The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2025

Volume 34, Pages 221-233

ICBASET 2025: International Conference on Basic Sciences, Engineering and Technology

Designing Intelligent Models with ARIMAANN for Visionary Forecasts

Morena Breshanaj
University of Vlora

Areti Stringa
University of Tirana

Abstract: It is essential to choose the right model that can explain the growth of tourism in Albania and, therefore, to make the right decisions and direct the flow of tourists. This research aims to compare and apply three forecasting models: ARIMA, Artificial Neural Networks (ANN), and the hibrid ARIMAANN model to forecast the number of international tourists in Albania. The outcomes indicate that the interaction between the two approaches, ARIMAANN, is the first model to explain 96% of the data variation and provides the minimum mean absolute percentage error (MAPE) of 21.6%. In order to enhance the model's precision, the refined model, ARIMAANN 21-24, was suggested, which excluded the pre-pandemic and pandemic periods. This adjustment resulted in significant enhancements where the accuracy was 0.99, and MAPE was 7.09 %, making it the most accurate forecast. The proposed model shows that tourism will keep increasing in the next five years. The most tourists are expected in August 2029, with 2.9 million international tourists. This research provides a predictive tool for policymakers, tourism operators, and government agencies to capitalize on the benefits of hybrid modelling to enhance sustainable tourism's strategic development and management. All data and analyses were processed in RStudio with the latest advancements in time series modelling.

Keywords: Tourism demand forecasting, ARIMA, ANN, ARIMAANN

Introduction

Tourism is one of the strongest sectors of the economy of almost any country. The number of international visitors has increased dramatically, from 25.2 million in 1950 to 439 million in 1990, to 1.4 billion in 2019 and 1.4 billion in 2024. (UN Tourism World Tourism Barometer | Global Tourism Statistics, 2025) As per the Global Economic Impact & Trends, tourism is one of the critical components of 10.4% of the global GDP and 10.6% of total employment. (WTTC: World Travel & Tourism Council). In the last World Tourism Report by the United Nations World Tourism Organisation, the number of international arrivals was 1.4 billion in 2024, a 99% recovery from 2019 and 11% growth from 2023 or 140 million more international tourists. (UN Tourism World Tourism Barometer:Global Tourism Statistics, 2025)

The Secretary General of the United Nations, Zurab Pololikashvili, noted, "In 2024, international tourism has rebounded fully from the pandemic, and in many countries, the number of arrivals and, particularly, revenue is even higher than in 2019. The growth is expected to maintain its direction through 2025, with high demand as a major force for socio-economic impact for mature and emerging markets. This is evidence of our role as a sector in the globalization of the sector and making the sector more sustainable in terms of people and the planet. This is why it is our responsibility as a sector to build on this and ensure that people and the planet are at the centre of development in tourism." (UN Tourism | Bringing the world closer). In 2024, Europe was the largest region in the world, with 747 million international arrivals, 1% more than before the pandemic and 5% more than in 2023. Europe's sub-regions have seen a rise in arrivals compared to the pre-pandemic figures, except Central and Eastern Europe.

As one of the European countries with the highest rate of growth in the tourism industry, Albania has also experienced an increase in the number of tourist visits and 11,700,340 million non-resident visitors in 2024, which is 45% more than in 2019. However, these figures need to be supported by sound models to ensure the sustainable management of the flows and the formulation of policies. This research proposes a combined model to predict the number of tourists visiting Albania. This study has applied an enhanced version of the hybrid model, ARIMAANN 21-24, to enhance the model's accuracy and exclude the impact of the COVID-19 pandemic. This approach has also helped attain a better fit for the recent trends in tourism and has lower forecast errors than the previous models. The improved model has provided more accurate forecasts for the next five years that correctly capture the tourism trends in Albania. The purpose of this research is to provide empirical evidence that can be used by the tourism industry, government agencies and investors to shape policy and decision-making towards the sustainable management and development of tourism in Albania and thus contribute to the national economy and defend the natural and cultural heritage of the area. Lastly, the implications of this study for tourism management and economic development are discussed, and the possible application of the model developed in this study to policy making and resource allocation are discussed. Hence, the accurate prediction of the trends in tourism is crucial because the stakeholders can prepare and manage their resources to address the challenges, develop the right marketing strategies, and establish policies that can help the growth of sustainable tourism in the region. The research carried out in this paper provides important findings that can be used to help Albania properly utilize its tourism resources, pursue good economic development and conservation, and ensure the sustainability of the population. The number of non-resident visitors to Albania for 2016-2024 is presented in the table below.

Table 1. Number of non-resident tourists in Albania for the period 2016-2024

	2016	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
January	173534	172757	232338	209520	249461	185392	194237	377211	489599
February	169556	185721	190531	201597	228941	158173	200997	309325	480539
March	218858	243462	254020	283910	95321	170279	275976	384028	564975
April	285902	359452	351610	410060	12188	241015	402051	614256	794129
May	358484	360082	394869	401531	16389	328742	604430	705079	973688
June	394361	437820	496137	630334	179594	604381	876056	1007033	1244967
July	864718	864023	1028182	1176810	387716	1158962	1403333	1770724	1818378
August	1124615	1142198	1449952	1556826	575559	1537191	1710041	2022754	2167665
September	406582	459168	664184	591414	377033	494332	761544	1104190	1146090
October	302834	379327	337603	387152	206439	298712	365237	696825	761636
November	210377	256216	246678	268176	148362	244061	358350	547221	604332
December	225690	257474	280699	288708	180815	267409	391565	616994	654342

We have presented a visual representation of these values in the interactive graph https://rpubs.com/Breshanaj/1265708, which shows the increasing trend and seasonality of the data. The graphs were built in the R programming language with the help of the " *plotly* " package.

