Leveraging Customer Generated Data on Social Media to Drive Business Solutions

Team: Yelp to the Help - *Benliang He, Harshith Byregowda Settyhalli, Suhas Alur and Tina Lu*

Business Understanding

Opening a restaurant is among one of the most common businesses many entrepreneurs choose to pursue. The idea of creating a unique experience where customers can indulge in a wide variety of cuisines is seen as both an exciting and appealing business venture to many. However, the restaurant business is notorious for being a highly competitive industry, making it extremely difficult for restaurant entrepreneurs to stay in business. According to a study about restaurant failure conducted by Ohio State University, 60% of restaurants do not remain in business after the first year and 80% of restaurants go under in five years (Feloni, 2014).

In order to be successful in the restaurant industry, it is critical for both prospective and current business owners to better understand what attributes contribute to the success of operating a restaurant. According to an article written by restaurant owner Hester Lacey, different factors such as quality of, cleanliness of restaurant, customer service, creating unique experiences, understanding and meeting market needs, etc. (Lacey, 2014). With such ambiguous definitions and meanings of successful attributes. Because of this, it is nearly impossible to explicitly identify what exact attributes drive the success or detriment of a restaurant. Most importantly, customer experiences at restaurants vary depending on countless different variables. However, there are many different methods of narrowing down what factors drive the success of restaurants directly from customer opinions.

With the increase of social media usage and online platforms, business owners now have the opportunity to utilize consumer generated data to understand how their businesses are perceived by customers. According to a survey conducted by the Social Times, 26% of Internet users like to voice their opinions on social media and 49% of Internet users share information about products or services in order to impact opinions about businesses and encourage action (Bennett, 2014). Consumers today are voluntarily writing reviews online and voicing their opinions of various restaurant on online platforms such as Facebook, Twitter, TripAdvisor, Google My Business, blogs and Yelp. Online reviews present a great opportunity for many restaurant owners because reviews are free sources for receiving direct and authentic feedback from customers.

We are utilizing Yelp reviews and ratings to understand the sentiments that drive good and bad ratings on restaurants. We then perform sentiment analysis on the customer generated data on social media to predict ratings customers would have given to restaurants. In result, we can better understand customers in order to help determine factors that drive success in the restaurant industry.

By applying data mining techniques to online reviews, restaurants can identify which words are most commonly used in reviews that are given high ratings and low ratings. These words can then be used to form common topics among reviews in different locations, which will then help business owners narrow down what topics are perceived positively when used in reviews and which are seen as unfavorable when used in reviews.

Ultimately, the sentiment scores of various words used in consumer generated social media data can be analyzed in order to predict what rating a review will be given to identify business features that are most indicative of success and can help restaurants devise sensible strategies to improve their own ratings.

Yelp is one of the most popular and well-known online sites for customers to rate and review local businesses. Users are given the option of rating businesses from a predefined scale of 1 to 5 stars and also have the opportunity to express their experience in words through open-ended text reviews. Many businesses often times organize their own listings on the Yelp website in order to facilitate conversations around their restaurants.

This data mining solution can be used to predict and recognize “friendly users,” which are people who tend to give good ratings in their reviews (average star >=4). Distinguishing this population can potentially be useful given that businesses will able to effectively target a positively influential market of customers. Most importantly, this data mining solution can help us understand the necessary features required to start a successful restaurant in a specific location.

Problem Statement

Utilize reviews to identify the sentiment of commonly used words in reviews to understand main drivers of good and bad ratings for a particular business in order to generate business strategies.

Case Study

By applying data mining techniques to the Yelp reviews dataset, businesses can leverage online textual data from social media sources to gain customer insights regarding their restaurants. Take for example the analytics start up, “Data Diggers.” The analytics firm specializes in collecting social media data and performing advanced data mining methods to gain insights for prospective businesses. Data Diggers was recently approached by a new restaurant, John Doe’s Restaurant. The new restaurant was facing issues with retaining customers and in result was losing business. Mr. Doe found that he was receiving a lot of negative reviews on Yelp, but did not have the time nor resources to read each individual review to understand why his customers were not enjoying their experiences. Most importantly, Mr. Doe was most interested in understanding the different factors that make a restaurant successful.

The Data Diggers team decided to help Mr. Doe by analyzing Yelp reviews to build a predictive model in order to see how different words used affected the number of stars rated for a review. The team first pulled data from Yelp and preprocessed the data in order to format it in a usable way. The team then calculated the sentiment scores for the words in the reviews based on their frequency of occurrence with extreme scores assigned to words that occur mostly in positive and negative words and neutral scores assigned to words mostly occurring neutral reviews. Afterwards, they then assigned each review an overall sentiment score, which was based on the sentiment scores of the words in that review. Afterwards, they then were able to obtain a training model for the Yelp dataset. They then used this same model to create a test model in order to apply to other reviews. The team found that their model performed with an exceptional predictive accuracy score for Yelp star ratings.

By utilizing free data sources, primarily Yelp reviews, the Data Diggers were able to find which words were used in positively rated reviews and negative words associated with poorly rated reviews. The team found from the negatively rated reviews that customers experienced very poor service. Mr. Doe had experienced a stint in time where his restaurant had high employee turnover rates, which then led to lack of stack. Customers then experienced long waiting times to be seated and served, which led to frustrated customers. The negative words used in reviews describing the service included words such as slow, sluggish, bad, untimely, undesirable, lagged, ignore, awful, failed, faulty, horrible, unpleasant, etc.

The Data Diggers also found positive insights about the quality of the food at Mr. Doe's Restaurant. From reviews that received high ratings of 3.5+ stars, the team found words used to predict high ratings pertained to the great food. Positive words used in highly rated reviews included words such as food, cuisine, dishes, spicy, comforting, delicious, yummy, good, wonderful, tasty, super, curry, spices, amazing, creative, original, authentic, etc.

