Journal of King Saud University – Computer and Information Sciences xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Journal of King Saud University – Computer and Information Sciences

Airline ticket price and demand prediction: A survey

Juhar Ahmed Abdella [a], Nazar Zaki [b,*], Khaled Shuaib [a], Fahad Khan [c]

[a] Department of Information Systems and Security, College of Information Technology, UAEU, United Arab Emirates
[b] Department of Computer Science and Software Engineering, College of Information Technology, UAEU, United Arab Emirates
[c] Computer Vision Laboratory, Linköping University, Sweden

ARTICLE INFO

Article history:
Received 3 September 2018
Revised 5 January 2019
Accepted 1 February 2019
Available online xxxx

Keywords:
Survey
Ticket price prediction
Demand prediction
Price discrimination
Social media
Deep learning line

ABSTRACT

Nowadays, airline ticket prices can vary dynamically and significantly for the same flight, even for nearby seats within the same cabin. Customers are seeking to get the lowest price while airlines are trying to keep their overall revenue as high as possible and maximize their profit. Airlines use various kinds of computational techniques to increase their revenue such as demand prediction and price discrimination. From the customer side, two kinds of models are proposed by different researchers to save money for customers: models that predict the optimal time to buy a ticket and models that predict the minimum ticket price. In this paper, we present a review of customer side and airlines side prediction models. Our review analysis shows that models on both sides rely on limited set of features such as historical ticket price data, ticket purchase date and departure date. Features extracted from external factors such as social media data and search engine query are not considered. Therefore, we introduce and discuss the concept of using social media data for ticket/demand prediction.

Contents

1. Introduction

The airline industry is considered as one of the most sophisticated industry in using complex pricing strategies. Nowadays, ticket prices can vary dynamically and significantly for the same flight, even for nearby seats (Etzioni et al., 2003; Narangajavana et al., 2014). The ticket price of a specific flight can change up to

[*] Corresponding author.
E-mail address: nzaki@uaeu.ac.ae (N. Zaki).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2019.02.001
1319-1578/© 2019 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

7 times a day (Etzioni et al., 2003). Customers are seeking to get the lowest price for their ticket, while airline companies are trying to keep their overall revenue as high as possible and maximize their profit. However, mismatches between available seats and passenger demand usually leads to either the customer paying more or the airlines company loosing revenue. Airlines companies are generally equipped with advanced tools and capabilities that enable them to control the pricing process. However, customers are also becoming more strategic with the development of various online tools to compare prices across various airline companies (Li et al., 2014). In addition, competition between airlines makes the task of determining optimal pricing is hard for everyone.

The last two decades have seen steadily increasing research targeting both customers and airlines. Customer side researches focus on saving money for the customer while airline side studies are aimed at increasing the revenue of the airlines. Conducted researches employ a variety of techniques ranging from statistical techniques such as regression to different kinds of advanced data mining techniques.

From the customer point of view, determining the minimum price or the best time to buy a ticket is the key issue. The conception of "tickets bought in advance are cheaper" is no longer working (William Groves and Maria Gini, 2013). It is possible that customers who bought a ticket earlier pay more than those who bought the same ticket later. Moreover, early purchasing implies a risk of commitment to a specific schedule that may need to be changed usually for a fee. The ticket price may be affected by several factors thus may change continuously. To address this, various studies were conducted to support the customer in determining an optimal ticket purchase time and ticket price prediction (Anastasia Lantseva et al., 2015; Chawla et al., 2017; Domínguez-Menchero et al., 2014; K. Tziridis et al., 2017; Li et al., 2014; Santana et al., 2017; T. Liu et al., 2017; T.Wohlfarth et al., 2011; V. H et al., 2018; William Groves and Maria Gini, 2013; William Groves and Maria Gini, 2015; Y. Chen et al., 2015; Y. Xu and J. Cao, 2017). Most of the studies performed on the customer side focus on the problem of predicting optimal ticket purchase time using statistical methods. As noted by Y. Chen et al. (2015), predicting the actual ticket price is a more difficult task than predicting an optimal ticket purchase time due to various reasons: absence of enough datasets, external factors influencing ticket prices, dynamic behavior of ticket pricing, competition among airlines, proprietary nature of airlines ticket pricing policies etc. Nevertheless, few studies have attempted to predict actual ticket prices with the work done by the authors in (Anastasia Lantseva et al., 2015; Domínguez-Menchero et al., 2014; K. Tziridis et al., 2017; Santana et al., 2017; T. Liu et al., 2017; V. H et al., 2018) as examples.

On the airlines side, the main goal is increasing revenue and maximizing profit. According to Narangajavana et al., 2014, airlines utilize various kinds of pricing strategies to determine optimal ticket prices: long-term pricing policies, yield pricing which describes the impact of production conditions on ticket prices, and dynamic pricing which is mainly associated with dynamic adjustment of ticket prices in response to various influencing factors. Long term-pricing policies and yield pricing are associated with internal working of the specific airline and do not help that much in predicting dynamic fluctuations in price. On the other hand, dynamic pricing enables a more optimal forecasting of ticket prices based on vibrant factors such as changes in demand and price discrimination (Malighetti et al., 2009). However, dynamic pricing is challenging as it is highly influenced by various factors including internal factors, external factors, competition among airlines and strategic customers. Internal factors consist of features such as historical ticket price data, ticket purchase date and departure date, season, holidays, supply (number of available airlines and flights), fare class, availability of seats, recent market demand and flight distance. External factors include features such as occurrence of some event at the origin or destination city like terrorist attacks, natural disaster (hurricane, earthquake, tsunami, etc.), political instability (protest, strike, coup, resignation), concerts, festivals, conferences, political gatherings and sports events, competitors' promotions, weather conditions and economic activities.

A diagram illustrating interactions between customers and airlines in determining dynamic pricing is given in Fig. 1. Generally, dynamic pricing can be considered as a game between the retailer and consumers where each party tries to maximize its own profit (Y. Wang, 2016). The airlines have the desire to increase their profit by selling as many tickets as possible with highest price. However, tickets have to be sold within limited time horizon as waiting for long time could incur more loss as a result of unsold seats. On the other hand, customers are hoping to buy tickets at the lowest price and keep on monitoring ticket prices across airlines until the ticket price drops. Moreover, customers arrive randomly, and the demand could vary from time to time. Therefore, to become profitable in such complex situations, airlines must dynamically adjust ticket prices based on the current demand, the behavior of customers, ticket prices given by competitors in the market and other internal and external factors (Y. Wang, 2016; Yiwei Chen and Vivek F. Farias, 2015). This dynamic adjustment of ticket prices in response to various influencing factors is known as dynamic pricing. The aforementioned studies (Malighetti et al., 2009; Narangajavana et al., 2014; Y. Wang, 2016; Yiwei Chen and Vivek F. Farias, 2015) explain that dynamic pricing is implemented by airlines as one of the most common price strategies. However, they do not discuss the different kinds of prediction methods that are utilized to implement dynamic pricing.

A significant number of research works exits that proposed prediction models for dynamic pricing in airlines which can be classified into two groups: demand prediction (Bo An et al., 2016; Bo An et al., 2017; Chieh-Hua Wen and Po-Hung Chen, 2017; Diego Escobari, 2014; H. Yuan et al., 2014; Jie Liu et al., 2017a,b; Mumbower et al., 2014) and price discrimination (Efthymios Constantinides and Rasha HJ Dierckx, 2014; Mantin Benny and Bonwoo Koo, 2010; Marco Alderighi et al., 2011; Steven L.Puller and Lisa M.Taylor, 2012). Early prediction of the demand along a given route could help an airline company preplan the flights and determine appropriate pricing for the route. Existing demand prediction models generally try to predict passenger demand for a single flight/route and market share of an individual airline. Price discrimination allows an airline company to categorize customers based on their willingness to pay and thus charge them different prices. Customers could be categorized into different groups based on various criteria such as business vs leisure, tourist vs normal traveler, profession etc. For example, business customers are willing to pay more as compared to leisure customers as they rather focus on service quality than price.

Despite the fact that there are several studies conducted on both sides, customer and airlines, no attempt has been made to present a literature survey and review of existing work. Therefore, the main goal of this paper is to present a comprehensive literature review of existing studies related to this topic which can be utilized by future researchers. We first classify and present existing studies into two categories based on their desired goal (customer side models and airline side models). We then group existing work based on the specific problem being addressed. Several issues have been discussed including data sources, features and various techniques employed for prediction. We believe that this is an important contribution for researchers who are aiming to work on this exciting area of research.

One of the results of our review indicates that existing models generally rely on limited number of features which are not effective enough in predicting ticket price. For example, customer side

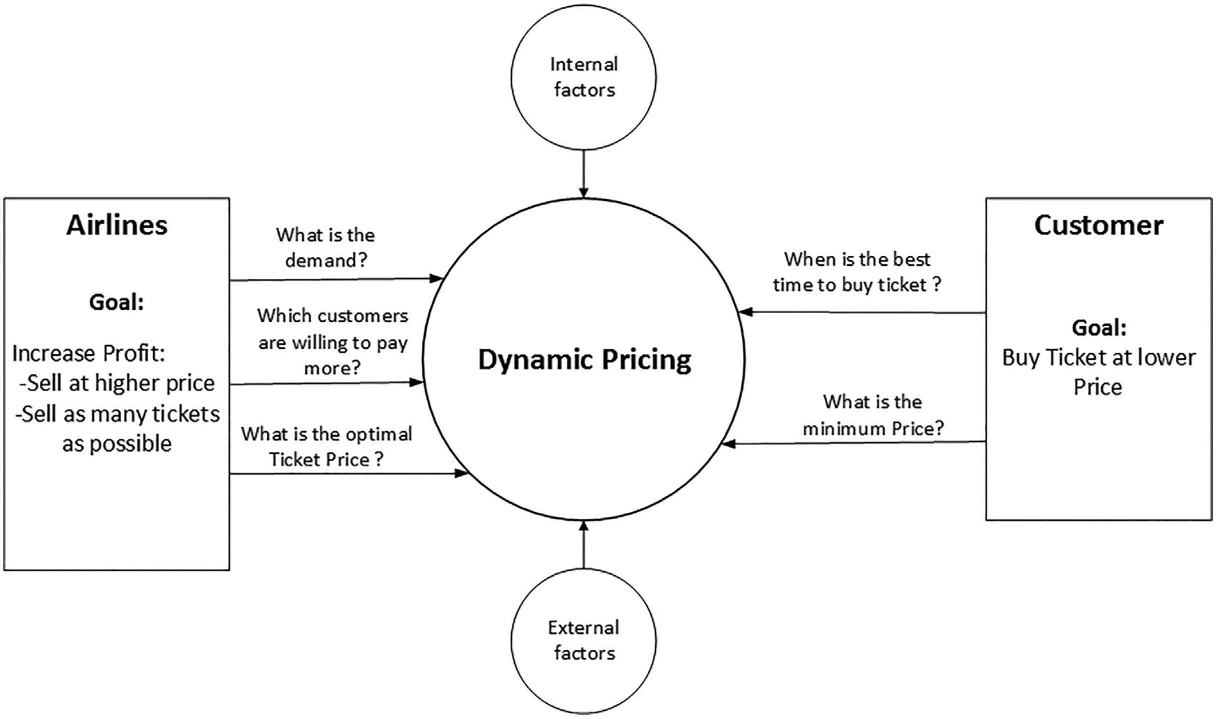
Internal factors
Airlines
Goal:
Increase Profit:
-Sell at higher price
-Sell as many tickets as possible
What is the demand?
Which customers are willing to pay more?
What is the optimal Ticket Price ?
Dynamic Pricing
When is the best time to buy ticket ?
What is the minimum Price?
Customer
Goal:
Buy Ticket at lower Price
External factors

Fig. 1. Dynamic Pricing.

models generally utilize restricted features extracted from historical ticket price data, ticket purchase date and departure date. In a similar way, airlines side models are also developed based on limited internal factors such as seasonality, holidays, supply (number of available airlines and flights), fare class, availability of seats, recent market demand, flight distance and competitive moves by other airlines etc. However, ticket prices and passenger demand can also be affected by many of the dynamic external factors mentioned earlier. Even though the attributes used by earlier researchers play a significant role in predicting ticket pricing/demand, the incorporation of these external factors could also lead to a better result.

