Tourism Demand Forecasting Using Stacking Ensemble Model With Adaptive Fuzzy Combiner

Selcuk Cankurt
Ala-Too International University

Abdulhamit Subasi (absubasi@effatuniversity.edu.sa)
Effat University  https://orcid.org/0000-0001-7630-4084

Research Article

**Keywords:** Artificial neural network (ANN), ANFIS, Multivariate time series forecasting, Stacking ensemble, Tourism demand forecasting

**DOI:** https://doi.org/10.21203/rs.3.rs-734279/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Tourism Demand Forecasting Using Stacking Ensemble Model with Adaptive Fuzzy Combiner

Selcuk Cankurt  
Faculty of Engineering  
Ala-too International University  
Bishkek, Kyrgyzstan  
E-mail: s.cankurt@iaau.edu.kg

Abdulhamit Subasi  
Institute of Biomedicine, Faculty of Medicine, University of Turku, 20520, Turku, Finland.  
E-mail: abdulhamit.subasi@utu.fi  
Department of Computer Science, College of Engineering, Effat University, Jeddah, 21478, Saudi Arabia.  
E-mail: absubasi@effatuniversity.edu.sa

Abstract  
Over the last decades, several soft computing techniques have been applied to tourism demand forecasting. Among these techniques, a neuro-fuzzy model, Adaptive neuro fuzzy inference system (ANFIS) has started to emerge. A conventional ANFIS model cannot deal with the large dimension of a dataset, and cannot work with our dataset, which is a composed of a 62 time-series, as well. This study attempts to develop an ensemble model by incorporating neural networks with ANFIS to deal with the large number of input variables for multivariate forecasting. Our proposed approach is a collaboration of two base learners, which are types of the neural network models and a meta-learner of ANFIS in the framework of the stacking ensemble. The results show that the stacking ensemble of ANFIS (meta-learner) and ANN models (base learners) outperforms its stand-alone counterparts of base learners. Numerical results indicate that the proposed ensemble model achieved a MAPE of 7.26% compared to its single-instance ANN models with MAPEs of 8.50% and 9.18% respectively. Finally, this study which is a novel application of the ensemble systems in the context of tourism demand forecasting has shown better results compared to those of the single expert systems based on the artificial neural networks.

Keywords: Artificial neural network (ANN); ANFIS; Multivariate time series forecasting; Stacking ensemble; Tourism demand forecasting.

1 Introduction  
The tourism sector has grown immensely over the past several decades, and it has became the largest and the fastest growing industry on the world. As tourism became the biggest social and economic phenomena, huge number of the researches were held related to the tourism studies. In the tourism studies, tourism forecasting has become one of the established areas of research. The essential aim of tourism demand forecasting is to help the public and private sectors for the planning and policy purposes. Tourism demand forecasting is one of the prerequisites to successfully and readily making the plans and

1 Compliance with ethical standards  
Conflict of interest: The authors declare that they have no conflict of interest.  
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
regulations for the future and to carry out them timely. That is why accurate tourism demand forecasting is heavily studied by the researchers.

Tourism demand forecasting is of great economic importance for both the public and private sectors. Due to their perishable value, tourism items such as unfilled airline seats, unoccupied hotel rooms and unused facilities cannot be stored (Archer, 1987). Thus, accurate forecasting of tourism demand has a great significance to the players of the tourism sector for their efficient management and plans (Pai & Hong, 2005). Short term tourism forecasting empowers the managers of the tourism sector in their operational decision such as inventory control, stock control, employing operative-staff, effective use of resources and facilities (Dogru, Sirakaya-Turk, & Crouch, 2017; Fernandez-Morales, Cisneros-Martínez, & McCabe, 2016).

The tourism sector is one of the most important sources of income, employment, and foreign exchange earnings in developing countries and rather Turkey. The motivation for this work is to make a contribution to Turkey by the help of the accurate tourism projection. Since tourism is the world’s largest industry and Turkey is one of the biggest players on the tourism market, any research on tourism and especially on the scope of Turkey yields great economic benefit and contributes to the tourism industry.

Before the 1990s, conventional regression methods dominated the literature on tourism demand forecasting, but this pattern shifted after the mid-1990s, as more researchers started using modern techniques such as seasonal autoregressive integrated moving average (SARIMA), extreme learning machine (ELM), support vector regression (SVR) with particle swarm optimization (PSO), long short-term memory (LSTM) (Wong, Song, Witt, & Wu, 2007; Li, Pan, Law, & Huang, 2017; Rossello & Sanso, 2017; Akin, 2015; Zhang, Jiang, Wang, & Sun, 2021). The literature on tourism demand forecasting methods has begun to be influenced by the significant developments in soft computing and related disciplines of the 1990s. The artificial neural network method was introduced to tourism forecasting in the late 1990s (Law & Au, 1999), later the artificial neural networks have increasingly used to forecast demands for tourism (Songa & Li, 2008; Law, 2000; Pattie & Snyder, 1996). Predominantly, intelligence techniques, such as support vector machines (SVM), fuzzy logic, genetic algorithms, swarm intelligence, artificial neural networks (ANN) have emerged in the literature on tourism forecasts (Songa & Li, 2008).

In recent years, newly developed soft computing techniques like fuzzy systems have been applied to various fields to implement an intelligent information systems, including forecasting (Morabito & Versaci, 2003; Anifowose, Labadin, & Abdulraheem, 2013; Anifowose, Labadin, & Abdulraheem, 2013; Anifowose & Abdulraheem, 2011). The ability of the incorporation between fuzzy systems and neural networks has been noticed and started to investigate since the early 1990s, which led to the development of the hybrid neuro-fuzzy systems. In the hybrid neuro-fuzzy model, fuzzy logic principles collaborate with the neural network techniques to meet the benefits of both in a single model (Azar, 2010). Individually, neural networks and fuzzy logic are powerful soft computing techniques for the universal approximations (Hornik, Stinchcombe, & White, 1989; Chen & Chen, 1993; Kosko, 1992; Wang & Mendel, 1992). Fuzzy logic and artificial neural networks are complementary technologies and appear to be a promising method in the design of intelligent systems for the universal function approximator. ANFIS is a new kind of soft computing tool used in the forecasting, which is one of the specific implementations of the neuro-fuzzy. In general speaking, ANFIS refers to the grid partitioned ANFIS, which is so-called conventional ANFIS. It is a successful hybrid model to implement a universal approximator (Jang & Sun, 1997). However, it cannot achieve the satisfactory training performance (training time) and forecasting performance (forecasting accuracy) without sacrificing either from the number of input variables or from the linguistic variables in the existence of the moderate number of attributes, and it cannot tackle with the large number of input variables at all. Recently, in a few studies, ANFIS techniques have been applied to tourism forecasting problems (Fernando & Turner, 2006; Chen, Ying, & Pan, 2010; Hadavandi, Shavandi, & Ghanbari, 2010; Fernando, Reznik, & Turner, 1999; Karaboga & Kaya, 2020).

