# Analysing Prospects for Expansion in Bangladesh's Domestic Aviation Industry using Machine Learning and Prediction Intervals

Nazmus Sakib[1], Mohammad Anwar Hosen[1], Nirjhar Gope[2], Burhan Khan[1], Bruce Gunn[1]
[1] Intelligent Systems Research and Innovation (IISRI), Deakin University, Waurn Ponds, Victoria, Australia
{s.sakib, anwar.hosen, burhan.khan, b.gunn}@deakin.edu.au
[2] Department of CSE, BRAC University, Bangladesh
nirjhargope@gmail.com

2025 IEEE International systems Conference (SysCon) | 979-8-3315-0818-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/SYSCON64521.2025.11014790

***Abstract*—Bangladesh, grappling with high population density and severe traffic congestion, has witnessed a surge in air travel demand due to inadequate transportation infrastructure. This trend has propelled the growth of the domestic aviation industry, with several private companies and a government-owned entity operating domestic flights. However, monopolistic tendencies and barriers to entry pose challenges for new market entrants. In response, this paper presents a comprehensive analysis of Bangladesh's domestic aviation sector, focusing on comprehensive analysis of growth opportunities. Over a four-month period, our study constructs a comprehensive dataset from diverse sources, including flight data, operational expenditures, passenger surveys, and interviews, to evaluate the scope and growth potential within the domestic aviation market. Leveraging machine learning algorithms and novel uncertainty quantification techniques like prediction intervals, we investigate the profitability of flights on different routes based on operational costs and ticket fares. Prediction intervals perform better than point predictions in this case, as they offer a range of likely outcomes, providing more valuable insights into potential profit margins, offering valuable guidance for new entrants seeking opportunities in the market. The dataset's trustworthiness is ensured through rigorous validation, with a trustworthiness score of 83.5%. Additionally, the dataset is available upon request to facilitate further research and collaboration.**

***Index Terms*—Prediction Intervals, Machine Learning, Aviation Industry.**

## I. INTRODUCTION

Bangladesh, with its rapidly rising population [1] and emerging metropolitan areas, presents significant transportation infrastructure challenges, including severe traffic congestion and mobility concerns [2]. In response, the aviation industry has emerged as a crucial alternative, offering faster and more efficient travel options. Despite its relatively small landmass, with high population density, emphasising the critical need for efficient transportation systems to support economic growth and societal well-being [3].

In recent years, Bangladesh's aviation sector has experienced significant growth, led by factors including growth in the economy and urbanisation, and a rising middle class with increased disposable income. However, this growth is not without its hurdles. Government regulations, competitive market dynamics, and currency fluctuations pose significant challenges. Additionally, new entrants encounter barriers due to the industry's heavy capital requirements [4] [5].

Despite these obstacles, private aviation companies in Bangladesh have shown resilience and strategic acumen, utilising innovative strategies to navigate challenges and solidify their market presence. This determination, coupled with the growing demand for air travel fueled by corporate culture and time-saving preferences, continues to drive the expansion of the domestic aviation industry [3] [4].

Furthermore, The economic landscape of Bangladesh offers both opportunities and challenges for the aviation sector. While government emphasis on infrastructure development and economic diversification supports industry growth, factors like rising fuel prices, and inflation present sustainability challenges [6] [7]. Escalating operational costs and regulatory complexities pose hurdles for new entrants amid established players. Despite these obstacles, innovative business models and operational efficiency strategies can enable newcomers to carve out niches and succeed in this dynamic market [8].

This scenario motivates us to investigate deeper into understanding the gaps within the aviation industry. Machine learning (ML) algorithms are extensively utilised for regression problems and forecasting predictions, including the widely accepted logistic regression (LR), support vector machine (SVM), random forest (RF), gradient boosting (GB), and multi-layer perceptron (MLP) [9].

We meticulously compiled a dataset spanning flight data, passenger demographics, ticket prices, operational costs, US dollar exchange rates, and profit margins over a four-month period, classified according to profit margin, enabling optimisation of classification performance through model accuracy, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) during training and testing. Our dataset focuses on a single model, the ATR 72-600 aircraft, widely used by all private operators in Bangladesh for domestic operations and short-haul international flights [10].

Additionally, employing prediction intervals offers a range of values based on actual and predicted values, aiding in assessing investment opportunities [11]. Given the substantial investment required in the aviation industry, often amounting to millions, leveraging these models and prediction intervals can help new entrants analyse their scope and identify valuable opportunities. Our analysis concentrates on three specific research points, aiming to provide actionable insights and contribute to the sustainable growth of Bangladesh's domestic aviation sector.

- Firstly, we have prepared a dataset covering flight information, passenger demographics, operational costs, currency exchange rates and profit margin.
- Secondly, the dataset is analysed by five machine learning algorithms to identify market growth.
- Lastly, Prediction reliability was enhanced using uncertainty quantification techniques like prediction intervals.

