Operational Performance Analysis for Brazilian Aviation System using XAI: A Case Study for Load Factor Analysis

Burak Cankaya
Embry-Riddle Aeronautical University, Daytona Beach, Florida, USA

Kamil Ciftci*
California State University, Fresno, Fresno, California, USA

Joao S. D. Garcia
Embry-Riddle Aeronautical University, Daytona Beach, Florida, USA

Nader Madkour
Lamar University, Beaumont, Texas, USA

Berna Eren Tokgoz
Lamar University, Beaumont, Texas, USA

Khem Poudel
Middle Tennessee State University, Murfreesboro, Tennessee, USA

Load Factor (LF) calculates the sold-to-available seats ratio and defines the plenitude of commercial flights. It is one of the primary key performance indicators (KPI) for airlines, and the LF of an airline route is one of the most critical unknowns when opening a new line. In addition, understanding the associated variables and their importance in predicting the LF is essential for making a managerial decision about the condition of a route. This study employed Naïve Bayes, Fast Large Margin, and Deep Learning models with Explainable AI (XAI) SHapley Additive exPlanations (SHAP) methods for local and global interpretations of how airlines could increase their LF. The Naïve Bayes model gave acceptable predictive performance results (98.1 % Mean ROC, 90.35% Precision, 77.08 Recall, and 91.5% Accuracy) and the most interpretable results. The preliminary results present relevant insights into understanding the factors that drive efficiency in the Brazilian aviation system.

---

*Corresponding Author. E-mail address: kamilciftci@csufresno.edu

I. INTRODUCTION

The airline business has a complex financial model structure to define profitability (Baldanza, 1999). The primary financial income of the industry is derived from sold tickets (McAfee and Te Velde, 2006). Ticket prices vary among seats, legs, and dates. The price can be defined through dynamic pricing strategies following auction theory (Ivaldi, Petrova, and Urdanoz, 2022). Load Factor (LF) is the number of paying passengers divided by the number of available seats on a specific flight leg. It is one of the key performance indicators (KPI) associated with airlines' profitability and general performance (Sinha, 2007; Szabo, Mako, Tobisova, Hanak, and Pilat, 2018; Anderson, Curtis, Moss, Smith, and Mey, 2017). Companies measure the LF for different routes, among other variables, to support route planning decision-making, and especially for low-cost carriers, the LF has more importance on reflecting the operational performance (Belobaba, 2009; Roberts and Griffith, 2019). Different routes require different aircraft with various flight ranges, fuel efficiencies, yields, type ratings, number of pilots, pilot bases, pilot motivations, and many other direct and indirect factors that impact the LF (Zeng and Rutherford, 2019). Even if it may be challenging to calculate precise revenue values for each flight independently, the LF is an essential factor associated with the companies' profitability (Daft and Albers, 2012). In this study, we evaluated domestic Brazilian airlines for ten years (2011-2021) to predict the LF of individual routes using Machine Learning (ML) methods. We analyzed and summarized the underlying patterns for success.

The dataset used in this research is made publicly available by the Brazilian National Civil Aviation Agency. It includes many variables such as flight stage and airline related to the origin and destination locations, number of paying and unpaying passengers, number of available seats, the total mass of cargo and mail carried, distance flown, and fuel consumed, among other variables. See (Marazzo, Scherre, and Fernandes, 2010 ; Agência Nacional de Aviação Civil, 2022) for the details of the dataset. With that, relevant metrics in addition to the LF were obtained, including average seat kilometer (ASK), revenue seat per kilometer (RPK), average ton kilometer (ATK), and revenue ton kilometer (RTK). Still, these variables are metrics related to the LF; these additional variables were not employed in the proposed predictive models to avoid overfitting because they might be related to the target variable (Cankaya, Eren Tokgoz, Dag, and Santosh, 2021)

We first prepared the data to achieve thrifty Artificial Intelligence/ Machine Learning (AI/ML) models. The data comes from open-sourced databases that the Federal Aviation Authority of Brazil releases monthly (ANAC, 2022). First, we downloaded the dataset, translated it into English, and cleaned it. Then we prepared additional variables from existing continuous variables, such as Distance_Long and Airtime_Long, by breaking them into categories, such as long and short distance and airtime, to give the algorithm more power to understand the patterns for prediction. The target variable categories are 0 (%72.7) and 1(%27.3), and the data is ready for prediction.

We applied ten-fold cross-validation to verify the findings of the ML algorithms in each fold. The data is broken into ten pieces; every piece is tested against the remaining nine pieces, and median of these pieces is used for final comparison. Then, we utilized variable selection algorithms to choose the best algorithm and optimal parameters for these algorithms. Naïve Bayes, Fast Large Margin, and Deep Learning algorithms were chosen due to the mixed categorical-continuous nature of the data. The results of each Naïve Bayes, Fast Large Margin, and Deep Learning have their power either by being simple, explanatory, or highly predictive, so we used the Information Fusion method to combine the findings of these methods and prepare the variable importance diagram.

The remainder of this study is organized into six sections. Section 2 provides a global background with the literature review. Section 3 shares the original data and how this study's final version was used. In addition, the methodologies used in this work are explained. Section 4 presents the numerical results of the performance measurement comparison of the models and selected algorithms are reported. Section 5 summarizes the main results and business findings. Finally, study limitations and future directions are summarized in Section 6.

II. LITERATURE REVIEW

Aviation companies aim to measure and enhance their performance through Key Performance Indicators (KPIs), offering valuable insights into various operations.

In literature, combinations of Key Performance Indicators (KPIs) are utilized as data variables for various purposes. For instance, Lozano and Gutierrez (2011) employed combinations of environmental impact, fleet cost, operating cost, and Revenue Tonne Kilometers in their multi-objective linear programming model within the Data Envelopment Analysis (DEA) approach to assess operational efficiencies of European airlines. Sakthidharan and Sivaraman (2018) developed a DEA approach utilizing linear programming models, where the total number of revenue-paying passengers and the total weight of freight carried were utilized as input factors. Additionally, total weight capacity, cost incurred per available seat kilometer, fuel cost per available seat kilometer, maintenance cost, cost per fleet ownership, and number of employees in airlines were considered to estimate the efficiency of airlines in India. In another study, Singh, Sharma, and Srivastava (2019) incorporated average seat per kilometer, average payload, average stage length, average fuel price, and airline ownership in their regression model to gauge the impact of these factors on operational cost efficiency for Indian Airlines.

These examples can be expanded given the extensive use of KPIs in literature. However, we concentrated on two essential topics within Aviation literature: LF in Aviation and ML applications for Aviation business metrics or KPIs.

2.1. LF in Aviation

The LF represents the percentage of occupied seats on an aircraft during a specific flight or over a given period, making it a vital metric in the aviation industry. Achieving an optimal LF is critical for airline profitability and operational efficiency (Smith and Tesone, 2019). Numerous studies have examined the LF and its implications for airline performance.

Barros, Liand, and Peypoch (2013) applied a B-convex optimization model to assess the technical efficiency of US airlines. They considered a combination of operational and financial indicators, including Total Revenue, Passenger numbers, LFs, and Revenue Passenger Mile.

In their study, Joo and Fowler (2014) utilized DEA to evaluate and contrast airlines operating in Asia, Europe, and North America. They analyzed KPIs including Revenues, Passengers, Revenue per Kilometers, and LF. However, they noted that LF did not emerge as a significant factor in Tobit regression analysis when explaining operational efficiency.

Choi, K. (2017) conducted an extensive analysis of the technical efficiency (TE), pure technical efficiency (PTE), and scale efficiency (SE) scores of 14 US airlines spanning from 2006 to 2015. The study also investigated the impact of mergers and acquisitions (M&A) on airline efficiency and identified key economic drivers affecting environmental variables. KPIs such as Cost per Available Seat Mile, Revenue per Available Seat Mile, Passenger Yield, and Load Factor (LF) were used in this study and one significant finding suggested that JetBlue Airways and Virgin America should implement new revenue management strategies to increase seat sales revenue and improve load factors.

Johnson and Thompson (2018) analyzed the impact of the LF on airlines' financial performance and they discovered a positive correlation between the LF and revenue, underscoring the importance of efficiently filling aircraft seats to enhance profitability. The authors emphasized the role of pricing strategies, network planning, and demand forecasting in optimizing LF.

Chen and Li (2020) investigated the relationship between LF and customer satisfaction, they found that higher LFs positively influence customer satisfaction, as passengers perceive full-capacity flights as more popular and reliable. This finding underscores the importance of maintaining high LF to enhance customer experience and loyalty.

2.2. ML Applications for Aviation Business Metrics

In recent years, ML techniques have gained prominence for analyzing large volumes of data and extracting valuable insights. Several studies have explored ML applications for Aviation business metrics and KPIs, offering improved decision-making and operational efficiency opportunities.

