A Prediction Model for Burlington Restaurant Ratings Based on Review Features

Caleb Oliveira
Data Science Undergrad
Caleb.Oliveira@uvm.edu

Emma O'Brien
Data Science Undergrad
Emma.R.OBrien@uvm.edu

ABSTRACT
We collected review data on Burlington restaurants from the popular review site, 'Yelp', collecting both the rating and text for each individual restaurant review, as well as the restaurant's cuisine type, location, and average rating. Upon examining our data we found some general trends regarding reviews in Burlington, notably that most restaurants have an average rating around 4 stars (median: 4, standard deviation: .5), that ratings have generally increased over time, and that certain streets seem to have higher rated restaurants than others. From here we performed sentiment analysis on every review in our dataset, training a logistic regression model to predict good or bad reviews based on word features. Using this model we were able to obtain a validation accuracy of 78%. Additionally we developed a random forest model using one hot encoding, based on not just the text reviews, but also the other variables in our dataset (cuisine type, location, data, price). This model required our data to be balanced, which we accomplished using random oversampling. This model achieved an 81% validation accuracy. Finally we looked at the most impactful features in our model and were able to identify trends in ratings. Notably nightlife had a negative impact on rating, as did certain locations such as Main Street. Conversely we found restaurants that offered brunch and were lower in price tended to receive higher ratings.

Keywords
Restaurant Reviews, Sentiment Analysis, Review Prediction

1. INTRODUCTION
Operating a restaurant is a challenge anywhere. In fact it has been estimated by the National Restaurant Association that roughly 80% of restaurants fail within their first 5 years. Therefore, it is essential for aspiring entrepreneurs and passionate chefs to be informed of the factors that maximize their chances of success. So what is it that determines the success of a restaurant? Most would be inclined to say the quality of the food, the dining experience, the service, and perhaps even the location or type of cuisine. In a city like Burlington, Vermont, which offers a diverse range of restaurants, we can examine each of these factors to determine which are most important to the general public.

2. RELATED WORK
In the study, "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review" [1]. Researchers looked at customer reviews from food service delivery companies to see if they could find ways to increase customer satisfaction. They used sentiment analysis, machine learning models, deep learning models, and explained artificial intelligence methods to predict customer sentiment. They found DL was the most accurate but lacked explainability. Although this study looked at a different kind of business, the overall objectives were very similar to ours.

Another study, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning" [2] looked at a dataset of 4000 restaurant reviews in a large city in Pakistan. They used sentiment analysis to determine if the review was negative or positive, then using text classification techniques classified the reviews by food taste, ambiance, service, and value. They used several algorithms for classification (including Naive Bayes Classifier, Logistic Regression, Support Vector Machine (SVM), and Random Forest). They found the most success with a random forest algorithm (95% accuracy). Since this study was very similar to ours, (and also used a logistic regression model), it inspired our approach when it came to choosing models.

3. METHODS
For our project we had to create a dataset of Burlington restaurant reviews, identify potentially impactful features, then create a model around these features to draw conclusions from.

3.1 Data Collection and Cleaning
We gathered a comprehensive list of 110 Burlington restaurants and compiled corresponding links to the popular review site Yelp. Using the python package 'Beautiful Soup' we scraped from the first 3 pages of reviews for each restaurant. After we stored the raw data and parsed through it to exact each individual review, along with its general information (date, restaurant location, price, and cuisine type), which we then added to a CSV file. After extensive cleaning, which involved removing leftover JavaScript and HTML tags, as well as correcting misspelled words, we were able to create a dataframe using the Pandas package and begin an initial exploration of the data.

3.2 Data Exploration and Modeling
Using the Pandas dataframe along with the MatPlotLib package we generated some visualizations and tables to examine our data. This included exploring the values that were contained within each variable, their distributions, and the relationship between rating and the other variables.

