

# Determining Attributes of Good Restaurants Using Review Text

Analysis for  
restauranters/diners

Noah-John Hizon  
11/1/22

# Outline

- Business Problem
- Data Understanding
- Classifier Model
- Classifier Results
- Recommendations
- Next Steps

# Introduction

Has a restaurant ever not lived up to expectations?

- Service
- Food Quality
- Atmosphere

How do you decide between restaurants when dining out?

- Proximity
- Type of cuisine
- Review based
- Word of mouth



# Business Problem

In an industry as sensitive as food service, reviews, whether on a customer level or a professional level can make or break a restaurant in a customer's mind. Owners with failing restaurants may be at a loss when trying to find avenues of improvement.

- Investment in new hardware, staffing, or ingredients may be costly and prove to not be worth it.
- Any combination of factors can lead people to leave a review on a business's page, difficult to isolate relevance from noise



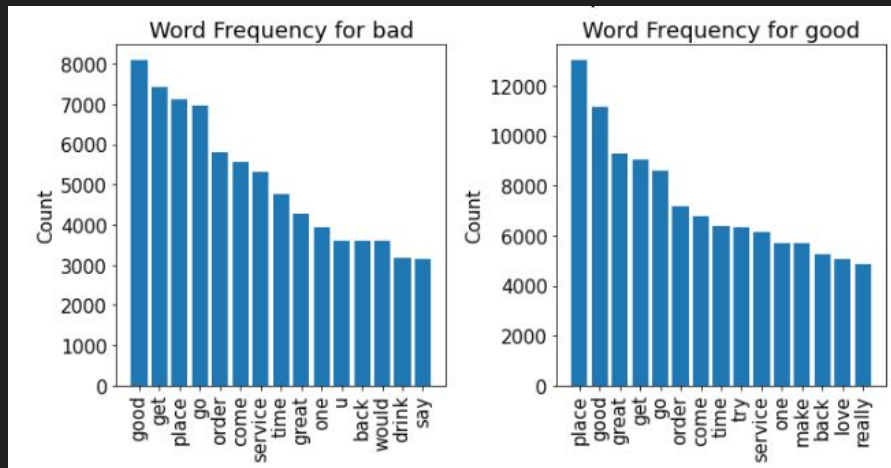
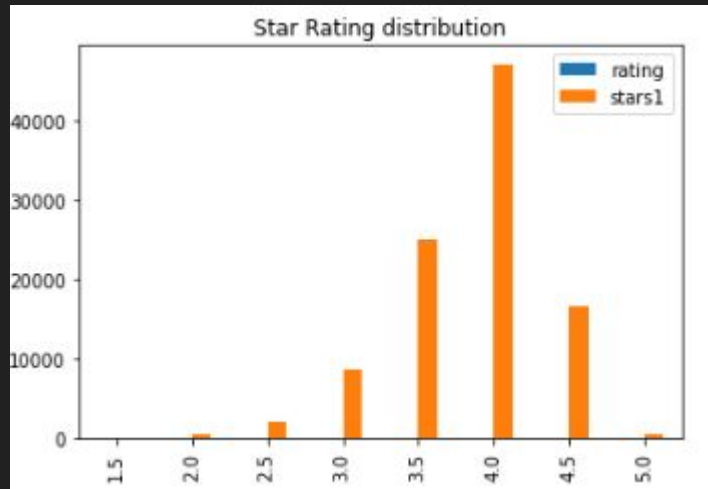
# Data Understanding

**SOURCE:** Yelp restaurant review data spanning 2010-2014 (Kaggle)

- Each row in the table represents a unique review written by a user for a specific business, with text stored in a singular string
- **Columns of interest: 'text', 'rating'**

## Initial Dataset Review Counts:

- ~700K total reviews before filtering
- ~50K users
- ~3,5K businesses



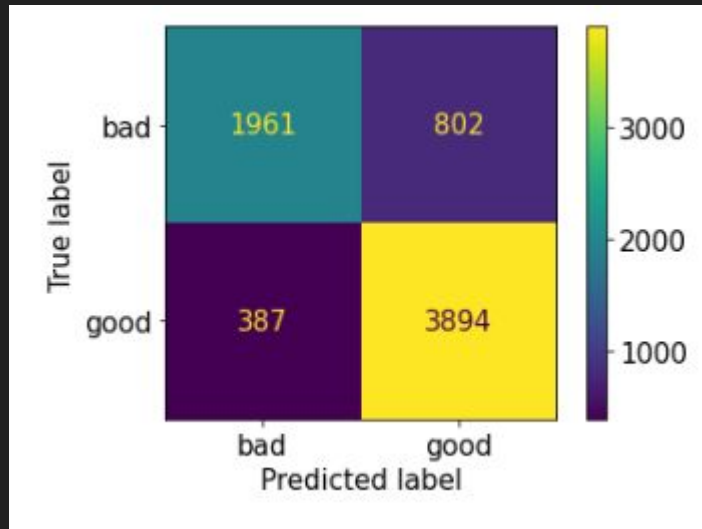
# Modeling

An iterative process of classifier model comparison was used to collectively improve classification a given restaurant review as “good” or “bad”