2.2.1 Flight Delay Prediction

Wang and Delahaye (2018) introduced a hybrid learning prediction model for estimating arrival times in terminal maneuvering areas. This model combines clustering preprocessing with a Multi-cell Neural Network to address short-term trajectory predictions. In another study, Schultz and Reitmann (2018) concentrated on aircraft boarding, crafting a validated stochastic boarding model to gather dependable aircraft status data. Additionally, they employed a Long Short-Term Memory model to understand passenger behaviors, integrating this data into their neural network approach to predict boarding times. Etani (2019) developed a flight prediction model for Japan, integrating various supervised machine learning techniques to forecast on-time arrivals by considering weather and flight data correlations. Wang, Tang, Tang, Yang, and Li (2019) conducted a study using ML algorithms to predict flight delays and optimize airline operations. The authors developed a predictive model that outperforms traditional methods by leveraging historical flight data. The ML approach enabled airlines to proactively manage delays, minimize disruptions, and improve overall service quality.

2.2.2 Sustainability

The sustainability of airline operations is another key focus area where machine learning methods are applied, particularly in predicting fuel efficiency and emissions. In an early study, Kapoor, Horvitz, Lauve, and Horvitz (2014) developed a wind forecasting model aimed at minimizing fuel usage while optimizing flight plans. Baklacioglu (2016) introduced a generic algorithm-optimized neural network for predicting fuel consumption using real flight data. Tian, Huang, Ye, and Yang (2019) utilized adaptive machine learning methods for air quality classification at airports, incorporating various factors such as airport configuration, flight operations, aircraft performance, and weather data. Zhu and Li (2021) proposed a novel deep-learning flight time prediction model to reduce excessive fuel loading for airline flights, demonstrating its potential to save fuel costs and minimize emissions. Wan, Zhang, Lyu, and Zhou (2022) suggested a combined adaptive weighting approach, integrating long-short term memory and extreme gradient boosting (XGBoost) models, to measure fuel consumption and flight emissions using real-time weather data for short-term emissions prediction. They compared the performance of their model with six other machine learning models.

2.2.3. Air Traffic Management

Reitmann and Schultz (2022) proposed an adaptive framework utilizing machine learning methods to optimize the air traffic management system, focusing on predicting boarding times and traffic patterns.

In another study, Midtfjord, De Bin, and Huseby (2022) developed a machine learning framework based on XGBoost to predict airport runway conditions and identify slippery levels. Their models were further enhanced with Explainable AI, providing valuable insights such as high accuracy and trustworthy information about runway conditions. They discussed the benefits of such information, including reduced fuel usage under certain conditions. In a recent study, Krishna, Sariki, Sigamsetty, Kumar, and Benil (2023) forecasted flight landings, improved safety measures, and enabled informed decision-making in complex situations for both autonomous and human-operated aircraft using Deep Learning algorithms such as long short-term memory (LSTM).

2.2.4. Various Studies and Aviation KPIs

Various machine learning studies have concentrated on different aviation operational and business metrics in the literature. For instance, there are studies on engine design and condition prediction (Malatesta and Yang, 2021; Matuszczak, Żbikowski, and Teodorczyk, 2021; weather forecasting (Li, He, and Paoli ,2020; Sim, Park, Park, Ahn, and Chan, 2018), wildlife strikes (Altringer, Navin, Shwiff, and Anderson,2021),flight heath status (Basora, Bry, Olive, and Freeman, 2021), ticket pricing(Alauddin and Ting, 2020; Wozny, 2022). However, relatively few studies specifically utilize machine learning applications to focus on aviation KPIs. One such study by An and Lee (2021) investigated the application of ML algorithms for demand forecasting. They developed accurate demand prediction models by analyzing historical passenger data and incorporating factors such as seasonality and economic indicators. These models enabled airlines to optimize capacity planning, flight scheduling, and resource allocation, ultimately enhancing operational efficiency.

This literature review has emphasized the significance of Aviation KPIs, explicitly focusing on the LF in Aviation and the applications of ML for Aviation business metrics or KPIs. The LF has been identified as a critical metric impacting airline financial performance and customer satisfaction. Optimizing the LF necessitates effective pricing strategies, network planning, and demand forecasting. Meanwhile, ML applications offer opportunities to leverage historical data and enhance decision-making in various areas, such as predicting flight delays and optimizing demand forecasting.

As the aviation industry evolves, integrating effective KPIs and ML applications will play a vital role in driving operational excellence, improving customer experience, and sustaining competitive advantage. Although the use of ML methods to investigate airline metrics is rising, this study fills the gap to understand better the various impact elements and how they interact to affect airline LF efficiency using Explainable AI techniques. More research is required to fill in these gaps and develop a more thorough understanding of the intricate interactions between many elements and their combined effects on the LF of passenger flights. The gap in the literature is specifically understanding and explaining the features that predict LF with Machine Learning. In addition, we utilize XAI to explain the local and global business questions that bring transparency and robustness to understanding the LF prediction patterns and considering operational frameworks of the airline business.

Thus, by using top-notch machine learning algorithms, we are contributing to the literature by;

a) We introduce a new airline efficiency metric that measures airlines' sustainable and operational efficiency index.

b) We present research data (availability and originality of the data) on Brazil's airline efficiency for 10 years.

c) We compare Naïve Bayes, Fast Large Margin, and Deep Learning methods by predictive performance and expansibility.

d) We used the SHapley Additive exPlanations (SHAP) method as XAI to explain the prediction and make local inferences to compare airlines to understand their LF efficiency on multiple routes.

e) We simulate real business scenarios and implement local and global interpretations of ML methods.

III. DATA DESCRIPTION AND METHODOLOGY

3.1. Data Description

The data came from a publicly accessible database, ANAC (2021), made available from January 2000 through 2021. The business circumstances, including the COVID-19 issue, were considered for our analysis using the period from January 2011 to August 2021. There were ten variables in total—two continuous variables, eight categorical variables, and 75173 data instances—in our dataset.

ANAC has been very transparent by releasing detailed KPIs regarding airlines and routes every month since 2000. Over 100 variables evaluate airlines' business performance, environmental sustainability, and economic sustainability. The original data was in Portuguese, and it was translated into English. The data used the International System of Units with units such as kg, meters, etc. This study investigated the LF, and the variables used in the original LF calculation could not be used because that would cause overfitting. We also only utilized known variables to the routes before the route was opened. These variables were the flight specifications such as the carrier, distance, airtime, departure and arrival states, departure and arrival region, year, and month. We did not use variables that could not be known before a flight, such as fuel consumption, available seat, revenue seat, etc. Distance variable was created as a binomial variable if the flight was equal or longer than 1000 km to be 1, 0 otherwise. The airtime variable is also created as a binomial variable; if the flight was equal to or longer than two hours, it is 1, 0 otherwise. Setting-up goals for KPIs in airlines is a common practice, and many airlines worldwide have 80% LF as their success threshold. We also used 80% LF as our success threshold and prepared the target variable as one if LF was equal or greater to 80% to be 1, 0 otherwise. The descriptive details about the dataset used in the study are given in Table 1.

TABLE 1. DESCRIPTIVE OF THE VARIABLES

Variable	Explanation	Data Type	Descriptive Statistics*	Percent Missing
LF	If >=0.80 flight is efficient otherwise flight is not efficient	Binary	0 (72.7), 1(27.3)	0
Destination State	Arrivals State	Nominal	Sao Paulo (21.9), Minas Gerais (8.9)	
Destination Region	The region of the destination of the flights	Nominal	South East (40.2), North East (20.8)	0
Distance	Distance by Kilometer	Numeric	849.6(587.9)	0
Distance_Long	Flights Long or Short (Greater equal or less than 1000 km)	Nominal	Short (75.6), Long (24.4)	0
Airtime	Airtime by Minutes	Numeric	1668(0.76)	0
Airtime_Long	Airtime Long or Short (Greater equal or less than 2 hours)	Binary	Short (0.74), Long (0.25)	0
Origin State	Departures State	Nominal	Sao Paulo (20.2), Minas Gerais (8.8)	0
Carrier	The airline carrier company	Nominal	C (29), A (24.6)	0
Year	The year of flights	Numeric	2015 (2.8)	0
Month	The month of flights	Numeric	6.3 (3.4)	0

* Descriptive Statistics: Bolded variables used in ML model. Binary- % of each category; Nominal- % of most common two categories; Numeric- average (standard deviation).

3.2. Methodology

Due to the predictive power and explanatory strength of ML methods, ML methods can be used to predict the business performance metrics in the airline business, which is an efficient method (Choi, O'Connor, and Truong 2019).

CRISP-DM is the standard process to ensure the quality of the analysis in this project (Wirth and Hipp, 2000). The CRISP-DM life cycle has six steps, as shown in Fig. 1. The first step is business understanding. Aviation is a deep subject area, and business understanding is essential to align the analysis to the right business objectives and choose the target variable accordingly (Rinehart, Smith, Spencer, 2014). The second step is data understanding, which is closely related to the first step, which is to use the correct data to aim at the target variable.