Using the LogisticRegression and CountVectorizer functions from the SkLearn package as well as a list of stopwords from the NLTK package, we conducted sentiment analysis analyzing the word features in each individual review. We created a new variable 'Customer Satisfaction', which classified reviews as 'good' or bad' based on whether they received a score greater than or equal to 4 stars. Ratings above or at 4 were marked as good, while lower ratings were marked as bad. Then we vectorized the text reviews using the top 100 words, excluding the stop words. Due to the unbalanced nature of our data we computed balanced weights using the class_weight function which were then used when training our model. After dividing our data into training and testing sets with an 80-20 split, we passed our review data into the logistic regression function. We used the vectorized text as the predictor variable and the customer satisfaction as the predicted variable. We included the hyperparameter 'class_weight' to incorporate the balanced class weights. We then evaluated the model, testing it on the validation data and creating a confusion matrix with the confusion_matrix function from Sklearn. Finally we examined the word features that were most and least impactful in our model.

To extend our analysis beyond individual reviews and examine the other variables in our data we created a random forest model that would make overall predictions for restaurants. This meant consolidating our dataframe to have just a single row for each restaurant. We did this by creating a new dataframe with the individual review text concatenated, creating a single entry with all the reviews for each restaurant. Once again, we created a new variable ‘Customer Satisfaction’ dependent on whether the restaurant's average rating exceeded 4 stars. To prepare the data for modeling we performed one hot encoding on the review text, restaurant address, and cuisine type. This was accomplished by vectorizing the data with ‘count_vectorization’ and creating a new column for each feature. In order to address the issue of unbalanced data we used random oversampling via the RandomOverSampler function from the imblearn package. This allowed us to generate a balanced dataset for model creation. We proceeded to train a random forest model on the training data using RandomForestClassifeier from sklearn and assessed its accuracy. Again we produced a confusion matrix using the validation data. In order to understand feature importance, we calculated SHAP values using the SHAP package, which allowed us to identify the most impactful features and their contributions.

4. RESULTS

Overall we were able to identify key trends in Burlington restaurants in relation to their ratings. We achieved successful predictive models for both performing sentiment analysis on individual reviews and predicting overall restaurant ratings based on the variables in our dataset. Furthermore we identified the variables that had the greatest impact on ratings.

4.1 General Exploration Figures

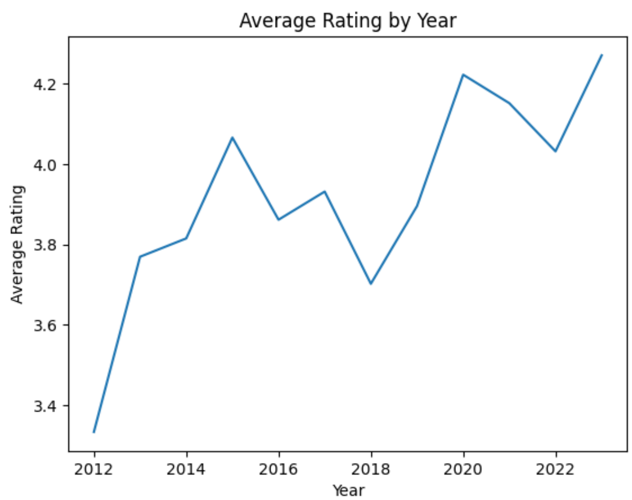

Figure 1: The average rating given for each year present in our dataset. Notice the general upward trend and consider that individuals are either more lenient on ratings or the quality of restaurants has increased.

