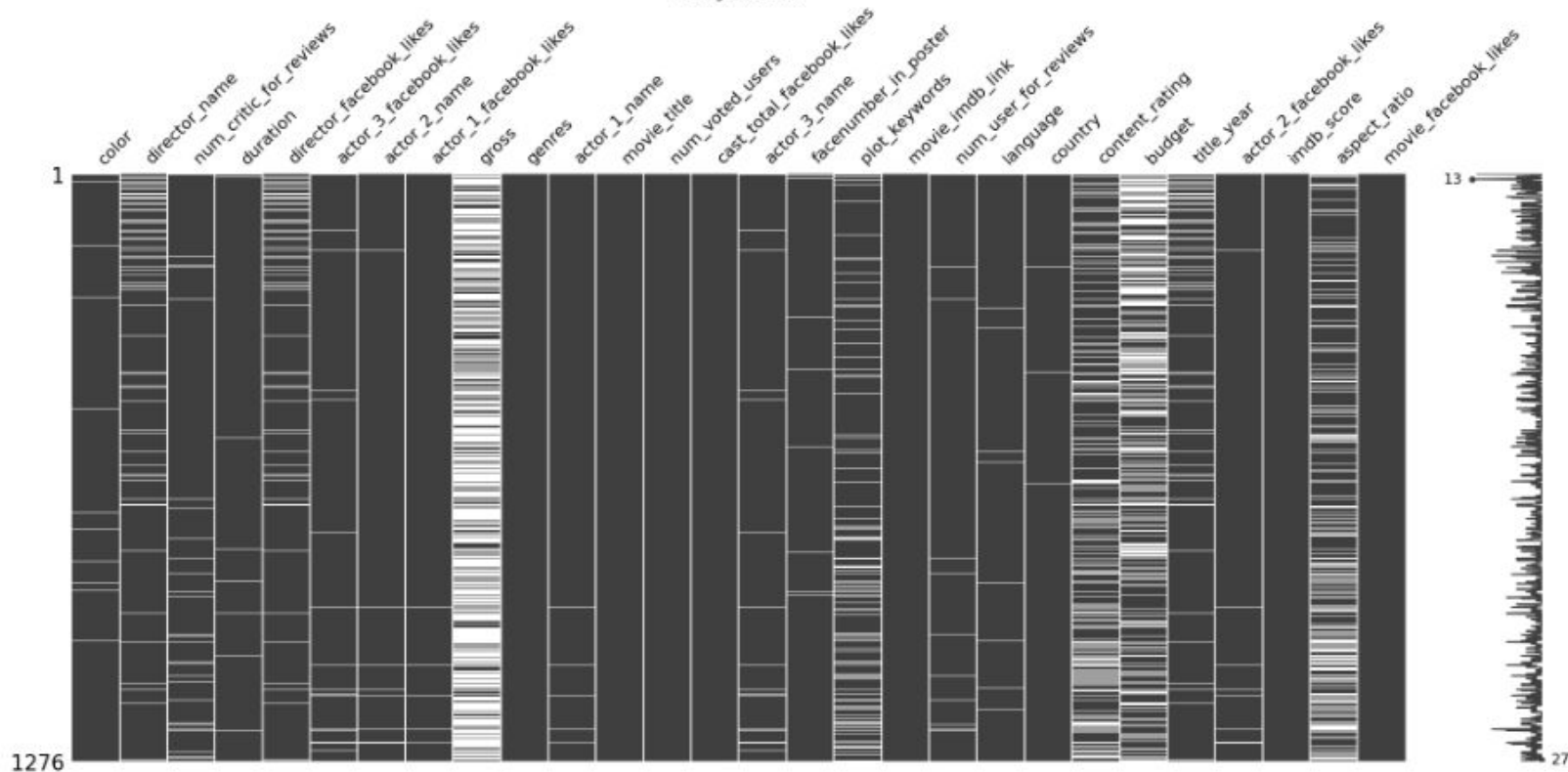
*Ne prend pas en compte l'inflation, donc plutôt inutiles en l'état, donne une fausse information.*



