Architecture

Restaurant Rating Prediction

Harsh Paseriya

Document Version Control

Date Issued	Version	Description	Author
01 / Aug / 2022	1.1	First Draft	Harsh Paseriya
03/Aug/2022	1.2	Added Workflow chart	Harsh Paseriya
03/Aug/2022	1.3	Added user I/O flowchart	Harsh Paseriya
05/Aug/2022	1.4	Added dataset overview and updated user I/O flowchart.	Harsh Paseriya
05/Aug/2022	1.5	Restructure and Recreate LLD	Harsh Paseriya

Contents

Document Version Control 1

1 Introduction 3

1.1 Why this Low-Level Design Document? 3

1.2 Scope 3

1.3 Constraints 3

1.4 Risks 3

1.5 Out of Scope 4

2 Technical specifications 4

2.1 Dataset 4

2.1.1 Dataset overview 4

2.1.2 Input schema 5

2.2 Predicting Rating 5

2.3 Logging 6

3 Technology stack 6

4 Proposed Solution 7

5 Model training/validation workflow 8

6 User I/O workflow 9

7 Performance 9

Abstract

Bengaluru being an 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 of restaurants it has therefore become important to study the demography of a location. In the world of rising new technology and innovation, healthcare industry is advancing with the role of Artificial Intelligence. Machine learning algorithms can help to early detection of the disease and to improve the quality of the life. This study demonstrates the how different classification algorithms can forecast the presence of the disease. Different classification algorithms such as Logistic regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine have been tested and compared to predict the better outcome of the model.

1 Introduction

1.1 Why this Low-Level Design 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

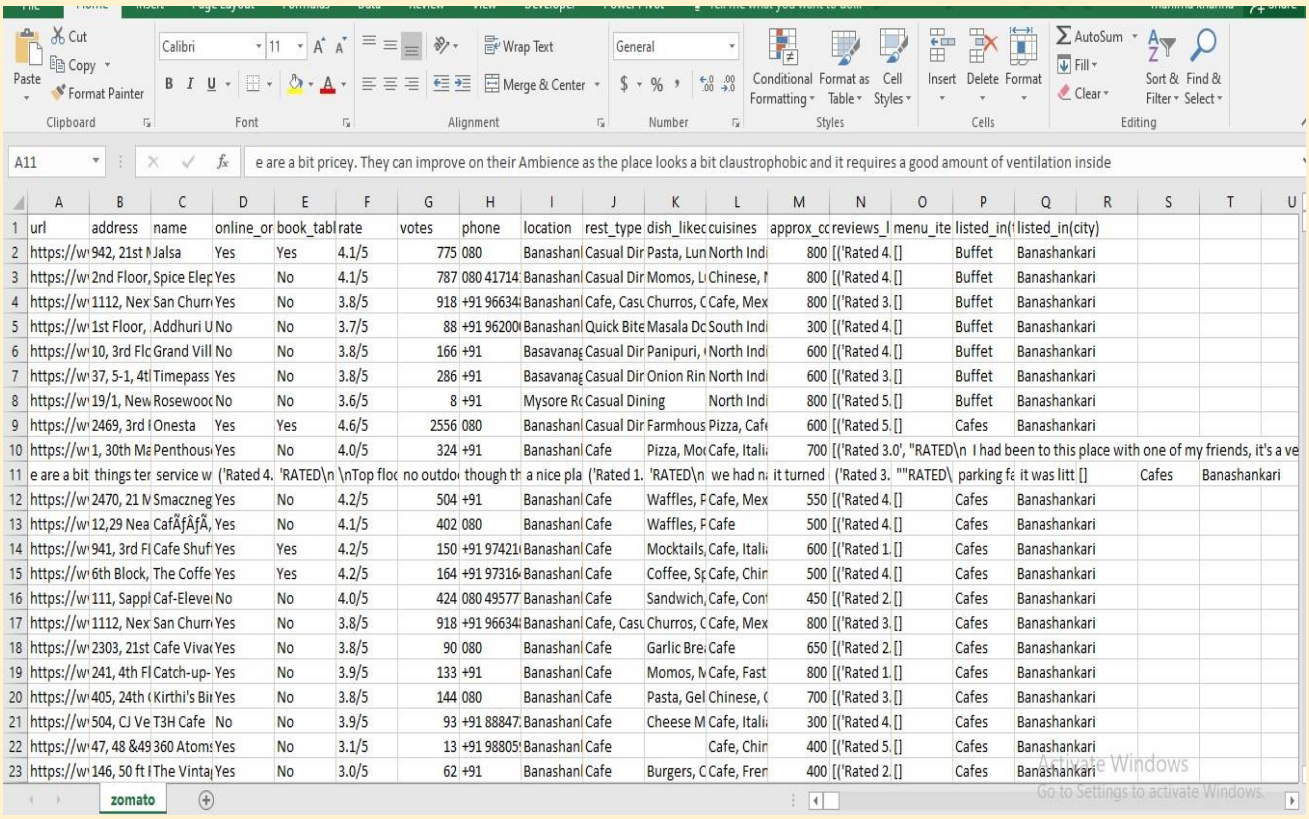
Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

