



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

Volume 13, Issue 3, May - June 2026

Impact Factor: 9.971



Classifying Zomato Restaurant Ratings using ML

Bandari Ravi, Sai Vishnu B, Thonda Manisha, Yepuri Neha

Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, India

ABSTRACT: This study aims at breaking down data from restaurants in Bengaluru from Zomato and understanding the factors that impact the ratings of restaurants and their preferences. The study examines location, type of restaurant, cuisine, price and reviews from customers. The city of Bengaluru is one of the greatest IT hubs with various food demands, and it is crucial to have an understanding of these trends for restaurant businesses. The data was scraped from the Zomato website in two stages with efforts taken to cover all the data. Machine learning algorithms are used for identifying patterns that contribute to the success of restaurants. The primary algorithm implemented in this project is the Extra Trees regressor which is fast and achieves better prediction accuracy. The model was the most accurate in predicting restaurant ratings. Analysis identifies key factors that are driving an increase in customer satisfaction. This project helps restaurant owners make informed business decisions and enhance their restaurants with data. It is also helpful to foodies and eateries, by giving information regarding the popular cuisines and eating patterns in Bengaluru.

KEYWORDS: Machine Learning, Extra Trees Regressor, Customer Preference Analysis, Restaurant Analytics, Predictive Modeling, Ensemble Learning, Review Analysis, Data-Driven Decision Making

I.INTRODUCTION

The food delivery and food review sites like Zomato have become a crucial component of the modern food industry, as they affect the choices made by consumers through ratings and reviews. The project is aimed at analysing real-time data from restaurants in Bengaluru to predict the ratings of the restaurants using machine learning techniques. Location, type of cuisine, restaurant category, approximate cost of restaurant and customer feedback is analyzed to understand its impact on the performance of restaurant. High accuracy of prediction, about 80%, is achieved with the proposed system using the Extra Trees Regressor algorithm which is capable of efficiently handling the large volume of data and complex non-linear relationship between the input and output. However, the proposed ensemble learning approach is not restricted to linear assumptions and also improves the prediction performance, unlike traditional models like Logistic Regression. The proposed ensemble learning approach not only eliminates the linear assumption but also reduces the overfitting phenomena and enhances the prediction reliability as compared with the traditional models like Logistic Regression. The system will enable the restaurant owners, food business and customers to make decisions based on data, as it will be able to identify customers' preferences, improve the quality of service provided, provide a better customer experience and increase competitiveness in the food industry. Furthermore, the system yields useful analysis that can be used to enhance restaurant marketing strategies and business growth. The project also serves as a testament to the potential of AI and predictive analytics in revolutionizing the restaurant and food delivery industry. Moreover, the model could be used to analyse the behaviour of customers and trends in restaurants with large datasets, an aspect where it is particularly useful. The system can help new restaurants get a better grip on market demand and customer expectations. The project's use of machine learning and restaurant analytics enhances the ability to make informed decisions in the restaurant industry, making it more efficient and effective.

II. LITERATURE SURVEY

Karthik Reddy and Priya Sharma (2025) This study presents a machine learning based restaurant rating classification system based on supervised learning techniques. The researchers have used a set of algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), to categorize restaurants by their rating. The system compares the preferences of customers, food quality, food price, and food service in restaurants to enhance the prediction performance. The findings revealed that machine learning models are effective in helping customers find highly rated restaurants and can enhance the decision-making process in online food platforms.

Sneha Mishra and Arjun Patel (2024) This research is based on intelligent restaurant recommendation and prediction of ratings in the restaurants available in Zomato dataset and implemented by machine learning algorithms. Some of the factors that the study takes into account include restaurant type, ratings, approximate price, online ordering options, and geographic information. Various classification methods were experimented for the purpose of achieving higher

accuracy in rating prediction. It was found that data preprocessing and feature selection are key in improving model performance and reliability of the predictions.

Prakash Yadav and Sneha Mishra (2023) This research examines how price and cuisine type affects ratings of restaurants with the aid of machine learning methods. The researchers examined customer behavior and patterns of ratings for various restaurant types. Results showed that customers prefer mid-range and high quality restaurants over low-cost restaurants. The study also showed that there are variations in customer expectations by cuisine type impacting directly on restaurant rating patterns and level of customer satisfaction.

