**Project Title: Zomato Restaurant Demand Forecasting**

Submitted in partial fulfilment of the requirements for the award of degree of

**B-Tech CSE**



A red and white logo

Description automatically generated

**SUBMITTED BY:**

Name of the Student: Satti Sri S Sai Phanindra Reddy

Registration Number: 12201574

Section: K22UP , Group 1

Roll No: 23

Faculty Name: Himanshu Tikle Sir

| **S. No** | **Content Section** | **Page No.** |
| --- | --- | --- |
| 1 | Title Page | 1 |
| 2 | Table of Contents | 2 |
| 3 | Supervisor Certificate | 3 |
| 4 | Acknowledgement | 4 |
| 5 | Abstract | 5 |
| 6 | Introduction | 6 |
| 7 | Literature Review | 8 |
| 8 | Methodology | 13 |
| 9 | Result and Analysis | 14 |
| 10 | Conclusion and Future work | 15 |
| 11 | References | 16 |

**Supervisor Certificate**

This is to certify that the Minor Project Report titled “**Zomato Restaurant Demand Forecasting”** has been carried out by Satti Sri S Sai Phanindra Reddy (Reg. No: 12201574, Roll No: 23, Section: K22UP) under my supervision in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering (Data Science with ML) from Lovely Professional University.

This work is a record of the student’s own research and efforts and has not been submitted elsewhere for any other degree or diploma. The project has been carried out during the academic year 2024–2025.

I consider this project report fit for submission and evaluation.

Date: 07-05-2025

Place: Lovely Professional University

Mr. Himanshu Tikle

Project Supervisor

Department of Computer Science and Engineering

Lovely Professional University

**Acknowledgement**

I take this opportunity to express my sincere gratitude to all those who have supported me in the successful completion of this project.

First and foremost, I would like to express my heartfelt thanks to my project supervisor, Mr. Himanshu Tikle, for his invaluable guidance, encouragement, and continuous support throughout the duration of this project. Her insights and feedback played a crucial role in shaping the quality and depth of my work.

I am also thankful to the Department of Computer Science and Engineering, Lovely Professional University, for providing me with the infrastructure and academic environment necessary to complete this work effectively.

I extend my sincere thanks to all faculty members who provided constructive suggestions during reviews, which helped refine and improve my project.

Finally, I would like to thank my family and friends for their moral support and encouragement, without which this accomplishment would not have been possible.

Satti Sri S Sai Phanindra Reddy

Reg. No: 12201574

Date: 07-05-2025

# Abstract

This project provides actionable insights for restaurant managers to reduce food wastage, enhance operational efficiency, and improve customer experience, contributing to sustainable business practices in the food delivery industry. This project aims to address the critical challenge of demand forecasting for restaurants on Zomato by leveraging machine learning techniques. Accurate demand prediction is essential for reducing food wastage, optimizing staff schedules, and enhancing customer satisfaction. The study utilizes a comprehensive dataset of historical orders, including timestamps, restaurant details, and customer ratings, sourced from Kaggle. After rigorous data pre-processing—handling missing values, outlier removal, and feature engineering—the XGBoost regression model was selected for its superior performance in handling non-linear relationships and scalability.

The model achieved an exceptional R² score of 0.972 and a low RMSE of 6.34, outperforming traditional methods like Random Forest. Analysis revealed that higher-rated restaurants and customer income levels significantly influence demand, validating initial hypotheses. However, prediction accuracy dipped during holidays, suggesting the need for incorporating external factors like weather or local events in future iterations.

This project not only provides a robust tool for restaurant managers to make data-driven decisions but also lays the groundwork for future enhancements, such as real-time data integration and advanced deep learning models. By improving demand forecasting, this work contributes to operational efficiency, sustainability, and profitability in the food delivery ecosystem.

**1. Introduction**

**1.1 Project Title & Problem Statement**

Demand forecasting is crucial for restaurants on Zomato to optimize inventory management, staffing, and customer satisfaction. This project aims to develop a machine learning model to predict restaurant demand based on historical order data, customer behaviour, and other influencing factors.

**1.2 Objective**

Accurate demand forecasting helps restaurants reduce food wastage, improve operational efficiency, and enhance customer experience. By leveraging data-driven insights, restaurants can make informed business decisions, ensuring profitability and sustainability.