In tourism demand studies, besides the single and hybrid forecasting methods, combined methods are also proposed. Since the technique of combining forecasts is introduced by the early works (Reid, 1969; Bates & Granger, 1969); it is the well-established area in the literature of forecasting. A
comprehensive review paper of (Clemen, 1989) and many other empirical papers, for example, (Ginzburg & Horn, 199; Zhang G. P., 2003; Lemke & Gabrys, 2010; Lin, Lin, Shyu, & Lin, 2012; Firmino, Neto, & Ferreira, 2013) support that combined forecasts can generally outperform the individual forecasts. As an extension to the combined models, many ensemble models are introduced and investigated mainly in the machine learning area (Dietterich, 1997; Anifowose, Labadin, & Abdulrahheem, 2017; Fatai Anifowose, 2015), such as stacked generalization (or stacking) (Wolpert, 1992), boosting (Schapire, 1990), bagging (Breiman, 1996), and voting. Ensemble is an approach to the combination of results, produced by many single forecasters. The key stimulus behind the ensemble techniques is to get a better overall performance than the independent single forecaster. Some of the most important findings of ensemble methods is to get a more stable and accurate prediction (Polikar, 2006). Several studies (Hansen & Salamon, 1990; Gheyas & Smith, 2011; Qian & Rasheed, 2004; Nanni & Lumini, 2009) have shown that ensemble methods perform better than their individual predictors. In the tourism sense, there have been several analytical works on the combination of forecasts (Wong, Song, Witt, & Wu, 2007; Fritz, Brandon, & Xander, 1984; Chan, Witt, Lee, & Song, 2010; Chen K.-Y. , 2011; Andrawis, Atiya, & El-Shishiny, 2011; Shen, Li, & Song, 2011). However, there has been a few ensemble approach (Chen & Jie, 2011; Cankurt S., 2016 ; Zhang, Jiang, Wang, & Sun, 2021) seen in the circumstance of the tourism demand forecasting, which is a recent emerged area in the machine learning.

The purpose of this study is to introduce the conventional ANFIS technique, to assess its major strengths and weaknesses, and to propose a stacking ensemble model on how the ANFIS model can be applied in a multivariate forecasting in the existence of the data set with the large amount of the input features. In this study, a type of heterogeneous ensemble model has been introduced and developed to make multivariate forecasting by employing ANFIS without sacrificing the number of input variables. The expected outcome of this approach is to get a better approximation; hence the number of the input variables will affect the representation of the phenomenon of interest.

The proposed ensemble model concerns the issues of: (1) at the base level, making the initial predictions by using the original data set and generating meta-data set in smaller dimension to make it suitable to use with ANFIS; (2) at the meta-level, combining the predictions of the base learners and improving the accuracy of them for the multivariate forecasting problems. More specifically, this paper aims to develop a stacking ensemble model for the fact that whether a data set in the large dimension can be applied to ANFIS.

This paper is organized as follows. Section 2 introduces the data used in this study, the theoretical background employed in this research, and explains classification models proposed and evaluated in this study. The evaluation of experimental results and comparison of classifier results obtained in this research are given in Section 3. The discussion about the results is given in Section 4. And lastly, the discussion and conclusion of the research is given in Section 5.

2 Methodology and Data Set

2.1 Data set

We have employed a subset version of the tourism data used in (Cankurt & Subasi, 2016), which covers the monthly time series of Turkey and its top 24 ranked tourism clients. Complete list of the monthly time series are: Wholesale Prices Index, US Dollar Selling, One Ons Gold London Selling Price USD, Hotel Bed Capacity of Turkey, CPI of leading clients of Turkey (namely Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Russian Federation), number of the tourists coming from the leading clients of Turkey (namely Germany, Russia, France, Ireland, Bulgaria, Georgia, Greece, Ukraine, Azerbaijan, Austria, Belgium, Denmark, Holland, England, Spain, Sweden, Switzerland, Italy, Norway, Poland, Romania, USA, Iraq, Syrian Arab rep.), Exchange rate of the leading countries of Turkey (Canadian dollar, Danish Krone, Norwegian Krone, Polish Zloty, Swedish Krona, Swiss Franc, Turkish Lira, British Pound, Russia Rouble), and Former
Tourists are attributes and Total Number of tourists is output. Also indexes of the year, month, and season are included the data set. These variables might help the forecasting models to recognize the seasonal pattern in the nature of the tourism data set.

2.2 Using ANFIS for the multivariate time series prediction problems

ANFIS is a recent and successful hybrid soft computing approach for the forecasting problems. Hence ANFIS accepts the multiple numerical inputs and produces a single output; it is well suited to the numerical prediction problems. If a better performance is desired, it might be one of the first nonlinear models to see if the performance level could be improved with it. However, one of the major drawbacks for ANFIS is unable to deal with the large number of input variables (The MathWorks, 2013).

The conventional ANFIS model, which employs grid search technique to build the initial fuzzy inference system (FIS) structure, is not able to flexibly adapt to the data sets with the large amount of the input variables. Like many numerical algorithms, ANFIS also suffers from the curse of dimensionality, which denotes such a circumstances that the cost of an optimal solution increases exponentially with the dimension. The grid partitioning approach generates rules by considering all possible combinations of membership functions and inputs, and produces a FIS structure based on a fixed number of membership functions. Therefore, this partitioning approach causes the excessive generation of rules when the number of inputs is moderately large, that is, more than five or six. The resulting number of the fuzzy rules increases exponentially with the number of input variables, which causes the "curse of dimensionality" (The MathWorks, 2013).

