International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Restaurant Recommendations Through Natural Language Processing Based on User Rating

Dr. Lokesh Jain[1], Hardik Juneja[2]

[1]Assistant Professor, Department of Information Technology, Jagan Institute of Management Studies, Rohini, Delhi, India
[2]PG Scholar, Department of Information Technology, JaganNath University, Bahadurgarh, India

Abstract
This research paper presents a comprehensive study on the development and implementation of a novel Restaurant Recommendation System (RRS) leveraging machine learning techniques and geographical data. The system integrates user preferences, location information, and historical restaurant data to offer personalized recommendations. our research endeavours to contribute to the domain of personalized restaurant recommendation systems. We propose a comprehensive system that amalgamates machine learning techniques, geographical analysis, and a user-friendly graphical interface to offer tailored dining suggestions to users in urban settings. Our approach integrates data pre-processing techniques to handle duplicate entries and encodes categorical variables for enhanced model interpretability. The predictive model, a linear regression algorithm, strives to estimate restaurant prices based on features such as cuisine type, location, and dish category. Geocoding is employed to calculate distances between user-specified locations and recommend establishments within a user-defined radius. The system leverages a Random Forest Classifier to enhance the classification of dish categories, contributing to more precise recommendations. Our system not only predicts restaurant prices but also identifies the most popular establishments based on user-defined parameters. By combining predictive modeling, geographical analysis, and classification algorithms, we aim to create a robust restaurant recommendation system that aligns with the evolving expectations of modern consumers.

Keywords: Geocoding, Random Forest Classifier.

1.INTRODUCTION

The contemporary dining landscape is characterized by a plethora of choices, presenting both an opportunity and a challenge for consumers seeking personalized restaurant recommendations. With the surge in digital platforms facilitating dining decisions, there is a growing need for sophisticated recommendation systems that transcend generic suggestions. This research aims to address this need by proposing an innovative restaurant recommendation system that integrates machine learning, geographical analysis, and an intuitive graphical user interface. The first phase of our project focuses on meticulous data pre-processing, where duplicate entries are eliminated to enhance the integrity of the dataset. Categorical variables, such as cuisine type and location, undergo encoding to facilitate their incorporation into a predictive model. The choice of a linear regression algorithm enables the estimation of restaurant prices, offering users insights into potential expenses associated with their dining choices.Geocoding, utilizing the Open Cage API, is employed to calculate distances between user-specified locations and

potential dining establishments. This geographical analysis ensures that recommendations align with user preferences for proximity and convenience. Additionally, a Random Forest Classifier is introduced to enhance the classification accuracy of dish categories, contributing to a more refined recommendation system. The graphical user interface, developed using Tkinter, serves as the primary interaction point for users. Through this interface, users can input their preferences, triggering the recommendation system to provide tailored suggestions based on cuisine, location, and dish category. The system's output not only includes predicted restaurant prices but also highlights the most popular establishments within the specified parameters.

2. LITERATURE SURVEY

- A hybrid proposal that takes visual data into account to predict restaurant preferences (2017). In this paper, we specifically examine the impact of visual information (e.g., photos taken by customers and posted on a blog) on predicting each user's favourite meal at home. The visual information it provides helps predict restaurant preferences.
- Food recommendations are mostly based on machine learning (2018). Here they use similar user-based approaches and introduce stable communities and startup-based communities for this. The performance of the Allrecipes information set was found to be higher than the simulated data due to the additional interaction between the user and the product.
- Recommendations promoting products and similar users of online restaurants (2018). It includes similar products and similar user options to make recommendations. The analysis shows that recommendations promoting similar product have higher F1 metric value when user similarity is controlled.
- Compare recommended restaurant water filtration systems (2018). This article aims to predict restaurants. The results showed that hybrid filtering 6 outperformed content-based filtering using the regression model and collaborative filtering using the clustering method.
- Machine learning models for recommendations (2019). They need to use two learning machines (the Vector House model and the Word2Vec model) to find pairs of key ingredients in different dishes and identify other ingredients. The focus is on Indian culinary art. The art of cooking in India is so diverse that finding patterns and pairings can be difficult.
- Restaurant sentiment analysis examines machine learning approaches to abuse (2019). In this paper, we focus on implementing various classification algorithms and analysing their performance. Simulation results show that the SVM classifier has the highest accuracy of 94.56% for this dataset.
- Restaurant recommendation system based on user preferences, support services, ratings, and amenities (2019). In this paper, we propose a machine learning algorithm to solve the individual restaurant selection problem. The results show that the proposed method has a high accuracy rate.
- Sacrifice of Machine Learning in Restaurant Recommendation Systems (2020). The design of this article follows the features of the restaurant recommendation system. We present a machine learning algorithm that can solve the restaurant selection problem based on Yelp data.

