Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 3 PAGE: 836-843 YEAR: 2024
DOI:10.17798/bitlisfen.1501209

Predicting the Number of Visitors with Artificial Neural Networks to Support Strategic Decision-Making for Science Centers

Ali ÇETİNKAYA[1*], Havva KIRGIZ[2], Ferzan KARA[3]

[1]*Konya Science Center*
[2]*Konya Science Center*
[3]*Konya Food and Agriculture University*
(ORCID: 0000-0002-7747-6854) (ORCID: 0000-0003-0985-024X) (ORCID:0009-0001-0057-5010)

Keywords: Science center, Prediction, Neural Networks, Visitor, Bayesian, Machine Learning.

Abstract
Accurately predicting visitor attendance has become increasingly vital for science centers to optimize operations, improve visitor experiences, and stay competitive in attracting and engaging global audiences. As the demand for advanced predictive analytics grows, this study explores the use of artificial neural networks (ANNs) to forecast visitor numbers at science centers. In order to achieve this objective, data pertaining to the number of visitors to the Konya Science Centre was utilized. By analyzing a dataset of ten input factors, such as weather conditions and past visitor behavior, the study develops predictive models capable of accurately estimating future attendance patterns. The best-performing model, utilizing Scaled conjugate gradient, 0.93739 for the training set, 0.90645 for the test set, 0.92376 for the validation and 0.93069 overall. These findings underscore the transformative potential of predictive analytics in science center management. Leveraging machine learning techniques, the study provides valuable insights into visitor preferences and behavior. This knowledge can empower science centers to make data-driven decisions, optimize resource allocation, and adapt their offerings to meet the evolving needs of their target audience. Ultimately, the study highlights how predictive analytics can enhance the long-term sustainability and global competitiveness of science center operations.

1. Introduction

Science centers are informal learning environments where the public can engage with science and technology and where science festivals, workshops, special day events and science shows are held for participants from different age groups. There are more than 3,000 science centers in the world. It is estimated that these centers welcome in excess of 200 million visitors each year [1]. Scientific and Technologic Research Council of Turkiye (TUBITAK) opens large and small scale science centers in many provinces within the scope of science and society activities. Konya Science Center, which is the first TUBITAK-supported science center among science centers, is a science and technology complex located on a total area of 100,000 m2 [2]. According to architectural analyses, Konya Science Centre has the feature of being an "icon building" for the city in parallel with the developing trends in the fields of design and construction technology [1].

To estimate the number of visitors to places such as museums, zoos, etc., many different methods are used [3-9]. In the case of science centers, there is a big gap in the literature for studies conducted for predicting the number of visitors using ANN. For this reason, a literature review will be given for the studies conducted for visitor number prediction of museums.

Normal et al. applied the CRISP-DM methodology to perform the research. They developed and built four different types of regression

*Corresponding author: *ali.cetinkaya@gmail.com*
Received: 14.06.2024, Accepted: 12.08.2024

models using R and its machine learning packages to model number of visitors. Predictions of visitor number were then generated from each of the four models, and forecast accuracy was measured. The extreme gradient boost model was found to be the most effective, with an average forecast accuracy of 93% and the lowest forecast variability when benchmarked against the actual visitor number from the test data set. [10].

The number of museum visitors in Turkey between 2001 and 2019 and the probable number of museum visitors in 2037 were estimated using time series. According to research, it can be said that number of visitors of museums and ruins have an increasing trend. The number of visitors is estimated to be 84.641.851 in 2037. It is concluded that the Box-Jenkins models, which is generated for the number of visitors to museums and ruins, were statistically significant ($p<0.05$). The MAPE value (7.8694) showed that the series contains highly usable estimates [6].

By selecting 3 different museums in Thailand, the average number of visitors was estimated according to the number of visitors of 3 museums with a newly created method. Four different methods were used in this study. The RF and XGBoost models have average accuracy of 56.99% and 58.08%, respectively. Regarding the proposed method, the highest average accuracy was 85.14 and 83.13% for proposed method D [11].

