© July 2026 | IJIRT | Volume 13 Issue 2 | ISSN: 2349-6002

Flight Price Prediction Using Machine Learning

Javeriya Makandar
Department Of Master of Application, Secab Institute of Engineering & Technology, Visvesvaraya Technological University, Belgavi, Karnataka, India

***Abstract*—Airfare price prediction has become a central focus in the airline industry, as accurate forecasting directly impacts both travelers and airlines. This study conducts a comparative analysis of several machine learning algorithms—Decision Tree (DT), Random Forest (RF), Logistic Regression, Linear Regression, and XGBoost—applied to flight price prediction.**
The proposed system leverages a data-set containing features such as airline name, source, destination, departure and arrival times, duration, number of stops, and journey date. To enhance prediction accuracy, prepossessing techniques including data cleaning, feature extraction, encoding, and normalization are employed. Logistic Regression is also adapted to handle binary classification tasks relevant to tourist ticket outcomes. Model performance is evaluated using the R2 score, and feature importance analysis is conducted to identify key factors influencing airfare pricing. By comparing the strengths and limitations of these algorithms, this study provides insights into the most effective approaches for flight price prediction. The findings aim to assist stakeholders—from travelers to airline operators—in making more informed decisions.

***Index Terms*—Flight Delay Prediction Machine Learning Explainable Artificial Intelligence (XAI) Hazrat Shahjalal International Airport Bangladesh Aviation.**

I. INTRODUCTION

In the dynamic and highly competitive airline industry, pricing strategies are constantly shaped by factors such as fuel costs, demand fluctuations, and operational efficiency. Accurate flight price prediction has therefore become a crucial tool—empowering travelers to make informed, cost-effective decisions while enabling airlines to optimize their revenue management systems.

This study presents a comprehensive analysis of three machine learning algorithms—Decision Trees (DT), Random Forests (RF), and Logistic Regression—aimed at evaluating their effectiveness in flight price prediction. Decision Trees are first examined for their simplicity and interoperability, as they can transparently capture pricing patterns. However, their limitations in modeling complex relationships necessitate further exploration. Random Forests, an ensemble learning technique, are then investigated for their ability to enhance prediction accuracy by aggregating multiple decision trees, thereby addressing the challenges posed by intricate pricing dynamics. Logistic Regression, a widely used classification algorithm, is adapted to the unique context of airfare prediction due to its strength in handling binary outcomes, which align with certain ticket pricing decisions. To assess the performance of these approaches, real-world flight pricing data is utilized, with metrics such as the R2 score providing quantitative evaluation. Additionally, feature importance analysis is conducted to identify the key variables influencing airfare.

A. Objective of the Project

The primary objective of this research is to evaluate and compare the effectiveness of three machine learning techniques—Decision Trees (DT), Random Forests (RF), and Logistic Regression—in accurately predicting flight ticket prices. By analyzing real-world airfare data and assessing algorithm performance, the study aims to identify the most efficient method for forecasting ticket prices. The insights derived are intended to enhance airline pricing strategies and empower travelers to make informed, cost-effective decisions.

B. Problem Statement

Current flight ticket pricing systems often fail to adapt to the rapidly changing dynamics of the airline industry. Existing models lack the sophistication to account for multiple influential factors, leading to suboptimal pricing outcomes for both airlines and passengers. This creates a pressing need for more advanced and reliable machine learning models capable of accurately forecasting airfare, thereby supporting revenue optimization for airlines and affordability for travelers.

C. Motivation

Accurate flight fare prediction is vital in today's competitive airline sector.
This research investigates the performance of DT, RF, and Logistic Regression algorithms in forecasting ticket prices. By comparing their metrics and conducting feature importance analysis on real-world data, the study highlights each method's strengths and limitations. Ultimately, these findings aim to guide stakeholders—both travelers and airlines—toward informed decision-making and optimized pricing strategies.

D. Scope of the Project

This project focuses on developing and evaluating machine learning models—specifically Decision Trees, Random Forests, and Logistic Regression—for flight price prediction. Real-world airfare datasets are utilized to assess model performance in terms of accuracy and inter pretability. Additionally, the outcomes of this research are expected to support airlines in refining pricing strategies and assist travelers in making better-informed booking decisions.

II. RELATED WORK

Recent studies have explored hybrid machine learning approaches to improve airfare prediction. With the aviation sector generating vast amounts of data, traditional prediction methods often struggle to capture the multitude of variables influencing ticket prices. Hybrid models, which integrate techniques such as Decision Trees, regression-based methods, and ensemble approaches like Random Forest and Gradient Boosting, have shown promise in addressing nonlinear relationships and complex pricing patterns. By leveraging the strengths of multiple algorithms, these models aim to deliver more reliable forecasts, thereby supporting both travelers and airlines in making better decisions and enhancing revenue management strategies.

The study *"A Machine Learning-Based Approach for Airfare Prediction: A Comparative Analysis"* (2021) offers a detailed evaluation of various ML algorithms applied to airfare prediction. It compares models based on predictive accuracy, computational efficiency, and scalability, using real-world airfare data and performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The research emphasizes the importance of feature engineering, prepossessing, and feature selection in improving model performance.

