Restaurants in Hoboken, NJ: Predicting review ratings based on review terms

Course Project: Term Paper

Authors: Daniel Hunt, [Sufyan Gulab Shaha](mailto:sshaha@stevens.edu)

Web Mining

BIA-660

May 2024

**Introduction:**

Hoboken is known for having the most bars per capita of any city in the United States. Adding the delis, coffee shops, pizza parlors, ice cream shoppes, and restaurants without liquor licenses, Hoboken has a massive list of restaurant options within its 1.25 square mile. This makes the industry extremely competitive for existing and prospective owners.

One way these owners can find an advantage over competition is by analyzing the landscape and customer feedback. First, these owners should look at the other establishments to see what type of cuisine they offer and how saturated their niche might be. This can help an aspiring owner find the right niche with the most potential for market share. Second, existing owners can look at customer reviews to see what elements are discussed most and which themes drive a higher rating. By understanding the themes that customers associate with a 5-star review, owners can try to incorporate these themes and elements to boost their ratings. Further, owners can implement controlled tests and market research to understand what people might say about new menu items or experiential changes. These tests can be processed to understand the likely impact on the overall rating of a restaurant on Yelp.

**Preliminary Literature review:**

Word of mouth advertising and marketing was first popularized by George Silverman in the 1970s which led to a series of experiments proving that groups can be influenced by the opinions of individuals in a variety of life situations (Erickson, 2005). Sorensen and Rassmusen found that both positive and negative book reviews in the New York Times Book Review influenced sales with positive reviews leading to a more significant bump in sales than negative reviews (2004). Brands and entrepreneurs have harnessed public opinions to introduce and sell their products to new audiences.

The restaurant industry is one industry that relies heavily on reviews. As of the late 70s/early 80s, a significant portion of the population (greater than half) reads restaurant reviews (Jolson and Bushman, 1978). That has only increased with the rise of social media and digital forums. There is plenty of literature on length of reviews and rating of reviews influencing sales (Li et al., 2020; Chevalier and Mayzlin, 2006). However, there are limited reports on what terms are most common in each level of rating.

Farronato and Zervas found there were certain hygiene factors that could be extrapolated from Yelp reviews to aid governments in regulating restaurants, but there are limitations in things that customers may not be exposed to (2022). However, we are reviewing how owners can influence the consumer experience, so the terms we identify should all have a relevant impact on customer experience.