Nowadays, social media sentiment analysis has become a good source of information for various data mining models. For example, social media data has been used for event prediction (A. Dingli et al., 2015; Arif Nurwidyantoro, and Edi Winarko, 2013; Hila Becker et al., 2012; Mario Cordeiro, 2012; Nikolaos Panagiotou et al., 2016; Takeshi Sakaki et al., 2010; Xiaowen Dong et al, 2015), competitor intelligence (Lipika Dey et al., 2011; Malu Castellanos et al., 2011; Martin Längkvist et al., 2014; Wu He et al., 2015), price prediction (A. Porshnev et al., 2013; J. Santos Domínguez-Menchero et al., 2014; L. Bing et al., 2014; L. Li and K. Chu, 2017) and tourist traffic flow prediction (R. Linares et al., 2015) and many more. A similar approach could be followed to extract useful social media information related to various external factors affecting airlines passenger demand and ticket price. For example, analysis of different twitter hash tags could give valuable information about the presence of an event at an origin/destination city, competitors' promotions, volume of tourist traffic flow, weather condition, economic activity etc. This in turns might allow us to predict the change in ticket price/demand. It is expected that a data mining model that utilizes information resulting from social media data would give better results than existing work in forecasting route demand and or ticket price. However, to the best of our knowledge, there is no existing work that utilizes social media data to predict route demand and or ticket price. Therefore, in this paper we also discuss the concept of using social media data to extract several external features that enable better ticket price prediction and demand forecasting.

The rest of the paper is organized as follows: We review customer side models in Section 2. Airline side modes are presented in section 3. We summarize existing work and provide a discussion on existing work in Section 4. Deep learning and social media data-based ticket/demand prediction is discussed in section 5. Section 6 concludes and summarizes the paper.

2. Customer side models

Despite the fact that various ticket pricing strategies are implemented by several airlines and Online Travel agencies (OTA), there are no adequate research papers available discussing this topic. This can be due to two reasons: First, ticket pricing strategies are highly business sensitive and remain proprietary of the owner company (Etzioni et al., 2003). Most airlines do not reveal their ticket pricing strategies because of competition with other airlines. Second, there is lack of publicly available datasets that could enable researchers to conduct their prediction effectively. As a result, researchers are obliged to rely on small datasets that are gathered using Web scrapping programs. Nevertheless, there exist limited works that came up with various techniques for ticket price prediction regardless of the limited resources available (Anastasia Lantseva et al., 2015; Chawla et al., 2017; Domínguez-Menchero et al., 2014; K. Tziridis et al., 2017; Li et al., 2014; Santana et al., 2017; T. Liu et al., 2017; T.Wohlfarth et al., 2011; V. H et al., 2018; William Groves and Maria Gini, 2013; William Groves and Maria Gini, 2015; Y. Chen et al., 2015; Y. Xu and J. Cao, 2017). The studies performed on the customer side can be roughly categorized into two: those that try to predict optimal ticket purchase timing (Etzioni et al., 2003; Li et al., 2014; T.Wohlfarth et al., 2011; William Groves and Maria Gini, 2013; William Groves and Maria Gini, 2015) and those that proposed solution to predict exact value of ticket price (Anastasia Lantseva et al., 2015; Domínguez-Menchero et al., 2014; K. Tziridis et al., 2017; Santana et al., 2017; T. Liu et al., 2017; V. H et al., 2018).

2.1. Optimal ticket purchase timing prediction

One of the pioneers on optimal ticket purchase timing prediction is probably the work done by (Etzioni et al, 2003). The authors proposed a model that advise the user whether to buy a ticket or to wait at a particular point of time. For each query day, the model generates a buy or wait signal based on historical price information. The model uses various data mining techniques such as Rule learning (Ripper), Reinforcement learning (Q-learning), time series methods, and combinations of these to achieve various accuracy levels. Q-learning and Ripper are used to predict the behavior of new flight data based on a set of training data while the time series method uses the moving average to forecast the price characteristics of a flight based on historical price data of the same flight. Around 12,000 historical ticket price data representing 41 departure dates for two routes was used for the analysis. The dataset has limitations in which the collection was done only starting from twenty-one days before the departure. Moreover, a constant seven days round-trip is considered. The features used by the model include flight number, number of hours until departure, current price, airline and route (origin and destination city). Simulation is used to measure the savings passengers gained due to each of these data mining methods. The saving (or loss) performance of a model is calculated by computing the cost due to the price difference between the ticket price at an earlier purchase point and the ticket price at the time recommended by the algorithm. The best accuracy (61.9% as compared to optimal saving) is achieved from the combination of all the techniques used in the study. According to (William Groves and Maria Gini, 2013), the model proposed by (Etzioni et al, 2003) has been implemented in real time for a popular ticket search website known as Bing Travel as "Fare Predictor" tool.

A closely related work to that of (Etzioni et al, 2003) is also proposed by (William Groves and Maria Gini, 2013) which predicts optimal ticket purchase timing and is in fact inspired by (Etzioni et al., 2003). However, unlike (Etzioni et al, 2003), (William Groves and Maria Gini, 2013) can forecast the optimal purchase time for all available flights across different airlines for a given departure date and route. Moreover, they use a dataset that is collected 60 days ahead of the departure date. The data was collected for a period of 3 months using daily price quotes from an OTA website from February 22, 2011 to June 23, 2011. Each query returned approximately 1,200 quotes for a single route from all airlines. The round trip was based on a constant 5 days round-trip. Two kinds of features were used for the analysis: Deterministic features and aggregated features. Examples of deterministic features include days to departure and quote day of week i.e. the number of different ticket prices available for a given flight on a specific route across different airlines). Aggregated features are features extracted from the historical data such as the minimum price, mean price and number of quotes. The minimum price, mean price and number of quotes are calculated for non-stop, one-stop and multi-stop for individual airlines and for all airlines. Moreover, a lagged feature computation is also used to consider the effect of time-delayed observations in prediction. Four kinds of regression techniques are used to generate a regression model for the analysis: Partial least squares (PLS) regression and three machine learning algorithms (Decision tree, nu-Support Vector Regression (nu-SVR) and Ridge Regression). The study in (William Groves and Maria Gini, 2013) used a similar approach as (Etzioni et al, 2003) for the performance evaluation. Based on experimental analysis for one route with 256 simulated purchases, PLS regression was found to be the best model with 75.3% saving as compared to the optimal one.

The same authors above also proposed another ticket purchase time optimization model in (William Groves and Maria Gini, 2015) based on various machine learning techniques. The machine learning methods consisted of REPTree classifier and four types of regression models (PLS regression, RepTree regression, Ridge regression and nu-SVR regression). Similar data as that of (William Groves and Maria Gini, 2013) is utilized but it was collected for 7 routes over 109 days with a total of 23.5 million quotes where each query for a single route gave 1,200 quotes on average. Further, they divided the 109 days dataset into 3 parts: 48, 20, and 41 days and used them as the training dataset, validation set, and test set, respectively. Other methods which were not used in (William Groves and Maria Gini, 2013) were also considered in (William Groves and Maria Gini, 2015) including user-guided feature selection method and handling specific customer preferences such as nonstop-only flights, flights with specific take-off time, or flights from a specific airline. The best result achieved is 69% (compared to the optimal) using PLS regression based on experimental analysis for 7 routes.

The authors in (T. Wohlfarth et al., 2011) proposed an optimal ticket purchase time optimizing model based on a special preprocessing step known as marked point processes (MPP), data mining techniques (clustering and classification) and statistical analysis techniques. The MPP pre-processing technique was suggested to convert heterogeneous price series data such as international, national, long and short flights, different providers (low cost and regular) into an interpolated price series trajectory that can be fed to an unsupervised clustering algorithm. Once the MPP step is completed, the model applies clustering followed by classification and statistical processing techniques on historical price data to develop price decrease event predictive rules. First, the price series trajectory is clustered into groups based on similar pricing behavior. Next, a price evolution model that estimates price change patterns up to departure date is defined for each cluster. For a new test dataset, a tree-based classification algorithm is used to select the best matching cluster and then the corresponding price evolution model defined for that cluster is used to predict the price decreasing event. The dataset used by this research is obtained from Liligo.com's historical price data collected for 28 days. It covers data for 6 routes from 9 airlines. Unlike others, this paper also considers round-trips for 3, 7 and 14 days. The set of features in the analysis include: departure station, arrival station, departure date, return date, provider, day of week, day of month, day of year and demand. The authors claim that the model achieved 55% performance as compared to (Etzioni et al, 2003). However, no details of performance evaluation steps were presented.

The study by (Domínguez-Menchero et al., 2014) suggested a model that predicts the optimal purchase timing based on non-parametric isotonic regression techniques for a specific route, time period and airlines. The model determines the maximum number of days users might wait before purchasing ticket without significant price increase and the daily money loss that comes from delaying the purchase. Two types of variables are considered for the prediction: price and date of purchase. The authors analyzed four routes for direct flights and one-stop flights based on a two-month period daily price information that is extracted 30 days prior to departure date. They found that purchasing a ticket up to 18 days prior to departure incurs no significant economic loss. The authors claim that the isotonic method is advantageous in that this effect cannot be achieved with other types of regression techniques (e.g. linear regression). The paper (Chawla et al., 2017) investigated the dependency of ticket price on certain factors and built and compared various types of prediction models that consult the user whether to buy the ticket or wait for some time for a particular flight on a given route and date of departure. The paper has considered five factors including those which have not received much attention from other studies: oil price, number of intermediate stops, number of days before departure, week day of departure and number of competitors on the route. Two different supervised learning approaches have been used to build these models: Regression based modeling and Classification based modeling. Several techniques have been considered under both categories. The model that was built based on the Naive-Bayes technique was found to be the most accurate. It was tested based on a data that was collected for 2 months and was able to achieve an accuracy of 84%.

Different from the preceding models that perform local (short period e.g. per day) optimal purchase timing prediction based on past static price data, the authors of (Y. Xu and J. Cao, 2017) proposed an optimal purchase decision support system that allows continuous recommendations of the optimal purchase timing for several days before departure date based on real time dynamic price features and multi-step prediction in addition to the historical price data. The system utilizes two single-step time series prediction models: MA (Moving Average) and CART (Classification and Regression Tree) regression as a basis for the multi-step prediction. The multi-step prediction is built recursively based on the results of single-step prediction. A statistical approach known as Linear Discriminant Analysis (LDA) and the Bayes classification scheme is employed to classify days before departure into "BUY" and "WAIT" labels and to generate the probability of each label occurring. A recent work by (Manan Dedhia et al., 2018) also proposed an optimal ticket purchase timing prediction system based on logistic regression machine learning technique. The model suggests users to buy ticket or wait for some time.

The summary of optimal ticket purchasing time prediction models is given in Table 1.

2.2. Ticket price prediction

All the studies discussed in the previous section provided a model that forecast optimal purchase timing for customers. However, predicting real-time flight prices was not considered. Understanding this gap, Y. Chen et al., (2015) proposed a model that predicts the lowest price available for a given itinerary (a specific flight on a given route for a particular departure date). To be more precise, given the current day, d_1, and a specific itinerary (r, d_n) identified by route r and departure date d_n, the model predicts the lowest prices available for consecutive days d_2, d_3… d_{n-1}, d_n where $d_1 < d_2 < d_3 < d_{n-1} < d_n$. However, the model considers only non-stop flights. Moreover, it is not possible to predict the price of a single flight as it works at the route level. An ensemble-based learning algorithm Learn++.NSE, is modified and trained to incrementally learn from past patterns of the price changes and to forecast future prices.