With the insights found from the analysis the Data Diggers performed, Mr. Doe was then able to understand why his restaurant was receiving such poor reviews and what specific factors contributed to the success or lack of success of his restaurant. He was then able to take action and respond to the demand for more staff members in order to reduce the waiting time for customers and therefore improve the quality of service given at his restaurant. With the insights found from the data mining analysis, Mr. Doe was also able to capitalize on his well-liked food and could then advertise his food to customers similar to those who gave high ratings.

Being an ambitious restaurant owner, Mr. Doe also decided to use the insights found from the Data Digger's analysis to open a second restaurant in a new location. Now knowing what customers are posting about in bad reviews, Mr. Doe can use this information to improve different factors that customers have written poor reviews about. He can also use the information gained from the analysis to understand different locations receive different reviews due to the diverse clientele presented in different parts of the country. Using the same data mining analysis methods, the Data Diggers found that positive reviews in the Las Vegas area focused on the atmosphere and ambience of restaurants. Customers wrote more positive reviews about their fun and entertaining experiences at restaurants in the Las Vegas area and wrote bad reviews regarding lack of uniqueness and distinctiveness.

The Data Diggers also used the same predictive sentiment analysis methodology to predict what social media users rated Mr. Doe's Restaurant on other various social media platforms, such as Twitter. The team searched Twitter for the hashtag "#MrDoesRestaurant," and was then able to scrape Twitter data use Twitter APIs. From there, they obtained a document of Tweets that mentions the restaurant and calculated the sentiment score associated with each tweet. They then applied the predictive model already developed to predict the rating that could be assigned to each tweet based on its sentiment score. This allows Mr. Doe to gain insights from not just Yelp reviews, but from open-ended Tweets generated by social media uses as well. This allowed Mr. Doe to obtain even more information regarding consumer sentiment and opinions about his restaurant, and was then able to reactively improve his restaurant based on the candid and honest feedback posted online.

Thanks to the Data Diggers, Mr. Doe was able to turn the business of his restaurant around and even open up a second location that was very successful. By using Yelp data to build a predictive model, the team was able to predict ratings of reviews and use the same model to predict the overall sentiment and infer ratings from social media data. Mr. Doe was most satisfied with the fact that the data and information extracted from social media was free, and that he only had to pay the labor fee to the analytics team. Utilizing Yelp and Twitter data was free of charge and can be used by any restaurant to determine sentiment expressed about the restaurant by the customers.

Data Understanding

The data that we have used in our predictive data mining solution is a collection of Yelp reviews from the Yelp Dataset Challenge. Within the Yelp dataset, we have five separate data files; check-in data, business data, tip data, user data and review data. Because the scope of our data mining solution only requires the reviews as well as information regarding restaurants reviews are written about, we have combined the business data and Yelp review data and are only using these two data sets. By selecting a subset of data sources for our project, we are able to reduce the amount of data used in our analysis. This allows us to select only relevant attributes needed and effectively utilize our data.

Given that the business dataset contains information about over 86,000 businesses and the review dataset contains over 2.7 million reviews, we have chosen a sample of the data to build our predictive model. In order to reduce the size of the dataset and complexity of our analysis, we are only considering restaurants in the United States that are present in the dataset. However, after performing exploratory data analysis, we found that we could further reduce the density of our dataset by analyzing reviews for a specific location within the United States. We discovered that most of the data in the dataset was comprised of reviews for restaurants in the state of Nevada. This is due to the fact that the state of Nevada is highly influenced by the popular tourist destination, Las Vegas. Thus, we have decided to restrict our analysis to the state of Nevada and in particular to the city of Las Vegas.

A single data instance is at a review level, which in our case is for a specific restaurant in the state of Nevada in the city of Las Vegas. For each review, we have the raw text of the review, a unique identifier for each restaurant, the address of the restaurant, hours of operation, restaurant category, city, state, latitude and longitude of the location, review count, average price range of the cost of a meal, dress attire, a binary category for restaurant features and number of stars a restaurant received. For the purpose of our analysis, we only worked with a subset of attributes which included business ids, the raw text reviews and the star ratings given to the restaurants.

The Yelp reviews that we worked with were raw text reviews in the form of strings that were sourced json files from the Yelp Dataset Challenge. The reviews were in no given format and included both uppercase and lowercase letters, numbers, symbols and characters in various languages. Our target variable that we predicted was the number of stars for a given review, which was in the form of a numerical value.

Data Preparation

In order to produce the format required for our data mining problem, we followed a number of practices in order to prepare our data. First and foremost, because our data was in the form of unstructured text, we first converted the symbol "!" to the word "exclamation." We did this in order to preserve the sentimental value associated with the symbol, which can be used to help determine the overall sentiment of a review. Next, we converted all of the text in the review documents to lowercase letters. This allowed us to work with consistent data throughout our analysis. We also removed any stop words that would be used in the reviews and this reduced the number of features that our final data set had.

Our next step we took in preparing our data was remove star ratings that had no value, otherwise known as a value of "0." Because the purpose of our data mining analysis was to predict star ratings of Yelp reviews, having reviews with an omitted rating would skew our analysis, which in result would jeopardize the integrity of our analysis. Therefore, in order to eliminate reviews with a star rating of 0, we first replaced the scores of 0 with "NA." We found that there were 94 Yelp reviews in our dataset with no star reviews. Finally, we removed the reviews with a rating of "NA."

