

# Restaurant Ratings:

Machine learning is a challenging task when predicting restaurant ratings. Restaurant ratings have become increasingly difficult to anticipate with the popularity of online restaurant review websites. The market is still at an early stage of this industry's saturation. The number of restaurants opening every day is increasing. Their competition with already established restaurants has become increasingly difficult. It is mainly due to the subjective nature of restaurant ratings, which depend on many factors, including the reviewer's preferences, the quality of the food, the ambiance, and the overall service of the restaurant. We will use machine learning to explore how restaurant ratings can be accurately predicted in this theory.

## Objective

The dataset contains all the reviews for the restaurants from customers. We use machine learning models for a particular restaurant to predict its rating. Several parameters need to be considered when predicting ratings, such as the average cost for two people, the availability of online ordering, foods, menu items, and most popular dishes.

## Data Collection

It is imperative to collect the necessary data for any machine-learning project. Data will be collected from restaurant review websites like Zomato for this project. In addition to the reviewer's name, the review's date, rating, and the text of the review, we will need to gather additional information relevant to

the review. In addition, we must collect information about the restaurant, such as its location, the type of cuisine offered, the price range, and reviews.

## **Dataset Understanding**

The dataset is in the form of a CSV file. Before pre-processing, there were 20 columns and 50 rows in the dataset, with a few columns having lesser values and all columns were objects.

Several null values were present in the data. We followed the following steps to resolve the missing values problem:

1. The first step was to remove any rows in the 'cuisines' column with null values.
2. In the second step, we dropped the 'address', 'phone', 'URL', and 'listed\_in(city)' columns since they contained many null values.
3. Renaming the columns 'approximate\_cost(for 2 people)' and 'listed\_in(type)' to 'average\_cost' and 'listed\_type' was the third step.

## **Models**

Our data was split into 80 percent training data and 20 percent test data after preprocessing. Our next step was to fit the data using Machine Learning models.

Various regression algorithms were used to build the model, including Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, K Nearest Neighbours (K-NN), ADA Boost, and XG Boost.

## **Conclusion**

This paper examined different algorithms and features to predict a restaurant's ratings. Many factors that could be controlled were taken into account when making predictions. Thus, the models that gave us the high  $R^2$  score and low mean absolute percentage error can successfully predict the ratings of the restaurants effortlessly.