Section II reviews the relevant literature on the aviation industry, highlighting trends, challenges, and opportunities. Section III outlines the methodology for constructing the dataset, including data collection, variables, and quality assurance, while also detailing the machine learning algorithms with prediction interval used for regression analysis. In Section IV, analysis results and insights are presented and discussed. Finally, Section V comprises concluding ideas as well as recommendations for further research.

## II. Literature Review

The aviation industry grapples with challenges in predicting journey costs due to its intricate nature and constant fluctuations. Machine learning frameworks have the potential to evaluate data and produce more accurate price estimates and revenue margins for aircraft operations based on operational and variable expenses. Kiraci et al. investigated into the financial elements influencing the profitability of low-cost airlines, offering insights for strategic financial planning. Kerdpitak et al. identified factors shaping consumers' choices to fly on budget airlines, advancing understanding of customer preferences in airline services [12]. Singh et al. utilised various algorithms like linear regression, random forest, and decision trees to make precise predictions about flight prices, with the decision tree method proving most accurate [13]. Tziridis et al. applied eight ML models to predict airline fares, introducing a novel approach to fare prediction [14]. Rajankar et al. evaluated the efficacy of regression techniques in predicting airline fares, with gradient boosting demonstrating the highest accuracy [15].

By estimating airline operating profitability using machine learning techniques, Choi, O'Connor, and Truong provided insight into the financial dynamics of airlines [16]. Gini & Groves devised a predictive model to identify the optimal time to purchase airline tickets, utilising Partial Least Square Regression (PLSR) [17]. Wang et al. developed a system to forecast average ticket costs based on source-destination pairings using machine learning methods like support vector machine and XGBoost, showcasing high prediction accuracy. These studies underscore how ML can revolutionize flight price forecasting, benefiting customers, airlines, and stakeholders alike [18].

Dataset preparation assumes paramount importance for empirical research and analysis in the aviation industry. A comprehensive dataset incorporating various parameters such as flight data, passenger demographics, ticket prices, and operational costs proves indispensable for comprehending market dynamics [12]. The meticulous preparation and validation of datasets ensure the reliability and trustworthiness of research findings, facilitating informed decision-making and policy formulation [18].

The ATR 72-600 model aircraft holds significance in Bangladesh's domestic aviation sector, offering efficiency and reliability for short-haul flights [19]. The ATR-72 models are preferred by US Bangla Airways, Air Astra, Novo Air, and all other new entrants because to their low operational maintenance costs, minimal investment requirements, and improved decoration quality. Understanding the operational characteristics and performance metrics of the ATR-72 becomes imperative for optimising flight operations and ensuring passenger satisfaction. Notably, among US Bangla Airways' fleet of 18 aircraft, 10 are ATR-72s [1], while all 3 of Air Astra's aircraft are of the ATR-72 model[2] [20].

Machine learning models serve as invaluable tools for analysing and predicting outcomes in the aviation industry. Logistic Regression (LR) , Support Vector Machine (SVM), Multi-layer Perceptron (MLP) , Gradient Boosting (GB), and Random Forest regression (RF) models find common application for forecasting purposes [9]. Leveraging historical data, these models identify patterns and relationships, enabling researchers to make informed predictions about future growth and possibilities in the aviation sector [9] [15] [21].

Uncertainty quantification techniques, such as prediction intervals (PI), play a pivotal role in assessing the reliability and robustness of machine learning predictions. By quantifying uncertainty and providing Lower Upper Bounds Estimation (LUBE) on prediction outcomes, these techniques offer valuable insights into the accuracy and reliability of forecasting models, empowering stakeholders to make informed decisions with confidence [11] [22]. The PI framework includes Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Average Width (PINAW), and Coverage Width-based Criteria (CWC). *PICP*, calculated by Eq.1, gauges the probability of true values within prediction intervals. *PINAW*, from Eq.2, compares interval width to mean actual values. *CWC*, Eq.3, assesses model performance [11] [22] [23].

$$PICP = \frac{1}{n}\sum_{i=1}^{n} I(y_i \in [L_i, U_i]) \tag{1}$$

Where $n$ is the number of observations, $y_i$ is the actual value, $L_i$ is the lower bound of the prediction interval, $U_i$ is the upper bound of the prediction interval, and $I(\cdot)$ is the indicator function.

[1]https://www.airfleets.net/flottecie/US-Bangla.htm
[2]https://www.airfleets.net/flottecie/Air%20Astra.htm

$$PINAW = \frac{1}{n}\sum_{i=1}^{n}\frac{U_i - L_i}{\bar{y}_i} \quad (2)$$

Where $\bar{y}_i$ is the mean of the actual values within the prediction interval.

