Prediction Model of National Visitors to the Moche Route in Peru based on Time Series and Neural Networks

Oscar Serquén[1], Roger Alarcón[2], Jessie Bravo[3], Carlos Valdivia[4], Janet Aquino[5]

[1][2][3][4][5]*Digital Transformation Research Group, Pedro Ruiz Gallo National University, Juan XXIII 395, Lambayeque, Perú*

Abstract

The pandemic affected all economic sectors in the world, one of the most critical being the tourism sector, which is why the institutions involved need to manage urgent actions for its reactivation; innovation and digital transformation using disruptive technologies are important. The objective of the study is to use machine learning based on neural networks and time series to predict the influx of national visitors on the Moche Route of Peru, becoming a contribution to the use of artificial intelligence in favor of the social and economic development of the region. A methodology composed of 4 stages was developed: (1) data collection, (2) model analysis, (3) model development, and (4) model evaluation. Open access data was used during the period from January 2011 to December 2019, applying a recurring predictive process to determine the data in the pandemic years, using the algorithm based on time series and neural networks, finally, evaluated its operation and the proximity of the prediction to the real data. In conclusion, the model presents optimal results for all the tourist attractions of the Moche Route, demonstrating its prediction effectiveness, allowing the entities in charge of the tourism sector to have a tool for planning tourist itineraries and the necessary resources to cope to future demand.

1. Introduction

The COVID-19 pandemic had an unprecedented impact on the tourism industry worldwide, paralyzing travel and considerably reducing tourist arrivals. The reactivation of tourism today is a very important issue for countries and their authorities who must adequately manage tourism demand and volume. This reactivation is using information technology, changing traditional tourism and allowing the information-based tourism industry to develop, generating innovation [1], likewise, [2] highlights the growing use of new technologies and artificial intelligence (AI) techniques in the tourism sector that allows improvement in speed, creativity and knowledge of the service, in the search of improving tourist satisfaction.

In Peru, tourism activity by 2023 will contribute 2.5% to the National GDP [3], projecting the arrival of 2.2 million foreign visitors and 34.3 million trips by domestic tourists. The income generated by domestic tourism from 2012 to 2022 still does not achieve the income achieved before the pandemic [4].

One of the most visited tourist routes is the Moche Route that integrates the regions of La Libertad and Lambayeque, presenting archaeological, natural, cultural and landscape attractions [5]. In these regions of the northern coast of Peru, some of the most important pre-Columbian civilizations: Moche, Chimú and Sicán, being promoted by the National Strategic Tourism Plan 2025 [6].

Similarly, the Peruvian state, in July 2023, enacted Law No. 31814 which promotes the use of AI within the framework of the national digital transformation process with the purpose of promoting the economic and social development of the country [7]. An important advance for AI

is open data, where, according to the United Nations Electronic Government Report 2022, Peru is located in first place in Latin America [8].

1.1. Tourism and digital transformation

In the era of digital transformation, AI has revolutionized numerous sectors, so [9, 10] point out that forecasting future trends is of utmost importance for managers and decision makers in different sectors and specifically tourism has been no exception. The complex characteristics of tourist arrival series, such as seasonality, randomness and non-linearity, make tourist arrival forecasting still a difficult task [11].

1.2. Predictive models in tourism

Predictive models are components of AI that serve as an analytical tool that uses historical data to predict future values. According to [12] there are several models for the prediction of tourists or visitors, among them: Random Forest (RF), neural networks (RN) and support vector machines (SVM), however, there is no single technique or algorithm that delivers consistent forecasts, but the results depend on the model and technique applied, the number of observations and the characteristics of the dataset. In research by [13], he notes that RNs have been shown to be particularly accurate in univariate time series forecasting environments.

In the study [14] it is analyzed and managed to predict tourism demand in Vietnam using a multilayer perceptron artificial RN prediction model. In the case of data during the COVID-19 pandemic, dummy variables were used to model the impact of the pandemic and the number of tourists expected. Other techniques are also used, as stated [15], through the analysis of historical tourism volume and convolutional neural network models, which also allowed estimating tourism demand during the COVID-19 pandemic.