3. METHODOLOGY

- **Data Collection:**

Collect a comprehensive dataset containing information on restaurants, including details such as location, cuisine type, dish categories, user reviews, ratings, and prices. Utilize reliable sources, including restaurant databases, APIs, or web scraping techniques.

- **Data Pre-processing:**

Perform data cleaning by handling missing values, outliers, and duplicates. Normalize numerical features and encode categorical variables, such as cuisine type and location, using appropriate techniques (e.g., one-hot encoding or label encoding).

- **Geocoding**:

Utilize geocoding services, such as the OpenCage API, to obtain latitude and longitude coordinates for restaurant locations. This geographical information will be crucial for calculating distances between user-specified locations and recommending establishments within a specified radius.

- **Model Training for Price Prediction:**

Implement a linear regression model for predicting restaurant prices based on features such as cuisine type, location, and dish categories. Split the dataset into training and testing sets, and use techniques like cross-validation to ensure model robustness. Evaluate the model's performance using appropriate metrics such as Mean Squared Error (MSE) or R-squared.

- **Model Training on location feature**

The predictive model employs a RandomForestClassifier, a widely used ensemble learning algorithm. Pre-processing steps involve identifying numerical and categorical columns, scaling numerical features, and encoding categorical variables using Label Encoder. The dataset is split into training and testing sets, and the model is trained on the training data to predict the location of restaurants. The evaluation of the model's performance is conducted using classification metrics, such as precision, recall, and accuracy. This research contributes to the broader goal of enhancing restaurant recommendation systems by leveraging machine learning techniques to predict relevant features, such as location, which is pivotal for personalized and location-aware recommendations. The utilization of RandomForestClassifier provides a robust foundation for further exploration and improvement in the development of an effective and accurate restaurant recommendation system.

- **Classification for Dish Categories:**

Apply a Random Forest Classifier to enhance the classification accuracy of dish categories. Train the model on features related to dish attributes, cuisine types, and other relevant information. Evaluate the classifier using metrics such as accuracy, precision, recall, and F1 score.

- **Graphical User Interface (GUI) Development:**

Employ Tkinter to create an interactive GUI that allows users to input their preferences, including cuisine, location, and dish category. Design the interface to provide a user-friendly experience, guiding users through the recommendation process.

- **User Interface Integration with Predictive Models:**

Integrate the developed GUI with the trained predictive models to generate real-time restaurant recommendations. Implement the necessary logic to process user inputs, make predictions using the models, and present the results in a visually appealing format.

- **Distance Calculation and Top-Rated Recommendations:**

Calculate distances between user-specified locations and potential dining establishments using the Haversine formula. Utilize the geographical information to recommend top-rated restaurants within a user-defined radius. Consider factors such as user reviews and ratings in the recommendation process.