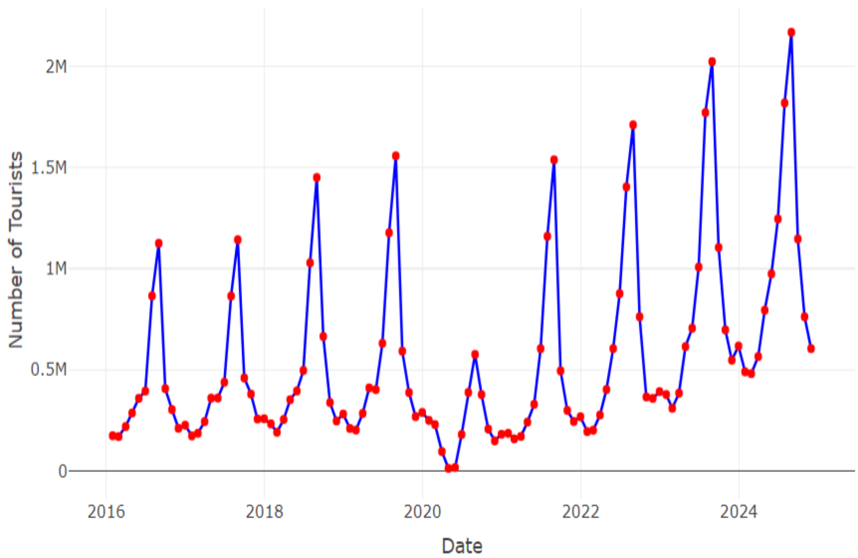

Figure 1. Number of non-resident tourists in Albania (2016-2024)

The interactive graph shows that the period with the least historical data is April 2020, which had 12,188 tourists. This is a sharp decrease from an increasing trend due to the global restrictions due to the COVID-19 pandemic. The highest monthly figure in these data is the 2,167,665 tourists recorded in August 2024. Because of the rapid growth in the number of tourists and the effects of seasonal variations, this study aims to analyze and predict future trends in tourism in Albania using advanced time series modelling techniques.

The second section outlines the methods applied in this study, and the third section gives some examples of applying the methods under study in different real-life situations. The fourth section explains how the ARIMA, ANN, and combined ARIMAANN models can forecast and model the number of non-resident tourists in Albania. The fifth section discusses the significant implications of the study, and the last section compares the findings of this study with other similar studies and presents the major conclusions.

Method

The approach used for this research in modelling and forecasting international tourist arrivals is based on a series of procedures to ensure that the modelling is done as effectively as possible. Figure 2 shows the strategy of the methodological approach, which describes the main phases from data preparation to model fitting and forecasting. All steps are implemented systematically to gain knowledge and improve the accuracy of the forecast.

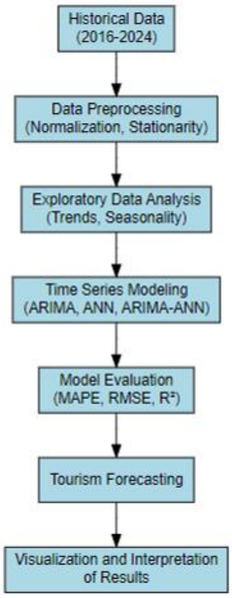

Figure 2. Structure of the methodology

This study's research design consists of formulating and comparing three statistical models for forecasting international tourist arrivals in Albania: ARIMA, ANN, and their combination ARIMAANN. The data set for this study includes monthly data from 2016-2024, which was collected from the Institute of Statistics of Albania (INSTAT) and divided into two sets: the training set, which is 80%, and the testing set, which is 20%.

The ARIMA(1,1,2)(0,1,2) [12] model is used to identify and model linear trends in time series data. In contrast, the ANN model uses a neural network with five inputs, two hidden layers, and one product to model nonlinear trends. The combined ARIMAANN method integrates the best features of both models to enhance the accuracy of the forecasts. This paper evaluates the models discussed based on accuracy metrics such as RMSE, MAE, MAPE and the R^2 coefficient of fit.

- Root Mean Squared $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$
- Mean Absolute Error $MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$

- Mean Absolute Percentage Error$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_t-\hat{y_t}}{y_t}\right| x100$
- R-squared$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t-\hat{y_t})^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2}$

On this note, the hybrid ARIMAANN21-24 method is the best as it is 96% fit to historical data and is more accurate in forecasting arrivals in the next five years. The modelling and analysis were done in RStudio, which provided a better visualization of the tourism dynamics and a reasonable basis for strategic planning in the Albanian tourism industry.

Literature Review

The ARIMAANN method combines the ARIMA (AutoRegressive Integrated Moving Average) model and ANN, an Artificial Neural Network. ARIMA models the linear components of a time series, while ANN models the nonlinear structure that ARIMA cannot capture. This method is adapted to increase the model's accuracy, especially for models with complex characteristics.

In 2003, Guangping Zhang, in his study “Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," introduced this combination. In this study, a hybrid methodology was introduced that combines both ARIMA and ANN models to take advantage of the unique power of ARIMA and ANN models in linear and nonlinear modelling (Mentes. & Yetkin, 2020). Experimental results with real data sets show that the combined model can effectively improve the forecast accuracy achieved by each of the models used separately (Zhang, 2003).

While Zhang (2003) first introduced the ARIMA-ANN model, subsequent studies such as those of Atesongun and Gulsen (2024) and Rahayu et al. (2023) etc. have shown significant improvements in this method in different time models. Adil Atesongun and Mehmet Gulsen studied a particular and interesting application of the ARIMAANN method in their article "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models" they have described in detail the functioning of the hybrid ARIMAANN method. They have applied this method to several issues, such as Turkish wheat yield, which was, in fact, the primary motivation for their article: sunspot observations, Canadian lynx counts, airline passenger data, and hourly electricity rates. They used the accuracy metrics MAD, MSE and MAPE to compare the hybrid model. These metrics have also been used in studies with other methods but with the same data, where the hybrid ARIMA-ANN method has resulted in the most successful. According to the metrics calculated in their study, the performance of the hybrid model is much higher than that of independent ARIMA or ANN. (Atesongun & Gulsen, 2024). These results indicate that the ARIMAANN model can provide higher accuracy for complex and variable data and may be suitable for forecasting tourism demand.

