High Level Design(HLD)

Restaurant Rating Prediction

Revision number: 2.0

Last date of revision: 30th Jan2023

Tushar Jain

Document Version Control

Data Issued	Version	Description	Author
17/01/2023	1.0	Initial version	Tushar Jain
30/01/2023	2.0	Finial version	Tushar Jain

Abstract

This project aims to perform an extensive Exploratory Data Analysis (EDA) on the Zomato dataset and build an appropriate Machine Learning Model to predict the ratings of Zomato Restaurants based on certain features. The project follows the classical machine learning process, including Data Exploration, Data Cleaning, Feature Engineering, Model Building, and Model Testing. Different machine learning algorithms will be tested to determine the best fit for the case. The ultimate goal is to build a solution that can accurately predict the ratings of the restaurants listed in the Zomato dataset.

Table of Content

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 include:

- Present all the design aspect and define them in detail
- Describe the User Interface being implemented
- Describe all hardware and software interfaces
- Describe the performance and requirements
- Include all design features and the architecture of the project
- List and describe the non-functional attribute like:
  - Security
  - Reliability
  - Maintainability
  - Portability
  - Re-usability
  - 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 (Navigation), 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

2. General Description

2.1. Product Perspective

The Restaurant Rating Prediction is a machine learning based model which will help us to predict the rating of the restaurant in Bangalore. The dataset also contains reviews for each of the restaurant which will help in finding overall rating for the place.

2.2. Problem Statement

The main goal of this project is to perform exploratory data analysis and later predict the rating of the restaurant.

2.3. Proposed Solution

1. Data Collection: The first step in this project is to collect the Zomato dataset, which contains information about various restaurants, their menus, and their ratings.
2. Data Exploration: The next step is to perform an extensive Exploratory Data Analysis (EDA) to get a better understanding of the data, including the distribution of the ratings, the relationship between the features, and any missing values or outliers in the data.
3. Data Cleaning: Based on the results of the EDA, the data will be cleaned and pre-processed to ensure that it is suitable for the machine learning algorithms.
4. Feature Engineering: The next step is to engineer new features or modify existing features to better capture the relationships between the features and the ratings.
5. Model Building: The cleaned and pre-processed data will then be used to build a machine learning model. Different algorithms will be tested to determine the best fit for the case.
6. Model Testing: The performance of the machine learning model will be evaluated through various testing methods, including cross-validation and hold-out testing.

7. Solution Implemenation: The best-performing machine learning model will be implemented as a solution to predict the ratings of Zomato Restaurants based on certain features.
8. Conclusion: The project report will provide a detailed account of the methodology, results, and conclusion of the project, making it a valuable reference for future work in this field.

2.4. Data Requirements

The dataset consists of a table with 56351 records and 17 features.
The given features are:

- url: contains the url of the restaurant in the zomato website.
- address: contains the address of the restaurants in Bengaluru
- name: contains the name of restaurant
- online_order: whether online ordering is available or not
- book_table:book table option is available or not
- rate: contains overall rating of all restaurants out of 5
- votes: contains total number of rating for the restaurant as of the above mentioned date
- phone: contains the phone number of restaurants
- 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
- review_list: list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer
- menu_list: 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

2.5. Tools Used

Python programming language and framework such as Numpy, Pandas, Scikit-Learn, Streamlit, Heroku, Git.

2.6. Constraints

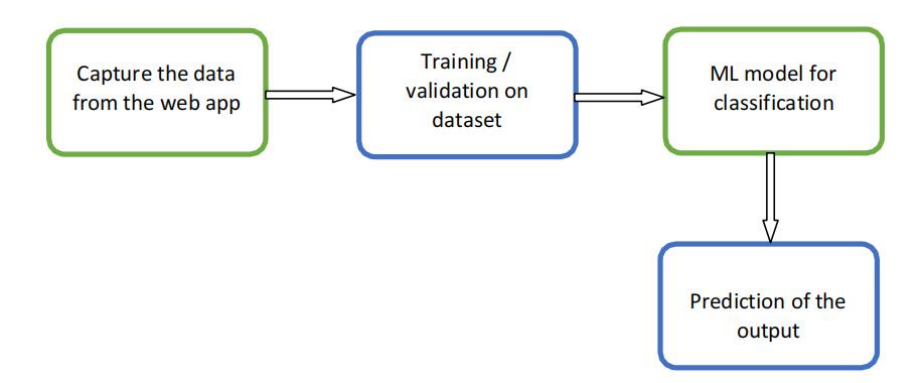
The restaurant rating prediction application must be user friendly, as automated as possible and users should not be required to know any of the workings.

