Flight Fare Prediction Using Machine Learning

Salini K B
Department Of Computer Application
Amal Jyothi College Of Engineering
Kanjirappally, Kerala
salinikb2023b@mca.ajce.in

Ms. Ankitha Philip
Department Of Computer Application
Amal Jyothi College Of Engineering
Kanjirappally,Kerala
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**

I. 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.

II. LITERATURE REVIEW

Wang et al. proposed a framework for airfare price prediction using machine learning techniques [1]. The authors utilized a combination of machine learning algorithms including decision tree, random forest, and gradient boosting, to predict airfare prices with high accuracy. Their results showed that the random forest algorithm outperformed other models with an accuracy of 85.53% for domestic airfare prediction and 87.63% for international airfare prediction.

Ratnakanth developed a deep learning-based system for predicting flight fares [2]. The author employed a Long Short-Term Memory (LSTM) model to predict airfare prices, achieving improved performance in terms of prediction accuracy compared to traditional machine learning models. The study also highlighted the importance of feature engineering in improving the performance of the deep learning models for flight fare prediction.

Subramanian et al. developed a machine learning-based approach to predict airline fares [3]. The authors compared the performance of different machine learning algorithms such as decision tree, random forest, and support vector regression and reported that the random forest algorithm performed the best in terms of prediction accuracy. This approach helped to capture more complex relationships between the features and the target variable, resulting in improved prediction performance.

Sharma and Carpenter provided an analysis of machine learning techniques for airfare prediction [4]. The authors discussed various machine learning algorithms used in airfare prediction and highlighted their strengths and weaknesses. their analysis provides a comprehensive overview of the current state-of-the-art in machine learning-based airfare prediction.

Champawat et al. proposed a system for Indian flight fare prediction [5]. The authors suggested the use of a hybrid model combining multiple machine learning algorithms such as random forest, decision tree, and support vector regression to achieve high prediction accuracy.

Joseph et al. used machine learning algorithms to predict flight ticket prices [6]. The authors compared the performance of various machine learning models such as linear regression, random forest, and support vector regression, and reported that the random forest algorithm outperformed the other models in terms of prediction accuracy.

Tian et al. conducted a survey on data analytics for air travel data [7]. The authors reviewed the literature on airfare prediction using machine learning techniques and provided new perspectives for future research directions.

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.

```
import mysql.connector
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.ensemble import RandomForestRegressor
import time
```

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

	id	c_id	Airline	flightNo	DepartureLocation	Arrivallocation	Departuredate	arrivaldate	seats	price	status
Edit Copy Delete	1	0	Vistara(VTI)	VTIVO1	Mumbai, Maharashtra,Chhatrapati Shivaji Internatio...	Bengaluru, Karnataka,Kempegowda International Airp...	2023-04-13 11:00:00	2023-04-13 14:40:00	60	30000	low
Edit Copy Delete	2	0	FlyBig(FLG)	FLG101	Chennai, Tamil Nadu,Chennai International Airport ...	Delhi,Indira Gandhi International Airport – DEL	2023-03-20 07:40:00	2023-04-20 11:40:00	60	40000	low
Edit Copy Delete	3	0	Alliance Air(LLR)	LLR40S	Kochi, Kerala,Cochin International Airport – COK	Delhi,Indira Gandhi International Airport – DEL	2023-04-20 13:20:00	2023-04-20 15:21:00	60	25000	low
Edit Copy Delete	4	0	Akasa Air(Aki)	AKJ101	Calicut, Kerala,Calicut International Airport – C...	Mangalore, Karnataka,Mangalore International Airpo...	2023-04-09 01:23:00	2023-04-09 03:24:00	60	60000	high
Edit Copy Delete	5	0	Airasia(AD)	IAD101	Hyderabad, Telangana,Rajiv Gandhi International Ai...	Jaipur, Rajasthan,Jaipur International Airport – J...	2023-04-11 07:30:00	2023-03-31 15:28:00	60	70000	high
Edit Copy Delete	6	0	Star Air(SDG)	SDG1W2	Thiruvananthapuram, Kerala,Thiruvananthapuram Inte...	Pune, Maharashtra,Pune Airport – PNQ	2023-04-02 08:40:00	2023-04-02 11:34:00	60	50000	high
Edit Copy Delete	7	0	Airasia(AD)	IAD123	Mumbai, Maharashtra,Chhatrapati Shivaji Internatio...	Thiruvananthapuram, Kerala,Thiruvananthapuram Inte...	2023-04-05 12:37:00	2023-04-05 15:37:00	60	70000	high