Rahayu et al. (2023) have presented in their study " Hybrid ARIMA-ANN Model for Solving Nonlinearity In Time Series Data" the ARIMA, ANN and hybrid ARIMA-ANN methods for estimating and forecasting the price of "shallots". They used daily data for the price of shallots for October 1, 2022 - September 30, 2023, which they divided into 90% training data and 10% testing data. From the metrics found, it was found that the hybrid ARIMA-ANN method is more accurate, with RMSE 2601 and MAPE 15.71%, respectively than the ARIMA model, which has RMSE 2888 and MAPE 17.51%. (Rahayu et al., 2023)

The research " Pemodelan ARIMAANN pada harga saham bank Mandiri" by authors Fadhillah et al. (2024) stock price data using the hybrid ARIMAANN method. This study used 493 daily data from January 2021 to December 2022. The authors built the ARIMA (0,1,1) model and the ANN model with four neurons in the hidden layer. The training and testing MAPE values for ARIMA-ANN resulted in 1.32% and 5.49%, values less than 10%, indicating that the ARIMA-ANN method is considered very good (Fadhillah et al., 2024).

The study “Application of Hybrid ARIMA and Artificial Neural Network Modeling for Electromagnetic Propagation: An Alternative to the Least Squares Method and ITU Recommendation P.1546-5 for Amazon Urbanized Cities" by Fraiha - Lopes et al. (2020) presents an empirical hybrid model integrating moving average (ARIMA) and artificial neural network (ANN), designed to estimate the propagation of electromagnetic waves in densely forested urban areas. The signal power intensity data was obtained through measurement campaigns in the Metropolitan Area of Belem (MAB) in the Brazilian Amazon (Fraiha- Lopes et al., 2020).

Authors Naheeda Perveen, Khadija Tariq and Hafiz Shabir Ahmad combined the ARIMA model and artificial neural network by adopting an equal-weight approach and a profit-weighted approach to capture the exchange rate's linear and nonlinear components. They developed a hybrid technique using the ARIMA-ANN model in their paper " Comparative Studies of Hybrid ARIMA and Artificial Neural Network (ANN) Techniques for Predicting Exchange Rate in Pakistan". The effectiveness of the models was analyzed for foreign exchange rate, imports and exports data and concluded that the hybrid techniques gave the best forecasting results (Perveen, et al., 2024).

A similar study to ours is the article "Hybrid Model of Seasonal ARIMA-ANN to Forecast Tourist Arrivals through Minangkabau International Airport", which uses the hybrid ARIMAANN method to model and predict number of tourists arriving at Minangkabau International Airport, Indonesia. First, they use a SARIMA (1, 0, 1) (1, 1, 0) 12 model to model the linear and seasonal components of the data. Then, an Artificial Neural Network (ANN) with a 2–2–2–1 architecture is applied to the SARIMA residuals to address the nonlinear patterns. The results show that the combination of SARIMA and ANN significantly improves the accuracy of the predictions, with a MAPE of 17.0963%, lower than that of SARIMA (17.1770%).

In the article "Tourism Demand Modelling and Forecasting: A Review of Literature" by Abdou et al. (2021), the authors conduct an in-depth review of different methods for forecasting tourism demand, analyzing 145 studies conducted from 1979 to 2020. They categorize the main methods into three groups: econometric models, time series models and artificial intelligence (AI) models. The study's findings show that while traditional methods such as ARIMA and econometric models help capture trends and seasonality, they cannot model the nonlinear complexity of tourism demand. On the other hand, artificial intelligence models, such as artificial neural networks (ANN) and Support Vector Machines (SVM), have great potential to improve forecast accuracy but can be more challenging to interpret. The authors emphasize that combining traditional methods with artificial intelligence through hybrid models such as ARIMAANN brings more accurate and reliable results in tourism demand forecasting. This review is important for our study, as it highlights the need for advanced approaches in tourism demand modelling, justifying the choice of the ARIMAANN model for forecasting the number of tourists in Albania.

The ARIMANN method has been widely applied in various real-life fields such as finance, energy and demand forecasting, tourism, etc. The choice of the ARIMAANN model in this study is applied after the performance of the ARIMA and ANN methods, which resulted in a need for improvement. To our knowledge, this method has no application in modelling and forecasting tourist flows in Albania. This lack of literature creates a space for exploring and evaluating the performance of this method in a new context. The results of this study will contribute to the scientific literature and will provide a valid model for the analysis of tourist flow forecasting in the country.

All these studies, which have been applied or not in the field of tourism, show the effectiveness of the ARIMAANN model in modelling complex data and seasonality, a characteristic that is essential for the analysis of tourist flows. Considering the results of the above works, our study aims to apply and evaluate this method in a field that has not yet been applied: tourism forecasting in Albania. This will help policymakers, tourism businesses, and government institutions improve the design of strategies for managing tourist flows and promoting Albanian tourism.

Modelling the Number of İnternational Tourists in Albania

Understanding and correctly predicting the number of international tourists visiting Albania is vital for effective planning, resource allocation, and policy-making in the tourism industry. As there are significant seasonal variations and more visitors today than before, it is imperative to use the right model that can capture both short-term fluctuations as well as long-term trends. Different forecasting methods can analyze historical data, each with advantages and disadvantages. This section examines the application of three techniques, ARIMA, ANN, and a combined ARIMAANN model, to explain the trends and forecast international tourist arrivals in Albania.