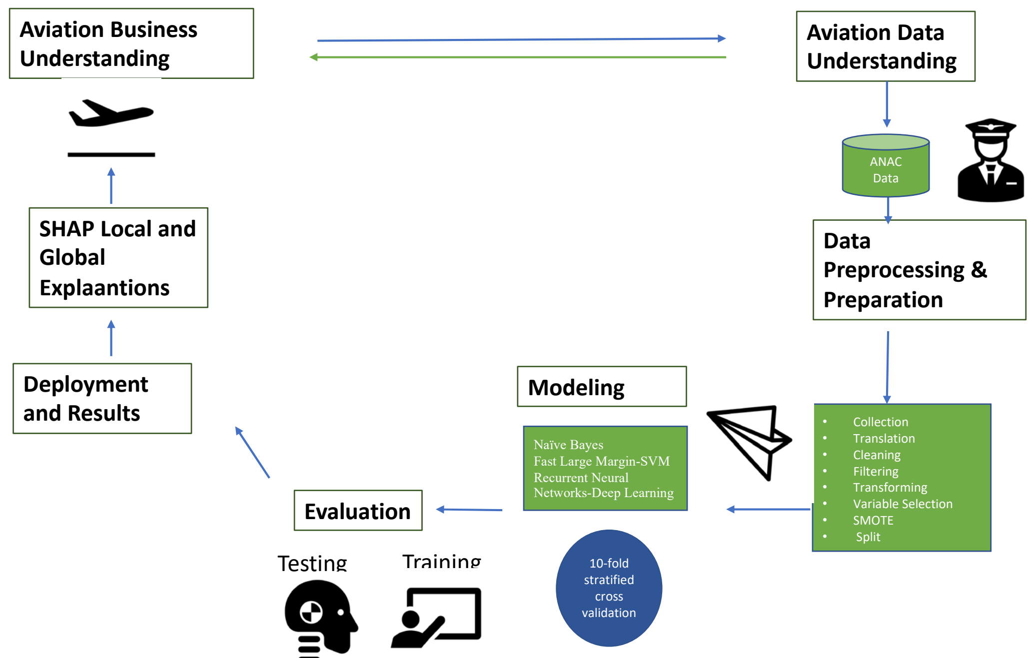

FIGURE 1. EXAMPLE FIGURE DEMONSTRATING CRISP-DM METHODOLOGY

3.2.1 SMOTE Data Balancing, Cross-Validation and Random Sampling.

Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the inequality between categories of the target variable to be one-to-one, which was 0 (%72.7), 1(%27.3) initially**.** The third step is data preparation, where necessary variables are chosen, missing data is treated, new variables are added if needed, and the information is filtered to aim the study's objective. Data preparation is one of the most time-consuming and vital steps for the best analysis. The fourth step is modeling, and the best-performing ML models, depending on the data types, are chosen at this step. After careful review and analysis, this study chose Naïve Bayes, Fast Large Margin, and Deep Learning algorithms. The data gets re-prepared according to the model results, such as data weight balancing and categorization using subject matter expertise (SME). The fifth step is to evaluate the models' results and convert statistical results to business findings and strategies (Ridgeway, Madigan, Richardson, and O'Kane, 1998). The sixth step is to deploy all three models, which could be used in the future for training and testing, and when new data comes, the model could test its prediction on the new data, which could save computational time and allow practical use of the model (Sheppard and Young, 2006).

One of the essential contributions of the study is using XAI methods. XAI methodology is an effective decision-making solution to fine-tune vital combinations in the business scenario and predict the likelihood of the created scenario for that particular carrier and flight route combinations using Shapley Additive Explanations (SHAP) local interpretations. SHAP are explainable ML algorithms used to explain prediction with local business scenarios. The SHAP explanation model gets the average likelihood of the scenario combinations. Additionally, if the event combination never happened, then it predicts the likelihood of the case if it happened and visualizes the results. The SHAP explanation function is a tremendous practical contribution to the state of the art, whereas previous ML models primarily ranked the correlated variables. The ML-based variable sensitivity SHAP explanation model enables us to understand deeper business relations, reverse engineer the prediction, and help us create remedy business insights (Cankaya, Topuz, and Glassman, 2023[a]; Banghart, 2017).

3.2.2. Predictive Models and Model Selection Criteria

In this section, we provide an overview of ML prediction models implemented in this study, review how data is processed for accuracy, and summarize how to evaluate the prediction models' performances. The predictive problem in this study focuses on the Explainability of the phenomenon of understanding the Load Factor that defines the aviation business performance of airlines. The most essential model selection criteria in this study are Explainability and robustness. We only use ten controllable variables, a mix of binary, nominal, and numeric variables. The context of the controllable variables is also essential. Origin, destination, and distance-related variables are critical for controlling the most efficient routes. The Fast Large Margin SVM model gave acceptable results by mainly using the binary and nominal variables, and the Deep Learning model showed promising results by using all variables. The naïve Bayes model gave the best results in terms of interpretability with the usage of Origin, destination, and distance-related variables, and its predictive performance given in Table 4 is also superior for accurate mean ROC and in-class precision. Considering model biases, the Naïve Bayes model is also the most transparent and favored its usage. The model fine-tuned data balance and the risk of underfitting or overfitting and confirmed the results for fitting and quality control of the predictions on testing data.

3.2.2.1. Naïve Bayes

The Naive Bayes is a classifier method which is a modified format of the Bayes Theorem. The Naive Bayes provides probability estimates for parameters with minimal training data and effort.

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)} \tag{1}$$

Where:
a is the targeted class, and b is predictor.
$P(a|b)$ indicates the probability of class a given predictor b.
$P(b|a)$ indicates the probability of predictor b given class a.
$P(a)$ is the probability of the class a.
$P(b)$ refers probability of the predictor b.

3.2.2.2. Fast Large Margin

The fast Large Margin model utilizes a fast margin learner using (Fan, Chang, Hsieh, Wang, and Lin 2008)'s proposed linear support vector learning method. While the Fast Large Margin model's performance and outputs are similar to a standard support vector machine (SVM), the Fast Large Margin model excels at handling large datasets. Due to this advantage, it is employed in this study.

3.2.2.3. Deep Learning

Deep Learning is an advanced open source that utilizes in-memory compression to handle big data and use ML algorithms to solve complex problems. While many ML algorithms, such as generalized linear modeling, Naïve Bayes, principal components analysis, k-means clustering, and word2vec, are available, several in-class algorithms at scale, such as distributed random forest, gradient boosting, and deep learning are implemented in the Deep Learning (Candel, Parmar, LeDell, and Arora, 2023).

Deep learning has become more popular as an algorithm choice for prediction accuracy since it offers better training stability, generalization, and scalability with larger datasets. The deep learning algorithm uses a multilayer, feed-forward structure in which the initial stage is called the input layer. In contrast, subsequent stages are known as multilayers of nonlinearity and linear regression layers, respectively. More specifically, "Bengio (2009)'s lost function, (2), is used as the learning algorithm:

$$L(W, B|\, J) \tag{2}$$

where W is the matrix of weighted collection layers, J is training samples while B represents column vector of biases for next layer". This study uses many features of Deep Learning algorithm, such as automatic data processing and standardization, the automatic turning of performance, various regulation procedures, and adaptive learning rate algorithm ADADELTA. For further details, we refer readers to (Candel, Parmar, LeDell, and Arora, 2023).

3.2.2 Cross-Validation and Random Sampling

We implemented a ten-fold cross-validation technique commonly applied in the literature to assess our predictive model's accuracy. In cross-validation, the data set is divided into ten equal parts, ensuring that the distribution of the response rate is maintained in each section through stratification. During each iteration, one portion is designated as the test set while the rest are used for training the model. Overall performance is determined by averaging the results of each iteration for the comparison. Utilizing the stratified parameter, which ensures the selection of samples in proportion, and the shuffle parameter, which promotes a well-mixed dataset, maintains randomization, and enables an effective k-fold cross-validation technique. Parameters and descriptions for cross-validation are given in Table 2.

TABLE 2. K-FOLD CROSS VALIDATION PARAMETERS

Parameter	Possible Range	Experimental Range	Type	Explanation
n_splits	[2, ∞)	10	Integer	It determines how many folds there are. It was set to 10 for this experiment to split the dataset into 10 smaller groups.
shuffle	[True, False]	True	Boolean	It shows whether to shuffle the data before dividing it into batches. To make sure that there is a healthy balance of data samples in each data subset, it was set to True.
stratified	[True, False]	True	Boolean	It shows the usage of stratified K-fold, which guarantees that each fold contains about the same percentage of samples from each target class as the entire set, is indicated. It was set to True for this experiment to preserve the distribution of classes.

3.2.3. Performance Evaluation

The main performance metrics used in the study are Mean ROC (3), Accuracy (4), Precision (4) and Recall (6). Since there are two conditions for the target variable, the predictions are categorized under true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The performance metrics are calculated from the confusion matrix as follows:

$$Mean\ ROC = TP\ /\ FP \tag{3}$$

$$Accuracy=(TP+TN)/(TP+TN+FN+FP) \quad (4)$$
$$Precision = TP / (TP + FP) \quad (5)$$
$$Recall = TP / (TP + FN) \quad (6)$$

3.2.4. Model Interpretation

The ML models predict a classification problem, and that prediction has a value for the subject matter area. In our case, it is the positive LF flights, and knowing these flights with high predictive performance will give us competitive power as an airline when we open a new route. The next level of information comes from Information Fusion, and it gives us the relative importance of the variables by orders and magnitudes while predicting the positive LF flight. We pay more attention to these flight specifications and use this order and magnitude for business inferences while opening/closing and changing route frequency. Another additional level of information we contribute through this study is SHAP local interpretations. The SHAP local interpretations display one feature's contributions to a particular occurrence.

