RESTAURANT RATING PREDICTION
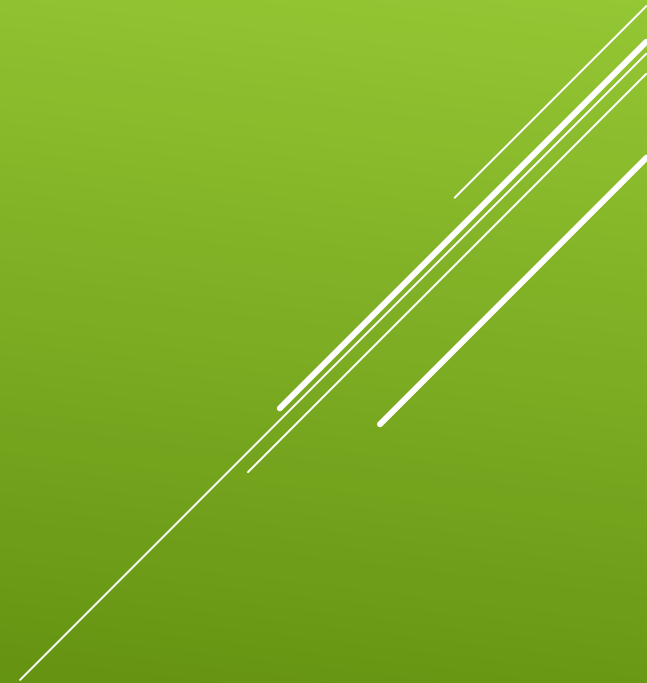
-NITESH KUMAR PANDEY

INTRODUCTION

Bengaluru Whenever we go for a food application to order some the 1st thing that comes to our mind is that we order food where we get quality food. To accomplish that, whether the restaurant can provide quality food or not, is we first look for the restaurant rating and what other customers have mentioned about the restaurant 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, studying the demography of a location has become important. In the world of rising new technology and innovation, the industry is advancing with the role of Artificial Intelligence. Machine learning algorithms can help early detection of the disease and improve the quality of life. This study demonstrates how different Regression algorithms can forecast the rating of restaurants so that one user can make a real decision whether to buy food from the restaurant or not as per their ratings. Different regression algorithms such as Decision Tree, Linear Regression, Random forest, etc have been tested and compared to predict the better outcome of the model.
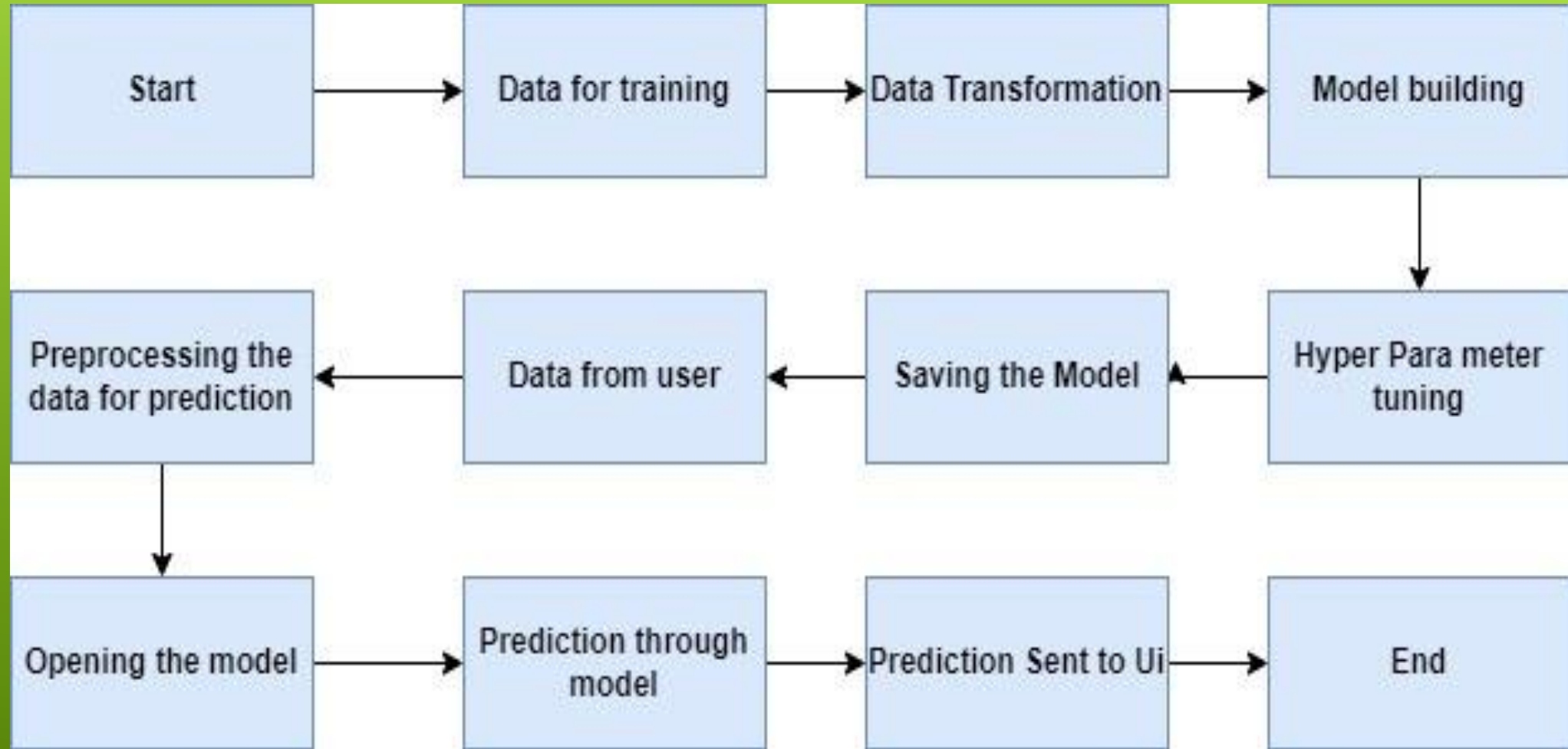
OBJECTIVE

