

# **SMARTBRIDGE EXTERNSHIP**

## **(APPLIED DATA SCIENCE)**

### **FLIGHT PRICE PREDICTION**

#### **PROJECT REPORT**

**Group No: 282**

#### **Team Members:**

**Nuchu Manikanta Yadav – 20BAI10374**

**Sai Krishna – 20BAI10370**

**Yeshwanth Reddy – 20BAI10371**

**Gaganeshwara Reddy – 20BCE10014**

# 1. INTRODUCTION

## 1.1 Overview:

The accumulated flight history will be used in our system to inform the development of a prediction model, which will be accomplished by employing machine learning methods. People will get an understanding from this system about the trends that prices follow, and it will also provide a forecast price value that they can refer to before purchasing flight tickets in order to save money. Customers may receive this kind of system or service if it is made available to them.

Our system will make use of the accumulated flight history to assist in the creation of a prediction model using machine learning techniques. This system will help people understand price trends and provide a forecast price value that they can use to compare prices before buying airline tickets in order to save money. If offered to customers, this kind of system or service may be obtained.

## 1.2 Purpose:

The purpose of flight price prediction is to forecast and estimate the future prices of airline tickets. It involves analyzing historical data, current market conditions, and various other factors to provide travelers with an estimate of how airfares might change over time.

Using Flight Price Prediction, we can achieve the following:

**Planning and Budgeting:** It helps in predicting flight prices allows travelers to plan their trips in advance and budget accordingly.

**Cost Savings:** Flight price prediction can help travelers save money by identifying periods when airfares are likely to be lower. It enables them to take advantage of price drops, discounts, or flash sales, allowing them to book their flights at the most economical prices.

**Decision Making:** When travelers have an idea of how airfares might change, they can make informed decisions about their travel plans. They can choose between different routes, airlines, or travel dates based on the projected prices

**Flexible Travel Arrangements:** With flight price prediction, travelers can be flexible with their travel arrangements. If prices are expected to be high during their intended travel period, they can adjust their travel dates to find more affordable options.

## 2. LITERATURE SURVEY

### 2.1 Existing Problem

There are several existing approaches to overcome this, some of them are:

**Historical Data Analysis:** This approach involves analyzing historical flight price data to identify patterns, trends, and seasonality. Historical data from multiple sources, such as airline ticketing systems, online travel agencies, or global distribution systems (GDS), can be used to train predictive models. Techniques such as regression analysis, time series analysis, or machine learning algorithms can be applied to extract insights and make predictions based on past price behavior.

**Big Data and Sentiment Analysis:** With the availability of vast amounts of data from various sources like social media, travel websites, and online reviews, sentiment analysis techniques can be applied to gauge customer sentiment and preferences. Analyzing customer opinions and feedback can provide insights into their price sensitivity and help in predicting price trends.

**Data Aggregators and Online Travel Agencies:** Some online platforms and travel agencies aggregate flight data from multiple sources and provide flight price prediction services to their customers. These platforms analyze historical data, real-time

market data, and user search behavior to estimate future prices and offer recommendations.

## **2.2 Proposed Solution**

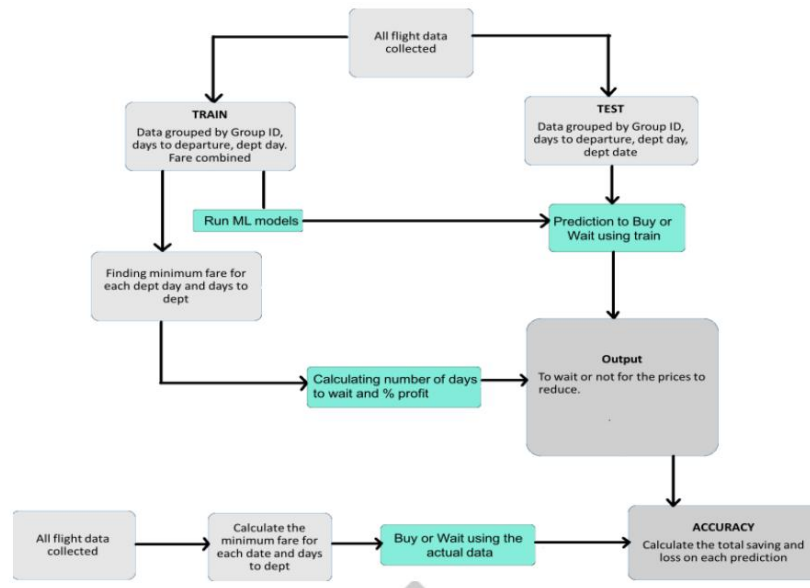
In this project, we are using some machine learning models for the prediction of flight prices. All the used models in the project are listed.

