Analysis Of Restaurant Cuisines Based on Locality, Dining And Delivery

Akshat Chourey
Computer Science And Engineering
PES University
PES2UG19CS028
akshatchourey123@gmail.com

G. Vijit Kumar Chary
Computer Science And Engineering
PES University
PES2UG19CS125
vijitkchary2001@gmail.com

Melavoy Nithin Reddy
Computer Science And Engineering
PES University
PES2UG19CS230
nithinreddy618@gmail.com

Nidhi Deshpande
Computer Science And Engineering
PES University
PES2UG19CS255
nidhidespande83@gmail.com

***Abstract*—The main purpose of this study was to examine the direct effects of restaurant location between dining and delivery service with the help of dining and delivery service review. Consumer surveys have indicated that people value the convenience and time savings of having their food delivered to them.In addition, many office-workers prefer the convenience of having food delivered to them rather than going out for lunch.**

***Index Terms*—Restaurant location, Dining services, Delivery services.**

I. Introduction

According to Deepinder Goyal, Zomato chief executive officer and co-founder told Tech Crunch that he expects to reach 10000 restaurants in Asian country throughout a couple of months. "We have a sales team of around three hundred in Asian country and 5000-odd advertisers... these partners understand the amount we tend to bring back them thus it's quite simple for the U.S.A. to launch this new service."

The impact that is made by the location significantly govern whether its delivery service or dining service will be popular as location is the only aspect of restaurant business which is very expensive to change which depend owner budget and restaurant popularity setting aside location aspect .In this work, we have chosen a restaurant based on Zomato dataset and have investigated to analyze how different cuisines are famous for their respective restaurants based on its location by analyzing dining review count and delivery review count.To address this question, we analyzed the Zomato restaurant data-set of Bangalore - also called The Silicon Valley of India, Bangalore has a population of 10 million and is the country's third most populated city. This paper aims to analyze the types of restaurants available in a particular locality of Bangalore and how the location impacts its dining reviews as compared to its delivery reviews. The rest of the paper is organized in the following manner. First, the literature is reviewed on restaurants, various factors affecting the success or failure of restaurants like location, cuisines by comparing its dining and delivery ratings. Then research methodology is used to analyze the data and graphically represent the empirical results.

An effective cuisines recommendation is a strategy that provides the opportunity to offer convenience to customers as well as to improve profits for restaurants.Different people have different tastes and certainly this has huge effect on what they would like to eat and where do they wanna dine.This is again impacted hugely by certain factors such as cost effectiveness and availability of the choices customer has.

II. Similar Approach

A. What Makes Restaurants to choose Delivery over Dining Services?

1.Convenience[1] is one of the benefits associated with outsourcing delivery services as it is more convenient from going there by yourself instead the food will be delivered at any time in the working hours in which most of the responsibility of picking up and food could be delegated to the online

delivery services or when someone want to avoid cooking the food by themselves.

2.Location [2]is the most important aspect of whether the restaurant will be a success or a failure.So because of this , restaurants decided to collaborate with online delivery service because the restaurant that is located in areas where some college or place where the crowd gathers in larger number could result in the restaurant location being less accessible to potential customers where they face large delays in reaching the restaurant. Hence, customers who visit the restaurant could encounter problem such as long traffic jams and limited availability of parking or under developed area.

Limitations: The main work done in this paper was comprehensive study of how success of a particular restaurant could depend on factors such as convenience and location.But what they failed to take into consideration is how varieties of food provided in those restaurants had similarly a huge impact too. And also there was no definitive quantitative study done in this paper. Hence we tried to work upon this and include the aspects were this paper failed.

B. Collaborative Study on User ratings on Restaurant Recommendation[3]

This paper used Collaborative filtering algorithm to train the model through the user's scoring information , and uses this model to predict the unknown data.The purpose is to predict ratings behavior of user who use the system, then according to the predicted scores to generate recommendation,to show the effectiveness of the proposed **AdvancedCF algorithm**, they compared the performances of the algorithm against **UCF**.Smaller the MEA value,better the performance.[4] They set the value of K,the k nearest neighbours, from 5 to 50, increasing 5 each time.They were able to get the accuracy in the range of 65-80 percent, which was quite impressive.

