Restaurants Rating Prediction using Machine Learning Algorithms

Atharva Kulkarni[1]

Student, M.Sc (BDA)

MIT-WPU

Pune, Maharashtra, India

Divya Bhandari[2]

Student, M.Sc (BDA)

MIT-WPU

Pune, Maharashtra, India

Sachin Bhoite[3]

Assistant Professor

Department of Computer Science, MIT-WPU

Pune, Maharashtra, India

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 main goal of this is to get insights on restaurants which people like visit and to identify the rating of the restaurant. With this article we study different predictive models like Support Vector Machine (SVM),Random forest and Linear Regression, XGBoost, Decision Tree and have achieved a score of 83% with ADA Boost.

Key Words: Pre-processing, EDA, SVM Regressor, Linear Regression, XGBoost Regressor, Boosting.

---

1. INTRODUCTION

Zomato is the most reputed company in the field of food reviews. Founded in 2008, this company started in India and now is in 24 different countries. Its is so big that the people now use it as a verb. "Did you know about this restaurant? Zomato it". The rating is the most important feature of any restaurant as it is the first parameter that people look into while searching for a place to eat. It portrays the quality, hygiene and the environment of the place. Higher ratings lead to higher profit margins. Notations of the ratings usually are stars or numbers scaling between 1 and 5.

Zomato has changed the way people browse through restaurants. It has helped customers find good places with respect to their dining budget.

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

2. RELATED WORK

Various researches and students have published related work in national and international research papers, thesis to understand the objective, types of algorithm they have used and various techniques for pre-processing and feature selection.

[1] Shina, Sharma S. and Singha A. have used Random forest and decision tree to classifying restaurants into several classes based on their service parameters. Their results say that the Decision Tree Classifier is more effective with 63.5% of accuracy than Random Forest whose accuracy is merely 56%.

[2] Chirath Kumarasiri's and Cassim Faroo's focuses on a Part-of-Speech (POS) Tagger based NLP technique for aspect identification from reviews. Then a Naïve Bayes (NB) Classifier is used to classify identified aspects into meaningful categories.

[3] I. K. C. U. Perera and H.A. Caldera have used data mining techniques like Opinion mining and Sentiment analysis to automate the analysis and extraction of opinions in restaurant reviews.

[4] Rrubaa Panchendrarajan, Nazick Ahamed, Prakhash Sivakumar, Brunthavan Murugaiah, Surangika Ranathunga and Akila Pemasiri wrote a paper on 'Eatery, a multi-aspect restaurant rating system' that identifies rating values for different aspects of a restaurant by means of aspect-level sentiment analysis. This research introduced a new taxonomy to the restaurant domain that captures the hierarchical relationships among entities and aspects.

[5] Neha Joshi wrote a paper in 2012 on A Study on Customer Preference and Satisfaction towards Restaurant in Dehradun City which aims to contribute to the limited research in this area and provide
insight into the consumer decision making process specifically for the India foodservice industry. She did hypothesis testing using chi-square test.

[6] Bidisha Das Baksi, Harrsha P, Medha, Mohinishree Asthana, Dr. Anitha C wrote a paper that studies various attributes of existing restaurants and analyses them to predict an appropriate location for higher success rate of the new restaurant. The study of existing restaurants in a particular location and the growth rate of that location is important prior to selection of the optimal location. The aim is to the create a web application that determines the location suitable to establish a new restaurant unit, using machine learning and data mining techniques.

3. 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 17Columns and 51,000 Rows

3.1 PreProcessing

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).

3.2 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)

4. 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.

1) Restaurant Rate Distribution

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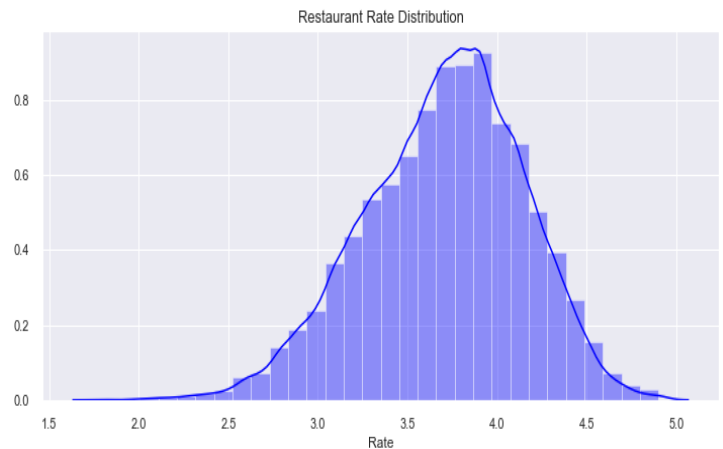
We can see that the number of restaurants with the rating between 3.5 and 4 are the highest. We will look into its dependencies further.