A recursive strategy is used to estimate multiple future prices iteratively i.e. the price from previous predictions is used to predict the next price in multiple steps. Features such as prices of the same itinerary, prices of recent itineraries before the target day, prices of itineraries with the same day of the week and price of itineraries with the same day of the month are used for the model. The model is tested on a daily price dataset extracted from an OTA company in China for 5 different international routes. The collection was made for more than 3 months (Feb 11 to Jun 01, 2015, for 110 days in total). For each day, the lowest prices of itineraries leaving in the next 60 days were recorded for every route, resulting in 110 × 60 total observations. Experimental results reveal that the model performs relatively better on diverse routes i.e. routes in which pricing behavior of different flights is completely independent from each other with different price level and variation magnitude having no universal pattern. A comparison between the Learn++.NSE, KNN and Passive-Aggressive (PA) is performed to see their relative performance. The model achieved the lowest mean absolute percentage error (MAPE) of 10.7% as compared to KNN (12.58) and PA (15.41%).

The authors in (Anastasia Lantseva et al., 2015) proposed a ticket prediction model based on an empirical data-driven Regression Model. The model predicts the price per kilometer for a given flight within 90 days before departure date. Two kinds of flights (local and international) were considered for the study based on data collected from two independent ticket price information aggregators (AviaSales and Sabre) in spring 2015. For local flights, they used flights from two Russian cities (Moscow and Saint-Petersburg) to 50 local Russian cities. Flights from the same two cities (Moscow and Saint-Petersburg) to 40 international destinations were considered for international flights with the domination of European cities. The minimum price for each flight per day was collected over a period of 75 days for AviaSales and 90 days for Saber. The features used for building the model include: city of departure, destination, ticket purchase date, departure date, ticket options with the price. Based on the proposed model, the authors compared the effect of early ticket purchasing on the price of tickets for local and global flights. It was found that early ticket purchasing has an advantage for international flights while local flight required additional investigations to reach concrete conclusions. The authors did not provide performance evaluations of the model. Moreover, the dataset used by (Anastasia Lantseva et al., 2015) was limited since it was collected over a short period and also for specific routes.

The study by (T. Janssen, 2014) utilized a linear quantile mixed regression model to predict the minimum ticket price that would occur within 60 days before departure. The method uses the quintile with low prices observation only instead of the whole price observations to predict the minimum price. Four variables are considered in the model: price, departure date, observation date, number of days left to departure and feature indicating day of week (weekend or weekday). The data used for the study consists of 2271 flights with a total of 126,412 records corresponding to a single route collected within 60 days before departure across 6 airlines. The dataset was limited to only one way trip leisure tickets with non-stop flights. The test results showed that the model performs well for shorter period before departure but tends to be inefficient as the number of days before departure increases.

The paper (K. Tziridis et al., 2017) compared the performance of eight state of the art regression machine learning (ML) models with respect to predicting airline ticket prices. The eight models considered include Multilayer Perceptron (MLP), Generalized Regression Neural Network, Extreme Learning Machine (ELM), Random Forest Regression Tree, Regression Tree, Bagging Regression Tree, Regression SVM (Polynomial and Linear) and Linear Regression (LR). Moreover, the paper attempted to identify the factors that have higher influence on airfare price prediction. The studied factors include departure time, arrival time, number of free luggage, days before departure, number of intermediate stops, holiday, time of day and day of week. The models were trained based on a dataset consisting of 1,814 flights for a single international route. The results revealed that the "Bagging Regression Tree" model outperforms the other models with accuracy of 87.42%, followed by Random Forest Regression Tree which achieved 85.91%.

Similar to (Y. Chen et al., 2015), an ensemble regression algorithm is proposed by (T. Liu et al., 2017) for predicting the lowest price available on a particular route for the days between the purchase date and a given departure date. The algorithm uses various kinds of machine learning techniques such as *k*-Nearest Neighbors, Random Forest and Bayesian as a base learner and feature clustering to build the ensemble learning model. Three kinds of features are considered by the model including historical ticket prices, a sig-

Table 1
Summary of Optimal Purchase Time Prediction Models.

Ref.	Addressed Problem	Dataset	Features	Computational Techniques Used	Performance Result	Remark
Etzioni et al., (2003)	Predicting optimal ticket purchase time	12,000 ticket price data collected over 41 day	Flight number, number of hours until departure, current price, airline and route	Rule learning (Ripper), Reinforcement learning (Q-learning), time series methods, and combinations of these	An average of 61.8% savings achieved as compared to optimal saving.	- Limitations in data set - Limited number of features - Only considers 7 days round-trip - Does not consider heterogeneous flights
William Groves and Maria Gini, (2013)	Predicting optimal ticket purchase time	Data collected for three months 60 days prior to departure date.	Days to departure, Quote day of week, minimum price, mean price, Number of quotes	PLS regression Decision tree,nu-SVRRidge Regression	75.3% saving for the as compared to optimal saving.	- Does not consider heterogeneous flights - Only considers 7 days round-trip
William Groves and Maria Gini, (2015)	Predicting optimal ticket purchase time	The same data as above but for 7 routes	Same as above but with the addition of user-guided feature selection in addition to lagged feature selection	Decision tree (RepTree), PLS regression, RepTree regression, ridge and nu-SVR	69% (compared to the optimal) using PLS regression based on analysis test for 7 routes.	- Does not consider heterogeneous flights - Only considers 7 days round-trip
T.Wohlfarth et al., (2011)	Predicting optimal ticket purchase time	Data collected for 28 days for 6 routes from 9 providers. It also considers round-trips for 3, 7 and 14 days	Departure station, arrival station, departure date, return date, provider, day of week, day of month, day of year, demand	Marked point processes (MPP) for Preprocessing, Clustering, classification andstatistical analysis	55% performance as compared to (Li et al., 2014)	No detail performance evaluation steps are presented.
Domínguez-Menchero et al., (2014)	Maximum number of days to wait before purchasing ticket	2 months daily price information extracted 30 days prior to departure date	Price and date of purchase	Non-parametric isotonic regression	–	No performance evaluation
Chawla et al., (2017)	Comparing machine learning algorithms for optimal ticket purchase time prediction	2 months data	Oil price, number of stops, number of days before departure, week day of departure and number of competitors	Regression based modeling and Classification based modeling	84% accuracy using Naive-Bayes technique	
Y. Xu and J. Cao, (2017)	Optimal purchase decision support system	–	Historical price data and real time dynamic price features	-Multi-step prediction that uses two single-step prediction models: Moving Average and Classification & Regression Tree.	–	

nal indicting whether the departure date is holiday or not and number of days before departure. The approach also dynamically adjusts features adaptively based on the context to get more accurate result. The signal that shows whether the departure date is holiday or not is among the features that is considered as context information. The model is trained and tested based on a dataset that consists of 19 different routes and spans three months period (92 days). The algorithms has been shown to perform better compared to the single base learners as indicated by mean absolute percentage error (MAPE) evaluation metrics where the error has been improved from (7% –12%) to (3.7% – 6%). Another research paper (William Groves and Maria Gini, 2011) developed a price prediction model called stacked prediction model by combining conventional machine learning algorithms. In order to determine which machine learning techniques to use, the authors first evaluated the performance of several machine learning algorithms including Random Forest, Decision Tree, Multilayer Perceptron, Support Vector Machines, *k*-Nearest Neighbours, AdaBoost and Gradient Boosting based on three evaluation metrics: R-squared (R^2), Mean Absolute Error (MAE) and Mean Squared Error (MSE). Random Forest and Multilayer Perceptron were the two best-performers. Therefore, the combination was made by applying the two best performing models (i.e. Random Forest and Multilayer Perceptron) by assigning them different weights. The model was trained based on an airfare data that consists of 51,000 records for a 7-day round trip nonstop flights of three domestic airlines collected 21 days prior to the departure date. However, international and multi-stop flights are not considered in the data. A total of predict twelve features were extracted from the data: airline, flight number, date of purchase, departure date, departure time, arrival time, fare class, number of stops, price, departure airport, arrival airport, and arrival date. The stacked prediction model was better than both Random Forest and Multilayer Perceptron with 4.4% and 7.7% respectively based on R^2 evaluation metrics.

Other recent works also exist on ticket price prediction. Boruah A. et al., (2018) attempted to predict ticket prices using a famous Bayesian estimation technique known as a Kalman filter. The paper proposed an algorithm that predicts the ticket price for a specific flight based on the linear model of the Kalman Filter. The model utilizes features derived from an observation of previous fares where the observed data is given as input in the form of a matrix similar to the linear Kalman Filter model. Yuling Li and Zhichao Li, (2018) designed and implemented a ticket price forecasting system using a combination of ARMA algorithm and random forest algorithm. The model is implemented using a Python language and SQL Server database.

Please cite this article as: J. A. Abdella, N. Zaki, K. Shuaib et al., Airline ticket price and demand prediction: A survey, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.02.001

Table 2
Summary of Ticket Price Prediction Models.

Ref.	Addressed Problem	Dataset	Features	Computational Techniques Used	Performance Result	Remark
Y. Chen et al., (2015)	Minimum Ticket Price Prediction	More than 3 months (110 days) data for 5 international routes.	Prices of the same itinerary, prices of recent itineraries before the target day, prices of itineraries with the same day of week, price of itineraries with the same day of month	An ensemble-based learning algorithm Learn++.NSE is modified and used	Mean absolute percentage error (MAPE) of 10.7% as compared to KNN (12.58) and PA (15.41%).	- Not possible to predict price for a flight - Does not consider multi-stop flights
Anastasia Lantseva et al., (2015)	Ticket Price per kilometer Prediction	Ticket price data collected for 75 days and 90 days for local and international flights.	City of departure, destination, ticket purchase date, departure date, ticket options with the price	Regression Model	Not given	- No performance evaluation presented. - The dataset set is limited
(K. Tziridis et al., (2017)	Comparing regression machine learning models for predicting airline ticket prices.	A dataset consisting of 1814 flights for a single international route	Departure time, arrival time, number of free luggage, days before departure, number of intermediate stops, holiday, time of day and day of week	Eight regression machine learning models	Bagging Regression: 87.42%, accuracy and Random Forest Regression Tree: 85.91%. accuracy	
T. Liu et al., (2017)	Predicting the lowest price available before departure date	Data consisting of 19 different routes and spans three months period (92 days).	Historical ticket prices, a signal indicting whether the departure date is holiday or not and number of days before departure	Ensemble model that uses techniques such as K-Nearest Neighbors, Random Forest and Bayesian	Improved the MAPE from (7% –12%) to (3.7% – 6%).as compared to the single model	
V. H et al., (2018)	Ticket price prediction	51,000 records of a 7-day round trip nonstop flights of three domestic airlines	Airline, flight number, date of purchase, departure date, departure time, arrival time, fare class, number of stops, price, departure airport, arrival airport, arrival date	A model called Stacked prediction model based on. Random Forest and Multilayer Perceptron model	4.4% and 7.7% better than Random Forest and Multilayer Perceptron respectively as measured by R^2	
T. Janssen, (2014)	Predict the minimum ticket price before departure	2,271 flights with a total of 126,412 records corresponding to a single route collected within 60 days before departure	Price, departure date, observation date, number of days before departure and day of week (weekend or weekday)	Linear quantile mixed regression model	Performs well for shorter period but is inefficient for longer period	Limited to one only way trip leisure tickets with non-stop flights

A summary for the discussed ticket price prediction models is shown in Table 2.

3. Airlines side models

Airlines side models represent studies targeting profit gained by airlines and OTAs. Two main categories of researches exist in the literature regarding this. The first group proposes demand prediction models (Bo An et al., 2016; Bo An et al., 2017; Chieh-Hua Wen and Po-Hung Chen, 2017; Diego Escobari, 2014; H. Yuan et al., 2014; Jie Liu et al., 2017a,b; Mumbower et al., 2014;) while the second group focuses on price discrimination (Efthymios Constantinides and Rasha HJ Dierckx, 2014; Mantin Benny and Bonwoo Koo, 2010; Marco Alderighi et al., 2011; Steven L.Puller and Lisa M.Taylor, 2012).

