Predicting Star Rating and Characteristics of Restaurants for Development Analysis

Name: Noshin Islam
Institution: *Department of Electrical and Computer Engineering,North Southth University*
Address: Dhaka, Banlgadesh
Email: noshin.islam@northsouth.edu

Name: Sazzad Hossain Sharif
Institution : *Department of Electrical and Computer Engineering,North Southth University*
Address: Dhaka, Banlgadesh
Email:sharif530@gmail.com

Name: Ismail Hossain
Institution : *Department of Electrical and Computer Engineering,North Southth University*
Address: Dhaka, Banlgadesh
Email: ismail.9595@gmail.com

***Abstract*— As restaurant business is at the highest peak in our country, mostly in our capital city, The authenticity of restaurant reviews has become more important than ever. We want to examine: the underlying factors that prompts a user to give 'x' amount of stars to a restaurant. There are many food ordering service aligned with restaurant review systems available in our country through digital medium: HungryNaki, Pathao, FoodPanda, FoodPeon etc. Our goal is to analyze the user ratings/reviews from a renowned system's database: by listing out the probable reasons as factors that results in a prediction algorithm which will be able to predict a user's rating. After a thorough investigation by using different models, we found SVM is slowest but gives higher accuracy and naïve bayes is fastest but lowest accuracy.**

Keywords— rating, reviews, restaurants, business, food, service

I. Introduction

Amid the most recent couple of years, reviews have turned out to be urgent to the accomplishment of an restaurant as each restaurant proprietor knows about the way that great surveys can support prevalence and benefit, while horrible reviews even have the capability of shutting organizations down. That is the reason it is critical for restauranteurs to comprehend the effect of review sites, for example, Yelp, Toptable or TripAdvisors and the job they play the achievement or ruin of a business.

Every time a restaurant is mentioned in a printed magazine or newspaper there is some sort of increase in sales. Reviews can produce a considerable measure of introduction for an eatery, be that as it may. Be they positive or negative, reviews and ratings convey learning of the eatery to the forthcoming benefactors consideration. This expanded introduction can prompt more deals, and more individuals discussing your eatery. While reviews that are antagonistic can be squashing hits to an eateries prominence, regardless they open the eatery to various individuals who might somehow not have thought about it. It is conceivable to enhance your eatery after a critics review.

In this project, we will utilize the Yelp public dataset/ Zoomta restaurants data/ manually collected dataset to analyze the star rating of restaurants to investigate the devlopment process. In particular, we will predict the star ratings of restaurants and find the most useful traits in determining their success.

This task is important as it will allow new businesses with limited customer input to have a better idea of how well they will perform in the long run

II. Motivation

A. *From a restaurant owner's perspective*

- People tend to look at the number of reviews and the quality when it comes to selecting the place they want to go. This is why websites such as yelp are held in such high regard
- The more number of reviews you have the better it is from a SEO
- Consistent reviews are a huge thing, How well you generally score will indicate your position in the local ratings chart on that site
- This increased exposure(positive or negative) can lead to more sales, and more people talking about your restaurant
- Ratings help a restaurant understand people's general thoughts and where it stands in the competition
- Both positive and negative rating impacts a restaurants business, thus learning about the underlying factors for a rating is very important

B. *From a Customer's perspective*

- Reviews/Ratings make the selection process easier for people who do not want to do all the research themselves and would rather trust an authority source to make their decision
- Customers tend to search for good food and positive experience, ratings help their selection process
- Apps like Foodpanda, lets you know about the food varieties available in a restaurant and if the pricing ranges are same in two or more restaurants for a particular food, people tend to depend on the rating of a restaurant
- If customers can be assured that the reviews given are authentic, they will trust on the service
- It is very crucial to find out the most prominent factors

III. Background

In the project report named "Predicting Success of Restaurants in Las Vegas", the authors introduce an innovative way to predict the success of restaurants in Las Vegas by predicting the star ratings of the restaurants and find out the features that are most useful in determining the success. By investigating the public dataset of Yelp, the authors analyses the restaurant attributes and unigrams and bigrams from reviews in yelp, to use them as key structures for classification among the restaurants. From the results, they found that the perceptron Neural Network is the highest performing algorithm for both feature extractors [1]

Researchers at Stanford University, used Yelp Dataset Challenge to predict how useful a review will prove to be to users The features we identified with the highest absolute weights were linked to how many of certain types of Yelp compliments the user had received. However, the feature that we identified as likely the most important was the number of Yelp friends the user had. Even though this feature often had a lower weight than some of the compliment features. The feature that we identified as likely the most important was the number of Yelp friends the user had. [2]

The project conducted by Chen Li and Jin Zhang [3] has two tiers. First, using a customer's review on a business to predict the star rating given by the customer. Second, using many customers reviews on a business and on different to predict this customer's likely ratings of different businesses. They found several classification algorithms to be best fitting. The SVM and SVR predict the best for polarity and 5-star classification respectively. So they decided to use SVM for their recommendation model.

Rakesh Chada and Chetan Naik [4] from stony brook university published a paper where they predict a review's star rating given just the review's text. The project is relied on two state-of-the-art techniques to achieve this task. The first technique was "Latent Dirichlet Allocation (LDA)" which is proposed by Blei in 2003 and the second technique was "Non-negative Matrix Factorization (NMF)". For dataset they used the Yelp Academic Dataset which consists of 1,125,458 reviews, 320,002 business attributes and 252,898 users.

F.M Takbir Hossain, Md. Ismail Hossain, Ms. Samia Nawshin [5] from Daffodil International University published a paper where combined user review texts which were collected from that website to build a model that can predict a review asserting positive or negative For data processing they have collected review data from local businesses. For data preprocessing they used Natural language Toolkit (NLTK) which is widely used for processing text data. In this paper they have used four algorithms, Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM) K nearest neighbor (KNN) and Linguistic Regression (LR) The best result has given Linguistic Regression (LR) with more than 77% accuracy.