3.2.4.1. Information Fusion and Sensitivity Analysis

The information fusion method is a sensitivity analysis method that involves multiple ML algorithms and gets a fair evaluation of variable importance. Each algorithm has a different preference for variables depending on the data type, quantity, and quality of the variable. To be fair for various data types, merging these predictions and reaching a decision on variable importance is better than using only one ML model's variable importance. Merging the predictions of multiple ML methods give more precise and reliable outcomes. Information fusion applied majority voting and weighed averaging of variables to conclude variable importance.

Information Fusion can be numerically defined as:

Let's define:

- $X = \{x_1, x_2, \ldots, x_n\}$ as the set of predictions from different models
- $y = \{y_1, y_2, \ldots, y_n\}$ as the set of weights for each model, weight can be coming from predictive performance.
- N_{if} is the merged Information Fusion.

We can formulate Information Fusion using the weighted average method as follows:

$$N_{if} = \sum (x_i * y_i) \textit{ for all i in [1, n],}$$
$$\textit{where } \sum y_i = 1 \textit{ for all i in [1, n].} \quad (7)$$

The weight y_i for the N_{if} can be decided depending on the model. While in some models, predictive model performance is more important, in other models, reliability is more critical, and in some models, computational speed is more important. y_i can be one or multiple factors depending on the predictive problem. In our problem setting, the weight is chosen to be the predictive performance and reliability due to the high performance and reliability nature of the research problem. More details of the Information Fusion model can be gathered from (Torra, 2003).

3.2.4.2. SHAP Local Interpretations

XAI is a crucial topic that brings value over predictions because we need more than predictions. Explaining a prediction is as valuable as making predictions in many research cases.

SHAP stands for SHapley Additive exPlanations (SHAP). This informative method is essential in the XAI journey because we know some complex, patterned predictive problems will need other methods. Even though SHAP does not explain causality, it brings local and global explanations that bring transparency and robustness to understanding the prediction patterns. The SHAP gets the explanation from the Shapley Values, and the SHAP Values get the average of the marginal contribution of the variable's errors in all permutation occurrences for a prediction and make local explanations depending on the averages. The SHAP is calculated from Shapley Values, and it is a game theory-based algorithm that can work with tree-based and ensemble models (Lundberg and Lee., 2017). The SHAP gives Shapley Values to each variable, which are important values for the specified classification problem. Then, Shapley Values are converted to partial dependence plots where they visualize how a variable affects the positive and negative relationship for the predicted class likelihood while others are kept constant. Being able to visualize the contribution of individual variables increases the transparency of the model. Evaluating SHAP values for input variables in local explanation tests the robustness of the model,

We can formulate SHAP values as

$$SP_i = \sum (|SQ|!\,(C - |SQ| - 1)! \,/\, C!) * [T(SQ \cup \{c_i\}) - V(SQ)] \quad (8)$$

Where:

- $C = \{c_1, c_2, I, c_n\}$ as the set of members of the coalition
- $T(c)$ as the total value of the coalition
- SP_i as the Shapley value of coalition member i
- $\varphi(c_i \mid SQ)$ as the marginal influence of coalition member i to a coalition S $(where\ S \subseteq C)$

In light of Štrumbelj and Kononenko (2014)'s initial Shapley Value formulization, we defined the Shapley value SP_i for each variable as:

The Shapley value comes from the coalition game theory. In this coalition game, we assume x number of players build a grand coalition, and the sum value of this coalition is ΔC. Next, there are smaller coalitions between variables which are called subset coalitions. The sum value of these smaller coalitions are defined as $SQ(where\ SQ \subseteq C)$. The coalition aims to share the sum value of the coalition fairly also by considering each variable's value in sub-coalitions. The value of each variable is called the perturbed value. The SHAP values are recalculated with perturbed values for each variable to calculate the sensitivity analysis.

IV. RESULTS

Results for the different predictive models are presented in this Section. Table 3 shows the weight by correlation for the model, which reflects the most potent original variables and created variables needed to predict the combined results using information fusion (Graefe, Armstrong, Jones Jr, and Cuzán, 2014). In the Brazilian aviation LF context, the distance of flights was the most relevant factor. This means the airlines already consider LF while deciding the number of departures and their timing during peak seasons. The number of departures and the flight distance followed the distance. Airlines are also distinguished from each other by LF, meaning there are better companies with better LF. The relative comparative weight by correlation diagram in Table 3 shows that distance is the most impactful variable in having a high LF. We evaluated other variables with our aviation subject matter knowledge; however, they were mainly characteristics connected with the load factor. It would be overfitting to use these variables. On the other hand, our goal in this XAI research is explainability to maximize the business inferences we targeted under our control. The finest XAI methods to suggest the optimum routes for increasing the Load Factor come from simple variables relating to origin, destination, and distance.

TABLE 3. RELATIVE WEIGHT BY CORRELATION

Attribute (A_n)	***Weight (w_n)***

Distance	0.34
Airtime	0.31
Year	0.263
Carrier	0.246
Dest_State	0.20
Orig_State	0.19
Dest_Reg	0.16
Orig_Reg	0.14

Correlation coefficients for each attribute in Table 3 show the relative weights of each variable while predicting the LF. It shows the distance of routes as the most crucial variable (0.34). Airtime (0.31) is also related to the distance for routes, but it might be different in changing weather conditions and the direction of flights where the earth's revolving impacts east-west direction flights. One good observation for the market is that the year of flight is essential, and the LF has improved steadily over time. New-generation aircraft and optimization models are used in day-to-day operations, and airlines improve their LF yearly. The following important information comes from states and regions. The states are more important than regions, and destinations are more critical than origins where Dest_State (0.20), Orig_State (0.19), Dest_Reg (0.16), and Orig_Reg (0.14) relative percentage.

Table 4 shows the model comparison with predictive performance measures to predict the LF success of routes. We have three models with good performance results. Two of these (Fast Large Margin and Deep Learning) models are ensemble models and excel in prediction but need to improve in explaining the predictions. The Naïve Bayes model is a Bayesian model, and it is successful for predictive performance and very explanatory in showing the patterns; it is suitable for XAI (Aslam et al., 2022). It is essential to make accurate predictions but more critical to understand the prediction patterns in aviation LF prediction. Airlines can make strategic route and frequency decisions by analyzing the patterns for successful LF. The Naïve Bayes model showed 91.5% accuracy, 98.1% Mean ROC, 77.08 in-class recall for high-LF, and 90.35 in-class precision for high-LF, which means the predictive results for both positive and negative classes are acceptable for making business inferences (Simsek, Dag, Tiahrt, and Oztekin, 2021).

TABLE 4. MODEL PERFORMANCE COMPARISON

	Fast Large Margin SVM	Deep Learning	Naïve Bayes
Accuracy	90.0	91.7	**91.5**
Mean ROC	98.2	97.9	**98.1**
Class Recall	72.52	86.3	**77.08**
Class Precision	92.9	83.9	**90.35**

The confusion matrix is also critical to showing model predictive performance because of the number of cases in which we make true and false predictions for both high LF and low LF cases. These are known as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). In some business practices, making mistakes in predicting the correct category can incur very high penalties, such as predicting aviation accident cases involving injuries. Other studies also value different performance measures as highly important., such as understanding general no-injury aviation incidents that cause delays (Cankaya, Topuz, Delen, and Glassman, 2023[b]). In our study, understanding and explaining high LF is the most critical metric. Accordingly, we evaluate the performance metrics in Table 5 by having 7531 True Positive cases versus 2339 False Positive cases. The number of repetitions of these predictions is also essential, as the percentages prove the model performance.

TABLE 5. CONFUSION MATRIX FOR NAÏVE BAYES MODEL

Critical component (X_n)	True 0	True 1	Class Precision
Predicted 0	25222 (FN)	2239 (TN)	91.85%
Predicted 1	804 (FP)	7531 (TP)	90.35%
Class Recall	96.91%	77.08%	

The confusion matrix shows the model's reliability and predictive power for different categories. In this problem setting the interested predicted category is cases the LF is over 80%. Explanations to explain these successful routes are more important.

Table 6 provides the performance parameters of the Naïve Bayes method. The Naïve Bayes algorithm acts with a Bayesian background. It tests conditional probabilities for the classes of variables. Fine-tuning the use of ML is one of the most essential steps. In this study, we chose greedy estimation with ten kernel density modes that use Laplace conversion; the minimum kernel bandwidth was 0.1 for our Naïve Bayes algorithm.

