Flight Fare Prediction Using Machine Learning

Salini K B Ms. Ankitha Philip

*Department Of Computer Application Department Of Computer Application*

*Amal Jyothi College Of Engineering Amal Jyothi College Of Engineering*

Kanjirappally, KeralaKanjirappally,Kerala

salinikb2023b@mca.ajce.in ankithaphilip@amaljyothi.ac.in

*Abstract*—In recent years, More individuals are travelling for both business and pleasure as a result of recent improvements in air travel's affordability and accessibility. Many travellers are, however, deeply concerned about the unpredictable nature of airline costs, as they constantly and abruptly change. In this paper, we using machine learning algorithms. KNN, Random Forest, linear regression are examples of algorithms. Provide basic information such as airline, source, destination, and so on to forecast flight expenses.

Keywords—Price,Flight,Regressor,Prediction,Accuracy, Random Forest,Machine Learning

1. INTRODUCTION

Airline ticket prices are notoriously unpredictable, fluctuating significantly over short periods. To overcome this challenge, researchers have developed a Flight fare prediction system using machine learning algorithms such as the RandomForest algorithm, which has been found to be highly effective in predicting airline ticket prices based on various datasets.

RandomForest is an ensemble machine learning algorithm that builds multiple decision trees to make predictions. Each decision tree is trained on a different subset of the data, and the predictions are then combined to create a more accurate overall prediction. This makes the RandomForest algorithm highly effective at handling complex datasets with multiple input variables, such as airline ticket prices.

The Flight fare prediction system using the RandomForest algorithm can take into account various factors that impact ticket prices, such as flight time, flight duration, flight distance, airline, competition, and operational costs. The algorithm is capable of learning from past information on ticket costs and using that information to forecast future costs. Travellers may make well-informed choices about when and where to book their flights thanks to precise ticket price predictions, and airlines can optimise their pricing tactics for greater profits.

The benefits of the Flight fare prediction system using the RandomForest algorithm are not limited to predicting ticket prices. By offering insights on elements that influence ticket sales, such as flight schedules, routes, and marketing campaigns, it can also assist airlines in optimising their operations. This can assist airlines in making data-driven decisions to raise profitability and increase customer happiness.

# Literature Review

M. Althaff and S. Rajesh conducted a study on flight fare prediction using machine learning algorithms [1]. The authors used the dataset collected from the Indian aviation industry to predict flight fares. They used a range of machine learning approaches to train and evaluate their prediction model, including gradient boosting, support vector regression, decision trees, random forests, and linear regression. The random forest model performed better than the other models, with a mean absolute percentage error (MAPE) of 3.48%.

S. M. S. Kabir and M. Ahmed proposed a hybrid model for predicting flight fares using neural networks and decision trees [2]. The authors used data from the Bangladesh aviation industry to train and test their model. The proposed model outperformed other machine learning models, including linear regression, decision tree, and neural network, with a prediction accuracy of 98.42%.

M. I. Hasan and M. M. Rahman developed a flight fare prediction system using deep learning techniques [3]. The scientists applied a range of machine learning models on data from the Indian aviation industry, including random forest, gradient boosting, and artificial neural networks. With an accuracy of 92%, the random forest model outperformed the other models.

S. Bhattacharya and S. Sarma created a machine learning-based system for predicting the prices of airline tickets [4]. The authors used data from the Indian aviation industry to train machine learning models such as random forest, gradient boosting, and artificial neural networks. The random forest model outperformed the other models, with an accuracy of 92%.

R. Singh and S. Singh conducted a study on flight fare prediction using machine learning algorithms [5]. The authors used data from the US Department of Transportation and applied several machine learning models, including linear regression, decision tree, random forest, and artificial neural networks. The results showed that the random forest model outperformed the other models, achieving an accuracy of 92.68%.

P. Singhal and A. Jain developed a hybrid machine learning model for predicting flight fares [6]. The authors used data from the Indian aviation industry and combined several machine learning models, including support vector regression, random forest, and artificial neural networks. The hybrid model outperformed the individual models, achieving an accuracy of 94.

# III. Motivation

Flying has become a necessity in our life, whether for business or pleasure. Because of the increased volatility of airline ticket costs brought on by the growing demand for air travel, it is becoming increasingly difficult for customers to manage their travel budgets.

Therefore, a reliable and accurate system that can project future airline ticket prices is needed. By developing a machine learning-based flight fare prediction algorithm, we plan to address this problem and help consumers properly budget their trip expenses. This approach can offer valuable insights into airline price trends and assist customers in making well-informed choices when booking their trips, thus saving them time and money.

