Data Analytics for Air Travel Data: A Survey and New Perspectives

HAIMAN TIAN and MARIA PRESA-REYES, Florida International University
YUDONG TAO, University of Miami
TIANYI WANG, Florida International University
SAMIRA POUYANFAR, Microsoft
MIGUEL ALONSO JR. and STEVEN LUIS, Florida International University
MEI-LING SHYU, University of Miami
SHU-CHING CHEN and SUNDARAJA SITHARAMA IYENGAR, Florida International University

From the start, the airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers. Consequently, this has ushered in substantial economic growth, both nationally and internationally. The airline industry produces vast amounts of data, capturing a diverse set of information about their operations, including data related to passengers, freight, flights, and much more. Analyzing air travel data can advance the understanding of airline market dynamics, allowing companies to provide customized, efficient, and safe transportation services. Due to big data challenges in such a complex environment, the benefits of drawing insights from the air travel data in the airline industry have not yet been fully explored. This article aims to survey various components and corresponding proposed data analysis methodologies that have been identified as essential to the inner workings of the airline industry. We introduce existing data sources commonly used in the papers surveyed and summarize their availability. Finally, we discuss several potential research directions to better harness airline data in the future. We anticipate this study to be used as a comprehensive reference for both members of the airline industry and academic scholars with an interest in airline research.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Applied computing** → *Economics*; Transportation;

Additional Key Words and Phrases: Airline, revenue management, big data

ACM Reference format:
Haiman Tian, Maria Presa-Reyes, Yudong Tao, Tianyi Wang, Samira Pouyanfar, Miguel Alonso Jr., Steven Luis, Mei-Ling Shyu, Shu-Ching Chen, and Sundaraja Sitharama Iyengar. 2021. Data Analytics for Air Travel Data: A Survey and New Perspectives. *ACM Comput. Surv.* 54, 8, Article 167 (October 2021), 35 pages.
https://doi.org/10.1145/3469028

This research is partially supported by NSF CNS-1461926 and NSF CNS-1952089.
Authors' addresses: H. Tian, M. P. Reyes, T. Wang, M. Alonso Jr., S. Luis, S.-C. Chen, and S. S. Iyengar, Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL 33199; emails: {htian005, mpres029, wtian002, malonsoj, luiss, chens, iyengar}@cs.fiu.edu; Y. Tao and M.-L. Shyu, Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124; email: {yxt128, shyu}@miami.edu; S. Pouyanfar, Microsoft, One Microsoft Way, Redmond, WA 98052; email: sapouyan@microsoft.com.

0360-0300/2021/10-ART167 $15.00
https://doi.org/10.1145/3469028

1 INTRODUCTION

The structure of the airline industry experienced significant changes after the Airline Deregulation Act of 1978, which removed the U.S. federal government control over many areas and sparked a renewed interest among economists in the determinants of airfares for individual market pairs. Consequently, the response to this deregulation was the establishment of a free market within the commercial airline industry. A similar deregulation process also happened to E.U. airlines but led to much smaller changes [134]. Beginning in 1990, several E.U. air transport liberalization and deregulation packages entered into effect. The third deregulation package, going into effect in 1993, removed most of the barriers that limited the E.U. airlines to access all air transport routes with unrestricted rate-making. Aside from aiming to improve the customer experience, the continuously growing business challenges still center on keeping up with market changes to further increase profits and productivity. Notably, studies from both economic and operations research contribute to exploring the connection between airfares and market-specific measures of demand, cost, and competition [8, 42].

Incorporating cutting-edge technology promotes an airline's path toward realizing total revenue optimization. For example, **Revenue Management (RM)** was first introduced into the airline industry in the 1970s, which promoted the use of differentiated pricing—charging passengers different fares for multiple booking classes [88]. Traditional RM derives marketing and pricing strategies based on economic theory and is popularly applied and developed in many industries to design the appropriate products that satisfy the customers' interests. An early example is the well-known static optimization model, which determines the booking limits only once, proposed by Belobaba in 1987 [17]. Furthermore, **Origin-Destination (O-D)** flow control prompted airlines to develop advanced approaches after understanding the basic RM controls, since fare class mix revenue gain came from the dynamic revision of booking limits. Dynamic revision often relies on human intervention, which is essential to account for unusual surges in demand due to special events.

Meanwhile, the airline industry is facing a complex and dynamic environment from the emergence of cost-competitive carriers vying for market share [24, 41]. For example, Southwest Airlines, the world's largest **Low-Cost Carrier (LCC)**, has been shown to dramatically increase the traffic and reduce the average fares at airports that the airline serves, as well as the new markets that it enters [163]. According to recent findings, LCCs control more than 24 percent of the entire market share; this change has revealed how customers' demands and expectations for cheaper fares are contiguously growing [24].

In 2007, a pioneering innovation known as ancillary pricing—revenue made from goods and services considered as secondary options to a company's primary product—was introduced. Studies focused on trends and impacts of ancillary revenue will often cite the case of Ryanair, a leading European LCC, as a catalyst of unbundled services and lower base fares [172]. The introduction of ancillaries revolutionized traditional RM. However, there is still much room for improvement to handle new challenges being ushered in by the internet era.

To directly compete with LCCs, **Full-Service Carriers (FSCs)** are urged to incorporate more innovative strategies based on understanding, predicting, and influencing customer behavior to further maximize its revenue and profits [58]. FSCs not only need to provide low fares to attract passengers, but they also need to keep regular fares to cover direct **Operating Costs (OCs)**. OCs are typically very high and volatile, and are often influenced by factors such as fuel, labor, and equipment maintenance, making RM increasingly complicated in such a dynamic cost environment. Therefore, bringing cutting-edge techniques to this industry is necessary to effectively and efficiently solve these issues and provide strategies to assist the decision-making process.

RM goes beyond inventory control, requiring a mix of various authorities and skills. Data analytics using **Artificial Intelligence (AI)** is the next step in the evolution of RM and its application to new areas, and it brings with it a set of **Machine Learning (ML)** technologies that can surpass traditional strategies. In the future, the airline industry needs to understand their passengers' requirements and make decisions that support other facets of the air travel experience, such as retailing and merchandising. These new requirements are going to be addressed by utilizing large volumes of air travel data that are generated from various components, with advanced technology ensuring robust analytics and providing more accurate and easy-to-understand information in real-time. Airlines can then leverage useful domain-specific knowledge, deliver causal analysis and actionable insights to all the stakeholders in the ecosystem.

Starting in 2000, applications of ML techniques, an essential part of data analytics, began to arise in the airline industry. ML applications promote more robust models and overcome the shortcomings existing in traditional approaches. For example, one approach is to use ML for creating pricing strategies across multiple market segments simultaneously. Then, traditional RM can be used to select prices that will be accepted for each product by responding to the demands of all the given markets, with many of these decisions made in real-time. Another example is the application of **Reinforcement Learning (RL)**, one of the most popular ML approaches. RL solutions begin by modeling the problem as a sequence of actions that yield a reward, depending on the outcome of said action. By modeling pricing strategies in this way, RL can provide more opportunities to generate optimal policies that were not achievable before. As the airline industry searches for more opportunities that will potentially boost revenue, more advanced and emerging ML techniques such as rule-based learning, tree-based learning, and deep learning will start to attract more attention.

While investigating the application of ML in the industry, two interesting questions arose: Are academic scholars aware of the recent changes in the airline industry? Similarly, which academic domains are dedicated to addressing long-standing business practice? To gain insight on these questions, an academic citation searching result from 1990 to 2017 was performed using the Web of Science. The results were generated by using the search term "airline" and limiting the occurrence of the term in either the paper's title, abstract, or list of keywords. The papers that were returned as a result of this query were divided into two categories: (1) traditional airline papers and (2) airline papers referencing **Computer Science (CS)** papers. Each category was counted separately. The results are shown in Figure 1. Upon closer inspection, very few papers in CS are referenced by airline researchers during the 1990s. However, CS papers have been increasingly cited over the last decade and are recently contributing equally or more than operation research papers.

Form 41 Finance Data from the Office of Airline Information of the **Bureau of Transportation and Statistics (BTS)** contains quarterly reported financial information on the largest certified U.S. air carriers. Figure 2 shows the FSCs that obtained the top-three annual operating revenues generated by domestic flights from 1990 to 2017. In the last year, all three airlines obtained over $23 billion in total revenue. Delta Airlines reached almost $29.7 billion in revenue. Three large spikes in revenue can be observed during 2009–2010, 2011–2012, and 2014–2016. This has been attributed to the mergers and acquisitions that happened within each airline [132]. Though the airline industry has experienced dramatic changes in recent decades, the outcomes show that the operating revenues have stayed stable, experiencing only moderate increases every year, even after the economic crisis of 2008. Thus, there is still room for improvement.

With a history of great successes in multiple disciplinary fields, AI is one of the hottest research directions in both academia and industry. This survey focuses on an overview of how data analytics have been applied in the airline industry for RM from different perspectives, including challenges, techniques, available data sources, and possible opportunities. To the extent of our knowledge,

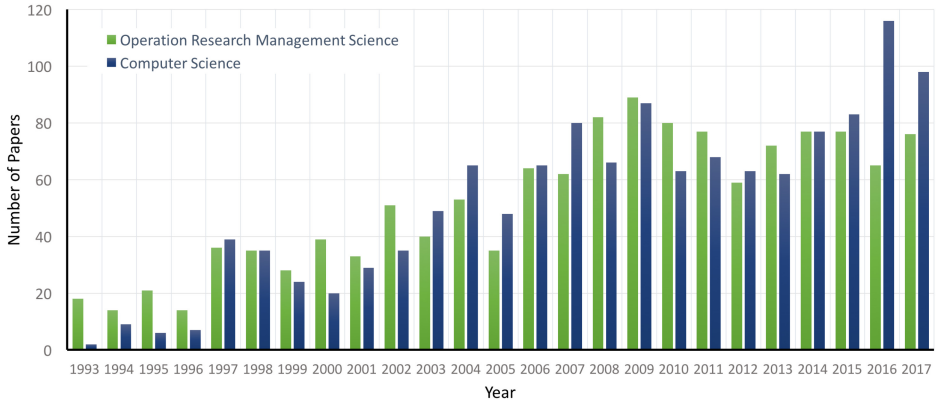

Fig. 1. Airline research papers that reference CS research papers from 1990–2017; source: Web of Science.

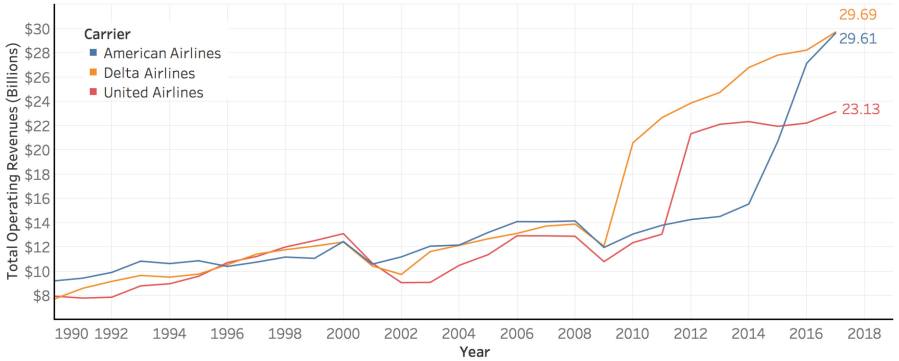

Fig. 2. Total operating revenues of the top-three FSCs from 1990 to 2017.

this article is the first comprehensive review on academic works that focuses on analytics using air travel data from two different points of view within the airline industry—the service provider and the customer. Other areas that employ air travel data for non-RM purposes are also included. Compared to previous academic surveys with a focus on a particular scope of RM [17, 65], this article distinguishes itself through its focus on different aspects of the airline industry by reviewing the existing work, research and applications, as well as presenting the authors' experience in this area. The covered topics are summarized in Table 1.

The rest of this article is organized as follows: In Section 2, the major problems addressed by the surveyed papers are briefly presented. Section 3 discusses various data analytics approaches and ML techniques that have been used in the airline industry to complement traditional statistical analytics. Since the airlines' data is mostly private and contain a lot of non-disclosure information, many public datasets and data sources are discussed in Section 4. As advanced ML methods and applications in the airline industry are still relatively new, Section 5 points out current challenges and potential directions for future research. Section 6 provides a conclusion to this article.

2 THE AIRLINE RESEARCH COMPONENTS

The research topics that were identified through the literature survey of airline papers are briefly described in Table 2. Taken individually, the topics are independent components. However, taken

Table 1. Topics Identified Regarding Airline Industry and Air Travel

Section #	Topics	Descriptions
2.1	Ticket Pricing	Finds correlation between the potential factors and the fare segments
2.2	Inventory Control	Discovers the optimal seat booking for each class to maximize airline revenue
2.3	Overbooking	Selects appropriate overbooking rate to mitigate the effects of cancellation
2.4	Demand Prediction	Forecasts the demand for air travel and its correlation with pricing
2.5	Simulation Tools	Evaluate and compare different RM models before deploying them into practice
2.6	Ancillary Pricing	Makes revenue by providing extra comfort
2.7	Service Improvement and Customer Experience	Capture and understand the customers' needs to gain the ability to be adaptive to the changing economic environment
2.8	Price Mining for Strategic Customers	Aims to minimize the cost of purchasing a flight ticket or constructing an itinerary
2.9	Connectivity	Studies the interactions among nodes in the air transport network by measuring their relationship
2.10	Air Traffic Management	Improves the accuracy, safety, and efficiency of air traffic control to support decision-makings for aircraft and airport operations

Table 2. Summary of Papers Categorized as Either Traditional or ML and Grouped by Publication Year Range

Year Range	Papers Using Traditional Methods	Papers Using ML Methods
Before 2000	Inventory Control: [17, 42, 87, 89, 97, 149, 159] Overbooking: [8, 16, 84, 130, 131, 137, 156] Demand Prediction: [76, 150] Service Improvement: [26], Simulation Tools: [171]	Demand Prediction: [7, 72] Service Improvement: [178]
2001-2006	Ticket Pricing: [29], Inventory Control: [103] Demand Prediction: [160] Simulation Tools: [32, 57, 141, 152, 153], Air Traffic Management: [62, 83]	Ticket Pricing: [52], Inventory Control: [67] Demand Prediction: [21, 170] Ancillary Pricing: [116]
2007-2012	Inventory Control: [119], Ancillary Pricing: [48], Air Traffic Management: [101]	Demand Prediction: [20, 37], Service Improvement: [50, 80, 151], Air Traffic Management: [14, 126, 180]
2013-Now	Ticket Pricing: [22], Inventory Control: [144] Overbooking: [179], Simulation Tools: [49] Service Improvement: [78, 102] Ancillary Pricing: [115, 145, 164] Air Traffic Management: [11, 104, 121]	Ticket Pricing: [45, 68, 110] Demand Prediction: [9, 98, 105] Ancillary Pricing: [39, 56, 111, 135, 158] Service Improvement: [3, 85, 138, 142, 161, 165, 176] Price Mining: [77, 107] Air Traffic Management: [5, 12, 13, 15, 33, 36, 38, 53, 61], [73, 79, 82, 94, 117, 118, 128, 140, 157, 166–168, 174, 181]

as a whole, the topics contribute to the improvement of the airline industry and have resulted in new profits. The most cited publications for each topic is also discussed.

