

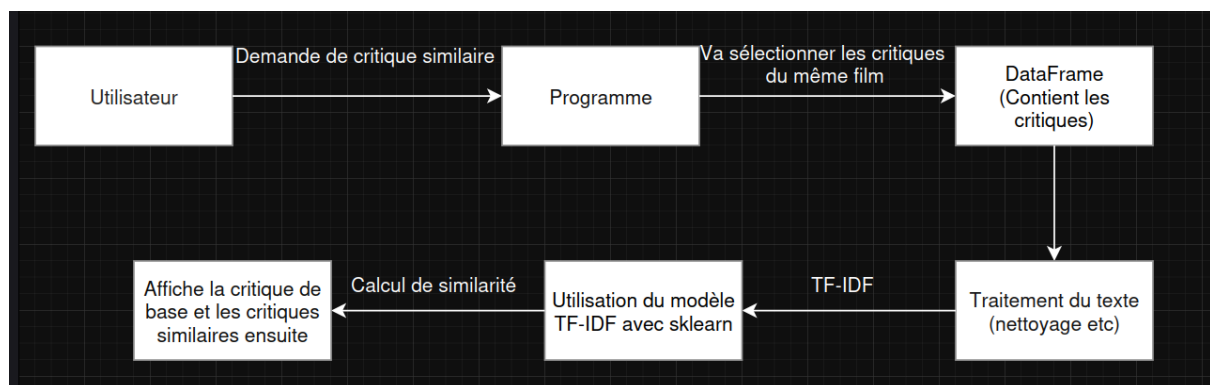
System Design – Recommandation de Critiques SensCritique

Le but de ce projet est de permettre à un utilisateur de trouver des critiques similaires à une critique en cours de lecture, mais uniquement pour le même film. L'approche retenue combine un filtrage par film et un calcul de similarité textuelle basé sur TF-IDF.

Pour donner plus de contexte avant de commencer lors d'un précédent stage à l'urssaf j'ai du faire des calculs de similarités (via l'algorithme de Levenshtein) et c'est la raison de pourquoi je connaissais déjà ce type de calcul avec la bibliothèque Sklearn car j'avais cherché plusieurs alternatives possible à ce moment là. (si vous souhaitez consulter mon rapport de stage à ce sujet je vous renvoie sur mon portfolio dans la catégories stage : <https://tharindu-p.github.io/portfolio/>)

Le système est conçu pour être simple, efficace et extensible puisqu'on peut ajouter d'autres csv de critiques.

Diagramme d'architecture



Description des composants

Bloc	Fonction	Choix technique
Utilisateur	Interagit via interface CLI ou futur front-end web	Simple, compatible avec Python
Nettoyage texte	Retrait des balises HTML, normalisation des espaces et suppression des stopwords français	BeautifulSoup + regex pour nettoyage, NLTK pour stopwords
TF-IDF vectorization	Transformation des critiques en vecteurs numériques	scikit-learn TfidfVectorizer, adapté pour petites/moyennes bases
Cosine similarity	Calcul de la proximité entre les critiques	Cosine similarity de scikit-learn permet de trouver les critiques les plus proches
Résultat – Top N critiques	Retour des critiques les plus similaires avec score	Affichage clair pour l'utilisateur

Utilisation de l'IA

- **README** : le texte du README a été amélioré avec l'aide d'une IA pour rendre le tout plus "propre" et agréable à lire.
- **Code Python** : l'IA a été utilisée uniquement pour suggérer certains imports et modules utiles, notamment :
 - typing pour les annotations de type
 - nltk pour les stopwords français
 - BeautifulSoup pour le nettoyage du HTML

Par contre aucune logique pour le calcul de similarité n'a été généré par l'IA.