*CWC*, which takes into consideration both underlying dataset variability and predicted accuracy, is expressed by:

$$CWC = PINAW + PICPe^{-\eta(PICP-\phi)} \quad (3)$$

where $\eta$ is the control parameter that magnifies the error (negative) for a low coverage probability (if *PICP* $\leq \phi$). This *CWC*, defined in Eq.3, is used as the cost function in the LUBE method for this study.

New entrants in the aviation market face challenges due to the dominance of established players. However, strategic planning and innovation can help them establish their presence. By focusing on under-served routes, offering unique services, and leveraging technology, new entrants can differentiate themselves and attract customers. Partnerships with other businesses or airlines can provide additional resources and support, enhancing their competitiveness [24] [25]. This analysis, which includes forecasting feasibility using machine learning and understanding the significance of the ATR 72-600 model aircraft, aims to provide insights for stakeholders and policymakers in Bangladesh's aviation market. With continued research and investment, the aviation industry can drive economic growth and development.

## III. Methodology

In this section, we outline the methodology employed for corpus creation and data collection, crucial steps undertaken to compile a comprehensive dataset, detailing the machine learning algorithms with prediction interval for analysing the profit margins and losses of domestic airline operators in Bangladesh.

### A. Corpus Construction

The dataset creation process employing a comprehensive methodology that combined physical surveys with online data collection methods. Trained agents conducted physical surveys at airports, engaging passengers from specific flights to gather detailed information. Concurrently, online tools were utilised to scrape data such as original flight date-times, flight dynamics, and economic data including the US dollar exchange rate. Additionally, interviews were conducted to gather information on flight ground operation costs and other variable costs necessary for calculating each flight's operational expenses [26] [27].

### B. Data Collection Process

This study presents an authentic and dedicated effort in compiling an exclusive dataset that meticulously analyses the profit margins and losses of domestic airline operators in Bangladesh. With 1198 records of domestic air travel, the dataset provides valuable insights into regional aviation dynamics. Direct customer data obtained from physical surveys are summarized in Table I.

TABLE I
Physical Survey Data Features of Airline Dataset

| Sl | Query | Details |
|---|---|---|
| i | Airlines Name | Operators supplying dataset |
| ii | Airlines Code | Special identification for airlines |
| iii | Date of Journey | Track flight dates |
| iv | Special Day | Tracking Holidays and weekends |
| v | Source | Flight starting point airport |
| vi | Destination | Flight final location airport |
| vii | Route No | Path e.g. DAC-CXB ( Dhaka- Cox Bazar) |
| viii | Departure Time | Flight start time |
| ix | Arrival Time | Flight end time |
| x | Ticket Cost | Single ticket price ( Customer Price) |
| xi | Number of Seats | Aircraft capacity (ATR 72-600 Model) |
| xii | Occupied Seats | Passenger occupancy estimate percentanges |

The data was gathered through scraping from various sources including FlightAware[3] and FlightRadar24[4] websites for details such as original flight time, taxi time, and altitude. Moreover, information regarding US dollar rate of date of travel or ticket purchase date, and conversion rate was obtained from the Bangladesh Bank website[5]. Furthermore, data was collected from online sources like ATR-72 website[6] and Airlines Cost Calculator[7] and surveys conducted among ground staff and accounts department workers to estimate variable costs for flights in Bangladesh. It's important to note that these values are approximations based on projections and may not be entirely precise. Online data ensembles and physical surveys on estimated variable cost are summarized in Table II.

### C. Data Pre-processing and Analysis

The study analysed into flight details, aircraft specifications, fuel costs, and ancillary expenses to grasp the financial dynamics of domestic flight operations. Online cross-referencing verified seat availability and aircraft models, while fuel costs and currency exchange rates were scrutinised for accuracy. Additionally, ancillary expenses like landing and waiting costs were factored in, providing a comprehensive understanding of financial dynamics. Rigorous filtering ensured data quality, with over 2000 initial data points yielding 1198 high-quality entries for analysis attaining 83.5% trustworthiness score of corpus annotators [28].

The analysis of airline profitability involved meticulous pre-processing of the data to ensure quality and compatibility with machine learning models. Data collection predominantly focused on economy seats, with peaks during holidays and

[3]https://www.flightaware.com/
[4]https://www.flightradar24.com/
[5]https://www.bb.org.bd/en/index.php/econdata/exchangerate
[6]https://www.atr-aircraft.com/aircraft-services/aircraft-family/atr-72-600/
[7]https://www.aircraftcostcalculator.com/AircraftOperatingCosts/

