

High Level Design(HLD)

Restaurant Rating Prediction

Harsh Paseriya

Document Version Control

Date Issued	Version	Description	Author
08/08/2022	1	Initial HLD-V1.0	Harsh Paseriya
08/08/2022	2	Final HLD-V2.0	Harsh Paseriya

Contents

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.....	7
2.6. Constraints.....	8
2.7 Assumptions.....	8
3. Design Details.....	9
3.1. Process flow.....	9
3.1.1 Model Training & Evaluation.....	9
3.1.2 Deployment Process.....	10
3.2. Error Handling.....	10

4. Performance.....	11
4.1. Reusability.....	11
4.2. Application Capability.....	11
4.3. Resource Utilization.....	11
4.4. Deployment.....	11
5. Conclusion.....	12
6. References.....	13

Abstract

Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. In the world of rising new technology and innovation, healthcare industry is advancing with the role of Artificial Intelligence. Machine learning algorithms can help to early detection of the disease and to improve the quality of the life. This study demonstrates the how different classification algorithms can forecast the presence of the disease. Different classification algorithms such as Logistic regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine have been tested and compared to predict the better outcome of the model.

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	
		AWS
Amazon Web Services		

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

The proposed solution for this project is Machine learning algorithms can be implemented to predict the rating of the restaurant. Considering various features like online order, book table, votes, rest type, cuisines, review as inputs from the web app, the implemented classification model will predict the output as rating of the restaurant.

Here, we have used Random Forest Classifier to predict the restaurant rating.

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

1. Baseline Model: Logistic Regression
2. Actual Model: Random Forest

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 restaurant in Bengaluru
- name: contains the name of the restaurant
- online_order: whether online ordering is available in the restaurant or not

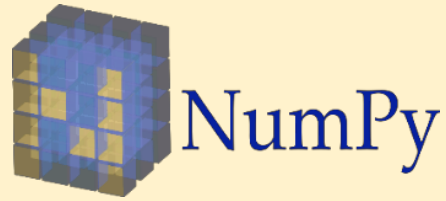
- **book_table**: table book option available or not
 - **rate**: contains the overall rating of the restaurant out of 5
 - **votes**: contains total number of rating for the restaurant as of the above mentioned date
 - **phone**: contains the phone number of the restaurant
 - **location**: contains 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

Name

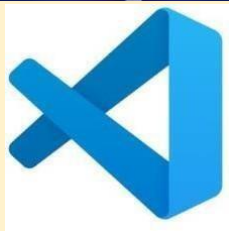
Description

2.5. Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask, Heroku, Git.



Flask



- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- Heroku is used for deployment of the model.
- Frontend development is done using HTML/CSS
- Python Flask is used for backend development.
- GitHub is used as version control system.

2.6 Constraints

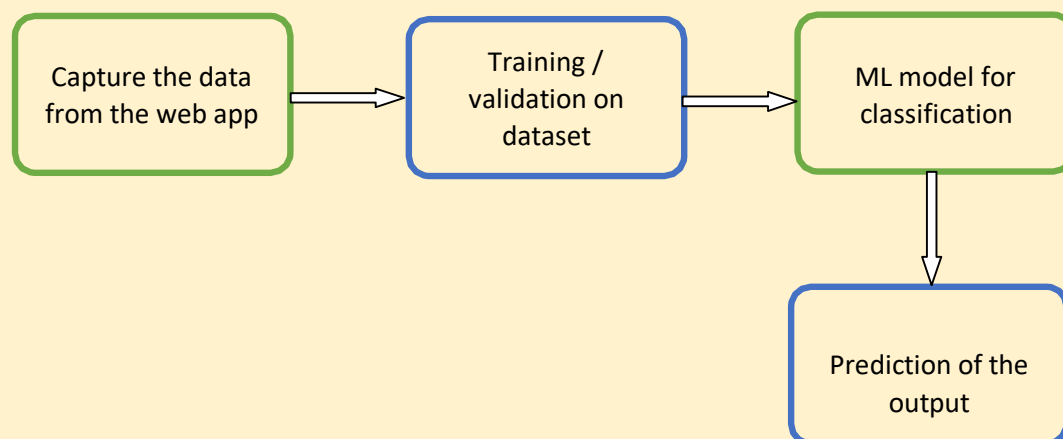
The restaurant rating prediction application must be user friendly, as automated as possible and users should not be required to know any of the workings.

