EAI Endorsed Transactions
on Tourism, Technology and Intelligence
Review Article EAI.EU

Machine Learning Applications in Smart Tourism: Overview, Research Challenges, and the Road Ahead

Premisha Premananthan[1], Hang Le[2], Minh-Hien T. Nguyen[3], M. Fahim[3], Trung Q. Duong[1,*]

[1]Memorial University, Canada
[2]Duy Tan University, Vietnam
[3]Queen's University Belfast, UK

Abstract

Tourism plays a vital role in stimulating economic growth, creating employment opportunities, and facilitating cultural exchange and mutual understanding between people of different backgrounds. However, as an industry facing increasing competition on a global scale, tourism needs to embrace technologies to adapt quickly to changes. The emergence of smart tourism offers a route to achieve a competitive edge. To this end, machine learning techniques are used to analyze the abundant data available, forecast trends, and provide best solutions. This paper provides an overview, challenges and future directions of research on machine learning techniques applied in smart tourism.

Received on 20 February 2024; accepted on 30 June 2024; published on 01 August 2024

Keywords: Analysis, Artificial intelligence, Machine learning, Smart tourism, Review

doi:10.4108/XX.X.X.XX

1. Introduction

Tourism plays an important role in each country's economy. Not only does tourism contribute to economic growth, it also expands the local job markets, shares the country's cultural values with other parts and maintains peace and harmony throughout the world.

The introduction of technologies within each sector entails vast changes towards smart applications. In this context, smart tourism comes with tremendous technological innovations to provide better customer experience through real-time personalization. The study in [1] focused on improving smart tourism through the Internet of Things (IoT) and machine learning techniques. The authors found a deviation between collecting datasets and analysing those data in an effective way. In order to resolve the gap, a smart tourism system using IoT and machine learning algorithms was implemented to customise travel recommendations for each user. A decision tree approach using environmental data acquired by wireless sensor networks to implement travel recommendations was proposed. As a result, the smart tourism service system was able to more correctly forecast typical congestion levels and outperform other methods. The study provided a theoretical foundation and served as a guideline for future smart tourism services.

On the other hand, machine learning is one of the most influential technologies that drives major industry development on a vast scale. Specially, the industries that involve a huge amount of data are the main beneficiaries. Machine learning techniques are used to predict future changes from large data sets, recommend new solutions within a shorter amount of time, and classify multiple groups without deep knowledge of the dataset. In tourism, a huge amount of data is flowing through multiple online platforms, which makes it very easy to grab data.

In this paper, we analyze and draw key points regarding machine learning techniques, their applications towards tourism sector improvements, challenges while developing machine learning models to solve the problems in tourism, and machine learning applications in smart tourism.

*Corresponding author. Email: ppremanantha@mun.ca

2. An Introduction to Machine Learning Techniques

Machine learning is one of the major component of artificial intelligence (AI) that aims at the implementation of algorithms that make computers learn from and develop forecasts or solutions based on data. The major goal is to enable machines to automatically learn patterns from data.

2.1. Unsupervised Machine Learning Techniques

Unsupervised approaches are useful when the labeled data is scarce or expensive to obtain. The ability to deliver accurate information on a particular tourist attraction is very important in tourism since there is a considerable amount of false information spreading over the internet. The authors in [2] worked on developing a separate platform to study the opinions regarding Phuket, the largest tourism destination in Thailand, by using latent Dirichlet allocation (LDA) algorithms. Through this system, anyone can easily understand the actual incidents without further knowledge of the particular place. This study allowed TripAdvisor's chat rooms to evaluate subjects using LDA topic extraction techniques in order to determine which topics were of interest to users and what trending topics were in comparison with the previous year. Then, in order to establish the parameters of the fascinating comment, they created a digital platform which specifies each of the issue along with the pertinent suggestion.

They also provide insights into sentiment without requiring manual annotation [3]. The study in [4] implemented a user recommendation system based on past travel histories. The user's reviews, personal information, and category tags used in the travel attractions were the focused data. After the preprocessing of this text data, the LDA algorithm was used to extract the topics and corresponding probabilities from the corpus of reviews. Using sentiment analysis, they valued each review in order to map the connections between tourist places. The study introduced a personalised recommendation system for a target user based on the attraction visit similarity, the similarity of review topic probability, and the similarity of review sentiment of other users.

Clustering is another technique used in unsupervised machine learning. The authors in [5] used fuzzy clustering analysis-related algorithms to evaluate the service quality of tourist places. Relying on the specifications of tourism services, this study implemented a confusion clustering model using fuzzy clustering. The authors used fuzzy comprehensive evaluation, which is based on the structural model, and compressed the idea of fuzzy operations. The majority of pertinent specialists and scholars acknowledge this type of assessment ideology and technology, which have evolved into new ways of ideology and routine approaches used by academia and even the relevant statistical sectors. By using layer-by-layer membership relations, multi-factor variables were bundled into a single comprehensive index at the end of the fuzzy comprehensive evaluation process.

2.2. Semi-supervised Machine Learning Techniques

In machine learning, semi-supervised learning is a relatively new technique. It is the part of machine learning techniques that is the combination of supervised and unsupervised learning, i.e., the use of labeled and unlabeled sample data simultaneously. The field of semi-supervised learning research is broadening these days. On the other hand, a variety of conventional supervised and unsupervised learning algorithms are continuously expanded or altered to enhance their capacity to learn in semi-supervised environments. In [6], a semi-supervised model was implemented to classify the Chinese traditional paintings Dunhuang with multidimensional features extricated by transfer learning. A large number of unlabeled samples were obtained using a small number of labeled datasets with a mixed method of active learning and label propagation following three stages: (i) initial stage, which is feature extraction, (ii) the classification methods, which include label propagation and support vector machine classifiers, and (iii) iterative algorithm, which includes active learning and label propagation. After several iterations, a more powerful classification learner was acquired.

With the development of the tourism industry, the number of information providers in this industry has also started to grow. However, there will still be a gap between the information needed by the tourists and the information available. The search system for tourism needs to be improved; rather than showing information from indexing websites, it should be fluctuant on the actual information about the tourist places. In order to overcome this issue, the study in [7] implemented Yet Another Two Stage Idea (YATSI), a recognition system based on semi-supervised learning using a naïve Bayes classifier. A component of information retrieval called named-entity recognition (NER) used unstructured text to elicit, gather, and portray data in a structured manner. The study suggested analyzing global entity candidates using extra features to increase the performance of the NER system. The selection amongst various classification algorithms applied to YATSI is important to get the best outcome.

2.3. Deep Learning Techniques

Deep learning is a very popular neural network-based technique that is implemented in an H2O framework to extract patterns and classes from information within data in a layered model. It works like a human

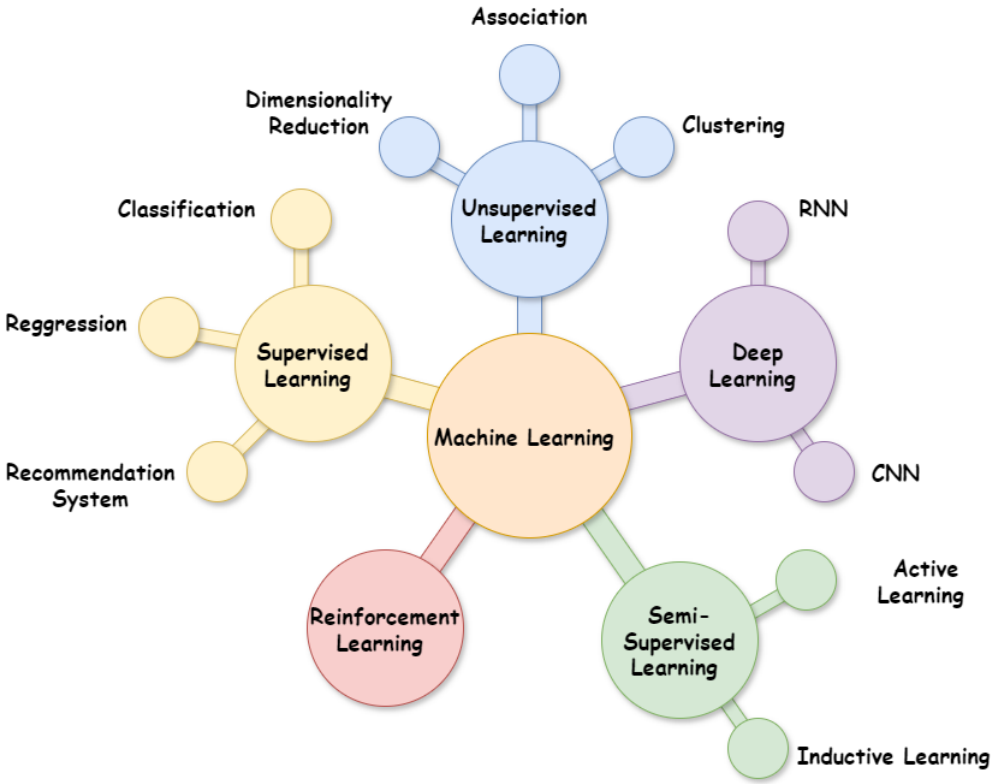

Figure 1. Machine Learning Techniques

brain where each neuron is modified based on the available information. It can also predict the output and act as a classifier. Deep learning excels at learning representations from complex and unstructured data, such as images, audio, and text.

Artificial neural networks (ANNs) are the foundation of deep learning. They have interconnected nodes called neurons, which are structured in layers. Those layers include an input layer that is compulsorily created, one or more hidden layers that can be customized based on our preferences, and an output layer. Weight is associated with each connection between neurons, which the network adjusts during training to optimize performance. One of the most common ANN types for text processing is the multi-layer perceptron design. The back-propagation technique is used during the training phase to determine the weights connected to each link. In [8], an ANN was used to explore the correlation between hotwords, which were recorded by the Google search engine, and the most visited tourist destinations. The ANN model was implemented using Google Tensorflow, with two hidden layers, each with 12 neurons, and the rectified linear unit as the activation function.

Accurate demand forecasting plays a vital role in the tourism industry in order to make strategic decisions and design relevant policies. Since tourism demand function is non-linear, deep learning will be a promising method for demand forecasting. In [9], a long short-term memory (LSTM) framework and gated recurrent unit (GRU) were used to focus on tourist arrivals in Morocco. These two models used two hidden layers, each of which contained 64 neurons. ANN and support vector regression (SVR) models were used as benchmarks to evaluate the LSTM and GRU models. The performance evaluation suggested GRU, LSTM, and a special type of recurrent neural networks (RNNs) as suitable for tourism demand prediction.

Convolutional neural networks (CNNs) are particularly effective for tasks involving images and spatial data. Convolutional layers can be utilized to automatically and adaptively learn spatial hierarchies of features. CNNs have been highly successful in tasks such as image classification, object detection, and image segmentation. Analyzing budgets and planning is the perfect way to measure the impact of tourism on society. Promotional videos play a vital role in promoting tourism. The authors in [10] predicted the social impact

of tourism in villages at a national level using promotional videos and images from the World Tourism Organization. A CNN model, with its deep learning architecture Visual Geometry Group (VGG)-16, was used to analyze and classify the collected dataset of video frames and images. The authors implemented the model using linear, Lasso, and Ridge regression models to analyze the socio-economic impacts, such as social media post views and GDP growth. This paper revealed a powerful correlation between image and video data and the countries' GDP growth rates. In addition, the proposed model is suitable for exploring the impact of image data as a predictive tool for studying tourism growth, particularly in rural areas.

2.4. Reinforcement Learning Techniques

Reinforcement learning is a very powerful algorithm where an agent sequentially learns and makes decisions by interacting with the environment without any supervision in order to maximize the rewards. This method is based on (i) exploration, which means the agent's interactions with the environment, and (ii) exploitation, which indicates the learning through and observations of how it responds, similarly to the trial-and-error method. In [11], a customized tourism route recommendation system based on reinforcement learning was deployed in Indonesia. The Markov decision process was used that involved using Q-learning to recommend routes to N-day travel destinations in the special region of Yogyakarta. The authors also considered the time interval and opening and closing hours of tourist attractions, and included time constraints into their methodology. In order to fulfill the user's unique intended goals and preferences, the technique also took into account a number of variables, including rating, cost, and journey time, as the cost functions used the multi-attribute utility theory. In several tests, the proposed approach was contrasted with the Firefly algorithm to evaluate its effectiveness and ascertain its optimality.

3. Role of Machine Learning in Tourism

Many real-world applications through research studies are adopted to overcome issues in the tourism sector. Specially after COVID there are lot of smart tourism related applications used in this domain. In this section, the major developments of applications in smart tourism will be discussed. Smart tourism aims at the development of tourism industry with information systems using high performance devices, cloud computing, artificial intelligence and administrative techniques. In the below discussion, the research studies contributed towards this development will be elaborated.

3.1. Automatic Classification in Tourism

Tourism was the world's third largest source of economic income after fuel and chemicals until COVID-19. The tourism sector's rehabilitation is pivotal for global economy now. In [12], the authors implemented a classification machine learning model using five machine learning methods, such as random forest, support vector machines, decision trees, artificial neural networks, and extreme gradient boosting, to predict whether a tourist is an international tourist or a non-international tourist. To this end, the online photos were collected to explore the international/internal nature of the tourist. Four main approaches used: (i) The online image's timestamp and geographical location showcasing whether the person is a resident or tourist; (ii) the person's native address(s): if the native address differs from the tourist place then the user is labeled a tourist; (iii) using a set of flexible detailors related to space and time while examining users' travel history according to their photographing line; and (iv) analyzing the geolocation in which the images were captured, i.e., tourist attractions, without reference to who uploaded them. This study aimed to provide a way to enhance tourist attractions and its organization by providing both qualitative and quantitative data related to travelers' travel manners and pursuits. Therefore, they used the image details, including geolocations, crowdsourced, and user-generated content, to contribute to the automatic classification of those above-mentioned types of user identification.

3.2. Recommendation Systems in Tourism

Following the successful implementation of recommender systems in the domain of e-commerce, and currently, in the tourism domain, researchers are expanding their methods for recommendations without spending more time and resources on a particular decision [4]. The study in [13] implemented multi-criteria ratings over single-criterion rating feedback, which was used to differentiate the same users in traditional collaborative filtering (CF) recommender systems in the past. The authors proposed a recommendation model for an existing Tripadvisor dataset using neuro-fuzzy inference systems and support vector regression (SVR). To reduce the dimensions of the data, they used principal component analysis. Self-organizing maps and expectation maximization were used as clustering techniques, which is the combination of multiple partitions of a dataset by a consensus function, yielding a final partition. Apart from clustering techniques, the authors proposed multi-criteria CF, a cluster ensemble approach, and the hypergraph partitioning algorithm to increase the accuracy of the recommendation system.

This evaluation model was proven to work, both online and offline, for current, existing Tripadvisor data.

TripMa is one of the real-world applications in smart tourism using machine learning techniques that overcomes the lack of knowledge gap about local tourist places in Sri Lanka. The authors in [14] proposed a system to suggest the best travel destinations with full package based on previous reviews, safety precautions, perfect time and stay of that place. This system was developed by using natural language processing with real-time data extraction. The authors continuously trained their model using the feedback from the previous users of the system, while real-time data accessibility helped to mitigate the risks of the tourists at the same tourist place.

Nowadays, the internet makes the data available and easily accessible. This is particularly true in the tourism industry, when planning a trip to a place or even any trip, there is plenty of data available on the internet. The issue, however, is whether that information is accurate or how frequently it is updated. To overcome this, the authors in [15] developed an intelligent travel route planning system using machine learning algorithms through big data, sensor techniques and internet map services. This is an iterative system between users and the system where the information is updated frequently, and this system's response time could be slower by 0.0005 seconds when every additional 100,000 tourists are added to the system. This proposed system was experimentally deployed from 2016 to 2020 in a province of China. Through this route planning system, the users can easily solve their preference conflicts, information overload and authenticity with a customized budget friendly trip package.

3.3. Demand Forecasting in Tourism

Accurate demand predictions are invaluable to both customers and businesses as well. In the empirical study in [16], a novel deep learning exploration of demand forecasting in tourism analyzed both advanced technology and the effects of multi-dimensional factors on human behaviors. Their research output provided an interpretation of tourism demand from both factors and temporal persistence patterns. The study was performed in two different tourist places before and during COVID-19. There are six factors that severely depend on tourism demand: accommodation, transportation, weather, travelogues, tickets, and date patterns. The authors proposed a three-stage method to analyze the factors that influenced tourism, which are tourism demand forecasting, temporal heterogeneous feature preparation, and forecasting results interpretation. First, the temporal fusion encoder-decoder network (TFED) acquired perceived past and familiar future features and predicted tourism demand. Then, Bayesian optimization was implemented with five major parts: the variable selection network, the gated residual network, long short term memory (LSTM) encoders and decoders, the interpretable multi-head attention layer, and the forecast output to discover optimal parameters.

Another study in [17] was conducted in demand forecasting to overcome the financial risk to cruise tourism in China. The authors implemented a least squares support vector regression model with a gravitational search algorithm (LSSVR-GSA) to improve forecasting performance. The big dataset with search query data from Baidu and economic indexes was used in this model. This proposed model outperformed selected mobile keywords and economic indexes and achieved the highest forecasting performance. The output showed that modeling with big data can achieve better forecasting performance compared to those using alternative data sources.

Another milestone of demand forecasting in tourism was to predict customer demands for space tourism. The study in [18] was analyzed within the United States in order to get their opinions about space tourism in the future. Space tourism will be expected to conquer 18–26% of the market for tourism between 2020 and 2030. The researchers primarily focused on three prominent areas, which are suborbital, orbital, and space tourism around the moon, based on public opinions regarding security, cost, quantity of passengers, training, and take-off locations, among other decision-maker factors. A multi-model machine learning prediction model was used to predict the output of the collected dataset of 2000 rows and 40 features. In this model, a decision tree, an extremely randomized ensemble classifier, a random forest algorithm, a gradient boosting classifier, and a multi-class model were used. In the tourism industry, these suggestions towards United States people's opinions on future moves towards mass space tourism were an extreme move.

To conclude this section on the applications of tourism and machine learning techniques, we analyze the challenges which are identified through the past studies. First of all, the privacy and security of the data is a concern. Smart tourism systems often collect vast amounts of personal data, including location, preferences, and behaviors which will potentially lead to unwanted tracking or profiling. This raises concerns about how this data is stored, used, and shared. Many smart tourism devices rely on internet connectivity, which can be vulnerable to hacking if not secured properly. Secondly, the shortage of skilled workforce can have significant implications for both the effectiveness of technology use and the overall quality of services. While the lack of users' technical skills is solved through the simplification

of the features, on the tourism providers' side, most of the domain knowledge people do not have all the technical knowledge to provide the smart tourism services and handle complex systems. Thirdly, most of the aforementioned applications depend on the wireless communications, which will required the seamless connectivity throughout the environment. Most of the adventure places do not have proper coverage strength, which can degrade the performance of the system. Last but not least, any Internet of Things (IoT) devices, especially those deployed in remote or hard-to-reach locations, rely on batteries. High power consumption can lead to frequent battery replacements or maintenance. In large-scale IoT deployments, managing energy usage across thousands or millions of devices can be complex, requiring sophisticated energy management solutions [19].

4. Conclusion

For analysis purposes, we can divide the the current tourism industry into two phases: before and after COVID-19. Because almost all industries faced huge changes due to COVID, tourism has had to make more specific adaptations to sustain itself. In this study, we have analyzed the machine learning trending techniques and tourism current issues and how they interconnects to produce a solution. Based on this analysis, particular gaps and challenges have been identified that will need further research. Firstly, there is an inconsistency of research to address the identified problems: to a given dataset and problem, mostly the researchers tried to apply single techniques. Even though the accuracy of the model was not satisfactory, research on the given problem and dataset was very rare. Secondly, there is a lack of publicly available large datasets in this industry. Some stakeholders had large datasets, but they were not in public. There has not been large multidimensional dataset models in tourism. The researchers mostly found issues in the before and after situations of COVID in individual tourist places.

This study has outlined the various applications of machine learning techniques in smart tourism, and serves as a general framework for future research towards a smart and sustainable tourism industry.

References

[1] Hui Ma. "Development of a smart tourism service system based on the Internet of Things and machine learning". In: *Journal of Supercomputing* 80 (5 Mar. 2024), pp. 6725–6745. ISSN: 15730484. DOI: 10.1007/s11227-023-05719-w.

[2] Ton Jaitrong and Aziz Nanthaamornphong. "Classification of Tourism-Related Topics in Phuket: A Case Study". In: *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT NCON)*. 2022, pp. 218–222. DOI: 10.1109/ECTIDAMTNCON53731.2022.9720391.

[3] Khan Md Hasib, Nurul Akter Towhid, and Md Golam Rabiul Alam. "Topic Modeling and Sentiment Analysis using Online Reviews for Bangladesh Airlines". In: Institute of Electrical and Electronics Engineers Inc., 2021, pp. 428–434. ISBN: 9781665400664. DOI: 10.1109/IEMCON53756.2021.9623155.

[4] Yu Hsiang Hsiao and Pei Yi Hung. "A Method of Evaluating User Preference Similarity for Destination Recommendations Based on Online Review Analysis". In: Institute of Electrical and Electronics Engineers Inc., 2023, pp. 452–455. ISBN: 9798350324228. DOI: 10.1109/IIAI-AAI59060.2023.00093.

[5] Junna Hou. "Service Quality Evaluation Algorithm of Tourist Scenic Spots Based on Fuzzy Cluster Analysis". In: *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*. 2022, pp. 1–6. DOI: 10.1109/ICMNWC56175.2022.10031898.

[6] Hui Cui et al. "Classification of Dunhuang Mural Image Based on Small-sample and Semi-supervised Learning". In: Institute of Electrical and Electronics Engineers Inc., 2023, pp. 653–657. ISBN: 9781665460033. DOI: 10.1109/ITNEC56291.2023.10082614.

[7] Khurniawan Eko Saputro, Sri Suning Kusumawardani, and Silmi Fauziati. "Development of semi-supervised named entity recognition to discover new tourism places". In: *2016 2nd International Conference on Science and Technology-Computer (ICST)*. 2016, pp. 124–128. DOI: 10.1109/ICSTC.2016.7877360.

[8] Jui Hung Chang et al. "Using ANN to Analyze the Correlation Between Tourism-Related Hot Words and Tourist Numbers: A Case Study in Japan". In: vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., July 2017, pp. 132–137. ISBN: 9780769563282. DOI: 10.1109/SC2.2017.27.

[9] Houria Laaroussi, Fatima Guerouate, and Mohamed Sbihi. "Deep Learning Framework for Forecasting Tourism Demand". In: Institute of Electrical and Electronics Engineers Inc., Nov. 2020. ISBN: 9781728159508. DOI: 10.1109/ICTMOD49425.2020.9380612.

[10] Kamuna Kipa et al. "A Data-Driven Approach to Predict Social Impact of Rural Tourism: Insights from UNWTO's Video Campaigns". In: *2023 IEEE International Conference on Big Data (BigData)*. 2023, pp. 6201–6203. DOI: 10.1109/BigData59044.2023.10386389.

[11] Muhammad Ilham Mubarak and Z. K. A. Baizal. "Tourism Route Recommendation Using Reinforcement Learning". In: *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*. 2023, pp. 1–7. DOI: 10.1109/I2CT57861.2023.10126347.

[12] Matan Mor, Sagi Dalyot, and Yael Ram. "Who is a tourist? Classifying international urban tourists using machine learning". In: *Tourism Management* 95 (Apr. 2023). ISSN: 02615177. DOI: 10.1016/j.tourman.2022.104689.

[13] Mehrbakhsh Nilashi et al. "A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques". In: *Computers and Industrial Engineering* 109 (July 2017), pp. 357–368. ISSN: 03608352. DOI: 10.1016/j.cie.2017.05.016.

[14] Lochana Vishwajith et al. "Machine Learning Technique based Trip Planning System - TripMa". In: *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*. 2022, pp. 1362–1369. DOI: 10.1109/ICOSEC54921.2022.9952110.

[15] Zhenzhen Yu and Shan Wang. "Smart Tourism Route Planning System Based on Machine Learning Algorithm". In: *2022 International Conference on Computers and Artificial Intelligence Technologies (CAIT)*. 2022, pp. 16–20. DOI: 10.1109/CAIT56099.2022.10072232.

[16] Shilin Xu, Yang Liu, and Chun Jin. "Forecasting daily tourism demand with multiple factors". In: *Annals of Tourism Research* 103 (Nov. 2023). ISSN: 01607383. DOI: 10.1016/j.annals.2023.103675.

[17] Gang Xie, Yatong Qian, and Shouyang Wang. "Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach". In: *Tourism Management* 82 (Feb. 2021). ISSN: 02615177. DOI: 10.1016/j.tourman.2020.104208.

[18] Rodrigo Mesa-Arango et al. "Missions and factors determining the demand for affordable mass space tourism in the United States: A machine learning approach". In: *Acta Astronautica* 204 (Mar. 2023), pp. 307–320. ISSN: 00945765. DOI: 10.1016/j.actaastro.2023.01.006.

[19] Toshima Makoondlall-Chadee et al. "Sustainable tourism, Technology and Internet 4.0: Opportunities and Challenges for Small Island Developing States". In: *2021 IoT Vertical and Topical Summit for Tourism*. 2021, pp. 1–4. DOI: 10.1109/IEEECONF49204.2021.9604866.