Limitations:The work done in this paper was quite definitive and the had a very well made model that generated quite a good accuracy.The paper was so focused on the exhaustive study of users' reviews and getting the score that they didn't put the importance of side features into consideration.Also they failed to take into account the features like cuisines and prices of the restaurants that play a huge role in recommendation system.

C. Restaurant Analysis and Prediction[5]

The focus of this paper was to make a model to predict the ratings of restaurants based on the information contained in the data set.The data was further processed by applying One-hot encoding to the categorical data.For prediction of values several Machine learning methods were used such Ridge [6] regression ,k-Nearest neighbours,Logistic regression,lasso regression and Random Forest regression. The accuracy of each algorithm was calculated for the testing data set.Out of these methods Random Forest Regression resulted in the best accuracy with accuracy coming out to be around 80 percent.

Limitations: The paper used several types of regression which although provide good accuracy, have some tweaks too. These include that these model do not work well if input data has errors, are susceptible to collinear problems, and work well only for low number of variables; which certainly was the case in the dataset this paper used. Hence we chose a dataset that had sufficient variables and had no such discrepencies.

III. Proposed Solution

Using Exploratory Data Analysis and Predictive statistics on the various restaurants available in Bengaluru,we would like to understand the basic preferences of customers in terms of cuisines on the basis of locality,prices and dining reviews for paricular restaurants.

The dataset[7] we will be using in our project is the Zomato Resaturants in Metropolitan Areas of India(Bangalore) from Kaggle(https://www.kaggle.com/aestheteaman01/zomato-restaurant-caf-dataset-pune-bangalore). This dataset contains 13 features and 5110 rows of which there are 4882 unique values.

A. Pre-processing

- Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

- Before getting on to the usage of the dataset we need to clean it.The datatset has a total of 1436 missing values. We impute these missing values with the mean.Also there were some values which were zero, we again inputed them with mean values.The following figure shows missing values:

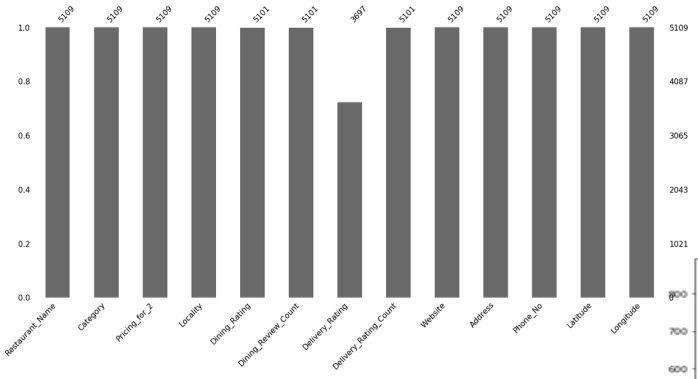

Fig. 1. Missing number of Values.

We choose to drop the Website,Address, Phone_No,Latitude and Longitude columns from the dataset as they were not giving us any meaningful data(redundant) that could have been worked upon.

Some features had to be modified so as to get a better understanding about them.Names of some columns such as Dining_Rating_Count, Dining_Rating etc. were fixed for easier readability.The duplicate rows were removed from the dataset.

B. Exploratory Data Analysis

Dataset Inspection. This phase includes inspection of the data,and visualisation by plotting various graphs between different attributes of the dataset and drawing conclusions from them.

- We visualise the number of restaurants at each of the particular locality in Bangalore.We can say some locality are food-hotspots as they have more amounts of restaurants as can be seen in Fig2.

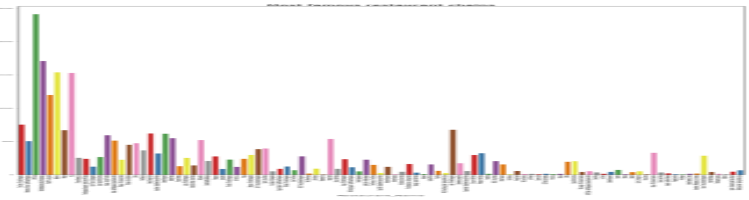

Fig. 2. Restaurants in each locality.

