Restaurant Recommender using Machine Learning

Sailaja Thota [1]
Professor, School of CSE
REVA University
Bengaluru, India
sailaja.thota@reva.edu.in

Shashank K S [2]
Student, School of CSE
REVA University
Bengaluru, India
shashankreddy0709@gmail.com

Sahil Ahmed Shaik [3]
Student, School of CSE
REVA University
Bengaluru, India
sahilahmedshaik63@gmail.com

Siddesh B R [4]
Student, School of CSE
REVA University
Bengaluru, India
siddesh2000br@gmail.com

Shashidhar Reddy [5]
Student, School of CSE
REVA University
Bengaluru, India
reddyshashidhar74@gmail.com

***Abstract*—Recommender systems widely use Sentimental Analysis in a variety of industries, including the cuisine industry. Food Quality, pricing, and service quality are the static parameters on which most recommendation systems solely depend. Extraction of food preferences from user remarks and Analysis of user needs leads to individualized guidance, filling a study gap in the literature. This paper suggests a context-aware recommender system that extracts users' food choices from their reviews and recommends restaurants based on their predilection. To group the names of food from user reviews and to show the restaurants to match their sentiments towards food. Finally, nearby top 5 open restaurants are given as output based on users' needs. For evaluation, we have used the Zomato data set available on Kaggle and applied various algorithms to build our model that provides content-based recommendations. As per the results, the suggested system can give suggestions like Restaurant, Rating, and Location to consumers with a high level of precision.**

Keywords—Sentiment Analysis, Restaurant Recommender, Food Preferences, Content-based suggestions, Ratings

I. INTRODUCTION

In this model, data is gathered on an end user's interests across a variety of categories, in order to help them make better selections. It is difficult for visitors and tourists to find a restaurant in unknown locations where there is a wide range of restaurants. Website and social media platform comments are increasingly regarded as a big source of data. The processing of these replies, as well as the analysis of their emotions, may indicate the user's gastronomic preferences in this scenario. Since human language is complex, people may use a variety of terminology to describe a concept, making it difficult to extract user preferences. As a result, sentiment analysis should use a semantic approach. People are sometimes perplexed as to which restaurant is best suited to their preferences as the number of eateries grows. People also have a hard time deciding where to eat and what cuisine to order, especially if they are new to the region. A web-based restaurant suggestion system called "Restaurant Recommendation System based on Machine Learning." The model's main goal is to recommend the best cuisine to eat at a specific place based on the users' culinary preferences. The system is aimed at users who want to dine at a restaurant. The program considers the users' meal preferences and ratings while making food recommendations. The program uses techniques like Random Forest, Random Forest with PCA, Density-based clustering, and K-Nearest Neighbour algorithms to give accurate and trustworthy content-based recommendations to users. The software gathers user feedback on a variety of culinary items and keeps it in a database. The program then makes recommendations to users based on their ratings.

The data collected from the Kaggle website is taken as a source of data for our dataset for the evaluation of the suggested recommender system. It consists of real time data about the restaurant's details located in different locations in Bangalore. The web application uses real time data collected in dataset for prediction of restaurants that actually exists in Bangalore as per Zomato records. The end

results are highly accurate that suggest to a user their preferred restaurants. The contributions are as follows:

- The sentimental analysis is required to suggest a model for analyzing end users' opinions to get user preferences. The accuracy of examining each user's comments to extract user preferences is significantly greater than other approaches.

- To boost the accuracy of collecting users' predilection, a technique that follows the semantic approach is used to group comments by their names.

- In order to propose a new system, there is a need for a relevant and accurate dataset. The dataset is obtained from the Kaggle website.

II. Related Work

In an existing system, a restaurant recommender system that is limited to the mobile environment uses only user location and restaurant previously visited history of the user to recommend the restaurants [1].

In a Hotel recommendation-based system that uses locations around the user to recommend restaurants, POIs databases are taken into consideration while predicting, similarities are computed between users' preferences and hotels in surrounding locations. recommends top-k hotels to the user [2].

The existing suggested method uses Location, Time, and Preference for a recommendation. where the recommendation score is computed considering various parameters such as end-users visiting trends, their taste of food, type of food, type of restaurant, restaurant operation timings, and distance from users' location. [3].

The Location-Based Mobile Environments recommender system uses foursquare data for recommendation where the location of the user was extracted from four data points located around the user. [4].