AutoRegressive Integrated Moving Average

ARIMA (AutoRegressive Integrated Moving Average) is one of the oldest methods of modelling and forecasting time series. However, at the same time, it is one of the best and has proved effective in many recent studies. It functions by using historical values (AutoRegresive), smoothing the series to make it stationary

(Integrated) and studying the errors or residuals using the moving average method (Moving Average). We used the "auto.arima" function in Rstudio to fit the ARIMA model to our data on international tourists' arrival in Albania. According to this procedure, ARIMA(1,1,2)(0,1,2)[12] was chosen as the best-fit model. The point forecasts and corresponding confidence intervals of the model selected by auto.arima are illustrated in the following figure.

```
Coefficients:
         ar1      ma1      ma2     sma1    sma2
      0.6531  -0.5172  -0.3340  -0.8235  0.1917
s.e.  0.1485   0.1663   0.1077   0.1175  0.1172

sigma^2 = 1.33e+10:  log likelihood = -1231.18
AIC=2474.37   AICc=2475.33   BIC=2489.63
```

Figure 3. ARIMA(1,1,2)(0,1,2)[12]

The parameters (1, 1, 2) are for the non-seasonal component and the parameters (0,1,2)[12] for the seasonal component.

- p = 1 indicates a first-order autoregressive (AR) component, which means that a linear combination of the previous value gives the current value. The value of this parameter tells us that our modelling will have the form$y_t = 0.6531\ y_{t-1} + \varepsilon_t$, y_t is the current value of the series at time t, $y_{t-1}$is the previous value and $\varepsilon_t$is the error at the time t(the difference between the predicted value and the actual value).
- d = 1 indicates that The differentiation process is used once to make the time series stationary. In the case of our series of the number of incoming tourists, the number of tourists increases every year, showing a clear upward trend. After the first differentiation, the limits of this series are transformed into differences between consecutive months, turning the series into a stationary series. The formula after this step will be$y_t - y_{t-1} = 0.6531(\ y_{t-1} - y_{t-2}) + \ \varepsilon_t$.
- q = 2 shows that The model uses error terms from the last two months in the moving average. Which will help correct the forecasts for the current month. The model is also restated as:

$$y_t - y_{t-1} = 0.6531(\ y_{t-1} - y_{t-2}) - 0.5172\ \varepsilon_{t-1} - 0.334\ \varepsilon_{t-2} + \ \varepsilon_t$$
$$y_t = 1.6531\ y_{t-1} - 0.6531\ y_{t-2} - 0.5172\ \varepsilon_{t-1} - 0.334\ \varepsilon_{t-2} + \ \varepsilon_t$$

- P = 0, the absence of a seasonal autoregressive component means that this model does not use past seasonal values to generate forecasts.
- D = 1, A seasonal difference is made to station the series as a seasonally unadjusted series. This is achieved by making the differences between each month and the previous year's respective month.
- Q = 2 shows that the model includes the impact of two preceding seasonal value $y_t$errors. In our case, since the series has a frequency of 12, the error of periods $t - 12$will be included in the modelling$t - 24$.
- [12] shows that our data are monthly, meaning the seasonality has a periodicity of 12.

The modelling expresses the combination of the above seasonal and non-seasonal components

$$y_t - 0.6531\ y_{t-1} - y_{t-1} + 0.6531 y_{t-2} - y_{t-12} + 0.6531\ y_{t-13} + y_{t-13} - 0.6531\ y_{t-14}$$
$$= \varepsilon_t - 0.5172\ \varepsilon_{t-1} - 0.334\ \varepsilon_{t-2} - 0.8235\ \varepsilon_{t-12} + 0.1917\ \varepsilon_{t-24}$$

Let us evaluate the statistical indicators for this method to understand how effective and accurate this model is. Table 2. Statistical metrics of ARIMA model accuracy

Table 2. Statistical metrics of ARIMA model accuracy

Model	**RMSE**	**MAE**	**MAPE**	R^2
ARIMA (1,1,2)(0,1,2)[12]	105171.1	67609.17	31.57071	0.9463824

According to the calculations of the above statistical coefficients, we can accept that this is a good model for modelling and forecasting the number of tourists in Albania since the MAE and RMSE are about 12% and 19% of the average value of historical data, values that categorize the implemented model as a good model. The value of the coefficient of determination $R^2 = 0.95$ shows that the model performs very well, representing about 95% of the data and leaving only 5% of the variation unexplained, but MAPE= 31.57071 is a value that leaves room for improvement. The modelling results and the forecast for the next 5 years according to the ARIMA method are graphically presented in Figure 4. (https://rpubs.com/Breshanaj/1268987)

