iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84391

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

Airline Ticket Price and Demand Prediction

Sahithi Varma Gottimukkala[1], Pravitha R Prasad[2]
[1]*MCA Student,* [2]*Assistant Professor, Department of MCA, AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam, Andhra Pradesh- 531173*

Abstract: ***Air ticket price prediction has become a pivotal research area within data analytics and machine learning. With the growth of airline networks and online booking platforms, navigating dynamic pricing to secure affordable tickets remains a major challenge for passengers. This paper presents a structured framework for predicting flight ticket prices using machine learning regression algorithms and deep learning techniques. Using a benchmark Kaggle dataset, categorical features were preprocessed using Label Encoding, and various predictive models including Linear Regression, Decision Trees, and Random Forest ensemble methods were trained and evaluated. We analyze key internal and external factors driving fare volatility and assess the performance metrics of each algorithm. The optimal regression model was selected based on predictive accuracy, offering a robust and scalable approach to dynamic flight fare estimation.***
Keywords: Label Encoding, Linear Regression, Decision Trees, and Random Forest, Air ticket price prediction, Kaggle.

I. INTRODUCTION

Airfare price prediction using machine learning is a data-driven approach to forecast the prices of airline tickets based on historical data and various factors that influence ticket prices. With the increasing availability of airline data and advancements in machine learning algorithms, this approach has gained popularity in the travel industry. The goal of airfare price prediction is to assist travelers, airlines, and travel agencies in making informed decisions about flight bookings. By accurately predicting airfare prices, travelers can plan their trips more effectively, taking advantage of lower prices, while airlines and travel agencies can optimize pricing strategies and revenue management. Machine learning algorithms analyze historical flight data, such as ticket prices, flight routes, departure times, airline carriers, seasonal patterns, and other relevant factors, to identify patterns and relationships. These algorithms can then be used to build predictive models that forecast future airfare prices.

Airfare pricing and seat demand operate within a highly dynamic, non-linear ecosystem governed by dynamic pricing algorithms, market competition, seasonal trends, and consumer booking patterns. For travelers, this extreme variance introduces budget uncertainty and booking inefficiency; for airlines, balancing yield management with optimal capacity utilization remains a fundamental challenge. Modern computational approaches—specifically supervised machine learning and predictive analytics—offer a robust solution to decode these complex pricing and demand dynamics.

The Airline Ticket Price and Demand Prediction System provides a dual-perspective framework designed to forecast both airfare fluctuations and route-level booking demand. By processing historical flight datasets, the system extracts critical temporal, spatial, and carrier-specific features—such as departure/arrival timing, flight duration, number of stops, and seasonal demand metrics. Using machine learning algorithms, including Linear Regression, Decision Trees, Random Forest the project builds trained models capable of identifying pricing patterns and demand spikes. Deployed through a lightweight Flask web application, the platform delivers real-time fare estimations and predictive insights, empowering passengers to optimize their purchase timing while offering foundational models for data-driven revenue management.

II. LITERATURE REVIEW

The paper begins with some broad information regarding machine learning, after which the authors further proceed to the methodology. The methodology consists of four-phase process that influences flight prices, collection of data from flight fare prediction dataset by MH, selection and evaluation of an accurate ML Regression model. Key to its success is the integrity of the system where consumers accurately represent the segments for which prices have been differentially determined. Today, airlines price tickets “as much as the customer and market will bear,” are according to consultant and former airline planning executive. Airlines also profile their customers to help them adjust prices [1].

This often means placing passengers into one of two groups: leisure or business. And the way each group is priced is very different. Most studies on airfare price prediction have focused on either the national level or a specific market. Research at the market segment level, however, is still very limited. We define the term market segment as the market/airport pair between the flight origin and the destination [2].

The airline dataset included the following eight characteristics: departure and arrival times, type of airline, number of stopages, source, destination and additional_information. The authors performed prediction using regression Machine Learning models that including, LGBM Regressor, Random Forest Regression Tree, and Decision Tree Regressor. Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model [3].

All in present a review of deep learning and social media data-based Airline ticket price prediction model. The authors introduce the current airline ticket pricing situation with the factors that affect ticket prices. A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees. A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships [4]. Random Forest Model is used for development since it outperforms other models such as LGBM Decision Tree Regressor and Neural Network in terms of data performance with a R squared score of 0.91, this prediction framework has a good level of accuracy [5].