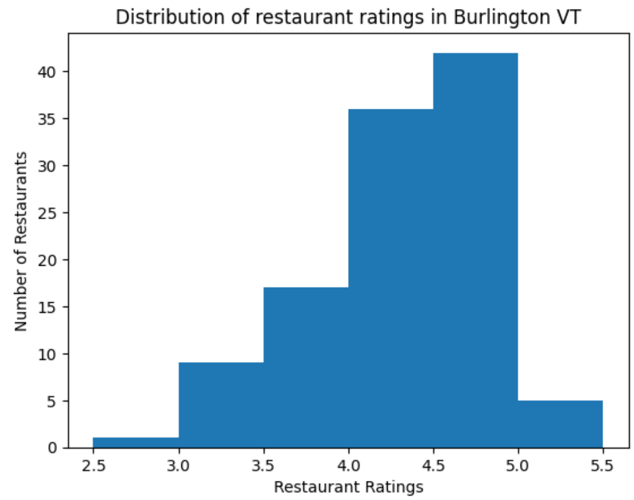

Figure 2: The distribution of restaurant ratings present in our dataset. Notice that most are distributed around the median of 4, with a standard deviation of just .53.

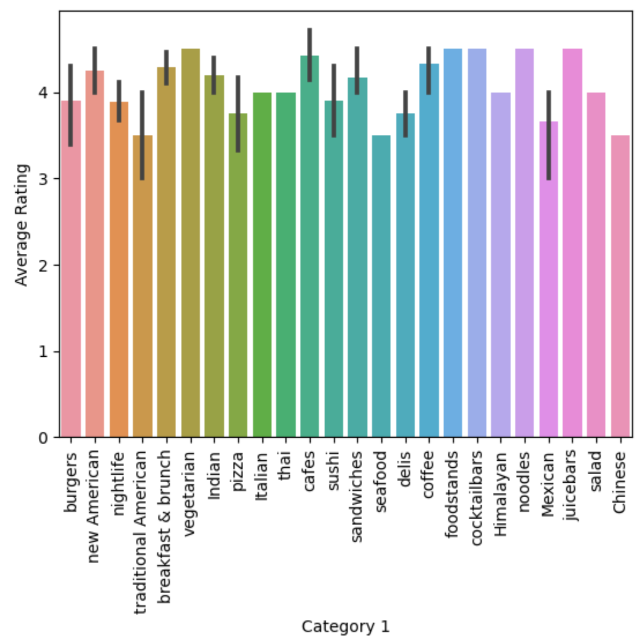

Figure 3: The average rating for each category of food in our dataset. Notice that while most hover around 4, there is still some variation.

4.2 Sentiment Analysis for Reviews with Logistic Regression Model

Our model achieved an 80% training and a 78% validation accuracy for predicting whether an individual review was positive (a rating of 4 or above) or negative (a rating below).

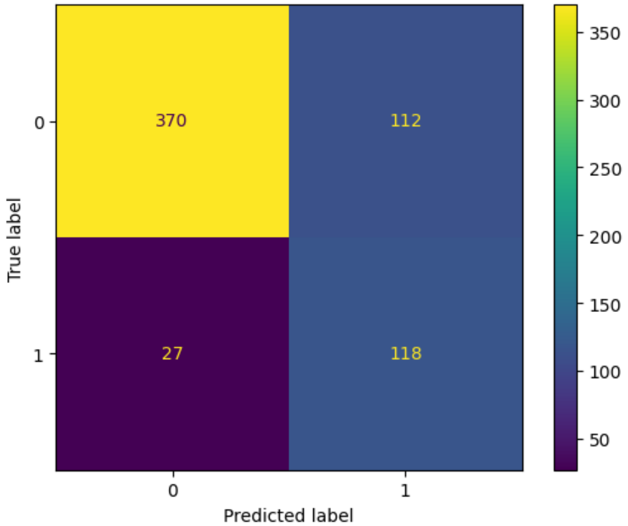

Figure 4: Confusion matrix for our model predictions on the validation data. Zero is a good rating while one is a negative.

	feature	coef
21	delicious	-1.944369
44	highly	-1.585731
2	amazing	-1.454004
31	excellent	-1.184753
48	love	-1.177975
..	...	...
27	even	0.536201
74	said	0.548959
98	went	0.596276
71	restaurant	0.607097
68	pretty	0.650434

Figure 5: The most impactful word features in our model. The negative coefficients correspond to a higher rating, while the positive coefficients correspond to a lower rating.