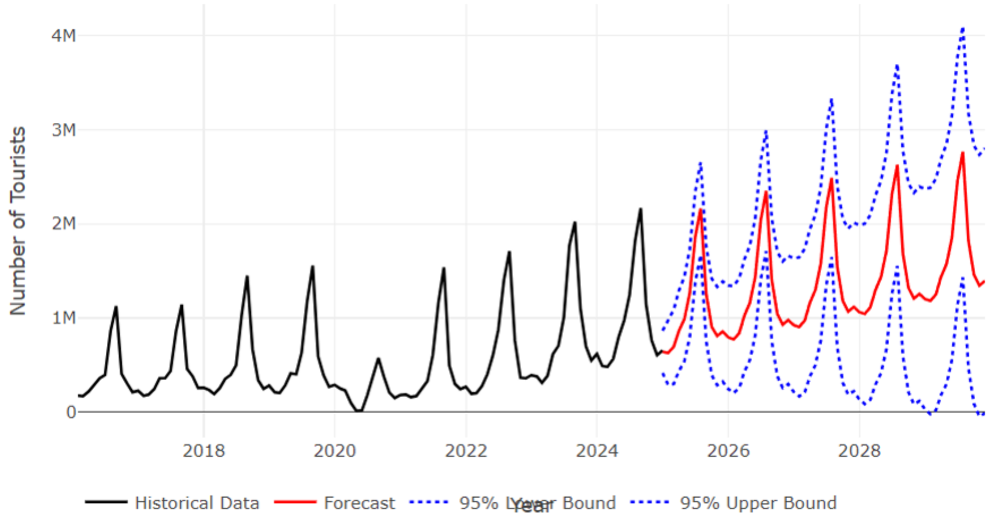

Figure 4. Forecasting tourist flow with the ARIMA method

Graphically, the optimistic forecast for the next five years is evident. According to the ARIMA (1,1,2)(0,1,2) model[12], the upward trend and seasonality of the monthly data on the number of non-resident tourists in Albania are expected to continue. The highest value, predicted according to this model, is expected to be recorded in August 2029 with a figure of about 2.8 million visitors. ARIMA works well in studying linear models but cannot study and process nonlinear data. We use the ANN method to fill this gap, which has shown efficiency in modelling nonlinear or complex data.

Artificial Neural Networks

The Artificial Neural Network (ANN) method is a powerful and advanced artificial intelligence and machine learning technique. It is based on imitating the structure and function of the human brain to analyze and process information (Breshanaj & Bakaj, 2024). An artificial neural network (ANN) is a computer simulation of the human brain in its most basic form (Fang et al., 2022). A normal brain can adapt to different environments and learn new things. The brain can evaluate partial, ambiguous, and ambiguous information and generate its own conclusions (Bharath et al., 2016).

Tourism time series usually have nonlinear data content and strong seasonality, which are handled quite well by artificial neural networks (ANN). We use the “neuralnet ” package in Rstudio to model our historical data on the number of non-resident tourists in Albania. The Artificial Neural Network model has three layers: input, hidden, and output. The input layer in this study uses five input values: the number of tourists in the previous month, the number of tourists in the past 2, 3 and 6 months, and the corresponding value from a year ago to understand historical trends and seasonality.

In the architecture of this model, we have built two hidden layers with 12 and 6 neurons, respectively. The first layer, with 12 neurons, receives information from the input layer, processes it by analyzing the relationships between the data, trends, and seasonality and, based on this information, creates the model. The sigmoid ("logistic") function is used in this layer to help model nonlinear data. With six neurons, the second layer processes the model obtained from the first hidden layer, making it more sophisticated and sending the information to the output layer. The interaction of both layers helps in improving the accuracy of the predictions. The output layer contains only one neuron, which forecasts the number of tourists for the next month. The function in this layer is linear, allowing the ANN to produce continuous values. In the figure below, the corresponding architecture of the ANN model is built.

After processing, modelling and forecasting data for international tourism in Albania with the ANN method in Rstudio, it turns out that the model has been adapted to 89% of the variance of historical data ($R^2 = 0.89$) and can deviate on average 22.7% from absolute values (MAPE=22.7).

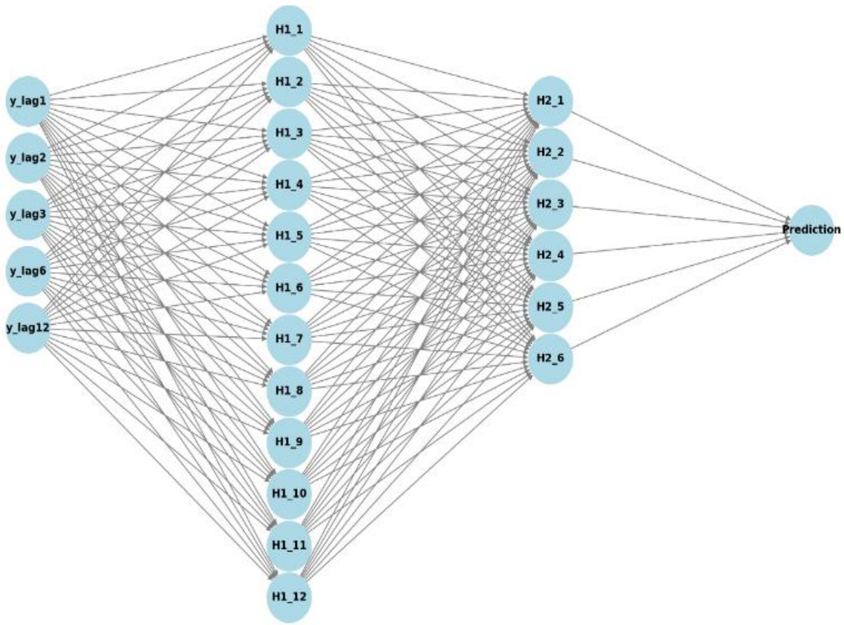

Figure 5. Architecture of the artificial network for tourism forecasting

Table 3. Statistical metrics of ANN model accuracy

Model	RMSE	MAE	MAPE	R^2
ANN	234026.4	187720.7	23.1	0.889

The ANN results show an $R^2 = 0.89$ and a MAPE = 23.1%, indicating that the model is effective in enhancing the forecast accuracy compared to ARIMA. The main advantage of ANN is that it is pretty flexible in capturing complex relationships and anomalies in the data, which helps forecast sudden changes in tourism. However, the model has a higher RMSE and MAE than ARIMA, which means there might be overfitting or lack of interpretability. The tables show that ARIMA has a higher accuracy in terms of RMSE and MAE, which indicates that its models are more robust to historical data. On the other hand, ANN has a lower MAPE, which indicates an enhancement in the capacity to model nonlinear variations in the data.In conclusion, there is no perfect method on its own. ARIMA is good at analyzing linear and seasonal data, while ANN is good at nonlinear patterns and anomalies. This has resulted in applying a hybrid ARIMAANN model that combines the best features of both models to develop a single model that can provide a better and more holistic forecast of tourism in Albania.

