Restaurant Recommendation

Project 2 : Kevin Lee, Ayush Agarwal, Yumin Zhang, Yi Xie

Dataset Overview

The project leverages the Yelp Dataset, focusing on restaurant reviews in Toronto from 2015 to 2019. The dataset comprises data from 5,471 restaurants, 44,485 users, and 253,050 restaurant reviews. Key features include:

1. Restaurant Details: Name, location, and category of each restaurant.
2. User Reviews: Text reviews and ratings (1–5 scale) provided by users.
3. Sentiment Scores: Polarity and compound scores derived from text reviews using Textblob and VADER for sentiment analysis.
4. Geographical Data: Location coordinates for restaurants to facilitate location-based recommendations.

Business Problem Statement

The goal is to enhance the dining experience by providing personalized restaurant recommendations. This system aims to benefit users by suggesting dining options tailored to their preferences and locations, thereby improving user satisfaction and engagement with food delivery apps like Deliveroo, GrabFood, and FoodPanda. For businesses, this translates to increased visibility, customer footfall, and potentially higher revenues, particularly for lesser-known establishments.

Goals

The technical objectives of this project are to:

1. **Develop a Robust Recommendation Engine**: Implement algorithms that analyze user preferences and dining history to make personalized restaurant suggestions.
2. **Incorporate Sentiment Analysis**: Utilize sentiment analysis of user reviews to refine the recommendation accuracy, going beyond mere rating scores.
3. **Enable Location-Based Recommendations**: Create a system that suggests restaurants based on the user's current location, addressing the user cold-start problem.

4. **Enhance User Experience**: Use topic modeling (LDA) to identify key themes in reviews, further personalizing recommendations based on user interests.

Methods to Be Used

To achieve these goals, the project will employ:

1. **Collaborative Filtering**: To identify similar users and recommend restaurants liked by those with similar tastes.
2. **Content-Based Filtering**: To recommend restaurants similar to those a user has previously enjoyed, based on features like cuisine type and sentiment scores.
3. **Natural Language Processing (NLP)**: For sentiment analysis and topic modeling, extracting meaningful insights from user reviews to influence recommendations.
4. **K-Means Clustering and Geospatial Analysis:** For grouping restaurants based on location and making proximity-based suggestions to users.

The system will iterate through feedback loops, continuously refining its recommendations based on user interactions and feedback, ensuring the relevance and accuracy of suggestions.