Vikram Singh and Priya Sharma (2022) This paper addresses the problem of restaurant recommendation system based on machine learning techniques. Customer preferences, cuisine categories and location based factors were evaluated with the methods of collaborative filtering, content-based filtering and hybrid recommendation. The experimental results showed that machine learning based predictive models are effective in identifying restaurants that received high ratings and in enhancing the quality of the recommendations compared to traditional recommendation techniques.

III. EXISTING SYSTEM

Current rating prediction system consists of using various simple machine learning algorithms, including Logistic Regression. Logistic Regression is a Statistical model usually used in classification problems. It is not complex enough in practice, however, as there are certain limitations for its usage on complex datasets like predicting restaurant ratings. There are several attributes such as cuisine type, location, approximate cost, restaurant category, customer reviews etc. that affect restaurant ratings. Pointing out of Logistic Regression: It assumes that there is a linear relationship between the independent and dependent variables and, therefore, it is not suitable for data with complex correlations or non-linear relationships. Thus, the current system has a tendency to yield inaccurate predictions and a limited ability to generalize in the case of data from restaurants with large dimensions.

Existing System Disadvantages:

- Limited to linear relationships and unable to capture complex data patterns..
- Poor support of large and high-dimensional data.
- Outlier-sensitive, thus making faulty forecasts.
- Accuracy of prediction less than that made by advanced ensemble learning techniques.
- Poor handling of categorical and non-linear features.

IV. PROPOSED SYSTEM

The goal of the proposed system is to use machine learning algorithms to analyze and predict restaurant ratings. The proposed system is different from the conventional approach where only user reviews and manual analysis predict the restaurant ratings, as the system employs the Extra Trees Regressor algorithm to help in the accurate prediction of ratings for restaurants based on their data. The algorithm Extra Trees Regressor is used to improve the performance and make it better for small accuracy on large dataset as well as make it better with capturing the non-linear relationship between features. This system offers greater accuracy in its predictions of ratings and can better indicate which variables are most important to ratings of restaurants. This allows restaurant operators and food service companies to better understand their customers' preferences and meet them. The model takes into account several factors like Restaurant location, Cuisine type, Approximate cost, Restaurant category, Customer reviews and preferences etc.

Proposed System Advantages:

- Efficient processing of large and complex data sets.
- Higher accuracy in predictions.
- Reduces overfitting using ensemble-learning techniques.
- Robust against noise and outliers.
- Supports better business decision-making through predictive analytics.