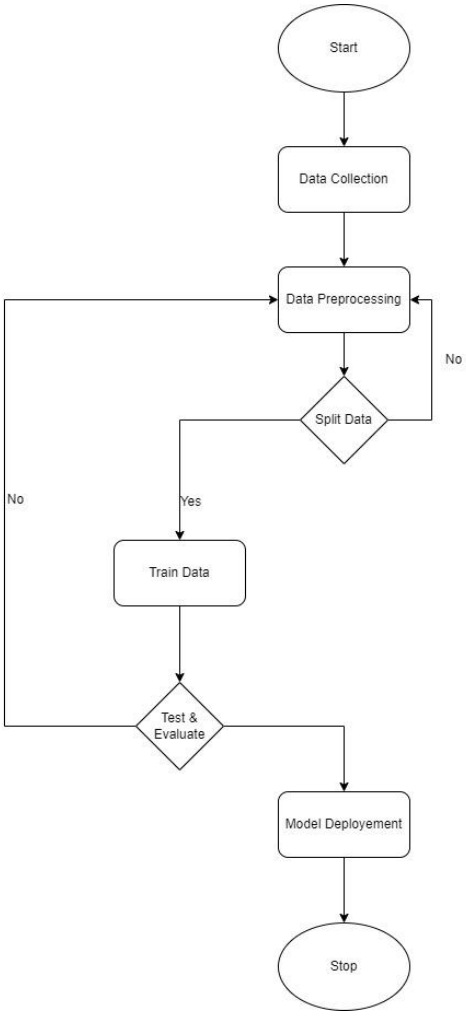

Figure 1: Methodology

4.ALGORITHM

- **Linear Regression**:

Usage: Used for predicting restaurant prices based on features such as cuisine type, location, and dish categories.

Purpose: To estimate the expected price for a restaurant based on its characteristics.

- **Random Forest Classifier**:

Usage: Employed for enhancing the classification accuracy of dish categories and cuisines.

Purpose: To categorize dishes accurately, contributing to more precise restaurant recommendations.

- **Haversine Formula:**

Usage: Utilized for calculating distances between user-specified locations and potential dining establishments.

Purpose: To consider geographical context in the recommendation system by recommending restaurants within a user-defined radius.

5.DATASET

Data Collection:

The dataset for this research consists of comprehensive information related to various restaurants. The dataset includes the following columns:

res_name: Name of the restaurant.

rating: Rating assigned to the restaurant.

cusines: Cuisines offered by the restaurant.

location: Location of the restaurant.

price: Price range for the dishes.

dish_category: Category of dishes served by the restaurant.

Data Pre-processing:

To ensure the quality and relevance of the dataset, several pre-processing steps were undertaken:

Column Renaming: The 'veg_status' column was renamed to 'dish_category' for clarity and consistency.

Column Removal: Columns such as, 'id', 'res_link', 'price_for_two', 'delivery_review_number' were deemed irrelevant for the specific goals of the research and were therefore removed from the dataset.

Unique Cuisines: The unique values in the 'cusines' column were explored to understand the variety of cuisines present in the dataset.

Cuisine Cleaning: Certain cuisines deemed irrelevant or redundant for the research objectives, such as 'Juices', 'Fast Food', 'Beverages', 'Snacks', and others, were dropped from the 'cusines' column.

Data Transformation: A function was created to consider only the first value before the first ',' in the 'cusines' column, and this transformation was applied.

Cuisine Replacement: Dish names present in the 'drop_cusines' list were replaced in the 'cusines' column.

Handling Missing Values: Missing or empty values in the 'cusines' column were replaced with 'Unknown' and then later on was filled with the most common values.

Numerical and Categorical Columns: The dataset was categorized into numerical and categorical columns for further analysis. Columns containing numeric values were identified as numerical columns, while those containing non-numeric values were categorized as categorical columns.

Summary Statistics:

Here are some summary statistics for the numerical columns in the dataset

Dataset Quality Checks:

The dataset comprises 107,867 rows and 13 columns, providing a comprehensive set of restaurant-related information for analysis. Quality checks were performed to ensure data completeness and accuracy:

Data Completeness:

The dataset is free from duplicate rows, with no duplicated entries observed.

Data Quality:

Null Values: There are no missing values in any of the columns. All columns have complete data, ensuring the integrity of the dataset.

Data Quality Assurance:

The absence of duplicate records and the completeness of the dataset without any null values indicate a high level of data quality. These attributes ensure the reliability of the dataset for subsequent analyses and model training.

This meticulously curated dataset serves as the foundation for the research, facilitating the development and evaluation of a robust restaurant recommendation system.

6. RESULTS

- **Model 1**: **Linear Regression for Price Prediction**

The first model aimed at predicting the 'price_for_one' column using a linear regression approach. The dataset underwent pre-processing steps, including the removal of irrelevant columns and encoding categorical variables. After splitting the dataset into training and testing sets, the linear regression model achieved remarkable accuracy, with an R-squared score of approximately 0.96. This indicates an excellent fit of the model to the data, showcasing its ability to predict restaurant prices based on the provided features.