3.1. Demand prediction

Among the recent work performed on route demand and market share prediction is the study done by (Bo An et al., 2016). The authors proposed a data mining technique designed for Maximizing Airline Profits (MAP) through prediction of total route demand and market share of an individual airline. They also suggested two algorithms (Bi-level Branch and Bound algorithm and Greedy algorithm) that find the optimum frequency allocation of flights for an individual airline while utilizing the route demand and market share predicted using the proposed prediction model. The proposed prediction model was an Ensemble Forecasting (MAP-EF) technique. It was developed based on existing route demand and market share prediction models, clustering techniques and game theoretical analysis. Several features were utilized for predicting market share and route demand. The features used include: ticket price, number of flights operated by an airline, airline past performance history (delay time, delay ratio, cancel ratio, average stop and safety), aircraft size, total seat, average price, population income and customer price index (CPI).

Unlike most other works, this work considers a broad set of routes (around 700 routes) across 13 airlines operating in those routes. The training dataset spans 10 years (40 quarters) while the testing set includes the first quarter of 2015 (a total of 9100 predictions). However, the prediction is performed quarterly and not for a short period of time which might not consider dynamic demand changes. Moreover, the routes considered are only national routes in the US. The data set is obtained from 4 publicly available data sources from US offices: The Bureau of Transportation Statistics (BTS), the Bureau of Economic Analysis (BEA), the National Transportation Safety Board (NTSB), and the U.S. Census Bureau (Census).

The model outperforms previous models based on three performance metrics: Pearson Correlation Coefficient (CC), R^2, and Mean

Absolute Error (MAE). The Correlation Coefficient was 0.95 for market share and 0.98 for demand as compared to 0.82 and 0.77 for previous models. However, the proposed model has higher time overheads in comparison with previous models because of the additional time for clustering and more advanced regression methods. The same authors above provided the extension of their work in another article (Bo An et al., 2017) where they introduce two new concepts on the basic Frequency-Based Profit Maximization algorithms in order to capture the conservative nature of airlines in deciding flight frequencies: bounded frequency and long-term profits. The tighter frequency bounds capture the scenario where airlines make only bounded changes to the frequencies even if it is more profitable to change the frequency by a large number. The second case expresses the case where airlines try to "drive off" other competing airlines and gain potential future profits by maintaining high frequency numbers which might not bring profit currently. Consideration of these two factors gave better profit maximization and proved that airlines are conservative in changing their frequencies and more concerned about long-term profits. The extension also investigated how optimal frequencies and profits can be calculated for the case where multiple airlines are strategic and independently change their frequencies (the original method assumes that only one airline is strategic in deciding its frequencies of a certain set of routes).

The decision of customers to buy a ticket for a given flight and route depends on various factors such as airlines' market share, customer membership (loyalty), and travelers' personal preferences of popular cities for destination and popular airlines for travel etc. (Jie Liu et al., 2017a,b). The article proposed a probabilistic framework model that enables to model airline customer travel preferences and to predict personalized airline passenger demand i.e. the destination and the airline an individual customer will choose. This is among the first works that proposes personalized air travel demand prediction. The approach utilizes Bayesian network-based topic model named Relational Travel Topic Model (RTTM) to model the preferences of customers and the characteristics of air routes and airline companies. Demand prediction is formulated using a Multiple Factor Travel Prediction (MFTP) framework that integrates multiple factors that influence the decision of customer's travel. Experiments are performed based on a 2-year passenger travel records of two cities in China (Beijing and Guangzhou,) consisting of more than 50 million flight records from more than 3 million customers with a total of around 550 air routes and 60 carrier companies. The data of the first year is used to train the models while the second-year data is used to test the models. The data is gained from airline reservation systems in the form of so-called passenger name records (PNRs). The PNRs contain the itinerary information of passengers and includes user-related information such as ID number, name, and gender, and flight-related information such as airline, origin and destination airport. Experiment results indicated that the proposed method is effective in demand prediction. A closely related work to that of (Jie Liu et al., 2017a,b) is proposed by Han-Tao Yang and Xia Liu, 2018. The paper came up with a model that predicts the airline passenger volume based on the daily passenger data of the airline for the route from Beijing to Sanya for the period between 2010 and 2017. The approach applied three types of prediction models: random forest, SVR and neural network. The result showed that the random forest prediction model achieved the highest accuracy with an MAPE of 4.18% followed by SVR: 6.87% and neural network: 12.38%.

The paper in (H. Yuan et al., 2014) develops a user behavior-based airlines ticket demand forecasting model to increase revenue of an OTA. The model estimates the effect of internal factors and external factors on the ticket sales market. It considers customer calls as internal factors and customer search engine query history as external factors. The model then compares historical weekly ticket price data fluctuations with the internal and external factors. Two search engine query keywords: "Ticket", and "Taobao Trip" were selected for external factors. The analysis is based on a 3-year customer call and ticket sales data collected from one OTA and search engine query data from Baidu (Dec. 2010 to Nov 2013). The data mining techniques used for the model include Neural Networks and two types of support regressions ($\mathcal{E}$-SVR and v-SVR). Experiment results indicated that external factors forecasted the most accurate prediction with Mean Absolute Percentage Error (MAPE) of 0.0466. Even though internal factors were also a good predictor of ticket sales with MAPE of 0.0491, the external factor gave better result than the combined effect of internal and external factors (0.0485). Prediction using only historical data gave the worst result (MAPE of 0.0522).

Other studies attempted to predict airline demand based on price elasticity. Price elasticity measures the degree to which a given flight is sensitive to price changes i.e. the extent to which changes in price will affect the demand. For example, a price elasticity of -1.5 indicates that 10% increase in ticket prices leads to a 1.5% decrease in demand. On the contrary, a 10% decrease in the ticket price leads to a 1.5% increase in demand. Price elasticity varies as a function of several factors. For example, it has been found that flights purchased on week days are more inelastic (less price sensitive) than flights purchased on weekends (María-Encarnación Andrés Martínez et al., 2017). Similarly, business class flights are more inelastic as compared to leisure class as business customers have less flexibility to change or cancel their travel date (Mumbower et al., 2014). In contrast, short distance flights are more elastic (more price sensitive) than long distance flights because of the availability of other travel options (e.g. bus, train, car etc.). Airlines use price elasticity information to determine when to increase ticket prices or when to launch promotions so that the overall demand is increased.

One of the latest studies conducted on this topic is the one presented in (Mumbower et al., 2014). In this paper, the authors forecast changes in airline demand based on flight-level price elasticity. The model first estimates price elasticity based on features such as the number of advanced bookings, departure day of the week, departure time of the day, booking day of the week, and competitor promotions. The number of booking for a flight for a given day is calculated as the number of seats which were "available" on that day but changed to "reserved" the next day and it was collected by tracking seat maps during the booking time. The estimated ticket price elasticity is then used to forecast the demand. Specifically, a linear regression method is employed to predict the number of bookings for a specific flight for a given departure date, route and number of days prior to departure date. The test data covers 13 flights for 4 routes across 21 departure dates (September 2, 2010 to September 22, 2010) with a total of 7522 bookings. However, the data is limited to one US domestic airline (JetBlue) and covers only non-stop flights. Moreover, it has been indicated that more than 25% of the observations were missing both price and demand information. The result showed that the estimated price elasticity of demand was -1.97. This price elasticity estimate showed close result with other similar previous studies. However, the extent to which this price elasticity is accurate in predicting the demand was not discussed.

A study by (Diego Escobari, 2014) investigated how demand changes with the number of days left before departure and found that the number of active consumers increases closer to departure date and also consumers become more price sensitive as time to departure approaches. The article in (Chieh-Hua Wen and Po-Hung Chen, 2017) made an attempt to predict changes in demand through examining the relationship between the purchase timing preferences of airline passengers and their characteristics. Unlike most other studies, this work is based on a survey data that consists of the customers' profile information (e.g. age, occupation, gender, education level, income) and flight information such as ticket price, purpose of travel, the frequency using low-cost airlines, airline, flight schedule, purchase date, weekly frequency of checking, airfare online, who paid the airfare, and travel companions. However, the data is limited in that it covers only a single route between Taiwan and Singapore and the respondents consist mainly of Taiwanese passengers. The results indicated that heterogeneous demand change patterns are observed for different types of passengers owing to the differences in their preferences of booking time. For example, demand for travelers who visit relatives and friends and users who frequently conduct on-line searches is high during early times and when fares are low. Whereas demand for people who travel for business purposes and demand for older people tend to increase few days prior to departure.

Among the most recent studies conducted on demand prediction are (Pan B. et al., 2018) and (Ali Mostafaeipour et al., 2018). Pan B. et al., (2018) proposed a long short-term memory (LSTM) based model to predict airline passenger demand. The traditional horizontal time series is used for short-term prediction (e.g. one day in advance) while a new vertical time series method is proposed for long-term prediction (e.g. half a month in advance). Ali Mostafaeipour et al., (2018) performed a study to predict air travel demand for all airports in Iran based on data provided by the Civil Aviation Organization of Iran from 2011 to 2015. Artificial neural networks are used to predict the air travel demand based on features such as income elasticity and population size of each area. Moreover, evolutionary *meta*-heuristic algorithms such as Bat and Firefly algorithms have been implemented in order to improve the performance of the neural network. The results indicated that the use of *meta*-heuristics algorithms increased the adaptation rate of neural network (NN) prediction and also increased the coefficient of determination from 0.2 up to 0.9.

The summary of Demand Prediction Models is given in Table 3.

3.2. Price discrimination

As indicated by several previous research (Mantin Benny and Bonwoo Koo, 2010; Marco Alderighi et al., 2011; Steven L.Puller and Lisa M.Taylor, 2012) airlines use various kinds of price discrimination mechanisms to charge customers different prices based on their willingness to pay for travel. However, most of the earlier studies are focused on testing a hypothesis to proof the existence of price discrimination and did not propose specific models or techniques for price discrimination. Moreover, mainly day dependent price discrimination was considered. The authors in (Steven L.Puller and Lisa M.Taylor, 2012) conducted research to check if there exists price discrimination based on the day of the week in which the ticket was purchased. A regression model is used to analyse ticket prices for the same flights purchased on different days of the week. The model considers controlling other factors which might affect the ticket price such as ticket restrictions (e.g. purchase deadline, travel restriction or duration of stay), factors that might influence the demand (e.g. the week of travel, time of the day) and the number of days before departure. The model is tested based on ticket transaction data collected for 85 US domestic routes across six major airlines (American, Delta, United, Continental, USAir and Northwest) for the fourth quarter of 2004. However, the data considers only nonstop round-trips.

Table 3
Summary of Demand Prediction Models.

