Restaurant rating predictor

Revision number – 1.0
Last date of revision- 13/10/21

Document Version control

Date issued	Version	Description	Author
13/10/21	1		Tanuj S

Contents

Abstract

Bengaluru is known as the ‘Silicon Valley of India’. And what comes with that tag, is a cosmopolitan city with new and upcoming restaurants in every part of the city. Restaurants from all over the world can be found here in Bengaluru. From United States to Japan, Russia to Antarctica, you get all type of cuisines here. Delivery, Dine-out, Pubs, Bars, Drinks, Buffet, Desserts you name it and Bengaluru has it. Bengaluru is best place for foodies. The number of restaurant are increasing day by day. Currently which stands at approximately 12,000 restaurants. With such an high number of restaurants. This industry hasn't been saturated yet. And new restaurants are opening every day. However it has become difficult for them to compete with already established restaurants. This work deals with rating restaurants in Bengaluru based on certain factors.

1 Introduction

1.1 Why this high level document?

The purpose of this high level design is to add the necessary detail to the current project description and represent a suitable model for coding . This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a higher level.

The HLD will:

- Present all of the design aspects
- Describe the user interface implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- List and describe the non-functional attribute.

1.2 Scope

The HLD documentation presents the structure of the system and the technology used .

1.3 Definitions

Term	Description
IDE	Integrated Development environment

2 General Description

2.1 Product perspective

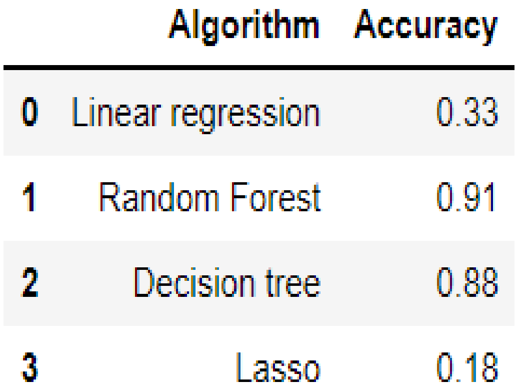
The restaurant based predictor is a machine learning model which will help us to predict the rating of a restaurant based on certain factors.

2.2 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.

2.3 Proposed solution

The solution proposed here is that we tested various algorithms to predict the rating . According to our testing , Random forest provides the best accuracy in predicting the rating. Before prediction, various steps like data cleaning , data preprocessing and EDA were performed in order to create the best model. Pipelines were also used in this project.



	Algorithm	Accuracy
0	Linear regression	0.33
1	Random Forest	0.91
2	Decision tree	0.88
3	Lasso	0.18

2.4 Further improvements

During the project , we realized that a lot of the data was missing in some columns . Imputing the missing values does help in most cases , but it never really explains the real scenario . Also , the model can be drastically improved if information regarding staff behaviour, waiting time would be available as service is so important in the hospitality industry.

2.5 Technical requirements

This document addresses the requirements for predicting the rating of the restaurant.

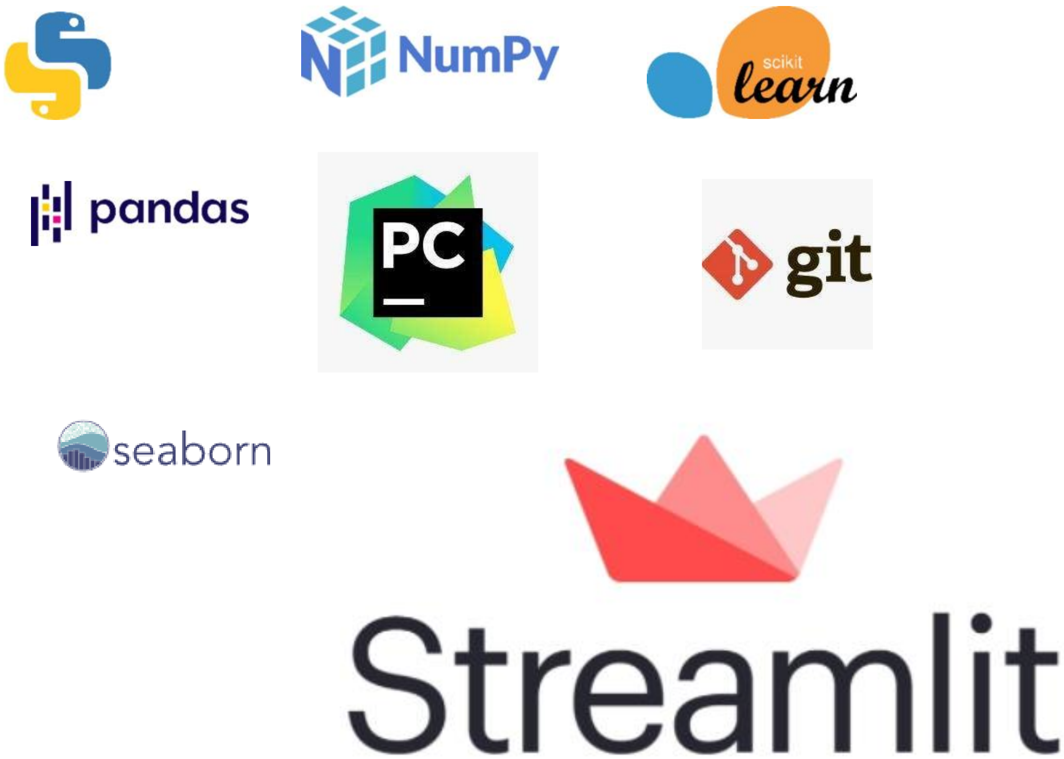
2.6 Data requirements

Data requirement completely depend on our problem statement.

- We need data that is balanced and has atleast 1000 entries.
- The development of a machine learning algorithm depends on large volumes of data, from which the learning process draws many entities, relationships, and clusters.

2.7 Tools used

Python programming language such as Numpy, Pandas , Scikit learn , TensorFlow , Pycharm, Streamlit were used . Heroku was used to deploy the model .