TABLE II
ONLINE DATA FEATURE OF THE AIRLINE DATASET

| Sl | Query | Details |
|---|---|---|
| i | Flight Time | Total time spent in the air |
| ii | Taxi Time | Approx. taxiing from the gate to the runway before takeoff and from the runway to the gate after landing |
| iii | USD Rate | Recording the USD exchange rate on flight date |
| iv | Altitude | Approx. cruising altitude typically maintained by the aircraft |
| v | Fuel Burn | Approx. fuel consumption efficiency of the aircraft |
| vi | Fuel Cost | Approx. calculated in US dollars based on prevailing fuel prices and consumption rates |
| vii | Operation Cost | Approx. operational costs incurred during the flight, including expenses such as maintenance crew salaries, and other operational expenditures |
| viii | Ground Cost | Approx. ground service expense estimate associated with the flight, including expenses for ground handling, landing fees, and other related services at airports |
| ix | Operation Margin | Approx. difference between total revenue generated from the flight and the total operational costs incurred. |

weekends. A crucial pre-processing step involved transforming categorical variables into numeric forms, achieved through encoding with an Ordinal Encoder for columns such as ‘flytime’, and ‘price’. Subsequently, a final dataset was curated, comprising essential features like ‘TicketpriceUSD’, ‘SeatOccupied’, and ‘margin’ for predicting airline profitability.

This preparatory step lays the groundwork for developing machine learning models to predict airline profitability margins. The data is meticulously cleaned, relevant features are selected, and categorical variables are properly encoded, ensuring its effectiveness for predictive modeling.

### D. Model Selection

The study’s focus is thus limited to investigating the feasibility of launching a new airline in Bangladesh using a comprehensive ensemble of machine learning models—each chosen for its special ability in capturing the nuances of different aviation industry strengths or shortcomings. This ensemble comprised Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Multi-layer Perceptron (MLP).

In selecting models for analysis, our approach prioritises maximising benefits across various facets crucial to understanding airline operations. Logistic Regression predicts categorical outcomes, offering initial insights into profitability determinants. Support Vector Machine (SVM) analysis explores complex dataset patterns, revealing factors such as the impact of fluctuating currency rates on operational costs and profitability in dynamic economic environments. We employ Gradient Boosting to capture complex data interplay and non-linear dynamics, providing valuable insights into operational efficiency and market demand. Random Forest Regression enhances model robustness, while Multilayer Perceptron (MLP) fits non-linear relationships for comprehensive airline operation insights. This multifaceted model selection strategy ensures a thorough understanding of airline profitability dynamics [9] [22].

In our model selection process, we prioritise maximising benefits across various aspects crucial to understanding airline operations. We utilise Logistic Regression for categorical outcome prediction, Support Vector Machine (SVM) analysis for exploring complex dataset patterns, Gradient Boosting for capturing data interplay, Random Forest Regression for model robustness, and Multilayer Perceptron (MLP) for fitting non-linear relationships. This multifaceted approach ensures comprehensive insights into airline profitability dynamics [11] [24] [25].

In the context of determining the financial investment of a new airline venture in Bangladesh, combining classical machine learning approaches with Prediction Interval (PI) performance provides an analytical framework. Traditional machine learning methods produce point forecasts, providing insight into possible profitability. Meanwhile, PI performance improves this analysis by providing ranges and associated confidence levels, increasing the depth and accuracy of insights for stakeholders to make financial decisions.

### E. Evaluation Metrics and Hyper-Parameter Tuning

The critical evaluation metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Precision, Recall, Area Under the Curve (AUC), and F1 Score, serve as pivotal indicators in assessing model performance. Precision, recall, and F1 scores gauge the model’s proficiency in correctly identifying profitable margins on flight routes. The model’s accuracy is determined by the coefficient of determination ($R^2$), which assesses the proportion of variance in the dependent variable predictable from the independent variables. It computes margin values based on fitted parameters, contrasting predicted outcomes with actual results to gauge accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y^i_{\text{true}} - y^i_{\text{pred}})^2}{\sum_{i=1}^{n} (y^i_{\text{true}} - \bar{y}_{\text{true}})^2} \tag{4}$$

The Eq.4 for $R^2$, the coefficient of determination (an accuracy score), compares actual ($y_{\text{true}}$) and predicted ($y_{\text{pred}}$) values to assess the proportion of variance explained by independent variables, with $\bar{y}_{\text{true}}$ denoting the mean of actual values and $n$ representing the number of data points.

Model accuracy measures how well predictions match actual observations. In airline profitability prediction, accuracy reflects the model’s ability to classify flight routes as profitable or not. Higher accuracy ensures more reliable predictions, vital for informed decision-making in airline operations. Moreover, leveraging cross-validation in hyper-parameter tuning enhances model optimisation, augmenting precision and robustness in airline profitability prediction.