Another location-based recommendation system that learns from Facebook comments and check-ins to track users' location and suggest users' how far they can travel by generating social-based trends to recommend the restaurants. [5].

Baidu map cloud service technique is implemented along with the recommender system to recommend restaurants to the new users in order to find the required restaurants of their choice which include features like location and cuisines. This system has less user involvement that would notify the user based on their movements and recommend restaurants [13].

III. Implementation

The front-end page is a user-friendly GUI that consists of login and register options to find preferred restaurants. The end-user will be recommended with the top 5 restaurants by considering the similarity in attributes and users' preferences. The system is more effective as recommendations are done using both collaborative and content-based. The closest restaurants could be suggested to the Consumer. The recommendation module turns into active, retrieving close by and eating places and rating them, primarily based totally on their properties, in line with the rankings generated.

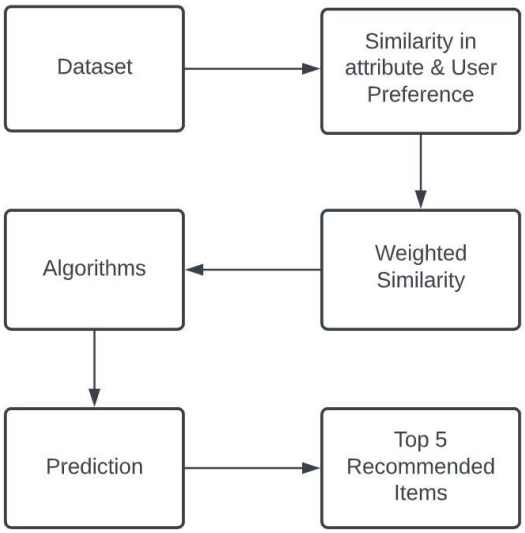

Figure 1: Methodology

Data is reviewed to see how much information we can gain from restaurant reviews. Various algorithms such as Random Forrest, K-means clustering, and hierarchical clustering were applied. And found that KNN method provided better results for building a model of user experience.

A. *Data set*

To suggest a recommendation model for users, a real time data which actually exists in a locality is needed for a web application. The Working dataset, collected in real time, contains user restaurant review ratings, restaurant ids, customers, location information, etc. The text from the restaurant review was analyzed by extracting usable features for developing a classification model which is obtained from Kaggle.

B. *Tools*

For data pre-processing, Jupyter Notebook was utilized, which is commonly used for text data processing. For front-end development, we used HTML and CSS.

C. *Methodology*

First, the data was processed to link users to reviews and reviews to businesses. The non-restaurant businesses were dropped from the dataset. The reviews were observed to see the maximum information that can be gained from a particular user.

- 51% of restaurants have ratings of more than 4.
- 44% of restaurants have 3-4 ratings
- Restaurants have 2-3 ratings less than 5%.

First, the data was analyzed to connect individuals to reviews. As a second step, the evaluations were examined to extract the required information for a specific user. Then, as a trial-and-error procedure, we employed several methods such as Random Forrest, K-means clustering, hierarchical clustering, and Random Forrest with PCA to observe the model's behavior. As our final algorithm, we examined the KNN technique. The application can be accessed from anywhere with an internet connection and is feasible to optimize when needed.

D. *Feature Selection*

Dimensionality reduction has been done with deep knowledge and an in-depth study of the proposed system.

Columns that are dropped are as follows:

- Name
- Place
- Rate
- Favorite Dish
- Reviews
- Type
- City

The columns appear to be significant, yet all of the same information may be found in less sophisticated columns.

E. Algorithms Used

Random Forest: As Random Forest is the simplest algorithm to implement, so Random Forest was the approach used as part of trial and error in model deployment, the model was trained and tested but the outcome expected was overfitting. Furthermore, the expected results were not matched hence proceeded with random future reduction using PCA [6].

Random Forest with PCA: The Principal Component Analysis approach, which detects the hyper-plane nearest to the data and projects it onto it, is the most used dimensionality reduction algorithm. The limitation here is that to select the best hyperplane with the least amount of information loss, we must maintain the greatest degree of variation. [6].

Hierarchical Clustering: This approach has a useful way of clustering using segmented observations that outputs the optimal number of groups falling under particular categories. The optimal groups of restaurants were determined, with no need for a pre-defining number of clusters that stands as a bonus over K-Means [7].

K-Means Clustering: To minimize the number of clusters and clumsiness in the model. K-means clustering is used where each cluster is assigned a data point (k) and the centroids for clusters are located. The cost of restaurant items is computed using the model K-means algorithm. [10].

