International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# Predicting Airline Ticket Prices Using Machine Learning

**Dr.N.Gopala Krishna***

**Gopu Iswarya[1], Bhavirisetty Sakunthala Narasimharao[2],**

**Bhahunandham Naga Durga Devi[3] , Bhupathi Dani[4] ,**

*[1,2,3,4] Student, Department of Computer Science and Engineering, Tirumala Engineering College*
*[*] Professor, Department of Computer Science and Engineering, Tirumala Engineering College*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** Airfare prices are constantly changing due to various factors like the time of flight, destination, and length of the trip. To help people figure out the best time to buy tickets, we are working on a system that will use machine learning to predict prices. By studying past flight data in India, we want to uncover pricing patterns and suggest the ideal time to purchase tickets. This project aims to confirm or dispel common beliefs about airlines by comparing prediction models and finding ways to save money on ticket purchases. Price trends can vary significantly based on various factors such as the route, month, day, time of departure, holidays, and the airline carrier. Competitive routes like major business destinations (Mumbai-Delhi) tend to see price increases closer to the departure date. On the other hand, routes like tier 1 to tier 2 cities (Delhi-Guwahati) have specific time frames when prices are lower. The data also shows that there are two main categories of airline carriers in India: economical and luxurious. Generally, the lowest-priced flights fall under the economical group. Moreover, the data confirms that specific times of the day tend to have higher prices. By including different routes in this project, there is a potential for substantial savings when buying domestic flight tickets in India.

***Key Words*:** Airline Ticket, Pricing strategies, Fare fluctuations, Competitor analysis, Market trends

## 1.INTRODUCTION

When purchasing a flight ticket, it is often recommended to buy in advance to avoid high prices. However, airlines do not always follow this rule. They may lower prices to attract more customers or raise prices when tickets are limited. Therefore, ticket prices can vary based on different factors. This project uses AI to analyze the price trends of flight tickets over time. Airlines have the flexibility to change ticket prices whenever they want. Travelers can save money by booking tickets at the lowest prices available. Those who fly often are usually familiar with price changes. Airlines use advanced Revenue Management strategies to implement different pricing tactics. Airlines often adjust ticket prices based on various factors like time of travel, season, and holidays. They may also update the header or footer on subsequent pages to show these changes. The main goal for airlines is to make a profit, while passengers want to find the cheapest rates possible. Many customers try to buy tickets well ahead of their travel date to avoid high prices as the departure day gets closer. But in some cases, this strategy doesn't always work, and customers may end up spending more than they need to for the same seat.

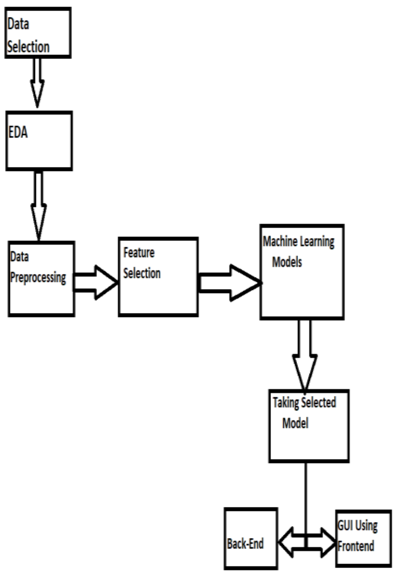
## 2. LITERATURE REVIEW

1. When evaluating the algorithm, a dataset is gathered, pre-processed, and used for data modeling to study differences in fares for travelers on restricted days. Machine learning algorithms are employed to accurately forecast airline fares, providing the precise cost of plane tickets at both the lowest and highest values.
2. By collecting airfare data from Aegean Airlines website, it was demonstrated that predicting flight prices based on historical fare information is feasible. The results of the experiment indicate that these models are effective in predicting airfare prices.
3. An algorithm called ACER is introduced for predicting airfare prices in a context-aware manner. This involves using machine learning techniques to address the time series nature of airfare forecasting. ACER combines context modeling, clustering, and ensemble methods to create an adaptive framework for regression. This approach aims to improve accuracy and efficiency in predicting airfare prices.
4. We created a machine learning model to forecast the average airfare price for each quarter within specific market segments. By merging data from U.S. domestic airline ticket sales and non-stop segment information from the DB1B and T100 datasets, our model utilizes feature selection methods to accurately predict quarterly airfare prices with an impressive adjusted R-squared score of 0.869.
5. Economic theory suggests that the way airline costs are passed on to consumers depends on whether the costs are specific to a certain airline or affect the entire industry. In monopolistic markets, airlines may pass on most, if not all, of a cost change to consumers, depending on how prices and demand are related.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