## IV. Results and Discussion

This section presents an in-depth analysis of various regression models utilised to assess the feasibility of a new airline venture in Bangladesh. Each model’s predictive performance

TABLE III
ACCURACY ($R^2$), MAE, MAPE AND RMSE ON TRAIN DATA

| Models | Train Data | | | |
|---|---|---|---|---|
| | $R^2$ | MAE | MAPE | RMSE |
| Logistic Regression | 0.9999 | 6.49 | 6.78 | 35.16 |
| Multilayer Perceptron | 0.9915 | 34.89 | 37.05 | 1.0773 |
| Support Vector Machine | 0.9871 | 430.37 | 221.83 | 1787.20 |
| Gradian Boosting | 0.9995 | 11.82 | 14.98 | 15.002 |
| Random Forest | 0.9801 | 72.33 | 64.17 | 94.26 |

TABLE IV
ACCURACY ($R^2$), MAE, MAPE AND RMSE ON TEST DATA

| Models | Test Data | | | |
|---|---|---|---|---|
| | $R^2$ | MAE | MAPE | RMSE |
| Logistic Regression | 0.9998 | 184.19 | 440.57 | 642.631 |
| Multilayer Perceptron | 0.9921 | 630.12 | 421.58 | 25.0456 |
| Support Vector Machine | 0.9842 | 415.33 | 132.34 | 652.969 |
| Gradian Boosting | 0.9982 | 319.22 | 377.34 | 56.321 |
| Random Forest | 0.9753 | 80.61 | 39.38 | 105.108 |

offering essential insights for stakeholders involved in the aviation industry.

## A. Predictive Performance of Machine Learning Models

We conducted simulations on five machine learning models, splitting the data into 80%-20% ratio for training and testing, with 10% of the training data allocated for validation. Table III and Table IV depict a presents overview of the regression models' effectiveness, focusing on prediction accuracy metrics for both training and test datasets, particularly MAE, MAPE and RMSE.

The analysis of flight profitability prediction models based on routes and date-time reveals varying levels of effectiveness and generalization across algorithms. While Logistic Regression, Gradient Boosting, and Random Forest models demonstrate high accuracy on both train and test datasets, indicating robust performance, multi-layer perceptron stands out for its better generalization and lower error metrics, suggesting its potential reliability in predicting flight revenue. Conversely, SVM displays lower accuracy and significantly higher error metrics, indicating challenges in accurately predicting flight revenue and potential over-fitting from the training phase. Overall, while Logistic Regression, Gradient Boosting, and Random Forest models show high accuracy, Multi-layer Perceptron emerges as a promising choice for its superior generalization and lower error rates, warranting further consideration in flight profitability prediction tasks.

Among the evaluated models, logistic regression, despite its strong performance in training but poor in testing, is considered less suitable due to its binary nature. multi-layer perceptron (MLP) and gradient boosting emerge as the preferred choices for flight margin prediction, owing to their high accuracy, low RMSE on unseen data, and ability to handle linear and non-linear relationships effectively, thus enabling precise fare forecasting and revenue maximization in the competitive airline market.

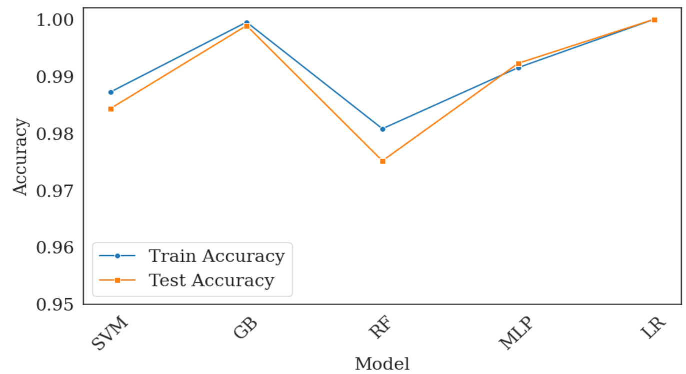


Fig. 1. Machine Learning Models Accuracy ($R^2$)

