high-level design [hld]

Restaurant Rating Prediction

October 3, 2022

[Company name]

[Company address]

# **Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Autor |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Contents**

[**Document Version Control** 1](#_Toc116135212)

[**Abstract** 3](#_Toc116135213)

[**1.** **Introduction** 4](#_Toc116135214)

[1.1. Why this High-Level Design Document? 4](#_Toc116135215)

[1.2. Scope 4](#_Toc116135216)

[**2.** **General Description** 5](#_Toc116135217)

[2.1. Product Perspective 5](#_Toc116135218)

[2.2. Problem Statement 5](#_Toc116135219)

[2.3. Proposed Solution 5](#_Toc116135220)

[2.4. Further Improvements 5](#_Toc116135221)

[2.5. Technical Requirements 5](#_Toc116135222)

[2.6. Data Requirements 5](#_Toc116135223)

[2.7. Tools Used 6](#_Toc116135224)

[2.8. Constraints 6](#_Toc116135225)

[2.9. Assumptions 7](#_Toc116135226)

[**3.** **Design Details** 8](#_Toc116135227)

[3.1. Process Flow 8](#_Toc116135228)

[3.1.1. Model Training and Evaluation 8](#_Toc116135229)

[3.1.2. Deployment Process 8](#_Toc116135230)

[3.2. Event Log 8](#_Toc116135231)

[3.3. Error Handling 8](#_Toc116135232)

[**4.** **Performance** 8](#_Toc116135233)

[4.1. Reusability 9](#_Toc116135234)

[4.2. Application Compatibility 9](#_Toc116135235)

[4.3. Resource Utilization 9](#_Toc116135236)

[Resource utilization is a vital area of project management that can sustain growth, increase profits, improve rating and enhance customer experience. 9](#_Toc116135237)

[4.4. Deployment 9](#_Toc116135238)

[A graphical user interface is created using python for the users to interact and predict the outcome. 9](#_Toc116135239)

[**5.** **Dashboards** 9](#_Toc116135240)

[5.1. KPIs (Key Performance Indicators) 9](#_Toc116135241)

[**6.** **Conclusion** 10](#_Toc116135242)

[**7.** **References** 10](#_Toc116135243)

# **Abstract**

In the era of technology and internet, a wide variety of services are going online. One of them is E-Commerce. E-commerce services are growing exponentially with new start-ups and well-established businesses. The performance of E-commerce businesses are measured in terms of rating and reviews provided by their customers/users. The collection of ratings and reviews can be used as an immensely powerful tool for any E-commerce business. Online reviews can reveal a lot about online store. A wealth of positive words can have a measurable impact on sales, driving purchases and creating a base of consumers who will stand behind you and your product. This digital feedback acts as an incredible way to increase traffic, improve PPC, and build your online reputation while increasing the confidence of consumers.

In this proposed project, we have predicted respective ratings of various zomato restaurants based on certain features using machine learning technology. This prediction will act as a decision-making tool for the restaurants to help themselves in budgeting, planning, and estimating future growth.

# **Introduction**

## Why this High-Level Design Document?

This High-Level Design (HLD) Document's main goal is to provide the existing project description with the extra depth it needs to describe an appropriate coding model. This paper may be used as a reference guide to show how the modules interact at a high level and are also meant to aid in identifying conflicts before coding.

**The HLD will:**

* Present all of the design elements and fully describe them.
* Explain the user interface that is being used.
* Explain the software and hardware interfaces.
* Describe the necessary performance standards.
* Include the project's architecture and design elements.
* List and explain the non-functional characteristics, such as:
  + Security
  + Reliability
  + Maintainability
  + Serviceability
  + Portability
  + Reusability
  + Application compatibility
  + Resource utilization

## Scope

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

# **General Description**

## Product Perspective

The Restaurant Rating Prediction solution is the Machine learning-based model used to predict the rating of the restaurant based on certain features which will act as a decision-making tool for the restaurants to help themselves in budgeting, planning, and estimating future growth.

