Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6928833, 13 pages
https://doi.org/10.1155/2022/6928833

Research Article

An Effective Approach to Promote Air Traveler Repurchasing Using the Random Forest Algorithm: Predictive Model Design and Utility Evaluation

Zhenyuan Zhang, Yanhong Chen, and Luning Liu

School of Management, Harbin Institute of Technology, 92 Xidazhi Street, Nangang District, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Yanhong Chen; chenyanhong@stu.hit.edu.cn and Luning Liu; liuluning@hit.edu.cn

Received 13 March 2022; Revised 28 July 2022; Accepted 5 August 2022; Published 9 September 2022

Academic Editor: Chung-Wei Kuo

How to promote air traveler repurchasing has become an important marketing strategy in airlines. However, because of the growing concern over user privacy, effectively and accurately delivering advertising to promote repurchasing has become more difficult. Here, we propose an effective framework based on machine learning to model the air traveler repurchasing and furthermore employ a field experiment to test the utility of a model framework. Specifically, we collected that this model framework is based on the random forest algorithm and compared with the conclusions of the other four algorithms, K-nearest neighbor, decision tree, support vector machine, and ExtraTree algorithms. The results show that the proposed model framework is better than the prediction results of the other algorithms. In addition, the proposed model framework was verified through a real case of an airline in China. This study will serve as a guide to analyze the repurchase behaviors of an air traveler and help airlines build a loyal air traveler base.

1. Introduction

The repurchase of travelers is the possibility of buying services or products from the same airlines, and the prediction of repurchase intention is an important index to measure travelers' repurchase behavior [1]. With the pervasiveness of smartphone and Internet penetration in our daily life continuously increasing, travelers are increasingly searching information from multichannel to make choices. However, on the one hand, owing to its intrusive nature in practice, there has been a growing concern over user privacy [2]. On the other hand, unlike traditional businesses, firms involved in mobile commerce find it difficult to obtain air traveler demographic or family background, because data are usually regarded as private [3]. Consequently, this gives rise to the question of how to model travelers' purchase intention and effectively deliver advertising to promote air traveler repurchasing.

In the context of mobile commerce, the traditional way for enterprises to carry out advertising-related marketing is to send mobile advertisements to all the travelers who have a purchase record. However, this method leads to high costs and inefficiencies, and even causes travelers aversion. Advertising-related marketing that targets travelers using time and location based on mobile technologies has been studied [4–7]. However, the premise of this advertising-related marketing is that travelers turn on the positioning of mobile phones or are willing to receive mobile marketing. Some travelers are extremely disgusted with such advertising and ignore it [8, 9]. Therefore, we proposed a model framework to predict the traveler repurchasing intention using machine learning based on the actual behavior data in airlines. The goals were to identify travelers with high purchase intentions in advance using machine learning and then to target these travelers using the proposed mobile advertising methodology.

Data-mining techniques are often employed to transform a large volume of data into valuable knowledge that may be used to support marketing decision-making [10, 11]. Many businesses utilize information mining technologies, such as data-mining techniques, to extract air traveler data to validate their strategy before implementation [12, 13]. Machine learning is an important data-mining technique. Machine learning and data-mining techniques for predictive purposes on an air traveler basis are often applied in air traveler relationship management and expert systems domains [14, 15], and predicting air traveler repurchases is a popular objective in these fields. In order to identify travelers with high purchase intention, in this study, the existing machine learning model has been proposed to predict the purchase intention of travelers in airlines.

A variety of predictive models for air traveler repeat purchasing have been developed over the past decades in traditional businesses. Nonparametric regression models, such as the k-nearest neighbor (KNN) model [16], negative binomial distribution models [17], logistic regression models [18, 19], Bayesian neural network learning models [20], time evolution models, such as time series models [21], and artificial bee colony (ABC) models [12], are commonly used. However, those methods and algorithms used to model in the traditional business context have to be modified for mobile commerce in airlines [10].

It is well known that random forest (RF), as an important algorithm in machine learning, is efficient and accurate prediction compared with the other machine learning models in many applications [16]. RF models improve the accuracy of regressions without greatly increasing the computational complexity. Additionally, RF models can explain the importance of thousands of variables [18]. RF is a classification algorithm well suited for microarray data, and it shows excellent performance even when most predictive variables are noise. It can be used when the number of variables is much larger than the number of observations and in problems involving more than two classes. Additionally, it returns measures of variable importance [4]. Because of these advantages, RF models have been selected to model on repurchase intentions in this study.

In this study, first, we compared RF models with the other machine learning methods, such as KNN, decision tree (DT), support vector machine (SVM), and extremely randomized trees (ET), to predict traveler repurchase intentions. The performance of RF was better than those of other machine learning algorithms. Then, we used a large-scale randomized field experiment to verify our research question, that is, after confirming traveler purchase intentions, does sending a promotional short messaging service (SMS) effectively promote traveler repurchases? This is an SMS-based advertising promotion organized by an airline in China. Three treatment groups and one blank control group were used to reduce the heterogeneity caused by the SMS content. Comparative analyses were performed among the treatment groups and between the treatment and control groups. By combining machine learning and experimental methods, we determined the main factors affecting the repurchase intentions of air travelers and effectively predicted their purchase intentions. Then, we determined that sending targeted SMS promotional ads after confirming travelers' purchase intentions effectively promoted traveler repurchases. This has practical significance to allow airlines to perform precision marketing in the mobile commerce context.

The remainder of this article is organized as follows: Section 2 describes the previous literature regarding air traveler repurchase, purchase predictions, machine learning, and SMS promotions. The methodology is described in Section 3, consisting of the problem statement, framework design, and evaluation. Section 4 reports the empirical results of the proposed algorithm's utility evaluation using a field experiment. Finally, conclusions and directions for further research are provided in Section 5.

2. Literature Review

2.1. Customer Repurchases and Purchase Intentions. Repeated purchase intention refers to the subjective probability/implied promise that an air traveler will continue to purchase a product from the same seller [1]. Traveler repeat purchase behavior has been a topic of interest in many contexts [22], such as in B2C e-commerce [1, 10, 23], on O2O platforms [24], in direct marketing [18, 20], and in mobile commerce [8, 25]. In our work, we focused on an airlines' direct marketing context, especially the low-cost carriers. The low-cost carriers are those that offer cheap tickets, while passengers need to pay for the auxiliary services (e.g., baggage checking and seat selection) [26]. Passengers of the low-cost carriers are usually low income or travelers. Hence, the low-cost carriers need to expand a wider group of travelers than full-services airline, in order to survive in the highly competitive field of aviation.

Most of the literature has focused on what drives the repeat purchase behavior of travelers [1, 23, 27–31]. Chiu et al. [1] indicated that both the utilitarian and hedonic values were positively associated with air traveler repeat purchase intention. A higher level of perceived risk reduced the effect of the utilitarian value and increased the effect of the hedonic value on repeat purchase intention. Bukhari et al. [30] developed a framework for measuring traveler willingness to ticket purchase intention through airlines' websites. Pandey and Srivastava [29] comprehensively studied the antecedents of the air traveler purchase intentions using chronological order and various parameters. Several studies [19, 32, 33] supported the contention that travelers expressing the intentions of making repeat purchases can be identified as loyal. Consequently, other studies focused on the impacts of loyalty programs on repeat purchase behavior [28, 34, 35]. While in this study, we focus on the antecedents that influence repurchase behavior of air travelers using real travel and individual behavior data. Besides, we also focus on how to send promotional information based on the predicted likelihood of purchase intentions.

