**Predicting Restaurant Rating Using Regression Analysis Approach**

Submitted by

221FA04503 G. Joseph Anand Kumar

221FA04006 N. Bala Sai

221FA04575 V. Leela Venkata Mani Sai

221FA04591 E. Sai Naga Lakshmi

Under the guidance of

**Sk. Sajidha Sulthana**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN’S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH**

**(Deemed to be UNIVERSITY)**

Vadlamudi, Guntur.

ANDHRA PRADESH, INDIA,

PIN-522213.



**CERTIFICATE**

This is to certify that the Field Project entitled “Predicting Restaurant Rating Using Regression Analysis Approach” that is being submitted by 221FA04503 - G. Joseph Anand Kumar, 221FA04006 - N. Bala Sai, 221FA04575 - V. Leela Venkata Mani Sai, 221FA04591 - E. Sai Naga Lakshmi;for partial fullfilment of Field Project is a bonafide work carried out under the supervision of **Mr. Pavan Kumar, Department of CSE**.

Guide Name & HoD,CSE Dean,SoCl

Signature

**DECLARATION**

We hereby declare that the Field Project entitled “Predicting Restaurant Ratings Using Regression Analysis Approach” is being submitted by 221FA04503-G. Joseph Anand Kumar , 221FA04006-N. Bala Sai , 221FA04575-V. Leela Venkata Mani Sai , 221FA04591-E. Sai Naga Lakshmi ,in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of **Mr. Pavan Kumar, M.Tech., Assistant Professor, Department of CSE**.

## **By**

221FA04503 G. Joseph Anand Kumar

221FA04006 N. Bala Sai

221FA04575 V. Leela Venkata Mani Sai

221FA04591 E. Sai Naga Lakshmi

Date:

**ABSTRACT**

Starting a restaurant now seems to be a task that is quite challenging, especially because of the competitive market environment that exists today. Consumers are much more exposed to different choices for where they will dine. Therefore, rating restaurants has become a powerful tool in gauging the quality of a restaurant since a good rating attracts customers. Every new restaurant owner would therefore be interested in knowing how likely it is that a particular venture would succeed.

This paper will present a method of reliable rating prediction for the restaurant. It stems from different factors, which reveal the characteristics of a restaurant, through which it allows an average rating to be predicted. Besides this, it helps in comparing all aspects of a restaurant, and so before launching a business there is more informed judgment involved. This reduces the risk of loss and time saving; otherwise, the whole process becomes calculated data-driven activity.

The ratings would be predicted through seven regression models during the analysis, and factors controlling the results are emphasized in planning. The performance metrics of the model are carefully assessed so as to select the most accurate model for future predictions. Therefore, the analysis would offer a robust framework for informed decisions that may lead toward success in the new restaurant establishments.

**KEYWORDS**

Restaurant Rating Prediction, Machine Learning, Regression Models ,Data Analytics, Business Optimization, Restaurant Planning, Customer Satisfaction, Random Forest, XG Boost, ADA Boost, Risk Mitigation, Predictive Analytics

**CONTENTS**

**1. Introduction**

1.1 Importance of Restaurant Ratings

1.2 Challenges in Restaurant Rating Prediction

1.3 Role of Regression Models in Prediction

**2. Literature Survey**  
 2.1 Existing Approaches for Rating Prediction  
 2.2 Sentiment Analysis and Collaborative Filtering  
 2.3 Machine Learning Techniques in Restaurant Rating Prediction  
 2.4 Research Gaps and Motivation

**3. Proposed System**  
 3.1 Overview of the Rating Prediction System  
 3.2 Dataset Details  
 3.2.1 Restaurant Attributes in the Dataset  
 3.3 Data Pre-processing  
 3.3.1 Handling Missing Data  
 3.3.2 Feature Encoding and Scaling  
 3.4 Regression Models Used  
 3.4.1 Linear Regression  
 3.4.2 Support Vector Machine (SVM)  
 3.4.3 Decision Tree  
 3.4.4 Random Forest  
 3.4.5 K Nearest Neighbors (KNN)  
 3.4.6 ADA Boost  
 3.4.7 XG Boost  
 3.5 Model Evaluation Criteria  
 3.5.1 Root Mean Square Error (RMSE)  
 3.5.2 Mean Absolute Percentage Error (MAPE)  
 3.5.3 R-Squared (R²) Score  
 3.6 Comparison of Model Performance

**4. Implementation**  
 4.1 Setting up the Environment for Experimentation  
 4.2 Preprocessing the Dataset  
 4.3 Training the Models

**5. Experimentation and Result Analysis**  
 5.1 Performance of Each Regression Model  
 5.2 Error Metrics and Interpretation of Results  
 5.3 Most Effective Model for Rating Prediction

**6. Conclusion**  
 6.1 Summary of Findings  
 6.2 Future Scope of the Work

**7. References**

**LIST OF FIGURES**

1. Name vs Number of Restaurants
2. Number of Restaurants Accepting Online Orders
3. Number of Restaurants with Book Table Facility
4. Location
5. Percentage of that restaurants placed in that location
6. Location vs Count
7. Restaurant Type
8. Type of Restaurant in Percentage
9. Average cost in restaurant for 2 people
10. Average cost for 2 People in Rupees
11. Top 10 dishes liked in Banglore
12. Average rating for Banglore Restaurants
13. Restaurant Type vs Rate
14. Top 10 cuisines in Banglore
15. Rate vs Online Order
16. True vs Predicted rate

**INTRODUCTION**

In the competitive world of hospitality, a restaurant's success is significantly shaped by customer reviews and ratings. With numerous dining choices available, these ratings serve as a crucial measure of a restaurant's quality and customer satisfaction. Positive reviews are vital for drawing in new customers, while negative feedback can discourage potential diners. Therefore, accurately forecasting a restaurant's future rating is essential for new businesses, offering valuable insights into performance based on factors like location, cuisine, and pricing.

Launching a new restaurant requires a considerable investment of both money and time. Entrepreneurs often grapple with assessing the potential success of their venture even before it opens. While traditional methods such as market research and customer surveys provide some insights, they often lack a data-driven, predictive approach. This is where machine learning and data analytics become essential.

This paper introduces a system that employs regression models to predict a restaurant's rating based on controllable factors before it opens. By examining various characteristics, such as online ordering options, table reservation availability, and the average cost for two diners, the system produces a predicted rating. Additionally, it facilitates comparisons among different restaurant aspects, empowering business owners to make informed decisions and reduce the risks tied to launching a new establishment.

**LITERATURE SURVEY**

Restaurant rating forecasting has gained significant attention in the last few years, driven by the need for data-driven decision-making in this fiercely competitive food and beverage industry. Extensive methodologies have been researched to address this challenge, including sentiment analysis, collaborative filtering, and a variety of machine learning approaches. This part outlines some of the key contributions in the literature to date and points out gaps that this paper addresses.

2.1 Current Rating Prediction Approaches

Several rating prediction-based opinion summary studies have been done. For example, Shibata et al. (2020) applied CSPD to predict the ratings given to hotels by customers and assigned sentiment scores to their reviews. Although this method promises the best rating prediction given existing reviews, it did not assist in the rating prediction of newly operating restaurants because these latter restaurants do not have reviews from customers. This is very key for newly opened restaurants, wherein no data exists for any of the restaurants.

Chanwisitkul et al. (2018), in another independent study, utilized the application of text mining methods for assessing customer satisfiability using hotel reviews. Although the research provided informative insights regarding customers' perceptions, the main purpose was to guide hotels in making suitable decisions rather than anticipating ratings. Besides, the method did not use other variables like location, food type, or price while dining that plays a crucial role when creating a rating prediction model.

Collaborative filtering has been observed as one of the approaches with the most rate prediction usage. Pradhan et al. (2016) developed a collaborative filter model that classified similar items and users to predict ratings for unrated items. However, the requirement of prespecified ratings for a similar item limits the applicability of the model in rating prediction where there is no history of the restaurant. Saha and Santra (2017) worked on the usage of text feedback coming from online restaurant reviews in extracting sentiment scores, and it is these scores which inform recommendation systems. Although this works well for established restaurants, by not taking into consideration any external variables such as geography and amenities available, the efficacy of the model in rating for new firms is suspicious.

This approach in addition mainly focused on pre-existing textual information, the influence of business characteristics such as reservation options or expenses for a pair of individuals.

2.2 Machine Learning Techniques for Rating Prediction

Recent works have successfully employed machine learning algorithms for rating predictions from structured data. Sivabalaselvamani and Soorya (2020) developed a system that could determine customer satisfaction in automated restaurants with the help of facial expressions by utilizing CNN. Though this is innovative, such a method was limited to the use of face-recognition systems within specific dining settings and could not afford to focus on traditional restaurant settings.

Shihab et al. (2018) applied multiple machine learning algorithms on Yelp data for predicting the best location for opening a restaurant. They came across 75 attributes determining the restaurant's success, and although they focus mainly on geographical factors, others are less considered. So, their work alone was insightful but could not lead to a complete solution in predicting restaurant ratings based on controllable attributes such as variation offered cuisines or services.

2.3 Research Gap and Motivation

There are several key limitations that the existing literature offers today. Most studies concentrate on rating forecasts of established businesses with a time-series review history or limiting the number of dimensions, including customer reviews and geography. This narrow scope limits them to newly opened restaurants where rating forecasts need an evaluation of controllable factors beforehand, including service options, price strategy, and menu options. Furthermore, despite the great investigation of many machine learning methodologies, there is still a gap in the comprehensive comparative study of various regression models intended for rating restaurant business based on the respective characteristics. To fill that gap, this study tries to assess and compare the effectiveness of seven different regression models based on easily manageable factors that influence the ratings of a particular restaurant. This gives an all-inclusive perspective with which to rate predictions including online ordering, table reservation availability, and average spend for two that can equip the new restaurants with information to make informed decisions. Emphasis is also placed in the existing literature on a sound predictive framework that entails considerations of qualitative as well as quantitative factors that the research is designed to address by incorporating machine learning and data analytics into the system.

**OBJECTIVE**

The primary objective of this paper is to develop a robust system that accurately predicts restaurant ratings based on controllable business factors before the restaurant is launched. By leveraging multiple regression models, the system aims to provide reliable predictions, helping restaurant owners make informed decisions regarding key aspects such as location, cuisine, pricing, online ordering options, and table booking availability.

Key objectives include:

Predicting Ratings: To predict the average rating of a new restaurant based on characteristics that can be controlled before its launch.

Model Comparison: To evaluate and compare the performance of seven regression models (Linear Regression, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbors, ADA Boost, and XG Boost) to identify the most accurate model for rating prediction.

Informed Decision-Making: To provide actionable insights to restaurant owners, enabling them to optimize features such as service offerings, cost structure, and location before investing in their business.

Risk Mitigation: To minimize the risk of investment failure by offering a data-driven approach that enhances the likelihood of business success in the competitive restaurant industry.

Enhancing Business Planning: To use machine learning techniques to transform restaurant planning into a calculated and well-thought-out process, improving the odds of success in a crowded market.

By achieving these objectives, the paper seeks to bridge the gap between restaurant planning and future business performance through predictive analytics.

**METHODOLOGY**

The report outlines the structured methodology for restaurant rating forecasts by using regression models, while focusing on the key business factors under the control of a restaurant owner. The approach has several pivotal stages- viz., data acquisition, preprocessing, feature selection, model development, and evaluation. This framework would ensure that the ratings thus generated are not only accurate but also practical and valuable resource inputs toward restaurant owners' decision-making processes.

**1. Data Collection**

The source of data for this study is a publicly available dataset on Kaggle, containing over 43,000 restaurants in Bengaluru, India. It contains various attributes, including the website, address, online ordering and table booking presence, average cost for two, type of restaurant, location, number of dishes liked, and customer rating. All these attributes have been taken as independent variables, but restaurant rating is taken as the dependent variable.

**2. Data Preprocessing**

Data preprocessing is an important step in the processing and preparation of raw data for further analysis. The following operations are performed:

Treatment of Incomplete Data: Columns with missing values are replaced with appropriate values. Rows have been dropped with more than 50% missing.

Renaming and Data Type Conversion: Renaming of column names makes it easier to understand, while data types are presented in an appropriate format (such as rating converted into float, numerical columns presented uniformly).

Feature Encoding: categorical variables-the availability of online ordering and table reservations, for example, use label encoding while location-based data uses ordinal encoding.

Feature Selection: Features like URLs, addresses, contact information, as all these do not contribute to the rating prediction are removed. Key Features that remained include restaurant type, location, online order availability, table booking availability, cuisine count, and approximate cost for two.

**3. Scaling Data**

The MinMax scaler is applied in order to standardize across these independent variables, bringing all the features to a uniform range, thus avoiding overinfluence by one variable on the model's predictions.

**4. Model Selection and Training**

Seven regression models are identified to be predicted. The models were selected, based on the diversified methods in handling linear as well as nonlinear relations within the efficacy of an analogous predictive effort:

Linear Regression: Assumes there is a linear relationship between independent variable and dependent rating variable.

SVM: SVM uses hyperplanes in order to detect the optimal fit in high-dimensional spaces.

Decision Tree: Constructs a tree-like model of decisions, dividing the data into subsets based on feature importance.

Random Forest: An ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting.

KNN: The rating is predicted by taking the average of the ratings assigned to similar restaurants in the dataset.

ADA Boost: a boosting algorithm that improves a weak model by giving more importance to the wrongly classified instances.

XG Boost: A highly efficient gradient boosting method that constructs models sequentially as corrections from the previous model.

The dataset is divided into training and testing subsets in an 80:20 ratio, whereby 80% of the data is allocated for the purpose of training the models, while 20% is set aside to evaluate their performance.

**5. Model Checking**

Based on three key performance metrics, here are the evaluation models.

Root Mean Square Error (RMSE): It measures the average magnitude of the errors between the actual and predicted ratings.

Mean Absolute Percentage Error (MAPE): Measures prediction accuracy as a percentage of the actual values. R-Squared (R²) Score: Indicates how well the independent variables explain the variability in the dependent variable.

These metrics allow to evaluate how precise and reliable is each model. The model is said to be effective when its R² score is close to 1, in combination with low values for both RMSE and MAPE. This comparative analysis among models allows to determine which one can effectively predict the ratings at restaurants.

**6. Final Model Selection**

As soon as the models get evaluated with the help of the above metrics, the best-performing model is determined. Models like Random Forest, XG Boost, and ADA Boost are anticipated to be better because they handle complex interactions among variables very efficiently. Next, the chosen model is applied for prediction on the test dataset.

**7. Summary and Analysis**

The findings are examined to ascertain the factors that most significantly affect a restaurant's anticipated rating. This analysis aids restaurant proprietors in comprehending how particular characteristics—such as geographical placement, pricing structures, and service alternatives—affect customer evaluations. Through the identification of the most efficient model and the principal elements that shape ratings, the system offers critical insights for restaurant strategizing.

The methodology ultimately provides a predictive framework that assists prospective restaurant proprietors in mitigating risks and in data-informed decision-making before the establishment of their enterprise. This methodology brings more accuracy to the forecast, lets for a closer comparison of the models, and is well-suited to discover actionables from the data.

**IMPLEMENTATION**

The implementation of the restaurant rating prediction system is divided into several stages, including environment setup, data preprocessing, model training, and evaluation. This section outlines the steps taken to implement the system, from setting up the necessary tools to running the machine learning algorithms and analyzing the results.

**1. Environment Setup**

The implementation is carried out in a Python environment, leveraging machine learning libraries for data processing, model building, and visualization. The development environment is set up using the following tools and libraries:

- Python 3.x: The primary programming language used for building the system.

- Jupyter Notebook: Used for interactive development and debugging.

- Scikit-learn: A popular library for implementing machine learning algorithms.

- Pandas: For data manipulation and preprocessing.

- NumPy: For numerical operations and data handling.

- Matplotlib/Seaborn: For data visualization.

To begin, all necessary libraries are installed using the package manager `pip`

**2. Data Preprocessing**

Before training the models, the raw data needs to be preprocessed to ensure that it is clean and formatted correctly. The following steps outline the preprocessing pipeline:

1. Loading the Dataset

2.Handling Missing Values

3. Feature Encoding

4. Normalization

5. Splitting the Data

**3. Model Building**

Seven different regression models are implemented to predict restaurant ratings. The models are trained using the preprocessed data and evaluated using test data. **4. Model Evaluation**

The performance of each model is evaluated using standard regression metrics such as Root Mean Square Error (RMSE) and R² Score. These metrics allow for a comparison of the models and help identify the best-performing algorithm.

**5. Comparison of Models**

The performance of all seven models is compared based on the evaluation metrics (RMSE and R² Score). The model with the highest R² score and the lowest RMSE and MAPE is selected as the best-performing model for restaurant rating prediction.

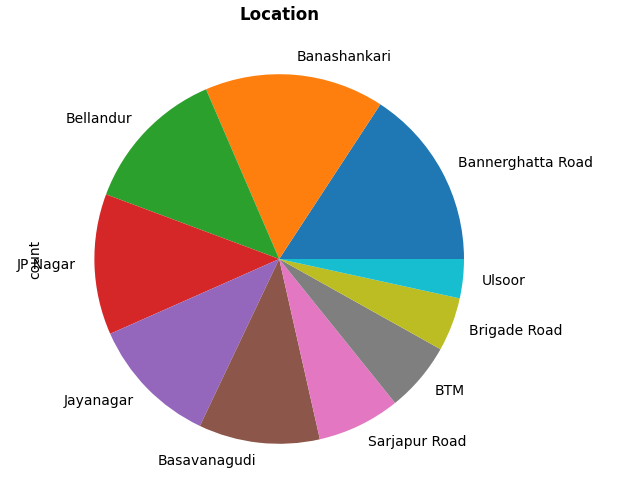
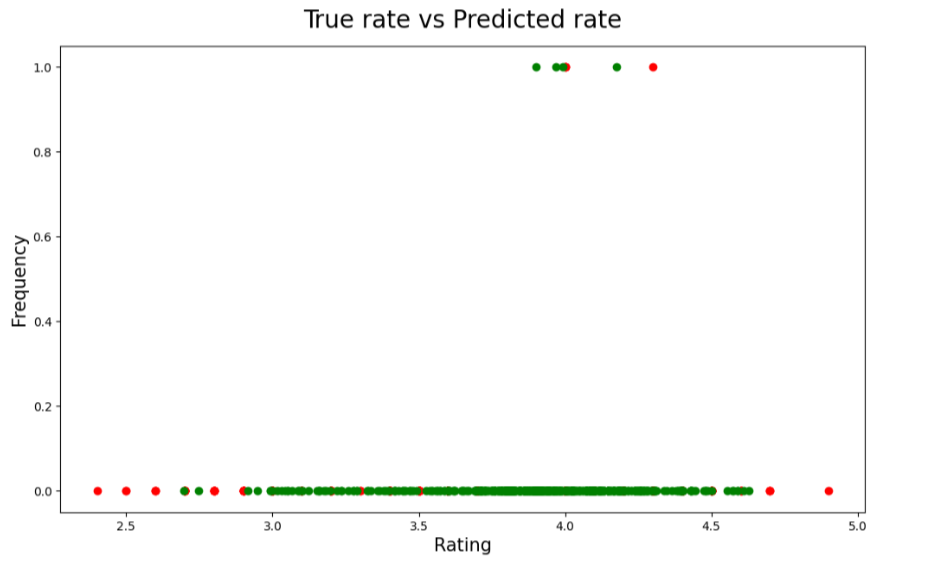
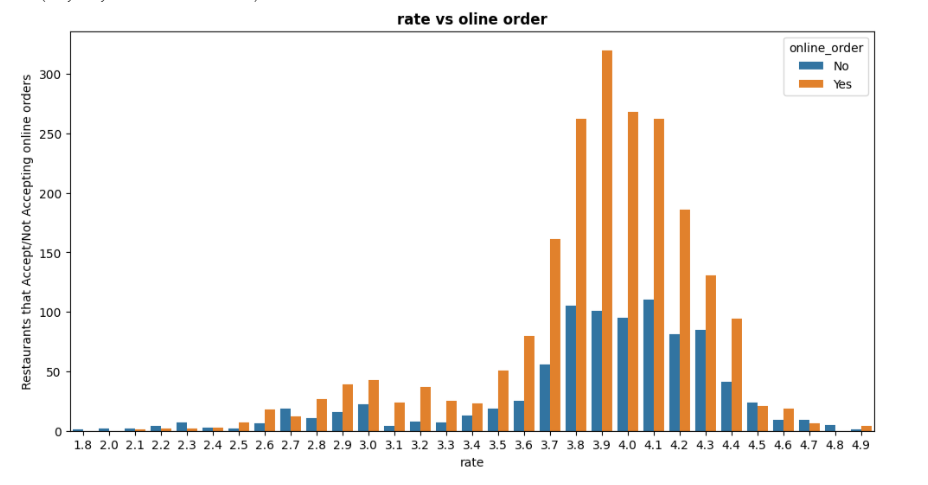
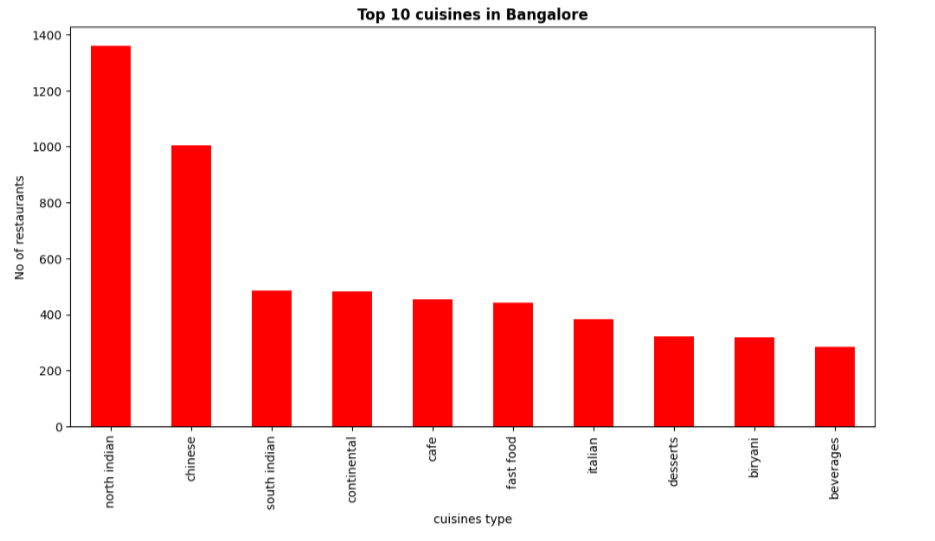
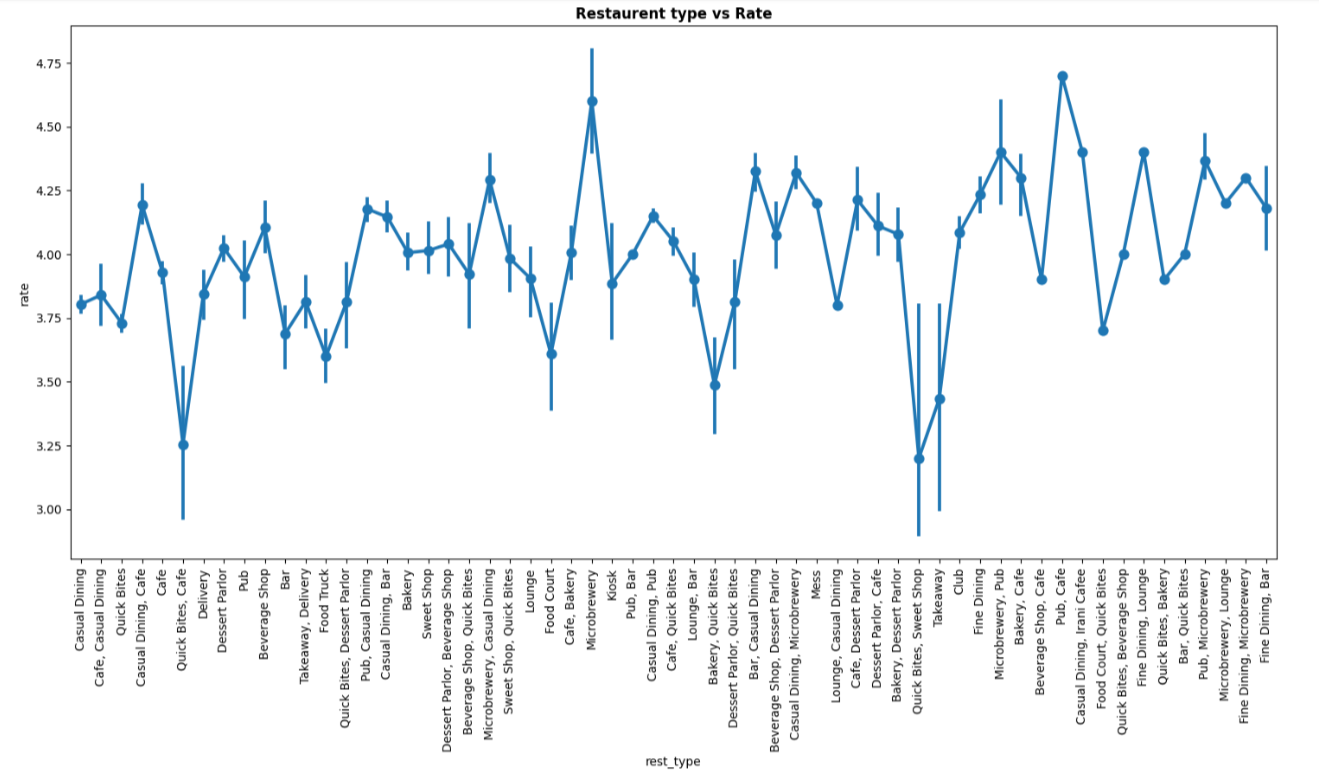
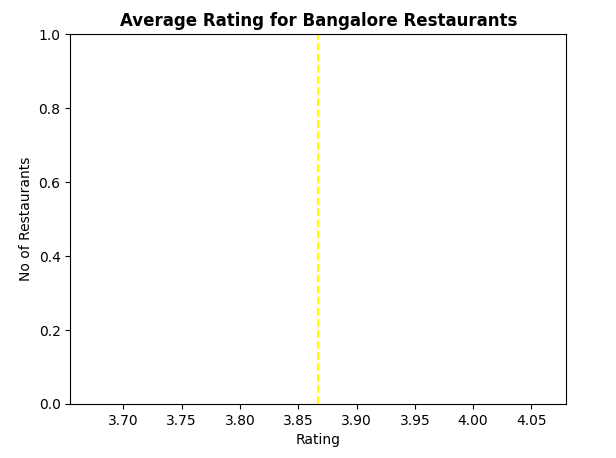
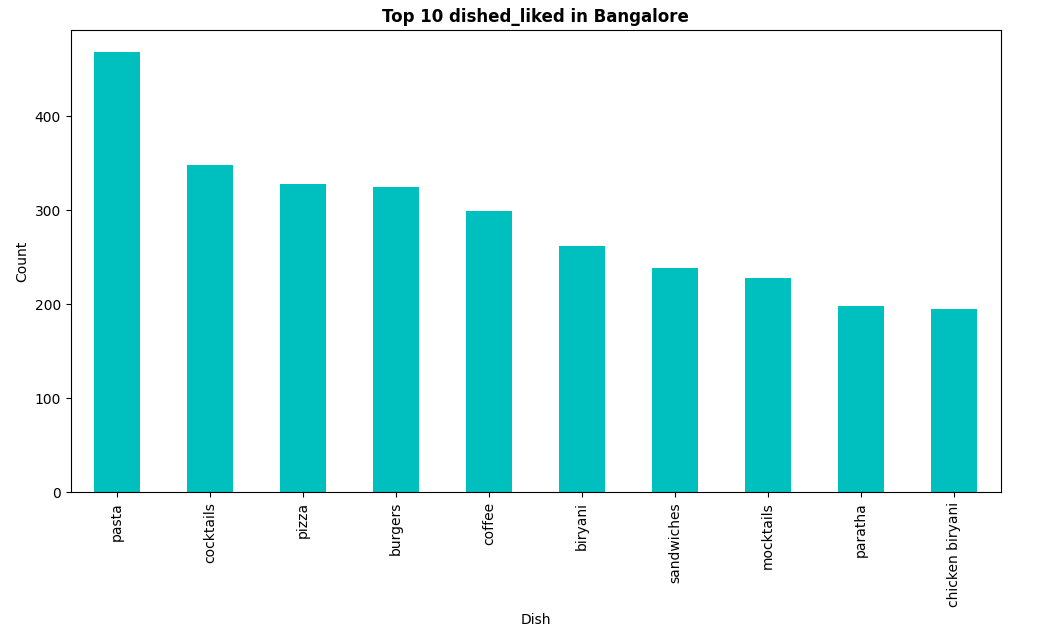
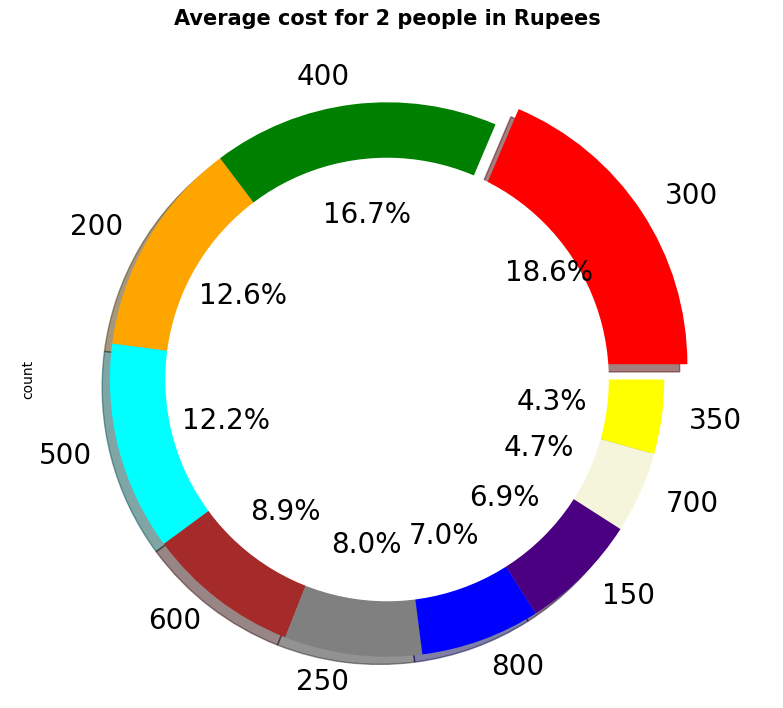
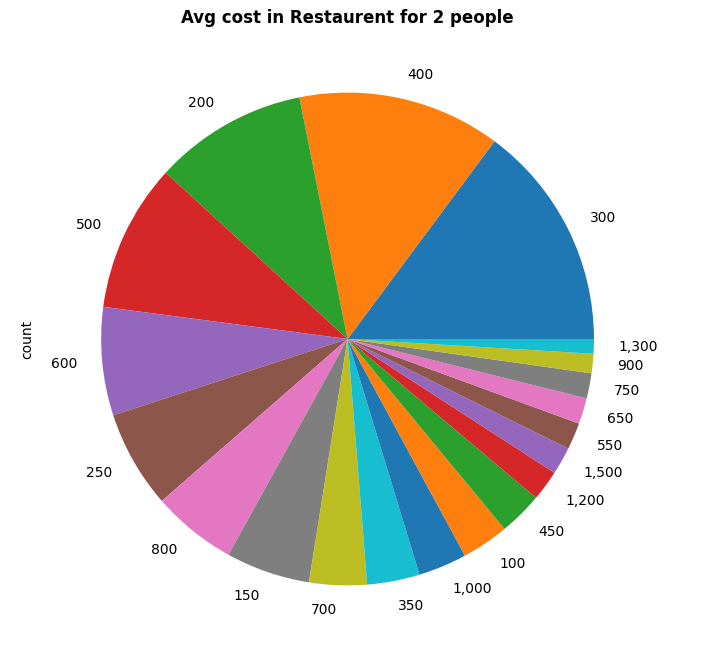
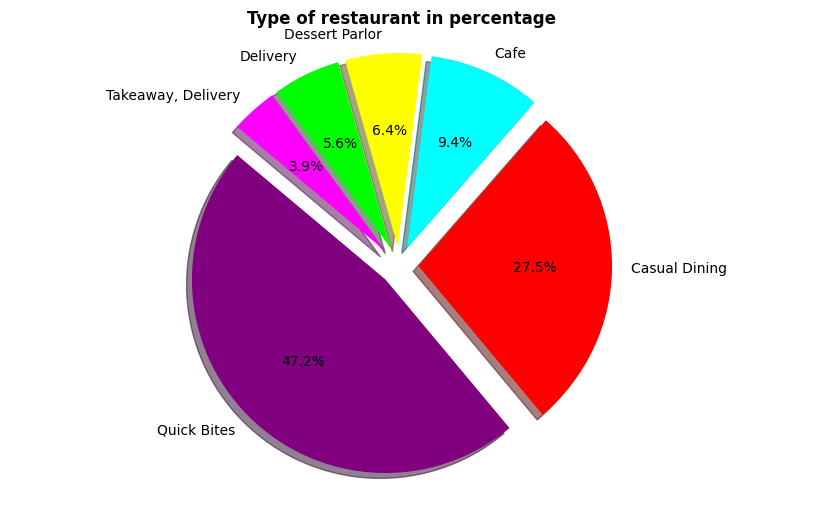
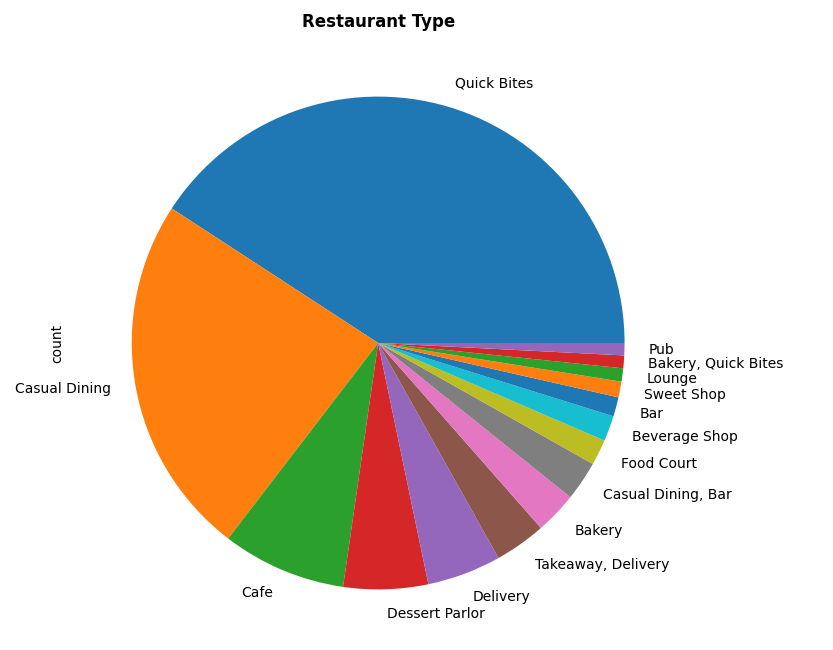
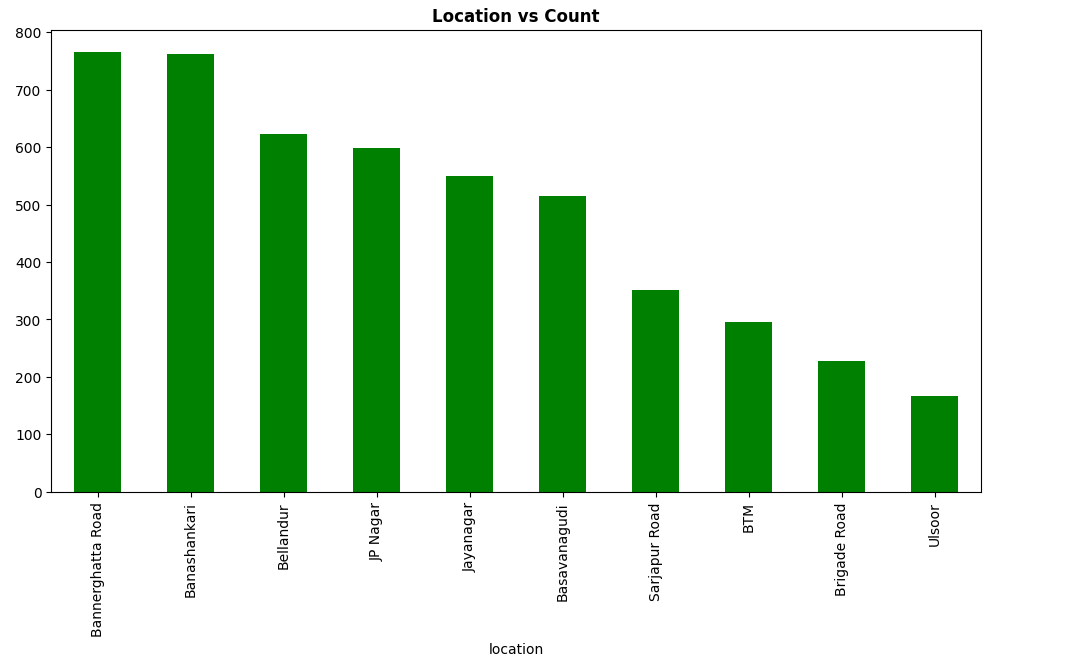
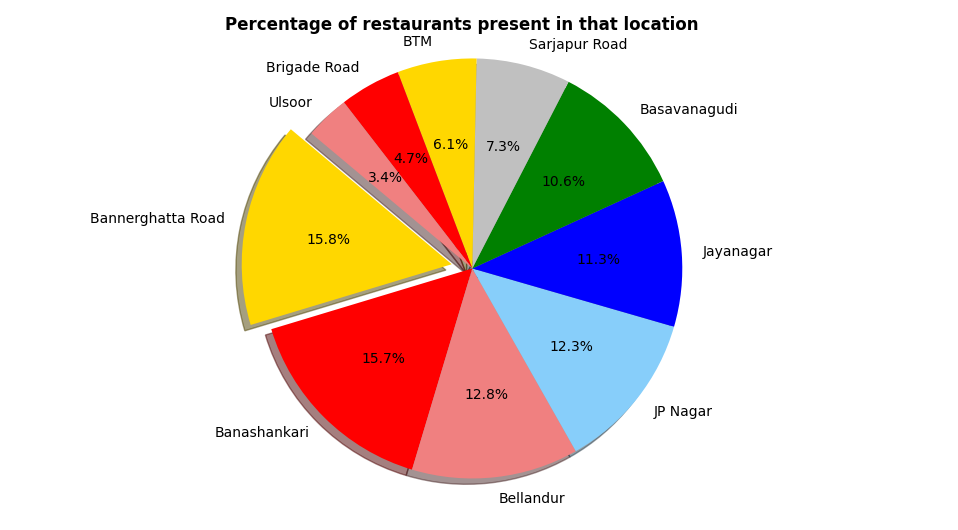
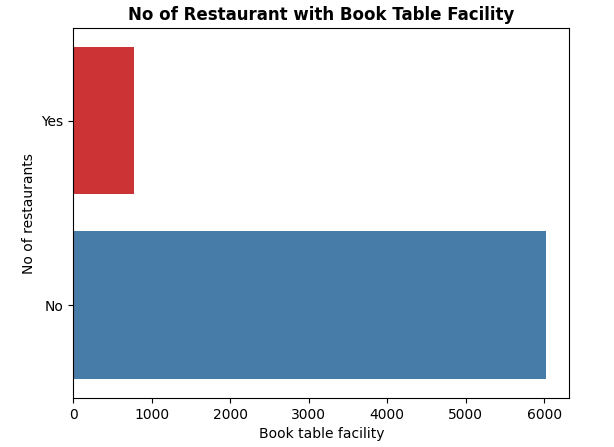
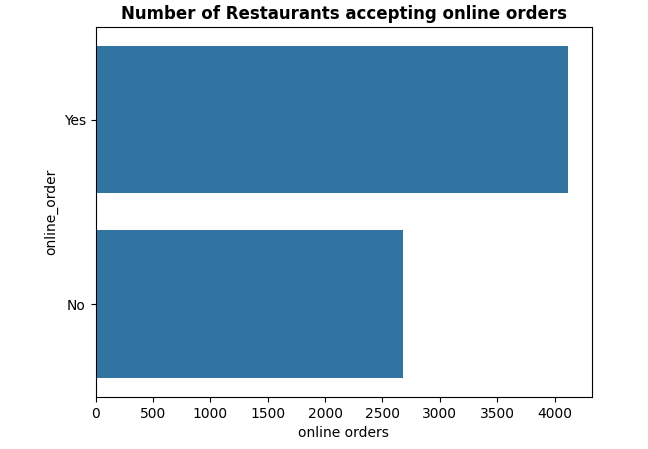
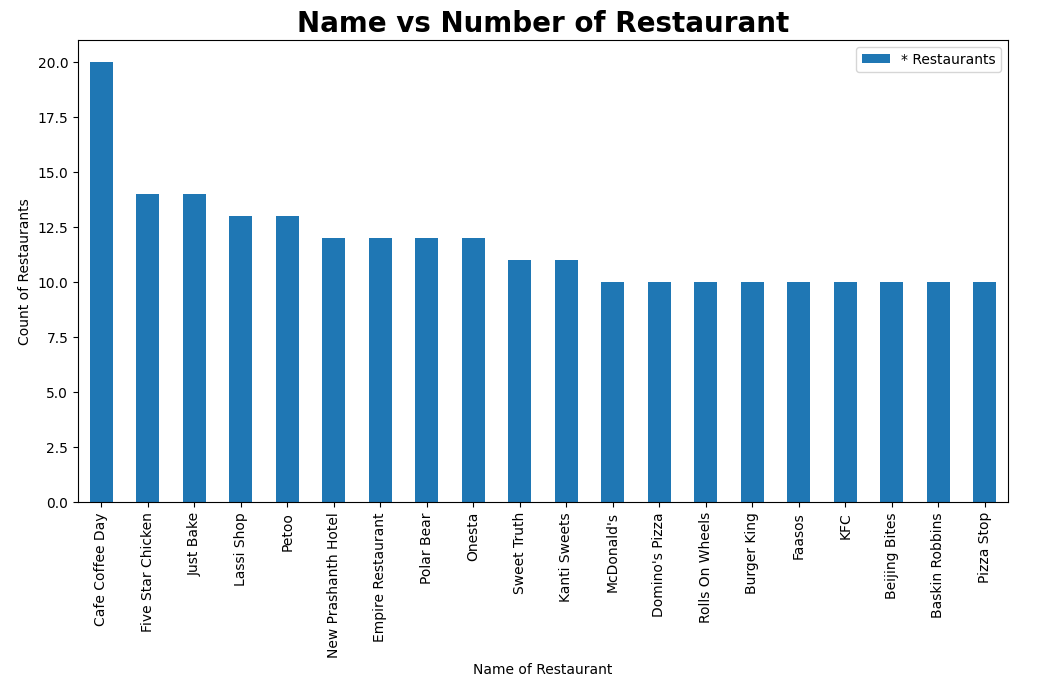
**6. Visualization of Results**

To better understand the results, visualizations are generated to show the distribution of predicted versus actual ratings, as well as the importance of various features in the model.

**7. Deployment**

Once the best model is identified, it can be deployed as part of a predictive system for restaurant owners. The system allows users to input key attributes of their restaurant, and the model will output the predicted rating, helping to guide business decisions before the restaurant is launched.

This implementation framework ensures that the system is robust, scalable, and capable of providing actionable insights for restaurant owners to optimize their business strategy.

**VISUALIZATION OF RESULTS**

**CONCLUSION**

This paper presents a thorough method for predicting restaurant ratings using machine learning regression models that focus on key business factors under the control of restaurant owners. By examining data related to aspects such as location, pricing, online ordering options, and table reservation capabilities, the system can provide accurate forecasts of a restaurant’s potential rating even before it opens. The study implements and compares seven different regression models—Linear Regression, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbors, ADA Boost, and XG Boost—to determine the most effective model for rating prediction based on performance metrics like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R² score. Models such as Random Forest, ADA Boost, and XG Boost demonstrate outstanding performance, yielding precise predictions with minimal error.

This research offers restaurant owners a valuable data-driven resource that can assist in optimizing business decisions and reducing financial risks. By forecasting ratings based on adjustable factors prior to a restaurant's launch, this system facilitates more strategic planning, thereby decreasing the chances of failure in a highly competitive market. Additionally, the study emphasizes the significance of utilizing machine learning techniques not only to anticipate customer satisfaction but also to aid restaurant management in making informed decisions regarding service offerings, pricing, and location. Future research could focus on broadening the model to include more diverse datasets and additional factors, further enhancing prediction accuracy and extending the system's applicability to a wider variety of restaurant environments.

**REFERENCES**

* **SUHAS** **SOMASHEKAR , SUHAS MALLESH** (2021) - Restaurant Rating Prediction Using Regression , Proceedings of the Fifth International Conference on Electronics, Communication and Aerospace Technology (ICECA 2021) , IEEE Xplore Part Number: CFP21J88-ART; ISBN: 978-1-6654-3524-6
* **J.** **Priya** **-** Predicting Restaurant Rating using Machine Learning and comparison of Regression Models , (2020) International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)
* **Ibne Farabi Shihab; Maliha Moonwara Oishi; Samiul Islam; Kalyan Banik; Hossain** **Arif -** A Machine Learning Approach to Suggest Ideal Geographical Location for New Restaurant Establishment , (2019)
* **Mara-Renata Petrusel; Sergiu-George Limboi -** A Restaurants Recommendation System: Improving Rating Predictions Using Sentiment Analysis , 2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)
* **Tugçe Bilen; Muge Erel-Özçevik; Yusuf Yaslan; Sema F. Oktug -** A Smart City Application: Business Location Estimator Using Machine Learning Techniques , DOI: 10.1109/HPCC/SmartCity/DSS.2018.00219
* **Ismam Hussain Khan; Md Habib Ullah Khan; Md Mamun Howlader -** An Intelligent Approach for Food Recipe Rating Prediction Using Machine Learning , 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) , DOI: 10.1109/CAIDA51941.2021.9425031
* **Neha Vaish; Nidhi Goel; Gaurav Gupta -** Machine Learning Techniques for Sentiment Analysis of Hotel Reviews , 2022 International Conference on Computer Communication and Informatics (ICCCI) , DOI: 10.1109/ICCCI54379.2022.9740876
* **Sandeep Bhatia; Chirag Arya; Soniya Verma; Devraj Gautam; Bharat Bhusan Naib; Abhishek Kumar -** Predict Success of a Zomato Restaurant using Machine Learning , 2023 World Conference on Communication & Computing (WCONF) , DOI: 10.1109/WCONF58270.2023.10235064
* **Yi Luo Xiaowei Xu -** Predicting the Helpfulness of Online Restaurant Reviews Using Different Machine Learning Algorithms: A Case Study of Yelp , 2019, 11(19), 5254; <https://doi.org/10.3390/su11195254>
* **Joshua -** Predicting Restaurant Rating using Machine Learning and comparison of Regression Models , 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) , DOI: 10.1109/ic-ETITE47903.2020.238