

**Flight Price Prediction**

Submitted by:

D.Mary Shereesha

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* Business Problem Framing

As the flight ticket prices so unpredictable and often cause anxiety when we would like to buy a ticket for our travel.

To solve this problem, which is actually caused by computers and algorithm behind for management of these in the systems, if we can unfold these algorithms and predict the price we can to some extent reduce the anxiety in the population.

* Conceptual Background of the Domain Problem

It is important to know that every airline has a complex computer system based on algorithms. The algorithms quickly adjust the ticket prices which is also called yield management. The bottom line: maximize income.

The system considers many factors among which are time of booking travel and travel demand.

* Review of Literature

The research was mainly on what are the factors which would help or determine the price tickets.

Each city has some special value associated with it

day of the week to book a flight

### Travel demand

#### Special events

#### Season

* Motivation for the Problem Undertaken

It’s a problem I have experienced and was always overwhelmed to book a ticket

**Analytical Problem Framing**

* Mathematical/ Analytical Modelling of the Problem

Built a model which could predict the rates, Random Forest Algorithm with factors like time, airlines, departure time/date/month, the travel duration was able to predict the rate to 99 accuracies

* Data Sources and their formats

What are the data sources mainly selected was Yatra and sky scanner from 5 different departure and arrival places was collected?

The data collected was mainly, Airlines, Arrival, Departure dates, time series going far from departure date etc

* Data Pre-processing Done

The main concentration was based on define the time series, analysing what are the main contributing factors and using them to feed into model.

* Data Inputs- Logic- Output Relationships

The main input was data scrapped from Yatra and sky scanner and was in the form of Excel.

* State the set of assumptions (if any) related to the problem under consideration

The assumption based on the literature search was, the factors for the price determination are Airlines, place of arrival and departure, the time/month/when is the departure date are the factors which are assumed to determine the price

* Hardware and Software Requirements and Tools Used

The software used is Jupiter Python notebook

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

The problem solving algorithm tested are Regression ML models to determined and predict the prices were

Random forest, Gradient boost regressor, Adaboost regressor, Linear Regressor

* Testing of Identified Approaches (Algorithms)

Random forest, Gradient boost regressor, Adaboost regressor, Linear Regressor

* Run and evaluate selected models

Random forest was selected as it showed 99 percent accuracy and also was 95% accuracy average cross validation CV.

* Key Metrics for success in solving problem under consideration

The Key Metrics is accuracy scores as it is a regression problem

* Visualizations

By Line plots visualizing the change of the price over time and Weekdays was used to understand the problem and factors of determination better

* Interpretation of the Results

From visualization determined hoe the prices determine in different Airlines

**CONCLUSION**

* Key Findings and Conclusions of the Study

Hahn Air systems are the costly airlines and change much with the Price , very constanly maintain the price over time

Srilankan Airlines and Emrites are the airlines which are costly when departure date is very close and are not available to book on the same day, As there was no data in April month available for booking

malaysia Airlines change fluctuate a lot with departure data and other factors

Hahn Air Technologies'-Price fluction is lot and prices are hicked on weekends

Srilankann Airlines-Fluctuates a lot when departure date is close and over weekends otherwise is more or less stable

Hahn Air Technologies quite stable and costly airlines but there prices are largely different over a point

Emrites is the Airlines, which are costly over weekend , very prominently

## 'Novaturas do not change its rice at all

### NonStop/Direct < 1 stop<2stop<=3stop

### this data also shows prices in Yatra are less compared to Skyscanner

* Learning Outcomes of the Study in respect of Data Science

# Best days to book a ticket are

1)Direct Destination flights

2)Weekdays

3)a month apart fro your departure day

* Limitations of this work and Scope for Future Work

It seem flight prediction is not a very complex issue, which a machine cannot achieved, As it is also determined by algorithms.