

Introduction

Problème

Trop de films disponibles → difficile de choisir (paradoxe du choix).

Objectif

Construire un moteur de recommandation qui répond à :

- ▶ Est-ce un bon film? (*Qualité*)
- ▶ Est-ce similaire à mes goûts? (*Sémantique*)
- ▶ Est-ce dans mon groupe de films? (*Clustering*)

Approche

Système hybride combinant plusieurs modèles de recommandation.

Méthodologie

Données

- ▶ Plus de 10 000 films

Prétraitement

- ▶ Nettoyage des données

Modèles

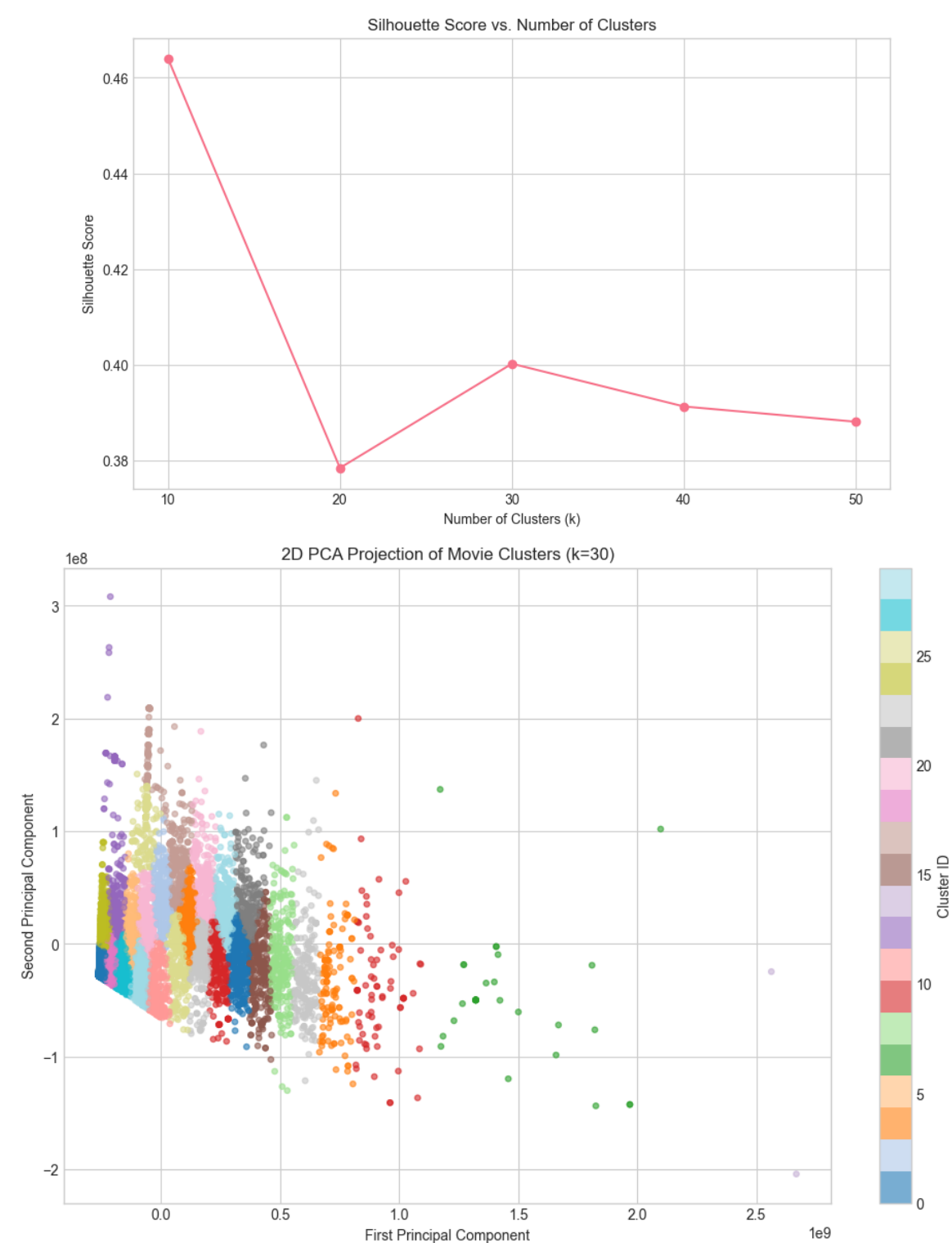
- ▶ Similarité sémantique
- ▶ Clustering de films
- ▶ Prédiction de la qualité

