




## Article

# Attention-Based STL-BiLSTM Network to Forecast Tourist Arrival

Mohd Adil <sup>1</sup>, Jei-Zheng Wu <sup>2,\*</sup>, Ripon K. Chakraborty <sup>3</sup>, Ahmad Alahmadi <sup>4</sup>, Mohd Faizan Ansari <sup>5</sup> and Michael J. Ryan <sup>6</sup>

<sup>1</sup> Department of Management Studies, NIT Hamirpur, Hamirpur 177005, India; adil.dms@nith.ac.in

<sup>2</sup> Department of Business Administration, Soochow University, Taipei 100, Taiwan

<sup>3</sup> Capability Systems Centre, School of Engineering & IT, UNSW Canberra at ADFA, Canberra, ACT 2610, Australia; r.chakraborty@adfa.edu.au

<sup>4</sup> Department of Electrical Engineering, College of Engineering, Taif University, Taif 21944, Saudi Arabia; aziz@tu.edu.sa

<sup>5</sup> Department of Applied Informatics, Silesian University of Technology, 44-100 Gliwice, Poland; mfansari2395@gmail.com

<sup>6</sup> Capability Associates, Canberra, ACT 2610, Australia; mike.ryan@ieee.org

\* Correspondence: jzwu@scu.edu.tw

**Abstract:** Tourism makes a significant contribution to the economy of almost every country, so accurate demand forecasting can help in better planning for the government and a range of stakeholders involved in the tourism industry and can aid economic sustainability. Machine learning models, and in particular, deep neural networks, can perform better than traditional forecasting models which depend mainly on past observations (e.g., past data) to forecast future tourist arrivals. However, search intensities indices (SII) indicators have recently been included as a forecasting model, which significantly enhances forecasting accuracy. In this study, we propose a bidirectional long short-term memory (BiLSTM) neural network to forecast the arrival of tourists along with SII indicators. The proposed BiLSTM network can remember information from left to right and right to left, which further adds more context for forecasting in memory as compared to a simple long short-term memory (LSTM) network that can remember information only from left to right. A seasonal and trend decomposition using the Loess (STL) approach is utilized to decompose time series tourist arrival data suggested by previous studies. The resultant approach, called STL-BiLSTM, decomposes time series into trend, seasonality, and residual. The trend provides the general direction of the overall data. Seasonality is a regular and predictable pattern which re-occurs at fixed time intervals, and residual is a random fluctuation that is something which cannot be forecast. The proposed BiLSTM network achieves better accuracy than the other methods considered under the current study.

**Keywords:** LSTM; BiLSTM; SII index; tourist arrival; forecasting; machine learning



**Citation:** Adil, M.; Wu, J.-Z.; Chakraborty, R.K.; Alahmadi, A.; Ansari, M.F.; Ryan, M.J.

Attention-Based STL-BiLSTM Network to Forecast Tourist Arrival. *Processes* **2021**, *9*, 1759.

<https://doi.org/10.3390/pr9101759>

Academic Editor: Chien-Chih Wang

Received: 4 September 2021

Accepted: 26 September 2021

Published: 30 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Tourism makes a significant contribution to the economies of many countries. During 2019, tourism contributed \$2750.07 billion, about 3.2%, to the global economy [1]. Stakeholders in the tourism industry—such as professionals, managers, government agencies, and transporters—require accurate data related to tourist arrivals and their demands to develop and maintain infrastructure [2], enhance services, and provide better experiences for tourists. Since such data is required in advance to ensure infrastructure and services are in place to meet demand, the forecasting of tourist arrivals is, therefore, a prime concern among practitioners and academics.

In tourism, forecasting methods can be broadly categorized into qualitative and quantitative. Qualitative methods largely depend on insights and past experiences of tourists [3]. Forecasting models are quantitative and are trained on previous data to

forecast future trends. In the case of tourism, the set of past data includes tourist arrival times, rates, and volumes, and other key factors—such as weather conditions, public transportation systems, and availability of the infrastructure—that help forecast future tourist arrivals and demands. With the advent of information technology, tourists make use of search engines to collect information such as the availability of hotels, popular places to visit, and weather conditions at the destination location [4]. Consequently, in recent years, search intensity indices (SII) data have been extensively used in tourist forecasting [5] to augment tourist arrival data, which has subsequently improved the performance of forecasting models [6].

Machine learning techniques such as “support vector regressor” (SVR) have shown significant performance improvement in forecasting data compared with the more traditional models such as “Autoregressive Integrated Moving Average” (ARIMA) [5]. However, many machine learning methods do not work without human intervention, relying on feature selection that requires human experts to select the key features that contribute to better forecasting performance. Since real-world datasets contain a large number of features, the manual selection of key features is a very tedious and time-consuming task. “Artificial neural network (ANN) is a machine learning method inspired by the human brain” [7], which does not require manual feature selection but is a more robust model for forecasting since it can automatically select features [6]. ANN has also been widely used for time series forecasting problems and has shown better results than other machine learning models. However, ANN suffers from an inability to remember long dependencies in sequences, affecting its forecasting performance [6].

Deep convolutional neural networks (DCNN) are an extension of ANN that has shown great success in image classification, object detection, and computer vision tasks [8]. The efficiency of DCNN is mainly due to its automatic feature extracting capability [9]. For sequence-based problems, particularly in natural language processing, other kinds of deep networks known as the recurrent neural network (RNN), long-short-term-memory (LSTM), and gated recurrent unit (GRU) have shown huge success because of their long-term learning dependencies. LSTM is still imperative in demand forecasting in varied domains such as tourist arrival forecasting “order demand based on short lead time” [6]. In tourism demand forecasting, Law et al. [6] suggested a deep learning model with an attention mechanism and highlighted the efficiency of the LSTM network in forecasting Macau tourism demand [6].

However, the main limitation of LSTM is that it remembers information in only one direction, i.e., “from left to right”. In [10], authors used the ARIMA model to forecast German tourist arrival in Croatia, and in [11] researchers used an econometric model to forecast Chinese tourist arrival. Law [12] used a neural network for tourism forecasting and Chang and Tsai [13] used a deep neural network to forecast the number of tourists. Similarly, authors in [14,15] have used a unidirectional LSTM network to forecast tourism arrival. However, all the previous studies were based on deep neural networks and used a unidirectional LSTM network that remembers only in one direction. In forecasting problems, accuracy can be considerably improved if the network can also accommodate information from both directions (i.e., from forward to backward and backward to forward) [16]. To reduce this gap in the tourism forecasting literature, we present, for the first time, an advanced version of the LSTM network—a bidirectional LSTM (BiLSTM) network—to forecast tourist arrivals. The bidirectional network trains the same input twice, once in the forward direction and once in the backward direction [17], which provides additional context information for the network, thereby improving its forecasting performance.

The effectiveness of the BiLSTM network has been examined in other forecasting problems. For example, the authors in [16] used a BiLSTM network for stock forecasting and showed that BiLSTM performs better than unidirectional networks. In [18] researcher made a comparative analysis of ARIMA, LSTM, and BiLSTM for financial time series and showed that BiLSTM outperforms LSTM and ARIMA. In [19] authors also demonstrate the effectiveness of BiLSTM network over LSTM network in stock forecasting. To display the

efficacy of BiLSTM network in tourism forecasting, we investigate the BiLSTM recurrent neural network in which the hidden layer is added in the reverse direction. This extra layer reads the same input from a backward direction to address the limitation of learning tendency only based on the immediately preceding pattern of the recurrent neural network to outperform the unidirectional LSTM network.

In sequence modeling, the attention mechanism has achieved tremendous success. The attention mechanism generates weights for the parts of the sequences that should receive more attentions and assigns an attention score to each element in the sequence. The attention mechanism allows the model to be more decipherable and enable the model to ignore irrelevant information while forecasting. In tourism forecasting, the attention mechanism is useful since it selects and gives more weight to those important factors in tourism. The proposed BiLSTM network with attention learns the “temporal relationship” among various factors and the relative significance of the factors in accordance with their influence on tourism demand.

The current study aims to examine BiLSTM deep neural network and test its efficacy in tourist demand forecasting. Along with the BiLSTM network, this study also incorporates the decomposition method, which has received significant attention from tourism researchers who have accommodated decomposition methods with DNN to improve accuracy [15]. Decomposition methods split the data into sub-datasets, which reduces the model complexity without adding more data. In line with [15], we used STL decomposition, which divides data into three sub-series: seasonality, trend, and residual. This component gives additional stationary data for forecasting, which further improves the overall accuracy. The goal of this study is to present a forecasting performance of BiLSTM as compared to the unidirectional LSTM network. This study used the monthly tourism arrival data obtained from Zhang et al. [15], which they achieved by extracting many SII intensities from search engines such as Google trends. One of this study’s significant contributions is to present the BiLSTM network in tourism forecasting along with attention mechanism and STL decomposition technique suggested by previous studies to further enhance the forecasting performance for tourism arrival.

