# Choix du sujet et du modèle

Le projet ‘Movie recommandation’ a été réalisé dans le cadre de la formation MLOPS. Notre groupe a entrepris de créer un moteur de recommandation de films qui recommanderait des films ayant de fortes chances d'être appréciés par l'utilisateur.

Supposons que quelqu'un ait regardé un film « beta » et qu'il l'ait aimé ! Que peut lui suggérer mon système de recommandation pour le prochain film ?

Il est possible de suggérer différents films sur la base de la similarité du contenu du film sélectionné, comme les genres, les noms des acteurs et de l'équipe, les mots-clés et toute autre métadonnée du film. Dans ce cas, il faudra utiliser un filtrage article-contenu.

Ou alors, il est également possible de comparer les métadonnées de l'utilisateur, telles que l'âge et le sexe, à celles des autres utilisateurs et suggérer à l'utilisateur des articles que des utilisateurs similaires ont aimés. Dans ce cas, il faudra utiliser un filtrage utilisateur-contenu.

L'ensemble de données movie-lens utilisé ici ne contient pas de données sur le contenu utilisateur. Dans un premier temps, nous allons donc construire un filtre élément-contenu (ici un film-contenu).

Il existe plusieurs façons de traiter ce sujet :

* la méthode basée sur la mémoire
* la méthode de filtrage collaboratif basé sur la mémoire
* la méthode de filtrage collaboratif basé sur un modèle

Dans les méthodes basées sur la mémoire, nous ne disposons pas d'un modèle qui apprend des données pour prédire, mais nous formons plutôt une matrice pré-calculée de similarités qui peut être prédictive. Les ensembles de données qui peuvent etre utilisé pour le filtrage du contenu d'un article sont movies.csv et tags.csv. Dans ce cas, nous pouvons collecter tous les tags donnés à chaque film par différents utilisateurs, ajouter les mots-clés du genre du film et former un cadre de données final avec une colonne de métadonnées pour chaque film.

Outre les métadonnées des films, nous disposons d'une autre source d'information précieuse : les données d'évaluation des utilisateurs. Notre système de recommandation peut recommander un film similaire à "Beta sur la base des évaluations des utilisateurs. En d'autres termes, quels autres films ont reçu des évaluations similaires de la part d'autres utilisateurs ? Il s'agit d'un exemple de filtrage collaboratif item-item. Vous en avez peut-être entendu parler sous la forme "Les utilisateurs qui ont aimé cet article ont également aimé ces autres articles". L'ensemble de données d'intérêt serait ratings.csv et nous le manipulons pour former des éléments en tant que vecteurs de notes d'entrée par les utilisateurs. Comme il y a beaucoup de votes manquants de la part des utilisateurs, nous avons imputé Nan(s) par 0, ce qui suffira pour notre filtrage collaboratif.

# Définition des métriques et exigences de performances

# Schéma d’implémentation



## Datasets

The following main data sources were used for this project.

* MovieLens
* MovieLens Tag Genome 🡺 no
* IMDb 🡺 no

## Tech Stack

Each part of the tech stack for this project was utilized to varying degrees by the different team member.

* MySQL - primary database
* IBM SPPS Modeler - analytics software for generating models
* Python - data pre-processing and data analytics implementation
* Amazon Web Services - cloud computing in conjunction with MySQL
* Weka - data mining tool for pre-processing data

# Récupération de nouvelles données