Forecasting the Airfare prices: Evidence from Online Touristic Agency Company

Dounias George — *University of Aegean*[*]

Chrysovalantis Vasilakis — *Bangor Business School, IZA*[†]

May 27, 2025

Abstract

We conducted research to find the best method for predicting airfare prices using a large dataset provided by an Online Touristic Agency. The data covered the period from 2011 to 2014 and included information on flight details, such as destinations, airlines, and ticket prices. Our goal was to create a helpful tool for internet users who want to know the best time to purchase flight tickets for a specific date.We employed various machine learning techniques, including linear regression models, neural networks, hybrid genetic algorithms, and support vector machines. After analyzing the results, we discovered that a hybrid model that combined neural networks and genetic algorithms yielded the most accurate predictions among all the techniques we tested.The findings of our study could offer a valuable service to the Online Touristic Agency, enabling them to provide a tool that helps customers make informed decisions about when to buy flight tickets. This tool could give the agency a competitive advantage and serve as a marketing strategy to attract more customers.

Keywords: Airplane tickets, forecasting, flights, big data

JEL Codes: C6,I140

[*]University of Aegean, Financial Engineering Department; e-mail: g.dounias@aegean.gr

[†]e-mail: c.vasilakis@bangor.ac.uk.

1 Introduction

Tourism is a cornerstone of the global economy, supporting 319 million jobs and contributing $8.8 trillion (10.4% of global GDP) in 2018 (Scott,2019, Scott,2022). By 2029, the World Travel & Tourism Council projects tourism to generate 421 million jobs globally (James,2021). In many developing countries, particularly in remote regions, tourism is a vital economic driver, heavily reliant on a steady influx of visitors (Bond, 1972, Shin, 2022). Air transport plays a critical role in facilitating this influx, yet its pricing dynamics pose significant challenges for both consumers and travel agencies. Airlines employ dynamic pricing strategies, often charging higher fares for last-minute bookings, typically associated with business travelers willing to pay a premium (Shin, 2022). Prices also fluctuate daily due to limited capacity and uncertain demand, with fares rising to prevent early sell-outs or falling after periods of low demand (Shin, 2022). This volatility raises a critical question: what is the most effective forecasting methodology for travel agencies to accurately predict airline ticket prices and provide reliable guidance to customers?

This study evaluates multiple forecasting techniques to identify the most accurate method for predicting airline ticket prices. We leverage a novel dataset from an online travel agency, covering customer searches for seven flight routes from 2011 to 2014. The dataset includes detailed information on search dates, departure and return dates, months, flight operators, proposed ticket prices, and the number of adult passengers. Using this data, we construct variables such as monthly, weekly, daily, and hourly dummies, the number of seats searched for specific routes, and seat types (e.g., economy or premium). We apply four machine learning techniques—Linear Regression (LR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), and a Hybrid Genetic Algorithm (HGA) as defined by (Tsakonaset al. 2025)—to determine the most accurate method for predicting ticket prices.

Our findings demonstrate that the HGA is the most effective forecasting method, achieving an average Mean Absolute Percentage Error (MAPE) of 11.57% and Root Mean Square Error (RMSE) of 12.77% across seven destinations, outperforming the benchmark of last year's price data. The SVM follows as the second-best method (average MAPE: 13.99%, RMSE: 13.65%), while ANN and LR yield higher errors (ANN: 14.77% MAPE, 13.27% RMSE; LR: 18.69% MAPE, 17.65% RMSE). Prediction accuracy varies temporally: Tuesday and Thursday are the most predictable days (HGA MAPE: 10.30% and 10.92%, respectively), while Wednesday

and weekends exhibit greater variability. January, November, and August show the lowest errors (HGA MAPE: 9.89%, 10.17%, 10.53%), whereas June and December have the highest (HGA MAPE: 17.07%, 19.30%) due to seasonal demand spikes. We also categorize airlines into low-cost carriers (LCCs) and high-cost carriers (HCCs). The HGA performs best for both, but LCCs are more predictable (average MAPE: 9.37%) than HCCs (average MAPE: 11.57%) due to their stabler pricing structures. These results underscore the HGA's robustness in handling complex pricing patterns, particularly for LCCs, and highlight the challenges of predicting HCC prices due to demand-driven volatility.

To our knowledge, this is the first study to leverage high-frequency, customer-level search data to evaluate machine learning techniques for airline ticket price forecasting. Unlike traditional airline datasets, which are often aggregated at monthly or quarterly levels (Goolsbee, 2008, Gerardi,2009, Berry, 2010), our dataset captures granular details, including search timing (e.g., morning, afternoon, evening), seat demand, and return flight information. This granularity enables a nuanced analysis of pricing dynamics, revealing that shorter trips and evening searches correlate with lower prices, while the number of customers searching for a flight significantly influences pricing. Our findings contribute to both the economic literature on airfare pricing (Tsakonas et al. 2025, Goolsbee 2008, Gerardi 2009, Berry 2010)and the application of genetic algorithms in forecasting (Chaturvedi 2008, Gonzalez 2015). By integrating ensemble learning and grammar-guided genetic programming, the HGA outperforms traditional methods, offering a novel solution for travel agencies to optimize pricing predictions and enhance customer decision-making.

Airline pricing is characterized by sophisticated dynamic pricing and revenue management practices. For instance, the Expected Marginal Seat Revenue (EMSR) approach [**?**] sets prices and booking limits based on route connections and estimated demand, adjusting fares dynamically to maximize revenue. Airlines are broadly categorized into full-service carriers (FSCs), such as Lufthansa and British Airways, and low-cost carriers (LCCs), such as Ryanair and EasyJet. FSCs typically exhibit volatile, high-frequency price changes, while LCCs maintain more stable, lower-priced trajectories (Etzioni2003). Common patterns include price increases two weeks before departure, peaking on the flight day (Etzioni,2003). Customer behavior modeling is also critical, with studies emphasizing consumer choice and rationalization as key factors in dynamic pricing (Bitran 2003, Elmaghraby 2003, Shen 2007). Economic analyses further highlight price dispersion as a result of

competitive frictions, cost variations, and heterogeneous demand (Varian 1980, Baye 2006, Dana 1999, Escobari 2007).

Our study advances this literature by demonstrating that search timing and customer demand are pivotal for securing lower fares. Specifically, we find that evening searches and shorter trips yield lower prices, offering actionable insights for consumers. Additionally, our application of the HGA extends the use of genetic algorithms, previously applied in energy and financial forecasting (Chaturvedi 2008, Gonzalez 2015), to the airline industry. By leveraging high-frequency data, we provide a more precise understanding of pricing dynamics compared to studies using lower-frequency data (McAfee 2006, Lazarev2013, Puller2015, Escobari2012, Clark2012. Our results inform both travel agencies and consumers, enabling better pricing strategies and booking decisions in a volatile market.

The paper is organized as follows: Section 2 describes the data, Section 3 presents the methodologies, Section 4 discusses the results, and Section 5 concludes.

The rest of the paper is organised as follows. Section 2 describes the data, section 3 presents the methodologies and section 4 discusses the results. Section 5 concludes.

