

HIGH LEVEL

DOCUMENT

**Restaurant Rating PREDICTION**

**January 23, 2023**

***Written by Sofana Benoutiq & Othmane Zoubairi***

**CONTENTS**

**DOCUMENT VERSION CONTROL**

**3 ABSTRACT**

**4 1) INTRODUCTION**

**5 1.1 WHY THIS HIGH-LEVEL DESIGN DOCUMENT?**

**5 1.2 SCOPE**

**1.3 DEFINITIONS**

**GENERAL DESCRIPTION**

**6 2.1 PRODUCT PERSPECTIVE**

**6 2.2 PROBLEM STATEMENT**

**2.3 PROPOSED SOLUTION**

**2.4 TECHNICAL REQUIREMENTS**

**2.5 DATA REQUIREMENTS**

**7 2.6 TOOLS AND TECHNOLOGIES USED**

**8 2.7 CONSTRAINTS**

**9 3.1 PROCESS FLOW**

**3.2 DEPLOYMENT PROCESS**

**3.3 EVENT LOG**

**3.4 PERFORMANCE**

**10 3.5 REUSABILITY**

**3.6 RESOURCE UTILIZATION**

**11 3.7 DEPLOYMENT**

**11 3.8 USER INTERFACE**

**12  Conclusion**

2 | P a g e

**DOCUMENT VERSION CONTROL**

|  |  |  |  |
| --- | --- | --- | --- |
| Date  Issued | Version | Description | Author |
| 22/12/2022 | 1.0.1 | Initial HLD – V1.0.1 | Sofana Benoutiq  &  Othmane  zoubairi |
| 15/01/2023 | 1.0.2 | Updated HLD  – V1.0.2 | Sofana Benoutiq  &  Othmane  zoubairi |
| 23/01/2023 | 1.0.3 | Updated HLD  – V1.0.3 | Sofana Benoutiq  &  Othmane  zoubairi |

ABSTRACT

**The basic idea of analyzing dataset is to get a fair idea about the factors affecting the establishment of different types of the restaurant at different places in different Cities, aggregate rating of each restaurant, With each day new restaurants opening the industry hasn't been saturated yet and the demand is increasing day by day.**

**Inspite of increasing demand it, however, has become difficult for new restaurants to compete with established restaurants.**

**Most of the people here are dependent mainly on the restaurant food as they don’t have time to cook for themselves. So that it es important to know which Restaurants in which Cities are good or not.**

4 | P a g e

  1) Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-level document (HLD) is to describe  the design of the project in detail which can be used as a  reference manual.

The HLD will:

• Present all the design aspects and define them in detail. • Describe the user interface being implemented. • Describe the software interfaces.

• Describe the performance requirements.

• Include design features and the architecture of the project.

1.2 Scope

The HLD document presents the entire structure of the project  in parts, such as the data ingestion, data pre-processing,  solution development, and the deployment part along with their  respective architectures. This uses non-technical to mild  technical terms which should be understandable to the

administrators of the system.

1.3 Definitions

|  |  |
| --- | --- |
| Term | Description |
| EDA | Exploratory Data Analysis |
| IDE | Integrated Development Environment |

5 | P a g e

2) General Description

2.1 Product Perspective

The Restaurant Rating Probability predictor is a machine  learning based on the use of GridSearchCV Random Forest Regressor model which will help us  to predict the probability Of restaurant rating based on the attributes of  the customer.

2.2 Problem Statement

The Restaurant Rating wants to enable and establish a viable business model to expand the customer base. One of the ways to expand the customer base is to introduce a new offerings. Base on Data Company we need to analyze

The Data of Clients to find who is going to give a good rating and who is going to give bad rating.

2.3 Proposed Solution

The solution proposed here is a web application, which takes  the details of the customer and those details will be taken by a  machine learning model in the backend, which will then predict the probability of good or bad rating and display it on the front-end page of  the user.

2.4 Technical Requirements

I used python version 3.7 with some important libraries to develop a machine learning model, which accurately predicts  the probability of restaurant ratings.

Then, the model is used as a back-end software for a front-end  web application which can be used by the users.