2.1. Theofanis Kalampokas et al. (2023)

In *"A Holistic Approach on Airfare Price Prediction Using Machine Learning Techniques"*, the authors compared traditional machine learning, deep learning, and quantum machine learning approaches. Reported accuracies ranged from 89% to 99% across different airline datasets. The study emphasized the critical role of feature engineering and data prepossessing in boosting model performance.

2.2. Fardeen Shaikh et al. (2023)

This work developed a machine learning framework for flight fare prediction using features such as arrival time, departure time, and airline information. Multiple regression algorithms were trained and evaluated, demonstrating that machine learning can effectively predict minimum airfare prices for specific routes.

2.3 Philip Wong et al. (2023)

Leveraging Spark-based machine learning models, this study analyzed large-scale flight pricing datasets from travel platforms.
Algorithms such as Gradient Boosted Trees and Random Forest were implemented to improve scalability and real-world applicability. The research also examined feature importance and model generalization.

2.4 Jaspreet Singh et al.

In *"Flight Fare Prediction Using Random Forest Regression"*, the authors investigated airfare prediction using Random Forest and other ML techniques.
Results showed that Random Forest achieved superior accuracy due to its ability to capture nonlinear relationships between flight parameters and ticket prices.

2.5 Hüseyin Korkmaz (2024)
This study compared multiple machine learning algorithms for airline ticket price prediction, highlighting the influence of demand fluctuations, competition, and seasonal trends on airfare pricing. The findings concluded that advanced machine learning models provide more reliable predictions for dynamic pricing systems.

III. PROPOSED WORK

This research adopts a structured machine learning framework integrated with explainable AI methods. The primary goal is to construct highly accurate predictive models while ensuring that their outcomes remain transparent and actionable for stakeholders such as airport authorities and airline operators.
The process begins with the acquisition of reliable historical flight and weather data, followed by rigorous preprocessing to improve data quality and consistency. This stage involves addressing missing or erroneous values, standardizing temporal attributes, and selecting critical features relevant to delay prediction. Feature engineering incorporates variables such as scheduled and actual departure times, airline codes, and day-of-week indicators to capture operational and temporal dynamics.

A range of supervised learning algorithms is trained and compared, including Support Vector Machines (SVM), Random Forest, Decision Tree, K-Nearest Neighbor (KNN), XGBoost. Model performance is evaluated using established classification metrics—accuracy, precision, recall, and F1-score—to ensure comprehensive assessment.

To strengthen inter pretability, Local Interpret able Model-Agnostic Explanations (LIME) is employed. LIME identifies the most influential features driving individual predictions, thereby clarifying the reasoning behind model outputs. This integration enhances stakeholder trust and ensures that predictive insights are not only accurate but also practical for real-world airport operations.