2) Approximate Cost of two people

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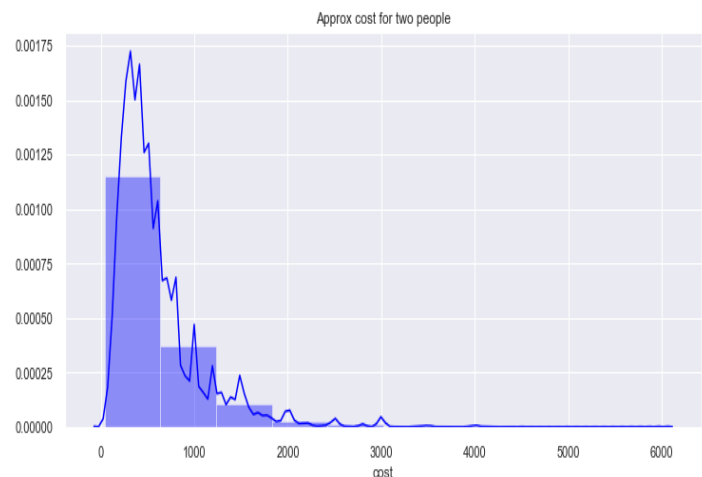
This is a graph for the 'Approximate cost of 2 people' for dining in a restaurant. Restaurants with this cost below 1000 Rupees are more.

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.
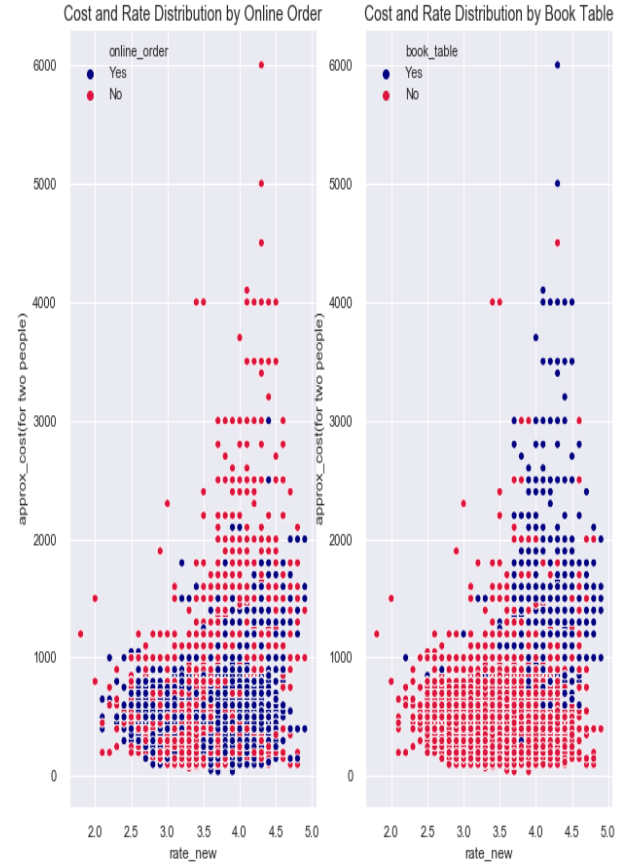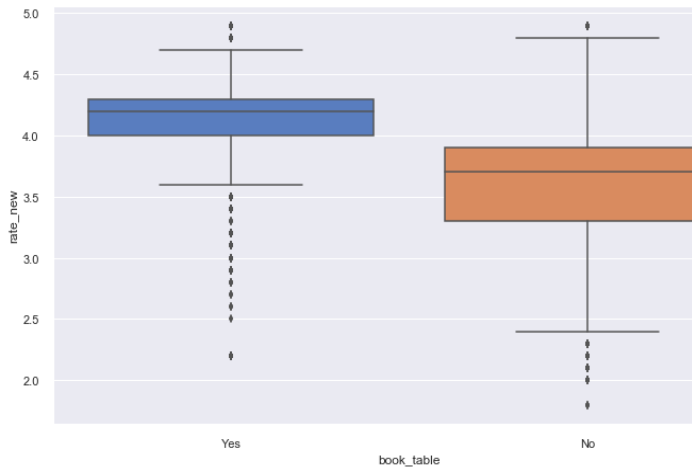
3) Online ordering with respect to Rating(Finding Outliers)
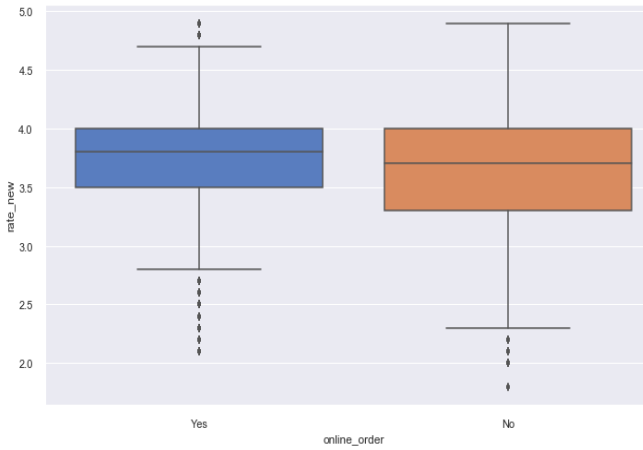
This graph just showcases the best restaurants in Bangalore along with their rating.

