RESTAURANT RECOMMENDER SYSTEM - Data Analytics Phase 1 Report

BASAVARAJ HARVI BHAVIKHA CHOPRA GITIKA VINAYKIYA NUMAN MOHAMMED ABBAS SRN: PES1UG19CS108 SRN: PES1UG19CS111 SRN: PES1UG19CS169 SRN: PES1UG19CS308

CSE Department CSE Department CSE Department CSE Department

PES UniversityPES UniversityPES UniversityPES University *Bengaluru, Karnataka*  *Bengaluru, Karnataka Bengaluru, Karnataka Bengaluru, Karnataka*

[basavaraj456vc2@gmail.com](mailto:basavaraj456vc2@gmail.com) [bhavikachopra1821@gmail.com](mailto:bhavikachopra1821@gmail.com) [gitikavinaykiya@gmail.com](mailto:gitikavinaykiya@gmail.com) [mohd123456789numan@gmail.com](mailto:mohd123456789numan@gmail.com)

***ABSTRACT:***

**Today, exploiting sentiment analysis has become popular in designing recommender systems in various fields, including the restaurant and food area. However, most of the sentiment analysis-based restaurant recommender systems only use static information such as food quality, price, and service quality. The analysis of users’ opinions and the extraction of their food preferences lead to the provision of personalized recommendations. In this paper, a context-aware recommender system is proposed that extracts the food preferences of individuals from their comments and suggests restaurants in accordance with these preferences.**

***Keywords—* Zomato, recommender systems, Data Cleaning, visualization, EDA, NLP, Sentiment Analysis, Clustering, tf-idf, Cosine similarity.**

# INTRODUCTION:

Today we live in a world where a new restaurant/cafe pops up every day. This gives the public a lot of options to choose from. There’s something for every taste bud. But one downfall to this is that having too many options can confuse the masses and they may end up going to the same place every time. There’s always understandable hesitation to try these places without dependable reviews.

Though the web world is a source of abundant valuable and useful information, it does introduce many concerns and can turn the decision-making process more tedious and complex. Hence, it’s mandatory that the information gets filtered and personalized concerning a particular user. Some of these restaurants also have multiple outlets in the same city and the food quality may differ, depending on the outlet. This is where the Restaurant Recommendation System comes into play. Our system helps the user select the restaurant according to his taste buds, location and approximate cost. Hence, it’s mandatory that the information gets filtered and personalized concerning a particular user.

The main objective of the work proposed in this paper is to enhance the user experience by analysing the reviews of restaurants and categorize them in some aspects so that a user can easily know about the restaurant. As a result, the accuracy in recommendation is low in these systems. In this paper, we augment the extraction of user’s preferences by sentiment analysis. The proposed recommender system is designed to first extract the users’ preferences from their textual comments and past history and then it suggests restaurants tailored to these preferences.

## LITERATURE REVIEW:

**1)** **Restaurant Review Classification and Analysis:**

The aim of this research paper was to analyse the restaurant reviews and present useful information that the ratings do not consider or overlook. Machine learning algorithms like Naïve Bayes and Logistic regression are used for first classifying the reviews in proper aspects then performing sentiment analysis on them. Summarization is done using genism and results are displayed using effective visualization techniques. The main objective of the work proposed in this paper is to enhance the user experience by analysing the reviews of restaurants and categorizing them in some aspects so that a user can easily know about the restaurant.

It is observed that Semantic orientation can also be used as a sentiment analysis model to classify reviews as 0 or 1 representing negative and positive respectively. The Naive Bayes model generally performs better than SVM. It deals with the aspect classification and pre-processing part for sentiment analysis. Feature generation is performed using TF-IDF. Finally, the dataset is prepared into vectors using the BOW vectorizer for sentiment prediction. The comparison shows that the Multinomial NB model performs with an accuracy of 86.6%, SGD Classifier performs with an accuracy of 85.8% and Random Forest performs with 82.5%.