**Final Model: Multinomial Naive Bayes, TF-IDF Vectorizer**

## Interpretation of Confusion Matrix:

- 3,894 reviews were correctly predicted as Good
- 1,961 reviews were correctly predicted as Bad
- 802 reviews were incorrectly interpreted as Good, when actual review type was Bad
- 387 reviews were incorrectly interpreted as Bad, when actual review type was good



## Confusion Matrix for Final Model

Model was able to classify a good/bad genre with an accuracy of 84%

Model is also relatively accurate in the face of testing data (83%)

# Recommendation System

To supplement the classifier, a recommendation system was created that appends self-reported user ratings and generates a list of recommendations based on the other users/ratings available in the dataset

- Inputs: user ID, number of restaurants to review, category

```
business_id name categories
29764 15FvoKE2-k6JpGVuhH9YVA Sushi Catcher Sushi Bars;Japanese;Restaurants
How do you rate this restaurant/business on a scale of 1-5, press n if you have not been :
2
business_id name \
92481 SPBZxmt8_nT30rNWnKHYKA Akaihana Sushi & Grill

categories
92481 Sushi Bars;Japanese;Restaurants
How do you rate this restaurant/business on a scale of 1-5, press n if you have not been :
1
business_id name categories
35796 6syfIU43FKGkcX2957Ga8A Sekai Sushi Sushi Bars;Japanese;Restaurants
How do you rate this restaurant/business on a scale of 1-5, press n if you have not been :
2
business_id name \
87008 V3ruBXjLGwiniPNPQOzRhiv Makino Restaurant

categories
87008 Sushi Bars;Japanese;Restaurants
How do you rate this restaurant/business on a scale of 1-5, press n if you have not been :
3
business_id name categories
86546 08IFR_ruWR96K3Q6sakI_g Sushi Creek Sushi Bars;Restaurants
How do you rate this restaurant/business on a scale of 1-5, press n if you have not been :
4
```

Out[45]:

| business_id            | categories  | name                       | recreating |
|------------------------|---|----------------------------|------------|
| 6X9iyuM2XdoCGT4q9qv5cA | Sushi Bars;Japanese;Restaurants                   | JJANGA Japanese Restaurant | 4.492273   |
| OwBPjUz2o0J5K3DzcHkBg  | Sushi Bars;Japanese;Restaurants                   | Soho Japanese Restaurant   | 4.491927   |
| HpaYCM_NCaul72LLXcC6SA | Tapas/Small Plates;Sushi Bars;Japanese;Restaur... | Yonaka Modern Japanese     | 4.489452   |
| sNBqulTaV3lbUWkzSUITpw | Sushi Bars;Japanese;Restaurants                   | Sakana                     | 4.489281   |
| Pgp3gbOQaJldyjqC9AOz6g | Sushi Bars;Restaurants                            | I Love Sushi               | 4.489053   |
| bpGFetX8muk0GxAT3Oea3Q | Sushi Bars;Japanese;Restaurants                   | Yummy Grill & Sushi        | 4.485786   |
| u6EUXOSFnxvzLI4D21bA   | Sushi Bars;Japanese;Restaurants                   | Harumi Sushi               | 4.485590   |
| i5Tq_jzm3osbY_uEJ79W1g | Seafood;Sushi Bars;Japanese;Restaurants           | Tokyo Boys                 | 4.480236   |
| QrRNFiSXmCo4pzVLAhA_bw | Steakhouses;Sushi Bars;Japanese;Restaurants       | Kanji Steak & Sushi        | 4.479729   |
| AYbUb5UcAngroJ6uJG7tLQ | Sushi Bars;Japanese;Restaurants                   | Kabuto                     | 4.475067   |

Recommendation System and  
Output



# Conclusion/Recommendations

From the model results, it can be seen that rating type (bad/good) can be predicted successfully using review text, with high accuracy in the face of new (test) data. Words/phrases that were most impactful for the performance of the overall classifier identified a restaurant's features and experiences, both good and bad.

- Value perceived by customer
- Service quality
- Food quality
- Time to receive food

## Recommendations for Improvement:

- Establish a French brigade system
- Establish a happy working culture and fair front-of-house compensation
- Optimize food spend while maintaining quality





## Next Steps

- More nuanced system of rating, similar to metacritic
- Utilize more filtering options in the recommendation system to mimic Yelp filters
- Create word cloud distributions for each restaurant recommendation for more user context
- Provide additional color to potential restaurant improvements, using additional feature flags provided within Yelp's dataset
- Replicate analysis for different restaurant category types; different kinds of restaurants may require different solutions

# Thank You!

Email: [noahiz917@gmail.com](mailto:noahiz917@gmail.com)

Github: noahiz917

LinkedIn: [linkedin.com/in/noah-john-hizon/](https://www.linkedin.com/in/noah-john-hizon/)