Density-Based Clustering: DBSCAN can discover the cluster with outliers as well as arbitrary-shaped clusters. The concept of DBSCAN is that a point belongs to the same cluster where other points are nearer to that same cluster. It is also slower in comparison owing to the neighborhood query for each item, and it has problems appropriately establishing the density threshold. [12].

K-Nearest Neighbour Algorithm: The Algorithm is a classification system that learns based on how similar a data point is to other data points. Because of its simplicity and precision, it has become one of the most well-known classification algorithms in the business. In K-NN, K is the number of nearest neighbors. The most crucial deciding element is the number of neighbors. The KNN algorithm predicts the values of new data points based on 'feature similarity.' This means that a value is assigned to the new point depending on how closely it resembles the points in the training set.

The KNN algorithm predicts the restaurants based on the cosine similarity. That would recommend the top 5 restaurants based on users' food predilections. [10].

IV. RESULTS

A. Exploratory Data Analysis

Deeper data visualization is performed on the dataset and obtained the following visualizations

- Number of restaurants providing Online orders

- Most famous restaurants chain
- Type of Service provided by the restaurant
- Favorite food counts
- Cost of booking tables in restaurants

1) Number of restaurants providing Online orders

Most of the restaurants provide an online order facility as people in Bangalore often order food online. The restaurants that do not provide online orders are rated very low compared to restaurants that provide online orders which are highly rated.

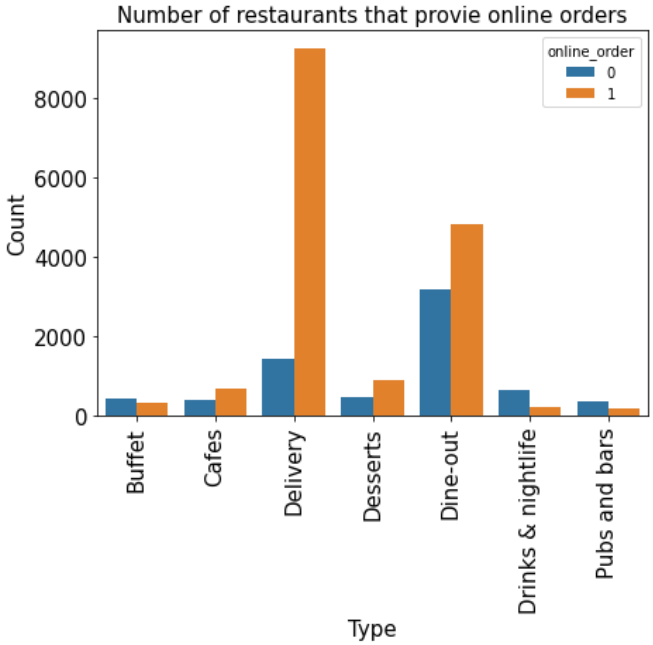

Figure 2: Number of restaurants providing online orders

2) Most famous restaurants

The list of most famous restaurants visited by customers is shown in Figure 3.

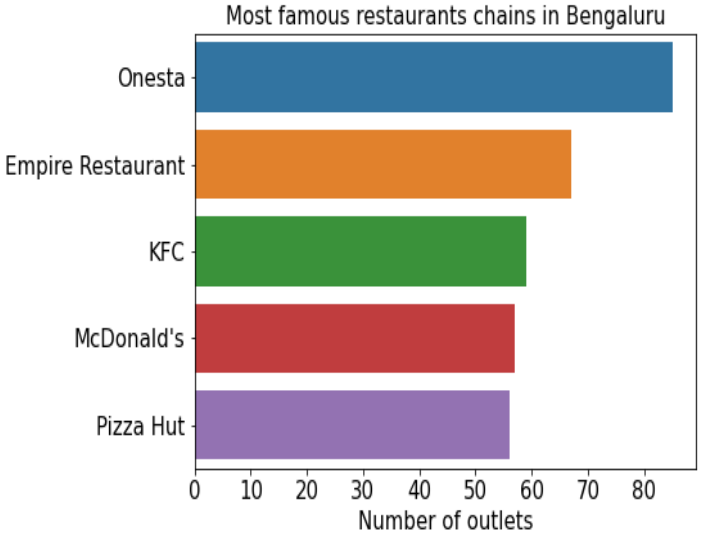

Figure 3: Most famous restaurants

3) Type of Service provided by the restaurant

We can observe that more than 10,000 restaurants provide delivery and up to 8,000 restaurants provide Dine-out services across Bangalore that as listed in Zomato dataset. Most restaurants focus on providing quality over quantity types of services.