After cleaning the data and formatting the reviews, we were then able to further prepare the data for data mining by creating document-term matrices. This was done to ultimately determine the frequency of words used in different both good and bad reviews. We accomplished this by using the "sklearn" library in Python, where we were then able to use the "CountVectorizer" function. We were then able to reformat individual reviews into matrices that consisted of words that appeared in a given review and the number of times that word appeared in the document.

Lastly, in order to format our data into the required form for our analysis, we normalized the star ratings. Because our target variable which was the number of stars given for a review is in the form of a numeric value, we normalized the star rating value for good practice. This allowed us to create a unit norm for our target variable that was then able to interpret the number of stars a review was given.

Modeling

Before we built the model, we had to process the data to be better analyzed using predictive models. Since we were dealing with textual data, which are the reviews from the customers on Yelp website, we carried out the following processing techniques to transform the textual data:

1. Transforming the textual data to lowercase
2. Removing punctuation marks such as exclamation, periods, commas and so on
3. We have not removed stop words such as “not” as they too provide sentiment to a review (negative sentiment)

Once the above mentioned processing has been performed, we then transformed the textual data into a review v/s words kind of matrix, wherein all the reviews under consideration were taken as rows and all the words in the corpus (combination of all the reviews) were taken as columns. Since, there would be a lot of words in the corpus, we decided to take only those words which had an occurrence of 100 words and above. The cells of the matrix now consist of occurrence of words in each review as a binary, 0 for if the word was present to 1 if the word was absent as shown in Table 1.

Table 1

	fantastic	fabulous	tasty	wonderful	yummy	good	...
Review 1	1	0	0	1	0	0	...
Review 2	0	0	1	0	1	0	...
...	...	...	...	...	...	...	...

Based on the above table, we then calculated the frequency of word occurrence across all the reviews along with the ratings of the reviews. Thus, the transformed matrix looked something similar to the one shown in table 2.

Table 2

fantastic	fabulous	tasty	wonderful	yummy	good	...	Rating
23	0	0	1,000	0	2,547	...	5
0	0	69	0	459	0	...	4
...	...	...	...	...	...	...	5

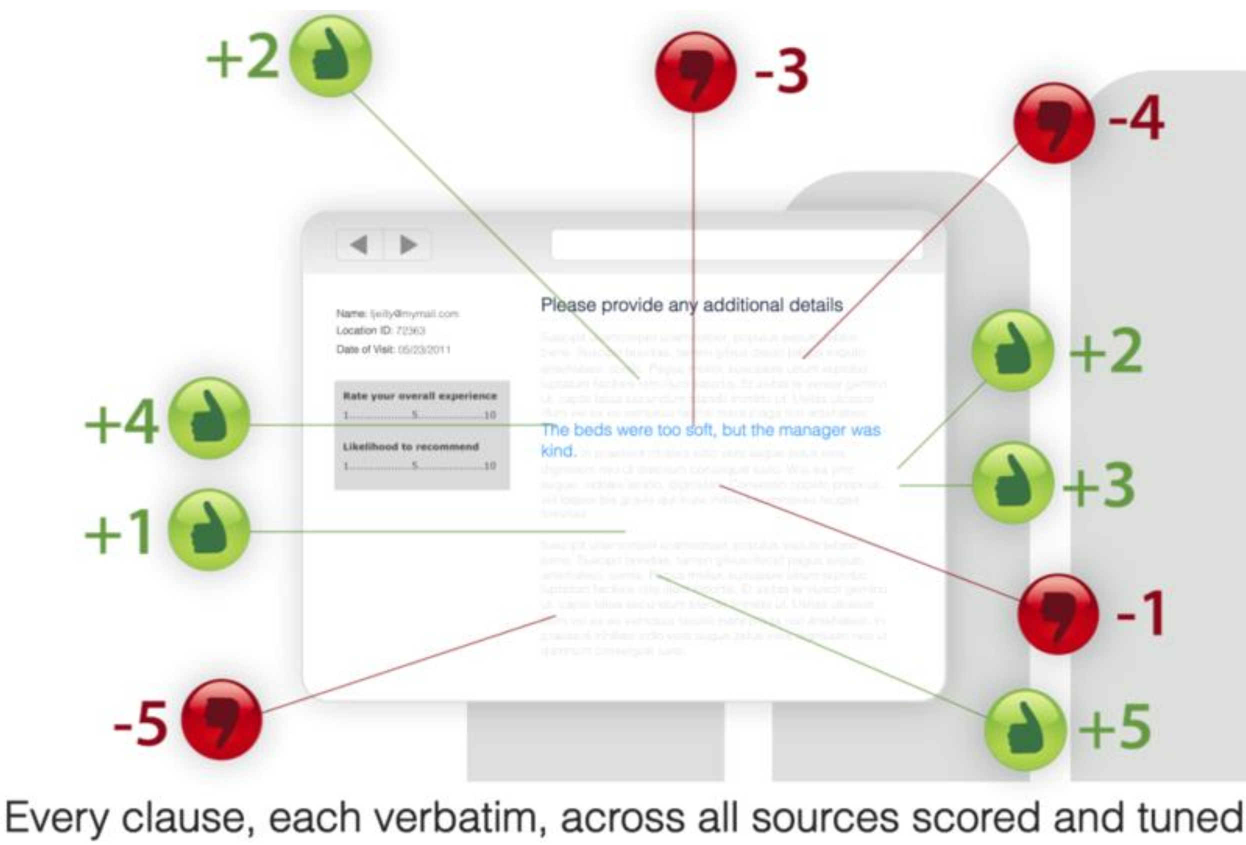
Calculating sentiment of the words

Then, for each word, we multiplied its percentage frequency in each document by the number of stars the reviewer ultimately assigned to the business (normalized to [-1, 1]), and summed these weighted frequencies across all documents to arrive at the word's sentiment score.

