Restaurants Rating Prediction using Machine Learning Algorithms

Vicky Malik, S.Prasad Babu Vagolu, Sunil Chandolu

*Abstract***:** *Restaurant Rating has become the most commonly used parameter for judging a restaurant for any individual. A lot of research has been done on different restaurants and the quality of food it serves. Rating of a restaurant depends on factors like reviews, area situated, average cost for two people, votes, cuisines and the type of restaurant. The project aim is to find out the relationship between the dependent and independent variable. Proposed project is a Machine Learning Regression problem which uses Restaurant Rating dataset. Based on various attributes like the food, quality, prize ambience of the restaurant it predicts the Restaurant Rating.*

Keywords **:** *Restaurant Rating, Random Forest Algorithm, Linear Regression, Machine Learning Algorithm.*

I. INTRODUCTION

In today's digitized modern world, popularity of food apps is increasing due to its functionality to view, book and order for food by a few clicks on the phone for their favorite restaurant or cafes, by surveying the user ratings and reviews of the previously visited customers. Restaurant Rating also provides columns for writing classified user reviews. Such sort of substance provided by web is named as client produced content. Client created content contains a great deal of significant and essential data about the food items and restaurant administrations. Since there is no control on the nature of this substance on the web and thus, these elevate fraudsters to compose counterfeit surveys to defame the restaurant administrations, to provide misguiding reviews, to generate irrelevant content regardless of the product or service, to advertise unrelated content, etc. These phony surveys anticipate clients and associations achieving genuine decisions about the product, services, and amenities of the restaurants or cafes. In this case, Review Analysis has become vital to generate authenticated and unbiased reviews which help in avoiding fraudulent activities used to promote business by publishing fake reviews.

Vicky Malik*, PG Student, Department of Computer Science, GITAM (Deemed to be University), Visakhapatnam Email: malikv925@gmail.com

S.Prasad Babu Vagolu, Assistant Professor, Department of Computer Science, GITAM (Deemed to be University), Visakhapatnam Email: prasadray@live.com

Sunil Chandolu, Assistant Professor, Department of Computer Science, GITAM (Deemed to be University), Visakhapatnam Email: kunny0306@gmail.com

Hereby in this paper we focus on mining customer reviews, authenticate them, classify them into positive and negative reviews, and find worthiness of the product.

Different machine learning algorithms like SVM, Linear regression, Decision Tree, Random Forest can be used to predict the ratings of the restaurants.

II. DATA SET DESCRIPTION

This is a kaggle dataset.
(https://www.kaggle.com/himanshupoddar/zomato-bangalore- restaurants).

It Represents information of Restaurants in the City of Bangalore. It contains 17 Columns and 51,000 Rows.

The dataset has the following attributes such as: Restaurant Name, Restaurant ID, City, Address, Cuisines, Cost for two people, has table booking, Has online delivery, Is delivering now, Switch to order menu, Prize range, Aggregate rating, Rating color, Rating text and votes.

So, for the restaurant to have a higher rating the customer rating plays an important role, and if the rating of the restaurant is higher it will also bring new customers to restaurants. The customer relationship plays a very important role for the success and profit in business.

Pre Processing

The Dataset contained 17 Attributes.

- Records with null values were dropped from ratings columns and were replaced in the other columns with a numerical value.
- Values in the 'Rating' column were changed. The '/5' string was deleted. For eg. If the rating of a restaurant was 3.5/5, it was changed to 3.5.
- Using LabelEncoding from sklearn library, encoding was done on columns like book_table,online_order,rest_type,listed_in(city).

III. FEATURE SELECTION

We did not use any feature selection algorithms but eliminated some columns due to available domain knowledge and thorough study of the system.

Dropped columns mentioned below:

- URL
- Address
- Dish_liked
- Phone
- Menu
- Review_list
- Location
- Cuisine

Some of these columns may look like they are important but all of the same information could be found in other columns with lesser complexity.