Using the predictions made for each review in our model, we were able to calculate the proportion of good and bad reviews for each restaurant in our dataset. With these predictions we able to predict the overall rating for the restaurants with 78% accuracy.

4.3 Random Forest Model for Predicting Overall Restaurant Rating

Our random forest model, which was trained on cuisine type, review text, price, and location achieved an 81% validation accuracy in predicting whether a restaurant's average review was positive or negative.

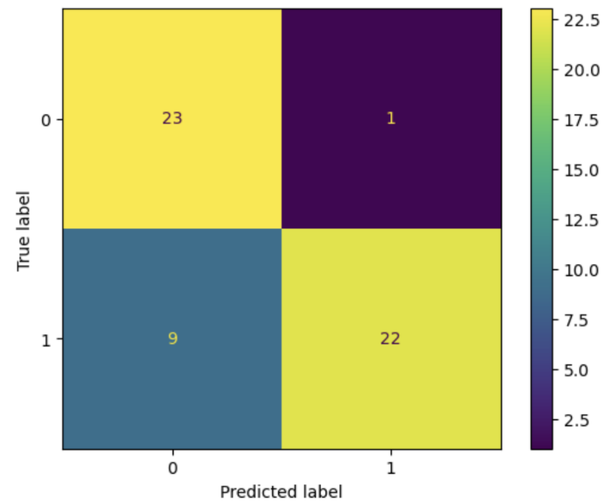

Figure 6: Confusion matrix for predictions made by random forest on validation data. Zero is a negative rating while one is a positive..

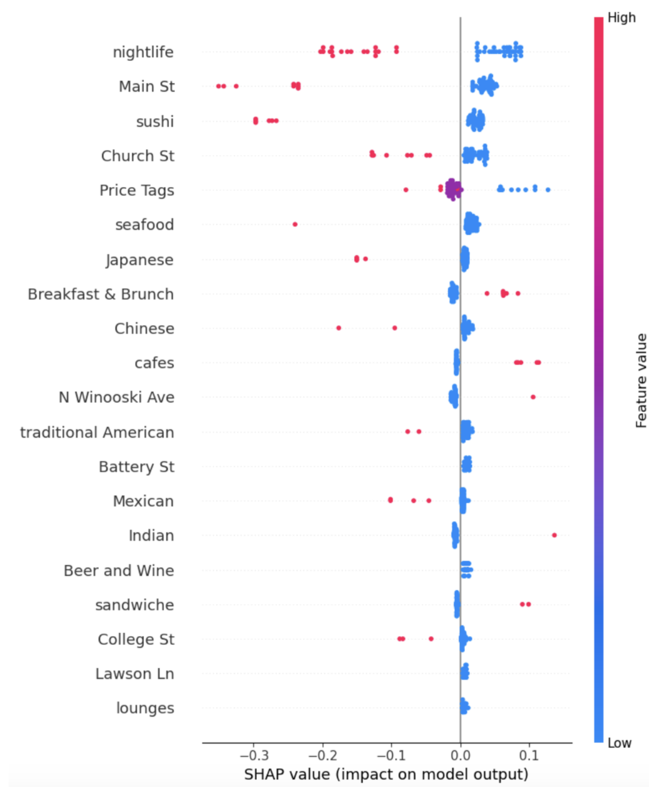

Figure 7: SHAP values for the most impactful features in the random forest model. Positive SHAP values lead to the model predicting that the restaurant will have high customer satisfaction. The red dots represent a restaurant with this feature, while blue dots occur when restaurants do not have this feature.