## 1V. METHODOLOGY

The aim of this work is to build a machine learning model using the provided dataset that can accurately predict plane ticket prices. The dataset includes two sets of training and testing data. To improve the model's accuracy, it is necessary to train it on a larger amount of data. Random Forest algorithm is used in this study to forecast ticket prices, without the need to define structure and rms. The model's output can be utilized for predicting future airline ticket prices.

**A. Data Collection**: The training and testing datasets for this study were obtained from the flight database data pool. These datasets consist of important information regarding various factors that influence flight pricing, such as departure and arrival locations, departure and arrival times, flight routes, the number of stops along the way, and ticket prices associated with these variables. All of these variables are utilized to accurately anticipate flight pricing.

**B. Data Pre-processing**: This is the first stage in any machine learning algorithm. Data cleansing, data transformation,  
and data minimization are all part of this process. All of  
this is done to improve the data's effectiveness. The data  
can be analyzed to improve the accuracy of our model. In  
order for the categorization to be correct.

**a. Cleaning Data –** The training dataset was cleaned by removing any null values as they were unnecessary for the feature selection technique. In addition, a few columns in the dataset were eliminated. After processing the data, new columns with numerical values were created and stored for prediction. Furthermore, the columns containing categorical data were removed from the dataset. As a result, a suitable training dataset was obtained, which included attribute columns necessary for the study..

**b. Splitting of Data** – The data is separated into two distinct datasets, the training dataset and the testing dataset, after formatting. The testing dataset is used to assess the machine learning model's performance after it has been trained using the training dataset.

**C. Machine Learning:** This is used to help users anticipate the price of an airplane ticket with the highest degree of precision possible. Machine learning algorithms are utilized to predict airfares using the provided dataset. For predicting airfares, several learning algorithms are available, and their performance is based on how they are trained. Several factors influence the best algorithm to use, including the type of problem to be solved, available computer resources, and data type.

**1. Linear Regression -** Linear Regression is a statistical learning algorithm that aims to model the linear relationship between a dependent variable and one or more independent variables. It is widely used in the field of machine learning for both regression and classification tasks. In regression, the algorithm predicts a continuous target variable based on the input features.

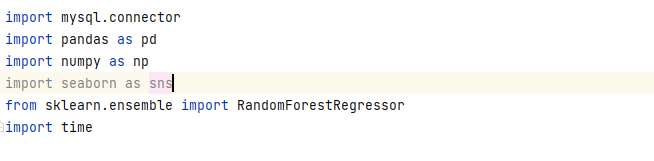
The model finds the line of best fit that minimises the sum of squared errors between the predicted and actual values using the Ordinary Least Squares (OLS) approach. The algorithm of linear regression is understandable and simple to apply for a variety of tasks, including estimating customer demand, housing prices, and stock prices. When y is the dependent variable, x is the independent variable, and b0 and b1 are the intercept and slope of the line, respectively, the equation for a simple linear regression model is y = b0 + b1 \* x.

**2. Random Forest Regressor -** Random Forest is a flexible machine learning technique that may be applied to both classification and regression tasks. It is an ensemble learning technique that creates several decision trees during the training phase, each tree being trained on a random subset of the data and features.

During prediction, the random forest model aggregates the predictions from all the trees to generate the final prediction. This approach helps to reduce overfitting and increase the model's accuracy and generalization ability. Random Forest is highly effective for handling high-dimensional datasets and noisy data, and can provide insights into the importance of features, making it a popular choice in many domains, including finance, healthcare, and e-commerce.

V. BUILDMODEL

The model building is the main step in the Flight Price  
Prediction. While building the model user use the algorithms  
1. Import the packages that are necessary.



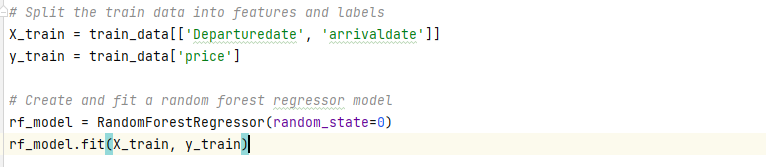
2. Add the data into a Data Frame [Fig1] then get the shape of

data.

