**Predict fare of airlines tickets**

**Business Objective**

This project aims to predict ticket prices for upcoming flights to help customers in selecting the optimum time for travel and the cheapest flight to the desired destination. A random forest regression model is applied to forecast the flight prices based on data.

**Data Description**

The dataset has 1 excel file,

* Data has 11 columns and 10683 rows

The data consists of the following attributes:

* Airline: Name of Airline
* Date of Journey: Date of journey of flight
* Source: Source of flight
* Destination: Destination of flight
* Route: Route of flight
* Dep\_Time: Departure time of flight
* Arrival\_Time: Arrival time of flight
* Duration: Duration of flight
* Total\_Stop: Total stops of flight
* Additional Information: Any information about flight. For example, flight includes meal or not

**Aim**

Build a machine learning model to predict the fare of airlines tickets by understanding and analysing features of data.

**Tech stack**

* Language - Python
* Libraries – numpy, pandas, matplotlib, seaborn, sklearn, pickle

**Approach**

1. Importing the required libraries and reading the dataset.

* Understanding the dataset

2. Exploratory Data Analysis (EDA) –

* Data Visualization

3. Feature Engineering

* Dropping of unwanted columns
* Removal of null values
* Checking for multi-collinearity and removal of highly correlated features

4. Model Building

* Performing train test split
* Linear Regression
* Decision Tree Regressor
* Random Forest Regressor
* mutual\_info\_regression

5. Model Validation

* Accuracy score
* r2\_score
* MAE
* MSE
* RMSE
* MAPE
* best\_params\_
* best\_estimator\_
* best\_score\_

6. Hyper parameter Tuning (RandomizedSearchCV)

7. Creating the final model and making predictions

8. Save the model with the highest accuracy in the form of a pickle file.