Low Level Design (LLD)

Restaurant Rating Prediction

Revision Number: 1.5
Last date of revision: 18/08/2021

Mohamed Illiyas I

Document Version Control

Date Issued	Version	Description	Author
8th August 2020	1.1	First Draft	Mohamed Illiyas
15th August 2020	1.2	Added Workflow chart	Mohamed Illiyas
16th August 2020	1.3	Added Exception Scenarios Overall, Constraints	Mohamed Illiyas
17th August 2021	1.4	Added user I/O flowchart, test cases	Mohamed Illiyas
18th August 2021	1.5	Added KPIs	Mohamed Illiyas

Contents

Abstract

Restaurant Industry is a highly growing sector all over the world. After online ordering facilities came into existence, people became excited and so the demand for restaurants keeps on increasing day by day. In India, Bangalore is one of the finest places with a wide variety of cuisines. As the number of restaurants has increased, it became difficult for the peoples to select particular good restaurant. So, the restaurant rating has become the most common used parameter for judging a restaurant. Rating of a restaurant depends on factors like reviews, area, locality, food, cuisines, average cost, etc. This project aims to understand the factors affecting the rating of restaurants. Machine learning model aids to predict the rating of restaurants. The model is built using various regression algorithms and ensemble techniques. The result of the project helps new restaurant in deciding their menu, cuisines, cost, type of foods, location, ambience, etc which will increase their business.

.

1 Introduction

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Restaurant Rating Prediction system. It will explain the purpose and features of the system, the interfaces of the system, what the prediction system will do, how the system will react to new external data. It describes the modules so that the programmer can directly code from the document. This document is intended for both the stakeholders (restaurant owners) and the developers of the system.

The main objective of the project is to predict the rating of the restaurant from the attributes like location, cost, cuisines, etc.

This project shall be delivered in two phases:

Phase 1: All the functionalities with PyPi packages.

Phase2: Integration of UI to all the functionalities.

1.2 Scope

This software system will be a Web application This system will be designed to predict the ratings out of 5 for the restaurants based on the input given by the user in the respective fields. More specifically, this rating will help the new restaurant owners to determine which food to be placed in their menus and they can utilize this for the improvement of their business and management. This system is designed to predict the ratings from Bangalore Zomato restaurants information such as demographics, cuisines, average cost, etc

1.3 Constraints

We will only be selecting restaurants in Bangalore.

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

2 Technical specifications

2.1 Dataset

url	address	name	online_order	book_table	rate	votes
https://www.zomato.com/bangalore/jalsa-banashankari?context=eyJzZSI6eyJlIjpbNTg2OTQsIjE4Mzc1NDc0IiwiI...	942, 21st Main Road, 2nd Stage, Banashankari, Bangalore	Jalsa	Yes	Yes	4.1/5	775
https://www.zomato.com/bangalore/spice-elephant-banashankari?context=eyJzZSI6eyJlIjpbIjU4Njk0IiwxODM...	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th Block, Kathriguppe, 3rd Stage, Banashankari, Bangalore	Spice Elephant	Yes	No	4.1/5	787
https://www.zomato.com/SanchurroBangalore?context=eyJzZSI6eyJlIjpbIjU4Njk0IiwiMTgzNzU0NzQiLCJ5MDkwLC.. .	1112, Next to KIMS Medical College, 17th Cross, 2nd Stage, Banashankari, Bangalore	San Churro Cafe	Yes	No	3.8/5	918
https://www.zomato.com/bangalo	1st Floor, Annakuteera,	Addhuri Udupi Bhojana	No	No	3.7/5	88

2.1.1 Restaurant dataset overview

Consists of a different table. Data table consists of the restaurant's public information such as address, phone number and most importantly we have the ratings given by customers on various orders.

It has 56000+ records with 17 features.

- 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 neighbourhood 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 neighbourhood in which the restaurant is listed

2.1.2 Input schema

Feature name	Datatype	Null/Required
Online order	String	Required
Book Table	String	Required
Votes	int	Required
Rest Type	String	Required
Cuisine	String	Required
Cost	int	Required
Location	String	Required
City	String	Required
Type of meal	String	Required

2.2 Predicting Ratings

- The system displays the choices of the restaurant types, meal types, cuisines.
- The User chooses the location and city.
- The User also selects whether the restaurants have online orders and table facilities.
- Then user gives required information such as cost and number of votes.
- The system presents these set of inputs to the user.
- The system then predicts that the rating of the restaurant given the above inputs.