2 Data

2.1 Data collection and variables

In this section, we present a description of our data set. We collect our data from a European Online Travel Agency and they consist of user flights having the following characteristics:1) One year and a half period from 01-June-2011 to 31/12/2014.), 21 routes in total regarding destinations throughout Europe and also from Europe to destinations in Asia. In our analysis, we consider only round trips. Both direct and stop-over flights were included. Stop-over flights were selected because in many instances are cheaper than direct flights. For this reason, stop-over flights were recorded separately. Furthermore, all the types of fare were included (i.e., first-class, business, economy) and we consider only adult passenger fares. Some certain constraints are derived from the feature availability of the e-sales websites restricted further the data set. Then, we "filter" the dataset such as to avoid Multiple-leg destinations (the search engine does not support his search feature) and any search

with more than 8 passengers. Table 1 describes the parameters for data collection and mining, which form the columns. The table shows the description of each parameter for both the search flights criteria (10 parameters) and the search results output (4 parameters).

2.2 Exploration of Data

In this subsection, we explore the data. We investigate the relevant and important properties of the data that would help to organize data preparation and model training later in a more effective way. Figure 1 shows the distribution of queries based on the category ticket. We observe that most of the tickets in all the requested are in the Economy class (41,52%).

Figure 2 (a) and (b) shows the frequency distribution and cumulative frequency distribution of the count of days from the date of search to the departure date. It is important to investigate the waiting time because it links the demand for the tickets and their prices (see Elmaghraby Keskinocak, 2003). It can be observed that the average waiting time on all routes are distributed approximately exponentially, which means that most users of the online service booking flights seek on a relatively short time horizon. As seen from the graph of the cumulative distribution, on average 50% of customers would like to travel to a maximum of one month, while 90% of users looking for tickets with a time horizon of 100 days or less.

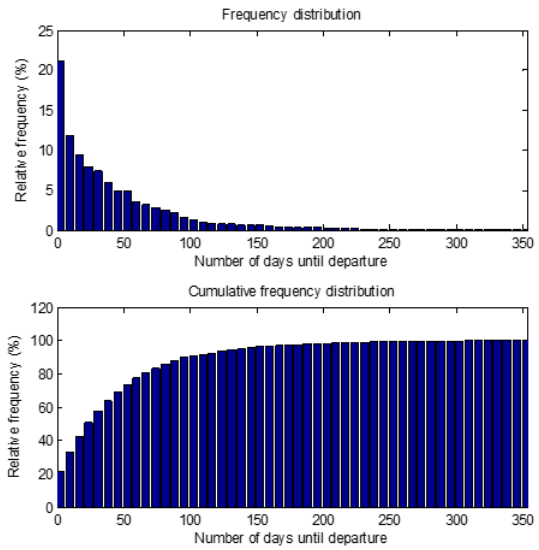

Figure 1: Frequency

In Table 2, statistical moments of the distribution of travel planning horizon for each route separately

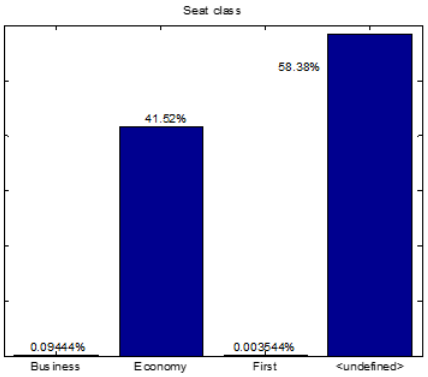

Figure 2: Frequency

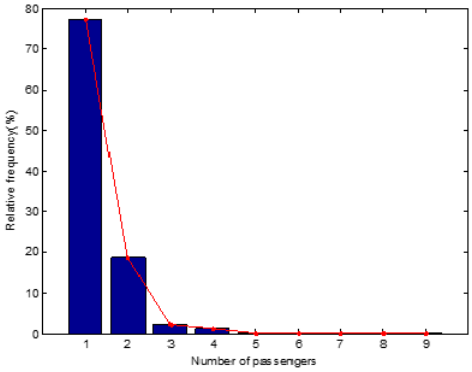

Figure 3: Frequency

are shown. The second column shows the maximum waiting time recorded for each route, while columns 3-7 indicate the 75th, 50th, 25th, 10th and 5th percentile of the distribution. In all routes, there were long term searches (approximately one calendar year). However, the bulk of users seems to plan their trip, 1 to 3 months before the departure date. For instance, 75% of flights on the route Moscow - Barcelona (MOW-BCN) has been searched at least 97 days before departure, and 25% of fares on the route Larnaca - Athens (LCA-ATH) has been searched 8 days before departure.

Figure 4 shows the distribution of searches based on the selected number of passengers. As shown by the statistical analysis of the sample, a significant number of airfares (more than 75%) involved one passenger, while only 18.65% of the tickets concerned two travellers.

From Figures 4 and 5 we see that a vast number of users (rate 97.95%) did not use the option of direct flight when looking for airfare. Approximately 58% of passengers do not choose class positions in their search for airfare, while 42% have a clear preference for economy class. In a very small percentage (0.09%, 0.003%) customers say at the outset that they prefer business or first class. This analysis revealed recurring patterns

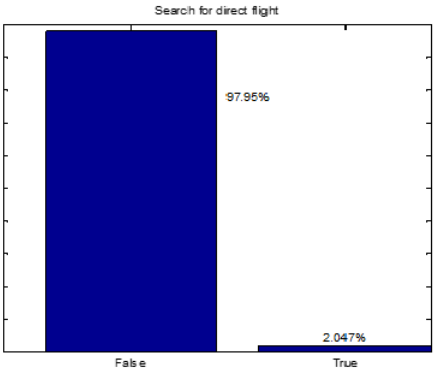

Figure 4: Frequency

Table 1: Determinants-seats positions

Route	Maximum time	Percentile 75	50	25	10	5	Percentage same-day departure
ATH-AMS	358	frame	37	14	5	2	2.96
ATH-LHR	359	67	32	12	5	2	2.86
ATH-LON	358	62	29	11	3	1	3.05
ATH-RHO	349	38	17	6	1	0	6.72
ATH-SKG	357	29	11	4	1	0	9.65
HER-ATH	357	32	14	4	1	0	9.14
HER-SKG	356	54	26	9	3	1	4.61
KHV-OVB	339	64	38	16	5	2	2.92
LCA-ATH	358	51	23	8	2	1	4.23
LCA-LHR	359	95	49	19	6	2	2.12
LCA-SKG	359	56	25	8	3	1	4.27
LCA-SOF	359	52	24	8	2	1	4.28
MOW-AER	355	37	16	5	1	0	8.19
MOW-BCN	359	97	47	20	7	4	1.46
MOW-BKK	359	90	45	18	6	3	2.06
MOW-KRR	352	32	13	4	0	0	10.88
MOW-MRV	358	36	16	5	1	0	9.43
MOW-PAR	359	92	47	20	7	3	1.89
MOW-PRG	359	90	46	21	7	4	1.57
OVB-PKC	313	66	41	19	6	2	2.36
RHO-ATH	352	35	16	6	1	0	6.96

Notes: ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels respectively.

and interesting correlations between the parameters of the problem that can be exploited to create a system for forecasting prices. Mostly, the analysis indicated the key parameters to be taken into consideration in such a system to be able to make estimates with high confidence. Besides the price formation mechanism, the study was extended to the systematic search patterns in price volatility. The latter determines the degree of certainty with which will be known the levels at which the fare will fluctuate on each route event in the future. The results of the statistical analysis showed that the values of certain routes are more volatile than others, so these routes pose a greater risk to passengers. Moreover, the average price dispersion of all routes is not constant but shows seasonal variations. It depends on the parameters of the search, such as the time that elapses until the departure. Also, we examine the relation of days of search before leaving and the lowest priced cost. The maximum number of days is 355 and the minimum is 0 (Figure 7). The mean is 55.1 days. The histogram of