III. EXISTING SYSTEM

The existing system uses traditional machine learning algorithms such as Linear Regression and Decision Tree for airline ticket price and demand prediction. These methods predict prices based on historical data but provide limited accuracy for complex airline datasets.

Disadvantages:

- Low accuracy
- Less feature compatibility

IV. PROPOSED METHOD

The proposed system predicts airline ticket prices and passenger demand using machine learning algorithms. Historical flight data, travel details, and market trends are used to improve prediction accuracy. The system helps users and airlines make better decisions through accurate price and demand forecasting USING Random forest algorithm.

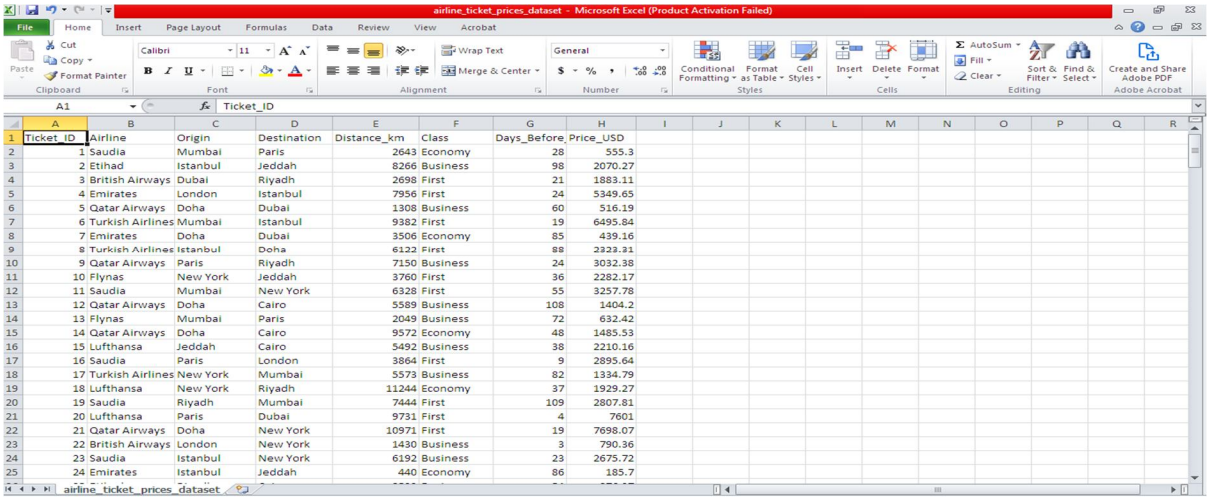
Advantages:

- High accuracy
- Better demand prediction
- High performance

V. METHODOLOGY

The proposed prediction framework follows a systematic machine learning pipeline designed to process raw flight data, train a tree-based ensemble model, and generate dual-currency price predictions. The process comprises the following distinct steps:

1) Data Acquisition: The system ingests a comprehensive air ticket price dataset (.csv format) containing historical travel details and ticket prices from kaggle.



Ticket_ID	Airline	Origin	Destination	Distance_km	Class	Days_Before	Price_USD
1	Saudia	Mumbai	Paris	2643	Economy	28	555.3
2	Etihad	Istanbul	Jeddah	8266	Business	98	2070.27
3	British Airways	Dubai	Riyadh	2698	First	21	1883.11
4	Emirates	London	Istanbul	7956	First	24	5349.65
5	Qatar Airways	Doha	Dubai	1308	Business	60	516.19
6	Turkish Airlines	Mumbai	Istanbul	9382	First	19	6495.84
7	Emirates	Doha	Dubai	3506	Economy	85	439.16
8	Turkish Airlines	Istanbul	Doha	6122	First	88	2323.31
9	Qatar Airways	Paris	Riyadh	7150	Business	24	3032.38
10	Flynas	New York	Jeddah	3760	First	36	2282.17
11	Saudia	Mumbai	New York	6328	First	55	3257.78
12	Qatar Airways	Doha	Cairo	5589	Business	108	1404.2
13	Flynas	Mumbai	Paris	2049	Business	72	632.42
14	Qatar Airways	Doha	Cairo	9572	Economy	48	1485.53
15	Lufthansa	Jeddah	Cairo	5492	Business	38	2210.16
16	Saudia	Paris	London	3864	First	9	2895.64
17	Turkish Airlines	New York	Mumbai	5573	Business	82	1334.79
18	Lufthansa	New York	Riyadh	11244	Economy	37	1929.27
19	Saudia	Riyadh	Mumbai	7444	First	109	2807.81
20	Lufthansa	Paris	Dubai	9731	First	4	7601
21	Qatar Airways	Doha	New York	10971	First	19	7698.07
22	British Airways	London	New York	1430	Business	3	790.36
23	Saudia	Istanbul	New York	6192	Business	23	2675.72
24	Emirates	Istanbul	Jeddah	440	Economy	86	185.7

