

## INDIVIDUAL PROJECT

### CMPE 256 Large Scale Analytics Summer-2019

**Project Title:** Yelp review recommendation

**Instructor:** Shih Yu Chang

**Name:** Sai Sarath Vattikuti

**SJSU ID:** 013821483

#### **Objective:**

The core idea is to recommend a restaurant such that you are more interested to like a restaurant if its reviews are similar to the reviews of the restaurants you already love.

**Summary:** Sometimes, you don't want Yelp to recommend "similar" restaurants per user history, but rather restaurants that worth exploring given your rating of the current restaurant. The problem with many recommendation systems including those in popular apps is their focus on \*similarity\* -- not the differences. Ultimately, however, I don't want to eat the same or have the same experience, but rather I would like to find a "surprisingly good" restaurant that may be a little different. This notion might be a bit hard to quantify, I would like to explore simple techniques for doing this using Yelp Data.

**Data collection:** Data is collected using the API provided by the yelp and by using the get request in the python script.

<https://www.yelp.com/developers>

**Data Pre-Processing:** The first step in the pre-processing is to remove the Null values and duplicates. The text data is translated and processed using lemmatization , stop words, snowball stemmer.

**Data Modelling:** I want to find a way to represent reviews using a bag-of-words representation. After doing so, we will also represent categories using a one-hot encoding representation. Such that,I can manipulate those representations to find similarities and differences while balancing the weights of the two. Note that the core idea assume that you are more likely to love a restaurant if its reviews are similar to the reviews of the restaurants you already had.

**Expected Result:** we compute the correlation distance of the average vectorized reviews to all the reviews, and compute the correlation distance between this category and all other categories.

S.no	Business_id	categories	name	stars
USER INPUT	aUrOyWFKxKeVXiFzwbTXSA	Vegetarian, Cafes, Vegan, Restaurants	The Green Owl Cafe	4.0
RECOMMENDATION	QXAEGFB4oINsVuTFxEYKFQ	Specialty Food, Restaurants	Emerald Chinese Restaurant	3.5