TABLE 6. NAIVE BAYES PARAMETERS

Parameter	Possible Range	Experimental Range	Type	Explanation
alpha	[0, ∞)	Laplace=1	Float	a method of correction that maintains zero probabilities of an event occurring using Laplace conversion
Estimation Mode	[None, List]	Greedy Estimation Mode	List/None	Kernel Density Estimation Mode.
Minimum Bandwith	(-+ ∞)	0.1	Real	Minimum Kernel Bandwith
Number of Kernels	[1, ∞)	10	Integer	Number of Kernels used

The model shows in Fig. 2 that high-distance (over 1000 km) flights have a 2.7 times higher likelihood of having successful LF. The figure visualizes the crucial impact of distance on the LF. Distance is the highest correlating variable in Table 3, and we can see that long-range flights in this business setting have the highest likelihood of having positive LF. Another discover from the study is that longer flights have a 1.28 times higher probability of success than shorter flights.

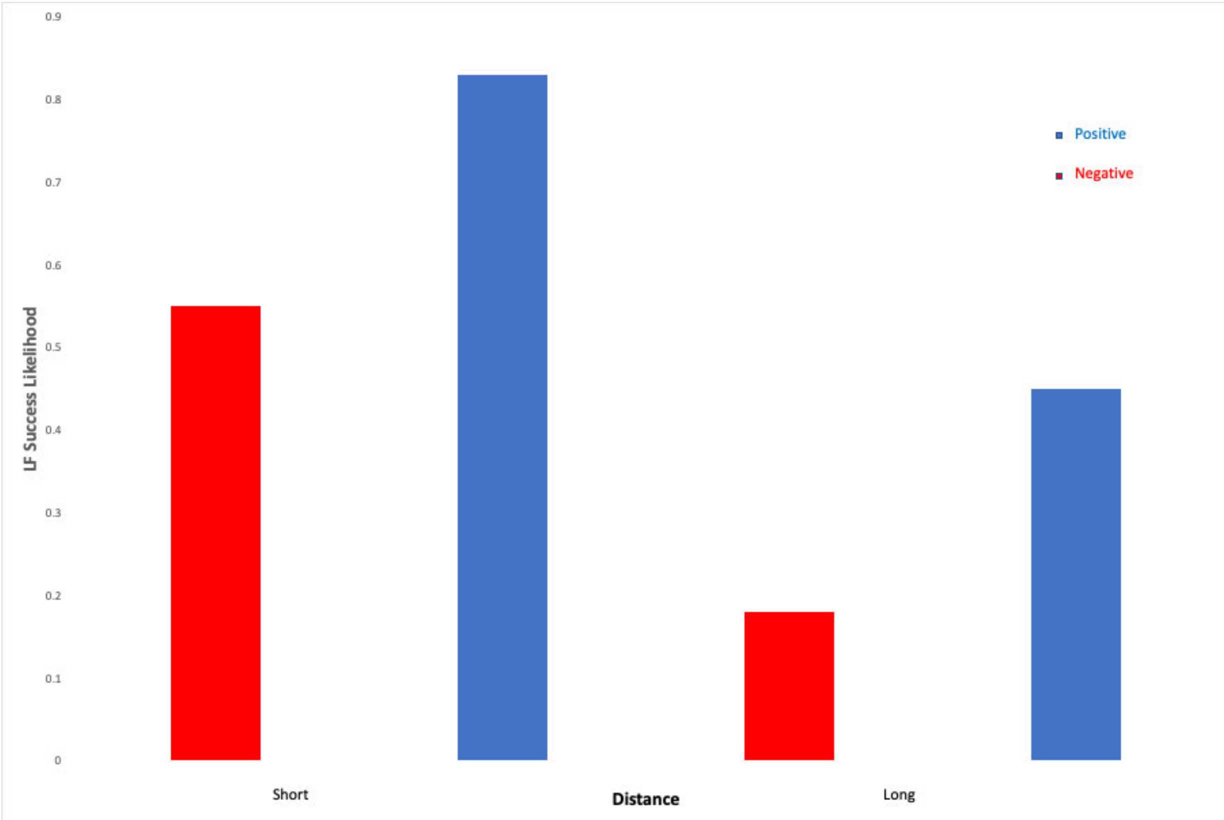

FIGURE 2. SHAP EXAMPLE FIGURE DEMONSTRATING LIKELIHOOD OF LF BY DISTANCE

Another aspect highlighted by the study is the differences between the regions. The information Fusion algorithm is a practical algorithm that uses multiple ML algorithms to combine their variable importance and give a combined ranking of variables. The variable comparison in Table 3 shows that the information fusion algorithm found that Destination regions had slightly more impact on LF than Origin regions. The LF already correlated with the number of flights. The likelihood of successful comparisons of the destination regions is shown in Fig. 3. The southeast (SE), northeast (NE), and central east (CE) regions have a high probability of LF success. These regions have a high population, economic value, and tourist movement that the aviation industry shaped around these destinations. Another finding is the difference between the rate of being a positive case and a negative case (P/N). N (2.3 P/N likelihood rate), NE (1.42 P/N likelihood rate), and CE (1.21 P/N likelihood rate) destined flights have a higher difference in being more likely to be positive. The P/N likelihood rate is highest for the North, but the number of flights is drastically lower in these regions, which minimizes the potential. The SE (1.07 P/N likelihood rate) has the most significant number of destinated flights. South, being 0.75, has the lowest P/N likelihood rate, showing the reconsideration of the load factor of that flight.

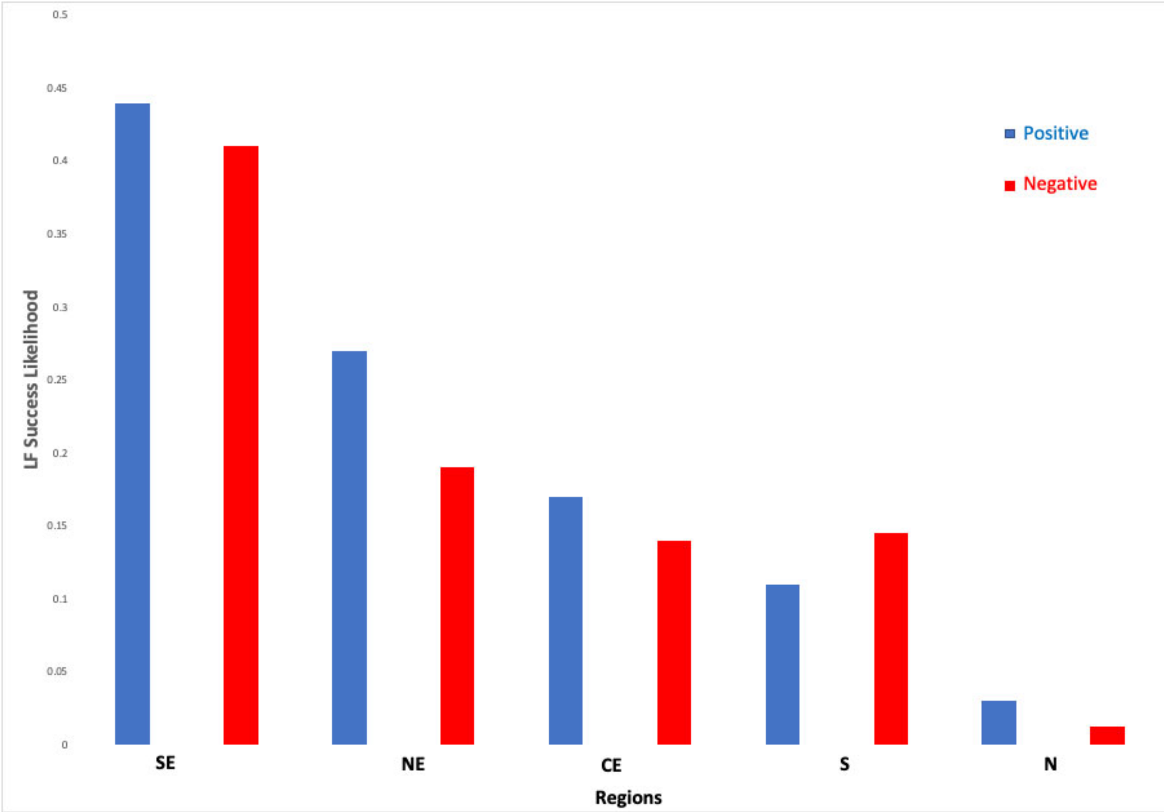

FIGURE 3. SHAP EXAMPLE FIGURE DEMONSTRATING THE LIKELIHOOD OF LF BY DESTINATION REGIONS

When we evaluate the likelihood of success by the origin states in Fig. 4, the states with high traffic, such as SP (1.23 P/N likelihood rate), DF (1.57 P/N likelihood rate), and BA (1.22 P/N likelihood rate), RJ (1.05 P/N likelihood rate), also have a high probability of success. On the other hand, even some states, such as MG (0.72 P/N likelihood rate), PR (0.69 P/N likelihood rate), and PA (0.35 P/N likelihood rate), have high traffic but are more likely to have negative cases. The flights in these states have spread to rural areas with relatively less frequent flights.