Ref. No	Addressed Problem	Dataset	Features	Techniques Used	Performance Result	Remark
Bo An et al., (2016)	Route Demand and Market share prediction	10 years (40 quarters) of data for 13 airlines and 700 routes	Ticket price, number of flights operated by an airline, airline past performance history (delay time, delay ratio, cancel ratio, average stop and safety), aircraft size, total seat, average price, population income, customer price index (CPI) and Nash equilibrium pricing calculated based on existing models.	Ensemble Forecasting technique based on existing route demand and market share prediction models, clustering techniques and game theoretic analysis	Pearson Correlation Coefficients of 0.95 for market share and 0.98 for demand	- The prediction is quarterly - The model has higher overheads - (Bo An et al., 2017) is an extension of (Bo An et al., 2016)
H. Yuan et al., (2014)	Airlines ticket demand forecasting	3 years customer call, ticket sales and search query data	Internal factors (number of customer calls), External factors: Two query key words: "Ticket", "TaobaoTrip" and historical ticket price data	Neural Networks and two types of support regressions ($\mathcal{E}$-SVR and v-SVR)	Mean Absolute Percentage Error (MAPE) of 0.0466	Data set is limited
Mumbower et al., (2014)	Demand prediction based on price elasticity	Data for 21 departure dates with a total of 7522 bookings	Number of advanced bookings, departure day of week, departure time of day, booking day of week, competitor promotions	Linear regression	Price elasticity of 1.97	- There is no performance evaluation - Non-stop flights only - More than 25% of the data were missing both price & demand
Jie Liu et al., (2017)	Predicting personalized airline passenger demand	2-year passenger records consisting of more than 50 million flight from more than 3 million customers	ID number, name, and gender, and flight-related information such as airline, origin and destination airport	Bayesian network technique to model the behavior of customers and a Multiple Factor Travel Prediction framework to predict Demand	F1-score ≈ 0.33	The performance evaluation is based on the first 5 top-ranked probabilistic predictions
Chieh-Hua Wen and Po-Hung Chen, (2017)	Predicting changes in Demand	Not specified	Uses related features (e.g. age, occupation, gender, education level, and income) and flight related information such as ticket price, purpose of travel, airline, purchase date etc.	Trigonometric Function	–	- The data is limited in that it covers only a single route and homogenous passengers - No detail performance evaluation

The information contained in the data included ticket price, the date of purchase, date of departure, the airline, route, flight number and service class. The test discovered that airlines charge 5% less price for similar tickets purchased on weekends as compared to tickets purchased on weekdays. This finding is in line with other studies such as (Mumbower et al., 2014) which concluded that customers who prefer to purchase tickets on weekends as leisure customers are more price elastic than customers who choose to purchase tickets on week days as business customers. The paper further investigated whether the weekend purchase effect is consistent with price discrimination or not using cross-sectional variation in route characteristics i.e. by testing the weekend effect for various routes serving different volume of leisure and business travellers.

A closely related work to that of (Steven L.Puller and Lisa M. Taylor, 2012) was done by (Mantin Benny and Bonwoo Koo, 2010). The authors investigated whether day of week dependent price discrimination existed or not. The authors performed an empirical analysis to test the claim that "price dispersion during weekends is larger than that during weekdays while the average price stays constant over all days of the week". The variables included in the equations governing the hypothesis are the number of days prior to the departure date and the day of the week. The data used for the hypothesis test is collected from Farecast.com website. It consisted of the lowest daily airfare history for 6 departure dates (each Wednesday between February 27, 2008 and April 2, 2008, and returning 7 days later) spanning 90 days prior to the departure date. The data was gathered for 1000 randomly selected routes across all airlines resulting in approximately 540,000 observations per day. The test result showed that a strong weekend effect exits in the dispersion of ticket prices, but not in the price level. Therefore, the study concluded that airlines implement day dependent dynamic pricing discrimination. Moreover, the study indicated that this weekend effect is likely driven by the different types of consumers who purchase tickets on different days of the week.

Researches indicate that consumer profiling is performed by airlines for price discrimination based on either direct information sources (e.g. consumer registration) or indirect sources such as cookie files. The paper (Efthymios Constantinides and Rasha HJ Dierckx, 2014) conducted a research to identify what kind of customer information is exploited by airlines to perform customer profiling. Specifically, the aim of the research was to determine whether price discrimination is based on cookie data customer profiling or other user profiling methods. Experiments revealed that customer information from other direct sources seems to be more important than cookie data for customer profiling. However, there are several limitations to the experiment. First, the dataset incorporates only four European airlines and a limited number of routes. In addition, the experiment was conducted for short period (one month before departure) and only working days were considered.

Price discrimination increases the revenue of the airlines. On the other hand, it could have negative effect on the consumer welfare of airline customers. A study by (David Liu, 2015) developed a model that estimates the demand for airline with the presence of price discrimination and also investigated the effect of intertemporal price discrimination on consumer welfare. Intertemporal price discrimination refers to the scenario where different consumers enter the market at different times allowing airlines to apply various types of price discrimination. The proposed system first models the behavior of strategic customers who decide the optimal time to buy ticket a based on their beliefs about future prices, search costs, and their probability of flying. Consumers' belief about future prices is modeled as a Markov process based on flight characteristics and current prices. Experiments performed on a dataset consisting of daily price and quantity observations on 55 routes showed that price discrimination increases the demand

Table 4
Summary of Price Discrimination Studies.

Ref.	Addressed Problem	Dataset	Features	Techniques Used	Test and Result	Remark
Steven L.Puller and Lisa M.Taylor, (2012)	Analysis of whether Day dependent Price Discrimination exists or not	1000 US domestic routes from 90 days before departure with nearly 540,000 observations daily.	Days of the week, day of week of travel, ticket restrictions, the demand characteristics of the flights, the number of days in advance that the ticket is purchased	Regression	They found that fares are 5% lower when purchased on the weekend	- It is only analysis does not suggest a method for price discrimination - Limited number of features are considered - Only for domestic routes
Mantin Benny and Bonwoo Koo, (2010)	Investigate if Price Discrimination exists based on the day of week		Time which is the number of days prior to the departure date, Different weekdays and Prices of tickets	Hypothesis Testing	Hypothesis testing and the hypothesis is significantly accepted	
Efthymios Constantinides and Rasha HJ Dierckx, (2014)	Identifying the type of information used for price discrimination	Four European airlines and a limited number of routes	Direct information sources (e.g. consumer registration) or indirect sources such as cookie files	Hypothesis testing	Customer data from direct sources is more important than cookie data for price discrimination	
David Liu, (2015)	Investigate the effect of inter-temporal price discrimination on consumer welfare	Daily price and quantity observations on 55 routes	Prices, search costs, and their probability of flying	Markov process based on flight characteristics and current prices	Price discrimination increases the demand and decreases consumer welfare	

and decreases consumer welfare for both leisure and business travellers.

Alexander Luttmann, (2018) investigated factors used by airlines to price discriminate passengers depending on directional price discrimination i.e. based on trip origin/destination. The research found out that airlines charge customers different prices based on their flight origin. The research also indicated that directional price discrimination is a result of differences in passenger price elasticities between route endpoints. Customers departing from cities where the passenger price elasticity of demand is comparatively lower are charged a higher price (higher income reduces the price elasticity of demand). According to the article, prices are found to be $0.18-$0.43 higher on average for each $1000 difference in average per capita income between origin/destination cities.

Most of the existing literature on price discrimination considers temporal effect (days before departure) as the main driver for price discrimination. Table 4 illustrates the Summary of Price Discrimination Studies.

4. Other studies related to ticket price and Demand.

Besides the customer side and demand side models discussed in the previous two sections, there are also several other researchers that have been mainly conducted to investigate the role of various factors affecting ticket prices and demand (Martijn Brons et al., 2002; Silke J. Forbes, 2008; Tomasz Szopiński and Robert Nowacki, 2015; María-Encarnación Andrés Martínez et al., 2017). The factors determining the price elasticity of demand such as economic, demographic and geographic determinants for airline passengers is analyzed in (Martijn Brons et al., 2002). Their research finding indicated that price elasticity increases with time and leisure customers are more price sensitive than business customers. The authors in (Silke J. Forbes, 2008) analyzed the effect of air traffic delays on airline prices and found that prices fall by $1.42 on average for each additional minute of flight delay. The influence of purchase date and flight duration on the dispersion of airline ticket prices is studied in (Tomasz Szopiński and Robert Nowacki, 2015). According to this paper, price dispersion increases closer to the departure date and longer flights cause less price dispersion. Another study (María-Encarnación Andrés Martínez et al., 2017) examined the determinants of airfare pricing including presence of low cost carriers in the market, market domination, market share, and type of destination and reached on a conclusion that market dominance and the presence of low cost airlines have strong effect on ticket prices.

5. Discussion and analysis of existing work

In this section, we summarize, discuss and identify the strengths and weaknesses of existing work and suggest future directions.

5.1. Overall evaluation

Dynamic pricing is one of the most common pricing strategies implemented by the airline industry to adjust ticket prices in response to various internal and external factors such as changes in demand, competitor promotions, ability of users to buy, availability of seats and others. Airlines need to predict changes in these factors to implement a dynamic pricing scheme that dynamically adjusts ticket prices to increase their profit. On the other hand, customers are also interested to forecast how ticket prices would change in the future to be able to buy tickets at lower prices. Therefore, researchers have developed various prediction models both for airlines and customers to help them deal with dynamic pricing. The two most common methods proposed for airlines are demand prediction and price discrimination which we collectively refer to as Airlines side models. Customer side modes involve optimal ticket purchase time prediction models and ticket price prediction models. There is a tradeoff between money saving by customer and increasing revenue by companies. As customers become more strategic by using customer side tools, it becomes more difficult for the airlines to apply dynamic pricing and to generate profit and vice versa. Therefore, there is a need for a prediction model that can predict the optimal ticket prices that can bring mutual benefit both for customers and airlines.

Based on what we have presented, we can infer that ticket price prediction and demand prediction research is at an infancy stage. There is room for improvements in several areas including predicting exact value of ticket prices/demand, dataset issues, limited the number of features, lacking generality, better prediction techniques and performance and complexity issues. The majority of researches conducted in this area do not predict the exact value of a ticket price or the demand. For instance, most of the studies related to the ticket price predict the optimal time to buy a ticket (Etzioni et al., 2003; William Groves and Maria Gini, 2015). These models work in such a way that for each ticket query the customer performs, the model generates a binary signal indicating either to buy or to wait. However, the models do not predict the exact value of a ticket price in advance. Moreover, the maximum performance achieved so far is 75% which is not always acceptable. Nevertheless, there are few studies which attempted to predict the exact value of ticket prices (Y. Chen et al., 2015). However, the used models in these studies suffer from computational overhead as it is computationally more intensive than predicting the optimal purchase time.

In the area of demand prediction, the most notable work (Bo An et al., 2016) predicts quarterly route demand but cannot work for short term prediction. The other models in (H. Yuan et al., 2014; Mumbower et al., 2014) suggested for demand prediction only estimate the percentage increment or decrement in demand for a flight based on price elasticity. Another important topic that is not yet explored well is related to the development of a price discrimination model. None of the previous studies propose a technique for price discrimination but they rather focus on proving the existence of price discrimination in airlines pricing strategies.

Lack of generality is also one of the weaknesses noticed among existing studies. The prediction models proposed so far work either at flight level or route level and do not support prediction at both levels simultaneously. Moreover, a model that combines prediction for different kinds of flights such as non-stop flights, multi-stop flights, round trips and one way trips etc. is not proposed yet. On the other hand, dataset issues, limitation in features and techniques employed are probably the most important issues and need to be discussed in details. Therefore, we look at each of these in a separate section.

5.2. Dataset issues

The lack of benchmarking data is one of the major obstacles for researches in this area. To the best of our knowledge, there is no publicly available datasets that sufficiently satisfies the needs of the majority of research to be conducted. The most common public dataset used by many earlier researchers is the one provided by the U.S. Department of Transportation (DB1B). However, this dataset only provides a 10% average of ticket price for different itineraries performed in each quarter for US domestic flights. Moreover, it does not specify the purchase date or departure date which makes it unsuitable for predicting short term and flight level ticket/demand prediction. Hence, researchers who utilized this dataset are those who worked on long-term (e.g. quarterly) and route level predictions. In addition to the dataset provided by DB1B, few research papers have also started to provide access to data on request (Mumbower et al., 2014) and (William Groves and Maria Gini, 2011). However, these two datasets are also restricted in that they cover data gathered via web-scraping for a short period of time and are not sufficient to address most of the needs of researchers. Thus, researches rely on data extracted from different websites using scrapping programs according to their needs. Our study indicates that sample data collected by researchers in this way is limited in many ways: it spans a short period of time (usually less than 3 months), not heterogeneous, and is also of small size not effectively representing the population. We refer to heterogeneous data as a sample data that covers different aspects of a flight such as national and international, non-stop and multi-stop flights, different providers (airlines), duration of rounds trips (3 day rounds trips, 5 day rounds trips, 7 day rounds trips, 2-week round trips etc.). Earlier studies focused on one aspect and ignored the other. For example, some used national flight data while others relied on data from international flights. Moreover, most of these extracted data are based on round trips with constant length of duration (5 or 7 days). However, according to (T. Wohlfarth et al., 2011), ticket price patterns depend on length duration of round trips. In addition, the average number of routes handled by previous models is small as large number of routes requires huge computational capacity. The work by (Bo An et al., 2016) claims that the number of routes considered was 700 based on quarterly data from DB1B. However, DB1B data is aggregated data which does not contain detail individual ticket prices, purchase date and departure dates. Therefore, in order to facilitate research in this area, we recommend all interested parties to work together to build a public dataset that will be accessible by all researchers across the globe.

