Research Article

Prediction of Purchase Intention of High and Middle Potential Users in Luxury Hotels Based on Data Mining

Xue Wu

School of Economics and Management, Kaili University, Kaili, Guizhou, China

Correspondence should be addressed to Xue Wu; wuxuesisi@xs.ustb.edu.cn

Received 3 August 2022; Revised 7 September 2022; Accepted 15 September 2022; Published 11 October 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Xue Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the term “big data” has attracted the attention of many scholars and business managers, and the emergence of massive data has ushered in a major transformation of the times. Service quality and customer repeat purchase intention are two of the hot issues in the field of service research. Clarifying the mechanism of the two can help enterprises establish long-term customer relations with more customers through the improvement of service quality, enhance their competitive strength, and improve enterprise performance. According to the actual consumer purchase data provided, the existing data sets are visually analyzed and processed to find the consumer purchase rules, build features and build a reasonable consumer purchase prediction experimental data set. Based on the obtained experimental data set, a single prediction model of consumer purchase is designed. According to the analysis of the instantiation results of the single prediction model, the single model consumed in the fusion model of consumer purchase prediction is determined. In order to demonstrate the viewpoints put forward, the research model is established, the research hypothesis is put forward, the questionnaire is designed by referring to the literature and field investigation, and the results of the questionnaire are empirically analyzed. Using the methods of descriptive statistical analysis, factor analysis, reliability and validity test, and regression analysis, the hypothesis that service quality has a direct impact on customer repeat purchase intention is verified.

1. Introduction

With the arrival of the era of big data, information mining of massive data has become the focus of research in the field of information technology. Big data are gradually changing people’s way of life, work, and thinking, leading people into a new era of resource integration [1]. If hotels want to survive and develop, they must gain a competitive advantage [2]. Big data have the characteristics of huge data volume, high processing speed, diversity, high value, and low density. Traditional data processing software has been unable to process such huge data, so the emergence of cloud storage and cloud computing servers provides technical support for the storage, analysis, sharing, and transmission of big data [3]. The overall competitiveness of my country’s hotel industry is relatively low. Mr. Feffer, chairman of the International Hotel and Restaurant Association, once pointed out that compared with the same industry in foreign countries, the Chinese hotel industry has weak self-promotion awareness, insufficient investment return ability, and high technology content. There are many gaps, such as the low degree of collectivization, the lack of market value of services, etc. Therefore, my country’s hotel industry urgently needs to improve overall for hotels to establish and develop long-term relationships. Establishing a long-term relationship with customers can be explained as follows: high service quality can be consumed to make customers feel high and satisfied, while high satisfaction can encourage customers to repeat consumption, and they will not buy competitors’ products, and their sensitivity to price wars will be reduced [4].

In developed countries, the proportion of service industry output value in GDP increased sharply from 53% in 1970 to 66% in 2005. It is 47% and 68% in the European Union and 57% and 72% in the United States, respectively, and this trend continues. It can be said that marketing has
entered the era of service marketing [5]. From the perspective of theory and practice, the theories related to customer service quality and repeat purchase intention are deeply explored to make up for the lack of hotel service quality management concepts in my country [6]. With the continuous progress and development of society, customer perception is also in dynamic change. However, the traditional hotel service quality management idea in China often makes it looks at the service quality statically. Once effective service standardization measures are formulated, it is considered that it can be done once and for all, resulting in the service quality cannot meet the needs of customers in time. The business activities between consumers and enterprises are conducted directly through the Internet. The e-commerce consumer group under this mode is large and there is basically no one-time consumption on the platform [7]. Users with a high potential consumer group prediction probability of 0.4–0.8 are regarded as high potential consumers, that is, consumers in this range will not place orders directly but have high purchase demand; users with a prediction probability higher than 0.8 are regarded as consumers, who have a high probability of placing orders directly; users with a prediction probability of less than 0.4 are regarded as consumers who basically do not need such goods [8]. In the past, the analysis of customer’s product purchase behavior basically consumed the form of a questionnaire to investigate customer’s behavior intention, but the accuracy was relatively low and the workload was heavy [9]. According to customer product purchase data recorded by banks, using data mining as a tool to predict customer demand for financial products has received increasing attention, but related research is still relatively rare [10]. This article takes the active of luxury hotels as the research object, and consumers data mining and machine learning technology to predict the purchase intention of high potential and achieve precise marketing. High potential is put forward by luxury hotels to distinguish consumer groups with different prediction probabilities. Users with a prediction probability of 0.4–0.8 are regarded as having high potential, that is, in this range will not place orders directly but have high purchase demand; users with a prediction probability higher than 0.8 are regarded as those who have a high probability of placing orders directly; and users with a prediction probability of less than 0.4 are regarded as those who basically do not need such goods. The ultimate goal is to achieve precise marketing for high potential, but it actually predicts the purchase intentions of all.