3. Design Details

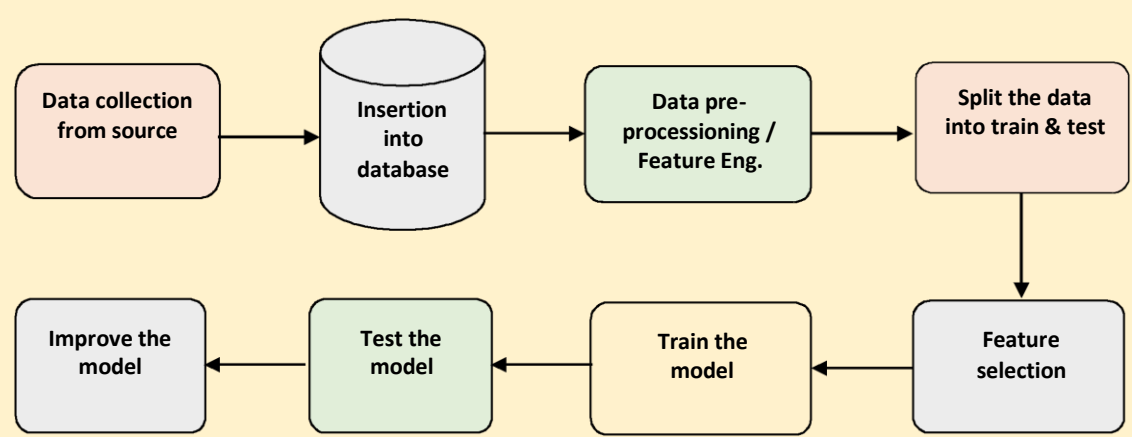
3.1 Process Flow

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

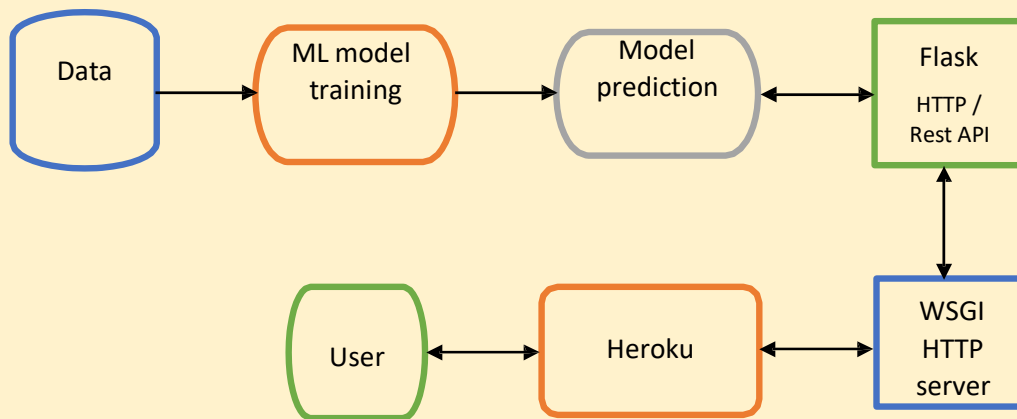
Proposed Methodology



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 Reusability

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

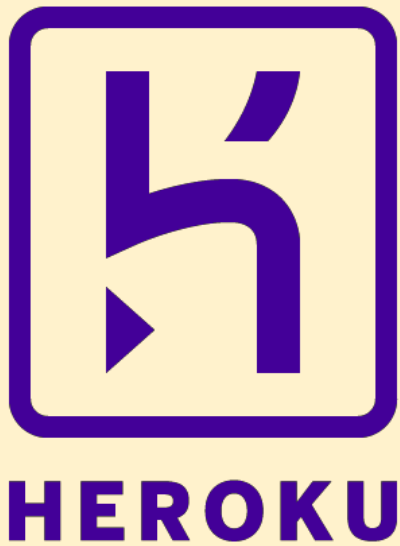
4.2 Application Compatibility

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 accuracies. The best among these models were the Random Forest model which shows an accuracy of 87%.

6. References

1. <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>
2. <https://www.analyticsvidhya.com/blog/2021/04/rapid-fire-eda-process-using-python-for-mlimplementation/>