Here models are trained based on historical flight data and use various features such as departure date, destination, airline, route, time of booking, and market conditions to predict future prices. Algorithms like linear regression, decision trees, random forests, gradient boosting, or neural networks can be employed to build these models.

Also, the models use Ensemble Learning technique which combines the multiple individual models, known as base models or weak learners, to create a more accurate and robust predictive model. It aims to leverage the diversity and complementary strengths of these models to make better predictions than any individual model alone.

## **3. THEORETICAL ANALYSIS**

### **3.1 Block Diagram**



## 3.2 Software Designing

Flight price prediction typically involves the use of various software components to process and analyze the large amount of data involved.

Some of the Software tools are:

**Data Processing Tools:** Flight price prediction involves processing and cleaning large datasets. Tools like Apache Hadoop, Apache Spark, or Apache Kafka are often used for distributed data processing, data ingestion, and data integration tasks. These tools help manage and process the vast amounts of flight data efficiently.

**Programming Languages:** Various programming languages are used for implementing flight price prediction models. Commonly used languages include Python, R, and Java. Python, with libraries like NumPy, Pandas, and Scikit-learn, is particularly popular due to its rich ecosystem of machine learning frameworks and libraries.

**Machine Learning Libraries and Frameworks:** There are several machine learning libraries and frameworks available that facilitate

the development and implementation of flight price prediction models. Popular libraries/frameworks include Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, and LightGBM. These libraries provide algorithms, tools, and utilities for training, evaluating, and deploying machine learning models.

**Data Visualization Tools:** Data visualization plays a crucial role in understanding and analyzing flight price data. Tools like Matplotlib, Seaborn, or Tableau are often used to create visualizations, charts, and graphs that help interpret and communicate the insights derived from the data.

## **4. EXPERIMENTAL INVESTIGATIONS**

While we go through the algorithms we employed (XGBoost, Random Forest, and Decision Tree) and also how they operate in our models, please read the discussion below.

### **A. Decision Tree**

The decision tree appears to be the most well-known and commonly employed categorization technique. A decision tree is a collection of nodes that resembles a diagram, for each junction indicating a test on the a characteristic and each branch indicating a test outcome, such that each node in a decision tree (terminal node) has a class label. A tree can be "trained" by dividing the resources collection into subgroups depending on a characteristic values test. This procedure is known as partitioning the data because it is performed iteratively on each derived subset. The recursion ends when all subgroups at a node have the same posterior probability, or when the split no longer adds additional value to the predictions. A decision tree is appropriate for experimental extracting knowledge since it does not need subject matter expertise or parameters configuration. Assume  $S$  is a collection of cases,  $A$  is a property,  $S_v$  is the subgroup of  $S$  with  $\text{Such } a = v$ , as well as  $\text{Value}(A)$  is the collection of all number of values of  $A$ , then

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

## B. Random Forest

A Random Forest is an ensemble approach that can handle simultaneous regression and classification problems by combining many decision trees using a technique known as Bootstrap as well as Aggregation, or bagging. The core idea is to use numerous decision trees to determine the final result instead of depending on personal decision trees. Random Forest's foundation learning methods are numerous decision trees. We arbitrarily choose rows and characteristics from the dataset to create sample datasets for each model. This section is known as Bootstrap. We simply have to understand the purity in our dataset, and we'll use that characteristic as the root of the tree which has the smallest impurity or, in other words, the smallest Gini index. Mathematically Gini index can be written as:

$$\begin{aligned} Gini\ Index &= 1 - \sum_{i=1}^n (P_i)^2 \\ &= 1 - [(P_+)^2 + (P_-)^2] \end{aligned}$$