Taking into account the above, the main objective of the research is to use machine learning based on neural networks and time series to predict the influx of national visitors in the Moche Region of Peru, constituting a contribution to the use of artificial intelligence in favor of the social and economic development of the country.

2. Materials and methods

Figure 1 describes the methodology applied in this research, which consists of four main stages: (1) data collection, (2) model analysis, (3) model development, and (4) model evaluation. The Python version 3 programming language and Google Colab storage were used for processing.
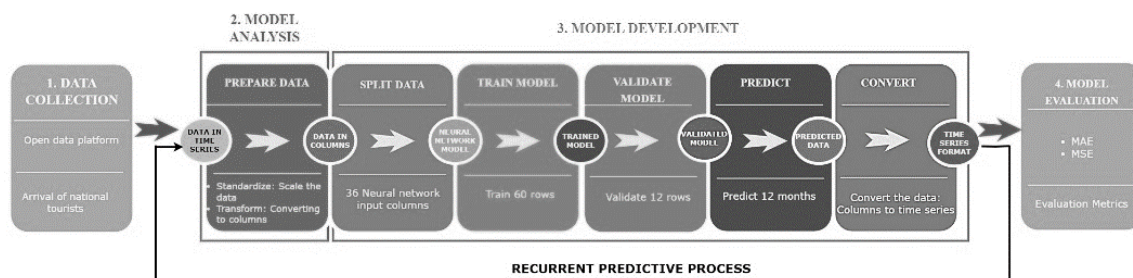

Figure 1: Applied Research Methodology

2.1. Data collection

The data used in the proposed model were divided into two dimensions, as described in Table 1, during the period between January 2011 and December 2019. The years from 2020 to 2022, the period of the COVID-19 pandemic, no data are recorded, so their treatment is explained in the development of the proposed model.

Table 1
Variables used in the prediction of the model.

Dimension	Variable	Description	Type
Temporal	DATE	Year and Month of visit	Numerical
Arrival of National Tourists	TOTAL_NATIONAL	Number of domestic tourists visiting tourist attractions on a monthly basis	Numerical

The **time** dimension represents the year and month within the evaluated period and the dimension **Arrival of National Tourists** are data extracted from the Internet [16], from the Open Data Platform of the Government of Peru, which expresses the number of national tourists who visited the tourist attractions of the Moche Route in Peru on a monthly basis. The tourist attractions of the Moche Route are located in the departments of La Libertad and Lambayeque, in the northern part of Peru, which are detailed in Table 1.

Table 2
Tourist attractions on the Moche Route.

Location	Tourist Attraction	Type
La Libertad	Huaca Arco Iris Archaeological Complex	Archaeological site
La Libertad	Huaca del Sol y la Luna Archaeological Complex	Archaeological site
La Libertad	Huaca el Brujo Archaeological Complex	Archaeological site
La Libertad	Chan Chan Site Museum	Museum
Lambayeque	Brüning National Archaeological	Museum
Lambayeque	Huaca Chotuna Site Museum - Chornancap	Museum
Lambayeque	Huaca Rajada Site Museum – Sipan	Museum
Lambayeque	Tucume Site Museum	Museum
Lambayeque	Sican National Museum	Museum
Lambayeque	Royal Tombs of Sipan Museum	Museum
Lambayeque	Pomac Forest Historical Sanctuary	Museum

2.2. Model Analysis

Preprocessing was carried out, for which the data was standardized by applying a scale with a range between -1 and 1, which allows the data to be scaled, without losing the originals, then the transformation of the same from rows to columns of 36 elements and 72 rows, which will serve as input data to the model.

The choice of 36 input elements provided better results when the model was applied, with respect to larger or smaller quantities. The pre-processed data maintained the time series structure, for which they were transformed into data required for the application of the selected algorithm.

The prediction algorithm used is time-series neural networks, which is a method of artificial intelligence that teaches computers to process data in a similar way to the human brain.

The type of neural network applied is feedforward (unidirectional or forward-propagating networks), made up of three layers, a dense layer, a flatten layer and a dense layer. In addition, the hyperbolic tangent activation function, applied in the dense layers, was used [17].