In the practical application, ANFIS has limits based on one outer and one inner parameter: (1) the number of input variables, which describes the phenomena of the forecast and (2) the number of the membership functions, which codes those variables into meaningful fuzzy sets. Employment of large numbers of membership functions helps distinctive mapping of the features into the fuzzy set. But the large number of membership functions practically can be used when you have enough size of training data to generalize the forecasting problem (The MathWorks, 2013). Also, the number of fitting parameters increases as the number of input variables increases. Using the conventional ANFIS with the data set in large dimension takes a long training time, making such analysis impractical or infeasible.

To use ANFIS for the multivariate time series prediction problems, the first thing should be considered is to keep the number of the inputs as much as smaller. In the most studies, this number is six or less. Therefore, we need to determine the most significant variables, which should be used as the input arguments to an ANFIS model. In the previous studies, to get rid of the curse of dimensionality, several techniques have been suggested: (i) empirically determining the subset of the features among the candidate features, for example, by using the exhaustive search, computationally small combinations of the large feature sets are enumerated and evaluated to find out the most suitable sub-features set, (ii) determining the sets of sub-features, which are smaller than the domain feature set based on the prior studies and the review of the domain experts, (iii) employing the pre-processing data methods such as the feature selection algorithms to choose a sub-features set from the candidates by eliminating the features with the low impact factors, (iv) hybridization of ANFIS and the clustering methods such as fuzzy c-means, subtractive clustering algorithms to partition the input space, and (v) in addition to them, we have proposed a stacking ensemble model.

H. Tabari et al. (Tabari, Kisi, Ezani, & Talaee, 2012) attempted to estimate the reference evapotranspiration value using ANFIS models. In their ANFIS models, they have employed empirically selected four, three, two and one input variables respectively from the six candidate features of the climatic data and compared them with SVM, MLR and multiple non-linear regression (MNLR) models and concluded that the ANFIS model with four input variables outperforms the others. M.-Y. Chen (Chen M.-Y. , 2011) proposed a hybrid model of particle swarm optimization (PSO) and ANFIS to predict business failures. He has selected the 5 variables out of the 13 input candidates for his proposed model. The remaining eight variables are discarded by conducting principal component analysis (PCA). Boyacıoğlu (Boyacıoğlu & Avci, 2010) proposed another ANFIS model to predict the stock price index return using nine input variables. Selection is done among 23 of financial and macroeconomic variables,
based on the review of domain experts and prior researches. E. Hadavandi (Hadavandi, Shavandi, & Ghanbari, 2010) proposed a hybrid model, which adopts ANFIS with the pre-processing data techniques for the feature selection and data clustering. He used stepwise regression analysis (SRA) to select the main variables to be utilized in the model and remove the ones with low impact factors, and then he used Self Organization Map (SOM) neural network in order to split the input data into sub-populations and decrease the complexity of the data space to achieve more consistent model.

2.3 Framework of the proposed ensemble method

The ensemble learning has two distinctive concepts: (1) implementation of multiple learners and (2) combining their predictions (Zhang & Ma, 2012). In general, there are two types of ensemble approaches for the combination of multiple predictive models: homogeneous and heterogeneous. The homogeneous model combines the various implementations of the same method, such as bagging (Breiman, 1996) and boosting (Schapire, 1990). The heterogeneous approach melds the implementations of the different or same type of the independent models, such as voting and stacking (Wolpert, 1992). Our proposed model is in the structure of the stacking approach as seen in Fig. 1. Original stacking method proposed by (Wolpert, 1992) is a two-layer model.

![Figure 1 A schematic view of the proposed ensemble model (k=62, n=2 herein)](image)

Original data (also called level-1 data) is presented to the base learners located in the first layer (also called level-1 layer) in a fashion of a sequential batch process. Outputs of the independent base learners constitute the meta-data set (also called level-2 data) in the small dimension. Number of the new generated data series is equal to the number of the base learners employed in the level-1. Level-2 data is fed into the meta-learner in the second layer (also called level-2 layer) which combines them and computes the final prediction.

Because of the curse of dimensionality, which arises from the grid input space partitioning method used by ANFIS, it cannot deal with the large number of inputs (The MathWorks, 2013). In the proposed model, number of the input variables for the ANFIS model is limited by the number of the base learners. In the first layer, firstly, all features are presented to the forecasters whom they are able to deal with the large number of the input variables. Secondly, they lead their outputs to the meta-learner as inputs. The ex-post forecasts are fed into ANFIS to build an intelligent system with the ability of tourist arrival forecasting. While the initial forecasts made by the ANN forecasters in the base learner section, ultimate
forecast and tuning made by the ANFIS forecaster in the meta-learner section. The new dimension of the input data for ANFIS is reduced the number of the employed forecasters in the base learner section. Therefore, ANFIS becomes possible to work with the data set with the large number of the input variables.

In our proposed method, the forecasters used in the base learner section even can function individually on our data set with the 62 attributes. But ANFIS used in the meta-learner section cannot function at all with the same data set. ANFIS can function with the given data set only if it is fed by the base learners. However, the initial forecasts made by the base learners can be improved by ANFIS considerably. While ANFIS is fed by the outputs of the primary forecasters, conversely it serves them by improving their forecasting accuracy. There is a mutual relationship between the meta-learner (ANFIS) and the base learners (two variations of the neural network model): (i) ANFIS is likely to improve the overall performance of the expert system. (ii) On the other side, the base learners will generate the information (meta-data set) in such a dimension, which ANFIS can deal with it. As a conclusion, hopefully ANFIS will contribute the overall system performance by combining the individual predictions of the base learners.

The expected advantages of our proposed model over the above approaches: (1) Generalization capability; because the resulting selection of the sub-features varies in the different data set, those models cannot be generalized easily to fit for the other data sets. Our model uses all the features at hand without doing any elimination or selection based on the specific conditions. (2) Accurate forecasting; while the above approaches are based on the selection of the sub-features by eliminating some of the input variables, our approach uses all the features in the given data set. If there is no redundant variables in the data set, hopefully more variables might contain the more necessary representative features for the interest of the forecasting problem. (3) Easy modelling; proposed model not necessarily requires any predetermined methods for the pre-processing of data such as the feature selection and clustering algorithms.