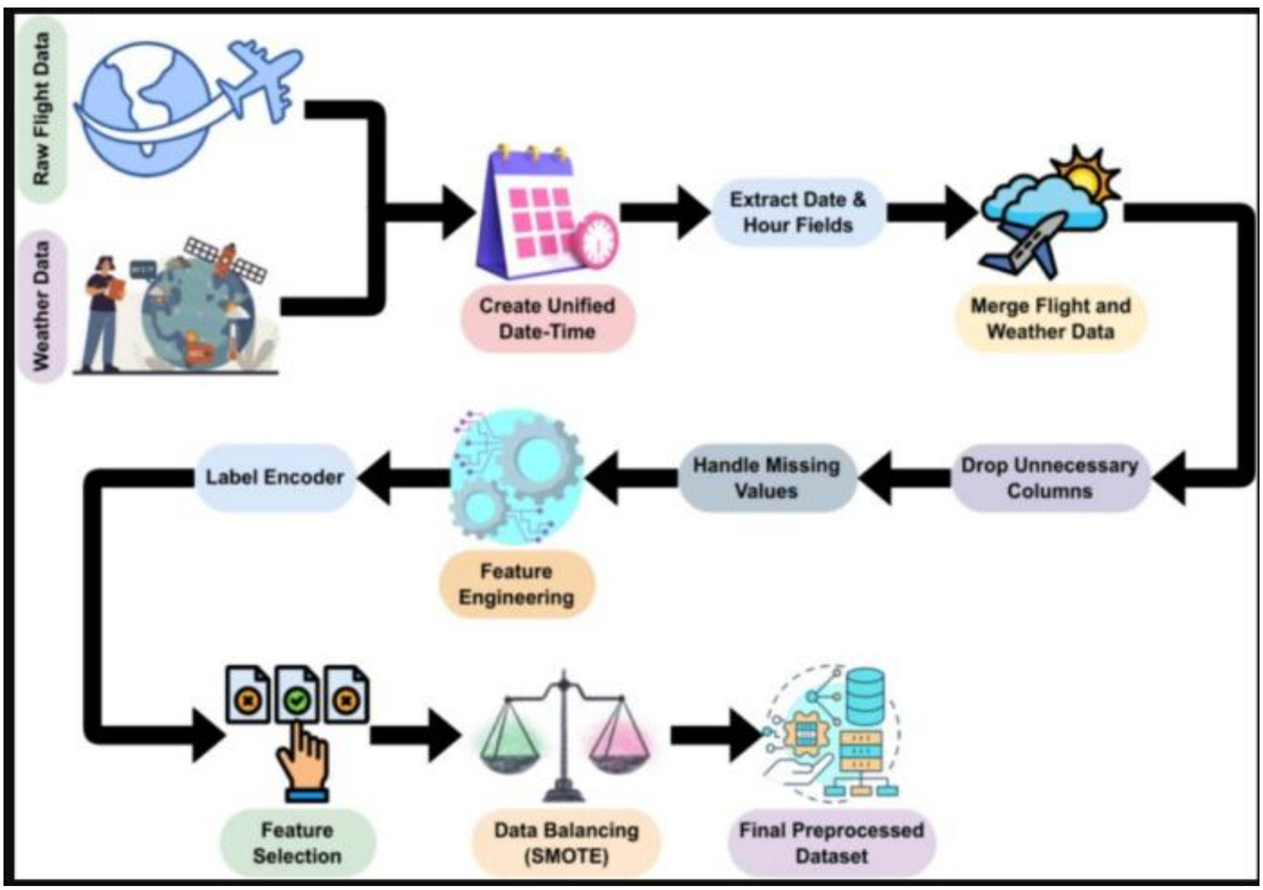


Fig. 1. Preprocessing workflow for flight delay prediction.

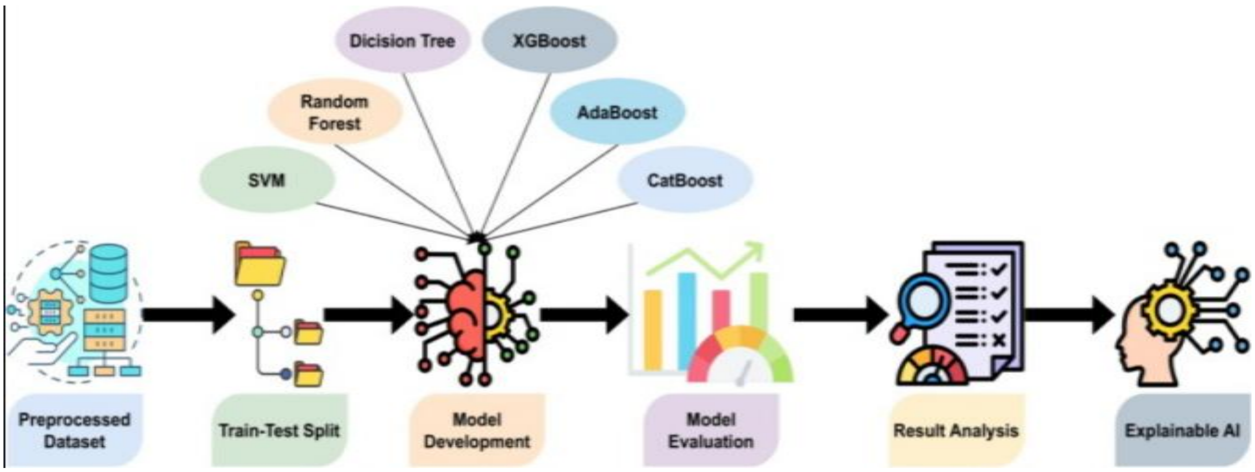


Fig. 2. Preprocessing workflow for flight delay prediction.

IV. METHODOLOGY

4.1 Data Collection
Urban flight pricing datasets were sourced from Kaggle and live flight aggregation services, incorporating variables such as airline, source and destination cities, flight duration, number of stops, departure and arrival times, travel class, and days to departure.

4.2 Data Preprocessing
Raw data was prepared for model training through key steps: handling missing or inconsistent values, encoding categorical variables (e.g., airline, source, destination) via One-Hot or Label Encoding, transforming date and time attributes into numerical features (e.g., extracting hours and days), and normalizing or scaling features to improve algorithm performance.

4.3 Decision Tree Regress-or:
A hierarchical structure that splits data based on optimal decision rules, offering transparency and inter pretability.

4.4 Random Forest Regress-or:
An ensemble method combining multiple decision trees to improve generalization and reduce over fitting.

4.5 Optional Models (Linear Regression/XGBoost):
Included for comparative analysis across algorithm types, providing additional insights into performance trade-offs. Evaluation Model performance is assessed using metrics such as the R2 score, alongside feature importance analysis. This comparative evaluation highlights the strengths and limitations of each algorithm in capturing pricing dynamics.

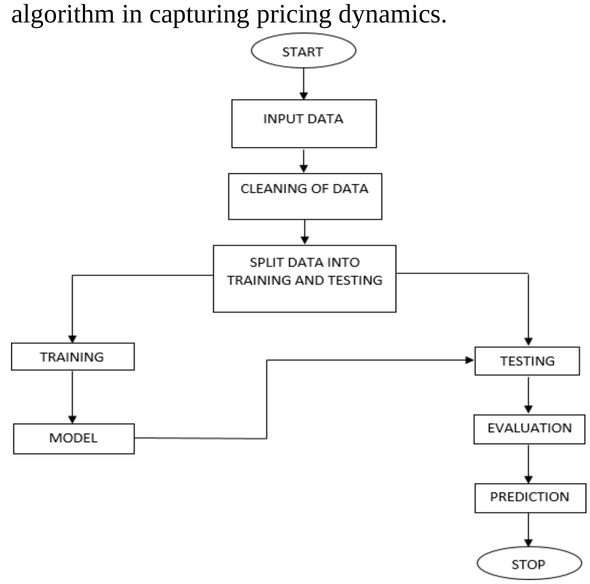


Fig3: Work flow of Methodology

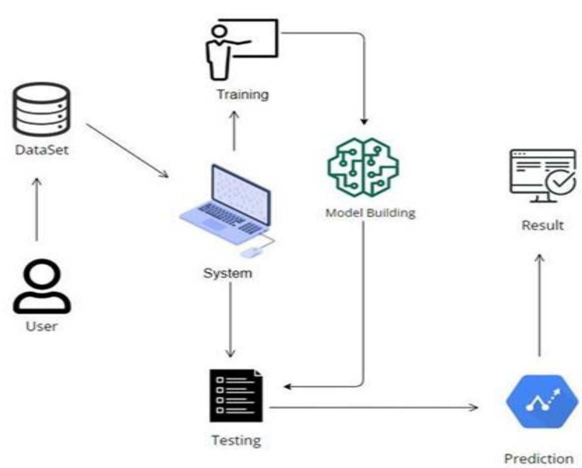


Fig4: Architecture of Project

V. IMPLEMENTATION

5.1 Decision Tree

The Decision Tree algorithm was employed to construct a reliable prediction model for airline ticket prices using features such as airline, source, destination, departure time, flight duration, and number of stops. Decision Trees are particularly effective for this task as they capture non-linear relationships and offer interpretability, enabling clear identification of attributes influencing pricing.

5.2 Random Forest

Random Forest was implemented to enhance prediction accuracy and robustness through ensemble learning. By aggregating multiple Decision Trees, Random Forest reduces overfitting and improves generalization, making it well-suited for modeling complex, non-linear interactions among variables such as travel date, airline type, route, and number of stops.

5.3 Logistic Regression

Logistic Regression was adapted to classify flight ticket prices into categories such as low, medium, and high.

Although traditionally used for classification tasks, it provides valuable insights into linear relationships between features and pricing outcomes.

5.4 Lasso Regression

Lasso Regression was implemented to build a robust and interpretable model with integrated feature selection. By applying L1 regularization, Lasso penalizes regression coefficients, driving less significant ones to zero and retaining only the most impactful predictors.

5.5 Multi-Layer Perceptron (MLP) Regression

MLP Regression, a type of feed forward neural network, was employed to capture complex, non-linear relationships between flight attributes and ticket prices. The model consists of input, hidden, and output layers, with back-propagation used to update weights during supervised training.

Optimizer such as Stochastic Gradient Descent and Adam were applied, while activation functions like ReLU and tanh introduced non-linear transformations.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

This section outlines the setup and evaluation of machine learning models designed to predict flight delays at Hazrat Shahjalal International Airport (HSIA). Model performance is examined using confusion matrices and explainable AI techniques to provide deeper insights into the decision-making process.

6.1 Experimental Setup

The experiments were conducted using CPU resources due to the unavailability of GPU support. Despite this limitation, efficient training was achieved through careful optimization of the learning process. Python libraries such as *Scikit-learn*, *XGBoost*, and were employed to ensure flexible implementation and reliable delay prediction.

Preprocessing steps—including feature scaling, data cleaning, and temporal standardization—were applied to prepare the data set for analysis. Hyper parameters were systematically tuned to maximize predictive accuracy while minimizing over fitting. This setup provides a robust foundation for evaluating model strengths and weaknesses across different performance metrics. Venue management in the aviation industry.

6.2 Implementation Details

This study employed a range of Python libraries and tools to manage data and construct predictive models. *Pandas* and *NumPy* were utilized for data manipulation and preprocessing, enabling efficient handling of missing values, feature transformations, and numerical computations. For model development and testing, *Scikit-learn* provided a versatile framework supporting algorithms such as Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). It also facilitated data preparation and hyperparameter optimization through *GridSearchCV*.

To incorporate advanced boosting techniques, *XGBoost* was applied for its speed and accuracy, while *CatBoost* was leveraged for its ability to effectively process categorical variables. Visualization of results and data distributions was achieved using *Matplotlib* and *Seaborn*, including the generation of confusion matrices and feature distribution plots. Although GPU resources were unavailable, the dataset size was manageable, and optimized CPU-based computations ensured smooth and efficient model execution.

6.3 Evaluation Criteria

Model performance was assessed using four widely adopted evaluation metrics: Accuracy, Precision, Recall, and F1-Score (Powers, 2011). These metrics, derived from the confusion matrix, provide detailed insights into prediction outcomes and error distribution.

In accordance with international aviation standards set by ICAO and IATA, a flight is classified as delayed if its actual departure time is 15 minutes or more later than the scheduled time. This definition aligns with global benchmarks, ensuring comparability with prior studies and industry practices.

A clear understanding of these metrics requires familiarity with the components of the confusion matrix, which illustrates true positives, false positives, true negatives, and false negatives.

