HIGH LEVEL DESIGN (HLD)

Restaurant Rating Prediction Based Machine Learning Models

Revision No.: 1.0

Last Date of Revision: 09/02/2023

Document Version Control

2

Date Issued	Version	Description	Author
09/02/2023	1	Initial HLD - V1.0	Amar G
15/02/2023	2	Updated HLD – V1.1	Rutuja D

Contents

Abstract

Restaurant Rating has become the most commonly used parameter for judging a restaurant for any individual. A lot of research has been done on different restaurants and the quality of food it serves. Rating of a restaurant depends on factors like reviews, area situated, average cost for two people, votes, cuisines and the type of restaurant.

The main goal of this is to get insights on restaurants which people like visit and to identify the rating of the restaurant. With this project we study different predictive models like Support Vector Machine (SVM), Random forest and Linear Regression, Ridge Regression and have achieved a score of 92.5% with Random Forest Regression.

1. Introduction

Zomato is the most reputed company in the field of food reviews. Founded in 2008, this company started in India and now is in 24 different countries. It is so big that the people now use it as a verb. "Did you know about this restaurant? Zomato it". The rating is the most important feature of any restaurant as it is the first parameter that people look into while searching for a place to eat. It portrays the quality, hygiene and the environment of the place. Higher ratings lead to higher profit margins. Notations of the ratings usually are stars or numbers scaling between 1 and 5.

Zomato has changed the way people browse through restaurants. It has helped customers find good places with respect to their dining budget.

Different machine learning algorithms like SVM, Linear regression, Decision Tree, Random Forest can be used to predict the ratings of the restaurants.

1.1 Scope

We can use better machine learning algorithm to analyse the sarcastic data.

This product can be used in software's like zomato to find out the ratings of the particular restaurant products.

It's a tool to find out the feedback for restaurants for self-improvement.

2. General Description

2.1 Problem Statement

The main goal of this project is to perform extensive Exploratory Data Analysis (EDA) on the Zomato Dataset and build an appropriate Machine Learning Model that will help various Zomato Restaurants to predict their respective Ratings based on certain features.

2.2 Approach

The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Try out different machine a learning algorithm that's best fit for the above case.

2.3 Results

We have to build a solution that should able to predict the ratings of the restaurants listed in the dataset.

2.4 Dataset Description

This is a kaggle dataset.

(https://www.kaggle.com/himanshupoddar/zomato-bangalorerestaurants).

It represents information of Restaurants in the city of Bangalore. It contains 17Columns and 51,000 Rows

2.5 Technical Requirements

2.5.1 Hardware Requirements

- System: Pentium IV 2.4 GHz.
- Hard Disk: 500GB
- RAM: 4GB
- Any desktop / laptop system with above configuration or higher level.

2.5.2 Software Requirements

- Operating System: Windows
- Front End: HTML5, CSS3
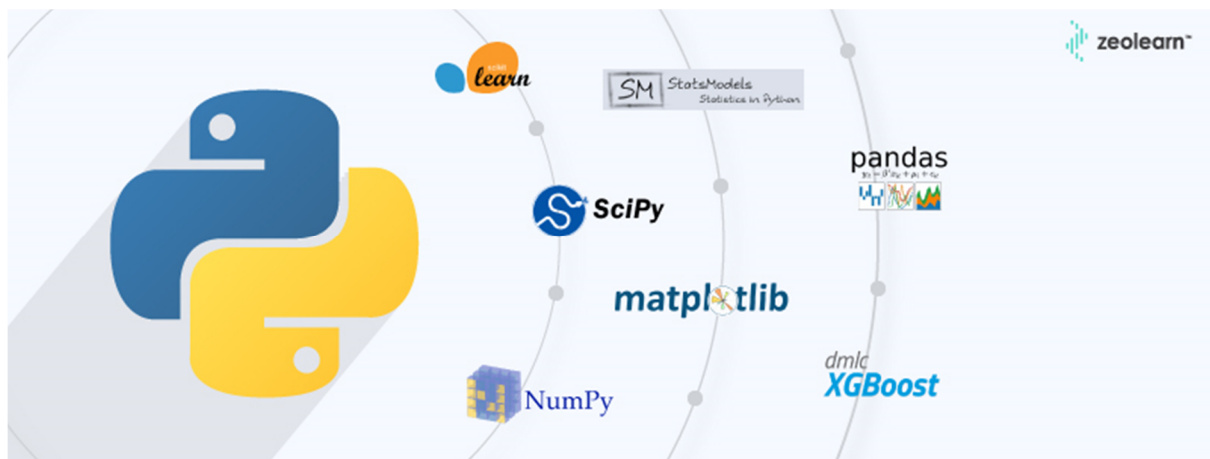- Back End: Python
- Dataset: Kaggle Dataset (.csv or .tsv files)

