

Restaurant Rating Analysis

A Key to opening a successful eatery in Bangalore

Meeta Kapoor
Computer Science Department
San Diego State University



Introduction

Bangalore, being the tech capital of India has a lot of people from different cultures. With that being said, this city has a lot of restaurants offering different cuisines. From big food chains to cafes serving local delicacies Bangalore has them all. Going by the data, Bangalore has approximately 12000 restaurants yet there's a heavy influx of new restaurants in the city. However, new businesses find it hard to compete with the well established restaurants. Increasing food costs, expensive real estate costs, shortage of required manpower are some of the Herculean tasks new restaurants have to face. This project focuses on analyzing the factors which play a critical role in making the restaurant business a success. Through this effort we have tried to answer questions such as, does the demography of the restaurant matter, is understanding the people of the locality important, will the theme of the restaurant have a impact on its business, and many more. We accomplish this using data collected from Zomato which is a Food Delivery, Dining and Restaurant Discovery Service

Nearby restaurants

11T	Chinatown	Greece, Rochester, New...	3.8
11T	Chili's	Greece, Rochester, New...	3.5
11T	Mark's Pizzeria	Greece, Rochester, New...	2.9
11T	Caraglio's Pizza	Greece, Rochester, New...	3.4

zomato

Problem Motivation

Be it dine-in or delivery, one cannot deny that Bangalore is the city for foodies. Personally, I am in awe with the wide variety of cafes in Bangalore and the amazing quality of coffee they serve. Based on the dining establishments and the fact that market has not yet reached saturation in Bangalore yet, I wish to open my own cafe there someday.

With this project, we wish to assist freshly opened eateries in comprehending the elements that contributed to a specific restaurant's success. Additionally, using the results of our investigation, we derive insights on a restaurant's overall rating, which can aid customers in choosing a restaurant with knowledge.

Approach

Understanding the dataset used for this analysis would be the first step towards our goal. We will perform data validations such as identifying missing values, redundant rows, etc. This would constitute the pre-processing stage of our project. Next, we would implement data visualization to answer certain questions about the restaurants. Lastly, we will predict the success ratings of restaurants using the feature values from the dataset by enhancing the features to aid our analysis. We will implement multiple regression algorithms to achieve the same and identify which gives the most accurate results.

Approach Details

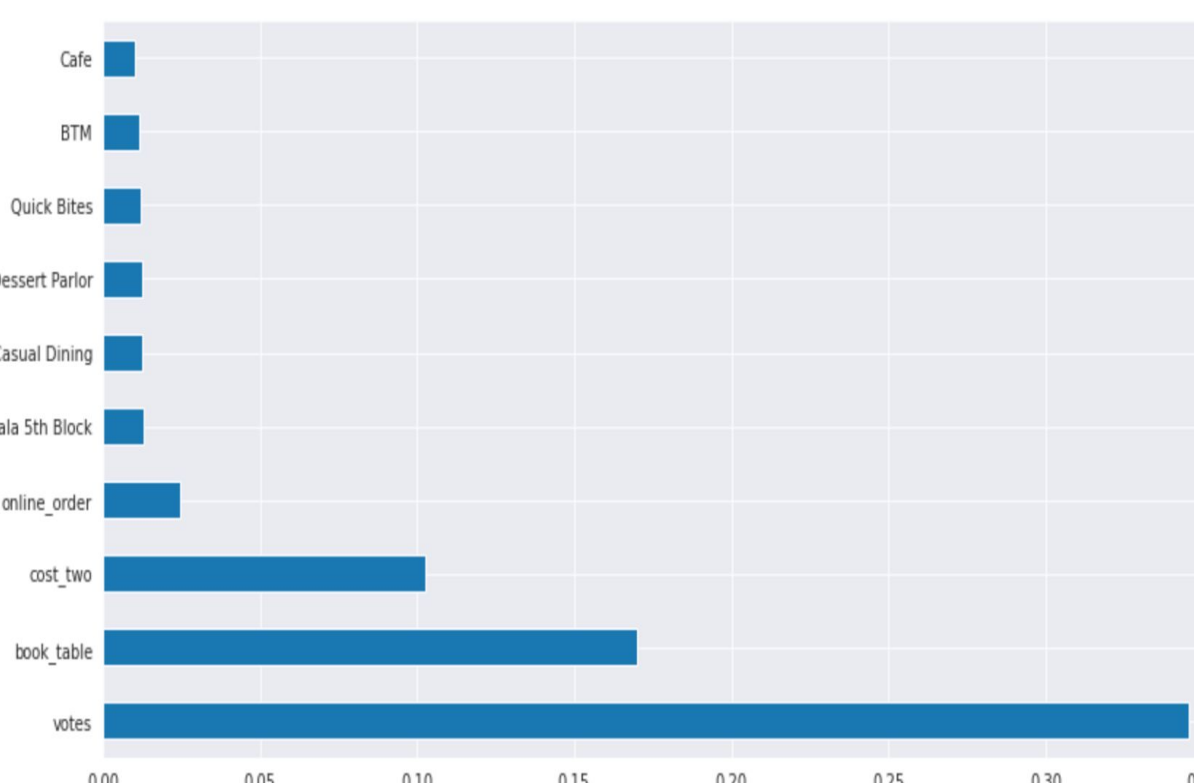
We use the following steps that play key role in the Life Cycle of any Data Science Project

