

# Zomato Restaurant Rating Prediction

The Zomato logo is centered on a red rectangular background. Surrounding the logo are three food items: a bowl of yellow rice with a star anise in the top right, a bowl of yellow rice with a star anise in the bottom left, and a plate of salmon with lemon slices and tomatoes in the bottom right.

**zomato**



## Objective

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.

## Benefits

1. Using organizational data into real world business use-case
2. Predicting Restaurant rating.
3. Optimum services provided by the restaurants.
4. Reduce risk of loss
5. Good understanding about the locality.



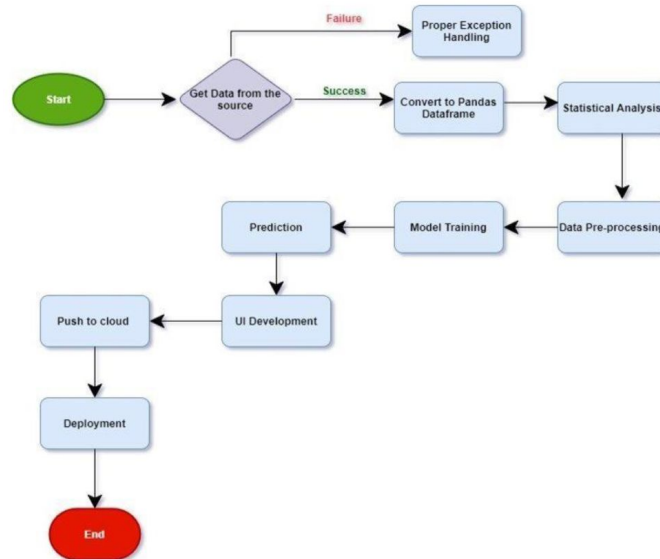
# Data Sharing Agreement

- Source of file  
<https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants?resource=download>
- Shape of the data is 51717x17
- 51717 rows
- 17 columns
- Column data types where:int64,object
- Where we use only these 9 features.

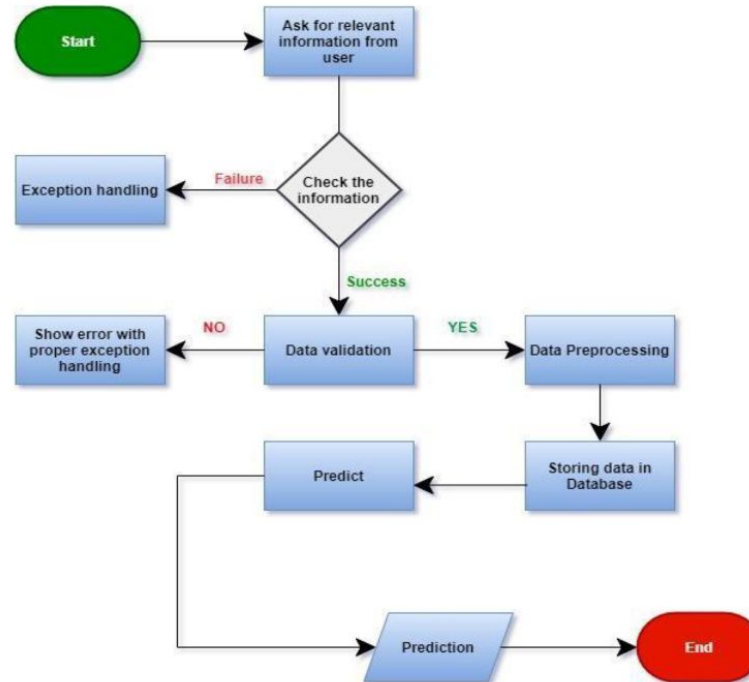
online_order	book_table	rate	votes	rest_type	cuisines	cost2plates	type	location
--------------	------------	------	-------	-----------	----------	-------------	------	----------

# Architecture

## Machine Learning Model



## Input Output Flow Diagram





# Model Training

Data Export from CSV:

- Loading CSV data using python pandas.

Data Preprocessing:

- Performing EDA to get insight of data like identifying distribution , outliers ,trend among data etc.
- Check for null values in the columns. If present impute the null values.
- Perform feature selection for ML model.



## Feature selection

In Feature selection we have selected the required feature from the dataset:

- Based on correlation of input variable with output variable.
- Based on common input feature which user can select.
- Input variable should not have same dependency on output feature (to avoid low accuracy on ML model).

Train and Test split:

- Train data is 70% of the whole data.
- Test data is 30% of the whole data.
- Data is randomly splitted in train and test.



## Model Selection:

As this is the regression problem use case we have used linear regression and followed by the other regression algorithms such as Ensemble algorithm. Where linear regression was not giving accuracy more than 26% so we use Ensemble algorithm such as Extra Tree Regressor and Random Forest among both Extra Tree Regressor was giving better result approximate (84%) accuracy.

## Prediction:

- Loading CSV data using Python Panda.
- Performing data Pre-processing.
- We used Extra Tree Regression algorithm for prediction.
- Based on Extra Tree algorithm, model is loaded and used to predict the outcome.
- Prediction of model is done on specific features.
- Our model is giving Approx 84% accuracy.
- We used MongoDB database to save users input data.





# Data Insertion In Database

Database: Database is created, scalable, flexible NOSQL format. This is linked with MongoDB Atlas, which is a cloud database service.

Collection: Collection is similar to table as we see in SQL database. Here collection has been created under 'user\_detail'. The records are independent to each other.



## Q&A

Q1) What's the source of data?

Source of file kaggle URL -

<https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants?resource=download>

Q 2) What was the type of data?


The data was the combination of numerical and object value.

Q 3) What's the complete flow you followed in this Project?

Refer slide 4 & 5.

Q 4) How logs are managed?

We are using different logs as per the steps we follow in validation and modeling like File validation log, Data insertion etc.,



Q 5) What techniques were you using for data pre-processing?

Visualizing relation of independent variables with each other.

Removing unwanted feature.

Cleaning data and imputing if null values are present.

Converting categorical data into numeric values.

Selecting feature.

Q 6) How training was done or what models were used?

Before diving the data in training and validation set we performed clustering over fit to divide the data into clusters.

As per the model the training and test data were divided.

Data is randomly splitted in train and test.



Algorithms like Linear Regression, Extra Tree Regression were used based on the accuracy final model was selected for prediction.

Saved that model.

Q 7) How Prediction was done?

The data files are shared by the client, then we have done some pre-processing techniques on data and trained data using Extra Tree Regressor algorithm and which is loaded and performed prediction.

Q 8) What are the different stages of deployment?

When the model is ready we deploy in the local environment.

Then project file transfer to AWS (EC2 Instance) using putty.

Finally the model has been deployed in AWS (EC2 Instance).