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

2 Technical specifications

2.1 Dataset



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	url	address	name	online_or	book_tabl	rate	votes	phone	location	rest_type	dish_likec	cuisines	approx_cc	reviews_l	menu_ite	listed_in(	listed_in(city)				
2	https://w	942, 21st M	Jalsa	Yes	Yes	4.1/5	775	080	Banashan	Casual Dir	Pasta, Lun	North Indi	800	[('Rated 4.	[]	Buffet	Banashankari				
3	https://w	2nd Floor,	Spice Elep	Yes	No	4.1/5	787	080 41714	Banashan	Casual Dir	Momos, Li	Chinese, I	800	[('Rated 4.	[]	Buffet	Banashankari				
4	https://w	1112, Nex	San Churri	Yes	No	3.8/5	918	+91 96634	Banashan	Cafe, Casu	Churros, C	Cafe, Mex	800	[('Rated 3.	[]	Buffet	Banashankari				
5	https://w	1st Floor,	Addhuri U	No	No	3.7/5	88	+91 96200	Banashan	Quick Bite	Masala Dc	South Indi	300	[('Rated 4.	[]	Buffet	Banashankari				
6	https://w	10, 3rd Flc	Grand Vill	No	No	3.8/5	166	+91	Basavanag	Casual Dir	Panipuri, (	North Indi	600	[('Rated 4.	[]	Buffet	Banashankari				
7	https://w	37, 5-1, 4tl	Timepass	Yes	No	3.8/5	286	+91	Basavanag	Casual Dir	Onion Rin	North Indi	600	[('Rated 3.	[]	Buffet	Banashankari				
8	https://w	19/1, New	Rosewooc	No	No	3.6/5	8	+91	Mysore Rc	Casual Dining		North Indi	800	[('Rated 5.	[]	Buffet	Banashankari				
9	https://w	2469, 3rd I	Onesta	Yes	Yes	4.6/5	2556	080	Banashan	Casual Dir	Farmhous	Pizza, Cafe	600	[('Rated 5.	[]	Cafes	Banashankari				
10	https://w	1, 30th Ma	Penthous	Yes	No	4.0/5	324	+91	Banashan	Cafe	Pizza, Mo	Cafe, Itali	700	[('Rated 3.0', "RATED\n I had been to this place with one of my friends, it's a ve							
11	e are a bit	things ter	service w	('Rated 4.	'RATED\n	\nTop floc	no outdo	though th	a nice pla	('Rated 1.	'RATED\n	we had na	it turned	('Rated 3.	""RATED\	parking fa	it was litt	[]	Cafes	Banashankari	
12	https://w	2470, 21 M	Smaczneg	Yes	No	4.2/5	504	+91	Banashan	Cafe	Waffles, F	Cafe, Mex	550	[('Rated 4.	[]	Cafes	Banashankari				
13	https://w	12,29 Nea	CafÃƒÂfÃ,	Yes	No	4.1/5	402	080	Banashan	Cafe	Waffles, F	Cafe	500	[('Rated 4.	[]	Cafes	Banashankari				
14	https://w	941, 3rd Fl	Cafe Shuf	Yes	Yes	4.2/5	150	+91 97421	Banashan	Cafe	Mocktails,	Cafe, Itali	600	[('Rated 1.	[]	Cafes	Banashankari				
15	https://w	6th Block,	The Coffe	Yes	Yes	4.2/5	164	+91 97316	Banashan	Cafe	Coffee, Sp	Cafe, Chin	500	[('Rated 4.	[]	Cafes	Banashankari				
16	https://w	111, Sappl	Caf-Eleve	No	No	4.0/5	424	080 49577	Banashan	Cafe	Sandwich,	Cafe, Cont	450	[('Rated 2.	[]	Cafes	Banashankari				
17	https://w	1112, Nex	San Churri	Yes	No	3.8/5	918	+91 96634	Banashan	Cafe, Casu	Churros, C	Cafe, Mex	800	[('Rated 3.	[]	Cafes	Banashankari				
18	https://w	2303, 21st	Cafe Viva(	Yes	No	3.8/5	90	080	Banashan	Cafe	Garlic Bre	Cafe	650	[('Rated 2.	[]	Cafes	Banashankari				
19	https://w	241, 4th Fl	Catch-up-	Yes	No	3.9/5	133	+91	Banashan	Cafe	Momos, M	Cafe, Fast	800	[('Rated 1.	[]	Cafes	Banashankari				
20	https://w	405, 24th (	Kirthi's Bir	Yes	No	3.8/5	144	080	Banashan	Cafe	Pasta, Gel	Chinese, (	700	[('Rated 3.	[]	Cafes	Banashankari				
21	https://w	504, CJ Ve	T3H Cafe	No	No	3.9/5	93	+91 88847	Banashan	Cafe	Cheese M	Cafe, Itali	300	[('Rated 4.	[]	Cafes	Banashankari				
22	https://w	47, 48 &49	360 Atom	Yes	No	3.1/5	13	+91 98805	Banashan	Cafe		Cafe, Chin	400	[('Rated 5.	[]	Cafes	Banashankari				
23	https://w	146, 50 ft I	The Vinta(	Yes	No	3.0/5	62	+91	Banashan	Cafe	Burgers, C	Cafe, Fren	400	[('Rated 2.	[]	Cafes	Banashankari				

