JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Restaura: Restaurant Analytics with Model Solutions

Navneeth S Holla
Department of Computer Science and Engineering
PES Univeristy
Bengaluru, India

Siddharth K Rao
Department of Computer Science and Engineering
PES Univeristy
Bengaluru, India

Ritvik Sanjeev Patil
Department of Computer Science and Engineering
PES Univeristy
Bengaluru, India

Vignesh Veeresh Anweri
Department of Computer Science and Engineering
PES Univeristy
Bengaluru, India

Suresh Jamadagni
Department of Computer Science and Engineering
PES Univeristy
Bengaluru, India

***Abstract*— With the advent of apps like google maps, Zomato, Swiggy, etc., it is possible to book tables and order food from our favorite restaurants with the click of a button. We can also rate and voice our opinions about restaurants through reviews on these platforms (google reviews, Zomato, Swiggy, etc.). These apps manage and handle a lot of restaurants and user data. Fortunately, apps like Swiggy and Zomato have made certain sections of their data public so that interested individuals and groups can analyze them. We have used this data to predict the success of a new restaurant and provide a real time restaurant rating system.**

This paper focuses on providing a success probability of a new restaurant. This is accomplished by using a machine learning model which is trained on past restaurant data. It considers the location of the restaurant, competition in the area, delivery option and many other factors for predicting the success rate. The paper also focuses on providing a real time leaderboard system for top performing restaurants in Bangalore. The leaderboard also provides the current performance of these restaurants based on ratings and reviews collected and consolidated bi-weekly. The main goal of this paper is gathering and analyzing restaurant data to provide value to aspiring entrepreneurs and passionate foodies.

I. Introduction

Apps like google maps, Zomato, Swiggy, etc., have changed the way we operate in our daily lives. It is possible to book tables and order food from our favourite restaurants with the click of a button. We can also rate and voice our opinions about restaurants through reviews on these platforms such as Google reviews, Zomato, Swiggy, etc. These apps manage and handle a lot of restaurants and user data. Fortunately, apps like Swiggy and Zomato have made certain sections of their data public so that interested individuals and groups can analyze them. We have used this data to:

i. Predict the success of a new restaurant
ii. Come up with a real time leaderboard system for top performing restaurants.

India has about 53,000 hotels and 70 lakh restaurants in the organized category and 2.3 crore restaurants in the unorganized sector [1]. With thousands of restaurants opening every year and an equal (if not more) number of restaurants closing their doors every year, there is a need for every restaurant to stay ahead of the competition. Thus, if you're an aspiring entrepreneur planning to open a restaurant, it is important to study the food industry, market, competition, location, cost and many such factors that will likely influence the success of your business. This paper aims on doing just that. We have built a machine learning model that learns from past data of all restaurants and provides the success probability of a new (or a hypothetical) restaurant. From the customer's perspective, quite often, the customer's experience of a restaurant changes from time to time. This can be due to several reasons such as change in staff, shortage of staff on a particular day, quality of food, etc. The second part of our paper aims at informing the users about how a restaurant is performing recently by gathering and taking the average of all the recent reviews and ratings

II. Literature Review

A. *Restaurant Market Analysis*

Research from 2018 shows the usage of attributes of current restaurants to suggest optimal and best location for higher success and growth rate while opening a new restaurant. Analysis of the restaurant market and its growth rate in a particular area is critical while choosing location while opening a new restaurant. Data mining and Machine Learning techniques are used while determining the location [2].

B. *Restaurant Review Analysis and Classification using SVM*

The paper classifies restaurant reviews using ML classification. Reviews can be about food quality, ambience, staff service etc. Reviews are classified using SVM. This can be useful while determining drawbacks and limitations of an area and how

to better the overall standard of food, services etc. Classification is performed using cloud infrastructure [3].

C. *Predicting the Helpfulness of Online Restaurant Reviews Using Different Machine Learning Algorithms*

Here, the enormous amount of text reviews and number ratings produced by customers on food review websites are analysed in a viable way. Website like Yelp, Zomato, etc. can use these techniques for better understanding of customers and provide them with improved restaurant recommendations. Restaurant owners can also use this information to understand their customer to a greater extent and exploit the competitive edge in the restaurant market [4].