2.4 Building blocks of the proposed ensemble model

In this study, neural network models are used as the base learner and ANFIS model is used as meta-learner for our multivariate tourism forecasting task. These models will be used to constitute building blocks of the ensemble model. ANN and ANFIS models, which are the proposed ensemble models are derived from, are summarized below.

2.4.1 Artificial neural networks approach

ANN consists of interconnected processing units (generally known as artificial neurons). Processing unit (Neuron) sums the weighted inputs and takes the net input through an activation function to normalize and produce a result (Jones, 2008). The equation of a simple neuron is given as:

\[ y_j = f \left( \sum_{i=1}^{N} w_{ij}x_i + b_j \right) \]  

(1)

The multilayer network architecture consists of two or more hidden layers, and one output layer. BP is one of the most popular approximation approaches for training the multilayer feedforward neural networks based on the Widrow–Hoff training rule. BP algorithm propagates one test case through the MLP in order to calculate the output and compute the error and then adjust the weights and the biases that minimize the sum of the square errors by propagation of the error back at each step (Bishop, 1995; Haykin, 1999).

The next step is to tune the related weights by utilizing the Eq. (2) considering the error before calculated for the node (whether hidden or output) (Jones, 2008).

\[ w_{ij} = w_{ij} + \mu_E y_i \]  

(2)
For the given error \( (E) \) and node output \( (y_i) \), we multiply by a learning rate \( (\mu) \) and add this to the current weight. The result is a minimization of the error at this node, while moving the output of the node closer to the expected output (Jones, 2008).

### 2.4.2 ANFIS - Adaptive Neural Fuzzy Inference System

ANFIS is introduced by J. S. R. Jang in 1993. ANFIS is a hybrid method of neural network and fuzzy system, which is an extension to Takagi-Sugeno fuzzy inference system. It uses TK fuzzy inference system to map the inputs and their corresponding output data, and employs the feed-forward neural network strategy and learning algorithms, which are borrowed from the artificial neural network theory to tune the parameters of TK-FIS automatically.

By adapting the antecedent parameters and consequent parameters for achieving the desired input-output mappings, the neuro-fuzzy inference system can be optimised. The network applies a combination of the least squares technique and the gradient descent approach of back propagation for training the parameters of membership functions and fuzzy rules of Sugeno-type fuzzy system (Jang & Sun, 1997; Jang, July 1991; Jang, May 1993; Sumathi & Paneerselvam, 2010).

### 2.4.3 ANFIS Architecture

ANFIS implements a Sugeno type fuzzy inference system in the form of five-layer neural network architecture. First-order Sugeno ANFIS architecture shown in Fig. 2 is composed of five layers, namely, a fuzzification layer, a product layer, a normalized layer, a defuzzification layer, and a total output layer. The functioning of each layer is defined as follows (Jang & Sun, 1997; Jang, July 1991; Jang, May 1993; Sumathi & Paneerselvam, 2010):

#### Layer 1 - Fuzzification Layer

Every node \( i \) in this layer is an adaptive node and outputs of layer 1 are the fuzzy membership grade of the input \( x \) (or \( y \)) respect to a node function:

\[
O^1_i = \mu_{A_i}(x) \quad \text{for } i = 1, 2 \\
O^1_i = \mu_{B_{i-2}}(y) \quad \text{for } i = 3, 4
\]

(3) \hspace{1.5cm} (4)
where \( A_i \) (or \( B_i \)) is a linguistic label (such as small or large) related to the corresponding node, and \( \mu_{A_i}(x) \) and \( \mu_{B_i}(x) \) are the membership functions. The usually employed membership functions are Bell shaped and Gaussian membership functions.

**Layer 2 - Product Layer**

In the second layer, every node is a fixed node. They are labeled with \( \Pi \), indicating that output of every node is the product of all the incoming signals:

\[
O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1, 2
\]

(5)

Every node output denotes the firing strength of a rule. Generally, any other T-norm operators, which perform fuzzy AND can be utilized as the node function in this layer.

**Layer 3 - Normalized Layer**

Every node in this layer is a fixed node. They are labeled with \( N \), indicating that normalized firing strengths. Output of the \( i^{th} \) node is calculated as the ratio of the \( i^{th} \) rule's firing strength to the sum of all rules' firing strengths:

\[
O_i^3 = \tilde{w}_i = \frac{w_i}{w_1 + w_2}
\]

(6)

**Layer 4 - Defuzzification Layer**

Every node in this layer is an adaptive node. The output of each node is simply the product of the normalized firing strength \( \tilde{w}_i \) from layer 3 and a first order polynomial which is given by:

\[
O_{4,i} = w_if_i = \tilde{w}_i(p_ix + q_iy + r_i)
\]

(7)

where \( \{p_i, q_i, r_i\} \) are referred to as consequent parameters.

**Layer 5 - Total Output Layer**

There is only one single node in this layer, which is a fixed node. It is labeled \( \Sigma \), indicating that the overall output. The overall output as the summation of all incoming signals is given by:

\[
O_{5,i} = \sum_i w_if_i = \sum_i \frac{w_if_i}{w_i}
\]

(8)

Nodes of the first and second layers are adaptive, and the others are the type of the fixed nodes (Jang, July 1991; Jang, May 1993; Sumathi & Paneerselvam, 2010).

### 2.5 Prediction performance metrics

To assess the performance of the individual neural network models and the proposed mutual ensemble techniques, the following three indices are employed: The mean absolute percentage error (MAPE), the root mean squared error (RMSE), and correlation coefficient \( R \), respectively. These error measurements are denoted in the following formulas, in which the \( n \) is the number of the test case and \( a_i \) is the observed value and \( p_i \) is the estimated value for the test case \( i \) (Witten & Frank, 2005).

\[
\bar{a} = \frac{1}{n} \sum_{i=1}^{n} a_i, \quad \bar{p} = \frac{1}{n} \sum_{i=1}^{n} p_i
\]

\[
S_a = \frac{1}{n-1} \sum_{i=1}^{n} (a_i - \bar{a})^2, \quad S_p = \frac{1}{n-1} \sum_{i=1}^{n} (p_i - \bar{p})^2
\]

\[
S_{pa} = \frac{1}{n-1} \sum_{i=1}^{n} (p_i - \bar{p})(a_i - \bar{a})
\]

(9)

**Mean Absolute Percentage Error (MAPE)**

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{a_i - p_i}{a_i} \right|
\]
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{a_i - p_i}{a_i} \right| \times 100 \quad (10)

It calculates the average of the absolute values of the percentage errors of a forecast.