Figure 1: Air ticket price dataset

2) Data Preprocessing: Missing values and anomalies are cleaned. High-cardinality categorical variables (e.g., Airline, Origin, Destination, Class) are transformed into numerical formats using Label Encoding.

3) Feature Selection: Key attributes influencing flight fares are extracted, including:
- Route Details: Origin, Destination, and Distance_km
- Travel Attributes: Airline and Class
- Timing Factor: Days Before Departure

4) Data Splitting: The preprocessed dataset is split into 85% training data for model fitting and 15% testing data for out-of-sample validation.

5) Model Training: The dataset is split into an 85:15 ratio (85% training set, 15% testing set) to ensure sufficient statistical representation during training while preventing target leakage during evaluation. The system utilizes a Random Forest Regressor, an ensemble machine learning algorithm that builds multiple decision trees during training and outputs the mean prediction (hat{y}) across all individual trees:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

Where B is the number of decision trees and Tb(x) represents the individual tree prediction for input feature vector x.

6) Model Evaluation: The trained Random Forest Regressor is benchmarked against unseen test data using standard regression metrics:
- Coefficient of Determination (R2): Quantifies the proportion of variance in flight prices explained by the features.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

- Mean Absolute Error (text {MAE}): Measures average absolute prediction error in actual currency terms.

\

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

7) Inference & Currency Conversion: The deployed pipeline accepts user-input flight parameters, passes them through the trained model, and outputs the predicted ticket price.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

