

**Zomato Restaurant Rating Prediction**

High Level Design

Domain: Machine Learning

Creator: Praveen Kumar

Date: 18.07.2023

Praveen Kumar

+91 7569993454

[talaripraveeenkumar53@gmail.com](mailto:talaripraveeenkumar53@gmail.com)

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Date issued | Version | Description | Author |
| June 28, 2023 | 1.1 | First Draft | Praveen Kumar |
| June 29, 2023 | 1.2 | Added unit test cases | Praveen Kumar |
| July 1, 2023 | 1.3 | Correcting font sizes | Praveen Kumar |

Contents

[Abstract 3](#_Toc110595993)

[Introduction 3](#_Toc110595994)

[What is High-Level Design Document? 3](#_Toc110595995)

[Scope 4](#_Toc110595996)

[General Description 4](#_Toc110595997)

[Definitions 4](#_Toc110595998)

[Product Description 4](#_Toc110595999)

[Problem Statement 4](#_Toc110596000)

[Proposed solution 5](#_Toc110596001)

[Further improvements 5](#_Toc110596002)

[Data requirements 5](#_Toc110596003)

[Tools used 5](#_Toc110596004)

[Hardware Requirements 7](#_Toc110596005)

[Constraints 7](#_Toc110596006)

[Assumptions 7](#_Toc110596007)

[Design Details 7](#_Toc110596008)

[Process Flow 7](#_Toc110596009)

[Event Log 8](#_Toc110596010)

[Error Handling 9](#_Toc110596011)

[Performance 9](#_Toc110596012)

[Reusability 9](#_Toc110596013)

[Application Compatibility 9](#_Toc110596014)

[Resource Utilization 9](#_Toc110596015)

[Dashboards 10](#_Toc110596016)

[KPIs (Key Performance Indicators) 10](#_Toc110596017)

[Conclusion 10](#_Toc110596018)

# Abstract

# The Zomato restaurant rating problem revolves around the task of predicting the rating of restaurants listed on the Zomato platform. Zomato is an online food delivery and restaurant discovery platform that heavily relies on customer ratings and reviews to assist users in making dining choices.

# The project involves utilizing machine learning techniques to analyze various factors that influence a restaurant's rating, including location, cuisine type, pricing, ambience, and customer reviews. By leveraging a large dataset of historical restaurant data from Zomato, the aim is to develop a system that accurately predicts a restaurant's rating. Introduction

## What is High-Level Design Document?

The goal of this HLD or a high-level design document is to add the necessary detail to the current project description to represent a suitable model for coding. This document 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 of design aspects and define them in detail
* Describe all user interfaces being implemented
* Describe the hardware and software interfaces
* Describe the performance requirements
* Include design features and architecture of the project
* List and describe the non-functional attributes such as security, reliability, maintainability, portability, reusability, application compatibility. resource utilization, serviceability

## Scope

The HLD documentation presents the structure of the system, such as 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

## Definitions

|  |  |
| --- | --- |
| Term | Description |
| IMTVP | Zomato Restaurant Rating Prediction |
| Database | Collection of the Information |
| Cloud | A data center full of services connected to the internet performing service |
| IDE | Integrated Development Environment |
| UI | User Interface |
| Anvil | A Python based UI builder |
| Heroku | A cloud service |

## Product Description

IMTVP is a Machine Learning based regression model which helps us to do predictive analysis of the traffic volume using certain features and parameters.

## Problem Statement

To create an ML based solution for predictive analysis of a person’s annual income and also deploy it in the form of a UI.

The aim to predict the traffic volume on specific weekday of month. This is basically regression problem.

## Proposed solution

Using all the standard techniques used in the life cycle of a Data Science project starting from Data Exploration, Data Cleaning, Feature Engineering, Model Selection, Model Building and Model Testing and also building a frontend where a user can fill their information in the form input and get the output instantly.

## Further improvements

This IMTVP can be easily embedded inside any website or an application and everybody can get quick answer by inputting required data on friendly user interface.

This can be further improved by training more data in the model. Data can be acquired from MN Department of Transportation

## Data requirements

* Restaurant Data: Information about the restaurants listed on Zomato, such as restaurant name, location, cuisine type, average pricing, opening hours, and other relevant attributes. This data helps identify the restaurants for which ratings are to be predicted. Date time. Weekday, month can be extracted from it.
* Restaurant Attributes: Additional attributes about the restaurants that may influence the ratings, such as the ambience, service quality, delivery options, hygiene standards, and other factors that users might consider while rating a restaurant.

These are the required parameters to feed into model.

## Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask, Anvil, and a few other libraries were used to build the whole model.



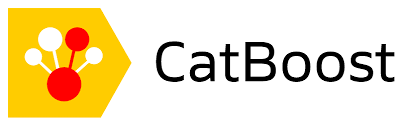












* For visualization tasks, matplotlib, seaborn and plotly were used
* For visualization tasks, matplotlib, seaborn and plotly were used
* Anvil and Flask were used for building the web application and server to run the code
* Apache Cassandra was used to storage and retrieval of data
* GitHub is used as version control system
* NumPy and Pandas were used to clean and interpret data
* Scikit-learn was used to cross validate and compare different models
* Cat Boost Regressor was used to build the final model

## Hardware Requirements

* Windows Server, Linux, or any operating system that can run as a webserver, capable of delivering HTML5 content.
* Minimum 1.10 GHz processor or equivalent.
* Between 1-2 GB of free storage
* Minimum 512 MB of RAM
* 3 GB of hard-disk space

## Constraints

The front-end must be user friendly and should not need any one to have any prior knowledge in order to use it.

## Assumptions

The main objective of this project is to implement the use case as previously mentioned (2.3 problem statement) for new dataset that comes through the UI. It is assumed that all aspects of this project have the ability to work together as the designer is expecting and also the data on which our model is trained is as correct as possible

# Design Details

## Process Flow

For accomplishment of the task, we will use a trained Machine Learning model. The process flow diagram is shown below:

**Data Preparation**

**Model**

**Development**

**Deployment**

**Deployment**

## Event Log

The system should log every event so that the user will know what process is running internally. Initial step-by-step description: 1. The system identifies at what level logging is required 2. The system should be able to log each and every system flow 3. Developer can choose logging method. You can choose database logging/ File logging as well 4. System should not hang even after so many loggings. Logging just because we can easily debug issues, so logging is mandatory to do.

## Error Handling

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

## Reusability

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

## Application Compatibility

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

## Resource Utilization

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





# Dashboards

As and when, the system starts to capture the historic/ periodic data for a user, the dashboards will be included display charts over time with progress on various indicators or factors.



## KPIs (Key Performance Indicators)

* Accuracy: This KPI measures how accurately the model predicts the restaurant ratings. It is calculated by comparing the predicted ratings with the actual ratings and determining the percentage of correct predictions.
* Response Time: This KPI measures the time taken by the predictive system to generate a rating prediction once the required restaurant attributes are provided. A lower response time indicates a more efficient and user-friendly system.
* User Satisfaction: Gathering feedback from users who utilize the predictive system can provide insights into their satisfaction levels. Surveys or user ratings can be used to assess the usefulness and accuracy of the predicted ratings and overall user experience.

# Conclusion

All in all, overall project architecture, design details, used technologies and performance were explained in detail. In summary, the Zomato restaurant rating prediction project contributes to enhancing the usability and effectiveness of the Zomato platform by providing users with a tool to estimate restaurant ratings. It assists users in making informed decisions, promotes transparency, and enhances the overall dining experience for Zomato users.