Fig 1-Table used for prediction

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

```
# Split the train data into features and labels
X_train = train_data[['Departuredate', 'arrivaldate']]
y_train = train_data['price']

# Create and fit a random forest regressor model
rf_model = RandomForestRegressor(random_state=0)
rf_model.fit(X_train, y_train)
```

```
C:\Users\User\PycharmProjects\seminar\venv\Scripts\python.exe C:/Users/User/PycharmProjects/seminar/price.py
Price is Low
Price is Low
Price is Low
Price is High
Price is High
Price is High
Price is High
Predicted flight prices:
[41150. 41900. 43650. 55500. 64900. 54500. 53600.]
Updated flight details:
   id  ...  price
0   1  ...  30000
1   2  ...  40000
2   3  ...  25000
3   4  ...  60000
4   5  ...  70000
5   6  ...  50000
6   7  ...  70000

[7 rows x 6 columns]
JPEG file created successfully.
```

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

```
# Update the flight price in the database based on the predictions
for i, prediction in enumerate(y_pred):
    if prediction > price_threshold:
        print("Price is High")
        update_query = f"UPDATE flights_tb SET status = 'high' WHERE id = {test_data.iloc[i]['id']}"
        cursor.execute(update_query)
    else:
        print("Price is Low")
        update_query = f"UPDATE flights_tb SET status = 'low' WHERE id = {test_data.iloc[i]['id']}"
        cursor.execute(update_query)

# Commit the changes to the database
cnx.commit()

# Print the predicted values and updated flight details
print('Predicted flight prices:')
print(y_pred)
print('Updated flight details:')
print(test_data)
```

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

```
import sklearn.metrics as metrics

print('MAE:', metrics.mean_absolute_error(y_test, prediction))
print('MSE:', metrics.mean_squared_error(y_test, prediction))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, prediction)))
#accuracy
metrics.r2_score(y_test,prediction)
```

```
MAE: 1192.4003631875319
MSE: 3987548.6916638864
RMSE: 1996.8847467152145
0.8300507626514256
```

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

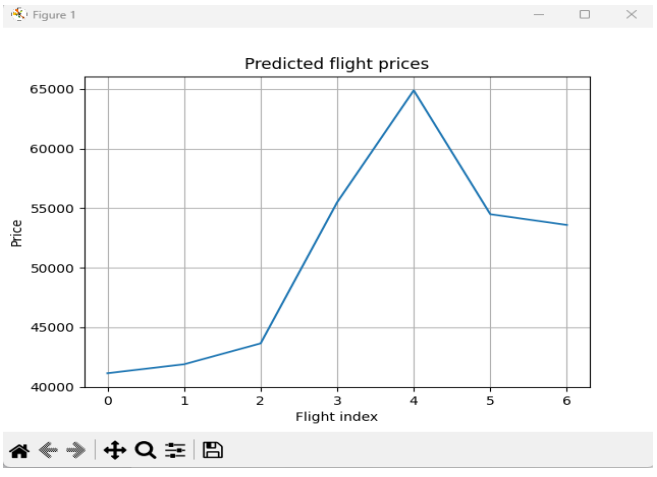

Fig2-Visualization of price variations.

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.

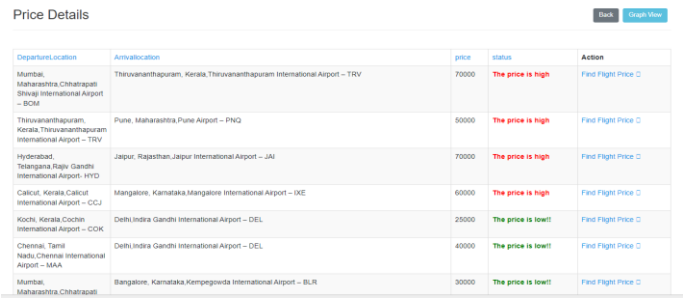

Fig 3 - Interface will show the price is High or Low.

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] Wang, T., et al. (2019). A framework for airfare price prediction: a machine learning approach. In 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 200-207). IEEE. https://doi.org/10.1109/IRI.2019.00041

[2] Ratnakanth, G. (2022). Prediction of flight fare using deep learning techniques. In International Conference on Computing, Communication and Power Technology (IC3P) (pp.308-313)https://doi.org/10.1109/IC3P52835.2022.00071

[3] Subramanian, R.R., Murali, M.S., Deepak, B., Deepak, P., Reddy, H.N., & Sudharsan, R.R. (2022). Airline fare prediction using machine learning algorithms. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 877-884). https://doi.org/10.1109/ICSSIT53264.2022.9716563.

[4] Sharma, L., & Carpenter, M. (Eds.). (2022). Analysis of machine learning techniques for airfare prediction. In Computer Vision and Internet of Things: Technologies and Applications (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003244165

[5] Champawat, J.S., Arora, U., & Vijaya, K. (2021). Indian flight fare prediction: a proposal. International Journal of Advanced Technology and Engineering Science, 9(3).

[6] Joseph, J., et al.: Flight ticket price predicting with the use of machine learning. Int. J. Adv. Trends Comput. Sci. Eng. 10(2), 1243–1246 (2021).

[7] Tian, Z., Li, Y., Li, J., Li, X.: A survey on data analytics for air travel data. IEEE Access 7, 14269-14279 (2019). https://doi.org/10.1109/ACCESS.2019.2892796