2.1 Ticket Pricing

Airline ticket pricing is a strategy that determines the number and type of fare products offered in different markets. The price for each product is set precisely, often using different sets of restrictions, resulting in limits on how many seats to sell at what prices. Since its beginnings, solutions to the airline pricing game have tended to focus on simplistic models, considering that theoretical game models are more complicated to solve [51]. Airline companies have the potential to increase their total revenue by applying different restriction levels to offer multi-fare class products. According to Reference [29], typical dynamic pricing products have three characteristics: (1) The products have fixed order and quantity; (2) The sales will end with a certain deadline; (3) The marginal cost of selling one additional item is low. As a classic example of the dynamic pricing game, ticket pricing has been a favorite topic in a variety of research societies for several decades. Traditional RM covers broader scientific domains, such as operation science, business, economics, engineering, and statistical analysis. More revenue can be achieved by managing to reach the

optimized pricing options. Those pricing strategies are designed based on service and restrictions controlled by a limited inventory.

Since the emergence of **online travel agents (OTAs)** and LCCs, airlines have been forced to take competitive awareness into ticket pricing, which not only includes many aspects related to the ticket itself, but also incorporates inventory control, customer behavior models, and other competitive attributes and parameters [127]. Also, maintaining a healthy operation with an expectation of improved economic benefit is challenging for a high-cost industry. Therefore, lowering the ticket price is not an easy decision, since determining pricing without restriction (either too high or too low) might result in a loss of customers or negative income. Sometimes, letting the flight depart with empty seats might be the better option.

Airlines use many strategies to manage their decision-making. However, a complete list of strategies and policies used by all airlines does not exist. Identifying customer and market segmentation groups are essential for better decision-making and yield management. Among different clustering techniques in ML, partitioning algorithm (centroid based clustering) has been widely used in the airline industry for segmentation. In Reference [123], the authors suggested that customers of airlines can be categorized by the ticket types they regularly buy or by their travel purpose (i.e., leisure or business travel). Using classical clustering such as K-means, they defined six customer segments based on travel behavior features. Examples of these segments include "medium amount of trips," "few weekend flights," "no return flights," and so on. Non-linear regression models and clustering analysis were leveraged in Reference [114] to conceptualize the pricing discount strategies used by airlines. The authors first selected several normalized variables, such as days before flight, trip time, capacity of the plane, and so on, to represent seat characteristics. Then, a four-cluster solution is identified by running a cluster analysis on those variables to discover the pricing strategies. These four clusters allow for a separate strategy for each group of fares from the least to the most expensive. In the end, an ordered logit model is built to describe the significant characteristics that contribute to the final discount fare. In general, K-Means is a simple and efficient clustering technique that is applicable and suitable for large-scale airline data. However, it needs prior knowledge about K (number of clusters) and is highly sensitive to outliers. Different from partitioning algorithms, hierarchical clustering is more informative, and it is easier to detect the number of segments by looking at the tree diagrams. This characteristic is important in different airline applications in which the number of clusters is unknown. Dai et al. [43] conducted a hierarchical clustering for market segmentation related to the survivability of the airline operators. The study segments the operators into seven clusters including "Local Inefficient," "Short Haul," "Small Efficient," "Domestic Efficient," "No Domestic," "Domestic Long," and "International Long." It also examined the relation of the load factor as the key parameter for discriminating between these segments to the survival of each operator cluster. However, the high time complexity of hierarchical clustering makes it unsuitable for large-scale data. In 2015, Piggott [122] applied a number of clustering algorithms to discover market segments and potential business passengers, including K-means, X-means (an alternative to K-means), **Expectation-Maximization (EM)**, and Hierarchical clustering. The results showed the effectiveness of the aforementioned clustering techniques for passenger segmentation and market analysis. In particular, EM algorithm generates the best clustering results in that study.

Beginning in the early 2000s, ML approaches, such as regression analysis and RL, have begun to provide analytical support for developing advanced pricing policies [45]. In some cases, simple **Linear Regression (LR)** may be inadequate when the relationship between the response and explanatory variables is complex and can not be explained by linear relationships. Thus, **Generalized Linear Models (GLMs)** have been applied to better analyze and describe airline industry data exhibiting non-linearity. According to Mumbower et al. [110], the presence of endogeneity

causes the final estimation to be biased. A **Two-Stage Least Squares (2SLS)** linear regression model was implemented to correct the limitation in **Ordinary Least Squares (OLS)** regression model to estimate the price elasticity. The flight, booking, and sale characteristics, as well as the competitor promotions, are considered to affect the 2SLS model when calculating price elasticity. While it has been shown that 2SLS can predict the amount of bookings for flights booked in advance when explanatory factors such as departure date and market are taken into account, their research is restricted to a specific sample size with over a quarter of price and demand details lacking.

RL techniques benefit by including an additional dimension of intuition into the pricing game to solve issues that were unsolvable by conventional strategies. In 2012, Collins and Thomas conducted a comparison work that used a dynamic airline pricing game as an example and examined different types of RL algorithms (i.e., Q-Learning, **State-Action-Reward-State-Action (SARSA)**, and Monte Carlo Learning) that work for an adaptable sequential airline pricing game [40]. In their experiments, both Q-learning and SARSA outperformed the Monte Carlo Learning approach. The extra dimensions that were added to the model while learning have the potential to aid the decision-making process, though it could only be used to understand the problem underlying the designed model. In 2013, they investigated how RL, especially SARSA, was utilized when considering complex customer behaviors [41]. While using SARSA, three new aspects regarding customer modeling were discussed: customer demand, customer choice, and the size of the market. The paper proved that SARSA has the ability to solve complex games in the airline industry and provide promising learning results as compared to the demand prediction model.

Regarding vertical relationships (e.g., producers and retailers), some other situations need to be considered in the airline industry [112]. Both airline companies (as producers) and OTAs (as retailers) occupy some degrees of market control. Although OTAs set their own airfare price, airline companies are the service producers and thus exert pricing control as well. There is a bargain space that will lead the final sticker price to be varied for the consumers who search for offers from different channels. The final price can be affected by the distribution of the market as a result of competition that exists among the products and the retailers. To better understand the distribution of airline ticket prices in those two aforementioned competitive market structures, Bilotkach and Pejcinovska [23] randomly selected 50 top U.S. domestic market-pairs and collected the fare quotes from three major OTAs. They concluded, through analysis by means of simple regression with a natural logarithm transformation applied to the dependent variables, the existing competition between agents is an essential factor to determine the price. It also creates competition between producers (different airline companies). In 2015, Bilotkache et al. [22] also found that, contrary to popular belief, a positive correlation exists between the load factor and the fare segments. Theoretically, the fare price should follow a steady increase, corresponding to the increasing load factor as the departure date approaches. However, by conducting instrument variable experiments, the results pointed to another probability that would potentially lead to a boost in revenue. Increasing the ticket price as the departure date gets closer does not guarantee a gain in profit, since it might stop the growth of the load factor. These theoretical models consider several key independent variables, such as potential peak demand periods, econometric limitations, correlation with stocks.

2.2 Inventory Control

Inventory control or inventory management is a critical topic in operations management that generally refers to the strategies maximizing the inventory usage of the company. In particular, inventory seat control in the airline industry is the process of discovering the optimal seat booking for each class to maximize airline revenue. The lack of efficient inventory control can lead to

customer dissatisfaction, as well as the reduction in profits and productivity. Over the past decade, the airline industry has started to leverage data analytics to improve pricing strategies and inventory management [27]. Generally speaking, seat inventory control techniques can be categorized into the following groups [103]: single leg and network based.

2.2.1 Single Leg. There are independent policies of booking control for different flight legs and the methods in this group are categorized as static and dynamic. Static techniques provide an optimal seat allocation at a specific point in time [28]. Littlehood [97], for example, suggested one of the first studies in the static strategy for a single leg flight with two fare classes, in which a low fare booking proposal should be accepted if it generates more revenue than the highest fare's projected revenue for the same seat. This idea was further extended by Belobaba [17] using the **Expected Marginal Seat Revenue (EMSR)** approach for multiple nested fare classes. The EMSR solution generates the degrees of protection for fare groups based on the amount of seats protected. Despite its effectiveness, this system is only capable of generating optimum booking limits for two fare classes. In a recent work [144], a static two-class overbooking model is proposed integrating both inventory control and overbooking techniques in RM. This model enhances the expected profit using a closed-form expression to find the best limit for overbooking, and a sensitivity analysis is conducted by modifying the parameters in the model. These parameters include refund, penalty, revenue, and denied boarding costs, as well as capacity and show-up probability.

Different from static solutions, dynamic techniques determine the policy over time instead of doing so at the beginning of the booking time. Discrete time dynamic programming models [89], and models that also consider overbooking, cancellation, and no-shows, are examples of dynamic solution methods. In addition, there are combined methods that link static and dynamic methods based on **Markov Decision Process (MDP)** [87]. Subramanian et al. formulated a complete MDP that integrated dynamic and static approaches to seat allocation on a single-leg flight with multiple fare classes [149]. The assumptions cover a wide range of features, such as cancellations, overbooking, and discounting. However, adding additional features increases the versatility of MDP. A simplified approach restricted to basic assumptions was introduced by Lautenbacher [87]. Later on, in 2002, Gosavii et al. [67] proposed a **Semi-Markov Decision Problem (SMDP)** for single leg revenue management. This was the first model designed based on RL in the airline industry, which outperforms previous methods such as EMSR for solving the RM problem. The authors considered the RL model because it is not only able to scale up to a large state space, but it can also handle the realistic factors mentioned above. They noted that most of the stochastic dynamic programming models can only accommodate a subset of the related features to make the model manageable. However, more features are needed to generate optimal or near-optimal results, such as random cancellations and random demand requests from multiple fare classes. In it, the authors used Q-learning to iteratively obtain the Q value for each state-action pair. The state-space is defined based on the latest purchasing request of a certain class with the updated status of the sold seats for each class and the specific remaining time until departure. The results presented in the paper showed an average improvement of 1.5% as compared to nested EMSR. Similarly, in another work [119], a simulation-based greedy grid-search algorithm was proposed to optimize the expected total revenue by considering the customer choice behavior in sequential multiple flights for seat inventory control. In early 2019, a deep Q-Learning framework was proposed to solve the RM problem in a single O-D market [136], which leverages the advanced deep learning and RL algorithms to approximate the Q function that solves inventory control and overbooking problems. Its nonlinearities were captured through the neural network to train an agent to make decisions on accepting or denying passenger's booking requests. However, this model is still a prototype that needs more improvements to handle the full scope of the real-world airline RM problems.

2.2.2 Network Based. In this group, revenue is maximized by considering all the flights in a network at the same time. Network inventory control is necessary in itineraries involving connecting flights. In such cases where the network effects are increased, the single leg inventory control is not sufficient. This group of methods started with mathematical programming approaches [63]. Curry [42] proposed a new mathematical programming approach in seat allocation using piecewise linear approximation to maximize the revenue. The work leveraged mathematical programming and marginal seat revenue to make it possible to handle larger problems and multiple O-D pairs, while considering the fare class nesting in seat inventory management.

In addition to the mathematical programming methods, there are several research studies on simulation-based methods for the problem of the network inventory control that can be categorized into model-based and model-free search techniques [119]. In Reference [19], a stochastic gradient algorithm combined with dynamic programming is employed by Bertsimas and De Boer to determine the optimal booking limits that optimize the expected revenue function. This expected revenue function is approximated by simulating the booking process. More accurate revenue differential estimates can be obtained by applying the revenues' average, acquired through booking limits, over a sequence of simulated booking request results. Finally, using dynamic programming, they significantly reduce the computation time. Gosavi et al. [66] presented a model-free search method to achieve the global optimal solution. They also identify the best booking limits using simultaneous perturbation integrated with simulated annealing, a popular meta-heuristic for discrete optimization.

2.3 Overbooking

Given the fact that a certain percentage of passengers for final pre-departure bookings might not show up, overbooking becomes an essential method to avoid the loss of revenue. Overbooking can also mitigate the effects of cancellation at any time before the departure. However, a high overbooking rate might lead to the risk of having not enough capacity to allocate all the passengers. Moreover, the airline carrier not only needs to compensate the extra ticket-holders financially, but it also risks losing the loyalty of their customers. Therefore, there are many research studies focusing on how an appropriate overbooking rate could be selected.

Early on, overbooking was studied mostly in operation science, where the optimal overbooking rate is obtained from a pre-built model. The first model of overbooking was proposed by Beckmann in 1958 [16], which provides the upper bound of the overbooking rate by optimizing the revenue loss from the unsold seats and an overbooking penalty. In 1961, Thompson [156] introduced stochastic techniques to solve the overbooking problem and proposed an incremental control approach to optimize the overbooking rate by applying a constraint based on the probability of having a passenger with no seat. This model provides a more realistic paradigm to solve the problem. Although stochastic models can help to determine the policy, customer demand is not considered in the model. As the usage of electronic reservation systems in the airline industry grow, demand prediction technique becomes feasible (see Section 2.4). By incorporating the demand prediction technique, Rothstein [131] proposed an overbooking model based on Markovian sequential decision processes. This model takes the passenger demand distribution into consideration and achieves the optimal revenue under the oversales constraint. Furthermore, a set of booking policy decision rules [137] is proposed by Schlifer and Vardi to find the optimal overbooking policy, based on the statistical models to characterize future demand and cancellation patterns. The proposed rules can also be extended to handle flights with two fare classes and two legs.

All the aforementioned models are static models and do not allow for optimizing the dynamic process of ticket cancellation. Kosten first proposed a dynamic approach in the continuous time

domain [84], but the computation was impractical, since a series of simultaneous differential equations need to be solved. To make it easier to solve the dynamic process problem, dynamic programming is first applied by Rothstein [130] and the model was tested with data collected from American Airlines. In the follow-up work, this model was extended to consider passengers in the first and economic class simultaneously [8].

Often, the historical no-show rate may not be available or the data is not completely reliable. In those instances, Zhang et al. proposed a fuzzy system to address to the overbooking problem [179]. The fuzzy system can provide an appropriate overbooking rate without historical no-show data. This method helps to mitigate the traditional models' dependency on the data and enables overbooking for the new routes without sufficient historical data.

2.4 Demand Prediction

The goal of demand prediction is to precisely forecast customer demands for air travel based on pricing. The common methodology is to build a regression model to identify the pattern between demand and a selected set of features, including the ticket price, income elasticity, CPI index, population of the region, and distance between the O-D pair. These features might be incorporated into the model directly or indirectly, i.e., the regression model can be trained by the statistics and entries of the historical data directly, while the features might be used to first model the implicit factors, such as popularity of the market [98], and then demand is computed based on these factors. Demand prediction is a critical process of a successful decision-making strategy for airline RM, since it bridges customer behavior and ticket pricing strategy.

Most demand prediction techniques operate at the product-level, i.e., the prediction is made for each O-D pair or each flight [160]. In early research on the demand for air travel [7, 76], log-linear models were built to estimate the demand in each time period, and the ticket price, the measure of incomes, the measures of the price of domestic goods, and the CPI index were considered as the variables. Apart from the macroeconomic factors considered in the model, other explanatory features have been explored, such as demographic [21] and geographic [20] factors. The population and the type of airport (large, medium, small hub, or non-hub), combined with the price and income, have been shown to have a significant influence on the passenger demand as well. In Reference [76], the relationship between the number of passengers and the ticket price as well as the relationship between the macroeconomic factors and the geographic factors are measured using the logarithmic-linear model. Delahaye et al. [44] demonstrated the integration between **Support Vector Machines (SVM)** and the traditional econometric approach, discrete choice-modeling, to study the customer's preference in regards to different flight routes. They also showed how the proposed technique could be applied to dynamic pricing optimization. Recently, **Artificial Neural Networks (ANNs)**, which can model demand as a non-linear function, have been proposed to further improve the prediction performance. In Reference [105], a **Multi-Layer Perceptron (MLP)** approach was applied to model the demand with respect to the passenger number and monthly flight number. Historical data were collected and used to train the MLP, which, as suggested by the authors, could be used to predict the future trend of the demand.