2.3. Model Development

The development of the model begins in the data division stage, selecting from the 72 existing rows, 60 rows for the training stage and 12 rows for model validation. The algorithm makes block predictions of 12-month national visitors, for which it takes input information from the previous

36 months. This process is repeated recurrently, using as data the information collected from 2011 to 2019 to predict the year 2020, subsequently the predicted data from 2020 was added as input data to the model and the year 2021 was predicted, the results obtained again served as input to the model, having aggregate data from the year 2020 and 2021, for the year 2022 the same process was carried out and the prediction for the year 2023 was completed.

2.4. Evaluation of the model

The evaluation of the model was based on the metrics: mean absolute error (MAE) and the mean square error (MSE). MAE is defined as a loss function calculated from the sum of the absolute differences between the expected value and the predicted variables and MSE measures the average of the squared errors, that is, the difference between the estimated value and the predicted value, the which were useful to examine the accuracy of the prediction model. In both metrics, the lower the calculated value, the better the prediction model obtained will be.

3. Results and Discussion

The neural network model, as seen in Figure 2, has an input layer of 36 neurons formed by the number of monthly visitors in blocks of 36 months, whose data was scaled in the data preparation process. As part of the neural network, a hidden layer is included, which allows prediction functions to be executed to generate an output layer, which corresponds to the prediction of month 37, still in its scaled form. This process is repeated taking the generated output as a new input to the neural network.
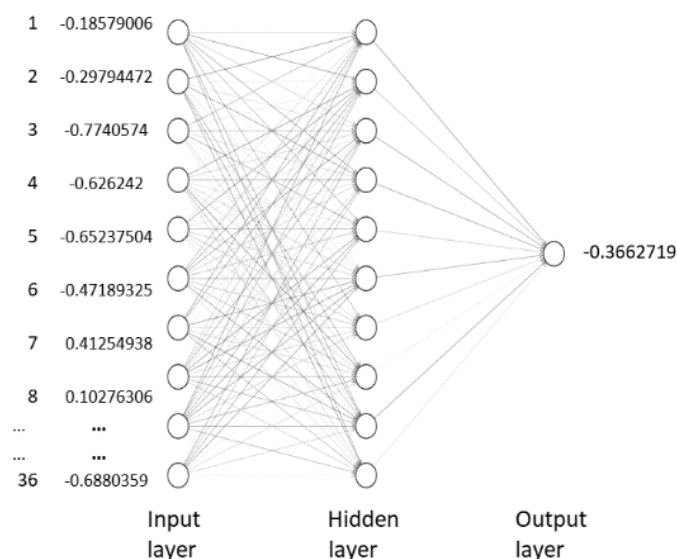

Figure 2: Developed neural network model

For the evaluation, a graphical and numerical comparison was made between the real data and the data predicted by the model, which correspond to the entire Moche Route separated by regions.