**2) Restaurant recommender system based on sentiment analysis**:

The proposed recommender system is designed to first extract the users’ preferences from their textual comments and past history and then it suggests restaurants tailored to these preferences. For this purpose, after doing the text mining and extracting the nouns in the user’s comments, unrelated nouns to the food domain are identified and filtered using the WordNet ontology. These nouns are then clustered together based on their semantic similarity. Afterwards, the sentences containing them are transmitted to the cluster and scored using sentiment analysis. Based on the scores earned, the cluster with the highest score reflects user preferences. Then, by converting the user’s preferences as well as the restaurant’s menu into vectors and calculating the similarity between them, the suitability of the restaurant for the user could be calculated. Finally, the restaurants that their menu is the most similar to user’s preferences are proposed to them. Here it extracts user preferences by analysing their opinions and refines the obtained list by sentiment analysis. The accuracy of extracting user preferences by analysing each user’s comments is far higher than methods such as TF-IDF. This type of recommendation is exploited in item-to-item collaborative filtering systems. Therefore, the proposed system is a hybrid filtering system. The proposed system also offers recommendations based on the current location of the user. For this purpose, the Google maps service is used to identify the current city of the user. The results have shown that making use of the Wu–Palmer criterion for clustering and the cosine criterion for calculating similarities between preferences and menus yields the best results regarding precision, recall and f-measure. The results also revealed that the proposed system can provide users with

92.8% precision.

**3) Aspect-Based Opinion Mining and Recommendation System for Restaurant Reviews:**

In this work, they propose a system that automatically generates personalized review recommendations using two different approaches. Firstly, drawing inspiration from Traditional collaborative filtering systems, the system generates user rating profiles and tailors the list of reviews to the preferences of each user. Secondly, we employ aspect-based opinion mining to identify the important features highlighted in each review. As opposed to traditional sentiment analysis, which provides an overall picture of whether a review is positive or negative, aspect-based opinion mining provides a fine-grain analysis of both sentiment and strength of the review. Given a predetermined set of features for a particular domain, results from aspect-based opinion mining can be used to rank/sort the reviews based on the aspect the users are most interested in.

Aspect Summarization:  This module aims at extracting the important features from each review, along with their polarity weight. To perform this, we employ the subjectivity lexicon] in order to map weak and strong positive and negative to numeric values (ranging from -4 to +4). which increase the intensity of the sentiment Each review is POS tagged. These three inputs are subsequently fed to our algorithm that generates as output an opinion score for each feature identified in a review. In this work we assume that the product aspects are predetermined.

Online Recommendations:  When a given user searches for a specific restaurant, the recommendation engine computes the similarity of the current user with all the reviewers of the particular business and ranks and presents the related reviews in descending order of similarity. As a result, each user will be presented with a different set of reviews for the same business.

**4) Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities:**

This paper proposes machine learning algorithms to resolve the issue of personalized Restaurant selection relying upon tripadvisor.com search data. All the information related to every hotel (viz, general details of hotel, comments, reviews, ratings) is stockpiled within a database. The facilities provided by the hotel are fetched from hotel database along with user’s comments that are being analysed and tagged using Natural Language Processing (NLP). The user reviews are parsed and necessary details (like features and views) are obtained. The summarized rating of a restaurant can be computed in terms of the collective reviews (be it neutral, positive or negative). The lexicon approach is applied to obtain divergence of sentiments whether Positive or Negative. Consequently, the features are merged according to user’s perspective and a score is generated for every sentence. Then the score of all the sentences are merged to generate a summarized score for a single review and thereafter using the database to store this sentimental outcome. Using the restaurant recommendation system, the basic hotel amenities are selected by the user and based on this parameter the matching hotels are then populated. The system then uses reviews and comments to analyse the hotel’s positive and negative aspects. Eventually, the higher-ranking hotel is then suggested to the user.

Various ML algorithms along with NLP are being proposed to resolve the issue of personalized restaurant selection.

We can observe that The NLP yield in greater accuracy compared to existing approaches. The Results depicts 92% accuracy with an error rate of 8%.

