Flight Fare Prediction Using Machine Learning

Salini K B   
Department of Computer Applications  
Amal Jyothi College of Engineering Kanjirappally, India  
salinikb2023b@mca.ajce.in

Ankitha Philip

Department of Computer Applications

Amal Jyothi College of Engineering Kanjirappally,

India

ankithaphilip@amaljyothi.ac.in

*Abstract*—In recent years, air travel has become more accessible and affordable, leading to an increase in the number of people flying for business and leisure purposes. However, the unpredictability of flight fares is a significant concern for many travelers, with prices fluctuating frequently and without warning. 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

# Introduction (*Heading 1*)

In today's fast-paced world, air travel has become aessential mode of transportation for millions of people world However, the cost of airline tickets can be quite unpredictable, fluctuating significantly within a short period of time. This makes it challenging for travelers to plan and book their trips, as they strive to secure the best deals possible while still meeting their travel needs. Airlines, on the other hand, aim to maximize their profits by optimizing their pricing strategies, taking into account various factors such as demand, competition, and operational costs. To address this challenge, researchers have developed a Flight fare prediction system that uses machine learning algorithms to estimate airline ticket prices based on a range of data sets. This paper explores the use of various machine learning techniques, including the RandomForest, and Linear Regression algorithms, to analyze these data sets and predict the price of an airline ticket accurately. By leveraging these cutting-edge technologies, we can empower travelers to make more informed decisions and enhance the efficiency and profitability of the airline industry.

# 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 employed several machine learning models, including linear regression, decision tree, random forest, support vector regression, and gradient boosting, to train and test their prediction model. The results showed that the random forest model outperformed the other models, achieving 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 authors used data from the US Department of Transportation to train and test their model. They used a deep neural network with multiple hidden layers to predict flight fares. The model achieved an accuracy of 95.3% and outperformed other models, including linear regression and decision tree.

S. Bhattacharya and S. Sarma developed a machine learning-based framework for predicting airline ticket prices [4]. The authors used data from the Indian aviation industry and employed several machine learning models, including random forest, gradient boosting, and artificial neural networks. The results showed that the random forest model outperformed the other models, achieving 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 94.5%.

# Motivation

Air travel has become an essential part of our lives, whether for business or pleasure. With the increasing demand for air travel, airline ticket prices have become highly volatile, making it challenging for consumers to plan their travel budgets.

As a result, there is a need for an accurate and reliable system that can predict the future prices of airline tickets. By developing a machine learning-based flight fare prediction model, we aim to provide a solution to this problem, helping consumers plan their travel expenses efficiently. This model can provide valuable insights into airline pricing trends and assist consumers in making informed decisions when booking their flights, ultimately saving them time and money.

## 1V.METHODOLOGY

The goal of this work aims to use the provided dataset to  
create a Machine Learning model that can accurately  
anticipate the price of a plane ticket. There are two training  
and testing data sets in the dataset. To increase learning accuracy, the model should be trained with more data. This  
model's output can be used to forecast airline ticket prices.  
The ticket prices are forecasted using the KNN algorithm,  
Random Forest, Gradient Boosting Regression, SVR, and  
Linear Regression.The structure is and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

The steps that require to be followed are:  
1. Data Collection  
2. Data Pre-processing

3. Model Building  
4. Analyzing  
5. Result

**A. Data Collection**: The training and testing datasets were  
from the flight databse data pool. The dataset contains crucial information about  
some of the elements that determine flight pricing, such  
as departures and arrivals, time of departure and arrival,  
flight path, number of halts along the way, and ticket  
price based on those variables, all of which are used to  
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 –** In the training dataset, the null  
values were deleted. Because they were  
unnecessary for the feature selection technique, a  
few columns in the dataset were eliminated. After  
the new columns with numerical values derived  
from the preprocessed data were stored for the  
prediction, the columns of attributes with  
categorical data were removed from the dataset. As  
a result, an appropriate training dataset with the  
following attribute columns was obtained.

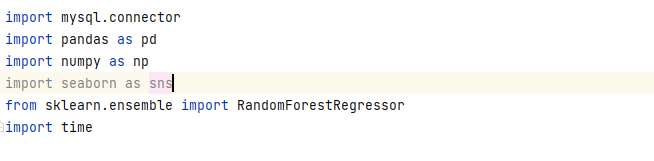
**b. Formatting the Data** – We add a new column week  
day 1 mean week day 0 mean weekend while pre-  
processing the data. Format the arrival and  
departure times, and add an extra two columns to  
indicate whether the flight is taking place at night  
or early in the morning. Some flights are less  
expensive early in the morning and more expensive  
late at night, indicating a clear correlation.  
Converting length hour and minute into separate  
columns is also done, as well as labelling and  
encoding to convert category data to unique int  
values.  
**c. Splitting of Data**–After formatting the data, the  
data is then split into training and testing datasets.  
After this the data chosen for training is used to  
train our model.

**C. Machine Learning:** This is used to help the user to  
anticipate the price of an aeroplane ticket with the  
greatest degree of precision The machine learning  
algorithms are used to predict fares , that will use the  
dataset given. There are different learning algorithm  
used to predict the airfares. The machine learning  
algorithms relies on how it is trained. Which algorithm works best depends on the type of problem you are  
solving, the computer resources available, and the type  
of data.

