Architecture

Restaurant Rating Prediction

Anuj Gupta

Document Version Control

Date Issued	Version	Description	Author
15 June 2023	v1	First Draft	Anuj Gupta

Contents

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 Low-Level Design Document?

The primary objective of analyzing the Zomato dataset is to gain insights into the factors that influence the overall rating of restaurants. This analysis also aims to understand the diversity of restaurant establishments in different locations, with a particular focus on Bengaluru. With over 12,000 restaurants in Bengaluru, the city offers a wide range of culinary experiences, featuring cuisines from around the world.

Despite the increasing demand for restaurants, the industry in Bengaluru remains highly competitive, making it challenging for new establishments to establish themselves. Many restaurants serve similar types of food, leading to intense competition. In this context, understanding the factors that contribute to a restaurant's success becomes crucial.

Bengaluru, being the IT capital of India, has a significant population that heavily relies on restaurant food due to time constraints. Consequently, studying the demography of different locations becomes essential. Analyzing the popularity of various types of food in specific localities can provide valuable insights. For instance, understanding if a particular locality predominantly prefers vegetarian food can help identify the demographic composition of that area, such as Jain, Marwari, or Gujarati communities who generally follow a vegetarian diet. By conducting such analyses using the available data, we can gain a deeper understanding of the preferences and characteristics of different localities.

By studying various factors, such as cuisine preferences, customer reviews, restaurant ratings, and demographic information, we can uncover meaningful patterns and insights. This analysis can help stakeholders in the restaurant industry make informed decisions regarding menu offerings, target demographics, and market positioning.

In summary, the analysis of the Zomato dataset allows us to explore the factors influencing restaurant ratings and gain insights into the diverse culinary landscape of Bengaluru. By examining these factors, we can better understand customer preferences, demographic characteristics, and the competitive dynamics within the restaurant industry.

1.2 Scope

This software system will be a Web application. This system will be designed to predict therating of the restaurant based on the input by the user.

1.3 Constraints

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

1.4 Risks

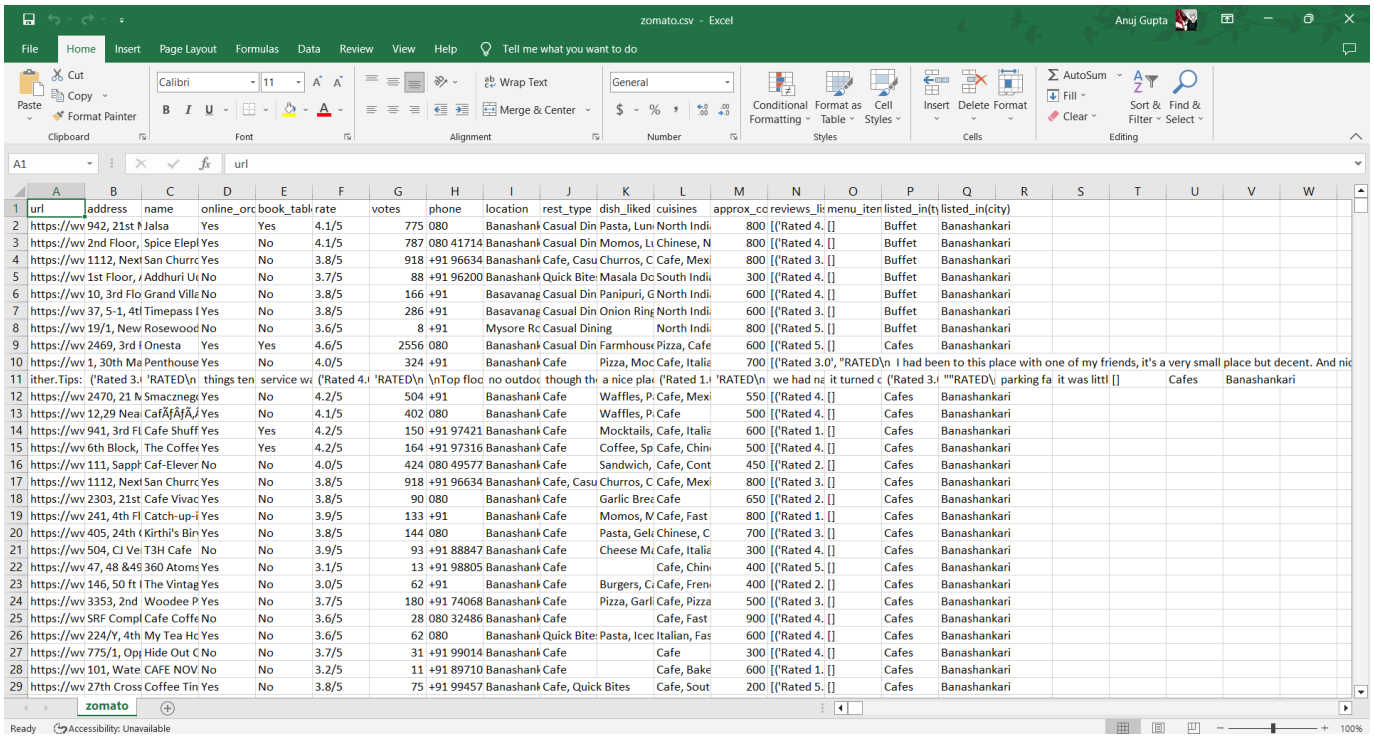
Document specific risks that have been identified or that should be considered.