## DATA CLEANING:

## ​

Fig1) total missing values present in the dataset

In the cleaning of our dataset, the following issues were identified:

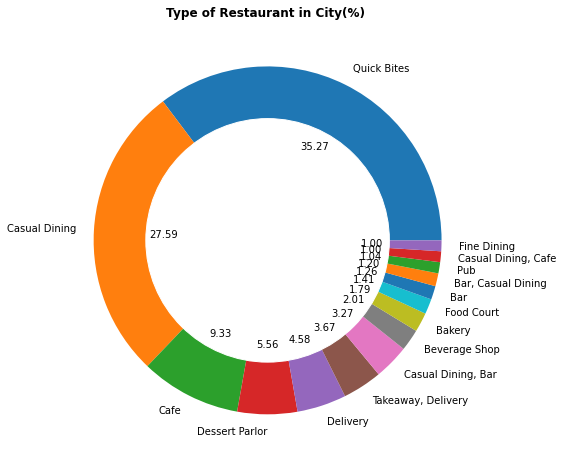
**Inconsistency in data formats**: Numerical data with commas in it (e.g.: the cost column was represented as “2,500” instead of 2500.00) which cannot be used for analysis, as well as numerical columns being represented as strings in the data (e.g.: the ratings column)

**Duplicate rows**: No duplicate rows found in the dataset.

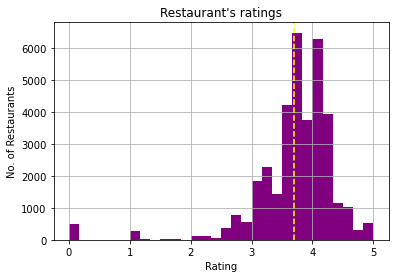
**Missing values**:

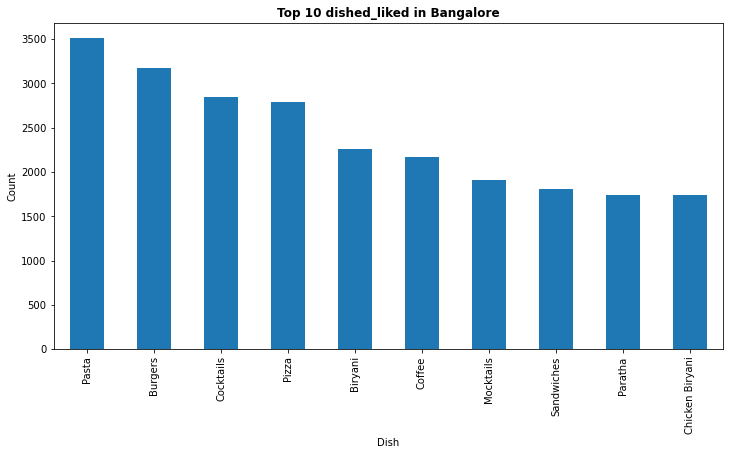
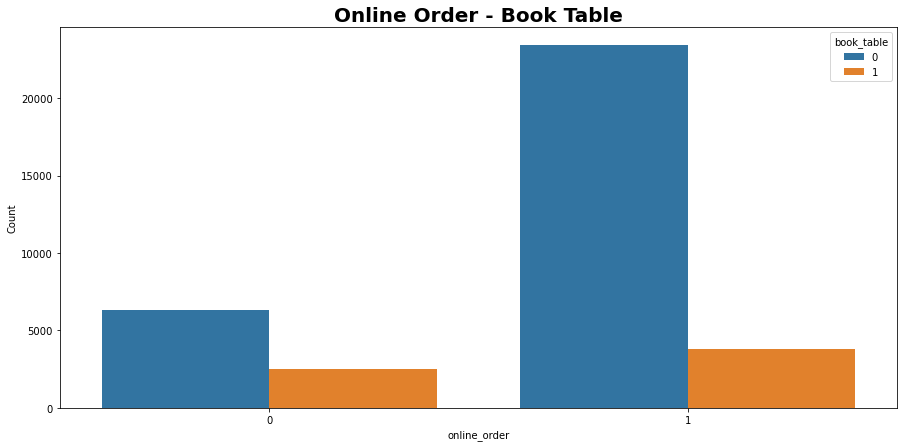
* **Dish liked**: The missing values present in this column were around 28,078. So we filled the missing value with menu item of that particular restaurant. If the menu item was also missing then it was filled with dish liked of other franchise of the same restaurant.
* **Ratings**: The user-given ratings in the data have missing values for certain number of rows in the data (7,775 missing values out of approximately 55,000 rows). In order to handle it, we extracted mean ratings from reviews column.
* **Approximate cost:** The missing values present in this column were around 346. As approx. cost had outliers, we replaced missing values with median.
* **Rest type**: The missing values present in this column were around 160.If the rest type is null, then its corresponding listed in type is taken then we group the dataset w.r.t listed in type, and then we take the mode of the rest type for the group-by dataset.

