Identification and Influence of Tourism Consumption Behavior Based on Artificial Intelligence

Jinxiao Duan
Faculty of Culture and Tourism, Jiyuan Vocational and Technical College Jiyuan 459000, Henan, China
E-mail: jinxiao_duan@hotmail.com

Keywords: artificial intelligence, tourism consumption behavior, machine learning, consumer preference

Received: May 16, 2024

Under the background of globalization, tourism has been widely concerned because of its remarkable promotion to economic development. With the rapid development of information technology, especially artificial intelligence (AI), the study of tourism consumption behavior has entered a new stage. This study explores the use of AI technologies such as machine learning and natural language processing to identify tourism consumption behavior and analyze its influencing factors, providing accurate market positioning and product promotion strategies for the tourism industry. Through in-depth analysis of online consumer behavior data and social media comments, this study utilizes random forest and logistic regression methods to identify and analyze the behavior characteristics and preferences of tourism consumers. The findings indicate that specific consumption behaviors and preferences are significantly related to consumers' personal characteristics and social and cultural background. The model evaluation results reveal an accuracy of 85% for the random forest model and 82% for the logistic regression model, demonstrating their effectiveness and complementary in tourism consumption behavior identification. Further analysis shows that high-income and high-budget consumers tend to prefer luxury travel options, while younger and frequent travelers lean towards budget-friendly choices. Emotional tendencies expressed in social media comments also significantly influence consumer preferences. This study fills the gap in existing research on intelligent identification of tourism consumption behavior and provides data-driven decision support for tourism enterprises, promoting the sustainable development of the tourism industry. The insights gained from this research enable tourism operators to optimize their marketing strategies and service offerings, ensuring a better match with consumer preferences.

Povzetek: Strojno učenje in obdelava naravnega jezika sta uporabljena za prepoznavanje turističnega vedenja in analizo dejavnikov, ki vplivajo nanj.

1 Introduction research background

In today's society, tourism has developed into a major driving force of the global economy, significantly impacting national economic growth, employment creation, and cultural exchanges. With the rapid advancement of digital technology, especially the application of artificial intelligence (AI) in data processing and behavior analysis, new growth opportunities have emerged in the tourism market. Understanding tourism consumption behavior is crucial for assessing market development and formulating effective policies. Identifying and analyzing consumer behavior patterns can help tourism enterprises optimize services, enhance customer satisfaction, and provide data support for governmental tourism strategies. However, tourism consumption behavior is highly complex and dynamic due to various factors, including economic conditions, social and cultural backgrounds, and personal preferences.

Traditional research on tourism consumption behavior primarily relies on questionnaires and face-to-face interviews. While these methods provide direct consumer feedback, they face limitations such as small sample sizes, long data collection periods, and complex analysis processes. In contrast, AI technologies, particularly machine learning and natural language processing, offer new perspectives and methods for studying tourism consumption behavior through the automatic processing of large-scale data and the ability to identify complex patterns. AI can extract valuable information from multi-source data such as social media, online reviews, and transaction records, uncovering consumer preferences and behavior trends through data mining and pattern recognition. This provides a scientific basis for personalized recommendations, market segmentation, and enhancing customer satisfaction with tourism products and services.

Therefore, exploring tourism consumption behavior based on AI has theoretical value and practical significance for guiding tourism enterprises and policymakers in optimizing decisions and promoting sustainable tourism development. This study aims to establish an effective behavior recognition model by analyzing the characteristics and influencing factors of tourism consumption behavior using advanced AI technology, offering new ideas and tools for formulating marketing strategies and improving tourism services.

Numerous detailed studies have been conducted, ranging from consumer behavior theories to environmental sustainability applications, and the emerging trend of AI in behavior analysis. Han discussed the influence of consumer behavior on environmental sustainability in the tourism and hotel industry, emphasizing the need to coordinate theory and practice to promote sustainable tourism practices [1]. Cui found a significant correlation between consumer satisfaction and their willingness to revisit, indicating that the quality of tourism experiences directly impacts tourism choices [2]. Manosuthi effectively predicted the revisiting intention of volunteer tourists by combining planned behavior theory with the normative activation model [3].

Wang analyzed how online review information influences consumers' purchasing decisions for tourism products from a knowledge network perspective, highlighting the importance of information processing on consumer behavior [4]. Manthiou and Kuppelwieser focused on consumers' reactions to slow travel, emphasizing the intrinsic value of slow travel and the promotion of environmental care [5].

In terms of technical applications, Guo demonstrated the effectiveness of deep learning models in identifying classroom behavior, showcasing AI's potential in behavior recognition [6]. Xie discussed AI's application in automatically identifying consumer behavior on e-commerce platforms, highlighting its role in enhancing customer experience and business processes [7]. Hu reviewed the application of deep learning in behavior recognition, analyzing how these technologies help understand complex consumer behavior patterns [8]. Bhatt explored machine learning in cognitive behavior analysis, emphasizing big data's potential in behavioral science [9]. Elayan studied the influence of behavior in the Internet of Things and the application of interpretable AI systems, highlighting how intelligent technology impacts consumer behavior [10].

1.1 Related works

Table 1: Comparison of previous studies in tourism consumer behavior

Methods Used	Datasets	Main Findings	Limitations
Consumer Surveys	Small-scale survey data	Impact on environmental sustainability	Small sample size, lack of real-time data
Logistic Regression	Tourist satisfaction surveys	Satisfaction linked to revisit intentions	No integration of advanced AI techniques
Norm Activation Model + Planned Behavior Theory	Volunteer tourist surveys	Predict revisit intentions	Limited consideration of diverse behaviors
Knowledge Network Analysis	Online reviews	Online reviews influence purchasing decisions	Struggled with large-scale data processing
Survey and Interviews	Slow travel consumer feedback	Intrinsic value and environmental concern	Limited generalization

Table 2: Comparison of current study and SOTA gaps

Aspects	Current Study	SOTA Gaps
AI Technologies	Machine Learning, NLP	Limited use of advanced AI
Data Types	Online reviews, social media, transaction logs	Smaller, less diverse datasets
Consumer Insight Depth	Comprehensive behavior and preference analysis	Less detailed consumer insights
Real-time Data Processing	Real-time insights from large-scale data	Limited real-time data analysis
Ethical Considerations	Data privacy, informed consent	Often not addressed

As shown in Table 1 and Table 2, Numerous detailed studies have been conducted, ranging from consumer behavior theories to environmental sustainability applications, and the emerging trend of AI in behavior analysis. Han discussed the influence of consumer behavior on environmental sustainability in the tourism and hotel industry, emphasizing the need to coordinate theory and practice to promote sustainable tourism practices. However, the study's limitation lies in its reliance on traditional survey methods, which may not capture real-time consumer behavior and preferences.

Cui found a significant correlation between consumer satisfaction and their willingness to revisit, indicating that the quality of tourism experiences directly impacts tourism choices. The limitation here is the lack of advanced AI techniques to provide deeper insights into consumer behavior and preferences, making the analysis less dynamic and comprehensive. Mouthiness effectively predicted the revisiting intention of volunteer tourists by combining planned behavior theory with the normative activation model. This study, however, did not account for the diversity of consumer behaviors, which limits its generalization across different tourist segments.