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



2.1.1 Dataset overview

Columns description

- 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
- dish_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: list of tuples containing reviews for the restaurant, each tuple
- 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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   url                          51717 non-null  object
 1   address                      51717 non-null  object
 2   name                         51717 non-null  object
 3   online_order                 51717 non-null  object
 4   book_table                   51717 non-null  object
 5   rate                         43942 non-null  object
 6   votes                        51717 non-null  int64
 7   phone                        50509 non-null  object
 8   location                     51696 non-null  object
 9   rest_type                    51490 non-null  object
 10  dish_liked                   23639 non-null  object
 11  cuisines                     51672 non-null  object
 12  approx_cost(for two people)  51371 non-null  object
 13  reviews_list                 51717 non-null  object
 14  menu_item                    51717 non-null  object
 15  listed_in(type)              51717 non-null  object
 16  listed_in(city)              51717 non-null  object
dtypes: int64(1), object(16)
memory usage: 6.7+ MB
```

2.1.2 Input schema

Feature name	Null/Required
Online order	Required
Book Table	Required
Rest Type	Required
Cuisine	Required
Cost	Required
location	Required
Listed_in(type)	Required

2.2 Predicting Rating

- The system presents the set of inputs required from the user.
- The user gives the required information.
- The system then predicts that the rating of the restaurant given the above inputs.

2.3 Deployment

- I used **Streamlit** for deployment.

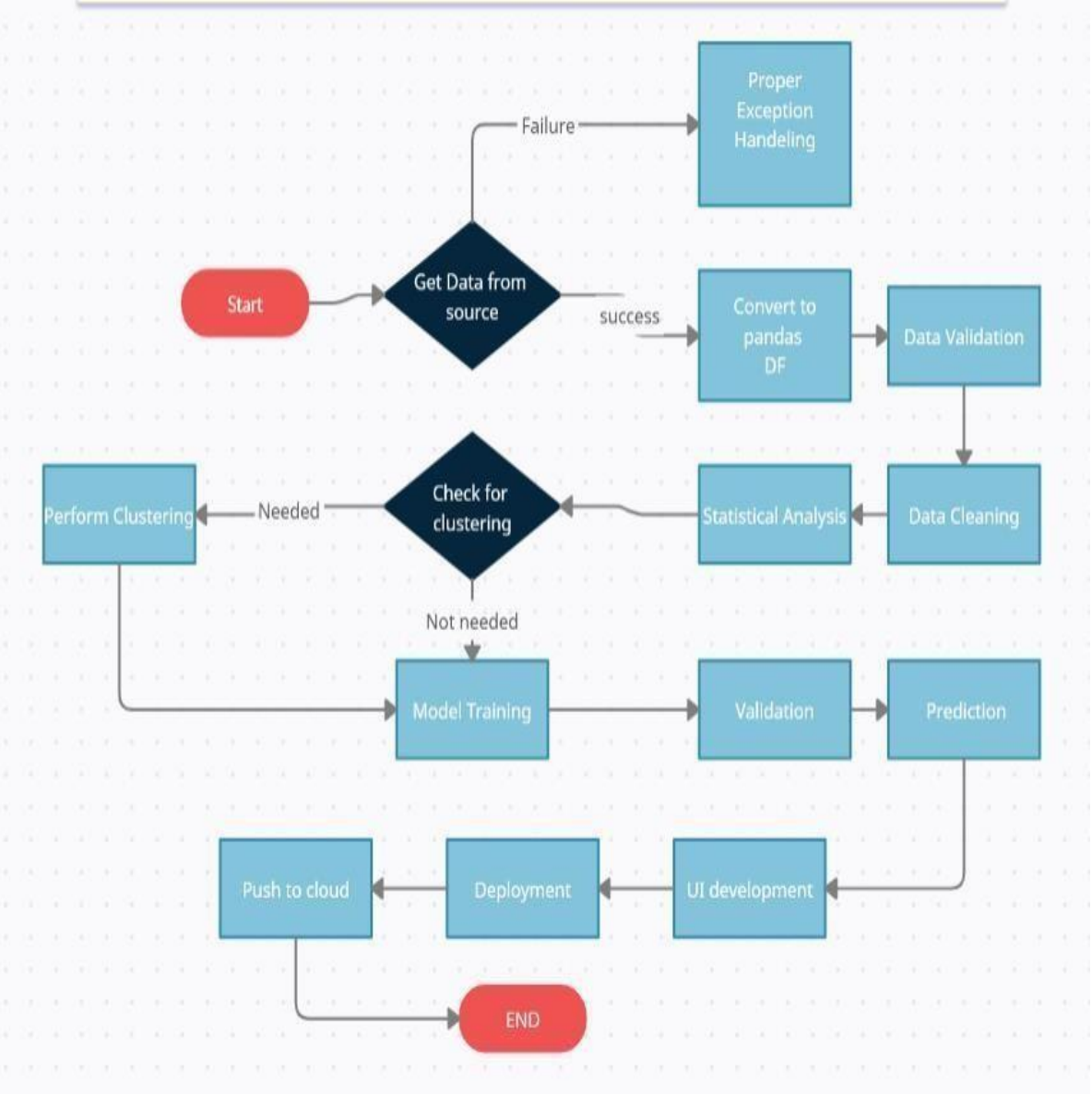
3 Technology stack

Front End	**Streamlit**
Backend	**Python, Streamlit**
Deployment	**Streamlit**

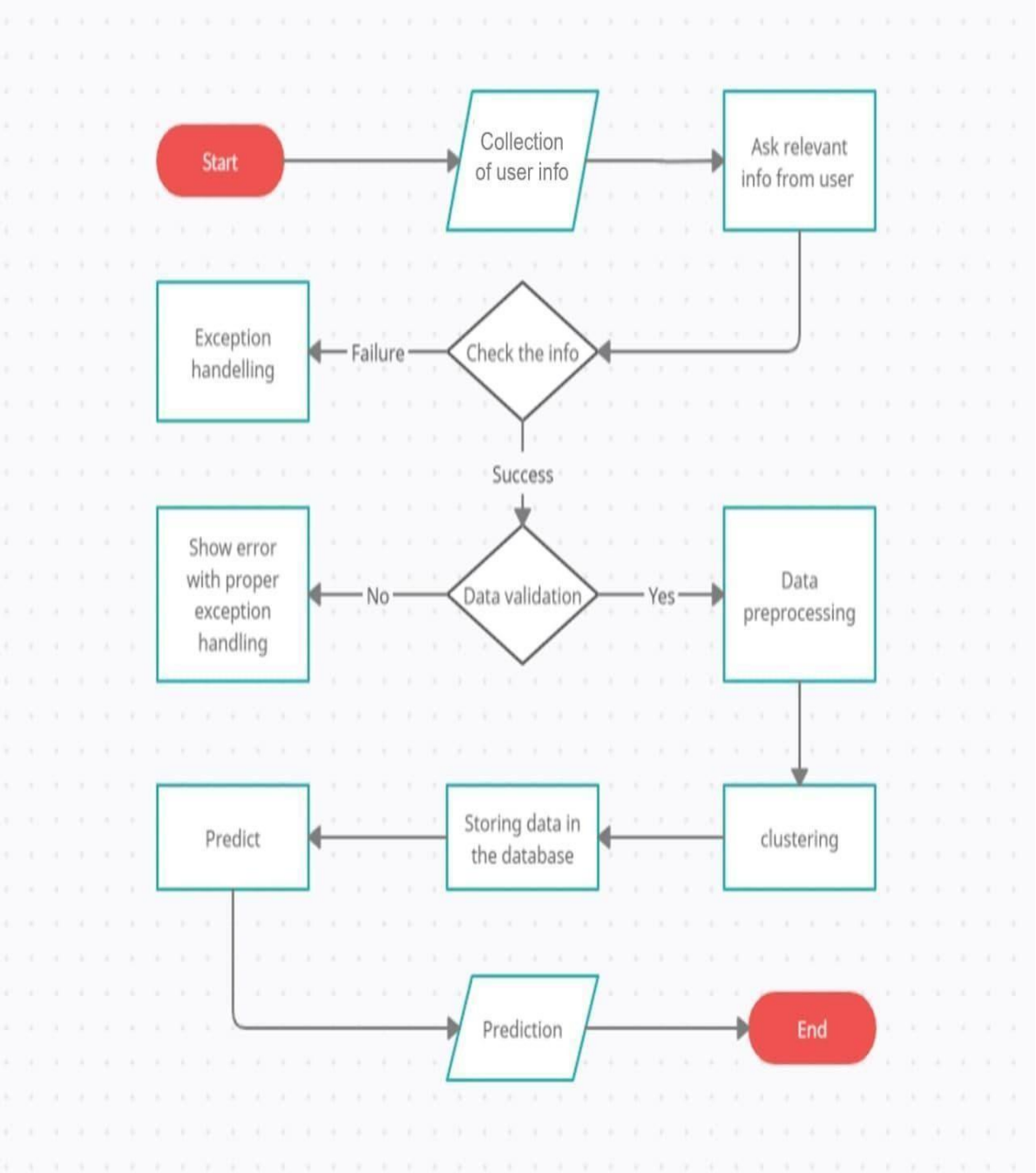
4 Proposed Solution

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, votes, rest type, cuisines, listed_in(type) as inputs from the web app, the implemented regression model will predict the output as a rating of the restaurant.
Here we tried different algorithms such as Random Forest, XGBoost, LGBM. The LGBM regressor gave us the best results.

5 Model training/validation workflow



6 User I/O workflow



7 Exceptional scenarios

Step	Exception	Mitigation	Module
15 June 2023	1.1	First Draft	Anuj Gupta

8 Performance

- The LGBM regressor gave us the best results and performed good.