D. *Sentiment Analysis of Food Reviews Using User Rating Score*

This study picks out the best product(restaurant) using sentiment analysis for the consumer. Text data in reviews are studied and classified as neutral, positive, negative based on a recommendation system. Sentiment analysis is the buzzword in today's technological industry, especially in the consumer market. System identifies not only opinions, but also features of the sentiment. Under natural language processing, binary classification of reviews and rating according to the sentiment is very widely used. This paper provides the optimal solution to picking the best classification algorithm with best prediction accuracy. For this purpose, the reviewer's weightage/score is considered [5].

III. Data Procurement

The data procurement is performed by two bots that scrape data from the Zomato API endpoints periodically. Taking scale into consideration, we confine our radius of interest to the city of Bangalore for now. As of now we aim to collect restaurant details every week and restaurant reviews and ratings daily. The two bots are built using python, with the help of libraries such as requests, json, csv, pandas etc. Bot scheduling takes place using libraries schedule and time. These bots will be hosted on a server, which will then be pipelined to our online analysis and ML model, to keep them up to date.

IV. Dataset Description

We obtain two datasets from our bot:

1. Restaurant Details

This data consists of all the information related to a restaurant, hotel, eatery etc.

Attributes consists of:

- **rest_id**: Each restaurant is given a unique ID to differentiate restaurant chains as well as restaurants with the same name.
- **name**: Name of the restaurant, hotel, eatery, etc.
- **dining_rating**: Rating of the restaurant based on customer experience of dining in the place.
- **dining_review_count**: Number of dine in reviews received by the restaurant.
- **delivery_rating**: Rating of the restaurant based on customer experience of taking delivery from the place.
- **delivery_review_count**: Number of delivery reviews received by the restaurant.
- **locality_name**: Name of the locality/area of where the restaurant is located.
- **locality_address**: Specific address of the restaurant.
- **cuisine**: Type of cuisines served by the restaurant, for example Continental, North Indian, Mexican, etc.
- **online_order**: whether online ordering is available in the restaurant or not.
- **book_table**: online table booking option available or not.
- **rest_type**: restaurant type.
- **cost**: Cost of dining in or delivery for two people in Rupees.
- **dining or delivery**: Tells us if a particular restaurant has only dine in, or only delivery or both.

2. Review Details

Consists of information related to the reviewer, the review and ratings given the reviewer.

Attributes consists of:

- **rest_id**: Unique ID of the restaurant. Same as that from the Restaurant Details dataset.
- **reviewText**: The review given by the user.
- **userReviewsCount**: Total number of reviews given by the reviewer on the website.
- **userFollowersCount**: Number of followers the reviewer has.
- **likeCount**: Number of likes the review has.
- **experience**: Whether the reviewer has had a dine in or delivery experience.
- **ratingText**: The total rating out of 5 is divided into 'horrible', 'bad', 'average', 'good', 'excellent'.
- **rating**: Rating given by the reviewer out of 5.

V. Exploratory Data Analysis

To come up with the machine learning model, one must understand the data first.

A. *Online Platform Conformity of Restaurants*

In this analysis, we look at restaurants in Bengaluru that conform to Zomato's online standards such as online order and table booking. There are some restaurants that facilitate the feature of ordering food online, while some do not. The pie chart below gives some insights about that. Similarly, few restaurants allow table bookings to be done online, while some do not. The pie chart shows the percentage of restaurants that provide the option of booking tables online vs the restaurants that do not provide the option.

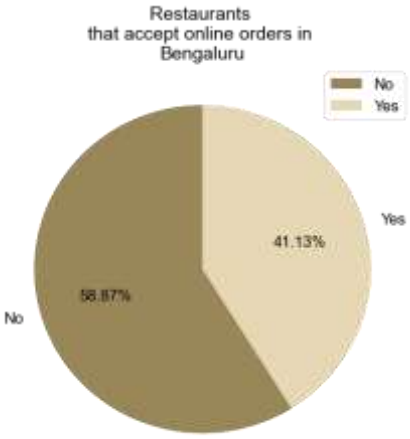

Fig. 1. Restaurants accepting online orders in Bengaluru

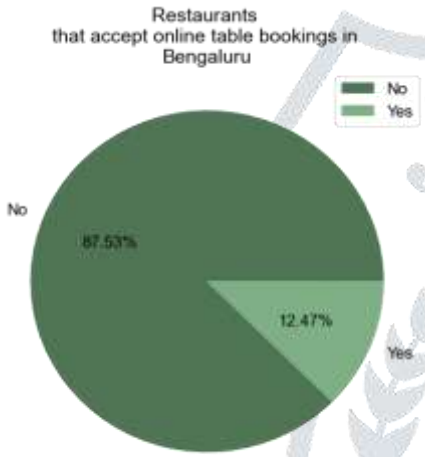

Fig. 2. Restaurants accepting online table bookings in Bengaluru

B. *Popular Cuisines and the Foody Areas of Bengaluru*

The restaurants in Bengaluru are known for offering a wide variety of cuisines. From the analysis done and the bar chart below shows the most popular cuisines in Bengaluru. The areas of Bengaluru having the greatest number of restaurants are termed as the Foody areas of Bengaluru.

