

## ***Project 2: Restaurant Recommendation System***

**Lara Clasen**  
**Spring Term 2020**  
**LLClasen.github.io**

### **Which Domain?**

*What domain is this data going to come from? Please list 10 references for your domain (with a brief annotation) to use to make sense of what you're doing with these data.*

My data set for Project 2 comes from the popular crowd-sourced review forum Yelp. Documentation for structure of the original data set can be found here: <https://www.yelp.com/dataset/documentation/main>.

### References

1. Rocca, B. (2019, June 12). Introduction to Recommender Systems. Retrieved from <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
  - a. Offers a detailed overview of the different recommendation systems that are available and in what situations you might use them.
2. Poudyal, R. (2018, June 27). Latent Factor based method in collaborative filtering. Retrieved from <https://medium.com/@rabinpoudyal1995/latent-factor-based-method-in-collaborative-filtering-77756a02f675>
  - a. This article reviews just one type of recommendation system, and uses generalized examples to explain how it works.
3. Das, S. (2015, September 24). Beginners Guide to learn about Content Based Recommender Engine. Retrieved from <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/>
  - a. This article reviews content-based recommender engines popular services such as Facebook to demonstrate its usage.
4. Mwititi, D. (2019, December 3). How to build a Simple Recommender System in Python. Retrieved from <https://towardsdatascience.com/how-to-build-a-simple-recommender-system-in-python-375093c3fb7d>
  - a. Simple recommendation systems offer a straightforward version of a recommender algorithm, which this article explains thoroughly.
5. Kirzhner, E. (2018, May 11). Machine Learning. Explanation of Collaborative Filtering vs Content Based Filtering. Retrieved from <https://codeburst.io/explanation-of-recommender-systems-in-information-retrieval-13077e1d916c>
  - a. This article is very helpful for my case as it takes the time to compare two types of more complex recommendation algorithms.
6. Ansari, A., & Kohli, R. (2000, August). Internet recommendation systems. Retrieved from [https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/385/Internet Recommendation Systems.pdf](https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/385/Internet%20Recommendation%20Systems.pdf)
  - a. An older article, this write-up offers some more numerical insight into the recommendation system.
7. Chauhan, N. S. (2019, July 9). Implement K-Nearest Neighbors classification Algorithm. Retrieved from <https://towardsdatascience.com/implement-k-nearest-neighbors-classification-algorithm-c99be8f14052>

- a. This article offers a valuable example of using the K-Nearest Neighbors classification algorithm.
8. Recommender Systems. (n.d.). Retrieved from <http://recommender-systems.org/>
  - a. A very basic yet useful resource for understanding different recommendation system types.
9. What is Yelp's recommendation software? (n.d.). Retrieved from [https://www.yelp-support.com/article/What-is-Yelp-s-recommendation-software?l=en\\_US](https://www.yelp-support.com/article/What-is-Yelp-s-recommendation-software?l=en_US)
  - a. A brief and somewhat vague overview of Yelp's recommendation system provided by the company itself.
10. Liao, K. (2018, November 19). Prototyping a Recommender System Step by Step Part 1: KNN Item-Based Collaborative Filtering. Retrieved from <https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea>
  - a. Another helpful resource for the utilization of the K-Nearest Neighbors algorithm, specifically for use with collaborative filtering.

## **Which Data?**

*What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.*

The dataset originally came as part of a challenge introduced by the company allowing data scientists to conduct research or perform analysis on the website's data in their own ways. The dataset can be downloaded from <https://www.kaggle.com/yelp-dataset/yelp-dataset>. This data set is a compilation of a large number of reviews and other business information available as an open subset of the website's available data. The data is made available as part of a challenge offered by Yelp, as well as simply to be used as a resource for students or other data scientists to use.

## **Research Questions? Benefits? Why analyze these data?**

*How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.*

With this data set I plan to build a restaurant recommender system. Recommendation systems are algorithms that work to offer relevant suggestions to the user based on a particular industry or product area. This data is important to analyze because it combines two huge pieces of the dining experience in today's world: recommendations and business information. It will allow users to obtain accurate recommendations for dining out without the requirement of doing anything but looking at their phone or using the app which utilizes such an algorithm.

## **What Method?**

*What methods will you be using? What will those methods provide in terms of analysis? How is this useful?*

There are a few different types of recommendation systems that can be used for this sort of data. A simple recommender is pretty basic and offer suggestions based on what the general public "likes". However, I'd like to get a bit deeper and so I want to build either a content-based or collaborative filtering recommendation system for this Yelp data. Both of these use the specific user's behavior to make suggestions. While content-based recommenders base their suggestions directly on user preferences for product features, collaborative filtering (as the name suggests) makes assumptions based on the user's

preferences as they relate to other user preferences. Because we have quite a bit of “content” to work with in this data set, we should be able to build either type of algorithm. I plan to use the k-Nearest Neighbors algorithm to begin the process of building a collaborative filtering recommender.

### **Potential Issues?**

*What challenges do you anticipate having? What could cause this project to go off schedule?*

Since we have worked with k-NN algorithms more than once, I don’t anticipate having much trouble in that respect. However, I don’t believe that I have experience with a content-based recommender and so I may run into some unanticipated trouble there. My main goal is to create a collaborative filtering algorithm so that is what I will focus on.

### **Concluding Remarks**

*Tie it all together. Think of this section as your final report’s abstract.*

In this second project I plan to build a recommendation system to predict a user’s preference in restaurant most effectively. I will use Yelp metadata to create a recommender that can work off of user behavior and business content to produce suggestions relevant to the user. This project and its resulting algorithm will be incredibly useful in a world where many people rely on reviews and recommendations before they will try things, especially new restaurants. By offering accurate recommendations, any application utilizing the algorithm offers a valuable service to their user.