

FLIGHT PRICE PREDICTION





16

05

22

OBJECTIVE

Anyone who has booked a flight ticket knows how unexpectedly the prices vary. The cheapest available ticket on a given flight gets more and less expensive over time. This usually happens as an attempt to maximize revenue based on

- Time of purchase patterns (making sure last-minute purchases are expensive).
- Keeping the flight as full as they want it (raising prices on a flight that is filling up in order to reduce sales and hold back inventory for those expensive last-minute purchases).

So, you have to work on a project where you collect data of flight fares with other features and work to make a model to predict the fares of flights.



16

05

22

DATA COLLECTIONS AND METHOD TO FOLLOW

1. Data collections using a Web-Scraping
 - a. Minimum 1500 Data with a must required data
 - i. Airline name
 - ii. Departure Time
 - iii. Arrival Time
 - iv. Number of Stops
 - v. Journey Time
 - vi. Destination & Price.
2. Data Analysis
3. Model Building

DATA DETAILS

SOURCE

Website

PAYTM.COM/FLIGHTS



DEPARTURE

City

DELHI



DATE

of travel

16TH MAY 2022



TRAVELLER(S)

Class

1 ADULT IN ECONOMY



DESTINATION CITIES

- LOCATION 1 Mumbai
- LOCATION 2 Goa
- LOCATION 3 Chennai
- LOCATION 4 Hyderabad
- LOCATION 5 Port Blair
- LOCATION 6 Thiruvananthapuram
- LOCATION 7 Tiruchirappalli
- LOCATION 8 Kochi



DESTINATION CITIES

- LOCATION 9 Kannur
- LOCATION 10 Vijayawada
- LOCATION 11 Visakapatnam
- LOCATION 12 Tirupati
- LOCATION 13 Tuticorin
- LOCATION 14 Coimbatore
- LOCATION 15 Patna
- LOCATION 16 Bhubaneshwar



FLIGHT

MAY 16 2022



FEATURES CONSIDERED

- Feature 1 Airline name
- Feature 2 Price
- Feature 3 Departure Time
- Feature 4 Arrival Time
- Feature 5 Journey Time
- Feature 6 Destination City



16

05

22

EXPLORATORY DATA ANALYSIS (EDA) AND VISUALIZATION

01. Univariate Analysis (the simplest form of analyzing data. “Uni” means “one”, so in other words, your data has only one variable)
02. Multivariate Analysis (is a set of statistical techniques used for the analysis of data that contain more than one variable)
03. Correlation of Dataset (is used to test relationships between quantitative variables or categorical variables)
04. Correlation with Target variable (with the target variable to know how the data is related)
05. Conclusion (Summary with the conclusion of all the analysis)



PREPARATION

MODEL TRAINING PHASES

- Get Data
- Clean, Prepare & Manipulate Data
- Train Model
- Test Data
- Improve



REGRESSION MACHINE LEARNING MODEL/S USED

- Linear Regression Model
- Ridge Regularization Model
- Lasso Regularization Model
- Support Vector Regression Model
- Decision Tree Regression Model
- Random Forest Regression Model
- K Neighbours Regression Model
- Gradient Boosting Regression Model
- Ada Boost Regression Model
- Extra Trees Regression Model

