

# **Rapport de Projet IA**

## **Recommandation d'oeuvre cinématographique**

## I. Introduction

Les systèmes de recommandation occupent une place essentielle dans divers domaines tels que le cinéma, l'e-commerce ou les plateformes de streaming, où ils facilitent la découverte de contenus adaptés aux préférences des utilisateurs. Dans ce projet, nous avons développé un système de recommandation de films en exploitant des techniques avancées d'apprentissage automatique, notamment les embeddings textuels pour analyser et représenter les résumés de films sous forme vectorielle, et le réseau de neurones UNet pour améliorer la qualité des données.

En parallèle, nous avons conçu un système de filtres incluant des critères tels que les genres, les collections et les décennies, permettant de personnaliser les résultats pour chaque utilisateur. L'objectif principal était de fournir des recommandations précises et pertinentes, tout en explorant des méthodes modernes de traitement des données et d'apprentissage automatique.

Pour rendre ces résultats accessibles, nous avons également développé un site web interactif et intuitif. Bien que cela ne fût pas une exigence initiale, cette plateforme a permis de démontrer efficacement les capacités du système et d'offrir une expérience utilisateur fluide et claire.

## II. Embeddings textuels

Les embeddings textuels sont des représentations vectorielles permettant de capturer la sémantique des textes. Pour ce projet, chaque résumé de film a été transformé en un vecteur dense à l'aide du modèle pré-entraîné *paraphrase-MiniLM-L6-v2*. Ce modèle, fourni par SentenceTransformers, produit des vecteurs de taille compacte (384 dimensions), facilitant la comparaison des résumés sans nécessiter de grandes ressources en calcul ou en mémoire.

Ces embeddings ont servi à mesurer la similarité entre les résumés des films et ainsi générer des recommandations pertinentes. Le choix de ce modèle a été suggéré par notre professeur, et aucune évaluation comparative avec d'autres modèles ou configurations n'a été réalisée, car ce choix s'est révélé satisfaisant pour répondre à nos besoins.

## III. Gestion des données

La gestion des données constituait une étape critique du projet. Nous avons travaillé avec deux fichiers CSV principaux :

- **main\_with\_compressed\_embeddings\_unet.csv** : contenant les embeddings des films.

- **movies\_metadata.csv** : regroupant les métadonnées (titres, genres, années de sortie, collections, etc.).

### A. Nettoyage des données

Le nettoyage des données a été nécessaire pour traiter des problèmes tels que :

- Les doublons.
- Les valeurs corrompues.
- Les champs incomplets, notamment pour certains genres ou collections ou encore résumés.

Ces étapes ont permis d'assurer la cohérence des données et de garantir des recommandations fiables.

### B. Préparation des filtres

À partir des données nettoyées, nous avons conçu des filtres spécifiques :

- **Genres** : Permet de sélectionner les genres de films préférés.
- **Collections** : Permet de regrouper les films appartenant à une même saga ou thématique.
- **Décennies** : Permet de cibler les films d'une période précise.

Ces filtres offrent une personnalisation accrue pour répondre aux préférences des utilisateurs.

## IV. Réseau de neurones UNet

UNet, un modèle de réseau de neurones convolutionnels initialement conçu pour la segmentation d'images, a été utilisé ici pour traiter les vecteurs unidimensionnels des films (métadonnées et embeddings).

### A. Structure du modèle

- **Phase d'encodage** : Les données passent par une série d'opérations convolutives (*Conv1d*), qui extraient progressivement les caractéristiques tout en réduisant leur dimension.
- **Phase centrale** : Les données encodées sont compressées dans une représentation compacte.
- **Phase de décodage** : Les convolutions transposées (*ConvTranspose1d*) reconstruisent progressivement les données tout en intégrant des connexions sautées (*skip connections*) pour préserver les informations essentielles.

### B. Défis techniques

L'entraînement du modèle UNet a nécessité environ 10 000 itérations, ce qui aurait été long et peu pratique sur des machines locales. Pour résoudre ce problème, nous avons utilisé Google Colab, qui offre un accès à des GPU puissants, réduisant considérablement le temps de calcul.

## V. Développement du site web

En complément du système de recommandation, un site web a été développé pour permettre aux utilisateurs de tester et de visualiser facilement les résultats.

### A. Technologies utilisées

- **Backend** : FastAPI, qui permet de créer une API REST performante pour gérer les données et les calculs de similarité.
- **Frontend** : Une interface simple et esthétique conçue pour offrir une navigation fluide et intuitive.

### B. Fonctionnalités principales

- Recherche de films via une barre intuitive.
- Application de filtres personnalisés (genres, collections, décennies).
- Présentation des recommandations basées sur les similarités cosinus calculées à partir des embeddings.

Bien que la création du site ne fût pas obligatoire, elle a permis de mettre en valeur le travail réalisé et d'en faciliter la démonstration.

## VI. Conclusion

Ce projet a combiné plusieurs techniques avancées pour développer un système de recommandation performant et intuitif. Les embeddings textuels, le réseau UNet et la gestion efficace des données ont permis d'obtenir des résultats précis et adaptés aux préférences des utilisateurs.

L'ajout d'un site web a enrichi le projet en rendant les recommandations accessibles et visuellement agréables. À l'avenir, des améliorations pourraient inclure l'intégration de modèles plus récents et l'ajout de fonctionnalités en temps réel pour une personnalisation encore plus poussée.