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.
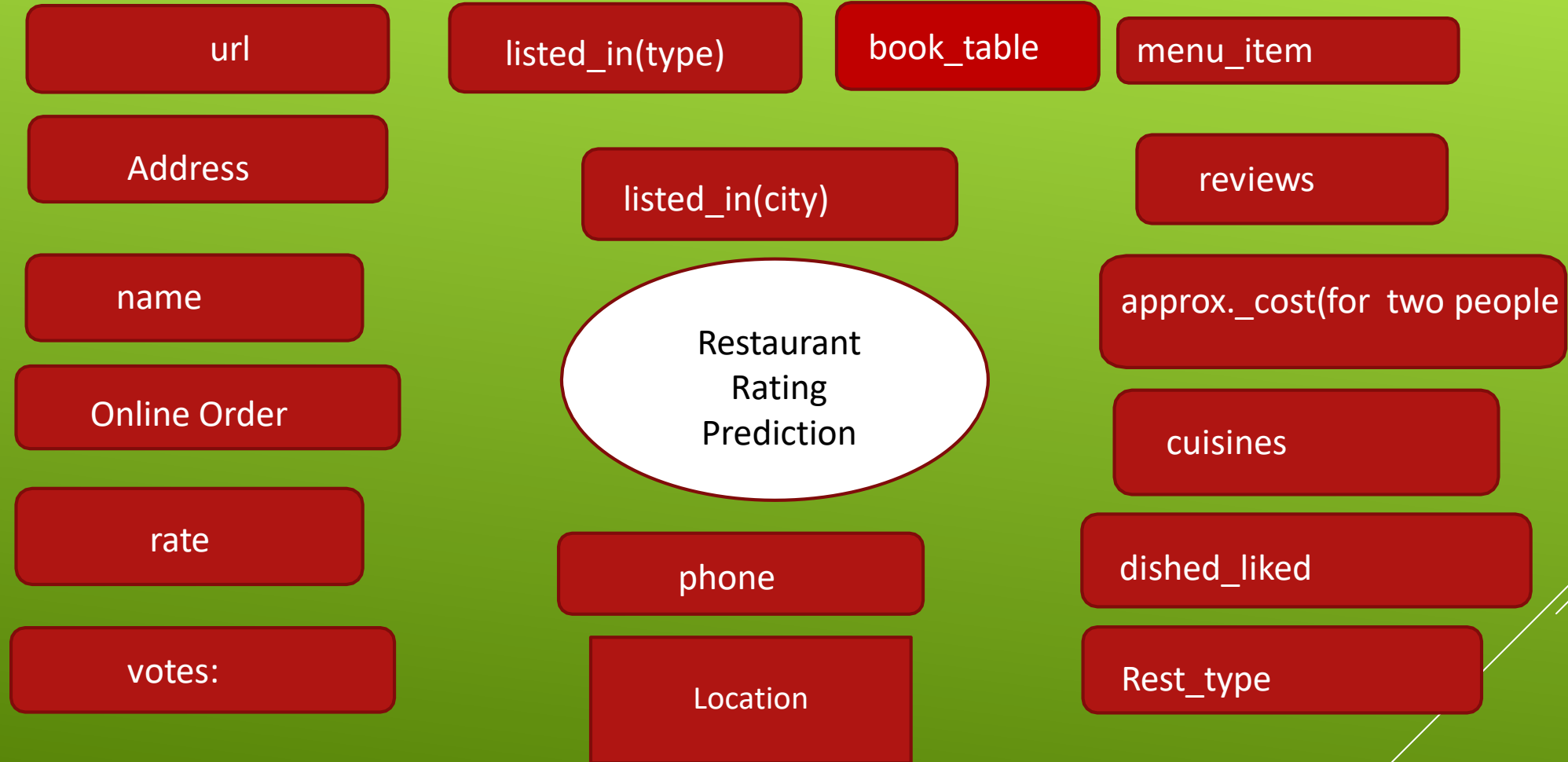
DATASET

