

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 misleading 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.

Revised Manuscript Received on April 30, 2020.

* Correspondence Author

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

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

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.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q		
1	url	address	name	online_or_book	tablrate	votes	phone	location	rest_type	dish	liked	cuisines	approx_cc	reviews	l	menu	listed_in	city
2	https://www.zomato.com/bangalore-restaurants/942-21st-1ajalsa	Yes	Yes	4.1/5	775	080	Banashan	Casual	Dir	Pasta, Lun	North	Indi	800	[[Rated 4.1]]	Buffet	Banashankari		
3	https://www.zomato.com/bangalore-restaurants/10-2nd-floor-spice-elixir	Yes	No	4.1/5	787	080	41714	Banashan	Casual	Dir	Momos, U	Chinese, I	800	[[Rated 4.1]]	Buffet	Banashankari		
4	https://www.zomato.com/bangalore-restaurants/1112-mex-san-churro	Yes	No	3.8/5	918	+91	96634	Banashan	Cafe, Cas,	Churros, C	Cafe, Mex		800	[[Rated 3.1]]	Buffet	Banashankari		
5	https://www.zomato.com/bangalore-restaurants/1st-floor-addhuri	No	No	3.7/5	88	+91	96200	Banashan	Quick	Bite	Masala Dc	South	Indi	300	[[Rated 4.1]]	Buffet	Banashankari	
6	https://www.zomato.com/bangalore-restaurants/37-5-1-4th-timepass	Yes	No	3.8/5	286	+91		Basavan	Casual	Dir	Panipuri,	North	Indi	600	[[Rated 4.1]]	Buffet	Banashankari	
7	https://www.zomato.com/bangalore-restaurants/19-1-new-rosewood	No	No	3.6/5	8	+91		Mysore	Ri	Casual	Dining	North	Indi	800	[[Rated 5.1]]	Buffet	Banashankari	
8	https://www.zomato.com/bangalore-restaurants/2469-3rd-1-onesta	Yes	Yes	4.6/5	2556	080		Banashan	Casual	Dir	Famous	Pizza, Caf		600	[[Rated 5.1]]	Cafes	Banashankari	
9	https://www.zomato.com/bangalore-restaurants/1-30th-mc-penthouse	Yes	No	4.0/5	324	+91		Banashan	Cafe	Pizza,	Moi	Cafe, Itali		700	[[Rated 3.0]]	[[RATED]]	n I had been to this plac	
10	ie are a bit things ter service w	[[Rated 4.1]]	[[RATED]]	n Top floor: no outdo though th'	a nice pla	[[Rated 1.1]]	[[RATED]]	n we had ni it turned	[[Rated 3.1]]	[[RATED]]	parking fe: it was litt	[[1]]						
11	https://www.zomato.com/bangalore-restaurants/2470-21-m-smaczeg	Yes	No	4.2/5	504	+91		Banashan	Cafe	Waffles,	F	Cafe, Mex		550	[[Rated 4.1]]	Cafes	Banashankari	
12	https://www.zomato.com/bangalore-restaurants/12-29-neo-cafe	Yes	Yes	4.1/5	402	080		Banashan	Cafe	Waffles,	F	Cafe		500	[[Rated 4.1]]	Cafes	Banashankari	
13	https://www.zomato.com/bangalore-restaurants/941-3rd-fl-cafe-shur	Yes	Yes	4.2/5	150	+91	97421	Banashan	Cafe	Mooltails,	Cafe,	Itali		600	[[Rated 1.1]]	Cafes	Banashankari	
14	https://www.zomato.com/bangalore-restaurants/6th-block-the-coffe	Yes	Yes	4.2/5	164	+91	97316	Banashan	Cafe	Coffee,	Sq	Cafe, Chir		500	[[Rated 4.1]]	Cafes	Banashankari	
15	https://www.zomato.com/bangalore-restaurants/111-sappal-cafe	Eleve	No	4.0/5	424	080	49577	Banashan	Cafe	Sandwich,	Cafe,	Com		450	[[Rated 2.1]]	Cafes	Banashankari	
16	https://www.zomato.com/bangalore-restaurants/1112-mex-san-churro	Yes	No	3.8/5	918	+91	96634	Banashan	Cafe, Cas,	Churros, C	Cafe, Mex		800	[[Rated 3.1]]	Cafes	Banashankari		
17	https://www.zomato.com/bangalore-restaurants/2303-21st-cafe-viva	Yes	Yes	3.8/5	90	080		Banashan	Cafe	Garlic	Bre	Cafe		650	[[Rated 2.1]]	Cafes	Banashankari	
18	https://www.zomato.com/bangalore-restaurants/241-4th-fl-catch-up	Yes	Yes	3.9/5	133	+91		Banashan	Cafe	Momos,	N	Cafe, Fast		800	[[Rated 1.1]]	Cafes	Banashankari	
19	https://www.zomato.com/bangalore-restaurants/405-24th-kirthi-s-bir	Yes	Yes	3.8/5	144	080		Banashan	Cafe	Pasta,	Gel	Chinese, C		700	[[Rated 3.1]]	Cafes	Banashankari	
20	https://www.zomato.com/bangalore-restaurants/504-cj-ve-t3h-cafe	No	No	3.9/5	93	+91	88847	Banashan	Cafe	Cheese	M	Cafe, Itali		300	[[Rated 4.1]]	Cafes	Banashankari	
21	https://www.zomato.com/bangalore-restaurants/47-48-8-49-360-atom	Yes	No	3.1/5	13	+91	98089	Banashan	Cafe	Cafe,	Chir			400	[[Rated 5.1]]	Cafes	Banashankari	
22	https://www.zomato.com/bangalore-restaurants/146-50-ft-the-vinta	Yes	Yes	3.0/5	62	+91		Banashan	Cafe	Burgers,	C	Cafe, Frer		400	[[Rated 2.1]]	Cafes	Banashankari	
23	https://www.zomato.com/bangalore-restaurants/3353-2nd-woodee	Yes	No	3.7/5	180	+91	74068	Banashan	Cafe	Pizza,	Gar	Cafe, Pizzi		500	[[Rated 3.1]]	Cafes	Banashankari	
24	https://www.zomato.com/bangalore-restaurants/srf-comp-cafe	Coffi	No	3.6/5	28	080	32486	Banashan	Cafe	Cafe,	Fast			900	[[Rated 4.1]]	Cafes	Banashankari	

