Low-Level Document

Restaurant Rating Prediction

Revision Number: 1.0
Last date of revision: 30 Jan 2023
Tushar Jain

Document Version Control

Date Issued	Version	Description	Author
30/01/2023	1.1	Finial Version	Tushar Jain

Abstract

This project aims to perform an extensive Exploratory Data Analysis (EDA)on the Zomato dataset and build an appropriate Machine Learning Modelto predict the ratings of Zomato Restaurants based on certain features. The project follows the classical machine learning process, including Data Exploration, Data Cleaning, Feature Engineering, Model Building, and Model Testing. Different machine learning algorithms will be tested to determine the best fit for the case. The ultimate goal is to build a solution that can accurately predict the ratings of the restaurants listed in the Zomato dataset.

Table of Contents

1. Introduction

1.1. Why this Low-Level Document?

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant, the establishment of different types of the restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry hasn't been saturated yet and the demand is increasing day by day. In spite of increasing demand, it however has become difficult for new restaurants to compete with established restaurants. Most of them serve the same food. Bengaluru is the 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 for restaurants, it has therefore become important to study the demography of a location. What kind of food is more popular in a locality. Does the entire locality loves vegetarian food. If yes, then is that locality populated by a particular set of people for eg. Jain, Marwaris, Gujaratis who are mostly vegetarian. This kind of analysis can be done using the data, by studying different factors.

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 rating of the restaurant based on the input by the user.

1.3. 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.

1.4. Risks

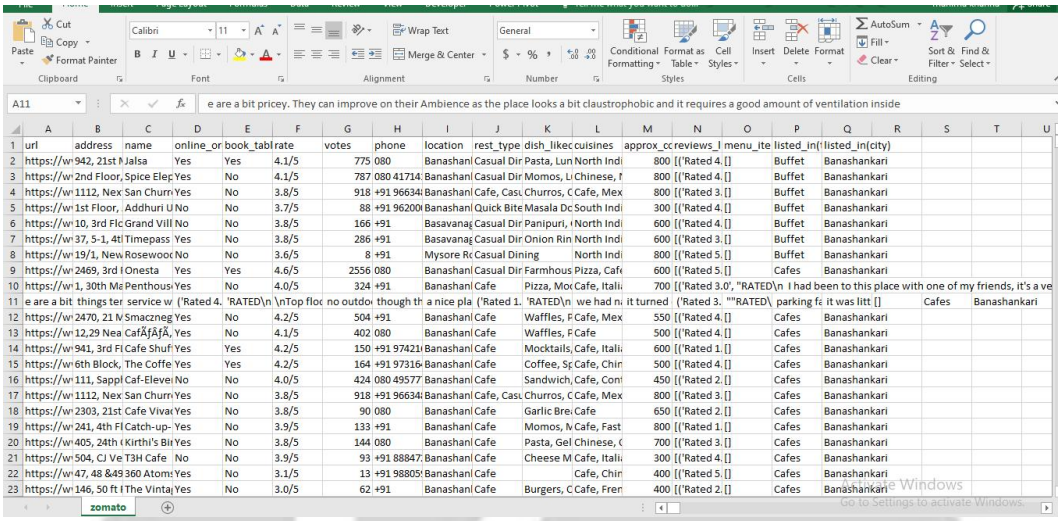
Document specific risks that have been identified or that should be considered.

1.5. Out of Scope

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

2. Technical Specification

2.1. Datasets



2.1.1. Datasets overview

The dataset consists of a table with 56351 records and 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 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

2.1.2. Input Schema

Feature Name	Null/Required
Online Order	Required
Book Table	Required
Votes	Required
Location	Required
Rest Type	Required
Cost	Required
Cuisine	Required
Type	Required

2.2. Predictive Rating

- The system presents the set of inputs required from the user.
- The user gives required information.
- 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