2.2. Air Traveler Purchase Predictions and Machine Learning. In traveler and marketing research, purchase predictions have received a great deal of attention [14]. The use of observed past air traveler behavior to understand their current and to predict their future behavior is a core idea behind traveler base analysis [36]. Additionally, predicting future air traveler behavior provides key information for efficiently directing resources in sales and marketing departments [14]. Specifically, determining the purchase intentions of air travelers to send targeted mobile advertising has significant managerial implications to allow airlines to use precision marketing.

Various approaches have been employed to model purchase decisions to understand air traveler repeat purchases. Logistic regression models and a structural equation model have been applied to determine which variables affect purchase decisions [22, 23, 37]. Three categories of response modelling have been proposed successively. The first category is purchase incidence modelling [14, 20, 38, 39], in which the purchase incidence is predicted within a fixed time interval (typically half a year). Other studies have investigated related problems dealing with both the purchase incidence and the purchase amount in a joint model [40]. A third alternative is to model interpurchase time through a survival analysis or (split-) hazard rate models [21]. However, historical of travelers in airlines has the characteristics of hug data features, serious missing values and much noise, the traditional models in data processing, high dimensional data, missing values data, and often being unable to obtain satisfaction of performance. Hence, new models based on the traditional models need to be developed.

Machine learning has been used to research air traveler repurchase intentions in traveler marketing [12, 14, 18, 20]. Baesens et al. [20] proposed that Bayesian neural networks offered a viable alternative for purchase incidence modelling in direct marketing. Kumar et al. [12] combined machine learning techniques (e.g., decision trees, SVM, and NN) and the artificial bee colony algorithm to predict online traveler repurchase intentions. Valecha et al. [18] proposed time-evolving RF classifiers to predict traveler purchasing behavior. Zheng et al. [13] developed a machine learning model framework for forecasting air transportation direct share. In this study, we developed a framework based on the RF algorithm and evaluate the validity of the framework model using a field experiment in airlines.

2.3. Mobile Marketing and SMS Promotions. Mobile marketing and SMS promotions have been highly studied and remain active areas of research. Mobile marketing enables market policymakers to target travelers by time and location through mobile technologies [6]. Therefore, previous empirical studies focused on how travelers respond to mobile marketing based on temporal and geographical targeting [4, 5, 7, 41]. Luo et al. [6] analyzed how different combinations of geographical and temporal targeting determined traveler responses to mobile promotions. Jahromi et al. [38, 42] demonstrated the effectiveness of sending mobile promotions to travelers that were near a competitor's location based on a randomized field experiment. Phang et al. [7] researched how the time of day influences users' IT behaviors through a field experiment, and they provided actionable guidance to practitioners in performing mobile time-based targeting. Some prior studies on traveler responses to mobile ads have examined contextual factors, such as geographic mobility [9], weather [43], product or user characteristics [44, 45], and hypercontextual targeting in conjunction with physical crowdedness [46]. However, it is rarely mentioned that sending promotional messages based on their purchase intention, although some studies have explored the positive effect of willingness to buy on actual buying behavior [1, 19].

SMS is a proven fast, effective, highly personalized, and low-cost way to reach travelers using mass mobile advertisements [2, 47]. SMS advertising provides a number of advantages to marketers, such as cost reduction, ubiquity, immediacy, and targeted message delivery [2]. Chen et al. [48] compared the efficacies of an SMS text message and a telephone reminder to improve attendance rates at a health promotion center, and they found that the SMS reminder was more cost-effective compared with the telephone reminder. Mohammed et al. [49] conducted a quasi-experimental study to assess the effect of a theory-driven mHealth intervention (health SMS messages) on the prevalence of malaria, while this research addresses how to effectively use machine learning algorithms and mobile marketing methods to promote travelers repurchase behavior in airlines.

3. Framework Design and Evaluation

3.1. Statement of the Problem. A well-known principle of economics, Pareto's principle, states that 80% of an airline's profits come from 20% of travelers [14]. Thus, appropriate retention strategies have strong benefits to airlines over acquisition approaches in marketing [12]. Traditional marketing methods are based on word of mouth or discount promotions. However, with the increasing development of the Internet and information technology, mobile marketing has become the main marketing tool of major industries [45–47]. The traditional way for enterprises to perform advertising marketing is to send mobile advertisements to all the travelers who have a purchase record. However, this method is costly and inefficient, and it even results in traveler aversion. Thus, how to use mobile marketing technology to promote products has become a major research topic [6, 7]. However, as awareness of traveler information security and privacy protection increases, advertising marketing based on travelers' personal location and mobile phone numbers has offended travelers [2]. In this era of big data, on the one hand, airlines generate masses of traveler behavior-related data. If the marketing decision-makers make full use of the data resources collected, then the airline will reap great benefits [12]. On the other hand, loyalty and intention towards the product/services play very vital roles in air traveler repurchases [18]. Thus, how to effectively and accurately identify travelers in need (e.g., travelers with high purchase intention) using data-mining technology to reduce traveler resistance to product advertising requires exploration.

3.2. Framework Design. The primary goal of this study was to establish a standardized framework that helps airlines delivering traceable information to effectively send SMS messages and promote repeat passenger purchases. The overall framework design is depicted in Figure 1 (framework design). Building this framework involved the following four main steps:

Step 1: Collected relevant and available features of passengers' traveler behaviors from a management information systems (MIS) (Figure 1, relevant features variables).

This data collection typically requires collaboration with airlines, and the available information items provided by the MIS, including demographic and purchase behavior characteristics, were sorted. This produced a very large amount of information that was not all related to traveler repurchase intentions. To reduce the algorithm's running time and improve its efficiency, it was necessary to identify important features that affect traveler repurchase intentions. Then, using a data preprocessing approach involving machine learning (e.g., OneHotEncoder or LabelEncoding), the classification features (e.g., departure and arrival cities) were encoded.

Step 2: Selected important features using RF (Figure 1, select features).

For the preprocessing dataset, we selected importance features using RF, which returns several measures of variable importance. The most reliable measure is based on the reduction of classification accuracy. When the value of a variable is randomly arranged in a node tree, this is the importance of measuring the variable [50]. Thus, we used RF models (e.g., scikit-sklearn) to preprocess the data and collect the important features. To reflect the importance of each feature, maps of relative importance were used.

Step 3: Selected the best machine learning algorithm to predict travelers' repurchase (Figure 1, machine learning algorithm selecting).

A variety of prediction models for air traveler repeat purchases have been developed over the past decades; however, some machine learning models are used because of their advantages, such as KNN, DT, SVM, ET, and RF. Each machine learning method may be selected for the following reasons:

(1) The KNN model has been extensively employed in many fields as a nonparametric regression model [37, 38]. It assigns the label of its nearest neighbor to an observation and determines the class through majority vote. Yu et al. [16] proposed a method combining RF and KNN to predict bus arrival times.
(2) During DT training, term features are commonly used, and each tree is built based on a random subset of features [28]. The dimensions of a dataset are the features. The DT is a special machine learning method derived from RF [19].
(3) SVM models represent a specific type of learning algorithm characterized by the capacity control of the decision function, the use of the kernel functions, and the sparsity of solutions [24]. Prior work has shown that SVMs [51] are excellent tools for predictive tasks.
(4) ET essentially consists of randomizing strongly both attribute and cut-point choices, while splitting a tree node. In extreme cases, it builds totally randomized trees having structures that are independent of the output values of the learning sample [52]. In addition to accuracy, the main strength of the resulting algorithm is computational efficiency.
(5) An RF model is constructed using a random vector of the data feature space [53]. Generally, RFs have shorter calculation times, and the problem of multicollinearity can be ignored. RFs are not sensitive to outliers and remain robust despite missing data. In addition, RF models can reduce overfitting [5, 16, 54]. Unlike a single decision tree, RFs are a combination of multiple decision trees. Each tree is a classification or regression of a certain set of features. The final results of an RF are obtained by voting on all the trees in the RF, which is superior to the single classifier model. Valecha et al. [18] proposed a time-evolving RF classifier to examine the relationship between the categories of traveler behavior and repurchase using survey data.

Step 4: Targeted mobile advertising using the best machine learning algorithm (Figure 1, consumers' repurchase). The traditional way of advertising is to target all the travelers who have a purchase record. However, this method is costly and inefficient, and it may even result in traveler aversion. Thus, using the optimal predictive repurchase model to calculate the traveler repurchase intention values, and sending SMSs to the travelers with high purchase intentions represents an attractive alternative.

3.3. Framework Evaluation. To assess the utility and usability of the framework, we utilized a case and tested it using the proposed algorithm with secondary data. The dataset on past purchase behavior at the order-line level had been collected from an airline in China. This allowed us, in close cooperation with domain experts and guided by the literature, to derive all the necessary purchase behavior variables for 14,532 air travelers. For each air traveler, these variables were measured in the period between 15 October 2016 and 15 October 2019.

An RF returns several measures of variable importance [4]. First, we used scikit-sklearn to preprocess the data, including the missing values, along with OneHotEncoder and LabelEncoding. Later, we selected 28 (Table 1) of 74 features, including the demographic and purchase behavior characteristics from a previous study, for an RF [50]. The importance of the 28 variables is shown in Figure 2. The correlation coefficients among the 28 variables are shown in

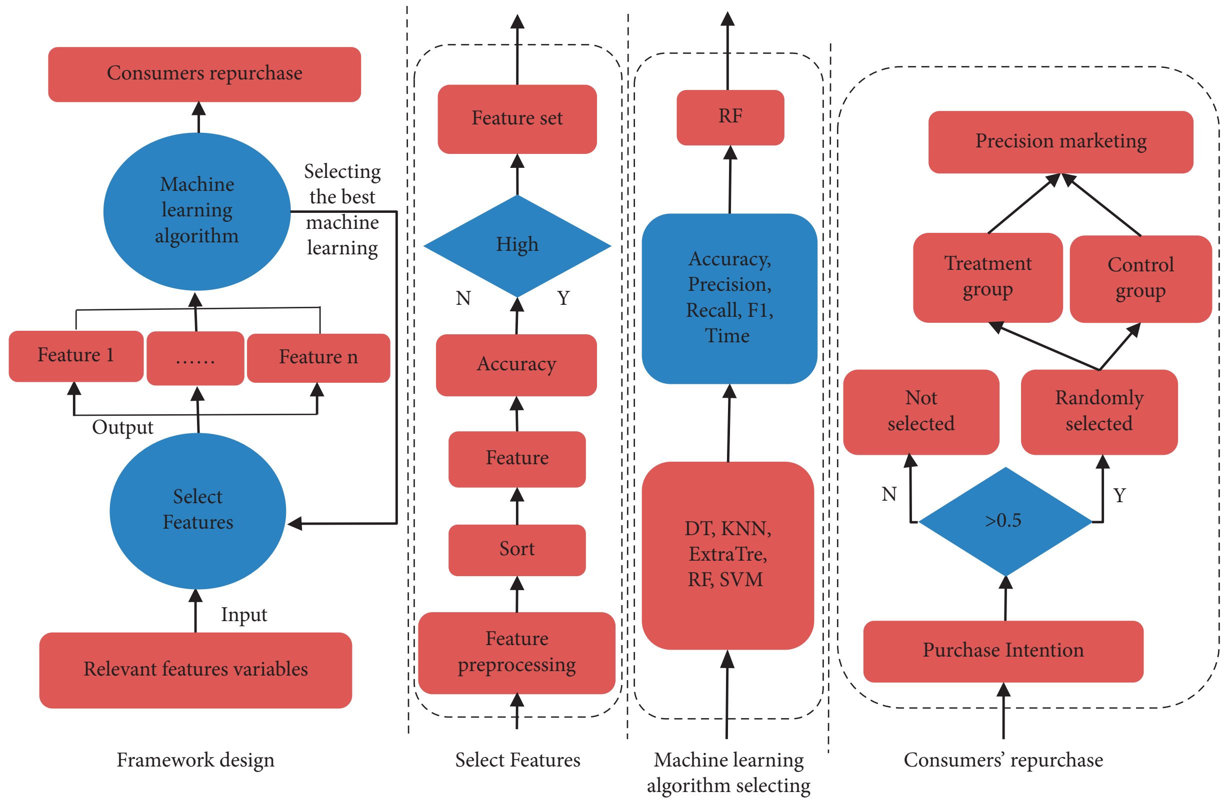

Figure 1: A framework for predicting traveler repurchases.

Table 1: The definitions of 28 feature variables.

Variables	Defines
Cust_type	Membership level (1–5); ranging from 1 (lowest level) to 5 (highest level) with the midpoint 3 (general level);
Status	Account status (0–2); elapsed: 0; slumber: 1; active: 2;
Is_lyreg	Is it a member account; nonmember: 0; member: 1;
Ly_status	Whether the member account is activated; nonactivated: 0; activated: 1; very active: 2
Gender	User gender; female: 0; male: 1;
Card_province	Province where the user ID is located;
Tel_province	Province where the user's mobile phone is located;
Order_buy_zh	Whether to buy economy class tickets; noneconomy class: 0; economy class: 1;
Order_language	The language preference of air tickets orders;
First_buy_diff	Difference between first purchase date and registration date;
Last_diff	Days between last two purchases;
Last_amount	The amount paid for the last purchase;
Flight_mile	Total flight distance;
Order_num_m24	Number of orders paid in the last 24 months;
Order_num_m12	Number of orders paid in the last 12 months;
Order_num_m6	Number of orders paid in the last 6 months;
Order_num_m3	Number of orders paid in the last 3 months;
Order_num_d30	Number of orders paid in the last 1 month;
Tkt_fee_all	Total payment;
Tkt_fee_m12	Total payment amount in the last 12 months;
Tkt_fee_m6	Total payment amount in the last 6 months;
Tkt_fee_d30	Total payment amount in the last 1 month;
Avg_daydiff	Average purchase frequency;
Avg_tktdiscount	Average discount;
Avg_tktfee	Average ticket price;
User_type	Type of membership (general member, silver member, gold member);
User_cycle	Member's life cycle;
Order_num_all	Total payment orders;

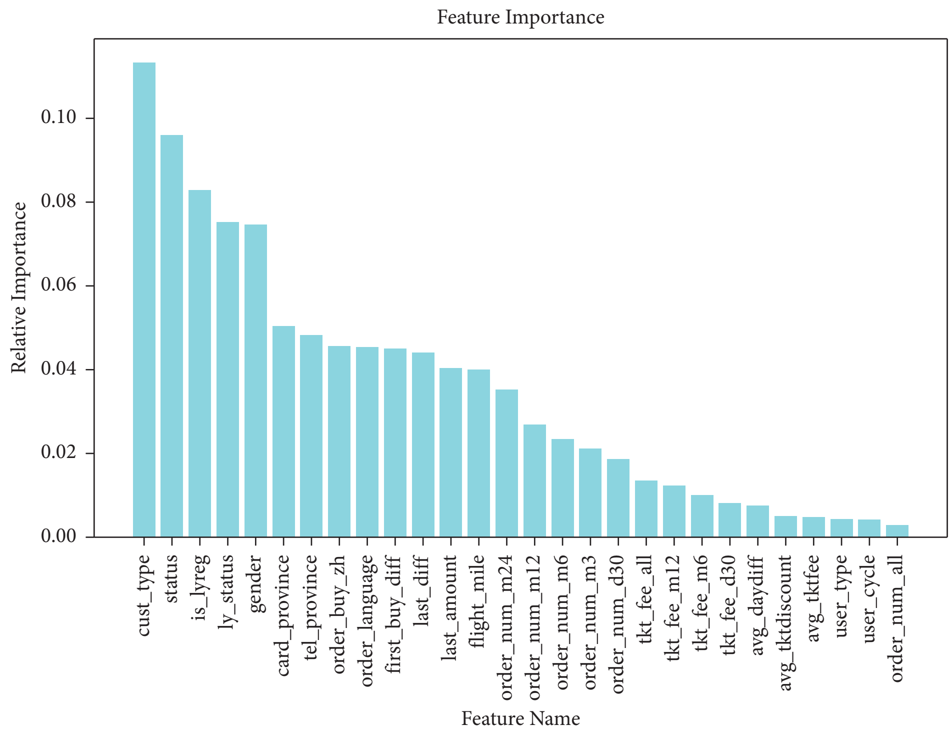

Figure 2: The importance of features.

Table 2: (70–30%) of data: models' performances.

Models	Accuracy	Precision	Recall	F1	Time (minutes)
KNN	0.8925	0.7882	0.7687	0.7721	10.7984
DT	0.8995	0.8028	0.7768	0.7888	0.4558
RF	**0.9224**	**0.8706**	**0.7957**	**0.8268**	**5.8581**
SVM	0.8800	0.7597	0.8153	0.7826	1 : 23.6638
ExtraTree	0.9157	0.8426	0.8006	0.8194	0.3730

Figure 3. The variable definitions of 28 features are depicted in Table 2.

From Table 2, Figures 2 and 3, we learned that among the factors that affect traveler repurchases, a membership level had the greatest influence, followed by account status. Thus, the higher the user's membership level, the greater the likelihood of repeated purchases [28]. This indicates a greater user loyalty to the airline. Account states include slumber, active, and elapsed. In the context of airlines, we found that the factors that influenced passenger ticket purchases again included sociodemographic features (e.g., gender, card_province, and tel_province) and purchasing behavior feature variables (e.g., order quantity and order fee). Airline managers should consider these feature variables to promote passengers to buy tickets in the future.

Various classification algorithms were used, including KNN, DT, SVM, ET, and RF. The prediction results of the five classification models were analyzed, and all the models were run on 70–30 training-testing partitions (Table 3). RF had the highest accuracy of 92.24% and precision of 87.06% on the training-testing partitions. The results confirmed that RF outperforms the other learning models for predicting traveler purchase intentions. The 5-fold and 10-fold cross-validations were also analyzed to demonstrate the robustness of the results (Figure 4).

4. Utility Evaluation

4.1. Experimental Preparation and Process. In collaboration with an airline in China, we conducted a large-scale randomized field experiment to understand whether sending SMS ads, after identifying traveler repurchase intentions, increased repeat purchases. The experimental design was divided into two stages: preparation and process.

Experimental preparation:

(1) RF was used to predict the purchase intentions of travelers who had purchased air tickets from this airline from 9 February 2019 to 9 August 2019.

(2) To reduce the cost, 80,000 travelers having a purchase intention greater than 0.5 were randomly selected. On the basis of their purchase intentions, the 80, 000 users were divided into four groups (A-A, A-B, B-A, and B-B).

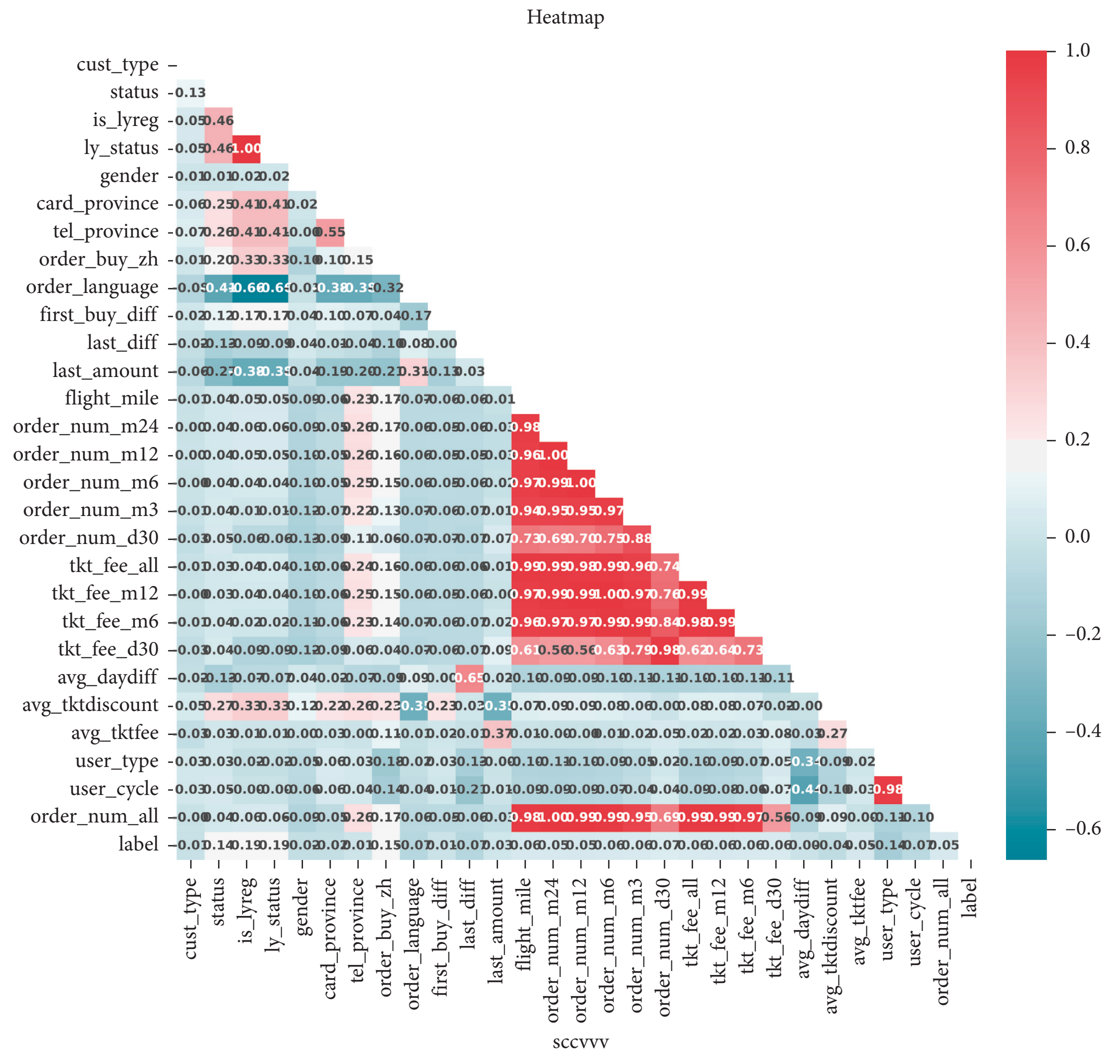

Figure 3: Correlation coefficient matrix.

Table 3: The data of "89 promotions" campaign.

Groups	Numbers of air travelers purchased	Quantity of orders	Numbers of tickets
A-A	4644	6494	10630
A-B	4518	5799	9582
B-A	4580	5987	9895
B-B	4498	5844	9614

(3) To reduce the deviation caused by the SMS content, we designed three different SMS content experimental groups—A-A, A-B, and B-A—and a blank control group (did not receive any promotional SMSs).

To eliminate heterogeneity caused by the time when the SMS ads were sent, we sent the promotional SMSs at two different times.

Experimental process:

(1) Sent promotional SMSs to the A-A and A-B groups at 10 am on 9 August 2019. The following SMS was sent to the A-A group: "Welcome to the "89 promotions" campaign. If you buy an air ticket now, you can enjoy a discount of 88 yuan. GreenWing members can enter the campaign 30 minutes in advance. Click! Return T to unsubscribe." The following SMS was sent to the A-B group: "The "89 promotions" campaign is limited to 3 days, and the air tickets are 9 yuan or less! GreenWing members can enjoy a maximum discount of 88 yuan. Click! Return T to unsubscribe."

(2) Sent promotional SMSs to the B-A group at 10 a.m. on August 12. The following SMS was sent to the B-A group: "The "89 promotions" campaign is back, and the air tickets are 9 yuan or less! There is no need to wait for the special low-price tickets in autumn and winter, and now, you have a chance to win ticket coupons! Click! Return T to unsubscribe."

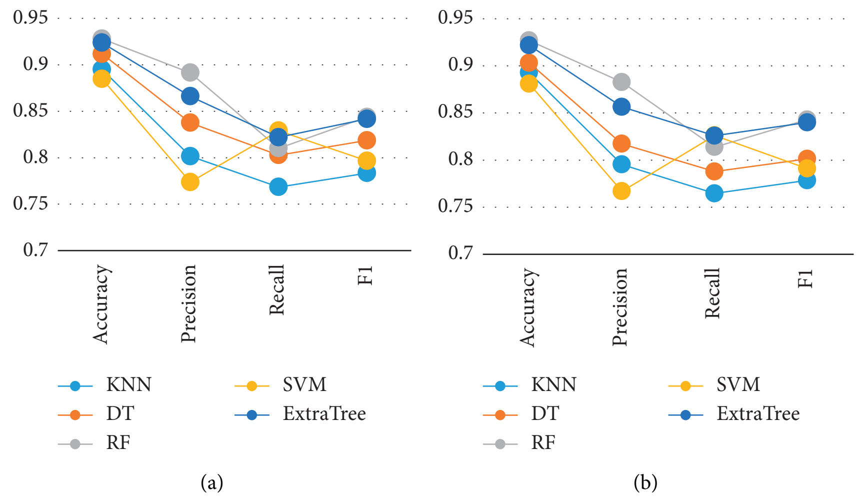

FIGURE 4: K-fold cross-validation. (a) 10-Fold presentation. (b) 5-Fold presentation.

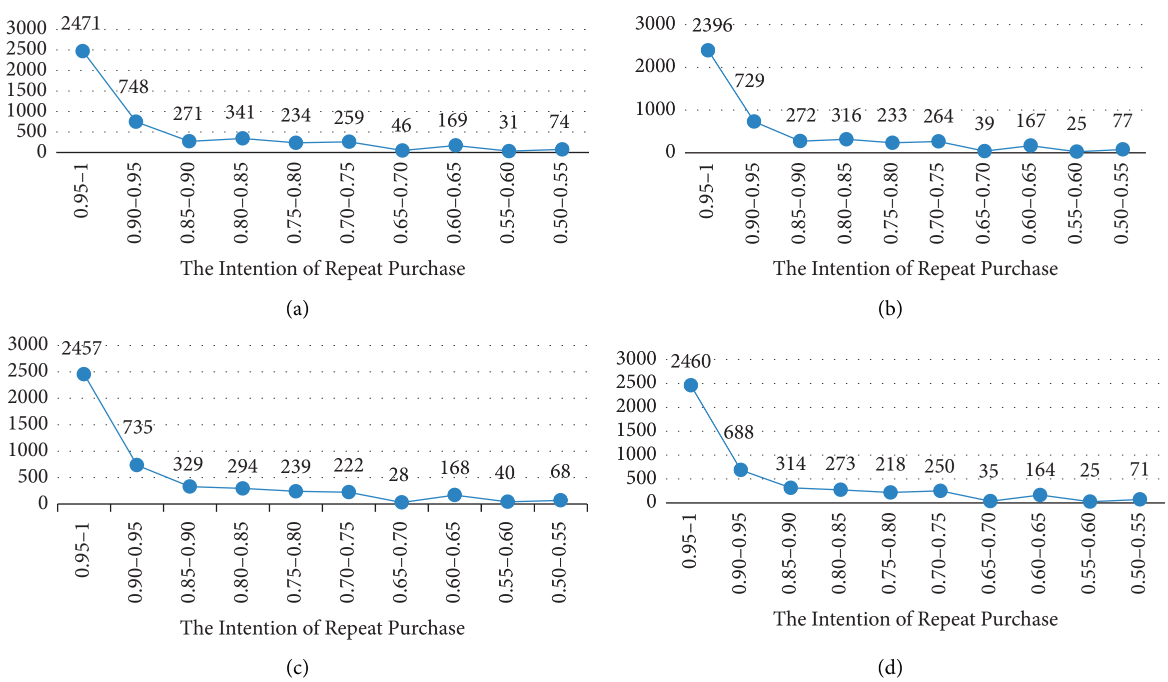

FIGURE 5: Numbers of air travelers purchased. (a) Numbers of customers purchased (A-A). (b) Numbers of customers purchased (A-B). (c) Numbers of customers purchased (B-A). (d) Numbers of customers purchased (B-B).

4.2. Results and Findings. The redemption period of promotional coupons is one week (from 9 August to 16 August) in “89 promotions” campaign. The dataset consists of 39,721 travelers who repurchase air tickets from 9 August to 16 August 2019 (Table 1).

First, we analyzed the number of travelers that purchased air tickets in each group (Figure 5). The number of travelers in each group (not only the treatment group but also the control group) decreased along with the repeat purchase intention.

Second, we analyzed the quantity of orders per group in the within-group and between-group comparisons (Figures 6 and 7). The quantities of orders in the different treatment groups (the A-A and A-B groups) decreased along with the decrease of purchase intention (Figure 6). However, the quantities of orders within purchase intention intervals did not differ much. Thus, even though the SMS promotions were sent at different times (A-A and B-A groups; Figure 7), the quantity of orders within purchase intention intervals did not differ significantly. However, the quantity of orders in the treatment group (A-A) was on average 35.8% greater than that in the control group (B-B). As the purchase intention increased, the quantity of orders differed more between the treatment and control groups.

Additionally, the number of air tickets sold also increased along with the purchase intentions. When the

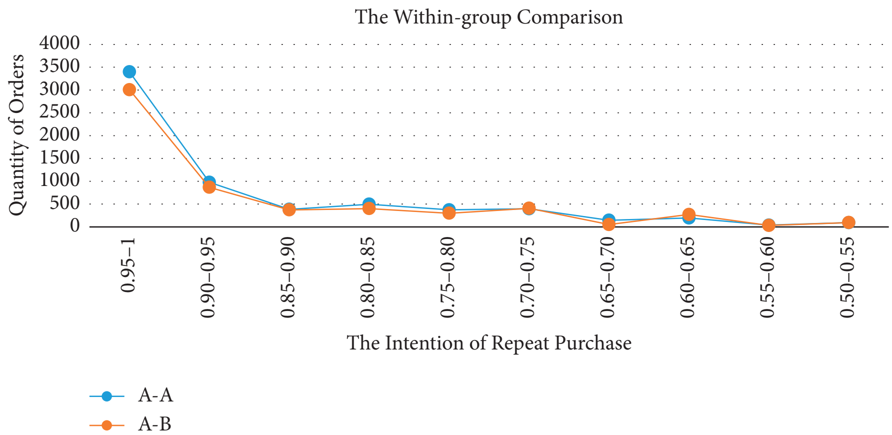

Figure 6: Comparison of A-A and A-B groups.

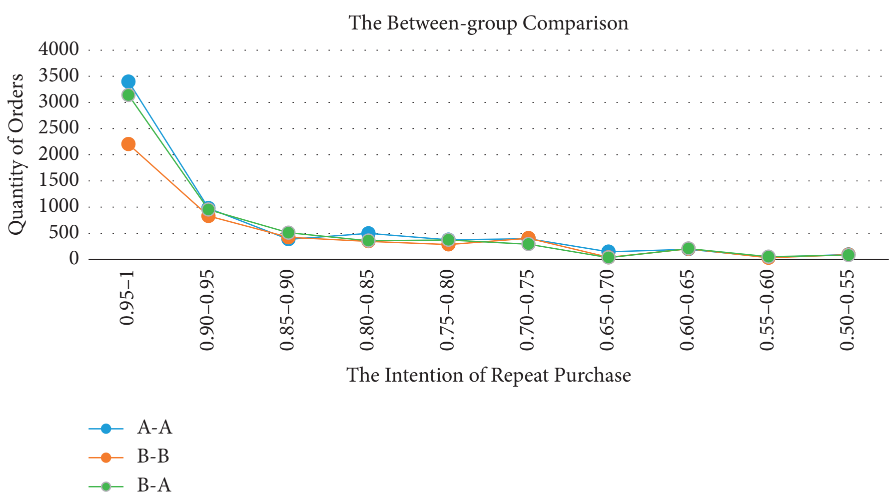

Figure 7: Comparison of A-A, B-A, and B-B groups.

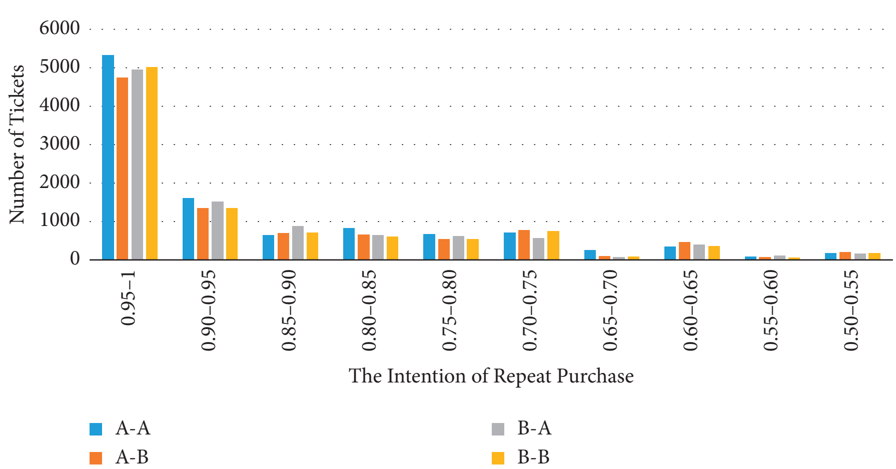

Figure 8: The number of tickets in four groups.

purchase intention was greater than 0.70, the difference in the quantity of air tickets bought between the experimental and control groups increased. On average, 100 more air tickets were purchased by the treatment group (A-A) than the control group (B-B; Figure 8).

5. Discussion and Conclusion

To sum up, this study focuses on how to effectively send promotional messages based on machine learning in airlines. More precisely, a predictions model of travelers' repurchase behaviors is proposed based on the random forest algorithm. In addition, a random field experiment was carried out to verify the validity of the proposed model framework. In other words, sending specific SMS promotional ads based on a traveler's intention to purchase can effectively increase traveler purchase behavior. Advertising allows businesses to introduce their products directly to potential travelers and measure their adjustments to purchase decisions on the products offered [2]. Therefore, how to effectively and accurately send mobile advertisements to potential air passengers to promote passengers' repurchase is an important and urgent problem for enterprises.

We used the RF algorithm to identify 28 important features that affect traveler repurchases in airlines. Then, the RF algorithm and four other machine learning algorithms, KNN, DT, SVM, and ET, were used to predict traveler purchase intentions. The results show that RF better predicted travelers' repeated purchase behaviors compared with the other algorithms. Finally, using the "89 promotion" campaign held by airlines, a random field experiment revealed that after determining the purchase intentions of travelers, sending SMS promotional ads to travelers with high purchase intentions effectively increased the likelihood of travelers' repeated purchases.

Through our model framework design and process, airlines can effectively use machine learning methods, especially RF methods, to mine historical purchase data and individual traveler characteristics. First, airlines can accurately identify potential traveler groups having high purchase intentions and the willingness to repurchase. Then, they can send targeted mobile advertising to promote repurchases. This will weaken traveler resistance to mobile advertising, thereby enhancing the airline's brand loyalty and revenues.

5.1. Theoretical Contribution. First, the current project joins a small but growing body of literature that examines the antecedents of traveler repurchases. A large amount of the previous literature focused on the impact of traveler consensus factors on traveler repurchases, such as perception [1, 55], attitude [27, 56], demographics [29], and trust [29, 31, 55]. Other research focused on nonconsensus factors, such as seat preselection behavior before [26], safety concerns [57], traveler satisfaction [34], and promotions and advertising [29, 58] from the perspective of the traveler. In our work, the antecedents of traveler repurchases were highly related to the travelers' previous purchase behaviors, such as the quantity purchased in the past 2 years, as well as the traveler' status and membership levels. We also analyzed the impact of previous purchasing behavior on traveler purchases' intentions, which is supplemental to the literature on the antecedents of traveler repurchases.

Second, our work makes several important contributions to the literature focused on the effects of marketing on traveler repurchases. Our research builds on individual consumption data recorded by an airline. We used five machine learning methods, KNN, DT, RF, SVM, and ET, to predict travelers' repeated purchase behaviors. There is much literature on predicting air traveler repeated purchase behaviors [8, 10, 14, 18–20], but literature based on using a machine learning model to make these predictions is still limited [8, 12, 15]. This is especially true for actual traveler sales date, because surveys or experimental data have often been used to study traveler repeat purchases [10, 18, 22, 26]. Besides, some studies have optimized other machine learning algorithms, such as artificial bee colony algorithm and deep neural network approach, to predict customer repurchase intention [12, 15]. However, these studies only use consumer data in specific environments to test their proposed optimization algorithms and do not study whether these algorithms can be well applied to the business environment. We not only propose an effective model framework to predict the traveler repurchasing using the random forest algorithm but also evaluate the validity of the model framework based on a field experiment in airlines.

Third, our work contributes to the literature on mobile advertising promotions, which enables the targeting of travelers by time or location using mobile technologies [6, 59–61]. However, people have more privacy concerns regarding location-tracking and location-based mobile marketing [2, 62, 63]. Therefore, push-based advertising (e.g., SMS) is popular with the marketing managers [64]. An effective method, in this work, is proposed to send profile-based push advertising. To be specific, it identifies travelers with high purchasing potentials using machine learning methods and then sends mobile advertisements to them. Furthermore, we used a combination of the machine learning model and experimental methods, which is a research method not generally used in this field. Our work offers a new perspective on studying the purchase intention of travelers and precision marketing.

5.2. Practical Contributions. Our work has significant managerial implications. Leveraging advertising for product promotion is an immensely popular marketing strategy [29]. Marketing managers in many fields rely on massive advertising campaigns and promotions to improve sales. Because advertising is becoming less effective, marketing managers should concentrate on selective promotions and services [8]. Our research shows that the random forest algorithm can be used to predict the purchase intention of travelers, identify those with high purchase intention, and then send promotional advertising targeted. This effectively improves the conversion rate of promotional information and the repurchase behavior of travelers. The framework model of our study can provide guidance to the precise marketing.

This work has guiding significance for the customer relationship management and maintenance. Machine learning model, especially RF model, is an effective and practical method to predict repeated purchase intentions of an air traveler. The results are very useful to organizations that are attempting to understand air traveler purchase intentions. Besides, it can effectively identify the high possibility of repurchase travelers. This is essential and fundamental for maintaining customer relationships. More precisely, a company can select the target air travelers on the basis of the FR forecast results. In the cut-throat competition scenario, predicting travelers' purchase intentions in advance, and then sending SMS promotions to travelers with high purchase intentions, can effectively increase revenue at little extra cost. In addition, the method that we proposed can increase traveler repurchases and increase corporate sales profit. Furthermore, it can reduce traveler aversion to mobile advertising. It is conducive to brand loyalty and traveler involvement with the airlines.

6. Limitations

Our findings discuss how to improve the repurchase behavior of an air traveler based on machine learning and field experiment methods. This study provides some guidance for customer relationship, marketing, and repurchase behavior of travelers in airlines, while there are still some limitations. First, the general 4 machine learning methods were compared with the framework model we proposed, and the result shows that the random forest algorithm has advantages in prediction accuracy and time in the field of airlines. However, there are many other machine learning algorithms (e.g., neural networks and genetic algorithms) that can be used to predict traveler repeat purchases. The scholars can study the promotion framework based on other algorithms models in the future. Furthermore, the dataset for the validation of our method was mainly collected from the aviation field. There are more areas (e.g., e-commerce and mobile commerce) in the marketing field in which traveler purchasing intentions need to be determined. Meanwhile, one-week response dataset of experimental subjects may be biased, and future research can collect data over a longer study period for the relevant research. Last but not least, our work sought to verify through randomized experiments that travelers with high purchase intentions will purchase again after receiving promotional SMSs. However, we did not verify the causality among them. Further research will focus on the causality between sending SMS promotional ads and the high likelihood of repurchases, or the effects of SMS contents on increasing the likelihood of repurchases.

Data Availability

The travelers' repurchase behavior data used to support the findings of this study have not been made available because these data belong to one of the airlines in China, and it is the commercial confidentiality.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the grants from the National Natural Science Foundation of China (#72034001 and #71974044), the Heilongjiang Provincial Natural Science Foundation of China (#YQ2020G004), and the Fundamental Research Funds for the Central Universities (HIT.HSS.DZ201905).

References

[1] C. M. Chiu, E. T. G. Wang, Y. H. Fang, and H. Y. Huang, "Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of utilitarian value, hedonic value and perceived risk," *Information Systems Journal*, vol. 24, no. 1, pp. 85–114, 2014.

[2] H. Kweon, D. Zhang, and L. Zhou, "User privacy in mobile advertising," in *Proceedings of the 18th Americas Conference On Information Systems 2012*, pp. 685–696, AMCIS, Washington, USA, August, 2012.

[3] S. H. Park, S. Y. Huh, W. Oh, and S. P. Han, "A social network-based inference model for validating customer profile data," *MIS Quarterly*, vol. 36, no. 4, pp. 1217–1237, 2012.

[4] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 1, pp. 3–13, 2006.

[5] Z. Fang, B. Gu, X. Luo, and Y. Xu, "Contemporaneous and delayed sales impact of location-based mobile promotions," *Information Systems Research*, vol. 26, no. 3, pp. 552–564, 2015.

[6] X. Luo, M. Andrews, Z. Fang, and C. W. Phang, "Mobile targeting," *Management Science*, vol. 60, no. 7, pp. 1738–1756, 2014.

[7] C. W. Phang, X. Luo, and Z. Fang, "Mobile time-based targeting: matching product-value appeal to time of day," *Journal of Management Information Systems*, vol. 36, no. 2, pp. 513–545, 2019.

[8] C. Lv, Y. S. Feng, and D. Y. Zhao, "Purchase prediction via machine learning in mobile commerce," *Natural Language Understanding and Intelligent Applications*, pp. 506–513, Springer, Berlin/Heidelberg, Germany, 2016.

[9] A. Ghose and S. P. Han, "An empirical analysis of user content generation and usage behavior on the mobile internet," *Management Science*, vol. 57, no. 9, pp. 1671–1691, 2011.

[10] J. Qiu, Z. Lin, and Y. Li, "Predicting customer purchase behavior in the e-commerce context," *Electronic Commerce Research*, vol. 15, no. 4, pp. 427–452, 2015.

[11] E. J. Kim, "Analysis of travel mode choice in seoul using an interpretable machine learning approach," *Journal of Advanced Transportation*, vol. 2021, Article ID 6685004, 13 pages, 2021.

[12] A. Kumar, G. Kabra, E. K. Mussada, M. K. Dash, and P. S. Rana, "Combined artificial bee Colony algorithm and machine learning techniques for prediction of online

consumer repurchase intention," *Neural Computing & Applications*, vol. 31, no. S2, pp. 877–890, 2017.

[13] X. Zheng, C. M. Liu, and P. Wei, "Air transportation direct share analysis and forecast," *Journal of Advanced Transportation*, vol. 2020, Article ID 8924095, 12 pages, 2020.

[14] A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, 2020.

[15] Z. Zhang, C. Ji, Y. Wang, and Y. Yang, "A customized deep neural network approach to investigate travel mode choice with interpretable utility information," *Journal of Advanced Transportation*, vol. 2020, Article ID 5364252, 11 pages, 2020.

[16] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 4, pp. 333–350, 2018.

[17] G. L. Grahn, "NBD model of repeat-purchase loyalty: an empirical investigation," *Journal of Marketing Research*, vol. 6, no. 1, pp. 72–78, 1969.

[18] H. Valecha, A. Varma, I. Khare, A. Sachdeva, and M. Goyal, "Prediction of consumer behaviour using random forest algorithm," in *Proceedings of the 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON*, pp. 1–6, IEEE, Gorakhpur, India, November, 2018.

[19] B. J. S. Mann and Rashmi, "Predicting repeat purchase behaviour: an Indian experience," *Asia Pacific Business Review*, vol. 6, no. 4, pp. 64–73, 2010.

[20] B. Baesens, S. Viaene, D. Van Den Poel, J. Vanthienen, and G. Dedene, "Bayesian neural network learning for repeat purchase modelling in direct marketing," *European Journal of Operational Research*, vol. 138, no. 1, pp. 191–211, 2002.

[21] N. Meade and T. Islam, "Using copulas to model repeat purchase behaviour–an exploratory analysis via a case study," *European Journal of Operational Research*, vol. 200, no. 3, pp. 908–917, 2010.

[22] J. Kim and M. Suk Kim, "Analysis of automobile repeat-purchase behaviour on crm," *Industrial Management & Data Systems*, vol. 114, no. 7, pp. 994–1006, 2014.

[23] Y. Choi, *A Structural Equation Model of the Determinants of Repeat purchase Behaviour of Online Grocery Shoppers in the uk*, Newcastle University, London, UK, 2013.

[24] D. Yi-sheng and D. Qiang, "Empirical analysis of customer repeat purchase forecast on O2O platform," *Economics And Management Innovations (ICEMI)*, vol. 1, no. 1, pp. 147–150, 2017.

[25] W. Huang and L. Li, "Predicting purchase proneness of anonymous user in mobile commerce," in *Proceedings of the Sixteenth Wuhan International Conference on E-Business*, pp. 366–375, Wuhan, China, May, 2017.

[26] R. C. Jou, Y. C. Chiu, and C. W. Kuo, "Low-cost carrier passengers' willingness to pay for the seat preselection service: a case study on the Taiwan-Japan route," *Journal of Advanced Transportation*, vol. 2021, Article ID 6699270, 14 pages, 2021.

[27] S. L. Chin and Y. N. Goh, "Consumer purchase intention toward online grocery shopping: view from Malaysia," *Global Business and Management Research: International Journal*, vol. 9, no. 4, pp. 221–239, 2017.

[28] S. Dlamini and N. B. Chinje, "The influence of loyalty programs on south african youth's repeat purchase behaviour," *Journal of Business & Retail Management Research*, vol. 13, no. 03, pp. 235–247, 2019.

[29] S. Pandey and D. S. Srivastava, "Antecedents of customer purchase intention," *IOSR Journal of Business and Management*, vol. 18, no. 10, pp. 55–82, 2016.

[30] S.M. F. Bukhari, A. Ghoneim, C. Dennis, B. Jamjoom, and B. Jamjoom, "The antecedents of travellers' e-satisfaction and intention to buy airline tickets online: a conceptual model," *Journal of Enterprise Information Management*, vol. 26, no. 6, pp. 624–641, 2013.

[31] Y. Fang, I. Qureshi, H. Sun, P. McCole, E. Ramsey, and K. H. Lim, "Trust, satisfaction, and online repurchase intention: the moderating role of perceived effectiveness of E-commerce institutional mechanisms," *MIS Quarterly*, vol. 38, no. 2, pp. 407–427, 2014.

[32] D. Bendall-Lyon and T. L. Powers, "The influence of mass communication and time on satisfaction and loyalty," *Journal of Services Marketing*, vol. 17, no. 6, pp. 589–608, 2003.

[33] J. Møller Jensen and T. Hansen, "An empirical examination of brand loyalty," *The Journal of Product and Brand Management*, vol. 15, no. 7, pp. 442–449, 2006.

[34] L. Meyer-Waarden and C. Benavent, "The impact of loyalty programmes on repeat purchase behaviour," *Journal of Marketing Management*, vol. 22, no. 1–2, pp. 61–88, 2006.

[35] A. Sharma and P. Bhardwaj, "Perceived benefits of loyalty programmes and their impact on purchase intentions of customers," *PRIMA: Practices and Research in Marketing*, vol. 6, no. 1and2, 2015.

[36] D. F. S. Zitzlsperger, T. Robbert, and S. Roth, "Forecasting customer buying behaviour 'controlling for seasonality," in *Proceedings of the ANZMAC Conference*, New Zealand, August, 2009.

[37] J. G. Cragg, "Some statistical models for limited dependent variables with application to the demand for durable goods," *Econometrica*, vol. 39, no. 5, pp. 829–844, 1971.

[38] A. T. Jahromi, S. Stakhovych, and M. Ewing, "Managing B2B customer churn, retention and protability," *Industrial Marketing Management*, vol. 43, pp. 1258–1268, 2014.

[39] A. Tizghadam, H. M. Khazaei, M. H. Y. Moghaddam, and Y. Hassan, "Machine learning in transportation," *Journal of Advanced Transportation*, vol. 2019, Article ID 4359785, 3 pages, 2019.

[40] N. Levin and J. Zahavi, "Continuous predictive modeling: a comparative analysis," *Journal of Interactive Marketing*, vol. 12, no. 2, pp. 5–22, 1998.

[41] P. J. Danaher, M. S. Smith, K. Ranasinghe, and T. S. Danaher, "Where, when, and how long: factors that influence the redemption of mobile phone coupons," *Journal of Marketing Research*, vol. 52, no. 5, pp. 710–725, 2015.

[42] N. M. Fong, Z. Fang, and X. Luo, "Geo-conquesting: competitive locational targeting of mobile promotions," *Journal of Marketing Research*, vol. 52, no. 5, pp. 726–735, 2015.

[43] D. Molitor, P. Reichhart, and M. Spann, "Location-based advertising: measuring the impact of context-specific factors on consumers' choice behavior," *SSRN*, Article ID 2116359, 2012.

[44] Y. Bart, A. T. Stephen, and M. Sarvary, "Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions," *Journal of Marketing Research*, vol. 51, no. 3, pp. 270–285, 2014.

[45] L. Xu, J. A. Duan, Y. J. Hu, Y. Cheng, and Y. Zhu, "Forward-looking behavior in mobile data consumption and targeted promotion design: a dynamic structural model," *Information Systems Research*, vol. 30, no. 2, pp. 616–635, 2019.

[46] M. Andrews, X. Luo, Z. Fang, and A. Ghose, "Mobile ad effectiveness: hyper-contextual targeting with crowdedness," *Marketing Science*, vol. 35, no. 2, pp. 218–233, 2016.
[47] T. H. Baek and M. Morimoto, "Stay away from me," *Journal of Advertising*, vol. 41, no. 1, pp. 59–76, 2012.
[48] Z. W. Chen, L. Z. Fang, L. Y. Chen, and H. L. Dai, "Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: a randomized controlled trial," *Journal of Zhejiang University - Science B*, vol. 9, no. 1, pp. 34–38, 2008.
[49] A. Mohammed, P. R. Acheampong, E. Otupiri, F. A. Osei, R. Larson-Reindorf, and E. Owusu-Dabo, "Mobile phone short message service (sms) as a malaria control tool: a quasi-experimental study," *BMC Public Health*, vol. 19, no. 1, pp. 1193–1212, 2019.
[50] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
[51] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling*, pp. 55–85, Springer, Boston, MA, 1998.
[52] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
[53] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[54] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *Statistics in Medicine*, vol. 22, no. 9, pp. 1365–1381, 2003.
[55] H. Ahasanul, K. Ali, and R. Sabbir, "Factors influencing buying behavior of piracy products and its impact to Malaysian market," *International Review of Business Research Papers*, vol. 5, no. 2, pp. 383–401, 2006.
[56] U. M. Dholakia and R. P. Bagozzi, "Mustering motivation to enact decisions: how decision process characteristics influence goal realization," *Journal of Behavioral Decision Making*, vol. 15, no. 3, pp. 167–188, 2002.
[57] A. M. Angulo, J. M. Gil, and L. Tamburo, "Food safety and consumers' willingness to pay for labelled beef in Spain," *Journal of Food Products Marketing*, vol. 11, no. 3, pp. 89–105, 2005.
[58] C. H. Chen and Y. W. Chang, "The impact of brand associations on brand equity," *Journal of Management*, vol. 18, no. 1, pp. 3–14, 2001.
[59] T. T. Lin, F. Paragas, and J. R. Bautista, "Determinants of mobile consumers' perceived value of location-based advertising and user responses," *International Journal of Mobile Communications*, vol. 14, no. 2, pp. 99–117, 2016.
[60] C. Bauer and C. Strauss, "Location-based advertising on mobile devices," *Management Review Quarterly*, vol. 66, no. 3, pp. 159–194, 2016.
[61] D. Molitor, M. Spann, A. Ghose, and P. Reichhart, "Measuring the effectiveness of location-based advertising: a randomized field experiment," *SSRN*, Article ID 2645281, 2020.
[62] O. Rafieian and H. Yoganarasimhan, "Targeting and privacy in mobile advertising," *Marketing Science*, vol. 40, no. 2, pp. 193–218, 2021.
[63] W. Shin and T. T. C. Lin, "Who avoids location-based advertising and why? Investigating the relationship between user perceptions and advertising avoidance," *Computers in Human Behavior*, vol. 63, pp. 444–452, 2016.
[64] L. H. Shan, T. A. Chin, Z. Sulaiman, and F. M. Muharam, "Effective mobile advertising on mobile devices," *Journal of Global Business and Social Entrepreneurship (GBSE)*, vol. 2, no. 4, pp. 164–177, 2016.