**Encoding of categorical and discrete variables**: The categorical variables (more specifically, binary variables) in the dataset that correspond to whether that branch offers online ordering and table bookings, are converted from string values (‘Yes’ and ‘No’) to integer values (0, 1).

The cuisines, dish liked that are listed as a comma-separated string of cuisines for each restaurant have been split up and converted to a form that can be easily evaluated as separated ****entities. (e.g.: Chinese, North Indian, South Indian” is represented as [’Chinese’, ’North Indian’, ’South Indian’]”).

**Reviews**: Since some of the reviews were not in English language, we encoded it using UTF-8 and then decode it in ASCII.

**DATA VISUALISATION**:

****Fig2.1). Observation: From the above graph we observe that mean rating is 3.606****

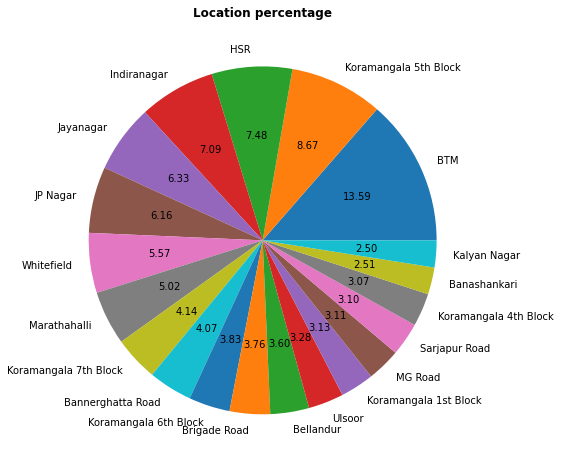
## Fig2.2) Observation:We can observe that top dishes prefered by Bangaloreans are Pastas,Burgers,etc.

.

Fig2.3) Observation: From the above graph, we can observe very few restaurants provide both Online order and Book table together(4k). So, if planning to open a new restaurant consider providing both the facilities

Fig2.4) Observation: No doubt about this as Bangalore is known as the tech capital of India, people having busy and modern life will prefer Quick Bites.



Fig2.5) Most loved cuisines by Bangaloreans are North and South Indian, Fast Foods etc.

## Observation: From the graph, We can conclude that Maximum no of restaurants are in BTM followed by Koramangala 5th block, HSR, INDIRA Nagar, so on.

**WORK TO BE DONE IN WEEKS AHEAD**:

* TF-IDF Based Approach and Cosine Similarity:

This content-based collaborative filtering approach is used to generate recommendations for the k most similar restaurants to a given restaurant. This approach treats each review for each restaurant as a single document.

* The cosine similarity is used as the similarity metric between the review vectors for each restaurant. Given the count of the number of recommendations k, the system outputs the k most similar restaurant recommendations based on the restaurant name that is entered by the user.
* Clustering will be done on dish liked and cuisines.
* Recommendation: Based on User’s past history we apply sentiment analysis on the reviews given by the user and select the dishes he liked and recommend those restaurants which provide similar dishes considering user’s location as well as their budget.

**REFERENCES**:

* <https://www.researchgate.net/publication/346937519_Restaurant_Review_Classification_and_Analysis>
* <https://reader.elsevier.com/reader/sd/pii/S2666827021000578?token=D4D5B884AAD4D02222E1D144A8E72070990F064AA0895FBFBAD98E11239D450DC666C30BE31FA45783B9B120D2BA2F9F&originRegion=eu-west-1&originCreation=20211015070510>
* <https://ieeexplore.ieee.org/document/8862048>
* <https://dl.acm.org/doi/abs/10.1145/2645710.2645716>
* <https://medium.com/@sumith.gannarapu/restaurant-recommendation-system-b52911d1ed0b>
* <https://medium.com/mlearning-ai/restaurant-recommendation-system-based-on-the-content-in-reviews-dfc3351004db>