



Symbiosis Skills and Professional University

Kiwale, Pune

PROJECT REPORT

On

“Flight Price Prediction”



Submitted by

Shyam Adsul

(2001206003)

ML-Batch-I

Under The Guidance of

Trainers' Name: Uday Royzada

STUDENT DECLARATION AND ATTESTATION BY TRAINER

This is to declare that this report has been written by me No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I aver that if any part of the report is found to be plagiarized, I shall take full responsibility for it.

Signature of student

Name of student: Shyam Adsul

Registration Number: 2001206020

Signature of trainer

Name of trainer: Uday Royzada

CERTIFICATE

This is to certify that the report entitled, “**Flight Price Prediction**” submitted by “**Shyam Adsul**” to Symbiosis Skills and Professional University, Pune, Maharashtra, India, is a record of bonafide Project work carried out by him under my supervision and guidance and is worthy of consideration for the completion of certificate course in ‘Machine Learning Engineer And AI Analyst’.

Signature of Trainer

Name of Trainer

Date: / 04 / 2021

Supervisor

Supervisor

Date:

ACKNOWLEDGEMENTS

I am very glad to take this opportunity to acknowledge all those who helped me in designing, developing and successful execution of my project on “Flight Price Prediction”.

I would like to express my sincere and deep gratitude to our project guide Prof. Uday Royzada, for his valuable guidance and suggestions.

INDEX

SR.NO	Index	Page no.
1	Acknowledgement	4
2	Abstract	6
3	Project overview	7
4	Business Understanding	8
5	EDA	9
6	Model building	11
7	Future Scope	13
8	Conclusion	14

Abstract

The airline implements dynamic pricing for the flight ticket. According to the survey, flight ticket prices change during the morning and evening time of the day. Also, it changes with the holidays or festival season. There are several different factors on which the price of the flight ticket depends. The seller has information about all the factors, but buyers are able to access limited information only which is not enough to predict the airfare prices. Someone who purchase flight tickets frequently would be able to predict the right time to procure a ticket to obtain the best deal. Many airlines change ticket prices for their revenue management. The airline may increase the prices when the demand is to be expected to increase the capacity. The airline implements dynamic pricing for the flight ticket. According to the survey, flight ticket prices change during the morning and evening time of the day. Also, it changes with the holidays or festival season. To estimate the minimum airfare, data for a specific air route has been collected including the features like departure time, arrival time and airways over a specific period. Features are extracted from the collected data to apply Machine Learning (ML) models. There are several different factors on which the price of the flight ticket depends. The seller has information about all the factors, but buyers are able to access limited information only which is not enough to predict the airfare prices.

Keywords - Machine Learning, Prediction, Random forest, Regression, etc.

Project Overview:

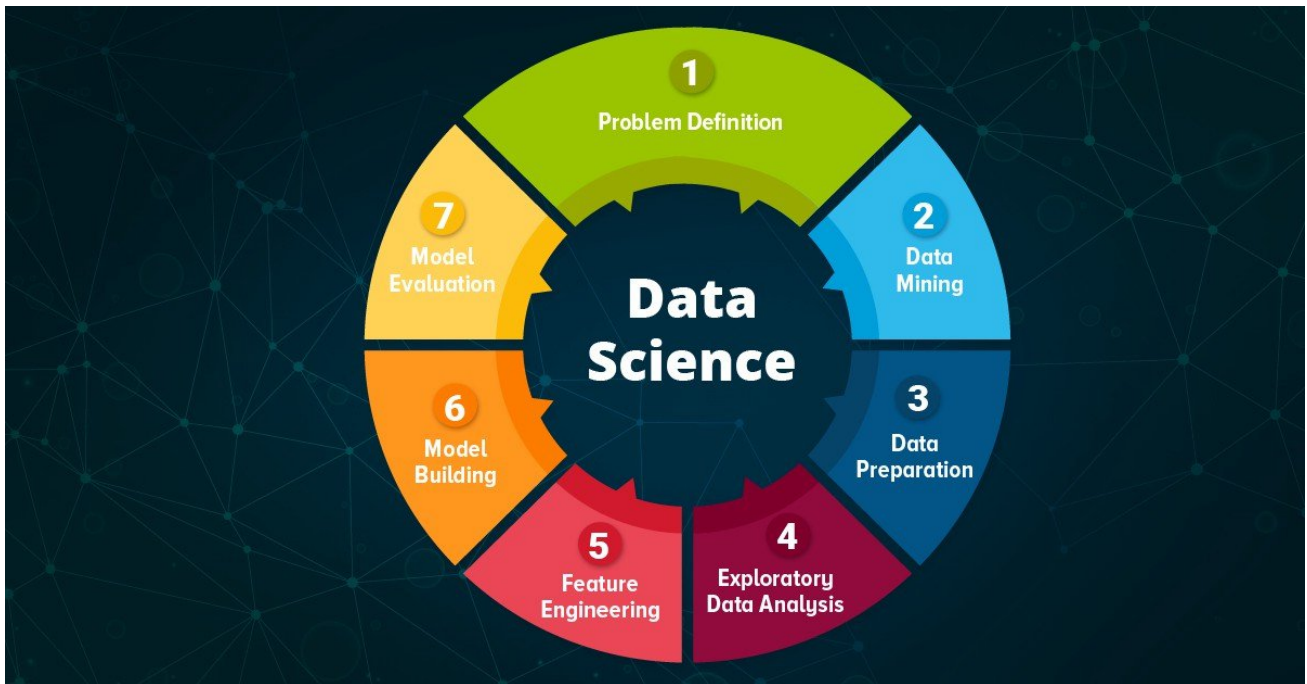
In this project, we will try to predict the prices of a flight ticket price based on different factors and The data set contains airlines name, date of journey, source, destination, route, departure time, arrival time, duration, total stops during journey, additional information, this kind of features are there. Based on this information we can use different machine learning models for prediction the price.

First, the data is acquired, collected and divided into training and test label. This data undergoes a preliminary analysis which includes uni-variate and bi-variate analysis. In the second stage, data per-processing is performed which takes care of missing and erroneous values in the dataset. In the third stage, the features are selected and modified to get the best results. In the fourth stage, feature transformation is used to convert categorical features to numerical features. In the fifth stage, using various algorithm techniques models are built and the results are evaluated. These results are communicated to the firm and finally, after approval the results are applied by the firm to generate a business model for next year. Using this method guarantees more accurate and better results.