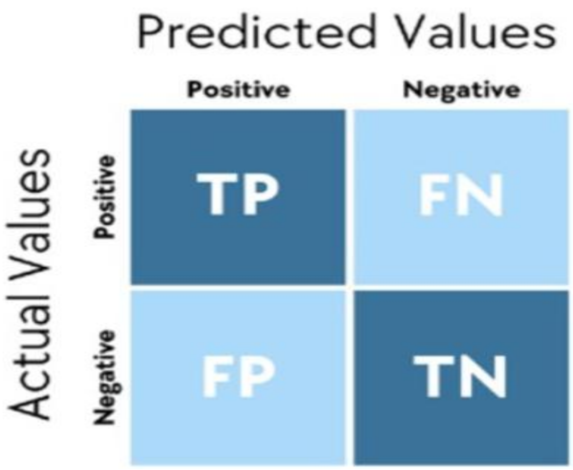


Fig. 5. Confusion matrix. Source: (Train in Data Blog., 2022).

A confusion matrix provides a structured tabular representation of a classification model's performance. It consists of four key components:

True Positives (TP):
Correctly predicted positive instances (e.g., predicted delay and actual delay).

True Negatives (TN):
Correctly predicted negative instances (e.g., predicted no delay and actual no delay).

False Positives (FP):
Incorrectly predicted positives (e.g., predicted delay but actually no delay).

False Negatives (FN):
Incorrectly predicted negatives (e.g., actual delay but predicted no delay). From these components, several evaluation metrics are derived:

Accuracy (Eq. 9):
Measures the proportion of correctly predicted instances (TP + TN) relative to the total number of predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision (Eq. 10):
Represents the proportion of true positives among all instances predicted as positive.
Precision =TP/TP+FP

Recall (Eq. 11):
Also known as sensitivity, it reflects the model's ability to correctly identify actual positive cases.
Recall=TP/TP+FN

F1-Score (Eq. 12):
The harmonic mean of precision and recall, providing a balanced measure of both.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

LIME explanation for flight delay prediction model CatBoost

LIME works by perturbing parts of the input data and observing how the model's output changes. Based on these variations, it constructs a simple local surrogate model that approximates the behavior of the complex classifier in the neighborhood of a specific instance. This process highlights the most influential features that contributed to a particular prediction, thereby making the decision-making process more transparent. As a model-agnostic technique, LIME can explain predictions from any trained model. By slightly modifying the input and tracking the corresponding output changes, it generates an interpretable local model—often linear—that is easier to connect back to the original model's reasoning for that instance (Ribeiro et al., 2016).

For example, Instance 42 in the dataset was classified as *on-time (Class 0)* with a confidence score of 1.00, indicating that the CatBoost model was highly certain about this prediction.

Table 1. LIME feature contributions for instance 42 (CatBoost model departure delay prediction).

Feature	Weight (Impact)	Interpretation
Month > 0.88	+0.13	Flights scheduled during these months were more likely to depart on time, likely due to favorable seasonal conditions and reduced congestion.
Flight Type (−1.43 < Flight-Type ≤ 0.71)	+0.08	Domestic or short-haul flights showed fewer delays, reflecting simpler operations and fewer turnaround dependencies compared to long-haul routes.
Day of the Week ≤ −0.99	−0.05	Certain days, possibly weekends or peak weekdays, were associated with slightly higher delays due to passenger volume and airport congestion.
Visibility ≤ 0.17	−0.04	Poor visibility increased the likelihood of delays, Consistent with safety Protocols that restrict takeoffs during Adverse weather.
Destination Airport (ICAO Code ≤ −0.75)	+0.04	Flights bound for less congested air ports tended to depart on time, suggesting smoother coordination and slot management.
Scheduled Minute ≤ −0.81	+0.03	Flights scheduled earlier within an hourly window were more likely to depart on time, indicating that minor operational delays accumulate as the day Progresses.

In the case of Instance 42, the CatBoost model classified the flight as *on-time (Class 0)* with a confidence score of 1.00, indicating complete certainty in its prediction. LIME analysis revealed that certain features contributed positively to this outcome, reinforcing the model's confidence. Conversely, features with negative weights—such as *Day of the Week* ≤ -0.99 and *Visibility* ≤ 0.17—slightly increased the likelihood of a delay. These minor influences may be attributed to operational patterns on specific travel days or reduced visibility due to weather conditions. However, their impact was insufficient to alter the overall prediction, and the model remained strongly confident that the flight would depart on time.

By applying LIME, this study adds a crucial interpretability layer to flight delay prediction. While previous research has largely focused on improving predictive accuracy, few studies have incorporated explanation tools to clarify *why* a particular decision was made.

The integration of LIME with CatBoost not only delivers robust predictions but also provides case-specific insights into the factors driving each outcome. This dual contribution enhances transparency, supports operational decision-making, and fosters greater trust among airline staff and planners in the reliability of machine learning models.

VII. DISCUSSION OF INSIGHTS

Our findings show that ensemble models such as XGBoost and CatBoost, combined with detailed weather and scheduling features, significantly enhance flight delay prediction at HSIA. With over half of daily departures delayed by more than 15 minutes and average delays exceeding two hours during monsoon months, this improvement is critical. LIME analysis identified meteorological factors (wind speed, temperature, visibility) and scheduling variables (airline, day, month) as key drivers of delays.

