## Swiggy Restaurant Analysis: Machine Learning Implementation Documentation

## 1. Introduction

The purpose of this analysis is to build a predictive model using machine learning to gain insights from Swiggy’s restaurant dataset. The primary focus is on predicting restaurant ratings based on various features such as delivery time, price, city, and cuisine type. The XGBoost model was chosen for its efficiency and high performance in classification and regression tasks.

## 2. Data Preparation

Before building the model, the dataset underwent the following preparation steps:

* **Data Cleaning**: Missing values were handled by either filling them with appropriate statistics (mean/median) or dropping incomplete rows.
* **Feature Encoding**: Categorical variables such as ‘City’, ‘Food Type’, and ‘Area’ were encoded using Label Encoding to convert them into numerical values.
* **Scaling**: Features like price and delivery time were standardized using StandardScaler to ensure that they were on the same scale, which is critical for many machine learning models.

.

## 3. Model Selection: XGBoost Classifier

The XGBoost model was selected due to its advantages:

* High efficiency and performance for structured/tabular data.
* Built-in handling for missing values and feature importance scores.
* Robust support for hyperparameter tuning, making it an ideal choice for optimizing model performance.

## 4. Model Implementation

**Splitting the Data**: The dataset was divided into training (80%) and testing (20%) sets using train\_test\_split from scikit-learn to evaluate the model performance effectively.

**Training the XGBoost Model**: The features selected included ‘Price’, ‘City’, ‘Area’, ‘Delivery Time’, and ‘Food Type’.

## 5. Hyperparameter Tuning

To optimize the model’s performance, hyperparameter tuning was conducted using RandomizedSearchCV and GridSearchCV:

* **RandomizedSearchCV**: This method was used initially to quickly search a wide range of hyperparameter values.
* **GridSearchCV**: A finer search around the best parameters identified by RandomizedSearchCV was performed for precise optimization.

## 6. Model Evaluation

The model’s performance was evaluated using several metrics:

* **Accuracy Score**: The model achieved an accuracy of 82%, indicating the proportion of correct predictions.
* **F1 Score**: The weighted F1 score was 0.79, accounting for the imbalance in the classes.
* **Classification Report**: Detailed metrics such as precision, recall, and F1 score for each class were generated to understand model performance across different classes.
* **Confusion Matrix**: A confusion matrix was plotted to visualize the model’s performance.

## 7. Insights and Interpretation

* The model showed strong performance in predicting ratings, especially in the ‘Excellent’ and ‘Good’ categories.
* Areas with high delivery times were often correlated with lower ratings, which matched the model’s predictions.
* The cuisine type and price also influenced ratings, with premium-priced restaurants generally receiving better ratings.6. Model Evaluation

## 8. Recommendations

Based on the model’s predictions and insights:

1. Swiggy should focus on optimizing delivery times in areas where delays are common to improve restaurant ratings and customer satisfaction.
2. Expanding partnerships with restaurants offering popular cuisines (e.g., North Indian) in high-density areas can maximize the positive impact.
3. Targeted marketing campaigns for restaurants with lower ratings and higher delivery times can help address customer complaints proactively.

## 9. Conclusion

The machine learning implementation for Swiggy’s restaurant dataset successfully demonstrated the model's ability to predict ratings based on multiple features like delivery time, price, and cuisine type. Hyperparameter tuning ensured optimal performance, and evaluation metrics confirmed the model’s effectiveness. These insights provide a strategic framework for Swiggy to enhance customer satisfaction and optimize operations.