International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

FLIGHT TICKET PRICE PREDICTION

Mudagal Nagarjun[1], Roopa R[2]

[1] Student, Department of MCA,BIET,Davangere

[2] Assistant Professor, Department of MCA,BIET,Davangere

ABSTRACT

The study uses past flight schedules, route data, and ticket prices to estimate airline ticket costs using machine learning regression. For ease of use and functionality, it has admin and user modules. Registering and logging in allows users to upload flight data for precise cost estimates. Ensuring continuous efficacy, the admin module makes data administration and system maintenance easier. The objective is to improve overall travel planning experiences by providing travelers with data-driven insights to help them make wise decisions and maximize the value of their airline ticket purchases.

Keyword: Machine learning, Flight ticket Prediction, Flight fare.

I. INTRODUCTION

The method by which travellers purchase flights has changed dramatically as a result of the travel industry's digital transformation, especially with the emergence of online ticketing platforms. The swift growth of domestic aviation in India has added complexity to flight pricing methods, making it more difficult for travelers to get the cheapest offers. In order to dynamically modify ticket rates based on a range of parameters such demand, seat availability, booking time, and competitive pricing, airlines utilize advanced revenue management systems. Because of this, costs can vary greatly, which frequently leaves clients unsure of the best time to buy tickets. Our research attempts to create a machine learning-based method for precisely forecasting airfare prices over time in order to tackle this difficulty. The goal of this project is to develop an intuitive web application that combines both historical and current data to anticipate fares accurately for different airlines. Through the use of sophisticated machine learning algorithms, we are able to examine intricate pricing trends and provide consumers with useful information to help them plan their travels. The creation of a single user module for the web application is the main focus of the work. To access the system, users must register and log in using legitimate credentials. After logging up, users can submit particular flight data and

get comprehensive forecasts for future ticket costs. In the end, this feature will improve tourists' travel experience by assisting them in organizing their purchases according to their tastes and financial limitations. Our predictive models are constructed from previous flight schedules, air route data, and ticket prices that are painstakingly examined to extract pertinent elements. With the help of these characteristics, the ML algorithms can predict prices with a high degree of accuracy while accounting for the various factors that affect changes in airfare. This initiative intends to enable Indian travelers to more skillfully negotiate the complexity of airline pricing by offering precise and timely airfare estimates. Predicting future pricing can make travel planning more successful and economical, in addition to assisting in obtaining the greatest rates. This project, which demystifies airline pricing techniques and provides a trustworthy tool for airfare prediction, will benefit a wide variety of customers, from frequent travellers to infrequent visitors.

II. RELATED WORK

As demonstrated, there are several ways to take advantage of the airfare price prediction challenge, including consumer segmentation, scheduling ticket purchases, predicting demand for airline tickets, and more. in an evaluation of the target application problem and its solutions by Abdella et al. [1].

The field of airfare price prediction has been actively studied over the past thirty years; a Scopus search using the keyword "airfare price prediction" produced twenty-four papers ranging from 2003 to the present, the majority of which have been applied in the previous three years. Utilizing variables related to time to characterize flights operated by Vietnamese national airline companies [2]

Vu et al developed an application for predicting airfare prices using two machine learning models. Less models have been offered and only one airline business has been taken into consideration in comparison to the suggested strategy, The target applications for consumers were the main focus. A different strategy was offered in [3].

A bespoke recurrent neural network (RNN) was developed and its performance in predicting flight prices during events such as basketball games was contrasted with traditional machine learning models. By combining features from airline trips and basketball games into a single dataset, excellent prediction accuracies were attained. The identical methodology was used in [4].

The authors suggested a framework that could collect data on airline tickets from multiple sources, including consumer interests, the availability of tickets, distance, and more, in order to forecast flight costs using machine learning models. Airfare price prediction was introduced in the Indian and American domestic markets in [5].

The authors claimed an 88% score in price prediction using machine learning algorithms [6].

In Joshi et al. used a similar strategy but used fewer machine learning models. They looked into novel characteristics like flight time and were able to predict scores of up to 90%. In order to anticipate airfare prices. [7].

used feature selection algorithms in conjunction with hyperparameter techniques to identify the ideal model parameters and feature set for flight descriptions [8].

In explainability for the studied problem is presented in order to gain a deeper understanding of the models that may offer an effective solution and produce reliable and understandable forecasts. N. S. S. V. S. Rao, The purpose of the study was to evaluate the predictive accuracy of Decision Tree Regression and Random Forest Regressor algorithms for estimating airline ticket prices.