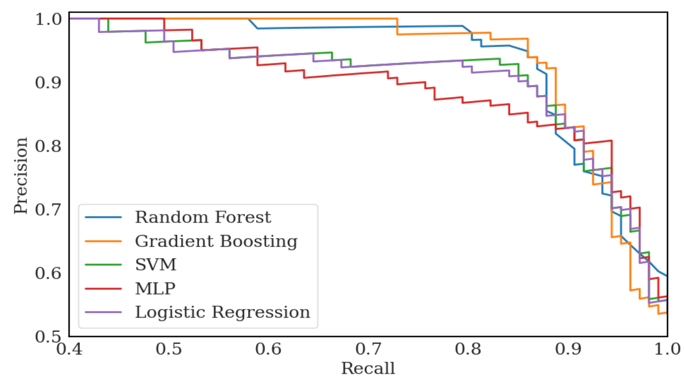


Fig. 2. Precision Recall Curve for the Machine Leaning Models

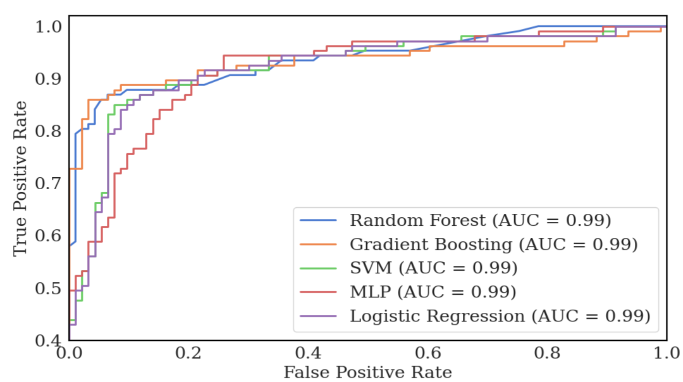


Fig. 3. ROC Curve for the Machine Leaning Models

In Figure 1, we compare the training and test data of five machine learning models, highlighting the consistent high accuracy of Gradient Boosting and Multi Layer Perceptron. Conversely, Logistic Regression shows higher prediction errors. Overall, Gradient Boosting and Multi-Layer Perceptron exhibit robust performance in predicting flight revenue. In Table V, results from testing machine learning models for flight

TABLE V
PRECISION, RECALL, AUC AND F1 SCORE OF THE ML MODELS

| Models | Precision | Recall | AUC | F1-Score |
|---|---|---|---|---|
| Logistic Regression | 0.999 | 0.999 | 0.999 | 0.9987 |
| Multilayer Perceptron | 0.9865 | 0.979 | 0.999 | 0.9972 |
| Support Vector Machine | 0.9852 | 0.9826 | 0.993 | 0.9837 |
| Gradian Boosting | 0.999 | 0.999 | 0.999 | 0.9982 |
| Random Forest | 0.999 | 0.999 | 0.999 | 0.9753 |

TABLE VI
MEAN PREDICTION AND ACTUAL PROFIT WITH DIFFERENT CONFIDENCE LEVEL

| Confidence | Lower Bound | Upper Bound | Actual | Predicted |
|---|---|---|---|---|
| 95% | -343.60 | 1493.712 | 776.9689 | 519.6464 |
| 90% | 12.438 | 1196.856 | 269.993 | 99.01004 |
| 85% | -220.0987 | 872.683 | 173.558 | 58.2222 |
| 80% | -151.177 | 992.1423 | 215.917 | -17.8663 |

fare predictions are summarised. Random Forest and Gradient Boosting show impeccable precision, recall, AUC, and F1-Score metrics. Multi-Layer Perceptron and Support Vector Machine exhibit slightly lower precision but still achieve commendable performance. Logistic Regression displays a minor discrepancy in F1-score, potentially indicating a marginal decline in balanced accuracy.

In Figure 2 and Figure 3, the precision vs recall curve and ROC curve respectively offer insights into the effectiveness of various models in flight fare prediction. Random Forest and Gradient Boosting excel in achieving optimal precision and recall, maintaining a balanced classification accuracy and coverage. Conversely, the multi-layer perceptron and support vector machine show minimal precision-recall tradeoff. Logistic regression exhibits high precision but sacrifices recall, indicating a conservative prediction approach biased towards lower-priced fares. The ROC curves illustrate the effectiveness of machine learning classifiers, with all models demonstrating remarkable sensitivity and specificity, evidenced by few false positives and a high true positive rate. With an AUC of 0.99, random forest, gradient boosting, SVM, MLP, and logistic regression showcase nearly flawless differentiation between higher and lower fares.

This study introduces the utilisation of the LUBE method, as detailed in Table VI, to generate predictive intervals via the MLP model. At a 95% confidence level, these intervals span from -343.60 to 1493.71, illustrating the range of expected margin predictions following the *CWC* approach. Within this range, an actual margin of 776.97 falls, showcasing the model's efficacy in capturing true margin values.

*CWC* cost function optimises prediction intervals' tightness while encompassing a broad spectrum of ticket prices, balancing precision with generalizability. This parallels the MLP model's quantiles, serving as a measure of uncertainty in fare prediction and enabling informed decision-making within a specified confidence framework. Both *CWC* and *MLP* aim to evaluate prediction intervals' reliability, ensuring a robust statistical foundation. Additionally, the comparison between actual and predicted flight fares, along with absolute error bars illustrated in Figure 4, underscores the MLP model's favorable fit. While errors may increase at higher fare values, suggesting uncertainty in predictions within that range, the model effectively supports a dynamic pricing strategy guided by confidence intervals, in alignment with CWC objectives.