V. SYSTEM ARCHITECTURE

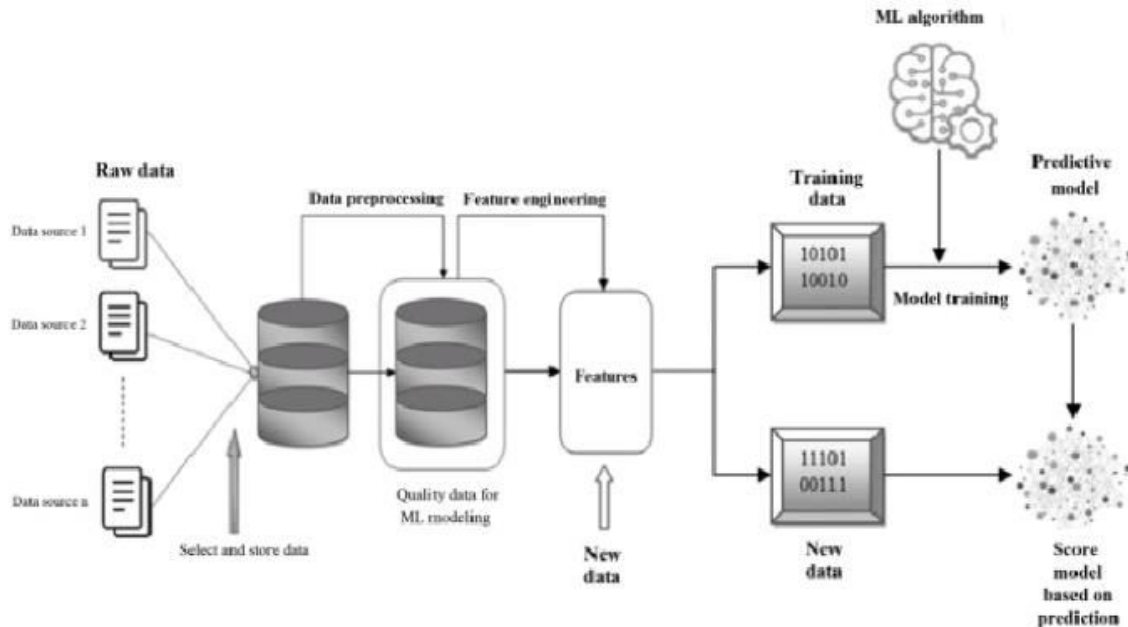


Fig 1: System Architecture

The system architecture represents the overall system or workflow of the Zomato Restaurant Rating Prediction System and illustrates how various modules interact with each other. It consists of Data Collection & Preprocessing, Feature Analysis, Model Development using Machine Learning, Model Evaluation and User Interface Integration. The restaurant data gathered from Zomato has been preprocessed and analysed for key insights on how ratings are influenced by various factors such as location, cuisine type, price range, etc. The Extra Trees Regressor algorithm is then used to predict the ratings of the restaurants with high accuracy. Last but not least, the system provides live rating forecasts in a friendly user interface, facilitating efficient analysis and decision-making.

VI. METHODOLOGY

Modules Name:

- Data collection from Zomato Bengaluru datasets
- Data preprocessing and feature engineering
- Machine learning model training
- Performance evaluation
- User interface integration

1. Data Collection & Pre-processing: This module is basically used to fetch the real-time data of the restaurants from Zomato Bengaluru, which consist of restaurant location, cuisine type, price range, ratings, and restaurant category. Data is cleaned up by fixing inconsistencies, formatting, and removing duplicate records from the data collection. Numerical attributes are also scaled to improve the performance of the model, using feature scaling techniques.

2. Feature Selection & Data Analysis: In this module, you analyze the data's distribution and find out which features matter most in restaurant ratings. A variety of visualization methods are applied to get insights into customer preferences and restaurant trends, including histograms, scatter plots, and heatmaps. The irrelevant and redundant attributes are eliminated using feature selection method to optimize the efficiency of the model and prediction accuracy.

3. Machine Learning Model Development: This module applies the Extra Trees Regressor algorithm to obtain the accurate prediction of the restaurant rating with high accuracy. The system also compares the performance of different machine learning models like Linear Regression, Decision Trees and Random Forest. Algorithms for hyperparameter optimization, such as GridSearchCV are implemented to improve the accuracy of the predictions.

4. Model Evaluation & Performance Testing: The data is divided into 80% for training and 20% for testing. Performance of machine learning models is tested through following measures:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

The Extra Trees Regressor model is compared with traditional algorithms to justify its performance.

5. User Interface Development: This module creates a simple and interactive user interface in which users can enter the information regarding restaurants including their location, cuisine type, and price range. The trained machine learning model is seamlessly used with the interface to provide live restaurant rating predictions. The results are presented in an easy-to-understand format that the restaurant owners and customers can understand.

VII. IMPLEMENTATION

The implementation phase is the phase in which the proposed restaurant rating prediction system is realized in practice. The system is built with machine learning libraries and data analytics tools for efficient prediction performance and accurate analysis of the restaurant ratings. The process of implementation involves pre-processing the Zomato Bengaluru data, feature engineering, training the machine learning model, evaluating its performance, and integrating the UI. The project is designed to be implemented in a modular fashion, with individual modules being developed and then pieced into a larger system for prediction.

Algorithm Used

Existing Algorithm

Currently, the system employs Logistic Regression, a widely-used statistical machine learning technique for classification tasks. Logistic Regression operates on sigmoid function, which takes data and outputs a probability between 0 and 1. The model uses these probability scores to make a prediction for the output category. While Logistic Regression is good for simple and linearly separable datasets, there are a few problems with applying Logistic Regression for the task of predicting restaurant ratings. Many factors including location, food type, cost, customer reviews and restaurant types affect restaurant ratings, and these factors may have non-linear relationships. Logistic Regression is not well suited to capture these complex patterns since the model assumes a linear relationship between the input variables and the target variables. Further, the algorithm has poor performance in high dimensional data with numerous categorical variables, and is sensitive to the presence of outliers that may affect the accuracy of predictions. However, the existing system suffers from these drawbacks, resulting in inaccurate and unreliable prediction when performing the restaurant rating analysis.