5.3. Features

Careful selection of appropriate features that can possibly affect prediction results is an important step towards building good prediction models. In this part, we summarize the set of features used in previous work and suggest additional features that are important for ticket/demand prediction. Table 5 shows the list of most features used by earlier studies mounting to around 30. The most commonly used feature among all was historical ticket price. Purchase day of the week and Departure day of the week were also widely used to predict ticket/demand. We can see from the table that existing models generally rely on limited number of internal factors to predict ticket price/demand. External factors are rarely considered. However, ticket prices and demand can also be affected by many external factors such as the global population mobility, presence of some event at the destination, volume of tourist traffic flow, weather condition, terrorist attacks, political instability, economic activity, natural disaster (hurricane, earthquake etc.), customer's sentiment about a destination city or the airline itself etc. For instance, evidence from (Bo An et al., 2016) and (H. Yuan et al., 2014) indicates that external factors play significant role in predicting ticket price/demand. The authors of (Bo An et al., 2016) utilized search engine query data to predict the number of ticket sales for an OTA company and external factors achieved more accuracy than internal factors. Moreover, (H. Yuan et al., 2014) included population income and customer price index as features to forecast demand and got good result. Even though many of the internal factors used by earlier researchers contribute a lot in predicting ticket pricing/demand, the incorporation of these external factors could lead to a more accurate result.

5.4. Techniques

A wide range of modeling techniques have been applied for ticket price/demand prediction. To summarize, we can classify the techniques employed so far into three categories. The first and the most commonly implemented set of techniques include simple hypothesis testing and regression techniques. Regression techniques perform well for relatively small size and homogenous dataset. However, they are not generally good enough to tackle complex models that make use of big and heterogeneous dataset. The second class of techniques used in previous research comprise of various kinds of data mining and machine learning techniques. Even though these approaches have the capability to handle heterogeneous data, acceptable accuracy levels were not achieved by using just a single data mining technique.

The third and latest approach applied before is ensemble learning that combines multiple individual data mining mechanisms to achieve better accuracy levels. Ensemble based learning methods are particularly employed by studies that predict exact values of ticket prices and demand (Y. Chen et al., 2015) and (Bo An et al., 2016). Ensemble learning based models have shown better accuracy than other methods. However, a relatively higher computational overhead has been observed in these types of approaches. It is noteworthy to observe that this computational overhead comes with relatively small datasets as mentioned earlier. Moreover, previous models mainly utilize features extracted from static and internal sources. However, as explained earlier, features extracted from external sources such as social media and websites are also a good source of information for better ticket/demand prediction. Therefore, future prediction models should incorporate features from both internal and external dynamic sources. Such kinds of models are expected to handle high computational overheads. Moreover, the models should also be able to support dynamic learning i.e. they should be able to learn incrementally from streaming real time data as data from dynamic sources update dynamically in real time. One of the most promising techniques for this could be deep learning. Deep learning is an emerging machine learning technique which is able to deal with handle huge data and dynamic learning.

6. Future directions

One of the future directions that has great potential to improve the ticket price and demand prediction is to use the latest and advanced machine learning techniques (i.e. deep learning) in conjunction with valuable social media-based data. Airline ticket prices/demand could be influenced by several dynamic factors which can be captured from social media data. This include occurrence of some event at the origin or destination city as mentioned earlier. Several previous studies proposed prediction models for various topics based on features extracted from social media including event prediction (A. Dingli et al., 2015; Arif Nurwidyantoro, and Edi Winarko, 2013; Chao Zhang et al., 2016; Hila Becker et al., 2012; Mario Cordeiro, 2012; Nikolaos Panagiotou et al., 2016; Q. Li et al., 2017; Takeshi Sakaki et al., 2010; Walther M. et al., 2013; Xiaowen Dong et al, 2015; Zhenhua Zhang et al., 2018) competitor intelligence (Lipika Dey et al., 2011; Malu Castellanos et al., 2011; Wu He et al., 2015), price prediction (A. Porshnev et al., 2013; L. Bing et al., 2014; L. Li and K. Chu, 2017) and tourist traffic flow prediction (R. Linares et al., 2015). In addition, features related to tracking competitors' promotions and sales helps an airline to estimate how ticket prices/demand would change in the future. The works done by (Lipika Dey et al., 2011) and (Wu He et al., 2015) indicate that social media based competitive intelligence can be applied to get features about competitor's

Table 5
Summary of Features Used by Previous Studies.

Feature	Optimal ticket purchase time prediction	Ticket Price Prediction	Demand Prediction	Price Discrimination	Papers Using the Feature
Flight Number	✓				Etzioni et al., 2003
Number of Days before departure	✓			✓	Etzioni et al., 2003; William Groves and Maria Gini, 2011; William Groves and Maria Gini, 2013; Steven L. Puller and Lisa M.Taylor, 2012; Mantin Benny and Bonwoo Koo, 2010
Quote Day of Week	✓			✓	William Groves and Maria Gini, 2013; T.Wohlfarth et al., 2011
Ticket Price history (minimum, maximum, mean, average, nash equilibrium price)	✓	✓	✓	✓	Etzioni et al., 2003; William Groves and Maria Gini, 2011; William Groves and Maria Gini, 2013; Anastasia Lantseva et al., 2015; Bo An et al., 2016; H. Yuan et al., 2014; Steven L. Puller and Lisa M.Taylor, 2012; Mantin Benny and Bonwoo Koo, 2010
Airline	✓				Etzioni et al., 2003; T.Wohlfarth et al., 2011
Route	✓				Etzioni et al., 2003
Number of Quotes per day	✓				William Groves and Maria Gini, 2011; William Groves and Maria Gini, 2013
Departure station	✓				T.Wohlfarth et al., 2011
Arrival Station	✓				T.Wohlfarth et al., 2011
Departure Date	✓	✓			T.Wohlfarth et al., 2011; Anastasia Lantseva et al., 2015
Return Date	✓				T.Wohlfarth et al., 2011
Departure Day of Week	✓		✓	✓	T.Wohlfarth et al., 2011; Mumbower et al., 2014; Steven L. Puller and Lisa M.Taylor, 2012
Departure Day of Month	✓				T.Wohlfarth et al., 2011
Departure Day of Year	✓				T.Wohlfarth et al., 2011
Demand (e.g. recent demand history)	✓			✓	T.Wohlfarth et al., 2011; Steven L.Puller and Lisa M.Taylor, 2012
Prices of (the same itinerary, recent itineraries, itineraries with the same day of week/month)		✓			Y. Chen et al., 2015
Departure City		✓			Anastasia Lantseva et al., 2015
Destination City		✓			Anastasia Lantseva et al., 2015
Ticket Purchase Date		✓			Anastasia Lantseva et al., 2015
Number of flights operated by airline			✓		Bo An et al., 2016
Airline performance (delay time/ratio, cancel ratio, average stop and safety) and capacity (aircraft size, total seat)			✓		Bo An et al., 2016
Population income			✓		Bo An et al., 2016
Customer price index (CPI)			✓		Bo An et al., 2016
Number of Customer Calls			✓		H. Yuan et al., 2014
Search Engine Query			✓		H. Yuan et al., 2014
Number of Advance bookings			✓		Mumbower et al., 2014
Departure time of day			✓		Mumbower et al., 2014
Purchase day of week			✓	✓	Mumbower et al., 2014; Steven L.Puller and Lisa M.Taylor, 2012; Mantin Benny and Bonwoo Koo, 2010
Competitor Promotions			✓		Mumbower et al., 2014
Ticket Restrictions (purchase deadline, travel restriction or duration of stay)			✓		Steven L.Puller and Lisa M.Taylor, 2012

products and promotions. It has also been shown that social media data can be used to model time series problems such as ticket price predictions. For instance, social media data has been used for real estate price prediction (L. Li and K. Chu, 2017), Stock market prediction (A. Porshnev et al., 2013; L. Bing et al., 2014), and many others. However, as far as our knowledge is concerned, there is no existing work that utilizes social media data to predict demand and or ticket prices.

Motivated by previous studies, we can think of various additional useful features from social media that can possibly forecast airlines passenger demand and or ticket prices. For example, sentiment analysis of different twitter hash tags could convey the presence of some event at a flight origin/destination city that improves the prediction of ticket price/demand. This kind of feature extraction might involve searching for special keywords or group of terms, determining the number of times they appear, understanding the location and the date, their context etc. Table 6 shows a sample of real user tweets from twitter social network, the possible search keywords that could be used and the effect of prediction on ticket prices/demand.

To determine if a given tweet belongs to some event or not, we need to first describe the event itself i.e. define the set of common words that can express the event well. For example, some of the terms that could frequently appear in a sports event are *match, game, competition, race, soccer, world cup, championship, tournament, fan, football, Olympics, baseball, swimming, basketball, cricket, cross country, golf, gymnastics, athletics, cup, stadium, names of clubs* (e.g. *Barcelona, Real Madrid, Manchester city etc.*), *there is, will play, will be playing, will take place etc*. The terms could be collected from various sources such as tweets related to the event, dictionaries or newspapers. To this end, we need to provide different weights to different terms as not all terms equally express an event. Moreover, the association between different terms also conveys different level of information about the event. The selected terms can be used as features for a given tweet. However, we know that a given tweet cannot contain all the terms specified for that event. Instead, it will contain only a subset of the selected terms only. Therefore, we represent the tweet data in a matrix form with each cell values set to either 0 or 1. If a tweet contains a given term, its corresponding value will be 1 otherwise 0. Based on this idea, we can prepare training dataset from real tweets manually labeled with the event name.

The features discussed above can then be used in conjunction with advance machine learning techniques for better prediction

Please cite this article as: J. A. Abdella, N. Zaki, K. Shuaib et al., Airline ticket price and demand prediction: A survey, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.02.001

Table 6
Example of feature extraction using real user tweets from twitter.

Real user tweets	Useful Key Words	Possible Effect on Price/Demand
Just wanted to leave this here: 5.5 to 6.5 earthquake predicted by Dutch for this week in Southern California: https://youtu.be/mOHFZs-0fZ0?t = 59m51s ...	"Earthquake", "Predicted", "California"	Ticket prices for flights departing from California will possibly increase
11/21/2017 — Pacific Northwest / Vancouver struck by M5.0 (M4.7) Earthquake as expected. Folks in Oregon/Washington be prepared!! Have a plan!!	"Earthquake", "Washington", "Prepare"	Ticket prices for flights to Washington might decrease
Thousands of Zimbabweans calling for, and celebrating the expected fall of President Robert Mugabe began marching towards his residence in Harare, #Zimbabwe, on Saturday, as the those opposed to #Mugabe's 37 years of a stronghold on power said he must go in #MugabeMustGo #Protest	"Harare", "Protest" "Thousands"	Demand and ticket prices to Harare might decrease
RT @makemoneyph United Airlines launches Hong Kong Airline Ticket Promotion http://ow.ly/1ozE71	"Airline Ticket" "Promotion" "Hong Kong"	The ticket price to Hong Kong might decrease

performance. Machine learning techniques are widely employed in many aspects of daily life, including automatic image/video search, autonomous driving, and recommendations on e-commerce web pages. Recently, deep learning techniques have revolutionized many areas of computer science including computer vision leading to dramatic performance improvements on a variety of traditional problems. Within deep learning, convolutional neural networks (Alex Krizhevsky et al., 2017) have received much attention recently and has been widely adopted by the computer vision community. These convolutional neural networks (CNNs) take a fixed sized RGB image or text as input to a series of convolution, local normalization and pooling operations (known as layers). Generally, the final layers in the convolutional neural networks are fully connected which are employed for feature extraction and classification. To the best of our knowledge, pre-trained word embedding and CNNs are yet to be explored for airline ticket price and demand prediction especially when considering external factors, including social media data and search engine query. As ticket price and demand prediction is a supervised learning problem, CNN can be utilized to classify images extracted from social media into popular destinations, future events, etc. This notion is illustrated in Fig. 2.