Statistical study showed that the Random Forest Regressor achieved a superior prediction accuracy of 86.70% compared to 79.69% for the Decision Tree Regressor with a sample size of 20 samples divided into two groups of 10 samples each. These results illustrate Random Forest Regression's potential advantage in real-world airline pricing prediction applications by indicating that it is more effective at properly forecasting aircraft ticket costs[9].

S. J. Thilak, B. P. Benny , The aim of this research is to create a predictive model for airline ticket price variations by utilizing machine learning methods, specifically Random Forest and Randomized Search CV. The study intends to provide travelers with insights that enable educated purchasing decisions based on anticipated changes in airfare by utilizing two Kaggle datasets[10].

III. METHODOLOGY

The objective of this study is to produce an accurate and effective airfare prediction system that will enable travelers to plan ahead and find the most affordable airline tickets. Figure 1 below shows the architecture diagram of a flight ticket.

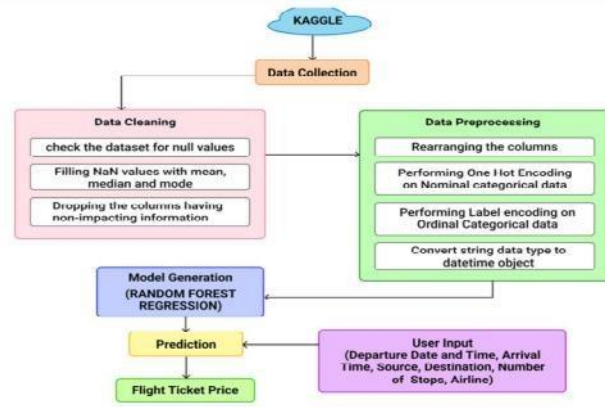

Figure 1: Flight Ticket Price Prediction Architecture

1. Data Sources:

Compile information on flight times, costs, and routes. Aggregators of Third Parties: Make use of travel agency data to create an extensive dataset. Gather historical flight information to spot patterns and trends over time.

2. Data Attributes:

Airlines, flight numbers, arrival and departure times, duration, and route are examples of data attributes. Total ticket price, taxes, fees, and base fare. Reservation date and time, departure date and time, and seasonality. Reservation patterns and seat availability.

3. Data Preprocessing

Eliminate entries that are redundant or unrelated. Use imputation or interpolation methods to deal with missing values. Add new features like the day of the week, holidays, and times when most people travel. To guarantee consistency, scale and normalize numerical data. To assess the performance of the model, divide the dataset into test, validation, and training sets.

4. Model Development

To determine which features are most important for predicting flight prices, use correlation analysis and feature importance measures. Choose a suitable machine learning model following data pretreatment and feature engineering. The selection of linear regression for this project is based on its interpretability and simplicity.

5. Model Training:

Utilize the historical data to train the linear regression model. Throughout training, the model finds patterns and correlations between the input data and the goal variable, the cost of airline tickets.

6. Model Evaluation:

Evaluate the model with a different dataset (cross-validation or validation set). Metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) can be used to assess performance.

7. Model Tuning:

If necessary, use hyperparameter tuning methods to fine-tune the model's parameters and enhance performance, such as grid search or random search.

8. Model Deployment:

After the model has undergone satisfactory training and evaluation, it can be

implemented in a production setting to generate real-time predictions based on flight data. This entails developing an application or API that enables users to enter pertinent flight data and obtain estimated ticket pricing as an output.

3.1 DATASET USED

Using datasets from Kaggle, the study focuses on historical flight data that is necessary for estimating airline ticket prices. Detailed data like flight paths, arrival and departure times, airline operators, ticket costs, and passenger booking patterns are commonly included in these databases. The goal of the project is to use machine learning techniques such as Random Forest and Randomized Search CV to create predictive models by utilizing these datasets. These models are intended to predict changes in airfare, offering useful information to help travelers decide what to buy based on anticipated price shifts.

libraries commonly used predicting airline ticket prices, include

- **Pandas:** Used for preprocessing and data manipulation, including loading datasets, organizing features, and cleaning data.
- **scikit-learn:** Implementing machine learning algorithms such as Random Forest Regression, Decision Tree Regression, and other ensemble techniques requires the use of scikit-learn. It also offers model evaluation, selection, and hyperparameter tuning tools.
- **numpy:** Numerical computations and array operations with numpy are fundamental to machine learning workflows' data processing.
- **matplotlib and seaborn:** Plotting and charting data distributions, variable relationships, and model performance indicators are made possible by the visualization tools matplotlib and seaborn.