The number of visitors was estimated by applying a GIS-based method for open-air museums in Denmark. The results of the correlation analysis must be read with considerable caution, as the model is based on data with very low spatial resolution, which can be mitigated by using higher-resolution demographic, economic, and social data, but at prohibitive costs. The data were analyzed using multiple linear regression (R2=0.94) [12].

Recent studies have focused on predicting visitor numbers using advanced machine learning techniques. In this context, a comprehensive prediction model was developed to identify key factors influencing the visitation patterns of both local and foreign tourists to these protected regions in Sarawak, Malaysia. The study employed a range of machine learning methods, including k-NN, Naive Bayes, and Decision Tree, to forecast whether visitor arrivals would be high, medium, or low. The dataset comprised visitor arrival information from 2015 to 2019, covering eighteen totally protected areas such as national parks, nature reserves, and wildlife centers. Performance evaluation based on accuracy, precision, and recall highlighted that the Decision Tree method, particularly using Gain Ratio, yielded the highest prediction accuracy for both local (80.65%) and foreign visitors (84.35%) [13].

Another research has focused on leveraging machine learning techniques to forecast visitor numbers to tourist attractions along the Moche Route in northern Peru. The data, spanning from January 2011 to May 2022. The research evaluated the performance of four machine learning algorithms: linear regression, KNN regression, decision tree, and random forest. The findings revealed that linear regression provided the most accurate predictions for both domestic and international tourists [14].

Utilizing monthly tourist data spanning 2012 to 2018, employed XGB and GM learning models to examine the tourist volume in Sanya over the past seven years. The forecast accuracy of these models is compared both demonstrating high prediction accuracy. Based on these fitting results, a combined model is proposed to predict the tourist volume, number of overnight transit tourists, and tourist income in Sanya, China for 2019 and 2020 [15].

The capacity to accurately forecast visitor numbers has become increasingly vital for science centers aiming to optimize operations and enhance visitor experiences.

In the digital era, the ability to predict visitor attendance plays a pivotal role in enabling science centers to efficiently allocate resources and compete globally in attracting and engaging audiences. Looking forward, the demand for advanced predictive analytics continues to surge, offering the potential to revolutionize the strategic decision-making processes of science centers worldwide.

In alignment with this vision, this study aims to explore the potential of artificial neural network (ANN) algorithms to forecast the influx of visitors in science centers. By examining factors such as weather conditions, special occasions, and historical visitor behavior, predictive models are developed to provide an accurate understanding of future attendance trends. The study utilizes extensive data analytics and cutting-edge machine learning techniques to offer science centers insights into visitor behavior and preferences.

While previous research has concentrated on forecasting visitor numbers in a variety of contexts, this study aims to differentiate itself in the field of science centers by striving to comprehensively optimize strategic decision-making processes. This study aims to conduct a forecasting study using artificial intelligence techniques with visitor data from Konya Science Centre, thus drawing a new route aiming to improve visitor experiences, promote sustainability and increase public interest in science and technology.

The following sections of this article summarize the methodology used (Section 2), present the findings and conclusions (Section 3), and conclude with reflections on the transformative potential of predictive analytics in science center management (Section 4).

The following research questions (RQs) were stated by taking into account the research gap on this topic:

RQ1: How can science centers and cultural institutions accurately forecast visitor numbers to optimize resource allocation and enhance visitor experiences in an increasingly dynamic and technology-driven world?
RQ2: What new techniques can be utilized to overcome the limitations of traditional methods in predicting visitor behaviors and supporting strategic decision-making processes in science centers?

2. Material and Method

Understanding visitor behavior in science centers requires a rigorous approach to data collection and analysis. In this section, we outline the methodologies used to collect, process, and analyze data.