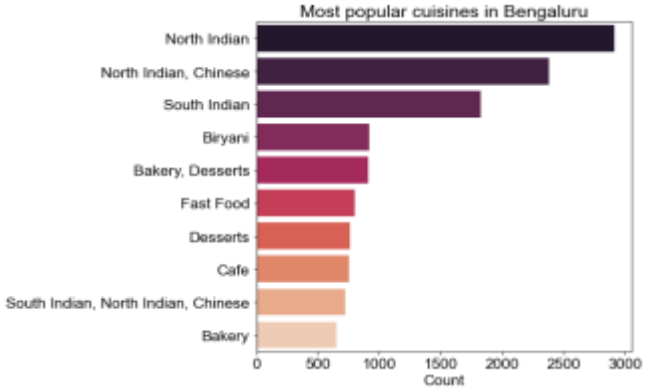

Fig. 3. Popular cuisines in Bengaluru

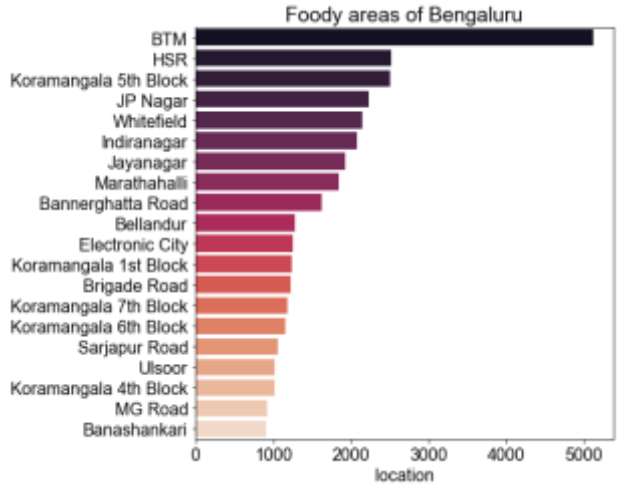

Fig. 4. Foody areas of Bengaluru

C. *Cost Distribution of Restaurants*

The cost distribution of restaurants in Bengaluru is found to be right skewed from the analysis done. It is found that many restaurants in Bengaluru offer food at a reasonable price.

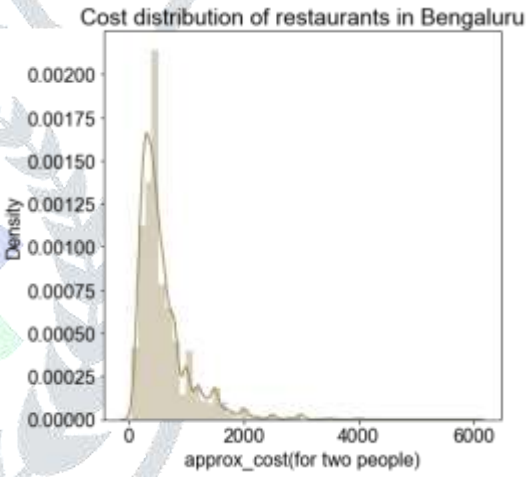

Fig. 5. Cost distribution of restaurants in Bengaluru

D. *Heat Map Analysis*

A Heat Map is a data visualization technique that shows the magnitude/density of a phenomenon using colour in a two-dimensional space. From the below heat map of South Indian restaurants across Bengaluru, we get to know how densely populated different areas of Bengaluru are with South Indian restaurants.

Fig. 6. Cost distribution of restaurants in Bengaluru

VI. Classification Model

To predict the success percentage of a restaurant, we train different classification models. The initial set of selected features include online_order, book_table, location, rest_type, cuisines, listed_in (type), listed_in (city), approx_cost. We created the multiple_types, total_dishes, and total_cuisines feature in a way of counting the food services offered by the restaurant. We build a complete pipeline for receiving the raw data with restaurant information and apply all the steps we select to make the data ready for training or prediction.