2.5.3 Language Specification

To develop the prediction model the following programming languages are used:
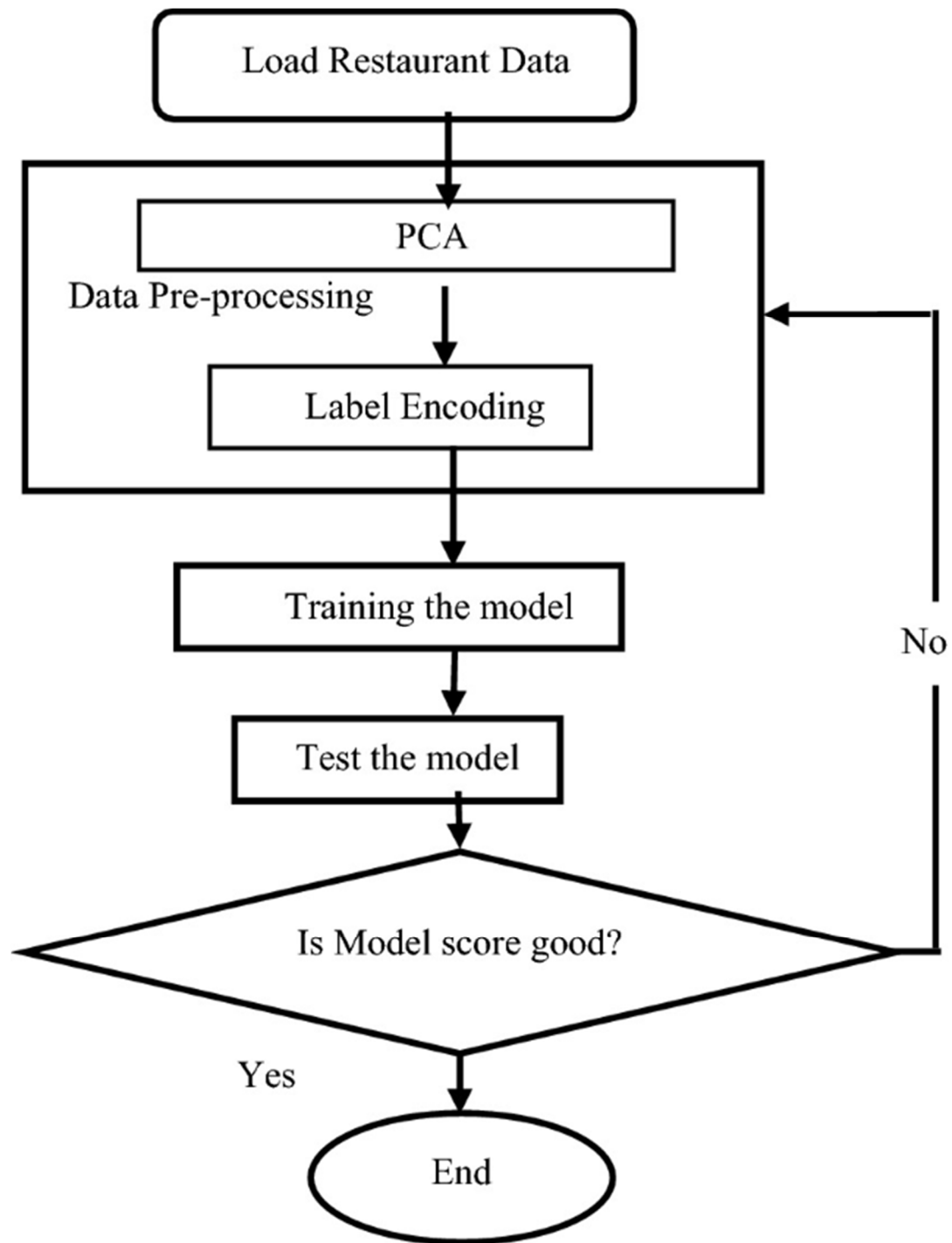
- Python
- SQL

