

# High Level Design (HLD)

## Restaurant Rating Prediction

Revision number: 2.0

Last date of revision: 30<sup>th</sup> Aug 2021

**Mohamed Illiyas I**

## Document Version Control

Date Issued	Version	Description	Author
21/08/2021	1	Initial HLD-V1.0	Mohamed Illiyas
30/08/2021	2	Final HLD-V2.0	Mohamed Illiyas

# Contents

Document Version Control .....	2
Abstract .....	4
1. Introduction .....	5
1.1. Why this High-Level Document? .....	5
1.2. Scope.....	5
1.3. Definitions.....	5
2. General Description .....	6
2.1 Product Perspective.....	6
2.2 Problem Statement .....	6
2.3 Proposed Solution .....	6
2.4 Data Requirements.....	6
2.5 Tools Used.....	8
2.6 Constraints.....	10
2.7 Assumptions.....	10
3. Design Details.....	11
3.1 Process flow.....	11
3.1.1 Model Training & Evaluation .....	11
3.1.2 Deployment Process.....	12
3.2 Error Handling .....	12
4. Performance.....	13
4.1 Reusability.....	13
4.2 Application Capability .....	13
4.3 Resource Utilization.....	13
4.4 Deployment.....	13
5. Conclusion.....	14
6. References.....	15

## Abstract

Restaurant Industry is a highly growing sector all over the world. After online ordering facilities came into existence, people became excited and so the demand for restaurants keeps on increasing day by day. In India, Bangalore is one of the finest places with a wide variety of cuisines. As the number of restaurants has increased, it became difficult for the peoples to select particular good restaurant. So, the restaurant rating has become the most common used parameter for judging a restaurant. Rating of a restaurant depends on factors like reviews, area, locality, food, cuisines, average cost, etc. This project aims to understand the factors affecting the rating of restaurants. Machine learning model aids to predict the rating of restaurants. The model is built using various regression algorithms and ensemble techniques. The result of the project helps new restaurant in deciding their menu, cuisines, cost, type of foods, location, ambience, etc which will increase their business.

# 1. Introduction

## 1.1 Why this High-Level Document?

The purpose of this High-Level Design (HLD) Document is to add necessary details to the current project description to represent a suitable model for coding. This model is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

**The HLD will:**

- Present all the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance and requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
  - Security
  - Reliability
  - Maintainability
  - Portability
  - Reusability
  - Application compatibility
  - Resource utilization
  - Serviceability

## 1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture, application flow (Navigations), and technology architecture. The HLD uses non-technical to mildly-technical term which should be understandable to the administrator of the system.

## 1.3 Definitions

TERM	DESCRIPTION
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment
Azure	Cloud service provided by Microsoft

## 2. General Description

### 2.1 Product Perspective

The Restaurant Rating Prediction system is a machine learning based model which will help us to predict the rating of the restaurant in Bangalore.

### 2.2 Problem Statement

The main goal of this project is to create a solution for finding this rating, and to implement the following use cases,

- To perform data pre-processing
- To perform Exploratory data analysis
- To predict ratings to improve new restaurants.

### 2.3 Proposed Solution

The proposed solution for this project is to find the ratings of the new restaurants by using Machine learning models. As the target is a continuous variable, we can use regression or decision tree models as the base model for this problem. From this baseline model, we can improve the performance by ensemble techniques.

However, drawing a baseline model is important since it tells us how well other models have performed compared to the base model. Here, the base model for the Restaurant Rating dataset is Linear Regression.

1. Baseline Model: Linear regression
2. Actual model: Random Forest

### 2.4 Further Improvements

The baseline model is further improved by using ensemble techniques like bagging (Bootstrap aggregation). We used Random Forest Regression.

## 2.5 Data Requirements

The dataset consists of a table with 54000+ records and 17 features.

- url: contains the URL of the restaurant in the Zomato website.
- address: contains the address of the restaurant in Bengaluru
- name: contains the name of the restaurant
- online\_order: whether online ordering is available in the restaurant or not
- book\_table: table book option available or not
- rate: contains the overall rating of the restaurant out of 5
- votes: contains total number of rating for the restaurant as of the above-mentioned date
- phone: contains the phone number of the restaurant
- location: contains the neighborhood in which the restaurant is located
- rest\_type: restaurant type
- dished\_liked: dishes people liked in the restaurant
- cuisines: food styles, separated by comma
- approx.\_cost (for two people): contains the approximate cost for meal for two people
- reviews: list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer
- menu\_item: contains list of menus available in the restaurant
- listed\_in(type): type of meal
- listed\_in(city): contains the neighborhood in which the restaurant is listed

## 2.6 Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask, Azure cloud, Git are used to build the whole model.







#### High-Level Design

- VS code is used as code editor.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Azure is used for deployment of the model.
- Frontend development is done using HTML/CSS/JS
- Python is used for backend development.
- Flask is used as Framework for python.
- GitHub is used as version control system.

## 2.6 Constraints

The restaurant rating prediction application must be user friendly, as automated as possible and users should not be required to know any of the workings.