As an example, say the word "good" shows up 5 times in a 5-star review, 4 times in a 4-star review, ..., and 1 time in a 1-star review, its sentiment score would be (5*1 + 4*0.5 + 3*0 + 2*(-0.5) + 1 * (-1)) / (5 + 4 + 3 + 2 + 1) = 0.33, indicating moderately positive sentiment.

As a couple extreme examples, if a word only appears in 5-star reviews, its sentiment score would be 1, and if it only appears in 1-star reviews, its score would be -1. What about highly frequent words that don't mean much (e.g., stop words)? As shown in the first example, showing up everywhere will dilute a word's overall effect. In an extreme example, if a word's usage is evenly distributed across all documents, its final score would be 0, indicating absolutely neutral effect.

Once, we calculated sentiment of all the words in the review, we then aggregated this data to get an average sentiment score for each review in our training dataset. A gist of the process is shown in the diagram below:



Source: https://img.revinate.com/image/upload/v1381744550/6a013483a348fd970c017ee3ce3853970d-800wi_hmitrg.png

Once the pre-processing was completed, we then had the following problem at hand:

1. Should we train the data for binary target variable - good rating (>3.5 rating) or bad rating (<=3.5 rating)

2. Should we use multi-class prediction to predict ratings from 1 to 5

We decided on trying both the methods and understand which gives a better performance. According to our business problem, understanding the customer sentiment was the primary focus, even if we were off by a rating of 0.5 would not influence our customer sentiment inference by much.

Models used to fit the training data:

Basic Models

1. Logistic regression

2. Logistic regression with regularization

3. Naive Bayes

4. Support Vector Machine

Ensemble Methods

Bagging Classifiers

- Bagging using decision trees

Boosting

- Gradient boosting

Since we were dealing with multiclass classification problem with numeric attributes, models such as logistic regression, support vector machine and naïve bayes work pretty well.

Naive Bayes

1. Extremely simple: If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models like logistic regression

2. Works well on a smaller dataset

3. If want something fast and easy that performs pretty well, Naive Bayes is the answer

4. Also, its main disadvantage is that it can’t learn interactions between features (e.g., it can’t learn that although you love movies with Brad Pitt and Tom Cruise, you hate movies where they’re together), but since we are not dealing with multiple features, this should not be an issue for us

Logistic Regression

1. It provides us with a lot of flexibility to regularize our model, and we don’t have to worry as much about our features being correlated, like we do for Naive Bayes

2. Also we have a nice probabilistic interpretation, unlike decision trees or SVMs, and we can easily update our model to take in new data (using an online gradient descent method)

Support Vector Machine

1. The key features of SVMs are the use of kernels

2. Absence of local minima

3. The sparseness of the solution

4. The capacity control obtained by optimizing the margin

5. Learning result is more robust

Bagging Classifier

1. Given that we only have one input variable; it doesn't quite make sense to use random forests since its random feature selection is of no use here. Without it, it is effectively reduced to simple bagging trees

2. But, we also tried using random forest model, just to see what accuracy it provides us with

Gradient boosting trees

One more ensemble tree-based model that we used was gradient boosting trees. Unlike bagging trees or random forests which focus on reducing variance using bootstrapping, gradient boost trees sequentially try to reduce bias using modified data (e.g., residuals). Thus, we expect gradient boosting trees to perform better than rest of the models.

These models are useful for the business for the following reasons:

1. Easy to train the models once the pre-processing of the textual data has been performed

2. Can be applied across any form of textual data, such as tweets, posts, blogs etc.

3. These models perform better for a binary classification, such as a good sentiment or a bad sentiment as compared to multi class classification, thus, if the business is interested in understanding just the overall sentiment of the customers, these models would provide a pretty good accuracy

Evaluation

For the multi class classification/prediction, we have considered our target variable to be numeric and predicted the numeric score, that is rating, using various models. We can look at the accuracy and confusion matrix, but instead of ROC curve, we can also look at numeric prediction evaluation metrics such as:

Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |p_i - a_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

Mean Absolute Percentage Error

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{p_i - a_i}{a_i} \right| = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{e_i}{a_i} \right|$$

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i)^2}$$

Having a MAE (mean absolute error) of 0.5 or less, would mean predicting rating within a range of +/- 0.5 of actual rating. This should be a good indicator of overall sentiment of the review being analyzed. This is because, having a review with a rating of 4 or 4.5 would both fall in similar category of a positive sentiment.

Based, on this, our models have the following performance metrics:

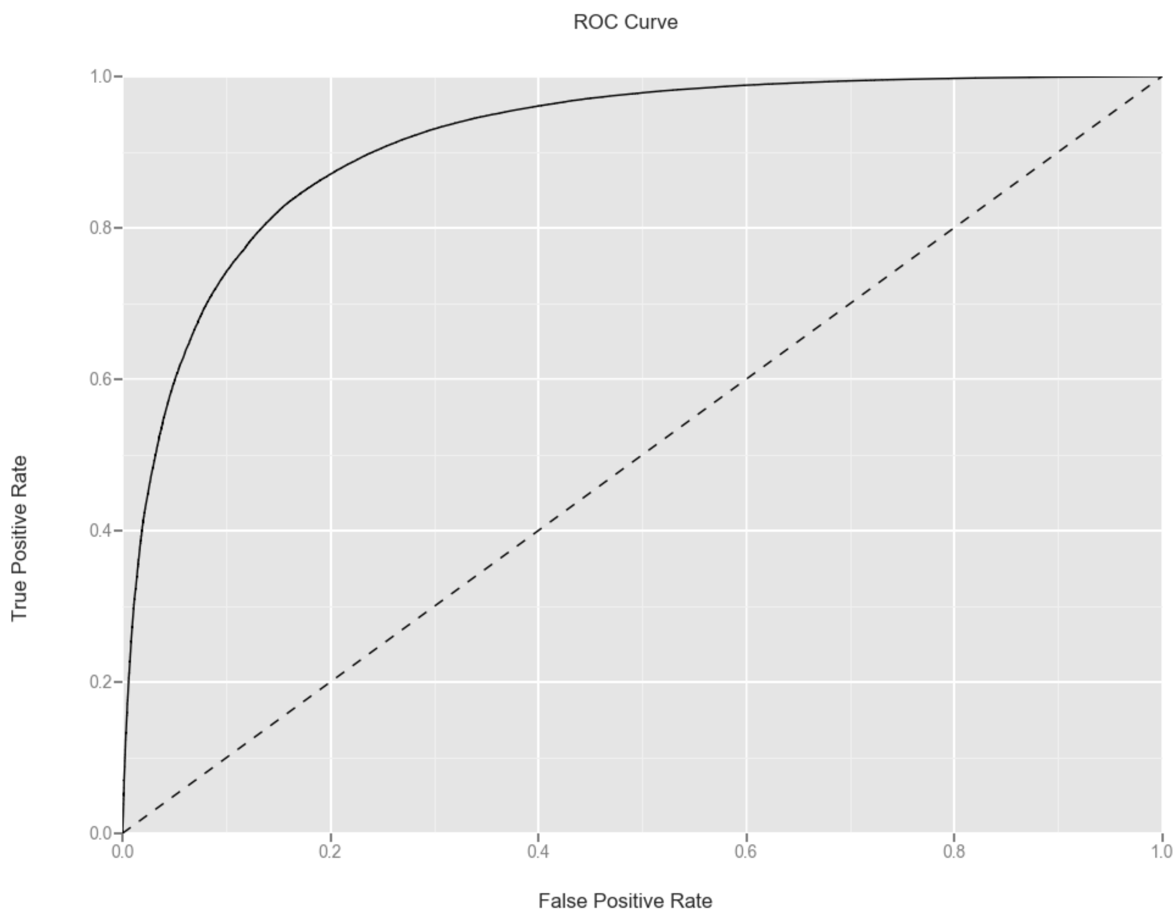
Model	Accuracy	MAE	MAPE	RMSE
Logistic Regression	51.59%	0.630	26.81%	0.993
Logistic Regression – l1	51.56%	0.631	26.74%	0.993
Bagging Classifier	42.92%	0.782	29.27%	1.149
Gradient Boosting Classifier	52.26%	0.595	22.22%	0.935

Next we transform our target variable into binary: 1 as positive for ratings greater than 3.5, and 0 as negative for lower ratings. Our models have much better performance for binary classification.

Logistics Regression:

The accuracy score: 83.93%

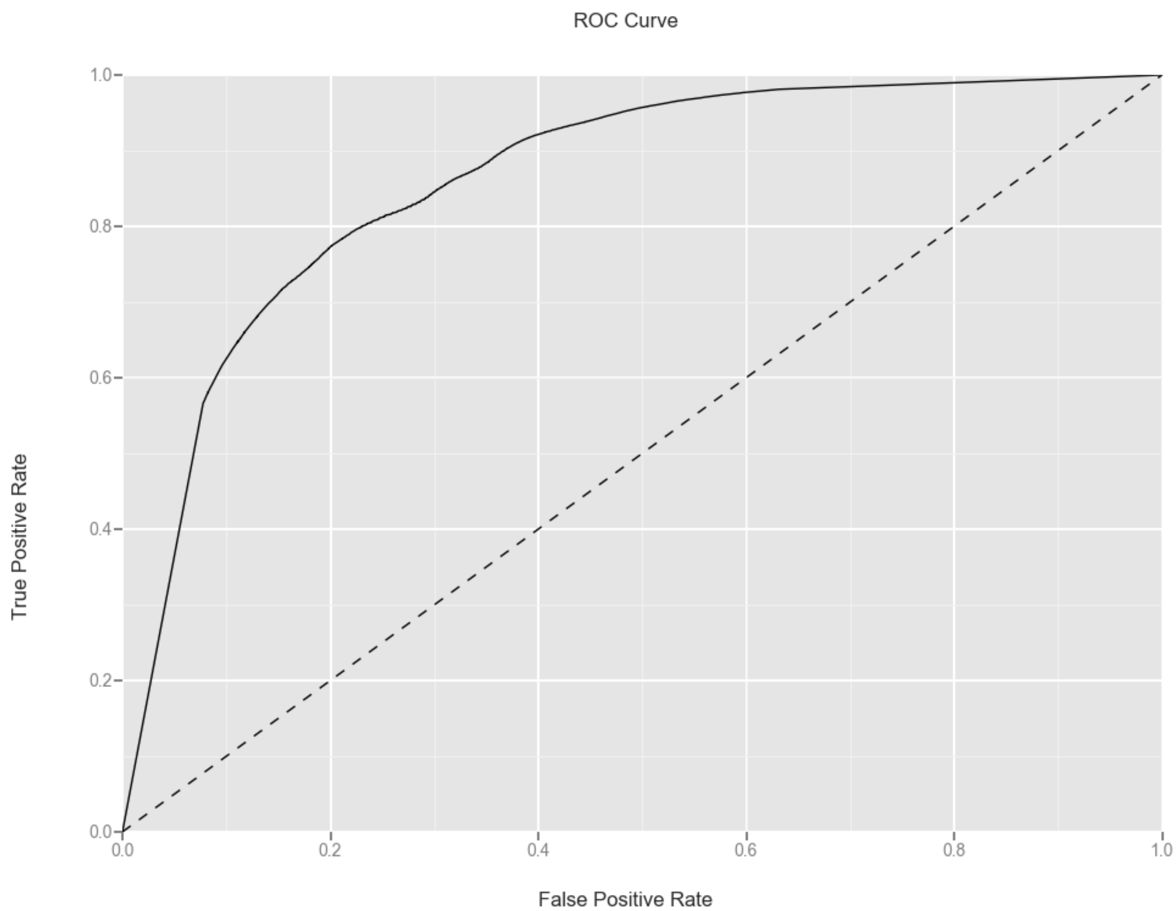
	Precision	Recall	F-measure	Support
Negative Rating (0)	0.72	0.83	0.77	71665
Positive Rating (1)	0.91	0.84	0.88	144689
Average / Total	0.85	0.84	0.84	216354