Wang analyzed how online review information influences consumers' purchasing decisions for tourism products from a knowledge network perspective, highlighting the importance of information processing on consumer behavior. The primary limitation is the struggle with large-scale data processing, which could be better addressed with advanced AI technologies. Malathion and Wiesenthal focused on consumers' reactions to slow travel, emphasizing the intrinsic value of slow travel and the promotion of environmental care. This study's findings are limited in their applicability to broader tourist behaviors and preferences.

In terms of technical applications, Guo demonstrated the effectiveness of deep learning models in identifying classroom behavior, showcasing AI's potential in behavior recognition. However, this research is more focused on educational settings and does not directly address tourism consumer behavior. Xie discussed AI's application in automatically identifying consumer behavior on e-commerce platforms, highlighting its role in enhancing customer experience and business processes. While relevant, this study is limited by its focus on e-commerce rather than tourism.

Hu reviewed the application of deep learning in behavior recognition, analyzing how these technologies help understand complex consumer behavior patterns. Despite its comprehensive review, the study lacks specific applications in tourism. Bhatt explored machine learning in cognitive behavior analysis, emphasizing big data's potential in behavioral science. The study is broad and does not focus on tourism-specific behavior, limiting its direct applicability to this field. Elayan studied the influence of behavior in the Internet of Things and the application of interpreter AI systems, highlighting how intelligent technology impacts consumer behavior. The limitation here is the focus on IoT, which does not fully cover the nuances of tourism consumer behavior.

Given the limitations of previous studies, there is a clear need for a comprehensive approach utilizing advanced AI technologies such as machine learning and natural language processing to analyze tourism consumption behavior. This study aims to fill this gap by providing real-time, detailed insights into consumer behavior and preferences, leveraging large-scale data processing capabilities. By addressing the shortcomings of prior research, this study not only advances the understanding of tourism consumer behavior but also enhances the practical applications of AI in the tourism industry, leading to more effective market positioning and product promotion strategies.

2 Theoretical basis and application status

2.1 Tourism consumption behavior theory

The theoretical study of tourism consumption behavior is fundamental to understanding the diversity and complexity of consumers' decision-making processes. This field integrates interdisciplinary perspectives, including psychology, sociology, and economics, aiming to uncover the internal motivations and external conditions driving tourism consumers' behavior [11-13]. Within the framework of tourism consumption behavior, motivation theory explains why individuals choose to travel, information search behavior clarifies how consumers gather and evaluate tourism-related information, and the decision-making process involves selecting specific tourism products or services.

Specifically, the decision-making process in tourism consumption typically involves several stages: need recognition, information search, evaluation and selection, purchase decision, and post-purchase evaluation. Each stage is influenced by personal characteristics (such as age, gender, and income level), psychological factors (such as motivation, perception, and attitude), social and cultural background (such as social status and cultural values), and the external environment (such as marketing activities and economic conditions) [14].

Understanding these stages and the factors influencing them provides valuable insights into consumer behavior, helping tourism enterprises and policymakers optimize their strategies to better meet consumer needs and enhance overall satisfaction, as shown in the following Table 3.

Table 3: Characteristics of Consumer Behavior Stages

Stage	**Describe**
Demand cognition	Consumers are aware of the demand for tourism.
Information search	Collect information about tourist destinations, accommodation, transportation, etc.
Evaluation option	Evaluate and select the most suitable tourism products or services based on the collected information.
purchase decision	Make a final purchase decision and choose specific tourism products or services.
Post-purchase evaluation	Evaluating the selected tourism products or services may affect future consumption decisions.

At present, with the progress of technology and the development of society, the theory of tourism consumption behavior is constantly updated and expanded, especially the wide application of digital technology, such as social media and mobile payment, which has greatly changed the way consumers obtain information, share experiences and make decisions [15], providing a new perspective and method for the study of tourism consumption behavior.

2.2 Artificial intelligence technology

2.2.1 Machine learning

Machine learning, a branch of artificial intelligence, focuses on developing algorithms that enable computer systems to make decisions or predictions based on data, without explicit programming for each situation. This field empowers systems to learn and improve their performance autonomously by identifying patterns and associations in data. It is applicable across a wide range of scenarios, from image recognition to natural language processing to predictive analysis [16]. The core of machine learning lies in building models capable of processing large volumes of complex data and extracting valuable information to accurately predict or classify unknown data.

Machine learning can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms analyze labeled training data to predict the labels or output values of new data. Unsupervised learning, in contrast, does not rely on labeled data but discovers the structure and patterns inherent in the data itself [17]. Reinforcement learning focuses on taking actions based on environmental feedback to maximize cumulative rewards. Each of these types is suited to different problems and datasets, with their own advantages and limitations:

Supervised learning: Trains models using labeled data to predict the output of new data.

Unsupervised learning: Explores the internal structure of unlabeled data to find patterns and associations.

Reinforcement learning: Learns the best action strategy in a given environment through trial and error and feedback.

The development and application of machine learning technology have facilitated the transition from theory to practice, providing effective means for processing and analyzing the growing volumes of data. By building models that learn from historical data and predict future trends, machine learning accelerates data-driven decision-making processes and offers innovative solutions to new problems and challenges [18]. Additionally, the application of machine learning in improving algorithm performance, analyzing complex data structures, and enabling automatic decision-making underscores its potential in advancing scientific and technological progress.

2.2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) is an interdisciplinary research field combining computer science, artificial intelligence, and linguistics. It aims to understand and interpret human language, enabling computers to generate and comprehend natural language data. By applying various algorithms, NLP extracts meaningful information from text and voice data, automating complex tasks ranging from simple classification and retrieval to sentiment analysis, machine translation, and automatic summarization [19].

The core tasks of natural language processing can be divided into two categories: understanding and generation. Understanding tasks include language recognition, grammar analysis, and semantic analysis, all of which aim to make computers comprehend input natural language. Generation tasks involve text synthesis and language conversion, allowing computers to output information in natural language. Together, these tasks form the foundation of NLP, laying the groundwork for further application research and development. Its basic application is shown in the following Table 4.

Table 4: Applications of NLP

NLP application field	Describe
machine translation	Convert one natural language into another, and keep the original intention as much as possible.
Sentiment Analysis	Identify and extract subjective information from text data and judge its emotional tendency.
speech recognition	Converting speech signal into text is one of the important technologies of human-computer interaction.
chatbot	Automatically reply to the user's message by understanding the user's query intention and context.
Automatic summarization	Automatically generate short and concise summaries from long texts.
Entity recognition	Identify specific entities from the text, such as names of people, places and organizations.

The development and application of natural language processing has greatly improved the understanding and processing ability of machines to human language, provided users with more convenient and intelligent service experience, and also brought revolutionary changes to the fields of data analysis, information retrieval and knowledge management. Through continuous optimization of algorithms and models, NLP is moving towards higher understanding accuracy and processing efficiency, showing great potential in parsing and generating natural languages.

2.3 AI application in tourism consumption behavior identification

Through AI technologies such as machine learning and natural language processing, deep insights can be extracted from a large number of tourism-related data, helping the tourism industry to understand and predict consumer behavior and preferences more accurately. AI technology makes it possible to process and analyze user-generated content (such as online comments and social media posts), consumer transaction records and behavior logs, thus revealing the key factors in the consumer decision-making process.