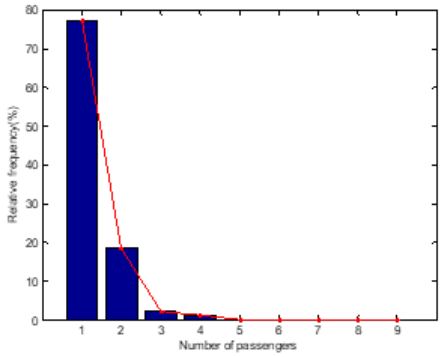

Figure 5: Histogram of searches based on the number of passengers

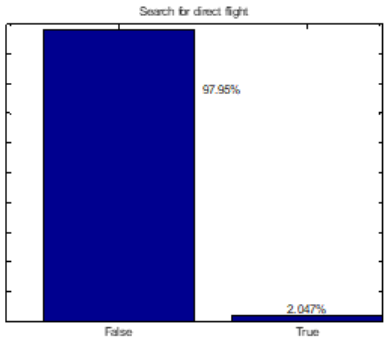

Figure 6: Histogram of searches based on the number of passengers

the frequencies follows. The density distribution is also calculated. Next, (Figure 8) we plot the logarithm of the days of search.

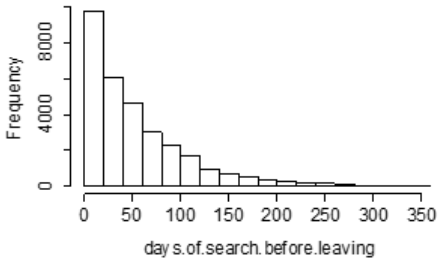

Figure 7: Histogram of searches based on the number of passengers

The summary of statistics is presented on the table 2 below.

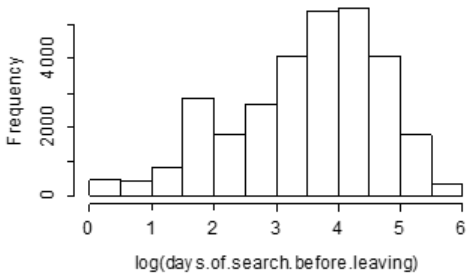

Figure 8: Histogram of searches based on the number of passengers

3 Methodology

The forecasting methods used in this paper are linear regression model

3.1 Linear Regression model

We use the OLS-fixed effect estimation to exploit the determinants of the Airfare ticket prices. The equation that we estimate is the following:

$$Y_{it} = a_i + a_t + X_{it} +_{it} (1) \tag{1}$$

Where Y is the price of the individual ticket bought for the route i at the day t. We include route/trip fixed effects and time dummies, X is other variables included as controls like dummies related to time of the day or whether it is a single ticket, the number of adults that they book the ticket, how many days before the departure date, the user searched for flights. ϵ_{it} is error term. We cluster heteroscedasticity adjusted standard errors at the trip-level to account for serial correlation and small standard errors (Bertrand, Mullainathan, and Duflo (2004).

3.2 Artificial Neural Network (ANN)

ANN have three layers for data modeling, namely, an input layer, an output layer, and hidden layers. The inputs and outputs are modeled through where s and s are connection weights, p is the number of input nodes and q is the number of hidden nodes. The output from the ANN is a non-linear function that maps the inputs to outputs with the help connection weights. ANN were applied for the forecasting using death rate and recovery rate as the input and cases growth as the output variable in R.

The second ML model that was developed is an Artificial Neural Network (ANN). ANNs are inspired by the human brain and the way hundreds of billions of neurons are processing information in parallel. A neural network consists of an input layer, the hidden layers, and the output layer. A really simple ANN is presented in figure 9.

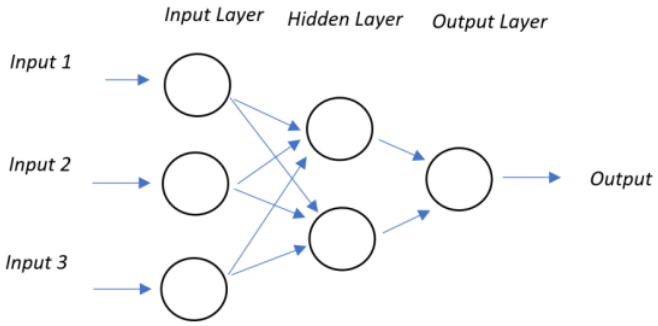

Figure 9: ANN structure

- Input Layer: The input layer receives the information defined by the user in order to process them and provide an output. The input shape, as in the number of input nodes, depends on the number of features, or explanatory variables, or predictors used in the model. When entering the input layer, a number of computations are executed in the nodes (circles) with the result being a set of weighted products for each of the input. These weighted products of the inputs are relative to the significance of their impact on the output.

- Hidden Layer: The output extracted from the input layer is the input of the first hidden layer. In the figure presented there is only one hidden layer, but it is easy to understand that each layer's output is the next layer's input. Hidden layers' nodes work in the same way described earlier. The impact of these layers is of great significance as it results in increased complexity and a better understanding of the hidden relationships between features. It is however possible to overfit the model by adding too many hidden layers. The number of nodes of each hidden layer is customizable. When trying to improve the performance of an ANN tuning the number of layers and nodes is a key task.

- Output Layer: The output Layer is the one responsible for the final predictions. It receives the results of the last hidden layer and compares them to a threshold value before deciding the output. In classification

problems, the number of output nodes matches the possible classes. The outcome from ANN is usually in a form of probability for each class. It is important to mention that all calculations happening in the neurons are determined by an activation function. In our case, we use a linear function called the Rectified Linear Unit, or ReLU. This function is easy to understand and has been consistently proven to result in easier training and relatively high-performance levels. As mentioned earlier tuning the number of hidden layers and nodes in each of them is a key aspect of improving the performance of an ANN. The process can be time consuming and there is always the possibility that the optimum values will not be discovered. In this research different optimizers, learning rates and numbers of layers and neurons were looked into. The best performance models were achieved by the 4 number of hidden layers and 20 and 75 nodes in the hidden layers for the one year and half training sample. All networks were trained with categorical cross-entropy as a validation loss function and evaluated on a 22% size evaluation set.

T

3.3 Support Vector Machines (SVM)

The Support Vector Regression (SVR), which is based on Vapnik's learning theory (Chen et al., 2015), has a design that is comparable to NN's layers that materialise the regression as a quadratic programming problem (Haykin, 1990). The SVR function is estimated by mapping non-linearities from the input space to the output space. In addition, SVR incorporates a cost function that acts as a parameter that allows the user to control the level of regularisation, thus maximising the model's ability to generalise. SVM are the machine learning techniques that is based on classification and regression algorithms and can be used for the forecasting purposes using regression method. SVM were implemented using e1071 package in R. The "linear" kernel were used along with "eps-regression" type from the parameters for the implementation of the method.

