FLIGHT PRICE PREDICTION BONUS-PROJECT
[CSL2050]

___

Suyash Bansal (B20CS076)

Abstract:

This paper reports our experience with building a ML model for predicting the prices of the flight depending upon various parameters. We have a dataset with some input features on which model is been trained. Here the ML model is developed in such a way that it is capable of judging the prices of flight tickets with some specific information provided to it. Here several of the ML algorithms is been used such as Random Tree Regression, linear regression etc. to train the model on input dataset.

Introduction:

This Flight Price prediction model is basically a model which predicts the upcoming prices of flights based on the previous trends obtained from the training dataset. Prices depends upon various factors like in the dataset as mentioned can be 'Airline' , 'Duration', 'Source' ,'Destination' and many such parameters as mentioned in dataset.

Dataset:

●It is named as Flight_price_dataset with the csv file format.

● It consist of total 10683 rows which are total datasentries with respect to the parameter on which price is estimated and total of 11 coloumns which are the features on which price of flights are given in the dataset.

●The dataset is been divided into the train and test dataset as test_size=0.3 so the 30 % is testing data and the remaining is the training data.

Methodology:

●Overview

    Algorithms used in this model are:
        Random Forest Classification
        KNN
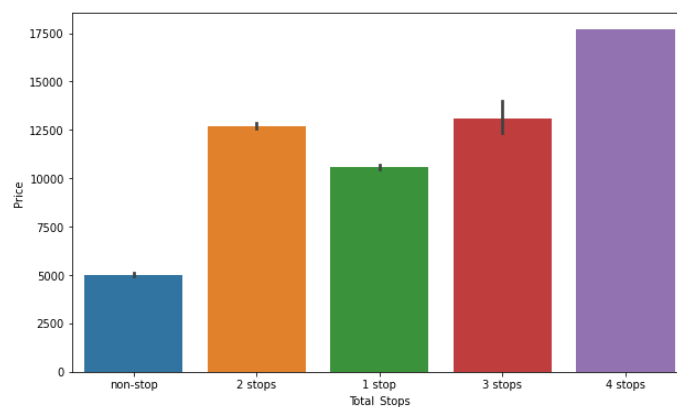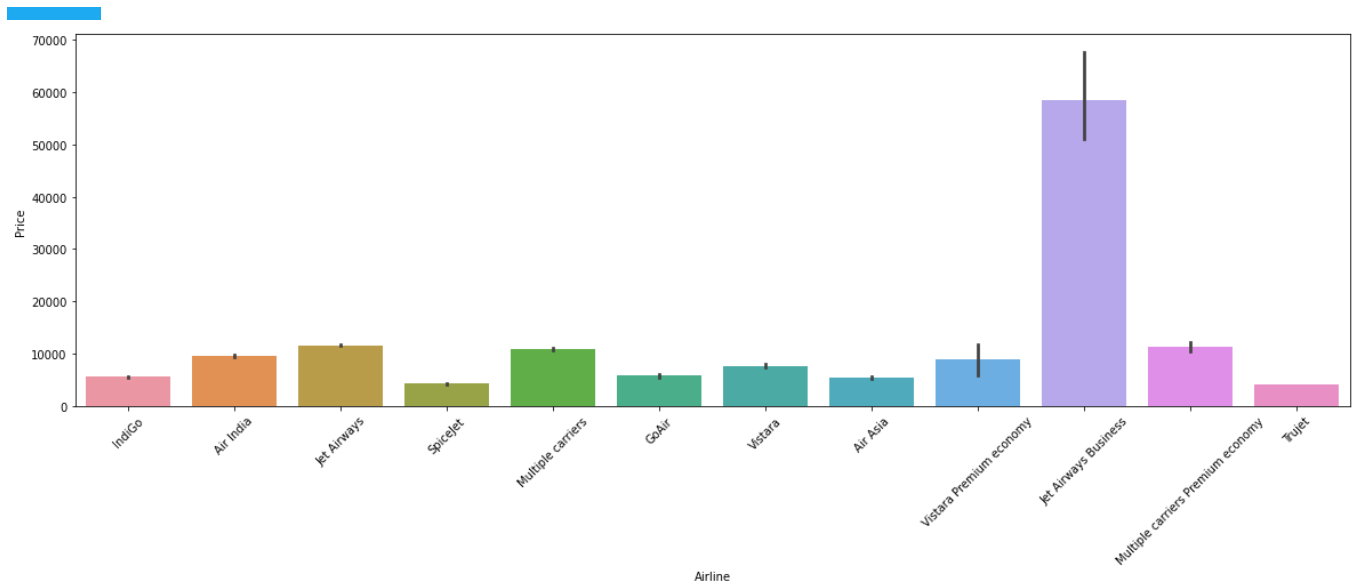   Logistic Regression
        SVM
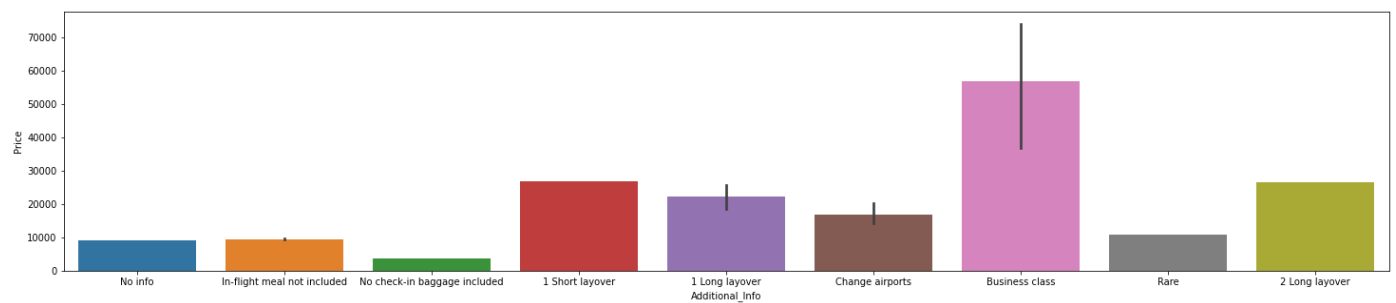   Decision Tree Regression