Narendra Gupta, Giuseppe Di Fabbrizio and Patrick Haffner [6] from AT&T Labs - Research, Inc. Published a paper where they predict ratings for service and product reviews. At the time of mining, they collect reviews of about 3,800 restaurants with an average of two reviews per restaurant containing around eight sentences per review from the website we8there.com. Here they use three models: numeric regression, ordinal regression and maximum entropy classification Models. Though the problem seems like a regression problem, maximum entropy classification models perform the best. Results show a strong inter–dependence in the way users rate different aspects.

Jiajie Shi, Muhan Zhao, Zhiling Liu [7] from University of California Published a paper about Rating Prediction and Recommendation of California Restaurants Based on Google Database. The object of this study was 292,465 reviews data and 45,536 restaurants located in California state. Then they plot some features like review text of time, numbers of reviews that the restaurant received and the length of reviews for each restaurant versus rating. Then they build up rating prediction model based on linear regression, latent factor model and latent semantic model. By computing the score for each user and restaurant, they find the restaurant with highest score, which is the one we would like to recommend to this user.

IV. DATA SOURCE

Initially, the dataset we selected to use in this project came from the Yelp Dataset Challenge in the form of JSON files. From this, the restaurants in Gilbert, a town in Arizona, USA were extracted, along with all the reviews associated with them. Given each restaurant, the initial data gave us the restaurant's characteristics as well as its reviews. The total amount of data for Gilbert included 128 restaurants. Total number of distinct reviews are 4,168. We experimented on this dataset to visualize how many restaurants in Gilbert city has ratings 4.5 and above,

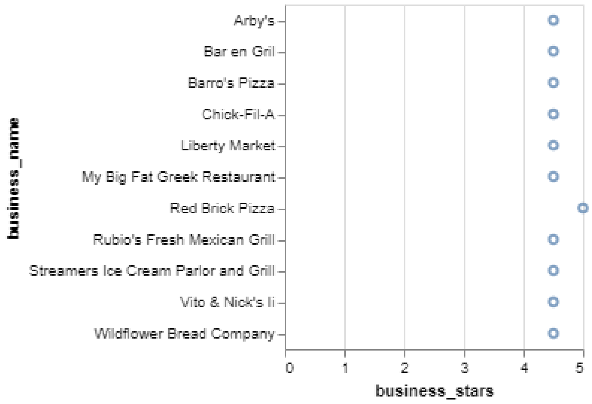

Figure 1. Restaurants having 4.5 stars and above

We realized soon enough that predicting star ratings doesn't really give us the inner dimensions of restaurant business development analysis. We wanted to collect a dataset which had both star-ratings and textual features as features. The training and test dataset is collected from the author of [8], which consists of around 8,000 and 2,000 data points.
They selected all the user reviews that has atleast one useful vote.

V. PREPROCESSING

In the initial dataset, the characteristics are all preprocessed to have numeric values, such as times being changed from 11:30AM to 11.5, True/False being changed to 1/0, etc. In addition, any feature for which the restaurant had no information was given the value of -1. The feature "categories", which was a list of categories that the restaurant fits into, i.e. Chinese, Korean, fast food, breakfast, barbeque etc. To better represent this information in our analysis, we create corresponding dummy variables for each of these hot tags.

We combined three separate JSON files, Business, Reviews and User. Restaurant reviews were preprocessed using the python nlkt natural language package. We performed a Parts-Of-Speech tagging of all words in all the reviews, and filtered out all the words that are adjectives and nouns.

But soon we were realizing that only restaurants in Gilbert City could have very biased reviews that is why we looked for datasets that had textual features of reviews, we found a dataset which had considered reviews with at least one useful vote.

The dataset we collected had 668 features. Features included unigrams, bigrams trigrams all having frequency greater than 300.

VI. FEATURES

In our method of using the dataset, we used 668 binary predictors (independent variables) and 8 binary outcome variables. The predictors include textual features in the form of unigrams, bigrams, and trigrams with highest frequencies. The text of each review is normalized at first, i.e. converted to lower-case, before the unigrams, bigrams, and trigrams can be extracted. The ngrams that contain punctuations are filtered out.
Total of 375 unigrams, 208 bigrams and 120 trigrams features. The authors tried out five classifiers, we also tested out the rating classifiers. We wanted to see if the reviews can predict the appropriate star that was assigned to it.

The eight binary outcome variables are:

1. IsFoodGood,
2. IsServiceGood,
3. IsAmbianceGood,
4. IsDealsGood,
5. IsPriceGood,
6. IsRatingGood (4-5 stars)
7. IsRatingBad (1-2 stars),
8. IsRatingModerate (3 stars)

Among these we chose IsFoodGood, IsServiceGood, IsRatingGood(4-5 stars), as these are the most anticipated questions from our point of view:

VII. ALGORITHMS

For this experiment, we decided to use:

1. Support Vector Machine (SVM) with RBF & Linear Kernal
2. Decision Tree
3. Naïve-Bayes
4. Random Forest

VIII. METHODOLOGY & EXPERIMENTAL RESULT

A. *SUPPORT VECTOR REGRESSION WITH LINEAR KERNEL*

Support Vector Regression was used to attempt to fit the data with a regression using a Linear kernel.
Our implementation is mostly in Python, making use of libraries such as Scikit-learn for machine learning, NLTK for text processing, We first imported our training and test datasets in the Python Development environment Spyder, using Anaconda Navigator. We used sklearn.svm package and imported the SVC class.
There was no additional feature scaling needed as it has already been done. Then we divided our training dataset and testing dataset into corresponding independent and dependent variables. We tested the independent variables to predict how they behave in a particular class. i.e. Can a training data containing textual features of a review that concludes to food being good, predict another similar review to be in the "IsFoodGood" class. Let's see by comparing the initial test result for the class and the predicted test result data

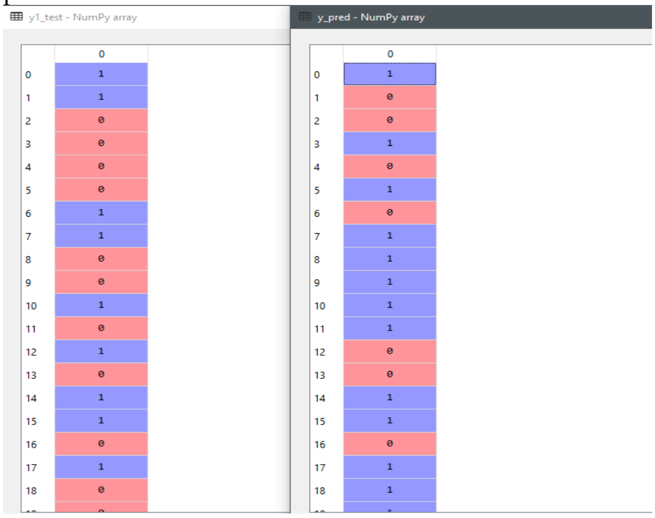

Figure 2.A glimpse of IsFoodGood classfier output variable result [(Kernal='Linear')(Testing values vs Machine predicted values)]

We also saw how it performs if we want to know about the service overview of a restaurant from the review.

