

DYNAMIC PRICE PREDICTION OF AIRLINES

Problem Statement:

Develop a machine learning model to predict the airline price using the features provided in the dataset.

Problem Overview:

Airline companies are following the dynamic pricing model. They have fixed number of seats in their plane. They are following the dynamic pricing strategy for selling seats to increase the revenue. The pricing will change based on the demand. The demand is dependent on many factors such as holiday, peak time, short/long route, number of stops, economy class etc,. Airlines have to model the demand, so that they can sell all the seats with more profit. Modelling the demand is critical because poor modelling may leave more seat empty or lead to less revenue.

The objective of this project is to model the airline price change using the different features. This predictive model will help the passenger to make purchasing decisions by predicting how air ticket prices will evolve in the future.

The following three files will be provided to work on this project

airfare.xlsx – It contains the characteristics of each airline timing, route and fare information

airdistance.csv – It contains the scrapped distance between different cities.

This dataset contains 10 following independent features to train a predictor (Price).

Airline: The name of the airline.

Date: The date of the journey

Departure Station: The source from which the service begins.

Arrival Station: The destination where the service ends.

Route map: The route taken by the flight to reach the destination.

Departure Time: The time when the journey starts from the source.

Arrival Time: Time of arrival at the destination.

Journey Time: Total duration of the flight.

Stops: Total stops between the source and destination.

Extra_Info: Additional information about the flight

Perform EDA and required Pre-processing:

Exploratory Data Analysis and Visualizations need to perform on the input features, to understand the characteristics of data. Proper pre-processing (treating NaN, cleaning, structuring, scaling/normalizing) will lead to better prediction. Statistical testing and correlation analysis will help to understand the significance of feature sets.

Apply Machine Learning Algorithms:

At the end of performing the EDA, one must get the clarity over which predictive Machine Learning algorithm can be applied on the data to predict the airline price. Apply Regression models to predict the airfare and measure the performance of model using appropriate metrics.

Feature Engineering

Perform feature engineering to understand the impact of each feature on prediction and build the model with only significant features. One can also think of extracting/deriving few more features from the data. Create some new feature using the domain knowledge with respect to this dataset and improve the model performance. Also, make use of the data file 'distance.csv' to improve the model performance. 'distance.csv' file have the information about the distance between source and destination station in kms.

Regularization

The prediction model output reliability can be improved using regularization techniques and parameter tuning. One can play with the different parameters for each algorithm, using GridSearchCV and RandomSearchCV packages to identify the optimal values which produce the maximum performance.

Create a Pipeline and Save Predictive Model:

Finally, the developed predictive model can be permanently saved in hard-drive with all the required pre-processing steps and whenever the new data to be tested, this saved model alone can be loaded and applied on test data to predict the output.

Submission Deliverables:

1. The complete script should be submitted in .ipynb format with all required outputs and visualizations.
2. PPT deck (4 to 8 slides) must be submitted to highlight the outcomes and significance of your project implementation.