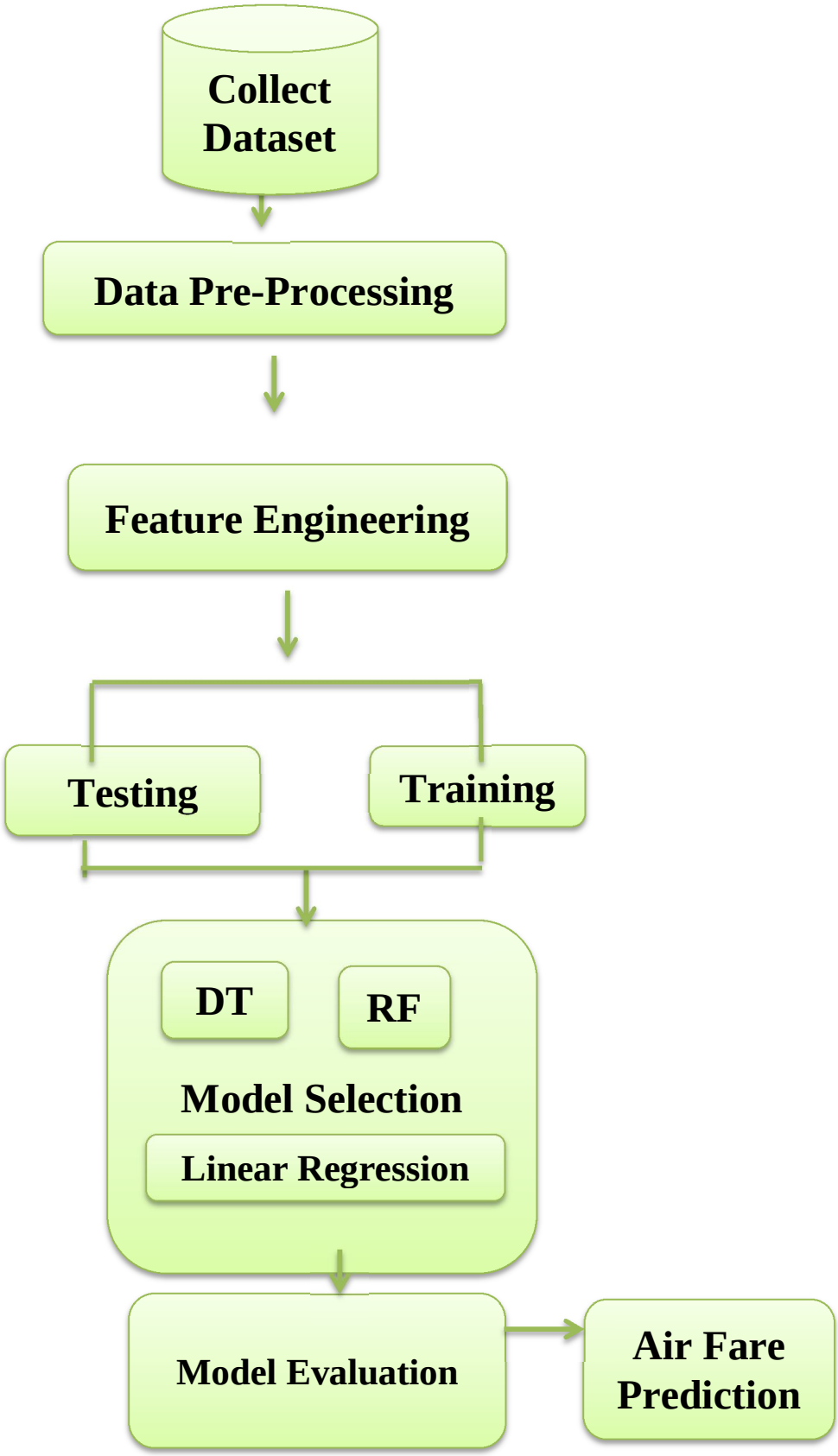


Figure 2: System Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

VI. MACHINE LEARNING ALGORITHMS

A. Linear Regression

A mathematical method for finding relationships between dependent and independent variables. Here's how to simulate target values using different predictors. Prediction of relationships between factors and causal analysis are important applications of this technique.

B. Random Forest

Supervised learning strategies include the well-known random forest machine learning algorithm. ML classification and regression can be used to solve the problem. Its theoretical foundation is the concept of ensemble learning, which combines different classifiers to solve difficult problems and improve model performance. In order to improve the predicted accuracy of the dataset, the Random Forest classifier, as its name suggests, "contains a number of decision trees on diverse subsets of the input dataset and takes the average." The random forest uses forecasts from all of the trees, rather than relying on just one, to predict the result based on which predictions earned the most votes.

C. Decision Tree

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute each branch represents the outcome of the test, and each leaf node represents a class label. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