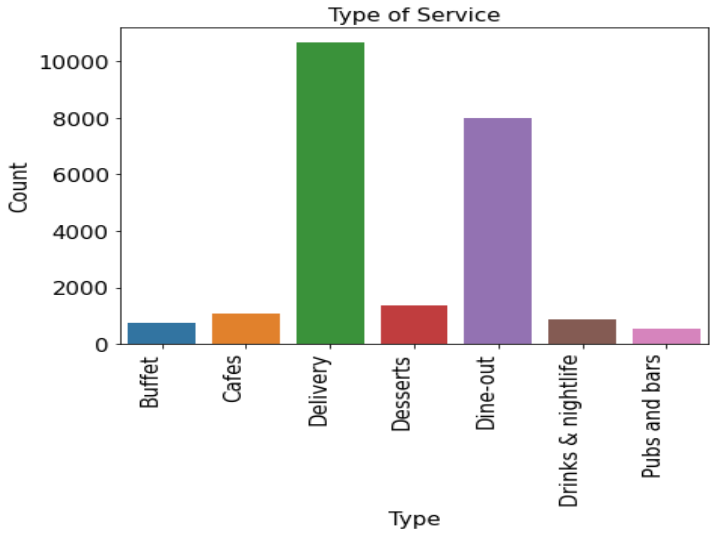

Figure 4: Type of Service provided by the restaurant

4) Favorite food counts

The majority of individuals wish to eat pasta, pizza, and cocktails in Bangalore. Most restaurants serve food-supported customer reviews so as to make customers to revisit their restaurants and create profits.

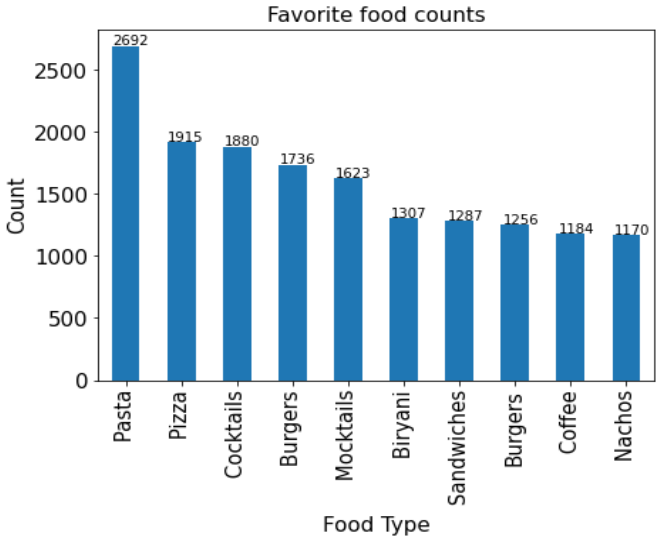

Figure 5: Favorite food counts

5) Cost of booking tables in restaurants

The average cost of the food is greater than ₹1000 for restaurants where the bookings were done online. Whereas the average cost of the food is lesser than ₹1000 for restaurants where the bookings were not done.