**Root Mean Square Error (RMSE)**

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (a_i - p_i)^2} \quad (11) \]

Our second accuracy metric will be root mean square error. It is the square root of the MSE, which is the mean of the sum of the squares of the prediction errors. This metric corrects the canceling out effects (Witten & Frank, 2005).

**Correlation coefficient (R)**

\[ R = \frac{S_{PA}}{S_P S_A} \quad (12) \]

The final measure which will be employed in our studies is the correlation coefficient (R). It measures amount and direction of a linear relationship between the actual values and the predicted values. The coefficient of determination ($R^2$) shows the percentage of the data which can be explained by the regression equation (Witten & Frank, 2005).

### 3 Experimental setup and results

In this study, a stacking ensemble model and two single models are created and tested on the tourism data set with 62 features for the forecast horizons of six and 12-months-ahead.

Two data sets are arranged as the composition of the input vectors (62 features) and the corresponding 12 and six-months-ahead outputs pairs, which are respectively defined by

\[ [x(t,1), x(t,2), \ldots, x(t, 62); y(t+12)] \text{ and } [x(t,1), x(t,2), \ldots, x(t, 62); y(t+6)] \].

Since our data set covers a wide range of values, logarithmic normalization is applied to the data. Logarithmic normalization enables us to decrease the range of the data set, while preserving the global view.

Overall ensemble models demonstrated in Fig. 3 are defined in the mathematical function notation by

\[ y_{t+12} = f_{ANFIS} (f_{ANN1} (X_t), f_{ANN2} (X_t)) \text{ and } \]
\[ y_{t+6} = f_{ANFIS} (f_{ANN1} (X_t), f_{ANN2} (X_t)) \quad (13) \]

where $d$-dimensional input $X \in R^d$, output $y \in R$ and $d = 62$.
Randomly selected 80% of data points (these become the training data set) are used for the training, while the other 20% are used as testing data set for validating the proposed ensemble model. Firstly, data is presented to the neural network models then the forecasting outputs of the single models are introduced to the ANFIS model as the input.

The neural network models and ANFIS, and related to them the stacking ensemble models are implemented using the Matlab Neural Network Toolbox and Fuzzy Logic Toolbox (The MathWorks, 2012).

### 3.1 Implementation and evaluation of the base learners of the ANN models

Firstly, we have developed and implemented the several feed-forward back-propagation ANN models on the basis of one-layer and three-layer neural networks and selection of their corresponding parameters. We have selected two ANN models, which demonstrate the best performances in each category (one-layer neural networks and three-layer networks). The performance of the ensemble model is highly depended on the diversification of the base learners (Sigletos, Paliouras, & Spyropoulos, 2005). To obtain the diversified base learners, we have selected the best ones of the two different categories.

In the most empirical studies, selection of the number of the hidden layer is varied one to three. One of the used results of Kolmogorov’s theorem for neural networks states that two hidden layers are sufficient for confident estimate of any complex nonlinear function. Actually, one layer in a network is enough to construct an approximation function (Bishop, 1995; Zhang, Patuwo, & Hu, 1998; Aslanargun, Mammadov, Yazici, & Yolacan, 2007). To maximize the diversification of the base learners, we have preferred one and three for the number of the hidden layers. There is no rule, which designates the optimum number of hidden neurons for any given problem. According to the prior studies, the number of neurons in the hidden layer can be up to (1) $2n + 1$ (where $n$ is the number of neuron in the input layer), (2) 75% of input neurons, or (3) 50% of input and output neurons (Lenard, Alam, & Madey, 1995; Patuwo, Hu, & Hung, 1993; Piramuthu, Shaw, & Gentry, 1994; Efendigil, Önüt, & Kahraman, 2009). The number of nodes in the hidden layers is optimized through trials.

The sigmoid transfer function has the ability to yield models with satisfactory accuracy (Choy, Lee, & Lo, 2003). While we have preferred to use the sigmoid transfer function for the nodes of the hidden layers, linear activation function is the only option, which can be employed in the output nodes for the regression problems.

The learning rate regulates the amount of the changes in weight and decreases the possibility of any weight oscillation during the training cycle. A learning rate between 0.05 and 0.5 achieve good results in most practical cases (Rumelhart, Hinton, & Williams, 1986). The momentum factor specifies the influence of past changes and current weight changes, and accelerates the learning time. In general, the momentum factor is close to 1, e.g., 0.8 (Rumelhart, Hinton, & Williams, 1986; Palmer, Montano, & Sese, 2006). When the specified number of epochs or learning rate is achieved, the network stops. Training continues until the performance with the validation set is satisfied or until the specified number of epochs is realized (Witten & Frank, 2005).

The first neural network model consists of three hidden layers with the nodes of 32, 15 and 7 (abbreviated as ANN (32, 15, 7)), and second one has only one hidden layer with 15 nodes (abbreviated as ANN(15)). They have trained with the settings of the learning rate 0.01 and momentum 0.8.

The results of the individual models of neural network forecasts are given in Table 1. Investigation and evaluation of those models (Table 1) showed that (1) ANN (15) has well accuracy with $R^2=0.928$, MAPE=9.18% and RMSE=264956 values, (2) ANN (32, 15, 7) model with $R^2=0.957$, MAPE=8.50% and RMSE=205097 values has the best accuracy among two individual neural network models in the use of the 12-months-ahead horizon. Moreover, the forecasting errors generally decrease with the forecast horizon: that is, errors for the 12-month forecast ahead were less than that of the six-month forecast. Frechtling (Frechtling, 2001) considers the forecasts with MAPE values of less than 10% as
highly accurate forecasting, between 10% and 20% as good forecasting. Individual performance of those single models appears to be satisfactory. But these traditional methods have achieved certain levels of success in the multivariate tourism forecasting. Furthermore, for the better forecasting performance, the further investigation has done based on the collaboration of the neural network models and the neuro-fuzzy modeling approach (ANFIS).