Other than CNNs, Recurrent Neural Networks (RNNs) (Jeffreyl Elman, 1990) analyze the text data word by word where the semantics of previously seen text is stored in the hidden layer. This hidden layer is of fixed size. Different to CNNs that are hierarchical, RNNs are sequential architectures and are shown to provide promising results on document-level sentiment classification (Duyu Tang et al., 2015). Two types of RNNs are commonly used: long short-term memory (LSTM) and gated recurrent unit (GRU). RNN, in particular LSTM is well suited the problem of predicting future ticket price and demand. Existing features such as current price, days to departure, quote day of week, departure airport, departure date, return date, provider, etc. can be combined with features which can be extracted from various sources including news media, search engines and social media. Suitable text analytics tools can be used to extract relevant information related to public feelings about a particular airline (sentiment analysis), popular destinations, events, environmental conditions, natural disasters, etc. For example, the study in (Bo An et al., 2016) has proven that external factors extracted from search engine query are good predictors of ticket sales demand. This idea is illustrated in Fig. 3.

In addition, existing ticket/demand prediction models are static models in which the number of samples and features are known ahead and do not change from time to time. Moreover, they rely on small datasets. On the contrary, machine learning models that are designed to work based on real time data such as social media are expected to deal with high dimensional and dynamic data whose exact number of samples and features is not known beforehand. Data from an online streaming system can be considered as big data as it comes in high volume and in various forms requiring high computational complexity. Moreover, such data is continuously growing i.e. both the number of instances and features may increase over time. The problem of streaming features and streaming instances is a recent topic that has attracted the attention of several researchers (Xindong Wu et al., 2010; Miguel García-Torres, 2016; Harshali D. Gangurde, 2014). Machine learning models which can support such kinds of behaviors are known as dynamic models. Therefore, social media-based ticket/demand prediction models should be dynamic models that can handle both computational complexity and vitality.

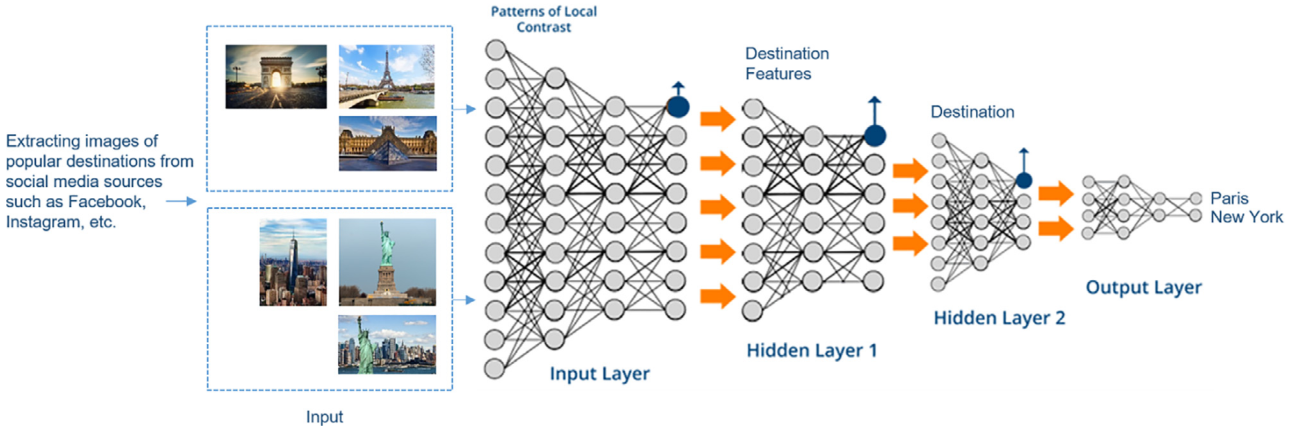
Extracting images of popular destinations from social media sources such as Facebook, Instagram, etc.
Input
Patterns of Local Contrast
Input Layer
Destination Features
Hidden Layer 1
Destination
Hidden Layer 2
Output Layer
Paris
New York

Fig. 2. Utilizing Convolutional Neural Network (CNN) to predict popular destinations, future events, environmental changes, etc. by classifying images extracted from various social media sources such as Instagram and Facebook.

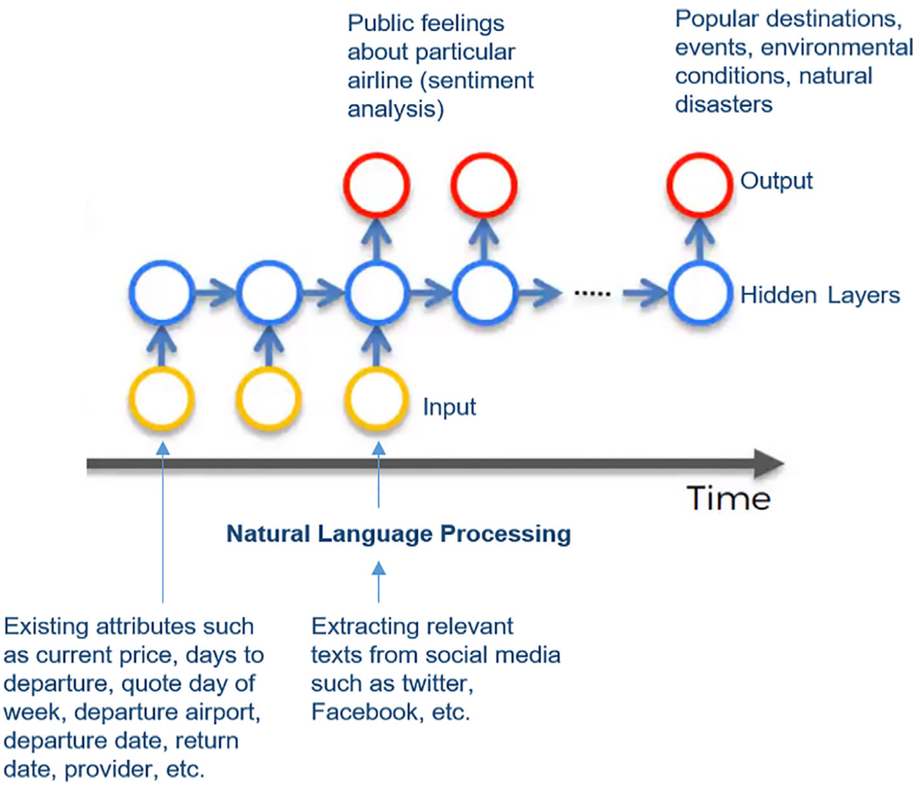
Public feelings about particular airline (sentiment analysis)
Popular destinations, events, environmental conditions, natural disasters
Output
Hidden Layers
Input
Time
Natural Language Processing
Existing attributes such as current price, days to departure, quote day of week, departure airport, departure date, return date, provider, etc.
Extracting relevant texts from social media such as twitter, Facebook, etc.

Fig. 3. Utilizing Recurrent Neural Networks (RNNs) to predict popular destinations, future events, environmental changes, public feelings about an airline (sentiment analysis) by analyzing text extracted from social media such as twitter.

Multiple events that can influence the airline demand/ticket price might occur in a particular airline destination city simultaneously. Different events affect traffic flow towards that location differently. Some events could lead to increase in demand while others might affect it negatively. For example, the occurrence of an earthquake in a particular city could decrease the demand towards that city. On the contrary, the presence of some sports event could increase the demand towards that city. The system proposed here first predicts if a tweet belongs to some predefined event that is considered by the airline as relevant to the airline business. Next, the tweets belonging to each event is clustered together as it helps to determine the magnitude of the event. A very big event would most probably receive large number of tweets than a small event. A weight coefficient is calculated for each event according to their magnitude. As mentioned earlier, the coefficient for some of the events could be positive while for the others, it could be negative. The total change in demand/ticket price is calculated based on the aggregate weight of the total events in the destination city. Therefore, an aggregate weight coefficient is calculated for all events. Once the total weight is known, it is used to estimate the increase/decrease in demand could be calculated.

7. Conclusions

In this paper, we presented a literature survey of ticket prediction and demand prediction models. We first presented an overview of dynamic pricing in airline industry which involves dynamic adjustment of ticket prices based on several internal and external factors. We explained the interaction between customers and airlines in deciding ticket prices dynamically. We then discussed two main previous research areas. Prediction models that are proposed to save money for the customer and those that are designed to increase the revenue of airline companies. Therefore, we classified existing models into customer side and airline side models based on their designed goals. We then summarized and discussed the strengths and weaknesses of existing work. Our analysis result showed that this research area has not been greatly explored and that there exist several aspects which need to be properly and thoroughly investigated including: performance issues, dataset issues, usage of dynamic external features such as social media data and search engine query. Therefore, we suggested and discussed a deep learning and social media data-based prediction model as the one of the most promising avenues of research going forward.

References

Dingli, A., Mercieca, L., Spina, R., Galea, M., 2015. Event detection using social sensors. In: 2015 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Rennes, 2015, pp. 35–41.

Porshnev, A., Redkin, I., Shevchenko, A., 2013. Machine Learning in Prediction of Stock Market Indicators Based on Historical Data and Data from Twitter Sentiment Analysis. In: 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, 2013, pp. 440–444.

Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey, 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.

Lantseva, Anastasia, Mukhina, Ksenia, Nikishova, Anna, Ivanov, Sergey, Knyazkov, Konstantin, 2015. Data-driven Modeling of Airlines Pricing. Procedia Comput. Sci. 66, 267–276. ISSN 1877-0509.

Arif Nurwidyantoro, Edi Winarko, Event detection in social media: a survey. In: International Conference on ICT for Smart Society (ICISS), Jakarta, 2013, pp. 1–5.

An, Bo, Chen, Haipeng, Park, Noseong, Subrahmanian, V.S., 2017. Data-driven frequency-based airline profit maximization. ACM Trans. Intell. Syst. Technol. (TIST) 8 (4), 61.

An, Bo, Chen, Haipeng, Park, Noseong, Subrahmanian, V.S., 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 (KDD '16). ACM, New York, NY, USA, pp. 421–430.

Zhang, Chao, Zhou, Guangyu, Yuan, Quan, Honglei Zhuang, Yu, Zheng, Lance Kaplan, Wang, Shaowen, Han, Jiawei, 2016. GeoBurst: Real-Time Local Event Detection in Geo-Tagged Tweet Streams. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR '16). ACM, New York, NY, USA, pp. 513–522.

Chawla, Bhavuk, Kaur, Ms Chandandeep, 2017. Airfare Analysis And Prediction Using Data Mining And Machine Learning. Int. J. Eng. Sci. Invention 6 (11), 10–17.

Wen, Chieh-Hua, Chen, Po-Hung, 2017. Passenger booking timing for low-cost airlines: a continuous logit approach. J. Air Transport Manage. 64, 91–99.

David Liu, 2015. A Model of Optimal Consumer Search and Price Discrimination in the Airline Industry.