Figure 3 shows the comparison between the years 2019 and 2022, of the 4 tourist attractions that correspond to the La Libertad region. It can be seen that the predicted data follow the trend and adjust to the actual data, it is also worth mentioning that as the year prediction progresses, a substantial improvement is observed between the prediction and the actual data.
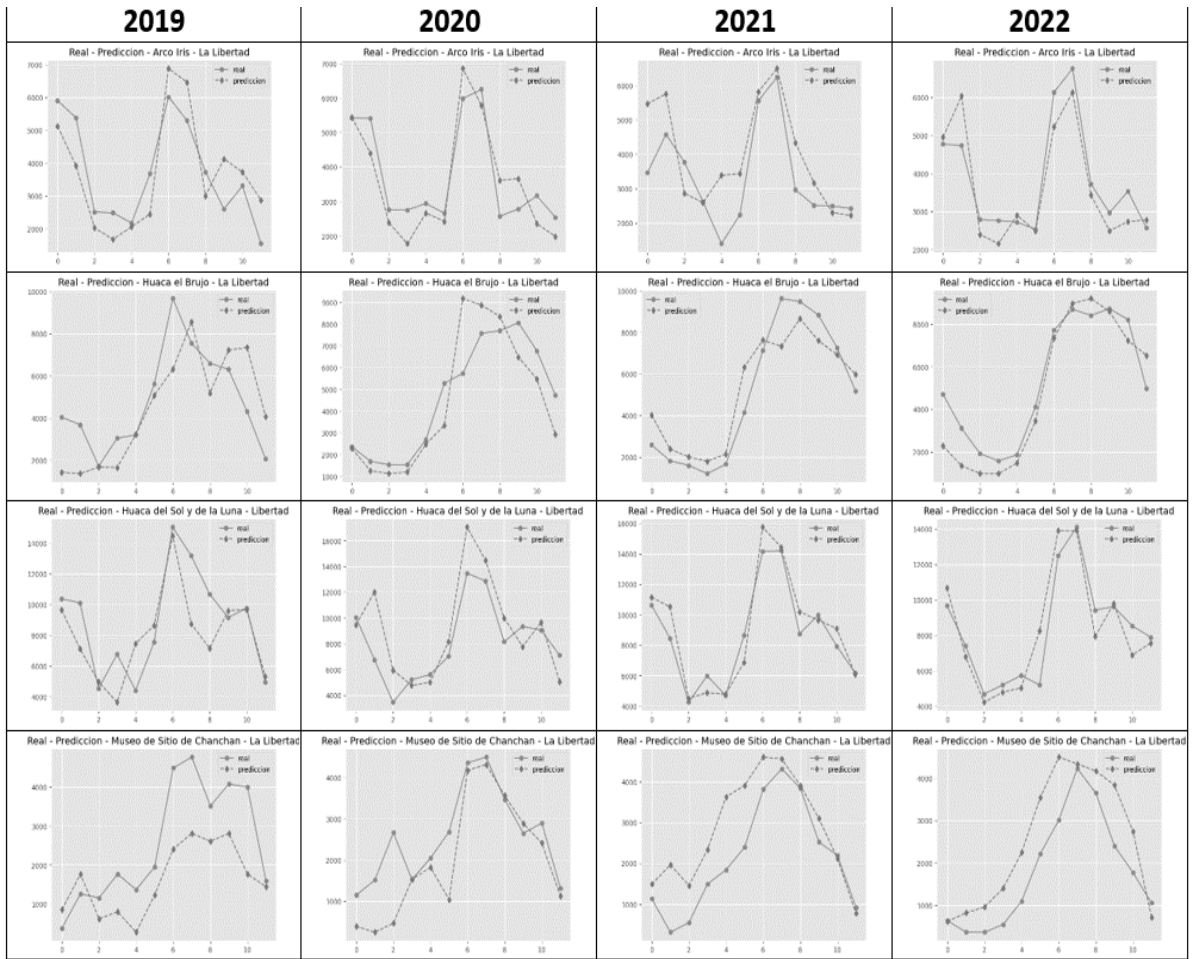
Figure 3: Comparison of Actual Values with Predicted Values - La Libertad.

Figures 4 and 5 show the comparison of the actual and predicted data for the 7 tourist attractions that correspond to the Lambayeque region, in which 4 attractions present values that fit very well with the data using a neural network model, while in 3 tourist attractions the predicted data vary significantly from the actual data, so it can be indicated that the model does not fit these values as well.
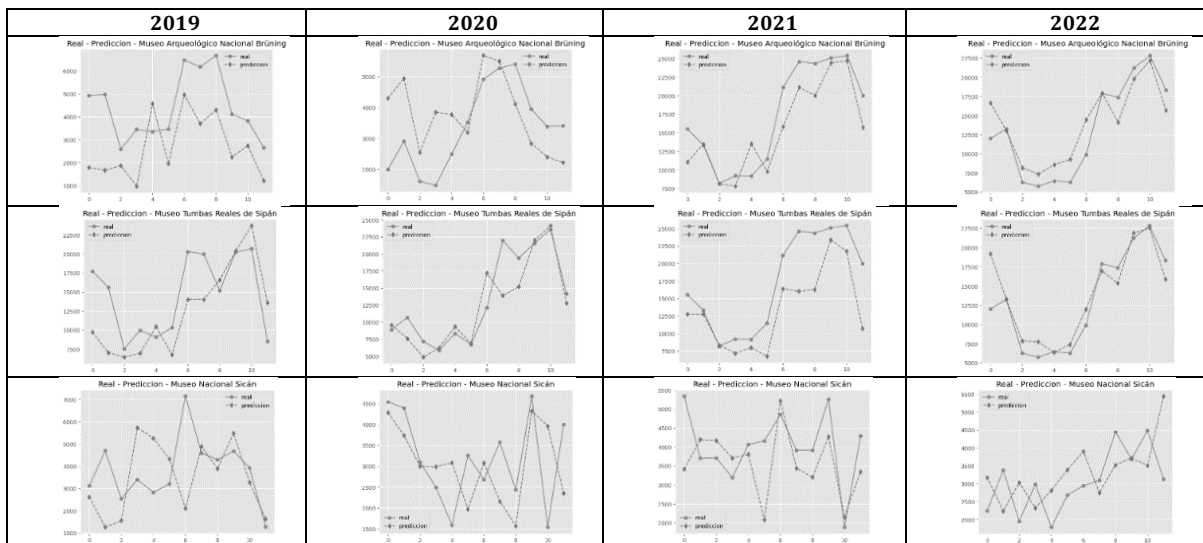