3. Design Detail

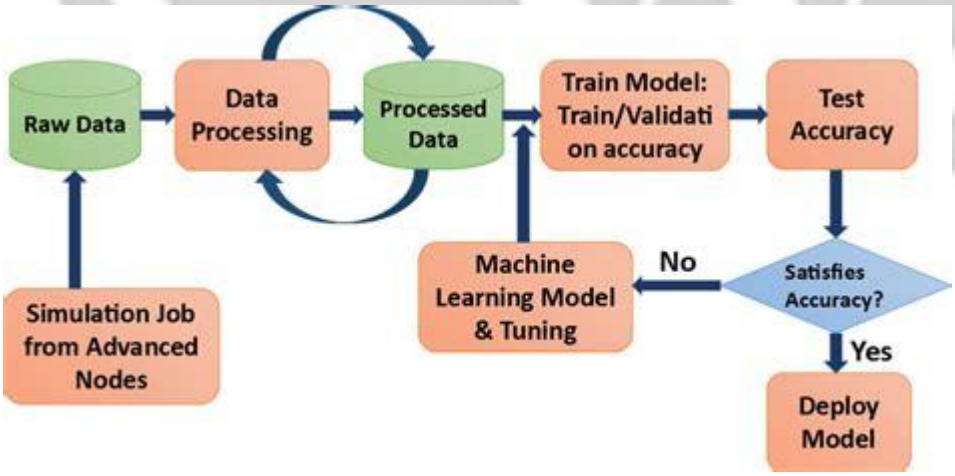
3.1. Process Flow

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

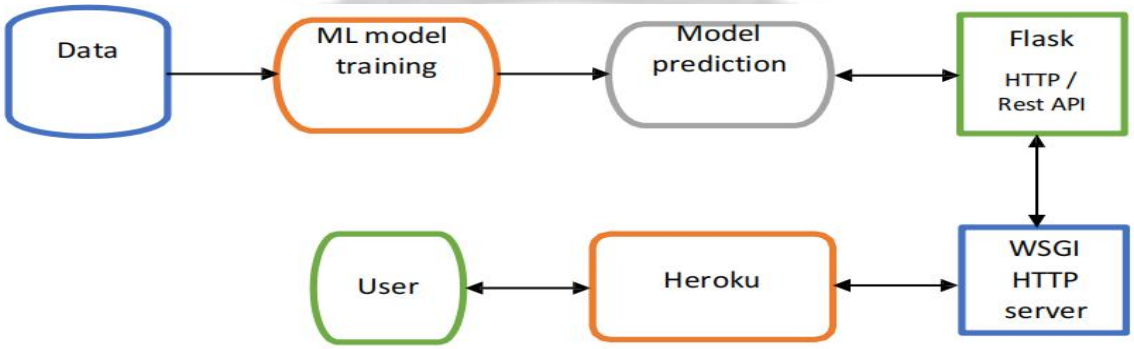
Proposed Methodology



3.1.1. Model Training and Evaluation



3.1.2. Deployment Process



3.2. 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 accuracy of the predicted output was seen at 87% using Random forest classifier. Other classification models such as logistic regression and decision tree have given good accuracy above 23% and 83% respectively.

4.1. Re-Usability

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

4.2. 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.

4.3. Resource Utilization

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

4.4. Deployment

5. Conclusion

In this project, EDA was performed showing various analytical results. About three machine learning models were built and each of the model shows different accuracy. The best among these models were the Random Forest model which shows an accuracy of 87%.

6. Reference

1. https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants
2. https://medium.com/analytics-vidhya/zomato-bangalore-restaurant-analysis-and-rating-prediction-101fd635ab15