## 3. PROPOSED SYSTEM

Forecasting airfare involves predicting future prices based on historical data. The Auto-Regressive Integrated Moving Average (ARIMA) method is commonly used for this purpose. AR and MA components of ARIMA help to account for positive and negative correlations in the data. These components work together to prevent overfitting and improve accuracy. Gordiievych proposed using the ARIMA model to create a system that can forecast airline ticket prices, aiding customers in their purchasing decisions. Another approach called Facebook-prophet is similar to STL decomposition, which breaks down time-series data into seasonal, trend, and residual components. When it comes to understanding and managing seasonal time series data, STL decomposition is considered to be more appropriate than traditional time series models in terms of control and interpretability. In Random Forest (RF), multiple decision trees are combined using ensemble learning. XGBoost (Extreme Gradient Boosting Decision Tree) is a more robust and effective machine learning method that can be utilized for problem identification and time series prediction. Wohlfarth integrated some of the latest supervised learning algorithms (such as classification tree (CART) and RF) at the beginning of the cluster to improve customer decision-making. They utilized CART to understand important rules and RF to assess the importance of each feature.




Our framework uses the DB1B and T-100 datasets along with macroeconomic data to predict the average quarterly airfare at the market segment level. The above figure provides an overview of the key components of our framework. During the data pre-processing stage, we clean all datasets, exclude any potentially erroneous samples, and then transform and combine them according to the market segment. The feature extraction module is responsible for extracting and generating handcrafted features that aim to characterize the market segment. The feature selection module aims to enhance the performance of the prediction model by analyzing the effectiveness of the features and eliminating any irrelevant ones. We then utilize the chosen features to construct our prediction model, which provides the predicted air ticket price as the output value.

- **Data Pre-processing:**

When working with datasets like DB1B and T-100, it is common to find duplicate information in various attributes. By merging tables directly, this can lead to multiple duplicate fields. Additionally, inaccuracies in the data reported by airlines may arise from human errors or currency conversion mistakes. Hence, having a well-designed data pre-processing workflow is essential for ensuring the accuracy of input data used to construct a machine learning model.

- **Extracting Features:**

Many features have been taken from the DB1B and T-100 dataset to showcase a particular part of the market segment. Additionally, to understand how the airline industry is tied to the overall economic conditions, several macroeconomic features are incorporated into the feature set. Table I details all the identified features during this process.

## ALGORITHMS

When discussing the algorithms we used in our models (XGBoost, Random Forest, and Decision Tree) and explaining how they work, please see the discussion below.

- **XGBoost:**

  XGBoost is a robust machine learning tool that uses gradient-boosted decision trees and is designed for scalability. It is widely used for handling regression, classification, and ranking tasks, thanks to its ability to enhance trees simultaneously. Having a strong grasp of fundamental machine learning concepts and methods is crucial to fully unlocking the potential of XGBoost. When predicting outcomes on unseen data, a supervised machine learning model trained on a labeled dataset comes into play. Decision trees are a method used to create a model that can predict a label by using a series of true/false questions based on features. XGBoost is a scalable technique that can help with this process. Decision trees are valuable for predicting categories, while regression with decision trees is good for forecasting continuous variables. An example provided below demonstrates how a decision tree can be used to predict the price of a house based on its square footage and number of bedrooms.

- **Decision Tree:**

It is a widely used categorization technique that is depicted as a diagram with nodes representing tests on characteristics and branches showing test outcomes. Each node, also known as a terminal node, is assigned a class label. To train a decision tree, the resources collection is divided into subgroups based on characteristic values tests through an iterative process called partitioning the data. This process continues until all subgroups at a node have the same posterior probability or when the split no longer contributes valuable predictions. This method is suitable for experimental knowledge extraction because it does not require expertise in the subject or parameter configuration.

Let's say we have a group of cases represented as S, where A is a property. Sv represents a subgroup of S with a specific value v, and Value (A) consists of all the different values of A.

- **Random Forest:**

In the world of data science, there is a powerful technique called Random Forest. This method is great for tackling both classification and regression problems at the same time. It works by combining multiple decision trees using a technique called Bootstrap and Aggregation, also known as bagging. Instead of relying on just one decision tree, Random Forest uses a large number of them to come up with the final result. To build a Random Forest model, we start by creating sample datasets for each individual tree by randomly selecting rows and features from the original dataset. This process is known as Bootstrap. We then look for the most pure characteristic in the dataset and use it as the starting point for each tree, aiming to minimize impurity or Gini index. This approach allows us to make more accurate predictions by leveraging the collective knowledge of many different decision trees. It's a smart and efficient way to deal with complex data sets and make sense of them.