2.5.4 Tools Used
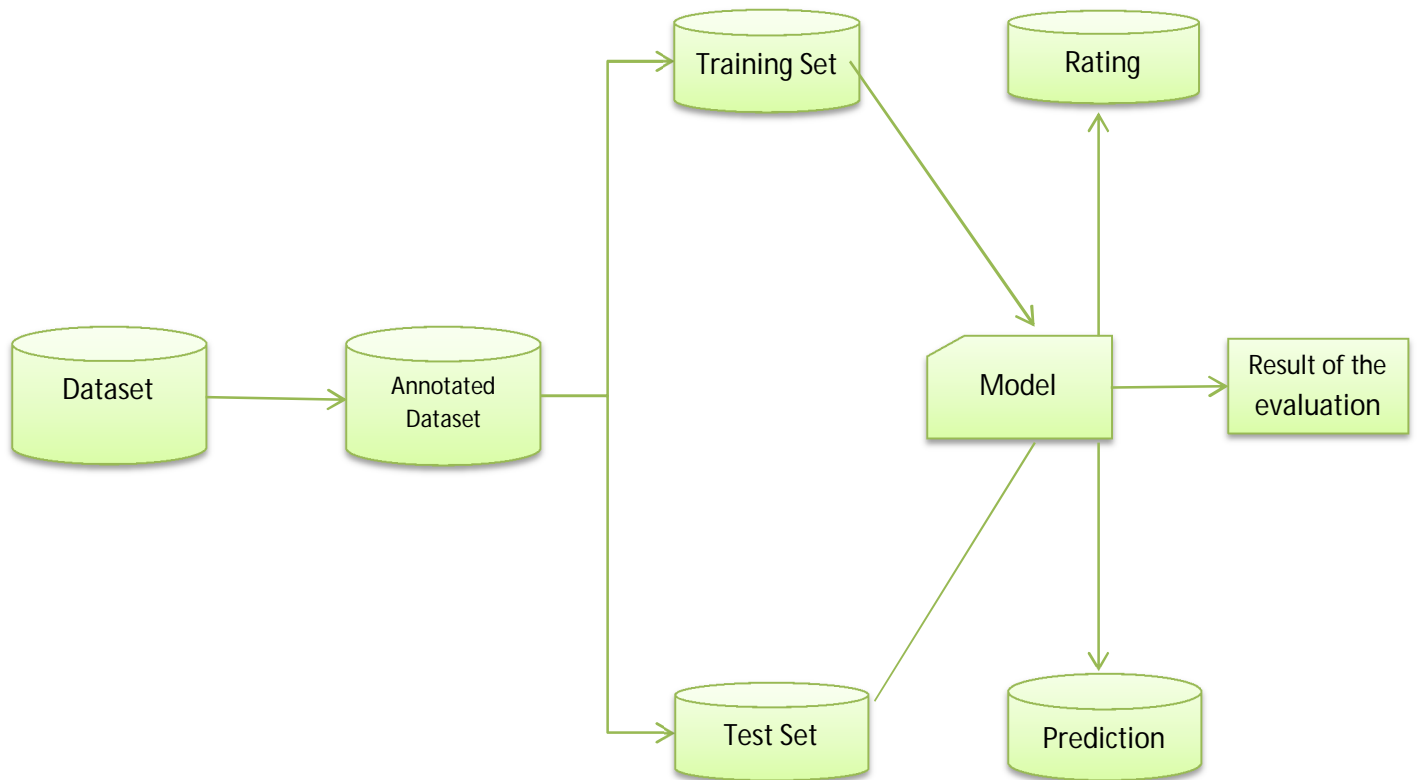
- Numpy
- Pandas
- Scikit learn
- Matplotlib
- Seaborn

3. Design Details

3.1 Process Flow:

Proposed Methodology

3.2 Model Training and Evaluation:

4. Data Pre-processing

4.1 PreProcessing

The Dataset contained 17 Attributes.

• Records with null values were dropped from ratings columns and were replaced in the other columns with a numerical value.

• Values in the 'Rating' column were changed. The '/5' string was deleted. For eg. If the rating of a restaurant was 3.5/5, it was changed to 3.5.

• Using LabelEncoding from sklearn library, encoding was done on columns like book_table,online_order,rest_type,listed_in(city).

4.2 Feature Selection

We did not use any feature selection algorithms but eliminated some columns due to available domain knowledge and thorough study of the system. Dropped columns mentioned below:

• URL

• Address

• Dish_liked

• Phone

• Menu

• Review_list

• Location

• Cuisine

Some of these columns may look like they are important but all of the same information could be found in other columns with lesser complexity.

The Columns being used are as follows:

• Name

• Online_order

• Book_table

• Votes

• Rest_type

- Approx. cost of two people

- Listed_in(type)

- Listed_in(city)

5. EXPLORATORY DATA ANALYSIS

A lot of effort went into the EDA as it gives us a detailed knowledge of our data. Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

- maximize insight into a data set;

- uncover underlying structure;

- extract important variables;

- detect outliers and anomalies;

- test underlying assumptions;

- develop parsimonious models; and

- determine optimal factor settings.
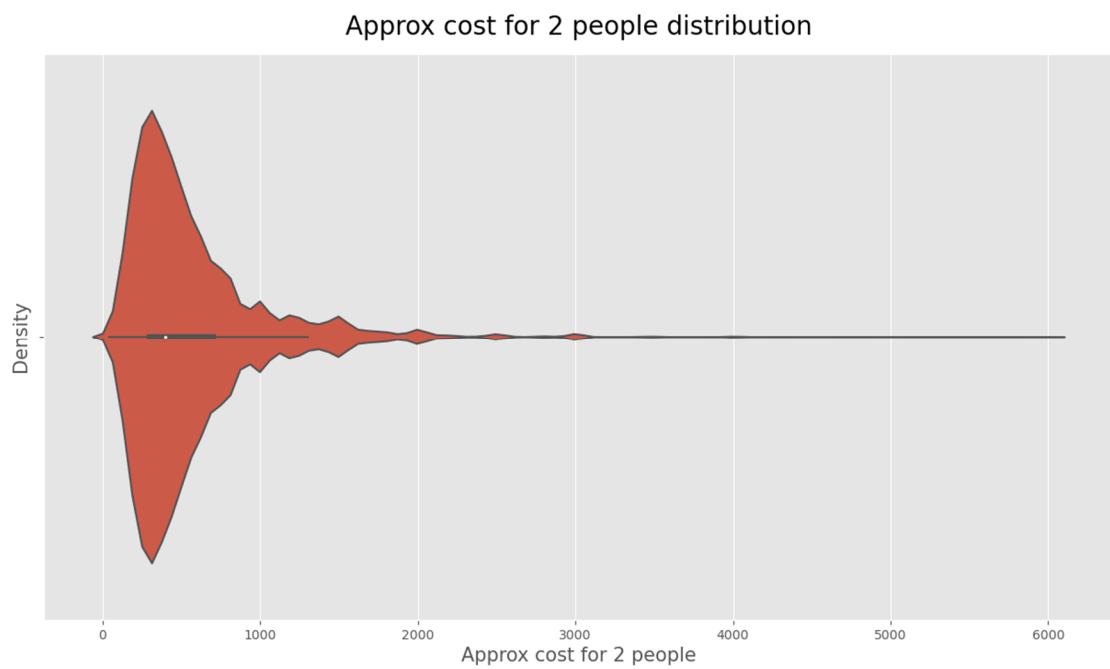
A) Distribution of Rate Column