**Objectives and expected contributions:**

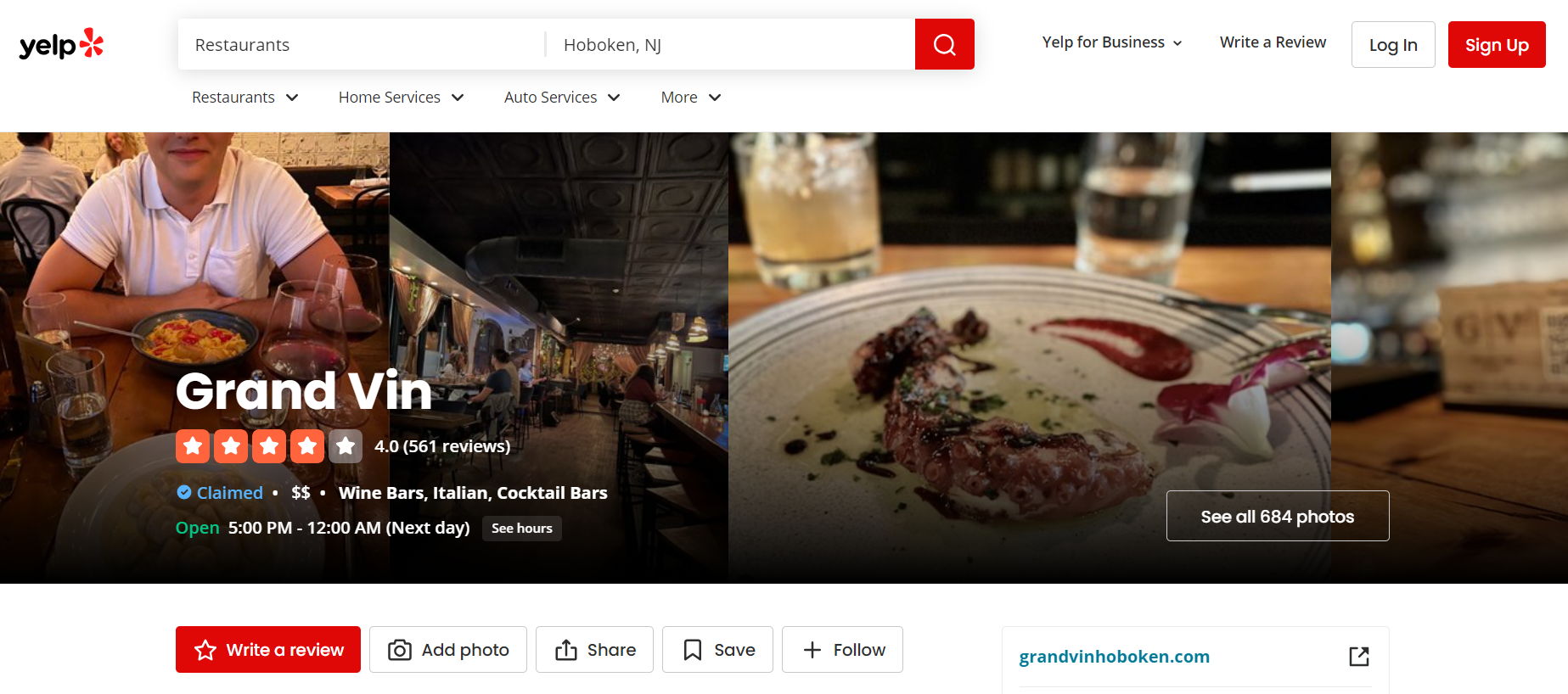
There are two objectives with this research:

1. To understand the sentiment and key elements of customer feedback ratings for restaurants in Hoboken.
2. To understand which cuisine is most saturated within Hoboken, and which cuisine is best rated as an aggregate category.

The contributions of this research will help inform restaurant owners how to prioritize business changes to influence customer satisfaction while informing owners of things that might lead to lower customer satisfaction. Lastly, the second objective sets an understanding of how saturated a food category is and what is the general sentiment for each type of category.

**Methodology:**

Public restaurant reviews are a popular way for customers to decide where to eat. Yelp, founded in 2004, has contributed over 200 million consumer reviews and ratings for local businesses. Its mission is to connect people with their local economy. The established website/application allows users to research restaurants and businesses based on other user feedback ratings, look at pictures of food, vibe, and menus, look up hours and amenities, and learn more about the owners.



*Figure 1.* Landing page for a restaurant based in hoboken.

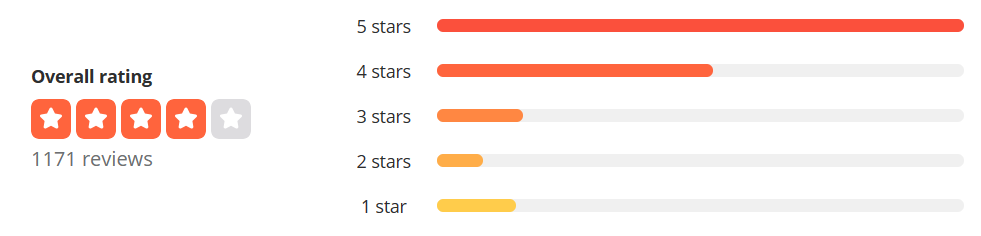
We designed a selenium based tool to navigate Yelp’s site geo-targeted to the Hoboken range. The code navigated through the list of most reviewed restaurants to extract links to each of the most reviewed restaurants. Once the links for top restaurants were collected, we leveraged a scrapy function to extract 10 pages of ratings and review copy in Yelp’s list order from each of the restaurant pages.

Once all rating and review data was extracted, we consolidated into a two column table for rating and review copy. To preprocess, we took several steps. First, when the review copy was extracted there were <br/> lines within the extracted text whenever a paragraph space was included in a review. All break elements were replaced with blank spaces. Next, there was some residual HTML language within some of the review text. It was the same string that was appearing in all reviews which indicated the language of the review. A line of code was included to remove this recurring HTML code. Finally, stopwords and punctuation were removed as they are not expected to influence the rating of the restaurant.