- We then try to find out the approximate cost of two based on the number of cuisines provided by the restaurants.Thus we can say that the more cuisines provided by the restaurants, the more the price increased indicating the quality of food increasing.

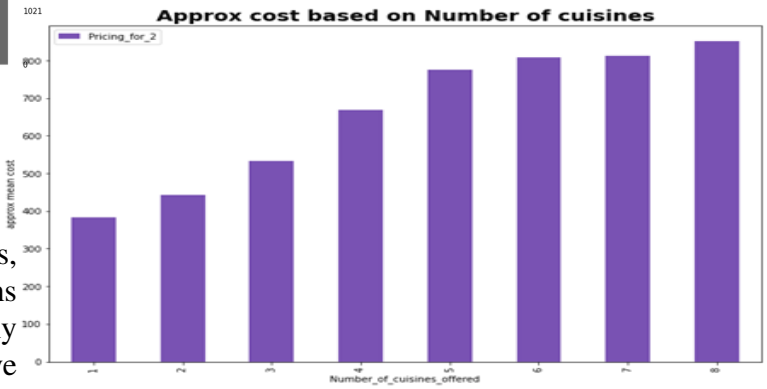

Fig. 3. Price For 2 for number of cuisines.

- We tried to visualise distribution of mean average cost at different location. Finding out in the process that average cost hovered around 800 which signified the costliness and quality of food and cuisines that were offered by the restaurants as can be seen in Fig4.

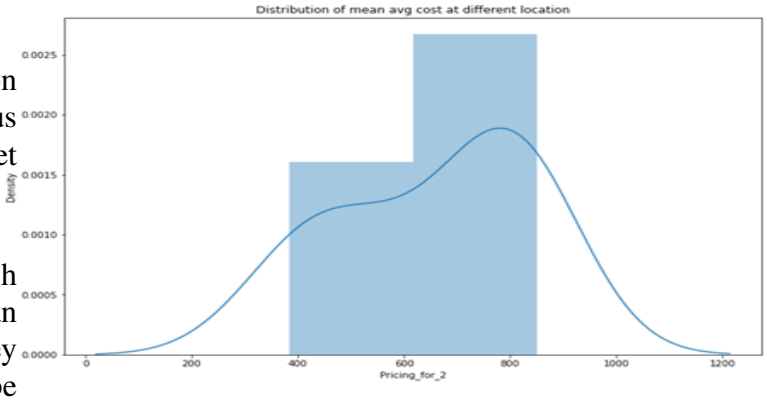

Fig. 4. Distribution of mean average cost at different.

- We then visualised the most popular cuisines in Bangalore on the basis of how many restaurants had included these cuisines in their menu and that had got good ratings on both dining and delivery.This could be seen in Fig. 5.

- Some restaurants had large number of diningreviews as well as delivery reviews indicating that particular restaurants were particularly famous in that location.

Fig. 5. Most popular cuisines in Bangalore.

- Different restaurants have different types of cuisines available.People we free to choose from an enormous ”menu” and had varieties of food available for them.

C. Models and output

The following models were used to get a definitive and predictive analysis of our work and understand the accuracy and precision of the analysis:

- **K-Nearest-Neighbors Regressor**
  We used K- Nearest Neighbor Regressor model which gave us the MSE value of:

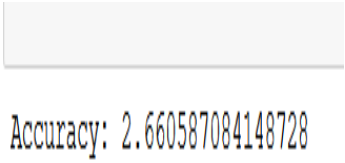

Fig. 6. Accuracy for KNNRegressor.

This MSE value is not desirable.Hence we can say this model is not appropriate for this type of analysis.

- **Stochastic Gradient Descent Classifier**

We used a classifier model in SGDClassifier to try to get a better accuracy for our analysis. In this model we tried to classify the number_of_cuisines_offered on the basis of Locality of the restaurant,DiningCount of the given restaurants,DiningRC,pricing_for_2 and Category or the cuisines the restaurant provided. The evaluation metric used was accuracy score which came out to be 37.6 percent as shown:

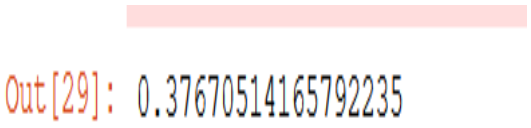

Fig. 7. Accuracy For SGDClassifier.