Fusion

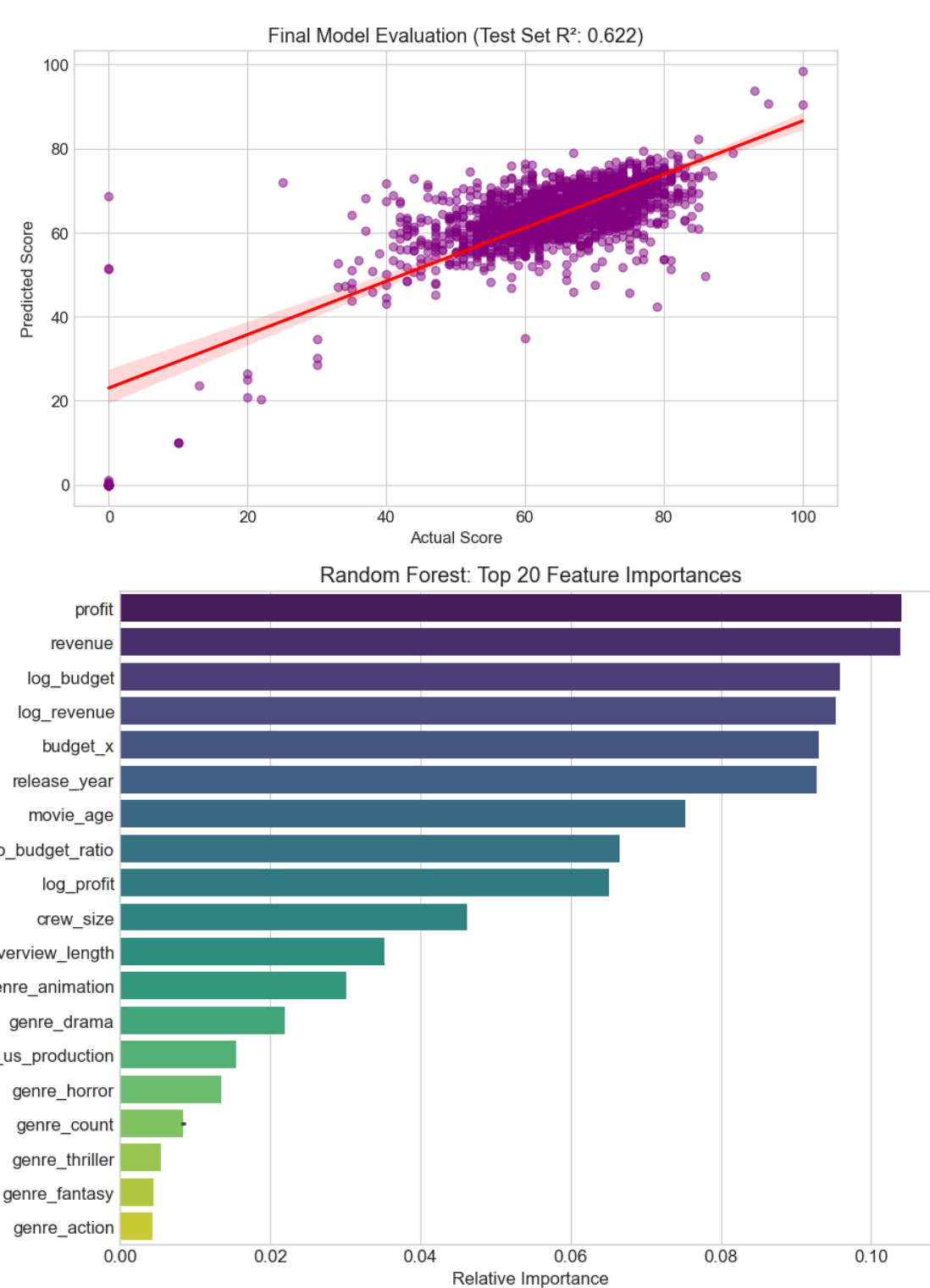
- ▶ Combinaison pondérée des scores

Modèle Non-Supervisé
(Clustering)