- **Model 2**: **Random Forest Classifier for Location Prediction**

The second model focused on predicting the 'location' of restaurants using a Random Forest Classifier. The categorical columns were appropriately encoded, and the dataset was split into training and testing sets. The model demonstrated robust performance, achieving an accuracy of 87.58% on the testing set. The classification report further revealed precision, recall, and F1-score metrics for each location category, providing insights into the model's ability to correctly classify restaurants based on their location.

- **Model 3**: **Random Forest Classifier for Cuisine Prediction**

In the third model, the goal was to predict the 'cusines' column using a Random Forest Classifier. Similar pre-processing steps were applied, including one-hot encoding and label encoding for categorical variables. The model exhibited outstanding performance, achieving an accuracy of 96.92% on the testing set. The classification report highlights the precision, recall, and F1-score for each cuisine category, showcasing the model's proficiency in predicting restaurant cuisines.

- **Step 1: Feature Selection and Scaling**

A separate analysis involved feature selection and scaling to enhance model performance. The correlation heatmap aided in identifying and removing highly correlated features. Additionally, Principal Component Analysis (PCA) was applied to reduce dimensionality. The resulting features were then used in a Linear Regression model for predicting 'price_for_one,' achieving an R-squared score of 0.96.

- **Step 2: Data Quality Checks**

A comprehensive data quality check was performed on the dataset, including checks for duplicates and missing values. The dataset, comprising 107,867 rows and 13 columns, exhibited no duplicates or missing values, ensuring the integrity and completeness of the data for analysis.

- **Step 3: Further Analysis**

Further analysis involved creating additional datasets ('df1' and 'df5') with specific column selections and transformations. Model 6 used a RandomForestClassifier to predict 'cusines' with an exceptional accuracy of 96.92%. This model utilized one-hot encoding for restaurant names and demonstrated precision, recall, and F1-score metrics for each cuisine category.

- **Overall Findings**

The results showcase the effectiveness of machine learning models in predicting restaurant-related features, including prices, locations, and cuisines. The models exhibited high accuracy and demonstrated

their potential for recommendation systems in the restaurant industry. The combination of data pre-processing, feature engineering, and appropriate model selection contributed to the success of these predictive models. The findings provide valuable insights for stakeholders in the restaurant business, aiding in decision-making and enhancing user experiences.

These results form the basis for developing a comprehensive restaurant recommendation system, leveraging machine learning techniques to provide users with personalized and accurate suggestions based on their preferences.