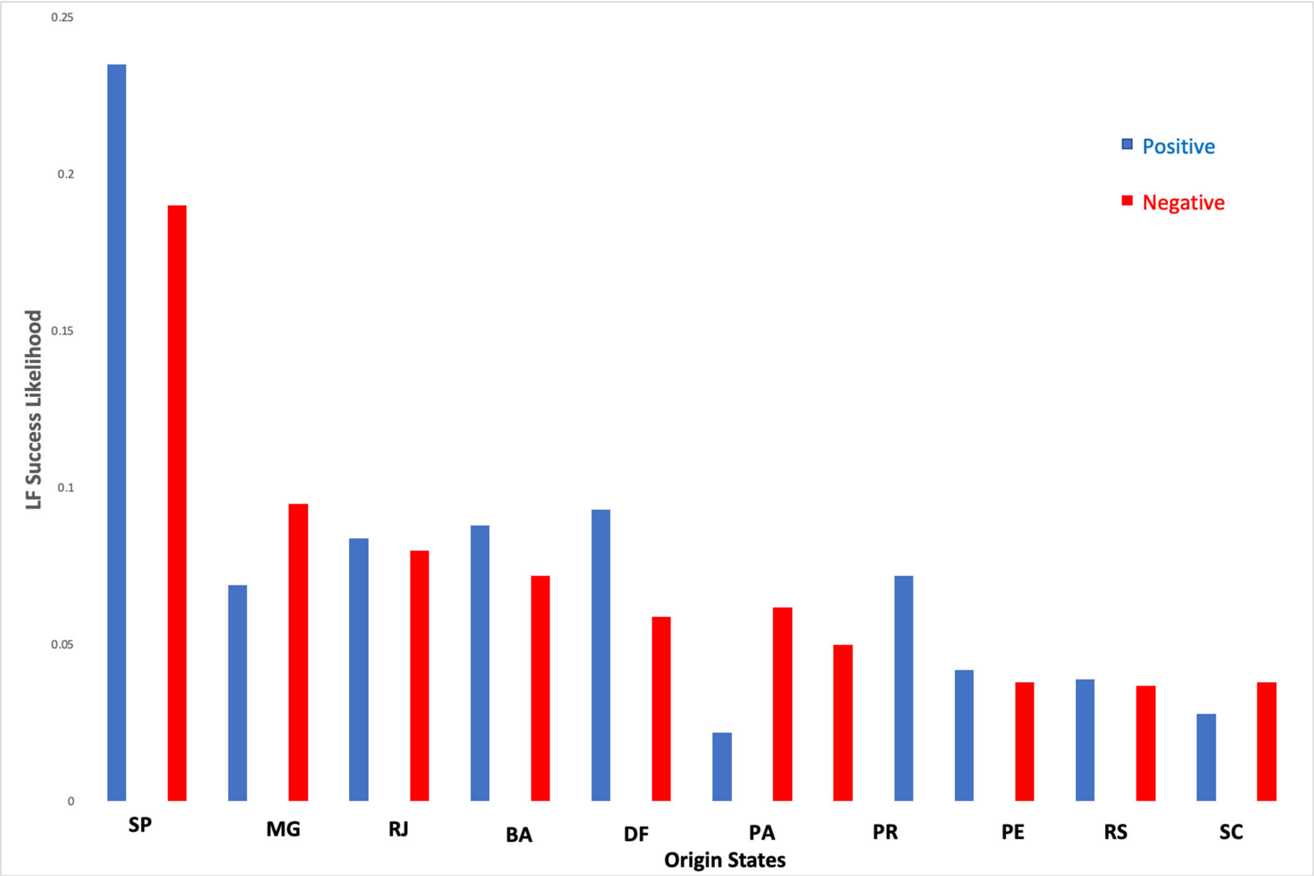

FIGURE 4. SHAP EXAMPLE FIGURE DEMONSTRATING THE LIKELIHOOD OF LF BY ORIGIN STATES

Another critical finding of our study is the difference in success likelihood between the operating companies, shown in Fig. 5. While some large companies have a higher likelihood of success, other large operators show less difference between the likelihood of being in positive and negative classes. These companies could utilize the SHAP explanation model in this study to dig deeper and find more prosperous routes visualized in Fig. 6.

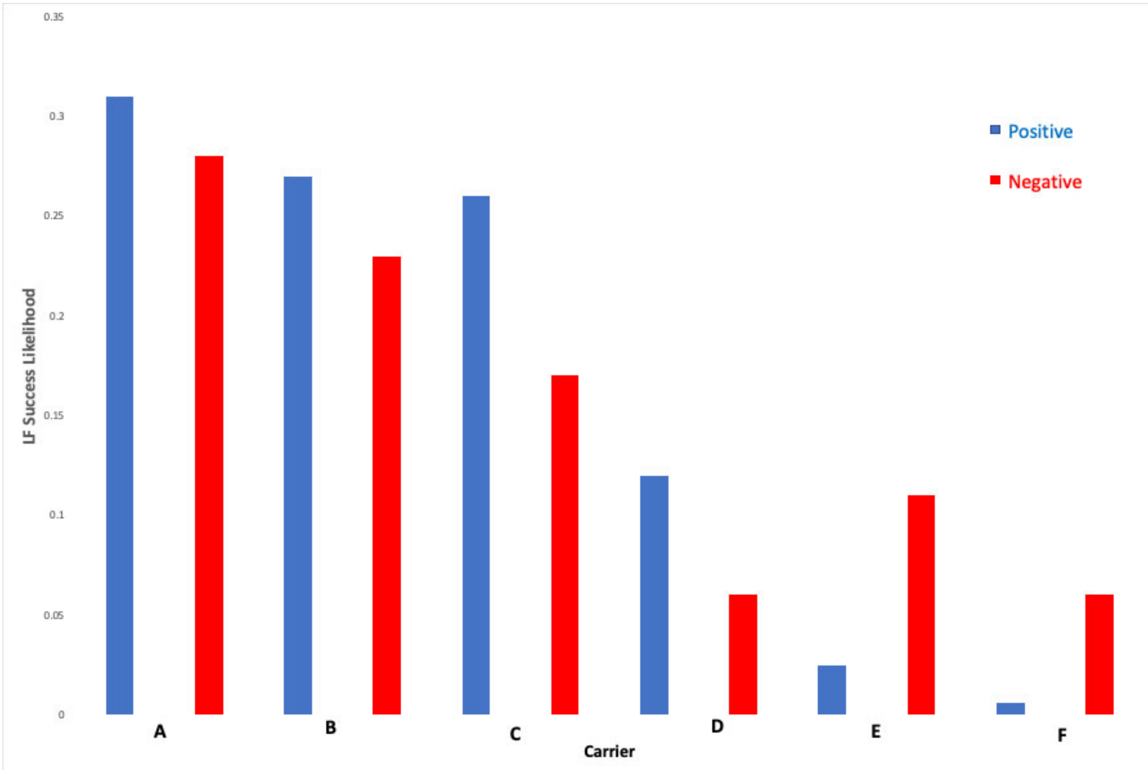

FIGURE 5. SHAP EXAMPLE FIGURE DEMONSTRATING LIKELIHOOD OF LF BY CARRIER

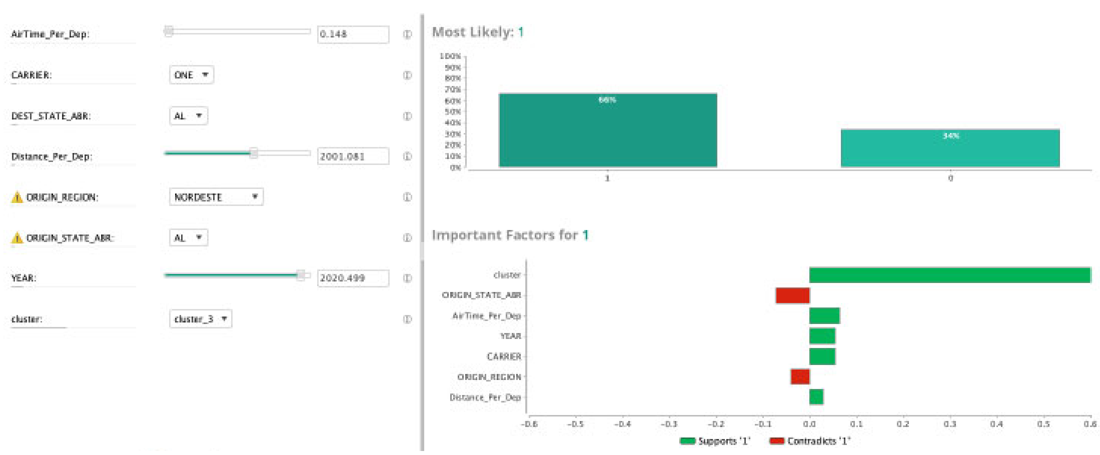

FIGURE 6. EXAMPLE FIGURE DEMONSTRATING SHAP SENSITIVITY ANALYSIS AND SHAP EXPLANATION

V. CONCLUSIONS

The study concludes that the flight distance positively correlates with the LF. This study implies the importance of flight distance and specific airlines, which are also related to airline loyalty and frequent flyer programs. As a business strategy, airlines should investigate the incentives they offer to these frequent flyers (Gao, 2020). When the flight length increases, people often choose to use other means of transportation for domestic travel. The destination region also significantly impacts the LF. In the decreasing likelihood of having a positive LF, Northeast, Central East, Southeast, South, and North are the areas with significant differences. The Brazilian aviation system is extensive and complex, and covering all factors associated with LF in this study would be impractical. Additional studies are recommended to support further inferences for business, as the data and algorithms presented may benefit the creation of business simulations and the understanding of such simulations. The business inference that longer flights have a 1.28 times higher probability of success than shorter flights is essential information for Brazil's airline industry. The inference for longer routes recommends that regional companies develop routes in different regions of the country. The overall airline industry worldwide also has correlating findings that companies lean toward higher-efficiency, longer-distance aircraft.