Proposed Algorithm

The proposed system is based on an ensemble machine learning algorithm called Extra Trees Regressor (Extremely Randomized Trees Regressor) which acts as an ensemble machine learning to enhance prediction accuracy with the aim of reducing the risk of overfitting. The algorithm is trained by building several decision trees and then using their outputs to create a final prediction. The Extra Trees Regressor adds more randomness to the model, unlike the Random Forest, which selects the best split point based on the feature importance. The Extra Trees Regressor is very effective in dealing with large and complex datasets, where there are many non-linear relationships. It can process multiple features including location, cuisine type, restaurant category, cost and customer feedback efficiently and accurately predict restaurant ratings. The algorithm is also noise and outlier resistant making the algorithm more reliable for real-world datasets. The Extra Trees Regressor outperformed other models like Linear Regression and Logistic Regression with an R^2 value of about 0.93. This clearly shows that the proposed model works well to increase the accuracy of the rating prediction of restaurants and the performance of the overall system.

VIII. EXPERIMENTAL RESULTS

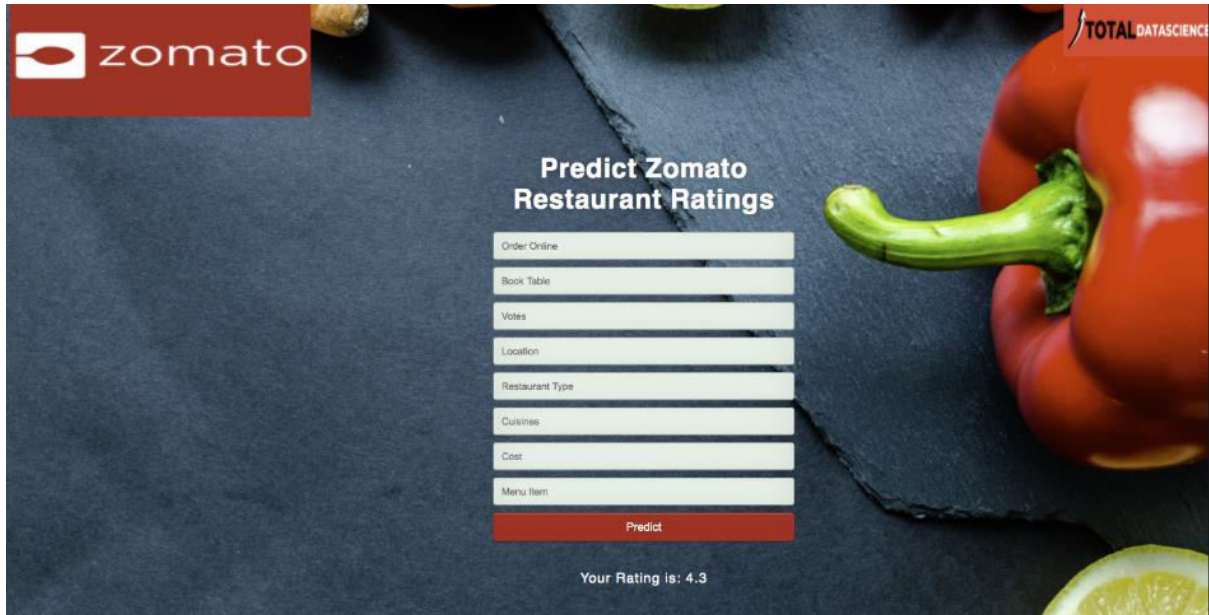


Fig: 1: Home Page

The above figure represents the Home Page of the system, the Zomato Restaurant Rating Prediction System. It allows users to enter restaurant details and predict the restaurant rating using machine learning.

