High-Level Design (HLD)

Restaurant Rating Prediction

Anuj Gupta

Document Version Control

Date issued	Version	Description	Author
15/06/2023	1	Initial-HLD-V1	Anuj Gupta

Content

Abstract

In today's era of online food ordering, one of the primary factors influencing our decision is the ratings and reviews of restaurants. Before placing an order, we often rely on the feedback from other customers to ensure that the restaurant provides high-quality food and timely service. This is particularly true in Bengaluru, one of India's top cities, where many individuals heavily rely on restaurant food due to their busy schedules. With such a high demand for restaurants, understanding the demographics of a location becomes crucial.

In this context, the role of Artificial Intelligence (AI) and machine learning algorithms is paramount. These technologies offer innovative solutions to simplify tasks and enable us to predict and forecast the future. By leveraging machine learning, we can predict the ratings of restaurants based on various factors, helping consumers make informed decisions about ordering food online.

This study showcases the application of different regression algorithms, including Random Forest, XGBoost, LGBM to forecast restaurant ratings. By analyzing historical data, including ratings and reviews, these algorithms can identify patterns and relationships that influence a restaurant's rating. Through comparison and evaluation, the algorithm that performs the best in terms of accuracy and predictive power is selected.

The integration of AI and machine learning in this study empowers consumers by providing them with predictive insights. By leveraging these insights, individuals can make informed choices about ordering food online from a specific restaurant. The goal is to simplify the decision-making process and ensure that customers receive high-quality food and service based on reliable predictions.

Overall, this study demonstrates how AI and regression algorithms play a crucial role in forecasting restaurant ratings. By harnessing the power of technology and data analysis, we can enhance the online food ordering experience in Bengaluru and enable individuals to make well-informed decisions based on ratings and reviews.

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 before 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 terms which should be understandable to the administrator of the system.

1.3 Definitions

Terms	DESCRIPTION
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment

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 restaurants which will help in finding the overall rating for the place.

2.2 Problem Statement

The main goal of this project is to perform extensive exploratory data analysis (EDA) and later predict the rating of the restaurant using the best machine learning alogirthm.

2.3 Proposed Solution

The proposed solution for this project is Machine learning algorithms that can be implemented to predict the rating of the restaurant. Considering various features like online order, book table, location, cuisines, cost_per_person as inputs from the web app, the implemented machine learning model will predict the rating of the restaurant. Here we tried Random Forest Regressor, XGboost Regressor, LGBM Regressor. The final model with best results was LGBM Regressor.

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 the total number of ratings 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 a meal for twopeople.
- **reviews**: list of tuples containing reviews for the restaurant, each tuple consists of two values, rating, and review by the customer.
- **menu_item**: contains the 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 frameworks such as NumPy, Pandas, Scikit-learn, Flask, AWS, Git.

- **Vscode** was used as IDE.
- For visualization of the plots, **plotly** was used.
- **Streamlit** is used for the deployment of the model.
- Frontend development is done using **Streamlit**.
- **Python** and **Streamlit** is used for backend development.
- **GitHub** is used as a version control system.

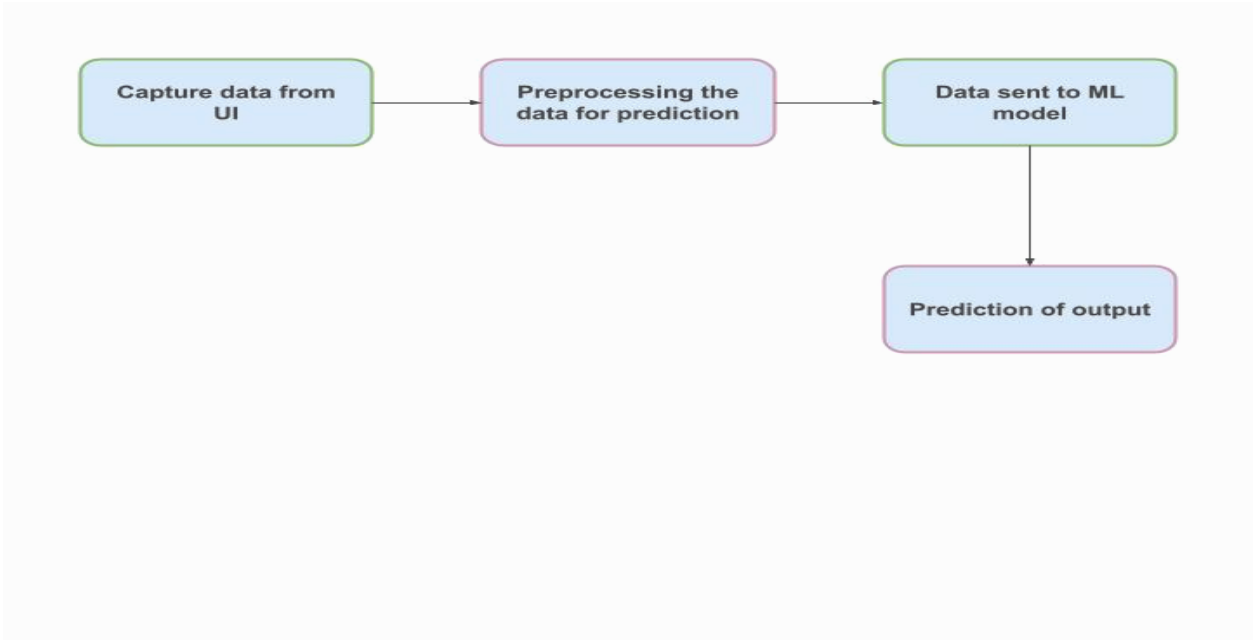
2.6 Constraints

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.

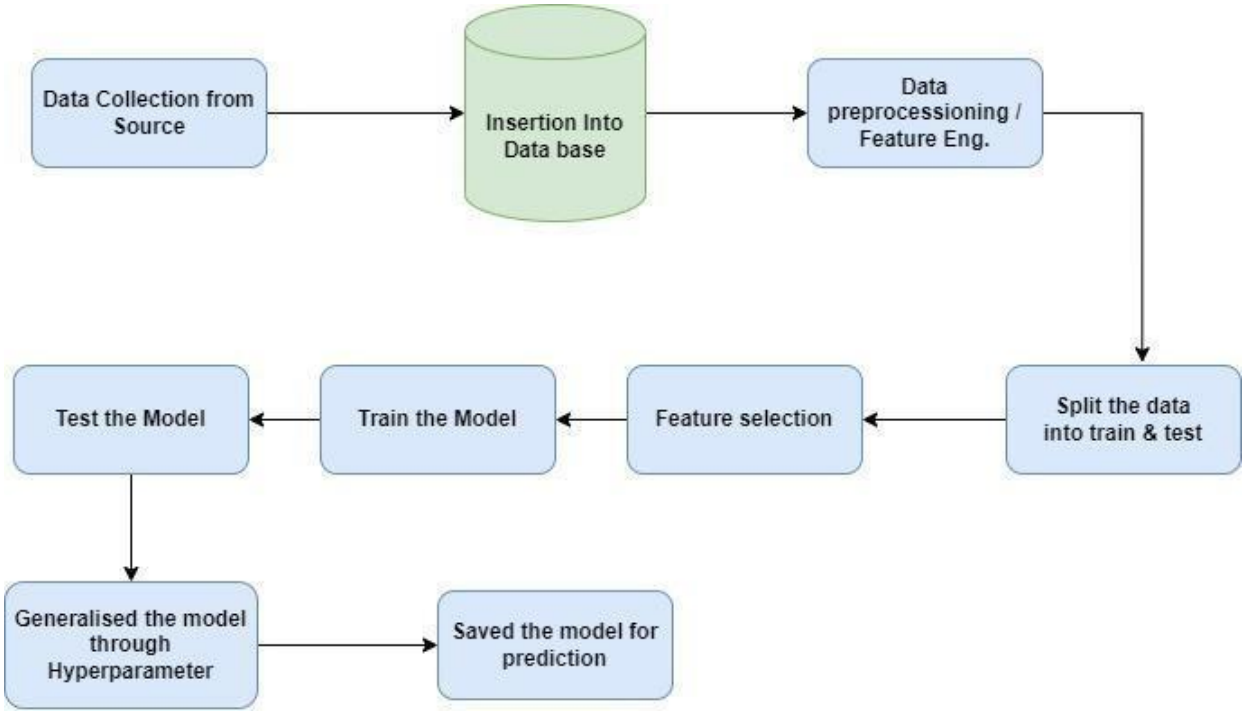
3. Design Details

3.1 Process Flow

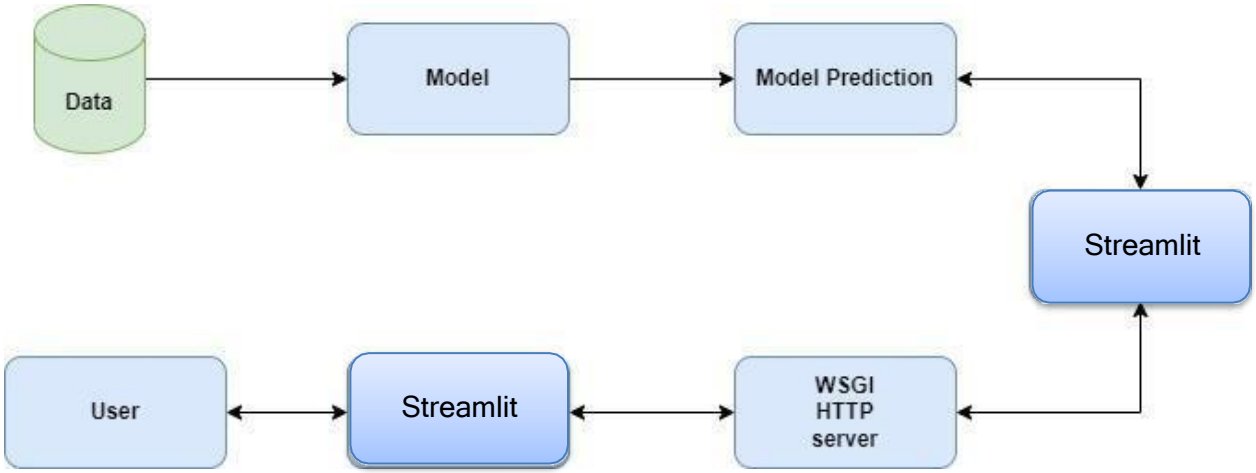
For predicting the rating of the restaurant, we will use a regression model. Below is the process flow diagram is as shown below:



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



4. Performance

We can observe that the LGBM Forest Regressor gave us the best results. Other machine learning models such as Random Forest, Xgboost and LightGBM Regressor were used but the best one was LGBM Regressor.

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 used as an interface between them. Each component will have its task to perform, and it is the job of Python to ensure the proper transfer of information.

4.3 Resource Utilization

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

4.4 Deployment

The project was deployed on streamlit. It is a deployment service that is used for deploying data-based apps like the project I worked on.

5. Conclusion

In this project, EDA was performed showing various analytical results. About three machine learning models were build. The best among these models was the LGBM regressor model.