Restaurant Rating Prediction

-MAHIMA KHANNA

Introduction

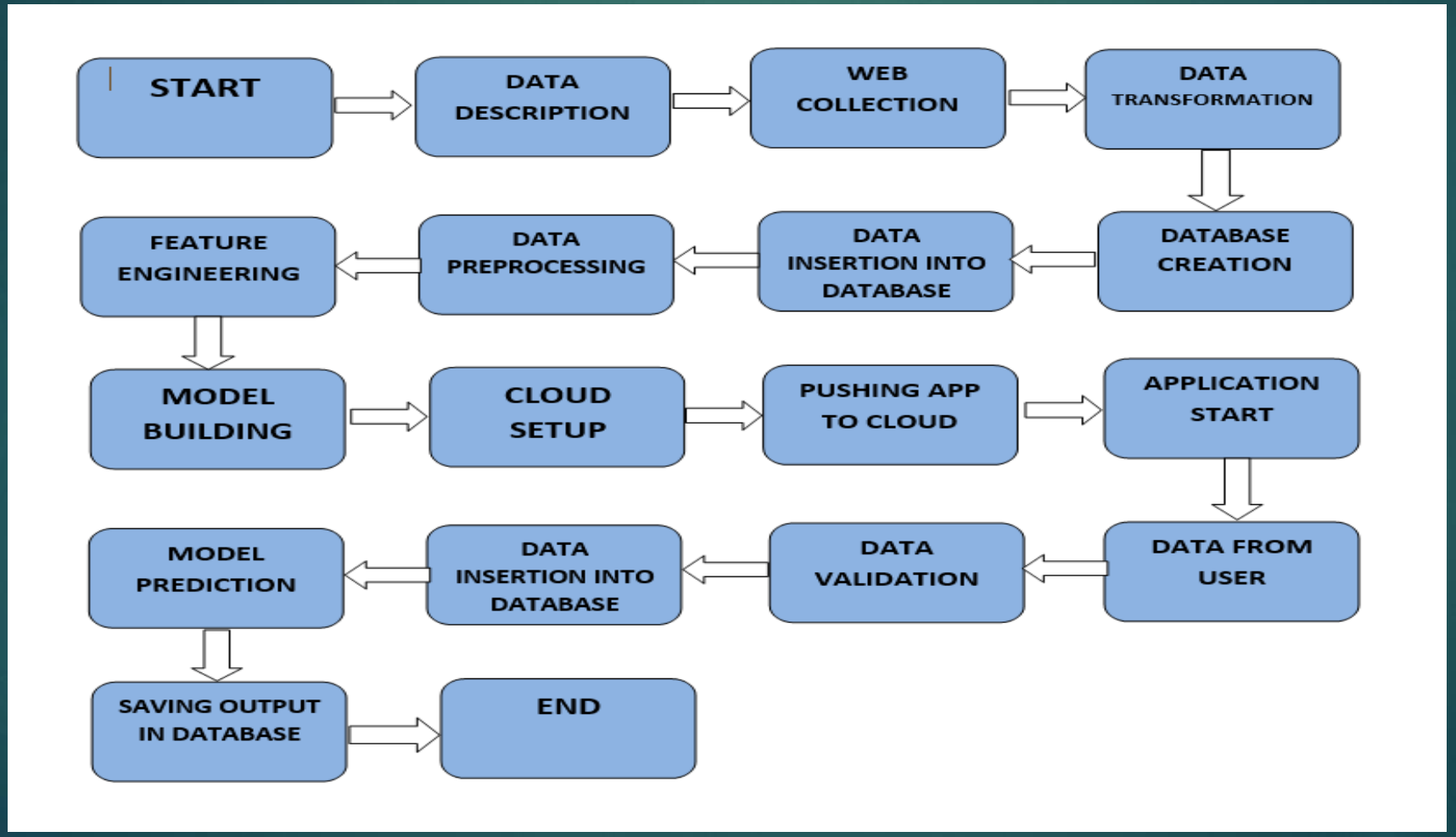
Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. In the world of rising new 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 forecasts the presence of the disease. Different classification algorithms such as Logistic regression, Random Forest, Decision Tree have been tested and compared to predict the better outcome of the model.