Support vector machines (SVMs) is a class of algorithms to effectively solve the problems of regression estimation, pattern recognition, and curse of dimensionality and overfitting in machine learning. The objective of support vector machines problem is to construct an optimal hyperplane that separates the two types of points with a maximal separation margin. The construction process of the optimal hyperplane is actually a process of solving the quadratic programming problem. In the binary classification case, the goal is to find an optimal

separating hyperplane, which is a line (explain as Figure 1). It separates the data into two categories and each class lays on either side of the line. The SVM is also called the maximal margin classifier, where the margin is the distance between the nearest training spots of each category

$$f = \sum_{i=1}^{l} w_i \phi_i(x) + b \tag{2}$$

Besides the classification problem, the SVMs can also solve the regression problem. The support vector regression (SVR) is an adaptation of support vector machines which "generates the regression function by applying a set of high-dimensional linear functions" (Liu and Hu, 2013). The objective of SVMs is to find a particular hyperplane in which separates these two classes with the minimum error while also making sure that the perpendicular distance between the two closest points from either of these two classes is maximized. Support vector regression (SVR) adopt the similar principle. It tries to minimize the distance between the predicted and the target outputs, the distance which is the forecast error. The SVR function is expressed as Equation (3-6).The equation (3-6) to equation (3-11) are cited from Tay and Cao (2001).

where f is the estimate target, $\phi(x)^l$=1 are the function of inputs. w_i=1 and b are coefficients and they are estimated by minimising the regularised risk function (3-7):

$$R(c) = \frac{1}{2}w^2 + \frac{1}{n}C\sum_{i=1}^{l} L(d, y) \tag{3}$$

$$L(d,y) = \begin{cases} |d-y| - \epsilon |d-y| \geqslant \epsilon, \\ 0, \qquad\qquad\qquad\qquad otherwise \end{cases}$$

(4)

The term $L(d, y)$ is the training error measured so-call by the –insensitive loss function. This self-explanatory function indicates the fact that it does not penalise errors below . $\frac{1}{2}w^2$ is the regularization term to mea-

sure function flatness. C is a regularised constant determining the trade-off between $L(d, y)$ and $\frac{1}{2}w^2$. Both $Candareuser - definedparameters.$

The slack variables ξ , xi^* are introduced to represent the distance from the actual value to the corresponding boundary values of ϵ -insensitive tube. With the introduction of slack variables, the Equation (2-2) can be retransformed as following constrained function: Minimise:

$$R(w, \xi, \xi^*)\frac{1}{2}w^2 + \frac{1}{n}C\sum_{i=1}^{l}(\xi, \xi^*) \tag{5}$$

subject:

$$w\phi(x) + b - d_i \leq \epsilon + \xi \tag{6}$$

$$d_i w\phi(x) - b_i \leq \epsilon + \xi \tag{7}$$

$$\xi^* \geq 0 \tag{8}$$

Then the Lagrange multipliers are introduced which satisfy the equality $\alpha_l x \alpha_i = 0$, $\alpha_l \geq 0$, $\alpha_i \geq 0$, $i = 1, 2 \ldots l.$ and are obtained by maximising the dual form of function (2- 4), which has the following form:

$$z(\alpha_l, \alpha_i) = \sum_{i=1}^{l} d_i(\alpha_l - \alpha_i) - \epsilon\sum_{i=1}^{l}(\alpha_l + \alpha_i) - \frac{1}{2}\sum_{i=1}^{l} d_i(\alpha_i - \alpha_i)(\alpha_j - \alpha_i)K(x, x_i) \tag{9}$$

Subject to:

$$\sum_{i=1}^{l}(\alpha_l + \alpha_i) = 0 \tag{10}$$

$$0 \leq \alpha_i \leq C$$

$$0 \leq \alpha_j \leq C$$

The later equation (2-1) becomes the following explicit form by introducing the Lagrangian multipliers α_l, α_i.

$$f(x, \alpha_l, \alpha_i) = \sum_{i=1}^{l} (\alpha_l - \alpha_i) K(x, x_i) + b \tag{11}$$

In this equation, $K(x, x_i)$ is called the kernel function. "There are some different kernels for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space" (Tay and Cao, 2001). Common examples of the kernel function including Linear, Polynomial, Radial Basis Function (RBF) and sigmoid.

3.4 Genetic Algorithm

In this paper, intelligent methodologies were employed to efficiently address the task of airfare forecasting. These methodologies, based on evolutionary computing, genetic programming, and ensemble learning, were implemented and evaluated through extensive experimentation across three phases:

Standard context-free grammar-guided genetic programming (GGGP) approaches Evolutionary ensemble models based on GGGP Grammar-guided evolutionary multi-island strategies, driven by statistical analysis and data subsampling Below is a brief description of the experiments conducted during each phase to identify the most suitable intelligent approach for airfare forecasting. This is followed by a literature review of the intelligent techniques that were tested.

Evolutionary computing is particularly adept at tackling complex optimization problems. Typically, evolutionary approaches are inspired by biological evolution, utilizing operations such as reproduction, mutation, and recombination. During evolution, a population of solutions is repeatedly evaluated using an appropriate fitness function. For more on evolutionary computing fundamentals, refer to [Back et al., 1997], [De Jong, 2006], and [Chiong et al., 2012]. One of the most efficient evolutionary computing methods is genetic programming, alongside other approaches such as genetic algorithms, nature-inspired optimization algorithms, and hybrid intelligent systems. Genetic programming represents solutions as a population of random programs that evolve over time through operations analogous to natural genetic processes (crossover, mutation, etc.). For more on the fundamentals of genetic programming, see [Koza, 1992; 1994; 1999; 2003] and [Banzhaf et al., 1998].

A notable advancement in genetic programming is the use of grammar formalisms as representation struc-

tures, which has demonstrated success in various real-world applications [Gruau, 1996; Dempsey et al., 2009; Keller and Banzhaf, 1996; O'Neil and Ryan, 2001; Tsakonas et al., 2006]. During the first phase of experimentation, evolutionary approaches were tested, with a focus on standard genetic programming algorithms. This led to the adoption of grammar-guided genetic programming (GGGP), specifically using Backus-Naur Form (BNF) for meta-syntax context-free grammar [O'Neil and Ryan, 2003]. The final model implemented a context-free grammar-guided genetic programming approach [McKay et al., 2010; Whigham, 1995; Sipser, 2012]. When tested on real data, this model achieved an average forecasting accuracy of approximately 80% compared to actual airfare prices for the following year.

In the second phase of experimentation, ensemble approaches were found to be more efficient for airfare forecasting. Ensemble models combine predictions from various models, with popular categories including bagging (fitting multiple decision trees), stacking (fitting multiple model types), and boosting (sequential correction of predictions) [Opitz and Maclin, 1999; Polikar, 2006; Rokach, 2010]. Ensemble models work by generating multiple forecasting mechanisms, which are then unified into a robust overall prediction. The trained rules are evaluated and fine-tuned using unseen test data, comparing forecasted airfare prices with those from the previous year.

The ensemble method employed in this phase was a two-level ensemble, with the base level utilizing hierarchical classification trees through GGGP, and the meta-level employing fuzzy rule-based systems also driven by GGGP. The final forecast was generated through a combination of 11 ensemble systems, using cascading, stacking, and grading techniques to enhance generalization. This approach was capable of handling nominal data and performed well in incremental learning. The average forecasting accuracy of this ensemble model exceeded 80%, slightly improving over the first phase.

In the third phase of experimentation, island models were introduced, a technique involving multiple applications of genetic algorithms [Tanese, 1989; Whitley, 1991; Gorges-Schleuter, 1991; Belding, 1995; Hornby, 2009; Izzo et al., 2012]. Island models generate multiple populations, promoting genetic diversity and global search by dividing individuals across islands. Selected individuals migrate between islands, helping avoid local minima. Island models can be parallel or serial, with each island housing its own genetic algorithm and subpopulation (deme) [Whitley et al., 1998; Folino et al., 2005]. Migration between islands is controlled by parameters such as

migration interval and migration size. The key advantage of island models is their ability to find higher-quality solutions than single-population genetic algorithms, owing to the diversity in both genotype and phenotype [Muhlenbein, 1991; Collins and Jefferson, 1991].

The most effective approach tested was a parallel five-island model using a canonical setup, where each genetic algorithm node adhered to a strict hierarchy and a BNF-grammar structure. Parallel island models outperformed serial models due to their broader search space and improved information exchange through migration. When tested on real airfare data, this five-island model achieved an average forecasting accuracy of nearly 85

The methodology applied to airfare forecasting is summarized in Figure 1. Data were collected from a major travel agency, consisting of electronic flight search queries for various destinations over a year. These data were initially classified by destination, then split into training and test sets by month. After statistical preprocessing and logarithmic transformation, the final training and test sets were created. The heart of the forecasting system evolved from a standard GGGP approach (Phase 1), to an ensemble GGGP approach combined with fuzzy rule-based systems (Phase 2), and ultimately to a multi-island genetic algorithm model (Phase 3). The system's generalization ability was validated using a second dataset from the following year, outperforming the previous year's ticket prices as a benchmark predictor.

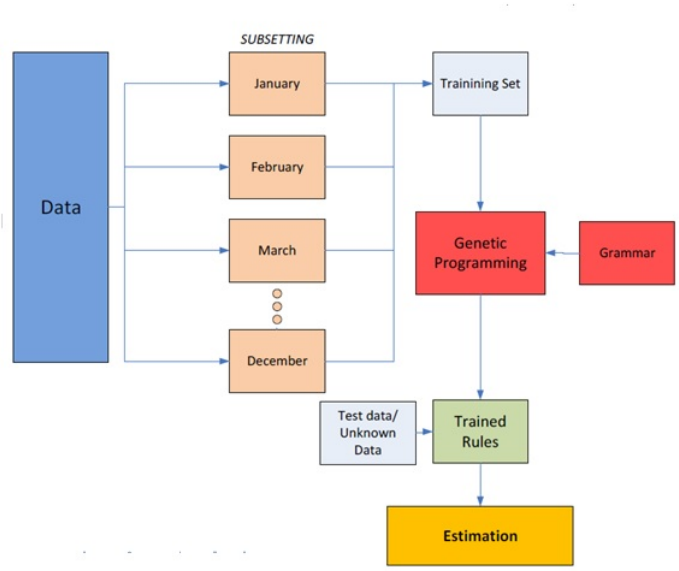

Figure 10: Method

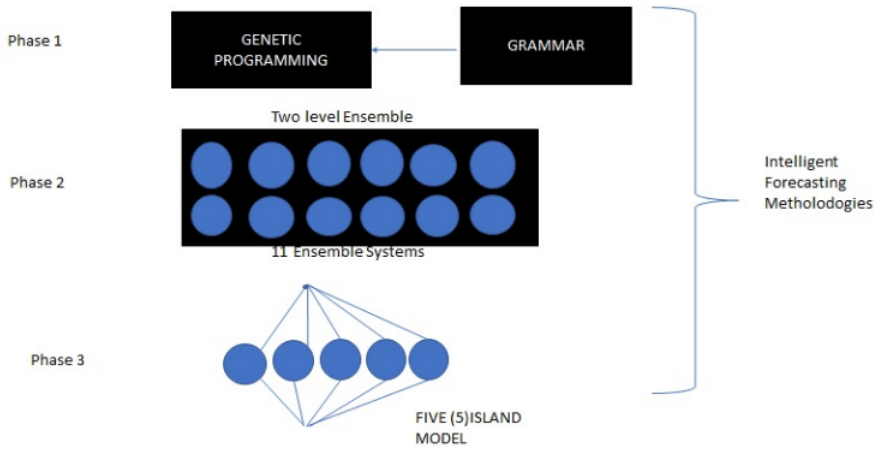

Figure 11: Grammar

4 Results

In forecasting, accuracy is the major although not the only factor that has been dealt with in much-related literature (Makridakis et al., 1982). The prediction performance is evaluated using the two statistical metrics: MAPE (mean absolute percent error), and RMSE (root mean square error). We define P as the predicted value for time t, A is the actual value at time t, and N is the number of predictions. The predict error is defined as $e_t = P - A$. MAPE measures the MAE in percentage terms, so it allows comparisons in units which are differents the actual observations time series.

$$MAPE = \frac{1}{N}\sum_{i}^{N}\frac{|e_i|}{y_t} \tag{12}$$

RMSE is a widely-used evaluation metrics in time series prediction problem (Kanas and Yannopoulos, 2001, Pantazopoulos et al., 1998, Brownstone, 1996) and its fumula is shown as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(e_i^2)} \tag{13}$$

We trained the system using the methodology outlined in previous sections on 21 routes. After completing the training phase, the model was tested through the tested sample. The results correspond to various scenarios set for seven destinations. These scenarios primarily involved searches ranging from one week to one month before the flight, reflecting typical customer behavior. For each destination, 20 to 25 different scenarios were

applied randomly, varying by dates, months, and days of the week.

As a benchmark for comparison, we used the previous year's search and price data, following expert guidance to ensure a fair comparison of prediction performance. Specifically, the average price from the same search period and similar flight date from the prior year was used. The benchmark model made minor adjustments for leap years or missing data to maintain accuracy. For instance, if comparing prices for Friday, November 21 of this year, the equivalent date from the previous year would be Friday, November 22, 2013. Adjustments were also made when weekdays in the current year corresponded to weekends in the previous year.

The results, presented in Table 2, show that the genetic algorithm model achieved a low predictive error, ranging from 9.4% to 21.7%, with an average error of 12.77% across the seven destinations compared to the rest of the method. Importantly, the system required training for each destination before being tested on different search dates. We need to highlight that the second best method for forecasting is Support vector machine which it is RMSE is on average 17% and MAPE is around 16.4%. Neural network model is slighty higher than SVM. On the contrary the highest RMSE and MAPE errors are on the linear regression model.

FLIGHT DESTINATION	GENETIC	ANN	LINEAR REGRESSION	SVM
	RMSE	RMSE	RMSE	RMSE
MOW-KRR	14.30%	17.20%	25.30%	14.80%
ATH-AMS	21.70%	25.20%	31.57%	21.95%
HER-SKG	8.80%	11.20%	20.65%	11.23%
LCA-SOF	13.60%	17.20%	18.65%	19.56%
RHO-ATH	9.80%	11.22%	15.25%	12.68%
LCA-SKG	9.40%	12.10%	15.36%	10.21%
LCA-ATH	11.10%	12.90%	21.98%	15.98%
AVERAGE	12.77%	13.27%	17.65%	13.65%

FLIGHT DESTINATION	GENETIC	ANN	LINEAR REGRESSION	SVM
	MAPE	MAPE	MAPE	MAPE
MOW-KRR	12.20%	18.50%	25.30%	15.21%
ATH-AMS	18.20%	26.30%	32.74%	22.36%
HER-SKG	6.70%	12.40%	21.00%	10.23%
LCA-SOF	12.10%	16.70%	19.02%	17.35%
RHO-ATH	9.50%	12.12%	16.30%	11.32%
LCA-SKG	9.40%	11.40%	14.23%	12.14%
LCA-ATH	11.30%	13.10%	21.21%	12.32%
AVERAGE	11.57%	14.77%	18.69%	13.99%

We analyzed the error between actual and predicted prices based on both days of the week, months and seasons. Notably, Tuesday emerged as the most predictable day, suggesting lower traffic and less fluctuation in

ticket prices in all the models. similarly, to previous results the model the best forecasting technique is the hybrid genetic algorithm and the second best is the SVM. Thursday followed closely in predictability in all the different models. It is important to highlight that the worst predictions are received by the linear regression model. Given the mechanisms of the determination of flight pricing is a complex process, it is obviously that more complex techniques could predict it better. The least predictable day was Wednesday, with significant deviations seen over the weekend (Friday through Monday), as expected. Weekend prices showed larger variations.

The months that have the lowest predicted error in all the different methods are January, November and August. Other hand, the months that they have the highest predicted errors are June and December. The methods that it is the most suitable to predict is the hybrid genetic algorithm and the second best is the support vector machine. The difference between these two techniques is around 6%. When examining seasonal trends, we observed that summer had more uncertainty than autumn, likely due to higher traffic leading to rapid price shifts. Errors were particularly high during summer holidays and around Christmas, likely due to sudden adjustments in airline pricing strategies. As ticket demand decreases, our price predictions become more accurate.

These findings are consistent with Law, Leung, Guillet, and Lee (2011), who found it difficult to identify short-term price fluctuations when comparing data across five online travel agencies (OTAs), advising customers to monitor prices instead of purchasing tickets too early. Similarly, Escobari, Rupp, and Meskey (2019) found that prices tend to be higher during office hours when demand is greater and lower in the evenings when demand drops.

Days	GENETIC	ANN	LINEAR REGRESSION	SV
Monday	13,49 %	18.24%	25.24%	15
Tuesday	10,30 %	17.29%	37.24%	13
Wednesday	17,04 %	18.23%	29%	18
Thursday	10,92 %	16%	22%	15
Friday	11,52 %	13.23%	18%	12
Saturday	13,12 %	14.76%	25%	13
Sunday	13,04 %	16.73%	23%	14
Seasons	ERROR BETWEEN ACTUAL AND PREDICTED PRICE			
Summer	12.72%	16.98%	28	18
Autumn	11.65%	14.54%	19	15
Winter	14.19%	19.86%	20	20
Spring	12.99%	15.01%	19	19

We then categorized the airlines into two groups: low-cost carriers and high-cost carriers. The reason is

Month	Genetic Algorithm	ANN	Linear Regression	SVM
June	17.07%	18.7%	27.12%	18.03%
July	12.56%	13.56%	18.97%	14.18%
August	10.53%	13.22%	20.32%	11.88%
September	12.35%	15.66%	22.44%	14.55%
October	12.45%	14.55%	24.50%	16.54%
November	10.17%	12.31%	15.22%	11.34%
December	19.30%	20.21%	28.45%	22.44%
January	9.89%	11.00%	12.00%	11.76%
February	14.13%	14.98%	15.47%	15.32%

that the low-cost carriers have different models of pricing from the high-cost carriers. The results remained robust, with the hybrid genetic algorithm consistently being the most effective forecasting method for both categories. However, we observed that the Mean Absolute Percentage Error (MAPE) was lower for all methods when applied to low-cost carriers. This is likely because low-cost airlines exhibit less price variation across different dates compared to high-cost carriers, making their prices easier to predict.

For high-cost airlines, the greater volatility in pricing, driven by factors such as demand spikes, premium services, and complex pricing structures, led to higher MAPE errors across all methods. Despite this, the hybrid genetic algorithm still outperformed the other forecasting techniques.

In summary, while the hybrid genetic algorithm remains the most accurate forecasting method across the board, price predictability is generally easier for low-cost carriers due to their more stable pricing patterns. Conversely, high-cost carriers present a greater challenge for prediction, reflected in the higher MAPE errors. Nevertheless, the results confirm the robustness of the forecasting approach, even when applied to different airline segments.

FLIGHT DESTINATION	GENETIC	ANN	LINEAR REGRESSION	SVM
	MAPE	MAPE	MAPE	MAPE
MOW-KRR	8.10%	10.40%	12.36%	9.02%
ATH-AMS	7.10%	15.12%	18.49%	9.21%
HER-SKG	4.20%	11.14%	16.25%	6.32%
LCA-SOF	9.30%	12.36%	12.45%	9.32%
RHO-ATH	4.30%	11.25%	14.23%	5.22%
LCA-SKG	5.20%	9.36%	10.32%	6.32%
LCA-ATH	8.20%	12.1%	11.23%	11.11%
AVERAGE	9.37%	14.77%	18.69%	13.99%

FLIGHT DESTINATION	GENETIC	ANN	LINEAR REGRESSION	SVM
	MAPE	MAPE	MAPE	MAPE
MOW-KRR	16.23%	19.36%	29.20%	17.14%
ATH-AMS	21.21%	29.21%	32.34%	21.99%
HER-SKG	12.23%	15.78%	34.28%	14.25%
LCA-SOF	14.21%	19.58%	22.24%	19.32%
RHO-ATH	11.78%	15.32%	19.54%	15.25%
LCA-SKG	12.12%	15.32%	19.24%	14.85%
LCA-ATH	11.30%	16.25%	25.98%	13.25%
AVERAGE	11.57%	14.77%	18.69%	13.99%

5 Services generated from Airfare forecasts

The competition in the airfare and touristic industry is high. Tourists wish to travel with low cost and high service and the companies want to maximize their revenues and as a result their profits. Our methodology developed in this paper provides the forecasted price in an interval of 80 to 90 per cent of the actual price. This tool can be used as a service from an online touristic agency to address competition and to be a marketing weapon to attract customers for it. More precisely, the customer could choose the dates, the route, the seat, the airline company, the day and the hour of the flight that he/she is interested in and find the price that is expected to pay. The software developed for this application is very easy to be used and in a short period, the customer can find suitable dates that offer him/her affordable prices for his travel. Then, once, the customer will have a clear picture of the airfare price, he/she book his or her flight tickets from the website of the online touristic agency. The latter service could be included in the cost of booking airline tickets or offered free. For instance, suppose a candidate customer wants to book his tickets to fly from Larnaca to Thessaloniki and the actual price is 100 euros. He is not aware of the actual price but he knows that last year for the same flight paid 82 euros. Then, he book, he will have a cost of 18 pounds. Using our approach (being on the website of the online agency), the price will be shown to be 93 euros. First, the customer will have less damage to his revenues by 7 euros. At the same time, the candidate customer will realise that price is higher than last year so he might choose another period that is more affordable compared to his budget. The e-ticket company will have more satisfied customers and meanwhile, the customer will continue booking than cancel his trip.

6 Conclusions

This paper investigated various machine learning methods to forecast airfare prices using large-scale flight data provided by an Online Travel Agency (OTA). After analyzing customer search data for seven routes across Europe and Asia, we found that the hybrid genetic algorithm consistently provided the most accurate predictions. The support vector machine (SVM) also performed well, with predictive errors close to those of the hybrid genetic algorithm.

A detailed analysis revealed that forecasting accuracy varies by day, month, and season. Tuesday emerged as the most predictable day, likely due to lower traffic, while the least predictable days were weekends, where price fluctuations were much larger. Seasonal trends showed that summer and holiday periods (particularly June, December, and Christmas) are more challenging to predict due to sudden price shifts, whereas months like January, November, and August exhibited lower prediction errors across models.

When dividing airlines into low-cost and high-cost carriers, the results remained robust. Low-cost airlines showed lower price variation, resulting in more accurate predictions and lower MAPE errors across all models. High-cost airlines, on the other hand, exhibited greater pricing volatility due to factors like premium services and complex pricing structures, leading to higher MAPE errors, though the hybrid genetic algorithm still outperformed other methods.

In summary, the hybrid genetic algorithm remains the best forecasting technique across different airline segments, and the SVM presents a strong alternative. The study suggests that further exploration of hybrid models, such as ensemble methods that combine linear regression and neural networks, could yield even better predictive performance. Techniques like XGBoost and decision trees might also offer new insights and help improve accuracy. Identifying the optimal forecasting method could allow OTAs to enhance their customer service and maintain a competitive edge in the highly dynamic airline industry.

The findings of this study have several important implications for online travel agencies (OTAs), airlines, and policymakers. By adopting advanced forecasting techniques such as the hybrid genetic algorithm, OTAs can significantly enhance the customer experience by offering more accurate price predictions. This allows customers to make better-informed decisions about their travel plans. Furthermore, this predictive service could be marketed as a premium feature, giving OTAs a competitive advantage, particularly in attracting

budget-conscious travelers.

In addition to benefiting OTAs, airlines can use these forecasting tools to improve their demand management strategies. Understanding trends in demand and pricing fluctuations, especially during peak periods like summer and holidays, enables airlines to optimize their pricing strategies and better handle price volatility. This is particularly valuable during times of high demand when rapid price changes are most common.

Accurate price forecasting also has the potential to strengthen market competitiveness. OTAs that integrate predictive pricing tools into their platforms can offer more tailored services, which not only enhances customer loyalty but also helps them attract more travelers. This is particularly beneficial for low-cost carriers, where customer decisions are highly influenced by price sensitivity. By providing more transparent and accurate price forecasts, OTAs can grow their market share.

Moreover, the research findings can support policymakers in regulating dynamic pricing within the airline industry. By gaining insights into the factors driving price fluctuations, regulators can establish guidelines that ensure fair pricing for consumers, particularly during peak seasons when prices are prone to sudden spikes. This can help mitigate the impact of dynamic pricing on consumers, promoting fairness in the marketplace.

The success of hybrid models in this study also highlights the importance of further research into combining machine learning techniques for more efficient price forecasting. Government and industry bodies could sponsor or encourage such research, fostering innovation in both the travel and aviation sectors. This could lead to the development of even more effective forecasting tools, benefiting both businesses and consumers alike.

In conclusion, the adoption of advanced machine learning techniques such as the hybrid genetic algorithm offers substantial benefits for OTAs, airlines, and policymakers. By improving price predictions, businesses can enhance customer satisfaction, optimize their pricing strategies, and maintain a competitive edge in the market, while also contributing to fairer and more efficient pricing practices across the industry.

7 References

Alba, E., Cotta, C., Troya, J. M. (1996), Evolutionary design of fuzzy logic controllers using strongly-typed GP, In Proceedings of 1996, IEEE international symposium on intelligent control (pp. 127–132) New York, NY

Baye, M., Morgan, J. and Scholten,P. (2006), Persistent Price Dispersion in Online Markets, in

Dennis W. Jansen (ed.), The New Economy and Beyond: Past, Present Future. Edward Elgar, London.

Belobaba P. Peter (1989), Application of a Probabilistic Decision Model to Airline Seat Inventory Control, Operations Research, March/April 1989 vol. 37 no. 2 p. 183-197.

Berry, S., and Jia, P.. (2010). Tracing the Woes: An Empirical Analysis of the Airline Industry. American Economic Journal: Microeconomics, 2 (3): 1–43.

Bilotkach, V., Gorodnichenko, Y., Talavera, O., (2010), Are airlines' price-setting strategies different? Journal of Air Transport Management 16(1), pp.1-6

Bitran G. Caldentey R, (2003), An Overview of Pricing Models for Revenue Management, MSOM Summer 5:203-229

Breiman L. (1996), Bagging predictors, Machine Learning, Vol. 24, pp. 124-140.

I. Bond, M.E., Ladman, J.R. (1972). Tourism: A Strategy for Development. Nebraska Journal of Economics and Business, 11(1), 37-52.

Button, K., Vega, H., (2007), The uses of the "temporal-fares-offered curve" in air transportation. Working Paper, George Mason University.

Cao, Q. Parry,M.(2009). Neural network earnings per share forecasting models: A comparison of backward propagation and the genetic algorithm, Decision Support Systems,47(1), pp.32-41.

Collins R, Jefferson D., (1991), Selection in Massively Parallel Genetic Algorithms. In L. Booker and R.

Belew, editors, Proceedings of the Fourth International Conference on Genetic Algorithms, Morgan Kaufmann, pp 249-256

Dana J.D. Jr. (1999), Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure, RAND Journal of Economics 30, pp. 632–660.

Ellison G. and S.F. Ellison, (2005), Lessons about markets from the internet, Journal of Economic Perspectives 19 (2), pp. 139–158

Elmaghraby W. Keskinocak P, (2003), Special Issue on E-Business and Management Science, Part 1 of 2: Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and

Future Directions, Management Science 49:1287-1309

Escobari, D., Gan, L., (2007), Price dispersion under costly capacity and demand uncertainty, NBER Working Paper No. 13075, National Bureau of Economic Research, Cambridge

Etzioni O. et al, (2003), "To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price", Dept. of Computer Science, University of Washington, presented in SIGKDD '03, August 2427,2003, Washington, DC, USA

Folino Gianluigi, Pizzuti Clara , Spezzano Giandomenico, (2005), GP Ensemble for Distributed Intrusion Detection Systems, Pattern Recognition and Data Mining, Lecture Notes in Computer Science Vol. 3686, pp 54-62

Gallego G. and G. van Ryzin (1994), Optimal dynamic pricing of inventories with stochastic demand over finite horizons, Management Science 40(8), pp. 999–1020

Gerardi, K.S., Shapiro, A.H., (2007), Does competition reduce price discrimination? New evidence from the airline industry, Working Paper No. 07-7, Federal Reserve Bank of Boston

Gerardi, K.S., Shapiro, A.H., (2009), Does competition reduce price discrimination? New evidence from the airline industry. Journal of Political Economy.117:1,pp. 1-37.

Gillen, D, Hazeldine, T., (2006), The new pricing in Canadian air travel markets: implications for competition and antitrust, Working Paper, University of British Columbia.

Goolsbee,A. and Syverson,C. (2008). How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines Get access Arrow The Quarterly Journal of Economics, Vol. 123, Issue 4,pp. 1611–1633.

Gruau, F. (1996), On using syntactic constraints with genetic programming, In P. J. Angeline, K. E.

Jinnear (Eds.), Advances in genetic programming, Cambridge, MA: MIT Press.

Gruau, F., Whitley, D., Pyeatt, L. (1996), A comparison between cellular encoding and direct encoding for genetic neural networks, In Koza, J.R., Goldberg, D.E., Fo- gel, D.B., Riolo, R.L., eds.: Genetic Programming 1996: Proceedings of the First Annual Conference, Stanford University, CA, USA, MIT Press (1996) 81–89.

Gupta,M. Rees, J., Chaturvedi,A. Chi, J. (2006). Matching information security vulnerabilities to organizational security profiles: a genetic algorithm approach, Decision Support Systems,41(3),pp. 592-603.

Janikow, C. Z., (1996), A methodology for processing problem constraints in genetic programming. Computers and Mathematics with Applications, 32(8), 97–113.

Jin F., Li Y., Sun S., Li H. (2020). Forecasting air passenger demand with hybrid ensemble approach. Journal of Air Transport Management, 83 pp. 1-18.

Koenigsberg O., E. Muller and N.J. Vilcassim (2008), EasyJet pricing strategy: should low-fare airlines offer last-minute deals?, Quantitative Marketing and Economics 6, pp. 279–297.

Koza, J.R (1992), Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press.

Koza, J. R., Bennett, F. H., Keane, M. A. (1999), Genetic programming III, San Francisco, CA: Morgan Kaufmann

Koza, J., Bennett, F., Andre, D., Keane, M.(2003): Genetic Programming III: Automatic Programming and Automatic Circuit Synthesis. Morgan Kaufmann, San Francisco.

Kuo, R. J. J.L. Liao, J.L. Tu, C. (2005). Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce, Decision Support Systems, 40(2). Pp. 355-274.

Kyriakakis, N..,Marinaki, M. , Marinakis,Y. (2021). A hybrid ant colony optimization-variable neighborhood descent approach for the cumulative capacitated vehicle routing problem, Computers Operations Research, 134, 105397.

Kyriakakis,N., Sevastopoulos, I., Marinaki, M., Marinakis,Y. (2022). A hybrid Tabu search – Variable neighborhood descent algorithm for the cumulative capacitated vehicle routing problem with time windows in humanitarian applications, Computers Industrial Engineering,164, 107868.

Lemke C., Gabrys B. (2010) Meta-learning for time series forecasting and forecast combination, Neurocomputing 73 (10), pp.2006-2016

Liu, Q., Serfes, K., (2006). Second-degree price discrimination and price dispersion: the case of the US airline industry. Working Paper, Drexel University

McAfee P.R. and V. te Velde (2007), Dynamic pricing in the airline industry. In: T.J. Hendershott, Editor, Handbook on Economics and Information Systems, Elsevier.

Mostafaeipour A., Goli A., Qolipour M. (2018). Prediction of air travel demand using a hybrid artificial neural network (ANN) with bat and firefly algorithms: A case study. The Journal of Supercomputing, 74. Pp. 5461-5484.

Montana, D.J (1995).: Strongly Typed Genetic Programming. Evolutionary Computation 3(2)

Muhlenbein . (1991), Evolution in Time and Space: The parallel Genetic Algorithm. In G. Rawlins, editor, FOGA -1, Morgan Kaufmann, pp 326-337

Opitz D., Maclin R. (1999), Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research 11, pp. 169-198

Pels E. and P. Rietveld (2004), Airline pricing behaviour in the London–Paris market, Journal of Air Transport Management 10, pp. 279–283.

Piga C. and E. Bachis (2007), Pricing strategies by European traditional and low-cost airlines: when is it the best time to book online?. In: D. Lee, Editor, Advances in Airline Economics: The Economics of Airline Institutions, Operations and Marketing, Elsevier, pp. 319–344.

Ribeiro,N. Jacquillat,A. Antunes, A. Odoni, A. Pita, J. (2018). An optimization approach for airport slot allocation under IATA guidelines,Transportation Research Part B: Methodological, Vol. 112, pp.132-156,

Riedel S., Gabrys B., (2005), Hierarchical multilevel approaches of forecast combination Operations Research Proceedings 2004, pp. 479-486

Ryan, C., Collins, J. J., O'Neil, M. (1998), Grammatical evolution: Evolving programs for an arbitrary language. In W. Banzhaf, R. Poli, M. Schoenauer, T. C. Fogarty, Genetic programming, Lecture notes in computer science. London, UK: Springer.

Sanchez, D., Boyacı,B., Zografos,K. (2020). An optimisation framework for airline fleet maintenance scheduling with tail assignment considerations, Transportation Research Part B: Methodological,Vol.133,pp.142-164.

Scott,D. Michael Hall, C., Gössling,S. (2019) Global tourism vulnerability to climate change, Annals of Tourism Research, vol.77, pp.49-61.

Scott,D. and Gössling,S. (2022). A review of research into tourism and climate change - Launching the annals of tourism research curated collection on tourism and climate change, Annals of Tourism Research,vol.95, 103409.

Sharma, J., Mitra, S. K. (2021). Asymmetric relationship between tourist arrivals and employment. Tourism Economics, 27(5), 952-970. https://doi.org/10.1177/1354816620910000.

Shen, Z.-J., X. Su, (2007), Customer Behavior Modeling in Revenue Management and Auctions: A Review and New Research Opportunities, Production Operations Management 16 (6) 713-728.

Shin, R.H., Kim, K., Chung,S. (2024) The Impact of Job Stress of the Cabin Crew on the Service Quality During COVID-19 era. Journal of Distribution Science. 22-5. 117-129.

Simumba, N. Okami, S. Kodaka, A. Kohtake,N. (2022) Multiple objective metaheuristics for feature selection based on stakeholder requirements in credit scoring, Decision Support Systems, 155, 113714.

Su X. (2007), Intertemporal pricing with strategic customer behaviour, Management Science 53, pp. 726–741.

Talluri K.T. and G.J. van Ryzin (2004), The Theory and Practice of Revenue Management, International Series in Operations Research and Management Science Vol. 68, Springer

Tsakonas, A., Dounias, G. (2002): Hierarchical Classification Trees Using Type-Constrained Genetic Programming. In: Proc. of 1st Intl. IEEE Symposium in Intelligent Systems, Varna, Bulgaria

Tsakonas ., Gabrys B. (2012), GRADIENT: Grammar-driven genetic programming framework for building multi-component, hierarchical predictive systems, Expert Systems with Applications, Vol.39, Issue 18, pp 13253-13266.

Tsakonas ., Gabrys B. (2013), A fuzzy evolutionary framework for combining ensembles, Applied Soft Computing, 2013, pp. 1800–1812.

Varian H. (1980), A Model of Sales, The American Economic Review 70(4), Sept. 1980, pp. 651-659.

Wen C.H., Chen P.H. (2017). Passenger booking timing for low cost carriers a continuous logit approach. Journal of Air Transport Management, 64, pp. 91-99.

Whitley Darrell, Rana Soraya, Heckendorn B Robert., The Island Model Genetic Algorithm: On Separability, Population Size and Convergence, Journal of Computing and Information Technology, 1998, Vol. 7, pp 33-47.

Zheng,S. Fu,X., Jiang, C. , Ge, Y. (2020). Airline investments in exclusive airport facilities: Timing decisions under demand ambiguity,Transportation Research Part B: Methodological, Vol. 139,pp. 343-363.