2.3 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.4 Deployment

1. Azure Cloud

3 Technology stack

Front End	HTML/CSS/JS/
Backend	Python Flask
Deployment	Microsoft Azure Cloud

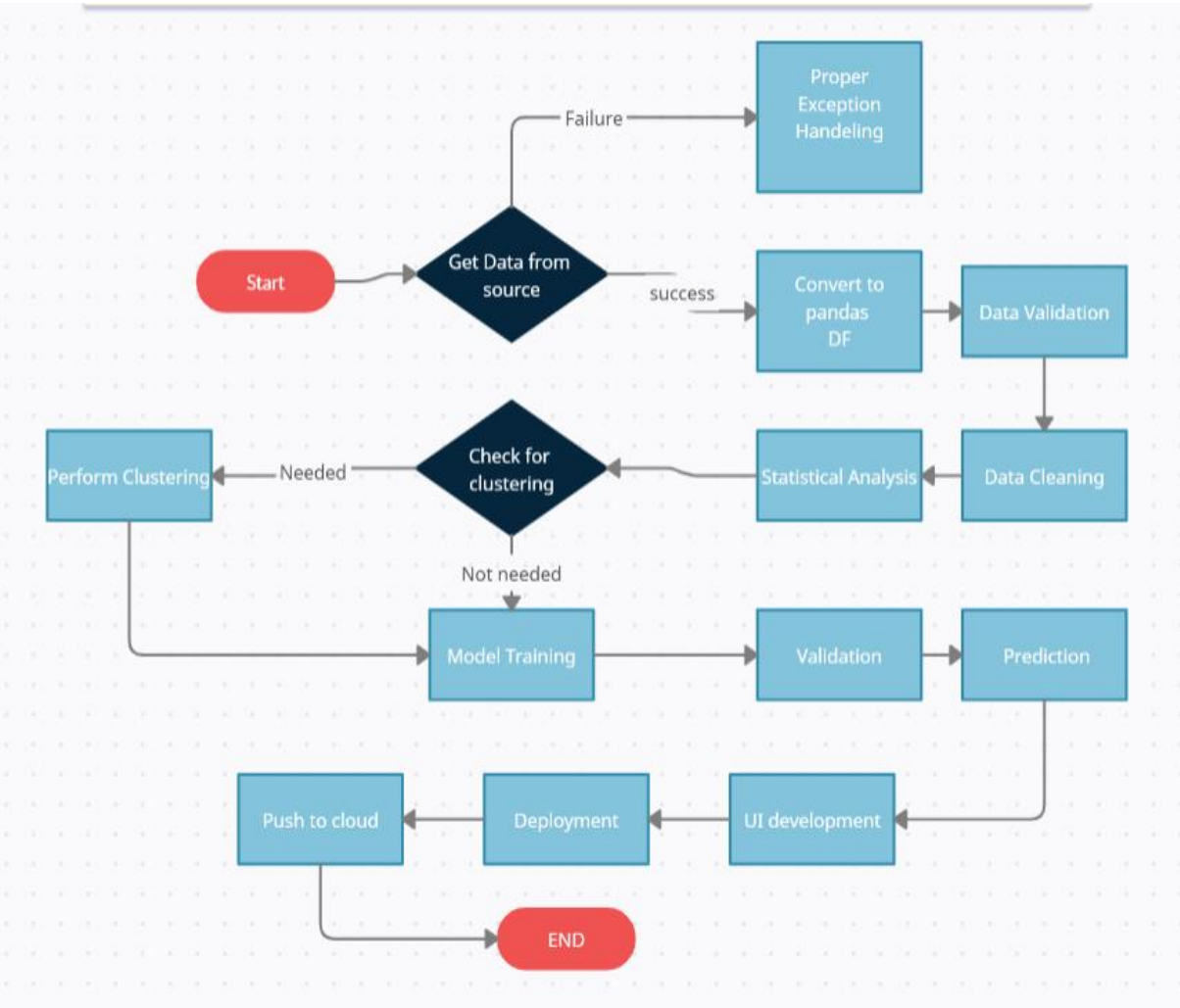
4 Proposed Solution

The proposed solution for this project is to find the ratings of the new restaurants by using Machine learning models. As the target is continuous variable, we can use regression or decision tree models as the base model for this problem. From this baseline model, we can improve the performance by ensemble techniques.