2.1. Data Collection

In this study, analysts played a key role in forecasting visitor numbers at science center using a variety of methods, including both Konya Science Center's visitor data between 2020 and 2023 and other types of data such as meteorological data. Analysts worked on collecting, processing and analyzing visitor data. Daily, monthly and annual visitor numbers were collected from the Konya Science Center. In addition, meteorological data from government agencies were used to understand the impact of weather conditions and special events on visitor flow.

2.2. Data Analysis

The data analysis phase involved creating and evaluating a neural network model to predict visitor numbers at science centers.

Feature engineering involved converting meteorological data into measurable measurements suitable for predictive modeling. This conversion process increased the accuracy and reliability of predictions by converting raw data into a format that can be easily interpreted by machine learning algorithms [16]. To achieve this, the following steps were taken:

Temperature data was converted from continuous values into discrete categories to simplify analysis and capture the varying impacts of different temperature ranges on visitor numbers. The categories were defined as follows:

- Very Hot: 30°C and above = 6
- Hot: 25°C - 30°C = 5
- Warm: 20°C - 25°C = 4
- Cool: 15°C - 20°C = 3
- Cold: 10°C - 15°C = 2
- Very Cold: Below 10°C = 1

Precipitation data, which can significantly influence visitor behavior, was also categorized to capture different levels of rainfall:

- Heavy Rain: 15 mm and above = 6
- Moderate Rain: 5 mm - 15 mm = 5
- Light Rain: 1 mm - 5 mm = 4
- No Rain: 0 mm (dry) = 3

Wind speed data was categorized to reflect varying degrees of wind conditions, which could affect the comfort and safety of visitors.

- Very Strong Wind: 20 km/h and above = 6
- Strong Wind: 15 km/h - 20 km/h = 5
- Moderate Wind: 10 km/h - 15 km/h = 4
- Light Wind: 5 km/h - 10 km/h = 3
- Calm: Below 5 km/h = 2

The model consisted of 9 input nodes representing factors such as weather conditions, events, and historical visitation patterns, and a single output node for estimated visitor numbers.

Input parameters: 9 input parameters were used for each ANN model. The input parameter distributions are as follows: holidays (two), seasons (four), meteorological data (three).

Target parameters: daily visitor numbers
For categorical variables such as holidays, and seasons, a binary coding scheme was adopted, where a value of 1 indicates the presence of the condition and a value of 0 indicates its absence (Figure 1). Meteorological data such as daily air temperature, precipitation amount, and

wind speed were divided into high, medium and low levels to facilitate analysis. These adjustments were made to optimize the modeling process and increase the accuracy of predictions. With the implementation of these improvements, the models were finalized and ready for analysis.

HOLIDAY	SCB	WINTER	SPRING	SUMMER	AUTUMN	HEAT	WIND	RAIN	NU. OF VISITORS
1	0	1	0	0	0	1	2	3	271
0	0	1	0	0	0	1	3	3	562
0	0	1	0	0	0	1	3	3	800
0	0	1	0	0	0	1	2	3	614
0	0	1	0	0	0	1	2	3	477
0	0	1	0	0	0	1	3	6	719
0	0	1	0	0	0	1	3	5	715
0	0	1	0	0	0	1	3	3	626
0	0	1	0	0	0	1	3	3	1.056
0	0	1	0	0	0	1	2	3	1.008
0	0	1	0	0	0	1	2	3	523
0	0	1	0	0	0	1	2	3	488
0	0	1	0	0	0	1	2	3	1.019
0	0	1	0	0	0	1	2	3	975
0	0	1	0	0	0	1	2	4	301
0	0	1	0	0	0	1	2	3	561

Figure 1. Input and target parameters.