6 | P a g e

 2.5 Data Requirements

For training and testing the model, I used the public set  available in Ineuron intership, “

**URL;**

<https://raw.githubusercontent.com/mrinalmayank7/datascience/main/zom.csv>

**Dataset Attributes:**

Customer details:

1. Restaurant Id: Unique id of every restaurant across various cities of the world
2. Restaurant Name: Name of the restaurant
3. Country Code: Country in which restaurant is located
4. City: City in which restaurant is located
5. Address: Address of the restaurant
6. Locality: Location in the city
7. Locality Verbose: Detailed description of the locality
8. Longitude: Longitude coordinate of the restaurant’s location
9. Latitude: Latitude coordinate of the restaurant’s location
10. Cuisines: Cuisines offered by the restaurant
11. Average Cost for two: Cost for two people in different currencies (local currency)
12. Currency: Currency of the country
13. Has Table booking: yes/no
14. Has Online delivery: yes/ no
15. Is delivering: yes/ no
16. Switch to order menu: yes/no
17. Price range: range of price of food
18. Aggregate Rating: Average rating out of 5
19. Rating color: depending upon the average rating color
20. Rating text: text on the basis of rating of rating
21. Votes: Number of ratings given

2.6 Tools And Technologies Used



• VS CODE TO CREATE PROJECT

Ipynb file is used for EDA and experimentation  with various ML algorithms with the help of pandas,  numpy, matplotlib, seaborn, sklearn   libraries.

• Jupyter was also used for the development and  deployment of the solution with logging. Used python  version 3.7 and libraries include logging, pandas, numpy,  scikit learn, flask, and HTML

* Github is used as a version control system. ‘
* Using Git to have acces to github to make uptade evry time you needed
* Deployed on the web using Gunicorn and Heroku.

2.7 Constraints

The Concrete Compressive Strength Prediction system must be  user-friendly, errors free and users should not be required to  know about any of the workings.

8 | P a g e

 3 Event Log

In this project, I used the “logging” library in both the  development and deployment stages, which keeps logging the  events at every step into the “ restaurant\_rating\_logs” files. One of the advantages  of event logging is, it makes debugging much easier, we can  directly go to that specific line of code, which has errors.

3.1 Performance

The ML-based Probability Of restaurant rating application is  used for predicting the Probability Of restaurant rating based on various  attributes of the customer. So, it should be as accurate as  possible, so that it will not mislead the Data.

Model retraining is very important to keep it relevant in order  to keep the model dynamic to changing times and customer behaviour.

3.2 Reusability

The code written and the components used have the ability to  be reused without any problem.

3.3 Application Compatibility

The different components or modules of this project use python version 3.7 as their interface between them. Each component has its own task to perform, and it is the job of the python version to ensure proper transfer  of the information.

3.4 Resource utilization

In this project, any task may use all the processing power  available in the system, until it is accomplished.

3.5 Deployment

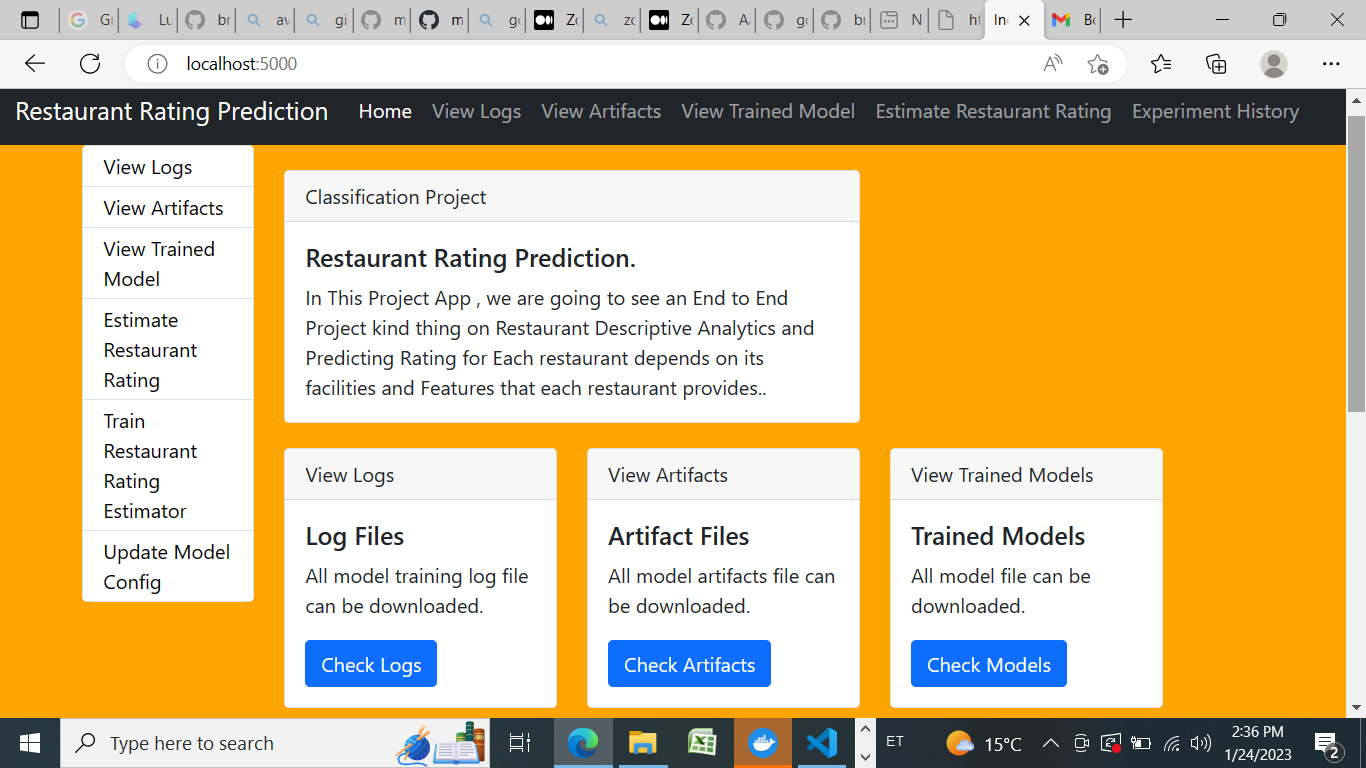
I deployed the application on the web using Heroku