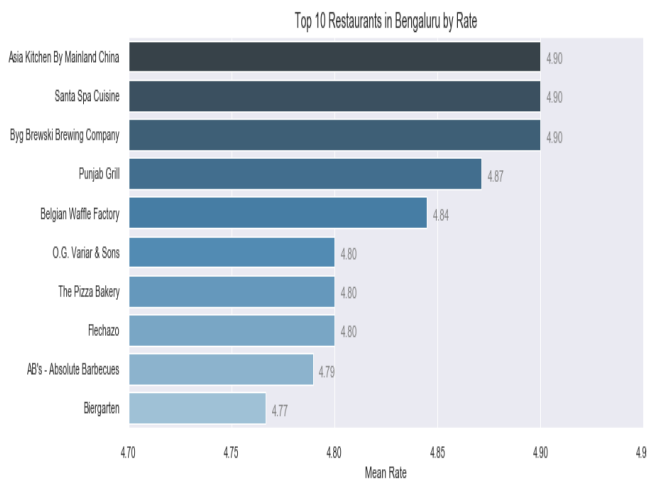6) Cost and Rate Distribution according to online ordering and booking table

4) 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.

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

4.1. Key Findings

5) Top Rated Restaurants

	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

5. RESULTS

Algorithms	Accuracy
Linear Regression	30%
KNN	44%
Support Vector Machine	43%
Decision Tree	69%
Random Forest	81%
ADA Boost(DT)	83%
XGBoost	72.26%
Gradient Boosting	52%

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 Ada Boost Regressor which gives the highest accuracy compared to others.

6. CONCLUSIONS

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.

Lot of researches have been made on factors which affect sales and market in restaurant industry. Various dine-scape factors have been analysed to improve customer satisfaction levels.

If the data for other citirs is also collected, such predictions could be made for accurate.

7. 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

[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] Rrubaa Panchendrarajan, Nazick Ahamed, Prakhash Sivakumar, Brunthavan Murugaiah, Surangika Ranathunga and Akila Pemasiri. Eatery – A Multi-Aspect Restaurant Rating System. Conference: the 28th ACM Conference

[5] 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/fef5/88622c39ef76dd773fcad8bb5d233420a270.pdf

[6] Bidisha Das Baksi, Harrsha P, Medha, Mohinishree Asthana, Dr. Anitha C.(2018) Restaurant Market Analysis.
International Research Journal of Engineering and Technology (IRJET)
Link: https://www.irjet.net/archives/V5/i5/IRJET-V5I5489.pdf