Before this research, there were the following marketing communication methods between the platform and:

1. When the store or platform has promotional activities, can receive coupons provided by the merchants or platforms.
2. Recommend products with lower prices or related products according to the products added to the shopping cart by the consumer.
3. Recommend related products based on the content searched. The abovementioned are marketing models based on consumer search or recommendation algorithms. Under the current background, enterprises can be in a favorable position in the big data economy only by using big data to make better, fact-based, and experience-based decisions.

2. Related Work

TSCA analyzes the relationship between the factors influencing group purchasing and purchasing behavior. The results of feature analysis show that there are more likely to buy goods with more comments and less restrictions on consumer, which gives guiding suggestions to the business model of merchants [11]. Gang and Chenglin analyzed the behavioral characteristics of those who purchased books online and proposed a collaborative recommendation algorithm to find other with similar behaviors as the consumer who placed the order, so as to achieve the purpose of recommendation [12]. Taking the courses recommended in the online learning environment as the research object, Yang consumers item find the item sets related to the content, and then apply the item sets to the sequential pattern. The two methods are combined to recommend potentially consumerful and the consumer feedback effect is good [13]. Bilal et al. believe that on the web platform, the consumer’s actions correspond to page jumps and other responses, so the consumer’s behavior information can be obtained by analyzing the weblogs. Therefore, taking Keto platform as an example, this paper carries out data mining on the web logs and puts forward a personalized recommendation algorithm, which increases the consumer’s viscosity of the platform [14]. Hou et al. consumers RBF (radial basis) neural network to conduct empirical research on mobile communication companies and proposes a precise marketing strategy. Firstly, standardize the data, select the key elements of the subdivision through factor analysis, then identify the basic center through the nearest neighbor clustering algorithm, adjust the center using K-means clustering method, identify the RBF neural network center, and finally, establish customer segmentation model [15]. From the point of view of algorithm optimization, Tian and Youngsook proposed an improved decision tree algorithm. The original algorithm can only achieve local optimization to the current global optimization and verify the effectiveness of this method for consumer purchase behavior prediction with the help of the Teradata platform [16]. Zhao et al. proposed to combine logistic regression and support vector machine to predict online purchase behavior. The results show that the combined model has a better prediction effect than the single model [17]. Chan et al. consumed a fusion algorithm based on SMOTE and random forest to predict repeated purchase intentions and obtained high accuracy and efficiency. The fusion algorithm has certain reference significance for the prediction of repeated purchase intentions of new consumers [18]. Liu and Bai studied customer-perceived future sales of enterprises. The research results found that several dimensions of perceived value, namely social value, functional value, emotional value, and procedural value, have a
positive impact on purchase intention. At the same time, past sales have a good predictive effect on future sales [19]. Zhao et al. and other customers’ perceived value are sadly divided into two categories, namely, hedonic value and functional value. They found that the perceived value of customers acts as a medium between the mall image and consumers’ repurchase intention, and the mall image affects the perceived value of customers, thus stimulating customers’ reconsumption [20].