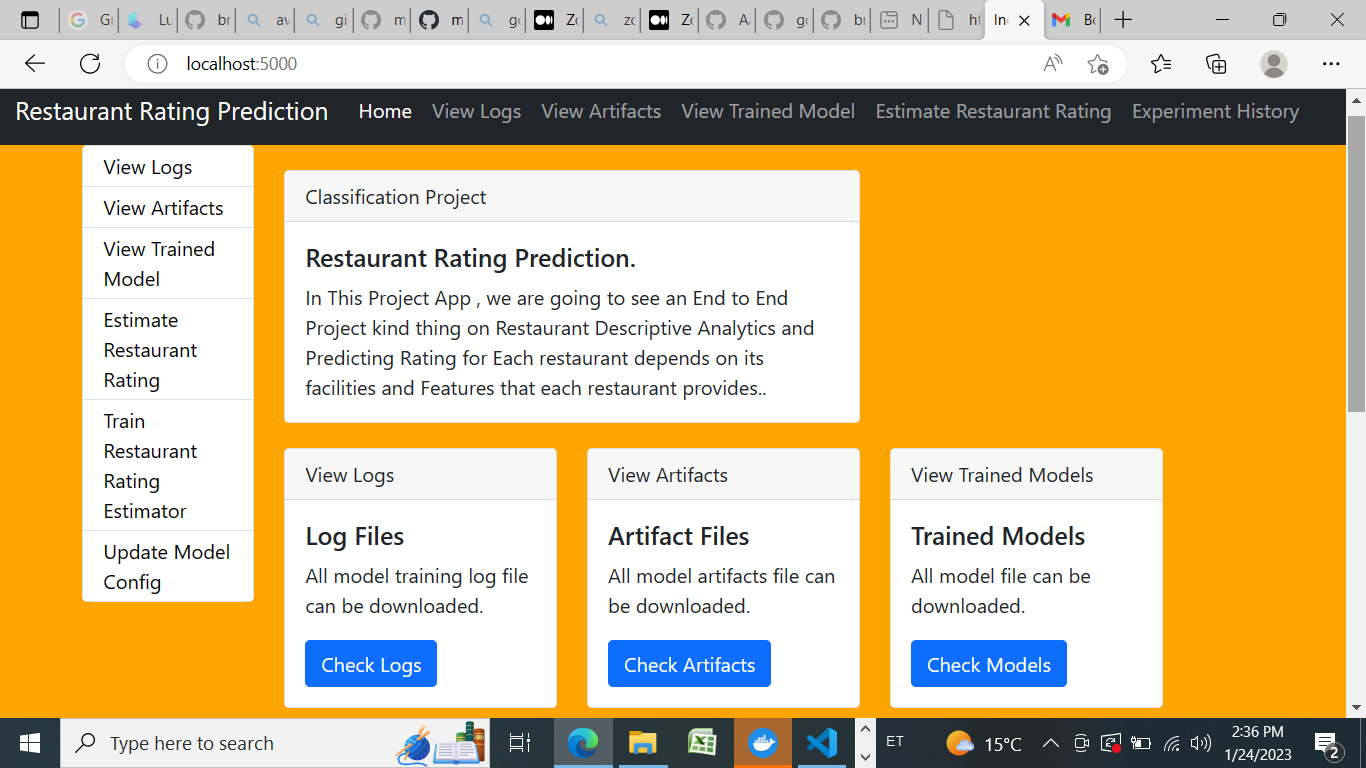
\*URL - https://- app-rest-rating-predherokuapp.com/

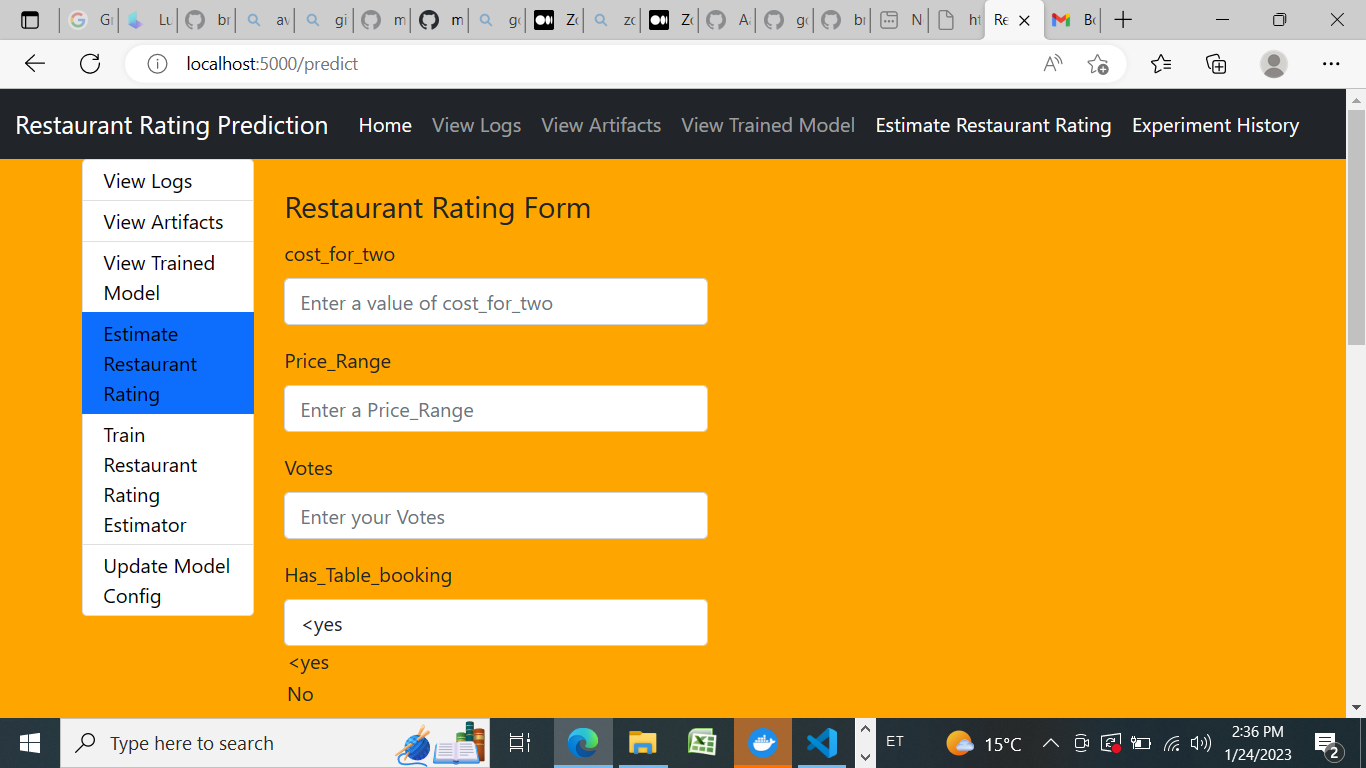
\*Deployemnt dosen work in Heroku Free Version if the size data more than 500 MB in free version

from heroku

3.6 User Interface







4 Conclusion

We collect data from CSV file and we tried to fill estimate values using related column.

Firstly we design a random model, a **random model** is something which randomly chooses values from 1.0 to 5.0, such model gives **MSE 2.12**.

We tried only 5 one-hot encoded features and try different models Random Forest Regressor was most learning model, so we tune model using **grid-search** technic, **minimal MSE** = 0.03485. Then we tried with 7 one-hot encoded features and try on different models. Again Random Forest regressor was winning the race. we achieved **MSE = 0.01404**.

Then we done some **Feature Engineering**, used response coded feature, but this time “Linear Regression” perform well than previous model, Random Forest Regressor is winning the race as usual. we achieved **MSE =0.00353**.

The Model Accuracy is hight 98.00% without underfitting