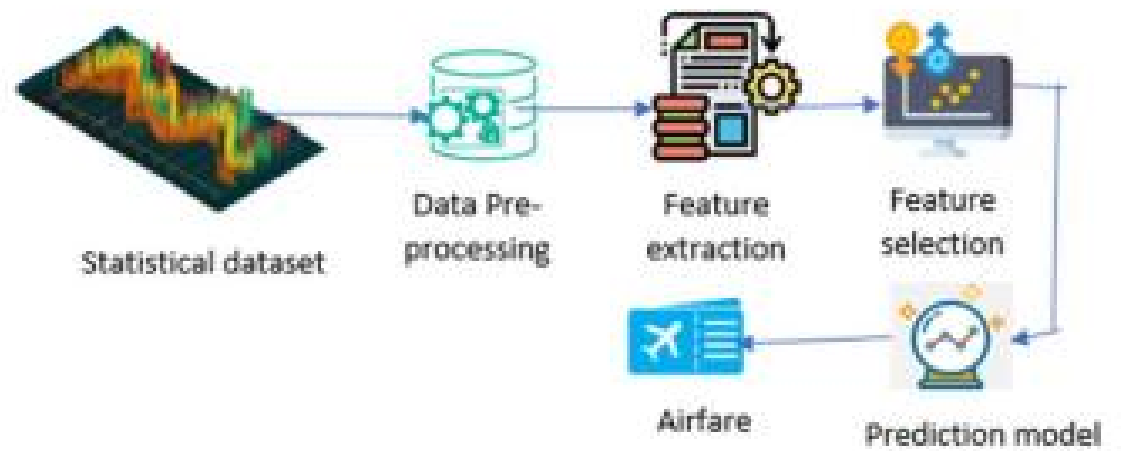
## C. XGBoost

XGBoost is an effective method for developing supervised regression models. Knowing as to its (XGBoost) goal function and baseline learners can help determine the truth of this proposition. This optimization problem has both a loss function and a regularization component. It makes a distinction between real and theoretical predictions, i.e. how far the model outputs deviate from the real amounts. In XGBoost, the most used standard error in regression problems is quarantine, whereas reg:logistics is used for classifications.

The formula may be used to compute the output value of each model

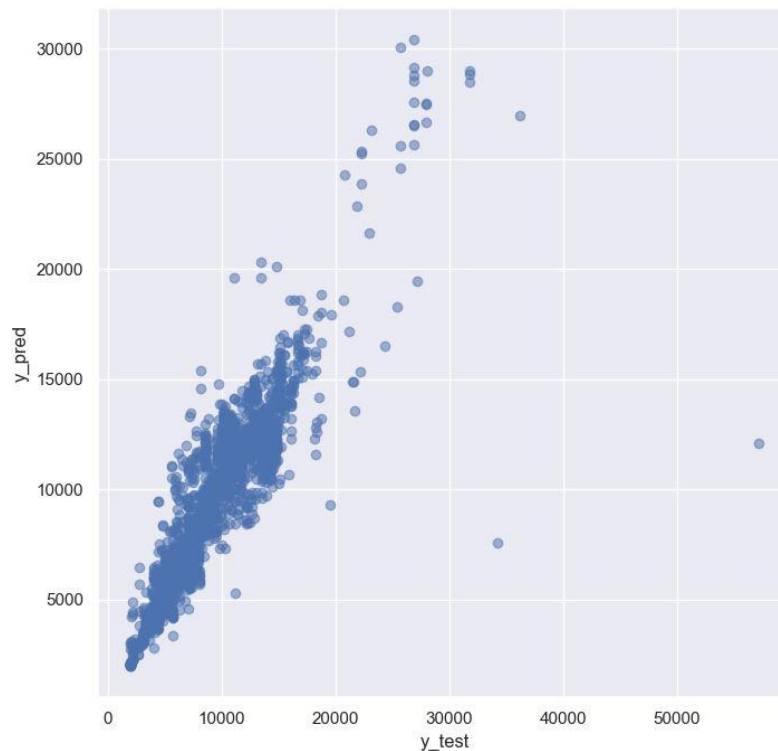
$$Output\ value = \frac{\sum Residual}{no.\ of\ Residual + \lambda}$$

## 5. FLOWCHART



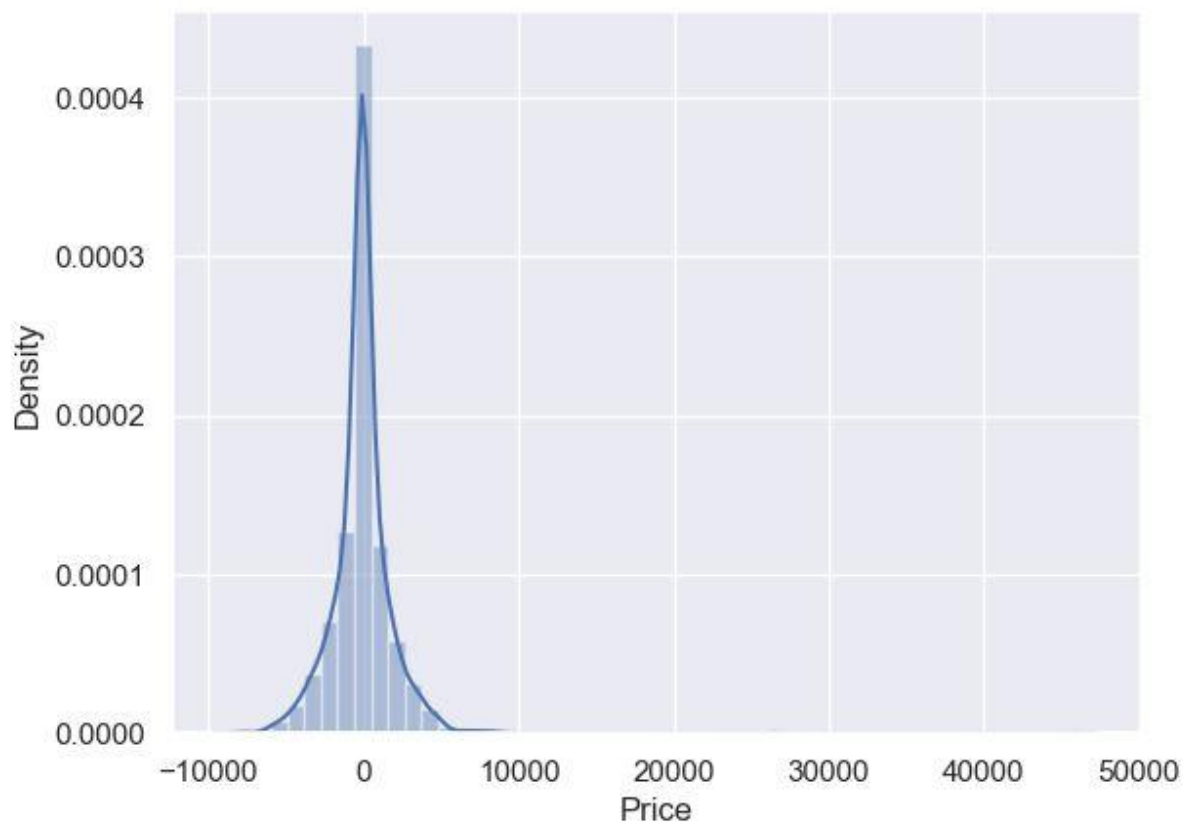
## 6. RESULT

First we plot the points of  $y_{\text{testing}}$  data and  $y_{\text{training}}$  data on x-axis and y-axis respectively.





We drew the graph using data visualization to highlight the significance of each attribute for predicting flight prices.



The accuracy we acquired with the current predictor is 79%.

