**INTRODUCTION**

A flight fare (otherwise known as an airfare) is the fee paid by a passenger for air transport and is made up of the charge for a passenger to fly from an origin to destination and includes the conditions, rules and restrictions for travelling on the airfare.

Flight fares are typically made up of fare and rule components that define the airfare product, services and price and includes: origin/destination pair, fare class, one-way/round-trip indicator, fare amount, validity dates, mileage and other rules.

**PROBLEM STATEMENT**

The aviation industry faces significant challenges in setting competitive and profitable flight fares. Airlines must balance the need to fill seats with the goal of maximizing revenue. Given the dynamic nature of flight prices, influenced by various factors such as demand, seasonality, time until departure, and market competition, predicting flight fares accurately is a complex task.

On the other hand, customers seek the best possible prices for their travel plans. Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, and it will be a different story.

**OBJECTIVE**

The objective of this project are :

* To develop a machine learning model that accurately predicts flight fares based on historical flight data.
* To predict flight fare price using machine learning algorithm which gives the best route to purchase ticket.
* To provide travelers with valuable insights for better planning and making decision on airlines to pick and which routes to take in order to get the greatest deal.
* To attract customers and maximize the sales

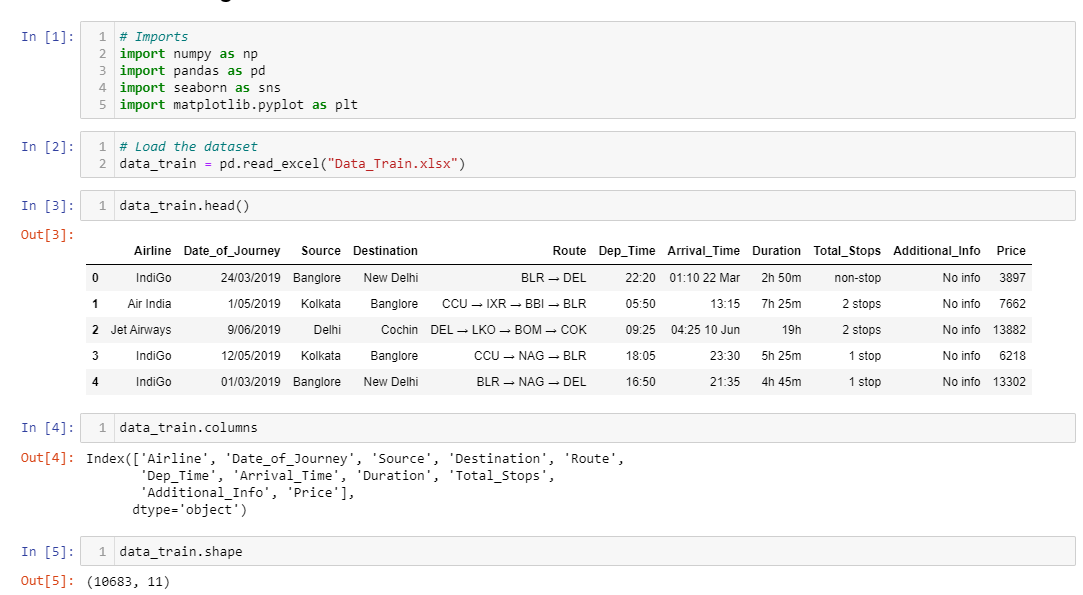
By achieving these objectives, the project aims to create a reliable tool for flight fare prediction, benefiting both airlines and customers in the competitive aviation market.

**METHODOLOGY**

Data Collection

The act of obtaining, acquiring, and combining the data that will be used to develop, test, and verify a machine learning model is known as data collection in machine learning. This step plays crucial role in implementation. Here data is collected from flight fare dataset which is imported from Kaggle. The dataset consists of both categorical data and numerical data. The categorical data includes source, destination, type of airline, additional info and numerical data includes arrival and departure dates, number of stops.

There are 11 columns (each represents a feature) and 10683 rows in this large dataset.