Since there are multiple airline carriers in the market, the aforementioned models can only provide insights into the relationship between passenger demand and various variables, but fail to deliver insight into the competition among different carriers. To overcome this issue, Hansen [72] introduced *n*-player non-cooperative game theory to investigate competition considering price, frequency, and number of stops. Each carrier is modeled as one of the players of the game, and their pricing strategies for each origin-determination market form the strategy space. The carriers in the game compete with each other for market share and reach equilibrium. This formulation of

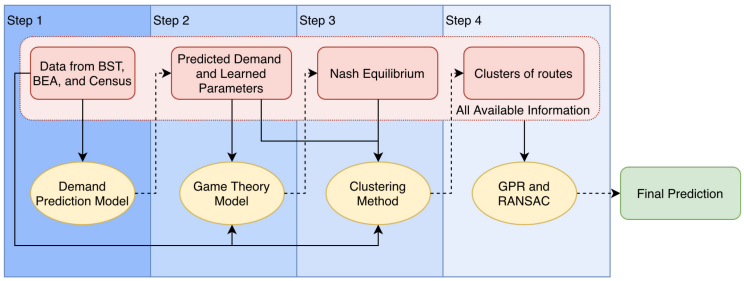

Fig. 3. The MAP architecture. Solid lines for input flows and dashed lines for output flows [9].

.

the n-player non-cooperative game can be solved by using the **Multinomial Logit Regression (Multinomial LOGR)**, obtaining the estimations of market share for the airline carriers in the market. Multinomial LOGR (a.k.a. Softmax Regression) is a type of GLM, which can be applied to the classification problems with more than two categories. Hansen models the market share of the flight service as the Multinomial LOGR form, where the market share of each type of flight service is predicted by the features of all flight services in the market. Particularly, the effects of carrier-specific factors (operational safety and delay duration) [150] and aircraft-specific factors (aircraft type and seat availability) [170] are added in the game to improve model performance.

However, the previously mentioned studies mainly focus on short-term analysis, which does not consider the long-term effects on the selected variables nor conduct time-series analysis on the data. Therefore, these models may fail to respond appropriately during non-stationary situations. Junwook and Jungho [37] applied Johansen co-integration analysis and a vector error-correction model for demand prediction and performed an analysis between demand and the NASDAQ index, along with price and income. This model was more robust and could respond to rapid changes in market data, such as the ones caused by the 9/11 terrorist attack.

Furthermore, An et al. [9] proposed the **Maximizing Airline Profits (MAP)** architecture, which integrates total demand prediction, competition analysis, and RM. This model is able to provide demand prediction to the airline carrier along with the demand-prediction-based ticket pricing for profit optimization. As shown in Figure 3, the price, income, and all of the related data are first used to predict the total demands of each route. Then, a clustering technique is applied after the competition analysis, which groups the routes into various categories. Due to the heterogeneity within the routes, the clustering technique can be helpful to separate different routes and improve the final profit optimization performance, which is generated by iterative **Gaussian Process Regression (GPR)** and **RANdom SAmple Consensus (RANSAC)**. In the MAP model, both demand prediction and competition analysis are performed based on an ensemble of different models to enhance the performance of each component.

Recently, with the development of big data techniques and autonomous recommendation systems, personalized demand analysis has been proposed to predict the future demands of each customer, which can provide more detailed information [98]. The travel topic model and relational travel topic model, for example, combine the latent factor model and the collaborative filtering method. This proposed model discovers both the travel preferences of airline customers and the categories of travel based on air routes and airline carriers. The **Passenger Name Record (PNR)** data are anonymized and used instead of the statistics of air travel to train the model. These components enable the model to have a strong capability to systematically analyze the market and individual customers to improve the prediction performance of the travel demands at the customer level.

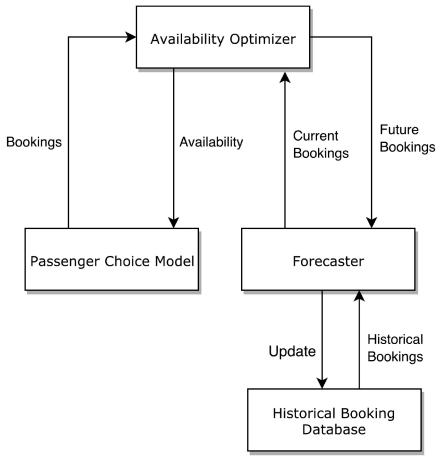

Fig. 4. Structure of the PODS framework.

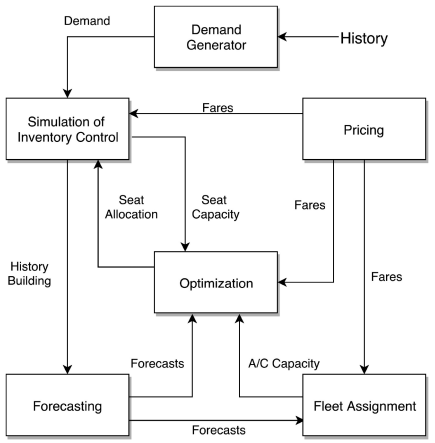

Fig. 5. Structure of Frank's Simulation Tool.

2.5 Simulation Tools

Since the early 1970s, when Littlewood [97] first proposed a solution for the airline RM problem, the airline industry has kept working on identifying optimal policies that could maximize their revenues. As demonstrated in Reference [141], an effective RM approach could improve airline earnings by up to 7%. However, RM methods are becoming increasingly sophisticated, and as a result, it is impossible for airlines to switch from one RM approach to another without incurring a considerable amount of monetary and time cost. Therefore, it is crucial to develop a practical methodology to measure and compare the performance of different RM techniques in the early stages of an airline's RM strategic timeline. Simulation techniques, emphasized by Talluri in Reference [153], are frequently utilized to evaluate the RM models by examining potential revenues in multiple scenarios. A simulation tool can draw a whole picture of the RM strategy under evaluation by including carefully modeled customer behavior and the firm's sales practice. A properly configured simulation environment can evaluate and compare different RM methods before deployment, allowing airlines to decide the RM method that best meets their needs.

The **Passenger Origin-Destination Simulator (PODS)** is one of the most popular simulation frameworks that model the air-travel demand. It was initially developed at Boeing and could be used to study the impact of changing the arrangement of routes and networks, fare products, flight schedules, and RM system capabilities, especially at the network-level [18]. PODS simulate the interaction between passengers and the airline company regarding their varied choices of airlines, routes, and fares. The simulation results include the choices of each passenger, the terminated traffic loads, and total revenues. These results are helpful resources for the researchers to analyze and evaluate different RM strategies [32, 55]. Figure 4 illustrates the structure of the PODS framework, which consists of four major components: a historical booking database, demand forecaster, seat availability optimizer, and passenger choice model. The historical booking database keeps all of the simulated booking records. They are the inputs to the demand forecaster, which predicts the customer demand for an upcoming flight or fare class. The demand forecaster also affects the seat availability optimizer, which sets up the booking limit for each fare class. Finally, the passenger choice model generates synthetic passengers with different characteristics, such as their willingness-to-pay, the costs incurred by choosing the departure/arrival window they do not prefer, adding legs to the itineraries, and restrictions associated with different fare classes. By

employing the optimized booking limits and synthetic passenger characteristics, the simulator generates passenger bookings with the highest-ranking path/fare class.

An event-driven stochastic simulation framework was proposed by Frank et al. [57] in 2006. As depicted in Figure 5, this framework investigates the impact of continuously adjusted fleet assignment on revenue. The fleet assignment module assigns a type of aircraft to a specific leg by considering the forecasted passenger demand from the demand generator module with certain constraints (e.g., the number of available slots). Booking requests are generated by a series of probability distribution models with the parameters inferred from the historical data. The pricing engine module calculates the fare based on passenger's demand segmentation, leg-level booking classes, and O-D level fare classes. The forecast and optimization component utilizes an additive pick-up prediction model. It uses exponential smoothing [171] and the EMSR method, combined with a heuristic bid price approach for the optimization task [152]. Simulation test results have demonstrated a positive impact of repeated fleet assignment on the revenue. Recently, Doreswamy et al. [49] introduced a new simulation analysis tool named **Airline Planning and Operations Simulator (APOS)**, which has been developed at Sabre Airline Solutions. In this work, APOS was used to test the impact of migrating from a leg-based RM system to an O-D based RM system. The simulation results show that revenue could increase from introducing the O-D system by up to 6.6%.

2.6 Ancillary Pricing Optimization

Ancillary revenue is the revenue made from goods and services considered as secondary options to a company's primary product. Ancillary products were first introduced in 2009, as airlines responded to the 2008 financial crisis [60]. Airlines quickly realized it is the perfect opportunity to make a decent amount of revenue by providing their customers the opportunity for some extra comfort. Ancillary revenue is a pioneering innovation development mainly stemmed from the LCCs business model. Only recently are traditional networks beginning to pick up on the new methodologies [115, 145]. Products that were traditionally part of the fare ticket are now unbundled, optional for an extra charge. In the case of baggage fees, the traditional approach was to include the first two pieces of checked luggage in the ticket purchase, unless these exceeded the overall weight limitation. However, these priced options for checked baggage are now one of the fastest growing items in a portfolio of unbundled products [115].

RM for ancillary products is still nowhere near reaching its full potential. In 2013, the **International Air Transport Association (IATA)** reported that, although airlines can create clear value for customers, the industry still had difficulty making an adequate level of profits [120]. Recently, according to IdeaWorks and CarTrawler,[1] airlines are paying more attention to optimizing their RM through the sale of ancillaries to improve their profitability. Fiig et al. [56] describes how current legacy IT systems delegate the airline's access of the data to content aggregators, limiting the airline's control over offer construction and consequently hiding the customer's identity. Some of the identified reasons that airlines may miss significant revenue opportunities include (1) customers having a limited access to the value of ancillary products until after the purchase of a ticket; (2) a reliance on fixed prices that are decided in advanced; and (3) lack of price point optimization, even prices differ according to the market, time of departure, or sales channel.

The application of advanced ML techniques has started to make an impact in ancillary revenues. Navitaire, a company that assists newly launched airlines with products that focus on reservations, revenue, and operations management, developed **Ancillary Price Optimization**

[1] http://www.ideaworkscompany.com/wp-content/uploads/2017/11/Press-Release-123-Global-Estimate.pdf.

(APO)[2] which applies a combination of A/B testing and ML techniques, making it a robust tool for high-dimensional data analysis and helps extract essential pricing features without the need of a customer choice model. Several low-cost/hybrid carriers have started experimenting with APO by dynamically adjusting the price for premium seats and checked luggage, since these products can generate higher sale volumes [56]. The process follows a cycle that starts with a collection of data that contains attributes for a price point, and the customer's decision—whether there was a purchase or not. APO then identifies important pricing attributes and selects a model among standard methods (i.e., **Decision Trees (DTs), Random Forest (RF)**, Regression, and ANNs). Through uplift analysis [125], the model that has the best performance and results in the largest benefit is then deployed to production. The model is continuously monitored to verify that it behaves according to expectations. Even after the deployment of the model, the cycle continues based on new experiments [56]. ML techniques have also been proposed to solve the problem of asymmetric dominance (or decoy effect) and drive conversion rates. In marketing, the decoy effect considers the consumer behavior as irrational. By adding an inferior offer, it generates a cognitive bias where other alternatives appear more appealing to the customer. **Pairwise Choice Markov Chains (PCMC)** have been proposed to overcome the limitations of traditional choice models and demonstrate the decoy effect occurring within the data. Lhéritier [90] proposed an improvement of the **PCMC based on neural networks (PCMC-Net)** and tested on airline choice. Furthermore, the decoy effects have recently been studied and airlines begin to value such strategy in ancillary services.[3] A recent study by González-Prieto [64] proposed a theoretical model of the decoy effect to enhance airline profitability by studying customers' purchasing process of ancillary services. Nevertheless, there is a lot of potential to conduct more comprehensive studies and apply advanced ML techniques in the evaluation of the decoy effect for ancillary revenue optimization.

According to an Amadeus report in 2011,[4] ancillary products can be segmented into four basic categories: (1) *unbundling fares (a la carte)* – service features that are part of the customer's trip, including in-flight food/beverages, checked luggage, reserved seats, WiFi access, and so on; (2) *commission-based* – opportunities from third-party sellers involving hotel, car rental, and travel insurance; (3) ***Frequent Flyer Program (FFP)*** – loyalty-based incentives that encourage customers to accumulate airline miles or points to be redeemed for future trips or other rewards; and (4) *advertising* – magazines available in-flight or sold in the airline's airport lounge.

2.6.1 Unbundling Fares (A la Carte). An increasing number of airlines are beginning to rely on *a la carte* pricing to generate a boost in revenue [115]. This category includes products for the customer's extra comfort during their trip along with punitive prices (change/cancelation fees). Research in this area mainly focuses on mining the customer's opinion, acceptance level, and **willingness to pay (WTP)** for ancillary products through conducted surveys. Some of the techniques applied to the survey data include applying Causal Inference methods, such as **Structural Equation Modeling (SEM)** to convey the connection between perception on price fairness and the implications towards an emotional response [39, 158]. Logistic Regression is the most commonly used technique in the airline industry to estimate the probability of observing different fare-seat management strategies [114] or perform sentiment analysis. Statistical Inference techniques such as a simple t-test to prove a hypothesis [164] are also used. There have also been studies that focus on two essential unbundling items, checked bags and premium seats [111]. Scotti et al. [135] applied OLS to determine if there is an influence between checked luggage fees and flight delays on the rate of customer complaints. The paper concludes no evidence was found of a possible negative

[2] https://navitaire.com/Styles/Images/PDFs/APO_Whitepaper.pdf.
[3] https://amadeus.com/en/insights/white-paper/the-importance-of-understanding-travelers-motivation.
[4] https://amadeus.com/en/insights/blog/ancillary-revenue-coming-soon-around-the-world.

relationship between the introduction of unbundled pricing and customer satisfaction. However, in the presence of endogeneity, where there is a correlation between the error and explanatory variable, the estimates made by OLS are biased.

2.6.2 Commission-based. Commission-based ancillaries allow for opportunities involving hotel, car rental, and travel insurance services provided by third-party organizations. The category is also known as dynamic packaging, where pricing, constraints, and the final decision are all determined in real-time based on online inventory [115]. Therefore, a system dedicated to creating a dynamic package requires automatic online configuration and collection of travel products and services into a package targeted to certain customer segments. Cardoso and Lange [31] evaluated three major OTAs with support for dynamic packaging (Expedia, Travelocity, and Orbitz) and took note of their interoperability problems. The architectures of dynamic packaging systems are complex and challenging because of the integration of unstructured data from different sources on the Web. Although studies have acknowledged the advantages of AI techniques for dynamic packages, there is yet to be a specifically proposed technique that has been proven to accurately infer the interest and preferences of airline customers through pattern mining [54].

2.6.3 Frequent Flyer Program. A point-based system or mileage program is a well-known incentive companies use to retain its customers and encourage loyalty. Studies in regards to this category often aim to identify what are the main factors that influence a passenger into signing up for a loyalty program. Wong and Chung [173] studied the loyal passenger's decision pattern by analyzing their personal characteristics, consumer behavior, and overall perception towards the quality of service. Examinations of the characteristics from loyal customers were done using a DT to identify the possible strategies that can be utilized to attract more attention. They made use of the C4.5 algorithm [124] to build DTs from a set of training data. The training dataset focused on the customers' personal information (i.e., gender, age), consumption features (i.e., airline membership, most frequent location for ticket purchase), and degree of satisfaction for various service items (including airport service, cabin facilities, etc.). The C4.5 DT is fitted to the data considering a target variable that defines a passenger as either loyal or disloyal. A similar study by Dolnicar et al. [48] investigated the main drivers of airline loyalty and identified best discriminative variables between loyal travelers and those who are not, which included frequent flyer membership, fares, carrier status, and the airline's reputation. Through the application of DTs that are trained on the survey data through recursive partitioning, the approach essentially generates a customer segmentation by recursively splitting the data into two subsets according to an explanatory variable. Therefore, customers from the same sub-groups convey a similar behavioral loyalty.