Practically, sharing real-time weather data with airlines and airport staff can enable proactive schedule adjustments and better resource allocation. The model also indicates that shorter domestic flights face fewer delays, suggesting opportunities for optimized slot assignments or staggered departures. By integrating explainable AI through LIME, the study not only strengthens predictive accuracy but also provides clear, case-specific insights, supporting data-driven operational decisions that improve efficiency and passenger satisfaction.

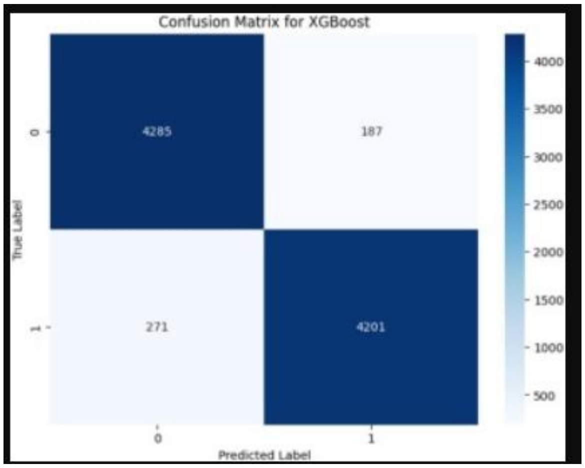


Fig. 6. Confusion matrix of the XGBoost model for flight delay prediction.

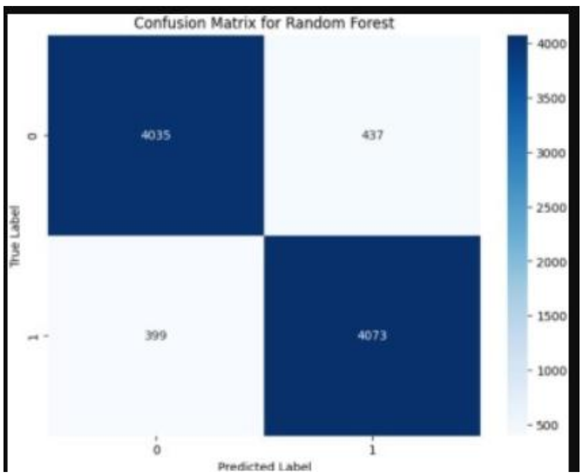


Fig. 7. Confusion matrix of the Random Forest model for flight delay prediction.

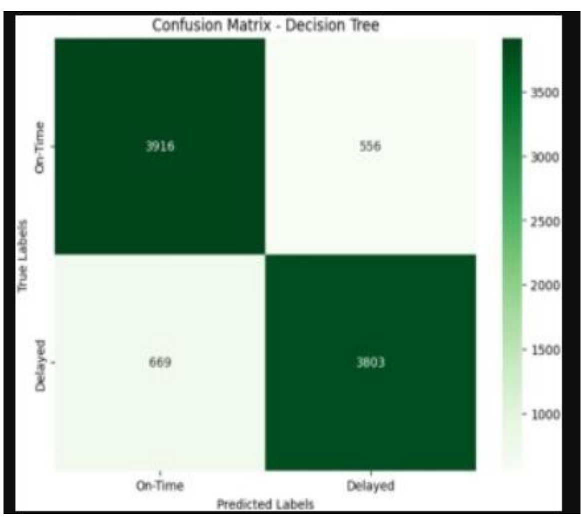


Fig. 8. Confusion matrix of the Decision Tree model for flight delay prediction.

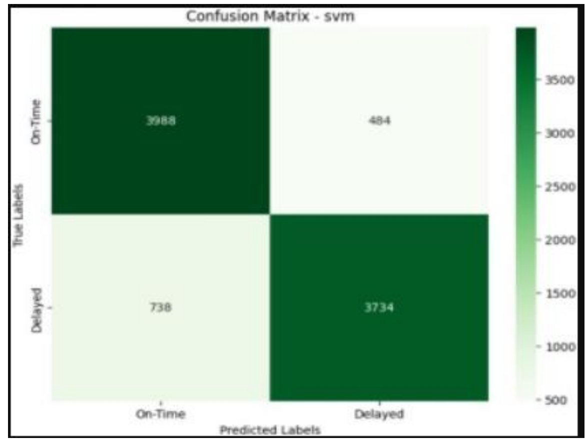


Fig. 9. Confusion matrix of the Support Vector Machine (SVM) model for flight delay prediction.

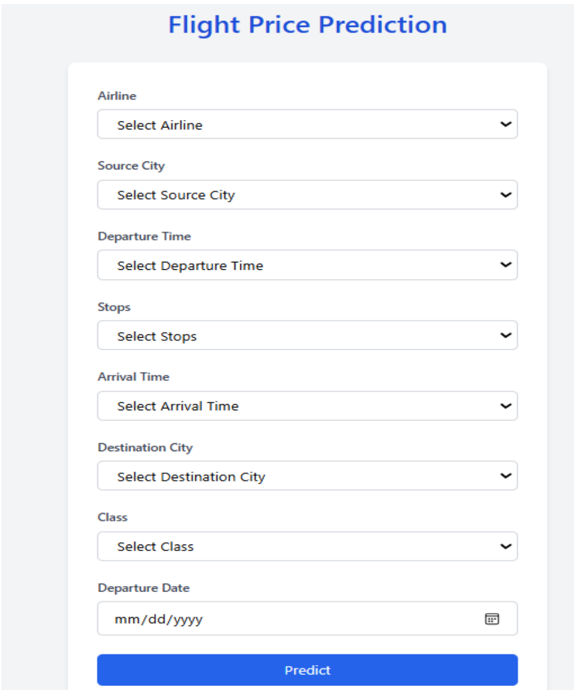
Finally, these findings have important policy implications. Airport authorities and policymakers could adopt explainable AI dashboards for operational monitoring, coordinate delay definitions with ICAO's 15-minute standard, and prioritize investments in real time weather monitoring and scheduling systems. By using predictive insights along with clarity, stakeholders can improve punctuality, enhance passenger experience, and make better operational decisions.