VII. RESULTS

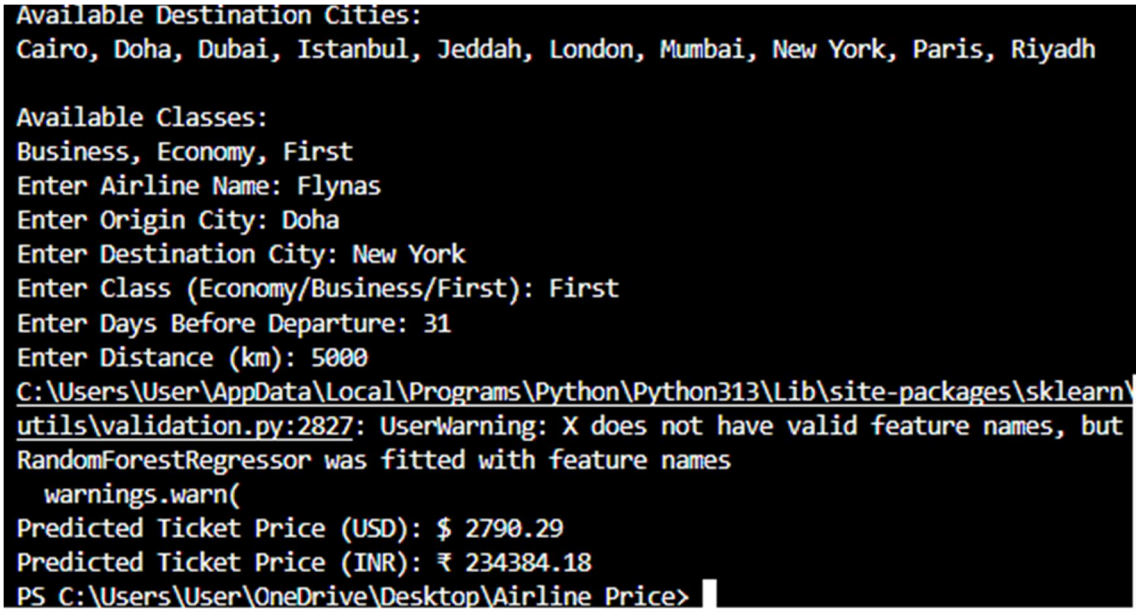


Figure 3. Airline fair Prediction

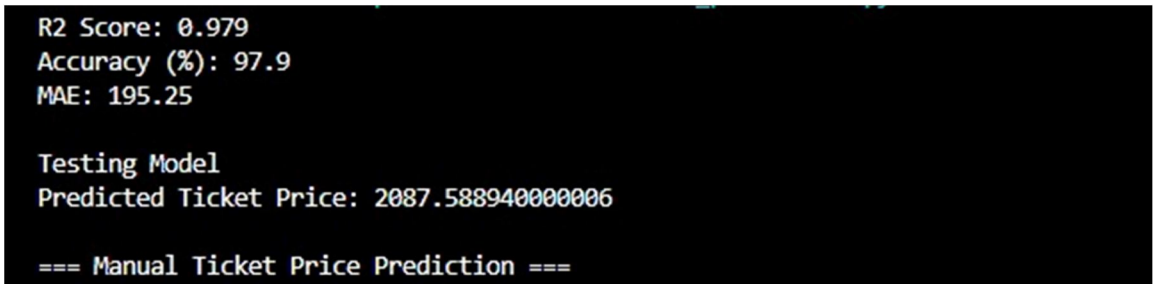


Figure 4. Random Forest Performance metrics

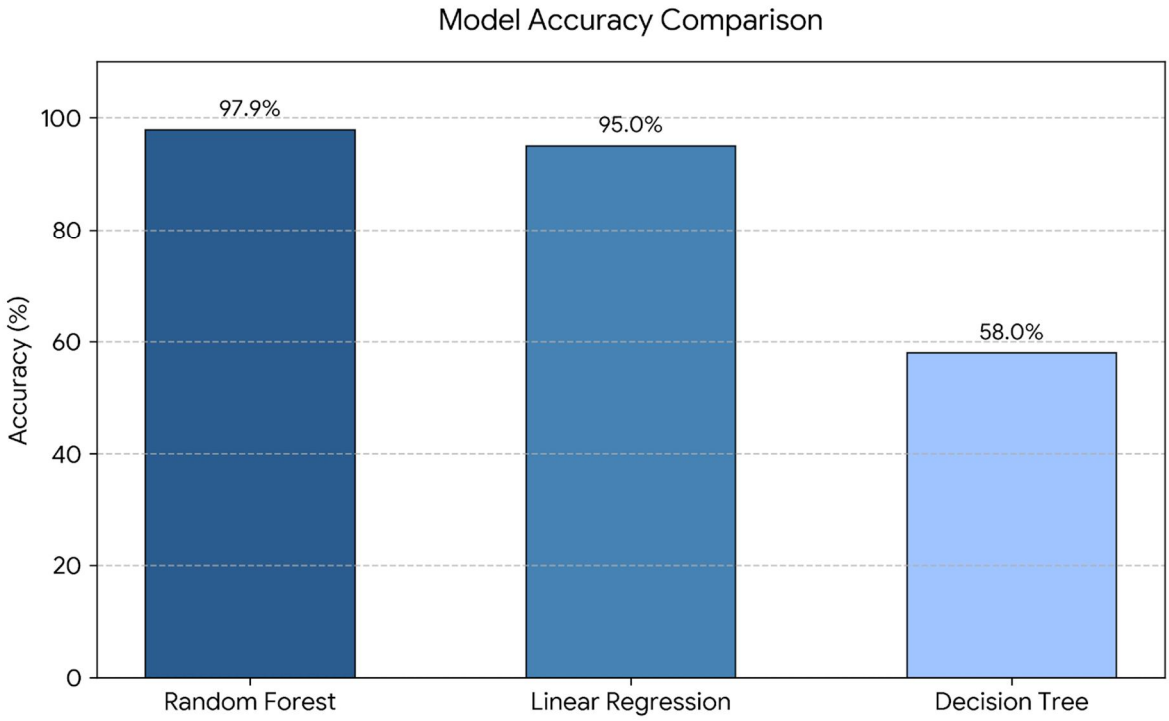


Figure 5. Accuracy Comparision

VIII. CONCLUSION

In conclusion, machine learning techniques can be effectively used for airfare price prediction. By analyzing historical data, features such as flight route, date, time, airline, and various other factors can be used to train machine learning models to predict future airfare prices. These models can provide valuable insights and assist both airlines and passengers in making informed decisions.

1) Accurate Predictions: Machine learning models can analyze large volumes of historical data and identify patterns and trends that humans might miss. This allows for more accurate predictions of airfare prices, helping both airlines and passengers plan their travel budgets effectively.