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,listedin(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.

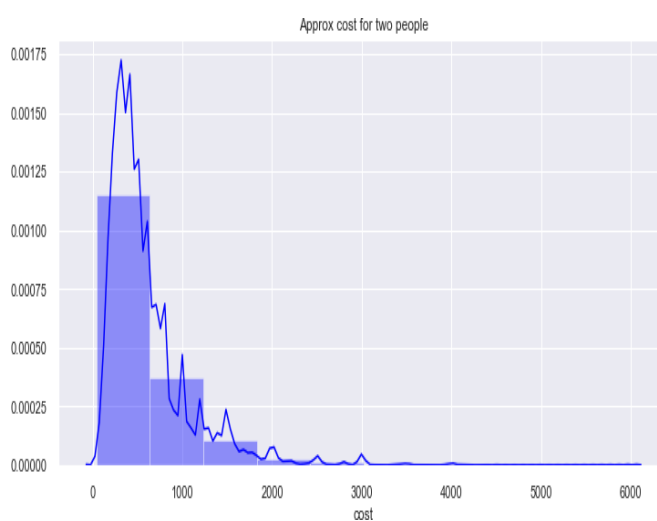
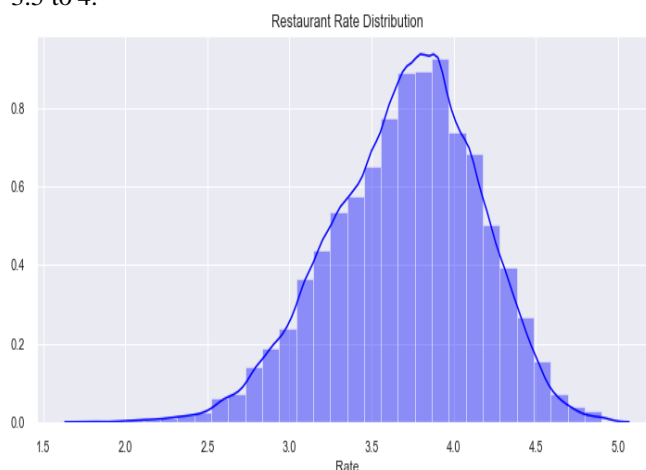
A. Restaurant Rate Distribution