Objective

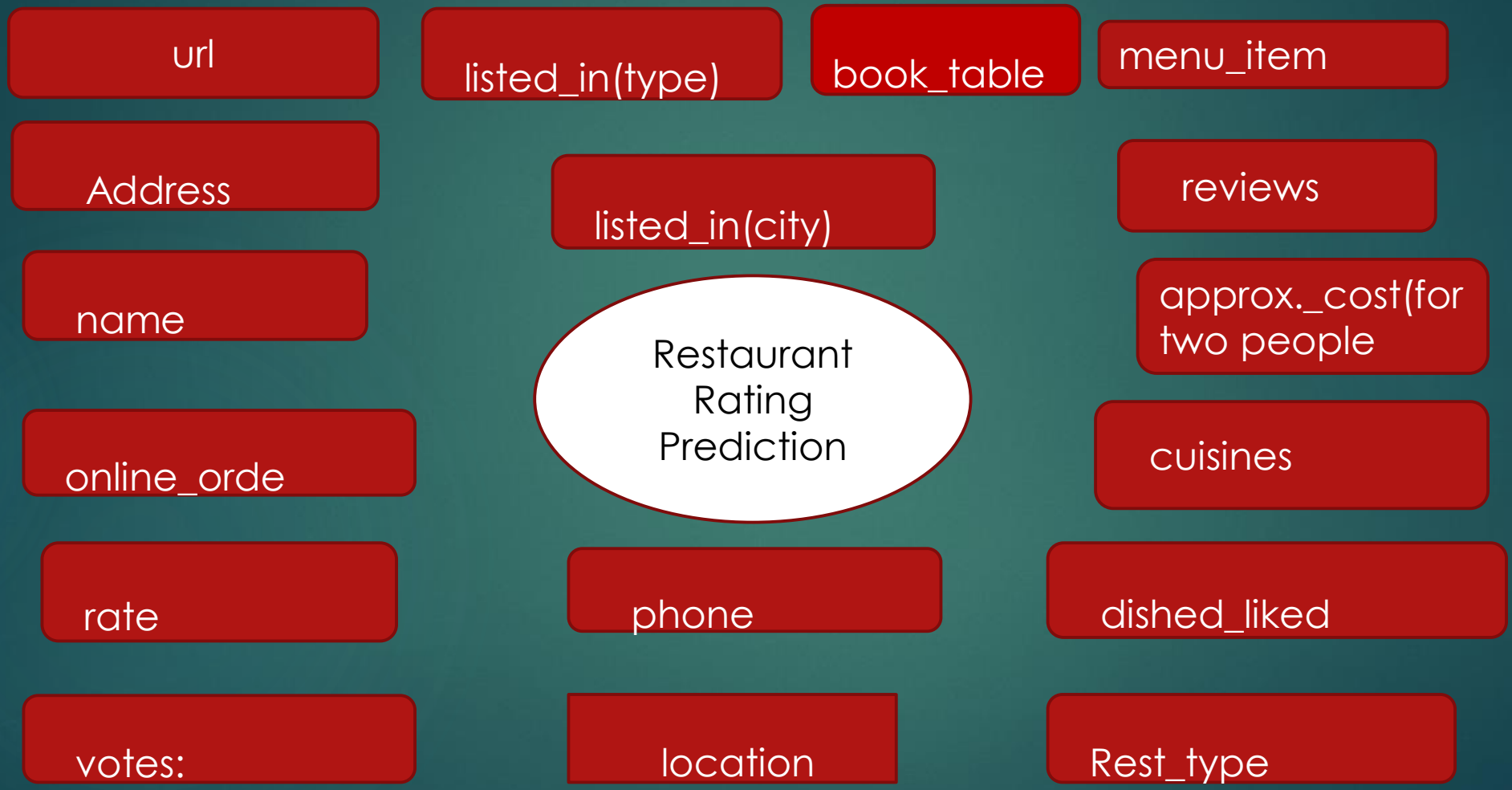
The Restaurant Rating Prediction is a machine learning based model which will help us to predict the rating of the restaurant in Bangalore. The dataset also contains reviews for each of the restaurant which will help in finding overall rating for the place.

The main goal of this project is to perform exploratory data analysis and later predict the rating of the restaurant.