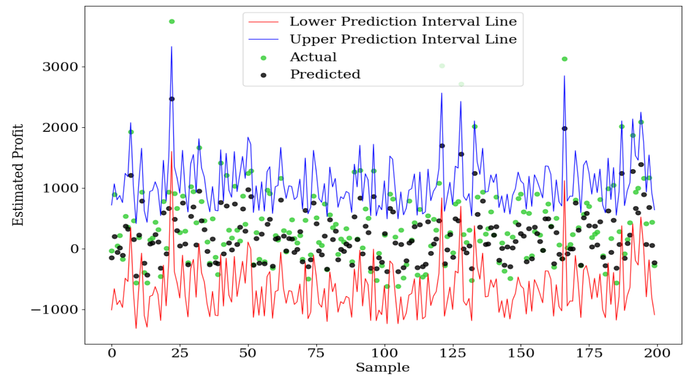


Fig. 4. Actual vs Predicted Values with Prediction Interval

## B. Discussion

Our research focuses on evaluating various models for flight fare predictions, with the Gradient Boosting model emerging as a standout performer, boasting a test accuracy of 99.8% and a RMSE on test data of 56.321, indicating close alignment with actual fare changes. Its precision and recall scores of 99.9% and an F1 score of 99.8% underscore its ability to accurately predict fare classifications, crucial for investment decisions. However, the multi-layer perceptron neural network closely rivals Gradient Boosting, with impressive precision and recall of 99.6% and 99.1% respectively, though with a slightly lower F1 score of 99.7%.

Inclusion of prediction intervals offers valuable insights into forecast reliability, providing investors with a range of projections under different confidence intervals. For instance, at a 95% confidence level, the model forecasts a fare of 519.6464, while the actual fare ranges from -343.60 to 1493.71, showcasing the model's cautious yet insightful predictions. This adaptability over various confidence intervals equips stakeholders in the aviation industry with a clear understanding of potential risks and rewards, strengthening strategic decision-making in Bangladesh's domestic aviation sector.

Employing dual data collection methods was crucial for research validity. Structured surveys gathered primary data from domestic flight passengers, enhancing internal validity. However, the study examined only a limited number of characteristics, leaving room for the inclusion of other significant predictors. While machine learning models excel in capturing complex patterns and trends, they may struggle with disruptions in market conditions. Historical data for training may limit their adaptability to sudden changes or unprecedented events, such as economic downturns or regulatory shifts. Therefore, while prediction intervals method with certain confidence offers powerful tool for profitability prediction, this should be used alongside human expertise and market insights to mitigate potential risks and uncertainties. Our research suggests that there's a favorable opportunity for new entrants to join the domestic airline market in Bangladesh, given the reasonable profit margins observed and the predictive capabilities of advanced models like Gradient Boosting and multi-layer perceptron neural networks.

## V. Conclusion and Future Direction

Our study identifies Gradient Boosting and Multi-Layer Perceptron as optimal models for predicting airline profitability in Bangladesh's aviation market, particularly for the widely used ATR 72 model. Models' accuracy and robustness make them valuable tools for forecasting financial margins and aiding decision-making processes. Techniques like prediction interval using lower upper bound estimation (LUBE) and simulated annealing provide insights into financial margin strategies, ensuring informed decision-making and sustainable growth.

Integration of customer demographics and external factors, alongside advanced machine learning techniques, could enhance airline fare prediction and revenue optimisation. Real-time data utilisation for dynamic pricing strategies and personalised pricing models could optimise revenue streams. Exploring global market expansion and analysing other small aircraft models like DASH-8 could enrich future research avenues. Additionally, augmenting the dataset with customer reviews based on specific routes could further enhance predictive models and drive innovation in airline revenue management.

## References


[1] N. Das, J. Chakrabartty, M. Dey, A. S. Gupta, and M. Matin, "Present energy scenario and future energy mix of bangladesh," *Energy Strategy Reviews*, vol. 32, p. 100576, 2020.

[2] M. A. Noor, S. Ashrafi, M. A. Fattah, S. R. Morshed, and S. Rahman, "Assessment of traffic congestion scenario at the cbd areas in a developing city: In the context of khulna city, bangladesh," *Transportation Research Interdisciplinary Perspectives*, vol. 11, p. 100435, 2021.

[3] N. A. A. Rahman, M. Shoeb-Ur-Rahman, and A. Hassan, "Tourism and air transport sustainability in bangladesh: The role of technology," in *Tourism Marketing in Bangladesh*. Routledge, 2020, pp. 42–50.

[4] G. Regmi, S. Shrestha, S. Maharjan, A. K. Khadka, R. P. Regmi, and G. Chandra Kaphle, "The weather hazards associated with the us-bangla aircraft accident at the tribhuvan international airport, nepal," *Weather and Forecasting*, vol. 35, no. 5, pp. 1891–1912, 2020.

[5] H. Chowdhury, T. Chowdhury, A. Rashedi, S. C. Banik, T. Khanam, R. Saidur, S. M. Sait, and M. A. Rosen, "Energy and exergy assessment with updated reistad estimates: A case study in the transportation sector of bangladesh," *Energy Science & Engineering*, vol. 9, no. 9, pp. 1349–1358, 2021.