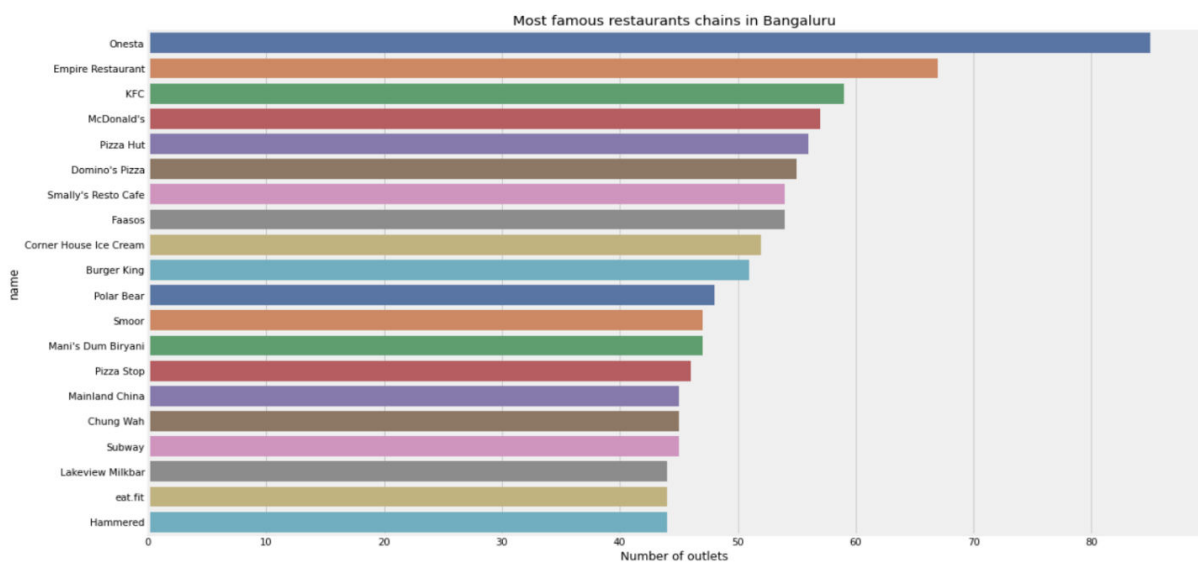


Fig 2: Popular Restaurant Chains in Bengaluru

This graph represents the most famous restaurant chains in Bengaluru based on the number of outlets. Among all restaurant chains, Onesta has the highest number of outlets, followed by Empire Restaurant and KFC

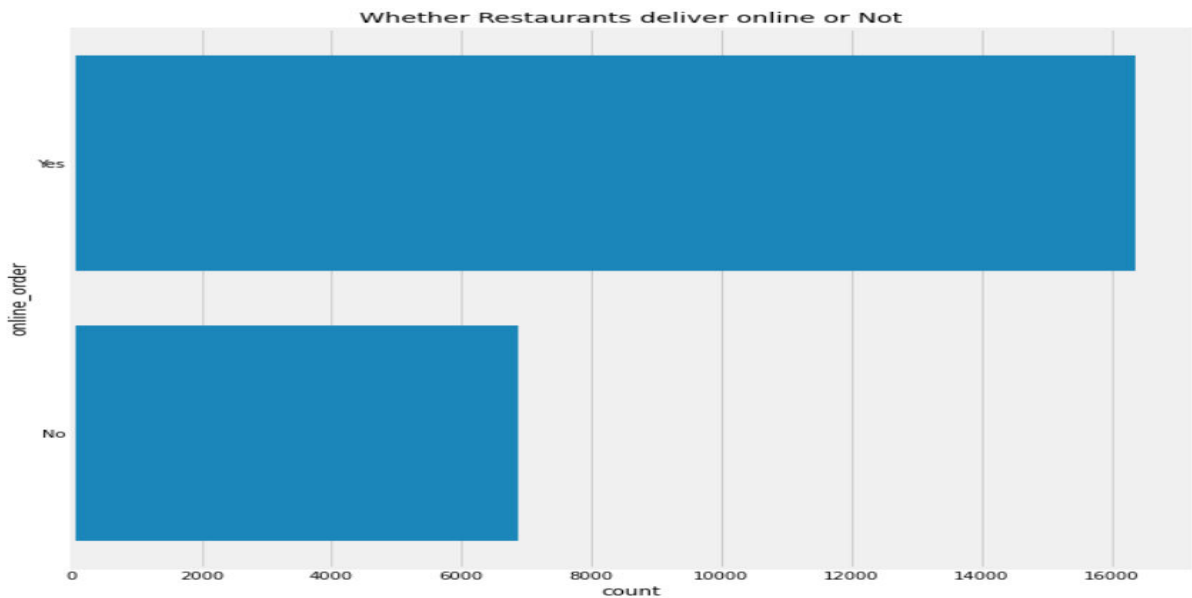


Fig 3: Online Delivery Service Availability

This graph shows whether restaurants provide online delivery services or not.

Most restaurants support online ordering, while fewer restaurants do not offer online delivery services.