Average Rating For Restaurants
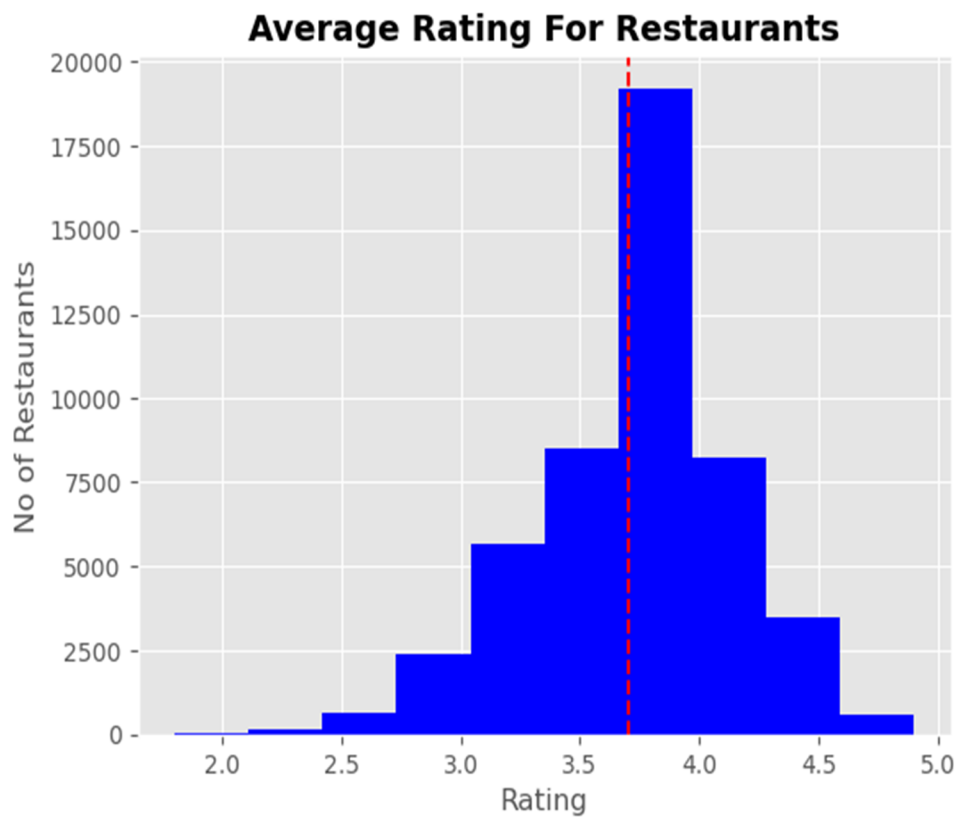
B) Restaurants with maximum number of orders