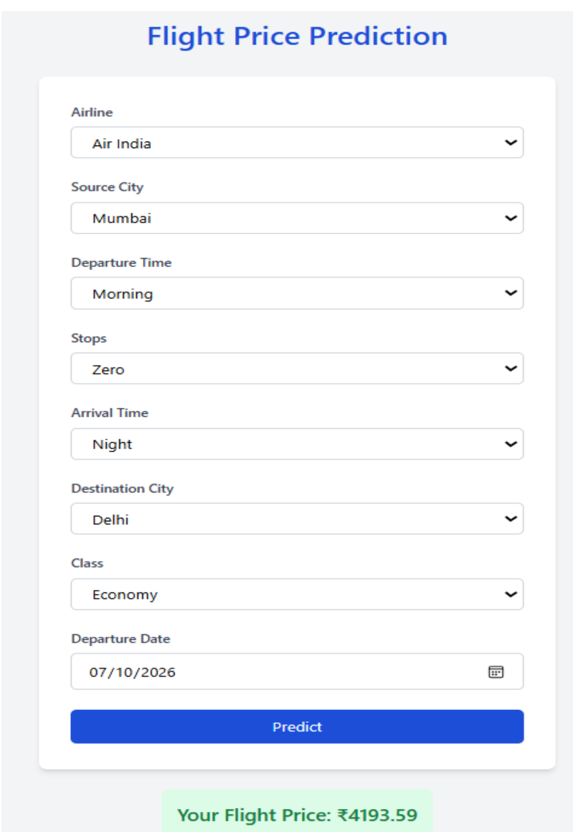
VIII. RESULTS

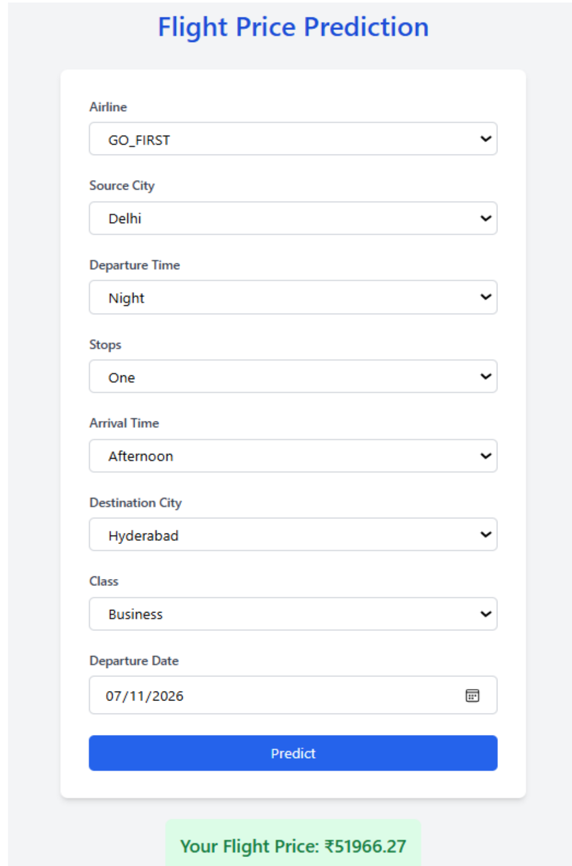
The flight ticket price prediction model was evaluated using key regression metrics, demonstrating strong predictive performance. The model achieved an R2 score of 0.8527, indicating that it was able to explain approximately 85.27% of the variance in ticket prices. This reflects a strong correlation between airfare and input features such as airline, source, destination, travel duration, and number of stops, with cash fare values also captured effectively. The Mean Absolute Error (MAE) was approximately ₹179.46, signifying that the predicted ticket prices deviated from actual prices by an average of ₹179. This level of error is considered low and acceptable for practical applications, given the dynamic nature of airfare fluctuations. Additionally, the Mean Squared Error (MSE) was recorded at 44,745.98, reinforcing the model's ability to minimize larger deviations, as MSE penalizes significant errors more heavily.

Overall, the model demonstrated robustness, accuracy, and generalization capability, making it suitable for real-world applications.

These results suggest that the system can be effectively utilized by travelers to identify optimal booking times and by airlines to refine dynamic pricing strategies, thereby enhancing decision-making and revenue management in the aviation industry.







IX. CONCLUSION

This study explored the application of three machine learning models—Decision Trees (DT), Random Forests (RF), and Logistic Regression (LR)—to forecast flight ticket prices in the dynamic airline industry. Each algorithm was critically examined for its strengths and limitations.