[6] S. Kar and P. Khandelwal, "Cross-hedging aviation fuel price exposures with commodity futures: Evidence from the indian aviation industry," *IIMB Management Review*, vol. 32, no. 4, pp. 389–401, 2020.

[7] R. A. Donkor, L. Mensah, and E. Sarpong-Kumankoma, "Oil price volatility and us dollar exchange rate volatility of some oil-dependent economies," *The Journal of International Trade & Economic Development*, vol. 31, no. 4, pp. 581–597, 2022.

[8] A. Horobet, M. L. E. Zlatea, L. Belascu, and D. G. Dumitrescu, "Oil price volatility and airlines' stock returns: evidence from the global aviation industry," *Journal of Business Economics and Management*, vol. 23, no. 2, pp. 284–304, 2022.

[9] N. Sakib, G. Shahariar, M. M. Kabir, M. K. Hasan, and H. Mahmud, "Assorted, archetypal and annotated two million (3a2m) cooking recipes dataset based on active learning," in *International Conference on Machine Intelligence and Emerging Technologies*. Springer, 2022, pp. 188–203.

[10] M. A. Khan, S. K. Ahmed, and S. Khan, "Sustainable electricity generation from hazrat shahjalal international airport, dhaka: A milestone for green energy in aviation," in *2021 5th Annual Systems Modelling Conference (SMC)*. IEEE, 2021, pp. 1–7.

[11] M. A. Hosen, A. Khosravi, H. D. Kabir, M. Johnstone, D. Creighton, S. Nahavandi, and P. Shi, "Nn-based prediction interval for nonlinear processes controller," *International Journal of Control, Automation and Systems*, vol. 19, no. 9, pp. 3239–3252, 2021.

[12] C. Kerdpitak, "The effects of innovative management, digital marketing, service quality and supply chain management on performance in cultural tourism business," *Uncertain Supply Chain Management*, vol. 10, no. 3, pp. 771–778, 2022.

[13] J. Singh, D. Gupta, and L. Sharma, "Analysis of machine learning techniques for airfare prediction," in *Computer Vision and Internet of Things*. Chapman and Hall/CRC, 2022, pp. 211–231.

[14] K. Tziridis, T. Kalampokas, G. A. Papakostas, and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1036–1039.

[15] S. Rajankar and N. Sakharkar, "A survey on flight pricing prediction using machine learning," *International Journal Of Engineering Research & Technology*, vol. 8, no. 6, pp. 1281–1284, 2019.

[16] W. Choi, M. B. O'Connor, and D. Truong, "Predicting the us airline operating profitability using machine learning algorithms," *International Journal of Aviation, Aeronautics, and Aerospace*, vol. 6, no. 5, p. 14, 2019.

[17] W. Groves and M. Gini, "An agent for optimizing airline ticket purchasing," in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 2013, pp. 1341–1342.

[18] T. Wang, P. Wu, Q. Ge, and Z. Ning, "Ticket prices and revenue levels of tourist attractions in china: Spatial differentiation between prefectural units," *Tourism Management*, vol. 83, p. 104214, 2021.

[19] A. Tropis, M. Thomas, J. Bounie, and P. Lafon, "Certification of the composite outer wing of the atr72," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 209, no. 4, pp. 327–339, 1995.

[20] S. Sultana *et al.*, "Economic growth of airlines industry: an overview of domestic airlines in bangladesh," *Journal of Management Research and Analysis*, vol. 5, no. 1, pp. 61–68, 2018.

[21] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937–1967, 2021.

[22] H. D. Kabir, A. Khorsavi, M. S. Rahman, M. A. Hosen, and S. Nahavandi, "Uncertainty quantification for the required fossil fuel generation in a smart grid," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[23] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE transactions on neural networks*, vol. 22, no. 3, pp. 337–346, 2010.

[24] P. Vlaar, P. De Vries, and M. Willenborg, "Why incumbents struggle to extract value from new strategic options: Case of the european airline industry," in *Strategic Management in Aviation*. Routledge, 2017, pp. 277–292.

[25] M. A. R. Sarker, C. G. Hossan, L. Zaman *et al.*, "Sustainability and growth of low cost airlines: an industry analysis in global perspective," *American Journal of Business and Management*, vol. 1, no. 3, pp. 162–171, 2012.

[26] M. Bugaj, T. Urminskỳ, J. Rostáš, and P. Pecho, "Aircraft maintenance reserves–new approach to optimization," *Transportation Research Procedia*, vol. 43, pp. 31–40, 2019.

[27] M. A. Camilleri and M. A. Camilleri, *Aircraft operating costs and profitability*. Springer, 2018.

[28] D. E. White, N. D. Oelke, and S. Friesen, "Management of a large qualitative data set: Establishing trustworthiness of the data," *International journal of qualitative methods*, vol. 11, no. 3, pp. 244–258, 2012.