

LES VAC' CINÉS! PRÉSENTENT



Team production :

Nos développeurs :

David, Gilles, Olivier, Romain et Leila

Nos webdesigner-UX :

David, Gilles, Olivier, Romain et Leila

Nos data analysts :

David, Gilles, Olivier, Romain et Leila

Product Owner :

Gilles

Scrum Master :

Leila

Sommaire



- 1. Outils et organisation**
- 2. Phase de cleaning**
- 3. API**
- 4. Visualisations et spécialisation**
- 5. Recommandations : choix pour le Machine Learning**
- 6. Notre appli**
- 7. Limites**
- 8. Next steps**

1. Outils et organisation



Pour ce projet nous avons utilisé des librairies comme : dash, numpy, pandas, plotly, sklearn, ect...

Les outils pour nous organiser : github, discord, trello, ect...

3 grandes phases de travail en sont ressorties :

- 1 Cleaning de la data + API**
- 2 Développement des modèles + Viz**
- 3 Implementation sur Dash + Tests**

2. Phase de cleaning

- dataset volumineux (infos dispersées dans plusieurs tables)
- filtrage : type 'movie', excl. genre Adult, films distribués en France (language Fr)
- liste tconst à conserver dans chaque table
- jointures et 'merge' tables des tables allégées
- drop colonnes : 'ordin', 'job', 'characters', 'endYear'.
- remplace les \N par des NaN.
- table 'overview' finale



3. API

- tmdb: id
- 'import request' et boucle
- colonnes récupérées : cf. photo
- perspectives: posters (affichage ML), origin_country (fonction ML).

	original_title	origina...	overview	budget	revenue	popularity	origin_country	poster_path	tmdb
	The Dark Knight	en	La suite ...	185000000.0	1004558444.0	125.591	US	https://image.tmdb.org/t/p/...	151
	Inception	en	Dom Cob...	160000000.0	825532764.0	130.588	US	https://image.tmdb.org/t/p/...	2720
	The Lord of the Rings: The Fellowship of th...	en	Le jeune ...	93000000.0	871368364.0	130.585	US	https://image.tmdb.org/t/p/...	12
	The Lord of the Rings: The Return of the King	en	Les armé...	94000000.0	1118888979.0	111.953	US	https://image.tmdb.org/t/p/...	12
	Interstellar	en	Dans un ...	165000000.0	701729206.0	140.424	US	https://image.tmdb.org/t/p/...	15733
	The Dark Knight Rises	en	Afin que l...	250000000.0	1081041287.0	127.466	GB	https://image.tmdb.org/t/p/...	4902
	The Lord of the Rings: The Two Towers	en	Après la ...	79000000.0	926287400.0	109.522	US	https://image.tmdb.org/t/p/...	12
	Django Unchained	en	Dans le s...	100000000.0	425368238.0	73.923	US	https://image.tmdb.org/t/p/...	6871

4. Visualisations et spécialisation



VIZ :

- dataset films distribués en France après 1920
- les stats sont incorporées à l'application.

SPÉCIALISATION ML :