Data were cleaned, inconsistencies and missing values were removed, and normalized for uniformity. Three optimization algorithms (Levenberg-Marquardt, Bayesian Regularization, and Scalar Conjugate Gradient) with various hidden layers (2 to 20) were used and trained iteratively for 10 epochs. These are the hyperparameter values used for the scaled conjugate: maximum number of epochs 1000; performance goal 0; minimum performance gradient 1e-6; maximum validation failures 6; epochs between displays 25. These are the hyperparameter values used for the Bayesian regularization maximum number of epochs to train 1000; decrease factor for mu 0.1; increase factor for mu 10; minimum performance gradient 1e-7; maximum validation failures 0. These are the hyperparameter values used for the Levenberg-Marquardt Maximum number of epochs to train 1000; performance goal 0; maximum validation failures 6; minimum performance gradient 1e-7; initial mu 0.001; increase factor for mu 10; maximum value for mu 1e-10.

Training, validation, and testing subsets were produced by randomly splitting the data in the ANN.

Each model was trained using the chosen algorithm and hidden layers, and the parameters were set to minimize the difference between predicted and actual visitor numbers. Measurements such as mean square error was used in the evaluation.

3. Results and Discussion

The analysis of results from this study provides valuable insights into the performance of various optimization algorithms and neural network architectures in predicting visitor numbers to science centers. By meticulously comparing the effectiveness of different model variants, configurations that offer precise predictions were pinpointed, illuminating the critical factors shaping visitor flows.

The employed ANN models consisted of one hidden layer, 9 inputs, and one target (Figure 2).

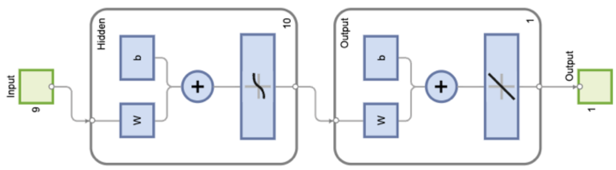

Figure 2. Neural network diagram.

Each model was executed 10 times for between 2 and 20 neurons to find the optimal number of neurons in its hidden layer. The best execution results were given in Table 1 for each algorithm.

Table 1. Analysis of the results

Algorithm	R	MSE	MAPE
Bayesian regularization	0.9281	16271.73	15.0594
Levenberg-Marquardt	0.9272	16311.58	13.5158
Scaled conjugate gradient	0.9306	15565.52	13.9231

Moreover, the best results' training, testing, and validation values are given in Fig. 2, Fig. 3, Fig. 4 for Levenberg–Marquardt, Bayesian regularization, and scaled conjugate gradient algorithms, respectively.

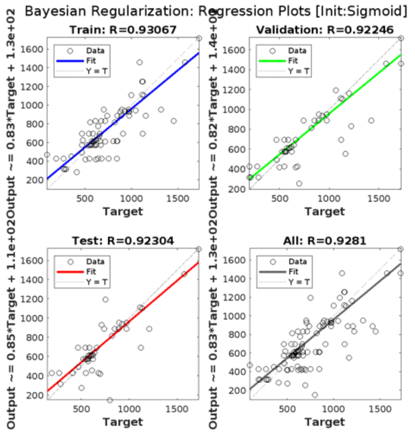

Figure 3. Best results of the Bayesian regularization algorithm (Layer 12).

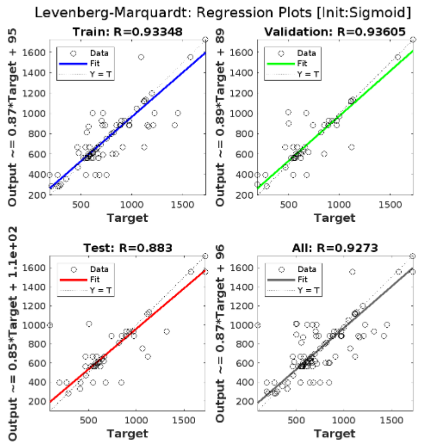

Figure 4. Best results of the Levenberg–Marquardt algorithm (Layer 11).

