RESTAURANT RATING PREDICTION

-M V S K KARTHIK

INTRODUCTION

Whenever we try to order some food online the first thing most of us do is check the ratings and reviews of the restaurant to confirm whether the restaurant serves quality food in time. To know whether the restaurant can provide quality food or not, one first looks at the restaurant rating and reviews about the restaurant and quality of food. Bengaluru is one the top cities in India. Most of the people here are dependent mainly on restaurant food as they might be busy in their own works. With such an overwhelming demand for restaurants, studying the demography of a location becomes 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 us to simplify the tasks by helping us to predict and forecast the fututre. This study demonstrates how different Regression algorithms can forecast the rating of restaurants so that one can make a decision whether to buy food(online) from a particular restaurant or not based on ratings and reviews. Different regression algorithms such as Decision Tree, Random forest, XGboost, have been tested and compared to predict the ratings and the algorithm which performed better was chosen.
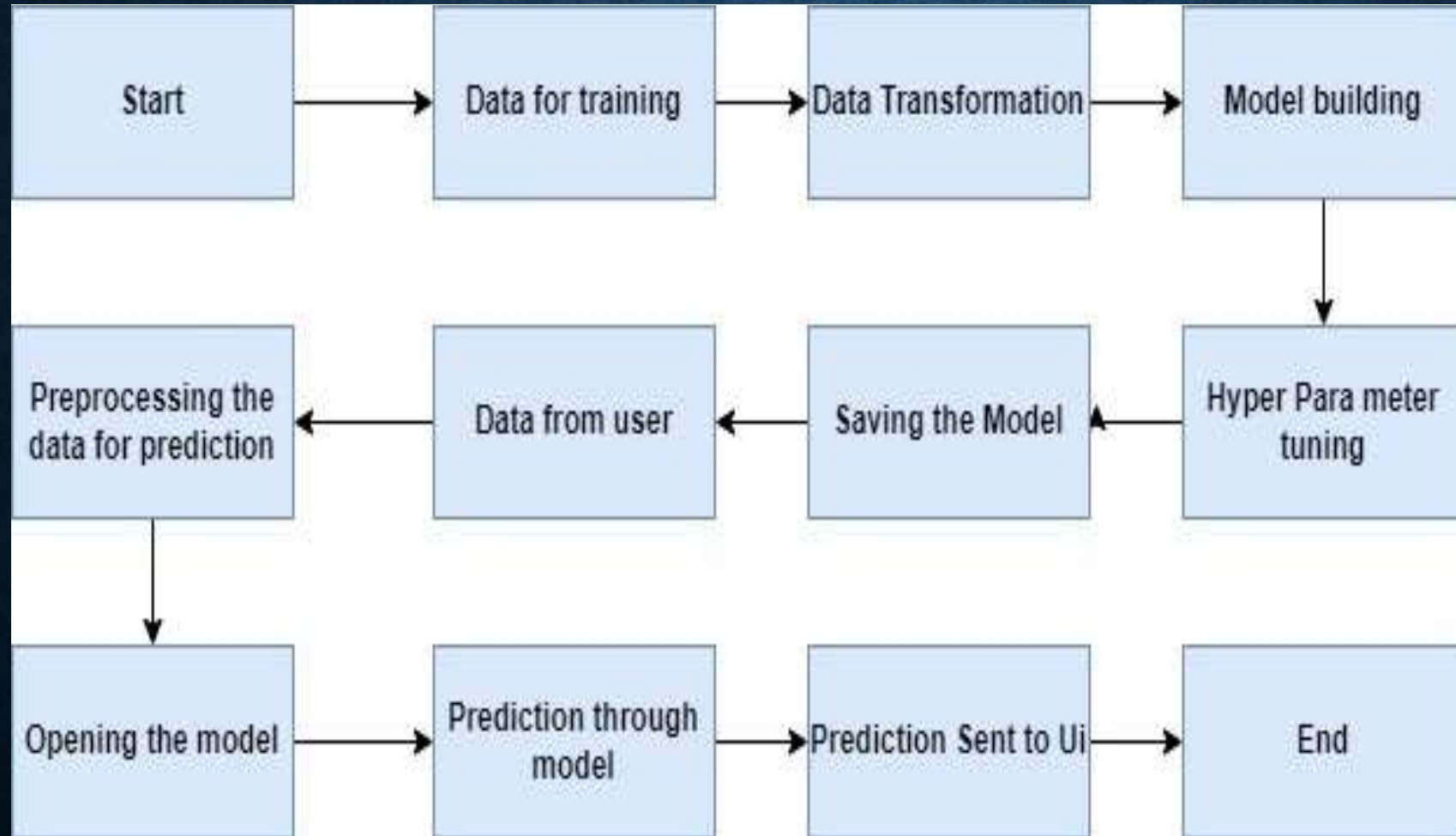
OBJECTIVE

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_orde

cuisines

rate

phone

dished_liked

votes:

location

Rest_type

DATA ANALYSIS

RANDOM FOREST REGRESSOR

INTRODUCTION

☐ 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**. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying oni ndividual decisiontrees.

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.