- films de 1980 à aujourd'hui
- ayant une note supérieure à 6

5. Recommandations : choix pour le Machine Learning

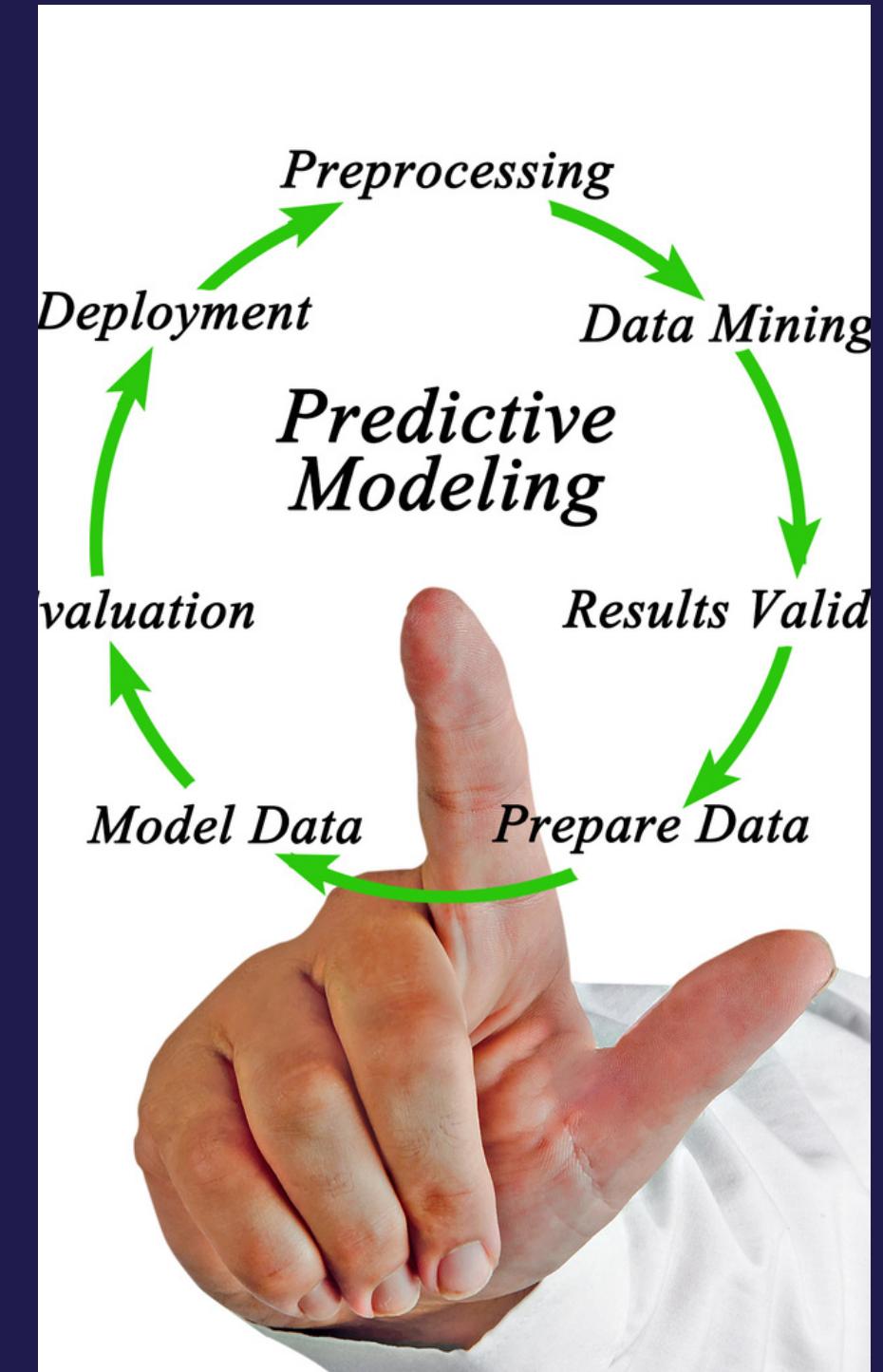
2 options de recommandations différentes pour plus de choix :

- Modèle KNN, 'K-Nearest Neighbor' :

classification basée sur des paramètres jugés pertinents par l'équipe.

- Modèle 'Cosine' :

pourcentage de similarité entre deux films basé sur des mots clé.



6. Notre appli :

Pour y accéder :
Vous êtes prêts??!

http://les_Vac_Cines.com



[https://github.com/Ochanman/
Dash_Reco_Imdb](https://github.com/Ochanman/Dash_Reco_Imdb)

The screenshot shows a movie recommendation application titled "Les Vac' Cinés!". At the top, there is a navigation bar with links: Accueil, Industrie, Acteurs & Actrices, Réaliseurs, Genres, Durée, Recommandations KNN, and Recommandations Cosine. On the left side of the main content area, there is a vertical sidebar featuring a logo with a film strip and clapperboard, and the text "YOUR MOVIES ON LINE". The central part of the screen displays a large, stylized title "Les Vac' Cinés!" in red and black, with a cartoon syringe character pointing towards it. Below the title, there is a view of Earth from space. The bottom of the screen shows the back of a theater audience seated in dark rows.

You vous êtes déjà demandé quoi regarder? Heureusement, il existe un service qui peut vous proposer des contenus qui vous sont proposés et qui vous aidera à trouver les films qui correspondent à vos recherches précises en insérant des mots-clés, des noms de films. Ce service vous est proposé par Gilles,

7. Limites



- API très utile mais téléchargement des données très long.
- Plotly , une visualisation 'basic' (versus les autres librairies).
- KNN et Cosine efficaces mais non pour de grands volumes.

8. Next steps :

- fonctions et graphs appelés grâce à un import de fichiers.py dans le script principal
- optimiser le KNN avec Randomized Search: configurer le modèle à partir d'hyperparamètres
- optimiser le modèle 'Cosine' via du NLP sur le synopsis des films
- proposition de la filmographie de l'acteur/actrice
- noter la pertinence des recommandations



'Dry as much as possible'!

Merci de votre attention !