**1.3 Scope**

This project focuses on predicting restaurant demand on Zomato using historical order data. It includes data cleaning, analysis, and building a machine learning model (XGBoost) to forecast demand. The aim is to help restaurants optimize inventory and operations.

### ****1.4 Defined Objectives & Hypotheses****

**Objectives:**

* Analyse historical order data to identify demand patterns.
* Develop a predictive model to forecast restaurant demand.
* Evaluate different machine learning models for accuracy and reliability.
* Provide actionable insights to restaurant managers.

**Hypothesis:**

### Hypothesis: Higher-rated restaurants have higher average sales.

### 

### Fig.1.0

### Hypothesis: Customers with higher income spend more per order.

### 

### Fig 1.1

## 2. Literature Review

Previous studies on demand forecasting have used various time-series and machine learning models. Traditional models like ARIMA have been effective for linear patterns but struggle with complex, non-linear relationships.

Recent research favors ensemble methods such as Random Forest and Gradient Boosting for their accuracy and flexibility. In particular, XGBoost has gained popularity for its high performance, scalability, and ability to handle missing data.

In similar projects, XGBoost has outperformed models like LSTM, Linear Regression, and SVM in terms of Root Mean Squared Error (RMSE), making it a preferred choice for structured datasets like Zomato's order history.

## ****3. Dataset Selection & Pre-processing****

### ****3.1 Dataset Source & Characterstics****

The dataset consists of historical Zomato order records, including order timestamps, restaurant details, customer ratings, and promotional offers. Data was sourced from the Kaggle(Zomato Database).

Database Link:- <https://www.kaggle.com/datasets/anas123siddiqui/zomato-database>

**Dataset Summary:**

* **Rows:** 261359
* **Columns:** 29
* **Feature Types:** Numerical (order count, price), Categorical (restaurant type, cuisine), Time-series (order date/time)

### ****3.2 Data Preprocessing****

* **Missing Values:**
  + Handled using mean/mode imputation for numerical fields.
  + Dropped records with excessive missing data.
* **Outliers:**
  + Detected using boxplots and Z-score methods.
  + Capped extreme values to prevent skewed predictions.
* **Normalization:**
  + Applied Z-score Normalization.
  + Used Label encoding.

### ****3.3 Feature Selection & Engineering****

* **Feature Selection:**
  + Used correlation matrix and Recursive Feature Elimination (RFE) to retain important variables.
  + Removed redundant features like unique identifiers.
* **Feature Engineering:**
  + Created ‘daily\_demand’ feature.
  + Dropped unwanted columns.