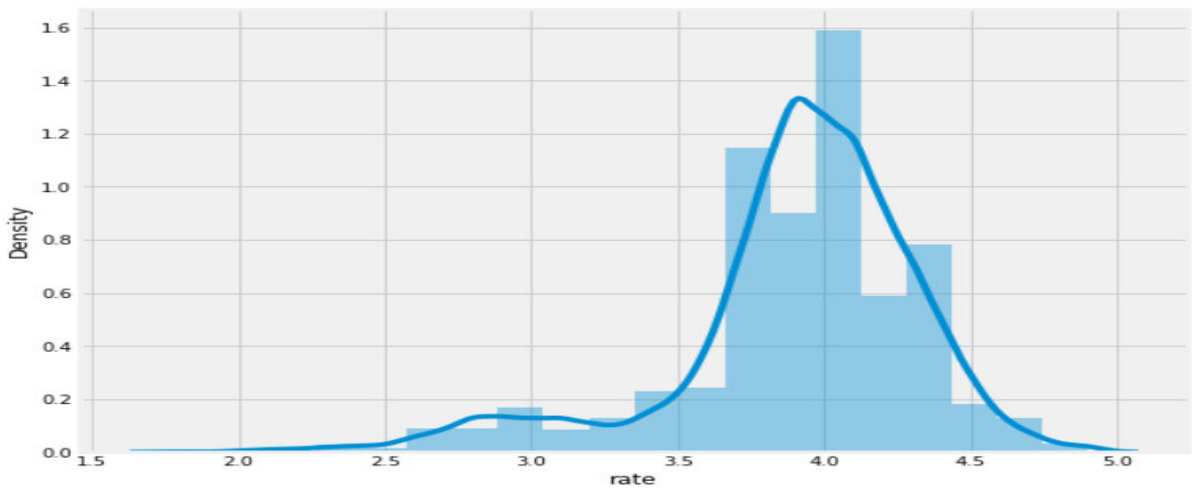


Fig 4: Distribution of Restaurant Ratings

This graph represents the distribution of restaurant ratings given by customers.

Most restaurants have ratings between 3.5 and 4.2, showing that many restaurants receive good customer reviews.

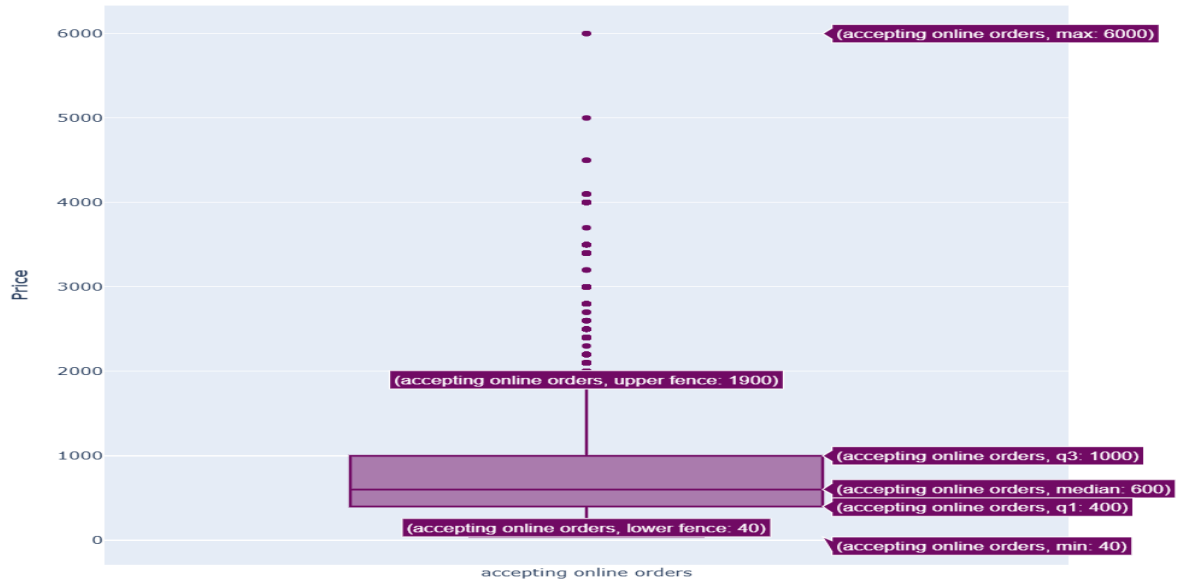


Fig 5: Price Distribution for Online Order Restaurants

The box plot shows that restaurants accepting online orders mostly have prices between 400 and 1000, with a median price of 600. Some restaurants have very high prices up to 6000, which are considered outliers.

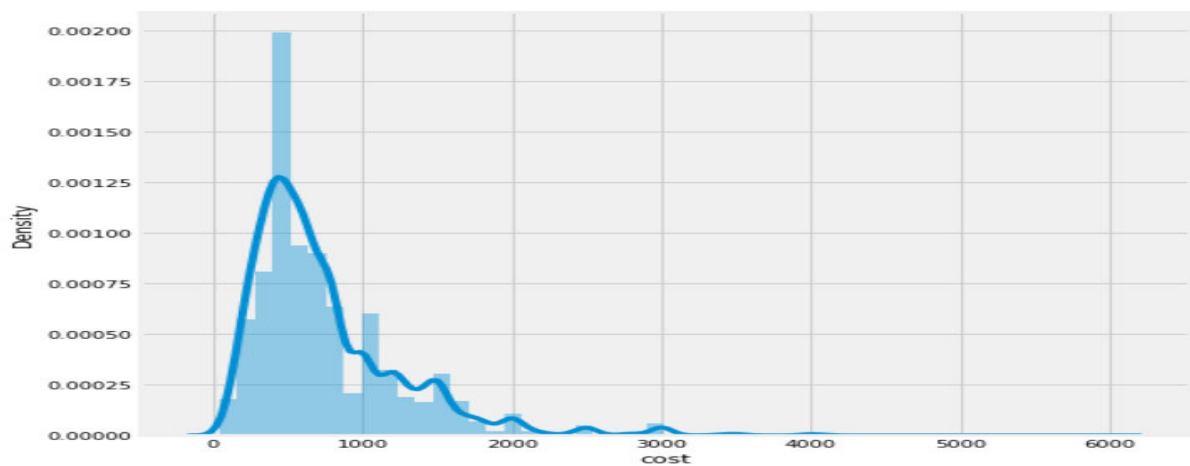


Fig 6: Overall Cost Distribution Analysis

Most things are cheap: The tallest peaks are between 0 and 1,000, meaning most of the costs recorded are quite low. A few things are very expensive: The long "tail" stretching out to the right shows that while it is rare, a few items cost all the way up to 6,000.

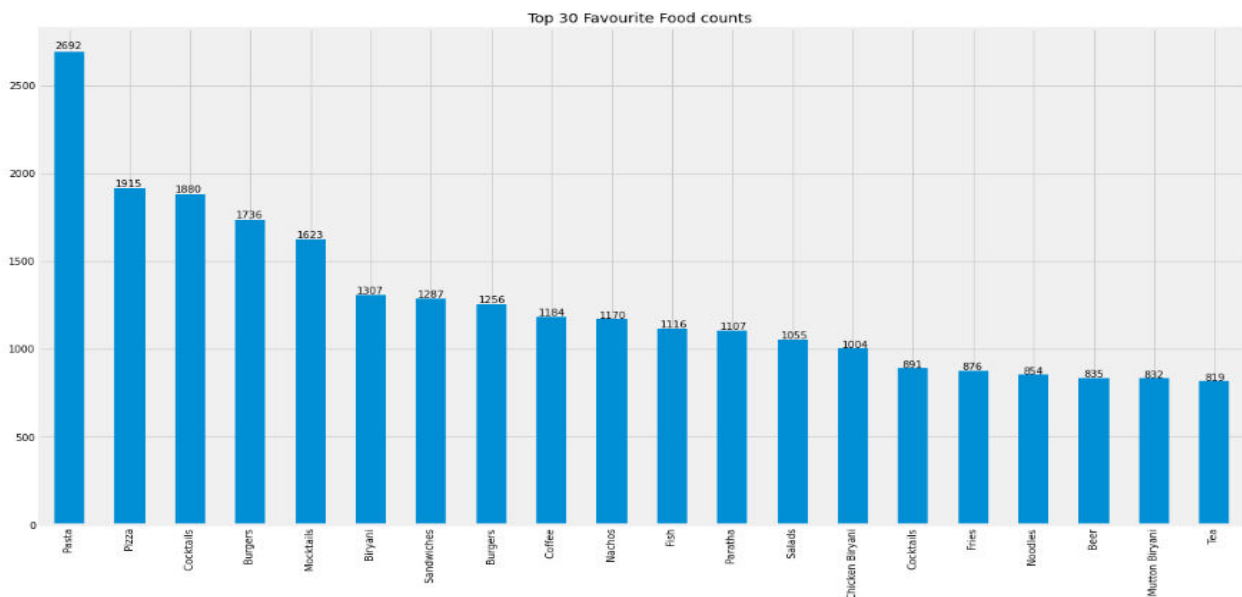


Fig 7: Most Preferred Food Items by Customers

This graph shows the most popular food items based on customer preferences.