- **Visual Representations:**

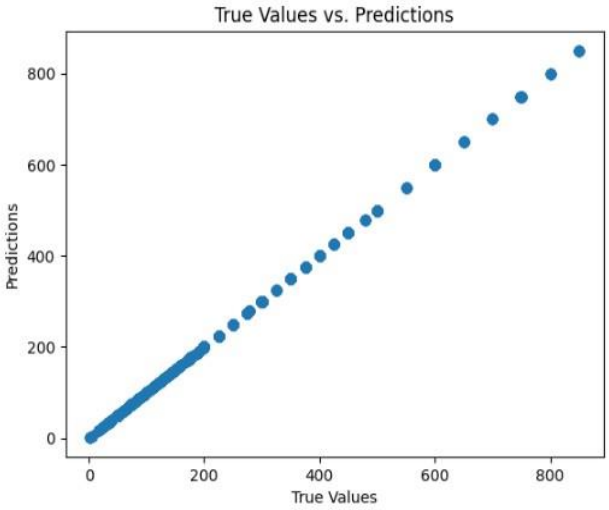

Figure 2: Linear Regression Model

Figure 3: Heatmap Linear Regression Model

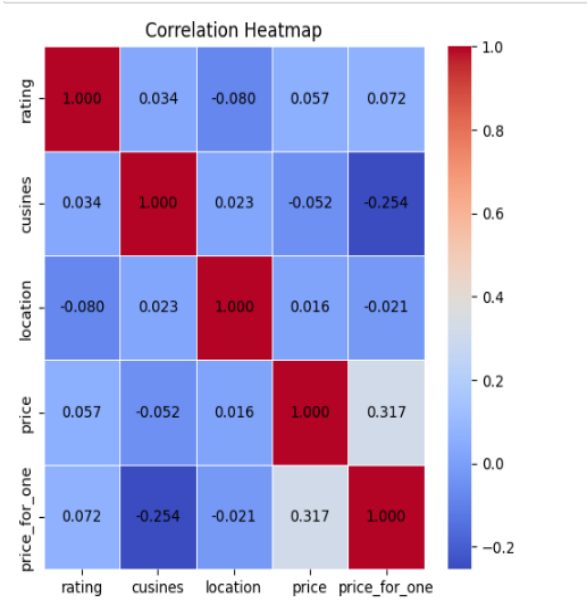

Figure 4: Heatmap Random Forest for Location

126	0.75	0.90	0.82	73
127	1.00	1.00	1.00	3
128	0.11	0.17	0.13	6
129	1.00	1.00	1.00	10
accuracy			0.88	21574
macro avg	0.79	0.76	0.77	21574
weighted avg	0.88	0.88	0.88	21574

Figure 5: Random Forest for Cuisine

	precision	recall	f1-score	support
0	1.00	1.00	1.00	37
1	1.00	0.67	0.80	3
2	0.91	0.91	0.91	132
3	0.92	0.93	0.92	720
4	0.78	0.54	0.64	13
5	0.00	0.00	0.00	5
6	0.14	0.33	0.20	3
7	1.00	1.00	1.00	25
8	0.50	0.35	0.41	20
9	0.88	0.88	0.88	132
10	0.99	0.93	0.96	261
11	1.00	1.00	1.00	74
12	0.88	0.88	0.88	2039
13	1.00	1.00	1.00	18
14	0.84	0.89	0.86	2552

Figure 6: Classification Matrix for Random Forest

28	1.00	1.00	1.00	38
29	1.00	0.99	1.00	475
30	1.00	1.00	1.00	3
31	1.00	1.00	1.00	110
32	1.00	1.00	1.00	39
33	1.00	1.00	1.00	2523
34	1.00	1.00	1.00	23
35	1.00	1.00	1.00	65
36	1.00	1.00	1.00	18
37	1.00	1.00	1.00	7376
accuracy			1.00	21574
macro avg	1.00	1.00	1.00	21574
weighted avg	1.00	1.00	1.00	21574

Figure 7: Random Forest for Location

	precision	recall	f1-score	support
0	1.00	1.00	1.00	289
1	1.00	1.00	1.00	225
2	1.00	1.00	1.00	231
3	1.00	1.00	1.00	244
4	1.00	1.00	1.00	27
5	1.00	1.00	1.00	90
6	1.00	1.00	1.00	17
7	1.00	1.00	1.00	2850
8	1.00	1.00	1.00	94
9	1.00	0.99	0.99	310
10	1.00	1.00	1.00	30
11	1.00	1.00	1.00	25
12	1.00	1.00	1.00	2246
13	1.00	1.00	1.00	259