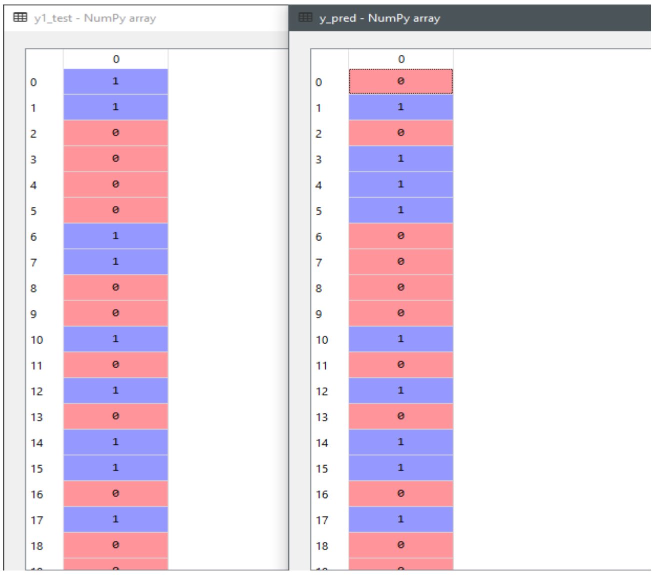

Figure 3. A glimpse of IsServiceGood classifier output variable result (Testing vs machine predicted values)

We computed the confusion matrix also known as error matrix to see how many the machine predicted wrong. From the above example we saw that among 1960 examples the machine predicted total 290+135=325 results wrong, which is not that bad.

Another example to classify a review into “IsReviewGood”, which is if the review falls into gave us results(Initial test vs Predicted values)

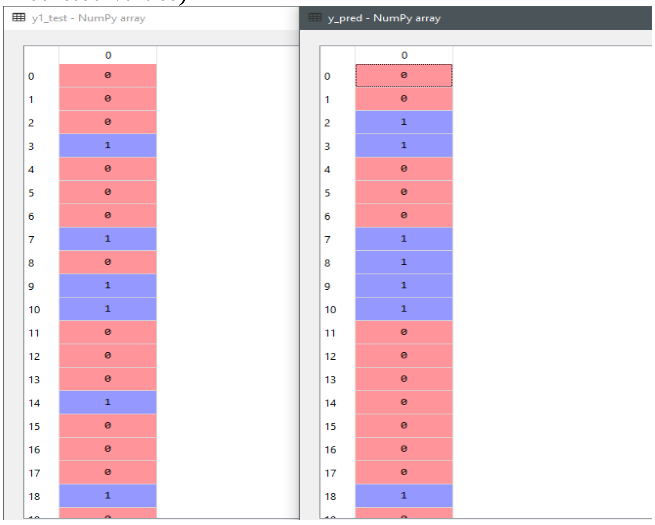

Figure 4. A glimpse of IsRatingGood classifier output variable result [(Kernal= ‘Linear’) (Testing values vs Machine predicted values)]

B. Decision Tree

We used Decision tree as a predictive model because it can take several input variables to predict one target variable. It is also a very popular one for pattern recognation. The principal algorithm for building decision trees called ID3, it’s a top-down, greedy search through the space of possible branches with no backtracking.

We took the similar examples/scenarios from Support Vector Machine:

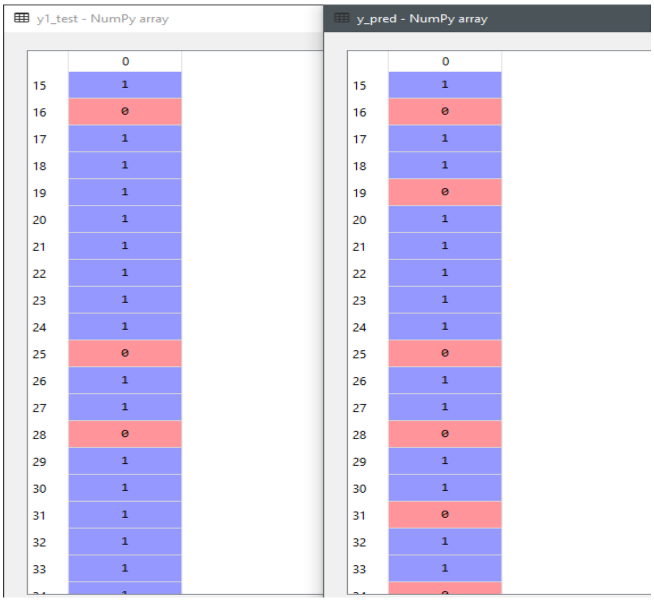

Figure 5. Glimse of IsFoodGood classifier output variable result [(Testing values vs Machine predicted values)]

We also saw how it performs for “IsServiceGood” classifer:

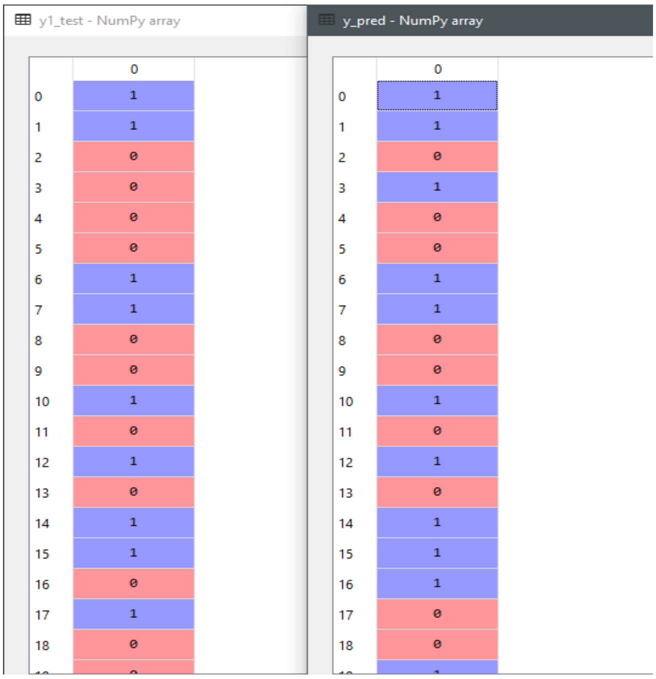

Figure 6. A glimpse of IsServiceGood classifier output variable result (Testing vs machine predicted values)

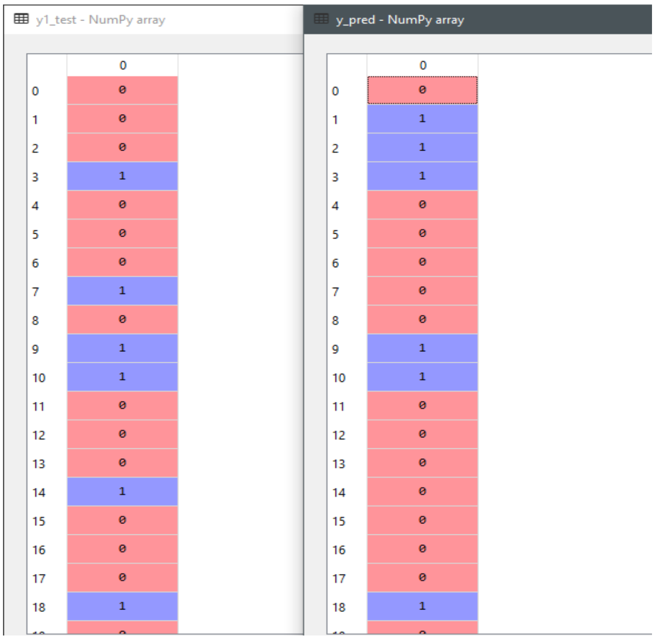

Figure 7. Fig: Glimpse of IsRatingGood classifier output variable result [(Testing values vs Machine predicted values)]

C. Random Forest

Random forests or random decision forests are an collaborative learning method for classification, regression and other responsibilities that functions by constructing a assembly of decision trees at training time and give and output of the class that is the mode of the classes or mean prediction of the individual trees.

We evaluated the similar scenarios:

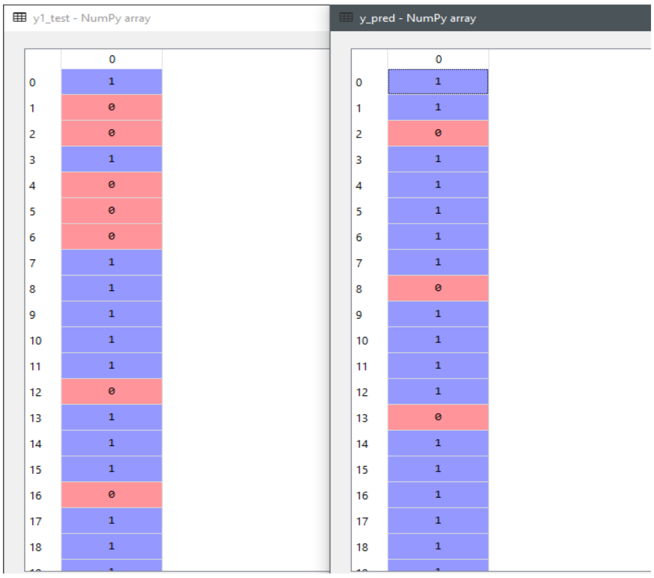

Figure 8. Glimse of IsFoodGood classifier output variable result [(Testing values vs Machine predicted values)]

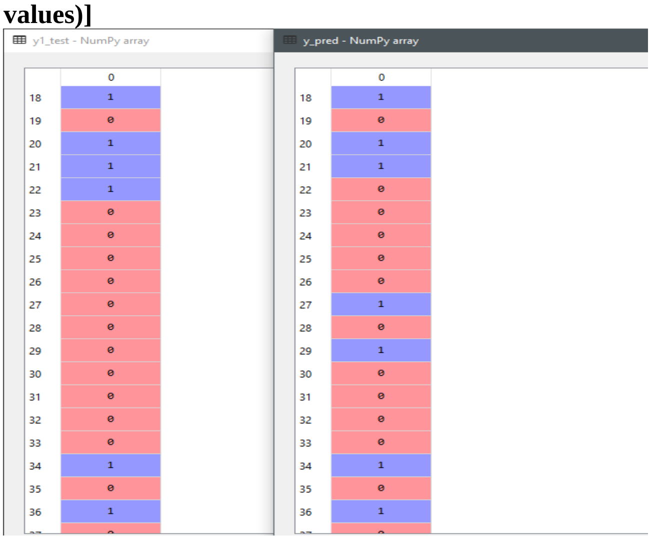

Figure 9. Glimse of IsRatingGood classifier output variable result [(Testing values vs Machine predicted values)]