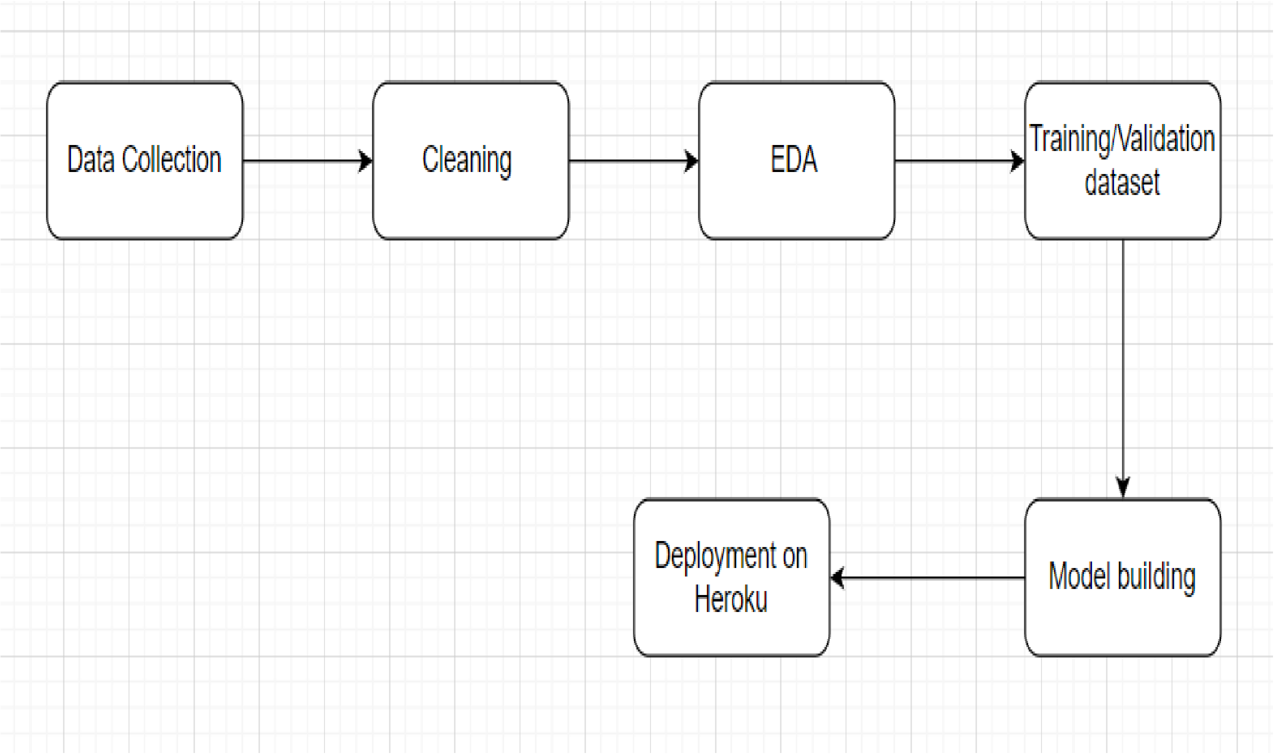


- Pycharm is used as IDE
- Data preprocessing , Cleaning and model building was done on Jupyter notebook.
- Model was imported using Pickle.
- Heruko was used to deploy the model.
- Streamlit was used for making the user-interface.
- Github is used as version control system.

