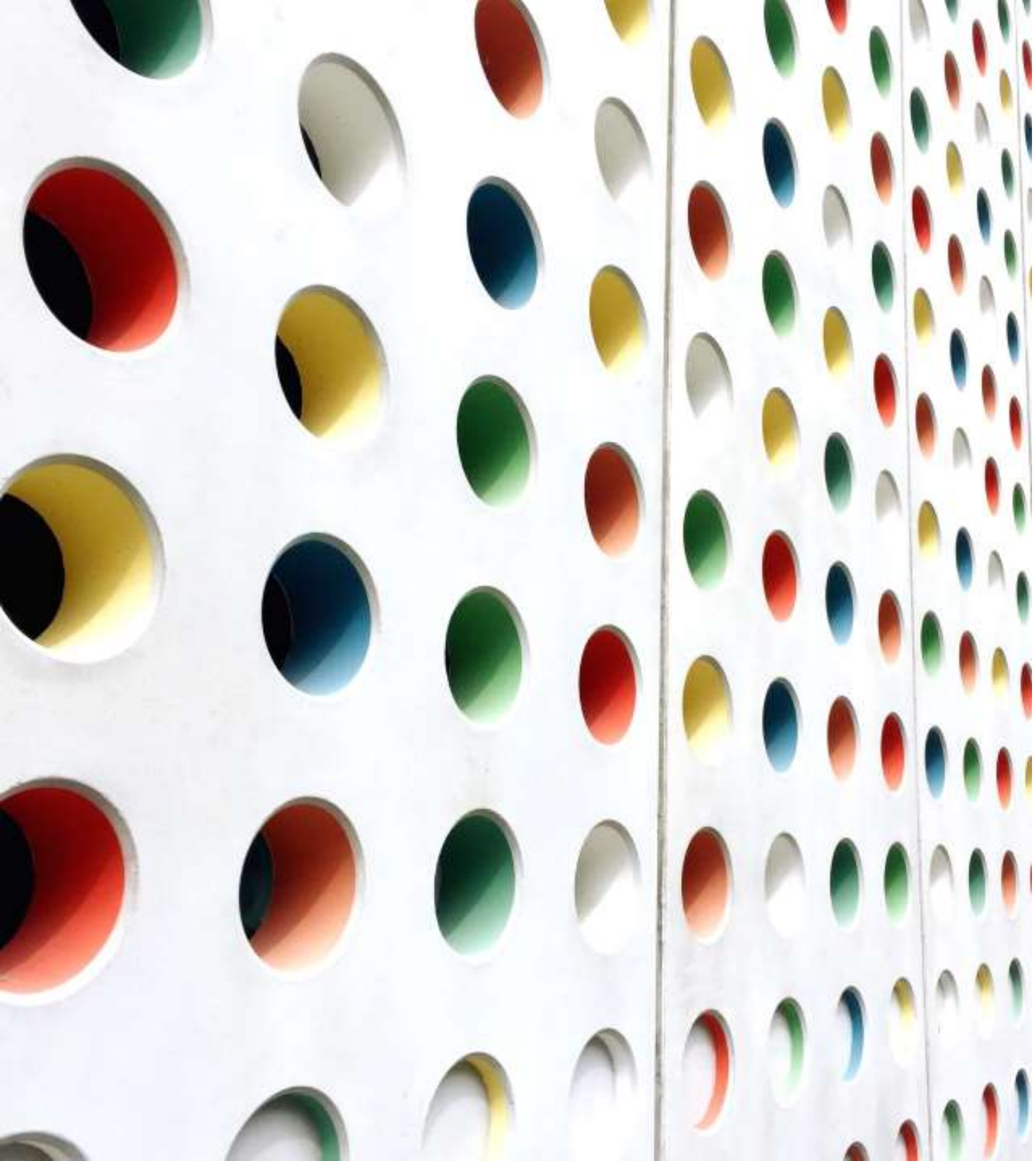


Restaurant Rating Prediction

RAHUL CHATTERJEE

Objective

Zomato is a popular restaurant search and discovery platform that provides information about various restaurants, menus, reviews, and ratings. The objective of restaurant rating prediction for Zomato can include the following:

1. Help users make informed decisions: Zomato's primary goal is to help users find the best restaurants based on their preferences. By predicting the rating of a restaurant, Zomato can help users make informed decisions about where to dine.
2. Improve user experience: Zomato can use the rating prediction model to improve user experience by providing personalized restaurant recommendations based on the user's preferences and predicted ratings.
3. Help restaurants improve: Zomato can use the rating prediction model to provide feedback to restaurants on how they can improve their ratings. This feedback can include information about food quality, service, ambience, and pricing.
4. Identify popular restaurants: By predicting the rating of a restaurant, Zomato can identify popular restaurants that are highly rated by customers. This information can be useful for users looking for popular restaurants in a particular location or cuisine.
5. Improve Zomato's credibility: Zomato's credibility as a restaurant discovery platform depends on the accuracy of its ratings and reviews. By developing an accurate rating prediction model, Zomato can improve its credibility and attract more users to its platform.

Overall, the objective of restaurant rating prediction for Zomato is to provide users with useful information, improve the user experience, help restaurants improve, identify popular restaurants, and improve Zomato's credibility.

Data Sharing Agreement

Number of Columns: 5

Column names:

1. Has Table Booking
2. Has Online Delivery
3. Average Cost for two
4. Price Range
5. Votes

Column data type: string, float, integer

Data Sharing Agreement

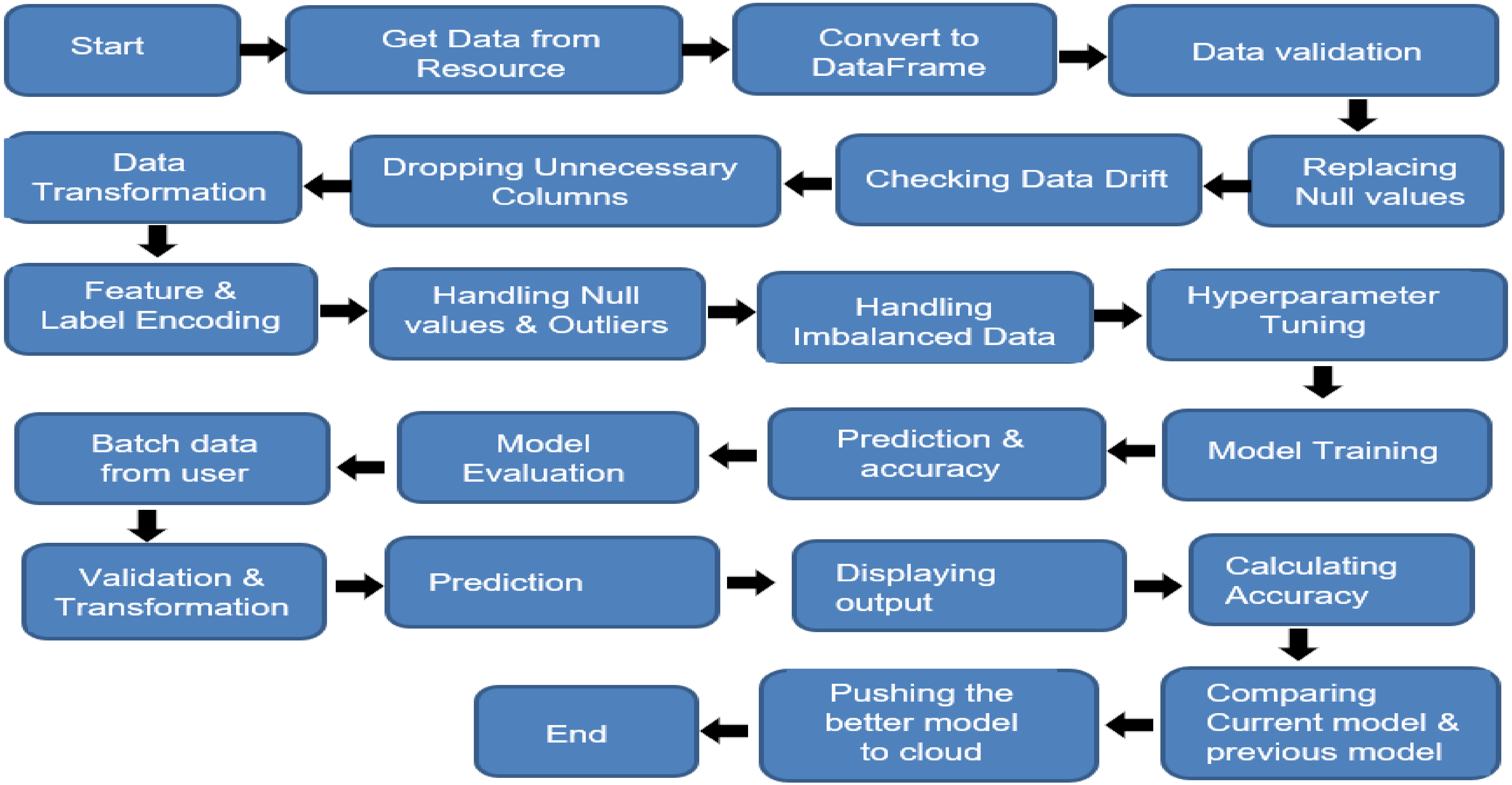
Number of Columns: 5

Column names:

1. Has Table Booking
2. Has Online Delivery
3. Average Cost for two
4. Price Range
5. Votes

Column data type: string, float, integer

Architecture



Start

Get Data from Resource

Convert to DataFrame

Data validation

Replacing Null values

Checking Data Drift

Dropping Unnecessary Columns

Data Transformation

Feature & Label Encoding

Handling Null values & Outliers

Handling Imbalanced Data

Hyperparameter Tuning

Model Training

Prediction & accuracy

Model Evaluation

Batch data from user

Validation & Transformation

Prediction

Displaying output

Calculating Accuracy

Comparing Current model & previous model

Pushing the better model to cloud

End

Data Validation and Data Transformation

Name Validation - Validation of files name as per the DSA. We have created a regex pattern for validation. After it checks for date format and time format if these requirements are satisfied, we move such files to "Good_Data_Folder" else "Bad_Data_Folder.“

Number of Columns – Validation of number of columns present in the files, and if it doesn't match then the file is moved to "Bad_Data_Folder.“

Name of Columns - The name of the columns is validated and should be the same as given in the schema file. If not, then the file is moved to "Bad_Data_Folder".

Data type of columns - The data type of columns is given in the schema file. It is validated when we insert the files into Database. If the datatype is wrong, then the file is moved to "Bad_Data_Folder".

