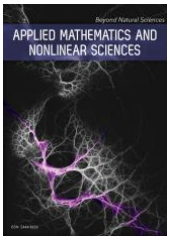
sciendo Applied Mathematics and Nonlinear Sciences, 10(1) (2025) 1-25



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on Precision Marketing and Smart Tourism Service Optimization of Online Marketing Driven E-commerce Platform Based on Big Data and Machine Learning

Aifang Zhang[1,†], Lingling Zhang[2]

1. Institute of Tourism Management and Rural Revitalization, Luoyang Vocational College of Cultural Tourism, Luoyang, Henan, 471000, China.
2. Institute of Marxism, School of Philosophy, Law & Politics, Shanghai Normal University, Shanghai, 200234, China.

Submission Info

Communicated by Z. Sabir
Received November 8, 2024
Accepted February 11, 2025
Available online March 19, 2025

Abstract

E-commerce platform occupies an important position in the modern economy, how to improve the precision marketing effect of the platform through online marketing has become the key to the development of enterprises. At the same time, the development of smart tourism services also puts forward higher requirements for personalized recommendations and precise marketing. The widespread application of big data technology and machine learning methods provides new opportunities, making it possible to optimize platform marketing and services through data-driven strategies. In this paper, we collect and process data from e-commerce users to construct an e-commerce user profile model. The model is used to accurately categorize users, and NSE precision marketing strategies are developed and implemented. Decision trees, random forests, support vector machines, and LightGBM algorithms are used to predict users' purchasing behavior and interest preferences. Meanwhile, the smart tourism service model was used to generate a list of Top-K attractions as a recommendation list for users to provide personalized tourism services. The empirical analysis results show that the online marketing strategy based on these techniques can effectively improve the user conversion rate and increase the overall revenue of the platform, and the number of orders on the e-commerce platform after the use of the platform increased from 138 to 245, which is an increase of 77.62%. Furthermore, the level of personalization of smart tourism services has been significantly enhanced. After structural equation analysis, it can be seen that the standardized coefficients of the influence of smart tourism marketing experience on behavioral intention and perceived value are 0.136 and 0.193 respectively while the P-value is less than 0.05, which indicates that the smart tourism marketing experience has a positive influence on tourists' behavioral intention and perceived value. The study shows that the combination of big data and machine learning provides a strong technical support for the optimization of e-commerce platforms and tourism services, and can help enterprises to achieve the precise marketing objectives and the enhancement of user service experience in the changing market environment.

Keywords: E-commerce user profiling; NSE precision marketing strategy; LightGBM algorithm; Smart tourism service.
AMS 2010 codes: 68T09

†Corresponding author.
Email address: Zhang198134@126.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2025-0426

sciendo OPEN ACCESS © 2025 Aifang Zhang and Lingling Zhang, published by Sciendo.

CC BY This work is licensed under the Creative Commons Attribution alone 4.0 License.

1 Introduction

In the new period, information technology has been comprehensively promoted and applied, especially the emergence and application of advanced technologies such as big data and artificial intelligence, which provide technical support for the innovative development of various industries in China. The tourism industry has likewise combined various advanced technologies and equipment with the traditional tourism mode to realize the intelligent development of tourism. Intelligent tourism fully combines the cultural industry and the tourism industry, making changes in the traditional tourism form [1]. A variety of new technologies and new ideas have emerged in the tourism industry, which not only bring into play the value of advanced technology in the tourism industry, but also enable tourists to get a more favorable experience in tourism and improve the overall development of the tourism industry [2-5].

As an emerging business model, tourism e-commerce has attracted the attention of the industry in earlier years. The information-intensive and information-dependent nature of the tourism industry and e-commerce have a natural adaptability, and tourism e-commerce after the combination of the two has had significant development in recent years [6-9]. E-commerce can help cultural and tourism enterprises to sell their products through online platforms and increase their sales performance [10-11]. Therefore, in the development process of cultural tourism enterprises, e-commerce should be comprehensively developed in combination with the development requirements of the times to rapidly accomplish the strategic objectives [12-13]. Among them, e-commerce precision marketing strategy is a marketing way for cultural tourism enterprises to realize measurable and low-cost of enterprises based on precise positioning [14-15]. Relying on the e-commerce precision marketing strategy of big data and machine learning, it can analyze different users in the target market in detail, and then realize the establishment of personalized marketing communication to different consumer groups in the segmented market according to different consumer psychology and consumer behavior characteristics [16-19]. In recent years, along with the development of today's high technology, this kind of precise, measurable, high investment and low-cost marketing is accepted by more and more enterprises [20-21].

This paper mines and collects e-commerce user data by deep combing user information from e-commerce platforms and constructing an ID mapping system. Based on the 6 major attributes of e-commerce users, such as behavioral attributes, social attributes, interest attributes, etc., user portrait features are constructed. Use the improved K-means algorithm to divide user categories, implement user population classification under precision marketing, derive the mathematical expression of NSE precision marketing strategy, and formulate the precision marketing strategy of e-commerce enterprises. Combining the support vector machine, decision tree, random Senri, and LightGBM algorithms, the user purchase behavior prediction model is constructed to predict the purchase behavior of e-commerce users. At the same time, the deep forest model is used to predict users' interest preferences. Build a personalized smart tourism service model to generate a list of recommended tourist attractions for users. Using the model constructed in this paper, the user conversion rate of the e-commerce platform is analyzed, and combined with structural equations, the service experience of smart tourism is evaluated.

2 Precision marketing modeling based on user needs

2.1 E-commerce Precision Marketing Implementation Plan for User Profiling

2.1.1 Collecting and processing e-commerce user data

Research all kinds of e-commerce display terminals and data sources, broaden the study of factors affecting e-commerce users, and deepen the combing and collection of user data from all e-commerce terminals and internal and external data sources. Formulate data cleaning rules and data standardization rules, according to the rules of data cleaning and processing, including duplication, vacancies, errors, inconsistencies and other issues of data processing, to ensure data accuracy. Research all the user identification in the data, including: imei, idfa, idfv, ID, cookie, email, phone and other identification fields screening, study the user's unified identification UID, build ID mapping system, mining and use of various types of associated ID, open the data barriers of various data sources, terminals, and businesses, and bring together the user's data in various channels to form The data form with the user's unified identity as the main key is shown in Figure 1.

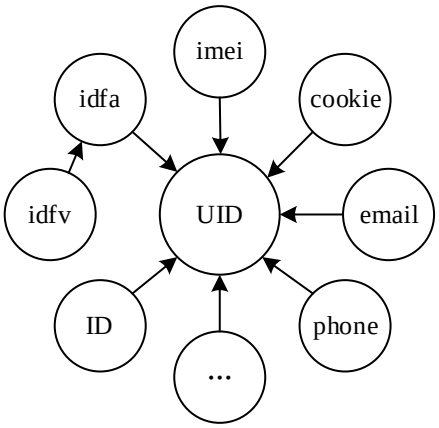


Figure 1. ID mapping system

2.1.2 Constructing an e-commerce user profile model

Construct e-commerce-oriented user portrait model, divide demographic attributes, social attributes, behavioral attributes, interest attributes, ability attributes, and psychological attributes into six dimensions based on e-commerce elements, construct the user portrait feature layer, form the user's performance in each dimension into a feature, and use the labeling rules and algorithms to realize the user marking, and label the user data. Build a complete labeling system, complete the construction of internal labels in a dimensional and hierarchical way, and study the two-level system of user labeling refinement division scheme.

The e-commerce user portrait is divided into 6 dimensions, 13 first-level labels, 92 second-level labels, and the model is represented as:

$$Persona = \{PO, SO, BE, IN, AB, PS\} \tag{1}$$

where each dimension corresponds to the first level of labeling and the model is represented as:

$$\begin{cases} P0 = \{NAC, SOC\} \\ S0 = \{SOF, SPF\} \\ BE = \{VIB, SHB, AFB\} \\ IN = \{LOI, SHI\} \\ AB = \{COA, PAA\} \\ PS = \{LIS, PEC\} \end{cases} \tag{2}$$

Each first-level label corresponds to a more refined second-level label, which is directly applied to the business and whose feature values have a high impact on the business.

2.1.3 Implementing population categorization from the perspective of precision marketing

Based on the e-commerce user profile, we will study the crowd classification problem from the perspective of precision marketing and gain insight into user classification. The first step is to study the main classification of the crowd, based on the e-commerce user profile behavioral attributes of the series of labels to divide the new user N, old user O two categories. The second step is to study the subcategorization problem of the crowd, further subdividing the old user O category based on the multi-class labels of the e-commerce user portrait, and applying the improved K-means algorithm to divide the old user O into two subcategories of stable user S and churn-prone user E. The third step is to study the subcategory population re-segmentation problem, combining the three categories of NSE users, using AP clustering method to obtain the fine-grained clusters, and completing the subcategory population re-segmentation.

Among them, the old user O category segmentation uses the improved K-Means algorithm, which is suitable for mining large-scale datasets, high efficiency, scalability, time complexity near linear, and suitable for the initial classification of users.The K-Means algorithm uses the distance as the similarity evaluation index, and the sum of squares of the errors from the sample points to the center of the category is used as the evaluation data for the goodness of the clustering, and the iterative method is used to make the overall categorization of the error sum of squares function is minimized. The details are as follows:

The old user data set is defined as $(X_1, X_2, \cdots, X_n)$, where each X_i is a feature vector of a user multidimension. The set of user group data is divided into two categories $S = \{S_1, S_2\}$, where the sum of squared distances of all elements in S_j to the category center U_k is $\sum_{x_i \in S_j} \|x_i - u_k\|^2$, where $\|x_i - u_k\|$ is the Euclidean distance. The objective of the classification is to minimize the weighted sum of the sum of squared distances of the two categories, and the objective function can be expressed as:

$$F(U) = \arg\min \sum_{k=1}^{2} \sum_{x_i \in /_j}^{2} \|x_i - u_k\|^2 \tag{3}$$

The binary variable $z_{ik} \in \{0,1\}$ is introduced to test into which category the data points should be classified. z_{ij} is defined as follows.

$$z_{ik} = \begin{cases} 1 & if \left\| x_i - u_k \right\|^2 = \min_{k=\in\{1,2\}}^{\min} \left\| x_i - u_k \right\|^2 \\ 0 & other \end{cases} \tag{4}$$

Clustering centers for:

$$u_k = \sum_{i=1}^{n} z_{ik} x_{ik} \Big/ \sum_{i=1}^{n} z_{ik} \tag{5}$$

Therefore the improved objective function is:

$$F\left(z,U\right) = \arg\min \sum_{k=1}^{2} \sum_{x_i \int_j}^{2} z_{ik} \left\| x_i - u_k \right\|^2 \tag{6}$$

Iterate the algorithm by minimizing the objective function as the optimization goal, the first iteration needs to be an optimization, from the set $\left(X_1, X_2, \cdots, X_n\right)$ consciously selected the distance between the two elements as the center of the two categories, the purpose is to pull away from the distance of the initial point, reduce the similarity between the clusters, to improve the efficiency and effectiveness of clustering, and then calculate the distance from each element in the set to the center of the two categories, and classify these elements into the distance to the the smallest one to the category for clustering. Recalculate the center of each category according to the results of the previous clustering. Repeat the above steps by constantly updating the centers $U = \left\{u_1, u_2\right\}$ of the two clusters until the classification results no longer change.

On the basis of NSE classification, various types of users need to be subdivided to form segmented populations. Here the AP clustering algorithm is used, which converts the user data into network nodes, calculates the clustering centers through two kinds of message passing between the attraction and belongingness of each node of the network, and continuously updates the attraction and belongingness of each node through iteration until multiple high clustering centers are generated to get the crowd segmentation, the advantage of the AP clustering algorithm is that there is no need to specify the number of clusters, it is insensitive to the selection of the initial value, and the clustering result is stable and conforms to the crowd segmentation scenario. The details are as follows:

After NSE classification one of the classes of user data set is defined as $\left(Y_1, Y_2, \cdots, Y_n\right)$, and the initial similarity matrix is generated using Euclidean distance:

$$s\left(i, j\right) = -\left\| y_i - y_k \right\|^2 \tag{7}$$

Set the initial reference degree P, $P_i = P\left(i\right)$ as the reference degree of Y_i, which refers to the reliability of using Y_i as the center of clustering. Generally set P as the median of the similarity values.

Calculate the attractiveness value between two user data:

$$r\left(i,k\right) = s\left(i,k\right) - \max_{k' \neq k} \left\{a\left(i,k\right) + s\left(i,k\right)\right\} \tag{8}$$

where $r(i,k)$ describes the degree to which the user data object y_k is suitable as a clustering center for the user data object y_i, and $a(i,k')$ describes the degree to which the user data object y_i is suitable for selecting the user data object y_k as its clustering center.

Calculate the value of the degree of belonging between two user data:

$$a(i,k)=\min\left\{0,r(k,k)+\sum_{i\neq i,k}\max\left\{0,r(i,k)\right\}\right\} \tag{9}$$

$$a(k,k)=\sum_{i\neq k}\max\left\{0,r(i,k)\right\} \tag{10}$$

The decay coefficient is introduced to update the iteration attraction $r(i,k)$ and attribution $a(i,k)$, and the calculation is terminated when the clustering centers no longer change over a number of iterations, and the individual clustering centers and individual classes are determined to obtain the segmented population.

2.1.4 Develop and implement precision marketing strategies for e-commerce companies

The three above types of NSE users have a life cycle relationship, and for specific user roles, there is an evolution and transition relationship between new, stable, and churn-prone users. Formulate NSE precision marketing strategy, and formulate three precision marketing strategies: "attracting new users", stabilizing user "customer protection" strategy, and "recall" strategy of easily losing users on the basis of e-commerce user portraits, so as to comprehensively improve marketing results.

The NSE precision marketing strategy is mathematically expressed as:

$$SalesEffect = NewUsers + StableUsers + EasilyLostUsers \tag{11}$$

2.2 Machine Learning Based User Buying Behavior

2.2.1 Machine learning algorithms

1) Support Vector Machine

Support Vector Machines, abbreviated as SVMs, are widely used in various classification problems, such as e-commerce platform user purchase prediction research, customer churn prediction research, etc. [22].

The basic idea of SVM is to find a separating hyperplane to separate the positive and negative samples. However, when looking for the separation hyperplane, there are many hyperplanes that can achieve the classification purpose, and the principle of support vector machine is to find the best separation hyperplane according to the principle of maximizing the classification interval. Figure 2 shows the schematic diagram of support vector machine.

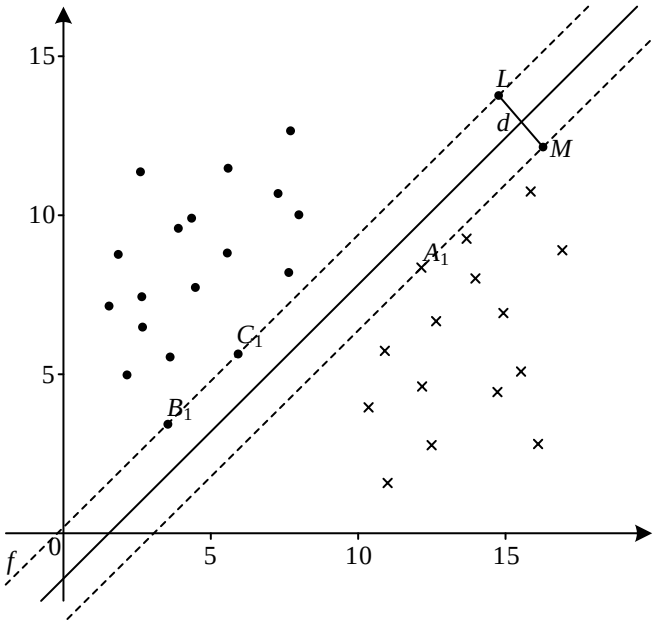


Figure 2. Schematic diagram of support vector machine

Taking the linearly differentiable support vector machine as an example, assuming that $\omega^T x+b=0$ represents the linear equation of the separation hyperplane, next, the best separation hyperplane is found by determining the values of ω and b, as shown in Fig. 2, and the plane represented by the black solid line in the figure is the best separation hyperplane to be found. After the separation hyperplane is found, the distance from the sample point to the plane needs to be calculated, and the formula for the distance is as follows:

$$\gamma^{(i)}=\frac{\left(\omega^T x^{(i)}+b\right)}{\|\omega\|} \tag{12}$$

Assuming that the positive and negative samples are denoted by 1 and -1, the points with the smallest distance from the hyperplane are called the support vectors, and the sum of the distances of the positive sample support vectors and the negative sample support vectors to the hyperplane is γ. It is necessary to find the values of ω and b such that γ is maximized, i.e., to find ω that satisfies the following equation.

$$\max_{\omega,b}\frac{2}{\|\omega\|} \tag{13}$$

To find the extreme values, you can use the Lagrange multiplier method, which translates into the following equation:

$$L\left(\omega,b,a\right)=\frac{1}{2}\|\omega\|^2+\sum_{i=1}^{m}\alpha_i\left[1-y_i\left(\omega^T x_i+b\right)\right] \tag{14}$$

After that, the derivatives of ω and b can be transformed into the following equation:

$$\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j \tag{15}$$

At this point, the problem of finding extreme values can be transformed into:

$$\max_{\alpha}\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j \tag{16}$$

The constraints are:

$$s.t.\sum_{i=1}^{m}\alpha_i y_i=0\left(\alpha_i>=0\right) \tag{17}$$

Support vector machine is a classical machine learning method, which has its unique advantages when applied to classification problems and has been widely adopted. The advantage of support vector machine is that it has strong applicability to complex datasets, even if the number of features is very large, it can still achieve good results, especially for high-dimensional data, it can accurately find the separation hyperplane through the mapping. However, Support Vector Machines also have some limitations, due to the complexity of the training process, when the data set is very large, the training effect of Support Vector Machines may be poorer, and it is easy to overfitting.

2) Decision Tree

Decision trees, abbreviated as DT, are also widely used in classification. From the name of decision tree, it is easy to understand that a decision tree is to get a definite result by continuously branching and making decisions with the help of the structure of a tree. The key to decision trees is how to build up the structure of the whole tree and how to make the right decisions at the branches [23].

The training process of the decision tree is relatively simple and very easy to understand the topmost circle in the decision tree is called the root node, the middle circle is called the internal node, and the rectangle is called the leaf node. First, the complete structure of the decision tree needs to be divided according to the research problem starting from the root, next, decisions are made for each node in turn, and according to the principle of loss minimization, the final decision result at the node can be determined by the information gain of each node, Gini coefficient and so on.

The information entropy is calculated by the formula:

$$H\left(x\right)=-\sum_{k=1}^{N}p_k\log_2\left(p_k\right) \tag{18}$$

The Gini coefficient is calculated by the formula:

$$Gini\left(D\right)=1-\sum_{k=1}^{N}P_k^2 \tag{19}$$

Decision trees are usually applied to classification problems, and they are less affected by missing values and outliers, as well as by the type of data, so that even if the data is not standardized, more accurate results can be obtained, which is a strong advantage compared with other methods.

Additionally, the training process of the decision tree is relatively simple, the calculation process is simple, easy to understand, and it is an extremely commonly used classification algorithm.

3) Random Forest

Random Forest, abbreviated as RF, is a common integrated learning algorithm and belongs to Bagging integrated learning [24]. In fact, Random Forest is a combination of numerous decision trees through Bagging integration, which can be applied to classification problems.

The basic principle of Random Forest is the integration of multiple decision trees into Bagging, which is based on the self-sampling method. First, a number of sub-datasets are extracted from the original dataset using the self-sampling method, which means that one sub-dataset is randomly extracted from the dataset each time there is a put-back. Then, several decision tree models are trained with each sub-dataset, i.e., as decision tree A, decision tree B, etc. in Fig. After that, the learner is trained with a random selection of features from the dataset. The output of these decision trees is used to generate the final output by voting. Due to the randomness of sampling when obtaining the data set, and the features input into the learner are also randomly selected, so for the same data set, the output of the Random Forest algorithm may be different, usually in the actual training, set the random seed that can solve this problem.

Random Forest algorithm in the training, each decision tree training do not affect each other, so the advantage of Random Forest is that its training efficiency is higher, the training speed is faster, for the existence of certain missing values of the dataset also has good applicability. However, Random Forest also has the disadvantage of being easily affected by outliers, when Random Forest is applied to noisy datasets with more outliers, it is likely to produce the problem of overfitting, so this problem should be paid attention to when applying the Random Forest algorithm, in addition, there are many parameters of the Random Forest, and if the parameters are not set appropriately, it will have a great impact on the accuracy of the results, therefore, when using the random forest algorithm, it is necessary to repeatedly adjust the parameters. In the Forest algorithm, the parameters should be adjusted repeatedly to find the optimal parameters as much as possible to enhance the accuracy of the model.

4) LightGBM

LightGBM is optimized for the high time and space complexity of XGBoost to solve the problems of slow training speed of large-scale samples, etc. The scalability of LightGBM is mainly manifested in three aspects: the use of smaller samples, fewer features, and less memory, which are respectively achieved by using Gradient-based One-sided Sampling (GOss), Mutually Exclusive Feature Bundling (EFB), and the histogram algorithm (Histogram) three techniques are implemented.

The gradient-based one-sided sampling algorithm improves model training speed by reducing sample size. The algorithm downsamples the samples with small gradient values to eliminate most of the small gradient samples, and the remaining samples with larger gradient values have a greater impact on the information gain, so the remaining samples are used to calculate the information gain. The specific steps are as follows:

(1) Sort the samples in descending order according to the absolute value of the gradient.

(2) Select $a\times100\%$ sample with the highest gradient value as a subset of the large gradient samples for the sorted sequence.

(3) Randomly select $b\times100\%$ samples among the remaining samples as a subset of small gradient samples.

(4) Merge the large gradient samples and the small gradient samples.

(5) Multiply the subset of small gradient samples by a constant factor of $(1-a)/b$ to offset the effect of the reduced sample size on the data distribution.

(6) Learn a new learner using the sampled samples described above.

Mutually exclusive feature bundling strategy improves the model training speed from the perspective of reducing the feature dimensionality, this algorithm binds mutually exclusive features in high dimensional features, thus reducing the number of features to improve the training speed. High-dimensional features are often mutually exclusive in sparse feature space, i.e., these mutually exclusive features cannot take non-zero values at the same time, so the mutually exclusive features can be reasonably bundled into a single feature, thus reducing the number of features and improving the computational efficiency, and if the two features are not completely mutually exclusive, i.e., the two features take non-zero values in order to strike a balance between the computational accuracy and efficiency, the degree of non-mutual exclusivity of features can be measured using the conflict ratio of the measure to evaluate whether to bundle features together or not.

The histogram algorithm is a substitute for XGBoost's pre-sorting algorithm, designed to address the issue of excessive number of split points. The basic idea of this algorithm is to split the features into boxes: divide the continuous feature values into k box, and construct a histogram with width k. When searching for the optimal segmentation point, only the discrete values of the histogram are searched, thus effectively reducing the number of segmentation points and lowering the storage and computation costs. Compared to the pre-sorting algorithm of XGBoost, the time complexity of LightGBM is greatly reduced. In addition, the histogram-based algorithm saves approximately seven times as much memory as the pre-sorting algorithm.

LightGBM grows the tree by using a leaf-by-leaf growth strategy, which selects the leaf nodes with the highest splitting gain to grow. Most tree models grow the tree by a per-layer growth strategy, i.e., by traversing the data once to split all leaf nodes in the same layer, a strategy that can be easily optimized for multiple threads. However, in practice, many leaf nodes have very low splitting gains, and splitting all leaf nodes indiscriminately increases the computational overhead. LightGBM employs a grow-by-leaf method to select the leaf node that has the greatest splitting gain among all leaf nodes to split. In addition, LightGBM restricts the depth of the tree to prevent overfitting by growing too deep decision trees, so the maximum depth parameter of the tree and the maximum number of leaf nodes parameter should be controlled during the model tuning process to avoid overfitting.

2.2.2 Modeling

For numerical data, due to the different ranges of values of each variable and different scale, it will cause bias to the analysis results, so for numerical data, it needs to be standardized, in this paper, the numerical variables in the dataset are standardized by using Z fractional standardization method,

so that the value of each variable is distributed in the range of -1 to 1. The formula of the Z -fractional standardization method is as follows:

$$y_i = \frac{x_i - \overline{x}}{s} \tag{20}$$

There are six variables in this dataset, which are “Chinese_subscribe_num”, “math_subscribe_num”, “add_friend”, “add_group”, “study_num”, and “city_num_l”. In the classification problem, the existence of sub-type variables will affect the calculation of label distance, so it is necessary to do one-hot encoding on these six sub-type variables, and the one-hot encoding will map the variables to the Euclidean space, so as to avoid the influence of the values of the variables on the calculation of distance.

2.3 User Interest Preferences

2.3.1 Deep forest based user interest prediction model

1) User interest vector characterization

Traditional item characterization usually involves converting different items into binary codes using numerical conversion. This straightforward characterization method, which converts items into numerical sequences of 0s and 1s, is obviously not entirely suitable for characterization of interest preference in users’ online behaviors. The algorithm proposed in this paper introduces the concepts of packets and examples into the characterization of user Internet logs, and views the prediction of user interests as a multiple-example learning problem. As a result, this paper proposes to use the Conceptual Vector Representation (CVR) algorithm to process the original dataset of the model and obtain the user interest vector representation as input for the training model.

Traditional machine learning models are divided into three main approaches when solving prediction problems: supervised learning, unsupervised learning, and reinforcement learning. 1997 saw the introduction of another machine learning method, multiple-example learning, which was proposed for the prediction of molecular activity by Dietteroch et al. It is different from supervised learning in which all examples in the training set are labeled. Unlike supervised learning where all examples in the training set are labeled and unsupervised learning where all examples in the training set are unlabeled, multiple example learning labels and defines the training data at both the package and example levels. For a given training set

$$\left\{\left(X_1, Y_1\right), \left(X_2, Y_2\right), \cdots, \left(X_n, Y_n\right)\right\} \tag{21}$$

where $X_i = \left\{x_{i1}, x_{i2}, \cdots, x_{it}\right\}$ denotes the set of points in space consisting of a finite number of points called packages, all points in space x_i are called examples, and Y_i denotes the labeling of the category to which the package X_i corresponds. It can be seen that in multiple-example learning, a single example has no labeling information, while a package composed of multiple examples has labeling information. Traditional supervised learning can be viewed as single-example single-labeling, and multiple-example learning as multiple-example single-labeling.

Definition 1 (Class determination of packages) A package is defined as a positive package if and only if there is at least one positively labeled example in the package, otherwise the package is a negative package.

This leads to the definition of the multiple example categorization problem:

Definition 2 (Multiple Example Classification Problem) For a given training set, $\left(X_i, Y_i\right)$ in the packet $X_i = \left\{x_{i1}, x_{i2}, \cdots, x_{it}\right\}$ presence of examples x_{ij} labeled positive, then Y_i take 1 to denote that X_i is a positive packet: $\left(X_i, Y_i\right)$ in the packet $X_i = \left\{x_{i1}, x_{i2}, \cdots, x_{it}\right\}$ all examples x_{ij} labeled negative, then Y_i take -1 to denote that X_i is a negative packet, and from this, a real-valued function $g\left(x\right)$ is found as a decision condition to obtain the decision function:

$$f\left(x\right) = sign\left(g\left(x\right)\right) \tag{22}$$

to predict the identification y corresponding to any example x in the space, so to predict the category labeling of an unknown packet $\tilde{X} = \left\{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n\right\}$, according to Definition 1 and Definition 2, the packet is a positive packet when and only when all examples $\tilde{x}_1, \cdots, \tilde{x}_n$ are positive, otherwise it is a negative packet, and the corresponding category labeling of the resulting packet $\tilde{X}$ takes the value $\tilde{Y}$ as:

$$\tilde{Y} = sign\left(\max_{i=1,2,\cdots,m} f\left(\tilde{x}_i\right)\right) \tag{23}$$

When analyzing the data of user Internet logs, it can be found that in reality, it is impossible for users to generate tagging behaviors for all of their Internet logs, and therefore, the prediction accuracy of the training model will be affected to some extent. In order to solve this problem, this paper introduces the concept of packages and examples, an idea derived from multiple example learning for recommendation processing of links in web pages.

2) Conceptual vector representation algorithm

Conceptual Vector Representation (CVR) algorithm is a multi-example learning algorithm based on degradation strategy. The core idea is to transform the multi-example dataset into a single-example dataset by changing the representation of the data packet so as to solve it. Referring to the idea of bag-of-words model for natural language processing and text information processing, the main work of CVR algorithm is two parts: 1) mining the concepts contained in all the examples, and 2) quantizing the examples in the packet onto concept clusters. The quantized data is then further processed using the R-pattern method.The CVR algorithm first organizes the original dataset into a training set consisting of packets $\left\{\left(B_1, Y_1\right), \cdots, \left(B_n, Y_n\right)\right\}$, $\left(B_i, Y_i\right)$ denoting the packet B_i corresponding to the label Y_i, where $B_i = \left\{x_1, x_2, \cdots, x_m\right\}$, i.e., packet B_i consists of m examples. Follow the steps below to change the representation of the packets:

Step 1: Use clustering method to divide all the examples into d clusters, where the packets to which the examples belong are not considered. After the division the result is expressed in the form shown in equation (24):

$$G=\{G_1,G_2,\cdots,G_d\} \tag{24}$$

Step 2: Obtain the initialized relational pattern RP of the packet by quantizing the examples of the packet to d cluster through equation (25), where $f=tf(g,B_i)$ is the number of examples belonging to cluster g in the computed packet B_i:

$$PB_i=\{(g,f)|g\in G, f=tf(g,B_i)\} \tag{25}$$

Step 3: Merge the initialized relational schema RPs with the same set of clusters as shown in equation (26):

$$RP_i=\{(g,P_i)|g\in G, P_i=B_{x_i}\oplus B_{x_i}\oplus\cdots\oplus B_{x_i}\} \tag{26}$$

Where, $x_1,x_2,\cdots,x_m\in[1,n]$, the initialized relational pattern RP of the packets connected by synthesis operator $\oplus$ has the same cluster set. At this time, the total number of packets is n, P_i the number of merged packets possessing the same set of clusters is $P_{num(B_x)}$, i.e., the number of packets connected by synthesis operator $\oplus$ is $P_{num(B_x)}$, and the support of P_i is obtained from equation (27):

$$Support(P_i)=\frac{P_{num(B_x)}}{n} \tag{27}$$

$$\beta(P_i)=\{(G_1,\omega_1),(G_2,\omega_2),\cdots,(G_d,\omega_d)\} \tag{28}$$

Step 4: The merged relational schema RP is represented as a standard form as shown in equation (28):

Where, $P_i\in RP$, ω_d denote the proportion of examples belonging to cluster G_d in the relational schema P_i to the total number of examples in the schema, calculated as shown in equation (29):

$$\omega_d=\frac{f_d}{\sum_{j=1}^{d}f_j} \tag{29}$$

where f_d is the number of examples belonging to cluster G_d in the merged P_i of the relational schema RP:

$$pr\beta(G)=\sum_{P_i\in RP,(G,\omega)\in\beta(P_i)}Support(P_i)\times\omega \tag{30}$$

Step 5: Calculate the importance of each cluster in the dataset as shown in equation (30):

Step 6: Represent the packet into concept vector form as shown in equation (31):

$$Bag_i = \left[N_1 \times pr\beta\left(G_1\right), N_2 \times pr\beta\left(G_2\right), \cdots, N_d \times pr\beta\left(G_d\right) \right] \tag{31}$$

Step 7: Train the learner using the collated training set.

The dataset representation is simplified by obtaining the concept vector representations of the items through the concept vector representation algorithm, and the multi-example dataset is transformed into a single-example dataset. After that it is used as a training set to train the learner.

2.3.2 User Interest Prediction Based on Deep Forest Algorithm

On the basis of the analysis and establishment of the user portrait labeling system, for the prediction of user dynamic interest labels, this paper proposes a user interest prediction algorithm based on the deep forest model and improves the model by using feature selection weights. The method utilizes the multi-granularity scanning of the deep forest to learn high-dimensional representations at a lower cost, and the number of layers of the cascade forest becomes a parameter that is automatically set according to the data situation through adaptive decision, which ensures the model complexity and accuracy while breaking through the limitation of hyperparameters.

From the analysis in the previous section, it can be seen that increasing the width and depth of the network structure has a significant effect on improving the performance of the model. For the deep forest model, increasing the width of the network is realized by increasing the types of base learners integrating the model in the two-part structure, while increasing the depth of the network is realized by expanding the number of training layers in the cascade forest structure.

2.4 Personalized Intelligent Tourism Service Modeling

First the application of the model is presented. The final model is obtained through model training, and the next step is to apply the trained model to a real task. It is trained to achieve high-efficiency and high-quality attraction recommendations. Given a user u who inputs a time constraint of m to travel in the recommender system and a candidate list consisting of attractions $S_1(u,l)$, the recommender system generates a TOP-k list of recommendations based on the recommendation formula 1, where the higher the result indicates that the attraction is more in line with the user's travel preferences. The result of Recommendation Equation $S_1(u,l)$ describes the extent to which the attraction meets the user's travel preferences while satisfying the user's time constraints. From there, a Top-k attraction is presented to the user as a recommendation list based on the recommendation score:

$$S_1(u,l) \propto (V_u + V_m)^T \cdot v_1 \tag{32}$$

Where V_u denotes the user-level implicit semantic feature vector of user u, V_m denotes the time-level implicit semantic feature vector of user's traveling, and v_1 denotes the attraction-level implicit semantic feature vector of all attractions. Next, the trained model is applied to the real attraction recommendation task, in which users generally do not choose attractions that they have visited before. Therefore, a TOP-k recommendation list is generated based on the demand of user u at time point m, and the set of P attractions that the user has visited before l_p and the candidate list of other attractions l according to the recommendation formula $S_2(u,l)$:

$$S_2(u,l) \propto (V_u + V_m + V_{lp})^T \cdot v_1 \tag{33}$$

Where V_u denotes the user-level implicit semantic feature vector of user u, V_m denotes the time-level implicit semantic feature vector of the user's tour, v_1 denotes the attraction-level implicit semantic feature vector of all attractions, and V_{lp} denotes the sum of implicit semantic feature vectors of all attractions in the attraction dataset that the tourists have already visited. It is found that the cold-start problem of attraction recommendation can be effectively solved by the MLS2vec model. When a brand-new user comes into the recommendation system, a TOP-k recommendation list is generated based on the user's travel time point m and the candidate list composed of attractions l using recommendation formula $S_1(u,l)$:

$$S_1(u,l) \propto V_m{}^T \cdot v_1 \tag{34}$$

3 Analysis of empirical studies

3.1 User conversion rate

3.1.1 Order trends

The user conversion rate of the e-commerce platform is analyzed using the precision marketing model constructed in this paper. Figure 3 shows the trend graph for the order. The data of the user order form is briefly analyzed, and the user order form includes key order information such as the date of placing an order, the region in which the order was placed, and the number of pieces of goods ordered. Since the goal of the contest is to analyze whether the target user placed an order in the tag month and the earliest order date, which is closely related to the information in the order table, analyzing the user order table is a top priority. The figure depicts the total number of orders placed by the target user on a daily basis between June 1, 2021 and April 30, 2022. As can be seen from the graph, there were several large fluctuations between July-August 2021, November-December 2021, December 2021-January 2022, and March-April 2022, which may be attributed to the use of the Precision Marketing Model coupled with the effects of the 618 Shopping Festival, the Double Eleven and Double Twelve Shopping Festivals, and the April Promotions. After the intervention of the Precision Marketing Model, the overall order trend after the start of the year 2022 is higher than the June to December 2021 period. The number of orders for 2021 and 2022 is 138 and 245 respectively, which represents a 77.62% improvement in the number of orders.

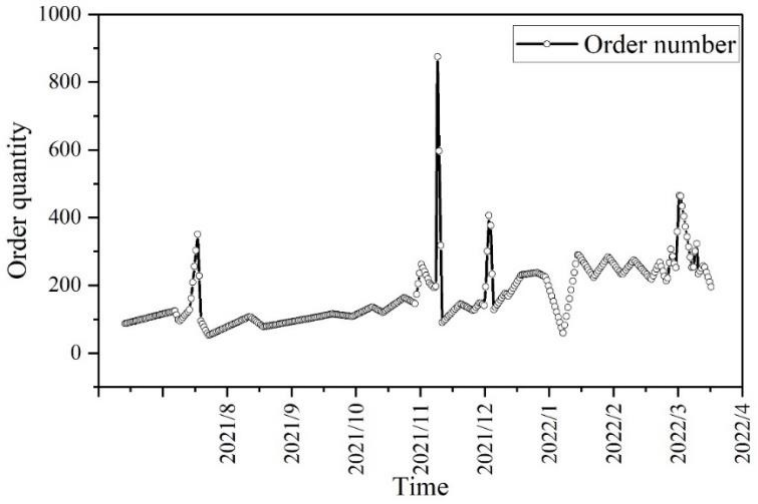


Figure 3. Order trend

3.1.2 Number of users purchased

Figure 4 shows the number of orders and purchasing users month by month, where the total number of orders and users within each month from May 2021 to April 2022 are calculated using month as the statistical dimension. It can be seen that the trends in month-by-month orders and month-by-month purchasing users are heading in a similar direction and can be combined and analyzed together, with two more notable increases between May 2021 and January 2022, in June 2021 and November 2021, when the number of month-by-month orders and purchasing users were 25,424 and 17,678, respectively, for the month of June. The number of orders and users purchased month by month in November was 46,634 and 31,357 respectively. The reason analysis is similar to Figure 3, which is considered as the effect of the precision marketing model. the number of users and the number of orders showed a significant growth trend from the beginning of February 2022 to April 2022, considering the reason is that the target users are selected from the part of the users who have purchased the target category of goods in the three months prior to the examination of the time period of May 2022, and therefore it is reasonable to see that the users' purchasing behaviors are active in these three months, presenting this trend of change with the The selection criteria of the target users are related.

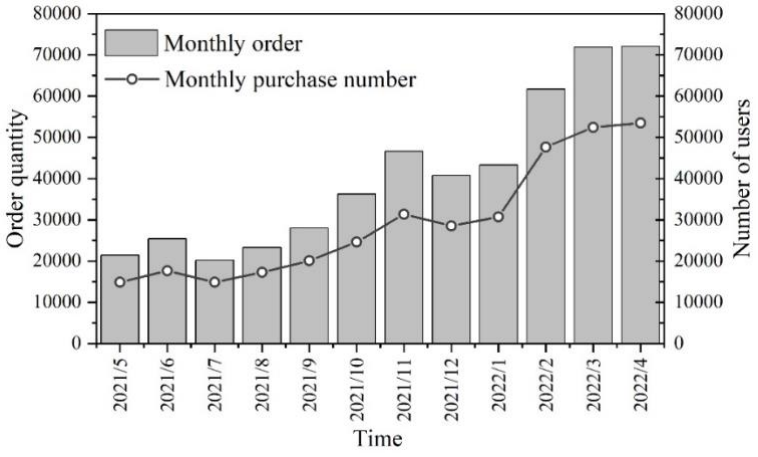


Figure 4. Monthly order and purchase number

Figure 5 shows the month-by-month average number of orders and products purchased, where the average number of orders and the average number of products purchased by users within each month from May 2021 to April 2021 are calculated using month as the statistical dimension. As can be seen from the figure, there is a small difference in the mean values between the months, with a distribution range of [1.33, 1.48] for the average number of orders and [2.03, 2.61] for the average number of products purchased.

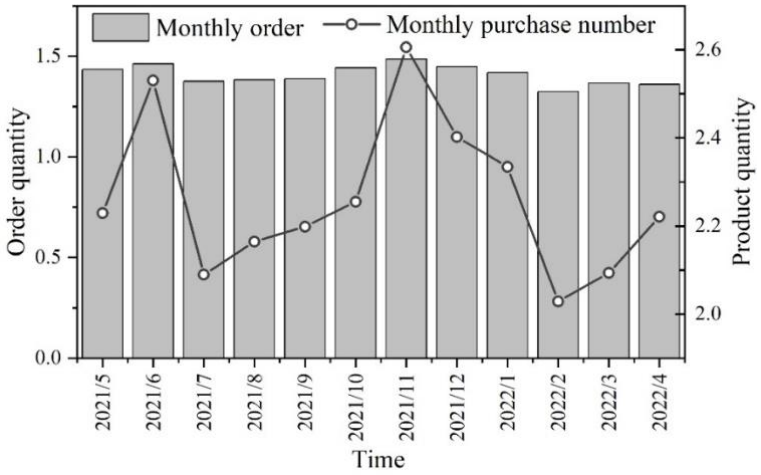


Figure 5. Monthly orders and purchase of products

3.2 Intelligent Tourism Service Experience

3.2.1 Modification of variables and research hypotheses

1) Variable Measurement

Table 1 shows the variable measurement, this paper in February 2023 in the CL tourist resort queuing area, parking lot, sitting area, restaurant and other areas randomly distributed questionnaires to tourists, the results of the recovery of field research questionnaires 400, according to the rules of judgment of the questionnaire data is invalid screened out valid questionnaires 298, and then through the microblogging group sent a message in the form of WeChat to WeChat friends to send out the e-questionnaire, the results of the recovery of online research 130 questionnaires, according to the judgment rule of invalid questionnaire data screened out 80 valid questionnaires, the final field research and online research effective questionnaires totaled 378, the effective rate of 71.32%.

Table 1. Variable measurement

Serial number	Latent variable	Variable	Describe
1	Intelligent travel marketing inspection	A1	Easy access route
		A2	The platform is easy to use and smooth
		A3	Learn about relevant activities
		A4	Platform understanding
		A5	Learn about relevant activities
2	Intelligent travel service experience	A6	Intelligent service is deeper in the scenic area
		A7	Get to your destination faster
		A8	Improve shopping efficiency
		A9	Improve meal efficiency
		A10	Quick response consulting and complaints
		A11	Check-in quickly
3	Intelligent maintenance experience	A12	Wifi coverage wide, network speed block
		A13	Quickly find rescue methods and equipment
		A14	Improve checkout efficiency
4	Intelligent integrated management experience	A15	Quick booking and check-in
		A16	Better arrangements
		A17	Feel fresh and interesting
5	Intelligent travel product experience	A18	The situational experience is interesting
		A19	Facilitate the acquisition of the program information
		A20	Interactive devices are fresh and interesting
		A21	Show fresh and interesting
6	Intelligent infrastructure experience	A22	Sightseeing safety
		A23	labeling
7	Perceived value	B1	Compared with the cost I need, it is worth it for me to travel with wisdom
		B2	Compared with the energy I need, it is worth the use of smart travel
		B3	Compared with the time I need to spend, it is worth it for me to travel with wisdom
		B4	Compared with the risks I need, it is worth the use of smart travel
		B5	Compared with the various costs I have to pay, intelligent travel generally meets my needs
8	Behavioral intention	C1	I will share my wisdom travel experience on the micro blog and the circle of friends
		C2	Smart travel will make me more likely to consume more products or services
		C3	If someone asks me for advice, I will recommend smart ways to travel
		C4	In the future, I will choose the smart tourist attractions

2) Revision of research hypotheses

According to the survey scale, the smart tourism experience is divided into five dimensions, which are smart tourism marketing experience, smart tourism service experience, smart maintenance and guarantee experience, smart comprehensive management experience, and smart tourism product experience, so it is necessary to make modifications to the research

hypotheses in this paper, and delete the influence of the smart infrastructure experience on the behavioral intention and perceived value of tourists, and the final research hypotheses of this paper are as follows:

H1: Smart tourism experience has a positive effect on tourists' behavioral intention

H1a: Smart tourism marketing experience has a positive effect on tourists' behavioral intention

H1b: There is a positive effect of smart tourism service experience on tourists' behavioral intention.

H1c: There is a positive effect of intelligent maintenance and guarantee experience on tourists' behavioral intention.

H1d: There is a positive influence of intelligent comprehensive management experience on tourists' behavioral intention.

H1e: There is a positive influence of smart tourism product experience on tourists' behavioral intention.

H2: There is a positive effect of smart tourism experience on tourists' perceived value

H2a: There is a positive effect of smart tourism marketing experience on tourists' perceived value

H2b: There is a positive effect of smart tourism service experience on tourists' perceived value

H2c: There is a positive effect of intelligent maintenance and guarantee experience on tourists' perceived value.
H2d: There is a positive influence of smart integrated management experience on tourists' perceived value.

H2e: There is a positive influence of smart tourism product experience on tourists' perceived value.

H3: Perceived value has a positive effect on tourists' behavioral intention

3.2.2 Descriptive statistical analysis

Through the descriptive analysis of the research participants on the questionnaire situation as shown in Table 2, to understand the mean and variance of each dimension and the overall, it can be found that the research participants of the smart tourism experience, perceived value, behavioral intention are biased towards positive affirmation, the mean value of the smart tourism experience, perceived value, and behavioral intention are 3.7359, 3.6534, and 3.9455, respectively. The very small value of all the question items, the very large value, skewness, and kurtosis are all within reasonable limits.

Table 2. The scores of each variable

Variable	N	Minimum value	Maximum value	Mean	Standard deviation	Degree of bias	Kurtosis
A1	378	1	5	3.845	0.614	-1.458	3.948
A2	378	1	5	3.876	0.637	-0.843	2.054
A3	378	1	5	3.765	0.763	-0.715	1.148
A4	378	1	5	3.861	0.641	-1.056	2.445
A5	378	1	5	3.834	0.628	-0.856	2.045
A6	378	1	5	4.086	0.715	-0.686	0.815
A7	378	1	5	3.971	0.539	-0.848	4.358
A8	378	2	5	3.982	0.557	-0.725	2.763
A9	378	1	5	3.763	0.799	-0.856	1.086
A10	378	1	5	3.768	0.731	-0.756	1.039
A11	378	1	5	3.715	0.937	-0.582	-0.015
A12	378	1	5	3.526	0.852	-0.456	-0.126
A13	378	1	5	3.715	0.775	-0.648	0.569
A14	378	1	5	3.706	0.827	-0.826	0.825
A15	378	1	5	3.746	0.772	-0.485	0.368
A16	378	1	5	3.708	0.826	-0.615	0.154
A17	378	1	5	3.625	0.883	-0.726	0.185
A18	378	1	5	3.587	0.854	-0.498	-0.049
A19	378	1	5	3.541	0.935	-0.425	-0.295
A20	378	1	5	3.689	0.848	-0.285	-0.157
A21	378	1	5	3.436	0.915	-0.352	0.069
A22	378	1	5	3.596	0.805	-0.648	0.308
A23	378	1	5	3.584	0.905	-0.429	0.428
B1	378	1	5	3.548	0.728	-0.728	1.265
B2	378	1	5	3.759	0.647	-0.625	1.169
B3	378	2	5	3.848	0.625	-0.648	-0.067
B4	378	1	5	3.454	0.732	-0.348	-0.018
B5	378	1	5	3.658	0.705	-0.386	1.785
C1	378	1	5	3.918	0.695	-0.758	0.658
C2	378	2	5	3.928	0.698	-0.563	1.265
C3	378	2	5	3.958	0.625	-0.578	1.715
C4	378	1	5	3.978	0.728	-0.848	1.325

3.2.3 Analysis of variance

Based on the previous literature analysis, it can be found that previous studies have shown that demographic variables affect the results of perceived value and behavioral intention. In this paper, the research object is tourists who have visited CL tourism resorts, whose gender, age, education, monthly income, and occupation may affect the significant difference of tourists' smart tourism

experience, perceived value, and behavioral intention; therefore, this paper adopts the independent samples t-test to conduct the analysis of the difference between the gender on the smart tourism experience, perceived value, and behavioral intention, and one-way analysis of variance (ANOVA) to conduct the analysis of the difference between the age , education, monthly income, and occupation on the differential analysis of smart tourism experience, perceived value, and behavioral intention.

1) Table 3 shows the independent samples t-test of gender differences on the relevant variables regarding the effect of gender on each variable. The independent sample t-test was used to confirm whether gender makes a significant difference in the means of each research variable. From the results of the T-test, it can be seen that the hypothesized equal significant values of variance for smart tourism experience, perceived value, and behavioral intention are 0.0945, 0.4596, and 0.1655, respectively, with a significance of P>0.05, which indicates that gender differences do not make a significant difference in smart tourism experience, perceived value, and behavioral intention.

Table 3. Independent sample t test of related variables for gender differences

Variable		Mean t test			
		T	Df	Sig. (double side)	Mean difference
Intelligent travel experience	Let's say the variance is equal	1.6548	383.452	**0.0945**	0.0856
	Let's say that the variance is not equal	1.6948	373.469	0.0915	0.0841
Perceived value	Let's say the variance is equal	0.7584	383.452	**0.4596**	0.0389
	Let's say that the variance is not equal	0.7669	381.656	0.4485	0.0385
Behavioral intention	Let's say the variance is equal	1.6348	383.452	**0.1655**	0.0945
	Let's say that the variance is not equal	1.6348	371.596	0.1565	0.0945
Variable		Standard error value	95% confidence interval of the difference		
			Lower limit	Upper limit	
Intelligent travel experience	Let's say the variance is equal	0.0496	-0.0152	0.1823	/
	Let's say that the variance is not equal	0.0458	-0.0152	0.1825	
Perceived value	Let's say the variance is equal	0.0526	-0.0625	0.1463	
	Let's say that the variance is not equal	0.0515	-0.0625	0.1485	
Behavioral intention	Let's say the variance is equal	0.0596	-0.0198	0.2158	
	Let's say that the variance is not equal	0.0548	-0.0195	0.2365	

2) Table 4 shows about the effect of age, education, monthly income, and occupation on each variable. One-way ANOVA was used to test whether each of the above variables would make a significant difference in tourists' smart tourism experience, perceived value, and behavioral intention, and it can be seen from the table that the grouping of age and education would make a significant difference in smart tourism experience, perceived value, and behavioral intention, while the grouping of monthly income and occupation had no significant difference in smart tourism experience (0.368, 0.529), perceived value (0.285, 0.936), and behavioral intention (0.485, 0.741) had no significant difference. Therefore, the age and education of tourists are used as control variables in this paper.

Table 4. Age, degree, monthly income, career impact on each variable

Variable		Age	Educational background	Monthly income	Occupation
Intelligent travel experience	F	3.559	4.926	1.095	0.718
	Significance	0.048	0.001	**0.368**	**0.529**
Perceived value	F	5.569	2.826	1.268	0.315
	Significance	0.002	0.034	**0.285**	**0.936**
Behavioral intention	F	6.345	5.756	0.866	0.548
	Significance	0.002	0.002	**0.485**	**0.741**

3.2.4 Impact pathway analysis

From the running results of the first-order dimensional structural equation model of the five segmented dimensions of smart tourism experience, perceived value, and behavioral intention, the path analysis table of the structural equation model of the segmented dimensions of smart tourism experience was compiled, as shown in Table 5, which shows the path of the influence of the five segmented dimensions of the smart tourism experience on perceived value and behavioral intention, and the specific results are as follows: the influence of the smart tourism marketing experience on behavioral intention and perceived value The standardized coefficients are 0.136 and 0.193 respectively, and the P-value is less than 0.05, which indicates that the smart tourism marketing experience has a positive influence on tourists' behavioral intention and perceived value, and the hypotheses of H1a and H2a are valid.

The standardized coefficients of the influence of smart tourism service experience on behavioral intention and perceived value are 0.148 and 0.187, respectively, with P-value less than 0.05, indicating that smart tourism service experience has a positive influence on tourists' behavioral intention and perceived value, and the hypotheses of H1b and H2b are valid. The standardized coefficients of the influence of smart maintenance and guarantee experience on behavioral intention and perceived value are 0.128 and 0.194 respectively, with P-value less than 0.05, indicating that the smart maintenance and guarantee experience has a positive influence on tourists' behavioral intention and perceived value, and the hypotheses of H1c and H2c are valid.

The standardized coefficients of the influence of smart integrated management experience on behavioral intention and perceived value are 0.259 and 0.175 respectively, with P-value less than 0.05, indicating that smart integrated management experience has a positive influence on tourists' behavioral intention and perceived value, and the hypotheses of H1d and H2d are valid.

The standardized coefficient of the influence of smart tourism product experience on behavioral intention is 0.069 respectively, and the P=0.318 value is greater than 0.05, indicating that the influence of smart tourism product experience on tourists' behavioral intention is not significant, and the H1e hypothesis is not established. The standardized coefficients of the influence of smart tourism product experience on perceived value are 0.168 respectively, P value is less than 0.05, indicating that smart tourism product experience has a positive influence on the perceived value of tourists, and H2e hypothesis is established.

Table 5. The path analysis of the wisdom travel experience segmentation dimension

Path			Normalization factor	S.E.	C.R.	P
Behavioral intention	←	Intelligent travel marketing inspection	**0.136**	0.075	2.215	0.024
Behavioral intention	←	Intelligent travel service experience	0.148	0.073	1.912	0.045
Behavioral intention	←	Intelligent maintenance experience	0.128	0.068	1.918	0.049
Behavioral intention	←	Intelligent integrated management experience	0.259	0.074	3.012	0.003
Behavioral intention	←	Intelligent travel product experience	0.069	0.059	0.915	**0.318**
Perceived value	←	Intelligent travel marketing inspection	**0.193**	0.082	2.718	0.008
Perceived value	←	Intelligent travel service experience	0.187	0.084	2.036	0.039
Perceived value	←	Intelligent maintenance experience	0.194	0.075	2.485	0.015
Perceived value	←	Intelligent integrated management experience	0.175	0.084	2.136	0.036
Perceived value	←	Intelligent travel product experience	0.168	0.063	2.069	0.048
Behavioral intention	←	Perceived value	0.365	0.074	5.136	***

Note: *** indicates that P is significant at levels less than 0.001.

4 Conclusion

In this paper, an e-commerce precision marketing model is constructed based on user needs and profiles. By integrating a variety of machine learning algorithms, it predicts the user's purchasing behavior and interest preferences. At the same time, a personalized smart tourism service model is established, and the practicality of the online marketing model is explored through empirical research in the two directions of user conversion rate and smart tourism service experience.

In terms of user conversion rate, after using the precision marketing model for personalized marketing, the number of orders increased from 138 in 2021 to 245 in 2022, an increase of 77.62%. The precision marketing model has a positive impact on the number of purchasing users, and the number of month-by-month orders and purchasing users in November 2021 are 46,634 and 31,357, respectively, after using the precision marketing model.

The smart tourism experience, perceived value, and behavioral intention of the research participants are all inclined to positive affirmation, and the mean values of smart tourism experience, perceived value, and behavioral intention are 3.7359, 3.6534, and 3.9455, respectively, which is a good evaluation of smart tourism services. The standardized coefficients of the influence of smart tourism marketing experience on behavioral intention and perceived value are 0.136 and 0.193, respectively, while P is less than 0.05, indicating that the smart tourism marketing experience has a positive influence on tourists' behavioral intention and perceived value.

Funding:

1) This article is a research outcome of the 2024 Henan Provincial Higher Education Teaching Reform Research and Practice Project titled "Research on the Reform of Practical Teaching in E-commerce Specialty from the Perspective of 'College-Academy' Collaborative Education" (Project Number: 2024SJGLX0905).

2) This article is the phase-in result of the major project of the National Social Science Foundation of China in the year of 2021, Research on Marxist Ideology of People's Democracy (Project No. 21&ZD007), and the commissioned project of the Marxist Theory

Research and Construction Project, Research on People's Democracy as the Proper Meaning of Comprehensively Constructing a Socialist Modernized Country (Project No. 2023MYB010).

References

[1] Mehraliyev, F., Choi, Y., & Köseoglu, M. A. (2019). Progress on smart tourism research. Journal of Hospitality and Tourism Technology, 10(4), 522-538.

[2] Hamid, R. A., Albahri, A. S., Alwan, J. K., Al-Qaysi, Z. T., Albahri, O. S., Zaidan, A. A., ... & Zaidan, B. B. (2021). How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. Computer Science Review, 39, 100337.

[3] Li, Y., Hu, C., Huang, C., & Duan, L. (2017). The concept of smart tourism in the context of tourism information services. Tourism management, 58, 293-300.

[4] Sigalat-Signes, E., Calvo-Palomares, R., Roig-Merino, B., & García-Adán, I. (2020). Transition towards a tourist innovation model: The smart tourism destination: Reality or territorial marketing?. Journal of Innovation & Knowledge, 5(2), 96-104.

[5] Rudwiarti, L. A., Pudianti, A., Emanuel, A. W. R., Vitasurya, V. R., & Hadi, P. (2021, May). Smart tourism village, opportunity, and challenge in the disruptive era. In IOP Conference Series: Earth and Environmental Science (Vol. 780, No. 1, p. 012018). IOP Publishing.

[6] Su, H., & Meng, X. (2021, May). Research on the development of rural E-commerce based on smart tourism. In Journal of Physics: Conference Series (Vol. 1915, No. 3, p. 032007). IOP Publishing.

[7] Leung, R. (2022). Development of information and communication technology: from e-tourism to smart tourism. Handbook of e-Tourism, 23-55.

[8] Brandt, T., Bendler, J., & Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. Information & Management, 54(6), 703-713.

[9] Yang, W., & Lin, Y. (2022). Research on the interactive operations research model of e-commerce tourism resources business based on big data and circular economy concept. Journal of Enterprise Information Management, 35(4/5), 1348-1373.

[10] Ye, B. H., Ye, H., & Law, R. (2020). Systematic review of smart tourism research. Sustainability, 12(8), 3401.

[11] Pathmanathan, P. R., El-Ebiary, Y. A. B., Yusoff, M. H., Aseh, K., Al-Qudah, O. M. A. A., Pande, B., ... & Bamansoor, S. (2021, June). The Benefit and Impact of E-Commerce in Tourism Enterprises. In 2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE) (pp. 193-198). IEEE.

[12] Hassannia, R., Vatankhah Barenji, A., Li, Z., & Alipour, H. (2019). Web-based recommendation system for smart tourism: Multiagent technology. Sustainability, 11(2), 323.

[13] Nemade, G., Deshmane, R., Thakare, P., Patil, M., & Thombre, V. D. (2017). Smart tourism recommender system. International Research Journal of Engineering and Technology (IRJET), 4(11), 601-603.

[14] Huda, C., Ramadhan, A., Trisetyarso, A., Abdurachman, E., & Heryadi, Y. (2021). Smart tourism recommendation model: a systematic literature review. International Journal of Advanced Computer Science and Applications, 12(12).

[15] Xia, W. (2022). Digital transformation of tourism industry and smart tourism recommendation algorithm based on 5G background. Mobile Information Systems, 2022(1), 4021706.

[16] Nan, X., & Wang, X. (2022). Design and implementation of a personalized tourism recommendation system based on the data mining and collaborative filtering algorithm. Computational Intelligence and Neuroscience, 2022(1), 1424097.

[17] Gamidullaeva, L., Finogeev, A., Kataev, M., & Bulysheva, L. (2023). A design concept for a tourism recommender system for regional development. Algorithms, 16(1), 58.

[18] Pei, Y., & Zhang, Y. (2021, April). A study on the integrated development of artificial intelligence and tourism from the perspective of smart tourism. In Journal of Physics: Conference Series (Vol. 1852, No. 3, p. 032016). IOP Publishing.

[19] Thananchana, A., Noinan, K., & Wicha, S. (2022, January). The designing of cultural-based tourism recommendation system with community collaboration. In 2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON) (pp. 510-513). IEEE.

[20] Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. Big Data Mining and Analytics, 4(1), 47-55.

[21] Sarkar, M., Roy, A., Agrebi, M., & AlQaheri, H. (2022). Exploring new vista of intelligent recommendation framework for tourism industries: an itinerary through big data paradigm. Information, 13(2), 70.

[22] Xin Wan,Xiaoling Cai & Lele Dai. (2024). Prediction of building HVAC energy consumption based on least squares support vector machines. Energy Informatics(1),113-113.

[23] Junxi Wu,Guoyan Zhao,Meng Wang,Yihang Xu & Ning Wang. (2024). Concrete carbonation depth prediction model based on a gradient-boosting decision tree and different metaheuristic algorithms. Case Studies in Construction Materialse03864-e03864.

[24] yaping Zhang. (2024). E-commerce Network Security Monitoring Based on HTM Algorithm and Random Forest. Computer Informatization and Mechanical System(6),37-39.