The likelihood of achieving a higher LF increases linearly over the years in all three models. This finding indicates that the overall LF is improving in the country's aerospace system. Increased LF tends to bring higher profits to companies and is expected to enhance competitiveness, resulting in higher operational efficiency and customer satisfaction.

The aircraft manufacturing sector's progress directly helps companies improve their LF. Recently, aircraft manufacturers have competed to produce more fuel-efficient aircraft that can fly mid-high ranges (Slocum, 2018). We classified flights longer than 1000 kilometers as long-distance flights, which can be accomplished with new-generation, fuel-efficient, high-efficiency engine aircraft. Long distances in this study had the highest LF. These new-generation, high-efficiency engine aircraft can accomplish these flights over 1000 kilometers, and we found that if these flights are operated by new-generation, high-efficiency engine aircraft, the LF on these flights will yield higher profits for the airlines. These new-generation aircraft, including domestically manufactured airplanes, could significantly enhance the earnings of Brazilian companies and regions. Additionally, these long-range regions are the most challenging lines to replace with another mode of transportation, such as ground or maritime. The current progress in the industry is expected to directly benefit the Brazilian aviation system.

The number of flights to a region is essential information; regions like the North have great potential for a higher P/N likelihood rate, but that also shows a limitation in the region's potential. Current high-flight traffic routes have already built the infrastructure and may help with achieving higher yields. The higher P/N likelihood rate is positive in high-frequency regions, while low-frequency areas often indicate the need for improvements in high-cost infrastructure for these regions that facilitate longer-term solutions to enhance flight frequency.

These differences among states could be due to a mismatch between the demand in that region and the fleet allocated to cater to that demand. There could be other motivating factors, such as higher ticket prices or government promotions, to foster trade in these areas. However, it is possible to forecast passenger demand in those regions and allocate an optimal scale of flights (Muroz Anguita and Díaz Olariaga, 2022). Flights to and from rural areas might be mixed cargo/passenger flights, which could rationalize the profitability of these flights.

The airline's name, which refers to the characteristics of that specific airline, is also important in defining success. Some airlines are generally very successful, while others are not in terms of the LF, but this does not directly correlate with overall business performance success. Each airline should be evaluated in-depth for its business model. Airlines A, B, C, and D are more likely to have favorable P/N rates. On the other hand, Airlines E and F are more likely to have a negative P/N rate. Companies with diverse loads, mixing freight with passengers, have a different LF than a low-cost airline. This P/N rate does not directly imply that these airlines are unprofitable. The profitability of an airline or airport is sometimes independent of LF. Variables such as ticket prices, operational costs, freight profitability, indirect costs, and indirect profits, along with the airport and airline business model, assess a flight's success. Local and federal governments influence the number of flights for various economic and political reasons. A flight might have a low LF but could have a very positive financial impact on the region or the airline. LF analysis needs to be integrated with other long-term business objectives and KPI analysis to make the ultimate decision regarding the success or failure of companies, airports, and regions. It is clear that some findings from the study are so complex that they can only be elucidated through the predictive and interpretable power of various ML models. Additional meta-studies that incorporate the interpretive power of other fields with superior ML algorithms can establish clearer relationships that benefit interested parties (Inan, 2022).

VI. LIMITATIONS AND FUTURE STUDIES

The original LF is a continuous variable; every company and route has its load factor goals. Changing the main parameters to increase LF, such as changing the aircraft type to serve that route, can take time and effort. It may take years to open a new route or even lease a new aircraft type to serve the suitable capacity for a particular route. Focusing on more than 0.8 LF in this study as the definition of success is a limitation. Companies can have dynamic LF goals and may change them gradually for different seasons. To benefit the predictive performance power of classification models, we choose a standard 0.8 threshold (Sun, Wandelt, and Zhang, 2022). Additional studies should focus on predicting low LF flights, and the findings of the studies should be combined to make a comprehensive commendation. In addition, LF is not the only KPI used to measure the success of companies. The operational costs and indirect benefits of flights may differ from the findings of the LF, so deeper reasoning is required to judge companies and routes.

Future studies could include more data about the economic activities of the routes. Additional data sources about the financial condition of the states and regions, patterns of tourist numbers, seasonality, and weather data would help improve the prediction quality and make the model more explainable. XAI methods are still exploratory methods, and the information from different models are not yet fully proven to be repeatable and robust. However, in this study, we aimed for some of the most transparent methods, such as Naïve Bayes, to have reliable business modeling.

An airline's success means more than LF, so additional prescriptive studies could be done to prepare a comparative KPI to measure the success of the regions and companies. Additional SHAP explanation scenarios could run to measure the success and, more importantly, understand the patterns that create success and fault lines. Also, other XAI methods, such as ICE and LIME, may be applied to the research problem, and their results could be benchmarked against each other to prove the expansibility quality, judging them using aviation SME. We believe the XAI field will be more mature shortly, and we will be able to benchmark more algorithms and start talking more about causalities.

REFERENCES

Agência Nacional de Aviação Civil(ANAC). “Metadados dos Dados Estatísticos do Transporte Aéreo. https://www.anac.gov.br/acesso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-estatisticos-do-transporte-aereo/48-dados-estatisticos-do-transporte-aereo, 2022, (last accessed June 20, 2023)

Alauddin, M. and Ting, C.Y., 2020. Digital click stream data for airline seat sale prediction using GBT. *International Journal of Engineering Trends and Technology*, pp.24-31.

Altringer, L., Navin, J., Begier, M.J., Shwiff, S.A. and Anderson, A., 2021. Estimating wildlife strike costs at US airports: A machine learning approach. *Transportation Research Part D: Transport and Environment*, *97*, p.102907.

An, J., and Lee, S., Application of machine learning algorithms for demand forecasting in the aviation industry. *Journal of Airline Management*, 36(2), 2021, 45-59.

Anderson, C., Curtis, T., Moss, R. F., Smith, M. O., & Mey, G., "Impact of US Low Cost Carriers Entrance to the Airline Market in Colombia, South America", *International Journal of Aviation, Aeronautics, and Aerospace* 4, no. 1 (2017): 1.

Aslam, N., Khan, I. U., Mirza, S., AlOwayed, A., Anis, F. M., Aljuaid, R. M., & Baageel, R., "Interpretable Machine Learning Models for Malicious Domains Detection Using Explainable Artificial Intelligence (XAI)", *Sustainability* 14, no. 12 (2022): 7375.

Baklacioglu, T., 2016. Modeling the fuel flow-rate of transport aircraft during flight phases using genetic algorithm-optimized neural networks. Aerospace Science and Technology, 49, pp.52-62.

Baldanza, Ben. "*Measuring airline profitability*", Handbook of Airline Operations, 1999, 147-159.

Banghart, Marc., "Identification of reverse engineering candidates utilizing machine learning and aircraft cannibalization data", *International Journal of Aviation, Aeronautics, and Aerospace* 4, no. 4 (2017): 5.

Barros, C. P., Liang Q.B., and Peypoch N., "The technical efficiency of US Airlines." *Transportation Research Part A: Policy and Practice* 50 (2013): 139-148.

Basora, L., Bry, P., Olive, X. and Freeman, F., 2021. Aircraft fleet health monitoring with anomaly detection techniques. *Aerospace*, *8*(4), p.103.

Belobaba, P., "Overview of airline economics, markets and demand", *The global airline industry* (2009): 47-72.

Bengio, Y., "Learning deep architectures for AI", *Foundations and trends® in Machine Learning* 2, no. 1 (2009): 1-127.

Candel,A.,Parmar, V.,LeDell, E., and Arora, A. “*Deep Learning with H2O*”, Apr 2023, http://h2o.ai/resources.

Cankaya, B., Tokgoz, B. E., Dag, A., and Santosh, K. C., "Development of a machine-learning-based decision support mechanism for predicting chemical tanker cleaning activity" *Journal of Modelling in Management* 16, no. 4 (2021): 1138-1165.

Cankaya, B., Topuz, K., and Glassman, A. M., “Business Inferences and Risk Modeling with Machine Learning; The Case of Aviation Incidents” *Proceedings of HICSS*, 2023[a], *11*, 1238.

Cankaya, B., Topuz, K., Delen, D., and Glassman, A., “Evidence-Based Managerial Decision-Making with Machine Learning: The Case of

Bayesian Inference in Aviation Incidents", *Omega*, 2023[b], 102906.