The remainder of the paper is organized as follows: Section 2 presents the prior work in tourism demand forecasting, followed by Section 3 explaining the methodology behind the proposed approach. Section 4 highlights the empirical findings and the performance. Lastly, the study culminates with Section 5, which covers implications, the conclusion, and some observations on limitations and potential future work.

## 2. Related Works

Tourist arrival forecasting can be formed under a time series analysis method that can be forecast using past data of tourist arrival to forecast future trends in arrival [10]. SII data shows tourists’ intention to visit a particular country, reflecting such indicators as tourist food interest, their activity plan to visits places of interest, weather conditions, and hotel and transportation availability. Since interested potential tourists themselves provide this information, SII data are effective indicators in forecasting tourism demand and have therefore been implemented in many tourism demand forecasting models. The relationships between tourist search queries for US cities and attractiveness have been discussed in [20]. Google trend indices have been used to forecast tourism demand in Hong Kong from nine countries [21]. In [14], the authors have shown that SII indicators reveal tourist preferences and identify changes in tourist preferences over time. The effectiveness of SII data has also been demonstrated in hotel occupancy forecasting in [22].

Tourism forecasting methods can be categorized into time series, econometrics, and artificial intelligence-based methods [23]. Time-series methods, such as ARIMA and its variants, are the most widely used time-series models [14,22–24], which are used to forecast future tourist arrival based on past observations and trends [25]. Song et al. [22] proposed an ARIMA model that includes seasonal parameters to forecast future trends. Baldigara and Mamula [10] presented an ARIMA model to predict German tourist arrivals in Croatia

and Lim et al. [26] used ARIMA with an explanatory variable model for multiple time series data to forecast international arrivals. Other simple time-series models such as the Persistence model (also known as the Naïve model) and exponential smoothing have been widely used in univariate time-series modeling. In fact, these models have been used as benchmark models for performance evaluation of other models [8,20]. Most of the time, series models, including variations of ARIMA models, assume that there is a linear relationship between past observations and future observations, which means that their performances suffer when data is nonlinear.

Econometric models aim to establish the association between tourist demand and demographic variables such as income and destination markets factors such as ease of going, transportation, and government policy for regional markets [27–30]. Econometric factors decide “economic growth” in tourism demands and offer more in-depth understandings for practitioners and policymakers [23,27]. Some contemporary/traditional models, for instance the error correction model [23] and the vector autoregression model, have been widely used in tourism demand forecasting. For example, Ognjanov et al. [11] used an economy model for Chinese tourism demand. Gunter and Onder [30] employed Bayesian factor augmented vector autoregression to predict Vienna’s tourism demand. Bangwayo-Skeete and Skeete [27] used a mixed data sampling model with Google search data to forecast tourism demand and claimed an improvement in forecasting performance. Song et al. [22] used an integration of different “econometric models” to forecast tourism demand. Asaaf et al. [31] proposed a Bayesian global vector auto-regression model to forecast the South Asian market’s tourism demand. They observed that their model performed well in forecasting the demand for South Asian tourists. Comparative studies have also been conducted between time series and econometrics models and time series methods have often been more accurate than econometric models [3].

With the advances in data generation methods, artificial intelligence methods have shown great success in many domains, including tourism demand forecasting [6]. Artificial intelligence methods comprise machine learning models such as SVM and deep learning models that include ANN and deep neural networks (DNN) [32]. Although ANN and SVM have been extensively used in tourism demand forecasting since the 1990s and have shown significant forecasting capability for tourism forecasting [5,31,32], these methods suffer from the lack of memory for long-term dependencies and the ability to forecast accurately. Given those limitations, researchers have recently started exploring advanced versions of neural networks—for instance, RNN, and LSTM and its associated variants—to forecast tourism demand since these can remember long-term dependencies. DLM with attention mechanisms for tourism demand forecasting was proposed by Law et al. [6], who achieved the best accuracy in forecasting Macau tourism demand. Attention mechanism automatically examines which feature is more relevant for the forecasting in the input data. On this basis, they assigned more significant weight to those features that significantly contribute to forecasting than those that were not of much importance.

Recently, decomposition methods have received significant attention from tourism researchers who have started incorporating decomposition methods with DNN to improve accuracy [15] further. Decomposition methods split the data into sub-datasets which reduces the model complexity without adding more data. Decomposition can be generated using filters, wavelet transformation, empirical mode decomposition (EMP) and “Seasonal and Trend Decomposition using Loess” (STL). Chen and Wu [33] applied EMP on tourist arrival in Taiwan. Singular Spectral Analysis (SSA) was used by Hassani et al. [34] for US tourist arrival and claimed increased performance. Silva et al. [35] used SSA and other decomposition with the neural network to forecast tourism demand for 10 European countries, namely Germany, Greece, Spain, Italy, Cyprus, Netherland, Austria, Sweden, and the UK. Similarly, Zhang et al. [15] used STL decomposition method with “duo attention deep learning model” (DADLM) and proposed a novel framework for tourism demand forecasting as STL-DADLM. STL decomposes the data into three sub-series, namely—seasonality, trend, and residual. These components provide additional stationary data for

forecasting and further improves the overall accuracy. Overfitting problem is a general problem in deep learning models, in which a model performs better on training data, but it does not generalize well on unseen and first-hand (new) data. Forecasting quality in tourism is affected by two factors: data volume, which is limited for complex deep learning model; and inclusion of irrelevant explanatory variables such as a lot of SII indices during model building [15]. To overcome these issues, the authors in [15] used STL decomposition for a limited data. The first layer performs feature selection without lag order selection, and the second layer does the reverse. This dual attention layer processes the equal amount of data by two feature engineering activities in parallel for explanatory variables and lag order in DADLM.

This section briefly outlines the review of the foundation literature about tourist arrivals and its forecasting methods, various models to forecast tourists' arrival, and related data. The existing literature indicates that previous researchers have used many methods to forecast the influx of the tourists, such as, time series-based method (ARIMA), econometric based model (e.g., error correction model or vector autoregression model, etc.), and most popular artificial intelligence-based methods (such as SVM, ANN, or DNN). The literature in the context of DNN highlights that previous methods/models have included LSTM network for tourist arrival forecasting. Information in the LSTM network traverses once during training from left to right. However, [25] shows that the BiLSTM network spans the input twice: once from backward to forward and then forward to backward during training, which provides additional context for forecasting and reduces the error rate by 37.78% compared with the LSTM network for time series problems. A comparative study has been made for ARIMA, LSTM, and BiLSTM in financial time series forecasting that further shows the improved forecasting performance of the BiLSTM network [18]. Joo and Choi [23] also showed the forecasting performance of LSTM and BiLSTM in case of stock prediction and shows the effectiveness of BiLSTM over the simple LSTM network. Addressing this research gap in tourism research and the dearth of studies in the literature, this study explores a BiLSTM network with an attention mechanism for tourism demand forecasting, which overcomes the limitations of the previously existing models in tourism research. Our study used a BiLSTM network that can remember information from both forward and backward directions with attention mechanism along with the STL decomposition method discussed in [15] to forecast tourist arrival that further achieved better accuracy than LSTM network.

### 3. Method

This research introduces the STL-BiLSTM deep neural network that achieves high accuracy in tourist demand forecasting. This section presents a detailed explanation of the proposed network and formulation of the tourist demand forecasting model.

#### 3.1. Problem Formulation

In tourism research, time-series study uses several factors, for instance, determinants and indicators, to forecast future tourism arrival. Tourism demand forecasting uses several multivariate time series factors from past data to forecast future tourism arrival volume.

Suppose vector  $[Y_t]_{t=1}^T = (Y_1, Y_2, Y_3, \dots, Y_t)$  represents the time series values for the tourism arrival volume at  $T$  time steps and the corresponding feature vector that forecast arrival time series data  $Y^T$  is  $[X_t]_{t=1}^T = (X_1, X_2, X_3, \dots, X_t)$ .  $T$  is the total time steps, such as the number of years, number of months, or number of days in the dataset. At time steps  $T$ ,  $X_t$  represents the various multivariate tourism factors such as the SII indicator, determinants; and  $Y_t$  represents tourist arrival.

Tourism demands forecasting takes tourism arrival  $[Y_t]_{t=1}^T$  and multivariate factors in time series  $[X_t]_{t=1}^T$  as an input to forecast the future tourist arrival volume  $[Y_t]_{t=T+1}^{T+\theta}$  with function  $\Phi$  forecasting future tourist arrival for future time steps  $\theta$ .

$$[Y_t]_{t=T+1}^{T+\theta} = \Phi([X_t]_{t=1}^T, [Y_t]_{t=1}^T), \quad (1)$$



where  $\theta$  is the future time steps that will be forecast by a function  $\Phi$ , since forecasting tourist demand is a nonlinear problem, function  $\Phi$  is a nonlinear function providing the relationship between tourist arrival input factors and output tourist arrival volume.