## Save Model using Pickle for further use in web app

```
[191] import pickle
      file = open('model1.pkl', 'wb')
      pickle.dump(reg_rf, file)

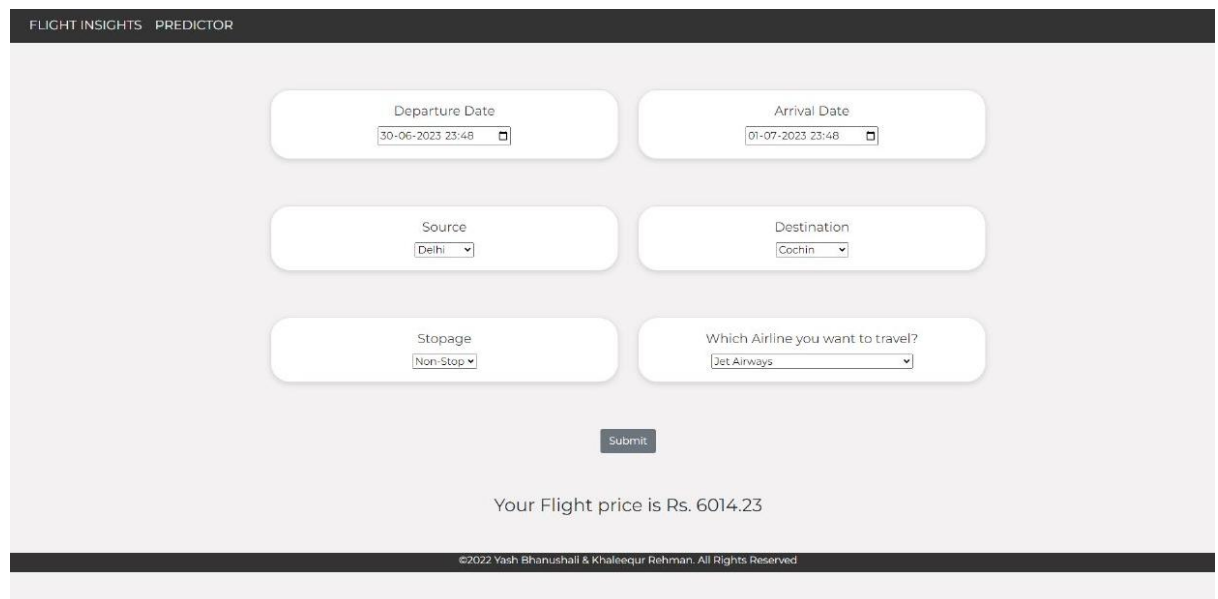
[191] model = open('model1.pkl', 'rb')
      forest = pickle.load(model)

[194] y_prediction = forest.predict(X_test)

[199] metrics.r2_score(y_test, y_prediction)

... 0.7969878237324193
```

Here are some pictures which show us the working of the Flight Price Predictor on web. We can select the departure and arrival dates as per flexible timings. And we also set our source location and destination (i.e., from and to, of our flight). Then we select whether the flight should have the stops or not of our choice. Finally, we choose the airways as per our wish.



The screenshot shows a web interface titled "FLIGHT INSIGHTS PREDICTOR". It features a form with six input fields arranged in three rows and two columns. The first row contains "Departure Date" (30-06-2023 23:48) and "Arrival Date" (01-07-2023 23:48). The second row contains "Source" (Delhi) and "Destination" (Cochin). The third row contains "Stopage" (Non-Stop) and "Which Airline you want to travel?" (Jet Airways). A "Submit" button is located below the form. The result displayed is "Your Flight price is Rs. 6014.23". At the bottom, a copyright notice reads "©2022 Yash Bharushali & Khaleequr Rehman. All Rights Reserved".

## 7. ADVANTAGES AND DISADVANTAGES

Flight price predictors can be useful tools for travelers to get an estimate of future flight prices and make informed decisions. However, like any tool, they come with both advantages and disadvantages. Here are some advantages and disadvantages of flight price predictors:

### Advantages:

**Cost savings:** Flight price predictors can help travelers find the most affordable flights by analyzing historical data and predicting future trends.

**Time-saving:** Instead of manually searching multiple airline websites or travel agencies for the best deals, flight price predictors provide a centralized platform where users can compare prices and find the most suitable options quickly and efficiently.

**Convenience:** Flight price predictors often offer additional features like price alerts, which notify users when prices drop or reach a desired threshold. This convenience allows travelers to stay updated on price fluctuations without actively monitoring flight prices.

**Disadvantages:**

**Inaccurate predictions:** Flight prices are influenced by various factors, including market demand, fuel costs, and airline pricing strategies.

**Limited coverage:** Flight price predictors may not have access to data from all airlines or travel agencies, resulting in limited coverage of flight options.

**Lack of real-time updates:** Predictors rely on historical data and algorithms, which may not capture sudden price changes or flash sales that occur in real time

**Influence on user behavior:** Flight price predictors can create a sense of urgency or false expectations among travelers, leading to impulsive decisions or disappointment when prices do not align with predictions.

.

Overall, flight price predictors can be valuable tools for travelers, but it's important to use them as aids rather than relying solely on their predictions. Travelers should consider other factors like personal preferences, travel flexibility, and alternate booking sources to make well-informed decisions.

## **8. APPLICATIONS**

Flight price predictors have several applications that can benefit travelers and the travel industry as a whole

**Price comparison:** Flight price predictors enable users to compare prices across different airlines, travel agencies, and booking platforms. This helps travelers find the best deals and identify the most affordable options for their travel needs.