3. Technology Stack

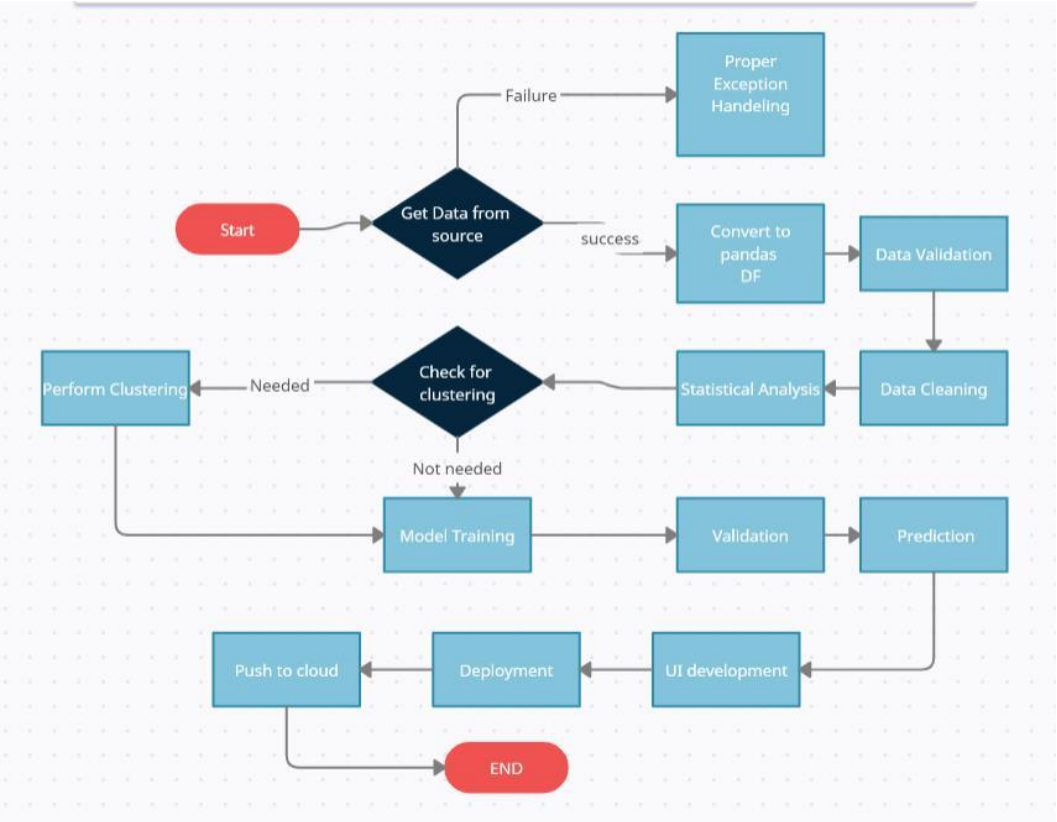
Web App	Streamlit
Machine Learning Model	Python/Jupyter Notebook
Deployment	Git/Heroku

4. Proposed Solution

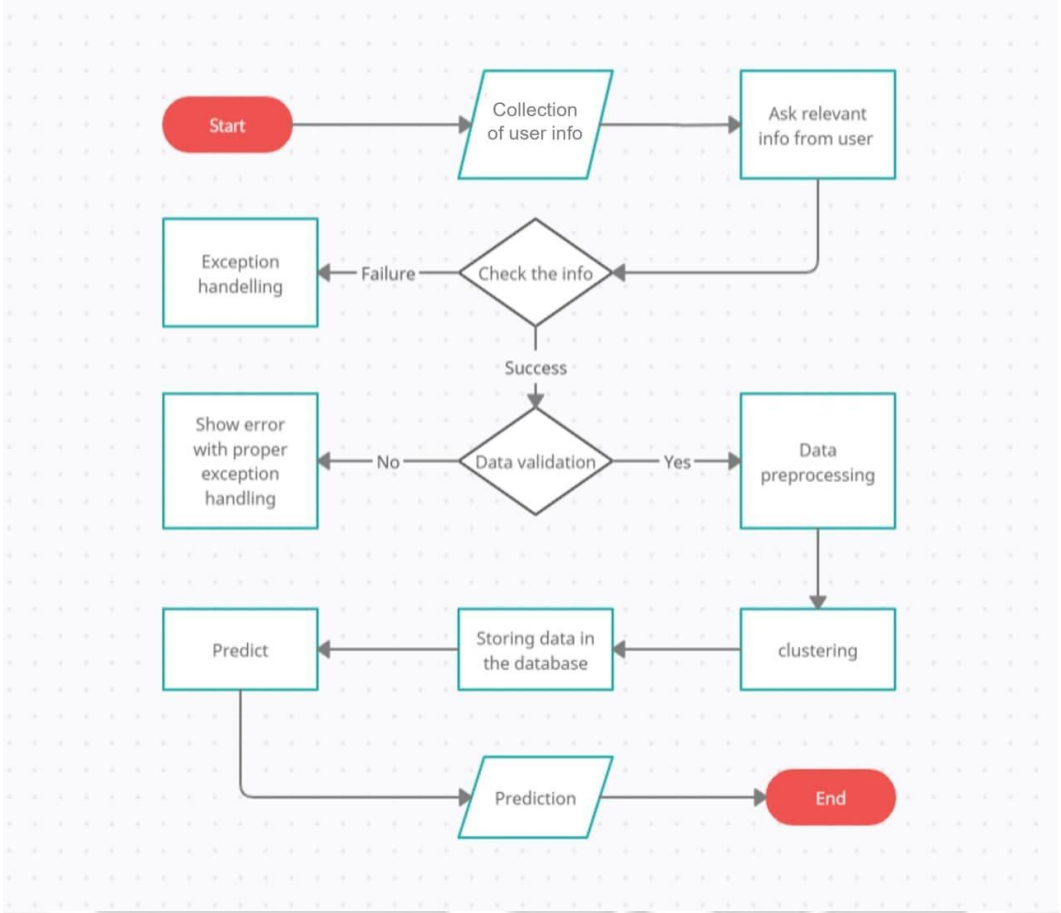
The proposed solution for this project is Machine learning algorithms can be implemented to predict the rating of the restaurant. Considering various features like online order, book table, votes, rest type, cuisines, review as inputs from the web app, the implemented classification model will predict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

5. Model Training/ Validation workflow



6. User I/O workflow



7. 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.