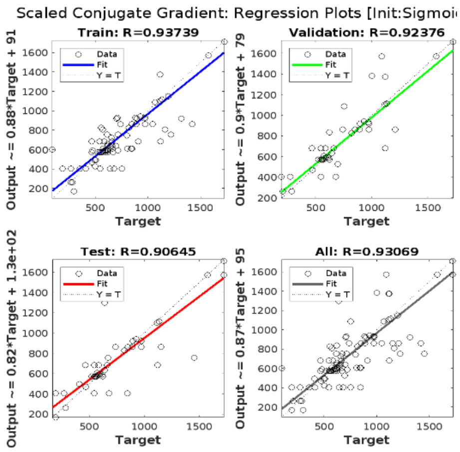

Figure 5. Best results of the scaled conjugate gradient algorithm (Layer 19).

The error histogram graphics are given in Fig. 6, Fig. 7, Fig. 8 for the three algorithms.

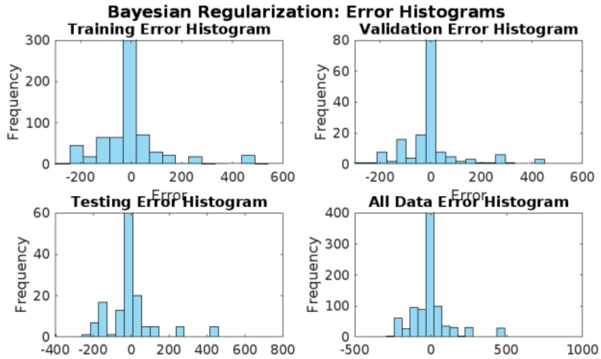

Figure 6. Bayesian regularization algorithm's error histogram values.

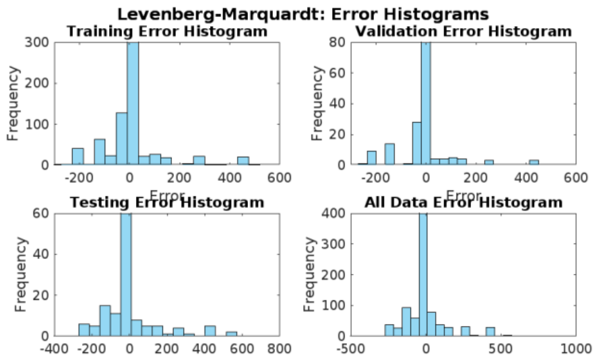

Figure 7. Levenberg–Marquardt algorithm's error histogram values.

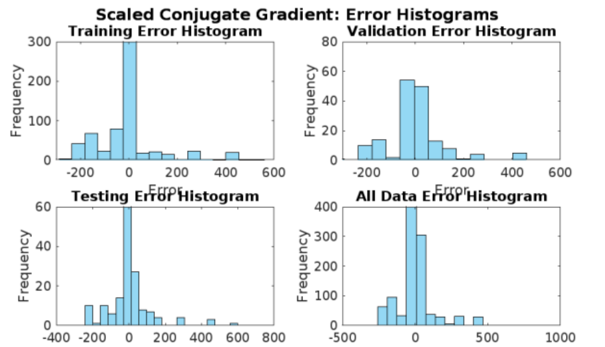

Figure 8. Scaled conjugate gradient algorithm's error histogram values

The error histogram, with blue bars representing training data, green bars for validation data, and red bars indicating testing data, offers a visual representation of outliers. These outliers are data points where the fit deviates significantly from the majority of the data.

The performance graphics are presented in Fig. 9, Fig. 10, Fig. 11 for the Levenberg–Marquardt, Bayesian regularization, and scaled conjugate gradient algorithms, respectively.