Once data was cleaned, we applied an TF-IDF function to generate a matrix and understand the importance of each word with the reviews and corpus. Finally, we built a text classification model by separating the data frame into two sets of information - 70% for training and 30% for testing. With the training data, we coded two classification models to map text to a rating - a Multinomial Naive Bayes model and Support Vector Machines model.

NB Model

A Multinomial Naive Bayes (NB) classification model predicts outputs based on probabilistic connections between features and labels. The assumption is that features operate independently of other features. The model estimated the probability of a particular rating based on the presence of terms within the review text. Given the dataset, we relied on Maximum Likelihood Estimates to dictate the model and set our alpha to 0. Further, we knew of the imbalance in the reviews. People who have bad experiences are more likely to post good reviews than people who have good experiences will post good reviews (Thomas, 2018). However, Yelp tends to see a higher quantity of positive reviews (3+ rating)(Figure 2.). It may be the community is more positive, or Yelp has designed its platform to incentivize more positive reviews.



*Figure 2. Example of a review distribution for The Cuban in Hoboken, NJ.*

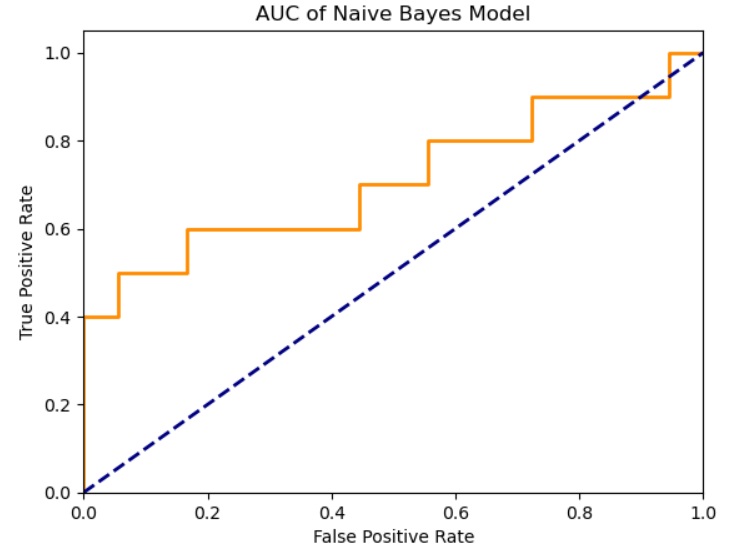
To offset this imbalance, we factored in a class weighting system to update the probabilities based on samples and class counts to balance the labeling frequency.

SVM Model

The second model we tested for this solution was a supported vector machine, a supervised learning algorithm. The algorithm is designed to classify data into classes and create maximum separation between each class using a vector as the divider. To address the stated imbalance issue, we incorporated class weights to make sure the ratings with fewer reviews received a higher weight against their probabilities. This meant a lift on 1 and 2 star review text.

Performance Testing

Once both models were trained, we assessed the performance of each using precision, recall, accuracy, F-Score, and AUC scoring. The most robust measurement for performance in this report was the AUC scores for each class. AUC scores are indicative of a model separating the classes and in our model these scores varied. For Naive Bayes, we saw moderate scores for one, two, three and five star ratings, but low scores for four star ratings.