Figure 4: Comparison of Actual Values with Predicted Values - Lambayeque.
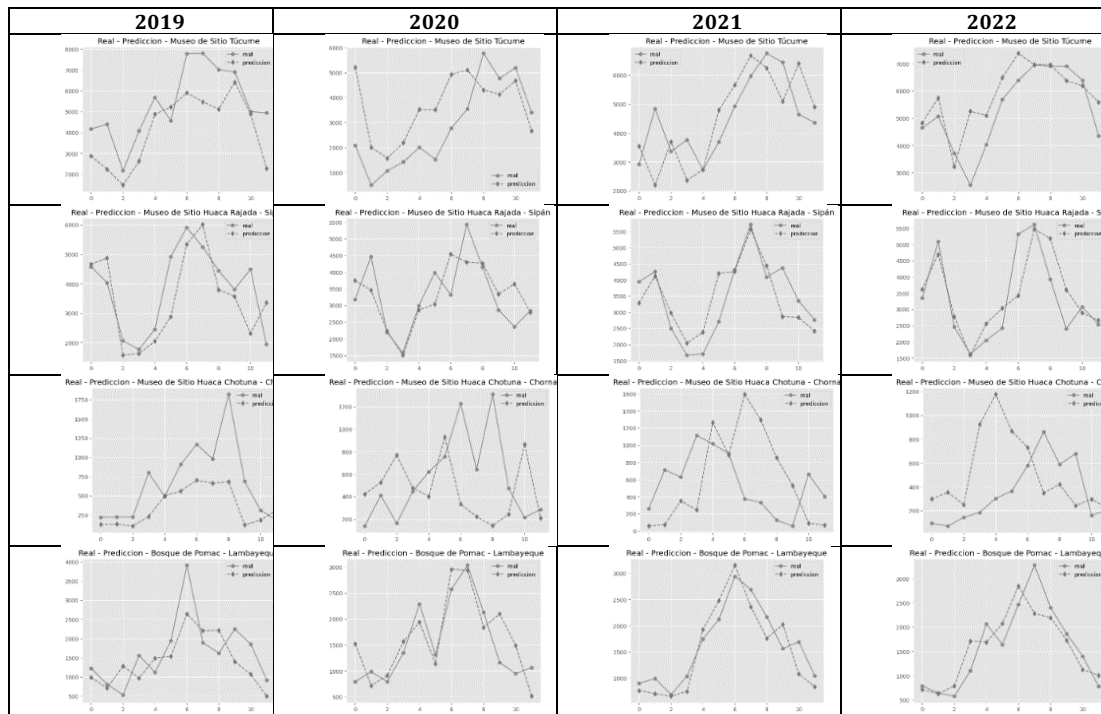
Figure 5: Comparison of Actual Values with Predicted Values - Lambayeque.

The MAE and MSE metrics were the performance measures used to evaluate the results obtained from the 11 different locations corresponding to the Moche Route.
Small values in the residual statistics, MAE and MSE, reflect better goodness-of-fit, i.e., values close to zero are considered a good prediction.

The accuracy of the predictions generated by the model was evaluated using a training set and a test set. A 5-year {Xt} training set (estimation period) and 4-year {Xt} test data (validation period) were considered and used for month-to-month prediction.