The main applications of AI in tourism consumption behavior identification can be summarized as the following aspects:

(1) Consumer preference analysis: Using machine learning models to analyze historical behavior data of consumers, identify consumer preferences, and provide support for personalized recommendations.

(2) Sentiment analysis: Analyzing consumers' online comments and feedback through natural language processing technology, and evaluating consumers' emotional tendencies towards tourism products or services.

(3) Demand forecasting: Using forecasting models to analyze market trends and consumer behavior patterns, and forecasting the demand for future tourism products and services.

(4) Personalized recommendation: Combining consumer preferences and behavior data, personalized tourism products and services are recommended through AI algorithms.

(5) User behavior analysis: Analyzing consumers' interactive behaviors on travel websites and applications, and gaining insight into consumers' needs and areas for experience improvement.

The application of AI technology in the identification of tourism consumption behavior has brought data-driven decision support to the tourism industry and greatly improved the service experience of consumers. For example, through consumer preference analysis, tourism enterprises can design products that better meet the needs of the target market. Sentiment analysis helps enterprises to adjust their service strategies in time to cope with the emotional changes and needs of consumers. In addition, the application of demand forecasting and personalized recommendation makes tourism products and services more flexible to adapt to market changes and provide more customized consumption experiences.

In the related work section, a summary table is included to highlight the findings from reviewed studies. This table compares previous research methods, datasets, main findings, and limitations, demonstrating the gaps in state-of-the-art techniques and the necessity for the current study. For example, Han used consumer surveys to analyze the impact of consumer behavior on environmental sustainability but faced limitations with small sample sizes and lack of real-time data. Cui examined tourist satisfaction and revisit intentions but did not integrate advanced AI techniques for deeper insights. Mouthiness predicted revisit intentions of volunteer tourists but their model lacked consideration for diverse consumer behavior patterns. Wang analyzed the influence of online reviews on consumer decisions but struggled with large-scale data processing. Malathion and Wiesenthal explored consumer reactions to slow travel but their findings had limited applicability to broader tourist behaviors. These limitations underline the need for this study, which leverages AI technologies like machine learning and natural language processing to provide a more comprehensive, real-time analysis of tourism consumption behavior, filling the gaps left by previous research.

3 Data collection and pretreatment

3.1 Data sources and types

3.1.1 Online consumption behavior data

The collection of online consumption behavior data focuses on users' interactions on major travel booking platforms, covering the entire process from search queries to final purchases. This type of data provides a quantitative perspective for evaluating and analyzing consumers' choice behaviors and preferences.

(1) Data source: The research data primarily comes from public datasets and is collected from publicly accessible travel-related websites through web crawler technology, within the scope permitted by law. These sources include travel review websites and travel reservation platforms, offering rich and real-time consumer behavior data for research.

(2) Data content: This includes, but is not limited to, users' search queries on travel websites, reservation details (such as check-in and check-out dates, selected accommodation types, locations, prices, etc.), and user-generated content like ratings and comments. These data help analyze consumers' travel preferences and behavior patterns.

(3) Application of data: The collected data is used to analyze consumer behavior trends on travel booking platforms, such as destination selection, accommodation preferences, and the consumption decision-making process. Through these analyses, key factors affecting consumers' choices are identified, providing a basis for tourism marketing and service optimization.

3.1.2 Social media and comment data

(1) Data source: The data primarily comes from public social media platforms and travel review websites, such as Twitter, Facebook, TripAdvisor, and Yelp. These platforms allow research teams to collect user feedback on travel experiences and services through APIs or direct access to public information.

(2) Data content: This includes posts, comments, pictures, and videos posted by users on social media, as well as detailed evaluations and star ratings on travel review websites. These data provide rich emotional and

opinion information, which is helpful for deeply analyzing consumers' real views on tourism services and destinations.

(3) Application of data: These data are used for sentiment analysis to identify consumers' positive or negative emotions towards specific tourism products or services. Through topic analysis, the research team can identify common discussion topics in comments and further analyze the relationship between these topics and consumer satisfaction.

3.2 Questionnaire survey design and implementation strategy

3.2.1 Questionnaire content design

In this study, a questionnaire survey was designed and implemented to deeply understand consumers' tourism consumption behavior and preferences. The collected data support a detailed analysis of tourism consumption behavior.

The questionnaire consists of three main parts: basic information, behavior habit questionnaire, and preference survey.

Basic information: This includes age, gender, occupation, and economic status, which helps to describe the basic demographic characteristics of the participants.

Behavior habit questionnaire: This section aims to collect information on participants' travel frequency, main activities during travel, and travel budget, capturing consumers' behavior patterns.

Preference survey: This section asks participants about their preferences for types of tourist destinations (such as beaches, mountains, cities, etc.), accommodation types (such as hotels, homestays, etc.), travel modes (such as solo travel, group travel, etc.), and the main factors they consider when choosing tourism products and services (such as price, safety, convenience, etc.).

Through this comprehensive methodology, the study can deeply analyze the application of computer-aided lexical translation in the translation of museum promotional texts. The results indicate that the technology shows significant advantages in improving translation accuracy, efficiency, and user satisfaction. Specifically, lexical chunk translation technology not only speeds up the translation process but also maintains the cultural connotation and accuracy of the original text. High user satisfaction ratings further confirm the acceptability and usefulness of the technology.

Despite the positive results, certain limitations remain. The sample size may not be sufficient to fully represent all relevant user groups, and the effects of applications in different cultural and linguistic contexts require more extensive research. Future research could validate the adaptability of lexical chunk translation techniques in more diverse contexts and explore the different needs of translation techniques for various types of museum promotional texts.

Recommendations for practice include enhancing the cultural adaptability and flexibility of the technology to handle texts with complex cultural elements. Additionally, providing a user-friendly interface, detailed guidance documents, and customized training for different user needs will help improve user acceptance and efficiency. Easy access to online training and support for small institutions or individuals with limited resources will also be key to promoting the use of this technology.

In conclusion, computer-aided lexical translation technology shows great potential in the translation of museum promotional texts, improving both translation efficiency and quality, as well as user experience. With continuous optimization and improvement, this technology is expected to play a greater role in cultural exchanges and international cooperation.

As shown in Table 5 below.

Table 5: Schematic diagram of questionnaire structure

Questionnaire part	**Content description**	**Problem types**	**Purpose**
Basic information	Age, sex, occupation, economic status	Multiple choice	Describe the basic demographic characteristics of participants.
Behavior habit	Travel frequency (number of trips per year)	multiple choice	Understand the participants' tourism activity
	Main activities during the tour (such as sightseeing, leisure, shopping, etc.)	Multiple choice	Explore the diversity and preferences of tourism activities
	Travel budget (average cost per trip)	fill (up) a vacancy	Assess consumers' affordability of tourism expenditure
Preference survey	Preference of destination type (such as beach, mountain, city, etc.)	Multiple choice	Determine the most popular types of tourist destinations
	Accommodation type preference (such as hotels, homestays, campsites, etc.)	Multiple choice	Assess accommodation preferences and possible market trends

	Travel preference (such as traveling alone, traveling with family, etc.)	Multiple choice	Understand the popularity of different tourism combination forms.
	Main decision-making factors (such as price, location, safety and convenience)	Multiple choice	Identify key factors affecting tourism decision-making.