This value again shows that this model was not suitable enough for the analysis.

- **Decision Tree Classifier**
  Finally we tried making the Decision Tree Classifier.Just like before we tried classifying the number_of_cuisines_offered on the basis of the different attributes.We observe a result as 65.8 percent as shown in Fig.8 :

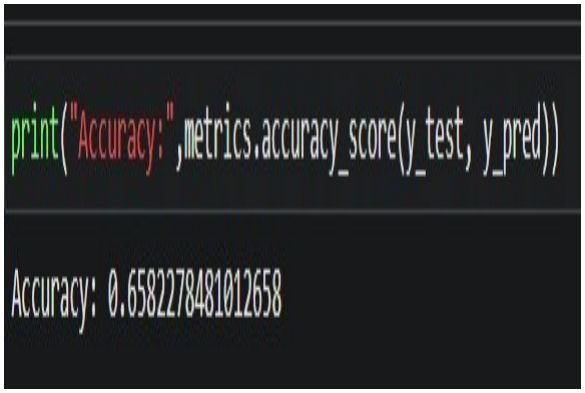

Fig. 8. Accuracy For DecisionTreeClassifier.

We can see that this is a fairly good classifying accuracy. But still Decision Trees suffer from instability if there are small changes introduced. Hence it is difficult to get a proper splitting attributes for the decision trees.

IV. Conclusion

After making and using different models for our dataset we were fairly able to understand the dataset and analyse it.Exploratory Data Analysis was performed to better understand the features of our dataset and further explore how they were related to other features.Some useful insights were gained using EDA and was certainly useful in better analysis and prediction on our dataset.

The models made were generally giving a low accuracy and MSE values, which indicate that the features were not as related to each other as had been initially thought of. This could be due to certain hidden factors such as availability of certain restaurants that provided those specific cuisines, mood and cravings of the customers and even the popularity of the restaurant in the locality.

The output accuracies indicate that the Classifier models worked better than Regressor model but certainly not good enough.This could indicate that model learning was not good fit for the given analysis.

Contribution: The work was equally divided among the team-members.Wherein Vijit and Nidhi were responsible for Pre-processing and EDA of the dataset to get insights about the features of the dataset. And Akshat and Nithin were responsible for making of various models and analysis and conclusion of our problem statement and hence get the final analysis.

V. References

[1] "TY - JOUR AU - Collier, Joel AU - Kimes, Sheryl PY - 2013/02/01 SP - 39 EP - 51 T1 - Only If It Is Convenient Understanding How Convenience Influences Self-Service Technology Evaluation VL - 16 DO - 10.1177/1094670512458454 JO - Journal of Service Research ER - "

[2] GohSee-Kong, NgSoo-Ryue, WongShiun-Yi, Chong-Lily "OUTSOURCING TO ONLINE FOOD DELIVERY SERVICES: PERSPECTIVE OF FI&B BUSINESS OWNERS"

[3] L. Li, Y. Zhou, H. Xiong, C. Hu and X. Wei, "Collaborative filtering based on user attributes and user ratings for restaurant recommendation," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2017, pp. 2592-2597, doi: 10.1109/IAEAC.2017.8054493.

[4] R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.

[5] Predicting Restaurants' Rating And Popularity Based On Yelp Dataset-http://cs229.stanford.edu/proj2017/final-reports/5244334.pdf

[6] Zomato Bangalore Restaurant Analysis and Rating Prediction-https://purnasaigudikandula.medium.com/zomato-bangalore-restaurant-analysis-and-rating-prediction-df277321c7cd

[7] Aman Kumar, "Zomato Restaurants in Metropolitan Areas of India." Kaggle, 2021, doi: 10.34740/KAGGLE/DSV/2572839.

[8] TY - BOOK AU - Adnan, Masrur AU - Sarno, Riyanarto AU - Sungkono, Kelly PY - 2019/09/01 SP - 121 EP - 126 T1 - Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm DO - 10.1109/ISEMANTIC.2019.8884282 ER -