The Columns being used are as follows:

- Name
- Online_order
- Book_table
- Votes
- Rest_type
- Approx. cost of two people
- Listed_in(type)
- Listed_in(city)

IV. EXPLORATORY DATA ANALYSIS

A lot of effort went into the EDA as it gives us a detailed knowledge of our data.

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

- Maximize insight into a data set;
- Uncover underlying structure;
- Extract important variables;
- Detect outliers and anomalies;
- Test underlying assumptions;
- Develop parsimonious models; and
- Determine optimal factor settings.

A. Restaurant Rate Distribution

We can see that the number of restaurants with the rating between 3.5 and 4 are the highest. We will look into its dependencies further.

B. Approximate Cost of two people

This is a graph for the 'Approximate cost of 2 people' for dining in a restaurant. Restaurants with this cost below 1000 Rupees are more.

This box plot helps us look into the outliers. We can also see that online ordering service also affects the rating. Restaurants with online ordering service have a rating from 3.5 to 4.

Restaurant Rate Distribution

Approx cost for two people
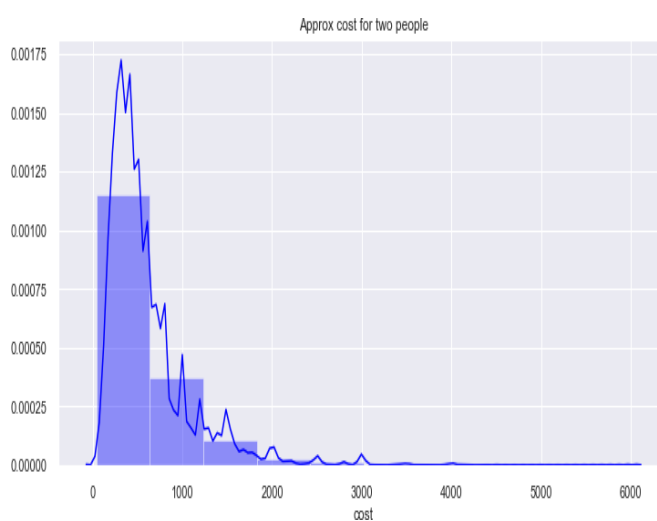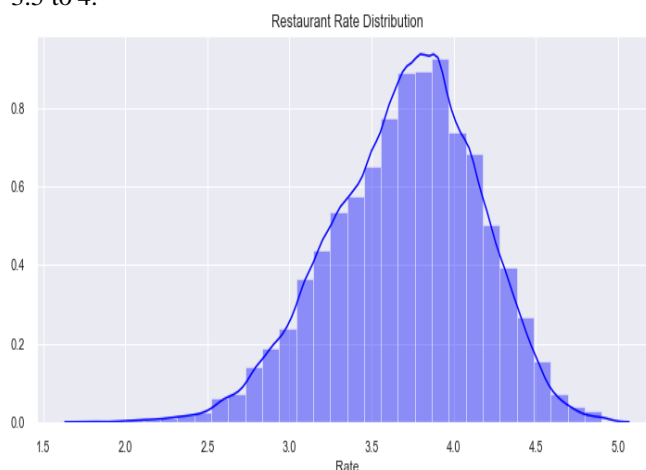
C. Online ordering with respect to Rating(Finding Outliers)
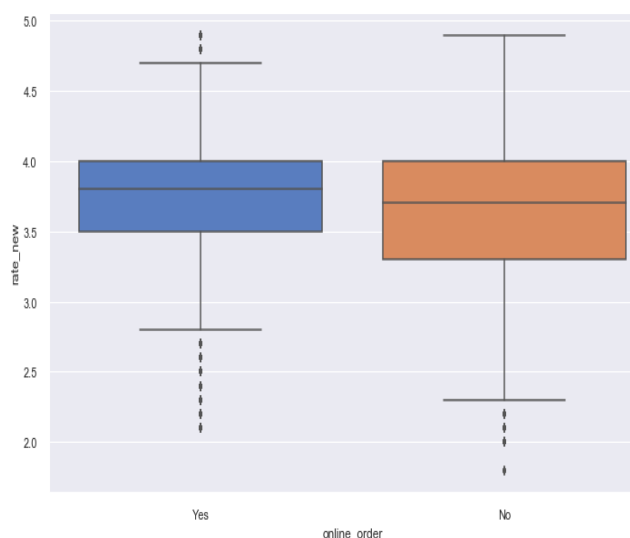
D. Booking table with respect to rating (Finding Outliers)