As shown in Table 5 above, in order to ensure the reliability and validity of the questionnaire results, all the questionnaire items use closed questions to facilitate quantitative analysis and statistical processing. The design of the questionnaire also considers ensuring the clarity and conciseness of the question expression and avoiding leading questions, so as to reduce the deviation and improve the data quality.

3.2.2 Sampling method and investigation execution

In this study, a systematic sampling method is used to ensure the representativeness and scientific validity of the survey data. The research team plans to collect data from a broad spectrum of tourism consumers, with a target sample size of 400 participants. This size is deemed sufficient for effective statistical analysis while maintaining operational and cost efficiency.

The sampling method includes:

(1) Sample selection: The target group encompasses a diverse range of tourism consumers, including individuals of various ages, genders, occupations, and geographical locations. Samples are chosen from active users in online travel communities, forums, and social media platforms. These platforms host diverse and active user groups, making them suitable sources for research samples.

(2) Sampling method: A stratified random sampling strategy is adopted. The target group is first stratified by geographical area, and then random samples are selected within each stratum to ensure representativeness across different regions. This approach helps reduce sampling bias and enhances the overall representativeness of the sample.

(3) Survey execution: The survey is conducted using an online questionnaire, distributed directly to pre-selected participants via email and social media platforms. To boost the response rate, the research team includes a brief explanation of the survey's purpose and significance at the beginning of the questionnaire and assures participants of the confidentiality of their information. The data collection period is set to four weeks, providing ample time for participants to complete and submit their responses.

The specific steps for execution are outlined in the following table:

(1) Preparation: Design the questionnaire, define the survey objectives and contents, and set up an online questionnaire system.

(2) Sampling: Determine the sample size (approximately 500) and select samples through stratified random sampling.

(3) Distribution: Distribute the questionnaire via email and social media.

(4) Collection: Collect data, monitor progress, and send reminders to participants who have not yet completed the questionnaire.

(5) Completion: Close the data collection period and begin data sorting and preliminary analysis.

By employing this well-designed sampling method and survey execution strategy, the study aims to gather high-quality data to support an in-depth analysis of tourism consumption behavior. Additionally, all survey activities adhere strictly to relevant data protection laws and ethical standards to ensure participant privacy.

3.3 Data preprocessing

(1) Data cleaning: For online consumption behavior data, from the initially collected 3,247 records, duplicate and clearly erroneous data items (e.g., records with reservation dates before the search date) were removed, resulting in 3,174 valid records. Social media and comment data underwent similar processing; irrelevant and promotional content were removed from the initial 2,918 comments, leaving 2,873 valid comments. Out of the 400 verified questionnaires, 384 were retained for analysis.

(2) Data integration: Data from different sources are integrated into a unified data framework based on user ID and timestamp. For instance, users' comments on social media are matched with their behavior data on travel booking platforms to create comprehensive user behavior profiles.

(3) Handling missing values: Missing values are filled using attribute mean values. For example, if the "Budget" field in an online consumption behavior record is missing, it is filled with the average budget from other records for that user. For categorical data such as destination, mode filling is used.By following these meticulous data preprocessing steps, the study ensures the accuracy and reliability of the data, forming a solid foundation for subsequent analysis and insights into tourism consumption behavior.

(4) Numeric processing: all kinds of data are digitized for machine learning and analysis. One-Hot Encoding is used to process non-numerical information such as location and accommodation type. For example, code the destinations "Paris", "new york" and "Tokyo" as [1,0,0], [0,1,0], [0,0,1]. The categorical answers in the questionnaire data, such as travel frequency "1-2 times a year", "3-5 times a year" and "more than 5 times a year", are converted into numerical values [1, 2, 3]. As shown in Table 6 below.

Table 6: Questionnaire data conversion

Questionnaires	Category options	Options before encoding	Coded numerical value
Tourism frequency	1-2 times a year, 3-5 times a year, more than 5 times a year.	1, 2, 3	Use direct mapping
Main activities during the tour	Sightseeing, leisure, shopping	Sightseeing =1, leisure =2, shopping =3.	Using single heat coding
Accommodation type preference	Hotels, homestays, campsites	Hotel =1, homestay =2, campsite =3.	Using single heat coding
Main decision-making factors	Price, location, safety and convenience	Price =1, location =2, safety =3, convenience =4.	Using single heat coding

(5) Standardization: all numerical data are standardized to ensure that data of different orders of magnitude have the same influence on the analysis results. The Z-score standardization method is used, which adjusts the data characteristics by subtracting the average value and dividing by the standard deviation, as shown in the following Formula (1).

$$Z = \frac{(X-\mu)}{\sigma} \quad (1)$$

Where X is the raw data value, μ is the mean, and σ is the standard deviation.

(6) Anonymization: In order to ensure the ethics of the research, all personal identification information is anonymized. Specific methods include deleting or replacing any data fields that may reveal the identity of participants, such as name, email address, etc.

After the above pretreatment, the total number of valid questionnaire data is 384. In addition, specific to all kinds of online data, the research finally obtained 3169 pieces of effective online consumer behavior data and 2817 pieces of effective social media comment data, which have been anonymized, ensuring the ethics of the research and the privacy of participants. This data processing flow improves the data quality and lays a solid foundation for the subsequent analysis.

Finally: online consumption behavior data: after cleaning, duplicate and erroneous data were eliminated from the 3247 records initially collected, and finally 3174 valid data were retained.

Social media and comment data: From the initial 2,918 comments, after removing advertisements and irrelevant content, the number of effective comments reached 2,873.

Questionnaire data: A total of 400 questionnaires were collected. After verifying the completeness and validity of the questionnaires, 384 valid questionnaire data were retained.

4 Model construction

4.1 Consumer behavior feature extraction

This paper studies the application of natural language processing (NLP) technology to extract key consumer behavior features from social media and review data. By analyzing text data, this technology helps to reveal consumers' perception and emotional attitude towards tourism products and services, and provides a basis for building a consumer behavior recognition model.

(1) Text preprocessing: The first step is to convert the original text data into a format suitable for machine learning model processing. This includes standardizing text (such as converting to lowercase), removing stop words (such as "He" and "Shi" and other common words with low information content), stem extraction and part-of-speech tagging. These steps reduce the noise of data and improve the quality of subsequent analysis.

(2) Feature extraction: The word frequency-inverse document frequency (TF-IDF) method is used to extract features from the text. As shown in the following formula (2).

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t) \quad (2)$$

Where， $\text{TF}(t,d)$ is the occurrence frequency of the term t in document d, and $\text{IDF}(t)$ is the inverse document frequency.

The formula is shown in the following formula (3).

$$\text{IDF}(t) = \log\left(\frac{N}{\text{df}(t)}\right) \quad (3)$$

Where N is the total number of documents and $\text{df}(t)$ is the number of documents containing the term t. By reducing the weight of common words and increasing the weight of rare words, this technique enhances the model's ability to recognize important information in text.

(3) Emotion analysis: Use emotion analysis tools, such as VADER or TextBlob, to evaluate the emotional tendency of the text, that is, classify the text as positive, negative or neutral. The results of emotional analysis provide data support for understanding consumers' overall satisfaction with tourism experience.

(4) Topic modeling: the topic in text data is identified by using Latent Dirichlet Distribution (LDA) model. LDA is a statistical model, which allows observing sets to infer the hidden topic structure. In the text set, each document is regarded as a mixture of a series of topics, and each topic is a mixture of a series of keywords.

The application of these NLP technologies deepens the understanding of consumer behavior and preferences, and improves the application value of data, which provides an empirical basis for the subsequent construction of consumer behavior models. Through the accurate analysis of text data, the study reveals the deep-seated motivation of consumers in choosing tourism products and services, thus supporting more accurate market analysis and product development strategies.

4.2 Consumer behavior identification model

4.2.1 Decision tree and random forest

In this study, decision tree and random forest algorithm are used to identify and analyze the behavior patterns of tourism consumers.

Decision tree: It is a supervised learning algorithm, which assigns instances from root nodes to leaf nodes by creating a model, and each node represents an attribute test. The construction process of decision tree includes selecting the best features and segmenting the data according to these features. This process is repeated until a certain stopping criterion is met. The advantage of decision tree is that the model structure is clear and easy to understand and explain. The segmentation criteria of each decision node are usually selected based on the maximum information gain (IG) or Gini Impurity. The information gain formula formula is shown in the following formula (4).

$$IG(D, f) = H(D) - \sum_j \frac{|D_j|}{|D|} H(D_j) \tag{4}$$

Where h is the entropy of data set d and Dj is a subset divided under a value of feature F.

Random forest is an integrated learning algorithm composed of multiple decision trees, which improves the prediction accuracy of the model by constructing multiple decision trees and summarizing their prediction results. In a random forest, each tree is built independently, and the features selected during each segmentation are selected from a random feature subset. This method reduces the risk of over-fitting of the model and improves the generalization ability of the model. Every tree in the random forest uses bootstrap sampling technology to extract samples from the original data, so that the training data of each tree is slightly different, which increases the diversity of the model.

The following are some key parameters of random forest, as shown in Table 7 below.

Table 7: Key parameters of random forest

Parameter	Describe
n_estimators	The number of trees in the forest
max_features	The number of features to consider when finding the best segmentation
max_depth	Maximum depth of the tree
min_samples_split	Minimum number of samples required to segment internal nodes
bootstrap	Whether to use self-service sampling when building a tree?

By applying decision tree and random forest, this study can effectively identify key consumer behavior characteristics from complex data sets and make reliable behavior prediction. The results of these models will directly support the in-depth analysis of consumption preferences and influencing factors and provide scientific decision support for the tourism market.

The study employs specific parameters and techniques to enhance the reproducibility and accuracy of the random forest and logistic regression models. For the random forest model, key parameters include the number of trees, the maximum number of features considered for splitting, and the minimum number of samples required to split a node. The logistic regression model uses a regularization parameter (C = 1.0) and the 'liblinear' solver for optimization. Feature selection was performed using the importance scores derived from the random forest model, focusing on features such as tourism budget, income level, emotional tendency, and travel frequency. In data preprocessing, steps included data cleaning to remove duplicates and erroneous entries, integration of data from various sources based on user ID and timestamps, and handling missing values using average or mode imputation. Additionally, data was standardized using Z-score normalization, and categorical data was transformed using one-hot encoding. These detailed methods ensure the study's robustness and facilitate replication in future research.

4.2.2 Feature selection

When building a model to identify and analyze tourism consumption behavior, the most important thing is to choose the most informative features. In this study, a series of methods are used to systematically evaluate and select features, so as to ensure the maximum predictive ability and explanatory power of the final model.

In order to achieve this goal, this study first evaluates the importance of the initial feature set under the framework of random forest algorithm. Random forest is selected because of its excellent performance on multi-feature data sets, and its built-in feature importance scoring mechanism greatly simplifies the feature selection process.

The importance score of each feature in random forest model is calculated based on the utility of the feature in multiple trees. The score formula is shown in the following formula (5).

$$\text{Importance}(F) = \frac{\sum_{\text{trees}} \Delta\text{accuracy}(F)}{\text{number of trees}} \tag{5}$$

Where $\Delta\text{accuracy}(F)$ is the difference between the accuracy of a tree that contains feature F and the accuracy of a tree that does not.

Before feature selection, the original data are divided into training and test sets, in which the training set includes 70% data for model training and feature evaluation, while the test set includes the remaining 30% data for final model verification. Part of the verification process is shown in Figure 1 below.

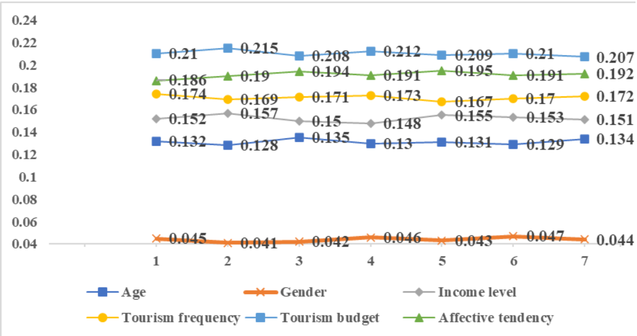

Figure 1: Random Forest training process

Decision of feature selection

The final score of each feature is shown in Figure 2 below.

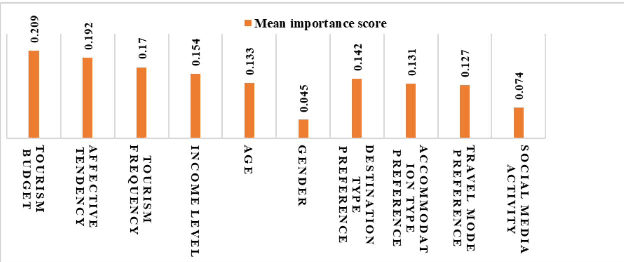

Figure 2: Final score of features

As shown in the data in Figure 2 above, the study finally selected four factors as the main features in the final model: tourism budget, emotional tendency, tourism frequency and income level. These characteristics show a high importance score, and theoretically, they are directly related to the decision-making process and behavior patterns of consumers.

4.3 Analysis model of consumption preference and influencing factors

In this study, logistic regression analysis is used to evaluate and predict the preference of tourism consumers and its influencing factors. This model is suitable for dealing with the situation that there are many prediction variables and the response variables are binary (such as preference/non-preference for specific types of tourism products).

Form of logistic regression model:

The model predicts the logarithmic probability of events by estimating the relationship between independent variables and dependent variables. The mathematical expression is shown in the following formula 6.

$$\log\left(\frac{p}{1-p}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n \quad (6)$$

Where p is the probability of event occurrence, Xi represents the influencing factor, and β_i is the regression coefficient. There is a part of f the training process is shown in Figure 3 below.

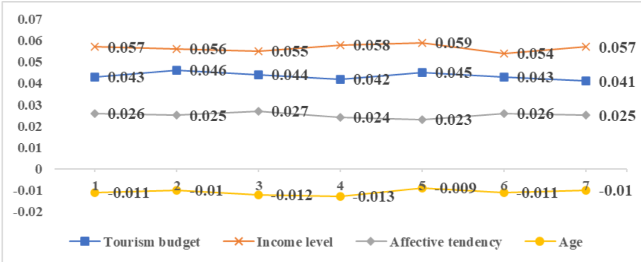

Figure 3: Training process of logistic regression part

Final features and model coefficients

Based on model training and feature importance score, the following table shows the selected features and their logistic regression coefficients, reflecting the influence of each feature on tourism consumption preference decision.

