



# Présentation – Système de Recommandation MovieLens

## 1. Introduction

- Le cinéma génère un très grand volume de contenu et les spectateurs sont souvent submergés par le choix.
  - Les systèmes de recommandation aident à proposer des films adaptés aux goûts de chaque utilisateur.
  - Notre projet utilise le jeu de données **MovieLens 100K** (943 utilisateurs, 1 682 films, 100 000 notes) pour générer des recommandations personnalisées.
- 

## 2. Objectif du projet

- Fournir les **5 meilleures recommandations de films** pour un utilisateur donné, en se basant sur ses évaluations passées.
  - Améliorer la satisfaction et la rétention des spectateurs.
  - Démontrer l'efficacité du filtrage collaboratif comme approche centrale.
- 

## 3. Compréhension des données

- **ratings.csv** → évaluations des films (notes de 0.5 à 5).
  - **movies.csv** → titres et genres des films.
  - Analyse exploratoire → genres les mieux notés : *Film-Noir*, *War*, *Documentary*.
- 

## 4. Modèle – Comment ça fonctionne ?

- Approche **filtrage collaboratif** via **SVD (Singular Value Decomposition)**.
  - Mise en place avec la librairie **Surprise**.
  - Le modèle apprend des interactions *utilisateur–film* sans connaissance du contenu.
  - Génère une estimation de note pour chaque film non vu par un utilisateur.
- 

## 5. Évaluation – Résultats obtenus

- **Métriques utilisées** : RMSE & MAE.
- **Résultat final** : RMSE  $\approx$  **0.95** → erreur inférieure à une étoile sur l'échelle 0.5–5.
- Exemple de recommandations pour l'utilisateur 200 :
  - 🎬 *Jumanji (1995)* – Adventure|Fantasy → préd. 3.50
  - 🎬 *Heat (1995)* – Action|Thriller → préd. 3.49

---

## 6. Limites

- **Cold Start** : difficile pour nouveaux films/utilisateurs.
- **Biais de popularité** : les films populaires dominent les suggestions.
- **Pas de contexte** : pas d'information temporelle ou démographique.

---

## 7. Recommandations et prochaines étapes

- Intégrer un **filtrage basé sur le contenu** (genres, acteurs, réalisateurs).
- Utiliser des métriques avancées comme **Precision@k**, **Recall@k**.
- Enrichir les données (interactions implicites : clics, visionnages).
- Tester sur un dataset plus grand (MovieLens 20M) pour vérifier la scalabilité.

---

## 8. Conclusion

- Le système de recommandation atteint une performance correcte ( $RMSE \approx 0.95$ ).
- Les prédictions sont globalement proches des notes réelles.
- Le modèle est fonctionnel et peut être amélioré par des approches hybrides.