zomato

2.1.1 Dataset 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
Rest Type	Required
Dish Liked	Required
Cuisine	Required
Cost	Required
Review	Required
Type	Required

2.2 Predicting 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

✓ HEROKU

3 Technology stack

Front End	HTML/CSS
Backend	Python Flask
Deployment	Heroku

4 Proposed Solution

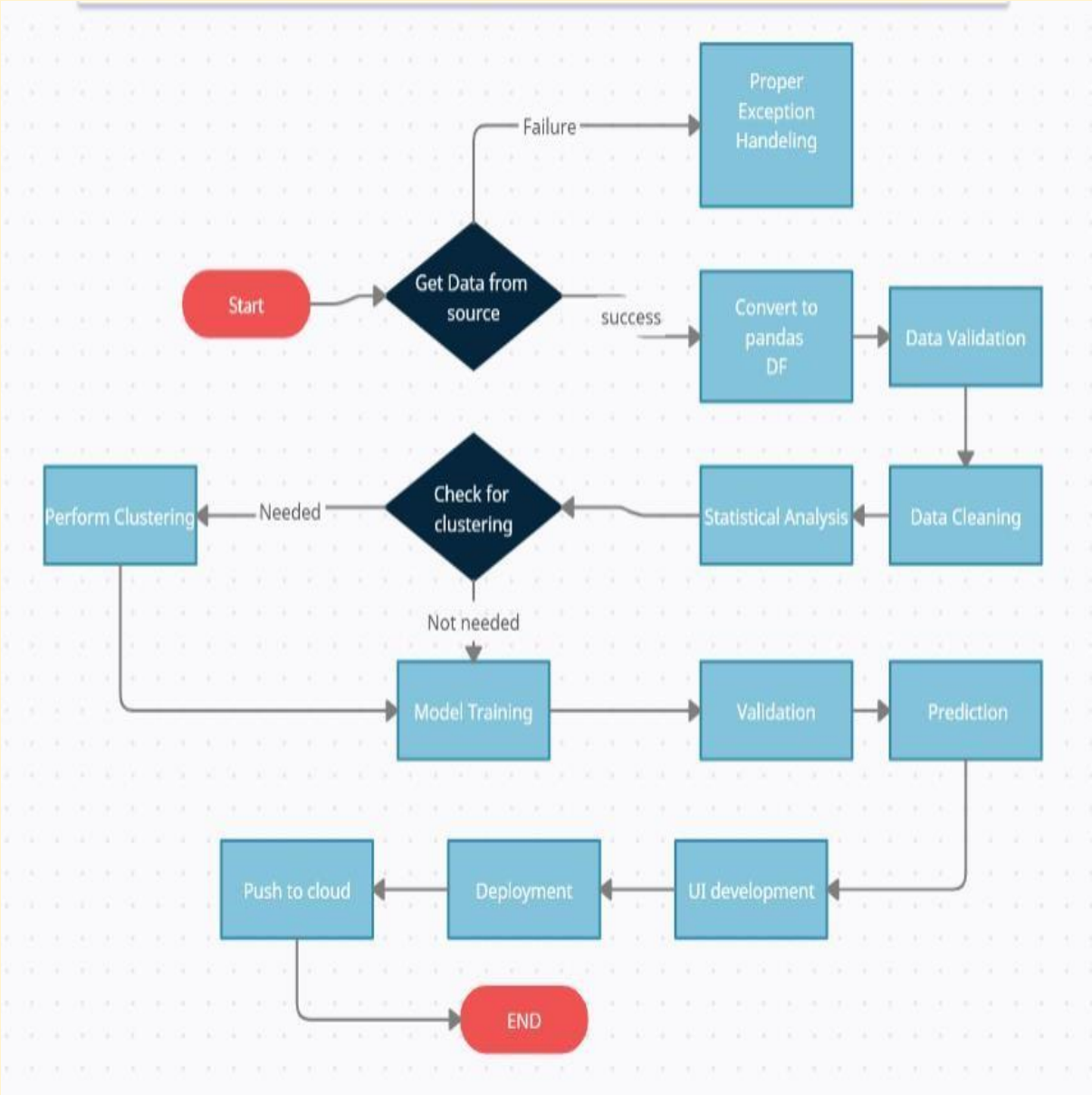
The proposed solution for this project is Machine learning algorithms can be implemented to predict the rating of the restaurant. Considering various features like online order, book table, votes, rest type, cuisines, review as inputs from the web app, the implemented classification model will predict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

Here, we have used Random Forest Classifier to predict the restaurant rating.