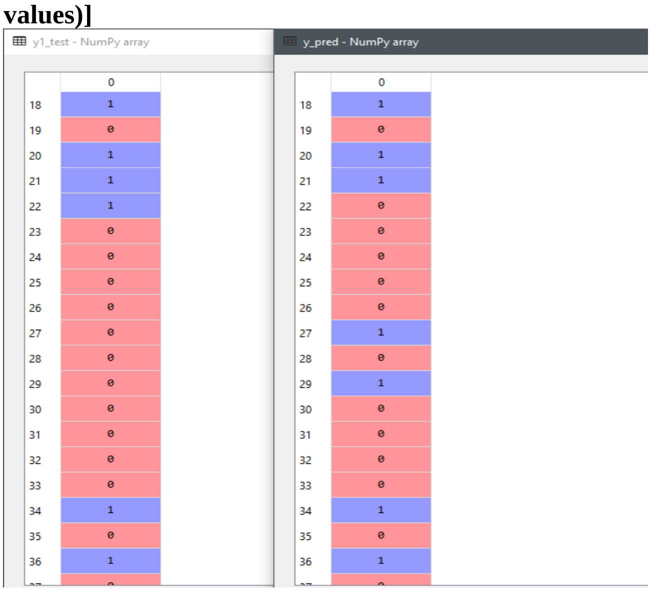

Figure 10. A glimpse of IsRatingGood classifier output variable result (Testing vs machine predicted values)

D. Naïve Bayes

The Naive Bayes Classifier technique is grounded on Bayesian theorem and is mostly suited for probles which has a large scale of input values. This classifier undertakes that the occurrence of a particular feature in a class is not related to the occurrence of any other feature.

We took Simliar Scenarios:

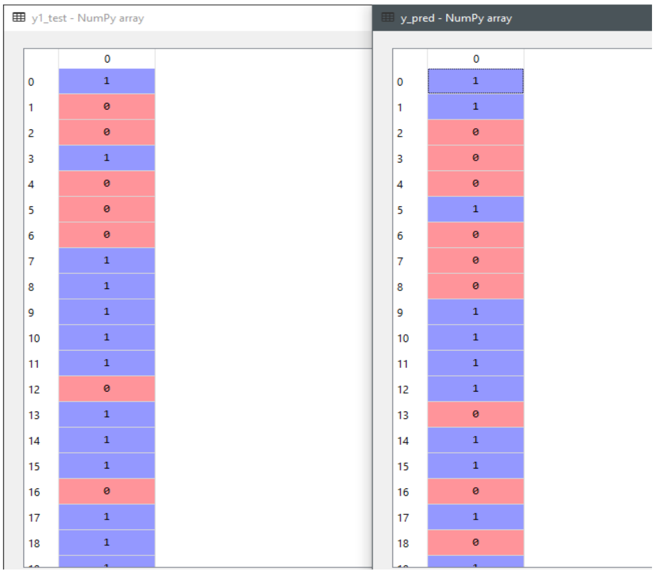

Figure 11. Glimse of IsFoodGood classifier output variable result [(Testing values vs Machine predicted values)]

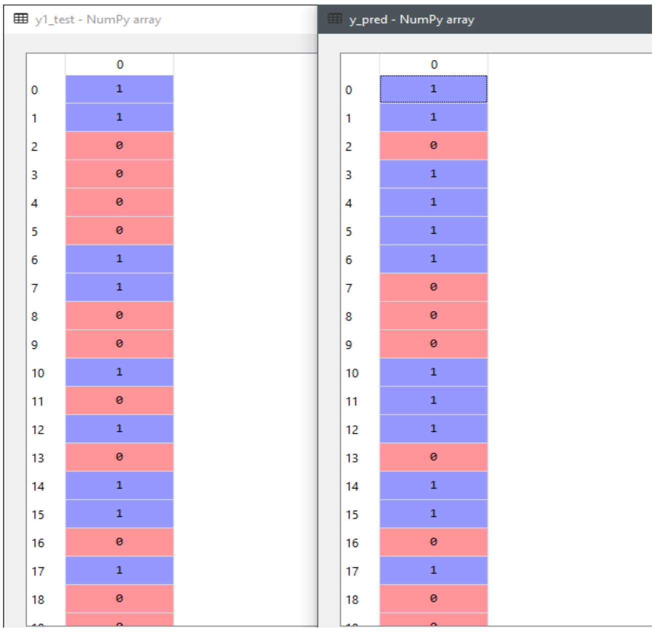

Figure 12. Glimse of IsServiceGood classifier output variable result [(Testing values vs Machine predicted values)]

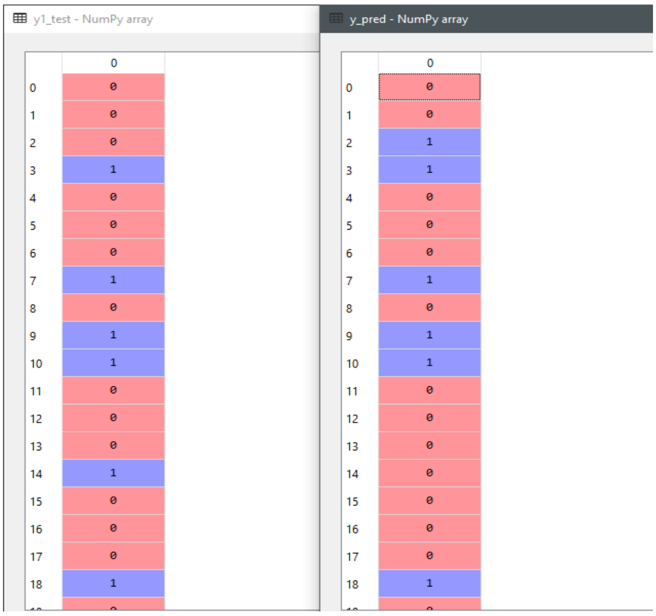

Figure 13. Glimse of IsRatingGood classifier output variable result [(Testing values vs Machine predicted values)]

IX. PERFORMANCE MEASURES

A. Confusion Matrix

We primarily focused on getting the confusion matrix for to observe how different models are reacting as individual classifier model.

