Restaurant Rating Prediction

BY: ANUJ GUPTA

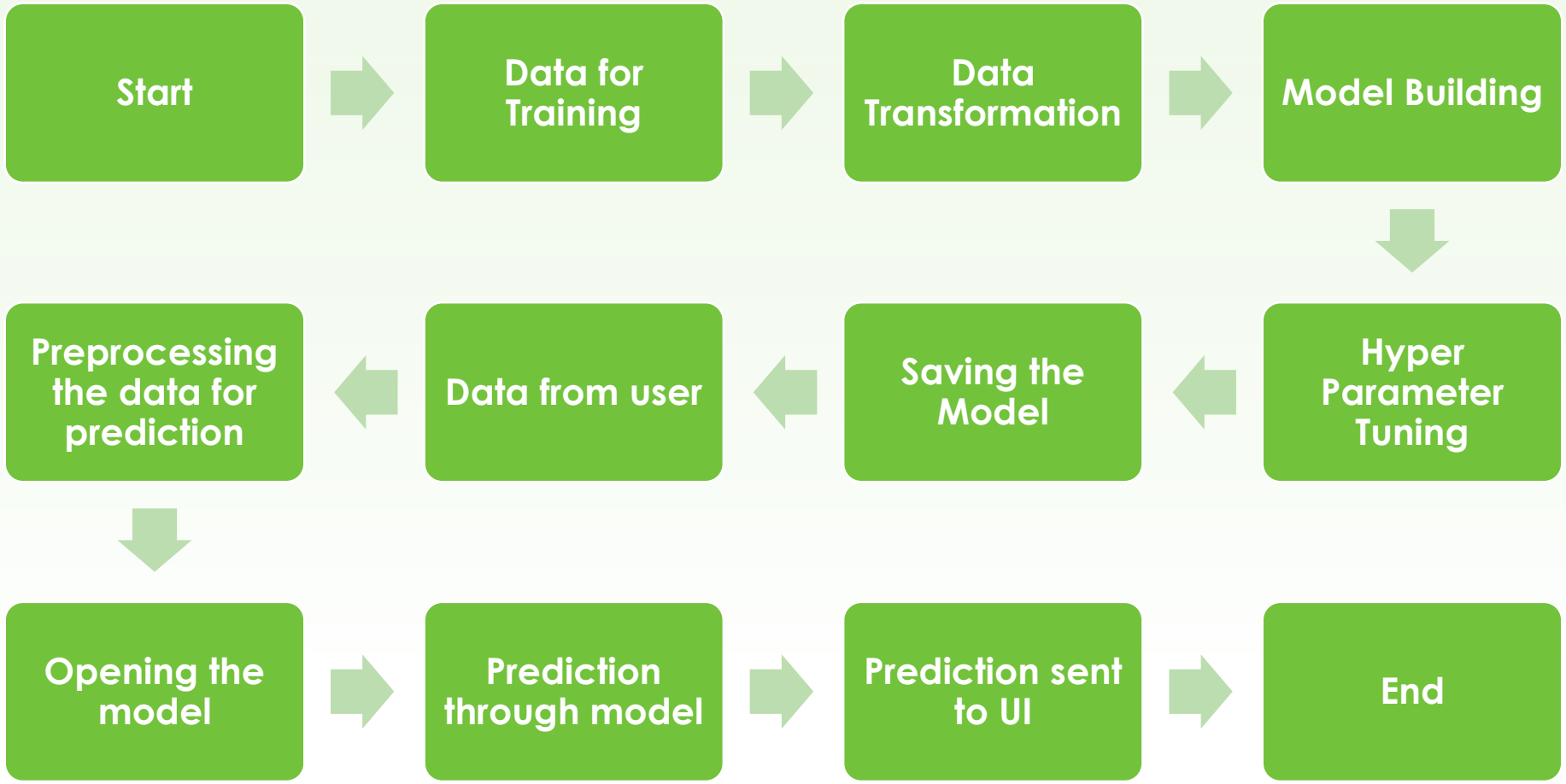
INTRODUCTION

- In today's era of online food ordering, one of the primary factors influencing our decision is the ratings and reviews of restaurants. Before placing an order, we often rely on the feedback from other customers to ensure that the restaurant provides high-quality food and timely service. This is particularly true in Bengaluru, one of India's top cities, where many individuals heavily rely on restaurant food due to their busy schedules. With such a high demand for restaurants, understanding the demographics of a location becomes crucial.
- This study showcases the application of different regression algorithms, including Random Forest, XGBoost, LGBM to forecast restaurant ratings. By analyzing historical data, including ratings and reviews, these algorithms can identify patterns and relationships that influence a restaurant's rating.

OBJECTIVE

- The objective is to harness the power of technology and data analysis, to enhance the online food ordering experience in Bengaluru and enable individuals to make well-informed decisions based on ratings and reviews.