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.

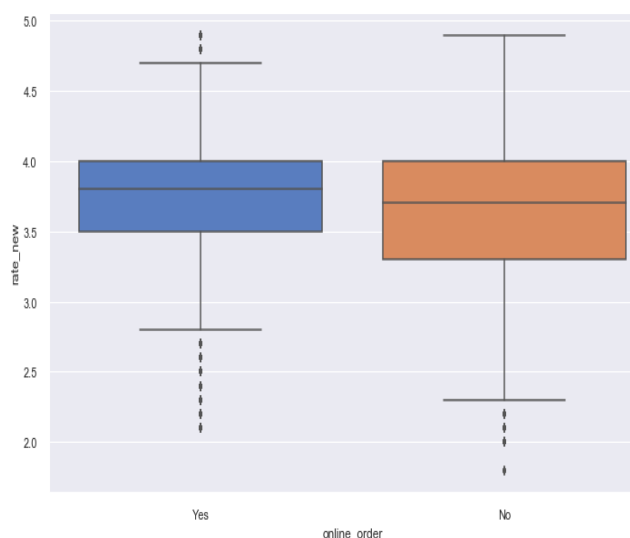
B. Approximate Cost of two people

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.

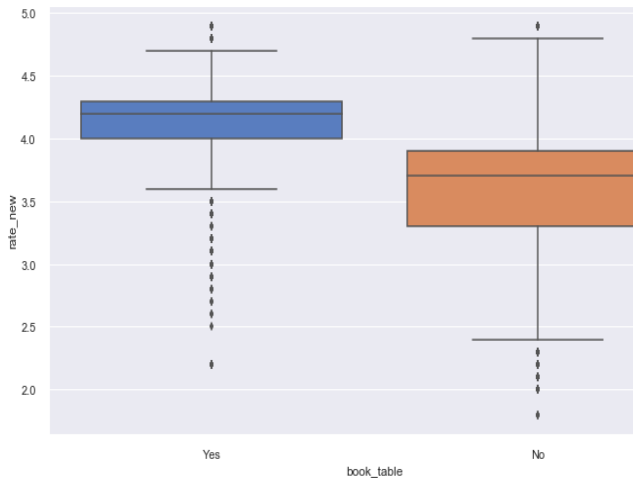
This box plot helps us look into the outliers. We can also see that online ordering service also affects the rating. Restaurants with online ordering service have a rating from 3.5 to 4.