Bagging Classifier

The accuracy score: 80.13%

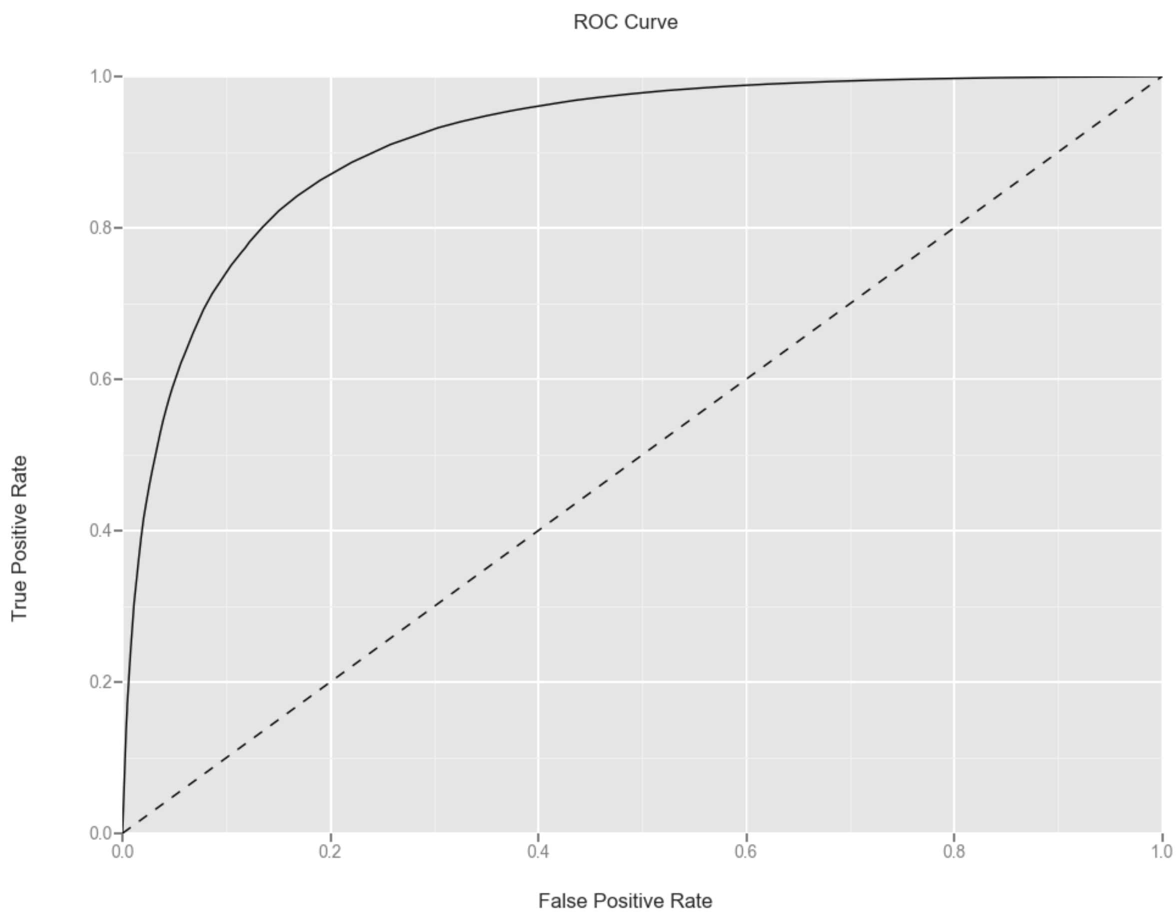
	Precision	Recall	F-measure	Support
Negative Rating (0)	0.65	0.73	0.68	71665
Positive Rating (1)	0.86	0.80	0.83	144689
Average / Total	0.79	0.78	0.78	216354



Gradient Boosting Trees

The accuracy score: 85.45%

	Precision	Recall	F-measure	Support
Negative Rating (0)	0.80	0.74	0.77	71665
Positive Rating (1)	0.88	0.91	0.89	144689
Average / Total	0.85	0.85	0.85	216354

