**Restaurant Rating Prediction Case Study Using Data Mining in Python**

In this case study, we will predict the future rating of a restaurant on Zomato using data mining techniques. The goal is to help Zomato decide whether to include or remove a restaurant from their app based on its predicted performance.

**Steps to Solve the Problem**

1. **Problem Understanding**
   * Zomato wants to predict the future rating of a restaurant based on historical data.
   * The prediction will help them decide whether to onboard or remove a restaurant from their platform.
2. **Data Collection**
   * Gather data from Zomato's database or publicly available datasets.
   * Features may include:
     + Restaurant ID
     + Location (City, Country)
     + Cuisine Type
     + Average Cost for Two
     + Has Online Delivery (Yes/No)
     + Table Booking (Yes/No)
     + Votes (Number of Reviews)
     + Aggregate Rating (Target Variable)
     + Other relevant features.
3. **Data Preprocessing**
   * Handle missing values.
   * Encode categorical variables (e.g., Cuisine Type, Location).
   * Normalize/scale numerical features.
   * Remove outliers.
4. **Exploratory Data Analysis (EDA)**
   * Analyze the distribution of ratings.
   * Identify correlations between features and ratings.
   * Visualize trends (e.g., ratings vs. cost, ratings vs. location).
5. **Feature Engineering**
   * Create new features if necessary (e.g., "Popularity" based on votes).
   * Select relevant features for modeling.
6. **Model Selection**
   * Use regression or classification algorithms to predict ratings.
   * Algorithms to consider:
     + Linear Regression
     + Decision Trees
     + Random Forest
     + Gradient Boosting (XGBoost, LightGBM)
     + Neural Networks (if data is large).
7. **Model Evaluation**
   * Use metrics like RMSE, MAE, or R² for regression.
   * For classification (e.g., predicting high/low rating), use accuracy, precision, recall, or F1-score.
8. **Deployment**
   * Deploy the model as an API or integrate it into Zomato's system.
   * Use the model to predict ratings for new or existing restaurants.