Data Visualization:

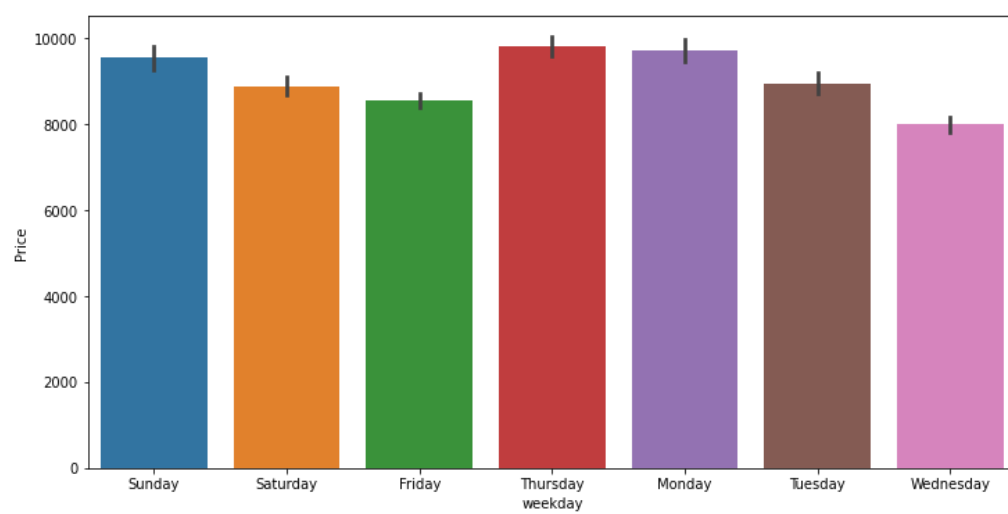Here several parameters vs price comparison is made using the bar plot as shown below
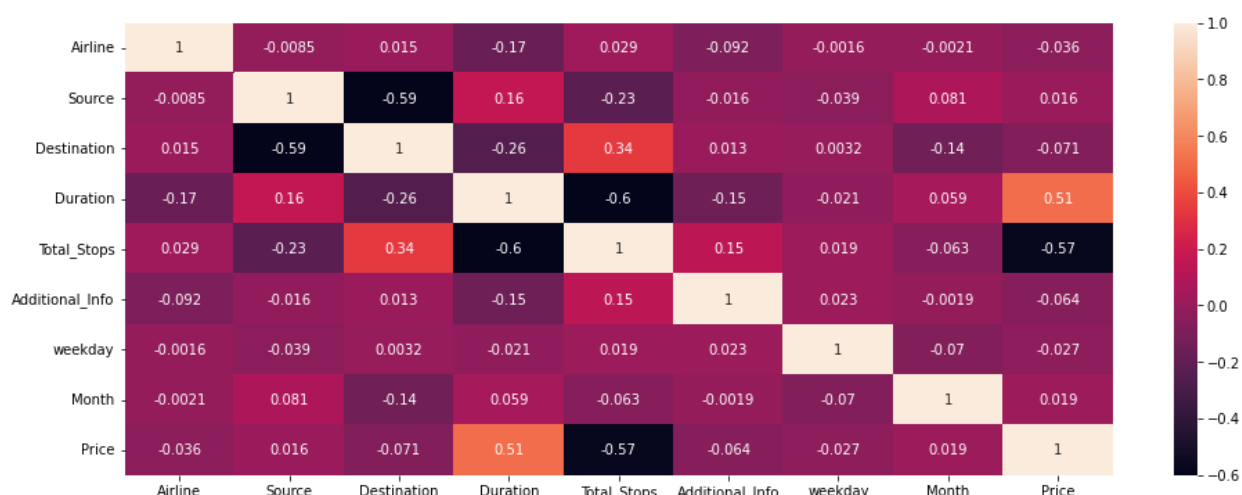

Price vs Total_stops

Price vs Airlines

Price vs Additional_Info

Price vs Weekday

●Heat_Map

It is created by normalization of features as shown in fig above diagnol values are 1 when matched with itself.

Performance Evaluation :

Here is the table of the accuracy with respect to the model selected for predicting the price of flight which are mentioned above in the report.

ML-Model	Accuracy
Linear Regression	15.840114 %
Lasso	34.617045 %
Random Forest Regression	81.651941 %
K Nearest Neighbors	68.208693 %
Support Vector Machines	2.1564 %
Decision Tree Regression	58.137 %

Conclusion:

From the above given accuracies we can clearly see that Random Forest Regression is having the highest accuracy score of nearly 81.65% when using 50 trees. In this on applying ensemble learning techniques we found that it increased significantly from 74 to 81 percent but later on it did not show much of progress after that it became almost constant.