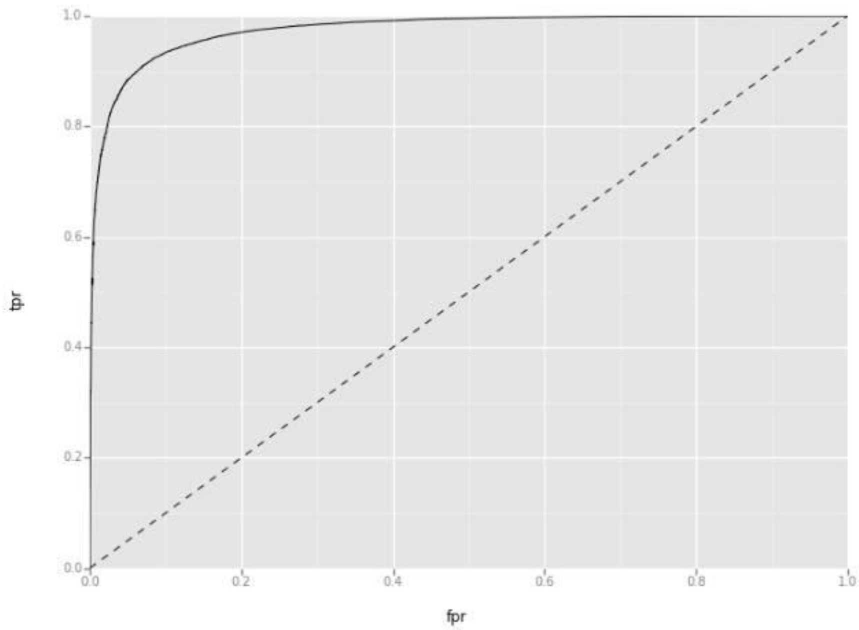


Support Vector Machine

The accuracy score: 85.62%

	Precision	Recall	F-measure
Overall	0.83	0.99	0.90

<u>ROC Curve</u>

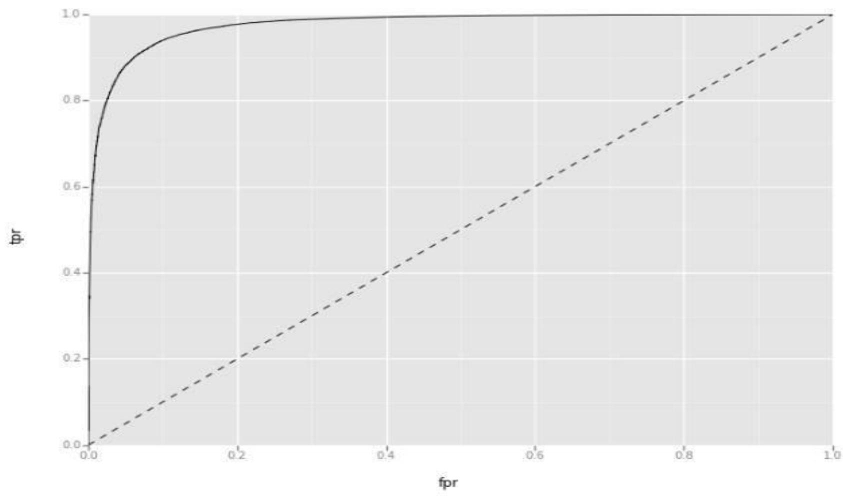


Naïve Bayes

The accuracy score: 91.36%

	Precision	Recall	F-measure
Overall	0.89	0.98	0.93

ROC Curve



Topic Modeling

What do people tend to write about in their reviews?

- *Topics to predict rating of reviews*

Since, we were dealing with textual data, we also decided to try out topic modeling and observe how well the topics of a review might be useful for predicting the rating of a review.

For this, similar to our previous pre-processing, we performed the following pre-processing steps:

1. Consider reviews with words greater than 100
2. Create a corpus of words from the review data from Yelp
3. Lowercase all the words in the corpus
4. Remove punctuations
5. Remove stop words
6. Build vectorized corpus
7. Apply linear dirichlet model to obtain topics, we have considered a total of 30 topics

Topics and their corresponding words

Topic 1: tacos, chips, mexican, taco, salsa, burrito, beans, nachos, asada, ordered

Topic 2: sauce, salad, flavor, fresh, dish, garlic, bread, served, cooked, tomato

Topic 3: sushi, roll, fish, fresh, rolls, ayce, tuna, eat, favorite, salmon

Topic 4: stars, service, star, staff, experience, server, friendly, five, attentive, reason

Topic 5: chicken, wings, fried, sauce, ordered, juice, buffalo, fries, hot, pickles

Topic 6: order, time, location, times, last, every, drive, first, still, few

Topic 7: smell, par, reviews, sub, donuts, star, cleanliness, bathrooms, alcoholic, resort

Topic 8: de, la, groupon, wynn, y, spoon, el, aria, sooooo, planet

Topic 9: service, amazing, best, definitely, delicious, vegas, time, recommend, everything, love

Topic 10: pizza, cheese, crust, slice, new, toppings, ordered, pizzas, delivery, sauce

Topic 11: lunch, service, happy, prices, hour, strip, quick, pretty, vegas, night

Topic 12: rice, chinese, japanese, egg, dishes, rolls, asian, fried, spring, hawaiian

Topic 13: friendly, staff, love, music, clean, vegan, service, super, new, live

Topic 14: yourself, favor, bc, cherry, farm, mr, met, venue, killer, flight

Topic 15: bar, beer, bartender, drinks, drink, beers, fun, game, selection, bartenders

Topic 16: manager, service, customer, asked, told, experience, called, rude, customers, business

Topic 17: buffet, price, better, quality, bad, worth, pretty, nothing, eat, money

Topic 18: service, minutes, time, order, wait, table, took, ordered, asked, server

Topic 19: lobster, old, year, wife, kids, son, last, night, daughter, server

Topic 20: hot, dog, taste, cold, dogs, wall, chili, ridiculous, eat, hole

Topic 21: soup, pho, noodles, bowl, broth, noodle, beef, spicy, tofu, meat

Topic 22: breakfast, coffee, eggs, bacon, brunch, toast, pancakes, french, egg, hash

Topic 23: burger, fries, burgers, cheese, ramen, shake, bacon, onion, ordered, truffle

Topic 24: sandwich, bread, sandwiches, cheese, turkey, pie, salad, pot, beef, slaw

Topic 25: dinner, wine, menu, dessert, meal, night, experience, ordered, table, dining

Topic 26: healthy, thumbs, god, massive, tourists, diet, training, fake, bowls, yuck

Topic 27: pork, rice, chicken, thai, fried, bbq, beef, sauce, ordered, tea

Topic 28: vegas, las, best, crepe, find, smoothie, better, city, trip, high

Topic 29: area, outside, inside, seating, menu, bar, tables, sit, patio, dining

Topic 30: steak, cooked, ordered, potatoes, medium, rib, salad, filet, rare, sides

Result of using logistic regression on the topics of reviews in order to obtain the prediction of the rating. For this, we used a binary prediction, using 5 rating as the positive sentiment and 1 rating as the negative sentiment.