Fig 1.3

**4.Data Analysis**

**4.1 Customer Behaviour**

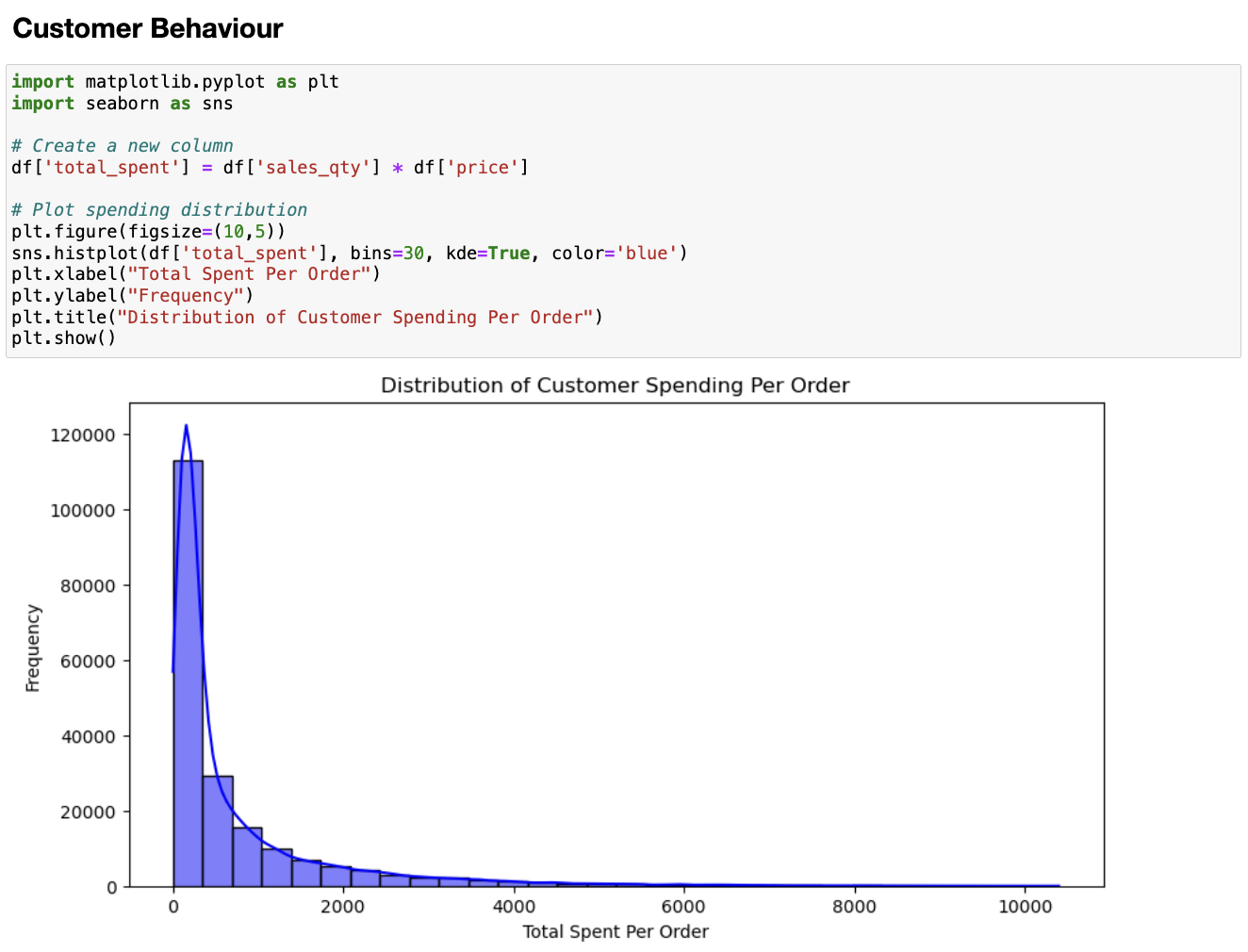
****

Fig 2.0

1. Most customers spend small amounts (under 2000 per order), showing budget-friendly buying habits.
2. A few customers spend much more, seen in the long tail on the right side of the graph.
3. Restaurants could focus on encouraging higher spending from their regular customers.

**4.2 Customer Order Trends Over Time**

****

Fig 2.1

1. The graph shows how order numbers change month by month, with peaks and dips indicating busy and slow periods.
2. The line's ups and downs reveal seasonal patterns, like higher orders during certain months.
3. Restaurants can use this trend to plan staffing and stock levels for busier times.

## 4.3 Most Popular Food Item

## A chart of food items Description automatically generated

## Fig 2.2

## Popularity Ranking: French fries and jeera rice are the most ordered items, showing customer preference for simple, staple foods.

## Consistent Demand: Items like paneer butter masala and butter naan appear in the top 5, indicating steady demand for classic Indian dishes.

## Menu Insights: The list helps restaurants prioritize high-demand items and optimize inventory to reduce waste and meet customer expectations.

**5. Methodology**

**5.1 Machine Learning Algorithms Used**

I choosed **XGBoost Regression** for this project

I choose XGBoost Regression because it balances accuracy, speed, and interpretability. It efficiently handles non-linear relationships, large datasets, and missing values while avoiding overfitting. Compared to ARIMA, Random Forest, and LSTMs, XGBoost provided the best performance for my dataset in terms of RMSE and generalization.