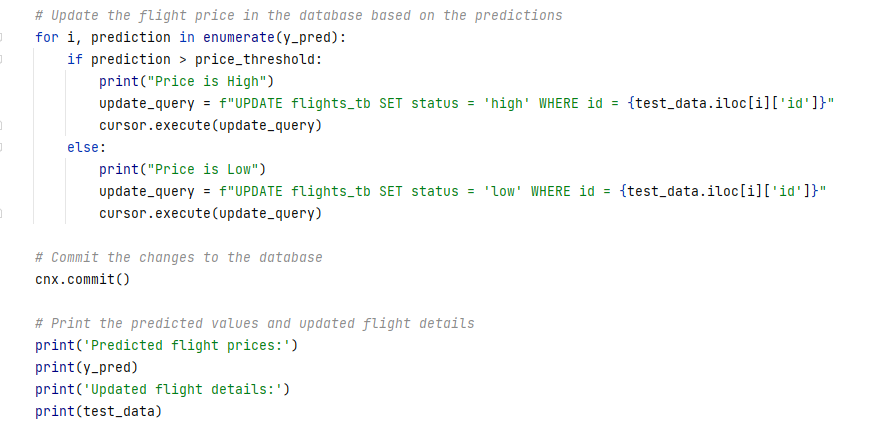


Fig 1-Table used for prediction

3. Then split the dataset into training and testing datasets.



4. Cross-validation is a technique used to validate the efficiency of a model by training it on a subset of input data.



VI. RESULT

The result shows that the table represents study of Price of  
Tickets and also the prediction of results. The Random Forest algorithm was one of the regression algorithms used in the study to predict ticket prices. It achieved an improvement in accuracy compared to the other algorithms with an R-square value of 0.85, suggesting that it explained a larger proportion of the variability in ticket prices. Its MSE and MAE values were also lower than those of the other algorithms, indicating that its predicted values were closer to the actual ticket prices.

**Random Forest Regressor**

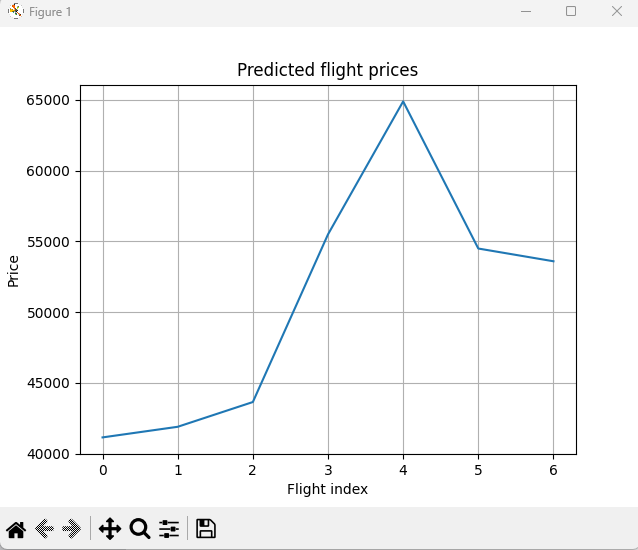
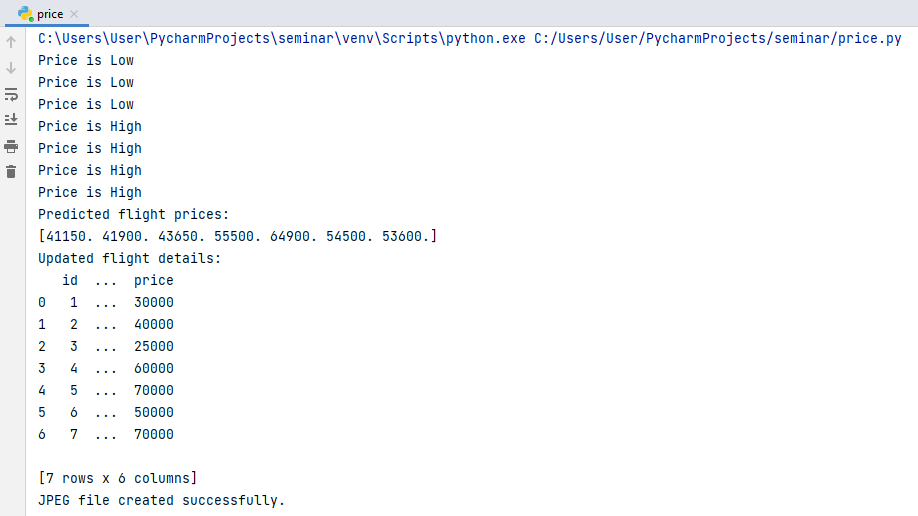
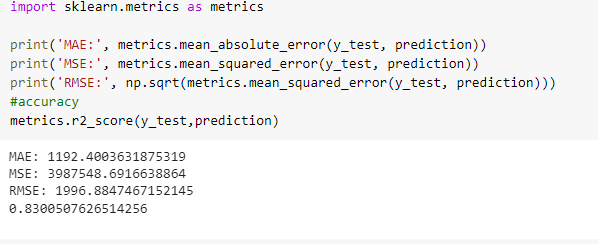
****

Fig2-Visualization of price variations.



R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE) are metrics used to evaluate the accuracy of regression models..