Data Preprocessing

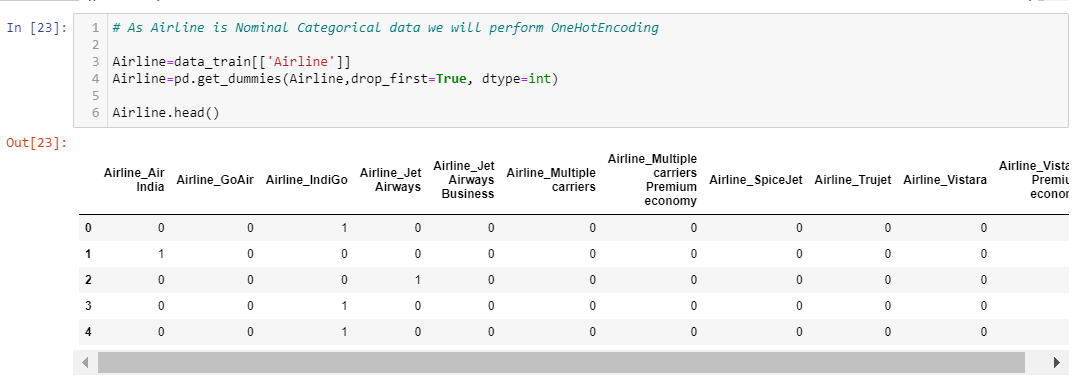
Data preprocessing means nothing but cleaning data, which can be used for model training and testing. By this step we can make our data useful for model training purpose. Data preprocessing involves cleaning, transforming, and preparing the data for data analysis. The sub steps involved in the data preprocessing are:

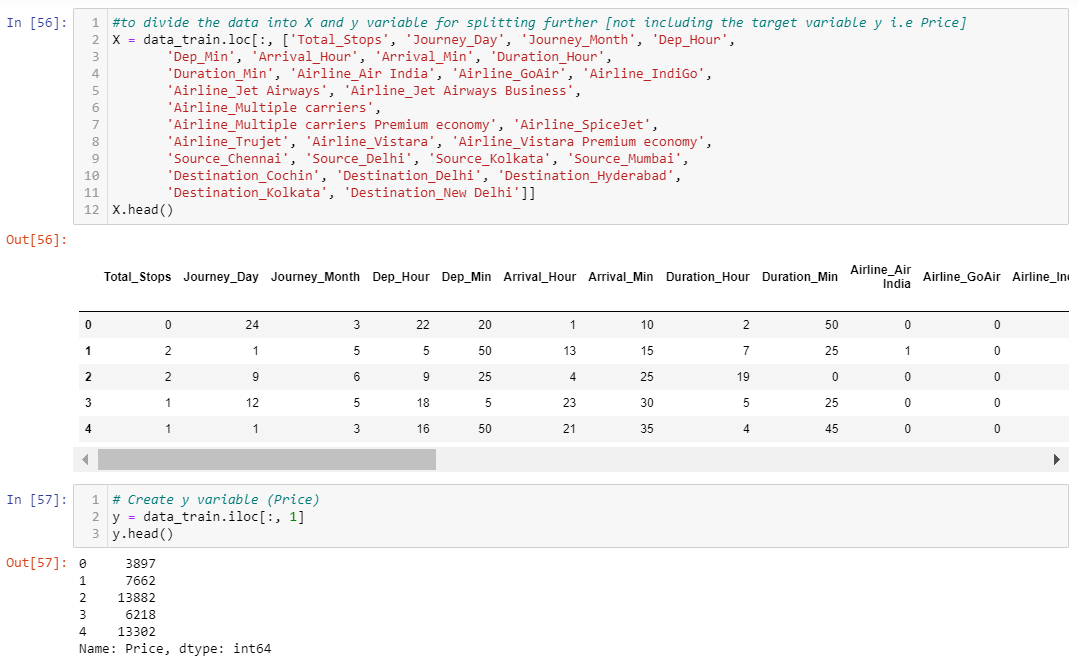
1. Data Cleaning: In this step the null values are removed, missing values are removed and if any duplicates are present that are also removed.





1. Feature selection and engineering: In this step the features of our model are extracted and all the relevant features are used for model training. In dataset it contains date of journey, arrival date, departure date columns and all the numerical values are extracted as Departure hour, departure minutes, arrival hour, arrival minutes, journey day, journey month. As dataset contains both categorical and numerical features, by using 'On hot encoding' method for nominal categorical data and 'label encoding' for ordinal categorical data was used to convert the categorical values to numerical values. The dataset consists of categorical variables like airline, source, destination, route, total number of stops and additional info.