3.2 Implementation and evaluation of the meta-learner using the ANFIS model

Outputs of the individual (ANN (32, 15, 7) and ANN (15)) models produced the metadata set which will be used as input for ANFIS. The ANFIS model was trained at 500 epochs of learning and with two membership functions in the form of a generalized bell-shape curve on each of two inputs, four altogether. There are 4 rules in the generated FIS matrix, 21 nodes in the network structure and the number of fitting parameters is 16, including 12 nonlinear parameters and four linear parameters.

It is crucial that the number of training data points should be several times greater than the number of the parameters being estimated, in order to accomplish good generalization capability (The MathWorks, 2013). Choosing the size of the membership function requires consideration of the trade-offs between the speed and accuracy or good generalization and over-fitting. Larger number of the membership function takes longer to train and to generate predictions. Also increasing the number of the membership functions, increase the number of the fitting parameters. That is why, you can use the large number of membership function, only if you have a large amount of the training data set. But usually in the real life application, it is not easy to collect a large amount of the training data. Since we have around 125 data points in our training data set, to avoid the over training (over fitting) problem, we have used only two membership functions for each input. In this study, the ratio between training data points and fitting parameters is about 7.9 (126/16) for the 12-months-ahead data set and 8.1 (130/16) for the six-months-ahead data set. Training of the ANFIS model and validation of it continue consequently until the specified epoch is reached. The final FIS is the one when the best performance on the validation set is obtained.

ANFIS uses the grid partitioning technique to produce initial membership functions, which are equally spaced and cover the whole input space (The MathWorks, 2013). After 500 epochs of training, new membership functions are produced by ANFIS. The initial and final MFs for the input 1 are shown in Fig. 4. Note that MFs after training have changed. Most of the fitting is done by the linear parameters while the nonlinear parameters are mostly for fine-tuning for further improvement (The MathWorks, 2013).
Fig. 5 displays error curves for training and checking in the measurement of the root-mean-square errors (RMSE). Changing of the training and checking errors can be observed easily in the metric of RMSE related to the epochs. From Fig. 5, we can observe that the RMSE becomes the minimum in checking data set at the epoch 115.
Fig. 6 shows the actual values, and the one predicted by ANFIS. The difference between the real tourism demand and the values estimated using ANFIS is very small when you consider the figures of the arrivals which are counted by the number of the several millions.

**Figure 5** Training and checking errors

**Figure 6** Comparison of the actual values with the results obtained from ANFIS
The results for the proposed ensemble models, which is the combinations of the ANN (32, 15, 7), ANN (15) and ANFIS are reported in Table 2. Results in Table 1 and Table 2 show that our proposed ensemble models have demonstrated significantly better forecasting performance compared to those of the single models (ANN (32, 15, 7) and ANN (15)) in the multivariate tourism demand forecasting. The finding also shows that the fuzzy systems have a potential to be used as a combiner due to its nature that can deal with the imprecise combination rules rather than fixed and exact rules defined by procedures of conventional algorithms.

| Horizon     | Type     | Model                  | MAPE | RMSE   | $R^2$ |
|-------------|----------|------------------------|------|--------|-------|
| 12 months   | Ensemble | $f_{\text{ANFIS}}(f_{\text{ANN(32,15,7)}}, f_{\text{ANN(15)}})$ | 7.26% | 192259 | 0.969 |
| ahead       |          |                        |      |        |       |
| 6 months    | Ensemble | $f_{\text{ANFIS}}(f_{\text{ANN(32,15,7)}}, f_{\text{ANN(15)}})$ | 9.30% | 218624 | 0.952 |
| ahead       |          |                        |      |        |       |

Table 2 Single Models

| Horizon     | Type | Model     | MAPE | RMSE   | $R^2$ |
|-------------|------|-----------|------|--------|-------|
| 12 months   | Single | ANN(32,15,7) | 8.50% | 205097 | 0.957 |
| ahead       |       | ANN(15)   | 9.18% | 264956 | 0.928 |
| 6 months    | Single | ANN(32,15,7) | 11.24% | 234387 | 0.940 |
| ahead       |       | ANN(15)   | 14.67% | 419409 | 0.840 |

The performances of the two ensemble models are evaluated for six and 12 months-horizon data sets. In total, we have developed and investigated three forecast models for each tourism demand data set in this study, these are, two individual forecasting models, which are ANN (32, 15, 7) and ANN (15), and one ensemble model.

Generally speaking, the individual models ((ANN (32, 15, 7) and ANN (15))) generate satisfactory forecasts, with the values of MAPE being 8.50% and 9.18% respectively. Our proposed model has generated more accurate forecasts, with the MAPE value of 7.26% which shows that the forecasting accuracy given by the ensemble model excels the individual models. A significant improvement in generalization performance has been observed in the use of the ensemble method. The performance of each of the ensemble models is highly consistent across the data sets with six and 12 months horizons. The ensemble model employing data set with 12-month horizon is the best-performing model compared to one employing data set with the six-months horizon.

4 Discussions

Mainly there are two key reasons to develop ensemble systems:

(i) to eliminate the risk of an unfortunate prediction of a single forecaster in some specific conditions and

(ii) to improve upon the performance of the single forecaster (Polikar, 2006).

The ensemble may or may not be superior over the performance of the best single forecaster in the ensemble, but it definitely decreases the overall risk of making a principally poor prediction (Polikar, 2006). Not only our proposed ensemble model intelligently and successfully combines the single forecasters of the neural network models, but also it significantly improves the overall accuracy. However, there are many other theoretical and practical reasons to develop the ensemble models. Our main motivation of developing the ensemble model is to make applicable our data with all features (62
Success of our model is due to the following reasons:

(i) if there is no redundancy in the overall feature set; the large amount of the features may have a more intrinsic challenge for the description of the forecasting problem,

(ii) instead of cutting out some features, which may result in throwing out some useful information, primary level forecasters convey the nature of all features though the layers of ensemble model,

(iii) because of the fuzzy constitution of the system, in the case of uncertainty, the proposed ensemble system still has an ability to generalize and deal with imprecise data,

(iv) instead of the conventional firm algorithms to produce the combination rules, they are adaptively extracted from the meta-data set by the fuzzy inference system.