467
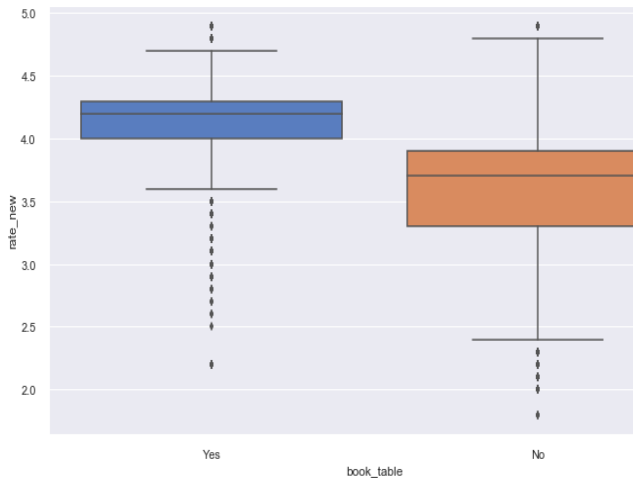
This box plot also helps us look into the outliers. This box plot is regarding how table booking availability is seen in restaurants with rating over 4.

F. Top Rated Restaurants

Top 10 Restaurants in Bengaluru by Rate

V. REVIEW CRITERIA

This graph just showcases the best restaurants in Bangalore along with their rating.

G. Cost and Rate Distribution according to online ordering and booking table

A very important scatter plot shows the correspondence between the cost, online ordering, bookings and rating of the restaurant.

4.1. Key Findings

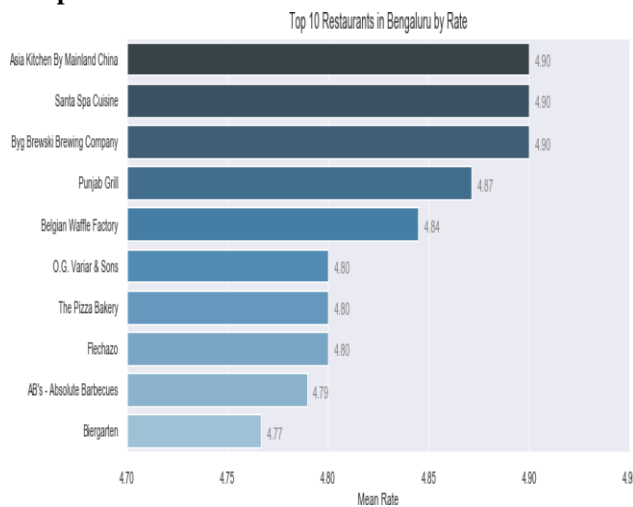|	Votes	approx_cost(for two people)	Rating
online_order			
No	367.992471	716.025190	3.658071
Yes	343.228663	544.365434	3.722440

	Votes	approx_cost(for two people)	Rating
Book_table			
No	204.580566	482.404625	3.620801
Yes	1171.342957	1276.491117	4.143464

VI. RESULTS

Algorithms	Accuracy
Linear Regression	30%
KNN	44%
Support Vector Machine	43%
Decision Tree	69%
Random Forest	81%

In this model, we have considered various restaurants records with features like the name, average cost, locality, whether it accepts online order, can we book a table, type of restaurant. This model will help business owners predict their rating on the parameters considered in our model and improve the customer experience.