1) Support Vector Machine:

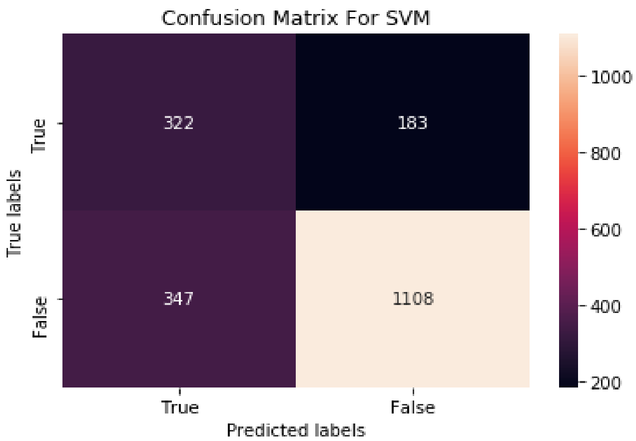

Figure 14. : IsFoodGood classfier confusion matrix result showing 347+183= 530 wrong results

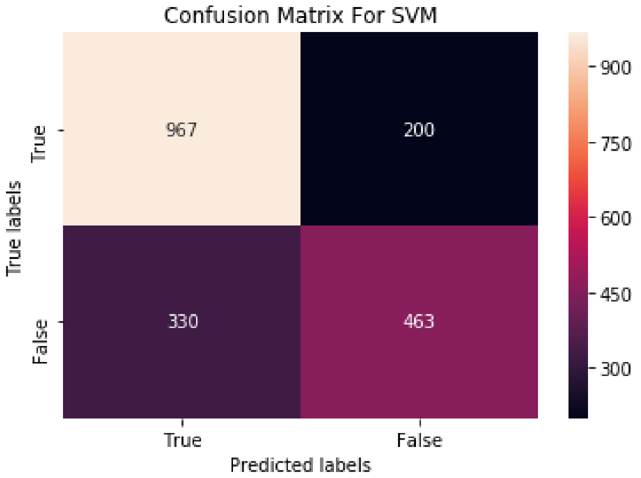

Figure 15. : IsServiceGood classfier confusion matrix result showing 200+330= 530 wrong results

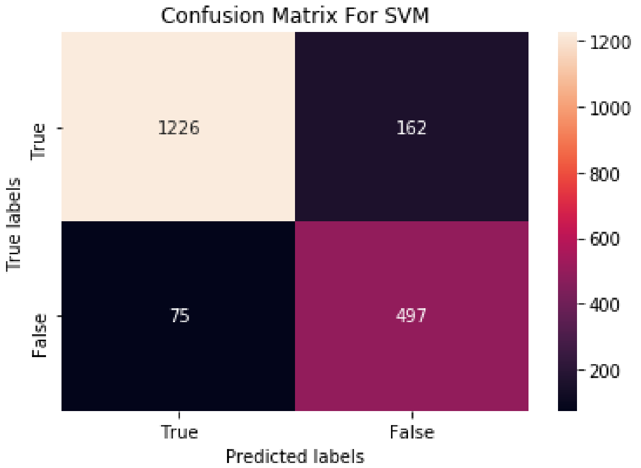

Figure 16. : IsRatingGood classfier confusion matrix result showing 162+75= 237 wrong results

2) Decision Tree:

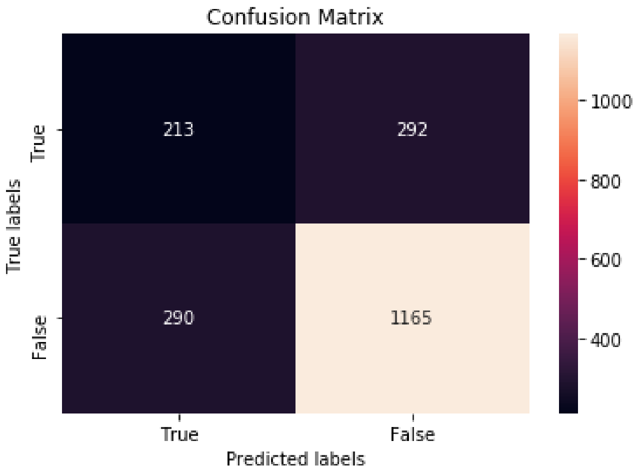

Figure 17. : IsServiceGood classfier confusion matrix result showing 292+290= 582 wrong results

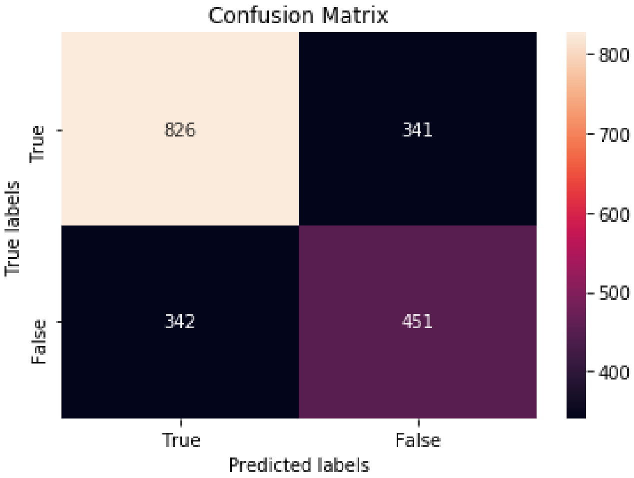

Figure 18. : IsServiceGood classfier confusion matrix result showing 341+342= 683 wrong results

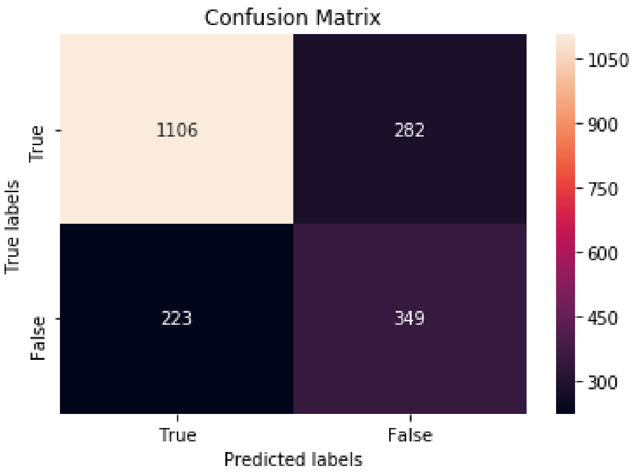

Figure 19. : IsRatingGood classfier confusion matrix result showing 282+223= 565 wrong results