2.6.4 Advertising. Advertising revenue is generated through the sales of ads funded by third-party organizations, including in-flight magazines, and fee-based products sample placements. According to a report by the *Wall Street Journal* in 2009, over 80% of passengers read the in-flight magazines for an average of 30 min per flight.[5] AI techniques can improve advertising for the airlines by creating personalized offers and a suggestion of products specifically targeted to individual customers [169].

2.7 Service Improvement and Customer Experience

Airline companies are under the constant pressure of balancing between cost-cutting and service improvement. Cutting cost alone without keeping satisfactory levels of service quality cannot guarantee a successful business [34]. The ability to quickly capture the opinions of customers about

[5] http://www.wsj.com/articles/SB10001424052748703819904574555701528290902.

their service experience and company products has become the indicator of how well an airline company can adapt to a fast-changing market. OTAs and social media networks have evolved into platforms for passengers to express their views on certain airline and airport companies. Consequently, significant amounts of data are produced in the form of product ratings, social media posts, and feedback. Therefore, there is a rise in demand for solutions that can handle data at this scale, such as automatic sentiment analysis, polarity detection, and opinion information extraction.

Sentiment analysis is considered to be the study of using text and **natural language processing (NLP)** techniques to systematically solve problems related to people's opinion, attitude, and emotions. Millions of messages, containing customer opinions about airline services and goods, are shared on social networking sites including Twitter and Facebook. As a result, numerous studies have proposed to classify the sentiment of social media posts in regard to a certain airline or airport services from social networks [85, 147] and dedicated air travel rating websites [175].

In general, sentiment classification approaches may be divided into three types: ML-based, lexicon based, and hybrid methods. The ML-based approach leverages different statistical algorithms to analyze the sentiment by following a typical text classification methodology that uses syntactic or linguistic features. Frequently used algorithms for the classification of customer's sentiment include **Naive Bayes (NB)**, LR, DT, RF, and SVM. The lexicon-based approach [151] will either manually or automatically build a dictionary with opinion terms to derive the polarity of each entity through term-matching. The hybrid approach integrates both methods, in which it applies ML classification to lexicon-based textual features [161].

2.7.1 Machine Learning-based Approaches. Supervised learning relies on having (1) a large enough quantity of the training data and (2) labels for each sample. The quantity of data especially affects the model's performance. To overcome the challenge of limited labeled data, Drury et al. [50] proposed a semi-supervised learning approach for sentiment classification. An iterative self-training process uses NB classifier as the base learner to label the training candidates that are selected by a high precision linguistic rule-based classifier. If the confidence score generated by the base learner is greater than a default threshold, then the data is labeled and added to the next iteration. Experiments conducted on user-generated reviews of airline meals[6] make use of randomly selected documents as training data. The proposed method has demonstrated promising results for classifying text documents into sentiment categories with a small training dataset. Another study focusing on sentiment analysis for airline reviews was carried out by Wan and Gao [165] in 2015. The study proposed an ensemble learning framework involving multiple supervised ML models, such as NB, SVM, Bayesian Network, DT, and RF. A total of 12,864 public tweets regarding 16 of the largest U.S. carriers were collected and their sentiment was classified into negative, positive, and neutral. The best result reaches an F1 score of 91.7%.

2.7.2 Lexicon-based Approaches. Lexicon-based approaches focus on the identification of opinion key terms that can convey sentiment in either desired (positive) or undesired (negative) states. Lexicon-based methods [151] use a handcrafted or automatically generated sentiment dictionary to match the opinion word list with the target corpus to determine sentiment polarity. The advantage of lexicon-based methods is that training data is not required. In Reference [102], airline-related tweets are classified as either service-related or product-related, where four specific airlines are studied. The lexicon dictionary consists of 20 terms, which means that a 20-dimension vector represents each tweet. The proposed model assesses the sentiment polarity of Twitter posts that contain information related to customer's feedback and experience on pricing and service quality.

[6] airlinemeals.net.

Recently, Kaur and Balakrishnan [78] developed an enhanced lexicon-based sentiment scoring system by analyzing the letter repetition patterns. The model was developed and tested using posts from the airlines' official Facebook pages. The experimental result shows that the proposed **Sentiment Intensity Calculator (SentI-Cal)** gains a significant edge over the traditional **Semantic Orientation Calculator (SO-CAL)** in terms of performance. The accuracy reached 90.7% compared to the baseline (58.33%).

2.7.3 Hybrid Approaches. The combination of ML- and lexicon-based methods has generated promising results for sentiment analysis applied to airline customer reviews. Khan et al. [80] proposed a notable sentence-level sentiment analysis framework for 700 reviews from Skytrax.[7] First, each sentence is labeled as negative or positive based on the result of the NB classifier using word-level features. Then, the labeled sentences are used to train an SVM classifier for detecting sentiment polarity. While a lexicon dictionary detects positive polarity in consumers' opinion, clustering analysis discovers the main topics involved in the discussion. Yee and Pei [176] incorporated text mining and clustering techniques on 10,895 tweets that hashtag (#) or mention (@) Malaysian airline companies. Subjects such as itinerary promotion and cancellation, customer service, and post-booking management were identified. Intuitively, sentiment polarities vary depending on specific topics or contexts. Therefore, opinion mining and summarization require functions that can detect both topic and sentiment together. More recently, Lacic et al. [85] crawled information from Skytrax to illustrate the rating features (i.e., lounge comfort, boarding time, seat legroom space, and cabin staff service quality) that have the most impact on travelers' satisfaction. Suffix tree clustering [178] is used to identify topics that cover reviews and additional features that are useful in the rating schema. The Hoeffding Tree algorithm has demonstrated to have both the highest accuracy and fastest training time when mining data in real-time compared to other classifiers. Smith et al. [142] implemented a user-centered human reinforcement topic modeling system that generates topics from airline service reviews interactively. The authors extended their original work by adding more user refinements. The experiment used the Kaggle Twitter U.S. Airline Sentiment dataset, which includes more than 14,000 tweets.

2.8 Price Mining for Strategic Customers

Traditional statistical approaches are broadly applied to industrial practices due to their simplicity and generally good performance. While airline industry stakeholders are exploring more pricing approaches to generate higher revenue, some research work aims to minimize the cost of purchasing a flight ticket for the customer's benefit. Traditional models that generate hand-crafted rules are straightforward compared to the rules generated automatically by ML algorithms. Therefore, the behavior of airfares change is getting more complicated. Early in 2003, Oren et al. [52] studied the pricing pattern for two specific markets and developed an algorithm to show when passengers should buy a ticket to minimize cost. Though the data they collected is limited, their multi-strategy data mining algorithm, Hamlet, has shown promising predictions. It incorporates RL, rule learning, and time series methods. In 2014, Li et al. [93] brought up the question of whether air-travel consumers are strategic. By creating a structural demand model of alternative pricing strategies, it not only proves strategic consumers exist, but also analyzes the potential consequences to the revenue. As the portion of strategic customers varies in different city-pair markets and booking to departure times, it tends to increase revenues in leisure markets. Other than the aforementioned techniques, rule-based feature selection methods explicitly designed for the objective are also widely adopted

[7]Website for airline, airport, and associated air travel traveler reviews: skytrax.com.

in the airline related studies. Lhéritier et al. [91] built a ticket choice model using RFs to generate the score of reduction in impurity for each feature. Groves and Gini [69] used the lagged feature computation strategy to prune certain features related to a later time window than the flight departure date time, based on the type of the flight. Recently, Mottini et al. [107] proposed leveraging **Recurrent Neural Networks (RNNs)** with attention to learn the conditional probabilities of a customer's choice. In their study, the decision-making process of the customers was modeled to predict the customers' preferred itinerary in different flight booking scenarios. In the proposed model, a variant of Pointer Networks [162] was implemented to select the preferred item among any provided set of inputs. A dense representation was first generated by normalizing and embedding a total of 15 features related to the flight, including Origin/Destination pair, price, and so on. Then, the Pointer Network utilizes an encoder-decoder structure based on **Long-Short Term Memory (LSTM)** networks [75] along with an attention mechanism to generate an estimated probability for an itinerary being selected out of all given itineraries. By combining all these techniques, the proposed model was able to represent a more complex relationship between the inputs and the outputs without relying on an assumption of independence, and achieved a higher prediction accuracy as compared to a conventional multinominal logit model. This is the first neural network structure that models discrete choice problems. One framework described in Reference [77] assists Skyscanner users by allowing them to create cheap and exclusive round-trip flight itineraries by combining outbound and inbound tickets from various airlines. After analyzing temporal patterns, the authors pointed out that when the departure approaches, a search considering combined airlines provides more competitive results, as the traditional single flight tickets get more expensive. The predictive itinerary construction is formalized as a supervised learning problem. Several popular models are utilized including LR, multi-armed bandit, and RF. Additionally, location information is represented by different types of embeddings. Experiments show further improvement by incorporating deep neural networks as compared to RF, regardless of the model complexity.

2.9 Connectivity

Developing a robust air transport network greatly facilitates the globalization of the markets, technology, and economic growth [86]. The air transport network connectivity is defined as the degree of connection of a specific node (airport, city, or even region) to all of its neighbors [30]. This is an essential measurement that demonstrates how well a node in the network is connected to the rest of the world and becomes an important domain in the airline industry. The network connectivity can be categorized either based on the measurement (e.g., node accessibility or betweenness) or the data source (supply or demand data) [148]. Supply data includes data relating to flight schedules, while the demand data provides information related to passengers. Both data sources are found to be used together in recent studies. Arvis and Shepherd [10] introduced the concept of **Air Connectivity Index (ACI)**, which sets a comprehensive guideline on how to interpret connectivity in the air transportation domain. The ACI is a metric to indicate the significance of a node in the global air transportation system. It measures the degree, closeness, and betweenness of nodes. Multiple attributes are defined for each node in the network to meet certain criteria: (1) The connection between a pair of nodes needs to be supported by a well-established transportation model; (2) Two nodes with the same connection to the rest of the network should have the same connectivity, regardless of the differences between their size, which can be measured by passenger volumes or amount of traffic; (3) The calculation of the metric should consider the full network, not only the target nodes and its immediate neighbors. The authors borrow the basic idea of biproportional structure from the generalized spatial interaction model framework [113] to develop an ACI model that satisfies the above-mentioned criteria. The model incorporates the concept of

attraction and impedance factors between a pair of nodes. Attraction is proportionate to the size, economic development of the origin, and destination pair. Impedance represents all the costs involved between the interaction of the pair, which includes travel time and distance. In practice, an attractive potential can be derived from the data such as the total number of flights, seat capacity, passenger flows, and cargo volumes. In addition, the impedance can be derived from geological distance from the O-D pairs. As a result, ACI can be categorized by both its measurement and data source.

Allroggen et al. [6] address the connectivity problem from a different angle. They focused on assessing the quality of air transportation between each pair of nodes by utilizing the "connection quality-weighting" approach and proposed the **Global Connectivity Index (GCI)**. GCI evaluates the connection quality of a specific airport/region by summing up the potential destination quality weighted by the attributes of each connection. The closeness between nodes is considered when creating the GCI. The absolute destination quality of a specific node is obtained by considering both the travel distance and the market potential, which is the total population of the accessible market. When compared with ACI, GCI models the connectivity factor by incorporating additional perspectives, such as the connection quality and destination levels.

A more recent work by Zhu et al. [183] defines the connectivity metric from the perspective of passengers. They model the air transport connectivity by using a multiplicative utility function that integrates three major components: seating capacity, trip duration, and flight transfer quality. The capacity factor is represented by the square root of the ratio of the seating capacity of a specific connection to the seating capacity of the benchmark aircraft. The velocity factor contains the flight duration, as well as the penalty for the time spent on the indirect flight. The transfer quality factor consists of time quality and service quality of indirect flights. The velocity factor represents the time spent on transfer flight, and the transfer quality factor takes into consideration the characteristics of the waiting lounge (i.e., comfort).

2.10 Air Traffic Management

In the previous subsections, we cover the vast majority of airline research topics that use customer- and/or market-oriented air travel data for RM, which are the major focus areas of this survey. In this subsection, we review other research topics that are not specific to airlines and customers.

Air Traffic Management (ATM) is another topic that also utilizes air travel data to assist general aircraft and airport operations. The increasing prevalence of LCCs along with the global rise of the middle-class have resulted in the rapid growth of the world's air traffic. **Trajectory-Based Operations (TBO)** uses time-based management to improve proactive forecasting of aircraft flows and minimize capacity-to-demand imbalances in the **National Airspace System (NAS)**. In the near future, TBO will be used as the core component of the ATM systems, which will significantly improve the accuracy, safety, and efficiency of air traffic control [61, 104]. As a result, during the past few years, aircraft trajectory prediction, a key technology in TBO, has attracted significant attentions in ATM research and recent modernization programs. Advanced data analytics techniques have been leveraged in this area to enhance the prediction accuracy in complex flight environments. Ayhan and Samet [13] proposed a stochastic trajectory prediction method based on **Hidden Markov Model (HMM)** and 4D trajectory data including 3D spatio-temporal parameters and weather features. Similarly, HMM has been used for online aircraft prediction of trajectories in Reference [117]. Zhao et al. [181] integrated multi-dimensional features of aircraft trajectories into deep LSTM to predict the aircraft trajectory. Bastas et al. [15] turned this problem into an imitation learning task and utilized the **Generative Adversarial Imitation Learning (GAIL)** framework [74] aiming to imitate experts "shaping" the trajectory. Finally, in Reference [118], **Conditional Generative Adversarial Network (CGAN)** was used for weather-based

trajectory prediction to alleviate the issue with limited data. Due to accidental events, such as traffic congestion and convective weather, the flight trajectories might need to be rerouted accordingly and thus reroute prediction becomes critical to enhance the accuracy of trajectory prediction, reduce the flying time, and improve the quality of airline operations. Since most of the reroutes are caused by severe weather, Michael et al. [101] developed a severe weather-modeling mechanism and integrated it with the airspace planning component to determine when and how the flight trajectory can be rerouted using a probability model. Recently, an ensemble of machine learning models, including DT, SVMs, and so on, has been utilized by Antony et al. [53] to predict flight reroute requests. Another important topic in ATM related to aircraft trajectory prediction is **Conflict Detection and Resolution (CD&R)** [155]. Many existing studies in CD&R used predicted aircraft trajectories to declare a conflict. Ayhan et al. [12] presented a framework based on their previous HMM trajectory prediction model [13] to detect the conflict. In Reference [33], a neural network binary classifier was used to discover which aircraft configuration might break the aviation regulation for minimum separation between in-flight aircrafts. The authors used a simulation tool to obtain the training dataset, which includes aircraft positions and each sample is labeled as a conflict or not. Finally, Wang et al. [168] developed a deep RL model for CD&R, based on the K-control actor-critic algorithm in which the agent generates a two-dimensional action to avoid the conflict.

Estimated time of arrival prediction, or flight delay prediction, is another application in air traffic control that provides critical information to facilitate ATM and the decision-making process. The flight delay prediction problem can be approached in a variety of ways. Zhang et al. [180] used a fuzzy LR model to evaluate the airport arrival delay and the estimated delay. The model takes into consideration both traffic and weather features. In another study, an asymmetric logit probability model was used as a tool to estimate the daily flight arrival delay probabilities [121]. It focuses on tackling the asymmetric nature of on-time and delayed flight frequencies. In recent years, ML techniques have been widely applied to predict the flight arrival and departure delays. For instance, RF is used to predict the estimated time of arrival in Reference [79], which combines the features from flight information, air traffic, and weather domains. In Reference [38], **Synthetic Minority Over-sampling Technique (SMOTE)** [35] was used to overcome the common data imbalance issue in flight on-time arrival data. Several ML algorithms were compared and RF was shown to demonstrate the best performance. Compared to previous studies, more weather-related features were added, such as visibility, snow depth, and various obscuration factors caused by different weather conditions. More advanced deep learning approaches are also used for flight delay prediction. For instance, Reference [82] applied RNNs on the daily on-time status data, and the model was capable of predicting the day-to-day delay status.