ARIMAANN

Complex real-world problems, in which nonlinearity is often present, can be successfully modelled using this technique (Perveen et al., 2024). It is universally stated that no single method is sufficient for problem-solving in every possible situation (Perveen et al., 2024). This has led researchers to combine different methods to achieve the best results for modelling or forecasting time series in various real-life problems. One of the most potent combinations has resulted in the ARIMAANN method, which manages data by dividing tasks between the traditional ARIMA and ANN artificial intelligence methods. The ARIMA model includes the linear trends and seasons in the time series data, and it works by identifying and extracting the regular patterns that can be expected to occur in the data set. The ANN model is designed to work with the nonlinear components of the time series, given its capability to learn and generalize from the data where the ARIMA model may be lacking. This way, the model can enhance the forecast precision to avoid the consequences of using a single method. This way, the data is first employed to build the ARIMA model to learn the linear structure of the data. Then, the residuals (the errors from the ARIMA model) are further analyzed using ANN to learn the nonlinear relationships in the data.

We have applied the hybrid ARIMAANN method to model and forecast the number of tourists in Albania. We have used the "ARIMAANN" package in Rstudio to model and forecast our monthly data. The accuracy metrics for this model show us that the model fits 96% of the data and can deviate on average 21.6% from the actual values (MAPE=21.6).

Table 4. Statistical metrics of the accuracy of the ARIMAANN model

Model	RMSE	MAE	MAPE	R^2
ARIMAANN	93012.66	64528.77	21.6	0.96

According to the ARIMA-ANN hybrid model, which was applied for modelling and forecasting the tourists in Albania, the result fit the historical data ($R^2 = 0.96$). Although the model is quite precise, the mean absolute error estimate (MAPE = 21.6%) suggests that other factors can significantly affect the variations in the number of tourists. Nevertheless, this model can be considered a potentially effective method of tourist flow analysis and forecasting in the Albanian context and can be helpful for decision-makers in the tourism industry. To enhance the precision of the forecasts, it is possible to try other combinations of methods or to use other factors that may impact tourist movement. In the graph below (https://rpubs.com/Breshanaj/1268516), built with "plotly" in R, we have reflected the monthly values of tourist arrivals in Albania and the forecast for the next five years.

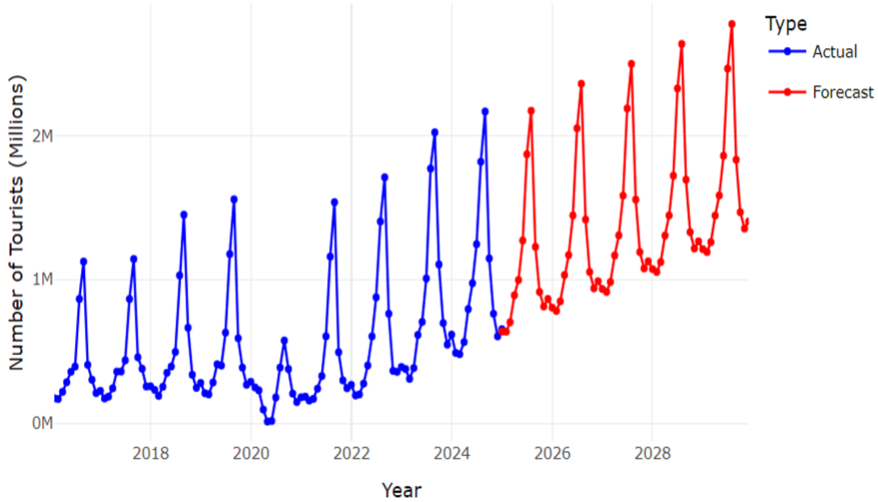

Figure 6. Forecasting tourist flow with the hybrid ARIMAANN method

The forecast graph shows a clear upward trend for the coming years, reflecting a continued expansion of the tourism sector in Albania. According to the ARIMAANN hybrid model, the projections show that the number of international arrivals will continue to increase, reaching a predicted peak of 2,775,862 tourists in August 2025. This result suggests a stabilization of tourism growth after the pandemic and a strong seasonal trend, with the summer months continuing to be the periods with the highest intensity of visitors. The modelling in this paper can aid in resolving decisions by all concerned in managing tourism demand and ensuring the highest quality of tourism.

ARIMAANN for the Period 2021-2029

To avoid the impact of the COVID-19 pandemic period (Sá et al., 2023), we applied the ARIMAANN model to the monthly data of non-resident tourist arrivals in Albania for 2021-2029. As shown in Table 5, this significantly improved model, referred to as ARIMAANN 21-24, showed a significant improvement in all performance metrics compared to the previous models. Statistical evaluations show an RMSE of 48761.5, an MAE of 39700.84, a MAPE reduced to 7.09%, and an R^2 of 0.99, suggesting an almost perfect fit to historical data and a reliable forecast for tourism for the coming years.