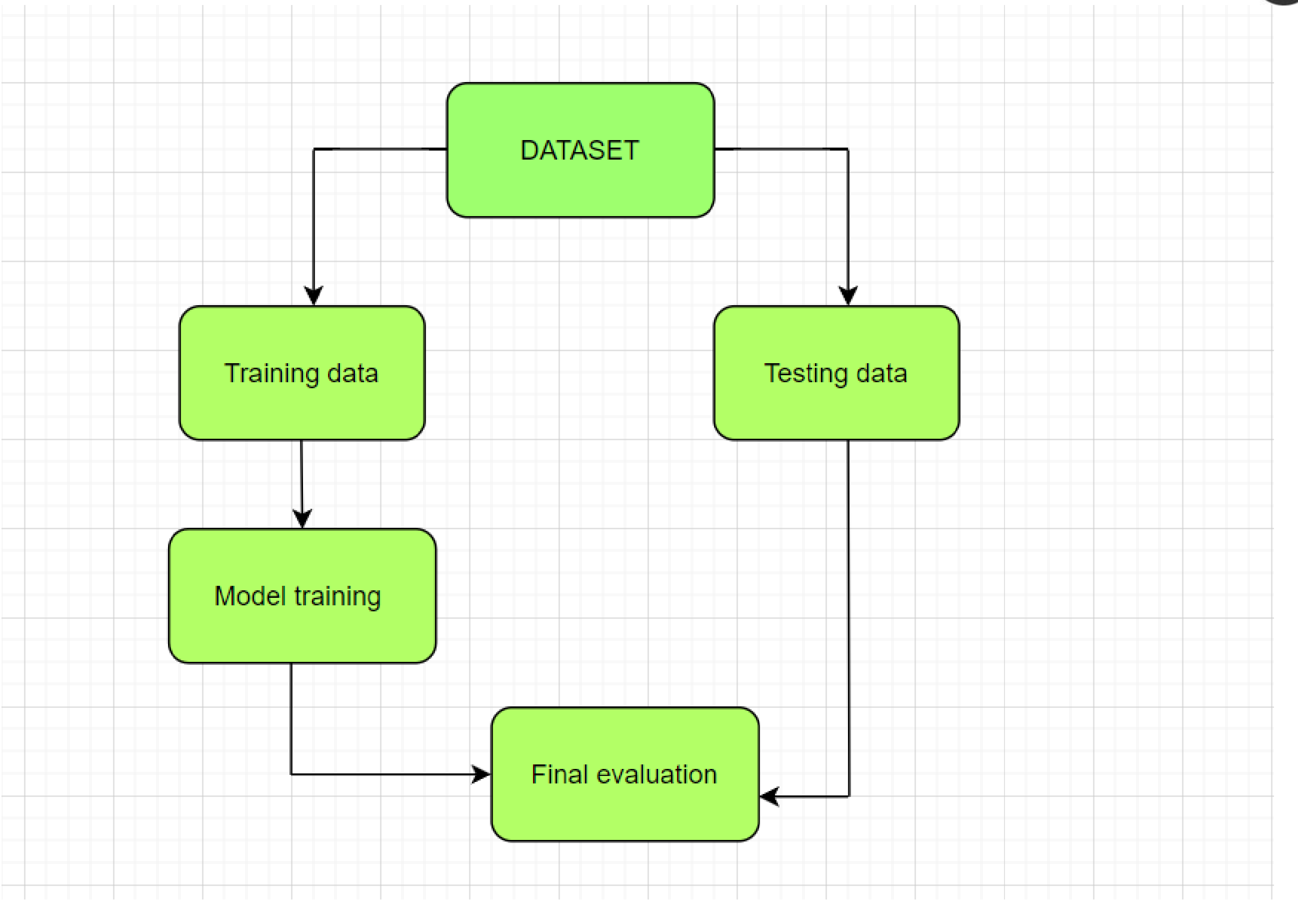
3 Design Details

3.1 Process flow

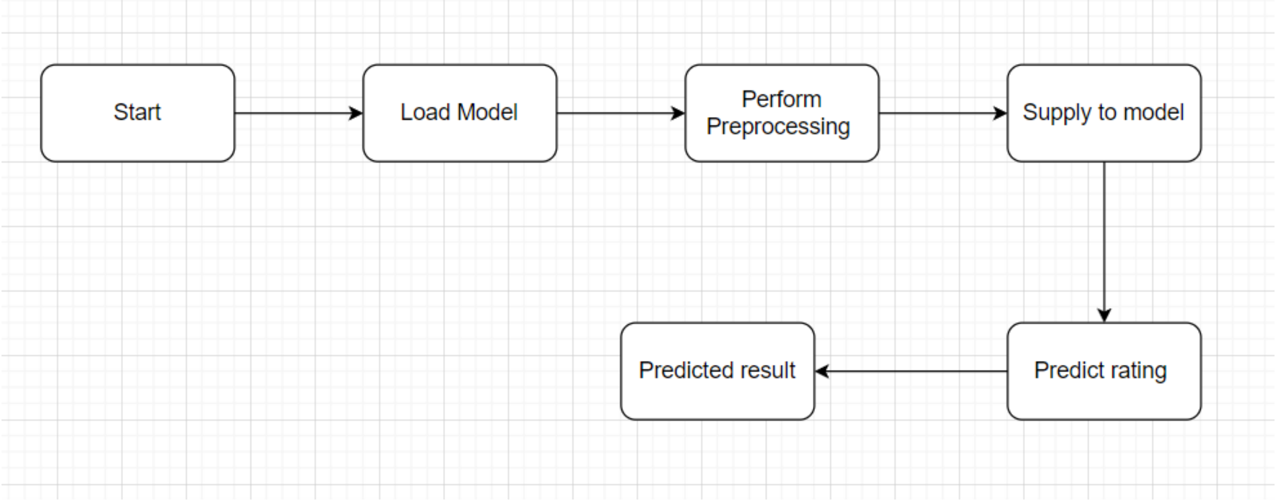
Below is the process flow diagram

- 

3.111 Model training and evaluation



3.1.2 Deployment process



4 Performance

This project aims to predict the rating as accurately as possible. This can be used to notify restaurant about their performance and what they can do to better their rating.

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different components used in this project will be using Python as an interface between them. Each component will have its own task to perform and it is the job of Python to ensure proper transfer of information.

4.3 Resource utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

5 Conclusion

In this paper, we have used various Machine Learning Algorithms such as Linear Regression, Random Forest and Gradient boosting regressor.

Through this project we trained our model based on the data of around 51,000 restaurants in and around the city of Bangalore . The rating can be an effective tool to judge and categorize the restaurants . The results indicate Random Forest has the maximum accuracy. Future work can also be done to increase the efficiency and performance of the proposed model in order to rating more accurately.