As the traffic volumes increase, to evaluate the system loads and allocate resources accordingly, the traffic density prediction at the sector-level becomes important for the ATM systems. Dynamic density was first proposed to determine how the traffic in all the sectors could be measured [83]. Due to the complexity of the air traffic system, David and Kevin [62] proposed to utilize neural networks to better model the traffic density. Furthermore, to utilize the spatial patterns of the sector-level traffic density, deep convolutional neural networks recently were developed to evaluate the traffic density and complexity [174]. In addition to estimating the density in real-time, Tian and Pan [157] proposed an LSTM model to predict the short-term traffic density with high accuracy.

With the increasing demand for air transportation with environmental considerations, many key challenges have been raised and caught lots of attentions at the airports, especially those busy hubs, to provide ground operations that support the on-time performance of flights. Airport runway configuration is an active and less-expensive approach for better airport capacity utilization

than airport expansion for increasing capacity. It tends to solve airport congestion by considering many factors that could also affect aircraft operations. According to References [11] and [126], weather conditions, airport demands, operator decision-making, coordination with surrounding airports, and runway characteristics are some of the key influencing factors for airport congestion. In the late 2010s, neural networks were popularly used to predict the runway configuration for maximizing the runway capacity. In Reference [5], a Multi-layer Artificial Neural Network model was presented by adopting different neural network techniques, including feed-forward back-propagation, recurrent back-propagation, and so on. Most recent research also illustrated the use of Convolutional Neural Networks in multiple airport systems [167]. The model used assembled grid weather forecast to obtain runway configurations and airport acceptance rates.

An accurate taxi time prediction is essential for both on-stand time prediction and take-off time prediction improvements and consists of the taxi-in time prediction and taxi-out time prediction. In 2010, Reference [14] leveraged RL to estimate runway taxi-out delays. The system models the taxi-out time prediction as an MDP. Ravizza et al. [128] presented a prediction model using a multiple LR analysis. The inputs are the airport layout and the historical taxi time. An airport layout was used to solve the airport ground movement problem, which could be generalized as a routing and scheduling problem. Relevant factors affecting the taxi time were identified, including the amount of traffic that could reduce the taxi speed, total taxiing distance to the gates, total amount of turning angels, and so on. Reference [140] studied the airport runway departure process and proposed a queuing model to estimate the taxi-out time distribution by attempting to forecast taxi wait times, queuing delays, and airport congestion levels. Other factors such as meteorological conditions, pilot behaviors, and system-wide taxi delays can also increase the uncertainties of making an accurate taxi time prediction. To measure those uncertainties and help comprehensive taxi planning, Chen et al. [36] employed the multi-objective fuzzy rule-based systems. Herrema et al. [73] focused on taxi-out time prediction models and identified the regression tree as the most efficient method with an average error of 1.6 minutes when comparing with the performance of using neural networks, RL, and MLP models. Recently, Li et al. [94] proposed a **wide-deep neural network model (WDM)** that works for both taxi-in and taxi-out time predictions, where the wide component is based on the GLM. The model involved the SMOTE method for data re-sampling, and its input contains both categorical and numerical parameters describing the weather conditions, runway configuration, aircraft type, taxiing distance, and so on. In another work, Wang et al. [166] first conducted a comprehensive review of the state-of-the-art taxi time prediction methodologies and reported RF as the best ML model that wins all the evaluation metrics, including accuracy, R^2, MAE, RMSE, and so on, when compared to the other models. The study also stated that it is critical to determine the values of the features for optimizing taxi time modeling accuracy and performance.

2.11 COVID-19 Impacts on the Airline Industry

The newly emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as COVID-19, was formally proclaimed a pandemic by the **World Health Organization (WHO)** on March 11, 2020. COVID-19 has had a major global impact on air travel mobility and the aviation sector in general. As a result, various studies have been published tackling the uncertainties brought by COVID-19, trying to make accurate forecasts of its impacts and proposing ways for rapid recovery post-pandemic. In the earlier months of COVID-19, Sobieralski [143] used time-series analysis to investigate the dynamics of recent global crises, including 9/11, SARS, and other disruptions, and examined the impact of instability shocks posed by COVID-19 on the U.S. airline labor. The analysis demonstrated that the effects of COVID-19 will continue for several years, and the airline workers will be the hardest hit. One major obstacle faced by the airlines is to leverage

their established methodologies and data analytics techniques, both during COVID-19 and during the recovery period. Many of the previously described traditional and machine learning methods, majorly trained and fine-tuned on historical data, struggled to overcome the difficulty of adapting to the new era due to the elevated schedule uncertainty and constantly evolving travel restrictions. Times of high uncertainty have pushed the airline companies to transition to more manual methods. Researchers are beginning to explore ways to reduce the traditional model's reliance on historical data, develop strategies that can pick up on the recent trends quicker, and integrate the domain knowledge generated by human revenue management experts [59].

3 SUMMARY OF DATA ANALYTICS TECHNIQUES IN CURRENT USE

Airline-related data are not only large in volume (number of instances), but also high in dimensionality (number of attributes). Most of the ML models are not designed to handle data with very large number of features (attributes), which will significantly decrease the model performance [182]. Moreover, irrelevant information in the data can also increase the difficulty of the learning process due to the additional training cost and the distraction of unrelated features. As a result, feature analysis is always introduced at an early stage to analyze, extract, and select relevant features from the raw data before translating it into an optimal representation that is tailored for each specific problem domain. Feature selection, either by hand or automatically, can (1) effectively reduce the impact of overfitting by reducing the dimensionality of the input vector; (2) improve the accuracy by including less misleading features; and (3) reduce training time, since the total number of parameters that the model needs to optimize is decreased. Several feature selection methods have been applied to airline-related tasks. For example, Berechman [4] developed a model using **Principle Component Analysis (PCA)** to determine the efficiency and quality of the airports. In another work, Gursoy et al. [71] used correspondent analysis, which is a special case of PCA, to analyze 15 features including on-time performance, mishandled baggage, customer complaints, and so on, to evaluate the service quality of the 10 major airlines in the US. Similarly, Dobruszkes [47] analyzed the European low-cost airline's network using PCA to select and reconstruct 75% of the features in the original data. The generated components were able to identify the features related to the importance of supply and exclusive routes. **Term Frequency-Inverse Document Frequency (TF-IDF)** and **Latent Semantic Analysis (LSA)** are two widely used strategies for text mining and sentiment analysis to classify relevant keywords and conduct dimension reduction with minimal knowledge loss. In Reference [70], TF-IDF was used to select important features from airline customer reviews by mapping the most frequent words to the feature collection and weight them across the entire corpus. However, TF-IDF lacks the ability to utilize semantic similarities between words, which makes it less ideal for tasks that involve complex semantic context. Gunarathne et al. [70] combined LSA with TF-IDF to generate a low-dimension feature space for further clustering of airline social media posts.

After the raw data is properly cleaned and the essential features are identified for a particular research topic, ML models can then be applied to solve the data modeling problems. Until this point in time, the airline industry has relied heavily on traditional ML techniques, particularly regression and MDP for RM. However, advanced techniques such as Neural Networks and RF models have been more frequently used in the ATM systems for a decade. Only recently, more advanced techniques such as Deep Neural Networks and **Deep Reinforcement Learning (DRL)** have been introduced to the airline RM domain. Table 3 summarizes the ML techniques used frequently to answer different airline research topic questions, as mentioned in the previous section. It provides an overview of how ML techniques in different categories have been utilized in the airline industry to model air travel data.

Table 3. Popular Machine Learning Models in Airline Industry

Categories	Models	Techniques/Papers
Supervised Learning	Regression	**Linear Regression:** Bilotkach and Pejcinovska [23], Mumbower et al. [110], Ravizza et al. [128], Scotti et al. [135], Zhang et al. [180] **Logarithmic-Linear Model:** Jung and Fujii [76] **Artificial Neural Network:** Ahmed et al. [5], Ali et al. [105] **Convolutional Neural Network:** Wang et al.[167], Xie et al. [174] **Deep Neural Network:** David and Kevin [62], Li et al. [94], Pang and Liu [118], Tian and Pan [157], Zhao et al.[181] **Hidden Markov Model:** Ayhan and Samet [13], Pan et al. [117] **Regression Tree:** Herrema et al. [73] **Random Forest:** Kern et al. [79], Wang et al. [166]
	Classification	**Logistic Regression:** Obeng and Sakano [114] **Multinomial LOGR:** Hansen [72] **Neural Networks:** Casado and Bermúdez [33] **Recurrent Neural Networks:** Michaela et al. [68], Kim et al. [82], Mottini and Acuna-Agost [107] **Decision Trees:** Dolnicar et al. [48] , Antony et al. [53], Fiig et al. [56], Wong and Chung [173] **Support Vector Machine:** Delahaye et al. [44]
Unsupervised Learning	Clustering	**Partitioning:** Obeng and Sakano [114], Pritscher and Feyen [123] **Hierarchical:** Dai et al. [43] **Distribution-based:** Piggott [122]
Reinforcement Learning	Markov Decision Process	**SARSA:** Collins and Thomas [40, 41] **Q-Learning:** Gosavii et al. [67] **Deep Q-Learning:** Shihab et al. [136] **Actor-Critic:** Wang et al. [168]

4 DATASETS AND DATA SOURCES

As companies collect more and more data, they develop effective techniques to tackle different objectives with the same ultimate goal, keep the business environment stable and running healthy—the airline industry is no exception. Collected data includes not only details on general performance, but also data on supply and demand, information that is not easily accessible to the public. A summary of popularly used datasets for airline related topics is listed in Table 4. This section describes datasets that mainly focus on airline functions along with research references that indicate how these datasets have been utilized in data analytic and ML approaches.

Public datasets in relation to the airline industry are limited, with academic research having to rely on either the data that is made available by the government or scrape their own data from public websites to test their hypotheses. The U.S. Department of Transportation website[8] is a rich source of information regarding airline operations, performance, and finance. BTS's Office of Airline Information maintains the Airline Origin and Destination Survey (DB1B), a 10% random sample of U.S. domestic carriers' ticket data. The DB1B dataset is used to determine patterns of air traffic, air carrier market share, and passenger flows. Each record contains information regarding a purchased ticket—origin and destination airports, miles the aircraft has flown, total fare, and whether the itinerary is a round-trip or one-way. DB1B data is beneficial for studies that determine the roles of aircraft characteristics in airlines' market share and demand [170], examine

[8] www.transtats.bts.gov.

Table 4. Popular Air Travel Datasets and Data Sources Used in Published Papers

Datasets/ Sources	Publi-cations	Content Focus					Number of...			Geographic Coverage	Time Range	Data Quality
		Fare	O-D Flows	Feed-back	PAX Profile	Trajec-tory	Airlines	Airports	PAX			
Access: Public												
DB1B/BTS	[9, 20, 170]	✓	✓				21	414	-	US	1993–present	Quaterly
T-100/BTS	[100]		✓				127	1K	-	US	1990–present	Monthly
Air-TCR/BTS	[135]			✓			31	360	-	US	1998–present	Monthly
OTA/Crawled	[23, 93]	✓					-	-	-	Multi-country	-	-
Skytrax/ Crawled	[80, 85, 175]			✓			492	895	-	Multi-country	2002–present	Daily
AirlineMeals/ Crawled	[50]			✓			745	-	-	Multi-country	2004–present	Daily
Data/OpenSky	[15, 133]		✓			✓	-	-	-	Multi-country	2013–present	Real-time
Access: Commercial												
MIDT/GDSs	[106–109]	✓			✓		-	-	-	Multi-country	-	-
OAG-Schedule/ OAG	[72, 100]		✓				1K	4K	-	Multi-country	1979–present	Real-time
Data/ Skyscanner	[77, 145]	✓					-	-	-	Multi-country	-	Real-time
ASDI/FAA	[13, 15]		✓			✓	-	-	-	North America	1991–present	Real-time
Access: Private												
PEK & CAN/ TravelSky	[98]		✓		✓		63	2	3M	Beijing & Guangzhou, China	(2 years)	-
Data/CAO.IRI	[105]		✓				-	42	-	Iran	2011–2015	-

the core of the air travel market's O-D structure and dynamics [20], and predict demand [9]. Air Carrier Statistics (T-100), also maintained by BTS, includes air passenger flows for U.S. domestic and international markets. BTS disseminates the **Air Travel Consumer Report (TCR)** disclosing flight delays, mishandled luggage, overbooking, and consumer complaints. The TCR is used in a study by Scotti et al. [135] to identify if there is a relationship between factors such as flight delays and mishandled luggage on the rate of consumer complaints. Researchers may crawl information from OTAs, such as Travelocity, Expedia, and Orbitz, a convenient resource for users to compare prices among different airlines. The top three OTAs have an overall 22% market share and 58% eyeball share in the U.S. market. With OTA data, researchers can conduct inter-market analysis utilizing time-series [93] and investigate potential discriminative factors from distributors for or against a specific airline [23]. For studies with a focus on customer feedback, Skytrax and AirlineMeals are two popular sources. Skytrax is a leading web resource collecting customer feedback and ranking concerning global airlines, lounges, and airports. Reviews and ratings crawled from both Skytrax and AirlineMeals have been utilized in sentiment analysis [50, 80], opinion mining about the airline service [175], and traveler satisfaction prediction [85].

Global Distribution Systems (GDSs) including Sabre, Amadeus, and TravelPort, create, maintain, and commercialize their own **Market Information Data Tapes (MIDT)**. Airlines pay millions of dollars each year to purchase this MIDT data made up of valuable **Passenger Name Records (PNRs)**. A PNR is a customer profile created at reservation time by air travel providers and includes passenger-specific details such as gender, carrier, origin, and destination airport. Amadeus publicized several studies proposing advanced DL techniques, including Pointer Networks and **Generative Adversarial Networks (GAN)**, and applying these techniques on a subset of their MIDT data, with records coming from approximately 420 airlines and over 93,000 travel agencies [106–109].

The **Official Airline Guide (OAG)** is a good resource of O-D flows information, providing the most comprehensive airline schedule and flight status from around the world. Both T-100 and OAG have been utilized in the prediction of monthly passenger volumes between directly

connected airports [100]. Another comparative web source known as Skyscanner contains records of user clicks on a chosen price point. Soyk et al. [145] assessed the performance for long-haul LCCs and developed a revenue model by combining the traffic, fare, load factor, and seat data from provided by Skyscanner. Karamshuk and Mathews [77] identified the factors that contribute towards a competitive combination itinerary, also using Skyscanner data.

Although some studies harnessed datasets that appear to be private, their described methodologies provide some valuable insights. TravelSky Technology Limited provided a private dataset to Liu et al. [98] of PNRs for travel records taking place during a two-year period from passengers of two cities in China: Beijing (PEK airport) and Guangzhou (CAN airport). The private flight data provided by **Civil Aviation Organization of Islamic Republic of Iran (CAO.IRI)** segmented by O-D where passenger counts, load factor, and the number of trips, has been used for the development of a model that can predict air travel demand [105].