Table 5. Statistical metrics of the accuracy of the ARIMAANN model

Model	RMSE	MAE	MAPE	R^2
ARIMAANN 21-24	48761.5	39700.84	7.09	0.99

The graph below shows the trend in the number of intern.ational tourists in Albania for the next five years. The forecasts confirm a progressive increase in tourism, with the highest points during the summer season, where the maximum number of visitors is expected to reach approximately 2.9 million in August 2029. https://rpubs.com/Breshanaj/1269266

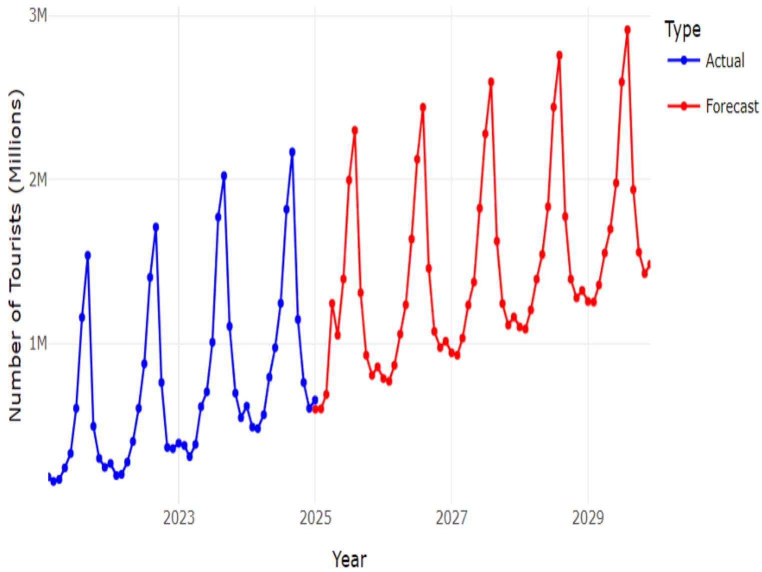

Figure 7. Hybrid ARIMAANN forecasting 2021-2029

These findings highlight the importance of hybrid models, where combining traditional methods with advanced artificial intelligence methods helps in strategic planning and careful management, becoming the foundation for sustainable growth in the tourism sector.

Results and Discussion

The results of this study clearly show that the models used to forecast the number of international tourists in Albania have significant differences in accuracy and appropriateness with historical data. To compare the ARIMA, ANN and ARIMAANN models, we use statistical indicators such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) (Narváez-Villa et al., 2021) and the coefficient of fit (R^2). The accuracymeasurement results are presented in the table below.

Table 6. Performance of ARIMA, ANN, and ARIMANN models

Model	RMSE	MAE	MAPE	R^2
ARIMA	105171.1	67609.17	31.57071	0.946
ANN	234026.4	187720.7	23.1	0.889
ARIMAANN	93012.66	64528.77	21.6	0.96
ARIMAANN 21-24	**48761.5**	**39700.84**	**7.09**	**0.99**

The combined ARIMAANN model, which integrates the two approaches, was the most accurate and best fitting of the three models, with the highest value $R^2$of 0.96 and fewer errors than the other models. This result shows that the combination of ARIMA, which models the linear and seasonal components of the data, and the nonlinear trends of ANN creates a better and more complete model for forecasting tourist flows.

The five-year future forecasts are optimistic and can be seen in all models applied in this study: ARIMA, ANN and ARIMAANN. Based on these models, it is predicted that the highest number of international tourists is expected during the summer months, and the peak is expected to be in August 2029, when the tourist could reach 2.8-2.9 million visitors. This result also supports the existence of a strong seasonality, where the summer months are peak seasons in terms of tourist arrivals while the winter months are low seasons. These findings are significant for policymakers, tourism managers, and government agencies as they can be used to effectively manage tourism dynamics concerning resource allocation and infrastructure development. The model suggested in this paper can also be employed to design marketing strategies and enhance the capacity of the tourism sector to address growing demands, thereby contributing to the sustainable development of the tourism sector in Albania.

Based on the study's findings, the hybrid ARIMAANN approach is a more complex and appropriate model for forecasting tourism demand than the application of individual models. While the ARIMAANN model has demonstrated a significant improvement compared to the individual ARIMA and ANN models, a more advanced approach, referred to as ARIMAANN 21-24, has shown even more accurate results by eliminating the effects of the COVID-19 pandemic. Compared to the previous version of the ARIMAANN model, the RMSE value has been reduced from 93012.66 to 48761.5; the MAPE has undergone an improvement from 21.6% to 7.09%, a value that categorizes the model as very good. This shows that the new model fits better with recent historical data, providing a more accurate and reliable forecast for tourist flows in Albania. Furthermore, the R^2 number has increased from 0.96 to 0.99, suggesting an almost perfect fit for the model with the data used for training. These findings are important to policymakers, tourism managers, and government agencies because they can assist in effectively managing tourism dynamics regarding resource allocation and infrastructure development. The model suggested in this paper can also be employed to design marketing strategies and enhance the capacity of the tourism sector to address growing demands, thereby contributing to the sustainable development of the tourism sector in Albania.

From the study's findings, the hybrid ARIMAANN21-24 approach is a more sophisticated and suitable model for forecasting tourism demand than individual models. The combination of classical approaches and artificial intelligence has led to a significant improvement in the accuracy of forecasts and a better fit to historical data, providing more accurate and reliable results.

Conclusion

Tourism is a rapidly growing industry in most countries (Saba et al., 2022) and requires more accurate modelling and forecasting of tourist arrival data for many purposeful decisions. (Tanzila Saba, Mirza Naveed Shahzad, Sonia Iqbal, Amjad Rehman, Ibrahim Abunadi, 2022) The results of our study clearly show that the hybrid ARIMAANN model has provided the best performance compared to the individual ARIMA and ANN models, confirming the conclusions of previous studies that support the power of combining traditional statistical methods with artificial intelligence.

In line with the studies of Zhang (2003), Atesongun and Gulsen (2024) and Rahayu et al. (2023), this study demonstrates that the hybrid approach provides a significant improvement in forecast accuracy, especially for time series with nonlinear trends and pronounced seasonality, as in the case of Albanian tourism. Also, the results of our study converge to the same conclusion as the findings of Musonera et al. (2021), highlighting the need for advanced models that combine statistical analysis and artificial intelligence algorithms to increase the reliability of forecasts.