Table 3 presents the results of the evaluation of the model based on the metrics for each of the different tourist attractions, observing the results it is found that in the year 2020 there are 3 tourist attractions out of the 11 that have very high values (highlighted in bold) in the MSE metric (Huaca el Brujo, Tucume Site Museum and Huaca Chotuna Site Museum - Chornancap) which could indicate a lower than expected forecast. For the year 2021, it is reduced to 2 tourist attractions with high values in this metric, also indicating a forecast below expectations. But in 2022 and 2023, all attractions have small values, indicating a better fit of the prediction in the model.

Finally, in the analysis of the year 2023, it can be seen that the Huaca del Sol y de la Luna, Brüning National Archaeological Museum and the Huaca Arco Iris present the lowest results, which indicates that the data of these attractions fit very well with the predictions made by the model. The rest of the tourist attractions also present a good prediction. Similar results can be seen for the MAE metric.

Table 3
Evaluation results using MSE and MAE as metrics.

	MSE				MAE			
Tourist Attraction	2020	2021	2022	2023	2020	2021	2022	2023
Huaca Arco Iris Archaeological Complex	0.0573	0.0493	0.0411	0.0400	0.1775	0.1650	0.1550	0.1465
Huaca el Brujo Archaeological Complex	0.1039	0.0697	0.0527	0.0586	0.2160	0.1987	0.1547	0.1788

Huaca del Sol y la Luna Archaeological Complex	0.0620	0.0629	0.0479	0.0287	0.1863	0.1775	0.1534	0.1182
Chan Chan Site Museum	0.0928	0.1093	0.0742	0.0702	0.2149	0.2347	0.1882	0.1939
Brüning National Archaeological	0.0982	0.0809	0.0515	0.0395	0.2251	0.1899	0.1605	0.1454
Royal Tombs of Sipan Museum	0.0503	0.0504	0.0492	0.0410	0.1568	0.1529	0.1574	0.1453
Sican National Museum	0.0889	0.0888	0.0951	0.0908	0.1804	0.1890	0.1820	0.2082
Tucume Site Museum	0.1017	0.0748	0.0745	0.0582	0.2355	0.1973	0.1853	0.1749
Huaca Rajada Site Museum – Sipan	0.0465	0.0487	0.0537	0.0488	0.1596	0.1578	0.1529	0.1506
Huaca Chotuna Site Museum - Chornancap	0.1345	0.2172	0.0955	0.0921	0.2280	0.3243	0.2115	0.1922
Pomac Forest Historical Sanctuary	0.0680	0.0756	0.0673	0.0576	0.2019	0.1976	0.1869	0.1660

According to the results of the forecasts of the 11 tourist attractions, the two tourist attractions with the best prediction results are described below.

In Figure 6, you can see the forecast of national visitors to the tourist attraction Brüning National Archaeological Museum, identifying that the trend of visitors is maintained over the years, the behavior of visitors between the months of September to November increases, while from December to June it has a tendency to decrease.
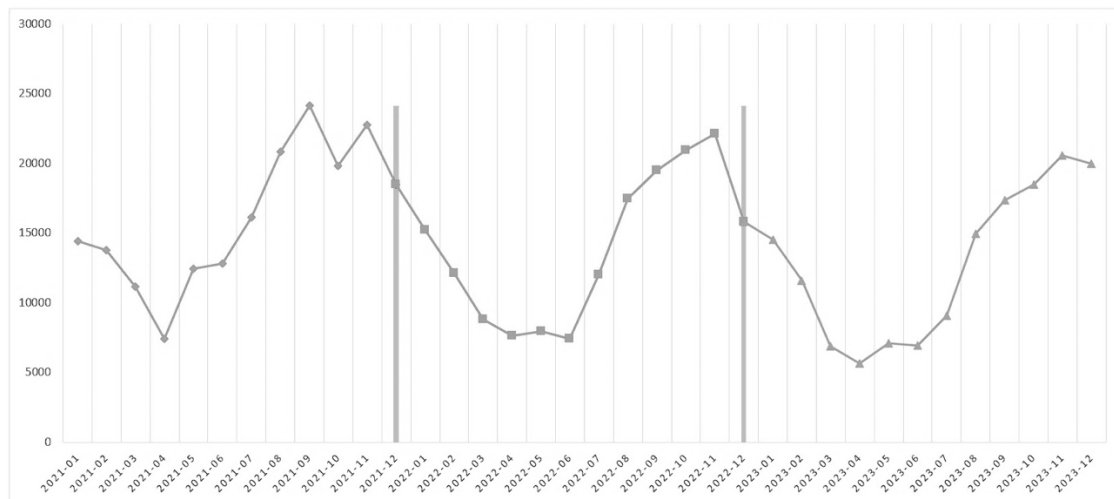

Figure 6: Forecast of national visitors at the Brüning National Archaeological Museum.

