

**FLIGHT PRICE PREDICTION**

Submitted by:

PAWAR VINAY PRASAD

**ACKNOWLEDGMENT**

Firstly, the support I received to complete this project would be from my:

Mentor/SME @ FlipRobo Technologies and

Data Trained Institute Inhouse Data Scientists and trainers.

Also referred to blogs and websites:

Stack Over Flow, Towards Data Science and Kaggle.

Professionals from Flip Robo Technologies and Data Trained Institute helped clear and give a clear picture on how to solve a particular problem.

**INTRODUCTION**

* Business Problem Framing

Aim is to develop a machine learning model which can predict flight prices.

* Conceptual Background of the Domain Problem

A small understanding on Airlines and the price for the fare fluctuations would do.

* Review of Literature

Scraped prices from makemytrip. com for different sources and destinations. This study will help predict flight prices accurately and help company to take necessary steps to maximize profits

* Motivation for the Problem Undertaken

The main motivation for this problem would be to predict flight prices accurately as the price fluctuate as the departure date nears.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

We have made sure all the columns do not have any type of symbols and commas so that we can easily pass them into a Machine Learning model. We have used labelencoder to encode the categorical columns and standard scaler to scale the data.

* Data Sources and their formats

We have scraped data from makemytrip.com website.

* Data Preprocessing Done

We removed index column and cleaned columns symbols and commas so that we can easily pass them into a Machine Learning model. We have used labelencoder to encode the categorical columns and standard scaler to scale the data

* Data Inputs- Logic- Output Relationships

If the departure date is far the price of the flight decreases. The price also depends upon the Airline and other factors.

* Hardware and Software Requirements and Tools Used
  1. Listing Requirements will start with a computer with i3 processing power and a dedicated GPU would be recommended as to process huge amount of data and save time. Anything more than i5 would be help.
  2. Use of Jupyter notebook is must, as this helps to carry out our whole project.
  3. Used Python language with Pandas, NumPy, Matplotlib, Seaborn. Pandas and NumPy helped in data importation and data wrangling. All the heavy work done was by using these two packages. Seaborn and Matplotlib was used to visualize and understand the data.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Used simple techniques that are essential and common for any project. Cleaning and visualizing of data, Pre-processing of data which included cleaning dropping of useless features. Lastly building a model by cross validating and by hyper parameter tuning.

* Testing of Identified Approaches (Algorithms)
* Decision Tree regressor.
* Random Forest regressor.
* AdaBoost regressor.
* Kneighbors regressor.
* Support Vector Regressor.
* Linear Regression.
* Gradient Boosting Regressor.

.

* Key Metrics for success in solving problem under consideration

We have used model score to understand how good the model has trained itself with the training data. We have used r2 score to understand the how close the data are to the fitted regression line.

* Interpretation of the Results

If the departure date is far and has more stop in route with long hour flight duration the price would be comparatively low.

**CONCLUSION**

* Key Findings and Conclusions of the Study

If the departure date is far and has more stop in route with long hour flight duration the price would be comparatively low.

Indigo would be the most preferred Airline. Majority of flights take off from Kolkatta and Chennai. Maximum flights land in New Delhi and Patna

* Learning Outcomes of the Study in respect of Data Science

This study would help us understand which flight airline and which route to get a very economical flight price to choose.

* Limitations of this work and Scope for Future Work

We would need to update the model with the ongoing market changes and need to make frequent updates to make sure we are up to date.