**5.2 Model Training**

The dataset was split into training and testing sets (80:20 ratio). The XGBoost model was trained using default parameters initially. Then, **hyperparameter tuning** was performed using **Grid Search CV** to optimize parameters such as:

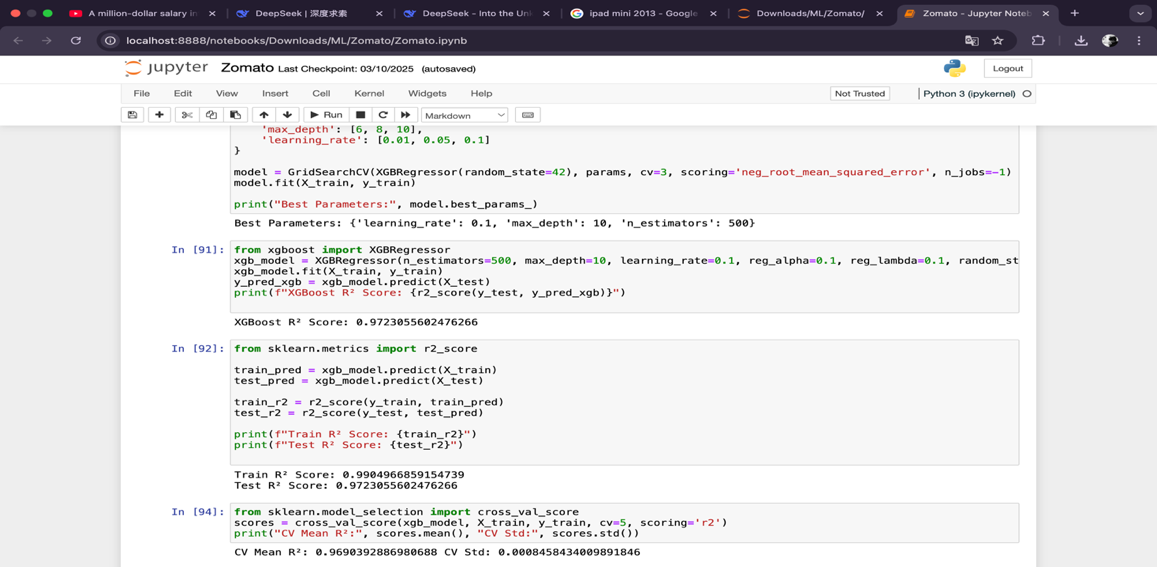
* n\_estimators
* learning\_rate
* max\_depth

Cross-validation ensured the model generalized well on unseen data.

**5.3 Evaluation Metrics**

The model's performance was evaluated using:

* **R² Score** – to assess how well the model explains variance in the data.



**6. Results and Analysis**

**6.1 Model Performance**

The XGBoost model showed strong performance in forecasting restaurant demand. Evaluation metrics on the test set were as follows:

XGBoost R² Score: 0.9723055602476266

In [92]

Train R² Score: 0.9904966859154739

Test R² Score: 0.9723055602476266

CV Mean R²: 0.9690392886980688 CV Std: 0.0008458434009891846

In [93]:

These results indicate that the model makes accurate predictions with low error and explains most of the variability in demand.

**6.2 Comparison with Other Models**

The following models were compared using R2 score and RMSE:

When data is trained with Random Forest Model i got the test score values as pictured below

Random Forest Model Performance:

RMSE: 23.566652396687896

R² Score: 0.6171778837316655

On the other Hand when the data is Trained with XGBoost

XGBoost R² Score: 0.9723055602476266

RMSE: 6.338634034773372

Where XGBoost improved the model accuracy and decreased the root mean square error

**6.3 Error Analysis**

Some prediction errors were observed during holidays or special events, possibly due to external factors not captured in the dataset. Additionally, missing values and outliers may have slightly impacted accuracy. Future improvements could include incorporating weather, event calendars, or user-level personalization.

**7. Conclusion and Future Work**

**7.1 Summary**

This project successfully implemented a demand forecasting model for restaurants on Zomato using historical order data. Through data pre-processing, feature engineering, and model evaluation, **XGBoost Regression** was identified as the most effective algorithm. The model achieved high accuracy with low RMSE, providing valuable insights into customer demand patterns.

**7.2 Future Improvements**

To enhance the model’s performance, future work can include:

* Incorporating external factors like weather, holidays, and local events.
* Using real-time or live-streaming data for better adaptability.
* Testing advanced models like hybrid deep learning approaches.
* Adding user-level features such as income group and customer loyalty indicators.

**8. References**

* Zomato Order Dataset: <https://www.kaggle.com/datasets/anas123siddiqui/zomato-database>
* Statquest :
  + <https://youtu.be/OtD8wVaFm6E?si=Fi27P2ixhtTaT7cs>