JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, 681-695 (2022)
DOI: 10.6688/JISE.202205_38(3).0011

Analyzing Tourist Behavior in Virtual Museums Using Intelligent Approach with Feature Selection

KRIT SRIPORN[1], CHENG-FA TSAI[2,*] AND PAOHSI WANG[3]
[1]*Department of Digital Technology*
Suratthani Rajabhat University
Surat Thani, 84100 Thailand
[2]*Department of Management Information Systems*
National Pingtung University of Science and Technology
Pingtung, 91201 Taiwan
[3]*Department of Food and Beverage Management*
Cheng Shiu University
Kaohsiung, 83347 Taiwan
E-mail: {krit.sri}@sru.ac.th; {cftsai}@mail.npust.edu.tw; {0627}@gcloud.csu.edu.tw

The Thai tourism sector has been significantly influenced by recent advances in multimedia technologies that help promote tourism, with Virtual Reality being exemplary. This study analyzes and predicts the behavior of tourists visiting the Virtual Chaiya National Museum to develop standards of quality for its products and services. Virtual Reality is useful in attracting tourists as it applies the concept of gamification using multimedia approaches. A sample of 580 tourists was used, and machine learning techniques were employed to predict the tourists' behavior. The results showed an accuracy of 99.48% using particle swarm optimization with random forest, followed by 99.45% using genetic algorithm search, and 99.13% using best-first search. Methods of feature selection were used with an Apriori algorithm to render the search for the rules governing tourist behavior more efficient. Particle swarm optimization with the Apriori algorithm yielded an effective confidence of 1.0. Tourist satisfaction with the quality of service at the Virtual Chaiya National Museum was 81.2%. The findings here can be used to choose the optimum dataset of features to create an effective prediction model and generate rules that can be used to describe tourist behavior.

Keywords: tourist behavior, game design, virtual reality, virtual museum, cultural tourism, machine learning, feature selection, arts and heritage, sustainable cities and communities

1. INTRODUCTION

Many applications make use of multimedia technology in a variety of areas, such as education. Multimedia technology impacts the productivity of instructional media in education [1]. Instructions delivered via innovative approaches to multimedia can also yield positive outcomes in schools as it advances logical and progressive improvements in educational training [2]. In addition, multimedia technology for tourism through films and advertisements can attract tourists and thus support the tourism industry. Systems are available within the tourism sector to provide advisory services that make use of multimedia information channels [3]. An example is Portugal's wine industry, which has developed tourism products by making use of Virtual Reality (VR) technology to offer globally competitive marketing that draws tourists [4]. The use of proactive tourism models for the

Received May 22, 2020; revised December 21, 2020 & February 23, 2021; accepted March 23, 2021.
Communicated by Hung-Yu Kao.

tourism industry can also increase travel revenue and mitigate harm to the marine environment by using virtual tours [5].

Thailand's National Tourism Policy seeks to promote Thai destinations to both domestic and foreign tourists to stimulate the economy of the country. As tourism strategy, it aims to develop quality standards for tourism products and services [6]. Moreover, VR is used increasingly commonly to disseminate information, and museums frequently use it to distribute information on the national culture [7]. VR is also an effective method to attract more tourists to museums by using image processing technology [8-10]. Nowadays, tourists are more interested in visiting museums because of the use of VR technology to provide VM. VR can utilize the opinions of tourists to design the VM to be more attractive to them. VR in case of national museum is important to promote for tourist. So, the development concept of researcher is to run the VM by using game design. Tourists feel fun when they can control the VM. The usage of the VM is similar to use computer game [11-13].

The VM as a part of the library contains a large amount of information from multidimensional data. VM using game design can attract users as well. The aim of the research study is to show the effectiveness of data mining and discuss its applications in the field of travel and tourism [14, 15]. The Department of Tourism in Gujarat would consider the use of machine learning to provide selected visitors with suitably-designed marketing materials, while decision trees can be sued to support a better understanding of visitor preferences in order to build stronger relationships with those visitors and boost visitor loyalty. This study is conducted using a sample group of 507 visitors [16]. Gabriel *et al.* [17] seeks to identify those tourist segments which exhibit similar perceptions of destination sustainability along with trustworthiness by employing a genetic algorithm. This research made use of a sample group of 438 respondents to an online survey. Krittipat *et al.* [18] used a machine learning approach in order to develop the association rule to ascertain activity patterns in the field of orchard tourism. The data were obtained from emails and surveys, and a total of 409 responses were used for analysis. Seonjeong *et al.* [19] investigated the use of VR, considering the potential industry applications of using VR features to influence consumers to exhibit desirable customer behavioral traits. A tracking application was used by Qiushi *et al.* [20] to study the movement trajectories of tourists and to gather data concerning travel itineraries and the behavioral characteristics of those tourists. The study then made use of machine learning to discover spatiotemporal behavior clusters among groups of tourists. Zhang *et al.* [21] made use of particle swarm optimization and support vector machine to study the behavior of urban eco-sports tourists.

The aim of this paper is to create classification model of predict tourist behavior, which can be applied to guide the selection of tourist to promote national museum with VCNM, and to use association rules to find the relational feature of VCNM for performance improvement of VCNM to promote national museum. VR in case of Chaiya National Museum is important to promote for tourist. So, the development concept of researcher is to run the virtual museum by using game design. Tourists feel fun when they can control the virtual museum. The usage of the virtual museum is similar to use computer game. There are several factors which are useful for study and can promote tourism in case of Chaiya National Museum that is analyzed by questionnaires from tourists to know about characteristics and significant features that can create and design the virtual museum by using machine learning techniques to find the characteristics of dataset influence to determine tourists' behavior and the relationships among features for helping the Virtual Chaiya

National Museum (VCNM) to become more interesting.

The research makes a significant contribution in providing VCNM for the heritage sector and by providing the necessary support to create VMs which are educational, fun, sustainable, simple to maintain, and can readily predict the behavior of tourists who visit the national museum by using machine learning with feature selection, which can be used to enhance the prediction model performance.

2. DATASET AND DESIGN OF THE VIRTUAL CHAIYA NATIONAL MUSEUM

2.1 Collected Dataset

This study examines the effectiveness of different methods of data classification when combining feature selection techniques with the wrapper method to make behavioral predictions concerning tourists to VCNM. "Weka" is used as software to process the data for machine learning through such process as data preprocessing, classification, clustering, association rule, and feature selection [22].

Table 1. Structure of the online questionnaire.

No.	Attributes	Values
1	Gender	{Male, female}
2	Age	{Below 25 years, 25–30 years, 30–40 years, 40–50 years}
3	Region	{South, north, central region, northeast, east, west}
4	Career	{Government officer, Business-man, Company employee, Stu-dent}
5	Have you ever visited the Chaiya National Museum?	{Ever, Never}
6	Have you ever played computer games?	{Ever, Never}
7	Have you ever used a system similar to the VCNM before?	{Ever, Never}
8	The VCNM is easy to use.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
9	The VCNM is convenient to use.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
10	The VCNM helps me see antiques more easily.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
11	The VCNM can display antiques clearly.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
12	Enjoy using the VCNM.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
13	The use of the VCNM is a good idea.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
14	Intend to use the VCNM.	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
15	Good interactive design	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}

No.	Attributes	Values
16	Overall satisfaction of using the VCNM	{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}
17	The VCNM can **promote tourism**. (Target label)	{Agree, Disagree}

Table 2. Abbreviations of the dataset of features.

The abbreviation	The features
Gender	Gender
Age	Age
Region	Region
Career	Career
Visit	Have you ever visited the Chaiya National Museum?
Computer game	Have you ever played computer games before?
Ever use	Have you ever used a system similar to the VCNM before?
Easy to use	The VCNM is easy to use.
Convenient	The VCNM is convenient to use.
Easy to see	The VCNM helps me see antiques more easily.
Display	The VCNM can display antiques clearly.
Enjoy	Enjoy using the VCNM.
Good idea	The use of the VCNM is a good idea.
Intend	Intend to use the VCNM.
Design	Good interactive design.
Promote	The VCNM can promote tourism.

Moreover, several data mining algorithms developed by the University of Waikato in New Zealand are used. The VCNM is generated by 3D blender [23], a well-known open-source tool for motion tracking, rendering, 3D modeling, and 3D game simulation. Only online questionnaires were used to gather information from 580 respondents on the VCNM as shown in Tables 1 and 2.

2.2 VCNM Design

The VCNM incorporates traditional concepts of the Chaiya National Museum through VR and game design. It features informed 2D and 3D images as well as details of the artifacts depicted therein, as shown in Fig. 1.

Fig. 1. The environment of the VCNM.

3. METHOD

The methods of this research can be divided into three parts as shown in Fig. 2. The first part is the feature selection method. The second part is classification. The third part consists of association rules to determine correlations among the attributes.

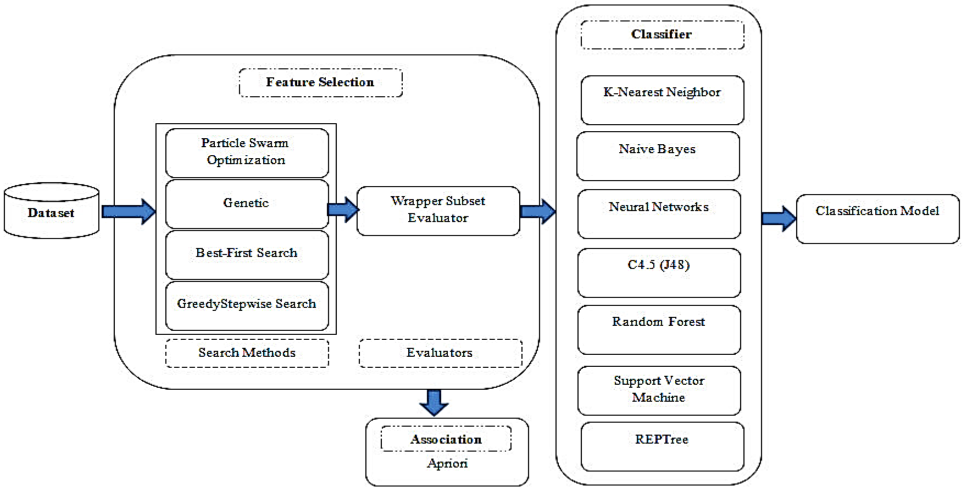

Fig. 2. The data analysis process.

3.1 Feature Selection

The process of feature selection involves choosing the optimum features of the dataset [24]. The chosen features can make the predictive modeling more accurate as they can be quickly processed. The process comprises four steps: generation of the subset, evaluation, criteria for stopping, and validation of the results.

3.1.1 Best-first search (BFS)

One effective way of performing this process is through the BFS together with depth-first search along with a heuristic function. This approach enables the selection of the best value by applying a greedy algorithm. In this case, the time efficiency is $O(bm)$, where b is the average number of nodes in each node, and m is the depth of the deepest tree [25].

3.1.2 Greedy stepwise search (GSS)

The use to the GSS offers a step-by-step approach to decision making that considers solutions that are initially available and compares them with those deemed the most valuable. The highly valued results are then applied to generate the solution to the optimization problem [26].

3.1.3 Genetic algorithm (GA) search

The GA works by first choosing a subset from a classification containing a binary string known as a “chromosome.” It uses fault tolerance when formulating models for challenging or complex data. It also uses natural imitation techniques, which begin by defining the problem in terms of genes and chromosomes, and then assigns a fitness function as the basis for finding answers, which consists of important parameters [27].

3.1.4 Particle swarm optimization (PSO)

The PSO searches for the target by using a large number of particles that search the space for the best target. Each particle remembers it's given position (x_i), direction, and velocity (v_i). In every cycle, the time (t) and velocity (v) of the particles are transformed using the best position of the data of each (p_i) as well as the best position of the data of all particles (p_g). C_1 and C_2 are the fixed rates or plus rates, and θ_1 and θ_2 are random numbers in the range 0–1. The velocity of each particle is then transformed as follows [28], as shown in the following equation:

$$v_i(t+1) = v_i(t) + C_1\theta_1(p_i(t) - x_i(t)) + C_2\theta_2(p_g(t) - x_i(t)). \tag{1}$$

3.2 Wrapper

A subset of the original features is selected. The method for evaluating the suitability of this subset is as follows: The wrapper method is applied. Moreover, the set of assessed values is used in the learning process to create a learning model and test a data model, which improves the performance of metaheuristics or feature subset selection algorithm to investigate the best feature subset [29].

3.3 Classification Algorithms

3.3.1 K-nearest neighbor (K-NN)

This method is classified as belonging to the lazy mode group, and uses the Euclidean distance function to determine the distance x, y in real values to measure similarity [22], as shown in the following equation:

$$k(x,y) = 1 - 2\sum_{i=1}^{m}\frac{(x_i - y_i)}{x_i + y_i}. \tag{2}$$

3.3.2 Naïve Bayes

The Bayes' theory for simple events assumes that A and B are events. Then, the conditional probability $P(A|B)$ according to Bayes' theory [22] is as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{3}$$

3.3.3 J48

J48 uses information gain or entropy reduction to select the optimal distribution. Suppose there is a variable X with a possible value K with probabilities p_1, p_2, p_3, p_4, …, p_k respectively. The minimum number of bits to send are called the entropy of X, where P_j

represents the proportion of the number of members in group j with the total number of members of the sample, and X represents attributes that measure entropy [22], as follows:

$$H(X) = -\sum_{i} p_j \log_2(p_j). \tag{4}$$

3.3.4 Neural network

The neural network has continuous features, because of which the characteristics of its input and output are in the range 0–1. The input (X_i) to the dataset is combined in a function, where W_{ij} represents the weight associated with input i to j, and takes i + 1 to j, and X_{ij} represents the input nodes i to j, and f represents the activation function, and b_i represents the bias using the following formula [30]:

$$y_i = f\left(b_i + \sum_{n}^{j=1} x_j \times w_{ij}\right). \tag{5}$$

3.3.5 Support vector machine (SVM)

The SVM has many advantages as it can handle the problem of over-fitting. Furthermore, it can transform lower-dimensional data into higher-dimensional data through the linear division model [22]:

$$f(x) = sign(W^T X + b). \tag{6}$$

3.3.6 Random forest

The random forest uses a supervised learning approach to learn a large number of datasets for classification. A learning curve is derived from the sampling function of the subset that replaces the original dataset of an identical size. This increases the efficiency of learning [31]. The random forest is a classification decision tree that uses criteria based on the sampling process using a precision probability technique.

3.3.7 REPTree

The reduced error pruning tree (REPTree) can speed-up the creation of regression trees by using pruning techniques, and can reduce pruning errors by applying variance reduction and information gain. REPTree is becoming increasingly popular in machine learning models [32]. Information gain is the basis for decision tree techniques, and the REP algorithm creates learning tree datasets.

3.4 Confusion Matrix

The classification result is shown in the confusion matrix, which is divided into two classes. Each value is shown in each row to indicate how much information is contained within the label's class. The model's output was estimated using these parameters using true positive (TP), which indicates the predictive results are positive and the real value is positive, and true negative (TN), which means the expected result is negative and the real value is negative. Moreover, false positive (FP) indicates a significant predictive outcome with a negative real value, while false negative (FN) indicates a negative predictive result with a positive real value. Eq. (7) was based on analysis of [33, 34], as shown in Table 3.

Table 3. Confusion matrix.

Confusion Matrix	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

The ratio of positives which are classified correctly (*i.e.*, the ratio of anyone that has a condition, which are accurately indicated the condition) is referred to as the true positive rate in Eq. (7) was shown below.

$$\text{TP Rate} = \frac{TP}{TP + FN} \tag{7}$$

The ratio of the number of negative events incorrectly classified as positive (false positives) and the total number of real negative events is used to measure the false positive rate in Eq. (8) was shown below.

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

The accuracy measures the ratio of expected to real values, regardless of whether the sample is positive or negative in Eq. (9) was shown below.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{9}$$

This study uses attributes 1-16th which serve as the feature input (x) in order to determine attribute 17th or class name "promote", which serves as the output (y) class, or target label (each element (label) in y consists of Agree or Disagree). The aim is to predict the opinion of visitors in using VCNM, which can be seen in Table 1. If T represents a set of the examples used for training, for which each is in the format (x, y) for which (x_1, …, x_k, y) x_k indicates the value for the kth attribute or feature associated with example x, and y serves as the corresponding label for the class. The gain in information for a given attribute k, can be explained in terms of entropy as shown, for any given value, v, which is taken by the attribute k [29].

$$S_k(v) = \{X \in T | x_k = v\} \tag{10}$$

3.5 Association Rules

Association rules are used to deal with latent relationships in data or to find patterns that are frequent. They can be written in the form of a set of items that results in another set of items.

Apriori algorithm

The Apriori algorithm is the basic method used to form a data relationship. It generates and monitors frequently occurring item sets of data in a transactional database. The correlation among the data must be consistent with the results of the minimum support and minimum confidence, which can be determined by the user [35].

$$Support = P(A \cap B) = \frac{\text{number of transactions containing both } A \text{ and } B}{\text{total number of transactions}} \quad (11)$$

$$Confidence = \frac{P(A \cap B)}{P(A)} = \frac{\text{number of transactions containing both } A \text{ and } B}{\text{number of transactions containing } A} \quad (12)$$

3.6 Tourists' Satisfaction

The average value of satisfaction with the VCNM when visiting the Chaiya National Museum was calculated by using $\sum$ is summation of the weighted mean value and N is the total number of respondents, which are variable of mean formula ($\bar{X}$). For standard deviation formula, *S.D.* is Standard deviation and X is the weighted mean, as follows [36]:

$$\bar{X} = \frac{\sum X}{N} \text{ and } S.D. = \frac{\sqrt{\sum\left(X - \bar{X}\right)^2}}{N-1}. \quad (13)$$

4. RESULTS

Table 4 describes the results of the dataset features obtained using a feature selection method suitable for each classification algorithm to increase the predictive potential of the model of tourists' behavior. Tables 6 and 7 illustrate the association rules for features of the dataset.

Table 4. The results of PSO combined with wrapper.

Measure/ PSO&Wrapper	The features
Random Forest	Region, Computer game, Ever use, Easy to use, Convenient, Easy to see, Good idea, Design, Promote
Neural Networks	Region, Computer game, Ever use, Convenient, Easy to see, Enjoy, Good idea, Design, Promote
J48	Career, Computer game, Ever use, Convenient, Good idea, Design, Promote
REPTree	Ever use, Convenient, Enjoy, Good idea, Design, Promote
SVM	Region, Convenient, Enjoy, Design, Promote
K-NN	Ever use, Easy to use, Convenient, Enjoy, Good idea, Intend, Design, Promote
Naïve Bayes	Gender, Age, Ever use, Easy to use, Convenient, Intend, Design, Promote

Table 5. Accuracy of classification models with and without feature selection.

Accuracy	Random Forest	Neural Networks	J48	REPTree	SVM	K-NN	Naïve Bayes
GSS & Wrapper	98.61	98.44	98.10	97.75	97.23	94.64	92.57
BFS & Wrapper	99.13	98.96	98.79	98.61	97.23	95.50	93.09
GA & Wrapper	99.45	99.13	98.96	98.79	97.23	95.85	94.64
PSO & Wrapper	99.48	99.48	99.30	98.79	97.23	96.02	94.99
Non-use feature selection	89.11	88.60	87.73	87.73	89.11	84.62	84.80

Table 6. Association rules of the dataset obtained through feature selection using PSO combined with wrapper: the 10 best rules.

No.	Frequent item set	Confidence	Support
1.	Enjoy = Neutral ==> Easy to see = Agree	1.00	0.53
2.	Convenient = Agree ==> Overall satisfaction = Agree	0.93	0.59
3.	Good idea = Agree ==> Overall satisfaction = Agree	0.90	0.57
4.	Computer game = Ever ==> Overall satisfaction = Agree	0.89	0.68
5.	Easy to see = Agree ==> Overall satisfaction = Agree	0.88	0.56
6.	Ever use = Never ==> Overall satisfaction = Agree	0.86	0.60
7.	Easy to see = Agree ==> Enjoy = Neutral	0.83	0.53
8.	Overall satisfaction = Agree ==> Computer game = Ever	0.77	0.68
9.	Overall satisfaction = Agree ==> Ever use = Never	0.69	0.60
10.	Overall satisfaction = Agree ==> Convenient = Agree	0.67	0.59

Table 7. Association rules of the dataset obtained without feature selection.

No.	Frequent item set	Confidence	Support
1.	Convenient = Agree ==> Promote = Agree	0.93	0.59
2.	Good idea= Agree ==> Promote = Agree	0.90	0.57
3.	Visit = Never ==> Promote = Agree	0.89	0.65
4.	Easy to use = Agree ==> Promote = Agree	0.89	0.59
5.	Computer game = Ever ==> Promote = Agree	0.89	0.68
6.	Easy to see = Agree ==> Promote = Agree	0.88	0.56
7.	Gender = Woman ==> Promote = Agree	0.87	0.55
8.	Ever use = Never ==> Promote = Agree	0.86	0.60
9	Ever use = Never ==> Visit = Never	0.79	0.56
10.	Promote = Agree ==> Computer game = Ever	0.77	0.68

Table 8. Mean and standard deviation of satisfaction among tourists with the quality of service of the Virtual Chaiya National Museum (*n* = 580).

No.	Feature	$\bar{X}$	SD	Level of satisfaction
1	Easy to use	4.08	0.54	Agree
2	Convenient to use	4.05	0.61	Agree
3	See antiques more easily	4.03	0.65	Agree
4	Antiques clearly	3.98	0.73	Agree
5	Enjoy	3.61	0.93	Agree
6	A good idea	4.07	0.57	Agree
7	Intend to use	3.98	0.90	Agree
8	Good interactive design	3.64	0.73	Agree
9	Overall satisfaction	4.06	0.80	Agree
Total	Average	3.94	0.71	Agree

5. DISCUSSION

Table 5 shows a comparison of the results of predictions of the tourists' behavior between the methods when the feature selection algorithm was used and when it was not. The random forest and the SVM obtained the highest accuracy, 89.11%. Neural networks yielded the second-highest accuracy. The parameters that may have a substantial effect on the effectiveness and efficiency of the search should be calibrated for every metaheuristic algorithm. Tuning of Parameter may allow a robustness and larger flexibility. The efficiency of finding key features of the dataset depended on the parameters of each method. The time consumption depended on the parameter values and dataset size [37], as shown in Table 9. As shown in Table 8, the VCNM obtained a mean value of 81.2% in terms of overall satisfaction for tourists.

Table 9. Parameters of each feature selection method.

Methods	Parameters
PSO	Population size, iterations, individual weight, inertial weight, social weight, and mutation probability
GA	Crossover probability, mutation probability, and population size
BFS	Directions forward, lookup cash size, search termination
GSS	Conservative forward search, threshold value, backward search, and producing a ranked list of attributes

Table 10. The best results of features of the dataset for each feature selection method.

Methods	The features
GSS & Wrapper	Career, Computer game, Easy to use, Convenient, Good idea, Design, Promote
BFS & Wrapper	Region, Visit, Computer game, Easy to use, Convenient, Easy to see, Enjoy, Good idea, Design, Promote
GA & Wrapper	Gender, Career, Computer game, Ever use, Convenient, Good idea, Design, Promote
PSO & Wrapper	Region, Computer game, Ever use, Easy to use, Convenient, Easy to see, Good idea, Design, Promote

A local search algorithm that begins from a possible solution that is then duplicated to a neighbor solution is called the GSS. So long as a neighborhood relation is expressed in the search space, it is feasible [26]. Compared with alternative approaches, the GSS obtained a lowest accuracy of 98.61%, as shown in Table 5.

An approach used to solve a problem faster when other approaches are too slow, or for getting an estimated solution when typical methods cannot find a precise solution is called BFS. This method can be introduced through a priority queue, a data structure that preserves the parameter in increasing order of F-values, within a general search system [25]. As shown in Table 5, with an accuracy of 99.13%.

The GA obtained an accuracy of 99.45%, and used 8 significant features as shown in Table 10. The values of the second to fourth methods are also shown in Table 5. Of the three key classes of processes (selection, crossover, and mutation). Typically, GA operators cannot create all possible solutions as the population density in the solution space is low. Using local search to find a better solution is a popular fix for GA operators [27].

PSO obtained a highest accuracy of 99.48%. As shown in Table 5, it used 9 significant features. The ideal particle had a strong impact on the entire swarm. The best solution obtained by PSO was not always certain as it does not use selection. The particles tended to be representative of the population for the entire run. Its origin determined a particle's path. PSO can be used to solving non-linear as well as linear problems, PSO has been used for developing robots to assist patients with bone fractures [28].

These metaheuristics have one particularly advantageous property, in requiring that the accuracy estimation process must carry of the cross-validation procedure [29] far more often when using small datasets than would be the case for large datasets. Small datasets demand shorter learning times, so the time taken for accuracy estimation, comprising both the running time for these metaheuristic algorithms.

6. CONCLUSION

The aim of this paper was to create a standard of quality for tourism products and services using VCNM, and to use machine learning techniques such as Random Forest classification model using PSO combine with wrapper as feature selection method to predict tourist behavior. It is important to determine key features of the VCNM that can promote tourism and encourage tourists to visit the place being simulated. The product of tourism is the VCNM, and the service offered is satisfaction as reflected in tourist behavior. The quality of the VCNM was reflected in its ability to show antiques easily to users without distortion. The details covered anastylosis, archaeological evidence, ancient architecture, inscriptions, ruins, excavations, eras, periods, and processes of reconstruction. Several machine learning techniques were used in this research. Feature selection was used to identify characteristics of the dataset used to determine tourist behavior. The second is the classification technique used to predict tourist behavior. Third, it was necessary to find the relationships among features. Features influence the satisfaction and behavior of tourists' through association rules. Fourth, the use of the VCNM simulation makes it possible to look around in the museum environment, and this can be used to promote museums. Fifth, the use of game design can help the VCNM be more interesting to users. Finally, the design of the user interface should be considered as it can make the VCNM more interesting and convenient.

The VCNM is user friendly, easy to use, and can save time and motivate tourists. It also motivated visitors to see artifacts in the Chaiya National Museum. In summary, develop tourism products or VCNM until it up to the standard and always concern and recognize to promote the Chaiya National Museum using VCNM. The curator and tourists had new experience by using VCNM that is support and promote conservation tourism follow by Thailand's national tourism policy.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the anonymous reviewers for their useful comments and suggestions for improving the quality of this paper, and we thank Ministry of Science and Technology, Taiwan, for supporting this research. This research was funded by the Ministry of Science and Technology, Republic of China, Taiwan, grant numbers MOST-107-2637-E-020-006, MOST-108-2637-E-020-003, MOST-108-2321-B-020-003, and MOST-107-2321-B-020-005.

REFERENCES

1. C. W. Park, D. Kim, G. Cho, and H. J. Han, "Adoption of multimedia technology for learning and gender difference," *Computers in Human Behavior*, Vol. 92, 2019, pp.

288-296.

2. H. Hartman and P. Johnson, “The effectiveness of multimedia for teaching drug mechanisms of action to undergraduate health students,” *Computers and Education*, Vol. 125, 2018, pp. 202-211.
3. Z. Xiang, “From digitization to the age of acceleration: On information technology and tourism,” *Tourism Management Perspectives*, Vol. 25, 2018, pp. 147-150.
4. J. Martins, R. Gonçalves, F. Branco, L. Barbosa, M. Melo, and M. Bessa, “A multisensory virtual experience model for thematic tourism: A port wine tourism application proposal,” *Journal of Destination Marketing and Management*, Vol. 6, 2017, pp. 103-109.
5. C. Martella, A. Miraglia, J. Frost, and M. van Steen, “Visualizing, clustering, and predicting the behavior of museum tourists,” *Pervasive and Mobile Computing*, Vol. 38, 2017, pp. 430-443.
6. Sirindhorn Anthropological Center Thailand, “The status of museums in Thailand,” http://www.sac.or.th, 2020.
7. J. Pallud, “Impact of interactive technologies on stimulating learning experiences in a museum,” *Information and Management*, Vol. 54, 2017, pp. 465-478.
8. P. Kyriakou and S. Hermon, “Can I touch this? Using natural interaction in a museum augmented reality system,” *Digital Applications in Archaeology and Cultural Heritage*, Vol. 12, 2019, pp. 1-9.
9. R. K. Napolitano, G. Scherer, and B. Glisic, “Virtual tours and informational modeling for conservation,” *Journal of Cultural Heritage*, Vol. 29, 2018, pp. 123-129.
10. J. Swarbrooke and S. Horner, “Main concepts in consumer behaviour, including models of consumer behaviour adapted for tourism,” *Consumer Behaviour in Tourism*, Elsevier Publisher, London, 2007, pp. 43-46.
11. T. Messaoudi, P. Véron, G. Halin, and L. de Luca, “An ontological model for the reality-based 3D annotation of heritage building conservation state,” *Journal of Cultural Heritage*, Vol. 29, 2018, pp. 100-112.
12. G. Younes, R. Kahil, M. Jallad, D. Asmar, I. Elhajj, and H. Al-Harithy, “Virtual and augmented reality for rich interaction with cultural heritage sites,” *Digital Applications in Archaeology and Cultural Heritage*, Vol. 5, 2017, pp. 1-9.
13. C. Andujar, A. Chica, and P. Brunet, “User-interface design for the ripoll monastery exhibition at the National Art Museum of Catalonia,” *Computers and Graphics*, Vol. 36, 2012, pp. 28-37.
14. I. Malegiannaki and T. Daradoumis, “Analyzing the educational design, use and effect of spatial games for cultural heritage: A literature review,” *Computers and Education*, Vol. 108, 2017, pp. 1-10.
15. E. Rowe, J. Asbell-Clarke, R. S. Baker, M. Eagle, A. G. Hicks, and T. Edwards, “Assessing implicit science learning in digital games,” *Computers in Human Behavior*, Vol. 76, 2017, pp. 617-630.
16. B. K. Manoj, S. S. Ravinder, S. S. Diler, and R. K. Dhaval, “Data mining for travels and tourism,” *Journal of Information and Operations Management*, Vol. 3, 2012, pp. 114-118.
17. I. P.-L. Gabriel, R.-S. Carla, R.-M. Felipe, and T. Eduardo, “A machine learning approach to segmentation of tourists based on perceived destination sustainability,” *Journal of Destination Marketing & Management*, Vol. 19, 2021, pp. 1-9.

18. P. Krittipat and N. Parinya, "Data mining approach for arranging and clustering the agro-tourism activities in orchard," *Kasetsart Journal of Social Sciences*, Vol. 39, 2018, pp. 407-413.
19. A. L. Seonjeong, L. Minwoo, and J. Miyoung, "The role of virtual reality on information sharing and seeking behaviors," *Journal of Hospitality and Tourism Management*, Vol. 46, 2021, pp. 215-223.
20. G. Qiushi, Z. Haiping, H. Songshan, Z. Fang, and C. Chongcheng, "Tourists' spatio-temporal behaviors in an emerging wine region: A time-geography perspective," *Journal of Destination Marketing & Management*, Vol. 19, 2021, pp. 1-13.
21. Z. Fan, "Prediction and evaluation of urban eco-sports tourism behavior using data mining technology," in *Proceedings of the 4th International Conference on Big Data and Computing*, 2019, pp. 68-71.
22. F. Eibe, A. H. Mark, and H. W. Ian, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, MA, 2011, pp. 145-147.
23. R. Hess, "Blender − a 3D modelling and rendering package," https://www.blender.org, 2020.
24. H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, 2005, pp. 491-502.
25. W. Lam, K. Kask, J. Larrosa, and R. Dechter, "Subproblem ordering heuristics for AND/OR best-first search," *Journal of Computer and System Sciences*, Vol. 94, 2018, pp. 41-62.
26. C. Eiras-Franco, V. Bolón-Canedo, *et al*., "Multithreaded and spark parallelization of feature selec-tion filters," *Journal of Computational Science*, Vol. 17, 2016, pp. 609-619.
27. D. Wang, H. Xiong, and D. Fang, "A neighborhood expansion tabu search algorithm based on genetic factors," *Open Journal of Social Sciences*, Vol. 4, 2016, pp. 303-308.
28. I. Sancaktar, B. Tuna, and M. Ulutas, "Inverse kinematics application on medical robot using adapted PSO method," *Engineering Science and Technology*, Vol. 21, 2018, pp. 1006-1010.
29. K. Ron and H. J. George, "Wrappers for feature subset selection," *Artificial Intelligence*, Vol. 97, 1997, pp. 273-324.
30. F. Schmitt, R. Banu, I. Yeom, and K. Do, "Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater," *Biochemical Engineering Journal*, Vol. 133, 2018, pp. 47-58.
31. G. Cano, J. Garcia-Rodriguez, A. Garcia-Garcia, H. Perez-Sanchez, A. Thapa, and A. Barr, "Automatic selection of molecular descriptors using random forest: Application to drug discovery," *Expert Systems with Applications*, Vol. 72, 2017, pp. 151-159.
32. H. Erdal and İ. Karahanoğlu, "Bagging ensemble models for bank profitability: An empirical research on turkish development and investment banks," *Applied Soft Computing Journal*, Vol. 49, 2016, pp. 861-867.
33. K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, "Analyzing lung disease using highly effective deep learning techniques," *Healthcare*, Vol. 8, 2020, pp. 1-21.
34. K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, "Analyzing malaria disease using effective deep learning approach," *Diagnistics*, Vol. 10, 2020, pp. 1-22.

35. D. T. Larose, “Association rules,” *Data Mining Methods and Models*, John Wiley & Sons, NJ, 2005, pp. 180-184.
36. H. O. Kiess, “Using statistics for inference and estimation,” *Statistical Concepts for the Behavioral Sciences*, Allyn & Bacon, Boston, 1996, pp. 135-156.
37. E. G. Talbi, “Population-based metaheuristics,” *Metaheuristics: From Design to Implementation*, John Wiley & Sons, NJ, 2009, pp. 253-255.

Krit Sriporn (彭肯瑞) is a Lecturer in the Department of Digital Technology, Suratthani Rajabhat University, Thailand. He has published well-known journal papers and conference papers in the data analytics field. His research interests include artificial intelligence, machine learning, and deep learning.

Cheng-Fa Tsai (蔡正發) is a Professor in the Department of Management Information Systems, National Pingtung University of Science and Technology, Taiwan. He has published many well-known journal papers and conference papers. He holds, or has applied for, fifteen U.S. patents and thirty ROC patents in his research areas. Dr. Tsai research interests are in areas of artificial intelligence, with emphasis on machine learning, and data mining.

Paohsi Wang (王寶惜) is an Assistant Professor in the Department of Food and Beverage Management, Cheng Shiu University of Taiwan. She has published journal papers and conference papers. Dr. Wang research interests are in the areas of hospitality, artificial intelligence, with emphasis on efficient machine learning, and deep learning.