We surprisingly obtained an overall accuracy of close to 89%. The result of the analysis has been provided below:

```
Result

Logistic rgression:

Call:  glm(formula = Star ~ ., data = topic_dist_train_1_5_df)

Coefficients:
(Intercept)       Topic1       Topic2       Topic3       Topic4       Topi
c5
    -0.4852       3.9029       3.6205       5.8495       4.3417       6.40
65
     Topic6       Topic7       Topic8       Topic9      Topic10      Topic
11
     5.6346       4.1862       1.1459       2.3335       2.9807       6.48
71
    Topic12      Topic13      Topic14      Topic15      Topic16      Topic
17
     1.2850       3.5980       3.5279       4.7605       7.6908       7.20
72
    Topic18      Topic19      Topic20      Topic21      Topic22      Topic
23
     5.9122       4.5987       5.4266       5.4167       3.6365       1.69
75
    Topic24      Topic25      Topic26      Topic27      Topic28      Topic
29
     7.0411      -1.3068       2.0934       2.4576       3.2785       5.60
07
    Topic30
     1.3254
```

```
Degrees of Freedom: 27692 Total (i.e. Null);  27662 Residual
Null Deviance:     110700
Residual Deviance: 39570     AIC: 88540

1 2 3 4 5 6
3 5 4 3 4 5
[1] 6957
Confusion matrix:    predictions
       1    5
  1 2836  681
  5  123 3317
```

Accuracy: [1] 0.8844329

Deployment

From the models, we received the sentiment scores for each word based on the Yelp reviews data. In the deployment, we applied the sentiment scores to the general customer generated textual data on social medias, which could be Twitter, Facebook or blogs, to predict the ratings for each message. In this project, we collected tweets from Twitter for a particular restaurant in a particular location to better understand customers, in order to identify main factors which could benefit our business.

Since both the Yelp datasets and tweets we collected are available to the public, and we didn't use any personal information from the internet users, we believe there is no ethical issues involved in our approach.

Due to the competition among restaurants, there are lots of fake reviews on Yelp. The fake reviews would lead to over counting the common words for a particular high rating or low rating when we calculate the average sentiment scores for those words. The sentiment scores of the reviews might also get affected. In our approach, we didn't include the detection of fake reviews. We assume that fake reviews contribute only a small proportion among all reviews, which shouldn't impact our modeling results too much.

The user group who generate the tweets are different from the customers who wrote reviews on Yelp, so there might be some discrepancies when we apply the same algorithm. Twitter users tend to have more subjective and emotional expressions due to the social media characteristic. Therefore, the messages we collected from Twitter might be less objective comparing to the Yelp reviews.

Another common situation is that the Twitter users sometimes talk irrelevant topics even though they use the hashtag of the restaurants' names. Therefore, assigning a rating to such tweets based on our algorithm might be misleading and unnecessary. Additionally, each tweet has a limit of 140 words, while Yelp reviews don't have such words limit. The tweets we collected might not be comprehensive enough for driving business values.

Just as the Yelp reviews, we also have the same problems when dealing with ambiguous words and linguistic differences. Different word orders and the use of negative words might also pass completely different messages, which cannot be captured in our models.

To further improve the effectiveness and accuracy of our algorithm regarding the linguistic difficulties, there are two alternative steps we could take next. We could either filter the reviews which are written in languages other than English, or translate all reviews into one standard language, English before applying models.

Bibliography

1. Does sentiment analysis work? A tidy analysis of Yelp reviews
http://varianceexplained.org/r/yelp-sentiment/

2. Predicting and Modeling Activity in the Evolving Yelp Network
http://snap.stanford.edu/class/cs224w-2012/projects/cs224w-017-final.pdf

3. Predicting Yelp Star Ratings Through Latent Topic Analysis
http://www.stanleyku.com/includes/sk_ms_project.pdf

4. Predicting Yelp ratings using textual reviews
http://www.runzemc.com/2014/12/predicting-yelp-ratings-using-textual-reviews.html

5. "Eat, Rate, Love" — An Exploration of R, Yelp, and the Search for Good Indian Food
https://www.springboard.com/blog/eat-rate-love-an-exploration-of-r-yelp-and-the-search-for-good-indian-food/

6. How Yelp Reviews Can Help Improve Patient Care
http://www.nytimes.com/2016/09/13/upshot/how-yelp-reviews-can-help-improve-patient-care.html?rref=collection%2Fsectioncollection%2Fupshot&_r=0

7. What (And Why) Do People Share On Social Networks?
http://www.adweek.com/socialtimes/what-why-social-sharing/500025

8. Food Network Chef Robert Irvine Shares the Top 5 Reasons Restaurants Fail
http://www.businessinsider.com/why-restaurants-fail-so-often-2014-2

7. 5 Tips for Cooking Up a Successful Restaurant Business
http://www.forbes.com/sites/hesterlacey/2014/02/24/5-tips-for-cooking-up-a-successful-restaurant-business/#2917ba534c61