AUC for class 1: 71.88%

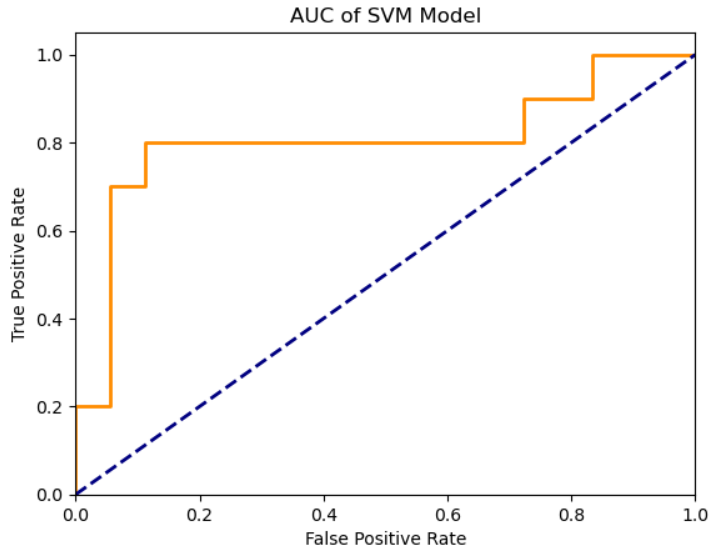
AUC for class 2: 71.88%

AUC for class 3: 68.00%

AUC for class 4: 46.94%

AUC for class 5: 71.11%

For the SVM model, there were slightly lower AUC scores. The model did a better job at separating three, four, and five star ratings, but it did not separate one and two star ratings very well:



SVM - AUC for class 1: 57.29%

SVM - AUC for class 2: 58.33%

SVM - AUC for class 3: 76.00%

SVM - AUC for class 4: 67.35%

SVM - AUC for class 5: 80.56%

Precision, recall and F-Score indicate the performance at a specific class threshold. The Naive Bayes model showcases low performances with weighted average scores around 30% (Precision - 30%, Recall - 36%, and F-Score - 29%). With the data assessed, our model proved moderately capable of separating ratings; however, the model did not effectively predict the ratings which means our model has potential and can identify some ratings, but not others. Four star ratings are particularly troubling due to high overlaps with five star and three star ratings.

The SVM model showcases even lower performance than the Naive Bayes model with weighted average scores ranging from 20% - 35% (Precision - 19%, Recall - 26%, and F-Score - 24%). One and Two star ratings received ‘0’ scores across precision, recall and F-Score. Five star reviews performed far better with 39% precision, 90% recall and 55% F-Score.

**Results and Discussion:**

In conclusion, the Naive Bayes model was the stronger model for this solution; however, the tool is not accurate enough to be seriously considered and applied by restaurant owners. We tested two random samples from a restaurant based on text in a two star review and a one star review. The models predicted the two star review to be a three and the one star review to be a five.

Looking at the the data, we can see trends in words that were most common and they are primarily specific menu items i.e. ‘pancakes’, ‘pork’, ‘paella’, ‘fish’, ‘mimosa’ or based adjectives i.e. ‘ juicy’, ‘soft’, ‘tender’, ‘delicious’. The tool can highlight which terms may be most influential in a review, but it struggles in the current form to understand sentiment and context around these terms to accurately place them in a rating category.

For now the tool should be used to see which terms are most discussed for a restaurant so new and existing owners can understand what items or descriptors they should make sure are positive and well received in hopes of gaining more positive reviews.

**References:**

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research, 43(3), 345–354. <http://www.jstor.org/stable/30162409>

Erickson, Kathleen.(2005) "The history of word of mouth marketing." Agri Marketing, vol. 43, no. 6, July-Aug. 2005, p. 42. Gale Business: Insights,

Farronato, C., & Zervas, G. (2022). Consumer reviews and regulation: Evidence from NYC restaurants. National Bureau of Economic Research. Working Paper 29715.

Jolson, M. A., & Bushman, F. A. (1978). Third-party consumer information systems: the case of the food critic. Journal of Retailing, 54(4), 63–63.

Li, H., Meng, F., Jeong, M., & Zhang, Z. (2020). To follow others or be yourself? social influence in online restaurant reviews. International Journal of Contemporary Hospitality Management, 32(3), 1067-1087. doi:<https://doi.org/10.1108/IJCHM-03-2019-0263>

Sorensen, A.T., & Rasmussen, S.J. (2004). Is Any Publicity Good Publicity? A Note on the Impact of Book Reviews.

Thomas, A. (2018). The secret ration that proves why customer reviews are so important: It’s what you don’t do that hurts you. Inc.com. Retrieved from [The Secret Ratio That Proves Why Customer Reviews Are So Important | Inc.com](https://www.inc.com/andrew-thomas/the-hidden-ratio-that-could-make-or-break-your-company.html).