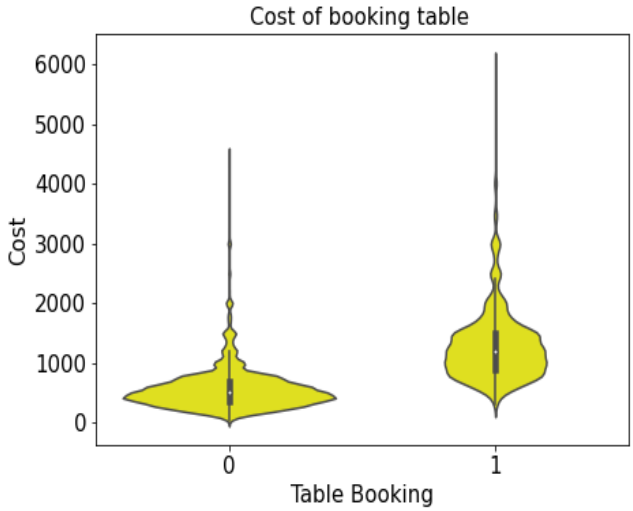

Figure 6: Cost of booking tables in restaurants

B. *UI-UX design*

Figure 7: Front Page

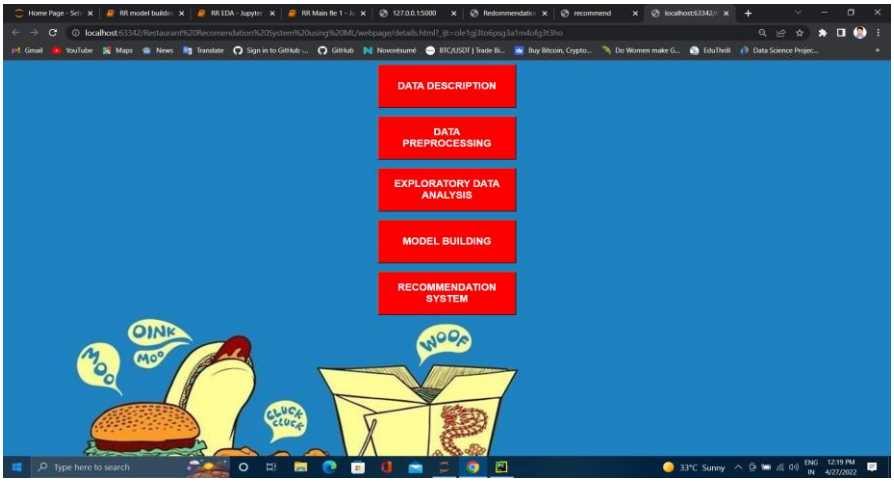

Figure 8: Model linkage page

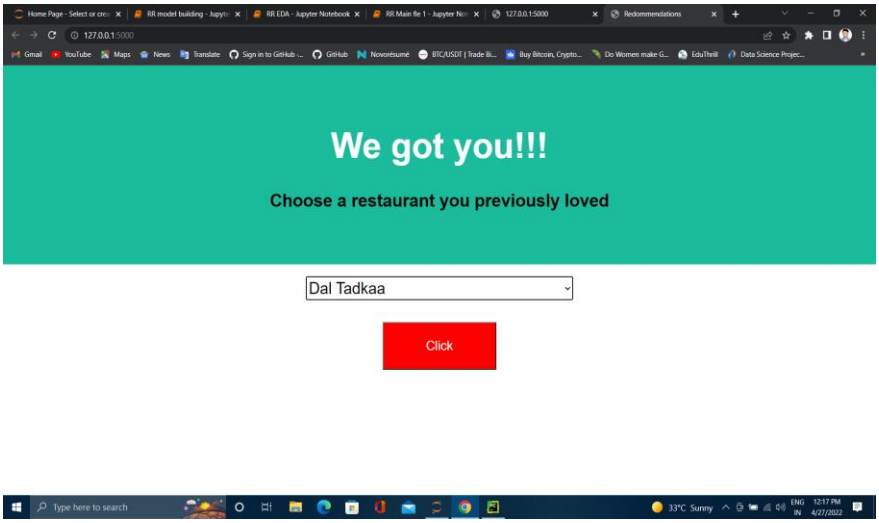

Figure 9: User selection page

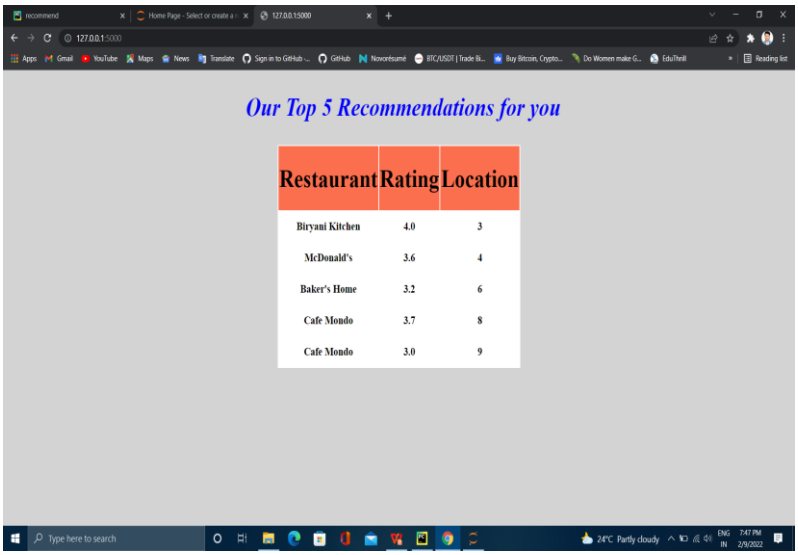

Figure 10: Recommendation page

C. *Inference Table*

Algorithm	**Insights**
Random Forest	The random forest model is overfitting
Random Forest with PCA	We got a good model after feature reduction with PCA.
K-Means clustering	Most restaurants are grouped under clusters, with average cost and average votes determined.
Hierarchical clustering	Determined and linked an optimal number of clusters.
Density-based Clustering	Outliers identified are determined.

K nearest neighbors	Top 5 Recommendations for users based on user preference using cosine similarity.

V. Conclusion

The model developed in this paper analyses the vast amount of data and provides the user with the top 5 recommendations based on the user's choice, considering restaurant type, ratings, and location. This model could be potentially very useful for a food lover to visit a good restaurant based on his\her choice. Hence Machine Learning techniques are used to implement the model. Thus, the complexity of finding good restaurants was significantly reduced as this model will recommend the top 5 restaurants. The current model was tested on Zomato data available on Kaggle and was found 96.7% accurate.

VI. References

[1] Anant Gupta, Kuldeep Singh, "Location based personalized restaurant recommendation system for mobile environments", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.

[2] Zhichao Chang, Mohammad Shamsul Arefin, Yasuhiko Morimoto "Hotel Recommendation Based on Surrounding Environments", Second IIAI International Conference on Advanced Applied Informatics(IIAIAAI), 2013.

[3] Md. Ahsan Habib, Md. Abdur Rakib; Mohammad Abul Hasan, "Location, time, and preference aware restaurant recommendation method",19th International Conference on Computer and Information Technology (ICCIT), 2016.

[4] Jun Zeng; Feng Li, Haiyang Liu, Junhao Wen, Sachio Hirokawa, "A Restaurant Recommender System Based on User Preference and Location in Mobile Environment", 5th IIAI International conference on Advanced Applied Informatics (IIAI-AAI),2016.

[5] Anu Taneja, Prashant Gupta, Aayush Garg, Akhil Bansa, Kawal Preet Grewal, Anuja Arora, "Social graph-based location recommendation using users' behavior: By locating the best route and dining in best restaurant "Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016.

[6] Ajesh A, Jayashree Nair, Jijin PS, "A Random Forest Approach for Rating–based Recommender System" International Conference on Advances in Computing ,Communications and Informatics(ICACCI) ,2016.

[7] Jevin D. West, Ian Wesley-Smith and Carl T. Bergstrom, "A Recommendation system based on Hierarchical clustering of an article-level Citation Network", IEEE Transactions on Big Data, April-June, 2016.

[8] F.M Takbir Hossain,Md.Ismail Hossain,Ms.Samia Nawshin,"Machine Learning Based Class Level Prediction of Restaurant Reviews",IEEE Region 10 Humanitarian Techology Conference(R10-HTC) , 21-23 Dec 2017.

[9] G Geetha, M safa, C Fancy, DSaranya,"A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System",National Conference on Mathematical Techniques and its Applications(NCMTA 18), January,2018.

[10] Donghi Wang,Yanchung Liang,Dong Xu,Xiaoyue Feng, Renchu Guan,” A Content-based recommender system for computer science publications”,Elsevier,Volume157,October,2018.

[11] R.M Gomathi, P. Ajitha, G.Hari Satya Krishna, I. Harsha Pranay,”Restaurant Recommendation system for user preferences and services based on Rating and Amenities”,International conference on Computational Intelligence in Data Science(ICCIDS),21-23 Februrary,2019.

[12] Atharva Kulkarni, Divya Bhandari,Schin Bhoite,” Restaurants Rating Using Machine Learning Algorithms”,International Journal of Computer Applications Technology and Research ,Volume 8-Issue 09,375-378,2019.

[13] Kunal Bikram Dutta,Aman Sahu,Bharat Sharma,Siddharth S.Rautarey,Manjysha Pandey,”Machine Learning-Based Prototypr for Restaurant Rating Prediction and cusine selection”,Advanced in Intelligent Systems and Computing(AISC),Volume 1166 ,2020.

Achmad Arif Munaji , Andi Wahju Rahardjo Emanuel,”Restaurant Recommendation System Based on User Ratings with Collaborative Filtering”,IOP Conference series Materials Science and Engineering,February,2021.