Flight Fare Prediction

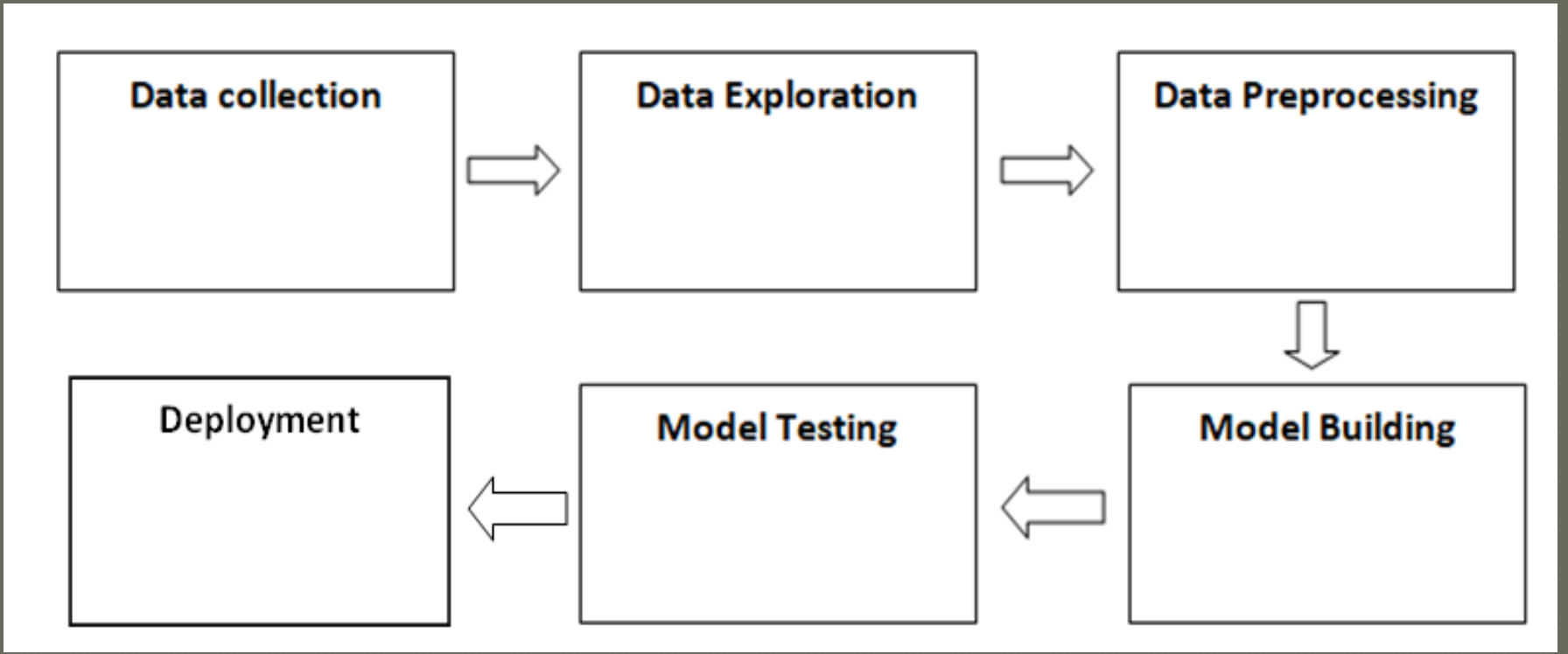
Objective:

The objective of the project is to create a solution to find the ticket price of the user's preferred flight using machine learning models

Benefits:

- Makes it easy for users to plan their trips.
- Helps to guess the range of the flight price for a preferred airline or route.

Methodology



Model Training:

- Dataset

  The dataset for the task is downloaded from Kaggle in csv format for model training

- Data Preprocessing

  - Performing data exploration to get insight of data like understanding trends in the data etc.
  - Duplicate rows were removed
  - Encoded categorical values so that they can be understood by the regession models.

Regression –

- After data preprocessing, the data is fed into each of the models for prediction.
- First the data into split into train-test sets in 80:20
- Then the training dataset was fed into different machine learning models like Linear regression, Random Forest, Decision tree, Extra trees regression, Gradient boosting and Ridge regression for training
- The models were then tested on the testing set to find the model with highest R2 score

Prediction

- It was found that Random Forest had the highest R2 score of among the proposed machine learning models
- The best model was pickled to be used in further cases and for deployment of the model as an API to be used for prediction.

Q & A:

Q1) What’s the source of data?

The data for training can be obtained from Kaggle in csv format

Q 2) What was the type of data?

The dataset was a combination of numerical and Categorical variables.

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

The project started with data collection, then data exploration, data preprocessing, model building, model training and finally deployment

Q 4) What techniques were you using for data pre-processing?

- Visualizing relation between the dependent and independent variables
- Cleaning data and imputing if null values are present.
- Converting categorical data into numeric values.
- Removing duplicate rows

Q 5) How training was done or what models were used?

- The dataset was split into train and test sets
- Categorical variables were encoded to be understood by the regression models
- Algorithms like Linear regression, Random Forest, Decision tree, Extra trees regression, Gradient boosting and Ridge regression were used for training.

Q 6) How Prediction was done?

After training the models, the models were tested with the testing data where the actual prediction is done and this also outputs the model with the highest R2 score and least Mean Absolute Error (MAE) and Mean Sqaured Error (MSE)

- Q 7) What are the different stages of deployment?
  - The model was deployed as an API using FastAPI where the user can input values to the relevant variables to predict the fare price of their preferred flight.