Figure 7 shows the forecast of national visitors to the Huaca del Sol y de la Luna archeological site, showing that the trend of visitors is maintained between 2021 and 2023. In addition, it can be seen that the months of July and August maintain high values for all years, while the months of April and May have the lowest values.
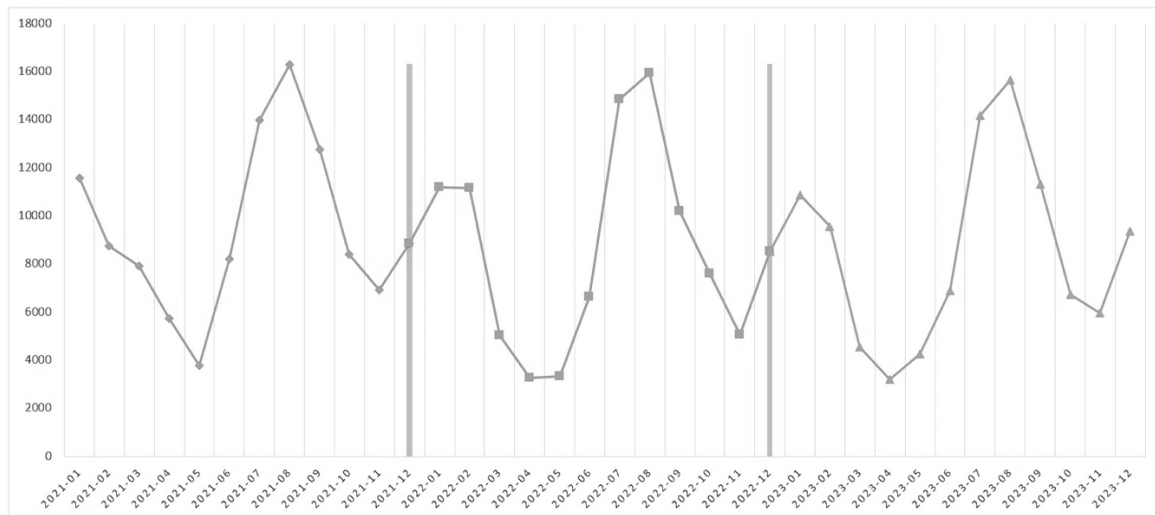
Figure 7: Forecast of national visitors in the Huaca del Sol y de la Luna Complex.

The results of the predictions made thanks to the application of computer science in tourism, through artificial intelligence, allow to understand the large sets of tourism data, analyze them through machine learning techniques to identify patterns and future trends in the prediction and management of tourism demand, allowing the optimal allocation of resources, improving the experience of tourists and contributing to more informed decision making at the business and government level.

4. Conclusions

This study presents a predictive model, based on neural networks and time series to determine the prediction of visitors to the Moche Route in the years 2020 to 2023, applying a recurrent predictive process.

In the recurring predictive process, the MSE and MAE metrics improved their results in each year evaluated, identifying in 2020 only three tourist attractions: Huaca el Brujo, Túcume Site Museum and Huaca Chotuna - Chornancap Site Museum with prediction below expectations; On the other hand, by 2021 it was reduced to two tourist attractions and finally, For the years 2022 and 2023, all the attractions present small values in the metrics, indicating a better fit of the prediction in the applied model.

The prediction model reached optimal values with the MAE and MSE evaluation metrics, which achieves significant results of the expected demand of national tourists for the tourist attractions of the Moche Route.

This research is framed as a contribution to Law No. 31814 on the use of Artificial Intelligence in the Peruvian state, since it will allow the entities in charge of the tourism sector to have a tool for the planning of tourist itineraries and the necessary resources to face the future demand.

