



Flight price prediction

Submitted by,

Nidhi Gupta

Internship 16

Introduction

Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. In this project we majorly targeted to uncover underlying trends of flight prices in India using historical data and also to suggest the best time to buy a flight ticket.

Remarkably, the trends of the prices are highly sensitive to the route, month of departure, day of departure, time of departure, whether the day of departure is a holiday and airline carrier

With a high probability (about 20-25%) that a person has to wait to buy a ticket, the scope of the project can be extensively extended across the various routes to make significant savings on the purchase of flight prices across the Indian Domestic Airline market

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 –

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

Data collection

we have used web scraping for scraping the data . the headers are as afollowing

1.Airl ine	2.Date_of_J ourney	Source	Destina tion	Ro ute	Dep_T ime	Arrival_ Time	Durat ion	Total_S tops	Additional _Info	Pri ce
---------------	-----------------------	--------	-----------------	-----------	--------------	------------------	--------------	-----------------	---------------------	-----------

Observation

- there are 13354 rows and 11 columns in the data set
- 10 columns are object type and 1 column is float type .
- we have null values in 3 columns i.e route, total stops, price.
- We have done value count in eda for every flight.
- We have encoded our data .
- Our data was absolutely skewed.
- We have power transform our data.

Models used

1. Random Forest Regressor
2. r2 score

Prediction

our r2 score is 98%

MAE: 16.709933739837158
MSE: 287159.707398108
RMSE: 535.8728462966826

In [96]:

