An Overview of Machine Learning Approaches for Airline Pricing and Fare Detection and Visualize the Insight

Amoolya J[1], Dr Yashpal Singh[2]
[1]Research Scholar, Dept. of CSE, Jain University, [2]Professor, Dept. of CSE, Jain University

Abstract: In order to assist airlines set pricing that are both competitive and profitable, a method known as "airline fare detection" involves examining and recognising patterns in airline fares. It is now possible to examine vast volumes of data and acquire important insights into the factors that affect airline fares thanks to machine learning (ML) techniques and data visualisation tools like Tableau. Airlines may find trends and patterns in price data by applying ML algorithms, such as which routes are in high demand, which airports are busiest, and when people want to fly the most. By using this data, pricing tactics can be improved and clients can receive more affordable rates. Airlines may comprehend their price data better and decide on their pricing strategy more intelligently by integrating ML approaches with data visualisation tools like Tableau. By giving clients more reasonably priced and practical travel options, they can remain competitive in a market that is continuously evolving.
Keywords: 1.Airline price , 2.Machine learning , 3.Random forest , 4.SVM, 5.Neural Networks, 6.Linear Regression, 7.XGBoost, 8.Mean absolute error (MAE), 9.Root mean squared error (RMSE).

1. **Introduction:** Pricing strategies are essential in determining an airline's profitability in the fiercely competitive aviation sector. It is a difficult effort to identify accurate airline fares since it requires assessing a large quantity of data from numerous sources, such as consumer behaviour, market trends, and rival pricing. It is challenging for airlines to set the best pricing in real-time due to the accuracy and efficiency issues with traditional approaches to airline fare recognition. A promising strategy to enhance airline fare recognition has developed in recent years: the application of machine learning (ML) techniques. Large data sets can be analysed by ML algorithms, which can also spot patterns and insights that are difficult for people to notice. Airlines are better able to manage their income, optimise pricing tactics, and analyse customer behaviour by utilising ML approaches. Machine learning (ML)-based airline fare detection has the potential to be advantageous for both carriers and passengers. For customers, lower prices and a better overall travel experience might come from accurate pricing tactics. Conventional methods of determining airline fares can lead to rigid pricing structures that do not take into account specific client preferences or demand trends. In contrast, ML approaches can examine a large quantity of data on consumer behaviour, travel habits, and market trends to create more individualised and dynamic pricing plans. In order to detect specific client preferences and adjust pricing accordingly, ML algorithms can examine user data, including search history and purchase habits. As an illustration, visualising the insights obtained from ML algorithms might assist airlines in deciding when to provide special fares or discounts that are most likely to appeal to particular passengers. [1].The safety of its customers, employees, and equipment has long been a priority for the aviation sector. Implementing flight data monitoring (FDM) programmes is one strategy that has earned more and more attention in recent years. During flight operations, these programmes gather, examine, and interpret data from aircraft systems and sensors. The obtained insights can be utilised to spot possible safety hazards, boost operational effectiveness, and cut expenses. [3]. The airline sector has always been a dynamic and competitive market, and for airlines to continue to be competitive and profitable, accurate ticket price forecasting is essential. Machine learning (ML), which has recently gained popularity as a promising technology for forecasting ticket costs, enables airlines to change their pricing strategies in real-time and gain an advantage over rivals. [6]. Traditional approaches to figuring out flight costs are typically restrictive and fail to take into account factors like customer demand, market trends, and

seasonality. Yet, machine learning (ML) has become a possible approach for predicting prices, giving airlines the ability to alter their pricing strategy in real-time and remain competitive. This study can help airlines improve their pricing practises and provide their customers more reasonable tickets, increasing customer satisfaction and boosting revenue.

2. **Literature Review:** as airlines strive to price their tickets competitively and maximise their revenue, airline fare detection is a crucial issue for the aviation sector. Machine learning (ML) algorithms have become a potential tool for precisely forecasting airline fares in recent years. This review of the literature seeks to give readers an overview of the current research on machine learning (ML)-based airline fare detection that also benefits users.[1]. The purpose of this study by Fernandes (2002) is to analyse the possible advantages of FDM programmes for airlines. The paper covers the deployment of FDM and its advantages for airlines after reviewing the relevant literature. The effects of FDM on aviation maintenance, flying safety, and accident avoidance are investigated by the author. The report also looks into the financial advantages of FDM, including as cost reductions via enhanced maintenance procedures, less fuel use, and improved flight planning. Overall, this report provides insightful information about the potential benefits of FDM programmes for airlines. This study can assist airlines in making educated decisions on the implementation of such programmes by giving a thorough analysis of the advantages of FDM, thereby improving aviation efficiency and safety.[2]. The relevant articles that were published between 2015 and 2020 were found by the authors using a systematic literature review process. They discovered that for estimating demand and ticket prices, the majority of studies used machine learning methods such neural networks, support vector regression, and random forests. The datasets collected and the approaches employed had an impact on the models' level of accuracy. For instance, for the prediction of ticket prices, one study showed a mean absolute error (MAE) of 5.68%, whereas a different study reported an MAE of 11.5%. Similar to this, the mean absolute percentage error (MAPE) for demand forecasting ranged from 1.2% to 9.8%. The accuracy of the models was found to be highly influenced by the quality of the data utilised in their development by the authors as well. Pre-processing methods such feature selection, normalisation, and outlier elimination were commonly used to improve the accuracy of the models. [3]. Using pre-processing approaches including data cleansing and feature engineering, the authors used a dataset of airline ticket prices gathered from various sources. To forecast the ticket costs, they next applied three machine learning models: Support Vector Regression (SVR), Random Forest, and Artificial Neural Network (ANN). They achieved an accuracy of 81% using random forest, with mean absolute error (MAE) values of 1172.6134, mean squared error (MSE) values of 4044048.9764, root mean square error (RMSE) values of 2010.9820, and R squared value of 81.0258. To ascertain the most important elements influencing ticket prices, the authors also carried out feature importance analysis. The findings demonstrated that the key variables in determining ticket pricing were the airline, the time of departure, and the length of the flight. [4]. The authors pre-processed the data by removing outliers and missing values from a dataset of historical airfare costs from a well-known travel website. The Bagging Regression Tree model has the highest accuracy, at 87.42%. With an MAE of 18.86, MSE of 754.58, and RMSE of 27.46, the findings demonstrated that the Gradient Boosting Regression model performed the best. With an MAE of 22.75, MSE of 1009.04, and RMSE of 31.75, the Decision Trees model had the most accuracy, while the Linear Regression model had the lowest accuracy with an MAE of 32.89, MSE of 1741.92, and RMSE of 41.72. To ascertain the most important elements influencing ticket costs, the authors also carried out feature significance analysis. The findings indicated that the most crucial variables were the flight's airline, the day of departure, and the number of days prior to departure. [5]. In compared to other models, such as LR SVM and Neural Networks, the Random Forest Model performs the best on the data, hence it is employed for development. With a R squared score of 0.869, this prediction system performs with great accuracy. The authors' trials show that their

ensemble model performs better in terms of accuracy and robustness than individual machine learning models and baseline techniques. For consumers, airlines, and travel agencies to forecast airfare costs and make wise travel selections, the proposed framework can be helpful.[6]. Based on historical data and a number of input features, such as airline carrier, departure and arrival airports, and trip dates, the authors utilise a variety of machine learning methods, such as linear regression, decision tree, random forest, and gradient boosting regression, to estimate flight prices. The random forest algorithm, which achieved an accuracy of 85% in the experiments done by the authors, outperforms the other machine learning algorithms examined. Additionally, the authors compare their strategy to other comparable works and show that, in terms of accuracy and forecast time, their methodology beats them. There isn't a clear explanation of the feature selection procedure in the study, and it's not clear if the authors have taken into account any novel features beyond those that have been discussed in the literature frequently. Moreover, the study makes no attempt to evaluate or analyse the findings or any drawbacks of the suggested strategy.[7]. The authors provide a machine learning model that predicts the cost of the airfare using input features including the departure city, arrival city, departure date, arrival date, number of stops, and airline carrier. Several machine learning methods, including Gradient Boosting Regression, Decision Tree Regression, Linear Regression, and Random Forest Regression, were compared for performance. According to the authors, the algorithm with an accuracy of 89.2%, gradient boosting regression, performed the best. The future potential for employing more sophisticated machine learning methods, such as deep learning algorithms, to boost the precision of airfare prediction is also covered in the research.[8]. To anticipate the cost of the airfare, a machine learning model uses input features including the date of travel, the airports at the source and destination, the number of days left in the trip, and the airline company. The effectiveness of different machine learning techniques, including linear regression, decision tree regression, and random forest regression, is compared. According to the authors, the Random Forest Regression technique delivered the best results with an accuracy of 94.14%. The research also considers the potential application of more sophisticated machine learning methods, such as deep learning algorithms, to boost the precision of airfare prediction in the future.[9]. Examine a number of machine learning models that have been applied to forecasting the cost and demand of airline tickets in the past, such as regression, clustering, and deep learning models. They talk about the advantages and disadvantages of these models as well as possible future research avenues. Regression models performed very well, according to the authors, who claim that the majority of the research under consideration were highly accurate at predicting demand and airline ticket prices. Nevertheless, they also draw attention to the need for greater study into how to include outside variables like the economy and weather in these models for more precise forecasts.[10]. The authors suggest a machine learning algorithm-based intelligent agent that forecasts the best time to buy airline tickets by taking into account a number of variables, including past ticket prices, the time until departure, and the possibility that the price will rise or fall in the future. According to the authors, their agency was able to save money in comparison to more conventional purchasing strategies, with typical cost savings between 10 and 30 percent. Including real-time data updates and combining numerous data sources are only two examples of potential future research topics they recommend for enhancing the agent's accuracy and effectiveness.[11]. The authors outline their approach, which includes gathering historical flight data, preprocessing and feature engineering, and using different machine learning techniques like Random Forest, XGBoost, and Gradient Boosting Regressor to train and evaluate the predictive models. The Gradient Boosting Regressor, which had an R-squared score of 0.87, was the one that the authors claim had the greatest accuracy in their tests. The study discusses the difficulty in effectively forecasting flight costs, which arises from the volatility and unpredictability of the airline sector. To further increase the predictive models' level of accuracy, the authors recommend pursuing future research avenues such adding real-time data and sentiment analysis from social media.[12]. They employ several machine learning models to forecast the likelihood of booking for various routes using a dataset of over 5 million flight searches and bookings done on a global distribution system. In order to determine which features are crucial for predicting bookings, they also create a new feature selection algorithm. The findings indicate that machine learning models beat conventional regression models in their ability to predict with accuracy the rate at which certain itineraries would be booked. Airlines may benefit from using the authors' methodology to enhance pricing and revenue management tactics in markets that are highly competitive.[13]. This study uses machine learning techniques to forecast airfare costs for popular tourist destinations in Turkey. With an R-squared value of 0.973, Gradient Boosting Regression (GBR) outperformed other regression models in the authors' performance comparisons, including Support Vector Regression (SVR), Random Forest (RF), and others. The authors claimed that the suggested approach may be expanded to other areas and nations and integrated into online travel agencies to provide real-time airfare forecasts.[14]. The writers gathered data from numerous sources, including Google Flights and Expedia, and they also extracted attributes including airline, route, and time of booking information. After that, they evaluated the effectiveness of different machine learning models as XGBoost, random forest, and linear regression. In comparison to other models, XGBoost fared better, with a mean absolute percentage error (MAPE) of 3.64%. Future work, according to the authors, might concentrate on adding more elements and assessing how well the framework works with actual data. They also suggest that the framework might be improved to accommodate dynamic pricing models used by airlines and to forecast airfare costs for multi-leg itineraries.[15]. The primary goal of the project was to use machine learning algorithms to forecast gasoline tankering in the aviation sector. To forecast petroleum tankering, the authors employed different machine learning models, including Decision Trees, Random Forest, and XGBoost. The study's dataset included information from 232 flights. Using criteria like Mean Absolute Error (MAE) and Mean Squared Error, the models' accuracy was assessed (MSE). With an MAE of 0.33 and an MSE of 0.19, the findings demonstrated that the XGBoost model performed better than the other models. According to the study's findings, fuel tanker predictions made using machine learning algorithms can result in significant cost savings for the aviation sector.[16]. For estimating the gate in time of scheduled aircraft and ensuring schedule compliance, the authors of the research used a variety of machine learning models, including Random Forest, Decision Tree, XGBoost, and Gradient Boosting. R-squared (R2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are a few metrics that are used to assess the models' correctness. The findings demonstrate that the XGBoost model works better than other models, with an accuracy of 87% for forecasting the gate in time of planned flights and 94% for schedule conformity. The study's findings demonstrate that machine learning models can be utilised to forecast scheduled airline gate times and schedule compliance. The models can assist airports and airlines in increasing operational effectiveness, decreasing delays, enhance customer satisfaction.[17]. investigates the limits of machine learning methods for forecasting airline flight block times. The accuracy and dependability of machine learning models for forecasting flight block duration are thoroughly examined by the authors. For estimating the flight block time, the authors of the research used a variety of machine learning models, including Random Forest, Gradient Boosting, and Support Vector Regression (SVR). A dataset with flight-related data, including route, aircraft type, and weather conditions, is used to train the models. The outcomes demonstrate the limitations of machine learning techniques in accurately forecasting flight block time. Because of their poor accuracy and reliability, the authors speculate that these models might not be appropriate for real-time operational decision-making.[18]. give a thorough analysis of the machine learning algorithms that were used to forecast the distribution of strategic flight delays. The distribution of strategic flight delays was predicted using a variety of machine learning models, including Gradient Boosting

Decision Tree (GBDT), Random Forest, Artificial Neural Network (ANN), and Decision Tree (DT). The models' accuracy is assessed using a variety of metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The findings demonstrate that the GBDT model outperforms competing models with an accuracy of 0.13 in forecasting the distribution of strategic flight delays. According to the study's findings, it is possible to accurately estimate the distribution of tactical flight delays using machine learning algorithms. The models can aid airports and airlines in improving customer satisfaction by reducing delays and better managing their operations. In order to increase the models' accuracy, the study additionally emphasises the significance of feature selection and model optimization.

3. **Conclusion:** The review study on machine learning (ML) models for airline fare detection demonstrates the tremendous potential of ML in precisely forecasting airline fares. The study's numerous machine learning (ML) models, such as regression models, classification models, and neural networks, are discussed in the report along with how well they forecast airline prices. The performance metrics of the ML models, such as mean absolute error (MAE), root mean squared error (RMSE), and correctness, are used to illustrate the conclusions drawn from the review study. The visualisations show that while some ML models, like neural networks, may have limitations in their ability to reliably anticipate airline prices, others may be able to do so. Also, the visualisations demonstrate that the feature selection and data quality processes have an impact on the ML models' accuracy, highlighting the significance of these elements in the development of successful ML models.

Reference

1. *Vaz Fernandes, R. "An analysis of the potential benefits to airlines of flight data monitoring programmes." (2002).*
2. *Abdella, Juhar Ahmed, Nazar Zaki, and Khaled Shuaib. "Automatic detection of airline ticket price and demand: A review." In 2018 International Conference on Innovations in Information Technology (IIT), pp. 169-174. IEEE, 2018.*
3. *Mukane, Janhvi, Siddharth Pawar, Siddhi Pawar, and Gaurav Muley. "Aircraft Ticket Price prediction using Machine Learning."*
4. *Tziridis, Konstantinos, Th Kalampokas, George A. Papakostas, and Kostas I. Diamantaras. "Airfare prices prediction using machine learning techniques." In 2017 25th European Signal Processing Conference (EUSIPCO), pp. 1036-1039. IEEE, 2017.*
5. *Wang, Tianyi, Samira Pouyanfar, Haiman Tian, Yudong Tao, Miguel Alonso, Steven Luis, and Shu-Ching Chen. "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, 2019.*
6. *Arjun, K.P., Rawat, T., Singh, R. and Sreenarayanan, N.M., 2022, December. Flight Fare Prediction Using Machine Learning. In Computational Intelligence and Smart Communication: First International Conference, ICCISC 2022, Dehradun, India, June 10–11, 2022, Revised Selected Papers (pp. 89-99). Cham: Springer Nature Switzerland*
7. *Subramanian, R. Raja, Marisetty Sai Murali, B. Deepak, P. Deepak, Hamsinipally Nikhil Reddy, and R. Raja Sudharsan. "Airline fare prediction using machine learning algorithms." In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 877-884. IEEE, 2022.*
8. *Biswas, Prithviraj, Rohan Chakraborty, Tathagata Mallik, Sk Imran Uddin, Shreya Saha, Pallabi Das, and Sourish Mitra. "Flight price prediction: a case study." Int. J. Res. Appl. Sci. Eng. Technol.(IJRASET) 10, no. 6 (2022).*
9. *Abdella, Juhar Ahmed, Nazar Zaki, and Khaled Shuaib. "Automatic detection of airline ticket price and demand: A review." In 2018 International Conference on Innovations in Information Technology (IIT), pp. 169-174. IEEE, 2018.*

10. *Groves, William, and Maria Gini. "An agent for optimizing airline ticket purchasing." In Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems, pp. 1341-1342. 2013.*
11. *Arjun, K. P., Tushar Rawat, Rohan Singh, and N. M. Sreenarayanan. "Flight Fare Prediction Using Machine Learning." In Computational Intelligence and Smart Communication: First International Conference, ICCISC 2022, Dehradun, India, June 10–11, 2022, Revised Selected Papers, pp. 89-99. Cham: Springer Nature Switzerland, 2022.*
12. *Hopman, Daniel, Ger Koole, and Rob Van Der Mei. "A machine learning approach to itinerary-level booking prediction in competitive airline markets." International Journal of Revenue Management 12, no. 3-4 (2021): 153-191.*
13. *Can, Yekta Said, Koray Büyükoğuz, Efe Batur Giritli, Mustafa Şişik, and Fatih Alagöz. "Predicting Airfare Price Using Machine Learning Techniques: A Case Study for Turkish Touristic Cities." In 2022 30th Signal Processing and Communications Applications Conference (SIU), pp. 1-4. IEEE, 2022.*
14. *Wang, Tianyi, Samira Pouyanfar, Haiman Tian, Yudong Tao, Miguel Alonso, Steven Luis, and Shu-Ching Chen. "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, 2019.*
15. *Yılmaz, İlker Güven, Elif Kartal, Zeki Özen, and Sevinç Gülseçen. "Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms." Journal of Aeronautics and Space Technologies 14, no. 1 (2021): 19-34.*
16. *Sahadevan, Deepudev, Palanisamy P Dr, Varun P. Gopi Dr, Manjunath K. Nelli Mr, and K. Mr. "Prediction of gate in time of scheduled flights and schedule conformance using machine learning-based algorithms." International Journal of Aviation, Aeronautics, and Aerospace 7, no. 4 (2020): 9.*
17. *Abdelghany, Ahmed, Vitaly S. Guzhva, and Khaled Abdelghany. "The limitation of machine-learning based models in predicting airline flight block time." Journal of Air Transport Management 107 (2023): 102339.*
18. *Wang, Ziming, Chaohao Liao, Xu Hang, Lishuai Li, Daniel Delahaye, and Mark Hansen. "Distribution Prediction of Strategic Flight Delays via Machine Learning Methods." Sustainability 14, no. 22 (2022): 15180.*