3) *Random Forest:*

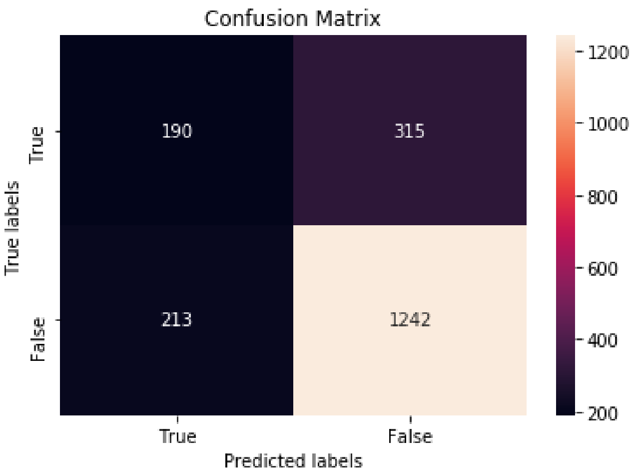

Figure 20. : IsFoodGood classfier confusion matrix result showing 213+315= 528 wrong results

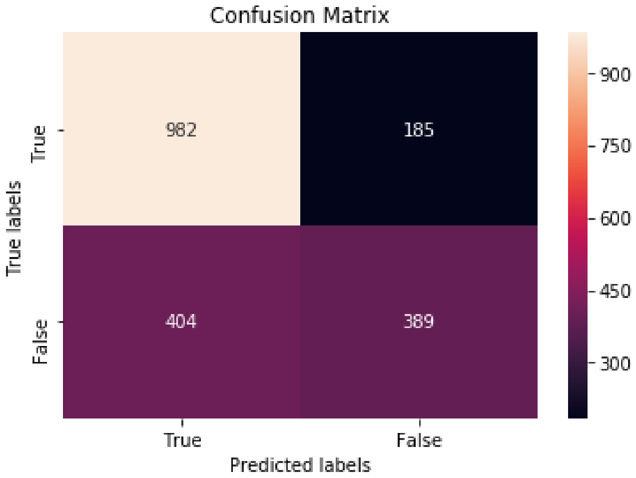

Figure 21. : IsServiceGood classfier confusion matrix result showing 185+404= 589 wrong results

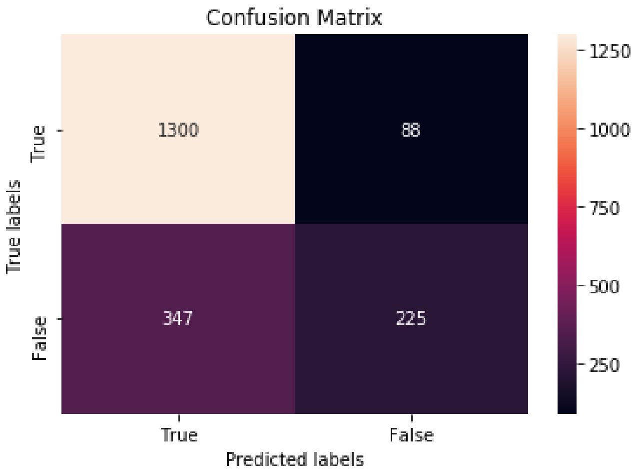

Figure 22. : IsRatingGood classfier confusion matrix result showing 347+88= 435 wrong results

4) *Naïve Bayes:*

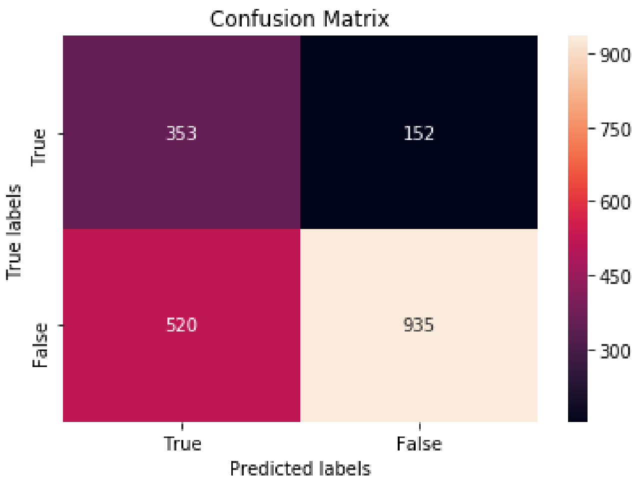

Figure 23. : IsFoodGood classfier confusion matrix result showing 520+152= 672 wrong results

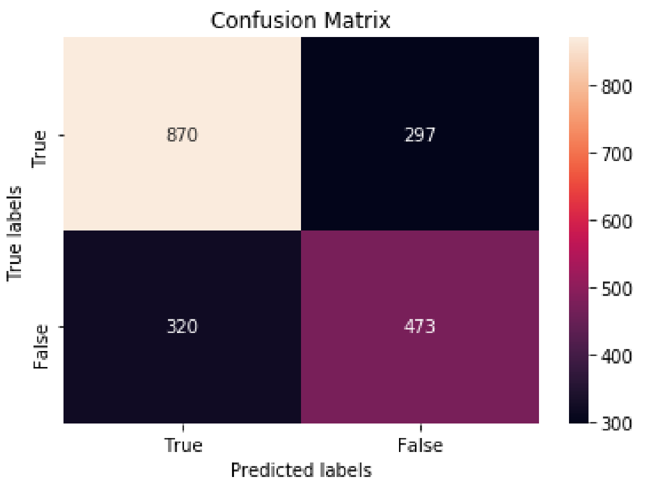

Figure 24. : IsServiceGood classfier confusion matrix result showing 297+320= 617 wrong results