C. Online ordering with respect to Rating(Finding Outliers)

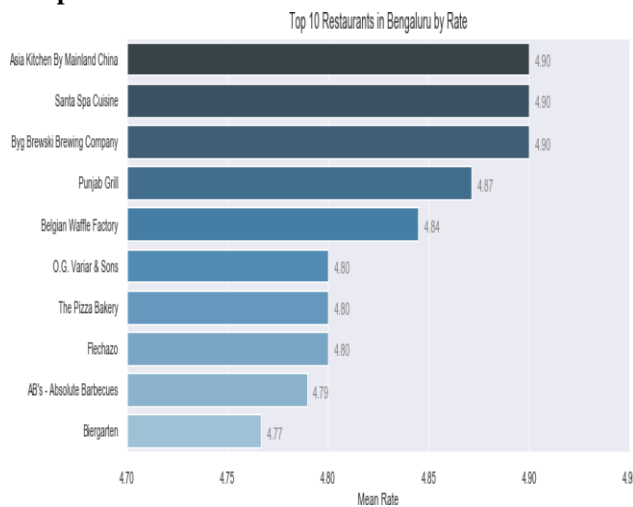


D. Booking table with respect to rating (Finding Outliers)



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



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

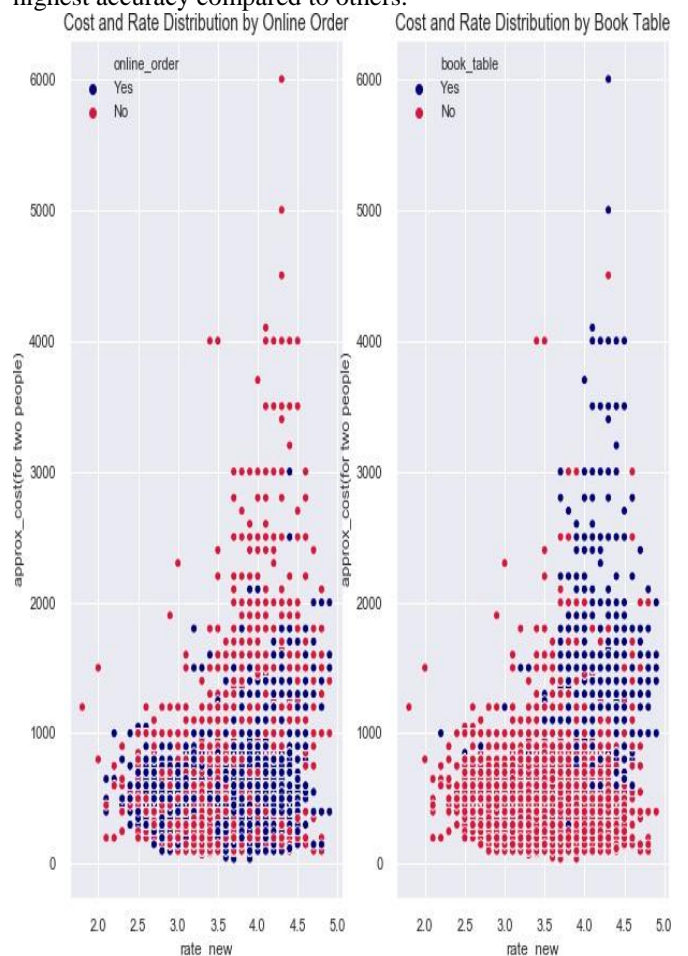
	Votes	approx_cost(for two people)	Rating
Book_table			
No	204.580566	482.404625	3.620801
Yes	1171.342957	1276.491117	4.143464

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 learning Machine Learning Techniques for Restaurant review. 2018 4th International Conference on Computing Communication and Automation (ICCCA) DOI:/10.1109/CCAA.2018.8777649
3. 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
4. Neha Joshi. A Study on Customer Preference and Satisfaction towards Restaurant in Dehradun City. Global Journal of Management and Business Research(2012) Link: <https://pdfs.semanticscholar.org/fe5/88622c39ef76dd773fcad8bb5d233420a270.pdf>
5. Bidisha Das Baksi, Harrsha P, Medha, Mohinishree Asthana, Dr. Anitha C.(2018) Restaurant Market Analysis.
6. International Research Journal of Engineering and Technology (IRJET)Link:<https://www.irjet.net/archives/V5/i5/IRJET-V5I5489.pdf>
7. f