Chen, S., and Li, J., "Load factor and customer satisfaction in the airline industry" *Journal of Supply Chain and Operations Management*, 2020, 18(3), 315-327.

Choi, K., "Multi-period efficiency and productivity changes in US domestic airlines." *Journal of Air Transport Management* 59 (2017): 18-25.

Choi, W., Mary B. O'Connor, and Truong, D., "Predicting the U.S. Airline Operating Profitability using Machine Learning Algorithms", *International Journal of Aviation, Aeronautics and Aerospace,* 2019, 6(5), 14.

Daft, J., and Albers, S., "Profitability analysis of low-cost long-haul flight operations", *Journal of Air Transport Management*, 2012, 19, 49-54.

Etani, N., 2019. Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data. *Journal of big data*, 6(1), p.85.

Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R. and Lin, C.J., "LIBLINEAR: A library for large linear classification", *the Journal of machine Learning research*, 2008, *9*, pp.1871-1874.

Gao, Y., "A conceptual framework for valuating airline frequent flyer program miles. *International Journal of Aviation, Aeronautics and Aerospace",* 2020, *7*(1), 11.

Graefe, A., Armstrong, J.S., Jones, R.J. and Cuz an, A.G., "Combining forecasts: an application to elections", *International Journal of Forecasting*, 2014, Vol. 30 No. 1, pp. 43-54.

İnan, Tolga, T., "Classification of Survivor/Non-Survivor Passengers in Fatal Aviation Accidents: A Machine Learning Approach", *International Journal of Aviation, Aeronautics, and Aerospace*, 2022, *9*(1), 8.

Ivaldi, M., Petrova, M., and Urdanoz, M., "Airline cooperation effects on airfare distribution: An auction-model-based approach", *Transport Policy*, 2022, 115, 239-250.

Johnson, R., and Thompson, A., "The impact of load factor on airline profitability: Evidence from US airlines", *Journal of Supply Chain and Operations Management*, 2018, 16(2), 127-138.

Joo, S. J., and Fowler K., "Exploring comparative efficiency and determinants of efficiency for major world airlines." *Benchmarking: An International Journal* 21, no. 4 (2014): 675-687.

Kapoor, A., Horvitz, Z., Laube, S. and Horvitz, E., 2014, April. Airplanes aloft as a sensor network for wind forecasting. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks* (pp. 25-33). IEEE.

Krishna, R., Sariki, T.P., Sigamsetty, S.M., Kumar, G.B. and Benil, T., 2023, December. Aviation Safe Landing Prediction Utilizing Long-Short Term Memory Algorithm. In *2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC)* (pp. 562-568). IEEE.

Li, S., Qin, J., He, M. and Paoli, R., 2020. Fast evaluation of aircraft icing severity using machine learning based on XGBoost. *Aerospace*, *7*(4), p.36.

Lundberg, S. M., and Lee, S. I., "A unified approach to interpreting model predictions", *Advances in neural information processing systems*, *30*.

Lozano, S., and Gutiérrez E., "A multiobjective approach to fleet, fuel and operating cost efficiency of European airlines." *Computers & Industrial Engineering* 61, no. 3 (2011): 473-481.

Malatesta, W.A. and Yang, B., 2021. Aviation turbine fuel thermal conductivity: A predictive approach using entropy scaling-guided machine learning with experimentalvalidation. *ACS omega*, *6*(43), pp.28579-28586.

Marazzo, M., Scherre, R., and Fernandes, E., "Air transport demand and economic growth in Brazil: A time series analysis", *Transportation Research Part E: Logistics and Transportation Review*, 2010, 46(2), 261-269.

Matuszczak, M., Żbikowski, M. and Teodorczyk, A., 2021. Predictive modelling of turbofan engine components condition using machine and deep learning methods. *Eksploatacja i Niezawodność*, 23(2), pp.359-370.

McAfee, R. P., and Te Velde, V., "Dynamic pricing in the airline industry", *Handbook on Economics and Information Systems*, 2006, 1, 527-67.

Midtfjord, A.D., De Bin, R. and Huseby, A.B., 2022. A decision support system for safer airplane landings: Predicting runway conditions using XGBoost and explainable AI. *Cold Regions Science and Technology*, *199*, p.103556.

Muros Anguita, J. G., and Díaz Olariaga, O., "Air passenger demand forecast through the use of Artificial Neural Network algorithms", *International Journal of Aviation, Aeronautics, and Aerospace*, 2022, *9*(4), 2.

Reitmann, S. and Schultz, M., 2022. An adaptive framework for optimization and prediction of air traffic management (sub-) systems with machine learning. *Aerospace*, *9*(2), p.77.

Ridgeway, G., Madigan, D., Richardson, T., and O'Kane, J., "Interpretable boosted naïve bayes classification", *In KDD*, 1998, pp. 101-104.

Rinehart, D., Smith, P., and Spencer, A., "SME-Defined Scenarios for Autonomy (SDSA): A method for exploring complex aviation system safety and performance", *IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*, 2014, pp. 5B2-1. IEEE.

Roberts, D., and Griffith, J. C., "A Tale of Two Airlines: A Comparative Case Study of High-Road versus Low-Road Strategies in Customer Service and Reputation Management", *International Journal of Aviation, Aeronautics and Aerospace,* 2019, *6*(2), 4.

Sakthidharan, V., and Sivaraman S., "Impact of operating cost components on airline efficiency in India: A DEA approach." *Asia Pacific Management Review* 23, no. 4 (2018): 258-267.

Schultz, M. and Reitmann, S., 2018. Consideration of passenger interactions for the prediction of aircraft boarding time. *Aerospace*, *5*(4), p.101.

Sheppard, J. M., and Young, W. B., "Agility literature review: Classifications, training, and testing", *Journal of Sports Sciences*, 2006, 24(9), 919-932.

Sim, S., Im, J., Park, S., Park, H., Ahn, M.H. and Chan, P.W., 2018. Icing detection over East Asia from geostationary satellite data using machine learning approaches. *Remote Sensing*, *10*(4), p.631.

Simsek, S., Dag, A., Tiahrt, T., and Oztekin, A., "A Bayesian Belief Network-based probabilistic mechanism to determine patient no-show risk categories", *Omega*, 2021, 100, 102296.

Singh, J., Sharma K. S., and Srivastava R., "What drives Indian Airlines operational expense: An econometric model." *Journal of Air Transport Management* 77 (2019): 32-38.

Sinha, D. , "Safety, profitability, and the load factor for airlines in the USA", *Economics Bulletin*, 2007, 12(6), 1-7.

Slocum, C., "How low-cost carriers are changing the growing East Asian travel market", *International Journal of Aviation, Aeronautics and Aerospace, 5*(5), 2018, 12.

Smith, M., and Tesone, D. V., "Optimizing load factor for airlines: An empirical investigation", *Journal of Air Transport Management*, 2019, 77, 73-81.

Sun, X., Wandelt, S., and Zhang, A., "Ghostbusters: Hunting abnormal flights in Europe during COVID-19", *Transport Policy*, 2022, *127*, 203-217.

Szabo, S., Mako, S., Tobisova, A., Hanak, P., and Pilat, M., "Effect of the load factor on the ticket price", *Transport problems*, 2018, 13.

Štrumbelj, E., and Kononenko, I., "Explaining prediction models and individual predictions with feature contributions" *Knowledge and Information Systems*, 2014, *41*(3), 647–665.

Tian, Y., Huang, W., Ye, B. and Yang, M., 2019. A new air quality prediction framework for airports developed with a hybrid supervised learning method. *Discrete Dynamics in Nature and Society*, *2019*.

Torra, V., *Trends in information fusion in data mining* (pp. 1-6). Springer Berlin Heidelberg, 2003.

Wang, Z., Liang, M. and Delahaye, D., 2018. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. *Transportation Research Part C: Emerging Technologies*, *95*, pp.280-294.

Wang, X., Tang, L., Tang, Z., Yang, Z., and Li, X., "A machine learning approach for flight delay prediction and optimization", *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(2), 868-880.

Wan, J., Zhang, H., Lyu, W. and Zhou, J., 2022. A Novel Combined Model for Short-Term Emission Prediction of Airspace Flights Based on Machine Learning: A Case Study of China. *Sustainability*, *14*(7), p.4017.

Wirth, R., and Hipp, J., "CRISP-DM: Towards a standard process model for data mining", *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, Vol. 1, pp. 29-39.

Wozny, F., 2022. The impact of covid-19 on airfares—a machine learning counterfactual analysis. *Econometrics*, *10*(1), p.8.

Zheng, X. S., and Rutherford, D., "*US-Latin America Airline Fuel Efficiency Ranking, 2017-2018*", https://theicct.org/wp-content/uploads/2021/06/ICCT_LatAm-airline-ranking_20191120.pdf, 2019, (accessed June 20, 2023).

Zhu, X. and Li, L., 2021. Flight time prediction for fuel loading decisions with a deep learning approach. *Transportation Research Part C: Emerging Technologies*, *128*, p.103179.