

Creating popular dishes list at Yelp by Machine Learning

Project Proposal

Shirley Zhu

- What is the problem I want to solve?

When people go to a new restaurant, they usually don't know what to order. Yelp reviews can usually provide some ideas about what they might like, but there are often too many reviews and it takes a long time to read the reviews and make a decision. This project leverages machine learning to figure out what everyone prefers to eat based on the restaurant's reviews.

- Who is my client and why do they care about this problem? In other words, what will my client do or decide based on this analysis that they wouldn't have done otherwise?

My client will be Yelp and the users of this feature will be the restaurants that list their business on Yelp and the customers who go to eat in these restaurants. Yelp can build this feature because users would love using Yelp to figure out what to order. Yelp can help restaurants to roll out its "popular dishes feature" to help the customers in deciding what to order from the restaurant with whom they aren't familiar. Good rated restaurants can have bad dishes and ok restaurants can have really good dishes. The restaurants can use this information to advertise their popular dishes and improve their not so good dishes, so the customers will be more satisfied after eating there. This will boost the business of the restaurants. Yelp can charge the restaurants for this feature, and Yelp takeout business will grow too.

- What data will I be using? How will I acquire the data?

The data will be from Yelp dataset website:

<https://www.yelp.com/dataset/documentation/main>

I will use the San Francisco Bay Area restaurants data from the website above.

- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

I will use Natural Language Processing to train a model to analyze the sentiment of the reviews associated with different dishes. The labels of the review text will be the rating score of that review, the higher the rating, the more positive the sentiment. The output will be the overall sentiment of the dish items. The top ranks of these items will for a list of popular dishes.

- What are your deliverables? Typically, this includes code, a paper, or a slide deck.

The deliverables will include code, a report and a slide summary.