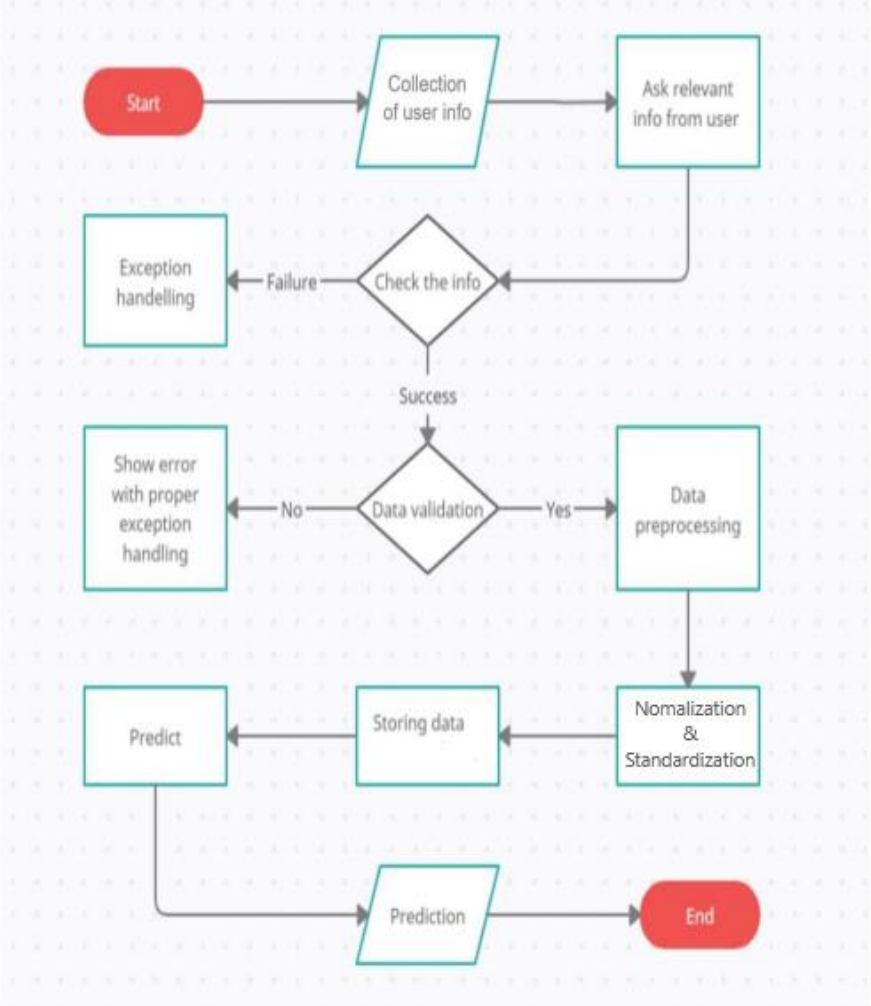
However, drawing a baseline model is important since it tells us how well other models have performed compared to base model. Here, the base model for Restaurant Rating dataset is Linear Regression.

1. Baseline Model: Linear regression
2. Actual model: Random Forest

5 Model training/validation workflow



6 User I/O workflow



7 Exceptional scenarios

Step	Exception	Mitigation	Module
8th August 2021	1.1	First Draft	Mohamed Illiyas
15th August 2021	1.2	Added Workflow chart	Mohamed Illiyas

8 Test cases

Test case	Steps to perform test case	Pass/Fail
Verify whether user is able to see input	1. Application is accessible 2. Check whether inputs available to user	Pass
Verify whether user is able to edit all input fields	1. Check whether inputs available to user 2. Check whether the inputs are editable or not.	Pass
Verify whether user gets Submit button to submit the inputs	1. Submit button functionality check	Pass
Verify whether user is presented with predicted results on clicking submit	1. Submit button functionality check 2. Redirecting to the output page	Pass

9 Key performance indicators (KPI)

- Comparison of R2 score of model prediction and real ratings.
- We observed that the base model gives 29% R2 score.
- After improvising the model with ensemble techniques, we get 81% R2 score