### 3.2. STL Decomposition

Data were collected from [15] and observed to have random variations in the arrival volume series  $[Y_t]_t = 1T$ . STL plays an important role in “time series analysis” when considerable seasonality is present in the data [15]. Seasonal smoothing is used to smooth the cyclic sub-series to determine the seasonal component. Further, lowpass smoothing is used to smooth out the estimated seasonal component. In the final stage, trend smoothing is used to find an estimation of the trend component. This process is repeated several times to improve the accuracy of the estimations of the components.

At a given time, step  $t$ , tourism arrival, can be calculated as a sum of the three series as in Equation (2).

$$[Y_t]_{t=1}^T = [P_t]_{t=1}^T + [S_t]_{t=1}^T + [R_t]_{t=1}^T, \quad (2)$$

where:

Trend =  $[P_t]_t = 1T$

Seasonality =  $[S_t]_t = 1T$

Residual or Noise =  $[R_t]_t = 1T$

Total arrival Volume =  $[Y_t]_t = 1T$

After STL decomposition, the trend series represent a global trend pattern, which is the stationary series throughout the series; and the seasonality as a constant component since it repeats with the same cycle. After applying STL, the total tourist arrival volume was decomposed into trend, seasonality, and residual at given time  $t$ . Forecasting of the total tourism arrival is calculated separately for trend and residual through a deep learning model. The seasonality component was computed according to the forecasting period with the persistence method. As a result, the tourism arrival volume was forecasted as three separate simple series, which achieves high accuracy without having additional data and avoiding the overfitting problem.

### 3.3. Data Standardization

Data standardization was applied to bring all features into the same range by scaling the data so that the features have zero mean and unit variance, as in Equation (3).

$$x' = \frac{x - \mu}{\sigma}. \quad (3)$$

where  $\mu$  is the mean,  $\sigma$  is the standard deviation and  $x$  is the feature vector.

### 3.4. LSTM Deep Neural Network

Since traditional neural networks cannot remember the previous state of the input, researchers proposed RNN, which has been shown to remember long-term dependencies. However, in reality, RNN fails because during back-propagation, network weights begin to vanish or explode, making the network unstable. Hochreiter and Schmidhuber [36] proposed a LSTM network, which addresses the long-term dependencies by introducing the cell state in the network, which stores the temporal information and three gates—the forget gate, input gate, and output gate through which information flows. LSTM network encodes inputs  $[X_t]_{t=1}^T$  to the set of hidden states  $[h_t]_{t=1}^T$ . The forget gate  $f_t$ , in Equation (4) decides which information needs to be excluded, looking at the previous hidden states  $h_{t-1}$  and the current state  $x_t$  [37]. In Equation (5), the following input gate decides which information needs to be stored in the cell state. The input gate has two layers: the first comprises sigmoid layers that determine what value needs to be updated and the second layer is tanh in Equation (6), which creates a vector of candidate values  $C_t$  that can be added to the cell state [30,38]. These two layers are combined to create an update to the cell

state. The last gate is an output gate  $o_t$ , in Equation (7) that dictates which information will be output. The output gate outputs the info in two layers: the first sigmoid layer decides which part of the cell state is to be output. Equation (8) shows the updated cell, which is the sum of the previous cell state's multiplication and the forget gate, and the multiplication of the input and the current cell state [36]. Then, the cell state is passed through the tanh function (see Equation (9)) and is multiplied with the sigmoid output.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (4)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (5)$$

$$C_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \quad (6)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (7)$$

$$C_t = f_t \times C_{t-1} + i_t \times C_t, \quad (8)$$

$$h_t = o_t \times \tanh(C_t). \quad (9)$$

In Equations (4)–(9),  $\sigma$  and  $\tanh$  are the activation functions that define a neuron's output in the neural network and  $W$  and  $b$  are the network parameters [6].

### 3.5. BiLSTM Deep Neural Network

In the study, we have combined a BiLSTM network with an attention mechanism. The BiLSTM network processes input in two ways: first, it processes information from the backward to forward direction, and then it processes the same input from forward to backward. The BiLSTM approach differs from unidirectional LSTM because the network runs the same input twice, i.e., from forward to backward and backward to forward direction, which preserves the extra context information that can be very useful in tourism demand forecasting to improve the network accuracy further. Two hidden states in the network are able to preserve information from the past and future. We used the STL decomposition implemented in [15] to make a comparison with our BiLSTM network. STL decomposition divides the original time series data into three different series: trend, seasonality, and a residual component. Decomposing the data provides a useful abstract model for thinking about time series generally and for better understanding of the problem during analysis and forecasting. The network takes input as the arrival volume and SII indicators that form a multivariate problem to forecast arrival volume in the future. In Figure 1, dense is a fully connected neural network layer, and softmax is a probability function that gives the probability. The network takes the arrival volumes, and SII factors as inputs and then attention is applied on those inputs. The inputs are then applied to the Bi-LSTM network, where the outputs from the LSTM network are fed into the dense layer in vector format. Outputs obtained from the dense layer are passed to the outputs layer, ultimately resulting in the arrival volume.

### 3.6. Attention Mechanism

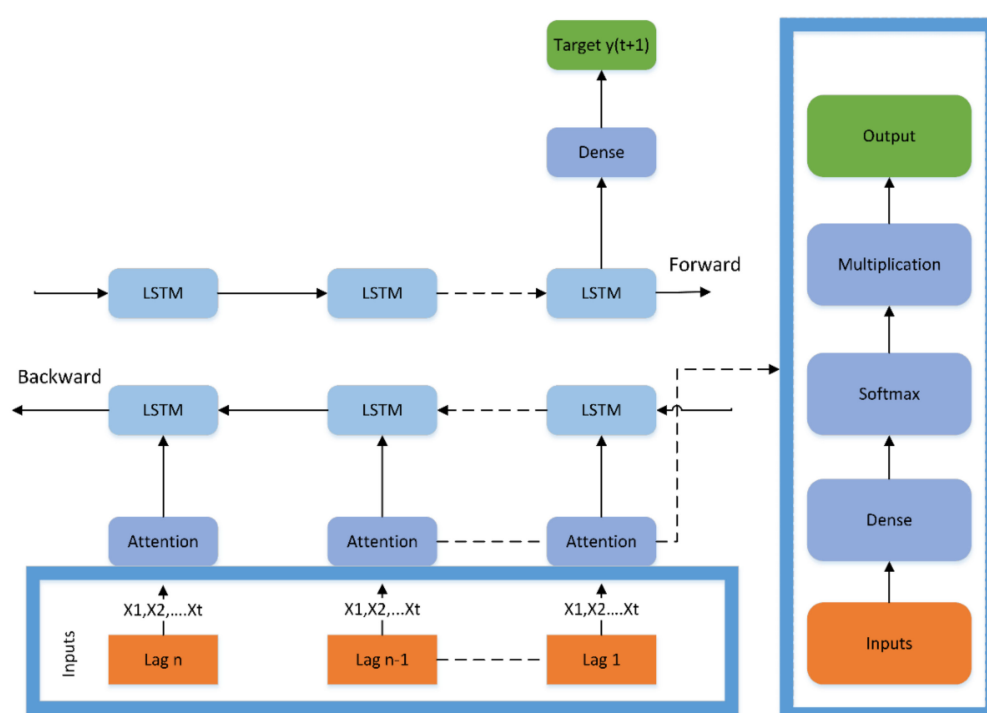
Two input attention layers are used, one for features and one for the time step dimension. The equation below shows the attention layers on input  $X^T$  [15]. In Figure 1, after input feature, attention mechanism is applied before the BiLSTM layer. This attention layer assigns more weights to those features that are important in tourism forecasting and fewer weights to those features that are not as important. After attention is applied, such features are sent to the dense layer (on the right side of Figure 1) where the network learns the relationships among different features. In the next step, the softmax function is applied, which provides the probability, and then multiplication is used with the feature vectors. The output from this layer is then sent to the BiLSTM network, where the network learns

the long-term dependencies and is then sent to the dense layer (left side in the figure) that provides the output as tourist arrival.

$$A^T = \text{softmax}(W^T \times [[X_t]_{t=1}^T]^T + b^T) \times [[X_t]_{t=1}^T]^T, \quad (10)$$

$$A^T = \text{softmax}(W^T \times [[X_t]_{t=1}^T]^T + b^T) \times [[X_t]_{t=1}^T]^T, \quad (11)$$

where input  $X^T$  contains the  $T$  time steps of the  $F$  features in a vector form. Equation (8) represents the attention in the time step dimension. The softmax function produces a  $F \times T$  dimension vector, which is a multiplication of time step  $T$  and  $n$  features vector  $F$  in the input vector. Equation (9) represents the attention to the feature vector that produces the  $T \times F$  vector after the softmax layer.  $W^T$ ,  $W^F$ ,  $b^T$ , and  $b^F$  are the parameters that are learned during the training.