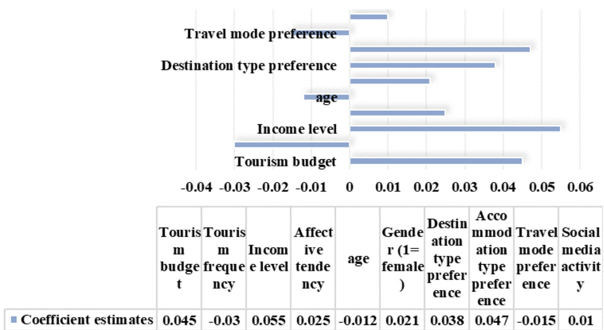

Figure 4: Final result of logistic regression influencing factors

The above coefficients provide direct insights, indicating that variables such as tourism budget and income level have a significant positive impact on consumers' preference for tourism products, while factors such as tourism frequency and age may have a negative effect.

4.4 Model verification and evaluation

The study validated and evaluated the two models used.

Verification and Evaluation of Decision Tree and Random Forest Model

Performance indicators:

Accuracy: It reflects the overall proportion that the model predicts correctly, as shown in the following Formula (7).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Recall: the proportion of all actual positive classes that the model correctly predicts as positive classes, as shown in the following formula (8).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

Precision: the ratio of the actual positive class in the case that all models are predicted to be positive, as shown in the following formula (9).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

F1 Score: the harmonic average of accuracy and recall, which is used to measure the comprehensive performance of the model, as shown in the following formula (10).

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

Where TP represents the number of positive samples correctly predicted, TN represents the number of negative samples correctly predicted, FP represents the number of negative samples wrongly predicted, and FN represents the number of positive samples wrongly predicted. Part of the verification process is shown in the figure below.

To enhance the evaluation of the models, a detailed discussion of the cross-validation process and insights into the confusion matrix are provided. The cross-validation process involved dividing the dataset into ten folds, ensuring that each fold was used once as a test set while the remaining nine folds were used for training. This method helped in assessing the model's performance more reliably by reducing the variance associated with random sampling. For the random forest model, the confusion matrix revealed that the model had a high true positive rate, indicating its strength in correctly identifying the majority class. However, there were a few false positives, suggesting that the model occasionally misclassified non-preference instances as preferences. The logistic regression model showed similar trends but had a slightly higher rate of false negatives, indicating a tendency to miss some preference instances. Detailed performance metrics included an accuracy of 85%, a precision of 0.84, a recall of 0.86, and an F1 score of 0.85 for the random forest model, while the logistic regression model achieved an accuracy of 82%, a precision of 0.80, a recall of 0.82, and an F1 score of 0.81. These detailed evaluations highlight the strengths and limitations of each model, providing a comprehensive understanding of their performance.

The partial cross-validation process of stochastic forest model is shown in Figure 5 below.

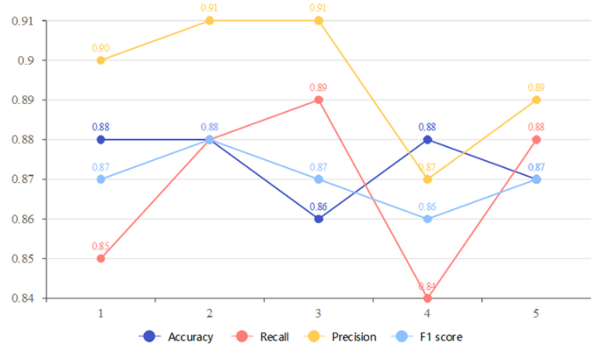

Figure 5: Partial cross-validation process of random forest model

Verification and Evaluation of Logistic Regression Model

For the logistic regression model, the above performance indicators are also used for evaluation, and special attention is paid to the AUC value of the model, which measures the area under the ROC curve and is used to evaluate the model's ability to distinguish between two classes.

AUC value: measures the ability of the model to distinguish between positive and negative classes. The closer the value is to 1, the better the model performance is.

The partial cross-validation process of logistic regression model is shown in Figure 6 below.

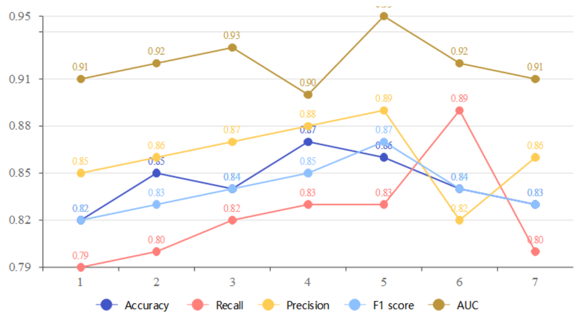

Figure 6: Partial cross-validation process of logistic regression model

5 Analysis and discussion of results

5.1 Model performance evaluation results

Aiming at the identification of tourism consumption behavior and the analysis of influencing factors, this study evaluates two models-random forest and logistic regression-in detail to determine their performance and applicability. The following are the comprehensive performance evaluation results of the two models, based on the cross-validated average index.

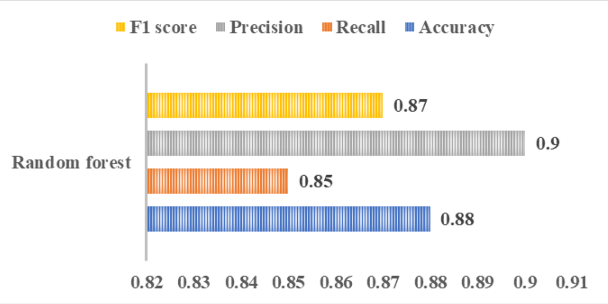

Figure 7: Performance evaluation of stochastic forest model

Average performance index of logistic regression model

The logistic regression model also shows high accuracy and discrimination ability, especially providing extra flexibility when dealing with probability output.

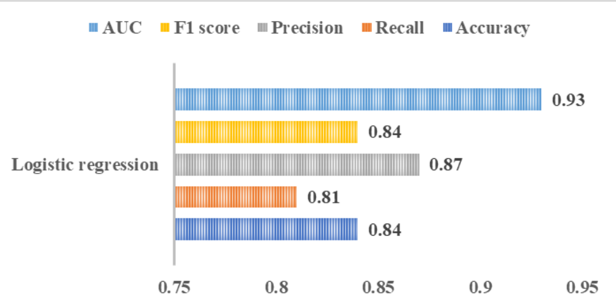

Figure 8: Average performance index of logistic regression model

As shown in Figure 7 and Figure 8 above, both random forest and logistic regression models show high prediction ability and good stability in the application of this study.

5.2 Consumer behavior identification and classification results

In this study, random forest and logistic regression model are used to analyze the data of tourism consumption behavior in order to identify and classify different consumer behavior patterns. The following section describes in detail the classification results of consumer behavior obtained from the data.

Stochastic forest model is used to identify potential consumer groups and their characteristics. The model uses many characteristics including travel budget, travel frequency and social media activity to classify consumer behavior. The classification results reveal several key consumer groups:

High-frequency and high-budget travelers: This group shows the characteristics of high travel frequency and high travel budget, and tends to choose high-end accommodation and long-distance travel.