- Preparing the cost and rate attribute from raw data
- Selecting initial features to be part of data preparation
- Creating new features based on original data
- Creating a target for using on training
- Splitting restaurants based on rated and non-rated ones
- Encoding the data for categorical features
- Filling the null data with the median for numerical features

The four methods that are used for training are:

- Logistic Regression
- Decision Trees
- Random Forest
- LightGBM

A. Logistic Regression

The logistic regression model is a statistical method for binary classification that can be generalized to multiclass classification. It is the transformation of a linear regression using the sigmoid function. The logistic function applies a sigmoid function to restrict the y value from a large scale to range between 0 and 1. The results obtained by using logistic regression are given in the table below.

TABLE I. Tabulated Results Of Logistic Regression

Accuracy	71.03%
Precision	74.34%
Recall	60.73%
F1	66.85%
AUC	77.77%
Total Time	0.0300

B. Decision Trees

A decision tree is a supervised learning technique that has a predefined target variable and is most often used in classification problems. This tree can be applied to either categorical or continuous input & output variables. The training process resembles a flow chart, with each internal (non-leaf) node a test of an attribute, each branch is the outcome of that test, and each leaf (terminal) node contains a class label. The uppermost node in the tree is called the root node.

TABLE II. Tabulated Results Of Decision Trees

Accuracy	72.52%
Precision	75.31%
Recall	63.77%
F1	69.06%
AUC	78.43%
Total Time	0.0260

C. Random Forest

Random Forest is an ensemble method, meaning that a random forest model is made up of a large number of small decision trees, called estimators, which each produce their own predictions. The random forest model combines the predictions of the estimators to produce a more accurate prediction.

Standard decision tree classifiers have the disadvantage that they are prone to overfitting to the training set. The random forest's ensemble design allows the random forest to compensate for this and generalize well to unseen data, including data with missing values.

Random forests are also good at handling large datasets with high dimensionality and heterogeneous feature types.

TABLE III. Tabulated Results of Random forest

Accuracy	76.26%
Precision	80.70%
Recall	66.57%
F1	72.96%
AUC	83.34%
Total Time	0.4480

D. LightGBM

LightGBM is a gradient boosting framework that makes use of tree based learning algorithms that is considered to be a very powerful algorithm when it comes to computation. It is considered to be a fast processing algorithm. While other algorithms' trees grow horizontally, the LightGBM algorithm grows vertically meaning it grows leaf-wise and other algorithms grow level-wise. LightGBM chooses the leaf with large loss to

grow. It can lower down more loss than a level wise algorithm when growing the same leaf.

TABLE IV. TABULATED RESULTS OF LIGHTGBM

Accuracy	73.61%
Precision	79.17%
Recall	61.25%
F1	69.07%
AUC	80.65%
Total Time	0.1470

VII. LEADERBOARD

The leader board displays the current and the varying trends among the restaurants in Bengaluru. There are two leader boards giving insights into how the restaurant is performing when it comes to delivery and dining. Considering a window period of two weeks, new details such as ratings and top 5 reviews are collected for each restaurant. These new details are only collected if it varies from the previously collected data for these restaurants.

The average rating for a particular restaurant is calculated using weighted averages rather than the traditional raw data averages. One of the main reasons to do this is to avoid any kind of fake reviews or accounts. It is essential that the rating of a restaurant should represent what the people experience at a particular restaurant [6].

The formula used to compute the average rating:

Weighted Rating = (v / (v + m)) * R + (m / (v + m)) *C

Here,

- R (Rating) = Average rating for the restaurant
- v (Votes) = Number of ratings for the restaurant
- m = Minimum votes required to be listed in the Top 50 (currently 25)
- C = The mean vote across the whole report

VIII. RESULTS

On the customer side, the website helps in finding the current top performing restaurants in Bengaluru. The technique of weighted average rating for the restaurants in Bengaluru is applied here to find the change in ratings. The bot used to collect the details of restaurants consolidates the performance of the restaurants bi-weekly.

