

# Catalogue de Données

## Introduction

Le catalogue de données est un élément essentiel de la gouvernance des données. Il fournit une vue détaillée et organisée de l'ensemble des données utilisées dans le projet de recommandation de films en temps réel. Ce catalogue assure la traçabilité complète des sources de données, des définitions de variables et des métadonnées associées, permettant ainsi une compréhension approfondie des données utilisées dans le projet.

## 1. Sources de Données

### 1.1 Données MovieLens

- Type de Données : Données de films, évaluations des utilisateurs
- Source : MovieLens
- Fréquence de Mise à Jour : Mensuelle
- Format : Fichiers CSV

### 1.2 Données Utilisateurs

- Type de Données : Informations sur les utilisateurs
- Source : Saisies par les utilisateurs
- Fréquence de Mise à Jour : À la demande
- Format : JSON

### 1.3 Données de Recommandations

- Type de Données : Résultats des recommandations
- Source : Modèle ALS
- Fréquence de Mise à Jour : Dynamique (à chaque interaction utilisateur)
- Format : JSON

## 2. Définitions de Variables

### 2.1 Données MovieLens

### Données Principales

1. movieId : Identifiant unique du film
2. title : Titre du film
3. release\_date : Date de sortie du film
4. genres : Genres du film
5. rating : Évaluation attribuée par l'utilisateur
6. timestamp : Horodatage de l'évaluation

### Données Additionnelles

7. IMDb\_URL : URL IMDb du film
8. genre\_count : Nombre de genres associés au film
9. release\_year : Année de sortie du film
10. genre : Liste des genres associés au film

### 2.2 Données Utilisateurs

1. user\_id : Identifiant unique de l'utilisateur
2. age : Âge de l'utilisateur
3. gender : Genre de l'utilisateur
4. occupation : Profession de l'utilisateur
5. zip\_code : Code postal de l'utilisateur
6. age\_group : Groupe d'âge dérivé

### 2.3 Données de Recommandations

1. user\_id : Identifiant unique de l'utilisateur
2. movie\_id : Identifiant unique du film recommandé

## 3. Métadonnées Associées

### 3.1 Données MovieLens

- Origine des Données : MovieLens
- Dernière Mise à Jour : [Date]

- Responsable des Données : [Nom]
- Description : Ensemble de données de films et d'évaluations d'utilisateurs provenant de MovieLens.

### 3.2 Données Utilisateurs

- Origine des Données : Saisies par les utilisateurs
- Dernière Mise à Jour : [Date]
- Responsable des Données : [Nom]
- Description : Informations détaillées sur les utilisateurs du système de recommandation.

### 3.3 Données de Recommandations

- Origine des Données : Modèle ALS
- Dernière Mise à Jour : Dynamique
- Responsable des Données : Système de Recommandation
- Description : Résultats des recommandations générées par le modèle ALS.

## Conclusion

Ce catalogue de données servira de référence centralisée, assurant ainsi une compréhension approfondie des données utilisées dans le projet de recommandation de films en temps réel. Il sera maintenu régulièrement pour refléter les évolutions et les mises à jour des données, renforçant ainsi la gouvernance et la transparence dans la gestion des informations.