Random forest regressor being ensemble technique tends to give more accurate result.This is because it works on principle ie number of weak learners when combined forms strong estimator.Even if one or more decision trees are prone to noise,overall results tend to be correct.

It has given us high accuracy as 93%.

MODEL TRAINING AND VALIDATION WORKFLOW
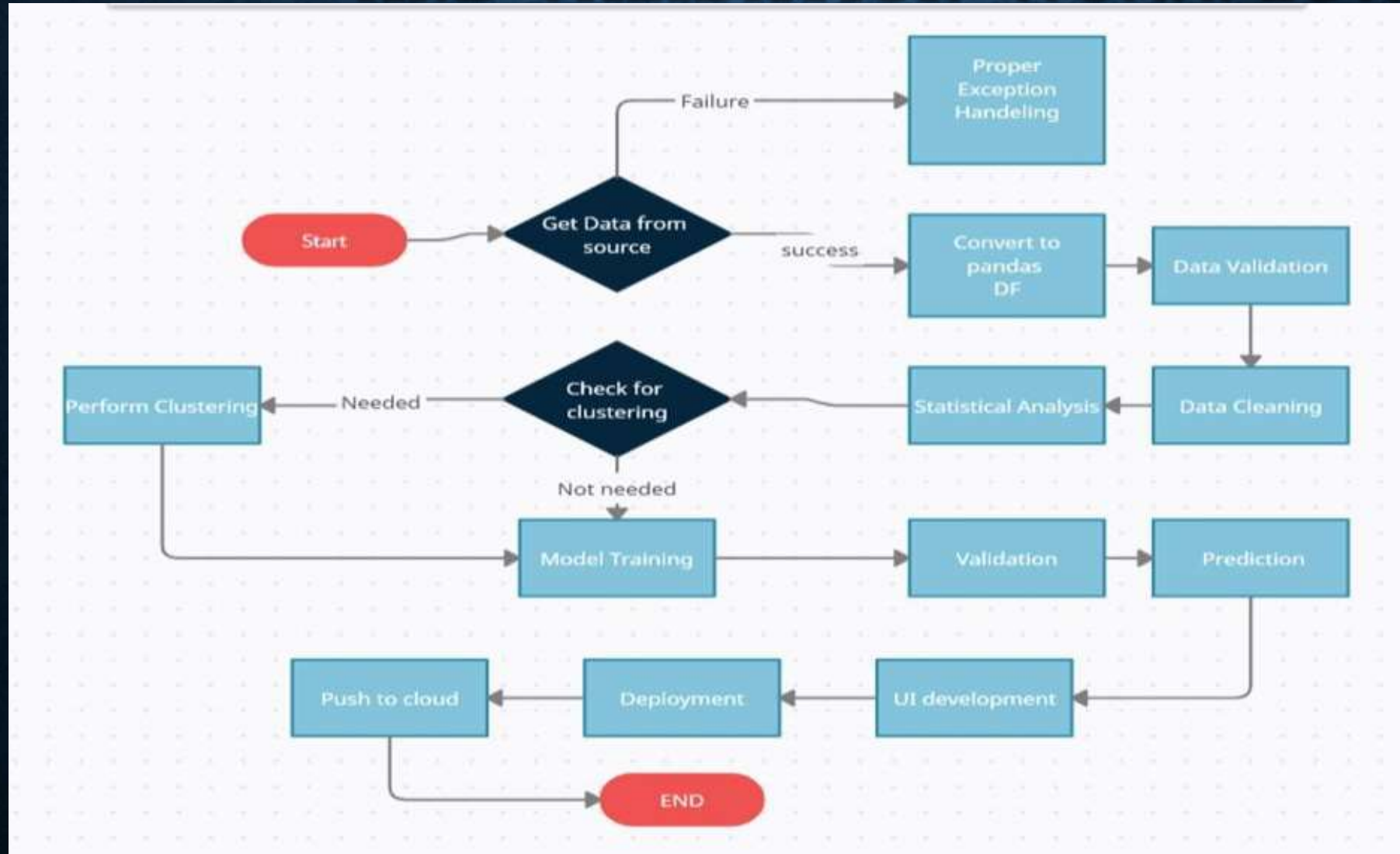
Data Collection

☐ Zomato Restaurant Data Set from Kaggle.

Data Pre-processing

☐ Missing values handling by Simple imputation (mode strategy)

☐ Other missing values which are less when compared to size of the dataset are dropped using dropna method

☐ Categorical features handling by ordinal encoding and label encoding

MODEL TRAINING AND VALIDATION WORKFLOW

Model Creation and Evaluation

- Various classification algorithms like XGboost, Random Forest, Decision Tree tested.

- Decision Tree and XGboost were given less results as comparison to Random forest regressor that what it has been chosen for the final model training and testing.

- Model performance evaluated based on r2 Score.

DEPLOYMENT

Model Deployment

- The final model is deployed using Streamlit

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.

Dropping null values.

- Visualizing relation between independent variables and dependent variables.

- Converting categorical data into numeric values.

Q5. How did you train your model?

- Firstly, exploratory data analysis was done among different variables.

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

- Decision Tree, Random Forest , Xgboost regressor were used among which Random forest regressor gave the highest accuracy.

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

☐ Random forest showed the highest accuracy of 93% 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