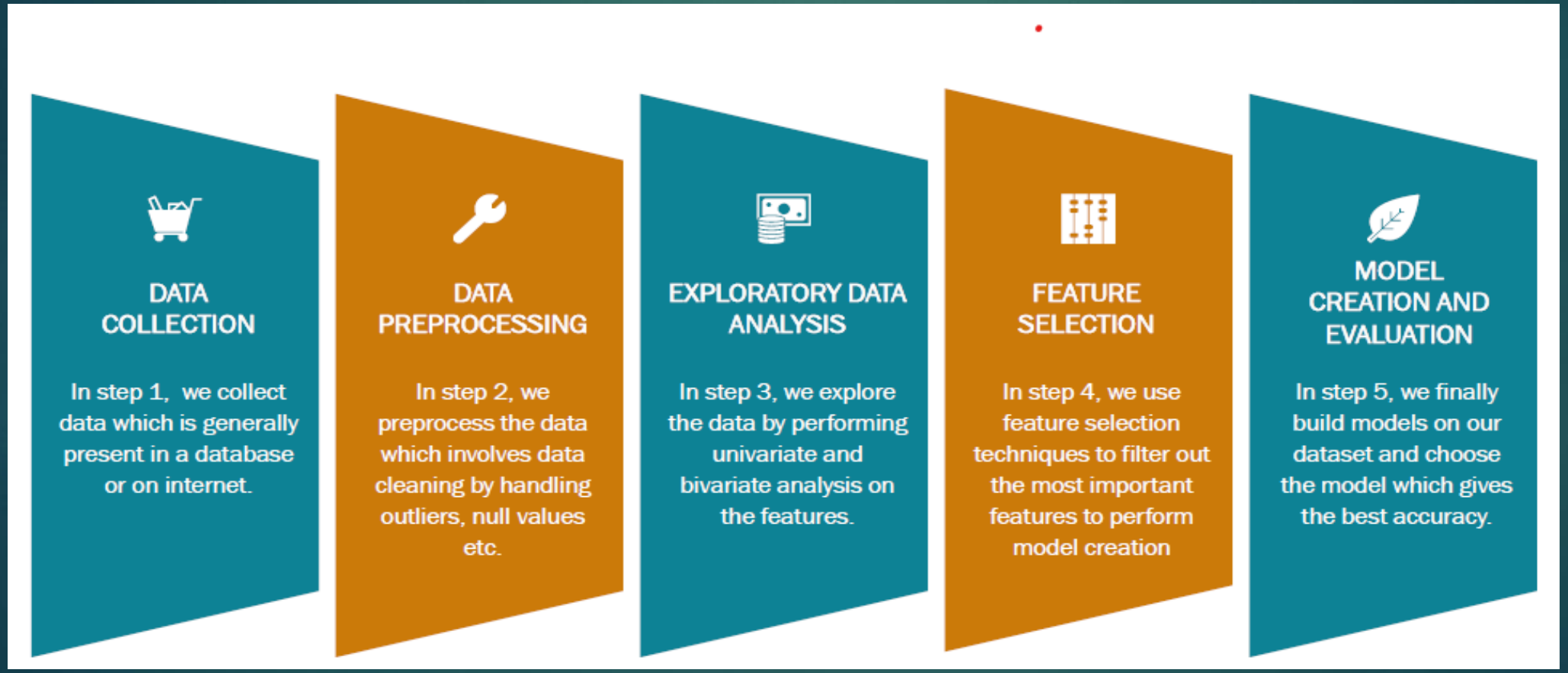
Architecture



Dataset


url
listed_in(type)
book_table
menu_item
Address
listed_in(city)
reviews
name
approx._cost(for two people
Restaurant Rating Prediction
online_orde
cuisines
rate
phone
dished_liked
votes:
location
Rest_type

Data Analysis



Random Forest Model

INTRODUCTION

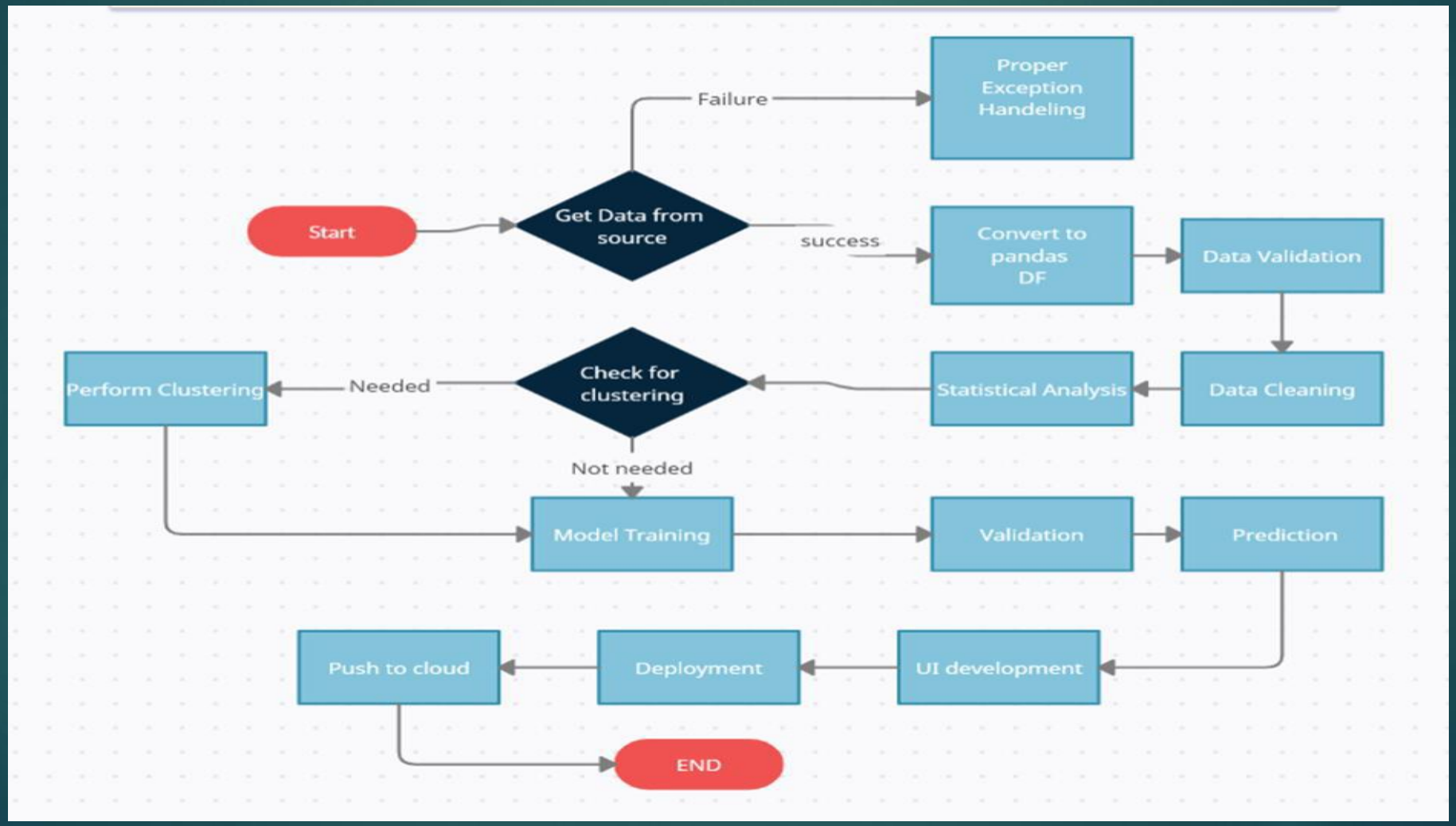
- Therandomforestclassifierisasupervisedlearningalgorithmwhichwecanuseforregressionandclassificationproblems.Itisamongthemostpopularmachinelearningalgorithmsduetoitshighflexibilityandeaseofimplementation.

- It is called Random Forest because it consists of multiple decision trees just as a forest has many trees. On top of that, it uses randomness to enhance its accuracy and combat overfitting, which can be a huge issue for such a sophisticated algorithm. These algorithms make decision trees based on a random selection of data samples and get predictions from every tree. After that, they select the best viable solution through votes.

- Random Forest Classifier being ensembled algorithm tends to give more accurate result. This is because it works on the principle i.e. number of weak estimators when combined forms strong estimator. Even if one or few decision trees are prone to noise, overall results would tend to be correct.

It gives us high accuracy as 87%.

MODEL TRAINING AND VALIDATION WORKFLOW



MODEL TRAINING AND VALIDATION WORKFLOW

Data Collection

- Zomato Restaurant Data Set from Kaggle.

Data Pre-processing

- Missing values handling by Simple imputation (median strategy)
- Outliers' detection and removal by boxplot and percentile methods
- Categorical features handling by ordinal encoding and label encoding
- Feature scaling done by Standard Scalar method

MODEL TRAINING AND VALIDATION WORKFLOW

Model Creation and Evaluation

- Various classification algorithms like Logistic Regression, Random Forest, Decision Tree tested.
- Random Forest, Decision Tree and Logistic regression were given better results. Random Forest was chosen for the final model training and testing.
- Hyper parameter tuning was performed.
- Model performance evaluated based on accuracy, confusion matrix, classification report.

DEPLOYMENT

Model Deployment

- The final model is deployed using on Heroku using Flask framework

Thank You