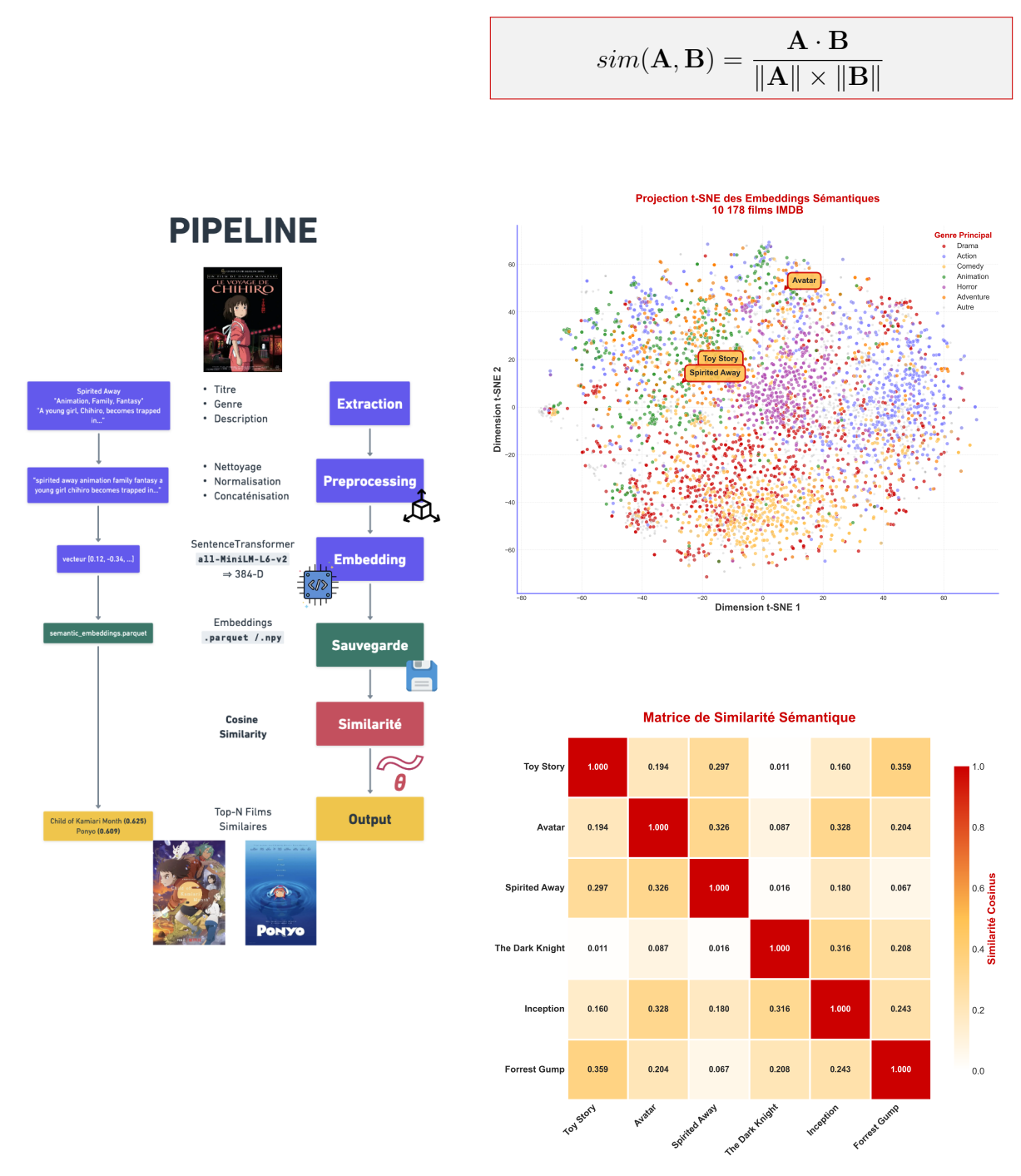
- ▶ Projection des films en espace numérique
- ▶ Regroupement par similarité statistique
- ▶ Sélection des clusters appréciés
- ▶ Calcul d'un score de recommandation

Modèle Supervisé
(Prédiction de Qualité)