url

listed_in(type)

book_table

menu_item

Address

listed_in(city)

reviews

name

Restaurant
Rating
Prediction

approx._cost(for two people

Online Order

cuisines

rate

phone

dished_liked

votes:

Location

Rest_type

RANDOM FOREST REGRESSOR

INTRODUCTION

- A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

- Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as **bagging**.

- Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap..

It gives us high accuracy as 90%.

MODEL TRAINING AND VALIDATION WORKFLOW
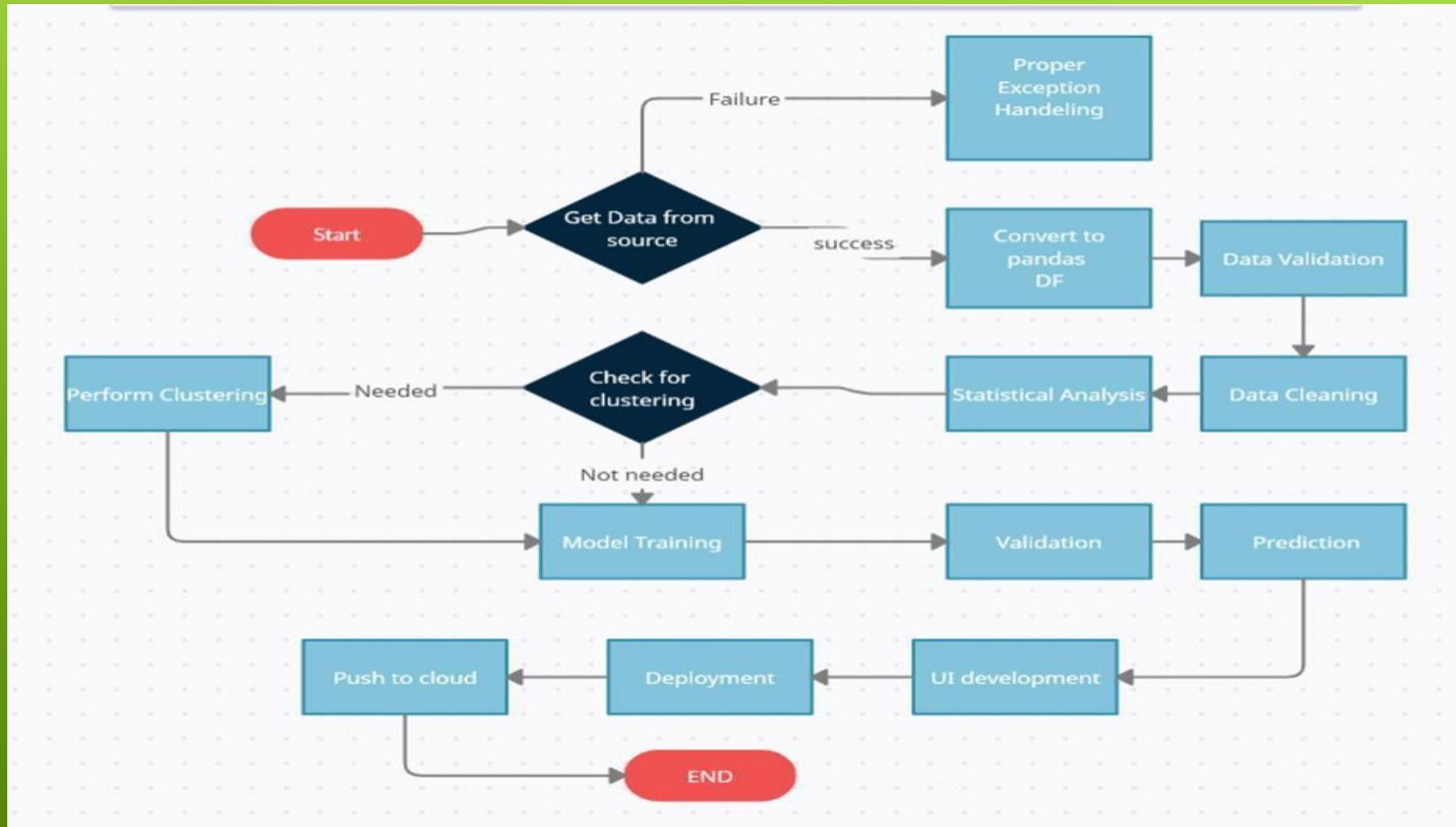
Data Collection

☐ 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

MODEL TRAINING AND VALIDATION WORKFLOW

Model Creation and Evaluation

☐ Various classification algorithms like Linear Regression, Random Forest, Decision Tree tested.

☐ Random Forest, Decision Tree and Linear regression were given less results as comparison to Extra tree regressor that what it has been chosen for the final model training and testing.

☐ Hyper parameter tuning was performed.

☐ Model performance evaluated based on r2 Score using Cross validation technique.

DEPLOYMENT

Model Deployment

☐   The final model is deployed using on AWS using Flask framework

FAQS

Q1.   What was type of data?

☐   It was a combination of numerical and categorical data.

Q2. How did you manage the null values in the dataset?

☐   Used dropna() attribute. Refer to code for better understanding.

Q3. What was the complete flow you followed in this project?

☐   Please refer to slide 4 for better understanding.

Q4. What were the techniques used for data pre-processing?

☐   Removing unwanted attributes.

- Removing outliers.

- Visualizing relation between independent variables and dependent variables.

- Converting categorical data into numeric values.

Q5. How did you train your model?

- Firstly, correlation was found among different variables.

- Then dataset was split into training and test size and different ML algorithms were used.

- Linear Regression, Decision Tree, Random Forest , Extra tree regressor were used among which Extra tree regressor gave the highest accuracy.

- DROPPING NULL VALUES.

Q6. WHAT WAS THE ACCURACY OF THE BEST MODEL OBSERVED?

- Random Forest Regressor showed the highest accuracy of 90% approx.

Q7. What challenges came up during the deployment of the model?

- An error in requirements.txt file can cause deployment failure.

- Requirement of correct and compatible versions of certain python libraries should be met.

- Cloning of Git repository should be done carefully. Use Git-LFS for files greater than 100mb.

THANK YOU