## 2.7 Assumptions

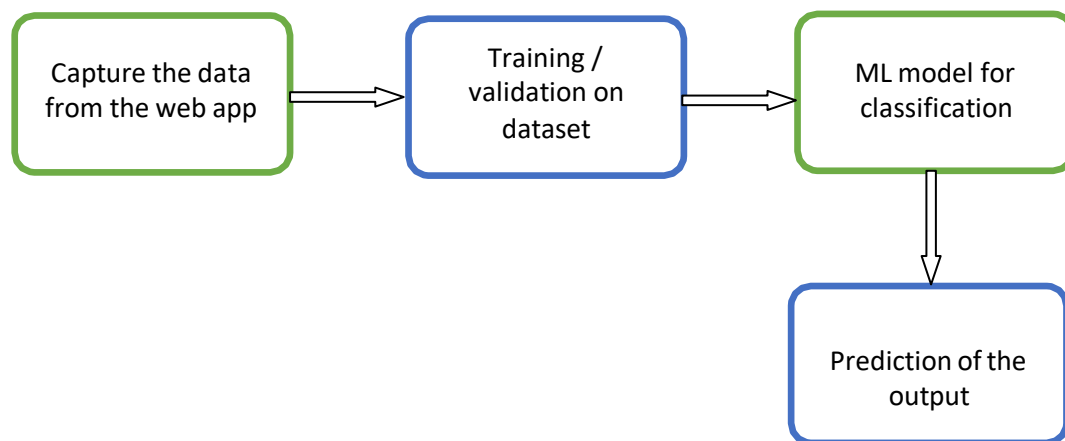
The main objective of the project is to implement the use cases as previously mentioned in the Problem statement. For new data set that comes from the user input which was captured in the user interface of the web application. Machine learning based algorithm is used for detecting the above-mentioned use cases based on the input data. It is assumed that all aspects of this project have ability to work together in the new environment.

## 3.Design Details

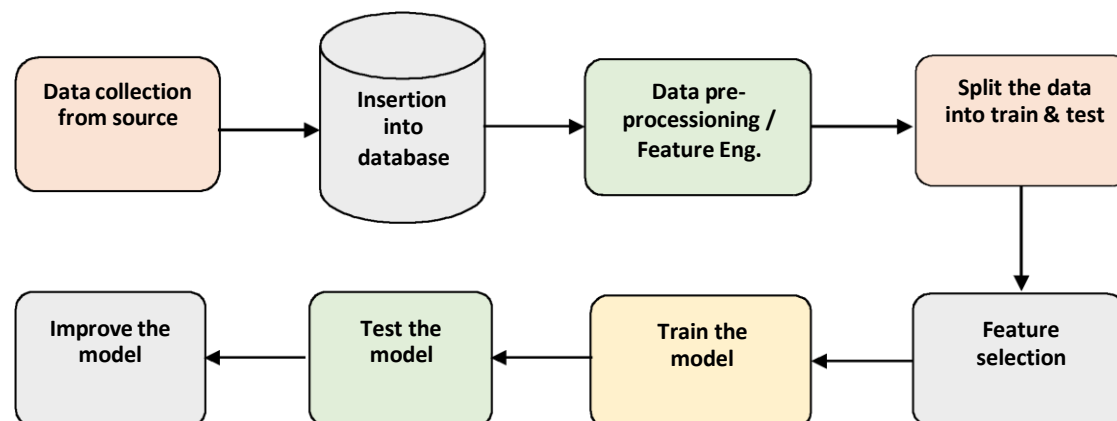
### 3.1 Process Flow

For predicting the rating of the restaurant, we will use regression model as the base model. Below is the process flow diagram.

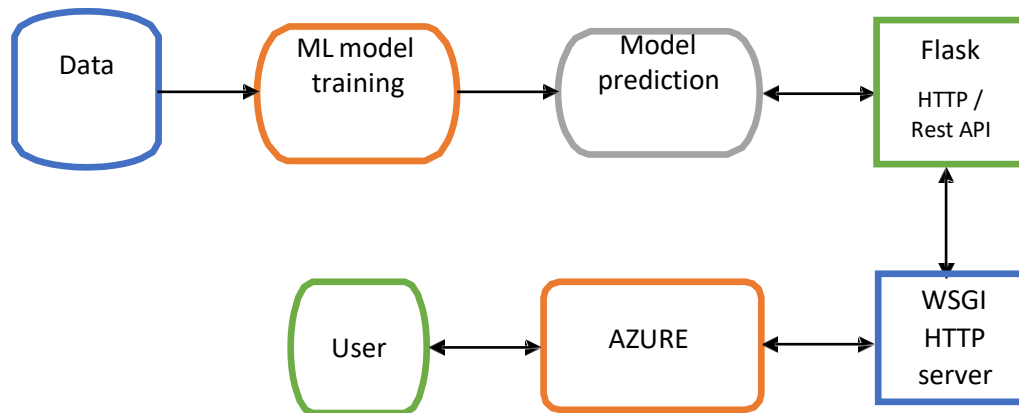
#### Proposed Methodology



#### 1.1.1 Model Training and Evaluation



### 1.1.2 Deployment Process



### 1.2 Error Handling

Should error be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage

## 4. Performance

We can observe that the  $r^2$  score of the predicted output was 33% for our base model linear regression. Decision tree perform better than base model. Its score is 78%. Further, improving the model with ensemble technique gives 81% score(Random Forest)

### a. Reusability

The code written and the components used should have the ability to be reused with no problems.

### b. Application Compatibility

The different components for this project will be using as an interface between them. Each component will have its own task to perform, and it is the job of the python to ensure proper transfer of information.

### c. Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

### d. Deployment



## 5. Conclusion

In this project, we have done data preprocessing and performed Exploratory data analysis and the data is interpreted using visualizations. Many machine learning algorithms similar to regression are applied to the data and most effective model is recognized and implemented. The best among those models were the Random Forest model which shows an accuracy of 81%.

## 6. References

1. <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>