**Figure 1.** BiLSTM network using attention mechanism.

Figure 1 presents a typical architecture used in this study to train the BiLSTM network. The first layer in the figure is the input layer, which takes the arrival volume and SII indicators as input in the network. In the second layer, attention mechanism is applied on the input. The attention mechanism assigns more to those input values which play pertinent part in prediction. The right side of Figure 1 exhibits the working of the attention mechanism, as in this layer input goes to the dense layer (also known as the fully connected layer) and output from this layer goes to the softmax function, which is a probability function. After that, this output is multiplied with input values, which assigns weight to input values. Further, afterwards the attention layers go to the BiLSTM layer, which processes the input from backward and forward direction to get the better representation of data as compared to the standard LSTM network. From there, data goes to a fully connected layer and finally it goes to the output layer, which gives the output.

#### 4. Empirical Study

In this empirical study, we forecast the tourist arrival in Hong Kong for the following reasons: (a) Hong Kong is an attractive destination for tourists; (b) a large amount of data is available pertaining to foreign tourist arrival; and (c) there is high seasonality in the data.



Further, tourism contributes significantly to the Hong Kong economy [39,40]. Therefore, accurate forecasting of tourist arrival is crucial for the government and other stakeholders to develop strategic policies to attract foreign tourists.

#### 4.1. Data Collection

We used the “HK2012–2018” dataset [15], which contains SII indicators collected from Google trends and tourist arrival volume collected from the Hong Kong Tourism Board (HKTb) [21] website that has records of monthly data of tourist arrival from a range of countries for up to 72 months. Following Zhang et al. [15], SII indicators were collected for six major countries: Australia, Philippines, Singapore, Thailand, United Kingdom, and United States. Although they are major sources of tourists, countries such as Japan, Taiwan, and South Korea were excluded because English search keywords were limited in Google for those countries. The authors in [15] defined several seed keywords (see Table 1) for the SII indicators in the chosen market within seven categories: recreation, shopping, lodging, tour, clothing, transportation, and dining. The authors recognized tourism-related keywords from the initial seed search keyword and collected data for each search query. They collected 96 keywords on the basis of their relevance for the chosen six countries in their study. Finally, with the help of a Python program they collected the SII data from Google trend.

**Table 1.** SII seed search keywords under the study.

Head	Keywords
Recreation	Hong Kong night clubs, Hong Kong night view, Hong Kong markets,
Shopping	Hong Kong shopping, shopping center in Hongkong,
Tourism	Hong Kong tour, Hong Kong travel, Outing in Hong Kong, Hong Kong visit
Lodging	Hotels in Hong Kong, budgeted accommodation Hong Kong
Clothing	Weather in Hong Kong, Temperature in Hong Kong
Transportation	Public transportation in Hong Kong, Hong Kong flights
Dining	Dining Hong Kong food, Hong Kong restaurant lodging

#### 4.2. STL Decomposition and Training

In the next step, STL decomposition was applied to the input data for all six chosen source markets. Trend, seasonality, and residual series were generated: Trend represents the global trend pattern, seasonality shows the constant component, and residual represents the local sensitivity for each country visiting Hong Kong. Before sending the trend and residual series to the BiLSTM network, we applied data standardization on all features to convert them to the same scale. The first row in Figure 2 depicts the general trend of tourist arrivals in Hong Kong from the United Kingdom and Australia—the y axis in the graph shows the total number of tourist arrivals, and the x-axis shows the year (the other source markets are not shown to save space). Cyclic trends can be observed in all source markets coming to Hong Kong. The three decomposed series for the United Kingdom and Australia are shown in Figure 2.

The trend series represents better stability as compared to the original arrival volume for the United Kingdom and Australia market demand in Hong Kong. The seasonality series show the constant cycle that occurs in specific periods. Local sensitivity represents the occasional and irregular events such as sudden earthquakes, coronavirus pandemic, and other events that are not known in advance, known as irregular events or residual. In Figure 2, the residual is shown for the United Kingdom and Australia. We used trend series, which represents the overall trend of tourist arrival after decomposition and local sensitivity events in the neural network training process. For the seasonality we used a persistence method. After STL, the total arrival volume is equal to the sum of trend, residual, and seasonality, as shown in Equation (3).

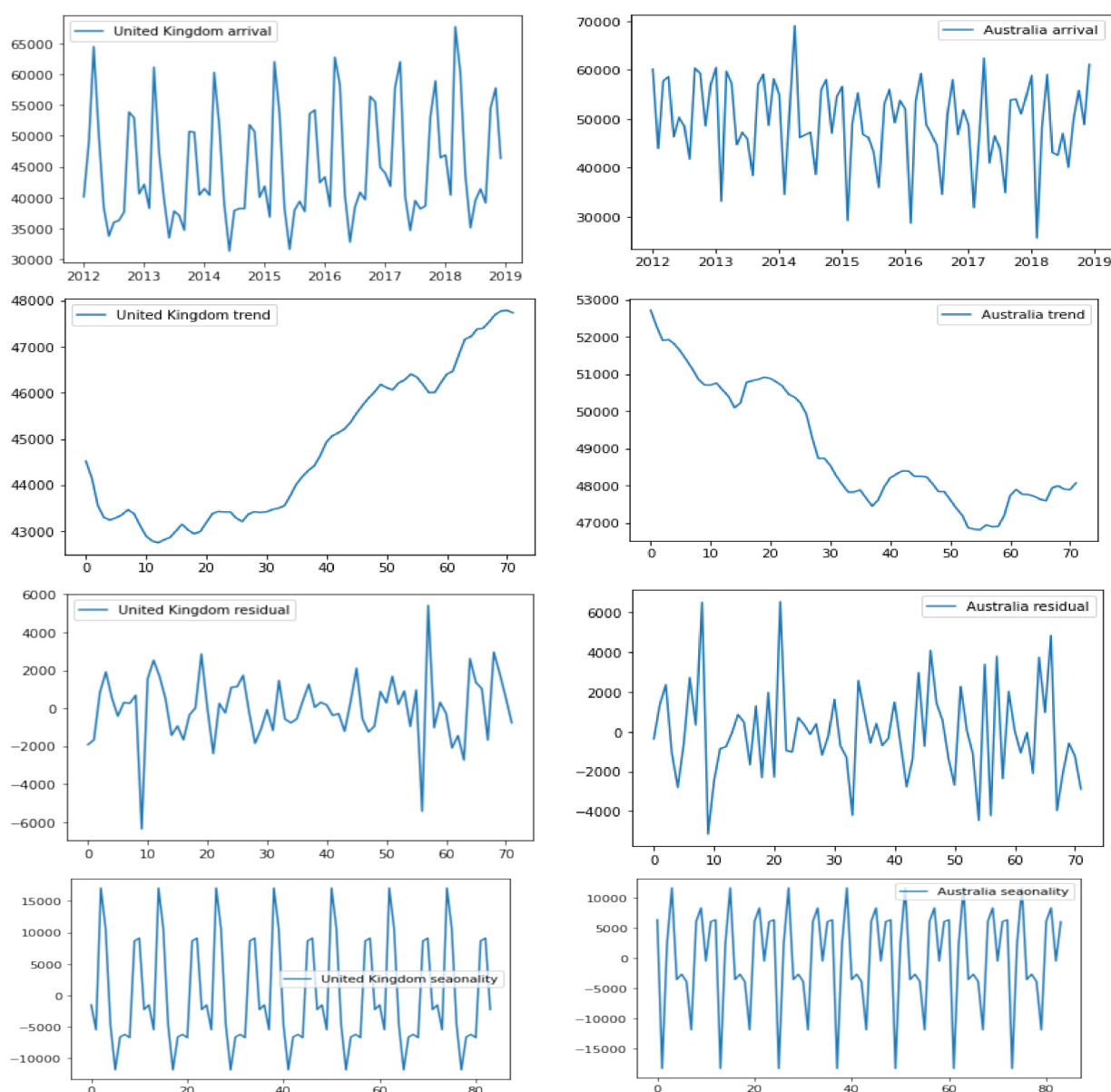


Figure 2. Trend, residual/error, and seasonality component after STL process.

#### 4.3. Performance Evaluation

To evaluate the performance of the proposed method compared with other methods for forecasting tourism demand, a walk-forward validation set up was used in this study to simulate the real-world environment. This section describes the validation steps and next, compares the one-step forecasting results for all methods. In real-world use, we would like to re-train the model again as new data is available so that the model has the opportunity to make better forecasts in each time step. We evaluate our model with the assumption that we first select the minimum number of input values to train the model, which is taken to be the window width if we use a sliding window.

Next, we have to decide whether the model will be trained on all data if it is available or trained on the latest available data. When an appropriate configuration has been chosen for the test setup, the model can be trained and evaluated in the following four steps:

- Select minimum samples in the window used to train the model;
- Make forecast for the next time step;
- Evaluate prediction;

- Increase window size to add the known values and repeat from step 1.

In the walk-forward validation setting, new arrival volume is available as input for the next month's forecast as in a real-world scenario. For this validation process, the persistence method, ARIMA method, and DADLM [15] methods (with and without STL) have been applied to serve as baseline models.

#### 4.4. Methods Investigated

In our study, following the validation setup, Persistence, ARIMA, DADLM, and STL-DADLM methods were used as baseline methods to compare with the forecasting performance of our STL-BiLSTM network. The Persistence method uses the value at the previous time step to forecast the next time step's value. ARIMA is a generalization of simple Auto-Regressive Moving Average, and DADLM is a deep learning-based method.

The ARIMA and Persistence model uses the arrival volumes for forecasting. These models take the past data to forecast the future tourist arrival volume. In contrast, DADLM and the BiLSTM network take the arrival volume and SII indicator data as their input to forecast the next month's arrival volume. DADLM and the BiLSTM network take the input in three dimensions: the first dimension is the number of training examples in the network, the second is the number of time steps, and the third dimension is the number of features, i.e., SII data and arrival volume for the corresponding months. The STL-BiLSTM network takes trend and residual series during training. For the one-step forecasting, data from 2012 to 2016 (80%) data was used as training data. The rest of the data from 2016 to 2018 (20%) data was used in step-by-step performance with walk-forward validation for all six countries to determine the six compared methods' performance. We divided all the data for training and test data using slicing operation in NumPy, first 48 months data was used as the training set and the last 12 months data was used as the test set.

#### 4.5. Sensitivity Analysis of Hyper-Parameters

The ARIMA model parameters are searched through the grid search, which gives the best ARIMA model for tourist demand forecasting. In our study, we set the range for  $p \in [0, 1, 2, 4, 6, 8, 10]$ , for  $d \in (0, 2)$  and for  $q \in (0, 2)$  as in the work of [41], where  $p$  is the number of lag observations in the model also known as lag order,  $d$  is the degree of differencing, i.e., how many times a number of observations are being differenced; and  $q$  is the moving average of observation also called the window size of the moving average. For the Singapore dataset, the best values of  $p, d, q$  for the ARIMA model are  $(0, 0, 2)$ , which were attained through grid search.

The BiLSTM network has a number of hyper-parameters such as the number of hidden layers, number of neurons in the LSTM cell, and activation. The dense layer is a neural network that also has hyper-parameters such as, number of hidden layer, number of neurons in the hidden layer, and activation function. When training a neural network, hyper-parameters include the number of epochs, batch size number, learning rate networks, and number of networks. Hyper-parameters of a neural network are important since they control the overall behavior of the training algorithm and have a notable impact on the performance of a model. Hyper-parameter tuning is a process to find the optimal combination of hyper-parameter that minimizes the loss function to give better results. To address the overfitting issue we used a dropout layer in the network. As the dropout layer automatically shuts some of the neurons in the network during training and therefore, it prevents the network from having an overfitting issue.

We set different values for all hyper-parameters and completed the hyper-parameters' sensitivity analysis to identify the optimal parametric values for the proposed BiLSTM model. Table 2 demonstrates the list of hyper-parameters of the BiLSTM method for sensitivity analysis in case of Singapore and the United Kingdom.

Figures 3 and 4 demonstrate the hyper-parameters sensitivity analysis for Singapore and the United Kingdom and demonstrate the effect of changes of hyper-parameter values on RMSE, MAE, and MAPE on the y-axis visually.

In Figures 3 and 4, the first rows show the number of epochs on the x-axis and performance values on the y-axis. In the United Kingdom case, the network can be observed to have achieved optimum values in 100 epochs, and for Singapore, optimal values have been achieved in 130 epochs. The second rows in both figures show the batch size number on the x-axis and changes in performance measures on the y-axis—the optimal value for Singapore is 4 and is 6 for the United Kingdom. However, the curve for Singapore and the United Kingdom shows a steep increase after batch sizes 4 and 6, respectively, and then a sudden steep decrease after 12. The third rows show the changes in all performance measures versus changes in the LSTM units. For Singapore, after 128 units, there is a steep increase in RMSE values compared to the MAE and MAPE but for the United Kingdom, RMSE increases, however MAE and MAPE values decrease for higher values of LSTM. The last rows in both figures show the number of dense units or hidden units in the dense layer (fully-connected layer) on the x-axis and corresponding changes in RMSE, MAE, and MAPE values on y-axis. In the case of the dense unit for both countries, the best result is obtained at 256 units. However, the United Kingdom graph shows almost the same zigzag curve for all performance metrics compared to Singapore. Hence, in Figures 3 and 4, the best combination of hyper-parameter for Singapore and the United Kingdom are attained through grid search and are (130, 128, 4, 256), (100, 64, 6, 256), respectively. The first values are the number of epochs; the second values are the numbers of LSTM units in the BiLSTM network; the third values show the batch number size, and the last values indicate the number of hidden units in the dense layer.

**Table 2.** Hyper-parameter list of the BiLSTM method used in the sensitivity analysis.

Country	RMSE	MAE	MAPE	Epochs	LSTM Unit	Dense Unit	Batch Size Number
Singapore	7156.45	6545.56	7.45	150	256	64	2
	7020.8	6410.96	7.22	130	128	256	4
	7126.45	6734.56	8.49	80	64	128	6
	7168.45	6674.56	8.31	100	32	32	12
	7234.56	6845.45	8.89	50	16	16	8
	7285.32	6898.78	9.11	30	8	8	10
United Kingdom	1555.65	1288.87	2.99	150	256	32	2
	1546.76	1295.21	3.1	130	128	128	4
	1538.66	1282.85	2.83	100	64	256	6
	1610.45	1334.34	3.78	80	32	64	8
	1676.89	1393.23	4.21	50	16	8	10
	1570.38	1350.65	3.76	30	8	16	12

Similarly, for all countries, optimal values of hyper-parameters have been attained through grid search in our proposed method. In the experiments for all methods including DADLM, STL-DADLM, BiLSTM, and STL-BiLSTM, Tables 3–5 exhibit the best values that have been achieved after grid search. Equations (12)–(14) show the performance measures used in this study.

#### 4.6. Optimization and Performance Metrics

An Adam optimizer used for the optimization of the network parameters was proposed for the attention-based BiLSTM network. Adam optimizer is preferred because it converges faster than other stochastic optimization methods [42,43]. Mean absolute percentage error is used as a loss function in our study as well as in [15] since it provides a percentage error that is easy to interpret. All the experiments were performed on

Windows 7 machine with 4 Gb RAM. The Tensorflow library is used to implement deep neural network.

We obtained results for all the six visitor countries, average error rate for *RMSE* in Equation (12), *MAE* in Equation (13), and *MAPE* in Equation (14). Evaluation metrics were calculated using actual tourism arrival and forecasted tourism arrival for all six countries.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y^{actual} - y^{pred})^2}, \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y^{actual} - y^{pred}|, \quad (13)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{(y^{actual} - y^{pred})}{y^{actual}} \times 100\%. \quad (14)$$

Tables 3–5 show the results of baseline methods and the proposed method (each with and without STL). ARIMA and Persistence baseline methods have been implemented according to [44,45] for our study. DADLM and STL-DADLM baseline methods were implemented based on the instruction in [15] and in their GitHub profile. Bold values show the better values as compared to the other methods, and the tables show that the BiLSTM network achieves better accuracy than DADLM for all six sources in all three metrics used in this study, which demonstrates that the BiLSTM network is able to utilize the additional context information achieved through processing the same input twice from forward to backward and backward to forward, leading to better performance [46,47]. Our decomposition-based method with the combination of bidirectional LSTM network, known as the STL-BiLSTM network, also outperforms the STL-DADLM method.

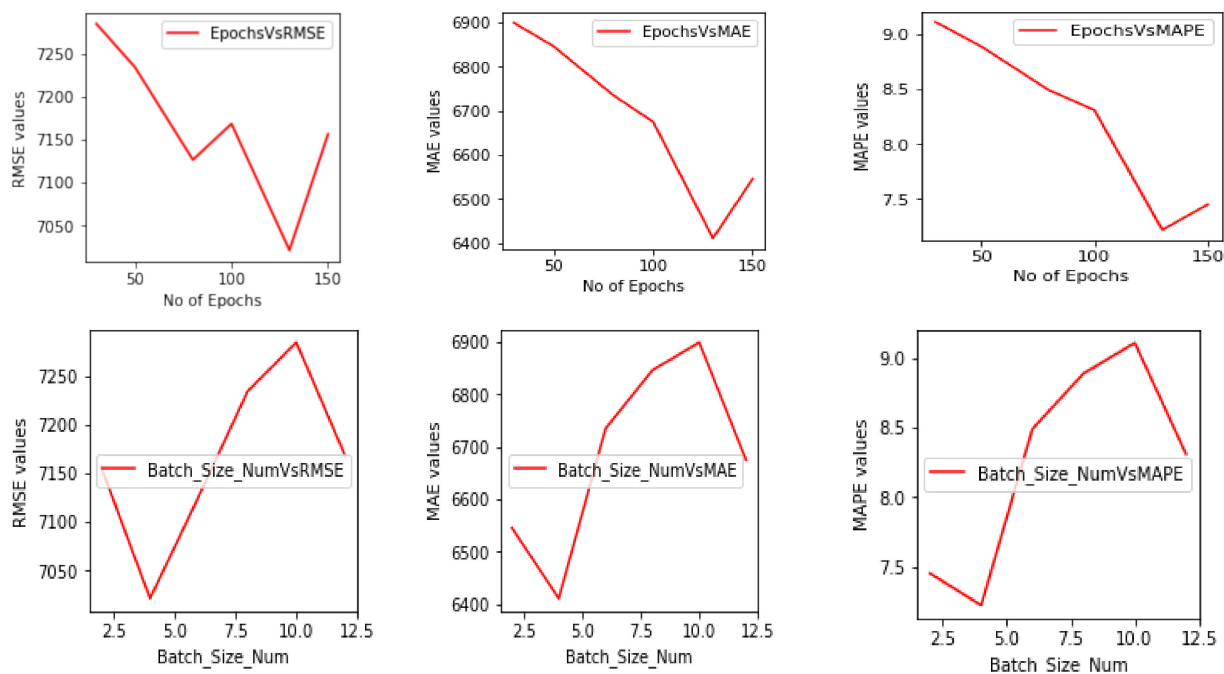
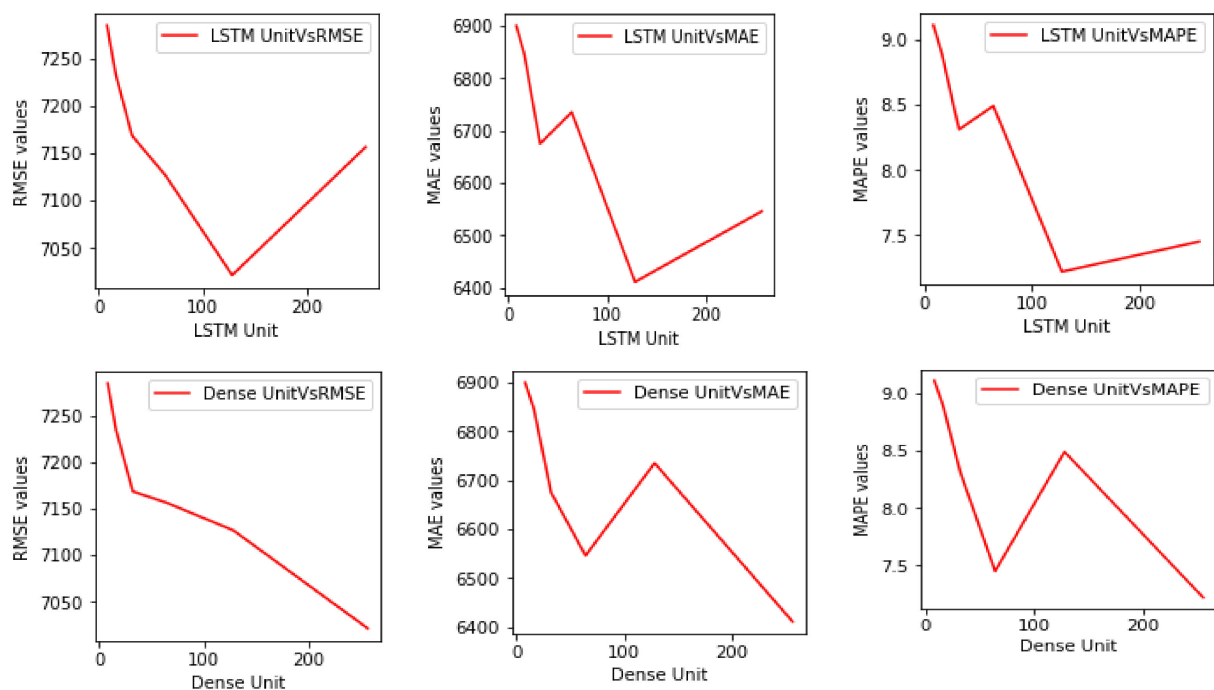


Figure 3. Cont.





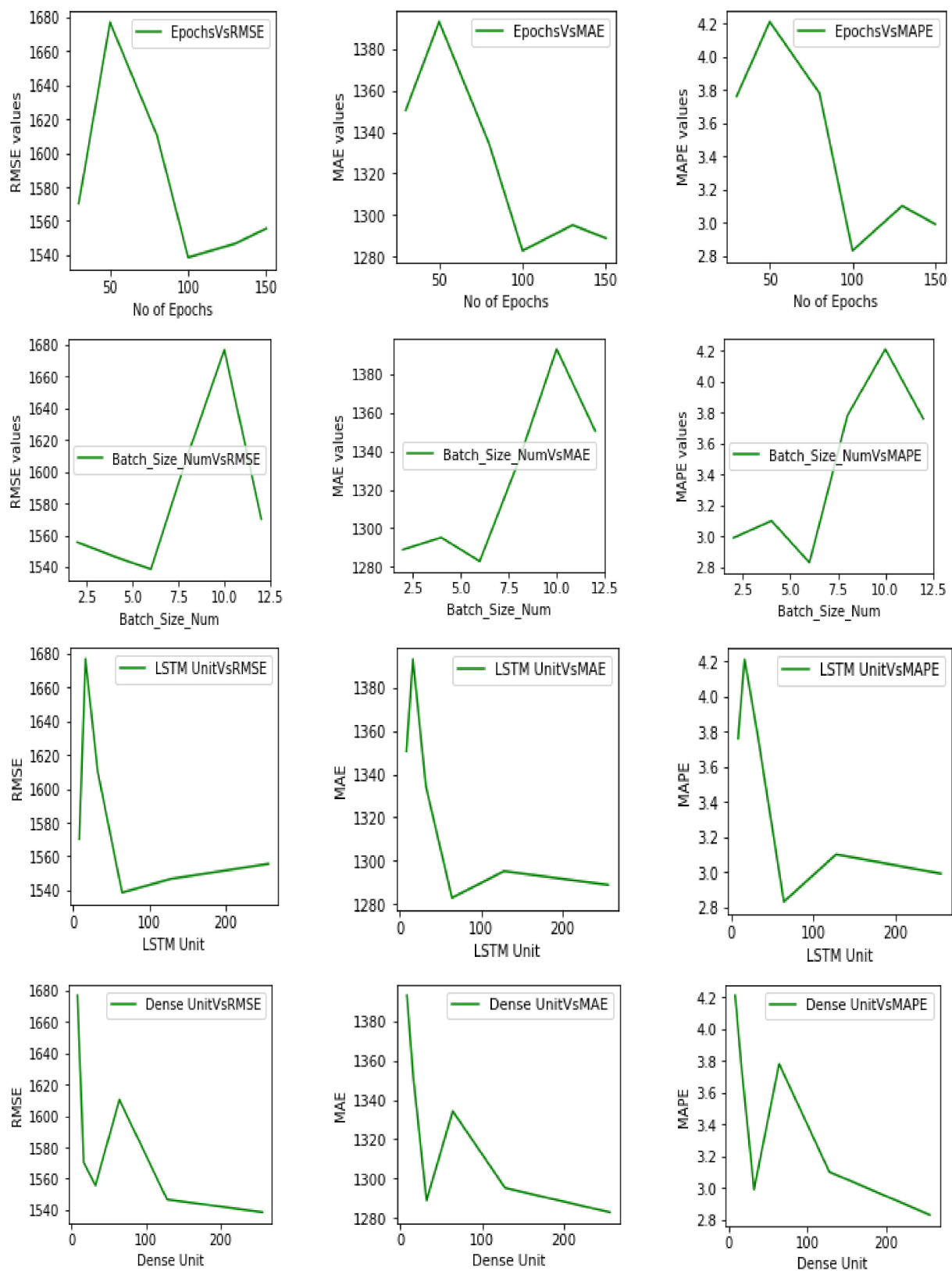
**Figure 3.** Sensitivity analysis of different parameters in case of Singapore arrivals in Hong Kong.

**Table 3.** RMSE comparison for selected forecasting methods for six source countries.

Models	Australia	Singapore	Thailand	United States	United Kingdom	Philippines
Persistence	13,660.56	22,602.51	15,610.02	30,734.47	16,904.61	12,721.11
ARIMA [14]	12,990.33	20,052.55	14,197.85	27,348.02	15,906.04	11,353.48
DADLM (LSTM) [32]	11,373.59	18,126.55	10,669.45	26,243.57	10,998.21	10,124.48
STL-DADLM (LSTM) [32]	2187.67	8429.56	4043.56	8412.45	2629.81	6745.43
BiLSTM	9524.56	17,399.61	8872.67	25,233.69	10,445.41	10,767.56
STL-BiLSTM [ours]	<b>1989.87</b>	<b>7020.20</b>	<b>3854.56</b>	<b>7260.22</b>	<b>1538.66</b>	<b>6478.67</b>

**Table 4.** MAE comparison for selected forecasting methods for six source countries.

Models	Australia	Singapore	Thailand	United States	United Kingdom	Philippines
Persistence	10,329.63	18,129.66	13,731.00	13,394.43	7869.87	10,714.66
ARIMA [14]	6851.40	14,527.07	12,043.80	12,043.80	6611.22	8622.58
DADLM (LSTM) [32]	8980.79	12,499.58	7543.36	24,450.54	8567.53	8088.05
STL-DADLM (LSTM) [32]	1823.19	7251.32	2067.43	7136.69	2327.16	6170.12
BiLSTM	7560.44	12,110.71	6534.78	23,234.56	8145.45	6770.45
STL-BiLSTM [Ours]	<b>1703.65</b>	<b>6410.96</b>	<b>1934.34</b>	<b>6251.36</b>	<b>1282.85</b>	<b>5898.06</b>



**Figure 4.** Sensitivity analysis of different parameters in case of United Kingdom arrivals in Hong Kong.

**Table 5.** MAPE comparison for selected forecasting methods for six source countries.

Models	Australia	Singapore	Thailand	United States	United Kingdom	Philippines
Persistence	22.20	30.62	31.71	14.25	16.30	18.82
ARIMA [14]	15.59	25.01	21.89	14.02	15.06	14.75
DADLM(LSTM) [32]	18.34	17.6	14.91	27.49	17.49	13.21
STL-DADLM (LSTM) [32]	3.98	9.12	7.23	6.28	5.23	9.13
BiLSTM	15.89	16.61	14.78	24.56	14.76	11.65
STL-BiLSTM [Ours]	<b>3.56</b>	<b>7.22</b>	<b>6.67</b>	<b>5.67</b>	<b>2.83</b>	<b>8.67</b>

## 5. Implications, Conclusion, and Limitation of the Study

### 5.1. Managerial Implication and Conclusion

The findings of the current study and the proposed model can be important to policymakers, proprietors, and managers in the tourism industry. The study presents a model by integrating BiLSTM and SII index data with attention mechanism and offers a robust method to forecast the tourist arrival with relatively higher accuracy [48]. Therefore, with accurate information in hand, the government and other stakeholders can plan the development of infrastructure, transportation, hotel booking and other resources in advance [49–51]. Moreover, the results will also enable managers/proprietors to design specific marketing strategies and communication messages to evoke a positive response from the tourists.

This study presents an attention-based BiLSTM network to enhance overall accuracy in forecasting tourist arrivals and their demands. For a case study, we used six countries—Australia, the Philippines, Singapore, Thailand, the United Kingdom, and the United States—as sources for forecasting tourism demand for Hong Kong. At the same time, we addressed the limitations of a standard LSTM network, which can only remember from left to right. Our proposed network can remember from left to right and right to left, which adds extra context to the network that can learn better representation for forecasting and improved network performance accuracy in tourism forecasting in the case of Hong Kong.

More specifically, by integrating attention mechanism and BiLSTM network and SII index data with tourist arrival, the study identifies individual SII index variables with the help of attention mechanism that affects the forecasting performance of BiLSTM network. Attention mechanism assigns more weights to those individual variables that empirically demonstrate a strong connection in forecasting, leading to improved overall forecasting performance of the BiLSTM network.

The proposed methods can benefit stakeholders involved in the tourism industry to make decisions and planning for their businesses. We compared ARIMA, Persistence, and DADLM methods (with and without STL) for all six countries. The proposed method outperformed all the other compared methods for all source markets except for Thailand with STL decomposition but, without STL, the method outperformed all methods for all six countries. This study, therefore, makes an essential contribution to tourism demand forecasting.

We present a bidirectional network that works better than the univariate LSTM network in Hong Kong tourism demand forecasting. To search for the best hyperparameter that can give the best-optimized parameter is a very complex and costly task in the case of deep neural networks. Future work will focus on identifying a bigger range of values in the hyper-parameter tuning process to achieve a better forecasting stability. Due to the ongoing coronavirus pandemic, the tourism industry is affected the most by the travel restrictions. The post-pandemic situation for the tourism industry will be dependent on government regulations and safety measures. It will depend on the government and tourism practitioners to respond after the post-Covid-19 crisis to attract tourists. Forecasting tourism

demands will also depend on several factors, including government policies, tourists' environmental attitude and information [52]. Therefore, there will be a need to include those factors in the study to forecast tourism demand accurately.

### 5.2. Limitations and Direction for Future Research

In designing this study, the authors have attempted to be methodological and scientific, yet this study has some limitations. Future studies may conceptualize in a manner that they can address these limitations. Limited hyper-parameters settings may have affected the generalization of the archived results, future researchers can attempt to diversify their study by including more values in hyper-parameters tuning. The second limitation is that this study presents the proposed methods based on the tourists' arrival in Hong Kong, thus, results may not generalize for other countries when replicating this study. Therefore, future researchers can include more countries and look for the effect of the proposed method for more than one country. Another limitation is that this study presents the sensitivity analysis for hyper-parameters for Singapore and United Kingdom. Future researchers can analyze all countries and see the effect model when changing hyper-parameters. Lastly, the current study has used a single method, i.e., LSTM, as the benchmark, hence, future researchers may compare the proposed method with a combination of Empirical Mode Decomposition (EMD) and BiLSTM or GRA-BiLSTM, which are quite popular methods of predictions.

**Author Contributions:** Conceptualization, M.A.; data curation, M.F.A. and R.K.C.; formal analysis, M.A., J.-Z.W. and M.F.A.; funding acquisition, A.A. and J.-Z.W.; investigation, M.A.; methodology, M.A., M.J.R. and M.F.A.; project administration, J.-Z.W. and A.A.; software, R.K.C., M.J.R. and A.A.; supervision, M.A.; writing—original draft, M.A. and M.F.A.; writing—review and editing, R.K.C., M.J.R., A.A. and J.-Z.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Taif University, Saudi Arabia, through the Taif University Researchers Supporting Project, under Grant “TURSP-2020/121” and the Ministry of Science and Technology, Taiwan (MOST108-2221-E-031-001-MY2; MOST110-2628-E-031-001) and the Center for Applied Artificial Intelligence Research, Soochow University, Taiwan (C-01).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank Sondoss El Sawah, Acting Director, Centre for System Capability, UNSW-Canberra at ADFA for her critical evaluation, numerous discussions, and helpful comments which helped in improving the overall quality of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

SII	search intensities indices
LSTM	Long Short-Term memory
BiLSTM	Bidirectional Long Short-Term Memory
STL	Seasonal and Trend Decomposition using Loess
SVM	Support Vector Machine
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Network
DNN	Deep Neural Network
DCNN	Deep Convolutional Neural Network
RNN	Recurrent Neural Network
GRU	Gated Recurrent Unit
EMP	Empirical Mode Decomposition
DADLM	Duo-Attention Deep Learning Model
HKBT	Hong Kong Tourism Board
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