Null values in columns - If any of the columns in a file have all the values as NULL or missing, we discard such a file and move it to "Bad_Data_Folder".

Data Insertion in Database

Database:- Data will be inserted into mongoDB database

Data will be coming form various sources and inserted into mongoDB database. In Data Ingestion step reading the data from mongoDB database and added into the feature_store folder path

Model Training:

Data Export from Db :

The accumulated data from db is exported in csv format for model training

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.

Encode the categorical values with numeric values.

Perform Standard Scalar to scale down the values.

Model Selection

Among all ML algorithm Decision Tree gives us the best result.
In every retraining step if we will get better result, then new model pushed into the model pusher step.

Prediction:

We can do batch prediction as well as singular prediction:

1. For batch prediction we need to provide the input file to predict
2. In singular prediction we can use our frontend web form

Q & A:

Q1) What’s the source of data?

The data for training is provided by the client in multiple batches and each batch contain multiple files

Q 2) What was the type of data?

The data was the combination of numerical and Categorical values.

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

Refer slide 5th for better Understanding

Q 4) After the File validation what you do with incompatible file or files which didn’t pass the validation?

Files like these are moved to the Achieve Folder and a list of these files has been

shared with the client and we removed the bad data folder.

Q & A:

Q 5) How logs are managed?

We are using different logs as per the steps that we follow in validation and modeling like File validation log , Data Insertion ,Model Training log , prediction log etc.

Q 6) What techniques were you using for data pre-processing?

Removing unwanted attributes
Visualizing relation of independent variables with each other and output variables
Checking and changing Distribution of continuous values
Removing outliers
Cleaning data and imputing if null values are present.
Converting categorical data into numeric values.
Scaling the data

Q & A:

Q 7) How training was done or what models were used?

Before diving the data in training and validation performed on top of the dataset, if everything is fine, then data will be trained by using Decision Tree Regressor.

Q 8) How Prediction was done?

The testing files are shared by the client .We Perform the same life cycle till the data is clustered .Then on the basis of cluster number model is loaded and perform prediction. In the end we get the accumulated data of predictions.

Q 9) What are the different stages of deployment?

When the model is ready we deploy it in Fire environment .Where SIT and UAT is performed over it.
Once We get Sign off from Fire we deploy in Earth and UAT is performed over it.
After getting the sign off from Earth we deploy in production