Actually, there is a two directional (mutual) utilization in this ensemble approach. ANNs implement the empirical risk minimization principle by minimizing the training error. Because of the search algorithm employed by multiple perceptron neural networks to converge to global solutions, the resulting search has always a change to get stuck in the local minima. This means the solution of multiple perceptron neural networks is not always unique, optimal, and global (Basak, Pal, & Patranabis, 2007). ANFIS can compensate the decision of the neural networks by combining them. On the other hand, conventional ANFIS itself cannot deal with a data set with 62 attributes, but neural networks can handle with this data set.

Furthermore tourism sector heavily influences with the uncertainty, because of the unpredictable factors such as, economic recessions, natural diseases, campaign and promotions. Soft computing approaches like neural network and neuro-fuzzy models can more successfully deal with the uncertainty than classical statistical and econometric models. The study shows that our proposed stacking ensemble model provides a promising alternative to the ANN for the tourism demand forecasting.

5 Conclusion

This study proposed the method in the framework of a stacking ensembles based on two base learners (neural network models) and a meta-learner (ANFIS) and examined the efficiency of combining forecasts in the tourism context. Furthermore, although the combined forecasts do not always outperform the best single forecasts, none of the ensembles examined in this study are not outperformed by any single model forecasts. Our proposed model considerably outperforms every individual forecaster employed in this study. This result implies that ensemble models are able to improve the performance of the single models, but also it suggests that they considerably reduce the risk of forecasting failure. To improve the accuracy and to obtain the reliability and consistency in the forecasting, our proposed model utilizes the results of the base forecasters according to adaptively generated combination rules against the risk of the forecasting failure made by the single forecasters. The conventional ANFIS models suffer from the curse of dimensionality in high-dimensional input spaces. This approach can easily handle the large amount of the input variables to feed ANFIS, with capability of the multivariate forecasting models. Additionally, research has found that ANFIS is a promising approach to combine and improve the prediction capabilities of the single models due to its adaptive fuzzy rule base nature. Experimental results indicate that the proposed method successfully improved forecasting performance of neural network models in the domain of multivariate tourism demand forecasting. The proposed ensemble model has a significant generalization ability in tourism demand forecasting. In conclusion, this result shows that ANFIS can be also properly used for the high-dimension data sets without discarding any input variables in the ensemble expert systems for yield numerical prediction. Although this study is done in the context of the tourism demand, the findings should be of use to researchers who are interested in the multivariate forecasting using ANFIS.
Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

6 References

Dogru, T., Sirakaya-Turk, E., & Crouch, G. I. (2017). Remodeling international tourism demand: Old theory and new evidence. Tourism Management, 60, 47-55.

Akin, M. (2015). A novel approach to model selection in tourism demand modeling. Tourism Management, 48, 64-72.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. International Journal of Forecasting, 27(3), 870-886.

Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2013). Prediction of Petroleum Reservoir Properties using Different Versions of Adaptive Neuro-Fuzzy Inference System Hybrid Models. International Journal of Computer Information Systems and Industrial Management Applications, 413-426.

Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2017). Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. Journal of Petroleum Science and Engineering, 480-487.

Anifowose, F., & Abdulraheem, A. (2011). Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization. Journal of Natural Gas Science and Engineering, 505-517.

Anifowose, F., Labadin, J., & Abdulraheem, A. (2013). A least-square-driven functional networks type-2 fuzzy logic hybrid model for efficient petroleum reservoir properties prediction. Neural Computing and Applications, 179–190.

Archer, B. (1987). Demand Forecasting and Estimation. Travel, tourism, and hospitality research. A handbook for managers and researchers pp. 77-85.

Aslanargun, A., Mammadov, M., Yazici, B., & Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. Journal of Statistical Computation and Simulation Vol. 77, No. 1, January , 29–53.

Azar, A. T. (2010). Adaptive Neuro-Fuzzy Systems. In Fuzzy Systems (p. 216).

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support Vector Regression. Neural Information Processing – Letters and Reviews, 11(10).

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. Operational Research Quarterly, 20(4), 451–468.

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford University Press.

Boyacioglu, M., & Avci, D. (2010). An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. Expert Systems with Applications, 7908–7912.

Breiman, L. (1996). Bagging Predictors. Machine Learning, 24(2), 123–140.

Cankurt, S. (2016 ). Tourism demand forecasting using ensembles of regression trees. IEEE 8th International Conference on Intelligent Systems (IS), (pp. 702-708).

Cankurt, S., & Subasi, A. (2016). Tourism demand modelling and forecasting using data mining techniques in multivariate time series: A case study in Turkey. Turkish Journal of Electrical Engineering & Computer Sciences, 24, 3388 – 3404.
Chan, C. K., Witt, S. F., Lee, Y., & Song, H. (2010). Tourism forecast combination using the CUSUM technique. Tourism Management, 31(6), 891-897.

Chen, K.-Y. (2011). Combining linear and nonlinear model in forecasting tourism demand. Expert Systems with Applications, 10368–10376.

Chen, M.-S., Ying, L.-C., & Pan, M.-C. (2010). Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system. Expert Systems with Applications, 1185–1191.

Chen, M.-Y. (2011). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. Information Sciences.

Chen, T., & Chen, H. (1993). Approximation to continuous functionals by neural networks with application to dynamical systems. IEEE Trans. Neural Networks, 4(6).

Chen, Z., & Jie, Z. (2011). Neural Network Ensemble for Chinese Inbound Tourism Demand Prediction. Scientia Geographica Sinica, 1208-1212.

Choy, K. L., Lee, W. B., & Lo, V. (2003). Design of an intelligent supplier relationship management system: A hybrid case based neural network approach. Expert Systems with Applications, 24, 225–237.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4), 559–583.

Dietterich, T. (1997). Machine-learning research: Four current direction.

Efendigil, T., Önüt, S., & Kahraman, C. (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. Expert Systems with Applications, 36, 6697–6707.

Fatai Anifowose, J. L. (2015). Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines. Applied Soft Computing, 483-496.

Fernandez-Morales, A., Cisneros-Martínez, J. D., & McCabe, S. (2016). Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. Tourism Management, 56, 172-190.

Fernando, H. P., & Turner, L. W. (2006). Neuro-Fuzzy vs Neural Network Forecasting. CAUTHE Conference.

