

Machine Learning Report

Section 1: Problem Identification.

Information is a key element of the making decision process for each transaction no matter the industry. The more information customer and companies have, the more value each transaction can get, and the more robust markets can turn. The value of information is high/expensive, and the ability to generate and understand it is a clear source of competitive advantage for companies and value capture for consumers.

The power of the information is a strong driver that makes leaders/managers able to change the course of a business, or even a market. In general, service markets are susceptible to information as an important element of the customer making decision process. We identified that phenomenon in the food market, where restaurants and its brands, can be extremely valued based on the ability to generate and spread positive information about them.

Restaurants (and its customers) are the objective of our analysis within the food market. They have always strived to deliver services that are fit to customers' desires. The analytical shift have added an edge that this market could leverage on, which is precisely the aim of our app.

The problem we are facing is the lack of clear, objective, real, fast, and customized information about restaurants within the state of Arizona, in the city of Phoenix. The solution we are recommending is our app called Bon Appetit which provides that information based on collaborative filtering, and sentiment analysis of user review data.

As mentioned previously, the information is valuable for business and customers. For this reason, our app approaches both users. The functionality is the same, but the implementation of the insights from the app are different based on the sphere of movement of each user.

Section 2: Analysis Method

Our app was developed by analyzing the yelp dataset . The analysis will use the yelp customer reviews text data as well as overall ratings data about the restaurants in Phoenix, Arizona.

Following are the analysis method we used:

- Descriptive statistics
- Collaborative filtering
- Sentiment analysis

We have 3 use cases with respect to the user of the app which produces results based on these above mentioned analysis.

a. Use Case 1 : Registered or unregistered customer/restaurant owner

The method for use case 1 is an extractive data method based on a descriptive analysis that is used to get information relative to specific criteria of interest. The need of information is a key element of the making decision process for restaurants and its customers. The food, service, location, ratings, and others; are insights of importance at the moment to decide. The descriptive analysis facilitates the making decision process by highlighting important criteria based on each user perspective. The present app provides the possibility to get access to that information generating a compiled review based on criteria of importance.

In the case when a customer wants to find the type of cuisine he/she wants, there is an option where he/she can input his/her choice using this the app will produce a customized result. In the case of a restaurant owner, the user can enter the input of interest and scan the environment understanding the market in terms of competitors, prices, type of cuisine, location, among others.

b. Use Case 2 : Registered customer only *

The method we used for use case 2 is collaborative filtering¹ of user ratings data for different restaurants to give the top 10 recommendation for a particular user.

From a marketing perspective, the utilization of recommendation is a useful technique to keep alive the customer over the time. The Customer Equity can be sustained and increased by this technique promoting the transactions between customers and restaurants. The customized recommendation using the name of the customer, and providing evidence of his/her previous reviews/behavior is a positive manner to keep customers alive.

In this case, a registered user login to our app then we show recommendations(filtering) based on the user's past review ratings as well as information we obtained from the preferences of other users(Collaborative).

c. Use Case 3 : Enterprise User

Here an enterprise like owner of a restaurant , marketing manager logs in to view the sentiments of people visiting the restaurant. We have utilized the concepts of Natural Language processing to identify the positive and negative sentiments.

¹ As the name suggest it is the filtering of a user's preferences by gleaning the preferences and tastes of many users(collaborative)

* This use case is under development, we have the algorithm in place, but we are having issues to integrate into Shiny

The app uses AFINN lexicon² give score to the single word(unigram) of reviews of each user for the particular restaurant, and uses NCR lexicon³ to get the most common words and uses BING lexicon⁴ to find the most positive and negative word and also utilize the same lexicon for the positive and negative word cloud.

There is a most common words word cloud, positive and negative words word cloud displayed the specific vocabulary people use to describe the restaurants, looking at which the owner can take business decisions.

App Layout

There are three tabs within the app one for each use case

In the first use case : Our app will take the preferences of individuals such as cuisines they want to eat, restaurant rating , zip-code and preferred pricing and will output pin-up locations on the map of the top restaurants nearby and hovering over these pin-ups will give the specific Name and cuisine type

In the second use case : The app takes an additional factor of User ID into consideration to provide very specific recommendations on some restaurants liked by that user based the user's previous ratings.

In the third use case : In this case the app takes Business ID as an input and provides an output with two word clouds one with the most common words used and the other with a positive and negative words in it with the negative on the top and the positive at the bottom. The output also shows tables for the three categories with top 10 common words , top 10 positive words and top 10 negative words.

Section 3: Data Description.

The yelp dataset consists of following json files:

² *Finn Arup Nielsen*: It is a vocabulary of words that has scores from -5 to 5 associated to it depending on the positivity or negativity of the word.

³ *NRC Emotion Lexicon from Saif Mohammad and Peter Turney* :It is a vocabulary of words that has emotions ranging from “anger”, “fear”, “joy” .. etc associated to it depending on the emotion it is indicative of.

⁴ *From Bing Liu and collaborators*: It is the vocabulary of words that has positive and negative flag associated to it depending on the positive and negative connotation of the word.

1. Business.json : This file consists data about the business entity like location details, facilities and overall rating of the place.
2. CheckIn.json : This file has the data on the number of checkins at a particular day and time by the customers who visit the restaurant.
3. Photos.json : This is a metadata of the restaurant photos uploaded by the customers.
4. Review.json : This file has the “text” review of the restaurant by the customer and individual star ratings provided by that user.
5. Tip.json : This file has some useful tips about the restaurant, provided by the customers.
6. User.json : This file provides information about the user who provides input to photos review and tips.

We scraped relevant data from the above json files required for our use cases

Our json of interests are enumerated below:

- I. Business.json
- II. Review.json
- III. User.json

A data frame was created using these json as the source with some predefined filters. The *business* data set was modified by combining cuisines based on intuitive similarity, i.e. restaurants that serve Argentinian or Brazilian food are combined in the South American cuisine category. Moreover, we reduced the size of the data by only using restaurants that are located in Phoenix, AZ.

Section 4: Link to the app.

<https://himanshi-sharma.shinyapps.io/final/>