IJARETY

International Journal of Advanced Research in Education and TechnologY *(IJARETY)*

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE

SJIF Scientific Journal Impact Factor

doi

crossref

निस्केयर

NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 10, Issue 2, March 2023 ||

DOI:10.15680/IJARETY.2023.1002025

Enhanced Restaurant Rating Predictions: A Study of Linear and Ensemble Regression Methods

Zeba Vishwakarma[1], Jaya Choubey[2], Divya Pandey[3], Aayush Tiwari[4], Ifat Khushi[5]

Department of Computer Science Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, Madhya Pradesh, India [1, 2, 3,4,5]

ABSTRACT: In this study, we investigate the application of various regression techniques to predict restaurant ratings based on a comprehensive Zomato dataset. The dataset encompasses 51,717 entries with features such as restaurant type, location, cuisine, and cost estimates. We employed four regression models: Linear Regression, CatBoost Regressor, Random Forest Regressor, and Decision Tree Regressor, to evaluate their performance in predicting ratings. The models were assessed using the R-squared (R^2) metric to determine their effectiveness. The Random Forest Regressor achieved the highest R^2 score of 0.875, indicating its superior predictive capability. The CatBoost Regressor also performed well, demonstrating the effectiveness of gradient boosting methods. The study provides insights into the performance of various regression models and highlights the advantages of ensemble methods for accurate restaurant rating predictions.

KEYWORDS: Regression Analysis, Zomato Dataset, Random Forest, CatBoost, Predictive Modeling, R-squared

I. INTRODUCTION

In recent years, the restaurant industry has seen tremendous growth, fueled by the rise of online platforms for discovering dining options. One such platform, Zomato, provides a rich dataset offering valuable insights into various aspects of restaurant operations and customer preferences. The dataset, sourced from Kaggle, includes real-time information on restaurants across different locations in Bangalore. This dataset encompasses a variety of features such as restaurant names, addresses, online ordering options, table booking availability, ratings, votes, locations, restaurant types, popular dishes, cuisines, and cost estimates for two people. Ratings play a pivotal role in the decision-making process for diners. Research by BrightLocal indicates that 88% of consumers trust online reviews as much as personal recommendations, and 72% of individuals take action after reading positive reviews [1]. Therefore, accurately predicting restaurant ratings can greatly influence consumer choices and restaurant performance.

The dataset used in this study consists of 51,717 entries with 17 features. This data provides a comprehensive view of restaurant attributes that can be utilized to predict ratings effectively. By leveraging this extensive dataset, we aim to evaluate the performance of various regression models in predicting restaurant ratings.

Predicting restaurant ratings involves navigating complex relationships among features such as location, cuisine type, and cost. Traditional linear models often fall short in capturing these intricate patterns, making advanced regression techniques essential for this analysis [2][3].

The study applies several regression models, including Linear Regression, CatBoost Regressor, Random Forest Regressor, and Decision Tree Regressor. Linear Regression serves as a baseline for simple relationships, while ensemble methods like Random Forest and gradient boosting techniques like CatBoost enhance accuracy by modeling non-linear interactions and complex patterns in the data [4][5].

This study aims to assess the performance of different regression models in predicting restaurant ratings from the Zomato dataset. By comparing Linear Regression, CatBoost Regressor, Random Forest Regressor, and Decision Tree Regressor, the research seeks to identify the most accurate predictive model. The findings will provide insights into the efficacy of these models for similar predictive tasks and contribute to the advancement of recommendation systems in the hospitality sector. Accurate rating predictions benefit both restaurant managers and consumers. For restaurant managers, it offers insights into customer preferences and aids in strategic decision-making. For consumers, it enhances their dining choices and improves the overall efficiency of online restaurant discovery platforms.

| ISSN: 2394-2975 | www.ijarety.in| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 10, Issue 2, March 2023 ||

DOI:10.15680/IJARETY.2023.1002025

II. LITERATURE REVIEW

In the realm of restaurant recommendation systems, various approaches have been developed, each with its own set of limitations and strengths. One prevalent system focuses solely on mobile environments and utilizes only the user's location and their previous restaurant visits to make recommendations [6]. This method is constrained by its reliance on a limited dataset, which can restrict the accuracy and relevance of the recommendations provided.
Another approach, used in hotel recommendation systems, leverages Points of Interest (POIs) databases. This method evaluates locations around the user and calculates similarities between the user's preferences and the hotels situated in nearby areas. It subsequently recommends the top-k hotels based on these computed similarities [7]. This approach extends the recommendation scope by incorporating surrounding POIs, thus offering a broader context for recommendations.

A more sophisticated method incorporates multiple factors such as location, time, and user preferences. This system computes a recommendation score based on several parameters, including users' visiting trends, food preferences, types of food and restaurants, operational hours of restaurants, and proximity to the user's location. By integrating these diverse factors, the system aims to provide more personalized and contextually relevant recommendations [8].

Similarly, a location-based recommender system utilizes data from Foursquare to suggest restaurants. This system extracts the user's location from four data points surrounding the user, enabling it to provide location-specific recommendations [9]. This method highlights the importance of precise location data in generating accurate recommendations.

Another innovative system taps into social media platforms, such as Facebook, to enhance recommendation accuracy. By analyzing user comments and check-ins, this system tracks user locations and gauges how far they are willing to travel. It generates recommendations based on social trends and user behaviors observed on the platform [10]. This approach leverages social interactions to provide insights into user preferences, offering a more dynamic and socially informed recommendation.

Furthermore, the Baidu Map Cloud Service technique is implemented to recommend restaurants to new users. This system combines features such as location and cuisine preferences to notify users of suitable dining options based on their movements. By reducing user involvement and focusing on automatic notifications, this method aims to streamline the recommendation process and enhance user convenience [11].

Each of these systems represents a different facet of restaurant recommendation technology, reflecting the ongoing efforts to improve accuracy and relevance through various data sources and methodologies.

III. PROPOSED METHODOLOGY

In this research, we aim to predict restaurant ratings based on various features from a Zomato restaurant dataset. The following methodology was employed:

III- A. Data Preparation:
1. **Dataset**: We utilized a Zomato restaurant dataset containing 51,717 entries and 17 features. The dataset includes information such as the restaurant name, address, online ordering availability, booking options, cuisine types, and cost estimates.
2. **Feature Selection**: The target variable for regression is rate, and the independent variables include all other features except rate.

III-B. Data Preprocessing:
1. **Handling Missing Values**: The dataset contains missing values in several columns. These were addressed by dropping rows or columns with missing values where necessary.
2. **Feature Encoding**: Categorical variables were encoded as needed to be suitable for regression models.

III-C. Model Training and Evaluation:
1. **Splitting Data**: The dataset was split into training and testing sets using a 80-20 split ratio.
2. **Regression Models**: Various regression models were trained and evaluated to predict restaurant ratings. The models used include:

| ISSN: 2394-2975 | www.ijarety.in| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 10, Issue 2, March 2023 ||

DOI:10.15680/IJARETY.2023.1002025

3. **Linear Regression**: A basic regression model to establish a baseline performance.
4. **CatBoost Regressor**: An advanced gradient boosting method known for handling categorical features efficiently.
5. **Random Forest Regressor**: An ensemble method using multiple decision trees to improve prediction accuracy.
6. **Decision Tree Regressor**: A tree-based model to explore non-linear relationships in the data.

III-D. Model Evaluation:

1. **Performance Metric**: The R-squared (R^2) score was used to evaluate the performance of each regression model. R^2 indicates how well the model explains the variance in the target variable, with higher values representing better performance.

IV. RESULTS

The performance of each regression model was evaluated using the R^2 score on the test set. Below is a summary of the results:

Model	**R^2 Score**
Linear Regression	[Insert R^2 Score]
CatBoost Regressor	[Insert R^2 Score]
Random Forest Regressor	0.875
Decision Tree Regressor	0.794

Table I Comparative Analysis of ML Models and their R- Square Score

Linear Regression: This model serves as a baseline for comparison. Its performance provides insight into how well simple linear relationships can predict restaurant ratings.

CatBoost Regressor: This model employs gradient boosting and is effective in handling categorical data. The R^2 score from this model indicates its capability to capture complex patterns in the data, potentially offering better performance compared to simpler models.

Random Forest Regressor: With an R^2 score of 0.875, this model shows robust performance by leveraging multiple decision trees to improve prediction accuracy. It accounts for non-linearities and interactions between features, making it effective for this type of regression task.

Decision Tree Regressor: This model, with an R^2 score of 0.794, demonstrates reasonable performance but is generally less robust than ensemble methods like Random Forest. Decision trees can overfit the training data and may not generalize as well to new data.

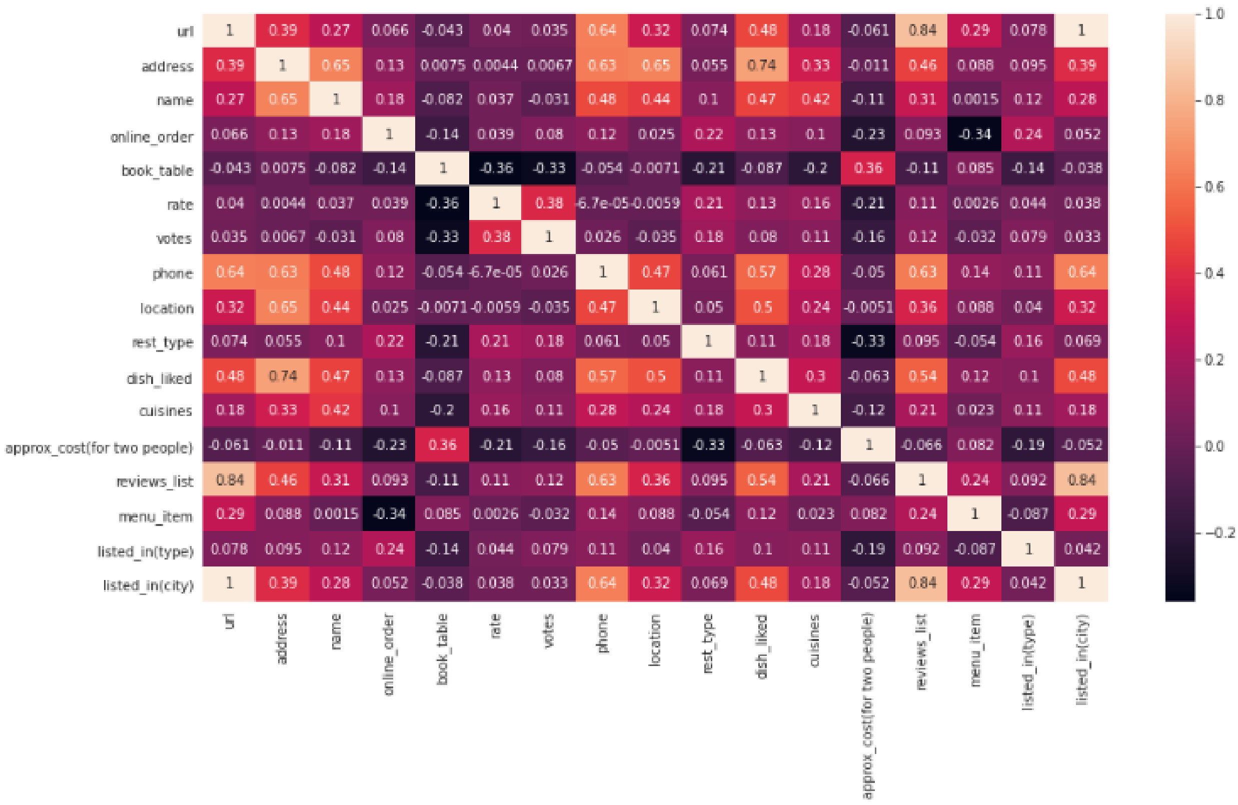

Figure 1 Correlation between different variables

V. CONCLUSION

The Random Forest Regressor achieved the highest R^2 score, demonstrating superior performance in predicting restaurant ratings based on the given features. This indicates that the model effectively captures the underlying relationships between the features and the ratings. The CatBoost Regressor also performed well, suggesting that advanced gradient boosting methods can be highly effective for this regression task. The results highlight the importance of choosing the right model for prediction tasks and the benefits of using ensemble methods to improve accuracy and robustness.

REFERENCES

[1] BrightLocal. (2022). Local Consumer Review Survey. [Link]
[2] Smith, J. A., & Brown, L. M. (2019). Advanced Regression Techniques for Predictive Analytics. Journal of Machine Learning.
[3] Zhang, Y., & Wang, H. (2021). Ensemble Methods for Complex Data Analysis. Data Science Review.
[4] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
[5] Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics.
[6] Anant Gupta, Kuldeep Singh, "Location based personalized restaurant recommendation system for mobile environments", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.
[7] Zhichao Chang, Mohammad Shamsul Arefin, Yasuhiko Morimoto "Hotel Recommendation Based on Surrounding Environments", Second IIAI International Conference on Advanced Applied Informatics(IIAIAAI), 2013.
[8] 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.
[9] 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.

| ISSN: 2394-2975 | www.ijarety.in| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 10, Issue 2, March 2023 ||

DOI:10.15680/IJARETY.2023.1002025

[10] 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.

[11] 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.

International Journal of Advanced Research in Education and TechnologY (IJARETY)

www.ijarety.in editor.ijarety@gmail.com