## Problem Statement

The main goal of this project is to perform extensive Exploratory Data Analysis (EDA) on the Zomato Dataset and build an appropriate Machine Learning Model that will help various Zomato Restaurants to predict their respective Ratings based on certain features.

## Proposed Solution

The proposed solution of this document is a supervised machine learning model of regression using the Random Forest algorithm. It is deployed with a piece of python code with a good-looking, OS adaptive user interface. Some of the attributes are chosen to predict the value of the rating based on their correlation with rating. Features like votes, cost and cuisines are used to predict the output value. All these chosen features are fed into number of models to predict the rating of restaurants out of which the best model is selected based on the outcome/performance.

## Further Improvements

This model is capable to be implemented further on any e-Commerce platform with an addition of more correlated features to predict the rating.

## Technical Requirements

This solution is portable and capable to run on any operating system and architecture. The user interface is ready to adopt the OS native look and feel.

This model is compatible to run on any type of processing unit. It can be cloud, Server, PC, Laptop, etc.

## Data Requirements

The data requirement is completely dependent on problem statement. Fields like name, url, address provided in the dataset are not necessary to solve the particular problem. It requires below mentioned certain fields to estimate the rating of the given restaurant :

|  |  |  |
| --- | --- | --- |
| Field | Detail | Constraint |
| Type: | Type of the given restaurant. It can be one of the 7 listed types. | Should be in the listed values |
| Location: | This is the location where the restaurant is situated in Bangalore. | Should be in the listed values |
| Votes: | It is the number of votes received by particular restaurant from its customers. | Should be an integer value |
| Cost: | Cost of the table for two members. | Should be an integer value |
| Cuisines: | Cuisines are cooking style of particular country/region available in the restaurant. | Should have text separated by coma (,) |
| Online Order | Does the restaurant provide the ease to order items online? | Checkbox |
| Book Table | Does the restaurant provide the service to book tables in advance? | Checkbox |

Table 1: Strict Data Requirements

## Tools Used

|  |  |
| --- | --- |
|  | Python Programming Language is used to structure the project with its basic libraries as os, logging, time, tkinter etc and some data science specialized libraries like numpy, pandas, matplotlib, seaborn, scikit-learn etc acting as paramount. |
|  | Jupyter Notebook is used as Python IDE for data cleaning and Research Work/Model Building. |
|  | PyCharm is used as an Integrated Development Environment (IDE) for the deployment. |
| Power BI Logo, symbol, meaning, history, PNG | Power BI is used to create interactive dashboard to draw meaningful insights from the restaurant information. |
| GitHub (@github) / Twitter | GitHub is being used for the project version control. |

* + - Python Library Pickle is used for the model and encoder saving and loading process.
    - Python’s tkinter is the user interface supporting library as it provide platform native experience.

## Constraints

To use the user interface major constraints are defined in the Table 1.

Furthermore, for the implementation of the model we are cautious about the model and encoder saved format.

## Assumptions

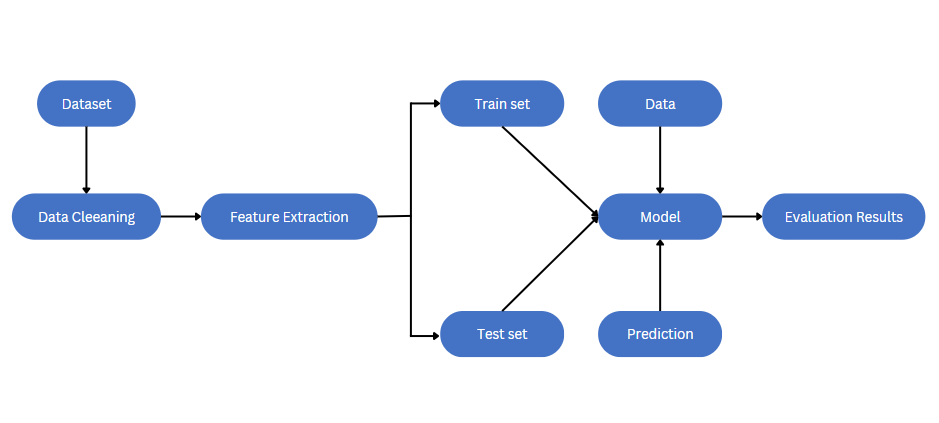
Below mentioned are few assumptions taken into account for model consistency:

* + - The rating will be predicted only if the cuisines entered by users are used from the dataset.
    - Location and restaurant type are selected only from the drop- down .

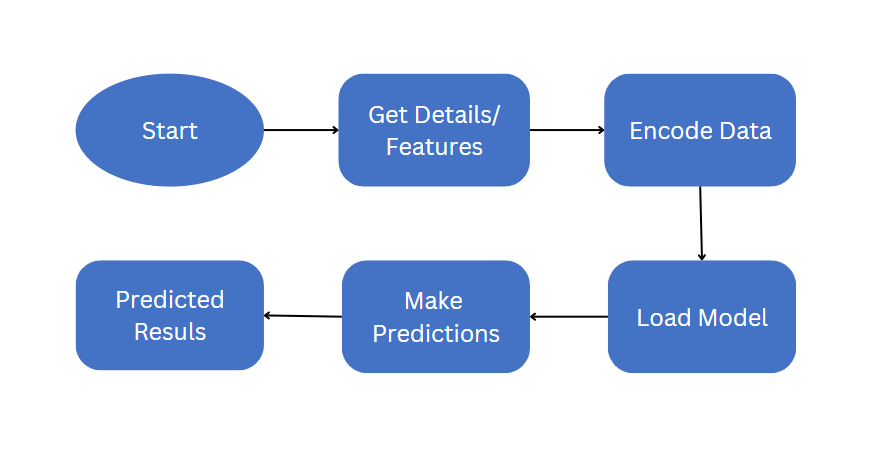
# **Design Details**

## Process Flow

## Model Training and Evaluation



## Deployment Process



## Event Log

Solution includes the logging feature. A separate directory is in the solution package where you access the logs of the system. At every step a log is created in those files. On every launch new log file will be created and will be name according to time.

## Error Handling

Error encountered by the solution will be shown as a prompt.

# **Performance**

The supervised machine learning model is used to predict rating of the restaurants. The data is first processed, analysed and visualized. Certain features are then selected based on analysis and several models were trained. The best model with high performance and accuracy is selected. Also, model retraining is very important to improve performance.

## Reusability

The python code implemented should have the ability to be reused with no problems.

## Application Compatibility

This model is compatible to run on any type of processing unit. It can be cloud, Server, PC, Laptop, etc. Python is used to ensure proper transfer of information.

## Resource Utilization

## Resource utilization is a vital area of project management that can sustain growth, increase profits, improve rating and enhance customer experience.

## Deployment

# A graphical user interface is created using python for the users to interact and predict the outcome.

# **Dashboards**

An extensive exploratory data analysis is required to draw insights from our data. Those insights helps us in decision making and growth of business. Based on the dashboard we can find our key performance indicators. We have performed exploratory data analysis of Zomato dataset in Power BI using suitable graphs and interactive dashboard. With the help of Dashboard, we can track KPIs.

## KPIs (Key Performance Indicators)

1. Cost of food and services charged by restaurants.
2. The inability of restaurant to deliver both online and offline orders.
3. The inability of restaurant to give its customers a option of booking table or not.
4. Cuisines provided by the restaurant.
5. Location of the restaurant.
6. Type of restaurant.

# **Conclusion**

The proposed machine learning model will predict rating of restaurants based on certain features. Rating acts as the most essential parameter for any restaurant in terms of its customer experience/feedback. The predicted rating will help the business in inspection, improving customer experience, growth of business and better decision making.

# **References**

* Dataset is provided by the Kaggle : [Zomato Bangalore Restaurants](https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants?resource=download)