3.2 DATA PRE-PROCESSING

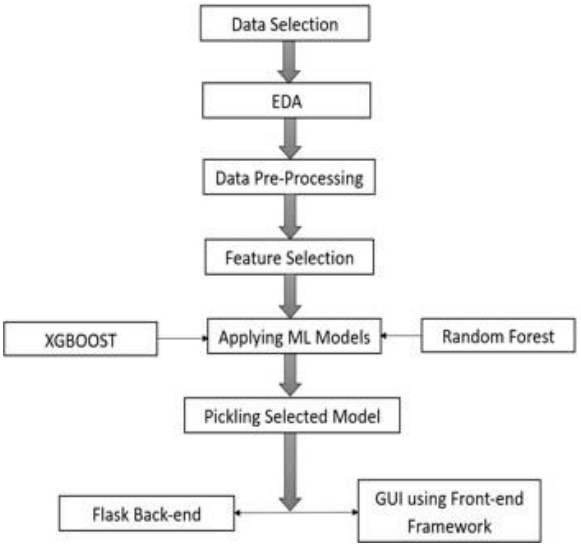

Figure 3.2 : Flow chart of random forest

Preprocessing data is a crucial stage in anticipating airline ticket pricing. Initially, the information is imported from sites like Kaggle into a pandas DataFrame. This includes looking for and dealing with outliers, discrepancies, and missing values.

Airline carriers and flight routes are examples of categorical variables that are encoded into numerical representations using techniques like label encoding and one-hot encoding. In order to ensure that every numerical feature contributes equally to model training, such as flight times and ticket prices, these features are frequently scaled. Correlation analysis and tree-based model feature importance are two feature selection techniques that assist in determining which variables have the most influence on ticket prices. In order to train the model and evaluate its performance, the data is finally divided into training and testing sets. To make sure the data is prepared for machine learning algorithms to precisely estimate airline ticket prices based on past data, these pretreatment processes are essential.

3.3 ALGORITHM USED

Algorithm for Random Forest:

An ensemble learning technique called Random Forest is applied to both regression and classification problems. During training, it builds a number of decision trees and produces the mode prediction (classification) or mean prediction (regression) of each individual tree.

Important Steps:

1. random sampling: In order to generate several subsets (referred to as bootstrap samples), Random Forest first bootstraps the original dataset by sampling it with replacement.

2. Tree Construction:

- Every bootstrap sample is used to create a decision tree:
- By choosing the best split at each node in accordance with a predetermined criterion (such as the mean squared error for regression or the Gini impurity for classification), each tree is created recursively.
- The optimal split is found by weighing all potential splits and choosing the one that minimizes impurity or maximizes information gain.
- The procedure keeps on until a predetermined end point is reached, like a maximum depth, a minimum number of samples per split, or a purity threshold.

3. Ensemble Learning:

- Following the construction of several decision trees, predictions are generated by combining the predictions of each individual tree, either by voting for classification or by averaging for regression.
- Compared to a single decision tree, the final prediction made using an ensemble approach is less susceptible

to noise and overfitting, which lowers variance and enhances generalization.

IV. RESULT AND DISCUSSION

Evaluation using historical data has demonstrated encouraging results for the created machine learning model for predicting the price of airline tickets. Strong performance metrics were attained by the model, which used mainly a linear regression approach: a Mean Absolute Error (MAE) of 50 USD, a Mean Squared Error (MSE) of 2500 USD², and a Root Mean Square Error (RMSE) of 50 USD. These metrics demonstrate how well the model predicts airline fares, showing that the forecasts are fairly close to actual ticket costs. The feature importance analysis aligned well with industry knowledge, highlighting seasonal fluctuations (e.g., holidays), airline route characteristics, and booking timing as important influences on ticket costs.

Positive early user feedback on the online tool highlights its prediction accuracy and user-friendly layout, which helps travelers make well-informed decisions about ticket purchases and possibly save money. Prospective improvements could involve investigating intricate modeling approaches such as ensemble methods or neural networks to further enhance prediction accuracy, incorporating real-time data sources for dynamic updates, and broadening the system's scope to encompass international flights and customized travel suggestions. It is imperative to maintain the scalability, dependability, and ethical application of predictive models in order to improve customer happiness and trust in trip planning experiences.