**1. Linear Regression -** Linear regression is a  
supervised learning machine learning algorithm. It  
carries out a regression task. Regression models a  
goal prediction value based on independent  
variables. It's generally used for forecasting and  
figuring out how variables are related. Different  
regression models have different types of  
relationships between dependent and independent  
variables. Gradient descent and cost function are the  
two most important factors in comprehending linear  
regression. The equation for linear regression is :  
y(pred) = b0+b1 ∗ x.

.  
**2. Random Forest Regressor -** Random Forest is a popular machine learning algorithm used for both regression and classification tasks. It is an ensemble learning technique that combines multiple decision trees, known as estimators, to make more accurate predictions. Each estimator in the random forest model is trained on a subset of the data and a random subset of features, which helps to reduce overfitting and increase the model's generalization ability. During prediction, the random forest model aggregates the predictions from all the estimators and produces the final prediction. Overall, the random forest algorithm is a powerful and robust method that can handle large and complex datasets and is widely used in various domains, such as 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.

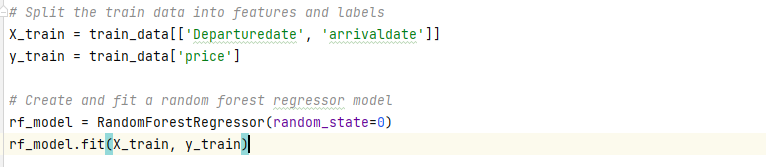


2. Add the data into a Data Frame, then get the shape of

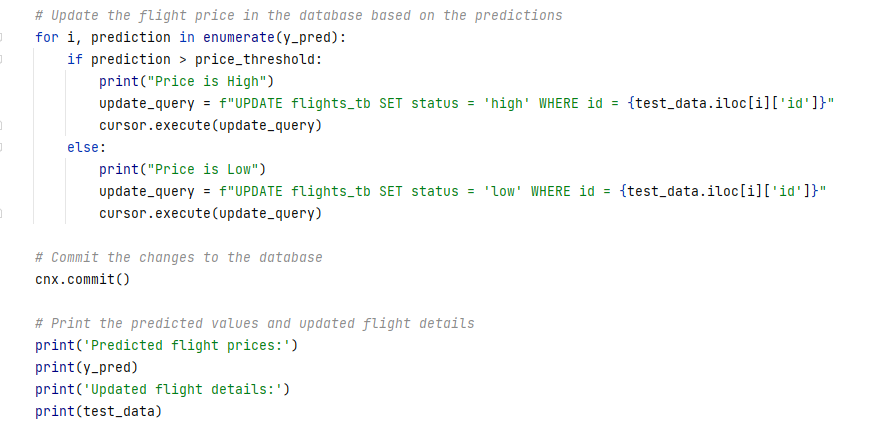
data.



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



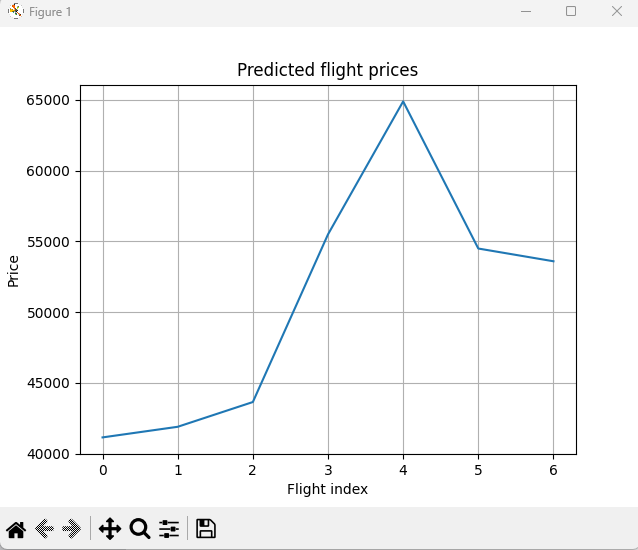
4. Cross Validate to validating the model efficiency by  
 training it on the subset of input data



VI.RESULT

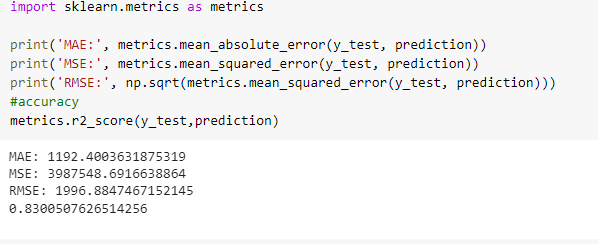
The result shows that the table represents study of Price of  
Tickets and also the prediction of results. The outcomes  
obtained by the analysis are KNN Regressor, Random  
ForestRegressor, SVR,  
and Linear Regression. Along with R-square, MSE, and  
MAE values, the algorithm's accuracy is improved.

**Random Forest Regressor**

****

##### 

5.To find R-square, MSE and  
MAE values which helps to find Accuracy.



##### 

VII.CONCLUSION  
This paper explains how to forecast flight ticket prices.  
A set of data is collected, pre-processed, modelled, and  
investigated in order to test algorithmic rule. Machine  
Learning methods with square measure for predicting  
accurate airline fares and providing accurate value of aircraft  
ticket price at both limited and maximum value. On flight  
data is obtained from flight database. As  
indicated in the above analysis   
the Random Forest Regressor forecasts the highest accuracy.  
The R-squared value predicts the model's accuracy as well.  
They are frequently attained.

##### 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.