Aside from the airline data, aircraft trajectories are highly crucial for air traffic control and there exist several public sources. The FAA's **Traffic Flow Management System (TFMS)** provides actual and planned aircraft positions during their flights in **Aircraft Situation Display to Industry (ASDI)** dataset. While the FAA provides data covering North America, EUROCONTROL provides similar data for flights in Europe. Data from FAA and EUROCONTROL have supported much air traffic control research [15, 181], but they are not publicly available and have limited access. Alternatively, **Automatic Dependent Surveillance Broadcast (ADS-B)** technology can be leveraged and used to collect the aircraft trajectories in real-time. OpenSky Network provides public access to the real-time and historical aircraft trajectory data for research purposes, which is the largest public aircraft trajectory database and now covers 40% of the flights [133]. There are many other commercialized ADS-B data sources with broader data coverage, including Flightradar24 and FlightAware. In addition to the trajectories, many other air traffic data, including **Terminal Area Forecasts (TAF)** data, FAA **Aviation System Performance Metrics (ASPM), Airport Collaborative Decision Making (A-CDM)**, and so on, have been used in various air traffic control research [11, 13, 15, 126].

Other data sources that may not have a direct focus on the airline industry but can be complementary and add values to various studies include weather data used as features in various models for air traffic control and management [79, 101, 118, 180], Twitter data utilized to evaluate customer service in the airline industry [102, 165], economic, income, wealth-related data such as **Consumer Price Index (CPI)**, and Census Data [9, 100].

5 FUTURE WORK

Data analytics and ML provide excellent opportunities for the airline industry to improve their products and operations. Despite the vast amounts of available data and the advanced tools that have been developed in the last decades, the airline industry applications, as well as ML techniques used in this domain, are still limited. In this section, we discuss possible future directions and avenues of data analytics research using data from the airline industry. Specifically, future directions are divided into airline applications and advanced ML techniques.

5.1 Airline Applications

5.1.1 Ancillary Service Optimization. Industries using RM have witnessed the significant boost in revenue that is being generated through the sale of ancillary services. The airline industry is a prime example, offering unbundled services such as reserved seats, priority boarding, in-flight food/beverage, checked luggage, and more. Figures 6 and 7 demonstrate the steady increase in revenue for three major air carriers in regards to two essential ancillary-based products: (1) checked

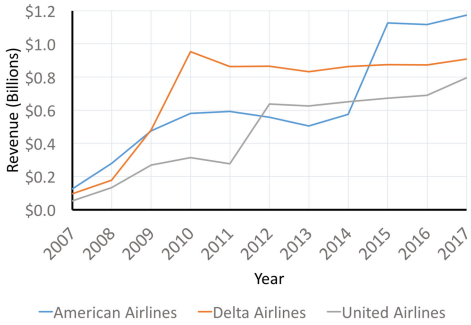

Fig. 6. Baggage fees.

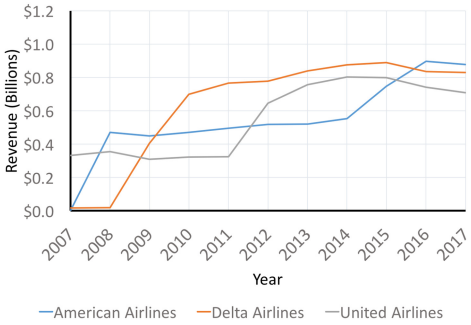

Fig. 7. Reservation change/cancelation fees.

baggage and (2) reservation change or cancelation.[9] However, the prices for these unbundled items are not optimized, causing airlines to miss significant revenue opportunities [56]. With the recent success of unbundled items, optimization models for ancillary revenue are getting an increasing amount of attention. A recent work by Navitaire, known as APO, is a good example of how the industry is starting to see the benefits of applying ML techniques to dynamically decide prices for ancillary products. As Identified by Fiig et al. [56], two possible research directions regarding the ancillary revenue optimization can be suggested for the future work: (1) mixed bundles and (2) correlated reservation prices.

Mixed Bundles: Companies can sell products separately (*a la carte*), as one entity (pure bundles), or as a component mix of two options (mixed bundles). The choice between these categories requires an internal pricing consistency. Mixed bundling has been proven to be an optimal approach to making more revenue rather than relying on pure bundling [2]. Nonetheless, the choice of bundles for maximizing revenue has been shown to be NP-hard when allowing more than two items in the same bundle [46]. Therefore, future research on mixed bundling should focus on more efficient and practical data-driven solutions for maximizing the revenue.

Correlated Reservation Prices: As of today, the fares for a customer's different itineraries are determined simultaneously. This is achieved through the advent of the fare adjustment theory [55]. The current challenge is to determine the prices for all correlated ancillary products simultaneously. A recent study by Bockelie and Belobaba [25] proposed an **Ancillary Choice Model (ACM)** by integrating passenger choice-models for the itinerary, fare class, and ancillary service. ACM serves to define the consumers' selection of specific ancillary services after deciding their preferred airline itinerary along with the fare class. Airlines have recently noted the benefits of employing the decoy effect to better predict customer choice and drive up sales. However, a comprehensive study is demanding to explain how the decoy effect can be modeled for the ancillary pricing optimization and analyze its benefits and limitations. In the future, ML and deep learning can be used to automatically model the passenger choices of ancillary services [107].

5.1.2 Recommender Systems. A recommender system can make suggestions that will appeal to users, giving them support during the decision-making process. It requires rich information on users, items, and the users' shopping patterns, along with domain knowledge of the underlying behavioral process that led to a specific decision. AI-supported recommendation systems are essential for web-applications and OTAs, and must rapidly and accurately infer these patterns from the users [54] in near real-time. By combining merchandising techniques (the way an offer is

[9]Data from BTS, Form 41 Financial Reports.

presented), psychological factors (how the user responds to the offer) and ML methods, companies can train models that generalize the process and automatically predict the users' preferred products and services [56]. A good example is the use of sentiment analysis on consumer reviews to identify the airline services users recommend. A study by Siering et al. [138] provides a look into which airline service consumers pay more attention to and how these factors explain a consumer's recommendation by making use of sentence-level sentiment polarity. Recently, Mottini et al. [109] introduced a novel benchmark on three choice-model methods (traditional, ML-based, and deep learning-based) to identify the most suitable approach for the recommender problem.

5.2 Advanced Machine Learning Techniques

Despite the great success of ML and DL in recent years and the increasing role of AI in many industrial applications, there are still considerably few ML techniques applied to the airline's applications. Nevertheless, many airlines have planned to leverage AI and more advanced ML techniques in the near future to address the existing challenges in dynamic pricing, ancillary service optimization, overbooking, EMSR, WTP, and costumers' feedback analysis.[10] An example of these techniques includes ensemble learning models for automatic air ticket pricing, as suggested by Abdella et al. [1]. In addition, more recently, **Multi-Agent Systems (MAS)** have attracted significant attention in the airline industry [129, 146]. MAS has been studied for two decades to tackle congestion problems in the transportation domain. However, there are still very few research studies using MAS to solve demand and capacity imbalance problems that are essential to solve in the airline industry. In the case of RM, deep RL, combining both DL and RL, has the capability of adapting to a dynamically changing environment. Airline RM could leverage techniques from Reference [96], where a deep multi-agent RL model is developed for maximizing the gross merchandise volume of the large-scale ride-sharing car fleet. Also, the idea of using RL for energy consumption scheduling [81] can be borrowed to solve the issue of dynamic pricing and energy consumption for the airlines. In terms of the importance of fairness in the airline marketplace, DRL can be considered in the future to develop dynamic pricing strategies [99].

Moreover, based on the publications described in Section 2.7.1, it is apparent there are limited numbers of advanced deep learning techniques for sentiment and opinion mining in the airline industry. Sentiment mining for airline reviews can benefit from advanced techniques proposed by Tang et al. [154], which utilized Convolutional Neural Network and LSTM networks on online reviews and even introduced a novel network that integrates the semantic representations of user and product information. Multi-aspect level sentiment analysis also contributes essential information for companies to gain a comprehensive understanding of the customers' perception of their services and products. The recently introduced attention mechanism has been advantageous for aspect-level representations of documents [92, 177].

DL also shows great potential in analyzing time-series data, which are commonly seen in the airline industry. Recently, Li and Cao [95] proposed to use LSTM to predict the tourism flow in local landscapes, while Silva et al. [139] proposed to apply autoregressive neural networks to estimate the tourism demand in multiple countries in Europe. Both methods achieve better performance than conventional methods. In the airline industry, the estimation of customer demand and O-D flow tackles very similar problems, and the performance of these problems can be potentially improved in the future with the help of DL. Moreover, other topics in this industry, such as ticket pricing and simulation tools, are built upon the demand and O-D flow prediction and thus benefit from deep neural networks as well.

[10] https://www.altexsoft.com/blog/datascience/7-ways-how-airlines-use-artificial-intelligence-and-data-science-to-improve-their-operations.

6 CONCLUSION

There is an increasing demand for computer science domain knowledge in the airline industry to conduct adequate data analysis. However, not much work has been done to consolidate this field across the various sectors of the industry. Thus, the airlines have yet to fully realize the benefits of embracing cutting-edge ML techniques for data analytics. To the best of our knowledge, this is the first comprehensive work that presents a multi-perspective look into the details of how ML is being applied to analyze airlines' data.

This article provides a detailed review of the state-of-the-art AI applications for data analytics in the most fundamental aspects of the airline domain. The authors investigate how the introduction of the free market stimulates the development of the airline industry after the Airline Deregulation Act of 1978; moreover, how computer science gradually becomes involved within the industry through the application of advanced ML techniques. Major studies involving critical components of the airline industry were identified along with frameworks and popular techniques in both traditional and ML domains. The most popular ML algorithms and models tested in the airline industry are presented. Moreover, a comprehensive list of datasets and data sources is provided. Although most of the sources in this list are public, several commercial or private data sources are also included based on the impact and frequency of their respective research references.

ML has the potential to deliver substantial impact based on the following challenges and future research directions:

- The traditional analytic approach has dominated the ancillary price optimization problem. Advanced ML methods can provide essential techniques to solve the mixed bundles and correlated reservation prices problem.
- ML recommendation system has been deployed in many fields, such as online retail and video streaming services, with great success. Air travel products and services can benefit the revenue generated through these sophisticated recommendation systems.
- With the help of advanced ML techniques (e.g., deep learning), the airline industry can leverage large amounts of data from different sources to generate better strategies and further improve the overall performance of their operations. Techniques such as deep customer opinion analysis and deep RL can assist the industry to better adapt to the dynamic market.

We hope this survey provides readers with a comprehensive understanding of the relationship between the airline industry and data analytics and sheds light on future research directions and opportunities.

ACKNOWLEDGMENTS

The authors would like to thank Tim Reiz (Chief Technology Officer), David Welborn (Business Intelligence Architect), and Diana Porro (Data Scientist) from Farelogix Inc. for providing input and support.

REFERENCES

[1] Juhar Ahmed Abdella, Nazar Zaki, and Khaled Shuaib. 2018. Automatic detection of airline ticket price and demand: A review. In *Proceedings of the International Conference on Innovations in Information Technology*. IEEE, Piscataway, NJ, 169–174.

[2] William Adams and Janet L. Yellen. 1976. Commodity bundling and the burden of monopoly. *Quart. J. Econ.* 90, 3 (1976), 475–498.

[3] Esi Adeborna and Keng Siau. 2014. An approach to sentiment analysis-the case of airline quality rating. In *Proceedings of the Pacific Asia Conference on Information Systems*. Association for Information Systems, Atlanta, GA, 363.

[4] Nicole Adler and Joseph Berechman. 2001. Measuring airport quality from the airlines' viewpoint: An application of data envelopment analysis. *Transport Policy* 8, 3 (2001), 171–181.

[5] Md Shohel Ahmed, Sameer Alam, and Michael Barlow. 2018. A multi-layer artificial neural network approach for runway configuration prediction. In *Proceedings of the International Conference on Research in Air Transportation*. ICRAT, 8.
[6] Florian Allroggen, Michael D. Wittman, and Robert Malina. 2015. How air transport connects the world—A new metric of air connectivity and its evolution between 1990 and 2012. *Transport. Res. Part E: Logist. Transport. Rev.* 80 (2015), 184–201.
[7] Gershon Alperovich and Yaffa Machnes. 1994. The role of wealth in the demand for international air travel. *J. Transport Econ. Policy* 28, 2 (1994), 163–173.
[8] Jens Alstrup, Søren Boas, Oli B. G. Madsen, and ReneVictor Valqui Vidal. 1986. Booking policy for flights with two types of passengers. *Eur. J. Oper. Res.* 27, 3 (1986), 274–288.
[9] Bo An, Haipeng Chen, Noseong Park, and V. S. Subrahmanian. 2016. MAP: Frequency-based maximization of airline profits based on an ensemble forecasting approach. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 421–430.
[10] Jean-François Arvis and Ben Shepherd. 2011. *The Air Connectivity Index: Measuring Integration in the Global Air Transport Network*. The World Bank, Washington, DC.
[11] Jacob Avery and Hamsa Balakrishnan. 2015. Predicting airport runway configuration: A discrete-choice modeling approach. In *Proceedings of the 11th USA/Europe Air Traffic Management Research and Development Seminar*. Federal Aviation Administration/EUROCONTROL, 23–26.
[12] Samet Ayhan, Pablo Costas, and Hanan Samet. 2019. A data-driven framework for long-range aircraft conflict detection and resolution. *ACM Trans. Spat. Algor. Syst.* 5, 4 (2019), 1–23.
[13] Samet Ayhan and Hanan Samet. 2016. Aircraft trajectory prediction made easy with predictive analytics. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 21–30.
[14] Poornima Balakrishna, Rajesh Ganesan, and Lance Sherry. 2010. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. *Transport. Res. Part C: Emerg. Technol.* 18, 6 (2010), 950–962.
[15] Alevizos Bastas, Theocharis Kravaris, and George A. Vouros. 2020. Data driven aircraft trajectory prediction with deep imitation learning. *CoRR* abs/2005.07960 (2020), 20.
[16] Martin J. Beckmann. 1958. Decision and team problems in airline reservations. *Econometrica* 26, 1 (1958), 134–145.
[17] Peter P. Belobaba. 1987. Survey paper—Airline yield management an overview of seat inventory control. *Transport. Sci.* 21, 2 (1987), 63–73.
[18] P. P. Belobaba and C. Hopperstad. 1999. Boeing/MIT simulation study: PODS results update. In *Proceedings of the AGIFORS Reservations and Yield Management Study Group Symposium*. AGIFORS, Atlanta, GA.
[19] Dimitris Bertsimas and Sanne De Boer. 2005. Simulation-based booking limits for airline revenue management. *Oper. Res.* 53, 1 (2005), 90–106.
[20] Dipasis Bhadra and Jacqueline Kee. 2008. Structure and dynamics of the core US air travel markets: A basic empirical analysis of domestic passenger demand. *J. Air Transport Manag.* 14, 1 (2008), 27–39.
[21] Dipasis Bhadra and Michael Wells. 2005. Air travel by state: Its determinants and contributions in the United States. *Pub. Works Manag. Policy* 10, 2 (2005), 119–137.
[22] Volodymyr Bilotkach, Alberto A. Gaggero, and Claudio A. Piga. 2015. Airline pricing under different market conditions: Evidence from European low-cost carriers. *Tour. Manag.* 47 (2015), 152–163.
[23] Volodymyr Bilotkach and Marija Pejcinovska. 2012. Distribution of airline tickets: A tale of two market structures. In *Pricing Behavior and Non-price Characteristics in the Airline Industry*. Emerald Group Publishing Limited, Bingley, UK, 107–138.
[24] John Bitzan and James Peoples. 2016. A comparative analysis of cost change for low-cost, full-service, and other carriers in the US airline industry. *Res. Transport. Econ.* 56 (2016), 25–41.
[25] Adam Bockelie and Peter Belobaba. 2017. Incorporating ancillary services in airline passenger choice models. *J. Rev. Pricing Manag.* 16, 6 (2017), 553–568.
[26] Brent D. Bowen, Dean E. Headley, and Jacqueline R. Luedtke. 1991. *Airline Quality Rating 1991*. Technical Report. Wichita State University.
[27] E. Andrew Boyd. 2007. *The Future of Pricing: How Airline Ticket Pricing Has Inspired a Revolution*. Palgrave Macmillan, New York, NY.
[28] Shelby L. Brumelle and Jeffrey I. McGill. 1993. Airline seat allocation with multiple nested fare classes. *Oper. Res.* 41, 1 (1993), 127–137.
[29] B. Burger and M. Fuchs. 2005. Dynamic pricing—A future airline business model. *J. Rev. Pricing Manag.* 4, 1 (2005), 39–53.