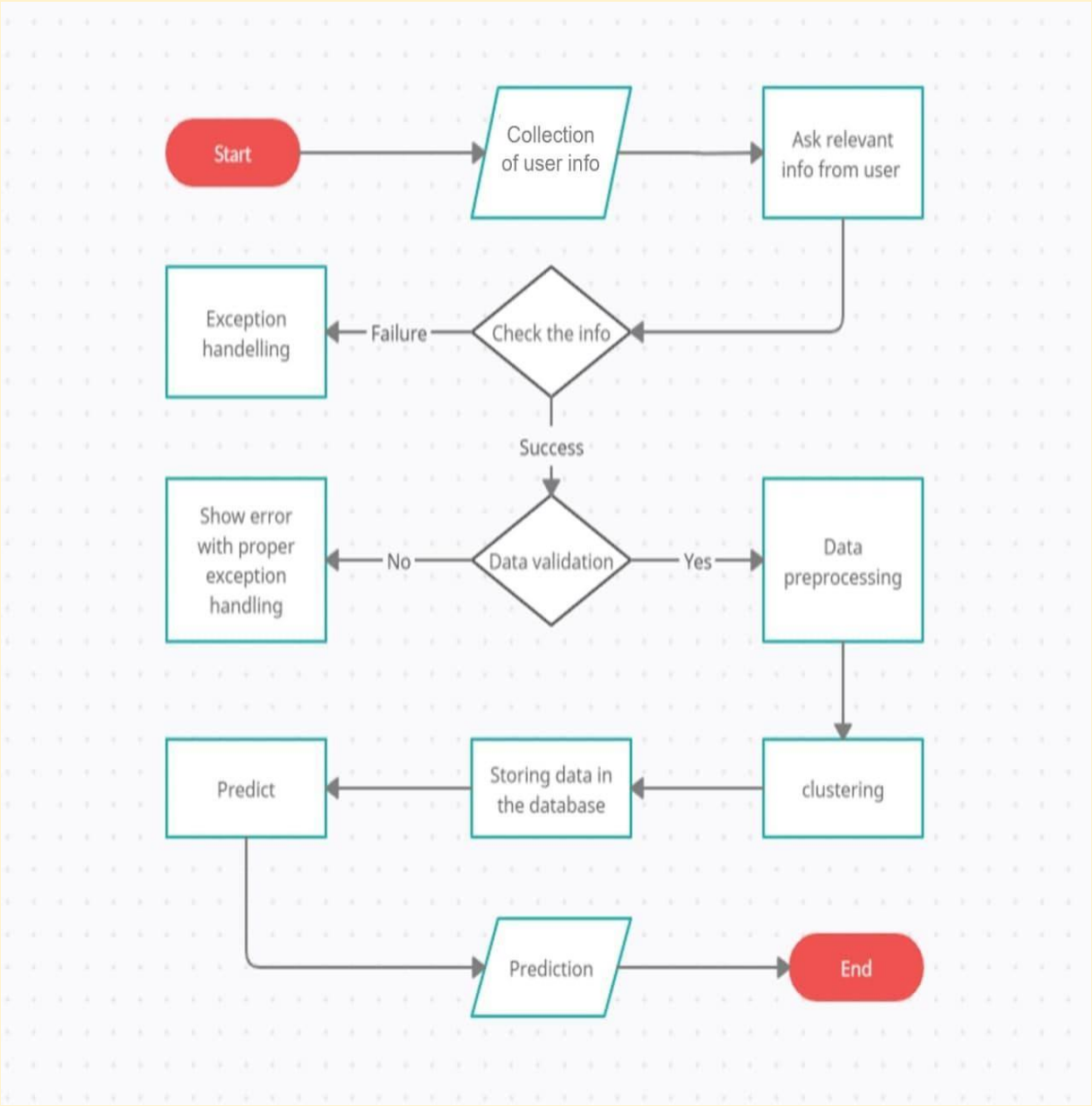
However, drawing a baseline model is important since it tells us how well other models have performed compared to base model. Here, the base model for Restaurant Rating dataset is Logistic Regression.

1. Baseline Model : Logistic Regression
2. Actual Model : Random Forest with ExtraTree Regressor

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.