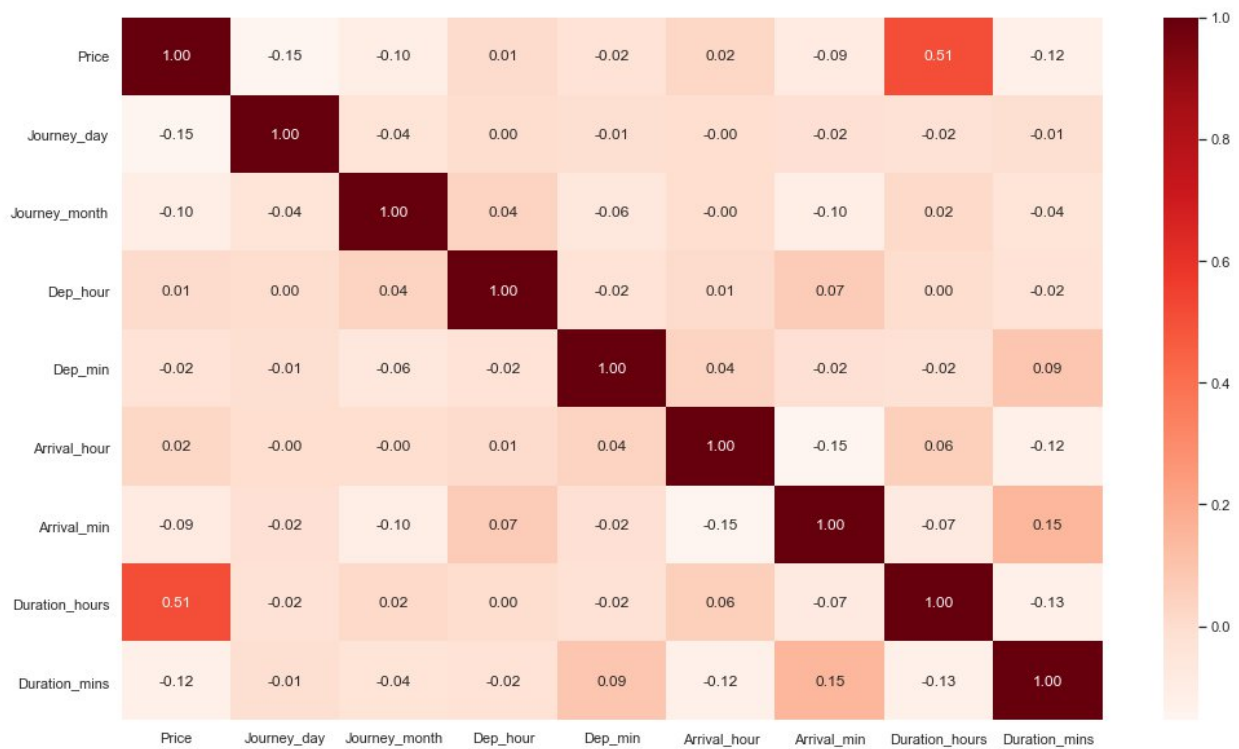
Business Understanding

Our goal for this project is predicting flight prices based on routes, destination, arrival time and date of journey this kind of information using predictions of prices can be given to customers. Generally the customers which books flight ticket repeatedly they know the when flight ticket prices are low and high. This project can be useful for those which firstly booking flight tickets.

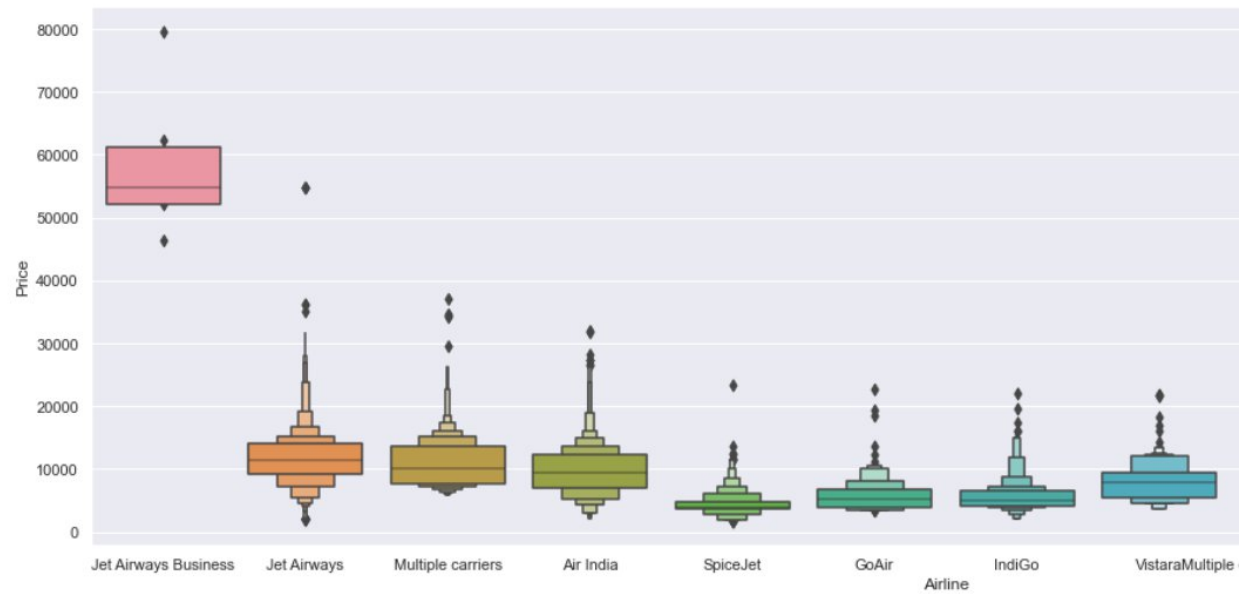
Exploratory Data Analysis (EDA)