[30] Guillaume Burghouwt and Renato Redondi. 2013. Connectivity in air transport networks: An assessment of models and applications. *J. Transport Econ. Policy* 47, 1 (2013), 35–53.
[31] Jorge Cardoso and Carola Lange. 2007. A framework for assessing strategies and technologies for dynamic packaging applications in e-tourism. *Inf. Technol. Tour.* 9, 1 (2007), 27–44.
[32] Emmanuel Carrier. 2003. *Modeling Airline Passenger Choice: Passenger Preference for Schedule in the Passenger Origin-destination Simulator (PODS).* Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA.
[33] Rafael Casado and Aurelio Bermúdez. 2020. Neural network-based aircraft conflict prediction in final approach maneuvers. *Electronics* 9, 10 (2020), 1708.
[34] Yu-Hern Chang and Chung-Hsing Yeh. 2002. A survey analysis of service quality for domestic airlines. *Eur. J. Oper. Res.* 139, 1 (2002), 166–177.
[35] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16 (2002), 321–357.
[36] Jun Chen, Michal Weiszer, Elham Zareian, Mahdi Mahfouf, and Olusayo Obajemu. 2017. Multi-objective fuzzy rule-based prediction and uncertainty quantification of aircraft taxi time. In *Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems*. IEEE, Piscataway, NJ, 1–5.
[37] Junwook Chi and Jungho Baek. 2012. A dynamic demand analysis of the united states air-passenger service. *Transport. Res. Part E: Logist. Transport. Rev.* 48, 4 (2012), 755–761.
[38] Sun Choi, Young Jin Kim, Simon Briceno, and Dimitri Mavris. 2016. Prediction of weather-induced airline delays based on machine learning algorithms. In *Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference*. IEEE, Piscataway, NJ, 1–6.
[39] Jin Young Chung and James F. Petrick. 2013. Price fairness of airline ancillary fees: An attributional approach. *J. Trav. Res.* 52, 2 (2013), 168–181.
[40] Andrew Collins and Lyn Thomas. 2012. Comparing reinforcement learning approaches for solving game theoretic models: A dynamic airline pricing game example. *J. Oper. Res. Soc.* 63, 8 (2012), 1165–1173.
[41] Andrew Collins and Lyn Thomas. 2013. Learning competitive dynamic airline pricing under different customer models. *J. Rev. Pricing Manag.* 12, 5 (2013), 416–430.
[42] Renwick E. Curry. 1990. Optimal airline seat allocation with fare classes nested by origins and destinations. *Transport. Sci.* 24, 3 (1990), 193–204.
[43] Yang Dai, Robert Raeside, and Austin Smyth. 2005. The use of load factors to segment airline operators. *J. Rev. Pricing Manag.* 4, 2 (2005), 195–203.
[44] Thierry Delahaye, Rodrigo Acuna-Agost, Nicolas Bondoux, Anh-Quan Nguyen, and Mourad Boudia. 2017. Data-driven models for itinerary preferences of air travelers and application for dynamic pricing optimization. *J. Rev. Pricing Manag.* 16, 6 (2017), 621–639.
[45] Arnoud V. den Boer. 2015. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surv. Oper. Res. Manag. Sci.* 20, 1 (2015), 1–18.
[46] Loc Do, Hady W. Lauw, and Ke Wang. 2015. Mining revenue-maximizing bundling configuration. *Proc. VLDB Endow.* 8, 5 (2015), 593–604.
[47] Frédéric Dobruszkes. 2006. An analysis of european low-cost airlines and their networks. *J. Transport Geog.* 14, 4 (2006), 249–264.
[48] Sara Dolnicar, Klaus Grabler, Bettina Grün, and Anna Kulnig. 2011. Key drivers of airline loyalty. *Tour. Manag.* 32, 5 (2011), 1020–1026.
[49] Goda R. Doreswamy, Aditya S. Kothari, and Sumala Tirumalachetty. 2015. Simulating the flavors of revenue management for airlines. *J. Rev. Pricing Manag.* 14, 6 (2015), 421–432.
[50] Brett Drury, Luís Torgo, and Jose Joao Almeida. 2011. Guided self-training for sentiment classification. In *Proceedings of the Workshop on Robust Unsupervised and Semisupervised Methods in Natural Language Processing*. ACL, 9–16.
[51] Hugh Dunleavy and Glen Phillips. 2009. The future of airline revenue management. *J. Rev. Pricing Manag.* 8, 4 (2009), 388–395.
[52] Oren Etzioni, Rattapoom Tuchinda, Craig A. Knoblock, and Alexander Yates. 2003. To buy or not to buy: Mining airfare data to minimize ticket purchase price. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 119–128.
[53] Antony D. Evans, Paul Lee, and Banavar Sridhar. 2018. Predicting the operational acceptance of airborne flight reroute requests using data mining. *Transport. Res. Part C: Emerg. Technol.* 96 (2018), 270–289.
[54] Pierluca Ferraro and Giuseppe Lo Re. 2014. Designing ontology-driven recommender systems for tourism. In *Advances onto the Internet of Things: How Ontologies Make the Internet of Things Meaningful*. Springer, Cham, Switzerland, 339–352.
[55] Thomas Fiig, Karl Isler, Craig Hopperstad, and Peter Belobaba. 2010. Optimization of mixed fare structures: Theory and applications. *J. Rev. Pricing Manag.* 9, 1–2 (2010), 152–170.

[56] Thomas Fiig, Remy Le Guen, and Mathilde Gauchet. 2018. Dynamic pricing of airline offers. *J. Rev. Pricing Manag.* 17, 6 (2018), 381–393.
[57] Michael Frank, Martin Friedemann, Michael Mederer, and Anika Schroeder. 2006. Airline revenue management: A simulation of dynamic capacity management. *J. Rev. Pricing Manag.* 5, 1 (2006), 62–71.
[58] Xiaowen Fu, Mark Lijesen, and Tae H. Oum. 2006. An analysis of airport pricing and regulation in the presence of competition between full service airlines and low cost carriers. *J. Transport Econ. Policy* 40, 3 (2006), 425–447.
[59] Laurie Garrow and Virginie Lurkin. 2021. How COVID-19 is impacting and reshaping the airline industry. *J. Rev. Pricing Manag.* 20, 1 (2021), 3–9.
[60] Laurie A. Garrow, Susan Hotle, and Stacey Mumbower. 2012. Assessment of product debundling trends in the US airline industry: Customer service and public policy implications. *Transport. Res. Part A: Policy Pract.* 46, 2 (2012), 255–268.
[61] Harris Georgiou, Nikos Pelekis, Stylianos Sideridis, David Scarlatti, and Yannis Theodoridis. 2020. Semantic-aware aircraft trajectory prediction using flight plans. *Int. J. Data Sci. Analyt.* 9, 2 (2020), 215–228.
[62] David Gianazza and Kévin Guittet. 2006. Evaluation of air traffic complexity metrics using neural networks and sector status. In *Proceedings of the International Conference on Research in Air Transportation*. ICRAT, 126–136.
[63] Fred Glover, Randy Glover, Joe Lorenzo, and Claude McMillan. 1982. The passenger-mix problem in the scheduled airlines. *Interfaces* 12, 3 (1982), 73–80.
[64] David González Prieto. 2017. *Enhancing the Profitability of Airline Tickets Purchasing Processes through Contextual Effects: A Study of Decoy Effect*. Technical Report. Universitat Politècnica de Catalunya, Barcelona, Spain.
[65] Ram Gopalan and Kalyan T. Talluri. 1998. Mathematical models in airline schedule planning: A survey. *Ann. Oper. Res.* 76 (1998), 155–185.
[66] Abhijit Gosavi, Emrah Ozkaya, and Aykut F. Kahraman. 2007. Simulation optimization for revenue management of airlines with cancellations and overbooking. *OR Spect.* 29, 1 (2007), 21–38.
[67] Abhuit Gosavii, Naveen Bandla, and Tapas K. Das. 2002. A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. *IIE Trans.* 34, 9 (2002), 729–742.
[68] Michaela Graf and Alf Kimms. 2013. Transfer price optimization for option-based airline alliance revenue management. *Int. J. Product. Econ.* 145, 1 (2013), 281–293.
[69] William Groves and Maria Gini. 2015. On optimizing airline ticket purchase timing. *ACM Trans. Intell. Syst. Technol.* 7, 1 (2015), 3.
[70] Priyanga Gunarathne, Huaxia Rui, and Abraham Seidmann. 2018. When social media delivers customer service: Differential customer treatment in the airline industry. *Manag. Inf. Syst. Quart.* 42, 2 (2018), 489–520.
[71] Dogan Gursoy, Ming-Hsiang Chen, and Hyun Jeong Kim. 2005. The US airlines relative positioning based on attributes of service quality. *Tour. Manag.* 26, 1 (2005), 57–67.
[72] Mark Hansen. 1990. Airline competition in a hub-dominated environment: An application of noncooperative game theory. *Transport. Res. Part B: Methodol.* 24, 1 (1990), 27–43.
[73] Floris Herrema, Richard Curran, Hendrikus Visser, Denis Huet, and Régis Lacote. 2018. Taxi-out time prediction model at Charles de Gaulle Airport. *J. Aerosp. Inf. Syst.* 15, 3 (2018), 120–130.
[74] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, 4572–4580.
[75] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computat.* 9, 8 (1997), 1735–1780.
[76] J. M. Jung and E. T. Fujii. 1976. The price elasticity of demand for air travel: Some new evidence. *J. Transport Econ. Policy* 10, 3 (1976), 257–262.
[77] Dmytro Karamshuk and David Matthews. 2018. Learning cheap and novel flight itineraries. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Cham, Switzerland, 288–304.
[78] Wandeep Kaur and Vimala Balakrishnan. 2018. Improving sentiment scoring mechanism: A case study on airline services. *Industr. Manag. Data Syst.* 118, 8 (2018), 1578–1596.
[79] Christian Strottmann Kern, Ivo Paixao de Medeiros, and Takashi Yoneyama. 2015. Data-driven aircraft estimated time of arrival prediction. In *Proceedings of the IEEE Systems Conference*. IEEE, Piscataway, NJ, 727–733.
[80] Aurangzeb Khan, Baharum Baharudin, and Khairullah Khan. 2010. Sentence based sentiment classification from online customer reviews. In *Proceedings of the 8th International Conference on Frontiers of Information Technology*. ACM, New York, NY, 317–331.
[81] Byung-Gook Kim, Yu Zhang, Mihaela Van Der Schaar, and Jang-Won Lee. 2016. Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Trans. Smart Grid* 7, 5 (2016), 2187–2198.
[82] Young Jin Kim, Sun Choi, Simon Briceno, and Dimitri Mavris. 2016. A deep learning approach to flight delay prediction. In *Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference*. IEEE, Piscataway, NJ, 1–6.

[83] Parimal Kopardekar and Sherri Magyarits. 2002. Dynamic density: Measuring and predicting sector complexity [ATC]. In *Proceedings of the 21st Digital Avionics Systems Conference*, Vol. 1. IEEE, Piscataway, NJ, 2C4–2C4.
[84] L. Kosten. 1960. Een mathematisch model voor een reserveringsprobleem. *Statist. Neerland.* 14, 1 (1960), 85–94.
[85] Emanuel Lacic, Dominik Kowald, and Elisabeth Lex. 2016. High enough? Explaining and predicting traveler satisfaction using airline reviews. In *Proceedings of the ACM Conference on Hypertext and Social Media*. ACM, New York, NY, 249–254.
[86] Tiruvarur R. Lakshmanan. 2011. The broader economic consequences of transport infrastructure investments. *J. Transport Geog.* 19, 1 (2011), 1–12.
[87] Conrad J. Lautenbacher and Shaler Stidham Jr. 1999. The underlying Markov decision process in the single-leg airline yield-management problem. *Transport. Sci.* 33, 2 (1999), 136–146.
[88] Richard D. Lawrence. 2003. A machine-learning approach to optimal bid pricing. In *Computational Modeling and Problem Solving in the Networked World*. Springer, Boston, MA, 97–118.
[89] Tak C. Lee and Marvin Hersh. 1993. A model for dynamic airline seat inventory control with multiple seat bookings. *Transport. Sci.* 27, 3 (1993), 252–265.
[90] Alix Lhéritier. 2019. PCMC-Net: Feature-based pairwise choice Markov chains. *CoRR* abs/1909.11553 (2019), 11.
[91] Alix Lhéritier, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. 2019. Airline itinerary choice modeling using machine learning. *J. Choice Model.* 31 (2019), 198–209.
[92] Cheng Li, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deep memory networks for attitude identification. In *Proceedings of the ACM International Conference on Web Search and Data Mining*. ACM, New York, NY, 671–680.
[93] Jun Li, Nelson Granados, and Serguei Netessine. 2014. Are consumers strategic? Structural estimation from the air-travel industry. *Manag. Sci.* 60, 9 (2014), 2114–2137.
[94] Nan Li, Qing-Yu Jiao, Lei Zhang, and Shao-Cong Wang. 2020. Using deep learning method to predict taxi time of aircraft: A case of Hong Kong Airport. *J. Aeron., Astron. Aviat.* 52, 4 (2020), 371–385.
[95] YiFei Li and Han Cao. 2018. Prediction for tourism flow based on LSTM neural network. *Proced. Comput. Sci.* 129 (2018), 277–283.
[96] Kaixiang Lin, Renyu Zhao, Zhe Xu, and Jiayu Zhou. 2018. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, NY, 1774–1783.
[97] Kenneth Littlewood. 1972. Forecasting and control of passenger bookings. *Airline Group Int. Feder. Oper. Res. Soc. Proc.* 12 (1972), 95–117.
[98] Jie Liu, Bin Liu, Yanchi Liu, Huipeng Chen, Lina Feng, Hui Xiong, and Yalou Huang. 2018. Personalized air travel prediction: A multi-factor perspective. *ACM Trans. Intell. Syst. Technol.* 9, 3 (2018), 30:1–30:26.
[99] Roberto Maestre, Juan Duque, Alberto Rubio, and Juan Arévalo. 2018. Reinforcement learning for fair dynamic pricing. In *Proceedings of the SAI Intelligent Systems Conference*. Springer, Cham, Switzerland, 120–135.
[100] Liang Mao, Xiao Wu, Zhuojie Huang, and Andrew J. Tatem. 2015. Modeling monthly flows of global air travel passengers: An open-access data resource. *J. Transport Geog.* 48 (2015), 52–60.
[101] Michael V. McCrea, Hanif D. Sherali, and Antonio A. Trani. 2008. A probabilistic framework for weather-based rerouting and delay estimations within an airspace planning model. *Transport. Res. Part C: Emerg. Technol.* 16, 4 (2008), 410–431.
[102] Fotis Misopoulos, Miljana Mitic, Alexandros Kapoulas, and Christos Karapiperis. 2014. Uncovering customer service experiences with Twitter: The case of airline industry. *Manag. Decis.* 52, 4 (2014), 705–723.
[103] Baskar Mohan. 2005. *Integrated Pricing and Seat Allocation for Airline Network Revenue Management*. Master's thesis. University of South Florida.
[104] Stéphane Mondoloni and Nicholas Rozen. 2020. Aircraft trajectory prediction and synchronization for air traffic management applications. *Prog. Aerosp. Sci.* 119 (2020), 100640.
[105] Ali Mostafaeipour, Alireza Goli, and Mojtaba Qolipour. 2018. Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: A case study. *J. Supercomput.* 74, 10 (2018), 5461–5484.
[106] Alejandro Mottini and Rodrigo Acuna-Agost. 2016. Relative label encoding for the prediction of airline passenger nationality. In *Proceedings of the IEEE International Conference on Data Mining Workshops*. IEEE, Piscataway, NJ, 671–676.
[107] Alejandro Mottini and Rodrigo Acuna-Agost. 2017. Deep choice model using pointer networks for airline itinerary prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 1575–1583.
[108] Alejandro Mottini, Alix Lheritier, and Rodrigo Acuna-Agost. 2018. Airline passenger name record generation using generative adversarial networks. *CoRR* abs/1807.06657 (2018), 1–9.
[109] Alejandro Mottini, Alix Lhéritier, Rodrigo Acuna-Agost, and Maria A. Zuluaga. 2018. Understanding customer choices to improve recommendations in the air travel industry. In *Proceedings of the Workshop on Recommenders in Tourism*. Central Europe Workshop Proceedings, 28–32.