Decision Trees proved valuable for their simplicity and interpret-ability, offering transparent insights into pricing factors, though they struggled to capture complex relationships within the data set. Random Forests, as an ensemble method, demonstrated superior predictive capability by aggregating multiple trees, thereby improving accuracy and generalization. Logistic Regression, while traditionally used for classification, was adapted to categorize ticket prices, providing efficiency and interpret-ability in modeling binary or categorical outcomes.

Model performance was evaluated using metrics such as the R2 score, complemented by feature importance analysis, which highlighted the key determinants influencing airfare.

X. FUTUREENHANCEMENT

Future improvements in flight price prediction can further strengthen the reliability and practical relevance of machine learning models in real-world applications. One promising direction is the integration of real-time data feeds, enabling models to adapt quickly to dynamic market conditions influenced by factors such as weather, seasonal demand, or sudden operational changes.

Another enhancement involves leveraging natural language processing (NLP) techniques to analyze customer reviews and sentiment data. This can provide valuable insights into traveler perceptions and preferences, which in turn can inform pricing strategies from a customer- centric perspective.

The development of hybrid models that combine multiple machine learning approaches—such as deep learning integrated with ensemble methods—offers the potential to significantly improve prediction accuracy by capturing both linear and non-linear relationships more effectively.

Finally, fostering collaborative efforts among data scientists and promoting the use of anonymize shared datasets can enhance collective analysis.

REFERENCES

[1] S. Albert and V. Rundshagen, "European airlines' strategic responses to the COVID-19 pandemic (January–May, 2020)," *J. Air Transp. Manage.*, vol. 87, Art. no. 101863, 2020, doi: 10.1016/j.jairtraman.2020.101863.

[2] V. Bilotkach, A. A. Gaggero, and C. A. Piga, "Airline pricing under different market conditions: Evidence from European low-cost carriers," Tourism Manage., vol. 47, pp. 152–163, 2015, doi: 10.1016/j.tourman.2014.09.015.

[3] D. Gillen and W. Morrison, "Bundling, integration and the delivered price of air travel: Are low-cost carriers' full-service competitors?" J. Air Transp. Manage., vol. 9, no. 1, pp. 15–23, 2003, doi: 10.1016/S0969-6997(02)00071-6.

[4] A. Gordiievych and I. Shubin, "Forecasting of airfare prices using time series," in Proc. 2015 Information Technologies in Innovation Business Conf. (ITIB), 2015, pp. 68–71, doi: 10.1109/ITIB.2015.7355055.

[5] W. Groves and M. Gini, "An agent for optimizing airline ticket purchasing (Extended Abstract)," presented at the Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS). [Online]. Available: https://www.ifaamas.org

[6] W. Groves and M. Gini, "A regression model for predicting optimal purchase timing for airline tickets," 2011.

[7] T. Janssen, S. Abbas, and A. C. van Riel, "A linear quantile mixed regression model for prediction of airline ticket prices," 2014.

[8] P. Malighetti, S. Paleari, and R. Redondi, "Pricing strategies of low-cost airlines: The Ryanair case study," J. Air Transp. Manage., vol. 15, no. 4, pp. 195–203, 2009, doi: 10.1016/j.jairtraman.2008.09.017.

[9] P. Malighetti, S. Paleari, and R. Redondi, "Has Ryanair's pricing strategy changed over time? An empirical analysis of its 2006–2007 flights," Tourism Manage., vol. 31, no. 1, pp. 36–44, 2010, doi: 10.1016/j.tourman.2009.02.007.

[10] C. Morlotti, M. Cattaneo, P. Malighetti, and R. Redondi, "Multi-dimensional price elasticity for leisure and business destinations in the low-cost air transport market: Evidence from easyJet," Tourism Manage., vol. 61, pp. 23–34, 2017, doi: 10.1016/j.tourman.2017.01.009.

[11] M. A. Abdelrahman and A. M. Abdelrahman, "Flight delay prediction using machine learning algorithms," 2020.

[12] AirHelp, Global Flight Compensation Report, 2023. [Online]. Available: https://www.airhelp.com

[13] M. Alfarhood, R. Alotaibi, B. Abdulrahim, A. Einieh, M. Almousa, and A. Alkhanifer, "Predicting flight delays with machine learning: A case study from Saudi Arabian Airlines," 2024.

[14] S. M. Al-Tabbakh and H. El-Zahed, "Machine learning techniques for analysis of Egyptian flight delay," 2018.

[15] J. D. Anderson, A History of Aerodynamics and Its Impact on Flying Machines. Cambridge, U.K.: Cambridge University Press, 2001.

[16] R. E. Bilstein, Flight in America: From the Wrights to the Astronauts. Baltimore, MD, USA: Johns Hopkins University Press, 2001.

[17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and

Data Mining (KDD). New York, NY, USA: ACM, 2016, pp. 785–794.

[18] Civil Aviation Authority of Bangladesh (CAAB), Annual Report, 2022. [Online]. Available: https://caab.portal.gov.bd

[19] Visual Crossing, "Historical Weather Data for Dhaka, Bangladesh," 2022. [Online]. Available: https://www.visualcrossing.com/weather/weather-dataservices/Dhaka/us/2022-08-01/2022-08-31/