

Article

Framework for Building Smart Tourism Big Data Mining Model for Sustainable Development

Ruoran Xu

School of Event and Economic Management, Shanghai Institute of Tourism, Shanghai 201418, China; xuruoran@shnu.edu.cn

Abstract: How to combine big data (BD) technology with specific applications in the tourism industry to achieve sustainable development in the tourism industry is a development issue that needs to be addressed in the tourism industry today. In order to promote the development of smart tourism, this text constructed a BD mining model for sustainable smart tourism. In this paper, based on tourism data from 2010 to 2021, a regression model and an exponential curve model are constructed to forecast passenger traffic, and a tourism spatial dimension model is constructed to build a tourism data table, pre-process the data and construct a data mining (DM) model using a SQL Server model. The experimental part of the study conducts experimental research on cities applying smart tourism DM technology in three areas: foreign exchange earnings from the city's tourism industry, jobs in the tourism industry and the development of tourism-related industries. The results showed that the application of smart tourism DM technology can improve the foreign exchange income (FEI) of urban tourism, increase employment in tourism and drive the development of tourism-related industries. Compared with 2010, the tourism FEI of the four cities would increase by more than 70% in 2021.

Keywords: smart tourism; data mining model; passenger flow forecast; development of tourism industry



check for updates

Citation: Xu, R. Framework for Building Smart Tourism Big Data Mining Model for Sustainable Development. *Sustainability* **2023**, *15*, 5162. <https://doi.org/10.3390/su15065162>

Academic Editor: Fadi Al-Turjman

Received: 13 January 2023

Revised: 6 March 2023

Accepted: 9 March 2023

Published: 14 March 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background and Significance

As a product of the development of modern technology, smart tourism has many impacts on all sectors of the tourism industry. For tourists, smart tourism can provide new tourism information services, destination information, consultation and the option to book tourism products without leaving home, which can greatly improve the service level of the tourism industry. For tourism destinations and other tourism operators, smart tourism enables them to promote tourism products for online marketing, attract tourists and conduct intelligent management of tourism destinations [1].

Data are an important element in building a smart tourism system because data support the construction of the service part of the smart tourism information system. The detailed analysis and DM of the tourism industry would improve the economic benefits of the tourism industry and increase the FEI and employment opportunities of the tourism industry. It finds problems and appropriate corrective measures according to the results of data analysis and mining and plays an important role in promoting the development of related industries [2].

1.2. Status

Smart tourism as a new field of research has been studied by a larger number of scholars. Shafiee used the steps of grounded theory as the analysis framework and proposed a new model of intelligent tourism destination. Jovicic, Dobrica Z. reviewed the evolution of the concept of major tourism destinations, with special emphasis on the concept of intelligent tourism destinations [3]. Ghorbani, Amir selected the intelligent dimensions in tourism organizations according to the situation of the tourism industry and their impact

on the development of intelligent tourism organizations [4]. Mehraliyev, Fuad combined qualitative and quantitative validation methods to determine the latest case study trends and identify the knowledge areas of intelligent tourism research [5]. Jasrotia, Aruditya discussed the two buzzwords of “smart city” and “smart tourism destination” and the relationship between them [6]. Del Vecchio and Pasquale showed how the massive social traffic BD provided by tourists can cultivate the value creation process of intelligent tourism destinations [7]. Current research on smart tourism is focused on the service and user side; while, from the data perspective, the concept and technology of BD is proposed, but the collection, processing, selection and integration of data is not addressed.

BD mining is used in tourism. Ardito, Lorenzo studied the role of BD in intelligent tourism [8]. Li Jingjing attempted to conduct a comprehensive literature review of different types of BD in tourism research [9]. Iorio, Carmela introduced a conceptual model digital tourism system which can handle various types of standard and non-standard tourism data [10]. Li Daming built a BD platform on account of tourist flow information by analyzing the spatial and temporal distribution characteristics of tourist flow in the scenic area. He proposed DM technology based on the dragonfly algorithm and hybrid kernel relevance vector machine algorithm to predict tourist flow in the dimension of space–time distribution [11]. Al Farani, Khalid used BD and AI technology to propose an architecture and conceptual framework of a tourism recommendation system in view of hybrid recommendation methods [12]. Liu Jun quantitatively measured the impact of climate change on hiking in 100 cities by using a mixed method including the generalized additive model and the piecewise regression model to analyze the BD generated by tourists [13]. Alaei, Ali Reza reviewed and evaluated the application of different sentiment analysis methods in the tourism industry, the use of datasets and their performance [14]. Rahmadian, Eko has a comprehensive understanding of the application of BD in sustainable tourism to solve various problems, and how BD supports decision-making in this case [15]. The current research on tourism DM focuses on the construction level of information systems, with corresponding integration studies for the construction of specific tourism GIS.

While the sustainable development of tourism has opened up many new opportunities for growth, at the same time, there are many uncertainties due to the broad scope of the industry, and problems such as overpopulation in some popular destinations and unattendedness in other emerging destinations persist. At the same time, due to the overall lack of macro control of the tourism industry, the quality of tourism services has declined, tourism-related lawsuits have increased significantly and serious safety issues have emerged. In addition to these challenges, the intelligent analysis of tourism information and the modelling and forecasting of tourism flows in existing tourism management systems are not sufficient to meet the growing demands in tourism management.

1.3. Contents

Based on existing research, this paper improves and refines tourism information management, creates a BD mining model to better predict and manage tourism data, improves the quality of tourism services and promotes the development of the tourism industry [16]. The structure of this paper includes an introduction to basic theory, tourism DM model creation and experimental validation. We first introduce the research on smart tourism and BD mining for sustainable development, and then construct a passenger flow prediction model. We then construct a tourism spatial dimension model and establish tourism data tables such as tourism, tourists, accommodation and entertainment. Finally, we establish the DM model, including data preprocessing, model building, dimension building and application, and DM model. The experimental part investigates and studies the data of cities after applying the smart tourism DM technology.

2. Sustainable Smart Tourism

Intelligent tourism is a very new topic in the tourism industry. It realizes the innovation of tourism services, continuously improves the value of tourism products and reduces the comprehensive cost of tourism products.

In addition to developing smart cities, the tourism administration also advocates smart tourism as the future goal of tourism. At present, tourism information services make tourism more intelligent through the integration and dissemination of new media and mass media. At the same time, it optimizes various management, services and marketing links to achieve high-level communication and quickly and effectively disseminate tourism information, which can provide tourists with high-quality and satisfactory services.

Smart tourism includes three levels: public service platform, application level and infrastructure. Public service platforms directly or indirectly provide services to tourists in the form of units or individuals, such as public management services, related services, consulting companies and tourism companies. Application level users are mainly travelers with Internet connection devices, including ultra-portable Internet terminals (such as tablets and smartphones). Infrastructure includes basic hardware, software and databases. On the basis of these hardware, there is a data center containing various travel information and a special organization responsible for managing and updating data [17]. The structure of smart tourism is displayed in Figure 1.

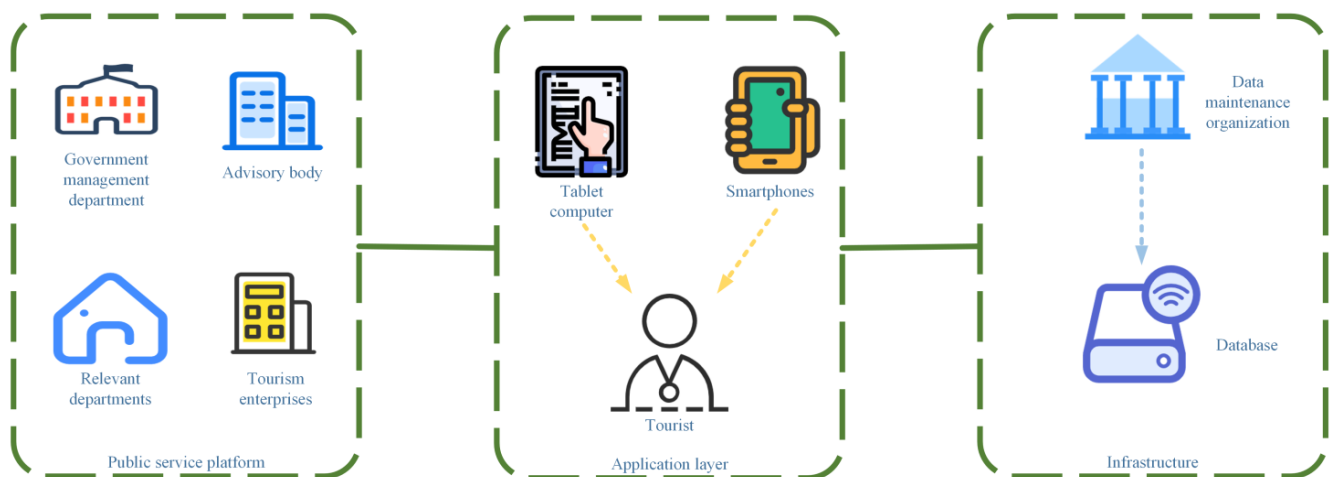


Figure 1. Smart tourism structure.

Smart tourism is embodied in product development and marketing, management and service, service feedback and other fields. By creating high-quality tourism products, people can better meet the needs of tourists and create greater environmental, cultural and social values. This can create greater economic value and support the sustainable development of the destination. Precise marketing makes the value of tourism products more obvious so that tourists can quickly and accurately obtain information about the value of tourism products and organize and display scientific information so that they can better plan travel and make decisions. It improves the management of enterprises and reduces the operating costs of enterprises and the economic costs of tourists. Through real-time communication and information sharing, it makes the tourism industry process smoother; tourists can easily review and enjoy the new experience of intelligent tourism, and operators can improve their ability of scientific analysis and feedback. Targeted improvements can be made to create tourism products that better meet the needs of tourists.

As a new industry, tourism is developing rapidly in modern society. In recent years, smart tourism has also been developing steadily, becoming extremely important to local tourism operators and tourism management departments. However, some tourism enterprises did not fully realize this and did not pay enough attention to it. Tourism enterprises need to know how to keep pace with the times and make use of appropriate opportunities

for change. Through smart tourism, tourism enterprises can analyze tourists' preferences and needs through relevant data. It determines target customer groups, provides personalized services that meet the needs of tourists, significantly improves service efficiency, reduces marketing costs and makes scientific decisions. Of course, tourism management departments would also be able to use the real-time information provided by intelligent tourism terminals and other services. It effectively implements appropriate deviation or remedial measures, and responds to tourists' complaints in a timely manner, providing a solid foundation for correct and appropriate actions.

Trains, planes and other means of transportation have security problems; because there are many people on them, which would attract the attention of some extremists. The intelligent travel transportation system provides detailed travel road information, and provides routes in case of traffic congestion in a timely manner. The system can also be used to select hotels, restaurants, entertainment programs, etc., through the intelligent travel mobile terminal. At the same time, tourists can plan their own walking routes according to their interest points and use information services, such as travel guides and shopping guides, when arriving at a beautiful place.

Of course, intelligent travel also promotes student travel by reducing barriers, increasing the amount of information and eliminating closures. When planning a trip, people can use their mobile devices to find the best itinerary, affordable accommodation, local snacks, famous tourist attractions, etc.

Smart travel provides people with complete travel services, enabling them to share travel experiences more widely, safely and in a timelier manner. Of course, if an accident occurs during travel, people can also control it through their smartphone. Smart mobile devices, including tourists and guides, organically link tourism enterprises with tourism management departments to improve people's tourism experience. It has also promoted the transformation and modernization of traditional tourism, so as to improve the international competitiveness and sustainable development of tourism.

3. Big Data Mining

3.1. Overview of DM

DM is a hot topic in the field of artificial intelligence and BD, which is also known as database development. DM is an important process by which to analyze and calculate a large amount of data in a database, and then discover previously hidden unknown data. DM itself has no clear goal. The results of data analysis are uncertain and can predict the future direction, which would have a profound impact on the direction of e-commerce.

DM is a new enterprise information processing technology that detects hidden, unrecognized or confirmed patterns, and provides better and effective modeling methods.

Because the computing performance of the computer was relatively low at that time, the analysis and processing efficiency of complex data was relatively low. Today, due to business automation in various industries, enterprise departments have produced a large quantity of unknown data, including hidden data. These data are no longer used for analysis but are the result of purely opportunistic business activities [18].

3.2. Common Methods of DM

Common methods of DM are displayed in Figure 2.

The classification method is mainly used to identify the common features of data object groups or groups in the database and classify them according to the specific classification model. Auto dealers can divide customers into different categories according to their car preferences, and then marketing personnel can send users advertising guides for different types of new cars according to their preferences, which greatly increases business opportunities.

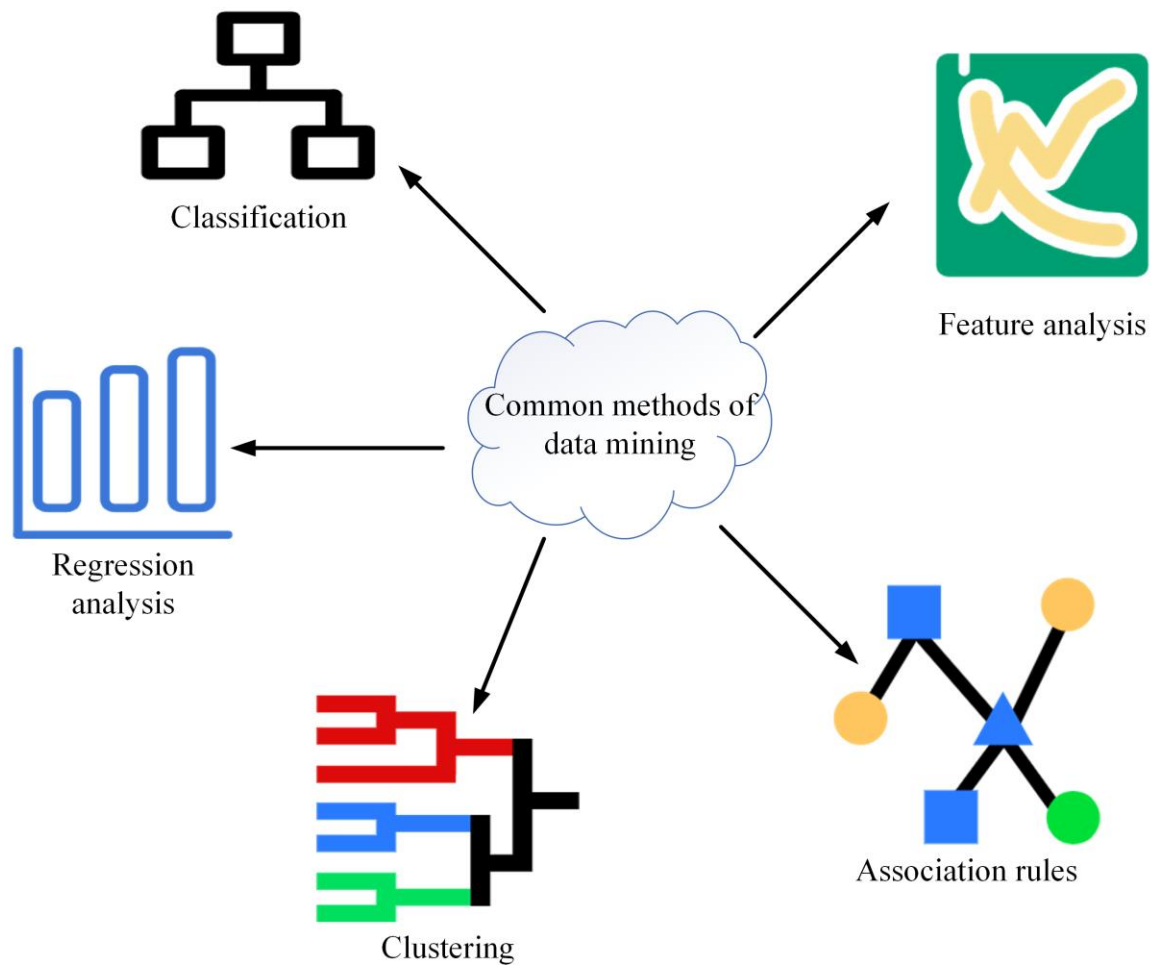


Figure 2. Common methods of DM.

Regression analysis mainly reflects the attributes of some attribute values in the database. It maps the generated columns to functions that use actual values to predict variables and verifies the correlation between variables or attributes.

The cluster analysis method includes classifying datasets into multiple categories according to differences and similarities.

The related rule method is a rule that describes the relationship between data items in the database; that is, other items of the same transaction, hidden relationships in the data or relationships between them can be exported according to the appearance of specific transaction items.

The feature analysis method is to extract the feature expression of these data from the dataset. These data are common features in the dataset. For example, marketing personnel can extract the key features of customer churn, determine the main reasons and characteristics of customer churn and prevent customer churn due to these reasons and characteristics.

Deviation contains much potentially interesting knowledge, such as classification exceptions, exception models and expected deviations. Unexpected rules often cover up huge benefits. Once the potential value of these exceptional rules is found, countless benefits would be obtained [19].

3.3. Main Technologies of DM

Traditional technology and advanced technology are usually two independent components of DM theory and technology. The object of DM is usually a variable with a large number of samples, which is used to simplify and modify the multivariate analysis

involved in advanced statistics. In particular, discriminant analysis using factor analysis for classification and cluster analysis using group division are commonly used in DM. The most commonly used DM methods are displayed in Figure 3.

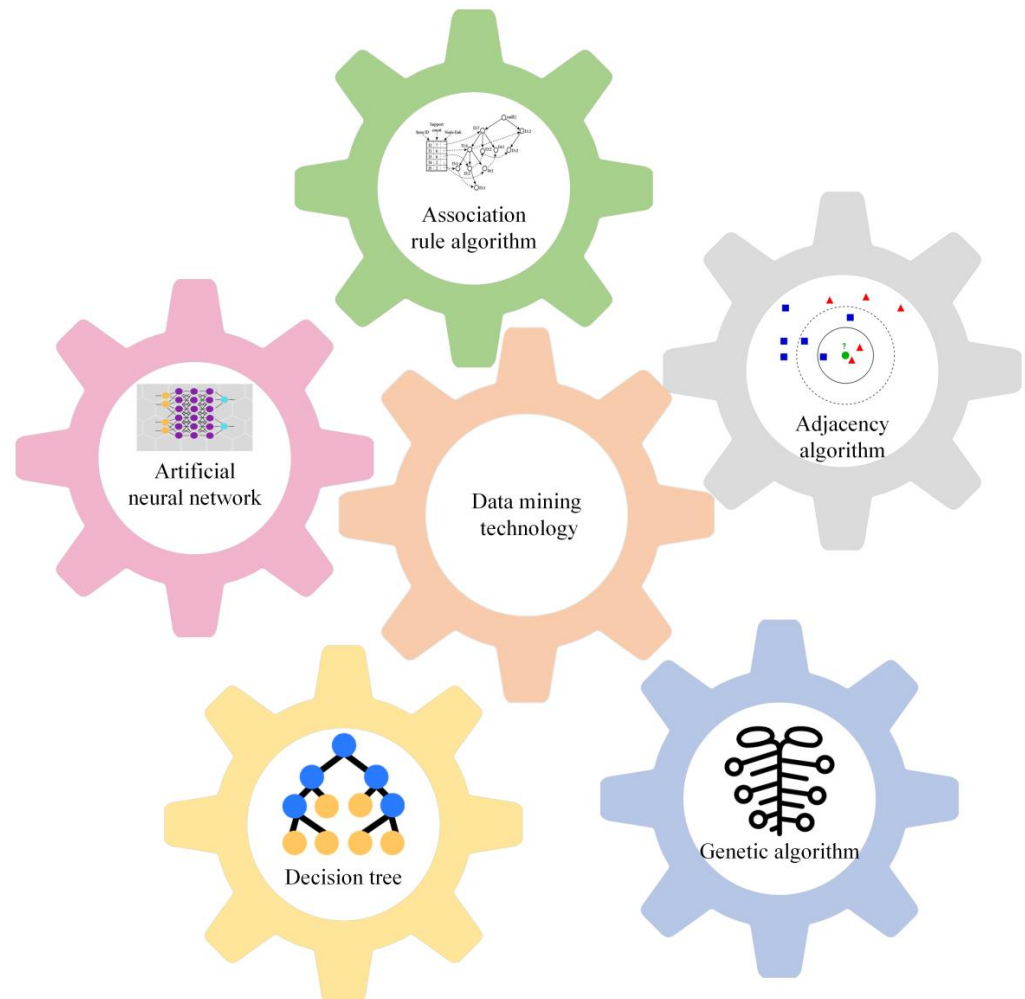


Figure 3. Common DM techniques.

Data mining technology mainly includes artificial neural network, decision tree, genetic algorithm, neighborhood algorithm and association rule algorithm. This paper selects the association rule algorithm for data mining.

Association rules are usually applicable to physical stores or e-commerce systems, mainly reflecting the relationship and dependency between things. We search the customer's purchase history or browsing records, finally understand the customer's habits by mining association rules, and obtain the similarity between related customers, such as the possibility of purchasing A and B products at the same time. By mining, adjusting the layout and designing perfect advertisements, we can ultimately increase the turnover of goods.

Association algorithm is an important algorithm in DM, and the typical algorithm is the Apriori algorithm. The first step of the Apriori algorithm is to search all common items in the transaction database using an iterative method. In other words, it can be understood that the collection is not less than the collection of all users. The second step is to use a set of public items to create rules that meet the user's minimum trust. Among them, viewing or identifying all common element sets is the core of the algorithm, which considers many aspects of the entire calculation [20].

3.4. Data Collection Techniques

Data collection includes three methods: on-site surveys, interview surveys and web-based collection. On-site collection includes GPS data collection, 3D panorama collection, questionnaire collection and visitor registration; methods of collecting data on the web include collecting official data and crawler technology data collection.

Web crawlers automatically search and access online information according to certain rules and are widely used in web search engines (or scripts). In addition to default text and images, web pages searched through browsers often have hyperlinks to information on other web pages, which allows web crawlers to access large amounts of data. In addition, as the Java programming language is cross-platform, Java web crawlers are well scalable and play an important role in search engine development [21].

Before preparing the design, the website can be browsed through a browser to view the HTML page elements to be found, import HTML tag elements and find valuable data. We store the successful crawl in a database, upload the content to a local database, filter non-important data and filter the most important data to quickly view important information [22].

3.5. DM Process

The process of DM is very complex, involving many operations. The DM process is displayed in Figure 4.

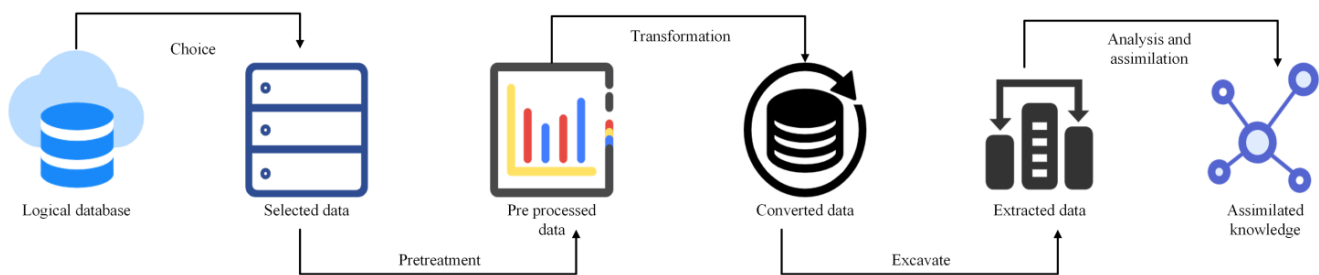


Figure 4. DM process.

The first step is to identify business objects. The results of data analysis are unpredictable, but the problems to be studied are very obvious. If people blindly analyze the data, the results would not be successful. We need to analyze the data quality and then prepare for the subsequent process to ensure the quality of data analysis results. This requires the description and evaluation of DM results. The analysis methods used are usually defined by DM operations, usually using current visualization technologies.

This chapter introduces the knowledge of DM, explains in detail the common methods of DM, as well as the current main techniques of DM, and finally makes a detailed introduction to the process of DM. Through the analysis of various techniques of DM, the association algorithm is finally selected as the DM method in this paper and applied in the next chapter.

4. Big Data Mining Model for Smart Tourism

The most pressing issue in the planning of tourism is how to effectively combine the spatial distribution of tourist attractions with the structure of their distribution in accordance with the consumer's desire to travel. The data provided in this paper come from the Tourism Statistical Yearbook, which is available in the relevant database on the website of the Bureau of Statistics. Crawler software will also be used to obtain tourism information from the official tourism website, including the name, rating and user reviews of a particular attraction. The user's comments are used to determine whether the user is a local order or an off-site order, and the season in which the user travels is distinguished by the timing of the comments.

4.1. Passenger Flow Prediction Model

4.1.1. Data Source

In view of the tourism industry data from 2010 to 2021, this study analyzed the changes and trends of the number of urban tourists over the years and used mathematical models to predict the number of tourists in the next few years. All the data provided in this study were from the Tourism Statistics Yearbook and can be viewed in the relevant databases on the website of the Statistics Bureau.

4.1.2. Regression Model Prediction

Assuming that the number of tourists in a city is increasing every year, the linear relationship can be used to describe the proportion of the number of tourists to a year. If the number of tourists follows the normal distribution, the regression model is as follows:

$$\hat{y} = a + bt. \quad (1)$$

In Formula (1), \hat{y} represents the predicted value and t is the time scale; a is the intercept of the regression line on the vertical axis and b is the regression coefficient, all of which are uncertain parameters.

a and b can be calculated by the least square:

$$b = \frac{n \sum t_i y_i - \sum t_i \sum y_i}{n \sum t_i^2 - (\sum t_i)^2}, \quad (2)$$

$$a = \frac{\sum y_i - b \sum t_i}{n}. \quad (3)$$

4.1.3. Exponential Curve Model

Based on these data, an exponential curve prediction model is established to predict the number of urban tourists.

The exponential curve equation is as follows:

$$\hat{y}_t = ab^t. \quad (4)$$

\hat{y} is the predicted value, a and b are unknown parameters and t is the time scale. If we take the logarithm at both ends of the formula and convert the variable, the converted linear equation is as follows:

$$\lg(y) = \lg(a) + \lg(b) * t. \quad (5)$$

$\lg(a)$ and $\lg(b)$ are calculated as follows:

$$\lg(b) = \frac{n \sum t_i \lg y_i - \sum t_i \sum \lg y_i}{n \sum t_i^2 - (\sum t_i)^2}, \quad (6)$$

$$\lg(a) = \frac{\sum \lg y_i - \sum t_i * \lg b}{n}. \quad (7)$$

4.2. Tourism Information DM

These data are analyzed and applied by enterprises to formulate policies and strategies and help formulate appropriate policies and programs. Tourism BD is the extraction and processing of such tourism-related information. This text analyzed the trends of different tourism sectors through in-depth research, including the development of scientific and reasonable management and operation models to achieve maximum social and economic benefits.

4.2.1. Tourism Spatial Dimension Model

The importance of tourism behavior lies in the fact that people travel or play in specific spatial areas. Coordinates are used to represent tourist attractions visited by tourists, and lines are used to represent tourist routes. Different tourists show different spatial choices when traveling. Figure 5 shows the different behavior patterns calculated by the tourism industry prediction system.

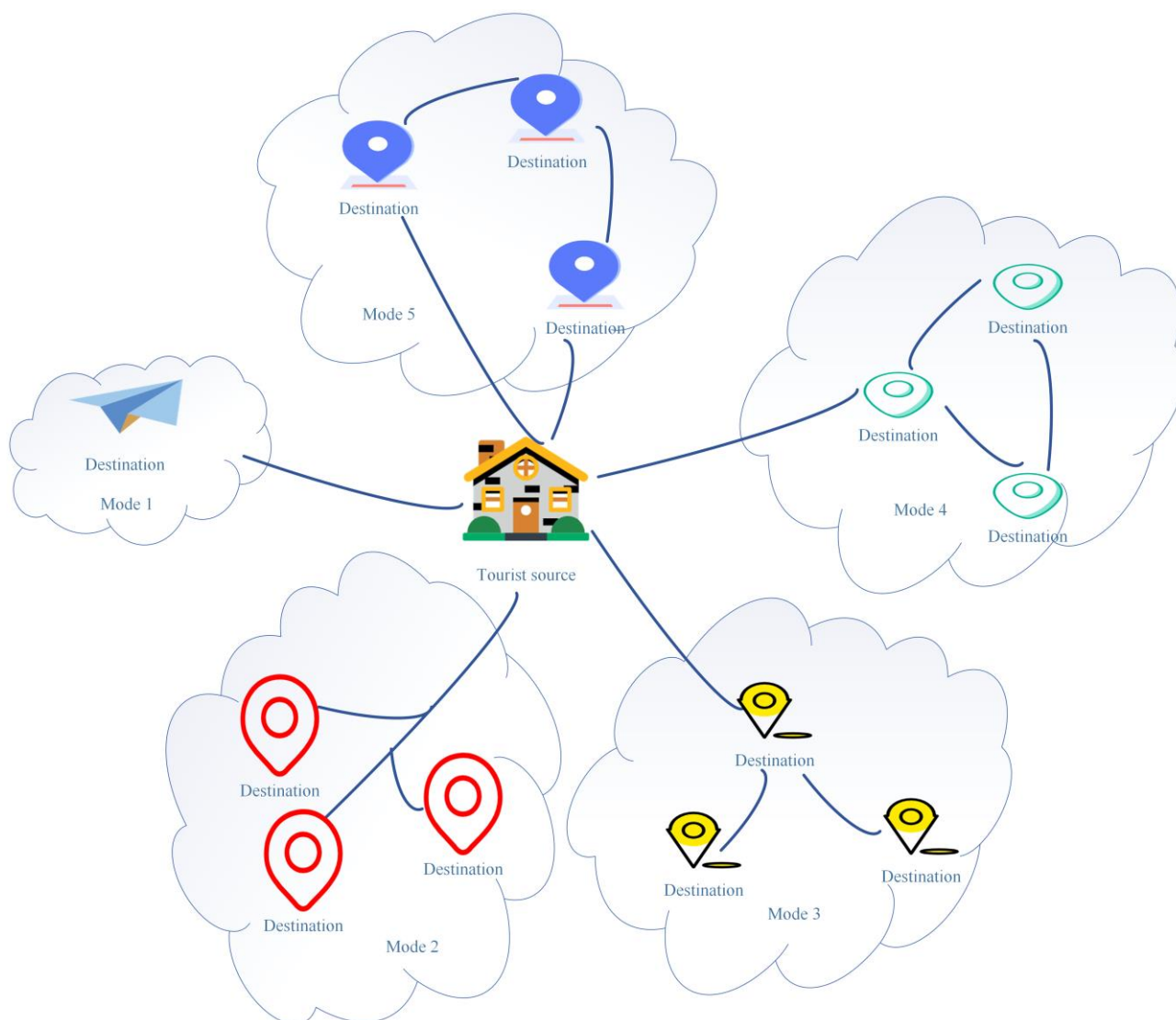


Figure 5. Tourism behavior mode.

The first is one-way tourism, which is the behavior choice of tourists for tourism activities. The second is linear tourism. Tourists look for multiple destinations on one route, but only choose one main destination. The third is the basic tourism. Although tourists have primary destinations, they also have secondary destinations, and finally reach the main destination. The fourth is circular tourism. Tourists divide the area into target areas to explore the tourist attractions and destinations in the area. The fifth type is tourism chain, which is a customer centered tourism network.

Spatial data are data that reflect the location, shape, size and characteristics of spatial unit distribution. They can be used to describe the real world goals characterized by location, quality, time and spatial relations. Spatial data form the basic structure of natural data, such as points and lines. Spatial data coding refers to the realization of graphic data,

images and spatial data structures. Each data source has a specific data structure, and the efficiency of data processing is often determined by the data structure.

4.2.2. Establishment of Tourism Data Table

In order to maintain resources, the database resources of the intelligent tourism information system include tourism, travel, food, shopping, entertainment, life, etc., as well as various digital resources according to the design principles of independent, safe, complete, standardized databases.

4.2.3. DM Model Establishment

A. Data preprocessing

The goal of data integration is to integrate data into data sources to eliminate semantic differences and store them in an integrated data format. Data specifications are identifiers of required datasets.

There are many reasons for data loss, so data needs to be filled in during preprocessing. Based on the existing information, some methods can be used to export the lost data. Even though it is relatively simple to replace global variables, it is usually meaningless. If people populate a numeric attribute with an average attribute value, there would be many averages, which may be too many. Attribute values of the same fill type based on all samples have advantages, but if the same type has the highest probability range value, special treatment is required. The other method is based on multi linear regression of the Bayesian derivative method or the decision tree method to determine possible values to fill in enough values.

B. Model establishment

The data storage is based on preprocessing, and the analysis model is created using Microsoft SQL Server 2000 Analysis SC Services.

The specific process of building the data mining model is as follows.

1. Organizing the data mining source data: templates are selected based on request topics, then event tables and size code tables are developed, using stellar and snowflake patterns. This relational database is not used to handle errors, but to prepare raw data for data mining. Once built, the corresponding indices are created by developing fact and size code tables. The data transformation service downloads data from different databases, retrieving and downloading data by extraction and normalization. When data are imported or exported, data packages are created to transform and load heterogeneous data for extraction needs;
2. Create data cubes: virtual databases are created for the objects stored in the service tree. The base table is defined as the index table for the multidimensional data cube. The index table above specifies how the indices are created, and the methods for creating the indices are called node attribute types and node attribute parameters. Determine the measurement levels and relationships specified in the multidimensional data cube measurement table, using the multidimensional data cube to define the tree structure. Determine the actual structural relationships of the multidimensional data cube and load the multidimensional data cube with data;
3. Build a server-side data mining model;
4. Association rules are applicable to one or more dimensions. Dimensions are the structural characteristics of data. They describe the structural hierarchy of data classification in data tables. Due to the decentralized data distribution, many applications find it difficult to find rules related to data details. After the concept hierarchy is introduced, it can be decomposed at a higher level. Higher level rules can be more general information; they can be shared by one user but may not be shared by another user. DM needs to provide these mining functions at multiple levels. The mining of multidimensional association rules can basically follow the framework of support and trust. There are two strategies to support association rules at one level.

The first is the unified minimum support. It uses the same minimum support at different levels. This is relatively easy for users and algorithms, but the disadvantages are also obvious.

The second is to gradually reduce support. Each level has a different minimum support, and the lower level has a lower minimum support. At the same time, the information that can be obtained from higher mining can also be used for filtering. Considering the minimal support for hierarchical allocation rules, the solution must be based on basic minimum support.

C. Establishment and application of dimensions

It determines the basic structure of the model and constructs the dimensions and cubes of system analysis. The dimension is the perspective from which people observe the real world. Decision analysis needs to observe and analyze data from different perspectives. The survey wizard would specify additional configurations for the data table to support cubes. The wizard prompts people to select a storage mode. After the wizard completes its task, people would return to the wizard to complete the design process. It can also improve performance by optimizing the data storage in the storage design wizard. In the tree directory in the left pane of the analysis manager below, people can select cube roles to create security features.

The dimensions defined here include age, region, gender, landscape, time and transportation.

The age range is divided into 15 years old, 15–30 years old, 30–40 years old, 40–50 years old, 50–70 years old and over 70 years old.

The regions are divided into China, North America, Latin America, Japan and South Korea, Southeast Asia, South Asia, the Middle East, North Africa, South Africa, Central Asia, Europe, etc.

There are 230 scenic spots in total, and the time includes the two dimensions of year and month.

Traffic measurement is mainly used to distinguish vehicles arriving at scenic spots, including cars, trains, planes, etc.

D. DM model

The system uses the SQL Server DM model to create a clustering model.

5. Smart Tourism DM Experiment

Total foreign exchange earnings from tourism is one of the most important indicators of the level of development of international inbound tourism in a country or region, and a comprehensive indicator of the country's or region's ability to generate foreign exchange from tourism. The number of jobs in the tourism industry and the foreign exchange earnings of related industries demonstrate the development of the tourism industry. In this text, four cities with smart tourism DM technology were selected to investigate their tourism-related data in 2010 and 2021 for data comparison. It included FEI in tourism, employment in tourism and FEI in related industries.

5.1. FEI of Tourism Industry

The FEI of tourism of the four cities in 2010 and 2021 is displayed in Figure 6.

Figure 6a shows the FEI of the city's tourism in 2010, and Figure 6b shows the FEI of the city's tourism in 2021. For City 1, the FEI of tourism in 2010 was 12.3 billion yuan, and in 2021, the FEI of tourism would increase to 25.7 billion yuan. For City 2, in 2010, the FEI of tourism was 24.2 billion, and in 2021, the FEI of tourism increased to 42.6 billion. For City 3, the FEI of tourism in 2010 was 8.7 billion yuan, and the FEI of tourism in 2021 would increase to 19.4 billion yuan. For City 4, the foreign exchange revenue of the tourism industry was 19.6 billion in 2010 and it increased to 50.3 billion in 2021. Tourism of the four cities has been greatly improved as a whole. Compared with 2010, the tourism FEI of the four cities would increase by more than 70% in 2021.

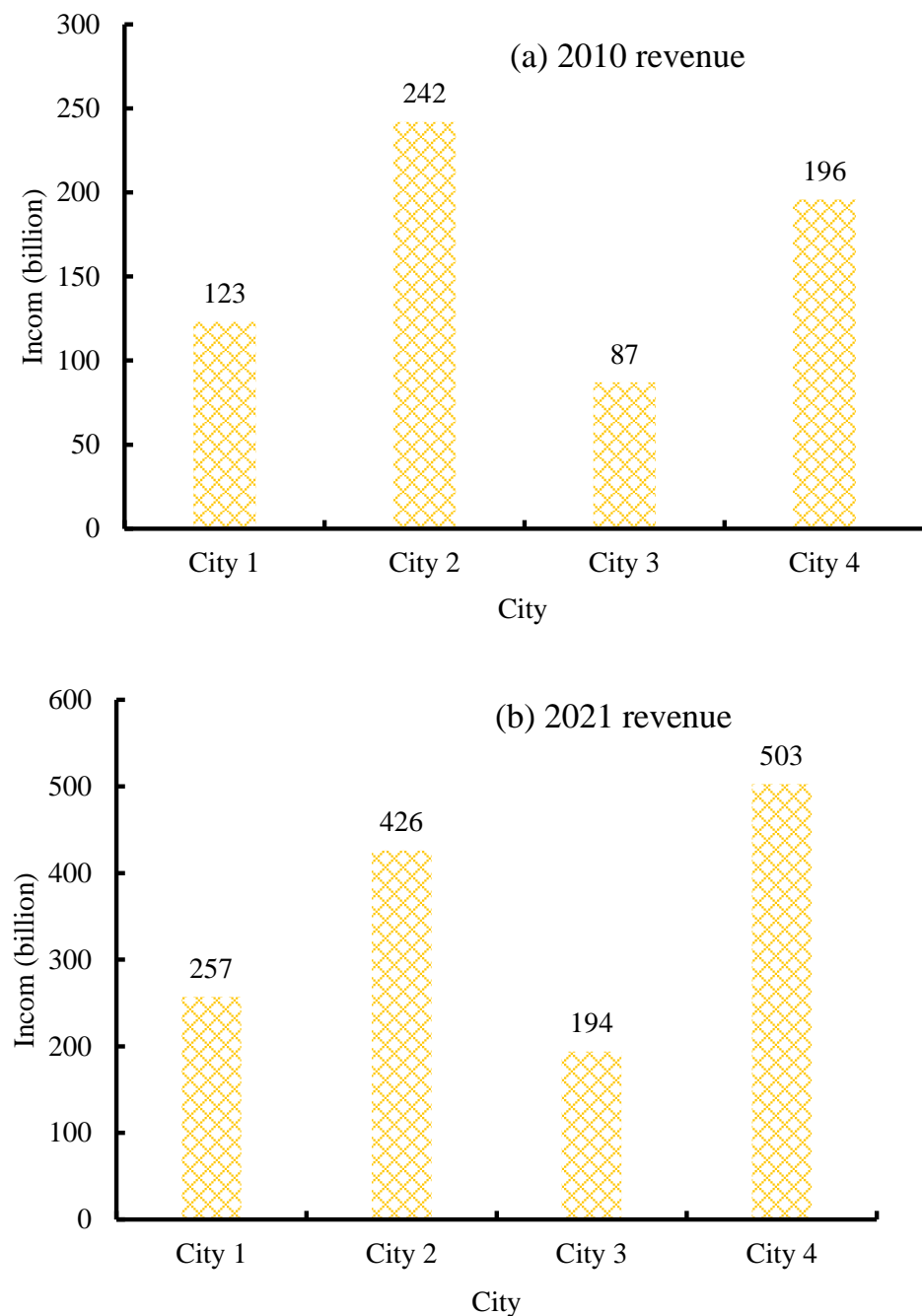


Figure 6. Foreign exchange earnings from the city's tourism sector. (a) FEI of urban tourism in 2010. (b) FEI of urban tourism in 2021.

5.2. Employment in Tourism

The tourism industry jobs in the four cities in 2010 and 2021 are displayed in Figure 7.

Figure 7a shows the urban tourism jobs in 2010, and Figure 7b shows the urban tourism jobs in 2021. In 2010, there were 1.07 million jobs in the tourism industry in City 1 and 1.54 million jobs in City 2. There are 690,000 tourism jobs in City 3 and 1,680,000 in City 4. By 2021, the number of tourism jobs in City 1 would increase to 2.14 million and that in City 2 would increase to 2.83 million. The number of tourism jobs in City 3 increased to 1.57 million and that in City 4 increased to 2.66 million. The city's tourism has been driven up, and the tourism industry has gradually increased employment.

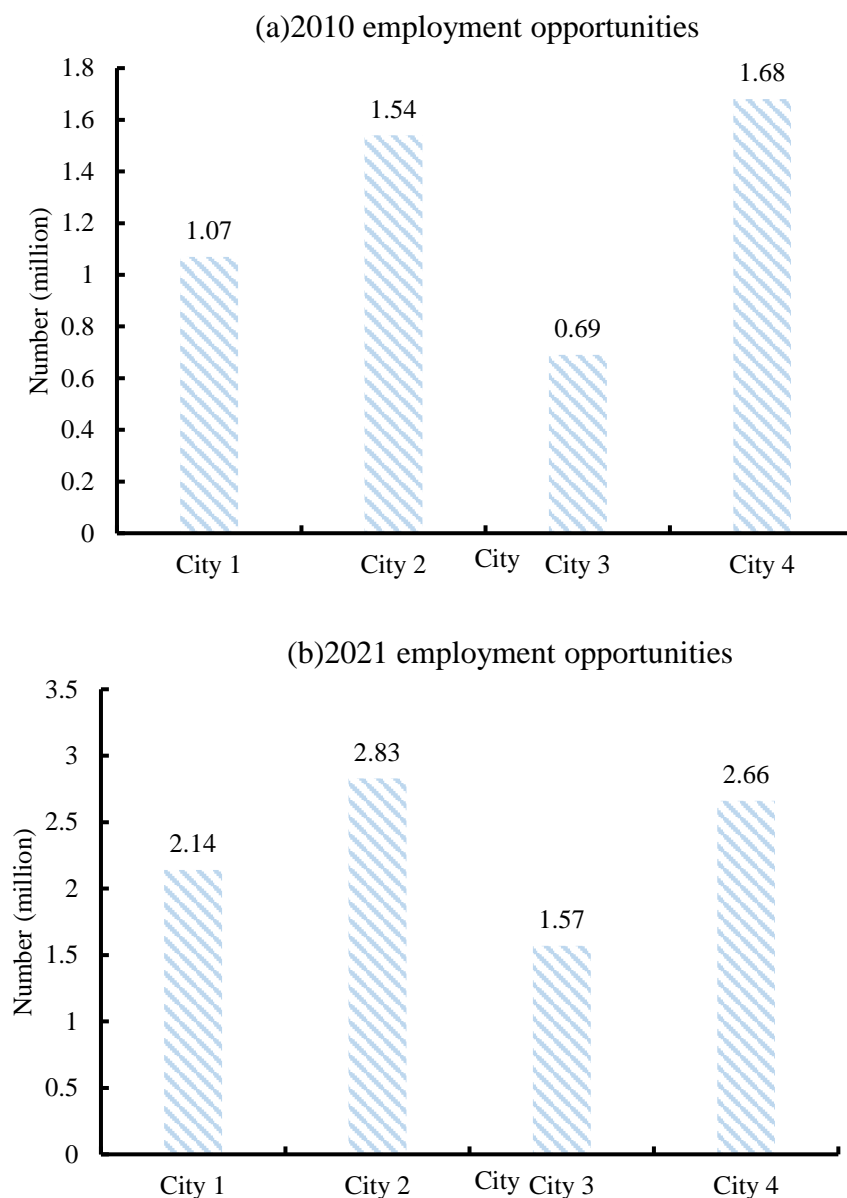


Figure 7. Jobs in the tourism industry sector in the city. (a) Jobs in the urban tourism sector in 2010. (b) Jobs in the urban tourism sector in 2021.

5.3. FEI of Related Industries

The FEI of tourism-related industries in the four cities in 2010 and 2021 is displayed in Figure 8.

Figure 8a shows the city's FEI from tourism-related industries in 2010, and Figure 8b shows the city's FEI from tourism-related industries in 2021. As for the accommodation, transportation and catering industries directly related to the tourism industry, it can be seen from the data that the FEI of tourism-related industries in 2021 would be much higher than that in 2010. With the application of smart tourism DM technology, the tourism industry has developed, driven the development of tourism-related industries, and the income of tourism-related industries has been greatly improved.

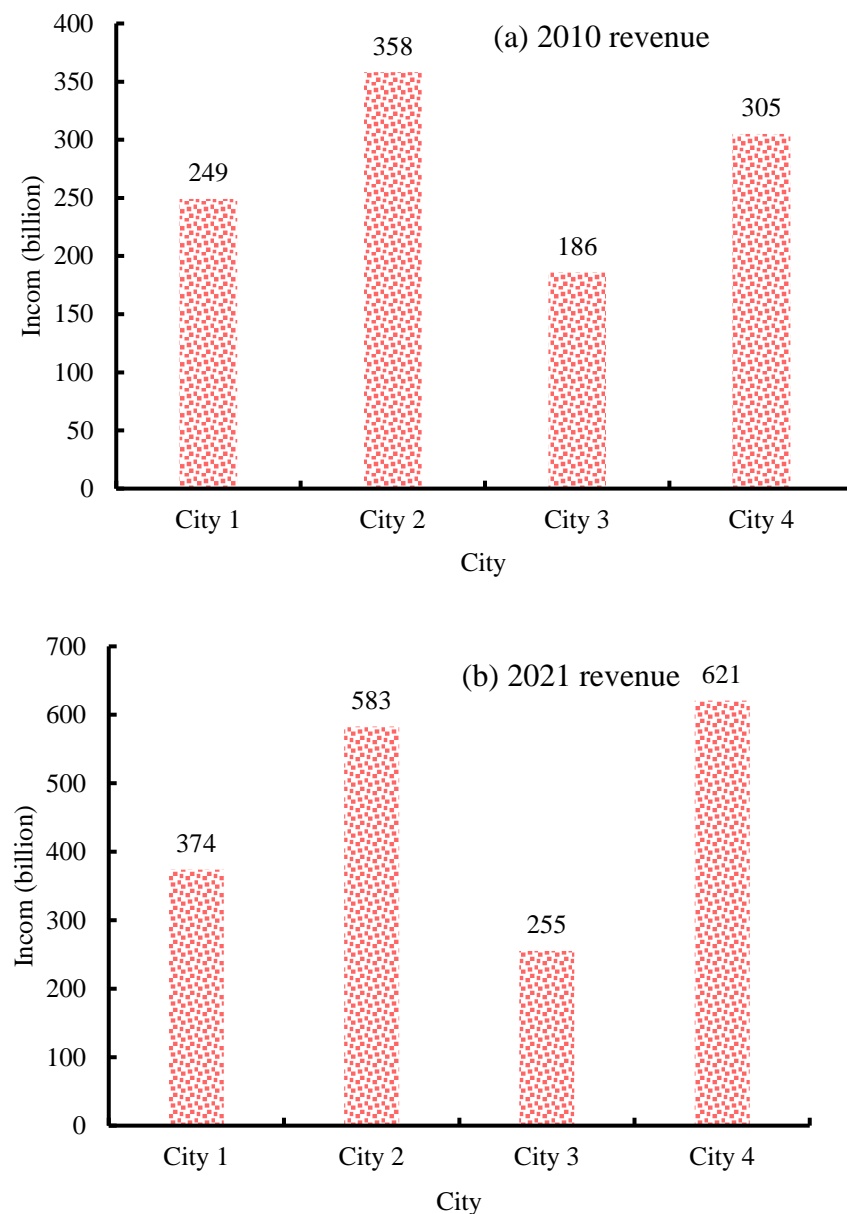


Figure 8. Foreign exchange earnings from tourism-related industries in the city. (a) Foreign exchange earnings from tourism-related industries in 2010. (b) Foreign exchange earnings from tourism-related industries in 2021.

6. Discussion

In this paper, based on tourism data from the Tourism Statistics Yearbook from 2010 to 2021, a tourism flow forecasting model is established by two methods, namely, regression model forecasting and the exponential curve model, to forecast the passenger flow. This paper also builds a tourism spatial dimension model based on tourists' tourism behaviour, collects tourism data and uses association rules to build a tourism DM model. The main work of this paper is as follows:

1. We analyzed the current state of research on sustainable smart tourism and BD mining, and to provide a detailed introduction to the DM process;
2. Tourism data from 2010 to 2021 were collected through the Tourism Statistics Yearbook and crawler software, and passenger flow forecasts were made using regression model forecasting and index model forecasting methods;

3. We constructed a tourism spatial dimension model according to the tourism industry behaviour of tourists, which can be used to recommend a combination of tourism locations and drive the passenger flow of surrounding tourist attractions;
4. We built tourism data tables according to different categories of tourism data, pre-process tourism data, used association rules for DM and used Microsoft SQL to build analytical models;
5. Four cities applying smart tourism DM technology were used as experimental objects to collect tourism data in 2010 and 2021 and conduct experimental research in three aspects: foreign exchange income from the tourism industry, employment in the tourism industry and foreign exchange income from related industries. The research results show that smart tourism DM technology can effectively promote the sustainable development of tourism in the city.

The research in this paper has solved the problem of excessive difference in passenger flow between scenic spots in different regions, predicted the passenger flow so that tourist attractions can prepare for tourism services in advance and avoided the decline of tourism service quality caused by the surge of passenger flow. Linkage of tourist attractions according to tourists' behavior and the formation of the tourism regional circle are conducive to improving competitiveness. The original tourism DM model is improved to better adapt to the growing demand for tourism management and improve the overall quality of the tourism industry.

7. Conclusions

Smart tourism is a future trend in today's tourism industry and a strategic need for the transformation and upgrading of urban tourism. The development of smart tourism technologies, applications and models, as well as DM for smart tourism, will help to improve the overall level and quality of urban tourism services. In order to improve the existing tourism DM model, better adapt to the growing needs of tourism management and achieve sustainable development of urban tourism, this paper carries out the construction of a smart tourism BD mining model. This paper DMs tourism data for sustainable smart tourism by collecting tourism data through the Tourism Statistics Yearbook and crawler software, constructs a passenger flow prediction model based on regression model prediction and index model prediction, establishes a tourism spatial dimension model for the intelligent recommendation of tourism locations and uses association rules to construct a DM model. The study shows that the application of intelligent tourism DM technology can greatly enhance the foreign exchange earnings of the tourism industry, improve employment opportunities in the tourism industry and promote the foreign exchange earnings of tourism-related industries. The tourism industry management system can improve the overall quality of the tourism industry, drive the sustainable development of the tourism industry, and promote the development of tourism-related industries. However, the intelligent tourism DM system in this paper only operates on small datasets, and due to the large amount of actual business data, there are more disruptive factors in a large system. A truly intelligent tourism DM system requires more testing and continuous optimization in real-world applications.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: There is no potential conflict of interest in our paper and all authors have seen the manuscript and its approved submission to this journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

References

1. Otowicz, M.H.; Macedo, M.; Biz, A.A. Dimensions of smart tourism and its levels: An integrative literature review. *J. Smart Tour.* **2022**, *2*, 5–19.
2. Sharma, P.; Meena, U.; Sharma, G.K. Intelligent Data Analysis using Optimized Support Vector Machine Based Data Mining Approach for Tourism Industry. *ACM Trans. Knowl. Discov. Data (TKDD)* **2022**, *16*, 1–20. [[CrossRef](#)]
3. Jovicic, D.Z. From the traditional understanding of tourism destination to the smart tourism destination. *Curr. Issues Tour.* **2019**, *22*, 276–282. [[CrossRef](#)]
4. Ghorbani, A.; Danaei, A.; Barzegar, S.M.; Hemmatian, H. Post modernism and designing smart tourism organization (STO) for tourism management. *J. Tour. Plan. Dev.* **2019**, *8*, 50–69.
5. Mehraliyev, F.; Chan, I.C.C.; Choi, Y.; Koseoglu, M.A.; Law, R. A state-of-the-art review of smart tourism research. *J. Travel Tour. Mark.* **2020**, *37*, 78–91. [[CrossRef](#)]
6. Jasrotia, A.; Gangotia, A. Smart cities to smart tourism destinations: A review paper. *J. Tour. Intell. Smartness* **2018**, *1*, 47–56.
7. Del Vecchio, P.; Mele, G.; Ndou, V.; Secundo, G. Creating value from social big data: Implications for smart tourism destinations. *Inf. Process. Manag.* **2018**, *54*, 847–860. [[CrossRef](#)]
8. Ardito, L.; Cerchione, R.; Del Vecchio, P.; Raguseo, E. Big data in smart tourism: Challenges, issues and opportunities. *Curr. Issues Tour.* **2019**, *22*, 1805–1809. [[CrossRef](#)]
9. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [[CrossRef](#)]
10. Iorio, C.; Pandolfo, G.; D'Ambrosio, A.; Siciliano, R. Mining big data in tourism. *Qual. Quant.* **2020**, *54*, 1655–1669. [[CrossRef](#)]
11. Li, D.; Deng, L.; Cai, Z. Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms. *Pers. Ubiquitous Comput.* **2020**, *24*, 87–101. [[CrossRef](#)]
12. Al Fararni, K.; Nafis, F.; Aghoutane, B.; Yahyaouy, A.; Riffi, J.; Sabri, A. Hybrid recommender system for tourism based on big data and AI: A conceptual framework. *Big Data Min. Anal.* **2021**, *4*, 47–55. [[CrossRef](#)]
13. Liu, J.; Yang, L.; Zhou, H.; Wang, S. Impact of climate change on hiking: Quantitative evidence through big data mining. *Curr. Issues Tour.* **2021**, *24*, 3040–3056. [[CrossRef](#)]
14. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment analysis in tourism: Capitalizing on big data. *J. Travel Res.* **2019**, *58*, 175–191. [[CrossRef](#)]
15. Rahmadian, E.; Feitosa, D.; Zwitter, A. A systematic literature review on the use of big data for sustainable tourism. *Curr. Issues Tour.* **2022**, *25*, 1711–1730. [[CrossRef](#)]
16. Ionescu, R.V.; Zlati, M.L.; Antohi, V.M.; Stanciu, S.; Burciu, A.; Kicsi, R. Supporting the tourism management decisions under the pandemic's impact. A new working instrument. *Econ. Res.-Ekon. Istraživanja* **2022**, *35*, 6723–6755. [[CrossRef](#)]
17. Nguyen, N.T.; Tran, T.T. Optimizing mathematical parameters of Grey system theory: An empirical forecasting case of Vietnamese tourism. *Neural Comput. Appl.* **2019**, *31* (Suppl. S2), 1075–1089. [[CrossRef](#)]
18. Olmeda, I.; Sheldon, P.J. Data mining techniques and applications for tourism internet marketing. *J. Travel Tour. Mark.* **2022**, *11*, 1–20. [[CrossRef](#)]
19. Ageed, Z.S.; Zeebaree, S.R.M.; Sadeeq, M.M.; Kak, S.F.; Yahia, H.S.; Mahmood, M.R.; Ibrahim, I.M. Comprehensive survey of big data mining approaches in cloud systems. *Qubahan Acad. J.* **2021**, *1*, 29–38. [[CrossRef](#)]
20. Wang, H.; Smys, S. Big data analysis and perturbation using data mining algorithm. *J. Soft Comput. Paradig. (JSCP)* **2021**, *3*, 19–28.
21. Kim, Y.-Y.; Kim, D.-S.; Kim, M.-H. Implementation of hybrid P2P networking distributed web crawler using AWS for smart work news big data. *Peer-to-Peer Netw. Appl.* **2020**, *13*, 659–670. [[CrossRef](#)]
22. Coban, O.; Ali, I.N.A.N.; Ozel, S.A. Towards the design and implementation of an OSN crawler: A case of Turkish Facebook users. *Int. J. Inf. Secur. Sci.* **2020**, *9*, 76–93.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.