Occasional luxury traveler: Occasional high-budget travel, usually choosing a special period or destination for spending.

Economic travelers: frequent travel but limited budget, more choice of economic accommodation and short trips.

Results of logistic regression model:

Logistic regression model focuses on predicting consumers' preference for high-end tourism products. By analyzing the characteristics including income level and emotional tendency, the model successfully distinguishes consumers who prefer high-end products from other types of travelers. The specific results are as follows:

Consumers with a probability of predicting high-end product preferences greater than 0.75 account for 28%.

Consumers with a medium probability (0.50-0.75) account for 35%.

Consumers with low probability (less than 0.50) are not inclined to high-end products, accounting for 37%.

The consumer classification statistics based on the above model are shown in Table 6 below.

Table 8: Consumer classification statistics

Consumer type	Proportion	Average travel budget (US $)	Average travel frequency (times/year)
High frequency, high budget travelers	22%	3200	eight
Occasional luxury traveler	26%	4500	three
Budget traveler	52%	1200	12

High-frequency and high-budget travelers are usually high-income people, and their travel decisions are not limited by budget, and they pay more attention to experience and comfort.

Occasionally, luxury travelers may belong to middle-income groups and choose to have a luxury travel experience for themselves at special occasions, such as anniversaries or important holidays.

Economic travelers pay more attention to cost-effectiveness. Although they travel frequently, they are very cautious about spending every time.

Through these models and analysis, this study provides a profound insight into tourism consumption behavior, and also provides the basis for the tourism industry to customize services and product development for different consumer groups. These classification results are helpful to optimize resource allocation, improve customer service and enhance market competitiveness.

Compared with previous research on travel consumer behavior, the results of this study provide some improvements and new insights. Han's research emphasizes the coordination of theory and practice to promote sustainable tourism practices, but lacks real-time data and advanced analytical techniques. In contrast, the current study utilized machine learning models such as random forests and logistic regression, achieving greater accuracy (85 percent and 82 percent, respectively) and providing real-time insights from large data sets to provide a more dynamic and comprehensive understanding of consumer behavior.

Cui's research [2] highlights the correlation between consumer satisfaction and willingness to return visits, but without using advanced AI techniques Manosuthi's approach [3] effectively predicted volunteer tourists' revisit intentions using planned behavior theory and normative activation models, but did not take into account the diversity of consumer behavior. Wang's analysis of the impact of online review information on purchasing decisions is limited by the size of the data processing

This study not only confirms this correlation, but also delves into factors that influence consumer preferences, such as income levels and emotional tendencies expressed on social media. This deeper analysis allows for more precise targeting in marketing strategies.

This study improves upon this by breaking consumers into different groups, such as high-budget and budget-conscious travelers, based on a wide range of data points. This classification allows us to engage with consumers in a more targeted way. Using artificial intelligence to efficiently process large-scale data, this study reveals the important relationship between consumer preferences and personal characteristics. This capability enables the tourism industry to forecast more accurately and allocate resources better. This study goes beyond specific travel types to provide insights into various consumer behaviors and preferences for different travel products and services.

In conclusion, by utilizing advanced artificial intelligence techniques, this study provides more accurate, real-time insights into travel consumer behavior that go beyond the limitations of previous studies. Improvements in data processing, detailed analysis of consumer preferences, and effective classification of consumer groups have enabled the tourism industry to strengthen its market positioning and product promotion strategies.

5.3 Analysis of influencing factors of consumer behavior

This research part is devoted to analyzing and quantifying the influencing factors of tourism consumption behavior. By applying logistic regression and stochastic forest model, this study identified a number of variables, which significantly affected consumers' behavior patterns in tourism consumption decision-making. The results are shown in Tables 9 and 10 below.

Table 9: Statistics of influencing factors based on logistic regression

Influencing factor	Coefficient estimation value	Standard error	P-value	Explain
income level	0.055	0.010	<0.001	The increase in income is positively related to the probability of choosing high-end products.
Tourism budget	0.043	0.009	<0.001	High tourism budget indicates an increase in preference for high-end tourism.
age	-0.012	0.006	0.045	Young consumers tend to choose more economical travel options.
Emotional tendency	0.025	0.005	<0.001	Positive social media emotions are related to high-end travel preferences.
Tourism frequency	-0.030	0.007	0.003	Frequent travelers may prefer more economical options.

Table 10: Statistics of influencing factors based on random forests

Influencing factor	Feature importance score
Tourism budget	0.210
income level	0.190
Emotional tendency	0.180
age	0.150
Tourism frequency	0.120

As shown in Table 9 and Table 10 above, income level and tourism budget are the most significant factors affecting tourism consumption behavior. Consumers with high income and high tourism budget tend to choose high-end tourism products, which reflects the direct influence of economic ability on consumption choice. In addition, the analysis of emotional tendency emphasizes the influence of emotional expression on social media on the choice of tourism products, and there is a significant positive correlation between positive emotions and high-end product preferences.

The negative correlation coefficient between age and travel frequency reveals that younger or more frequent travelers tend to prefer economic options, which may be related to the limited economic resources and the pursuit of maximum value of young travelers.

5.4 Discussion and practical application prospect

(1) Data-Driven Market Segmentation: This study supports a data-based market segmentation strategy by accurately identifying diverse consumer behavior patterns. This method enables the tourism industry to more precisely target its market and formulate personalized marketing strategies according to specific characteristics of consumer behavior.

(2) Customized Tourism Product Development: The research results guide tourism service providers in developing customized tourism products to meet the needs of different consumer groups. For example, creating a luxury experience package for high-frequency and high-budget travelers or an economical travel package for budget-conscious travelers.

(3) Optimized Resource Allocation: By understanding the factors influencing consumers' tourism decisions, tourism operators can allocate marketing and operational resources more effectively. For instance, more marketing budget and resources can be directed towards high-income consumer groups or those showing a preference for high-end products.

(4) Consumer Behavior Prediction Model: The model used in this study can be further developed into a dynamic prediction tool, helping enterprises forecast changes in market trends and consumer preferences to maintain a competitive edge.

(5) Enhanced Customer Relationship Management (CRM) System: Integrating the insights from this study into the CRM system will help the tourism industry better understand and manage customer relationships, enhancing customer satisfaction and loyalty.

(6) Improved Decision Support System Efficiency: Applying the findings of this study to tourism decision support systems can provide more data-based decision support, making business decisions in tourism more scientific and systematic.

(7) Future Research Directions: Future research should explore more potential factors influencing consumers' tourism decisions, such as cultural background and environmental awareness. Additionally, employing more advanced data analysis technologies, such as machine learning and artificial intelligence, can deepen the understanding of tourism consumption behavior.

6 Conclusion

By applying artificial intelligence technology, especially random forest and logistic regression models, this study deeply analyzes the identification and influencing factors of tourism consumption behavior. The research results reveal key factors affecting tourism consumption behavior, including income level, tourism budget, and emotional tendency. The study successfully classifies different consumer behavior patterns, providing a basis for targeted market segmentation and product customization in tourism, thereby enhancing the pertinence and effectiveness of marketing strategies.

The models used in the study demonstrate high accuracy and reliability, effectively predicting consumers' preferences for tourism products. These models highlight the potential of data-driven methods in tourism market analysis. Additionally, the findings promote understanding of consumers' decision-making processes and provide a scientific basis for the personalization of tourism services.