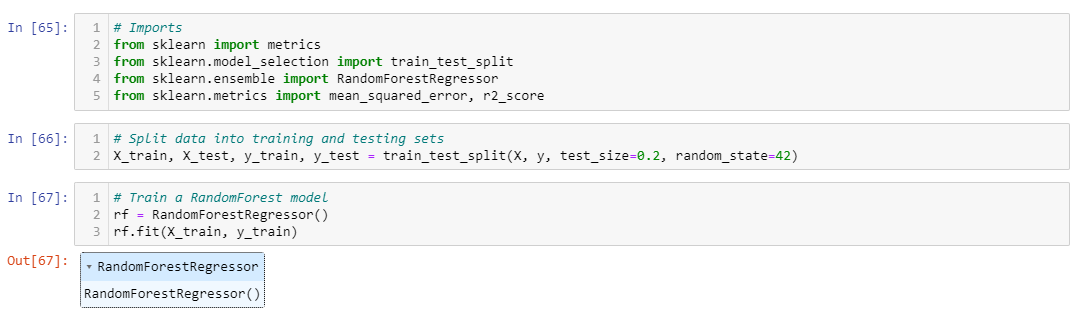


1. Data Splitting: This step involves splitting our data into two parts for training and testing purpose. For model training 80 percent of data was used by using Random Forest regressor model was trained.

The machine learning algorithms are:

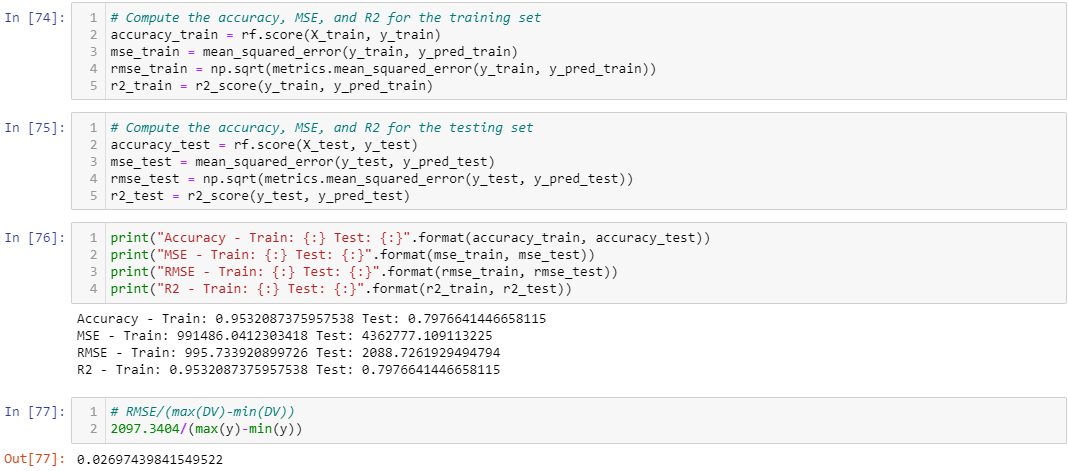
Random Forest Regressor

Random Forest regressor uses multiple decision trees to perform regression tasks. It is an example of ensemble learning. Random forest is a Supervised Learning algorithm which uses ensemble learning approach for classification and regression. Decision trees are sensitive to the specific data on which they are trained. If the training data is changed the resulting decision tree can be quite different and in turn the predictions can be quite different. Also, Decision trees are computationally expensive to train, carry a big risk of overfitting, and tend to find local optimal because they can’t go back after they have made a split to address these weaknesses, we turn to Random Forest.



1. Model evaluation: This is an important step in our project, as it helps us to measure the performance and accuracy of our model. Test data is used for model evaluation. Here, we employed Cross-validation for model evaluation. This method divides the data into k-subsets, called folds. the model is trained on k-1 folds and tested on the remaining fold. this process is repeated k times, so that each fold is used as a test set once. The average performance across all k-folds is reported as the final result. The metrics that are used for model evaluation purpose are:

* Root Mean Squared Error (RSME): It gives the root of the average squared difference between the actual values and the predicted values for a regression problem.
* Mean Absolute Error (MAE): It gives the absolute difference between the actual values and predicted values. The higher negative mean values indicate the better performance of model.
* R-Squared: This metric measures how well the regression model fits the data, by comparing it to a baseline predict the mean value. It shows how much variation in the data is explained by the model.



**CONCLUSION**

Flight fare prediction using Random Forest leverages the power of ensemble learning to provide accurate and robust fare predictions. By properly collecting, preprocessing, and analyzing historical flight data, a Random Forest model can significantly enhance the ability to predict future flight prices. Ultimately, the choice of the algorithm should be guided by experimentation and validation using historical flight price data.

The result of the research will potentially is being used to help stakeholders identify the factors that cause flight tickets to be expensive. Stakeholders can maintain price stability by understanding the factors that cause flight price inflation, allowing people to keep enjoying airline services.

Hence, we have successfully built the Flight Price Prediction Machine Learning Model to predict the price of flights which will help us to select the best possible travel route and to reach our destination according to our own demand and utility.