2) Pricing Strategy Optimization: Airlines can utilize machine learning models to optimize their pricing strategies. By considering various factors such as demand, seasonality, competition, and customer behavior, airlines can adjust their fares in real-time, maximizing their revenue and profitability.

3) Personalized Pricing: Machine learning can enable personalized pricing based on individual customer preferences and behaviors. This can help airlines tailor offers to specific customer segments, enhancing customer satisfaction and loyalty.

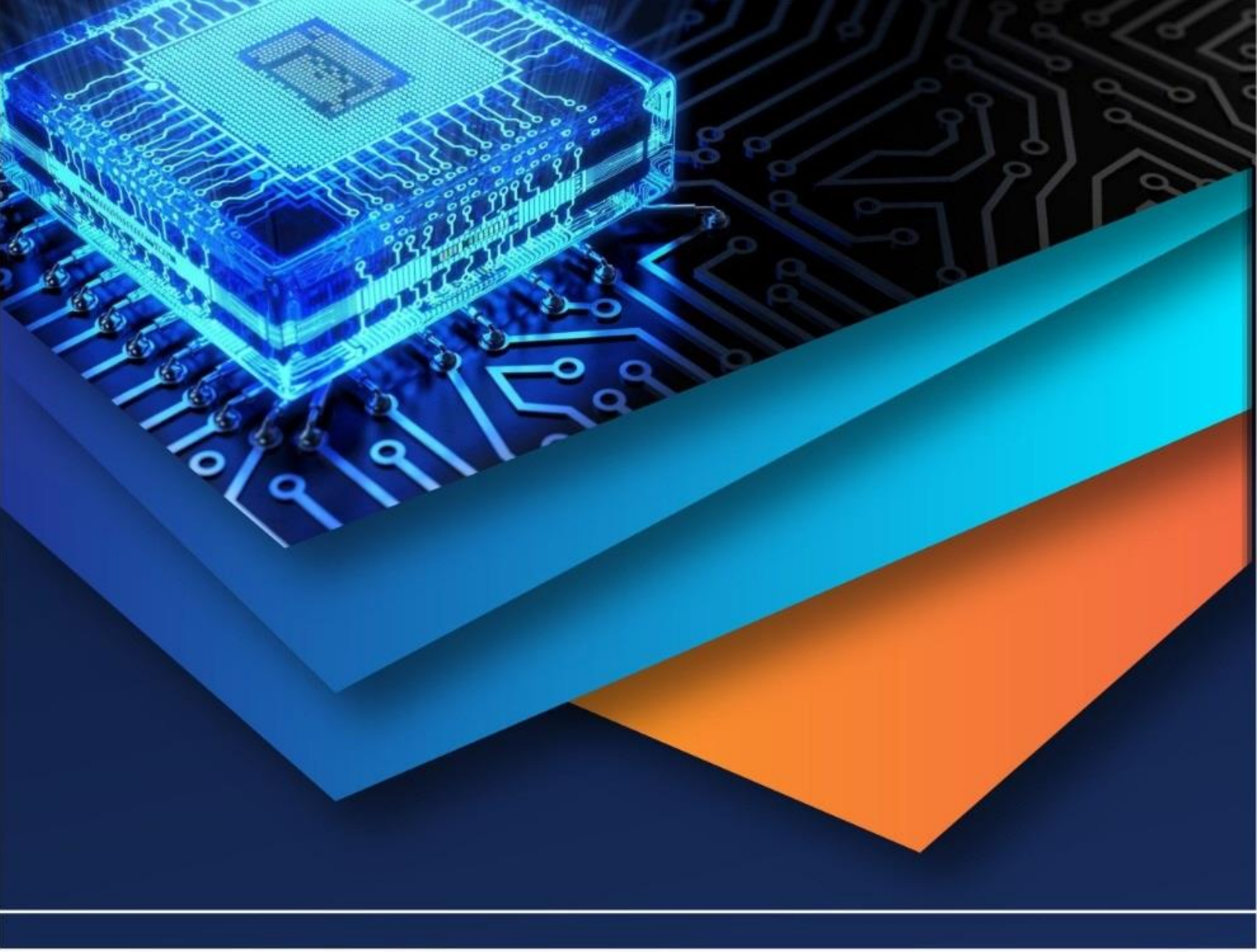
4) Demand Forecasting: Machine learning models can assist in predicting future demand for flights. By analyzing historical booking data and external factors like events, holidays, and economic indicators, airlines can optimize flight schedules and capacity planning.