Different algorithms were used but in the end the final model is selected on Random Forest Algorithm which gives the highest accuracy compared to others.
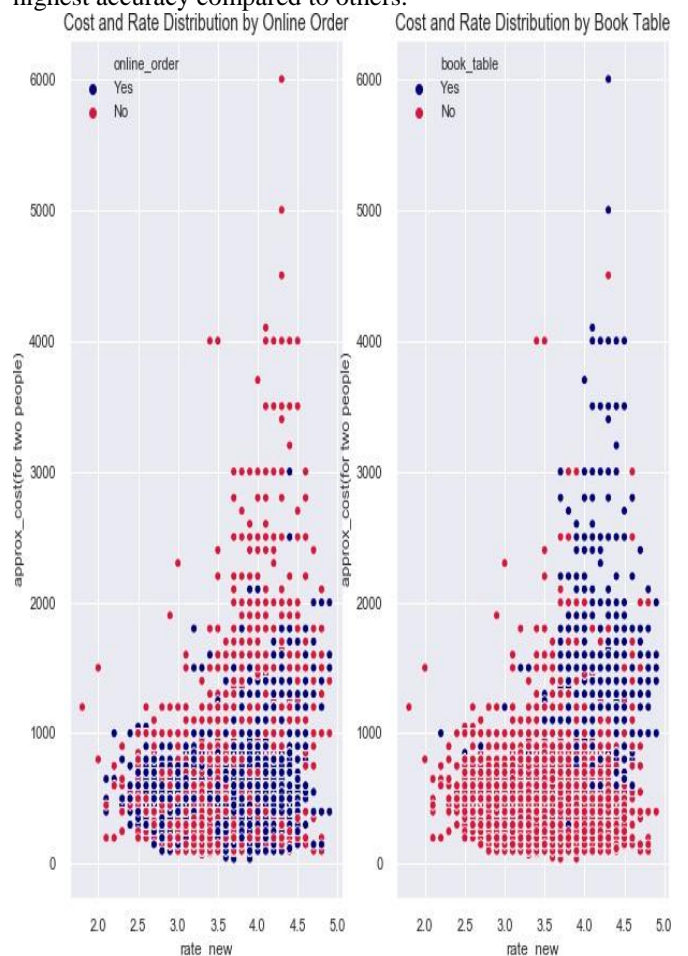
VII. CONCLUSIONS

This project performs both multinomial classification in terms of rating prediction and binary classification in terms of popularity change prediction. In this model, we have considered various restaurants records with features like the name, average cost, locality, whether it accepts online order, can we book a table, type of restaurant.

This model will help business owners predict their rating on the parameters considered in our model and improve the customer experience.

This paper studies a number of features about existing restaurants of different areas in a city and analyses them to predict rating of the restaurant. This makes it an important aspect to be considered, before making a dining decision. Such analysis is essential part of planning before establishing a venture like that of a restaurant.

REFERENCES

1. Chirath Kumarasiri, Cassim Faroo,"User Centric Mobile Based Decision-Making System Using Natural Language Processing (NLP) and Aspect Based Opinion Mining (ABOM) Techniques for Restaurant Selection". Springer 2018. DOI: 10.1007/978-3-030-01174-1_4 Computational Intelligence and Applications (ICCIA), Beijing, 2017, pp. 542-546. doi: 10.1109/CIAPP.2017.8167276
2. Shina, Sharma, S. & Singha ,A. (2018). A study of tree based machine
3. learning Machine Learning Techniques for Restaurant review. 2018 4th International Conference on Computing Communication and Automation (ICCCA) DOI:/10.1109/CCAA.2018.8777649
4. I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, 2017, pp. 542-546. doi: 10.1109/CIAPP.2017.8167276
5. Neha Joshi. A Study on Customer Preference and Satisfaction towards Restaurant in Dehradun City. Global Journal of Management and Business Research(2012) Link:
6. https://pdfs.semanticscholar.org/fef5/88622c39ef76dd773fcad8bb5d 233420a270.pdf
7. Bidisha Das Baksi, Harrsha P, Medha, Mohinishree Asthana, Dr. Anitha C.(2018) Restaurant Market Analysis.
8. International Research Journal of Engineering and Technology (IRJET)Link:https://www.irjet.net/archives/V5/i5/IRJET-V5I5489.pdf