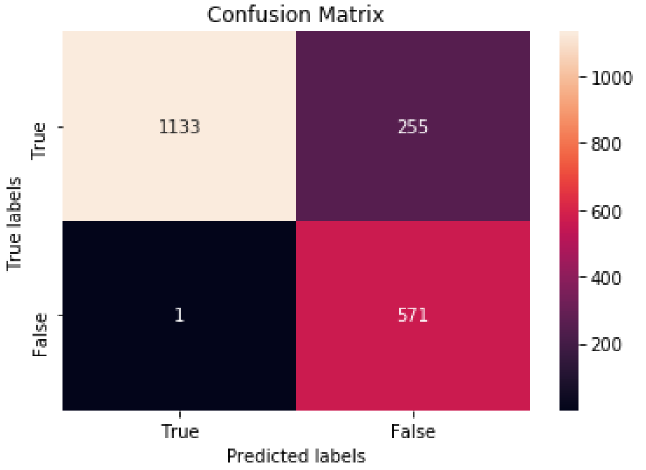

Figure 25. : IsRatingGood classfier confusion matrix result showing 255+01= 256 wrong results

B. Accuracy Measurement:

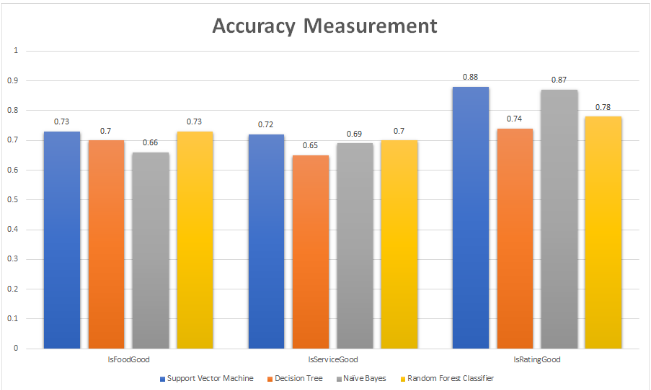

Figure 26. Accuracy measurement among different classifiers using the algorithms

C. 10-fold cross validation:

We used 10-fold Cross validation.to estimate classification models by segregating the original example into a training set to train the model, and a test set to evaluate it. The original dataset is randomly partitioned into 10 equal size subsamples.

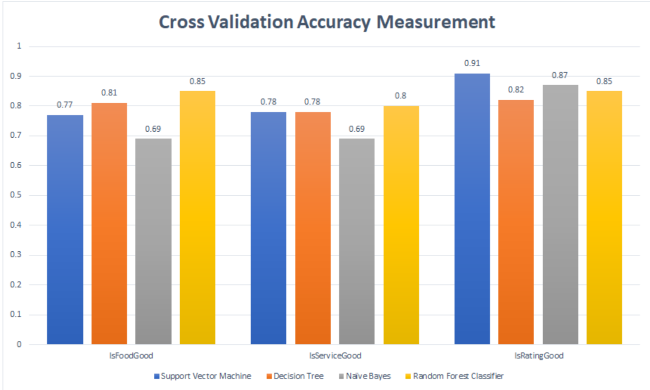

Figure 27. 10 fold cross validation Accuracy measurement variations

X. Discussion

We recognised our problem as a classification problem and we tested out several different algorithms to see how the different classifiers are behaving in each of them.

Algorithms	IsFoodGood Error rate	IsServiceGood Error rate	IsRatingGood Error rate
SVM	0.25	0.28	0.12
Decision Tree	0.3	0.35	0.26
Random Forest	0.34	0.31	0.13
Naïve Bayes	0.27	0.3	0.22

Figure 28. Error rate of different classifiers

We computed the error rate by using the formula:

1-accuracy = error rate

It can be seen that, for IsFoodGood classifier, SVM worked the best (0.25 error), the second best proved out to be naïve bayes(0.27 error). For IsServiceGood classifier, SVM again proved to be the best (0.28 error) and second is Naïve Bayes (0.3 error). Lastly, for IsRatingGood classifier, SVM (0.12 error) again proved to be the best for rating prediction from the reviews among other algorithms, but here Random Forest proved to be the second best (0.13 error)

All the classification machine learning models had very minimum amount of errors than usual. It can be safely said the error we anticipated to be happening was far less in reality among the different classification algorithms. Support Vector Machine and Naïve Bayes proved to be the best algorithms for restaurant characteristics (Food, service) classification, whereas Random Forest and Support Vector Machine proved to be best restaurant rating classifier.

When we computed accuracies through 10-fold cross validation we found that, SVM is slowest but higher accuracy, naïve bayes is fastest but lowest accuracy.

Random forest almost very good accuracy and fast. Decision tree is fast but didn’t perform that well in accuracy.

XI. Conclusion

Our goal throughout this experimentation was to analyze how restaurant characteristics and ratings work based on the user reviews. This experiment could be a great help for restaurants or any business in general to visulaize their development process. Yelp Datset Challenge has been a great resource for Machine Learning and Natural Language Processing investigation. Future work could be predicted as how unsupervised algorithms are going to be performing in the situation.

XII. References

[1] Sang Goo Kang and Viet Vo, “Predicting Success of Restaurants in Las Vegas,” 2016, Stanford University

[8] Hitesh Sajnani, Vaibhav Saini, Kusum Kumar , Eugenia Gabrielova , Pramit Choudary, Cristina Lopes, “Classifying Yelp reviews into relevant categories”, 2015, University of California, San diego, USA

[2] Ben Isaacs, Xavier Mignot, Maxwell Siegelman, “Predicting Usefulness of Yelp Reviews” , 2014, Stanford University

[3] Chen Li and Jin Zhang, “Prediction of Yelp Review Star Rating using Sentiment Analysis”, 2014, Stanford University

[4] Rakesh Chada , Chetan Naik , “Data Mining Yelp Data - Predicting rating stars from review”, 2014, Stony brook university.

[5] F.M Takbir Hossain, Md. Ismail Hossain, Ms. Samia Nawshin, “Machine Learning Based Class Level Prediction of Restaurant Reviews”, 2010, Daffodil International University.

[6] Narendra Gupta, Giuseppe Di Fabbrizio and Patrick Haffner, “Capturing the stars: predicting ratings for service and product reviews”, 2010, AT&T Labs - Research, Inc, Florham Park, NJ 07932 – USA.

[7] Jiajie Shi, Muhan Zhao, Zhiling Liu, “Rating Prediction and Recommendation of California Restaurants Based on Google Database”, 2017,University of California, San Diego, USA