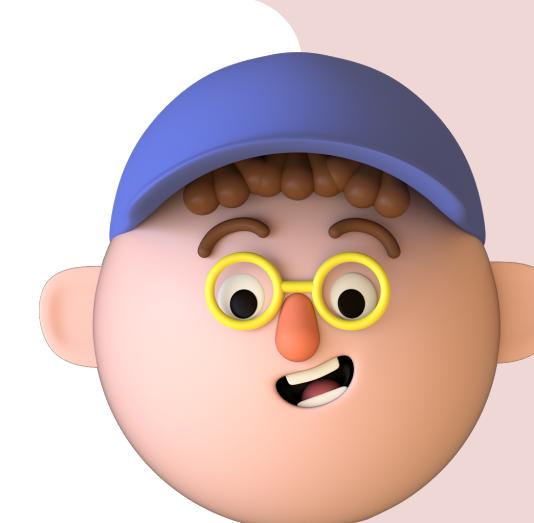


RESULT

LINEAR REGRESSION

Model

R2 SCORE: 40.23



RIDGE REGULARIZATION

Model

R2 SCORE: 40.34



LASSO REGULARIZATION

Model

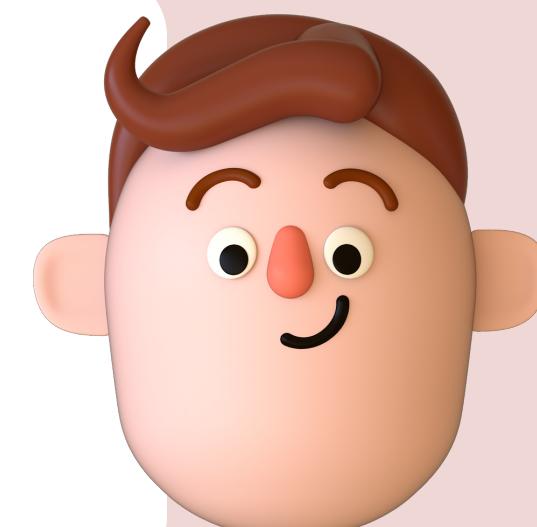
R2 SCORE: 40.24



SUPPORT VECTOR REGRESSION

Model

R2 SCORE: 40.88

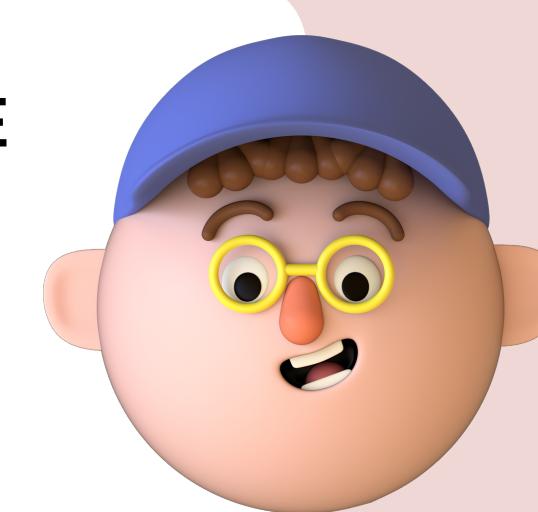


RESULT

**DECISION TREE
REGRESSOR**

Model

R2 SCORE: 25.41



**RANDOM FOREST
REGRESSOR**

Model

R2 SCORE: 39.66



**K NEIGHBORS
REGRESSOR**

Model

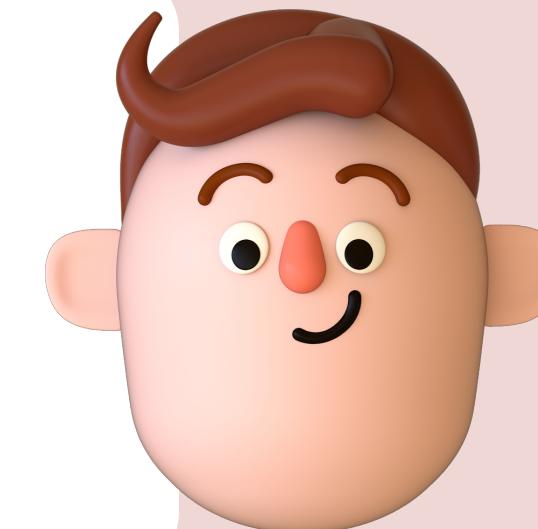
R2 SCORE: 38.36



**GRADIENT BOOSTING
REGRESSOR**

Model

R2 SCORE: 33.33



RESULT

ADA BOOST
REGRESSOR

Model

R2 SCORE: 41.76



EXTRA TREES
REGRESSOR

Model

R2 SCORE: 70.33



BEST MODEL CHOOSE TO HYPER TUNING

EXTRA TREES REGRESSOR MODEL

AFTER HYPER PARAMETER TUNING WE GOT THE
RESULT OF R2 SCORE: 71.818



KEY FINDINGS

Below are major points that plays important role in predicting the price.

- Destination City
- Departure Time
- Number of Stops
- Date of Travel
- Journey time





CONCLUSION

As we have seen, the prediction is showing a similar relationship with the actual price from the scrapped data set. This means the model is predicted correctly and it could help airlines by predicting what prices they can maintain. It could also help customers to predict future flight prices and plan the journey accordingly because it is difficult for airlines to maintain prices since it changes dynamically due to different conditions. Hence by using Machine Learning techniques we can solve this problem. The above research will help our client to study the latest flight price market and with the help of the model built he can easily predict the price ranges of the flight, and also will help him to understand Based on what factors the fight price is decided.





LEARNING OUTCOMES OF THE STUDY IN RESPECT OF DATA SCIENCE

The visualization part helped me to understand the data as it provides a graphical representation of huge data. It assisted me to understand the feature importance, outliers/skewness detection, and comparing the independent-dependent features. Data cleaning is the most important part of model building and therefore before model building, I made sure the data is cleaned.

Have generated multiple regression machine learning models to get the best model wherein I found Extra Trees Regressor Model being the best based on the metrics I have used.





ACKNOWLEDGMENT

Have used various sources from google and other external websites for reference and to build a better model as well as to understand the dataset and aviation industry in a fair usage without impacting the service provider.



THANK YOU!

Reallygreat Tour Guide