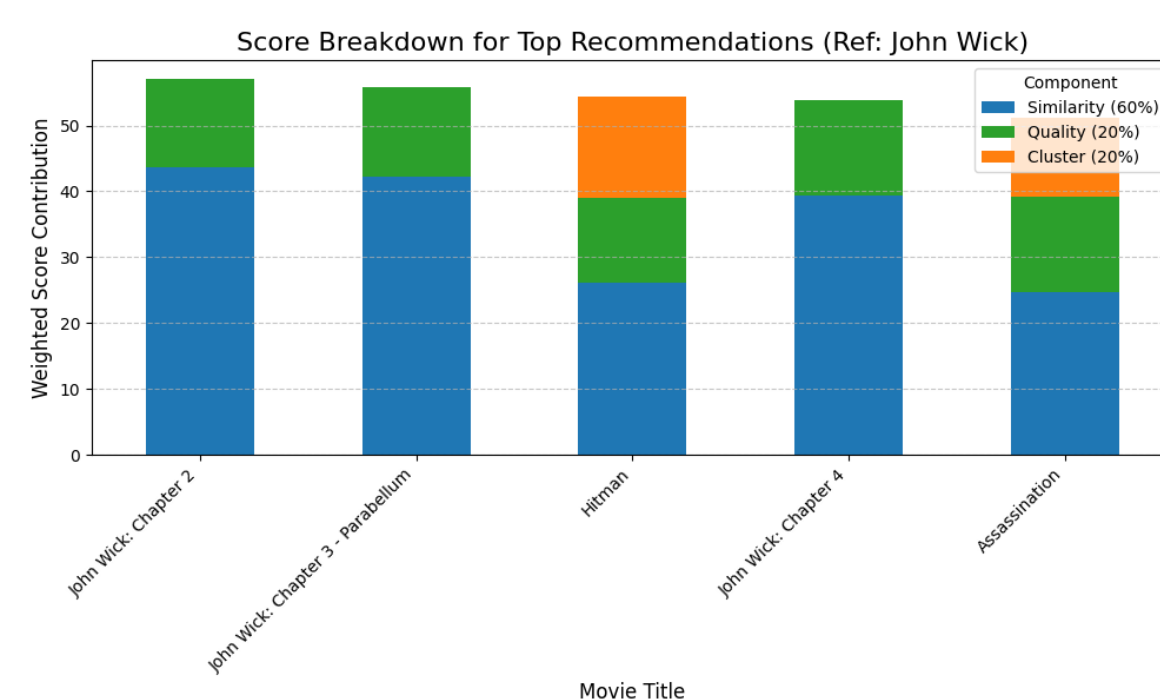
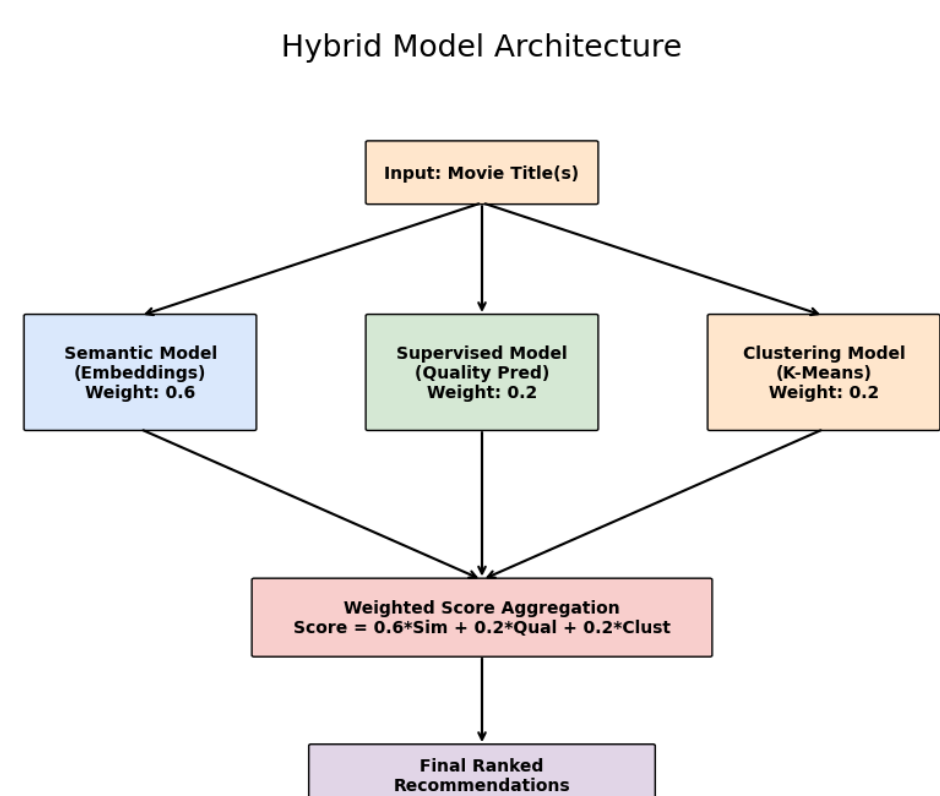
- ▶ Prédiction du score de qualité d'un film (0-100)
- ▶ Modèle supervisé : Random Forest Regressor
- ▶ Variables utilisées : année, durée, budget, revenus, genres
- ▶ Performance : $R^2 \approx 0.62$ (± 10 points pour $\sim 95\%$ des films)

Modèle Sémantique
(Transfer Learning)

- ▶ Embeddings textuels via all-MiniLM-L6-v2 (vecteurs **384-D**)
- ▶ Mécanisme d'**attention** pondérant titre, genres et description
- ▶ **Similarité cosinus** pour identifier films aux histoires proches



Modèle hybride



Discussion et conclusion

Limitations

- ▶ Dépendance à la base de données incluse dans les modèles
- ▶ Pondérations du modèle hybride fixées manuellement

Améliorations

- ▶ Apprentissage des pondérations du modèle hybride à partir de rétroaction de l'utilisateur
- ▶ Utilisation d'une API pour avoir accès à plus de données
- ▶ Recommandations à partir de descriptions libres (films non présents dans la base)
- ▶ Interface utilisateur permettant la recherche de films

Références / References

[1] Ashpal Singh. *IMDb Movies Dataset*. Kaggle, 2023. <https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>