Figure 8: Classification matrix for Random Forest

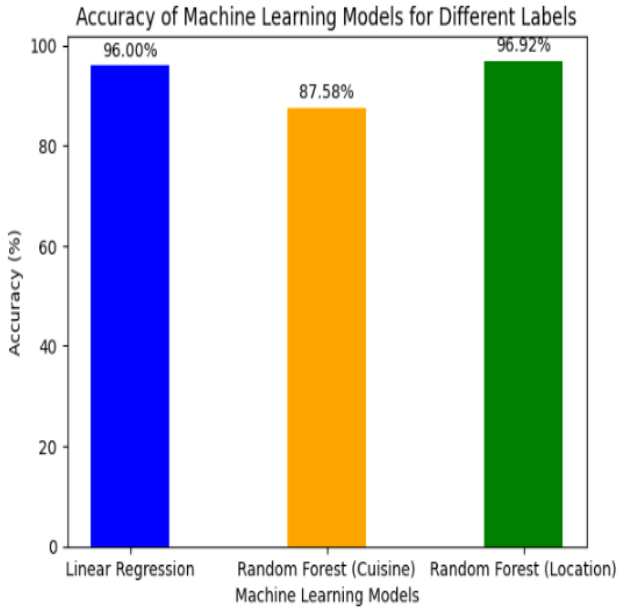

Figure 9: Accuracies of Model Used

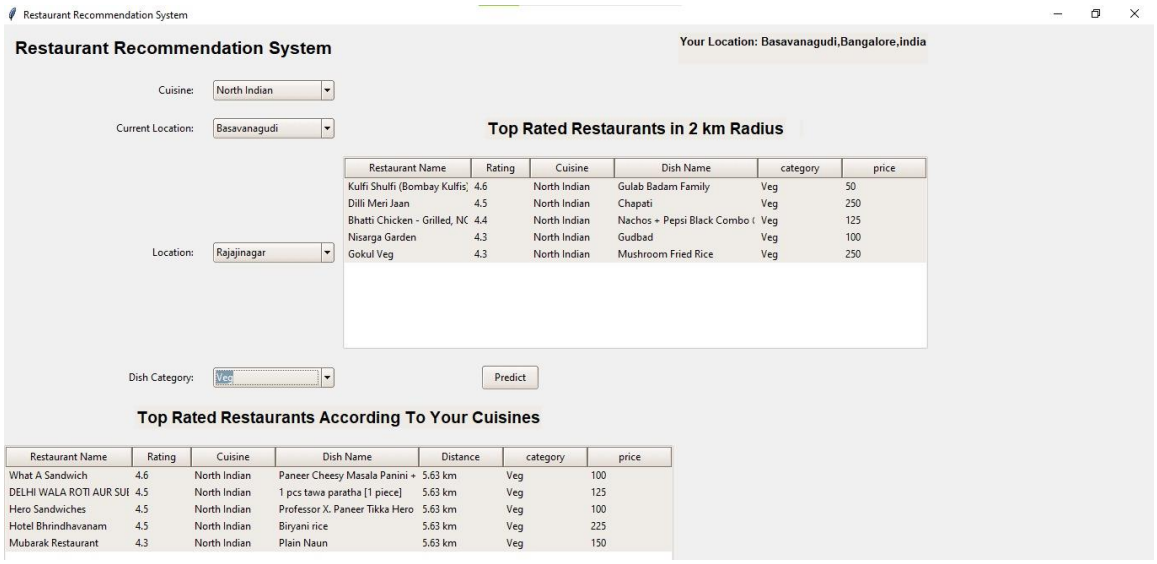

Figure 10: Proposed Model representation

7. CONCLUSION

The main purpose of this research is to create a restaurant recommendation that customers can use as an application using machine learning and the web interface. The app is designed to allow users to guess which restaurants are available and find the area's own famous dishes. The app allows customers to get ratings.

Popularity based and collaborative filtering makes recommendations better so that all users can easily guess restaurants using the app. Generally, users want nearby restaurants. We solved this problem by adding restaurants to the database. This way, our machine learning algorithms can easily predict restaurants based on the customer's current location.

This restaurant recommendation app will give users a better experience in searching for restaurants in nearby areas in a short time. This will reduce the user's effort and make the time more convenient.

REFERENCES

1. Yanai, K., Maruyama, T., & Kawano, Y. (2014). A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients. International Journal of Interactive Mobile Technologies, 8(2).
2. Lee, T., & Soatto, S. (2011, June). Learning and matching multiscale template descriptors for real-time detection, localization and tracking. In CVPR 2011 (pp. 1457-1464). IEEE.
3. Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. Education and Information Technologies, 25, 2635-2664.
4. Das, D., Sahoo, L., & Datta, S. (2017). A survey on recommendation system. International Journal of Computer Applications, 160(7).
5. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. Recommender systems handbook, 257-297.
6. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. Decision support systems, 74, 12-32.

7. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), 734-749.
8. Toledo, R. Y., Alzahrani, A. A., & Martinez, L. (2019). A food recommender system considering nutritional information and user preferences. IEEE Access, 7, 96695-96711.
9. Gomathi, R. M., Ajitha, P., Krishna, G. H. S., & Pranay, I. H. (2019, February). Restaurant recommendation system for user preference and services based on rating and amenities. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-6). IEEE.
10. Jain, L., Katarya, R., & Sachdeva, S. (2019, November). Opinion Leader discovery based on text analysis in Online Social Network. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 446-450). IEEE.
11. Jain, L., Katarya, R., & Sachdeva, S. (2019, November). Opinion Leader discovery based on text analysis in Online Social Network. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 446-450). IEEE.