

Mini-Projet: Real time context aware Recommendation for tourism using sentiment analysis

1. Introduction

La planification des visites touristiques, des sorties gastronomiques ou des activités de shopping est influencée par divers facteurs dynamiques comme la position géographique, les conditions météorologiques, l'heure de la journée, et même les avis des utilisateurs disponibles sur les réseaux sociaux. Ce projet vise à développer un système de recommandation en temps réel pour suggérer des lieux (restaurants, boutiques, sites touristiques, etc.) en tenant compte de ces paramètres changeants, tout en analysant les sentiments exprimés sur les réseaux sociaux à propos des lieux recommandés.

Objectif du projet :

Créer une application Big Data intégrant un système de recommandation personnalisé et temps réel, qui :

- Analyse les données spatio-temporelles pour adapter les recommandations
- Intègre des analyses de sentiments (depuis des datasets Yelp, foursquare, Reddit ou Twitter) pour évaluer la satisfaction des utilisateurs vis-à-vis des lieux recommandés.
- Fournit des recommandations contextualisées en fonction des préférences des utilisateurs, des tendances locales et des données en temps réel.

2. Outils de travail

Kafka, Spark et BD Nosql (i.e :Mongo DB), Mlflow, DVC, Airflow, webapp/Power BI, etc.

3. Étapes du projet

Acquisition des données :

- Collecter les données d'une API ou d'une base de données existante (Google Maps API, OpenWeatherMap API).
- Extraire les sentiments depuis les commentaires (reviews)
- Stocker ces données dans une base NoSQL pour une gestion efficace.

Analyse de Sentiments :

- Utiliser des modèles pré-entraînés pour évaluer les sentiments exprimés dans les reviews. (n'oublier pas la préparation des données)
- Générer un score de popularité pour chaque lieu basé sur les sentiments cumulés et récents des utilisateurs.

Préparation des données

- Nettoyage et prétraitement de données
- Génération des caractéristiques à partir des métadonnées

Modélisation des Recommandations :

Comparer les résultats de deux techniques :

la première :

- **Filtrage collaboratif** : utiliser une technique de factorisation de matrice (ALS, SVD, etc) pour Identifier des lieux populaires auprès d'utilisateurs similaires.
- **Filtrage basé sur le contenu** : Utiliser les caractéristiques des lieux pour faire correspondre avec les préférences utilisateur.
- Une méthode qui fusionne ces approches avec des techniques **d'hybridation** pour de meilleures performances.

La deuxième :

une technique récent (voir arxiv.org) ex : Two tower architecture, GNN, Mamba, Attention, etc.

Intégration du Temps Réel :

- Configurer une pipeline Kafka pour traiter les flux de données en temps réel (nouveaux reviews, météo, positions GPS).
- Intégrer Spark Streaming pour analyser et mettre à jour les recommandations au fil du temps.

Visualisation :

Développer un tableau de bord interactif pour afficher :

- Les recommandations actuelles.
- Les tendances des lieux les plus populaires dans la zone géographique de l'utilisateur.
- Une carte en temps réel des avis et sentiments sur les lieux proches.

4. Livrables du projet :

- Rapport de 12 pages max
- Présentation 15 min
- Code-source

5. Consignes :

- Le travail d'équipe est fortement recommandé (quardinome max)
- La mission de chaque membre de l'équipe doit être bien identifiée et équilibrée par rapport au autres membre.
- Dresser un planning prévisionnel avec les taches nécessaires pour l'accomplissement de projets avec le responsable.
- Décomposer les taches « separation of concern » pour favoriser le travail en parallèle
- Réaliser des documents d'interfaçage pour se mettre d'accord sur les points d'intersections entre les tâches
- Utiliser un outlis (ex. Gira, trello, etc.) pour le Suivi de réalisation de projet. (Data Driven Scrum)