- Exploratory Data Analysis - Understanding the available data is the first and crucial step of Data Science Cycle. This is where EDA or Exploratory Data Analysis comes into picture. EDA is used to derive meaning from the available data, identify anomalies or deviations by plotting informative visualizations. We have implemented EDA on Bangalore restaurants dataset to derive relations between the ratings and factors like votes, location and so on.
- Feature engineering - This would entail replacing null with mean values and modifying categorical data to numerical data that contributes to the analysis. We convert features like online order facility, locations the restaurant serves to etc into integer values.
- Feature Selection - We calculate the importance scores for all the features and recognize the features that play a crucial role in prediction. From the Feature importance scores we understand that Votes, table reservations, cost for two etc are main features that help in making restaurants a successful business.
- Model Building - We have used various regression models like Linear Regression, Random Forest Regressor & Decision Tree Regressor that form the core part of our analysis to predict the restaurant rating and choose the most accurate model.

Feature Correlation heatmap



Feature Importance Scores



Results

From our feature engineering & selection steps, we understand that the “Table reservations”, “Cost for two” & “Online order availability” features are highly correlated to “Rating” and have high importance scores, making them the features that impact the analysis most.

Linear Regression Algorithm

R-Square score: 33.431
Mean Absolute error: 3.7
Accuracy : 93.04

Random Forest Regressor Algorithm

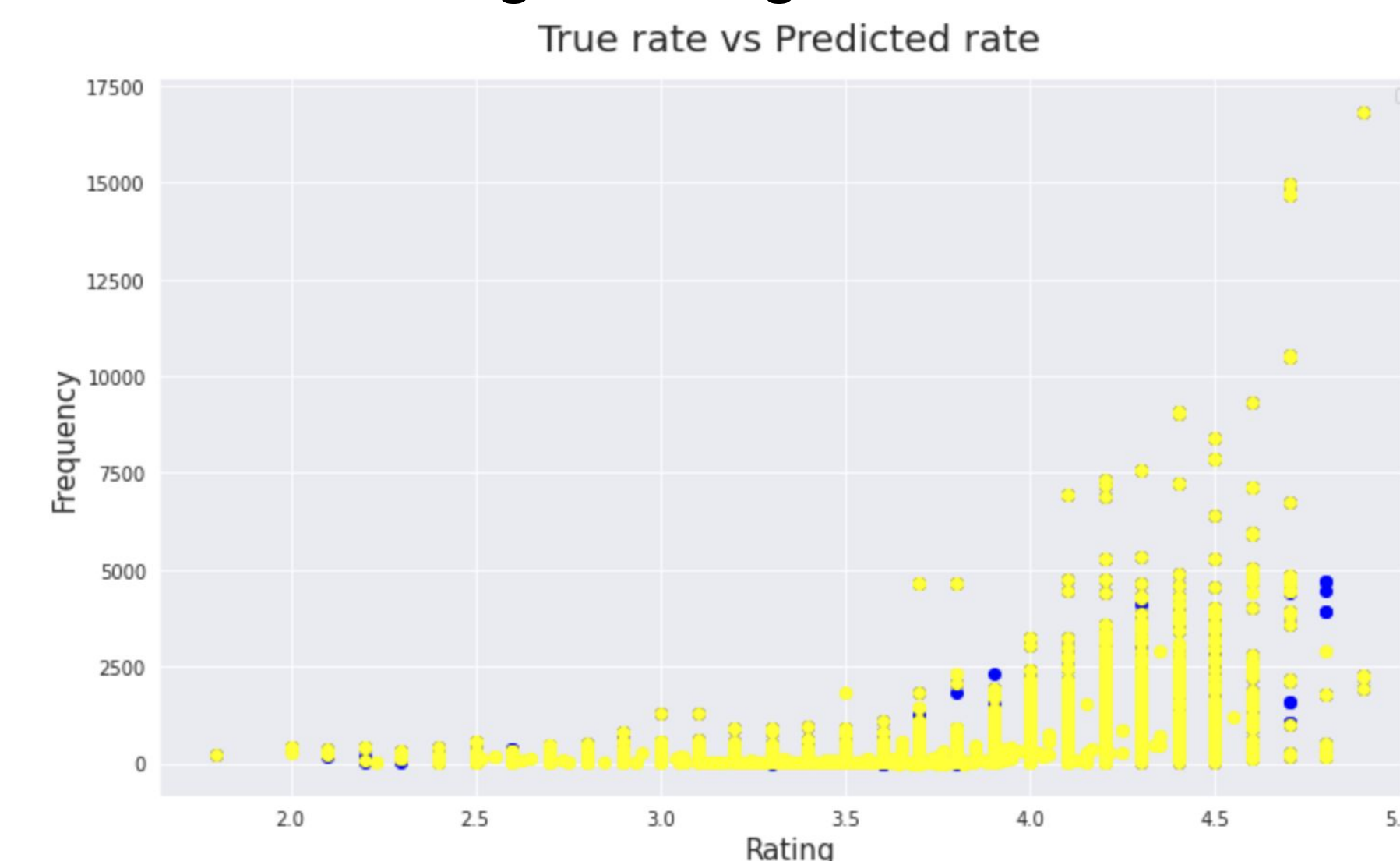
R-Square Score: 86.7407
Mean Absolute Error: 3.7
Accuracy : 98.87 %.

Decision Tree Regressor Algorithm

R-Square Score: 90.376
Mean Absolute Error: 3.7
Accuracy : 98.51 %.

Conclusion

Decision Tree Regressor Algorithm - BEST Solution!!



With the values derived from the algorithms (mentioned above), we can see that Random Forest Regressor has the best R-square score. However Decision tree Regressor algorithm is the most Accurate. This goes to show that a higher R-square Score isn't necessarily good for the regressor models. Going by the figures, we can conclude that Decision Tree regressor is the best module for our predictions.

References

- [1] Ahmed, Ahmet & Ali, 2018, Sudan, [A Decision Tree Algorithm Combined with Linear Regression for Data Classification](#)
- [2] Hai Wang & Fei Hao, 2012, Solan, [An efficient linear regression classifier](#)
- [3] Jitendra & Rita, 2017, Tiruchirappalli, [Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression](#)