[110] Stacey Mumbower, Laurie A. Garrow, and Matthew J. Higgins. 2014. Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. *Transport. Res. Part A: Policy Pract.* 66 (2014), 196–212.
[111] Stacey Mumbower, Laurie A. Garrow, and Jeffrey P. Newman. 2015. Investigating airline customers' premium coach seat purchases and implications for optimal pricing strategies. *Transport. Res. Part A: Policy Pract.* 73 (2015), 53–69.
[112] Serguei Netessine and Robert A. Shumsky. 2005. Revenue management games: Horizontal and vertical competition. *Manag. Sci.* 51, 5 (2005), 813–831.
[113] Ann Norman. 2004. Spatial interaction modelling: A regional science context. *J. Econ. Lit.* 42, 3 (2004), 986.
[114] Kofi Obeng and Ryoichi Sakano. 2012. Airline fare and seat management strategies with demand dependency. *J. Air Transport Manag.* 24 (2012), 42–48.
[115] John F. O'Connell and George Williams. 2016. Ancillary revenues: The new trend in strategic airline marketing. In *Air Transport in the 21st Century*. Routledge, London, UK, 195–220.
[116] Ignacio Olmeda and Pauline J. Sheldon. 2002. Data mining techniques and applications for tourism internet marketing. *J. Trav. Tour. Market.* 11, 2–3 (2002), 1–20.
[117] Yongzhen Arthur Pan, Mario A. Nascimento, and Joerg Sander. 2019. Online stochastic prediction of mid-flight aircraft trajectories. In *Proceedings of the ACM SIGSPATIAL International Workshop on Computational Transportation Science*. ACM, New York, NY, 1–10.
[118] Yutian Pang and Yongming Liu. 2020. Conditional generative adversarial networks (CGAN) for aircraft trajectory prediction considering weather effects. In *Proceedings of the AIAA Scitech Forum*. AIAA, 1853.
[119] Changkyu Park and Junyong Seo. 2011. Seat inventory control for sequential multiple flights with customer choice behavior. *Comput. Industr. Eng.* 61, 4 (2011), 1189–1199.
[120] Brian Pearce. 2013. *Profitability and the Air Transport Value Chain*. Technical Report. International Air Transport Association.
[121] Jorge Vicente Pérez-Rodríguez, José María Pérez-Sánchez, and Emilio Gómez-Déniz. 2017. Modelling the asymmetric probabilistic delay of aircraft arrival. *J. Air Transport Manag.* 62 (2017), 90–98.
[122] James Piggott. 2015. *Identification of Business Travelers Through Clustering Algorithms*. Master's thesis. University of Twente.
[123] Lisa Pritscher and Hans Feyen. 2001. Data mining and strategic marketing in the airline industry. *Data Mining Market. Applic.* 39 (2001), 39–48.
[124] J. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Francisco, CA.
[125] N. J. Radcliffe. 2007. Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Market J. Direct Market Assoc. Anal. Counc.* 1 (2007), 14–21.
[126] Krishnakumar Ramamoorthy and George Hunter. 2012. An empirical airport configuration prediction model. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference*. AIAA, 4488.
[127] Richard Ratliff and Ben Vinod. 2005. Future of revenue management: Airline pricing and revenue management: A future outlook. *J. Rev. Pricing Manag.* 4, 3 (2005), 302–307.
[128] Stefan Ravizza, Jason A. D. Atkin, Marloes H. Maathuis, and Edmund K. Burke. 2013. A combined statistical approach and ground movement model for improving taxi time estimations at airports. *J. Oper. Res. Soc.* 64, 9 (2013), 1347–1360.
[129] Luis Reis, Ana Paula Rocha, and Antonio J. M. Castro. 2018. An electronic marketplace for airlines. In *Proceedings of the International Conference on Practical Applications of Agents and Multi-agent Systems*. Springer, Cham, Switzerland, 60–71.
[130] Marvin Rothstein. 1968. *Stochastic Models for Airline Booking Policies*. Ph.D. Dissertation. New York University, New York, NY.
[131] Marvin Rothstein. 1971. An airline overbooking model. *Transport. Sci.* 5, 2 (1971), 180–192.
[132] Megan S. Ryerson and Hyun Kim. 2014. The impact of airline mergers and hub reorganization on aviation fuel consumption. *J. Clean. Product.* 85 (2014), 395–407.
[133] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, and Matthias Wilhelm. 2014. Bringing up OpenSky: A large-scale ADS-B sensor network for research. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, IEEE, Piscataway, NJ, 83–94.
[134] Moritz Ferdinand Scharpenseel. 2001. Consequences of EU airline deregulation in the context of the global aviation market. *J. Int. Law Bus.* 22, 1 (2001), 91–116.
[135] Davide Scotti, Martin Dresner, and Gianmaria Martini. 2016. Baggage fees, operational performance and customer satisfaction in the US air transport industry. *J. Air Transport Manag.* 55 (2016), 139–146.
[136] Syed Arbab Mohd Shihab, Caleb Logemann, Deepak-George Thomas, and Peng Wei. 2019. Towards the next generation airline revenue management: A deep reinforcement learning approach to seat inventory control and overbooking. *CoRR* abs/1902.06824 (2019), 9.
[137] E. Shlifer and Y. Vardi. 1975. An airline overbooking policy. *Transport. Sci.* 9, 2 (1975), 101–114.

[138] Michael Siering, Amit V. Deokar, and Christian Janze. 2018. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decis. Supp. Syst.* 107 (2018), 52–63.
[139] Emmanuel Sirimal Silva, Hossein Hassani, Saeed Heravi, and Xu Huang. 2019. Forecasting tourism demand with denoised neural networks. *Ann. Tour. Res.* 74 (2019), 134–154.
[140] Ioannis Simaiakis and Hamsa Balakrishnan. 2016. A queuing model of the airport departure process. *Transport. Sci.* 50, 1 (2016), 94–109.
[141] Göran Skugge. 2004. Growing effective revenue managers. *J. Rev. Pricing Manag.* 3, 1 (2004), 49–61.
[142] Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. ACM, New York, NY, 293–304.
[143] Joseph B. Sobieralski. 2020. COVID-19 and airline employment: Insights from historical uncertainty shocks to the industry. *Transport. Res. Interdisc. Perspect.* 5 (2020).
[144] Murati Somboon and Kannapha Amaruchkul. 2016. Combined overbooking and seat inventory control for two-class revenue management model. *Songklanakarin J. Sci. Technol.* 38, 6 (2016), 657–665.
[145] Christian Soyk, Jürgen Ringbeck, and Stefan Spinler. 2018. Revenue characteristics of long-haul low-cost carriers (LCCs) and differences to full-service network carriers (FSNCs). *Transport. Res. Part E: Logist. Transport. Rev.* 112 (2018), 47–65.
[146] Christos Spatharis, Theocharis Kravaris, George A. Vouros, Konstantinos Blekas, Georgios Chalkiadakis, Jose Manuel Cordero Garcia, and Esther Calvo Fernandez. 2018. Multiagent reinforcement learning methods to resolve demand capacity balance problems. In *Proceedings of the Hellenic Conference on Artificial Intelligence*. ACM, New York, NY.
[147] Fie Sternberg, Kasper Hedegaard Pedersen, Niklas Klve Ryelund, Raghava Rao Mukkamala, and Ravi Vatrapu. 2018. Analysing customer engagement of Turkish airlines using big social data. In *Proceedings of the 7th IEEE International Congress on Big Data*. IEEE, Piscataway, NJ, 74–81.
[148] Pere Suau-Sanchez, Augusto Voltes-Dorta, and Héctor Rodríguez-Déniz. 2015. Regulatory airport classification in the US: The role of international markets. *Transport Policy* 37 (2015), 157–166.
[149] Janakiram Subramanian, Shaler Stidham Jr., and Conrad J. Lautenbacher. 1999. Airline yield management with overbooking, cancellations, and no-shows. *Transport. Sci.* 33, 2 (1999), 147–167.
[150] Yoshinori Suzuki. 2000. The relationship between on-time performance and airline market share: A new approach. *Transport. Res. Part E: Logist. Transport. Rev.* 36, 2 (2000), 139–154.
[151] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computat. Ling.* 37, 2 (2011), 267–307.
[152] Kalyan Talluri and Garrett Van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Manag. Sci.* 50, 1 (2004), 15–33.
[153] Kalyan T. Talluri and Garrett J. Van Ryzin. 2006. *The Theory and Practice of Revenue Management*. Vol. 68. Springer Science & Business Media, New York, NY.
[154] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 1422–1432.
[155] Jun Tang. 2019. Conflict detection and resolution for civil aviation: A literature survey. *IEEE Aerosp. Electron. Syst. Mag.* 34, 10 (2019), 20–35.
[156] H. R. Thompson. 1961. Statistical problems in airline reservation control. *Oper. Res. Quart.* 12, 3 (1961), 167–185.
[157] Yongxue Tian and Li Pan. 2015. Predicting short-term traffic flow by long short-term memory recurrent neural network. In *Proceedings of the IEEE International Conference on Smart City/SocialCom/SustainCom*. IEEE, Piscataway, NJ, 153–158.
[158] Sven Tuzovic, Merlin C. Simpson, Volker G. Kuppelwieser, and Jörg Finsterwalder. 2014. From "free" to fee: Acceptability of airline ancillary fees and the effects on customer behavior. *J. Retail. Consum. Serv.* 21, 2 (2014), 98–107.
[159] Garrett Van Ryzin and Jeff McGill. 2000. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Manag. Sci.* 46, 6 (2000), 760–775.
[160] Garrett J. van Ryzin. 2005. Future of revenue management: Models of demand. *J. Reven. Pricing Manag.* 4, 2 (2005), 204–210.
[161] G. Vinodhini and R. M. Chandrasekaran. 2014. Opinion mining using principal component analysis-based ensemble model for e-commerce application. *CSI Trans. ICT* 2, 3 (2014), 169–179.
[162] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. Curran Associates, Inc., Red Hook, NY, 2692–2700.
[163] Timothy M. Vowles. 2001. The "southwest effect" in multi-airport regions. *J. Air Transport Manag.* 7, 4 (2001), 251–258.

[164] Blaise Waguespack and Tamilla Curtis. 2015. Ancillary revenue and price fairness: An exploratory study pre and post flight. *Int. J. Aviat. Manag.* 2, 3–4 (2015), 208–225.
[165] Yun Wan and Qigang Gao. 2015. An ensemble sentiment classification system of Twitter data for airline services analysis. In *Proceedings of the IEEE International Conference on Data Mining Workshop*. IEEE, Piscataway, NJ, 1318–1325.
[166] Xinwei Wang, Alexander E. I. Brownlee, John R. Woodward, Michal Weiszer, Mahdi Mahfouf, and Jun Chen. 2021. Aircraft taxi time prediction: Feature importance and their implications. *Transport. Res. Part C: Emerg. Technol.* 124 (2021), 102892.
[167] Yuan Wang and Yu Zhang. 2021. Prediction of runway configurations and airport acceptance rates for multi-airport system using gridded weather forecast. *Transport. Res. Part C: Emerg. Technol.* 125 (2021), 103049.
[168] Zhuang Wang, Hui Li, Junfeng Wang, and Feng Shen. 2019. Deep reinforcement learning based conflict detection and resolution in air traffic control. *IET Intell. Transport Syst.* 13, 6 (2019), 1041–1047.
[169] David Warnock-Smith, John F. O'Connell, and Mahnaz Maleki. 2017. An analysis of ongoing trends in airline ancillary revenues. *J. Air Transport Manag.* 64 (2017), 42–54.
[170] Wenbin Wei and Mark Hansen. 2005. Impact of aircraft size and seat availability on airlines demand and market share in duopoly markets. *Transport. Res. Part E: Logist. Transport. Rev.* 41, 4 (2005), 315–327.
[171] Richard Robert Wickham. 1995. *Evaluation of Forecasting Techniques for Short-term Demand of Air Transportation*. Technical Report. Massachusetts Institute of Technology.
[172] Hartmut Wolf, Peter Forsyth, David Gillen, Kai Hüschelrath, and Hans-Martin Niemeier. 2016. À la carte pricing to generate ancillary revenue: The case of Ryanair. In *Liberalization in Aviation*. Routledge, UK, 207–216.
[173] Jehn-Yih Wong and Pi-Heng Chung. 2008. Retaining passenger loyalty through data mining: A case study of Taiwanese airlines. *Transport. J.* 47, 1 (2008), 17–29.
[174] Hua Xie, Minghua Zhang, Jiaming Ge, Xinfang Dong, and Haiyan Chen. 2021. Learning air traffic as images: A deep convolutional neural network for airspace operation complexity evaluation. *Complexity* 2021 (2021).
[175] Boyan Yao, Hua Yuan, Yu Qian, and Liangqiang Li. 2015. On exploring airline service features from massive online review. In *Proceedings of the International Conference on Service Systems and Service Management*. IEEE, Piscataway, NJ, 1–6.
[176] Bee Yee Liau and Pei Pei Tan. 2014. Gaining customer knowledge in low cost airlines through text mining. *Industr. Manag. Data Syst.* 114, 9 (2014), 1344–1359.
[177] Yichun Yin, Yangqiu Song, and Ming Zhang. 2017. Document-level multi-aspect sentiment classification as machine comprehension. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 2044–2054.
[178] Oren Zamir and Oren Etzioni. 1998. Web document clustering: A feasibility demonstration. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 46–54.
[179] Chunxiao Zhang, Congrong Guo, and Shenghui Yi. 2014. Airline overbooking problem with uncertain no-shows. *J. Appl. Math.* 2014 (2014).
[180] Jing Zhang, X. H. Xu, Fei Wang, and Dong-xuan Wei. 2010. Airport delay performance evaluation based on fuzzy linear regression model. *J. Traff. Transport. Eng.* 10, 4 (2010), 109–114.
[181] Ziyu Zhao, Weili Zeng, Zhibin Quan, Mengfei Chen, and Zhao Yang. 2019. Aircraft trajectory prediction using deep long short-term memory networks. In *Proceedings of the 19th COTA International Conference of Transportation Professionals*. American Society of Civil Engineers, 124–135.
[182] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. 2017. Machine learning on big data: Opportunities and challenges. *Neurocomputing* 237 (2017), 350–361.
[183] Zhenran Zhu, Anming Zhang, Yahua Zhang, Zhibin Huang, and Shiteng Xu. 2019. Measuring air connectivity between China and Australia. *J. Transport Geog.* 74 (2019), 359–370.

Received May 2019; revised April 2021; accepted May 2021