Restaurants with Maximum No of orders

13
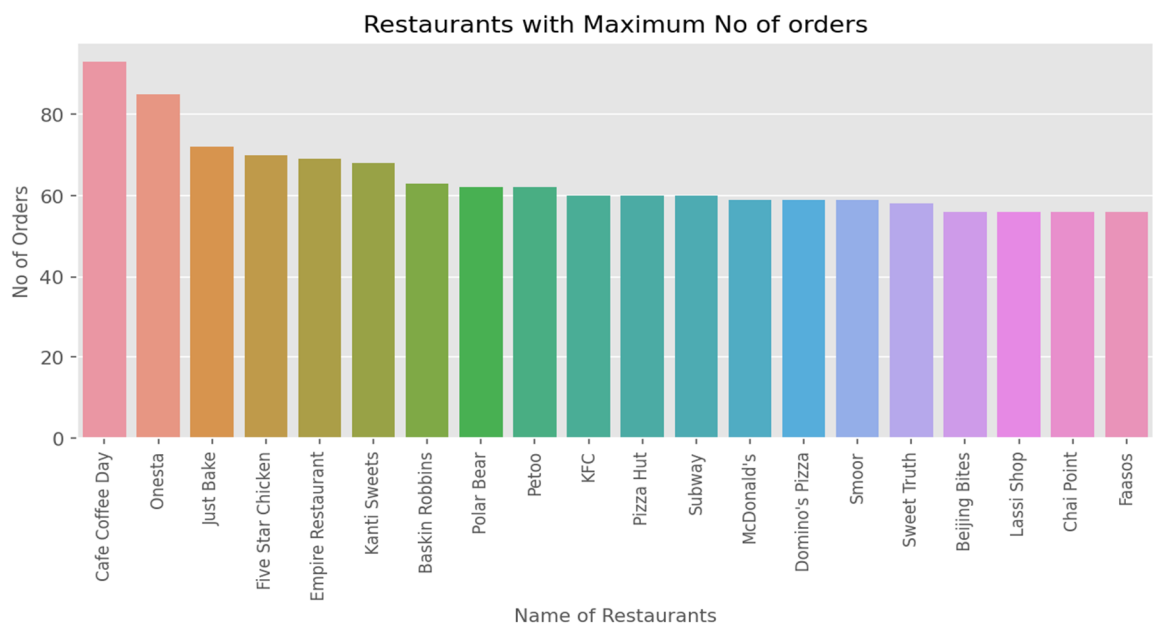
c) Approximate cost for two peoples

Approx cost for 2 people distribution

14

6. Results:

Algorithms	Accuracy
Linear Regression	27.21
Ridge Regression	27.21
Random Forest Regression	92.53
Support Vector Regression	40.33

In this model, we have considered various restaurants records with features like the name, average cost, locality, whether it accepts online order, can we book a table, type of restaurant. This model will help business owners predict their rating on the parameters considered in our model and improve the customer experience.

Here, we can see that the score of Random forest regression Algorithm is higher compared to other. Hence, we used Random Forest Regression Algorithm to predict restaurant ratings.

```
3. Random Forest Regressor

In [287]: from sklearn.ensemble import RandomForestRegressor

In [288]: rfr = RandomForestRegressor()
          rfr

Out[288]: RandomForestRegressor()

In [289]: rfr.fit(x_train,y_train)

Out[289]: RandomForestRegressor()

In [290]: print(rfr.score(x_train,y_train))
          print(rfr.score(x_test,y_test))

          0.9894347179832386
          0.9253274191497383

In [291]: rfr_pred = rfr.predict(x_test)

In [292]: from sklearn.metrics import r2_score

          rfr_score = r2_score(y_test,rfr_pred)*100
          print("Accuracy score for LR :",rfr_score)

          Accuracy score for LR : 92.53274191497383
```

7. Conclusions:

This project studies a number of features about existing restaurants of different areas in a city and analyses them to predict rating of the restaurant. This makes it an important aspect to be considered, before making a dining decision. Such analysis is essential part of planning before establishing a venture like that of a restaurant.

Lot of researches have been made on factors which affect sales and market in restaurant industry. Various dine-scape factors have been analysed to improve customer satisfaction levels.

If the data for other cities is also collected, such predictions could be made for accurate.

8. Preferences:

- https://colab.research.google.com
- https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants
- https://www.tutorialspoint.com/machine_learning_with_python/machine_learning _with_python_classification _algorithms_random_forest.htm
- Atharva Kulkarni,Divya Bhandari,Sachin Bhoite.A study of Restaurants Rating Prediction using Machine Learning Algorithms. International Journal of Computer Applications Technology and Research, 2019, p.377- 378.DOI: 10.7753/IJCATR0809.1008