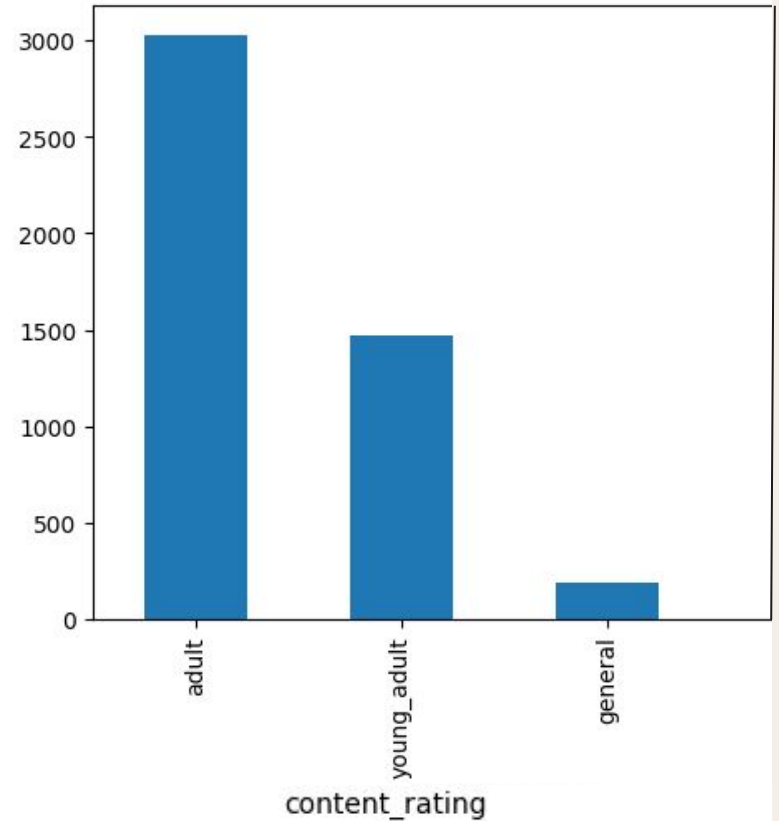
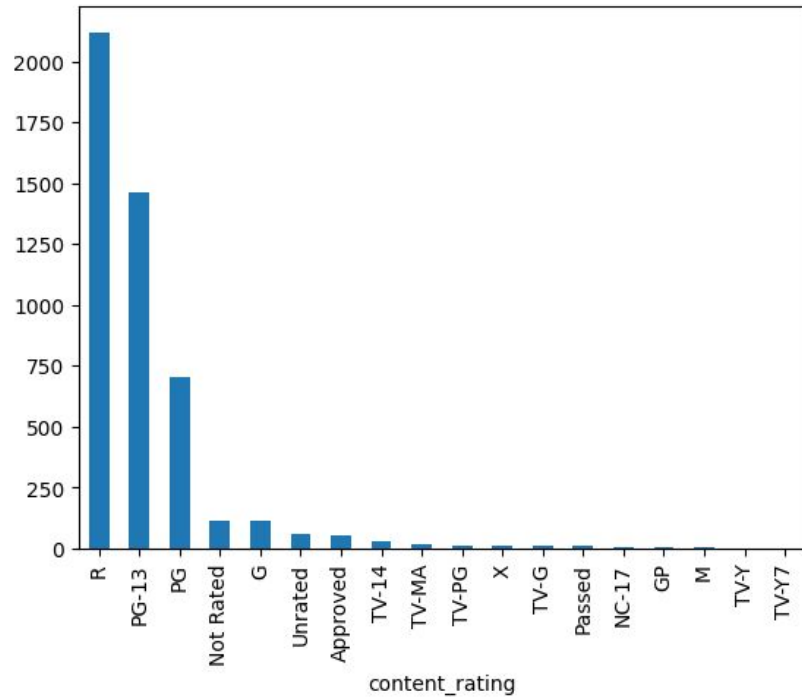