5) Dynamic Pricing: Machine learning algorithms can support dynamic pricing, allowing fares to be adjusted in real-time based on supply and demand dynamics. This can benefit airlines and passengers, ensuring optimal pricing and maximizing seat occupancy.

REFERENCES

[1] P. Malighetti, S. Paleari and R. Redondi, "Pricing strategies of low-cost airlines: The Ryanair case study", Journal of Air Transport Management, vol. 15, no. 4, pp. 195-203, 2009.
[2] P. Malighetti, S. Paleari and R. Redondi, "Has Ryanair's pricing strategy changed over time? An empirical analysis of its 2006-2007 flights", Tourism Management, vol. 31, no. 1, pp. 36-44, 2010.
[3] W. Groves and M. Gini, "A regression model for predicting optimal purchase timing for airline tickets", Technical Report 11-025, 2011.
[4] W. Groves and M. Gini, "An agent for optimizing airline ticket purchasing", 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), pp. 1341-1342, May 06-10, 2013.
[5] M. Papadakis, Predicting Airfare Prices, 2014. T. Janssen, A linear quantile mixed regression model for prediction of airline ticket prices, 2014.

[6] R. Ren, Y. Yang and S. Yuan, Prediction of airline ticket price, 2015.
[7] Hüseyin Korkmaz (2024) Prediction of Airline Ticket Price Using Machine Learning Method Journal of Transportation and Logistics, Vol. 9, No. 2, pp. 205–218, 2024. DOI: 10.26650/JTL.2024.1486696. This paper compares multiple ML algorithms for airline fare prediction and identifies key factors affecting ticket prices.
[8] Yilong Wang, Shengrun Zhang, Xiaowei Tang, Chongheng Zhang (2025) Airline Ticket Price Prediction Based on CNN-LSTM Hybrid Model Journal of Beijing Jiaotong University, Vol. 48, Issue 5. Published January 2025. This study uses deep learning (CNN-LSTM) for airfare forecasting.
[9] CheniSruneethi& I. Udaykiran (2025) Flight Price Prediction Using Machine Learning International Journal of Scientific Research in Computer Science, Engineering and Information Technology, May 2025.
[10] SK Fhysuddin, Pallam Tejasri, Avula Swapna Sri, et al. (2025)Intelligent Fare Dynamics Modeling Using Machine Learning for Adaptive Airline Pricing ForecastsInternational Journal of Data Science and IoT Management System, Vol. 4, No. 2, 2025. The paper compares Linear Regression, Decision Tree, and Random Forest models for ticket price prediction.
[11] Muzaffer Ertürk, Murat Emeç, Ayşe Atılgan Sarıdoğan, Nabi Küçükgergerli (2025)Large-Scale Airline Ticket Price Prediction Using Ensemble Machine Learning AlgorithmsNişantaşıÜniversitesiSosyalBilimler Dergisi, June 2025. Uses a dataset of over 300,000 Indian flight records and ensemble ML techniques.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY