Journal homepage: http://iieta.org/journals/isi

# Assessing Tourism Prediction Models: A Comparative Study of SARIMA, Random Forest, and LSTM, Considering Nonlinear Trends and the Influence of COVID-19

Samya Bouhaddour<sup>1, 2\*10</sup>, Mohammed Sbihi<sup>10</sup>, Fatima Guerouate<sup>10</sup>, Chaimae Saadi<sup>10</sup>

<sup>1</sup> Laboratory of Analysis Systems, Processing Information and Industrial Management, The Higher School of Technology of Sale, Mohammed V University, Rabat 8007, Morocco <sup>2</sup> SmartiLab Laboratory, Moroccan School of Engineering Sciences (EMSI), Rabat 10000, Morocco

Corresponding Author Email: Samyabouhaddour@research.emi.ac.ma

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

### https://doi.org/10.18280/isi.290631

ABSTRACT

Received: 25 March 2024 Revised: 4 September 2024 Accepted: 25 September 2024 Available online: 25 December 2024

### Keywords:

machine learning, deep learning, forecasting, non-linear data, Singapore, tourism, Seasonal Autoregressive Integrated Moving Average (SARIMA), RF, LSTM

# This study presents an analytical exploration of Time Series Forecasting (TSF), comparing econometric, machine learning, and deep learning models. Using a comprehensive dataset spanning from January 2008 to August 2023, including the COVID-19 pandemic period, monthly data from Singapore were analyzed. The study further decomposes the data into three testing periods, each reflecting Singapore's evolving travel conditions during the pandemic. The research evaluates the performance of each model under these diverse conditions, demonstrating significant improvements in forecasting accuracy through various data decompositions. Evaluation metrics, including the coefficient of determination (R<sup>2</sup>) and Root Mean Square Error (RMSE), quantify model performance and highlight promising results in specific decompositions. The study emphasizes the effectiveness of different TSF models in accurately forecasting time series data, with empirical results favouring the deep learning model.

### 1. INTRODUCTION

A time series (TS) consists of a sequence of data points or observations that are collected at regular, predetermined time intervals. These data points are typically organized in chronological order, with each data point corresponding to a specific moment in time. Time series data can cover a wide range of time intervals, from seconds to extended periods, depending on the context. Time series forecasting (TSF) is a methodology used to predict future outcomes based on historical data that has been chronologically arranged. This process is part of the broader field of time series analysis, which includes various techniques for extracting insights from time-ordered data. In TSF, the focus is on generating predictions or forecasts by identifying patterns and trends in historical data [1]. Making accurate and reliable forecasts remains one of the most challenging research topics in time series analysis and forecasting [2, 3]. Two popular categories of forecasting models are econometric models and artificial intelligence (AI) models [4].

Econometric models, which employ statistical and mathematical techniques to analyze data and relationships, have maintained their popularity since 2005. These models continue to serve as commonly used benchmarks for evaluating forecasting performance and comparing their predictive accuracy with emerging models. Econometric models primarily focus on the analysis of data, aiming to understand and quantify relationships using statistical and mathematical tools [5]. They are widely used in various disciplines, including tourism, to make informed predictions and draw conclusions based on empirical data and relevant theories. These forecasts help in resource allocation, setting priorities, and assessing potential risks.

The incorporation of Artificial Intelligence is increasingly being utilized in predicting tourist arrivals due to the prevalence of big data and the complex nature of tourismrelated datasets [6]. Tourism plays a crucial role in driving economic growth. These forecasts provide critical insights for tourism professionals and researchers.

The continuous advancements in AI, including machine learning (ML) and deep learning (DL) models, have showcased their ability to capture intricate patterns and nonlinear relationships within the data [4, 7]. This is particularly valuable for tourism forecasting, given the complexity of factors such as seasonality, economic trends, and shifting tourist preferences. AI models have brought about a significant transformation in the domain of forecasting tourist arrivals. By utilizing sophisticated algorithms and data analytics [6], these methods analyze historical data encompassing tourism patterns, weather conditions, economic indicators, and relevant variables. Subsequently, ML models excel at providing accurate predictions of future tourist arrivals [8]. What sets them apart is their adaptability and capacity to learn from new data, thereby continually enhancing their forecasting accuracy continually. Globally, ML is playing an indispensable role in optimizing the management and strategic planning of tourist destinations. With advancements over conventional AI models, DL has been widely used for prediction. DL has emerged as a powerful tool for predicting tourist arrivals within the framework of time



series analysis [9]. This advanced ML approach leverages neural networks with multiple layers to extract intricate temporal patterns and dependencies from historical data [10]. DL models can offer remarkably accurate forecasts of future tourist arrivals. What sets DL apart is its ability to automatically uncover latent features and adapt to complex, non-linear relationships within time series data.

It's important to recognize that the emergence and growth of AI owe much to the foundation laid by econometric models. The rigorous statistical and mathematical framework established by econometrics provided valuable insights into modeling data and understanding relationships, which served as a basis for the development of more advanced AI techniques. As a result, AI has not replaced econometric models but rather complemented them, offering new opportunities to enhance forecasting accuracy and gain deeper insights from data, particulary in tourism, where predicting tourist arrivals can benefit from AI-driven approaches [11]. The precision of predictions can vary significantly between econometric and AI models, with both offering valuable approaches for forecasting.

### 2. STATE OF THE ART

Econometric models, rooted in established statistical and economic principles, are frequently chosen for domains with relatively stable relationships, providing robust and interpretable results. Notably, among these econometric models, ARIMA (Auto-Regressive Integrated Moving Average), conceived in the 1970 [12], has emerged as a particularly valuable tool for forecasting tourist arrivals [13, 14]. On the other hand, AI models, particularly ML and DL, excel in capturing complex trends and nonlinear patterns, making them well-suited for dynamic and unpredictable data.

Tovmasyan [15] employed the ARIMA model to forecast travel patterns in Armenia during the COVID-19 epidemic. Time series data from 2015 to 2020 were analyzed, and the results showed a significant drop in tourism from 1,894,377 in 2019 to 375,216 in 2020. The average annual growth rate for 2021–2023 is projected to be 12.81%, 13.42%, and 13.66%, respectively, indicating a positive trend in the recovery of tourism. Despite these optimistic forecasts, the research acknowledges several limitations. Notably, it does not include data from 2020 because of the sudden onset of the pandemic, which might distort the results if not appropriately adjusted for context. Although this research provides valuable insights into tourist forecasting, it also emphasizes the need for continuous data gathering and methodological improvement.

Makoni et al. [16] utilized the SARIMA model to project international tourist arrivals in Zimbabwe, yielding precise predictions with favorable RMSE and MAPE values. Nonetheless, the research has several limitations. The forecasts do not account for potential disruptions caused by global events such as COVID-19, which could significantly affect tourism trends. Additionally, the model assumes data stationarity, which may not fully capture the dynamic nature of tourism influenced by changing market conditions and external factors. The SARIMA approach, while effective, may not capture all the complexities of contemporary forecasting techniques, which could enhance accuracy and adaptability.

Qiu et al. [17] conducted a study on predicting tourist arrivals for 20 countries using various predictive models. They found that the SARIMA, ETS, and STL models were the most accurate among single models. While stacking models, which combine multiple methods, generally provide higher accuracy, the study found that combining five single models was optimal for their specific dataset. A key limitation of this research is that the optimal number of models for stacking may not be universally applicable, potentially affecting the generalizability of the results. Additionally, the study might not fully account for external factors which could impact the accuracy of predictions in real-world scenarios.

In a comparative study by Bouhaddour et al. [13], the effectiveness of the SARIMA and PROPHET models in forecasting tourism in Singapore was evaluated using historical data on arrivals. The first model is known for its ability to capture both seasonal and non-seasonal trends, whereas the second model is designed to handle changes in trends and seasonality. The study employed performance metrics such as the mean absolute error and root mean squared error to assess the models' forecasting accuracy. The results indicated that the PROPHET model provided more accurate predictions than the SARIMA model. These findings are significant for enhancing tourism demand forecasting and assisting policymakers and industry stakeholders in making better-informed decisions.

Wu et al. [18] introduced an innovative forecasting approach by merging Seasonal Autoregressive Integrated Moving Average (SARIMA) with Long Short-Term Memory (LSTM), resulting in a hybrid model that integrates both econometric principles and deep learning capabilities. This innovative method was created specifically to forecast daily visitor arrivals in Macau Special Administrative Region, China, quickly. Long-term patterns in time series data are efficiently captured by LSTM, an AI technique renowned for its nonlinearity. The hybrid model of SARIMA + LSTM outperformed other predictive algorithms by combining the predictive powers of SARIMA with the capacity of LSTM to reduce residuals. The tourist industry's best practices are advanced by this study, particularly with regard to making accurate daily arrival estimates.

Bouhaddour et al. [19] present the PROPHET-BGP-FNN model as a superior approach for forecasting daily tourist arrivals in Hawaii, outperforming traditional models such as SARIMAX, PROPHET, and LSTM in terms of predictive accuracy. By combining PROPHET's trend and seasonality detection with the BGP-FNN's ability to capture complex nonlinear patterns, the model achieves lower MAE and mean MAPE values, demonstrating its effectiveness in handling the nuances of tourism data. However, the study is limited by its focus on Hawaiian data, which constrains the model's applicability to other regions and longer time series. Despite these constraints, the PROPHET-BGP-FNN model holds promise for enhancing decision-making in tourism and other sectors reliant on accurate forecasting.

This study offers a novel and comprehensive approach to Time Series Forecasting (TSF) by integrating and comparing econometric, ML, and DL models using an extensive dataset from Singapore that spans from January 2008 to August 2023, including the COVID-19 pandemic period. Unlike many studies that exclude the COVID-19 period from their datasets to simplify predictions, this research retains this complex period in the analysis. By including the pandemic period, the study acknowledges its impact on tourism forecasting and aims to provide a more accurate and realistic evaluation of model performance. The study's distinctiveness lies in its methodological rigor, particularly the decomposition of data into three distinct testing periods that reflect Singapore's evolving travel conditions during the pandemic. This approach allows for a nuanced evaluation of model performance across varying conditions, which is crucial for capturing the effects of unprecedented disruptions like COVID-19.

The research advances the field by demonstrating how different TSF models, including advanced DL models, can be evaluated under diverse conditions to enhance forecasting accuracy. By employing a robust set of evaluation metrics. such as the coefficient of determination (R<sup>2</sup>) and Root Mean Square Error (RMSE), the study provides new insights into the performance of these models across different phases of the pandemic. The empirical results reveal that DL models show particularly promising outcomes, suggesting that these models may offer superior forecasting accuracy in complex and dynamic environments. This contribution is significant as it highlights the effectiveness of integrating multiple TSF methodologies and adapting them to evolving conditions, while also emphasizing the importance of including all relevant periods in the dataset to avoid skewed predictions. This approach offers valuable guidance for improving tourism demand forecasts and aids decision-making for policymakers and industry stakeholders.

The manuscript adheres to a well-structured organization. Section 3 offers an elaborate elucidation of the methodology. Section 4 describes the procedure, including data preparation, model implementation, and a detailed analysis of the empirical results. The concluding observations are presented in Section 5.

### 3. METHODOLOGY

### 3.1 Data description

The dataset comprises total monthly arrivals, sourced from both the Civil Aviation Authority of Singapore and the Singapore Tourism Analytics Network (STAN). It encompasses the monthly influx of inbound tourists to Singapore from January 2008 to August 2023 [20] as depicted in Figure 1. An examination of the annual inbound tourist arrivals during this period unveils discernible trends and patterns.



Figure 1. Number of inbound tourists' arrivals to Singapore and monthly grow rate in arrivals (2008-2024) Source: Civil Aviation Authority of Singapore and Singapore Tourism Analytics Network (Stan).

Table 1. Summary statistics of the datasets

Variable	Year	Std. Deviation	Minimum	Maximum
	2008	53,821.44	745,794	922,569
	2009	75,105.03	689,935	972,233
	2010	74,969.81	857,387	1, 127,581
	2011	83,701.71	990,118	1,273,870
	2012	84,208.00	1,051,348	1,360,536
	2013	96,500.97	1,176,142	1,483,520
	2014	97,047.54	1,098,626	1,408,400
Tourist amizala	2015	122,674.57	1,131,976	1,519,233
Tourist arrivals	2016	133,575.98	1,144,063	1,622,405
	2017	96,676.52	1,320,050	1,632,147
	2018	99,917.57	1,406,985	1,732,899
	2019	108,465.61	1,463,547	1,802,594
	2020	506,232.28	750	1,688,102
	2021	22,034.71	10,029	92,796
	2022	321,804.43	57,174	931,441
	2023	167,727.46	931,679	1,419,634

Over the years, there has been a steady increase in the number of visitors arriving in Singapore, with a significant surge in tourist arrivals in 2018 and 2019. However, the year 2020 saw a drastic decline in tourist numbers due to the global COVID-19 pandemic, which led to travel restrictions and a reduction in tourism activities worldwide. Nevertheless, a promising recovery was observed in 2023, signifying the gradual resurgence of tourism in Singapore.

The datasets underwent exploratory data analysis (EDA) using Python programming. The objective of this analysis was to extract insights from the tourism data that go beyond formal modelling. Notably, a sharp decrease in tourist arrivals and a subsequent plunge in tourism demand were observed in 2020 due to the impact of the coronavirus pandemic. This downturn resulted in significant job losses, severe economic challenges, and the closure of numerous businesses.

From 2008 to 2019, the annual mean tourist arrivals consistently rose, reaching its peak in 2019 at approximately 1.59 million visitors. This upward trend highlights Singapore's growing popularity as a tourist destination during this period. The year 2020 marked a significant deviation from this trend due to the unprecedented impact of the COVID-19 pandemic. The number of tourist arrivals plummeted dramatically to around 228,536, reflecting the severe disruption caused by the pandemic on global travel. However, subsequent years, especially 2023, show a rebound in tourist arrivals, indicating a hopeful recovery for Singapore's tourism industry. Table 1 showcases the descriptive statistics for yearly inbound tourist arrivals in Singapore from 2008 to 2023.

## 3.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA model is a fundamental econometric model in time series analysis, particularly effective in understanding and forecasting data with seasonal patterns [21]. The architecture of SARIMA is defined by three main components: autoregressive (AR), differencing (I) order, and moving average (MA). Additionally, SARIMA incorporates seasonal autoregressive (SAR) and seasonal moving average (SMA) components to model the seasonal component of the data [7]. The SARIMA model components are [14, 22, 23]:

• Autoregressive (AR) terms (p): The autoregressive component represents the relationship between an observation and several lagged observations (previous time points). The term "p" represents the order of autoregressive terms, indicating how many lagged observations are included in the model.

• **Differencing (I) order (d):** The differencing order represents the number of differences needed to make the time series stationary. Stationarity is an important assumption in time series modeling. If the data is not stationary, differencing is applied to remove trend and seasonality.

• Moving Average (MA) terms (q): The moving average component represents the relationship between an observation and the residual error from a moving average model applied to lagged observations. The term "q" represents the order of moving average terms, indicating how many lagged residuals are included in the model.

• Seasonal Autoregressive (SAR) terms (P): Similar to the autoregressive terms, SAR terms represent the relationship between an observation and several lagged observations at the seasonal intervals. The term "P" represents the order of seasonal autoregressive terms.

• Seasonal Moving Average (SMA) terms (Q): Similar to the moving average terms, SMA terms represent the relationship between an observation and the residual error from a moving average model applied to lagged observations at the seasonal intervals. The term "Q" represents the order of seasonal moving average terms.

• Seasonal differencing (D): The seasonal differencing order represents the number of differences needed at the seasonal intervals to achieve stationarity.

To effectively employ the SARIMA model, ensuring data stationarity is of paramount importance. Stationarity in a time series context refers to a state where crucial statistical properties such as mean and variance exhibit consistent stability throughout the entire observational period. Achieving stationarity often entails the judicious use of differencing, a process involving computing the differences between consecutive data points. This technique aids in removing trends and seasonal components, rendering the data stationary and thereby enabling a more accurate modeling process [24, 25].

**Strengths:** SARIMA excels in capturing both seasonal and non-seasonal variations in time series data. Its ability to model periodic fluctuations and trends makes it highly effective for tourism forecasting, where seasonal peaks and troughs are significant. The model's structured approach to dealing with seasonality ensures accurate predictions based on historical trends, which is essential for understanding and forecasting tourist arrivals [14, 16, 19].

**Limitations:** Despite its strengths, SARIMA has limitations, including the need for extensive parameter tuning and its assumption of linear relationships within the data. This can limit its effectiveness in scenarios where data exhibits non-linear patterns or sudden changes.

**Suitability for Tourism Forecasting:** SARIMA is particularly suitable for tourism forecasting because it can effectively model the regular seasonal patterns seen in tourist data. The model's performance has been notably strong in this study, reflecting its capability to provide accurate forecasts even when the dataset includes the disruptive COVID-19 period.

### 3.3 Random Forest (RF)

A Random Forest (RF) has grown in popularity due to its high reliability and practical application in various fields [26]. This model combines classification and regression trees with the bagging method to improve accuracy. As a machine learning model, RF is exceptionally well-suited for making predictions in the context of time series data, particularly for forecasting tourist arrivals. Random Forest excels in this domain by leveraging its ability to capture complex relationships and patterns within time series data. When applied to tourist arrivals, it can analyze historical arrival patterns, seasonal trends, and various other factors affecting tourism.

Random Forest is an ensemble learning method that combines multiple decision trees to create a robust and accurate predictive model [26]. The architecture of a Random Forest model is defined by several key components and parameters [8, 26]:

• Ensemble of Decision Trees: Random Forest is an ensemble model, combining multiple decision trees to form a robust predictive model.

• Random Subset of Features: At each node of a decision

tree, a random subset of features is considered for making splits. This randomness adds to the model's diversity.

• **Bootstrap Sampling:** Random Forest uses bootstrap sampling (with replacement) to create different subsets of the training data for training each decision tree.

• Voting or Averaging for Predictions: For regression tasks, predictions from all trees are averaged to produce the final prediction. For classification, each tree votes for a class, and the majority class is chosen as the prediction.

**Strengths:** RF is effective at modeling non-linear relationships and interactions between variables, which is crucial for understanding the multifaceted nature of tourism data. Its ensemble approach improves robustness and reduces overfitting, making it reliable for handling diverse data features and capturing intricate patterns in tourist arrivals [25-27].

**Limitations:** RF's complexity can lead to reduced interpretability of the model, as the ensemble of decision trees may obscure the influence of individual predictors. Additionally, RF can be computationally intensive, especially when working with large datasets.

**Suitability for Tourism Forecasting:** RF is suitable for forecasting tourism data due to its ability to manage complex interactions and diverse features. Its performance was competitive in this study, demonstrating its capability to predict tourism trends effectively, although it was outperformed by SARIMA in certain aspects.

### 3.4 Long Short-Term Memory (LSTM)

LSTM is a cutting-edge deep learning model renowned for its proficiency in handling and understanding sequences of data [19, 28]. LSTM is a specialized type of recurrent neural network (RNN) designed to address the limitations of standard RNNs in modeling long-term dependencies within sequential data. Its unique architecture allows it to capture and retain crucial information over extended sequences, making it particularly effective in various applications involving timeseries and sequential data [29].

The fundamental architecture of an LSTM unit consists of several key components, notably a cell state, an input gate, a forget gate, an output gate, and a set of activation functions [14]. The cell state serves as a conveyor belt, allowing information to flow through the unit with minimal alteration. The input gate regulates the amount of new information to be stored in the cell state, while the forget gate controls the extent to which existing information should be discarded. The output gate then filters the information to be output based on the current context [30].

The advantage of LSTM lies in its capacity to retain longterm memory, which is essential for analyzing sequences with gaps or delays [19, 31]. By controlling the flow of information through the cell state and gating mechanisms, LSTM effectively mitigates the vanishing or exploding gradient problems that hinder conventional RNNs. Its architecture, governed by the cell state and a system of gates, allows for the modeling of long-term dependencies by regulating the information flow within the unit.

LSTM's adeptness in capturing and retaining crucial information over extended sequences has positioned it as a foundational element in advancing the field of deep learning and its applications in various technological domains [18].

Strengths: LSTM networks excel in modeling temporal sequences with long-term dependencies, making them well-

suited for capturing complex and evolving trends in tourism data. Their ability to remember and utilize information over extended periods helps in forecasting scenarios were historical data spans significant gaps or disruptions [32].

**Limitations:** LSTM networks are computationally demanding and require careful tuning to avoid overfitting. Their complexity and the need for extensive training time can be a drawback, particularly when working with limited computational resources.

**Suitability for Tourism Forecasting:** LSTM is highly suitable for forecasting tourism data, especially in contexts with significant temporal dependencies. The model's advanced capabilities in handling long-term patterns were evident in this study, though **SARIMA** demonstrated superior performance in certain aspects, particularly in the context of seasonal forecasting.

### 3.5 Models evaluation

In this section, we present the metrics used to evaluate the forecasting accuracy of the econometric, machine learning (ML), and deep learning (DL) models. The two primary metrics used in this study are Root Mean Squared Error (RMSE) and the coefficient of determination ( $R^2$ ).

### **Root Mean Squared Error (RMSE):**

The RMSE is used to evaluate how well a model's predictions align with the actual data. Lower RMSE values indicate a better fit, meaning the difference between predicted and actual values is smaller [24]. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$
(1)

where:

- *n* is the number of data points (observations);
- $\hat{y}_i$  is the predicted value for the *i*-th observation;
- $y_i$  is the actual (observed) value for the *i*-th observation.

### **Coefficient of Determination (R<sup>2</sup>):**

The coefficient of determination, or  $R^2$ , is a statistical metric that indicates the proportion of variance in the dependent variable explained by the independent variables in a regression model [33]. It serves as a measure of how well the model fits the data. The  $R^2$  value ranges from -1 to 1, where:

• R<sup>2</sup>=1 indicates a perfect fit, meaning the model explains all the variability in the dependent variable,

 $\bullet$  R<sup>2</sup>=0 indicates that the model explains none of the variability around the mean of the dependent variable,

 $\bullet$  R<sup>2</sup><0 suggests the model is worse than a simple model based on the mean.

The formula for calculating  $R^2$  is:

$$R^{2}=1-\frac{SS_{res}}{SS_{tot}}$$
(2)

where,

•  $SS_{res}$  is the sum of squares residual (the squared differences between observed and predicted values),

•  $SS_{tot}$  is the total sum of squares (the squared differences between observed values and the mean of observed values).

### 4. RESULTS AND DISCUSSION

### 4.1 Data repartition

In order to construct and refine the model, the dataset was divided into two segments: a training set and a testing set. This data partitioning was based on the observed trend in tourist arrivals, characterized by fluctuations reflecting both upward and downward movements. These fluctuations were particularly pronounced during the challenging period of the COVID-19 pandemic (Table 2).

### 4.2 Implementation of SARIMA

The initial step involves identifying the SARIMA model with suitable values for p, d, q, P, D, and Q. In time series modeling, it is uncommon for a statistical model to perfectly capture the data-generating process; some information is inevitably lost. The objective, however, is to minimize this loss of information as much as possible. Akaike's Information Criterion (AIC) quantifies the amount of information lost by a given model. The AIC is mathematically represented by Eq. (3):

$$AIC = -2 * ln(likelihood) + 2 * (p + P + q + Q) + d + D)$$
(3)

The AIC penalizes models with more parameters by adding a term 2(p+P+q+Q+d+D), which encourages simpler models while ensuring a good fit to the data. The final step in the modelling process is to assess the residuals to confirm that the SARIMA model provides the best fit.

Based on the data presented in Table 3, the SARIMA model (3, 2, 0) (1, 0, 1) has the lowest AIC value. Therefore, it is the optimal model, offering the best fit for forecasting tourism receipts in Singapore.

The analysis of the residuals from the SARIMA model (Figure 2) provides valuable insights into the model's adequacy and the behavior of the residuals. It is evident from this analysis that the residuals show minimal temporal dependence, as indicated by the near-zero autocorrelation values across various lags. This suggests that the residuals are not exhibiting significant patterns of correlation over time, which is a positive indication that the model has captured the underlying dynamics of the data well.

Table 2. Data partitioning by period and rationale

	Period	Why?	Percentage (Train/Test)
	January	In 23 January 2020,	
(a)	2008 -	Singapore reported its	760/ / 240/
(a)	January	first coronavirus case	1070 / 2470
	2020	[34].	
		In 24 March 2020,	
	January	Singapore implemented	
(b)	2008 -	suspension of arrivals	78% / 22%
	March 2020	and travel restrictions on	
		March 23, 2020 [34].	
		In September 2021,	
	January	Singapore had gradually	
	2008 -	eased travel restrictions	070/ / 120/
(c)	September	and set up "air travel	8/% / 13%
	2021	corridors" with certain	
		low-risk countries [34].	

 Table 3. Akaike information criterion (AIC) for different

 SARIMA models

SARIMA Model	AIC
(3, 2, 0) (1, 0, 1)	4,295.35
(1, 0, 1) (1, 1, 1)	4,296.87
(0, 1, 0) (1, 0, 1)	4,296.99
(2, 0, 0) (1, 0, 1)	4,297.67
(2, 2, 0) (0, 0, 0)	4,342.41
(3, 3, 0) (0, 0, 0)	4,344.53

Additionally, the residuals fluctuate within a relatively narrow range, approximately between -150,000 and 150,000, which is noteworthy given that the original dataset's values range up to nearly 2,000,000. The limited variability in the residuals' magnitudes, relative to the data's scale, suggests that the model has effectively captured a significant portion of the variability in the observed data.

These findings indicate that the SARIMA model has effectively accounted for temporal dependencies in the data, resulting in residuals that are largely uncorrelated and within a reasonable magnitude range compared to the original dataset. This reinforces the adequacy of the SARIMA model for the given time series data.



**Figure 2.** Residuals from SARIM (3,2,0)(1,0,1)

 
 Table 4. Evaluating SARIMA model performance with different data decompositions

Data Decomposition	RMSE	<b>R</b> <sup>2</sup>
Data decomposition 1	177,861.98	0.89
Data decomposition 2	128,057.66	0.92
Data decomposition 3	125,798.53	0.92

Drawing on the findings presented in Table 4, it is clear that the adapted SARIMA model utilizing the second and third data compositions performed exceptionally well in predicting tourist arrivals, surpassing the performance of the first data composition. This outcome reaffirms the reliability of SARIMA as a robust model for time series forecasting. Specifically, the models trained on the second and third training sets achieved the highest  $R^2 = 0.92$ , demonstrating a close match between forecasted and actual values.

These results highlight the importance of training models with ample and diverse data, particularly data that accounts for unexpected events like the COVID-19 pandemic, which enhances model accuracy. As emphasized in the previous study [35], researchers should focus on developing predictive models capable of handling unforeseen circumstances. In conclusion, the SARIMA model trained on the dataset that includes the COVID-19 pandemic period emerges as the most effective predictor, based on the results presented.

### 4.3 Implementation of RF

In this part, we present an implementation utilizing the Random Forest machine learning model for making predictions on a tourist time series dataset. The dataset was preprocessed, and the Min-Max Scaler was applied for normalization. The dataset was then divided into training and testing sets as showed in Figure 2. The dataset was then divided into training and testing sets, with appropriate data formatting for training using a specified look-back value. Subsequently, the Random Forest model was constructed and trained with 100 estimators.

 Table 5. Evaluating RF model performance with different data decompositions

Data Decomposition	RMSE	R <sup>2</sup>
Data decomposition 1	588,680.35	-0.27
Data decomposition 2	597,280.84	-0.67
Data decomposition 3	141,197.12	0.90

Considering the results presented in Table 5, the first data decomposition, accounting for 76% of the dataset, exhibited a higher RMSE and a negative R2 of -0.27. This suggests that the model faced challenges in accurately predicting tourist arrivals during this phase. The negative R-squared value indicates that the model did not fit the data well in this decomposition, due to the abrupt changes and unpredictability brought about by the onset of the pandemic. In the second data decomposition, covering 78% of the dataset, the RMSE remained high, indicating notable prediction errors. The R2 value further decreased to -0.67, signifying a weaker fit of the model to the data. This decline in predictive performance attributed to the lingering impact of the pandemic, with tourism still struggling to recover fully. However, in the third data decomposition, encompassing 87% of the dataset, the model showcased significantly improved performance. The RMSE notably decreased, indicating more accurate predictions compared to the earlier decompositions. The R2 value of 0.90 revealed a strong fit of the model to the data. This indicates that the model captured the underlying patterns and trends effectively during a phase of recovery and stabilization in the tourism industry, possibly after the pandemic's peak impact. Overall, these results underscore the importance of considering the temporal dynamics of tourism data, especially during exceptional events like the COVID-19 pandemic. The variations in predictive performance across different data decomposition phases highlight the necessity of adapting forecasting models to changing patterns and behaviors in the tourism sector. As tourism evolves over time, models need to be recalibrated and refined to accommodate these changes effectively, ensuring accurate predictions and informed decision-making for stakeholders in the tourism industry.

### 4.4 Implementation of LSTM

In the third part of our analysis, we utilized the LSTM (Long Short-Term Memory) model, a type of recurrent neural network known for its effectiveness in capturing long-term dependencies in sequential data [32]. The LSTM architecture consisted of a key LSTM layer with 4 units, followed by a Dense layer with a single unit. This LSTM layer played a crucial role in enabling the model to understand and recognize patterns over different time intervals [36]. The choice of using 4 units was determined through empirical experimentation, and this parameter can be further adjusted based on the complexity and characteristics of the dataset.

To prepare the data for the model, we applied a scaling technique known as Min-Max scaling. This method transformed the original data values (X) into a normalized range of [0, 1], using the minimum and maximum values of the variable. This step is fundamental for enhancing the model's convergence. The formula for Min-Max Scaling is given by Eq. (4):

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$
(4)

where,

X: is the original value you want to scale;

X<sub>min</sub>: is the minimum value of the variable;

 $X_{max}$ : is the maximum value of the variable;

 $X_{\text{scaled}}$ : is the scaled value in the range [0, 1].

The dataset was carefully divided into separate training and testing sets, as shown in Table 2. These sets were then converted into sequential time series, where each data point was linked to its preceding value. This transformation was essential, as it allowed the model to capture temporal patterns inherent in the data.

The LSTM model, with its custom architecture and carefully chosen parameters, was designed to effectively understand the intricate temporal dependencies within the data. The model was trained over 40 epochs with a batch size of 1. For optimization, we used the Mean Squared Error (MSE) loss function and the Adam optimizer. After training, the model was tested on both the training and testing datasets. Table 6 presents a comparative overview of the LSTM model's performance, showcasing the RMSE (Root Mean Squared Error) and R<sup>2</sup> values for each data decomposition. The results reveal varying levels of predictive accuracy based on the data decomposition, illustrating the model's ability to capture subtle patterns in tourist arrivals.

 Table 6. Evaluating LSTM model performance with different data decompositions

Data Decomposition	RMSE	R <sup>2</sup>
Data decomposition 1	228,903.83	0.79
Data decomposition 2	163,087.67	0.87
Data decomposition 3	89,125.37	0.96

From the results presented in Table 6, the LSTM model's performance during the pandemic years is especially notable. The third data decomposition demonstrated outstanding accuracy, with the lowest RMSE of 89,125.37 and a high  $R^2$  value of 0.96. This indicates that the LSTM model was able to effectively capture the intricate patterns in the data, even

during the tumultuous period marked by the pandemic.

### 4.5 Discussion

Figures 3-5 provide a comparative overview of the predicted trajectories for tourist arrivals generated by the SARIMA, RF, and LSTM models, in contrast with the actual value trajectory. The analysis of the SARIMA, RF, and LSTM models offers a nuanced understanding of how these forecasting approaches handle temporal patterns, seasonality, and disruptions such as the COVID-19 pandemic. Each model has its own strengths and weaknesses, and their performances provide valuable insights for the tourism industry in Singapore.



Figure 3. Actual tourist arrivals and predictions of all models for the first data decomposition



Figure 4. Actual tourist arrivals and predictions of all models for the second data decomposition



Figure 5. Actual tourist arrivals and predictions of all models for the third data decomposition

Each trajectory corresponds to a specific data decomposition representing different phases of the dataset. The SARIMA model, representing the econometric approach, demonstrates notable accuracy across all data decompositions, effectively capturing the underlying patterns based on statistical analysis. SARIMA stands out for its ability to capture seasonal trends and temporal dependencies effectively. The model's low RMSE and high R<sup>2</sup> values across different data decompositions indicate that it successfully models the inherent seasonality and regular patterns in tourist arrivals. The ACF and PACF analysis of the residuals supports this, showing that the residuals are largely uncorrelated, suggesting a good fit. However, SARIMA's performance during the pandemic highlights a limitation: while it can handle regular seasonal variations, it may struggle with abrupt, nonlinear changes. The model's inability to fully account for the sudden shifts in tourist behavior due to the pandemic demonstrates its challenge in adapting to such extraordinary events. This suggests that while SARIMA is robust for general forecasting, it may require additional adjustments or complementary models to manage unpredictable disruptions effectively.

The RF model, representing a machine learning approach, showcases varying levels of accuracy across different data decompositions. It struggles to predict accurately during the initial phases, particularly in the first data decomposition, due to the abrupt changes and uncertainties introduced by the pandemic. However, it showcases improved performance in the later phases, aligning with the stabilization of the tourism industry. The RF model's ability to capture complex nonlinear relationships makes it a valuable tool in understanding tourism trends. However, its performance varied significantly across different data decompositions. The high RMSE and negative R<sup>2</sup> values during the pandemic years indicate that RF struggled with the sudden and unpredictable changes in tourist patterns, likely due to insufficient adaptation to the new data dynamics introduced by the pandemic. In contrast, the improved performance in the post-pandemic period reflects RF's strength in modeling complex patterns once the data stabilizes. This underscores RF's capability to handle intricate data structures and interactions but also its sensitivity to the quality and continuity of data.

Remarkably, the LSTM model, representing deep learning approaches, consistently performs well across all data decompositions, indicating its ability to comprehend and forecast tourist arrivals even during challenging times. Specifically, it excels in the third data decomposition, showcasing a high level of accuracy and effectively capturing the patterns during the recovery phase post-pandemic. The LSTM's capacity to model long-term dependencies and intricate patterns in the data makes it a powerful tool for time series forecasting, particularly in domains with evolving trends and uncertainties, such as the tourism industry. These findings underscore the resilience and adaptability of the LSTM model, alongside the complementary strengths of both econometric and machine learning approaches, in capturing and predicting tourist arrivals. This makes them promising tools for accurate and insightful forecasting in the volatile tourism sector. The LSTM excels across all data decompositions, demonstrating its robust performance even during the turbulent pandemic period. The low RMSE and high R<sup>2</sup> values, particularly in the post-pandemic recovery phase, illustrate LSTM's effectiveness in capturing long-term dependencies and adapting to evolving patterns. LSTM's ability to model intricate temporal patterns and learn from sequential data makes it particularly suited for forecasting in volatile environments. This adaptability is crucial for managing the dynamic nature of tourism data, especially in the wake of unprecedented disruptions like the COVID-19 pandemic.

The implications for the tourism industry in Singapore, based on the performance of SARIMA, Random Forest (RF), and Long Short-Term Memory (LSTM) models, highlight the need for a nuanced approach to forecasting and planning. The SARIMA model proves effective in capturing seasonal trends and regular patterns in tourist arrivals, making it a valuable tool for understanding recurring seasonal variations. However, its performance during the pandemic illustrates its limitations in adapting to sudden, nonlinear disruptions. On the other hand, LSTM's ability to handle complex, dynamic changes makes it a superior choice for forecasting during volatile periods. Its consistent accuracy across different phases, including the postpandemic recovery, underscores its effectiveness in managing evolving patterns and abrupt changes in tourist behavior.

In practice, integrating the strengths of both SARIMA and LSTM can offer a more robust forecasting framework. While SARIMA's seasonal modeling capabilities provide a solid foundation for understanding long-term trends, LSTM can address the immediate and unpredictable shifts in data patterns, enhancing the model's responsiveness to extraordinary events. This combined approach can support more informed decisionmaking, enabling better resource allocation and strategic planning in the tourism sector.

Adapting to future disruptions will require models that not only account for routine fluctuations but also offer resilience against sudden changes. The LSTM model's demonstrated capacity to handle abrupt disruptions, such as those caused by the COVID-19 pandemic, highlights its value in preparing for and managing future uncertainties. By leveraging the complementary strengths of different forecasting approaches, the tourism industry can develop a more comprehensive and flexible strategy, ensuring better preparedness and adaptability in the face of both regular and exceptional challenges.

Overall, the comparative analysis of SARIMA, RF, and LSTM models emphasizes the importance of a multifaceted forecasting approach. Utilizing these insights allows for more effective tourism management, particularly as the industry continues to recover and adapt to new challenges.

### **5. CONCLUSIONS**

In this study, we conducted a comprehensive analysis of time series forecasting techniques to predict tourist arrivals, a critical aspect of the tourism industry's dynamics. The study evaluated three key approaches: SARIMA, representing econometric modeling; RF, epitomizing machine learning; and LSTM, embodying deep learning. Each model was assessed across distinct data decompositions that reflected different phases of the dataset, including the challenging period of the COVID-19 pandemic.

The SARIMA model, grounded in econometric principles, demonstrated significant accuracy in capturing temporal patterns during pre-pandemic phases. It excelled in identifying seasonal trends and established patterns. However, it faced challenges in adapting to the abrupt disruptions caused by the pandemic and exhibited diminished performance during the subsequent recovery phase. This limitation highlights the need for models that can better handle sudden changes and evolving trends.

The RF model showcased resilience by gradually adapting to the dynamic patterns introduced by the pandemic. Its performance improved as the tourism industry began to stabilize, revealing enhanced accuracy and the model's ability to capture complex, nonlinear relationships. Nevertheless, RF struggled during the initial phases due to the pandemic's unpredictability, revealing its limitations in forecasting during periods of significant disruption.

In contrast, the LSTM model, representing deep learning capabilities, consistently exhibited superior predictive performance across all data decompositions. Its ability to capture intricate temporal dependencies and adapt to evolving patterns, even during the volatile pandemic period, underscores the potential of deep learning in time series forecasting. The LSTM model provided robust predictions, particularly during the recovery phase post-pandemic, demonstrating its adaptability to changing industry dynamics.

### **Key Findings and Contributions**

The study highlights the distinct strengths of the SARIMA, RF, and LSTM models. SARIMA excels in seasonal pattern recognition, RF adapts to evolving trends, and LSTM effectively handles complex, dynamic changes. Each model has its unique advantages, suggesting that combining them could enhance forecasting accuracy.

The findings emphasize the importance of model adaptability, particularly in the face of unprecedented disruptions like the COVID-19 pandemic. LSTM's performance during the pandemic and recovery phases illustrates the value of deep learning in forecasting volatile and evolving data.

### **Practical Recommendations**

For tourism forecasting, especially in the aftermath of disruptions, integrating SARIMA's seasonal modeling with LSTM's dynamic pattern recognition could provide a comprehensive forecasting approach. This combined methodology would leverage SARIMA's ability to model established trends while utilizing LSTM's capability to adapt to sudden changes and complex patterns.

### **Implications for Tourism Policy and Planning**

The findings suggest that while SARIMA provides a solid foundation for capturing seasonal trends, LSTM offers superior performance in handling dynamic and unpredictable changes. For comprehensive forecasting, integrating LSTM's capabilities with SARIMA's seasonal modeling could provide a more robust approach, especially in the context of recovering from disruptions. Decision-makers should consider incorporating models that offer both robustness and flexibility to develop resilient forecasting strategies.

Accurate forecasting is critical for strategic planning in the tourism sector. LSTM's consistent performance across various phases, including recovery periods, indicates that it can support more informed decision-making, enabling better resource allocation and strategy development. SARIMA's strengths in seasonal trend modeling should be leveraged for long-term planning, while LSTM can offer insights into more immediate and volatile changes.

### Adaptation to Future Disruptions

The ability of LSTM to handle abrupt changes and evolving trends underscores its potential value in preparing for future disruptions. For effective tourism management, it is crucial to adopt models that can quickly adapt to sudden shifts in data patterns, ensuring resilience and flexibility in forecasting. Combining the strengths of different models can provide a more comprehensive forecasting framework, enhancing the tourism industry's ability to navigate both routine fluctuations and extraordinary events.

In summary, our comparative analysis illuminates the complementary strengths of diverse forecasting methodologies. The econometric SARIMA model excels in understanding pre-pandemic patterns, machine learning RF adapts to changing trends, and deep learning LSTM excels in capturing complex, evolving patterns. The choice of modeling technique should be aligned with the specific needs of the analysis, considering the dynamics of the tourism industry and its reaction to notable disruptions. This study recognizes the

value of flexible models in a constantly changing environment, opening the door for improved forecasting accuracy and wellinformed decision-making in the tourism sector. By leveraging the complementary strengths of SARIMA, RF, and LSTM, stakeholders can enhance forecasting accuracy and develop more resilient and adaptive tourism strategies, better equipped to navigate both predictable trends and unexpected disruptions.

### REFERENCES

- [1] Albeladi, K., Zafar, B., Mueen, A. (2023). Time series forecasting using LSTM and ARIMA. International Journal of Advanced Computer Science and Applications, 14(1): 313-320. https://doi.org/10.14569/IJACSA.2023.0140133
- [2] Sen, J., Chaudhuri, T.D. (2016). An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice: A comparative study of the Indian consumer durable and small cap sectors. arXiv preprint arXiv:1605.03930. https://doi.org/10.48550/arXiv.1605.03930
- [3] Sen, J., Chaudhuri, T.D. (2017). A time series analysisbased forecasting framework for the Indian healthcare sector. arXiv preprint arXiv:1705.01144. https://doi.org/10.48550/arXiv.1705.01144
- [4] Abdou, M., Musabanganji, E., Musahara, H. (2021). Tourism demand modelling and forecasting: A review literature. African Journal of Hospitality, Tourism and Leisure, 10: 1370-1393. https://doi.org/10.46222/ajhtl.19770720.168.
- [5] Yang, Y., Wong, K.K. (2012). A spatial econometric approach to model spillover effects in tourism flows. Journal of Travel Research, 51(6): 768-778. https://doi.org/10.1177/0047287512437855.
- [6] De Jesus, N.M., Samonte, B.R. (2023). 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(3): 816-823. https://doi.org/10.14569/IJACSA.2023.0140393
- [7] Amine, A.M., Khamlichi, Y.I. (2024). Optimization of intrusion detection with deep learning: A study based on the KDD Cup 99 database. International Journal of Safety and Security Engineering, 14(4): 1029-1038. https://doi.org/10.18280/ijsse.140402
- [8] Abdualgalil, B. (2020). Tourist prediction using machine learning algorithms. 3rd International Conference on Sustainable Globalization (2020), pp. 111-121.
- [9] Song, H., Qiu, R.T., Park, J. (2019). A review of research on tourism demand forecasting: Launching the annals of tourism research curated collection on tourism demand forecasting. Annals of Tourism Research, 75: 338-362. https://doi.org/10.1016/j.annals.2018.12.001
- [10] Li, Y., Cao, H. (2018). Prediction for tourism flow based on LSTM neural network. Procedia Computer Science, 129: 277-283.

https://doi.org/10.1016/j.procs.2018.03.076

- [11] Hassani, H., Webster, A., Silva, E.S., Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis. Tourism Management, 46: 322-335. https://doi.org/10.1016/j.tourman.2014.07.004
- [12] Makridakis, S., Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. Journal of Forecasting,

16(3): 147-163. https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

- [13] Bouhaddour, S., Saadi, C., Bouabdallaoui, I., Guerouate, F., Sbihi, M. (2023). Tourism in Singapore, prediction model using SARIMA and PROPHET. In AIP Conference Proceedings, 2508(1): 0131288. https://doi.org/10.1063/5.0131288
- [14] He, K., Ji, L., Wu, C.W.D., Tso, K.F.G. (2021). Using SARIMA–CNN–LSTM approach to forecast daily tourism demand. Journal of Hospitality and Tourism Management, 49: 25-33. https://doi.org/10.1016/j.jhtm.2021.08.022
- [15] Tovmasyan, G. (2021). Forecasting the number of incoming tourists using ARIMA model: Case study from Armenia. Marketing I Menedžment Innovacij, (3): 139-148. https://doi.org/10.21272/mmi.2021.3-12
- [16] Makoni, T., Mazuruse, G., Nyagadza, B. (2023). International tourist arrivals modelling and forecasting: A case of Zimbabwe. Sustainable Technology and Entrepreneurship, 2(1): 100027. https://doi.org/10.1016/j.stae.2022.100027
- [17] Qiu, R.T., Wu, D.C., Dropsy, V., Petit, S., Pratt, S., Ohe, Y. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team. Annals of Tourism Research, 88: 103155. https://doi.org/10.1016/j.annals.2021.103155
- [18] Wu, D.C. W., Ji, L., He, K., Tso, K.F.G. (2021). Forecasting tourist daily arrivals with a hybrid Sarima– LSTM approach. Journal of Hospitality & Tourism Research, 45(1): 52-67. https://doi.org/10.1177/1096348020934046
- [19] Bouhaddour, S., Saadi, C., Bouabdallaoui, I., Sbihi, M., Guerouate, F. (2023). A novel hybrid approach for daily tourism arrival forecasting: The PROPHET-Bayesian gaussian process-forward neural network model. Ingénierie des Systèmes d'Information, 28(4): 833-842. https://doi.org/10.18280/isi.280404
- [20] Singapore Tourism Analytics Network (Stan). https://stan.stb.gov.sg/content/stan/en/home.html.
- [21] Li, S., Chen, T., Wang, L., Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. Tourism Management, 68: 116-126. https://doi.org/10.1016/j.tourman.2018.03.006
- [22] Samal, K.K.R., Babu, K.S., Das, S.K., Acharaya, A. (2019). Time series based air pollution forecasting using SARIMA and prophet model. In proceedings of the 2019 international conference on information technology and computer communications, Singapore, pp. 80-85. https://doi.org/10.1145/3355402.3355417
- [23] Lu, W., Li, J., Li, Y., Sun, A., Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. Complexity, 2020(1): 6622927. https://doi.org/10.1155/2020/6622927
- [24] Singh, B., Henge, S.K., Mandal, S.K., Yadav, M.K., Yadav, P.T., Upadhyay, A., Iyer, S., Gupta, R.A. (2023). Auto-regressive integrated moving average threshold influence techniques for stock data analysis. International Journal of Advanced Computer Science and Applications, 14(6): 446-455. https://doi.org/10.14569/IJACSA.2023.0140648
- [25] Asghar, M.Z., Rahman, F., Kundi, F.M., Ahmad, S. (2019). Development of stock market trend prediction system using multiple regression. Computational and

Mathematical Organization Theory, 25: 271-301. https://doi.org/10.1007/s10588-019-09292-7

- [26] Andariesta, D.T., Wasesa, M. (2022). Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: A multisource Internet data approach. Journal of Tourism Futures, pp. 1-17. https://doi.org/10.1108/JTF-10-2021-0239
- [27] Feng, Y., Li, G., Sun, X., Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. Procedia Computer Science, 162: 628-633. https://doi.org/10.1016/j.procs.2019.12.032
- [28] Hrich, N., Azekri, M., Khaldi, M. (2024). Application of LSTM for redundancy detection in MCTS: Enhancing test precision. Ingénierie des Systèmes d'Information, 29(4): 1573-1579. https://doi.org/10.18280/isi.290430
- [29] Lu, W., Rui, H., Liang, C., Jiang, L., Zhao, S., Li, K. (2020). A method based on GA-CNN-LSTM for daily tourist flow prediction at scenic spots. Entropy, 22(3): 261. https://doi.org/10.3390/e22030261
- [30] Dou, R., Meng, F. (2022). Prophet-LSTM combination model carbon trading price prediction research. Second International Symposium on Computer Technology and Information Science (ISCTIS 2022), 12474: 477-487. https://doi.org/10.1117/12.2653785

- [31] Carmel, V., Akila, D. (2020). A survey on biometric authentication systems in cloud to combat identity theft. Journal of Critical Reviews, 7(3): 540-547. https://doi.org/10.31838/jcr.07.03.97
- [32] Salamanis, A., Xanthopoulou, G., Kehagias, D., Tzovaras, D. (2022). LSTM-based deep learning models for long-term tourism demand forecasting. Electronics, 11(22): 3681. https://doi.org/10.3390/electronics11223681

[33] Bonab, A.A.G. (2022). A comparative study of demand forecasting based on machine learning methods with time series approach. Journal of Applied Research on Industrial Engineering, 9(3): 331-353. https://doi.org/10.22105/jarie.2021.246283.1192

- [34] "World Health Organization." https://www.who.int.
- [35] Adebiyi, A.A., Adewumi, A.O., Ayo, C.K. (2014).
   Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics, 2014(1): 614342.
   https://doi.org/10.1155/2014/614342
- [36] Son, K., Byun, Y., Lee, S. (2018). Prediction of visitors using machine learning. In 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Bangkok, Thailand, pp. 138-139. https://doi.org/10.1109/ICIIBMS.2018.8549960