# Données manquante

Missing Data Matrix



# Content rating

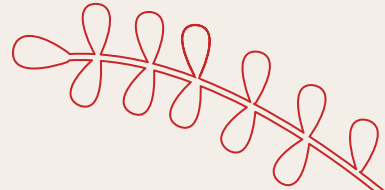




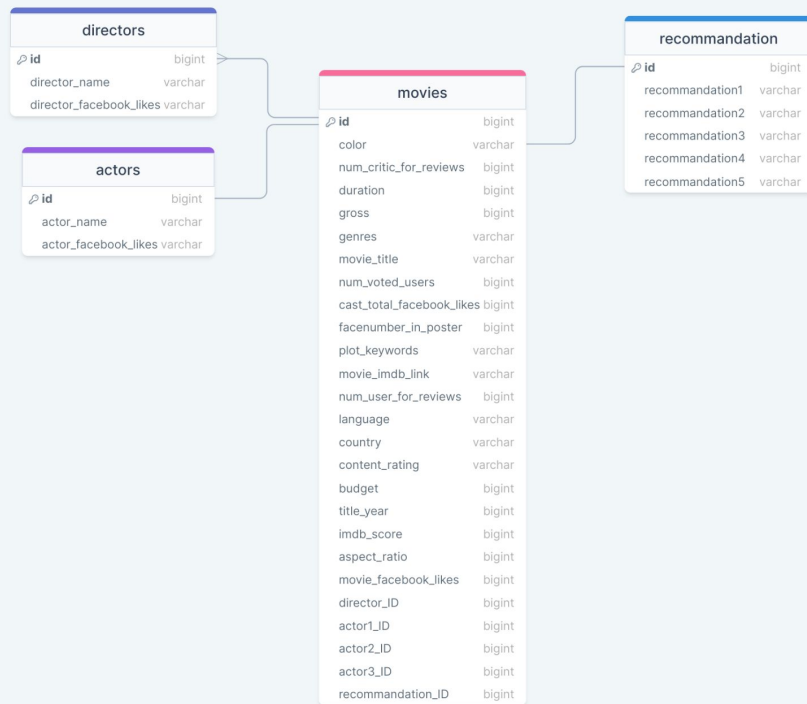
03

**BDD**

01010101  
01010101



# Architecture



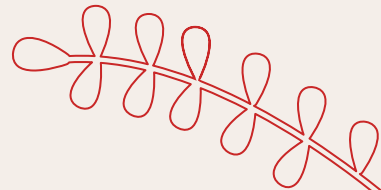


04

Prédiction



01010101  
01010101

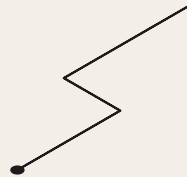


# Recommandation

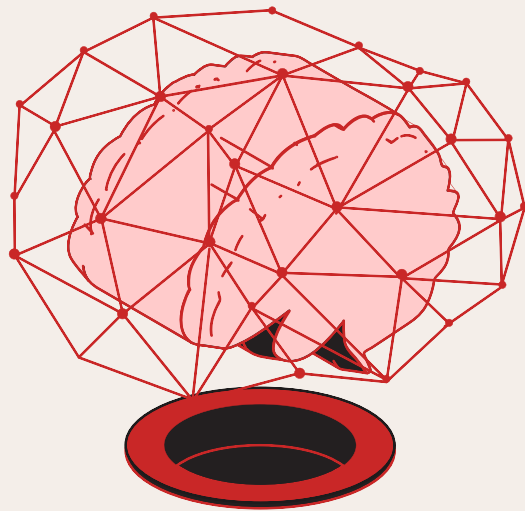


**TF-IDF vectorizer** : quantifie l'importance des termes en prenant en compte leur rareté, et les transforme en vecteurs

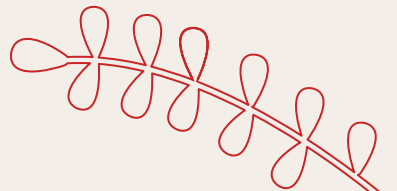
**Cosine similarity** : Mesure utilisée pour déterminer la similarité entre des vecteurs



# 05 Demo



01010101  
01010101

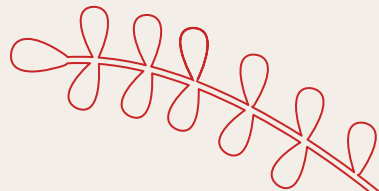




05

# Perspectives d'Amélioration

01010101  
01010101





# Perspectives



## Données consommation utilisateurs

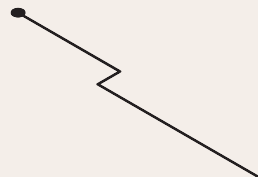
Permettrait de passer sur un système de recommandation de Collaborative Filtering, puis Hybrid

## Valeurs nulles

Web scraping pour combler les valeurs nulles souvent présentes sur le site Imdb)

## Reinforcement Learning

Pour continuer à améliorer les performances



# Merci !

Des questions ?