References

[1] Kong, Y.: Real-time processing system and Internet of Things application in the cultural tourism industry development. Soft Computing. 27, 10347–10357 (2023). https://doi.org/10.1007/s00500-023-08304-8

[2] Dangwal, A., Kukreti, M., Angurala, M., Sarangal, R., Mehta, M., Chauhan, P.: A Review on the Role of Artificial Intelligence in Tourism. Presentado en Proceedings of the 17th INDIACom; 2023 10th International Conference on Computing for Sustainable Global Development, INDIACom 2023 (2023)

[3] El Peruano: Titular del Mincetur: Actividad turística contribuirá este año con 2.5% al PBI nacional [Entrevista], https://elperuano.pe/noticia/214558-titular-del-mincetur-actividad-turistica-contribuira-este-ano-con-25-al-pbi-nacional-entrevista

[4] SIT-MINCETUR: Sistema de Inteligencia Turística, https://www.mincetur.gob.pe/centro_de_Informacion/mapa_interactivo/index.html

[5] Bravo, J., Alarcón, R., Valdivia, C., Serquén, O.: Application of Machine Learning Techniques to Predict Visitors to the Tourist Attractions of the Moche Route in Peru. Sustainability (Switzerland). 15, (2023). https://doi.org/10.3390/su15118967

[6] PENTUR: Plan Estratégico Nacional de Turismo del Perú-PENTUR, https://www.gob.pe/institucion/mincetur/informes-publicaciones/22123-plan-estrategico-nacional-de-turismo-del-peru-pentur.

[7] El Peruano: Ley que promueve el uso de la Inteligencia Artificial en favor del desarrollo económico y social del país, http://busquedas.elperuano.pe/dispositivo/NL/2192926-1.

[8] Vilches, C.: Biblioguias: Desde el gobierno digital hacia un gobierno inteligente: UN E-Government Survey, https://biblioguias.cepal.org/gobierno-digital/un-egovernment-survey.

[9] Salehi, S.: Employing a Time Series Forecasting Model for Tourism Demand Using ANFIS. Journal of Information and Organizational Sciences. 46, 157–172 (2022). https://doi.org/10.31341/jios.46.1.9.

[10] Kirtil, İ.G., Aşkun, V.: Artificial Intelligence in Tourism: A Review and Bibliometrics Research. Advances in Hospitality and Tourism Research (AHTR). 9, 205–233 (2021). https://doi.org/10.30519/ahtr.801690.

[11] Liang, X., Wu, Z.: Forecasting tourist arrivals using dual decomposition strategy and an improved fuzzy time series method. Neural Comput & Applic. 35, 7161–7183 (2023). https://doi.org/10.1007/s00521-021-06671-7.

[12] De Jesus, N.M., Samonte, B.R.: AI in Tourism: Leveraging Machine Learning in Predicting Tourist Arrivals in Philippines using Artificial Neural Network. International Journal of Advanced Computer Science and Applications. 14, 816–823 (2023). https://doi.org/10.14569/IJACSA.2023.0140393.

[13] Semenoglou, A.-A., Spiliotis, E., Assimakopoulos, V.: Data augmentation for univariate time series forecasting with neural networks. Pattern Recognition. 134, (2023). https://doi.org/10.1016/j.patcog.2022.109132.

[14] Nguyen, L.Q., Fernandes, P.O., Teixeira, J.P.: Analyzing and Forecasting Tourism Demand in Vietnam with Artificial Neural Networks. Forecasting. 4, 36–50 (2022). https://doi.org/10.3390/forecast4010003.

[15] Wu, B., Wang, L., Zeng, Y.-R.: Interpretable tourism demand forecasting with temporal fusion transformers amid COVID-19. Applied Intelligence. 53, 14493–14514 (2023). https://doi.org/10.1007/s10489-022-04254-0.

[16] datosTurismo, http://datosturismo.mincetur.gob.pe/appdatosTurismo/Content1.html.

[17] Rivas-Asanza, W., Mazon-Olivo, B., Mejia, F.: Capítulo 1: Generalidades de las redes neuronales artificiales. Presentado en junio 29 (2018).