The steps in the working procedure are as follows:
1. Pick R observations at random from the training data set.
2. Develop classification trees for the sample points selected.
3. Select a number A to indicate the classification trees in the system.
4. Repeat steps 1 and 2.
5. Find the projections for current data elements in each decision tree and assign the samples to the class with the best scores.

The construction steps of the Random Forest Algorithm are as follows:
1. Data Pre-processing
2. Application of the Random Forest Algorithm to the dataset for training
3. Forecasting the result of the study 4. Construction of Confusion matrix 5. Exploring the possible result.

**DESCRIPTION OF THE SYSTEM ARCHITECTURE**

1. First, we input all the important information into the software we built using Python.
2. Next, we preprocess the data using the training module.
3. Then, we extract features and select the ones that match the training data.
4. Finally, the system makes predictions and displays the results.

## 5. RESULTS

The proposed approach involves creating a predictive system for flight fares by utilizing machine learning to differentiate flights using the given dataset.

## 4. CONCLUSION

The dataset is used to apply machine learning algorithms to predict flight ticket prices dynamically, with the aim of offering the lowest possible cost to customers. The data is gathered from websites that sell flight tickets, which may limit the amount of information available. The accuracy of the model is evaluated using R-squared values generated by the algorithm. Having access to more data in the future, such as current seat availability, would improve the accuracy of the predictions. In conclusion, a process has been developed to predict airline ticket prices, and the validity of these predictions is supported by past trends and our own analysis.

This study explored the prediction of airline prices using machine learning models. Airfare data from Aegean Airlines was collected from the web and used to predict flight prices based on past fare data. The results demonstrated that machine learning models are effective in predicting airfare prices. Factors such as data collection and feature selection were found to play a crucial role in airfare prediction. The study also identified key features that have the most impact on airfare prediction accuracy. In addition to the selected features, there are other factors that could further enhance prediction accuracy.

In the future, we could potentially expand this research to forecast airfare prices for all destinations served by the airline. Conducting further experiments with larger airfare datasets is crucial. Nonetheless, this preliminary study demonstrates the potential of Machine Learning models in assisting consumers with purchasing airfare at the most opportune times.

## 6. REFERENCES

1. Sarao, Parwaz and Samanta, Pushpendu, Flight Fare Prediction Using Machine Learning (October 20, 2022). Available at SSRN: https://ssrn.com/abstract=4269263 or http://dx.doi.org/10.2139/ssrn.4269263
2. Ankita Panigrahi1 , Rakesh Sharma2 , Sujata Chakravarty3 , Bijay K. Paikaray4 and Harshvardhan Bhoyar 5 “Flight Price Prediction Using Machine Learning “ https://ceur-ws.org/Vol-3283/Paper90.pdf
3. Abdella, Juhar Ahmed, et al. "Airline ticket price and demand prediction: A survey." Journal of King Saud University-Computer and Information Sciences 33.4 (2021): 375-391.
4. S. Chakravarty, P. Mohapatra, P. K. Dash, (2016), Evolutionary Extreme Learning Machine for Energy Price Forecasting, International Journal of

Knowledge-Based and Intelligent Engineering Systems, 20, 75-96

5. Aditi Sharma “Prediction of Flight-fare using machine learning “ in Conference: 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP) Nov 2022
6. R. Ren, Y. Yang and S. Yuan, “Prediction of airline ticket price.” University of Stanford, 2014.
7. Gordiievych and I. Shubin, "Forecasting of airfare prices using time series," 2015 Information Technologies in Innovation Business Conference (ITIB), 2015, pp. 68-71, doi: 10.1109/ITIB.2015.7355055.
8. Lantseva, K. Mukhina, A. Nikishova, S. Ivanov, and K. Knyazkov, “Data-driven modeling of airlines pricing.” Procedia Computer Science, 2015, 66, 267-276.
9. S. L. Puller and L. M. Taylor, “Price discrimination by day-of-week of purchase: Evidence from the U.S. Airline Industry,” Journal of Economic Behavior & Organization, vol. 84, no. 3, pp. 801–812, 2012.
10. J. Abdella, M. N. Zaki, K. Shuaib and F. Khan, “Airline ticket price and demand prediction: A survey.”, Journal of King Saud University-Computer and Information Sciences, 33(4), 375-391, 2021.
11. K.D.V.N.Vaishnavi, L. Hima Bindu, M. Satwika, K. Udaya Lakshmi, M. Harini, N. Ashok “FLIGHT FARE PREDICTION USING MACHINE LEARNING” EPRA International Journal of Research & Development (IJRD)
12. https://medium.com/geekculture/flight-fare-prediction-93da3958eb95