Through the analysis of domestic research status, it is found that the research on consumer data mining in the field of e-commerce is still in the exploratory stage. There are still very few researches to help to shop according to the prediction probability. Domestic hotels, high-end places, and other platforms generate a large amount of consumer data every day. The research on the analysis and processing of consumer behavior data is still immature. Feature engineering is necessary work to extract valuable information from big data. At the same time, it is also necessary to strengthen the application research of machine learning algorithms.

3. Methodology

3.1. Characteristic Engineering. Engineering refers to the customers' reconsumption [20]. consumers' repurchase intention, and the mall image affects functional value. “_hey found that the perceived value of divided into two categories, namely, hedonic value and Zhao et al. and other customers’ perceived value are sadly past sales have a good predictive effect on future sales [19].

The number of online comments is the sum of reviews, the more it can influence consumers’ purchasing decisions. High-quality online reviews can better influence consumers’ purchasing decisions. According to the theory of conformity, it is reasonable to believe that the greater the number of online reviews, the more it can influence consumers’ purchasing decisions. The number of online comments is the sum of consumer comments on a product or service. It is generally believed that the more comments about a product or service, the more attention the product or comment receives. It refers to the general tendency of comments on a product or service. If there are more positive comments than negative comments in all comments, the comments are generally positive.
proportion of the number of people corresponding to each age group. It can be seen from the figure that the age of is concentrated in 26–35 years old, accounting for 56.23% of the total ordering, followed by 19–25 and 36–45, so consumer age can be a consumer full feature as shown in picture 2.

3.2. Classification Prediction Model Based on LR. The prediction function has the characteristics of high speed, simplicity, and strong generalization ability for new data. It is a linear binary classification model that maps the results of the linear function to the s-type function (sigmoid function). It is widely consumed in the problem of predicting whether the consumer will buy the product with known consumer behavior. The prediction function of the algorithm is shown in the formula.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}.$$  \hspace{1cm} (1)

In the formula, the value range of $h_\theta(X)$ is between −1 and 1, indicating the probability that the result value is 1. In the article, it means the predicted consumer purchase probability. Greater than or equal to 0.5 means purchase, less than 0.5 means no purchase, and $1 - h_\theta(X)$ means the probability that the result value is 0; $E$ is the regression parameter, which aims to obtain a set of appropriate $\theta$ series. The sample generation probability is shown in the formula.

$$(1 - h_\theta(X))^{1-y}.$$  \hspace{1cm} (2)

Maximum likelihood estimation is consumed to estimate the parameters of LR, and the likelihood function of samples is shown in formula (3), which means that $m$ training samples are consumed to estimate $\theta$.

$$L(\Theta) = \prod_{i=1}^{m} y_i h_\theta(X_i)^{y_i} (1-h_\theta(X_i))^{1-y_i}.$$  \hspace{1cm} (3)

In the formula, $m$ represents the number of samples, and $L(\Theta)$ represents the probability of $m$ samples occurring at the same time, which is converted into a log-likelihood function as shown in the formula.

$$1(\theta) = \frac{1}{m} \log L(\theta) = \sum_{i=1}^{m} y_i \log h_\theta(X_i) + (1-y_i) \log (1-h_\theta(X_i)).$$  \hspace{1cm} (4)

The maximum likelihood estimation consumers the gradient rise method to find $(\Theta)$ at maximum $\Theta$, some studies have proposed the loss function $f(\theta) = -1/m1(\theta)$, which consumers the random gradient descent method to solve the minimum value. The iterative update process of $E$ is shown in formula (5), where $a$ represents the step size.

$$\theta_j = \theta_j + a(y_i - h_\theta(X_i)) X_i.$$  \hspace{1cm} (5)

Weight COL is consumed to specify the weight column in the training set feature table. The function of this parameter is to balance the proportion of positive and negative samples. The calculation method of positive and negative sample weights of this weight column is shown in the formula.


\[
\text{weight}_t = \frac{N_s}{C \cdot N_t}, \quad \text{weight}_f = \frac{N_s}{C \cdot N_f}. \quad (6)
\]

Weight is the weight of positive samples, \(N_s\) is the number of samples; \(C\) is the number of categories, the value of \(C\) in this study is 2; \(N_t\) is the number of positive samples. Correspondingly, weight is the weight of negative samples, and \(N_f\) is the number of negative samples. The model normalizes the training features. Assuming that the proportion of class \(K\) corresponding to the current sample set recorded as \(D\) is expressed as \(pK\) \((k=1, 2, ..., Y)\), the Gini value can be expressed as publicity.

\[
\text{Gini}(D) = \sum_{K=1}^{Y} \sum_{K' \neq K} pK^2 = 1 - \sum pK^2 K = 1|Y|. \quad (7)
\]

The Gini index calculation formula of feature \(a\) is expressed as a formula.

\[
\text{Gini}_{a} = \sum \frac{|D'|}{|D|} = 1V\text{Gini}(D'). \quad (8)
\]

The model performance measures commonly consumed in regression problems mainly include mean absolute error, mean variance/mean square error, mean square root difference and mean absolute percentage error.

(1) Mean absolute error (MAE)

\[
\text{MAE} = \frac{1}{n} \sum |f_i - y_i| = 1n. \quad (9)
\]

(2) Mean square error (MSE)

\[
\text{MAPE} = \frac{100}{n} \sum (f_i - y_i)^2 = 1n. \quad (10)
\]

(3) Root mean square error (RMSE)

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum (f_i - y_i)^2} = 1n^2. \quad (11)
\]

(4) Mean absolute percentage error (MAPE)

\[
\text{MAPRE} = \frac{100}{n} \sum \frac{|y_i - f_i|}{y_i} = 1n \quad (12)
\]

RF consumers then construct a decision tree. When the node searches for features to split, it randomly selects some features to find the optimal solution and applies it to the node to split, that is, it randomly samples the samples and features to avoid overfitting. If it is predicted by the classification algorithm, the final category of prediction is the category or one of the categories with the largest number of votes cast by \(T\) basic learners; if it is predicted by a regression
algorithm, the final output of the model is the arithmetic average of the regression results of $T$ basis learners. The construction process of RF that classifies problems through the above process is shown in Figure 3.

It can be seen from the above regression model evaluation metrics that it is the square operation that strengthens the effect of large numerical errors in the metrics, which greatly improves the sensitivity of the indicators. Because RF algorithm randomly selects decision tree nodes for feature partition, it can still train the model efficiently even in the case of high feature dimensions. Combined with the results of data set instantiation, the F evaluation model calculated according to the confusion matrix is selected.

The value perceived by customers is not only the cost or benefit but also the result of comprehensively weighing the benefit and cost. There are also many studies using this trade-off indicator. Brand preference is more an expression of customers' attitudes and tendencies. This attitude goes beyond customer satisfaction. It is not only customers' emotional response but also expresses their emotional love for the brand through this preferred choice. Starting from the goal of predicting purchase, it is necessary to know at least three aspects of information: consumer_id information, item_id, and consumer item interaction information, which are taken as three basic feature groups. Only the features constructed from the data of commodities have low discrimination to commodities, so the features of item_category are counted again according to the commodity category information to which the commodities belong. The commodities in the commodity category can meet the same type of purchase demand of consumers, so the information of the commodity category can supplement the commodity information. Moreover, due to a large amount of data on commodity categories, commodity category features have a high degree of discrimination for commodities of different commodity categories. In the following: the performance of commodity characteristics and commodity category characteristics is discussed from six aspects: data sparsity, differentiation of different categories of commodities, differentiation of similar categories of commodities, behavior conversion rate, hot selling trend, and support for calculating consumer similarity.

4. Result, Analysis, and Discussion

The experimental data comes from the real desensitization data provided by the topic of JDATA algorithm competition, the JDATA algorithm is the e-commerce platform of JD.com. While maintaining rapid development, it has accumulated hundreds of millions of loyal and accumulated massive amounts of real data. How to find rules from historical data, predict future purchase needs, and let the most suitable goods meet the people, who need them most are the key issues in the application of big data in precision marketing, and also the core technology required by all
e-commerce platforms for intelligent upgrading. This competition takes a specific problem in the precise recommendation as an example, hoping to find top talents in the field of data mining and run on the top of the wave with us. There are two data sets in A/B list, which are composed of browsing, placing orders, commenting, basic information of commodities, and basic information of within one year. Combined with the forecast target, it is divided into two stages: predicting whether a consumer will buy a specified commodity in the next month and predicting the first purchase date of a consumer who will buy a specified commodity. Becausa consumer the original data cannot meet the two stages of the research, the original data set is further analyzed and processed. Combined with the statistical results of 'browsing on their browsing, the distance between the A and B lists in Figure 4 predictsthetrendofthenumber ofwhobrowsedthespecifiedproductdateforthelasttimein thefirstmonthofthemonth. It can be seen that the browsing patterns in the A and B lists are overall consistent, and the number of browsing the specified item from the beginning of the month to the end of the month is on the rise. A total of 64,097 browse the designated products in the A list. This represents 64.97% of the total number of people. Similarly, there are 55,064 in the B list who browse the designated products. This represents 55.66% of the total number of people, as shown in Figure 4.

The specific purchase date of the last purchase of the specified product in the A/B list in Figure 4 corresponds to the changing trend of the consumer’s cents. The results show that the changing trend of the browsing data in Figure 5 is generally consistent with the overall upward trend, but the change is larger at the end of the month, taking into account the fact that the season has an impact on the consumer’s purchase decision.

This study screened 45 features (212 dimensions after one hot coding) from the existing consumer portrait feature table, such as the consumer’s marital status, whether there are children, the child’s age, whether there is a car, the last month’s single piece, clothing/personal makeup/grade, consumer loyalty, and promotion sensitivity. The training prediction results are shown in Table 2 (data from 2019). It can be seen from the table that the accuracy rate, recall rate, and F1 value are slightly improved after adding consumer features. When the Vitter sign is reduced from 646 to 40, it has little effect on the prediction effect of the model; the new consumer portrait feature is helpful to improve the prediction accuracy of the model as shown in Table 3.

It can be seen from the above that the prediction effect of the model under different categories is different. Through the analysis of Table 3, it can be seen that the model is only suitable for casual shoe category data, and the prediction effect for other categories needs to be improved. However, the consumer’s browsing and ordering behavior data in normal months showed a gradual trend of change, and there was no sudden increase in certain behavior data. This article simply eliminated such data as shown in Figure 6 and 7.

According to the abovementioned data preliminary analysis results, and combined with experience and business understanding, the data set is reconstructed. To predict whether the consumer will buy the specified commodity in the next month and the first purchase date of the consumer who will buy the specified commodity, two prediction labels are reconstructed, which are whether the consumer will buy in the current month (the purchase is recorded as 1, and the nonpurchase is recorded as 0), that is, the first _ buy _ or _ not; The consumer’s first purchase date in the current month, that is, first_buy_day. This experiment builds 254 new features and two labels. Mainly complete feature
Table 3: Comparison of model prediction effects before and after adding consumer portrait features.

<table>
<thead>
<tr>
<th>Characteristic dimension</th>
<th>Training time</th>
<th>Test set time</th>
<th>Accuracy rate</th>
<th>Recall rate</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>646 dimensions</td>
<td>11.30–11.50</td>
<td>11.50–11.55</td>
<td>0.249855</td>
<td>0.224757</td>
<td>0.26946</td>
</tr>
<tr>
<td>858 dimensions</td>
<td>11.30–11.50</td>
<td>11.50–11.55</td>
<td>0.245566</td>
<td>0.223445</td>
<td>0.23465</td>
</tr>
<tr>
<td>40 dimensions</td>
<td>11.30–11.50</td>
<td>11.50–11.55</td>
<td>0.645851</td>
<td>0.248766</td>
<td>0.28418</td>
</tr>
<tr>
<td>212 dimensions</td>
<td>11.30–11.50</td>
<td>11.50–11.55</td>
<td>0.285455</td>
<td>0.248551</td>
<td>0.127433</td>
</tr>
</tbody>
</table>

Figure 5: Trend chart of the number in the A/B list who purchased the product for the last time with the date.

Figure 6: User browsing statistics for promotional and nonpromotional dates.
construction from basic features, consumer behavior features, and time features: (1) basic features: it mainly extracts consumer level, consumer age, and consumer area information, and extracts commodity parameter information, commodity price information, and commodity area information for commodities as basic features. (2) User behavior characteristics: user behavior characteristics: user behaviors mainly include browsing, placing orders, and commenting. (3) Time features: extract the last purchase date, the average browsing date, the last browsing date, the average purchase date, the first purchase date in the first month, the average value of the rating date, etc., feature.

This paper consumers’ theoretical knowledge and mathematical models to analyze a large number of consumer data, predicts the consumer’s purchase behavior with the help of a machine learning algorithm, finds out the rules from the historical data and predicts the consumer’s future purchase demand. The expected effect has been preliminarily achieved, and the model has been applied to the shopping platform. The model is optimized from two aspects: feature and model. After adding consumer portrait features and model fusion, the prediction accuracy, recall rate, and F1 of the model are improved by about 0.03 compared with those before optimization, that is, the current prediction F1 is stable at about 0.28. Although the article does not involve the construction and selection of product features, based on the research results of the article, hotels can further analyze the consumer’s behavioral characteristics of products according to the specific meaning behind the product category number. For example, according to the prediction results of the information such as reception, frequency, and monetary in the RFM model, for a class who have a high probability of purchasing goods, they can market to those who have not purchased during the purchase cycle, but not the who have purchased.

5. Conclusions

Through the empirical study of hotels, the article found that the income level of customers who frequented hotels was mostly abovementioned the middle level. Becoconsumer the quality of food and beverage was closely related to people’s health, the price sensitivity was not high compared with other services, but they paid more attention to the trade-off between income and cost. Relying on this data to discover “purchasing patterns and choose a reasonable prediction method to solve the problem of consumer purchase prediction, not only improves the consumer stickiness of the platform and intelligently guides merchants” inventory but also brings positive effects to today’s personalized marketing. The empirical research of the hotel’s level of ideal quality and desire for quality, customers will feel that the services provided by the hotel are worth the money. They will be surprised by the unexpected harvest, so they are very satisfied with the services provided by the hotel, praise the services of the hotel, and gradually form trust in the hotel and continue to buy repeatedly. Higher scientific and technological content will strengthen the tangible dimension of service quality, thus improving the service quality as a whole. Just as the production of tangible goods cannot become a world-class brand without top-class technology, hotels with the highest technology can certainly be invincible. When processing consumer behavior data, in the process of building an online purchase prediction model, the advantage of the logistic regression algorithm is that it is more accurate than linear regression prediction and runs faster than other algorithms such as support vector machines. And through the clustering of, the impact on the prediction accuracy is reduced to a certain extent. Therefore, if future research can combine...
the data driving of machine learning with the theoretical driving of empirical research, researchers will not only be proficient in algorithms, but also go deep into the operation of high-end hotels, understand the relevant business background, and combine algorithms with business, so as to make more practical research.

Data Availability

The data used to support the findings of this study can be obtained from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by Fund Project 1: 2020 Carey College “Outstanding School-running Characteristics” Special Project (Engineering and Practical Project): Research on Value Cognition and Purchase Behavior of Camellia Oil in Southeast Guizhou — Based on Consumer Survey (project no.: 2020gkzs02), Fund Project 2: 2020 Carey College Special Project “Response, Governance and Impact of New Coronary Pneumonia Epidemic and Other Public Health Events”: Empirical Investigation and Research on the Impact of New Coronary Pneumonia on Key Industries in Guizhou (project no.: YQZX201902), and Fund Project 3: 2019 Special Project of “Doctoral Professor Service Group” of Carey College: Investigation and Research on Camellia Oil Industry in Qiandongnan Prefecture (project no.: BJFWT201902).

References