Task 1 Report – Predict Restaurant Ratings

Internship: **Machine Learning Internship – Cognifyz Technologies**
Intern Name: **Aditya Raj**

The goal of this task is to predict restaurant ratings based on various features such as cost, location, cuisines, votes, and services offered. This involved data preprocessing, building regression models, evaluating them, and analyzing feature importance.

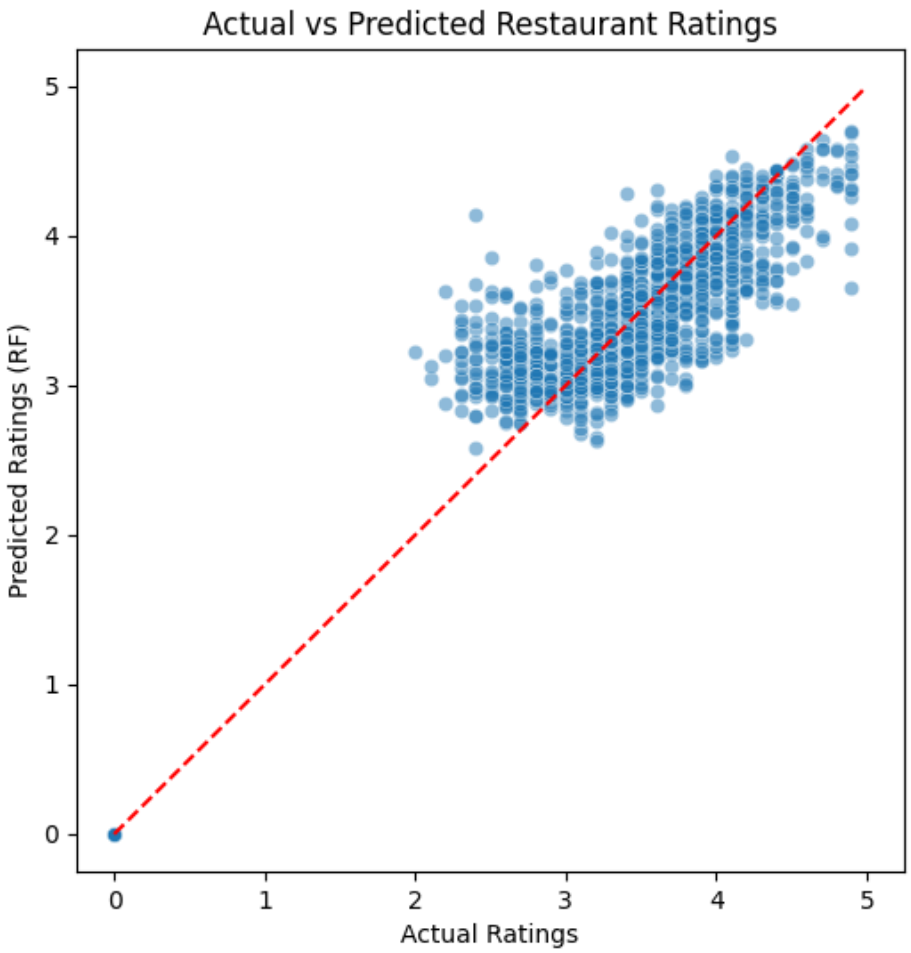
The dataset contained restaurant details such as name, city, cuisines, cost, votes, and ratings. Target variable: **Aggregate rating**.

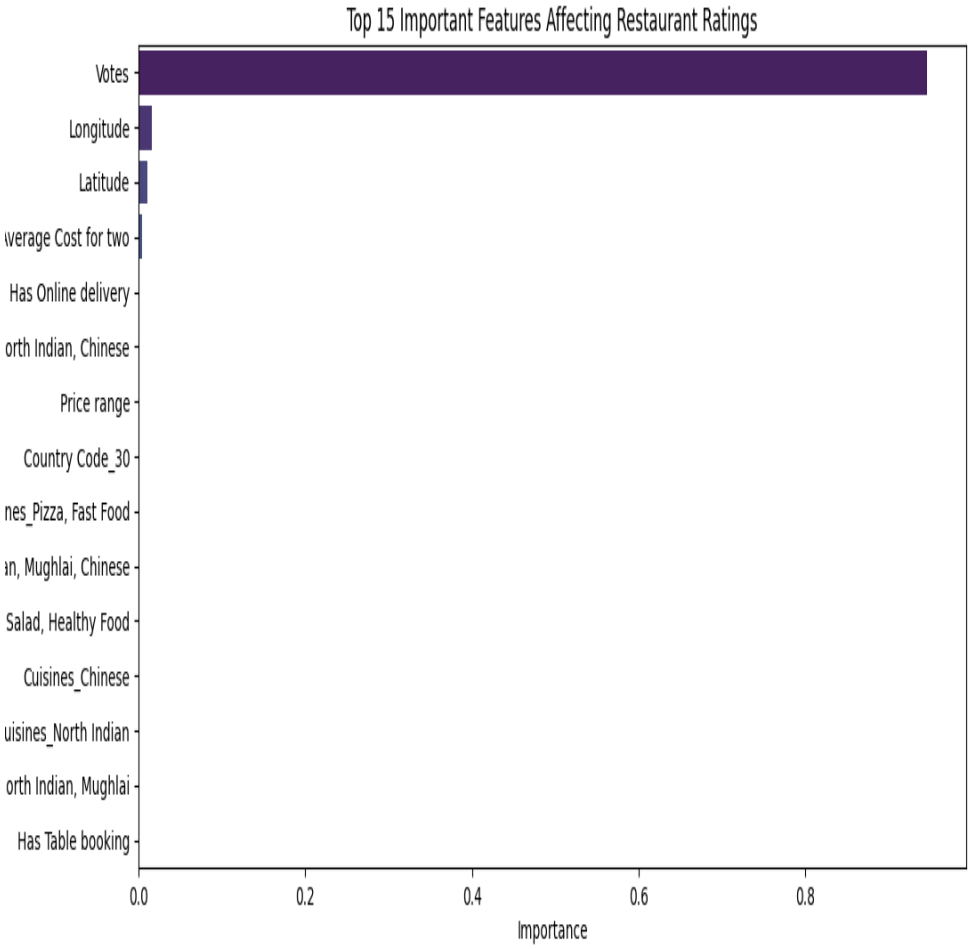
Steps followed:
- Data preprocessing (drop irrelevant columns, handle missing values, encode categorical features)
- Split into training (80%) and test (20%)
- Trained **Linear Regression** and **Random Forest Regressor**
- Evaluated with R^2 and RMSE metrics

Results:
Linear Regression $\rightarrow$ $R^2 = 0.16$, RMSE = 1.37
Random Forest $\rightarrow$ $R^2 = 0.96$, RMSE = 0.30

Random Forest performed significantly better than Linear Regression, showing that non-linear relationships exist in the dataset.

Feature Importance:
- Votes emerged as the most important factor.
- Cost, cuisines, and location had smaller impact.
- The number of votes strongly affects the credibility of ratings.


Actual vs Predicted Restaurant Ratings
Predicted Ratings (RF)
Actual Ratings
0
1
2
3
4
5



Conclusion:
Random Forest Regressor is the most suitable model for predicting restaurant ratings. Votes strongly influence ratings, while cost and cuisines have moderate impact. This task enhanced understanding of preprocessing, regression models, metrics, and feature importance.