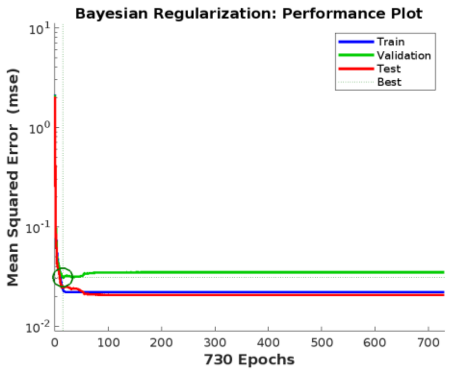

Figure 9. Bayesian regularization algorithm's performance values

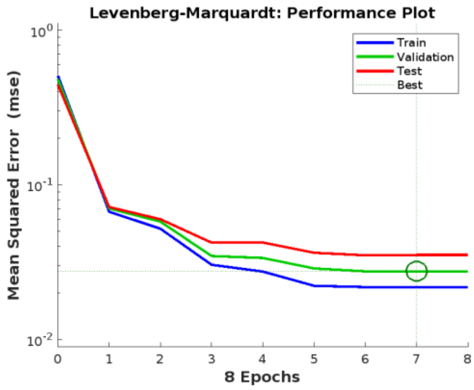

Figure 10. Levenberg–Marquardt algorithm's performance values

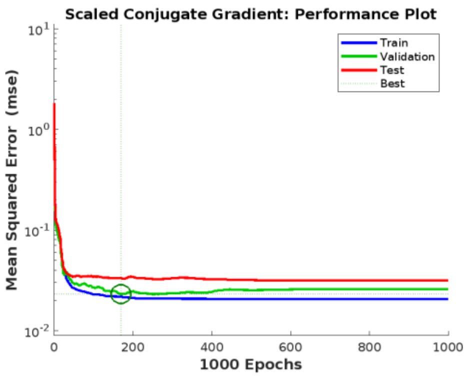

Figure 11. Scaled conjugate gradient algorithm's performance values

Among the ANN training algorithms, it has been observed that the scaled conjugate gradient algorithm gives the best effect in layer 19 of the hidden layer. The results of the scaled conjugate gradient algorithm are as follows: Training R=0.93739, Test R= 0.90645, Validation R=0.93069, All R=0.93069.

4. Conclusion and Suggestions

In conclusion, this study highlights the important role that predictive analytics plays in the management of science centers, particularly in predicting visitor numbers to optimize resource allocation and improve visitor experiences. Through the use of artificial neural network (ANN) algorithms, this research found that: The effectiveness of predictive models in accurately predicting future attendance trends by taking into account a variety of factors such as weather conditions, special occasions, and historical visitor behavior.

The integration of comprehensive data analytics and advanced machine learning techniques has provided insights into visitor behavior and preferences, allowing science centers to make informed, data-driven decisions and adapt their offerings to meet changing audience needs. These findings illuminate the transformative potential of predictive analytics in shaping the landscape of visitor engagement strategies, thereby contributing to the enduring sustainability and prosperity of science center operations.

As science centers continue to adapt to changing technological and societal dynamics, adopting advanced analytical methodologies and leveraging the capabilities of ANN algorithms will continue to be imperative. Predictive analytics remains an indispensable tool to guide strategic decision-making and ensure continued relevance in an increasingly competitive environment.

In the study conducted for 3 museums in Sinop, which is included in the literature, the best result was achieved with the Levenberg - Marquardt method with an R value of 0.99 [3]. Our study was conducted specifically for the science centers, and the best result was obtained with the scaled conjugate algorithm with an R value of 0.93.

This study is of significant value to science centers in their operational activities, such as determining the days on which a science center will be closed based on the number of visitors at any given time or identifying the most suitable days for planning science events. The lack of visitor estimates for science centers using this method in the literature makes the study more important. It would be beneficial for science center curators and for visitors or schools who wish to plan or book a visit to the museum if the calendar heat map could be displayed on the museum website. Such a feature would provide visitors with information that could influence their decision to visit the museum.

It might be worth considering whether the proposed method could be applied to other science centers as well. It would also be helpful to bear in mind that all science centers must be in a similar situation, such as similar weather conditions.