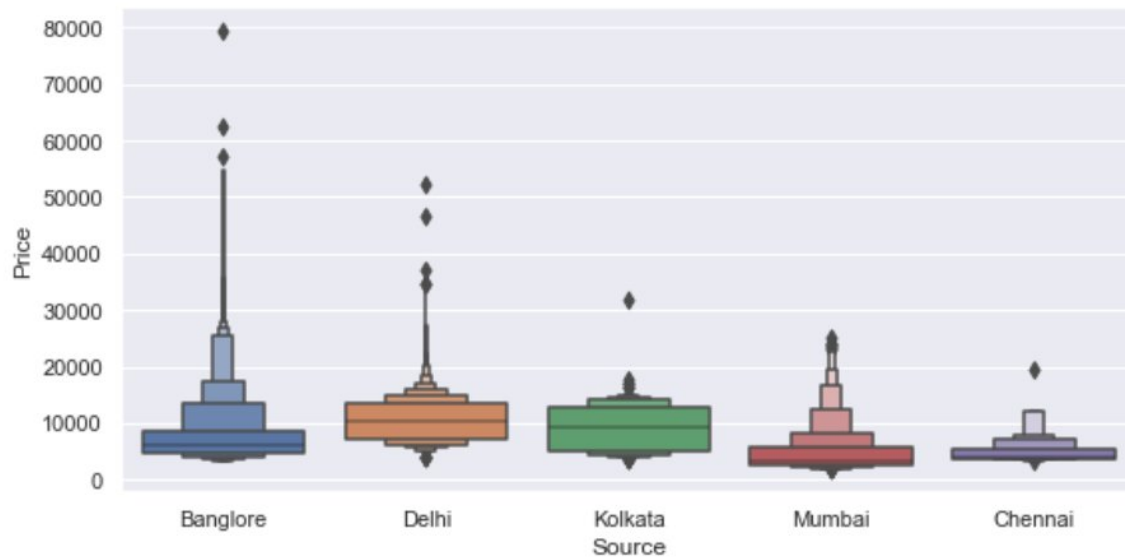
1. Checking the correlation matrix



2. Using Catplot Checking the highest price



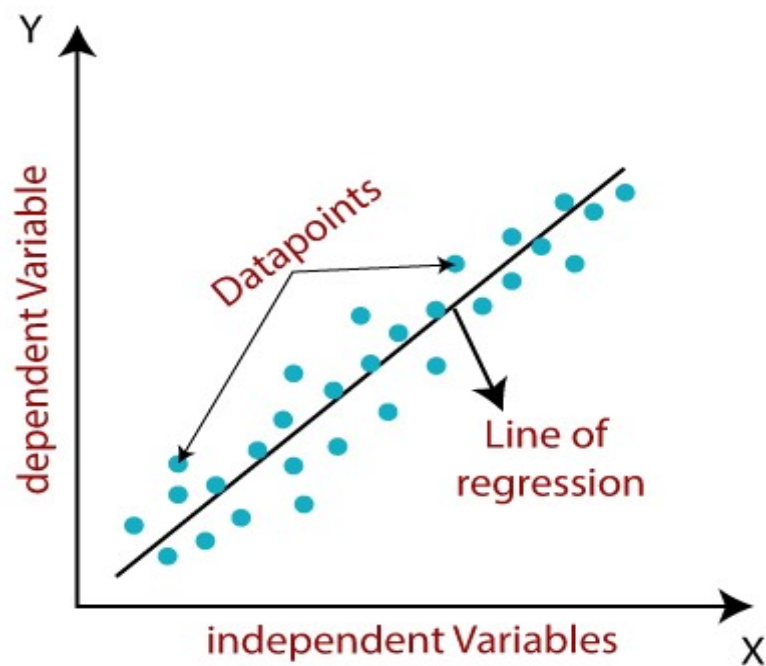
3. Checking Median



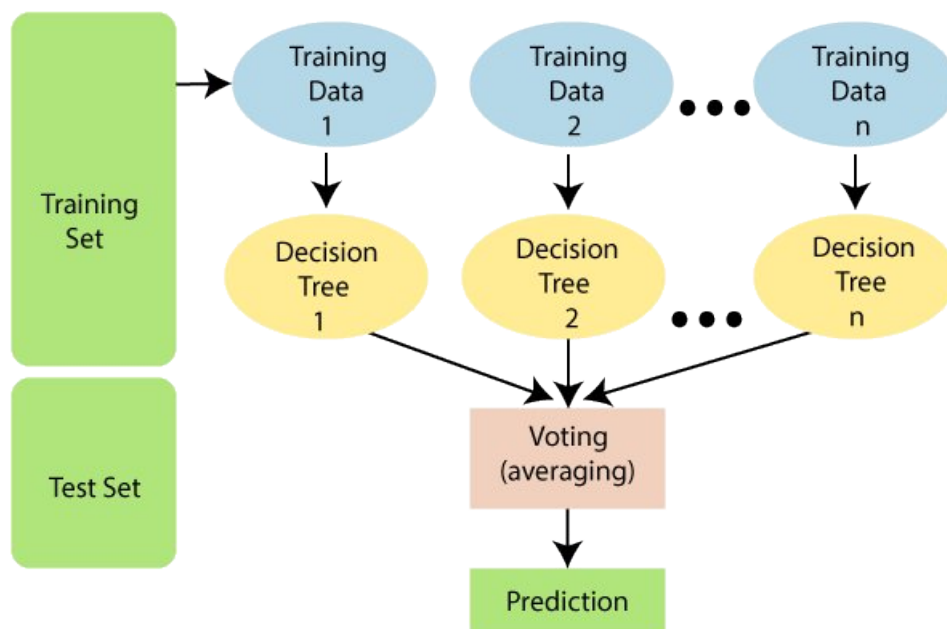
Model Building

1. LOGISTIC REGRESSION

Build the model using some variables from the dataset as independent variables to predict flight price prediction.



2. RANDOM FOREST



2.1 Feature Selection

Finding out the best feature which will contribute and have good relation with target variable. Following are some of the feature selection methods,

- 1.heatmap
- 2.feature_importance_
- 3.SelectKBest

2.2 Fitting model using Random Forest

- 1.Split dataset into train and test set in order to prediction with respected to X_test
- 2.If needed do scaling of data
Scaling is not done in Random forest
- 3.Import model
- 4.Fit the data
- 5.Predict with respected to X_test
- 6.In regression check **RSME** core
- 7.Plot graph

2.3 Hyper-parameter Tuning

1. Choose following method for hyperparameter tuning
 1. **RandomizedSearchCV - Faster than GridSearchCV**
 2. **GridSearchCV**
2. Assign hyper-parameters in form of dictionary
3. Fit the model
4. Check best parameters and best score

FUTURE SCOPE

- Data is collected from the websites which sell the flight tickets so only limited information can be accessed. The values of R-squared obtained from the algorithm give the accuracy of the model. In the future, if more data could be accessed such as the current availability of seats, the predicted results will be more accurate.

CONCLUSION

- For the prediction of the flights ticket prices different prediction models are tested for the better prediction accuracy. So to get result with maximum accuracy regression analysis is used as well as random forest is used among this two algorithm random forest will perform better than linear regression this is inferred from the studies, the feature that influences the prices of the ticket are to be considered.

Bibliography

Online References

<https://www.kaggle.com/datasets>

<https://www.geeksforgeeks.org/>

<https://www.anaconda.com/>