Fernando, H., Reznik, L., & Turner, L. (1999). Neuro-Fuzzy Forecasting of Tourism to Japan. Australian Tourism and Hospitality Research conference. Adelaide, Australia.

Firmino, P. R., Neto, P. S., & Ferreira, T. A. (2013). Correcting and combining time series forecasters. Neural Networks, 50, 1–11.

Frechtling, D. C. (2001). Forecasting Tourism Demand: Methods and Strategies. Oxford: Butterworth-Heinemann.

Fritz, R., Brandon, C., & Xander, J. (1984). Combining time series and econometric forecast of tourism activity. Annals of Tourism Research, 11, 219–229.

Gheyas, I. A., & Smith, L. (2011). A novel neural network ensemble architecture for time series forecasting. Neurocomputing, 3855–3864.

Ginzburg, I., & Horn, D. (199). Combined neural networks for time series analysis. Adv. Neural Inf. Process, 6, pp. 224–231.

Hadamani, E., Shavandi, H., & Shavandi, A. (2010). Hybridization of Adaptive Neuro-Fuzzy Inference System and Data Preprocessing Techniques for Tourist Arrivals Forecasting. Sixth International Conference on Natural Computation.

Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(10), 993–1001.

Haykin, S. (1999). Neural Networks: a comprehensive foundation. (2nd ed.). Prentice Hall.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multi-layer feedforward networks are universal approximators. Neural Networks, 2, 359-366.

Jang, J.-S. R. (July 1991). Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm. Proc. of the Ninth National Conf. on Artificial Intelligence (AAAI-91), (pp. 762-767).

Jang, J.-S. R. (May 1993). ANFIS: Adaptive-Network-based Fuzzy Inference Systems. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), pp. 665-685.
Jang, J.-S. R., & Sun, C.-T. (1997). Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall.

Jones, M. T. (2008). Artificial Intelligence: A Systems Approach. Infinity Science Press LLC.

Karaboga, D., & Kaya, E. (2020). Estimation of number of foreign visitors with ANFIS by using ABC algorithm. Soft Computing, 7579–7591.

Kosko, B. (1992). Fuzzy systems as universal approximators. Proceedings of the IEEE International Conference on Fuzzy Systems, (pp. 1153-1162). San Diego, CA.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting.

Law, R., & Au, N. (1999). A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong. Tourism Management.

Lemke, C., & Gabrys, B. (2010). Meta-learning for times series forecasting and forecast combination. Neurocomputing, 2006–2016.

Lenard, M. J., Alam, P., & Madey, G. R. (1995). The application of neural networks and a qualitative response model to the auditor’s going concern uncertainty decision. Decision Sciences, 26(2), 209–227.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. Tourism Management, 59, 57-66.

Lin, C.-C., Lin, C.-L., Shyu, J. Z., & Lin, C.-T. (2012). The ANFIS System for Nonlinear Combined Forecasts in the Telecommunications Industry. IJCA Journal, 37(12).

Morabito, F. C., & Versaci, M. (2003). Fuzzy neural identification and forecasting techniques to process experimental urban air pollution data. Neural Networks, 16, 493–506.

Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert systems with applications, 36, 3028-3033.

Pai, P., & Hong, W. (2005). An improved neural network model in forecasting arrivals. Annals of Tourism Research, 32(4), 1138-1141.

Pattie, D., & Snyder, J. (1996). Using a Neural Network to Forecast Visitor Behavior. Annals of Tourism Research.

Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Classification using neural networks. Decision Sciences, 26(6), 749–779.

Piramuthu, S., Shaw, M., & Gentry, J. (1994). A classification approach using multilayered neural networks. Decision Support Systems, 11(5), 509–525.

Polikar, R. (2006). Ensemble Based System in Decisions Making. IEEE Circuits And Systems Magazine, 1531-636.

Qian, B., & Rasheed, K. (2004). Stock market prediction with multiple classifiers. The University of Georgia.

Reid, D. (1969). A Comparative study of time series prediction techniques on economic data, Ph.D. thesis (University of Nottinghamam).

Rossello, J., & Sanso, ́A. (2017). Yearly, monthly and weekly seasonality of tourism demand: A decomposition analysis. Tourism Management, 60, 379-389.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, & J. L. McClelland, In Parallel distributed processing: explorations in the microstructure of cognition Cambridge (pp. 318–362). MA: MIT Press.

Schapire, R. E. (1990). The Strength of Weak Learnability. Machine Learning, 5(2), 197–227.

Shen, S., Li, G., & Song, H. (2011). Combination Forecasts of International Tourism Demand. Annals of Tourism Research, Vol. 38, No. 1, 72–89.

Sigletos, G., Paliouras, G., & Spyropoulos, C. D. (2005). Combining Information Extraction Systems Using Voting and Stacked Generalization. Journal of Machine Learning Research, 6, 1751-1782.

Songa, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism Management, 203–220.
Sumathi, S., & Paneerselvam, S. (2010). Computational intelligence paradigms: theory & applications using MATLAB. Boca Raton: CRC Press Taylor and Francis Group.

Tabari, H., Kisi, O., Ezani, A., & Talaei, P. H. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. Journal of Hydrology, 78–89.

The MathWorks, I. (2012). MATLAB and Fuzzy Logic Toolbox™ Release 2012b. Natick, Massachusetts, United States.

The MathWorks, I. (2013). Fuzzy Logic Toolbox™ User’s Guide. United States.

Wang, L.-X., & Mendel, J. (1992). Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. IEEE Transactions on Neural Networks, 3(5).

Witten, I. H., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.). Morgan Kaufmann.

Wolpert, D. H. (1992). Stacked Generalization. Neural Networks, 5(2), 241–259.

Wong, K. K., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? Tourism Management, 1068–1078.

Zhang, C., & Ma, Y. (2012). Ensemble Machine Learning: Methods and Applications. Springer.

Zhang, C., Jiang, F., Wang, S., & Sun, S. (2021). A new decomposition ensemble approach for tourism demand forecasting: Evidence from major source countries in Asia-Pacific region. International Journal of Tourism Research.

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

Zhang, G., Patuwo, E., & Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting, 14, 35–62. Prentical Hall.