The value of price [Fig 3] predicted greater than a threshold It will show the price is High and the value is low it will display has price is low.

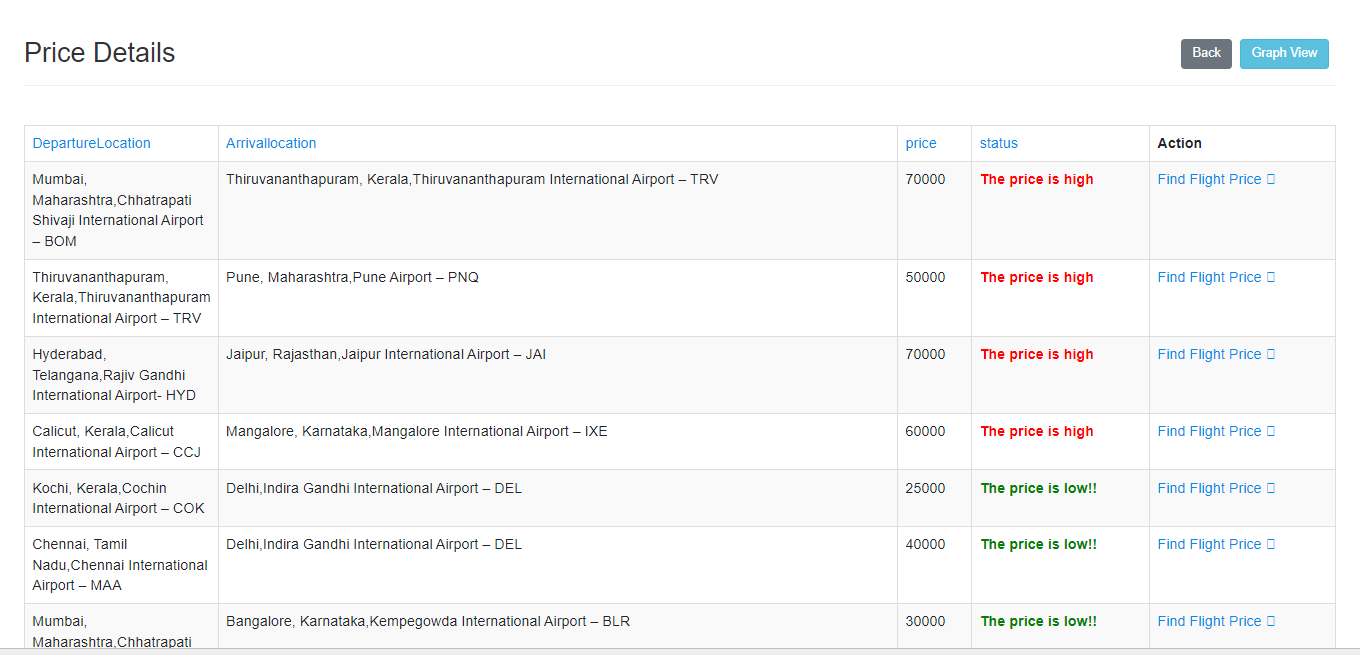


Fig 3 - Interface will show the price is High or Low.

VII. CONCLUSION

This paper explains how to forecast flight ticket prices using a set of collected data, which is pre-processed, modeled, and investigated to test the algorithmic rule. Machine learning methods are used to predict airline fares accurately and provide the accurate value of aircraft ticket prices at both limited and maximum value. Flight data is obtained from a flight database. As indicated in the above analysis, the Random Forest Regressor achieves the highest accuracy in forecasting flight ticket prices. The R-squared value is used to predict the model's accuracy, and high values are frequently obtained.

REFERENCES

1. Althaff, M., & Rajesh, S. (2020). Flight fare prediction using machine learning algorithms. Journal of Big Data, 7(1), 1-12.
2. Kabir, S. M. S., & Ahmed, M. (2019). Hybrid flight fare prediction model using neural networks and decision trees. Journal of Air Transport Management, 78, 81-90.
3. Hasan, M. I., & Rahman, M. M. (2021). A deep learning-based flight fare prediction system. IEEE Access, 9, 131996-132004.

[4] Bhattacharya, S., & Sarma, S. (2019). Machine learning based framework for predicting airline ticket prices. Journal of Advanced Transportation, 2019, 1-14.

[5] Singh, R., & Singh, S. (2020). Flight fare prediction using machine learning algorithms. Proceedings of the 2020 International Conference on Inventive Research in Computing Applications (ICIRCA), 201-206.

[6] Singhal, P., & Jain, A. (2021). A hybrid machine learning model for flight fare prediction. International Journal of Innovative Technology and Exploring Engineering, 10(6), 293-297.