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Contributions of the authors

A.Ç., H.K. and F.K. contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. A.Ç. and H.K. methodology; F.K. and A.Ç. software; A.Ç. and H.K. data collection; A.Ç. writing—original draft preparation; A.Ç. and H.K. writing—review and editing; H.K. and F.K.; visualization.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] E. Y. Kuyrukçu, "The importance of iconic buildings for city image: Konya Science Center example," *ICONARP International Journal of Architecture and Planning,* vol. 6, no. 2, pp. 461-481, 2018.
[2] TÜBİTAK. "About the Call for Large Scale Science Centres." https://bilimmerkezleri.tubitak.gov.tr/Icerik/buyuk-olcekli-bilim-merkezi-projeleri-cagrisi-hakkinda-146 (accessed 11.06.2024.
[3] Ö. Alcan, M. Demir, and Y. Alcan, "Prediction of the Numbers of Visitors at the Sinop Museums by Artificial Neural Networks," *Karadeniz Fen Bilimleri Dergisi,* vol. 9, no. 1, pp. 70-81, 2019.
[4] M. Çuhadar, İ. Güngör, and A. Göksu, "Turizm Talebinin Yapay Sinir Ağları İle Tahmini Ve Zaman Serisi Yöntemleri İle Karşılaştırmalı Analizi: Antalya İline Yönelik Bir Uygulama," *Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi,,* vol. 14, no. 1, pp. 99-114, 2009.
[5] S. Wahyuni, A. D. Nasution, and H. Hermansyah, "Optimization Of Data Mining In Predicting Tourist Visits At The Deli Serdang Museum," in *International Conference on Sciences Development and Technology,* 2023, vol. 3, no. 1, pp. 115-120.
[6] S. Öztemiz and M. A. Tekindal, "Forecasting the number of visitors of the museums and ruins by Using time series analysis: The case of Turkey," *Türk Kütüphaneciliği,* vol. 35, no. 2, pp. 232-248, 2021.
[7] J. Beresford, "Mind the Gap: Prediction and Performance in Respect to Visitor Numbers at the New Acropolis Museum," *Museum & Society,* vol. 12, no. 3, pp. 171-190, 2014.
[8] G. Trinh and D. Lam, "Understanding the attendance at cultural venues and events with stochastic preference models," *Journal of Business Research,* vol. 69, no. 9, pp. 3538-3544, 2016.
[9] E. S. Silva, H. Hassani, S. Heravi, and X. Huang, "Forecasting tourism demand with denoised neural networks," *Annals of Tourism Research,* vol. 74, pp. 134-154, 2019.
[10] N. Yap, M. Gong, R. K. Naha, and A. Mahanti, "Machine learning-based modelling for museum visitations prediction," in *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, 2020: IEEE, pp. 1-7.
[11] W. Limpornchitwilai and P. Suksompong, "Analysis of visitor data prediction of National Science Museum Thailand (NSM)," Thammasat University, 2021.

[12] B. Ejstrud, "Visitor numbers and feasibility studies. Predicting visitor numbers to danish open-air museums using GIS and multivariate statistics," *Scandinavian Journal of Hospitality and Tourism,* vol. 6, no. 4, pp. 327-335, 2006.

[13] A. Z. Abang Abdurahman, W. F. Wan Yaacob, S. A. Md Nasir, S. Jaya, and S. Mokhtar, "Using Machine Learning to Predict Visitors to Totally Protected Areas in Sarawak, Malaysia," *Sustainability,* vol. 14, no. 5, p. 2735, 2022.

[14] J. Bravo, R. Alarcón, C. Valdivia, and O. Serquén, "Application of Machine Learning Techniques to Predict Visitors to the Tourist Attractions of the Moche Route in Peru," *Sustainability,* vol. 15, no. 11, p. 8967, 2023.

[15] X. Liu, Y.-n. Chen, Z. Qiu, and M.-r. Chen, "Forecast of the tourist volume of Sanya city by XGBoost model and GM model," in *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2019: IEEE, pp. 166-173.

[16] G. Dong and H. Liu, *Feature engineering for machine learning and data analytics*. CRC press, 2018.