## References

1. WTTC. *Travel and Tourism: Economic Impact 2019*; World Travel & Tourism Council (WTTC): London, UK, 2019.
2. Chan, Y.-M.; Hui, T.-K.; Yuen, E. Modeling the Impact of Sudden Environmental Changes on Visitor Arrival Forecasts: The Case of the Gulf War. *J. Travel Res.* **1999**, *37*, 391–394. [\[CrossRef\]](#)
3. Witt, S.F.; Witt, C.A. Forecasting tourism demand: A review of empirical research. *Int. J. Forecast.* **1995**, *11*, 447–475. [\[CrossRef\]](#)
4. Volchek, K.; Liu, A.; Song, H.; Buhalis, D. Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tour. Econ.* **2018**, *25*, 425–447. [\[CrossRef\]](#)
5. Cai, Z.J.; Lu, S.; Zhang, X.B. Tourism demand forecasting by support vector regression and genetic algorithm. In Proceedings of the 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, 8–11 August 2019; pp. 144–146.
6. Law, R.; Li, G.; Fong, D.K.C.; Han, X. Tourism demand forecasting: A deep learning approach. *Ann. Tour. Res.* **2019**, *75*, 410–423. [\[CrossRef\]](#)
7. Mishra, M.; Srivastava, M. A view of artificial neural network. In Proceedings of the 2014 International Conference on Advances in Engineering and Technology Research (ICAETR-2014), Unnao, India, 1–2 August 2014.
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497. [\[CrossRef\]](#)
9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Baldigara, T.; Mamula, M. Modelling international tourism demand using seasonal ARIMA models. *Tour. Hosp. Manag.* **2015**, *21*, 19–31. [\[CrossRef\]](#)
11. Ognjanov, B.; Tang, Y.; Turner, L. Forecasting International Tourism Regional Expenditure. *Chin. Bus. Rev.* **2018**, *17*. [\[CrossRef\]](#)
12. Law, R. Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tour. Manag.* **2000**, *21*, 331–340. [\[CrossRef\]](#)
13. Chang, Y.; Tsai, C. Apply deep learning neural network to forecast number of tourists. In Proceedings of the 31st International Conference on Advanced Information Networking and Applications: Workshops (WAINA), Taipei, Taiwan, 27–29 March 2017; pp. 259–264.
14. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3478–3487. [\[CrossRef\]](#)
15. Zhang, Y.; Li, G.; Muskat, B.; Law, R. Tourism Demand Forecasting: A Decomposed Deep Learning Approach. *J. Travel Res.* **2020**. [\[CrossRef\]](#)
16. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
17. Rønning, O.; Hardt, D.; Sogaard, A. Sluice resolution without hand-crafted features over brittle syntax trees. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2018; Volume 2.
18. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292. [\[CrossRef\]](#)
19. Siami-Namini, S.; Tavakoli, N.; Siami Namin, A. A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM. *arXiv* **2019**, arXiv:1911.09512v1.
20. Yang, X.; Pan, B.; Evans, J.A.; Lv, B. Forecasting Chinese tourist volume with search engine data. *Tour. Manag.* **2015**, *46*, 386–397. [\[CrossRef\]](#)
21. PartnerNet. Hong Kong Tourism Board. Available online: <https://partnernet.hktb.com/en/home/index.html> (accessed on 21 November 2020).
22. Song, H.; Qiu, R.T.; Park, J. A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Ann. Tour. Res.* **2019**, *75*, 338–362. [\[CrossRef\]](#)
23. Joo, I.-T.; Choi, S.-H. Stock price prediction model based on bidirectional LSTM circulatory neural network. *J. Korea Inst. Inf. Electron. Commun. Technol.* **2018**, *11*, 204–208. [\[CrossRef\]](#)
24. Song, H.; Li, G. Tourism demand modelling and forecasting—A review of recent research. *Tour. Manag.* **2008**, *29*, 203–220. [\[CrossRef\]](#)
25. Prabhakaran, S. ARIMA Model—Complete Guide to Time Series Forecasting in Python. 2021. Available online: <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/> (accessed on 28 August 2021).
26. Lim, C.; McAleer, M.; Min, J.C. ARMAX Modelling of International Tourism Demand. *Math. Comput. Simul.* **2009**, *79*, 2879–2888. [\[CrossRef\]](#)
27. Bangwayo-Skeete, P.F.; Skeete, R.W. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tour. Manag.* **2015**, *46*, 454–464. [\[CrossRef\]](#)
28. Brownlee, J. How to Grid Search ARIMA Model Hyper Parameters with Python. Available online: <https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/> (accessed on 28 August 2021).
29. Long, W.; Liu, C.; Song, H. Pooling in Tourism Demand Forecasting. *J. Travel Res.* **2018**, *58*, 1161–1174. [\[CrossRef\]](#)



30. Gunter, U.; Önder, I. Forecasting city arrivals with Google Analytics. *Ann. Tour. Res.* **2016**, *61*, 199–212. [\[CrossRef\]](#)
31. Assaf, A.G.; Li, G.; Song, H.; Tsionas, M.G. Modeling and Forecasting Regional Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model. *J. Travel Res.* **2018**, *58*, 383–397. [\[CrossRef\]](#)
32. Brownlee, J. How to Create an ARIMA Model for Time Series Forecasting in Python. 2017. Available online: <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/#:~:text=A%20popular%20and%20widely%20used,structures%20in%20time%20series%20data> (accessed on 16 May 2021).
33. Chen, H.K.; Wu, C.J. Travel Time Prediction Using Empirical Mode Decomposition and Gray Theory: Example of National Central University Bus in Taiwan. *Transp. Res. Rec.* **2012**, *2324*, 11–19. [\[CrossRef\]](#)
34. Hassani, H.; Webster, A.; Silva, E.S.; Heravi, S. Forecasting, U.S. Tourist arrivals using optimal Singular Spectrum Analysis. *Tour. Manag.* **2015**, *46*, 322–335. [\[CrossRef\]](#)
35. Silva, E.S.; Hassani, H.; Heravi, S.; Huang, X. Forecasting tourism demand with denoised neural networks. *Ann. Tour. Res.* **2018**, *74*, 134–154. [\[CrossRef\]](#)
36. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
37. Wang, C.-C.; Chien, C.-H.; Trappey, A. On the Application of ARIMA and LSTM to Predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements. *Processes* **2021**, *9*, 1157. [\[CrossRef\]](#)
38. Yang, Y.; Pan, B.; Song, H. Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data. *J. Travel Res.* **2013**, *53*, 433–447. [\[CrossRef\]](#)
39. Correia, A.; Kozak, M.; Ferradeira, J. Impact of culture on tourist decision-making styles. *Int. J. Tour. Res.* **2010**, *13*, 433–446. [\[CrossRef\]](#)
40. Qiu, H.; Fan, D.X.F.; Lyu, J.; Lin, P.M.C.; Jenkins, C.L. Analyzing the Economic Sustainability of Tourism Development: Evidence from Hong Kong. *J. Hosp. Tour. Res.* **2018**, *43*, 226–248. [\[CrossRef\]](#)
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
42. Pan, B.; Yang, Y. Forecasting Destination Weekly Hotel Occupancy with Big Data. *J. Travel Res.* **2016**, *56*, 957–970. [\[CrossRef\]](#)
43. Choi, H.; Varian, H. Predicting the Present with Google Trends. *Econ. Rec.* **2012**, *88*, 2–9. [\[CrossRef\]](#)
44. Brownlee, J. How to Make Baseline Predictions for Time Series Forecasting with Python. 2017. Available online: <https://machinelearningmastery.com/persistence-time-series-forecasting-with-python/#:~:text=The%20equivalent%20technique%20for%20use,conditions%20for%20a%20baseline%20forecast> (accessed on 19 April 2021).
45. Kulendran, N.; Wilson, K. Modelling Business Travel. *Tour. Econ.* **2000**, *6*, 47–59. [\[CrossRef\]](#)
46. Song, H.; Dwyer, L.; Li, G.; Cao, Z. Tourism economics research: A review and assessment. *Ann. Tour. Res.* **2012**, *39*, 1653–1682. [\[CrossRef\]](#)
47. Gunter, U.; Önder, I.; Smeral, E. Scientific value of econometric tourism demand studies. *Ann. Tour. Res.* **2019**, *78*, 102738. [\[CrossRef\]](#)
48. Wan, S.K.; Song, H. Forecasting turning points in tourism growth. *Ann. Tour. Res.* **2018**, *72*, 156–167. [\[CrossRef\]](#)
49. Li, G.; Song, H.; Witt, S.F. Recent Developments in Econometric Modeling and Forecasting. *J. Travel Res.* **2005**, *44*, 82–99. [\[CrossRef\]](#)
50. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [\[CrossRef\]](#)
51. Adil, M.; Ansari, M.F.; Alahmadi, A.; Wu, J.-Z.; Chakraborty, R.K. Solving the problem of class imbalance in the prediction of hotel cancellations: A hybridized machine learning approach. *Processes* **2021**, *9*, 1713. [\[CrossRef\]](#)
52. Sadiq, M.; Adil, M. Ecotourism related search for information over the internet: A technology acceptance model perspective. *J. Ecotourism* **2021**, *20*, 70–88. [\[CrossRef\]](#)