**Price alerts and notifications:** Users can set up notifications for specific routes or desired price thresholds, and they will receive alerts when prices drop or reach the specified criteria. This allows travelers to take advantage of discounted fares and book flights at the optimal time.

**Business travel optimization:** For business travelers, flight price predictors can help optimize travel expenses and budgets. They can assist in finding cost-effective options for regular business trips and identifying the most suitable times to book flights based on projected price trends.

**Revenue management for airlines:** Airlines can use predictive models to optimize their revenue management strategies. By analyzing historical data and demand patterns, airlines can adjust prices, seat inventory, and flight schedules to maximize profitability.

## 9. CONCLUSION

In this case study, a few machine learning models were looked at to forecast the typical flight cost at the level of the business segment. We trained the training data with training data, and we tested it with test data. From these records, a number of traits were extracted. Using attribute selection techniques, our proposed model can calculate the quarterly average flight price. Most prior research on large-scale flight price prediction relied on conventional statistical methods, which have their own limitations in terms of underlying issue estimates and hypotheses.

In the case where, neither database offers precise data on ticket sales, including such information as departure and arrival times and days of the week. This framework might be extended in the future to include information about how much airline tickets cost, which can provide more information about each area, including the timestamps of entry and exit, seat assignment, covered auxiliary items, and other details. It is possible to create a more reliable and complete daily and even daily flight price forecast mode by combining such data.

Additionally, a market sector's flight prices may change if there is a significant increase in business travelers brought on by some unusual

events. In order to supplement our forecasting models, incident data will be gathered from a variety of sources, such as social media platforms and media outlets.

In order to improve current models and achieve the best design for airline price prediction, we will also examine specific technological models, such as Deeper Learning techniques.

## 10. FUTURE SCOPE

The future scope of flight price predictors is promising, with potential advancements and improvements in several areas. Here are some potential future developments and areas of growth for flight price predictors:

**Multi-modal travel integration:** As travelers increasingly opt for multi-modal journeys involving flights, ground transportation, and accommodations, flight price predictors could expand their scope to incorporate and analyze data from various travel modes. This would enable travelers to compare and optimize their entire travel itinerary for the best overall cost and convenience.

**Enhanced accuracy:** Improving the accuracy of flight price predictions is a key area of focus. Advances in data analytics, machine learning, and predictive algorithms can lead to more precise forecasts by incorporating a wider range of data sources.

**Real-time updates:** Integrating real-time data feeds and updates into flight price predictors can provide users with the most up-to-date information on fare fluctuations, flash sales, and last-minute deals. This would allow travelers to make timely decisions and take advantage of sudden price drops.

**Personalization and customization:** Future flight price predictors may offer more personalized recommendations based on individual traveler preferences and historical booking patterns. By considering factors such as preferred airlines, seating preferences, loyalty programs, and past

travel behavior, the predictors can tailor recommendations to each user's specific needs.

## **11.BIBLIOGRAPHY**

### **References:**

- J. Stavins, "Price discrimination in the airline market: The effect of market concentration," *Review of Economics and Statistics*, vol. 83, no. 1, pp. 200–202, 2001.
- B. Mantin and B. Koo, "Dynamic price dispersion in airline markets," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 6, pp. 1020–1029, 2009.
- P. Malighetti, S. Paleari, and R. Redondi, "Has Ryanair's pricing strategy changed over time? an empirical analysis of its 2006–2007 flights," *Tourism Management*, vol. 31, no. 1, pp. 36–44, 2010.
- T. H. Oum, A. Zhang, and Y. Zhang, "Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets," *Journal of Transport Economics and Policy*, vol. 7, no. 2, pp. 171–192, 1993.
- <https://www.ijedr.org/papers/IJEDR1804041.pdf>
- [https://www.researchgate.net/publication/335936877\\_A\\_Framework\\_for\\_Airfare\\_Price\\_Prediction\\_A\\_Machine\\_Learning\\_Approach](https://www.researchgate.net/publication/335936877_A_Framework_for_Airfare_Price_Prediction_A_Machine_Learning_Approach)
- [https://smartinternz.com/Student/guided\\_project\\_info/524381#](https://smartinternz.com/Student/guided_project_info/524381#)