Escobari, Diego, 2014. Estimating dynamic demand for airlines. Econ. Lett. 124 (1), 26–29.

Domínguez-Menchero, J. Santo, Rivera, Javier, Torres-Manzanera, Emilio, 2014. Optimal purchase timing in the airline market. J. Air Transport Manage. 40, 137–143.

Duyu Tang, Bing Qin, Ting Liu, 2015. Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015.

Efthymios Constantinides, Rasha H.J. Dierckx, Airline price discrimination: a practice of yield management or customer profiling? In: 43rd EMAC Conference Anonymous Paradigm shifts and interactions, Valencia, Spain, 3- 6 June, 2014.

Etzioni, Oren, Rattapoom Tuchinda, Craig A. Knoblock, Alexander Yates, To buy or not to buy: mining airfare data to minimize ticket purchase price. In: 9th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, USA, August 24-27, 2003, 119-128.

Yuan, H., Xu, W., Yang, C., 2014. A user behavior-based ticket sales prediction using data mining tools: an empirical study in an OTA company. In: 2014 11th International Conference on Service Systems and Service Management (ICSSSM), Beijing, 2014, pp. 1–6.

Gangurde, Harshali D., 2014. Feature Selection using Clustering approach for Big Data. Int. J. Comput. Appl., Proceedings on Innovations and Trends in Computer and Communication Engineering ITCCE (4), 1–3.

Dedhia, Manan, Jadhav, Amit, Jagdale, Rahul, Palkar, Bhakti, 2018. Optimizing Airline Ticket Purchase Timing. Int. J. Recent Innovation Trends in Comput. Commun. (IJRITCC) 6 (4), 296–298.

Boruah, A., Baruah, K., Das, B., Das, M.J., Gohain, N.B., 2018. A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter. In: Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, Springer, Singapore, vol. 714, pp. 191–203, 2018

Li, Yuling, Li, Zhichao, 2018. Design and implementation of ticket price forecasting system 1967, 040009.

Pan, B., Yuan, D., Sun, W., Liang, C., Li, D., 2018. A Novel LSTM-Based Daily Airline Demand Forecasting Method Using Vertical and Horizontal Time Series. Lecture Notes in Computer Science, 11154. Springer, Cham, pp. 168–173.

Mostafaeipour, Ali, Goli, Alireza, Qolipour, Mojtaba, 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), 5461–5484.

Han-Tao Yang, Xia Liu, 2018. Predictive Simulation of Airline Passenger Volume Based on Three Models. In: Data Science. ICPCSEE 2018. Communications in Computer and Information Science, Springer, Singapore, vol. 902. pp. 350–358, 2018

Luttmann, Alexander, 2018. Evidence of directional price discrimination in the US airline industry. Int. J. Ind Organiz. https://doi.org/10.1016/j.ijindorg.2018.03.013.

Becker, Hila, Iter, Dan, Naaman, Mor, Gravano, Luis, 2012. Identifying content for planned events across social media sites. In: Fifth ACM international conference on Web search and data mining(WSDM '12). ACM, New York, NY, USA, pp. 533–542.

Santos Domínguez-Menchero, J., Rivera, Javier, Torres-Manzanera, Emilio, 2014. Optimal purchase timing in the airline market. J. Air Transport Manage. 40, 137–143.

Elman, Jeffreyl, 1990. Finding Structure in Time. Cogn. Sci. 14 (2), 179–211.

Liu, Jie, Liu, Bin, Liu, Yanchi, Chen, Huipeng, Feng, Lina, Xiong, Hui, Huang, Yalou, 2017a. Personalized Air Travel Prediction: A Multi-factor. Perspect. ACM Trans. Intell. Syst. Technol. (TIST) 9 (3), 30.

Tziridis, K., Kalampokas, T., Papakostas, G.A., Diamantaras, K.I., 2017. Airfare prices prediction using machine learning techniques, 25th European Signal Processing Conference (EUSIPCO). Kos 2017, 1036–1039.

Bing, L., Chan, K.C.C., Ou, C., 2014. Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements. In: 2014 IEEE 11th International Conference on e-Business Engineering, Guangzhou, 2014, pp. 232–239.

Li, L., Chu, K., 2017. Prediction of real estate price variation based on economic parameters. In: International Conference on Applied System Innovation (ICASI), Sapporo, 2017, pp. 87–90.

Li, Jun, Granados, Nelson, Netessine, Serguei, 2014. Are consumers strategic? Structural estimation from the air-travel industry. Manage. Sci. 60 (9), 2114–2137.

Lipika Dey, Sk Mirajul Haque, Arpit Khurdiya, and Gautam Shroff, 2011. Acquiring competitive intelligence from social media. In: Proceedings of the 2011 joint workshop on multilingual OCR and analytics for noisy unstructured text data, ACM, NY, USA, 2011.

Malighetti, Paolo, Paleari, Stefano, Redondi, Renato, 2009. Pricing strategies of low-cost airlines: The Ryanair case study. J. Air Transport Manage. 15 (4), 195–203.

Malu Castellanos, Umeshwar Dayal, Meichun Hsu, Riddhiman Ghosh, Mohamed Dekhil, Yue Lu, Lei Zhang, Mark Schreiman, 2011. LCI: a social channel analysis platform for live customer intelligence. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (SIGMOD '11). ACM, New York, NY, USA, 2011, 1049-1058.

Benny, Mantin, Koo, Bonwoo, 2010. Weekend effect in airfare pricing. J. Air Transport Manage. 16 (1), 48–50.

Alderighi, Marco, Cento, Alessandro, Piga, Claudio A., 2011. A case study of pricing strategies in European airline markets: The London-Amsterdam route. J. Air Transport Manage. 17 (6), 369–373.

Andrés Martínez, María-Encarnación, Navarro, José-Luis Alfaro, Trinquecoste, Jean-François, 2017. The effect of destination type and travel period on *the* behavior of the price of airline tickets. Res. Transportation Econ. 62, 37–43.

Mario Cordeiro, Twitter event detection: combining wavelet analysis and topic inference summarization. Doctoral symposium on informatics engineering, Porto, 2012.

Brons, Martijn, Pels, Eric, Nijkamp, Peter, Rietveld, Piet, 2002. Price elasticities of demand for passenger air travel: a meta-analysis. J. Air Transport Manage. 8 (3), 165–175.

Längkvist, Martin, Karlsson, Lars, Loutfi, Amy, 2014. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recogn. Lett. 42, 11–24.

García-Torres, Miguel, Gómez-Vela, Francisco, Melián-Batista, Belén, Marcos Moreno-Vega, J., 2016. High-dimensional feature selection via feature grouping: A *Variable* Neighborhood Search approach. Inf. Sci. 326, 102–118.

Mumbower, Stacey, Garrow, Laurie A., Higgins, Matthew J., 2014. Estimating flight-level price elasticities using online airline data: a first step toward *integrating* pricing, demand, and revenue optimization. Transportation Res. Part A: Policy Practice 66, 196–212.

Narangajavana, Yeamduan, Garrigos-Simon, Fernando J., García, Javier Sanchez, Forgas-Coll, Santiago, 2014. Prices, prices and prices: A study in the airline *sector*. Tourism Manage. 41, 28–42.

Nikolaos Panagiotou, Ioannis Katakis, Dimitrios Gunopulos, 2016. Detecting events in online social networks: definitions, trends and challenges. In: Solving Large Scale Learning Tasks: Challenges and Algorithms, Lecture Notes in Computer Science, Springer, vol. 9580, 2016.

Li, Q., Nourbakhsh, A., Shah, S., Liu, X., 2017. Real-Time Novel Event Detection from Social Media. In: 2017 IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, 2017, pp. 1129–1139.

Linares, R., Herrera, J., Cuadros, A., Alfaro, L., 2015. Prediction of tourist traffic to Peru by using sentiment analysis in Twitter social network. In: 2015 Latin American Computing Conference (CLEI), Arequipa, 2015, pp. 1-7

Santana, Everton Jose, Mastelini, Saulo Martiello, Barbon Jr, Sylvio, 2017. Deep Regressor Stacking for Air Ticket Prices Prediction. XIII Brazilian Symposium on Information Systems: Information Systems for Participatory Digital Governance, 25–31.

Forbes, Silke J., 2008. The effect of air traffic delays on airline prices. Int. J. Ind. Organiz. 26 (5), 1218–1232.

Puller, Steven L., Taylor, Lisa M., 2012. Price discrimination by day-of-week of purchase: evidence from the US airline industry. J. Econ. Behav. Organ. 84 (3), 801–812.

Janssen, T., 2014. A linear quantile mixed regression model for prediction of airline ticket prices. Radboud University.

Liu, T., Cao, J., Tan, Y., Xiao, Q., 2017. ACER: An adaptive context-aware ensemble regression model for airfare price prediction. In: 2017 International Conference on Progress in Informatics and Computing (PIC), Nanjing, 2017, pp. 312–317.

Wohlfarth, T., Clemencon, S., Roueff, F., Casellato, X., 2011. A Data-Mining Approach to Travel Price Forecasting. In: 2011 10th International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, 2011, pp. 84–89.

Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo, 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web(WWW '10). ACM, NY, USA, 2010, 851-860.

Szopiński, Tomasz, Nowacki, Robert, 2015. The influence of purchase date and flight duration over the dispersion of airline ticket prices. Contemporary Econ. 9 (3), 353–366.

Vu, V.H., Minh, Q.T., Phung, P.H., 2018. An airfare prediction model for developing markets. In: 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 765–770.

Walther, M., Kaisser, M., 2013. Geo-spatial Event Detection in the Twitter Stream. In: Serdyukov, P. (Ed.), Advances in Information Retrieval. ECIR 2013, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, p. 2013.

William Groves, Maria Gini, 2011. A regression model for predicting optimal purchase timing for airline tickets. Technical report, University of Minnesota, Minneapolis, USA, Report number 11-025, 2011.

Groves, William, Gini, Maria, 2013. An agent for optimizing airline ticket purchasing, in International conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

Groves, William, Gini, Maria, 2015. On optimizing airline ticket purchase timing. ACM Trans. Intell. Syst. Technol. (TIST) 7 (1).

He, Wu., Harris, Wu., Yan, Gongjun, Akula, Vasudeva, Shen, Jiancheng, 2015. A novel social media competitive analytics framework with sentiment benchmarks. Information Manage. 52 (7), 801–812.

Dong, Xiaowen, Mavroeidis, Dimitrios, Calabrese, Francesco, Frossard, Pascal, 2015. Multiscale event detection in social media. Data Min. Knowl. Disc. 29 (5), 1374–1405.

Xindong Wu, Kui Yu, Hao Wang, Wei Ding, Online streaming feature selection. In: 27th international conference on machine learning (ICML-10), Johannes Fürnkranz and Thorsten Joachims (Eds.). Omnipress, USA, 2010.

Chen, Y., Cao, J., Feng, S., Tan, Y., 2015. An ensemble learning based approach for building airfare forecast service. In: 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, 2015, pp. 964-969.

Wang, Y., 2016. Dynamic pricing considering strategic customers. In: 2016 International Conference on Logistics, Informatics and Service Sciences (LISS), Sydney, NSW, 2016, pp. 1–5.

Xu, Y., Cao, J., 2017. OTPS: A decision support service for optimal airfare Ticket Purchase. In: 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1363-1368.

Yiwei Chen, Vivek F. Farias, 2015. Robust Dynamic Pricing With Strategic Customers. In: Proceedings of the Sixteenth ACM Conference on Economics and Computation (EC '15). ACM, New York, NY, USA, 2015, pp. 777–777.

Zhang, Zhenhua, He, Qing, Gao, Jing, Ni, Ming, 2018. A deep learning approach for detecting traffic accidents from social media data. Transportation Res. Part C: Emerging Technol. 86, 580–596.

Please cite this article as: J. A. Abdella, N. Zaki, K. Shuaib et al., Airline ticket price and demand prediction: A survey, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.02.001