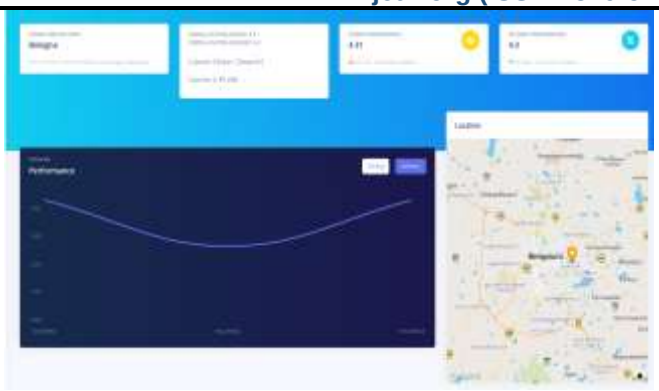

Fig. 7. Restaurant performance details (bi-weekly)

The website shows the locations of top performing restaurants in Bengaluru.

Fig. 8. List of top performing restaurants in Bengaluru (bi-weekly)

When it comes to the restaurant owner side, the website provides an easy interface for the user to input their restaurant parameters. Machine learning models are implemented to predict the success percentage of an unrated restaurant. Along with the success percentage, a detailed report of the factors affecting the success percentage of the users restaurant is generated.

Fig. 9. Form to fill in the restaurant parameters

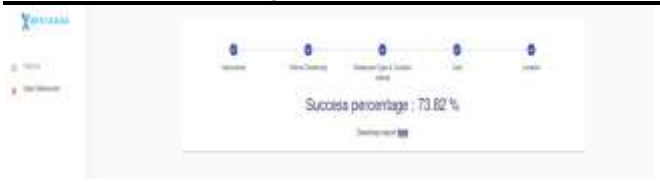

Fig. 10. Success percentage of a new restaurant

Main focus is to provide a GUI based web application to provide immersive user experience. There is a clear indication that differentiates the restaurant side and the customer side.

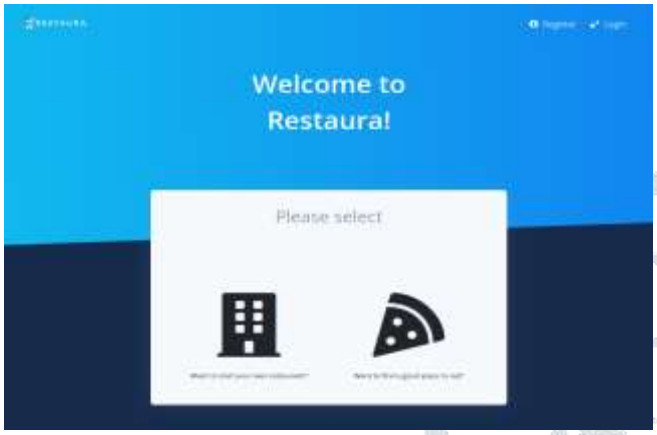

Fig. 11. Website home page showing the restaurant and the customer side

IX. Conclusion

Restaura is an application that predicts the success percentage of a new restaurant that a restaurant owner might plan on opening soon. This application also provides the customers with information about the currently well doing restaurants in the town or a particular area, based on various parameters collected from the restaurant and using a weighted average method to find the increase or decrease in the ratings.

This application tries to solve the issues faced by the new restaurant owners and keeps the customers updated about the current trends related to restaurants popularity.

X. Future Work

Some of the other features we plan on adding is to use certain NLP techniques along with sentiment analysis to take text reviews of restaurants and deduce the reason for the increase or decrease in the performance of a restaurant. We focus on improving scalability to handle operations outside Bengaluru as well and extend to other data sources like blogs, social media posts, tweets, etc.

XI. References

[1] [Online]Available: https://timesofindia.indiatimes.com/business/india-business/70-of-hotels-restaurants-could-close-in-45-days-warns-fhrai-after-no-relief-from-govt/articleshow/75808858.cms

[2] B. D. Bidisha, P. Harrsha, Medha, A. Mohinishree and C. Anitha, “Restaurant Market Analysis”, vol. 5, 2018.

[3] W. Veda, R, Onkar and P. Aruna, “Restaurant Review Analysis and Classification Using SVM”, pp. 49-52, 2019.

[4] Y. Luo and X. Xu, “Predicting the Helpfulness of Online Restaurant Reviews Using Different Machine Learning Algorithms: A Case Study of Yelp”, 25 September 2019

[5] K. Sornalakshmi, “Sentiment Analysis of Food Reviews Using User Rating Score”

[6] [Online] Available: https://help.imdb.com/article/imdb/track-movies-tv/weighted-average-ratings/GWT2DSBYVT2F25SK?ref_=helpsect_pro_2_8#