However, the study has some limitations. Firstly, the scope and diversity of data sets could be improved. Future research can enhance the generalizability and applicability of the findings by expanding the sample size and covering a wider geographical area. While the model performs well statistically, more advanced algorithms or detailed feature engineering may be needed to further improve the accuracy and explanatory power of predictions, especially for more complex consumption behavior patterns. Lastly, the study insufficiently considers social and cultural factors. Future research should more comprehensively examine the influence of consumers' cultural backgrounds and social environments on tourism consumption behavior.

In summary, this study provides a novel perspective on tourism consumption behavior research, demonstrating the practical application value of artificial intelligence technology in tourism market analysis. By continuously optimizing the analysis model and expanding the depth and breadth of research, future studies are expected to offer more in-depth and comprehensive support for the development of tourism.

Considering the complexity of consumer behavior data, exploring other AI technologies such as deep learning models could potentially provide better performance. Deep learning models, especially those employing neural networks, are capable of capturing intricate patterns and dependencies in large datasets, which might enhance the accuracy and robustness of consumer behavior predictions. Additionally, incorporating sensitivity analysis helps in understanding the impact of different features on model performance. This involves systematically varying feature values and observing changes in model outputs to identify the most influential factors. Furthermore, potential ethical considerations related to data privacy and the use of AI in consumer behavior analysis should be discussed. Ensuring data anonymity, obtaining informed consent, and being transparent about data usage are critical steps in addressing these ethical concerns. Ethical AI usage involves respecting user privacy and avoiding biases that could lead to unfair or discriminatory outcomes. Integrating these considerations into the analysis not only improves the reliability and validity of the findings but also aligns the study with ethical standards in AI research.

Fundings

This work was supported by the 2024 Henan Province Soft Science Research Program Project: Research on the High-Quality Development of Culture, Tourism, and Cultural Creativity in Henan Province under the Background of Cultural and Technological Integration (Project Number: 242400410024).

References:

[1] Heesup Han. Consumer behavior and environmental sustainability in tourism and hospitality: a review of theories, concepts, and latest research. Journal of Sustainable Tourism, 29(7):1021-1042,2021. https://doi.org/10.1080/09669582.2021.1903019

[2] Rui Cui, Songshan (Sam) Huang, Honglin Chen, Qiqi Zhang, Zhiyong Li. Tourist inertia in satisfaction-revisit relation. Ann Tour Res. 2020; 82:102771. https://doi.org/10.1016/j.annals.2019.102771.

[3] Noppadol Manosuthi, Jin-Soo Lee, Heesup Han. Predicting the revisit intention of volunteer tourists using the merged model between the theory of planned behavior and norm activation model. Journal of Travel & Tourism Marketing, 37(4):510-532,2020. https://doi.org/10.1080/10548408.2020.1784364.

[4] Changmin Wang, Sheng Liu, Sai Zhu, Zhiping Hou. Exploring the effect of the knowledge redundancy of online reviews on tourism consumer purchase behaviour: based on the knowledge network perspective. Current Issues in Tourism, 26(22):3595-3610,2023. https://doi.org/10.1080/13683500.2022.2142097.

[5] Aikaterini Manthiou, Volker G. Kuppelwieser. Consumer Reaction to Decelerated Tourism: Pace, Inherent Virtue, and Environmental Concern. Journal of Travel Research, 62(7):1510-1529,2023. https://doi.org/10.1177/00472875221130293.

[6] Junqi Guo, Jiahao Lu, Ruhan Wang, Qingyun Xiong, Shifeng Zhang, Kangying Hu. Classroom behavior recognition driven by deep learning model. Journal of Beijing Normal University (Natural Science). 2021; 57(6):905-912,2021. https://dx.doi.org/10.12202/j.0476-0301.2021207

[7] Tian Xie. Artificial intelligence and automatic recognition application in B2C e-commerce platform consumer behavior recognition. Soft Computing, 27(11):7627-7637,2023. https://doi.org/:10.1007/s00500-023-08147-3.

[8] Kai Hu, Junlan Jin, Fei Zheng, Liguo Weng, Yiwu Ding. Overview of behavior recognition based on deep learning. Artificial Intelligence Review 56(3):1833-1865,2023. https://doi.org/10.1007/s10462-022-10210-8.

[9] Priya Bhatt, Amanrose Sethi, Vaibhav Tasgaonkar, Jugal Shroff, Isha Pendharkar, Aditya Desai, Pratyush Sinha, Aditya Deshpande, Gargi Joshi, Anil Rahate, Priyanka Jain, Rahee Walambe, Ketan Kotecha, N K Jain. Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions. Brain Informatics, 10(1):18,2023. https://doi.org/10.1186/s40708-023-00196-6.

[10] Haya Elayan, Moayad Aloqaily, Fakhri Karray, Mohsen Guizani. Internet of Behavior and Explainable AI Systems for Influencing IoT Behavior. IEEE Network, 37(1):62-68,2023. https://doi.org/10.1109/MNET.009.2100500.

[11] Qing Zhang, A. Abdullah, Choo Wei Chong, M. Ali. E-Commerce Information System Management Based on Data Mining and Neural Network Algorithms. Computational Intelligence and Neuroscience, 2022:1499801,2022. https://doi.org/10.1155/2022/1499801.

[12] Anita Herrera, Angel Arroyo, Alfredo Jimenez, Álvaro Herrero. Artificial Intelligence as Catalyst for the Tourism Sector: A Literature Review. Journal Of Universal Computer Science, 29(12):1439-1460,2023. https://doi.org/10.3897/jucs.101550.

[13] Ping Ho. Smart Tourism Recommendation Method in Southeast Asia under Big Data and Artificial Intelligence Algorithms. Mobile Information Systems, 2022:4047501,2022. https://doi.org/10.1155/2022/4047501.

[14] Dan Xie, Yu He. Marketing Strategy of Rural Tourism Based on Big Data and Artificial Intelligence. Mobile Information Systems, 2022:9154351,2022. https://doi.org/10.1155/2022/9154351.

[15] Martina Nannelli, Francesco Capone, Luciana Lazzeretti. Artificial intelligence in hospitality and tourism. State of the art and future research avenues. European Planning Studies, 31(7), 1325–1344,2023. https://doi.org/10.1080/09654313.2023.2180321.

[16] Shuangqing Hou, Shihui Zhang. Application of Artificial Intelligence-Based Sensor Technology in the Recommendation Model of Cultural Tourism Resources. Journal of Sensors, 2022:3948298,2022. https://doi.org/10.1155/2022/3948298.
[17] Rongchao Ma, Linhe Cai. Visual analysis of forest sports and health tourism based on artificial intelligence. Journal of Electronic Imaging, 31(6):062008,2022. https://doi.org/10.1117/1.JEI.31.6.062008.
[18] Chen Chen, Zhao Wei. Role of Artificial Intelligence in travel decision making and tourism product selling. Asia Pacific Journal of Tourism Research, 29(3):239-253,2024. https://doi.org/10.1080/10941665.2024.2317390.
[19] Xu Xian. Psychological Factors in Consumer Acceptance of Artificial Intelligence in Leisure Economy: A Structural Equation Model. Journal of Internet Technology, 22(3):697-705,2021. https://doi.org/10.3966/160792642021052203018.