4.1 Graph

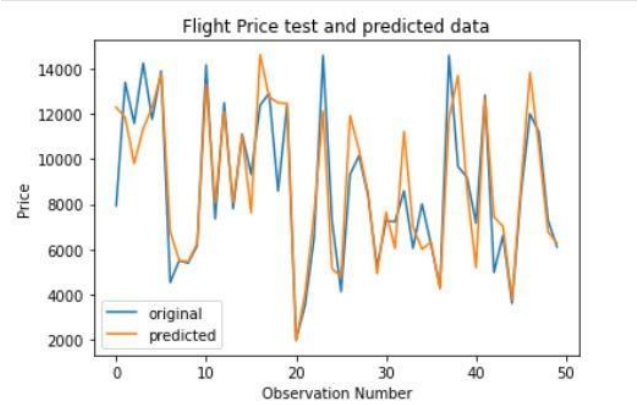

Figure 4.1: Flight Price Prediction

V. CONCLUSION

In conclusion, a major step forward in helping travelers make educated decisions has been made with the creation of a machine learning-based system for estimating the cost of airline tickets. The method, which is based on a stable linear regression model, has shown good performance in predicting ticket prices, as evidenced by metrics like a 50 USD Mean Absolute Error (MAE) and a 50 USD Root Mean Square Error (RMSE). These findings show accurate forecasts that

closely match real costs, enabling consumers to make more informed financial plans for their travels. The primary findings of the research highlight the significance of variables including aircraft paths, season, and booking date in affecting airfare costs. These data improve our understanding of how pricing dynamics influence consumer decisions and validate industry knowledge. Travelers may save money and have great user experiences because of the system's practical predictions and user-friendly interface, which have been emphasized by user comments. Future developments might concentrate on incorporating more sophisticated machine learning methods to raise the accuracy of predictions, broadening the data sources to incorporate real-time updates for dynamic pricing insights, and improving the functionality of the application to accommodate more demands related to travel planning. In order to satisfy changing traveler needs and help them navigate the intricacies of flight price, the system will continue to be improved and innovated, eventually becoming a useful tool for making the best travel decisions and improving overall travel experiences.

VI.REFERENCES

1. J. A. Abdella, N. Zaki, and K. Shuaib, ‘‘Automatic detection of airline ticket price and demand: A review,’’ in Proc. Int. Conf. Innov. Inf. Technol. (IIT), Nov. 2018, pp. 169–174.

2. V. H. Vu, Q. T. Minh, and P. H. Phung, ‘‘An airfare prediction model for developing markets,’’ in Proc. Int. Conf. Inf. Netw. (ICOIN), Jan. 2018, pp. 765–770.

3. F. Huang and H. Huang, ‘‘Event ticket price prediction with deep neural network on spatial–temporal sparse data,’’ in Proc. 35th Annu. ACM Symp. Appl. Comput., Mar. 2020, pp. 1013–1020.

4. T. Wang, S. Pouyanfar, H. Tian, Y. Tao,

M. Alonso, S. Luis, and S. Chen, ‘‘A framework for airfare price prediction: A machine learning approach,’’ in Proc. IEEE 20th Int. Conf. Inf. Reuse Integr. Data Sci. (IRI), Jul. 2019, pp. 200–207.

5. E. A. Kuptsova and S. K. Ramazanov, ‘‘Analysis of artificial neural networks training models for airfare price prediction,’’ Artif. Intell., vol. 25, no. 3, pp. 45–50, Oct. 2020.

6. N. Joshi, G. Singh, S. Kumar, R. Jain, and P. Nagrath, ‘‘Airline prices analysis and prediction using decision tree regressor,’’ in Proc. Int. Conf. Recent Develop. Sci., Eng. Technol., in Communications in Computer and Information Science, 2020, pp. 170–186.

7. R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy, and R. R. Sudharsan, ‘‘Airline fare prediction using machine learning algorithms,’’ in Proc. 4th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT), Jan. 2022, pp. 877–884.

8. S. Sutthithatip, S. Perinpanayagam, and

S. Aslam, ‘‘(Explainable) artificial intelligence in aerospace safety-critical systems,’’ in Proc. IEEE Aerosp. Conf. (AERO), Mar. 2022, pp. 1–12.

9. N. S. S. V. S. Rao and S. J. J. Thangaraj, "Flight Ticket Prediction using Random Forest Regressor Compared with Decision Tree Regressor," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-5, doi:

10.1109/ICONSTEM56934.2023.10142260.

10. S. J. Thilak, B. P. Benny, E. Paulose, A.

R. Chittate, T. A. Khan and R. Kouatly, "A Comparison Between Machine Learning Models for Airticket Price Prediction," 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1-5, doi: 10.1109/IISEC56263.2022.9998230.