4.4 Chi-Squared Statistics for Variables

Variable 1	Variable 2	P-Value	Evidence for Relationship
Rating	Location	.465	No
Rating	Price	.22	No
Rating	Review Text	.455	No
Rating	Cuisine Type	.0716	No
Price	Cuisine Type	.04	Yes

5. DISCUSSION

Examining the most impactful features in our logistic regression model, we observe that the top features associated with positive reviews are positive words such as, 'love', 'amazing', and 'delicious'. However, the features most associated with lower reviews are somewhat unexpected. We found words with seemingly neutral meanings, like 'said', 'restaurant', and 'pretty' were the most impactful in predicting a negative rating. This is likely due to the limited size of our dataset. Even though we selected from the the top 100 words in reviews, the data is still relatively small, so a few negative reviews with a higher use of these words can influence our model. As a result, few conclusions can be drawn from these features, although our model is reasonably effective at making predictions.

Our random forest model provided us with better insights. Although our Chi-Squared tests indicated no strong evidence of a relationship between any of our variables and ratings, we see that certain values were impactful on our predictions. Notably, Main Street had a negative contribution when predicting rating, while restaurants with the lowest price associated with them had a positive contribution. Looking at the restaurant categories, keywords like 'brunch' and 'cafe' were positive, while 'nightlife' was negative.

The only relationship we identified through a Chi-Squared test was the relationship between price and location. This makes sense considering some locations are regarded as desirable and associated with higher income parts of the city.

Our biggest limitations likely came from the relatively small size of our dataset. Although we were able to compile over 3000 reviews across 110 restaurants, there are certain restaurant categories that are represented only a few times. As a result it's hard to make meaningful predictions about the success of restaurants on a street when certain streets only have one or two restaurants. Additionally certain categories of business ideally would have been analyzed independently from one another. For example, one of our most impactful features in our random forest model was the negative association with restaurants labeled as 'nightlife', but this just means that bars and clubs tend to get lower reviews than other types of restaurant. However, the lack of data made it hard to compare these sub-categories. Ultimately, certain values that were represented only a few times may have had a disproportionately large impact on our predictions, an issue that could have been avoided with a more extensive dataset.

Another concern comes from the fact that rating does not necessarily predict a restaurant's success. To make conclusions about that we would need to include other variables that are not publicly available, like earnings and years in business. As well as historic values to see what caused past restaurants to fail.

6. FUTURE WORK

Given our limitations it would make sense for further work to be conducted in a location with more restaurants in order to generate a larger dataset. Once a more expansive dataset was created it would be possible to compare restaurants within the same category and make more specific predictions about their success. Furthermore it would be nice to acquire other metrics of measuring restaurant success like earnings and years in business to make up for the limitations of review data.

7. CONCUSSION

Restaurant operation comes with many challenges but also carries the potential of significant success. Although our project was limited to just our own community, it is important for restaurant owners everywhere to understand and consider the factors that give them the best chance of success. In our town it seems like an affordable brunch location might be a good idea if you are looking to get into the business. While our project was small in scope and ultimately limited, it does raise some very real questions that business owners have to consider and has provided some insights into the Burlington restaurant-goer.

8. ACKNOWLEDGMENTS

Many thanks to our Professor Dr. Nick Cheney and TA Csenge Petak for a successful semester.

9. REFERENCES

[1] Adak, Anirban, Biswajeet Pradhan, and Nagesh Shukla. 2022. "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review" Foods 11, no. 10: 1500. https://doi.org/10.3390/foods11101500

[2] K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300098

About the authors:

Caleb Oliveira is a current Data Science undergrad at the University of Vermont, with plans to continue onto a graduate program. In his personal time he is currently trying to obtain a perfect 300 score in Wii Sports Bowling, his current high score stands at a very impressive but also frustrating 298.

Emma O'Brien is currently finishing her junior year in the Data Science B.S. program at the University of Vermont. In her spare time, she enjoys scoring very highly in the game Wii Play Fishing, against all odds. She also enjoys traveling and yoga.