In addition to the higher accuracy of the ARIMAANN21-24 model, this study has shown that the trend of increasing international tourist arrivals in Albania will continue in the coming years, with a peak predicted during the summer months. In line with existing studies on hybrid models, this study shows that the selection of the training period significantly impacts the accuracy of the forecasts. Through ARIMAANN 21-24, it is demonstrated that by eliminating the effects of exceptional periods such as the COVID-19 pandemic, a significant reduction in forecast errors and an increase in the fit to recent tourism data in Albania can be achieved. This model's improvement of MAPE and RMSE reinforces the importance of using the most appropriate training periods and hybrid methods to optimize future forecasts in the tourism sector. These results have important implications for strategic tourism management, helping policymakers and tourism operators to improve resource allocation and infrastructure development more efficiently. Although the hybrid model has provided promising results, it cannot capture all the factors that affect tourism, such as sudden political, economic, or climatic changes.

This study leaves room for further studies, where the integration of external influencing factors in the field of tourism or the application of other hybrid methods, such as SARIMA + LSTM**,** ARIMA + XGBoost, or Wavelet Transform + ARIMA + ANN, etc., could further improve the forecast of tourist flows in Albania. These results can be generalized to predict tourist arrivals in any country or region with a complicated data model.

Recommendations

The results of this study highlight the efficiency of hybrid models, especially ARIMAANN 21-24, for enhancing the accuracy of tourism demand forecasting. Policymakers, tourism operators, and government agencies should include such models in their decision-making process for better resource management, infrastructure, and marketing strategies. The interaction between academic institutions, government, and tourism organizations is vital in encouraging innovation and ensuring that the forecasting models are relevant to the practice. Future work should also incorporate other external factors like political risk, climate change and economic factors to enhance the precision of the forecasts. Also, applying advanced hybrid models such as SARIMA + LSTM, ARIMA + XGBoost, or Wavelet Transform + ARIMA + ANN can produce even better results. The forecasts obtained from the models should be used as a basis for the formulation of sustainable tourism policies that can support the economic development of the country while at the same time protecting the environment and improving the quality of services offered to tourists. These suggestions are meant to assist in the growth of a strong and strategic tourism sector in Albania.

Scientific Ethics Declaration

* The authors declare that the scientific, ethical, and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest.

Funding

* We thank the "Ismail Qemali" University of Vlora for supporting this work. A special recognition also goes to the Faculty of Technical and Natural Sciences, which provides an inspiring academic environment and encourages the development of scientific research.

Acknowledgements or Notes

* This article was presented as a poster presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Trabzon/Türkiye on May 01-04, 2025.

References

Abdou, M., Musabanganji, E., & Musahara, H. (2021). Tourism demand modelling and forecasting: A review literature. *African Journal of Hospitality, Tourism and Leisure*, *10*(4), 1370-1393.

Atesongun, A., & Gulsen, M. (2024). A hybrid forecasting structure based on arima and artificial neural network models. *Applied Sciences, 14*(1671), 1-10.

Breshanaj M., & Bakaj, A. (2024). Application of ANN in the tourism industry in Albania. *The Albanian Journal of Economy and Business*, 85-102.

Fadhillah, R., Kusnandar, D., & Huda, N. M. (2024). Pemodelan Arima-ann pada harga saham bank mandiri. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, *13*(1), 117-126.

Fang, H., Tu, Y., Wang, H., He, S., Liu, F., Ding, Z., & Cheng, S. S. (2022). Fuzzy-based adaptive optimization of unknown discrete-time nonlinear Markov jump systems with off-policy reinforcement learning. *IEEE Transactions on Fuzzy Systems*, *30*(12), 5276-5290.

Fraiha- Lopes, R. L., Fraiha, S. G., Gomes, H. S., Lima, V. D., & Cavalcante, G. P. (2020). Application of hybrid ARIMA and artificial neural network modelling for electromagnetic propagation: An alternative to the least squares method and ITU recommendation p. 1546-5 for Amazon urbanized cities. *International Journal of Antennas and Propagation*, *2020*(1), 8494185.

Kukreja, H., Bharath N., Siddesh C., & Kuldeep, S. (2016). An introduction to artificial neural network. *International Journal of Advance Research and Innovative Ideas in Education, 1*(5), 27-30.

Mentes, A. & Yetkin, M. (2020). An application of soft computing techniques to predict dynamic behaviour of mooring systems. *Brodogradnja: An International Journal of Naval Architecture and Ocean Engineering for Research and Development*, *73*(2), 121-137.

Narváez-Villa, P., Arenas-Ramírez, B., Mira, J., & Aparicio-Izquierdo, F. (2021). Analysis and prediction of vehicle kilometers traveled: A case study in Spain. *International Journal of Environmental Research and Public Health*, *18*(16), 8327.

Perveen, N., Tariq, K., & Ahmad, H. S. (2024). Comparative studies of hybrid ARIMA and artificial neural network (ANN) techniques for predicting exchange rate in Pakistan. *Bulletin of Business and Economics (BBE)*, *13*(2), 630-636.

Sá, J., & Luís, A. L. (2023). Industrialization is not the solution for creating wealth. *The Journal of Social, Political, and Economic Studies*, *47*(3/4), 324-332.

Saba, T., Shahzad, M. N., Iqbal, S., Rehman, A., & Abunadi, I. (2022). A new hybrid SARFIMA-ANN model for tourism forecasting. *Cmc-Computer Materials and Continua*, *71*, 4785-4801.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.

Author(s) Information

Morena Breshanaj
University "Ismail Qemali"
Vlore, Albania
Contact e-mail: *morena.breshanaj@univlora.edu.al*

Areti Stringa
University of Tirana
Tirana, Albania

To cite this article:

Breshanaj, M., & Stringa, A. (2025). Designing intelligent models with ARIMAANN for visionary forecasts. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 34,* 221-233.