

# High Level Design for Zomato Rating Prediction

Prepared by

**Vikram Nayyar**

**vikramnayyar@live.com**

October 08, 2021

## **Abstract**

In the past six years, over 1400 new eateries have opened. Evidently, the population's willingness to visit restaurants; has phenomenally increased. Subsequently, numerous restaurant categories have emerged at several locations. Furthermore, they offer several cuisines, price ranges, services and offers. Therefore, predicting the customer's liking is remarkably challenging. To overcome the limitation, an app is developed that predicts the rating of a restaurant. This is based on \_\_\_ user inputs. The prediction model was trained and validated using various data science models. The best validated model was developed to form a Streamlit app.

# 1. Introduction

## 1.1 Purpose

The document provides a high level technical description of the Restaurant Rating Prediction Application. The application was developed to accurately predict a restaurant's rating.

The document also provides essential insights to overcome any difficulties in application development.

## 1.2 Scope

The document is aimed to be a foundational basis for technical insights of the application. The targeted audiences are

- Development Team
- Marketing Team
- Administration

## 1.3 Definitions

- **GUI**- Graphical User Interface is the front end of the application. This is an interface between the user and the model.
- **Model**- This is a data science model; created after training on a suitable dataset.
- **Prediction**- The output of the model; that represents the outcome.
- **User**- The customer using the application.

## 2. General Description

### 2.1 Proposed Solution

A general product design is represented in figure 1. Based on the particular restaurant, a user enters desired inputs. GUI accepts the inputs; and assigns them values according to the model dictionary. Succeeding inputs are fed to the model.

The model processes the inputs and sends the predictions to the GUI. GUI displays the output to the user.



**Fig. 1.** General design of the data science application.

### 2.2 Wireframe

Wireframe of the product's UI is depicted in figure 2. In order to predict rating, the application expects 6 user inputs. Each input is properly labeled and is clearly annotated with the user selection. This prevents any confusion in selection. Please note; usual values are pre-selected for each user field. This is for the user's convenience.

The user selections are followed by the predict button. Clicking this button; inputs the user selections to the model, and the predicted rating is displayed below the predict button.

Rating Prediction for a Bangalore Restaurant

1. Does restaurant provide online orders?

☒ Yes

☐ No

2. Does restaurant provide booking?

☒ Yes

☐ No

3. How many votes have restaurant received?

324

4. What is approximate cost of restaurant?

800

5. What is restaurant's location?

Select

6. What is restaurant's type?

Select

Predict

Prediction Statement

**Fig. 2.** Wireframe of the data science application.

## 2.3 Future Scope

As the product is based on Bangalore restaurants, this can be extended to a number of cities.

## 2.4 System Environment

- **Programming Language:** Python 3.8.8
- **Numerical Calculations:** Numpy, Scipy
- **Plots:** Matplotlib, Seaborn
- **Model Training:** Scikit Learn, XGBoost, LightGBM, CatBoost
- **Application Interface:** Streamlit
- **Deployment:** Heroku

## 2.5 Constraints

Previously discussed wireframe; depicts a suitable layout and describes necessary inputs. Therefore, referring to the wireframe will prevent any oversight. Given a number of user inputs, the application should be prefilled with usual values. This will be convenient for the user.

New market trends emerge constantly. Therefore; after a suitable period, the model is required to be updated. This will maintain the model accuracy.

## 3. Design Details

### 3.1 Process Flow

The process flow diagram is shown in figure 3.

#### 3.1.1 Data Acquisition & Preprocessing

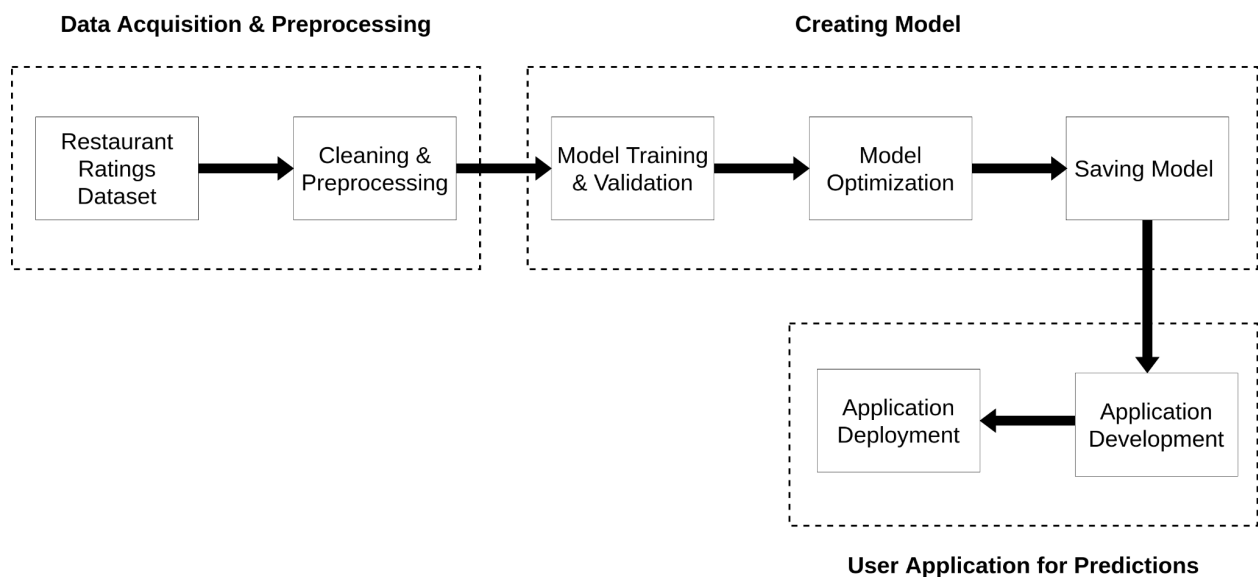
Bangalore restaurants dataset was acquired from Zomato. This was cleaned and suitably prepared to run a number regression models.

#### 3.1.2 Modeling

A number of models were trained and tested. The model with highest accuracy was selected. The parameters of the selected model were optimized to finally save the model.

#### 3.1.3 User Application

The model requires 6 different inputs to make rating prediction. These inputs are user/restaurant specific. Therefore, to accept different inputs from user; an application was developed. Saved model was imported in this application.



**Fig. 3.** Process flow description of application.