Projet Web Scraping : Analyse de commentaires The Fork



Votre équipe a été recrutée par les services d’hygiène en restauration. Le but du projet est d’aider ces services à repérer les restaurants à risques grâce à l’IA pour leur permettre d’intervenir au plus tôt. Pour cela, vous allez devoir récolter des commentaires sur des restaurants grâce au web scraping, pour ensuite pouvoir les analyser grâce à l’IA et ainsi repérer les commentaires qui indiquent un problème dans un restaurant.

Vos tâches :

* Créer un script de web scraping permettant de récupérer des commentaires positifs et négatifs sur le site de réservation [The Fork](https://www.thefork.fr/).
  + Données à récupérer : nom des restaurants, adresse des restaurants, catégorie des restaurants, commentaires, notes des commentaires.
* Stocker les données récupérées dans une BDD (type mongoDB ou SQL)
* Traiter le texte de ces commentaires pour en faire un dataset d'entraînement pour un modèle de machine learning
* Entraîner un modèle de machine learning permettant de classifier les commentaires

Modalités :

* Taille du groupe de travail : binôme
* Début projet : Mercredi 13/12/2023
* Fin projet : Vendredi 15/12/2023
* Présentations : Vendredi 15/12/2023 (après-midi)

Livrables :

* Présentation orale de votre projet
* Notebooks ou Scripts de scraping et insertion des données dans la BDD.