

Final Project Proposal  
Vikas Sinha  
7/7/2019

For the final project, I would like to implement a Hotel recommendation system based on data from Expedia. The problem is currently posted on Kaggle as a competition at <https://www.kaggle.com/c/expedia-hotel-recommendations/overview>

For this project I would aim to evaluate my Recommender model using a portion of the training data that is withheld from the model.

The problem requires use of information other than explicit ratings, since the information relates to user behavior such as searches or clicks executed by the user. Description below is copied from Kaggle:

Expedia has provided you logs of customer behavior. These include what customers searched for, how they interacted with search results (click/book), whether or not the search result was a travel package. The data in this competition is a random selection from Expedia and is not representative of the overall statistics.

Expedia is interested in predicting which hotel group a user is going to book. Expedia has in-house algorithms to form hotel clusters, where similar hotels for a search (based on historical price, customer star ratings, geographical locations relative to city center, etc) are grouped together. These hotel clusters serve as good identifiers to which types of hotels people are going to book, while avoiding outliers such as new hotels that don't have historical data.

Your goal of this competition is to predict the booking outcome (hotel cluster) for a user event, based on their search and other attributes associated with that user event.