

Retraction

Retracted: Construction of Tourism Management Engineering Based on Data Mining Technology

Journal of Electrical and Computer Engineering

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Electrical and Computer Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Z. Ma, "Construction of Tourism Management Engineering Based on Data Mining Technology," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 1982462, 11 pages, 2022.

Research Article

Construction of Tourism Management Engineering Based on Data Mining Technology

Zhichen Ma 

College of Marxism, Xi'an Fanyi University, Xi'an, Shaanxi 710105, China

Correspondence should be addressed to Zhichen Ma; mazhichen2014025@xafy.edu.cn

Received 11 April 2022; Revised 30 April 2022; Accepted 7 May 2022; Published 8 June 2022

Academic Editor: Wei Liu

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

In order to improve the construction quality of tourism management projects, this paper applies data mining algorithm to tourism management, and analyzes the SMOTE algorithm. According to the improvement direction, this paper proposes two improved algorithms, KM-SMOTE and RM-SMOTE, and uses the clustering algorithm to preprocess the minority data set. Moreover, on this basis, this paper establishes clusters and obtains cluster centers. The deficiencies of fuzzy positive and negative class boundaries can be effectively solved by oversampling with the cluster center as the base point, and in the case of appropriately expanding the reasonable data interpolation method, the area space of boundary interpolation can be shrunk, and the performance of the algorithm can be improved. It can be seen from the simulation test research that the tourism management system based on data mining proposed in this paper can play an important role in tourism management and effectively promote the improvement of tourism management efficiency.

1. Introduction

In the context of the rapid development of information technology, people's living habits and ways of thinking are constantly affected by Internet informatization. At the same time, tourism has become an important flavoring agent in people's life and a "new fashion" for people to relieve the pressure of life and work. At the same time, the number of tourism companies and the number of tourists also continued to increase. In this form, the effective application of big data to tourism management has become an inevitable trend of tourism development at this stage. In the context of the development of the tourism market, various new phenomena in the tourism industry emerge one after another. How to comprehensively collect, organize, improve, supervise, and control the information behind these phenomena is a dilemma facing the tourism industry at this stage. Moreover, how to deal with the problems behind the information and improve the tourism market system requires the application of big data in tourism management. Through the collection and analysis of various problems in the tourism market through big data, the management

defects of tourism enterprises can be discovered in time, and the enterprises can better improve service quality and increase benefits in the process of making up for shortcomings and tapping their own advantages. At the same time, in order to ensure the service quality of tourism enterprises to meet the needs of tourists, it is necessary for tourism enterprises to realize the diversification of services themselves, so as to steadily enhance their own value.

The application of big data to manage the tourism industry can effectively collect and summarize information such as the number of tourists and tourists' travel preferences. Moreover, on the premise of improving the application of the corresponding software and performing model construction and verification, it can ensure that the acquired data is close to being effective. Through the information analysis of these near-effective data, the result obtained is the authenticity data required for the development of tourism [1]. In addition, tourism enterprises can use this information to further optimize their management and operations, facilitate real-time monitoring and detection, and determine the effective operation results of enterprises, and can also play a guiding role in the infrastructure of tourist

destinations [2]. Based on this, the application of big data for tourism management can greatly promote the development of tourism, showing good tourism development trends and diversified structures. On the basis of ensuring the directionality of tourism services, it can carry out a full range of services to maximize the realization of tourists' subconscious tourism goals, thereby improving the service level of the tourism industry and keeping tourism management in line with the pace of social development [3].

The application of big data in the analysis and prediction of tourism links enables high-quality analysis, prediction and improvement of tourism links through existing tourism data and relevant information such as tourists' needs, so as to ensure that tourism links meet the actual needs of tourists. In addition, by applying the relevant principles of statistics, the tourist information and travel time can be investigated, and the forecast model of tourist demand can be set according to the obtained survey results. At this stage, there are mainly the following ways to predict tourist demand: The first is the structural model. This model takes the actual needs of tourists as the premise, and introduces factors that affect the needs of tourists to set up planning. The second is the trend extrapolation model. This model mainly refers to the past tourist demand, so as to effectively predict the future tourist demand and lay the foundation for future tourism services. The third is the simulation model. This kind of model effectively organizes and merges the structural model and the trend extrapolation model, so as to judge the real situation of each link of tourism. Therefore, the simulation model is more consistent with the actual tourism situation. The fourth is the stereotyped model. This model is based on the exploration and research of tourism services, to find out the defects, and to ask experts to answer the existing problems. The setting of the stereotype model can effectively promote the further improvement of the tourism industry.

In order to improve the construction effect of tourism management project, this paper applies data mining algorithm to tourism management, and constructs an intelligent tourism management model.

2. Related Work

Literature [4] analyzes the impact of advertising on the demand, pricing and profits of tourism supply chain node enterprises, establishes three tourism supply chain models, and finally concludes that the travel operator model is more effective than the other two models. The pricing strategy and profit of channel members to make it profitable. Literature [5] discusses the tourism supply chain cooperation model between large scenic spots and small scenic spots. The model structure of the tourism supply chain has experienced the development process from a simple chain structure, to a network structure including indirect suppliers, and then to a new type of tourism supply chain under the background of the Internet in the new era [6]. Generally speaking, there will be a core node enterprise in the supply chain, and the core node enterprise will drive other franchise node enterprises to realize the continuous value-added of the entire supply chain through division of labor and cooperation. Literature [7]

believes that travel agencies have special functions in the supply chain. They can not only directly contact consumers, but also act as intermediaries for related enterprises such as accommodation, catering, transportation, and scenic spots. This characteristic determines that the travel agency can control the collection of tourism information, the scheduling of tourism flow, and the settlement of funds at the same time, which directly determines that it has the conditions to grow into a core enterprise in the tourism supply chain [8]. Literature [9] introduced the supply chain model into the tourism industry, with travel agencies as the main body, combined with the characteristics of tourism products, organically integrated the core ideas of supply chains into the tourism industry, and constructed a tourism supply chain with travel agencies as the core. Literature [10] believes that travel agencies should be the central coordinator of the entire tourism supply chain, connecting tourism supply companies and tourism consumers, and promoting the formation of a tourism consumption system. Literature [11] believes that the tourism supply chain with travel agencies as the core has brought many problems to the development of tourism, and it is not in line with the new situation of tourism development. It discusses and builds a tourism supply chain with tourist attractions as the core model. Literature [12] pointed out that in the Internet environment, it is necessary to build a tourism supply chain model with scenic spots as the core. Literature [13] believes that the developing tourism e-commerce website will become the real core of the tourism supply chain. Literature [14] combined with the principle of maximizing the perceived value of tourists, and proposed that the tourism market dominated by individual tourists should build a tourism supply chain with network intermediaries as the core.

On the whole, the research on the development strategy management of tourism enterprises has received more attention. Literature [15] cuts in from the brand strategy perspective of tourism enterprises, expounds the internal connection between the core competitiveness and brand strategy of tourism enterprises, believes that tourism brand strategy is the core of tourism enterprise development strategy, and expands the vision of tourism enterprise strategy research.

From the perspective of marketing, there are two main aspects to study tourism enterprise management. One is the analysis of tourism enterprise marketing creative planning and the analysis of specific marketing methods such as blog marketing, Weibo marketing, and network marketing in the context of the current era; the second is to study the process of tourism enterprise marketing. Exploring solutions to ethical issues. Literature [16] analyzes the positive role of blog marketing, and proposes implementation suggestions for tourism enterprises to use blog as a marketing approach. Literature [17] pointed out the emergence of immorality in tourism enterprise marketing, and believed that the primary problem and key to solving the immorality in tourism enterprise marketing activities is to build a tourism enterprise marketing ethics evaluation model to conduct tourism enterprise marketing ethics evaluation. Evaluate. Combined with the background of the era of big data, literature [18]

pointed out the new development opportunities of online marketing for tourism enterprises, which is in line with the trend of the times.

Literature [19] pointed out the relationship between the contents of the enterprise management innovation system, and pointed out that the systematization of research content and the quantification of methods should be emphasized in the future. Literature [20] analyzes the influence of knowledge economy on tourism enterprises in business management, puts forward innovative strategies for tourism enterprise management under the environment of knowledge economy, and points out that tourism enterprises must strengthen the management of tourism enterprises if they want to achieve greater development in the next century. Research on knowledge economy, increase the emphasis on knowledge, information and talents, continuously improve its comprehensive innovation ability, and adjust management strategies. Hou Xueyan innovatively analyzed the ecological attributes of tourism enterprises, emphasizing that the implementation of tourism enterprise management strategies from this perspective should cultivate ecological concepts and build a tourism enterprise ecological management operation mechanism. Literature [21] pays attention to the fact that tourism enterprises should interact with the internal and external environments in the face of the current crisis and rebirth, and actively adapt to the environment, and then continue to shape the environment or make effective adjustments. The characteristic management flexibility will inevitably become the trend of tourism enterprise management.

3. Improved Data Mining Algorithm

At present, the research on classification optimization of imbalanced data mainly focuses on the data level and the algorithm level, that is, improving the classification algorithm and reconstructing the data. Data reconstruction is crucial for imbalanced data classification. In particular, the combination of data reconstruction and classification algorithms has become the mainstream at this stage. Through the summary and analysis of the research results of imbalanced data set classification, it is concluded that in the research of imbalanced data, data reconstruction, that is, data level research, is very effective.

The KM-SMOTE algorithm is proposed by introducing the K-means clustering algorithm to preprocess the minority class and simply modifying the corresponding interpolation formula.

In the SMOTE algorithm, since each interpolation is associated with the sample point data and its K nearest neighbors. Due to the large randomness of the SMOTE algorithm oversampling interpolation, if the interpolation result is not ideal, the oversampling operation will blur the positive and negative class boundaries of the data class.

We need to improve it for the case of fuzzy positive and negative class boundaries. The clustering algorithm is used to preprocess the minority class data set, and on this basis, the cluster is established, the cluster center is obtained, and the oversampling operation is performed with the cluster center

as the base point, which can effectively solve the shortage of fuzzy positive and negative class boundaries.

The effect of preprocessing is shown in Figure 1(a). It can be seen from Figure 1(a) that we use the K-means algorithm to perform data preprocessing on the minority class, and obtain three clusters and cluster centers. The three clusters are c_1, c_2, c_3 . Next, we take cluster: as an example to perform data oversampling operation, as shown in Figure 1(b).

From Figure 1(b), we can see that through the connection between the cluster center and the data samples in the cluster, an association is established, and the samples are synthesized on the connection line between the cluster center and the data samples in the cluster. The synthesized samples are always in the clustering area and will not cross the boundary, which sets the sampling rules for the synthesized samples, which can effectively control the oversampling process. A more obvious effect after interpolation is shown in Figure 1(c).

As can be seen from Figure 1(c), the feature attributes of the KM-SMOTE algorithm determine its application significance. Before interpolation, a judgment operation needs to be performed on the data, and the synthesized samples can effectively avoid invalid interpolation, reduce the probability of blurring the boundaries of positive and negative classes, and maintain the distribution pattern of minority data.

3.1. Algorithm Core. The core of the algorithm is mainly analyzed from three parts: determining the boundary point of the minority class, judging the dangerous point, and correcting the oversampling formula.

3.1.1. Determining the Boundary Points of the Minority Class.

For the imbalanced dataset S, it is defined as formula (1):

$$S = \{(W_1, v_1), (W_2, v_2), \dots, (W_m, v_m)\}. \quad (1)$$

Among them W_i is the instance in dataset S and $v_i \in \{1.0, 0.0\}$, $1 \leq i \leq m$ is the class label of W_i . P is the minority class set, $v_j = 1.0$, $1 \leq j \leq |P|$, Q is the majority class set, $v_l = 0.0$, $1 \leq l \leq |Q|$.

For the minority class instance, it obtains its K nearest neighbors, and judges the category to which the K nearest neighbors belong. If there is a majority class, the minority class is a boundary minority class data sample, which is included in the boundary data sample set R, and R is expressed as shown in formula (2):

$$R = \{(X_1, v_1), (X_2, v_2), \dots, (X_t, v_t)\}. \quad (2)$$

Among them, X_i represents the boundary point instance, and v_i is the class label. While recording R, the minority class instances and the majority class neighbors in R are recorded in the data sample set T.

3.2. Judging the Danger Point. After clustering the minority class, it is necessary to judge the number of data samples of the boundary minority class in the data samples of each

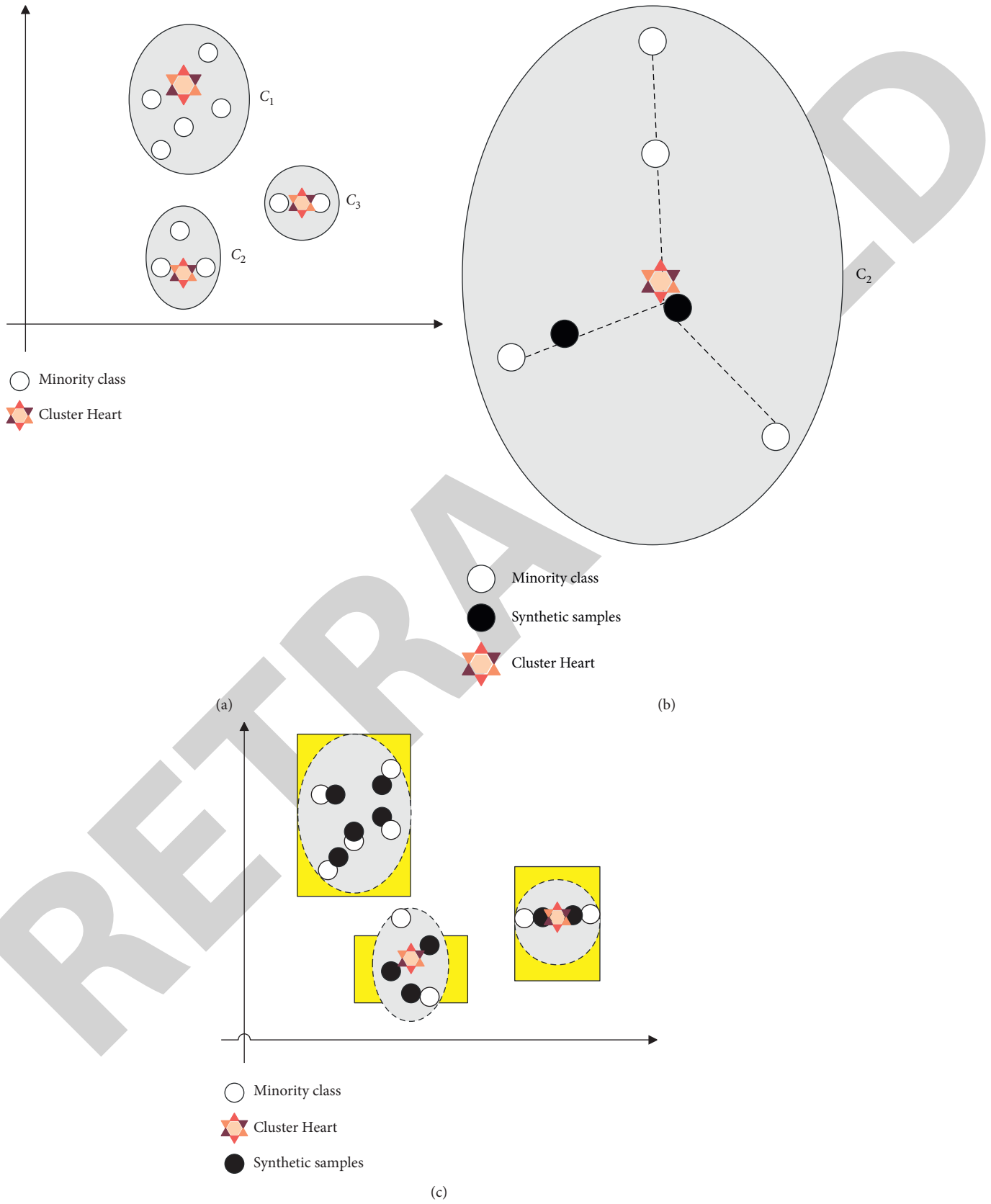


FIGURE 1: Processing process of data clustering. (a) Preprocessing for a minority class (b) Oversampling of cluster c_2 . (c) Synthetic samples of the KM-SMOTE algorithm.

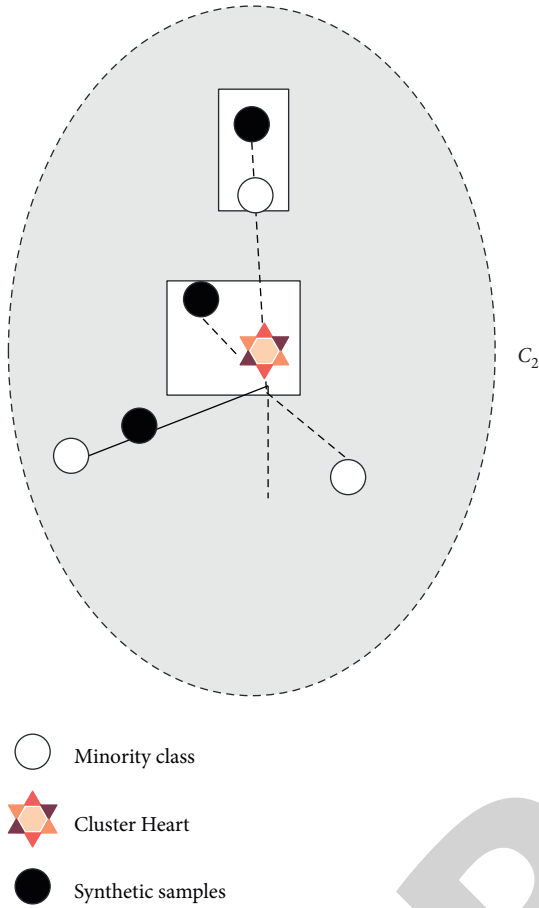


FIGURE 2: KM-SMOTE oversampling samples.

cluster. If it is greater than 1, the cluster c_i needs to be judged again.

It judges the Euclidean distance d_{uq} between the cluster center u_i and the boundary minority class instance X and the Euclidean distance d_{up} between the K nearest neighbors of the minority class instance X from u_i to the majority class neighbors.

If there is $d_{uq} < d_{up}$ and the K nearest neighbors of X are all majority classes, the point X is judged to be a dangerous point of clustering, the point is deleted, and the clustering center is recalculated.

After that, it performs and so on until there is no such situation, and finally determines the clusters and cluster centers.

3.3. Correcting Oversampling Formula. After preprocessing the data with the clustering algorithm, oversampling operation needs to be performed. For the improved SMOTE algorithm, we also need to modify the corresponding oversampling formula. The two oversampling formulas mentioned below are introduced in detail.

This article first introduces the oversampling formula that directly performs the oversampling operation.

Referring to the formula, in the value formula of the SMOTE algorithm, the oversampling operation is to insert

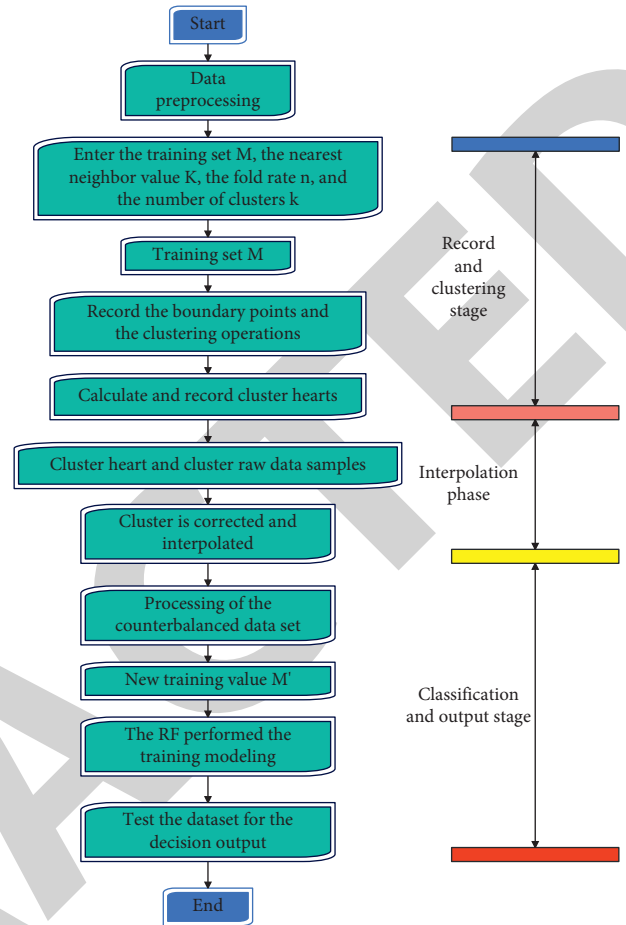


FIGURE 3: Flowchart of the KM-SMOTE algorithm.

new data based on the selected sample points. Considering the characteristics of the improved algorithm, the base point of the oversampling value formula is revised, and the base point is set as the cluster center of the cluster. Therefore, the value formula of the KM-SMOTE algorithm is shown in formula (3) below [22]:

$$X_{mnw} = u_i + \text{rand}(0, 1) * (X - u_i), \quad i = 1, 2, \dots, k. \quad (3)$$

Among them, X_{new} is the newly interpolated sample, u_i is the cluster center, and X is the original sample data in the cluster where u_i is the cluster center. $\text{rand}(0, 1)$ represents a random number between 0 and 1, and k is the number of clusters.

Formula (3) refers to the SMOTE algorithm. This interpolation method makes the interpolation space too small and easily leads to overfitting. Therefore, equation (3) is revised in this paper.

Next, this paper introduces the revised oversampling operation formula.

For the oversampling formula of the KM-SMOTE algorithm, this paper considers the Euclidean distance from the cluster center to each cluster data sample, and obtains the maximum Euclidean distance.

The set of Euclidean distances from the cluster center u_i of cluster c_i to each data point in the cluster is D , then D is expressed as formula (4):

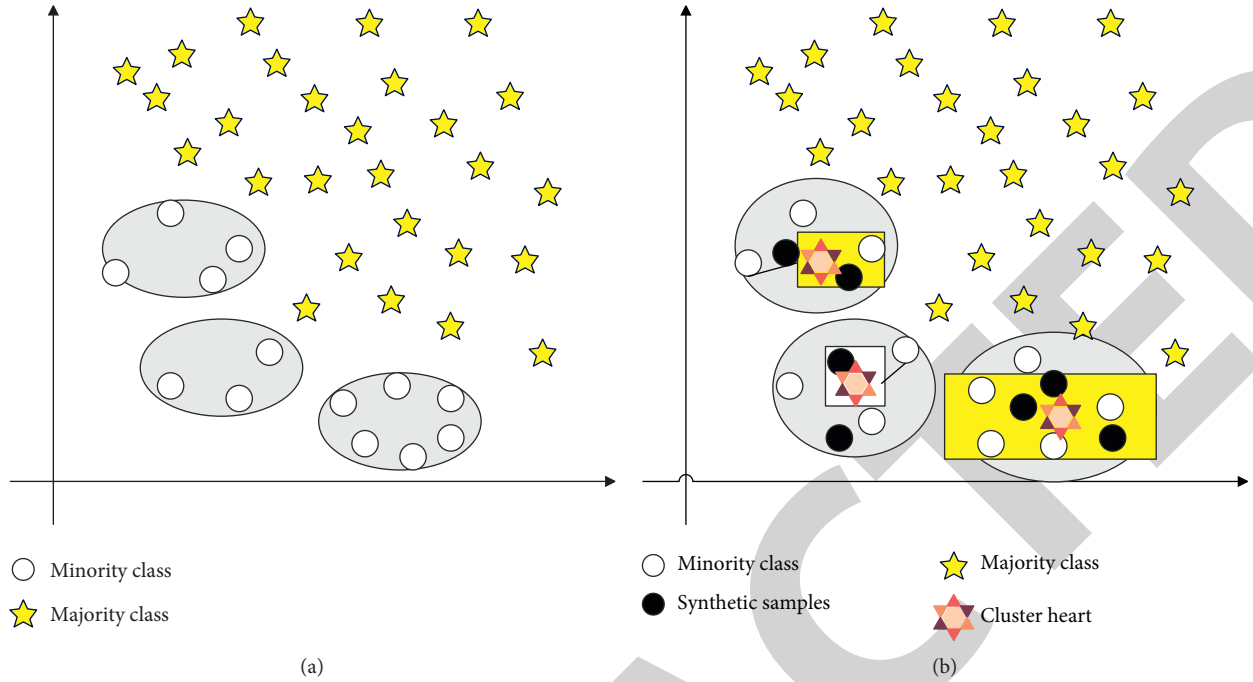


FIGURE 4: Data clustering and interpolation results. (a) Clusters formed by minority classes (b) RM-SMOTE interpolation result.

$$D = \{d_{up1}, d_{up2}, d_{up3}, \dots\}. \quad (4)$$

Among them, d_{upj} represents the Euclidean distance, $1 \leq i \leq k$ represents the number of clusters, and j represents the number of data sample points in cluster i . After obtaining the Euclidean distance, the maximum Euclidean distance D_{\max} is taken, and the relevant formula is shown in the following (5):

$$D_{\max} = \text{MAX}\{d_{up1}, d_{up2}, d_{up3}, \dots\}. \quad (5)$$

After the maximum Euclidean distance is obtained, the relationship between the Euclidean distance d_{upj} between the judged cluster center and the data sample point X and the maximum Euclidean distance is determined. Moreover, the multiples H_{ij} of the data sample d_{upij} and the maximum Euclidean distance D_{\max} are calculated, as shown in formula (6):

$$H_{ij} = \frac{D_{\max}}{d_{upij}}. \quad (6)$$

Taking an integer for H_{ij} is shown in formula (7) below:

$$H = [H_{ij}]. \quad (7)$$

The new interpolation formula obtained is shown in formula (8):

$$X_{\text{new}} = u_i + \text{rand}(0, H)^* (X - u_i), \quad i = 1, 2, \dots, k. \quad (8)$$

Among them, X_{new} is the newly interpolated sample, u_i is the cluster center, X is the original sample data in the cluster with u_i as the center, and $\text{rand}(0, H)$ represents a random number between 0 and H .

From Figure 2, we can see that we extend the interpolation range of the oversampling algorithm within the

cluster, and the range is on the extension line between the cluster center and the data, but it never exceeds the cluster interval. It can be seen from the algorithm design that the synthetic samples of the KM-SMOTE algorithm have the following two characteristics:

- (1) All synthetic samples are on the line connecting the cluster center and the data samples.
- (2) All synthetic samples are within the interval of relevant clusters.

According to the specific content of the above algorithm description, the process of the KM-SMOTE algorithm is mainly divided into the following stages: boundary point determination, clustering, judgment, interpolation, classification output, etc.

The main flow of the algorithm is shown in Figure 3.

Because the evaluation method of the G -means criterion depends on two metrics, the accuracy of the classifier to classify the minority class and the accuracy of the classifier to classify the majority class. We can express it succinctly with two values a^+ and a^- .

The classification accuracy of minority class sample is shown in formula (9):

$$a^+ = \frac{TP}{TP + FN}. \quad (9)$$

The classification accuracy of the majority class samples is shown in formula (10):

$$a^- = \frac{TN}{TN + FP}. \quad (10)$$

The calculation formula of the overall classification performance indicator G -means is shown in formula (11):

$$G - \text{means} = \sqrt{a^+ \times a^-}. \quad (11)$$

The classification effect of the classifier is evaluated by the value of G -means. The larger the value, the better the classification effect.

The evaluation criteria commonly used in imbalanced datasets are given below, where the precision is calculated as shown in formula (12):

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (12)$$

The formula represents the ratio of correctly classified positive samples to all samples classified as positive.

The recall calculation formula is shown in formula (13):

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (13)$$

The formula represents the ratio of correctly classified positive samples to the actual positive samples.

The calculation of the F value is shown in formula (14):

$$F - \text{value} = \frac{(1 + \beta^2) \times \text{Recall} \times \text{Precision}}{\beta^2 \times \text{Precision} + \text{Recall}}. \quad (14)$$

F -value represents the combination of precision and recall, where $\beta = 1$. When Recall and Precision tend to the maximum value at the same time, the value of F -value tends to the maximum value. The larger the F -value, the better the classification effect of imbalanced data.

In order to expand the interpolation range of oversampling and make the algorithm more applicable, we design another algorithm.

The design idea of this algorithm is to introduce the idea of h -dimensional spherical space to limit the interpolation. In the case of appropriately expanding the reasonable data interpolation method, the area space of boundary interpolation is reduced to improve the performance of the algorithm.

The steps of preprocessing data by clustering are shown in Figure 4(a). After a minority class sample is preprocessed by the clustering algorithm, three clusters are obtained. The proposed RM-SMOTE algorithm calculates the Euclidean distance between the cluster center and the clustered data samples, and obtains a times the maximum value as the spherical radius, and performs random interpolation in the spherical area. Figure 4(b) presents an example result of the interpolation.

From Figure 4(b), we can see that all synthetic data are interpolated within the region defined by the set radius. Moreover, we can see from the figure that the interpolation does not have certain rules, it is not fixed on the data connection and extension line, but is randomly distributed in a spherical body limited by a fixed radius.

This article focuses on the calculation of the maximum Euclidean distance. u_i represents the cluster center of cluster c_i , and X_i represents any data sample in the cluster. In addition, the number of attributes of the minority data set sample is E , then the data sample X_i can be represented by attributes, as shown in formula (15):

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iE}). \quad (15)$$

Among them, $(x_{i1}, x_{i2}, \dots, x_{iE})$ is the attribute value of the E attributes of X_i , as shown in formula (16):

$$u_i = (u_{i1}, u_{i2}, \dots, u_{iE}). \quad (16)$$

Among them, u_i represents the cluster center of cluster c_i , and $(u_{i1}, u_{i2}, \dots, u_{iE})$ represents the E attribute values of the cluster center u_i .

For the Euclidean distance set from the cluster center u_i of cluster c_i to each data point in the cluster is D , then $D = \{d_{upi1}, d_{upi2}, d_{upi3}, \dots\}$, $1 \leq i \leq k$ represents the number of clusters.

To find the maximum Euclidean distance D_{\max} , the relevant formula is shown in formula (17):

$$D_{\max} = \text{MAX}\{d_{upi1}, d_{upi2}, d_{upi3}, \dots\}. \quad (17)$$

3.4. Correcting Oversampling Interpolation Formula. If the new data to be synthesized is X_{new} , then there is $X_{\text{new}} = (x_{\text{new}1}, x_{\text{new}2}, \dots, x_{\text{new}E})$. According to the theory

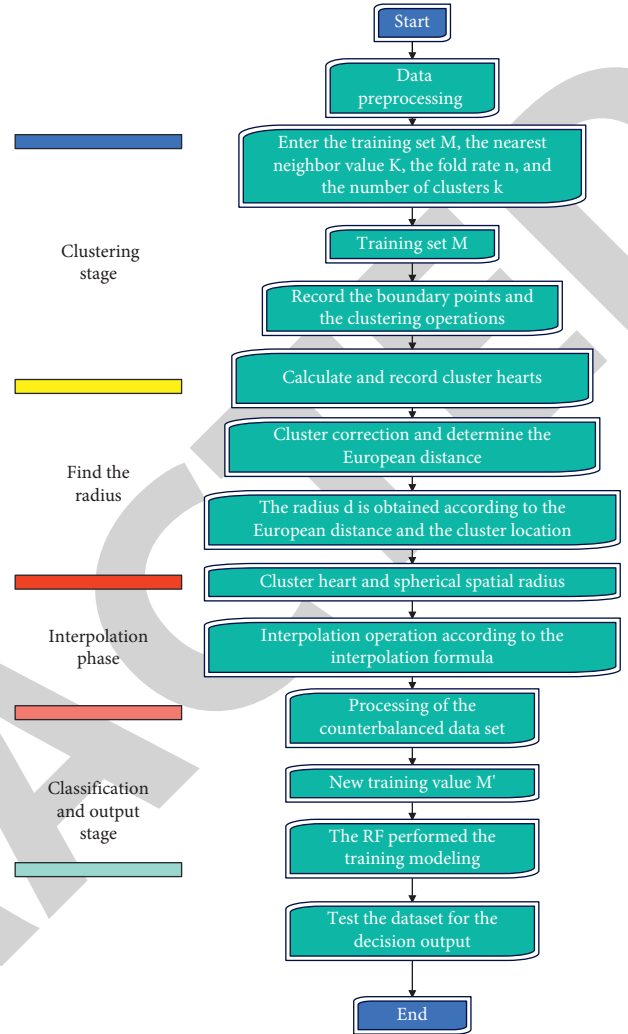


FIGURE 5: RM-SMOTE algorithm flow chart.

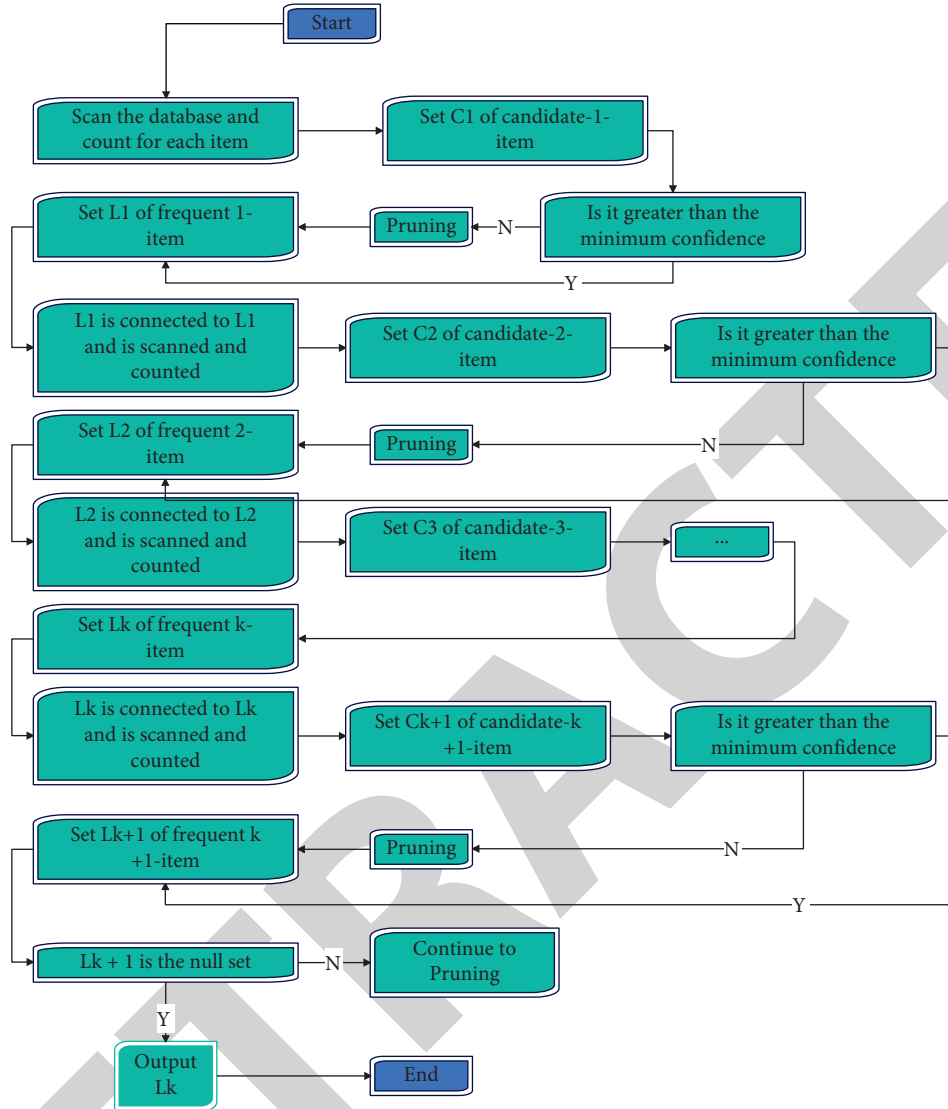


FIGURE 6: Intelligent data process of tourism management engineering construction.

of the RM-SMOTE algorithm, the randomly generated synthetic data designed in this paper must meet the conditions shown in equations (18)–(20).

$$\|X_{\text{new}} - u_i\| \leq D_{\text{max}}. \quad (18)$$

Among them, $\|X_{\text{new}} - u_i\|$ represents the Euclidean distance from the synthetic data X_{new} to the cluster center u_i . As mentioned above, D_{max} is the maximum Euclidean distance from all cluster data samples to the cluster center u_i .

$$x_{\text{new}j} = u_{ij} + \text{rand}(0, 1) \times (b_j - a_j), \quad 1 \leq j \leq E. \quad (19)$$

Among them, $x_{\text{new}j}$ represents the attribute value of the j -th attribute of the synthetic sample X_{new} , $\text{rand}(0, 1)$ represents a random number between (0, 1), and $(b_j - a_j)$ satisfies the conditions of formula (20).

$$\begin{aligned} a_j &= u_{ij} - |x_{\text{max}j} - u_{ij}|, \\ b_j &= u_{ij} + |x_{\text{max}j} - u_{ij}|, \quad 1 \leq j \leq E. \end{aligned} \quad (20)$$

Among them, $|x_{\text{max}j} - u_{ij}|$ represents the absolute value of the attribute difference of the j -th attribute between the data X_{max} that obtains the maximum Euclidean distance and the cluster center u_i .

For different datasets, the extent of spherical space can be expanded or reduced accordingly, such as oversampling based on average Euclidean distance. This is for Literature only, and a brief introduction is given below.

The mean Euclidean distance D_{avg} is obtained. If it is assumed that there are f data samples in the cluster, the formula is as shown in formula (21):

$$D_{\text{avg}} = \frac{1}{f} \sum_{j=1}^f d_{\text{up}ij}. \quad (21)$$

The relationship a between D_{avg} and D_{max} is obtained, the formula is shown in formula (22):

$$a = \frac{D_{\text{avg}}}{D_{\text{max}}}. \quad (22)$$

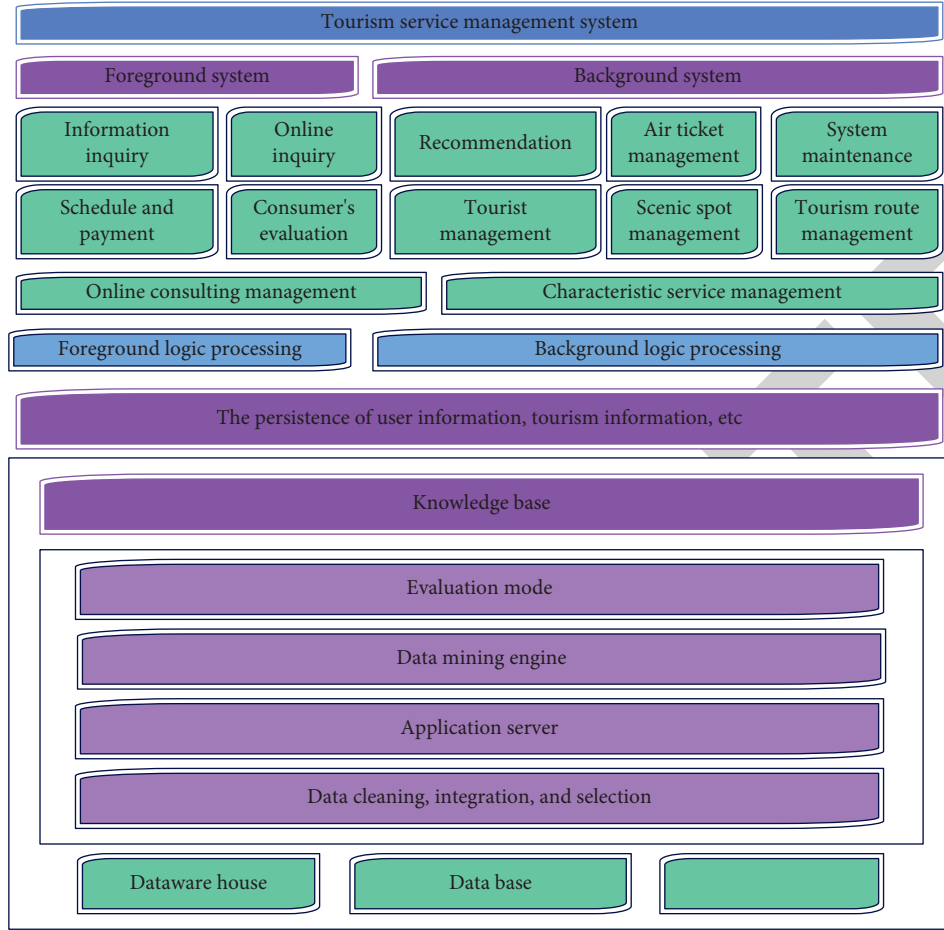


FIGURE 7: Tourism management system based on data mining.

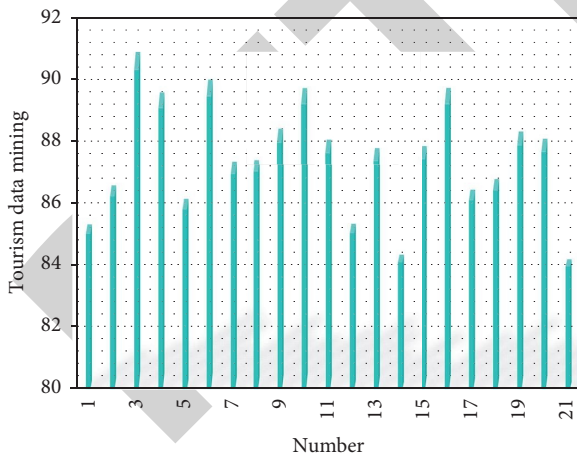


FIGURE 8: Tourism data management effect of tourism management system based on data mining.

Then, the synthetic data meets the requirements shown in formulas (23)–(25):

$$\|X_{\text{new}} - u_i\| \leq D_{\text{avg}} = a \times D_{\text{max}}, \quad (23)$$

$$x_{\text{new}j} = u_{ij} + \text{rand}(0, 1) \times a \times (b_j - a_j), \quad 1 \leq j \leq E, \quad (24)$$

$$\begin{aligned} a_j &= u_{ij} - |x_{\text{max}j} - u_{ij}|, \\ b_j &= u_{ij} + |x_{\text{max}j} - u_{ij}|, \quad 1 \leq j \leq E. \end{aligned} \quad (25)$$

The meaning of each variable is the same as the meaning of the variable oversampled based on maximum Euclidean distance.

The RM-SMOTE algorithm is subdivided into the following parts: finding neighbors, recording boundary points, clustering operations, judging and correcting clusters and cluster centers, finding radius, oversampling, and classification.

The flowchart of the RM-SMOTE algorithm is shown in Figure 5.

4. Construction of Tourism Management Engineering Based on Data Mining Technology

The intelligent data process of tourism management engineering construction proposed in this paper is shown in Figure 6 below. First, the algorithm finds frequent 1-itemsets, denoted as L1. Then, the algorithm finds the result of the candidate item set C2 through L1, determines the items in C2, and mines L2, that is, the frequent 2-item set. Then, the

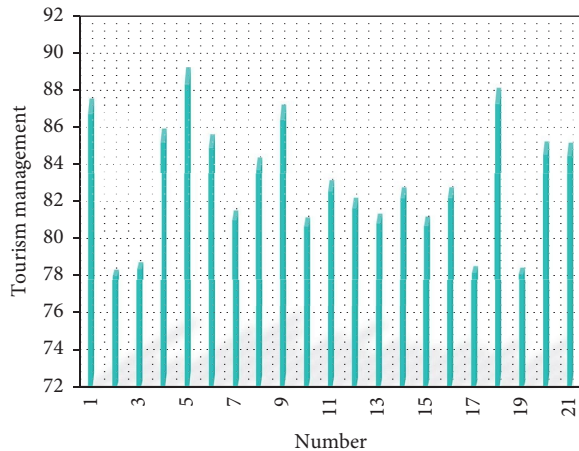


FIGURE 9: Auxiliary tourism management effect of tourism management system based on data mining.

algorithm continues to loop in this way until no more frequent k -itemsets can be found. The database is scanned once for each dig. Data mining algorithms take advantage of the property that all frequent itemsets and their non-empty subsets are also frequent. In other words, when a k -itemset candidate is generated, if a subset of a candidate item in $(k-1)$ -itemset is not included, then this candidate item can be deleted directly without comparing it with the support.

The system uses the popular four-tier architecture in the application software structure. Compared with the two-tier architecture and the three-tier architecture, this structure separates the business logic layer and the persistence layer, which increases the independence between programs and makes it easy to expand. At the same time, the security of the system is relatively improved in the use of this structure. On the basis of the above algorithm, a tourism management system based on data mining is constructed, as shown in Figure 7.

After the above system model is obtained, the effect of the system model in this paper is verified. This paper uses the simulation test to process the tourism data and evaluate the tourism management effect of the tourism management system constructed in this paper, and obtain the results shown in Figures 8 and 9.

From the above research, we can see that the tourism management system based on data mining proposed in this paper can play an important role in tourism management and effectively promote the improvement of tourism management efficiency.

5. Conclusion

In order to improve tourists' satisfaction with tourism, reasonable and effective tourism forecasting should be carried out. Tourism forecast can make the connection of the tourism industry chain more complete, point out the direction for the decision-making of tourism management, and help relevant government functional departments to carry out a comprehensive and complete overall planning of tourism resources. In the past tourism management, the

analysis and prediction of tourism links were mainly carried out by the corresponding staff based on their own work experience. This leads to many problems, which not only affect the tourist experience of tourists, but also bury some hidden safety hazards and threaten the life and property safety of tourists. In order to improve the construction effect of tourism management project, this paper applies data mining algorithm to tourism management. It can be seen from the simulation test research that the tourism management system based on data mining proposed in this paper can play an important role in tourism management and effectively promote the improvement of tourism management efficiency.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declare no competing interests.

Acknowledgments

This study is sponsored by 1. Research project on Theory and Practice of Higher Education in Shaanxi Province in 2021 (cooperative project). "Research on the path of integrating Shaanxi red resources into college students' Ideological and Political Education" (No: 2021HZ0707); 2. 2021 general special scientific research plan project of Shaanxi Provincial Department of Education. "Research on the path of integrating Shaanxi red resources into college students' Ideological and Political Education" (No: 21JK0164).

References

- [1] J. Bulchand-Gidumal, M. Sigala, R. Rahimi, and M. Thelwall, "Big data and innovation in tourism, travel, and hospitality. Managerial approaches, techniques, and applications," *Zeitschrift für Tourismuswissenschaft*, vol. 13, no. 2, pp. 309-310, 2021.
- [2] F. Chen, J. Zhang, Z. Wang, S. Shi, and H. Liu, "Passenger travel characteristics and bus operational states: a study based on IC card and GPS data in Yinchuan, China," *Transportation Planning and Technology*, vol. 42, no. 8, pp. 825-847, 2019.
- [3] M. Chen, J. Yang, L. Hu, M. S. Hossain, and G. Muhammad, "Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 14-20, 2018.
- [4] P. D. Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from social big data: implications for smart tourism destinations," *Information Processing & Management*, vol. 54, no. 5, pp. 847-860, 2018.
- [5] S.-K. S. Fan, C.-J. Su, H.-T. Nien, P.-F. Tsai, and C.-Y. Cheng, "Using machine learning and big data approaches to predict travel time based on historical and real-time data from Taiwan electronic toll collection," *Soft Computing*, vol. 22, no. 17, pp. 5707-5718, 2018.
- [6] I. Gallego and X. Font, "Changes in air passenger demand as a result of the COVID-19 crisis: using Big Data to inform

- tourism policy,” *Journal of Sustainable Tourism*, vol. 29, no. 9, pp. 1470–1489, 2021.
- [7] Y. Han and Y. Kim, “A study of measuring traffic congestion for urban network using average link travel time based on DTG big data,” *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 16, no. 5, pp. 72–84, 2017.
- [8] Z. H. A. O. Juanjuan, X. U. Chengzhong, and M. E. N. G. Tianhui, “Big data-driven residents’ travel mode choice: a research overview,” *ZTE Communications*, vol. 17, no. 3, pp. 9–14, 2019.
- [9] A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, “Verification and validation techniques for streaming big data analytics in internet of things environment,” *IET Networks*, vol. 8, no. 3, pp. 155–163, 2019.
- [10] B. Li, M. C. Kisacikoglu, C. Liu, N. Singh, and M. Erol-Kantarci, “Big data analytics for electric vehicle integration in green smart cities,” *IEEE Communications Magazine*, vol. 55, no. 11, pp. 19–25, 2017.
- [11] C. Llorca, J. Molloy, J. Ji, and R. Moeckel, “Estimation of a long-distance travel demand model using trip surveys, location-based big data, and trip planning services,” *Transportation Research Record*, vol. 2672, no. 47, pp. 103–113, 2018.
- [12] P. Nitu, J. Coelho, and P. Madiraju, “Improvising personalized travel recommendation system with recency effects,” *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 139–154, 2021.
- [13] C. M. Peak, A. Wesolowski, E. Zu Erbach-Schoenberg et al., “Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data,” *International Journal of Epidemiology*, vol. 47, no. 5, pp. 1562–1570, 2018.
- [14] J. Ren, X. Luo, L. Dong et al., “Analysis on spatial-temporal features of taxis’ emissions from big data informed travel patterns: a case of Shanghai, China,” *Journal of Cleaner Production*, vol. 142, no. 2, pp. 926–935, 2017.
- [15] P. Shi and Y. Long, “A research on complementary resources input decision of the travel agencies co-creating the O2O model,” *Tourism Science*, vol. 31, no. 2, pp. 55–68, 2017.
- [16] H. Tian, M. Presa-Reyes, Y. Tao et al., “Data analytics for air travel data: a survey and new perspectives,” *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–35, 2021.
- [17] A. I. Torre-Bastida, J. Del Ser, I. Laña, M. Ilardia, M. N. Bilbao, and S. Campos-Cordobés, “Big Data for transportation and mobility: recent advances, trends and challenges,” *IET Intelligent Transport Systems*, vol. 12, no. 8, pp. 742–755, 2018.
- [18] C. J. Wang, C. Y. Ng, and R. H. Brook, “Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing,” *JAMA*, vol. 323, no. 14, pp. 1341–1342, 2020.
- [19] H. Wang, X. Tang, Y. H. Kuo, D. Kifer, and Z. Li, “A simple baseline for travel time estimation using large-scale trip data,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–22, 2019.
- [20] R. Wang, J. X. Hao, R. Law, and J. Wang, “Examining destination images from travel blogs: a big data analytical approach using latent Dirichlet allocation,” *Asia Pacific Journal of Tourism Research*, vol. 24, no. 11, pp. 1092–1107, 2019.
- [21] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, “Big data analytics in intelligent transportation systems: a survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2018.
- [22] D. H. Huang, D. Liu, M. Wen, X. L. Dong, M. Wen, and X. H. Zhao, “A clustering method of Gas load based on FCM-SMOTE,” in *E3S Web of Conferences* vol. 257, EDP Sciences, Article ID 01032, 2021.