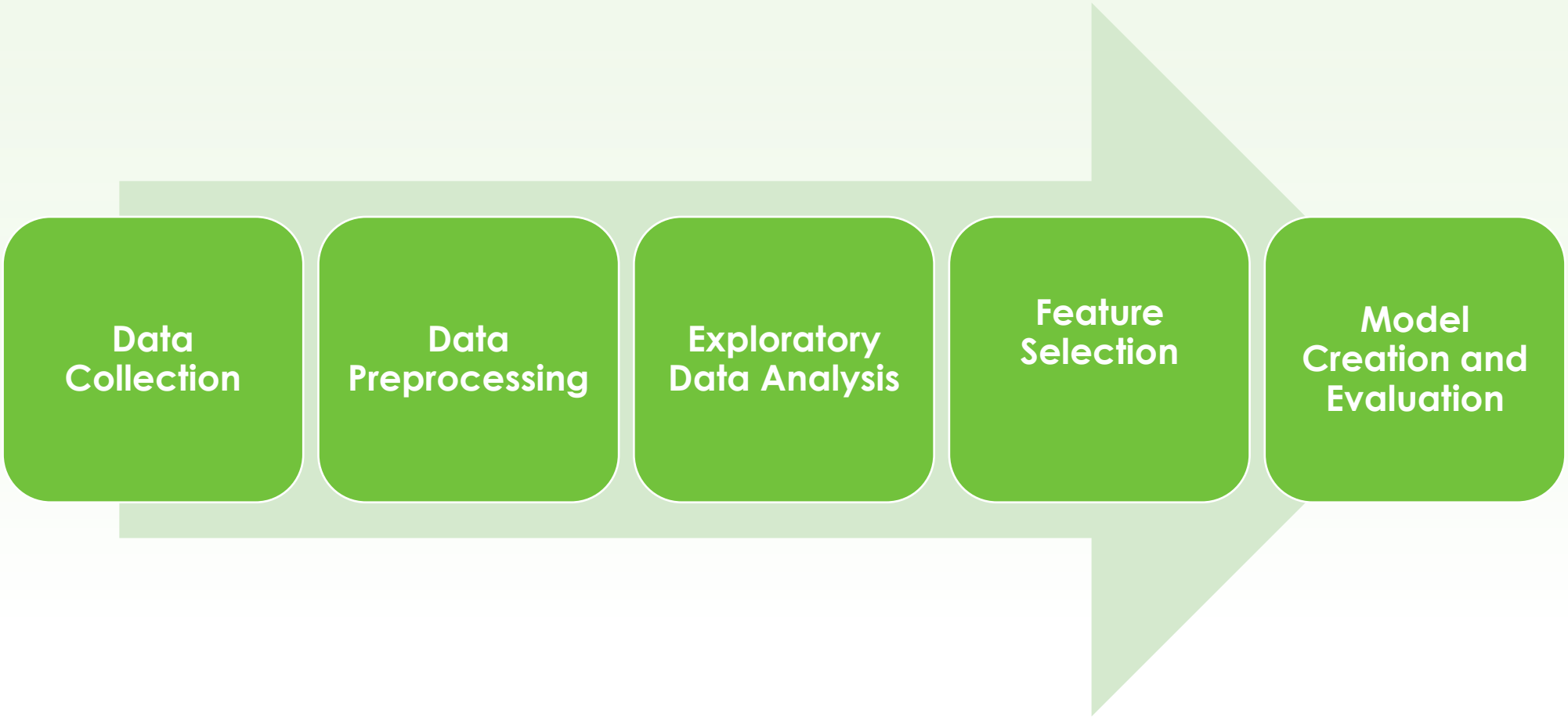
ARCHITECTURE



DATASET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   url                          51717 non-null  object
 1   address                      51717 non-null  object
 2   name                         51717 non-null  object
 3   online_order                 51717 non-null  object
 4   book_table                   51717 non-null  object
 5   rate                         43942 non-null  object
 6   votes                        51717 non-null  int64
 7   phone                        50509 non-null  object
 8   location                     51696 non-null  object
 9   rest_type                    51490 non-null  object
 10  dish_liked                   23639 non-null  object
 11  cuisines                     51672 non-null  object
 12  approx_cost(for two people)  51371 non-null  object
 13  reviews_list                 51717 non-null  object
 14  menu_item                    51717 non-null  object
 15  listed_in(type)              51717 non-null  object
 16  listed_in(city)              51717 non-null  object
dtypes: int64(1), object(16)
memory usage: 6.7+ MB
```

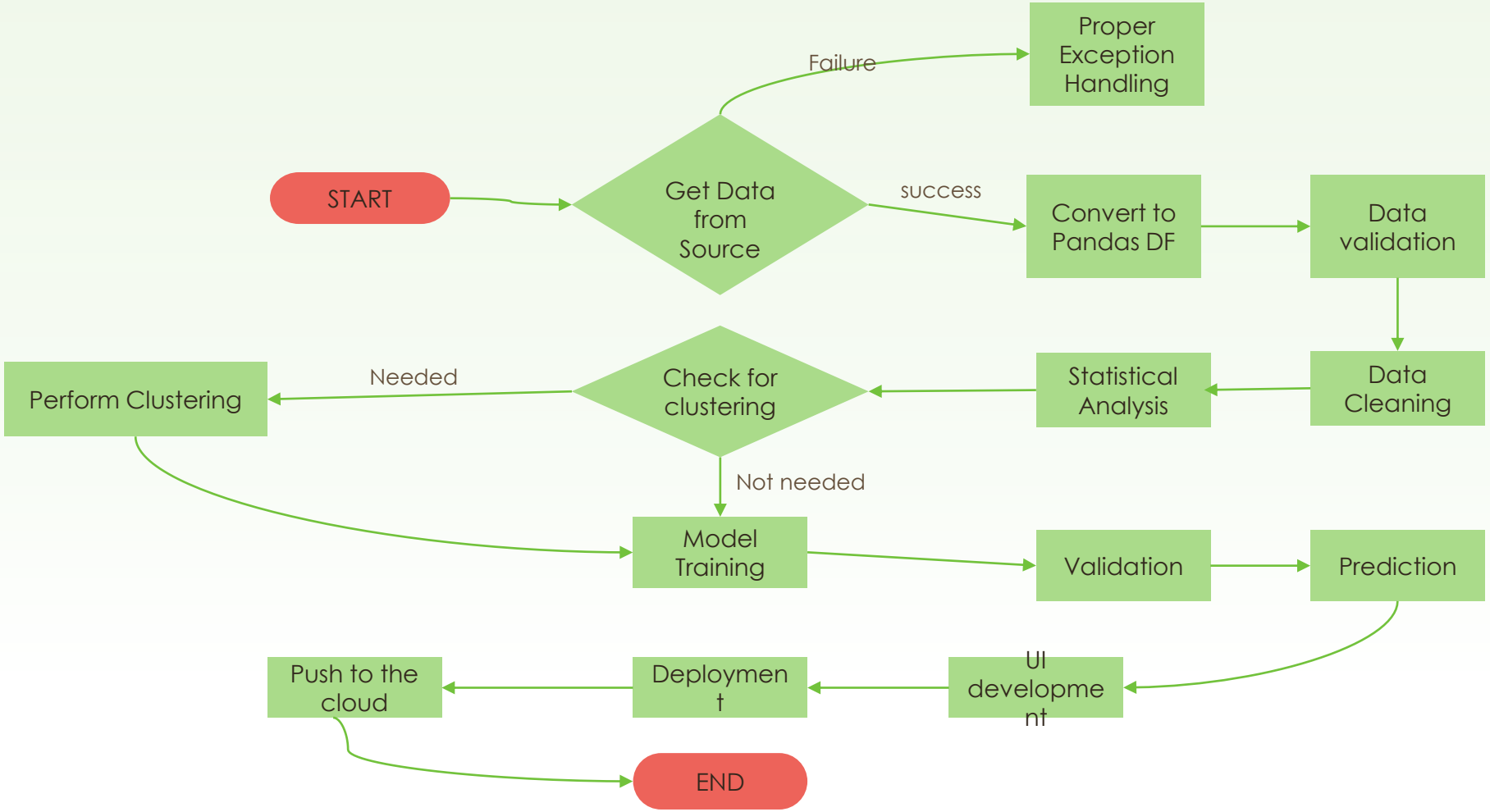
DATA ANALYSIS



LGBM REGRESSOR

- LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be efficient and provides high-performance implementations of gradient boosting algorithms. LightGBM is commonly used for regression and classification tasks and is known for its speed and accuracy.
- In the project, LGBM Regressor was used. It provides an interface to train and use gradient boosting regression models using the LightGBM algorithm.
- By using the LGBMRegressor, the project leveraged the benefits of LightGBM's efficient and accurate gradient boosting algorithm for regression tasks. The hyperparameter optimization process helped find the best combination of hyperparameters, and the trained model was saved for future use.

MODEL TRAINING AND VALIDATION WORKFLOW



MODEL TRAINING AND VALIDATION WORKFLOW

Data Collection : Zomato Bangalore Restaurants dataset.

Data Pre-processing:

- Data Cleaning.
- Data Integration.
- Data Transformation.
- Feature Selection.
- Feature Engineering.
- Data Normalization.
- Handling Imbalanced Data.
- Data Splitting.
- Data Scaling.

MODEL TRAINING AND VALIDATION WORKFLOW

Model Creation and Evaluation.

- Various machine learning algorithms like Random Forest, XGboost, LGBM.
- **LGBM regressor** performed the best as compared to random forest regressor, xgboost regressor.
- Model performance was evaluated based on **negative MSE (Mean Squared Error).**

Model Deployment

- The final model was deployed using **Streamlit.**

THANK YOU