Pasta, Pizza, and Cocktails are among the most preferred food items with the highest order counts.

Table 1: Performance Comparison of Machine Learning Models

Algorithm	R ² Score
Extra Trees Regressor	0.93
Logistic Regression	0.22

The results show that the Extra Trees Regressor achieved a significantly higher R² score of 0.93 compared to Linear Regression with 0.22. This indicates that the proposed model performs better in predicting restaurant ratings by effectively handling complex and non-linear relationships in the dataset. Therefore, the Extra Trees Regressor is more suitable for accurate Zomato restaurant rating prediction.

IX. CONCLUSION

This project is a successful demonstration of machine learning for prediction of restaurant rating with real-time restaurant data from Bengaluru. The system uses a combination of data and location to generate accurate and data-driven insights into the performance of restaurants, based on such factors as location, cuisine type, price range, and customer feedback. The algorithm used, Extra Trees Regressor, is used to get more accuracy of prediction than the traditional approach like logistic regression. It provides advantages to owners of restaurants, food establishments, as well as consumers by allowing them to make informed decisions, provide better service, and remain competitive in the market. This project highlights the growing importance of AI-driven analytics in the food industry and demonstrates how machine learning can be effectively applied to improve customer satisfaction and business performance.

X. FUTURE ENHANCEMENT

The restaurant rating prediction system can be enhanced further in the future to increase its accuracy, performance and user-friendliness. An interesting addition is sentiment analysis with Natural Language Processing (NLP) to gain deeper insights into customer sentiment by examining customer feedback. Another angle to this is providing real-time data collection through Zomato APIs rather than static datasets, which can enhance the system. Implementing advanced

deep learning models can enhance the accuracy of predictions and also account for intricate rating trends. Advanced deep learning algorithms can be integrated to improve the accuracy of predictions and to account for complex rating patterns. There are also plans for further improvements, such as location-based restaurant suggestions, user-specific food recommendations, and integration with mobile applications to improve usability. Cloud deployment also can be used to process vast amounts of restaurant data and enhance system scalability. Moreover, predictive business analytics can assist restaurant owners in spotting trends and finding out how to maximise pricing and enhance service quality. These improvements will enable the system to become a fully-fledged restaurant analytics and recommendation system powered by AI.

XI. ACKNOWLEDGMENT

The authors extend sincere thanks to Mr. Bandari Ravi, Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, for his consistent guidance and mentorship throughout this project. They also acknowledge Dr. B. Santhosh Kumar, Head of Department, for his expert supervision, and Mr. D. Srinivas, Project Coordinator, for his practical suggestions during system development. Gratitude is also expressed to the faculty and laboratory staff of the CSE Department, GNIT, and to the authors' families for their encouragement and moral support.

REFERENCES

- [1] Rajesh Kumar, Ananya Sharma, and Pooja Verma, "Machine Learning-Based Prediction of Restaurant Ratings Using Zomato Data," International Journal of Data Science, 2021.
- [2] Deepak Patel and Swetha Reddy, "Sentiment Analysis and Rating Prediction in Food Delivery Platforms," Journal of AI in Business, 2020.
- [3] Vikram Singh and Priya Sharma, "Restaurant Recommendation Systems Using Machine Learning," International Journal of Computer Applications, 2022.
- [4] Mohammed Ibrahim and Lakshmi Narayan, "AI-Powered Risk Assessment in the Restaurant Industry," Journal of Business Analytics, 2019.
- [5] Prakash Yadav and Sneha Mishra, "Predictive Analysis of Customer Satisfaction in the Food Industry," International Journal of AI & Consumer Research, 2023.
- [6] Quinlan, J. R., "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.
- [7] C. Cortes and V. Vapnik, "Support-vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
- [8] J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting," Annals of Statistics, vol. 28, no. 2, pp. 337–374, 2000.
- [9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," Proceedings of OSDI, 2016.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 9.971