Capstone project proposal
Hotel Recommender system

1. Problem to solve

According to the 2016 Tripadvisor trip barometer, most travelers decide their destination based on the airline tickets, and after that they look for a convenient accommodation, which accounts for the most time spent once the destination is decided. A personalized recommender system that knows the clients' preferences and experiences can have a significant impact in the hotel search process, facilitating their decision-making steps and reducing the user efforts, thus improving the user experience.

In this project, I will implement two hotel recommender systems capable of suggesting the hotels may fit the user better based on either previous stays and based on people similar users. Both results will we provided with a meaningful score. A model-based cold-start solution will be provided.

The system will be made publicly available as a website.

2. The client

The clients are the travelers / users who are looking for an accommodation on an already known destination. The goal is to attract as many users to the platform as possible, so advertising revenues increase. As of now, I do not have access to traffic data on any of the main search engines, so it will not be possible to show an improvement.

3. Data

A large dataset of hotel information and ratings is readily available. The one selected contains almost 1000 hotels from 4 different cities. The data lacks information about the hotel star range and the price. That information will be completed by scraping the search engines for a specific date – even though a single date may not be fully meaningful, it can be taken as a proof of concept. The dataset selected is not very large because of the need to search for additional information within a reasonable timeline.

A virtual set of clients will be generated to begin with. This will be based on academic articles describing relationships and principles involved in booking decisions. A database of users with their feedback will be created and completed as users sign up.

4. Approach

A dual recommender system with a specific cold start approach.

Will use NMF as collaborative filtering

Will use a Bayesian classifier for content-based filtering

5. Deliverables

Name	Tool	Date
Datasets	Scraping / self-generated	14 Aug
NMF code	Sklearn	16 Aug
Bayesian code	Sklearn	20 Aug
Basic website	Django	21 Aug
Final website	Django	23 Aug
Go live	TBD	TBD