



Tourism demand forecasting using stacking ensemble model with adaptive fuzzy combiner

Selcuk Cankurt¹ · Abdulhamit Subasi^{2,3}

Accepted: 15 December 2021 / Published online: 10 January 2022
© The Author(s) 2022

Abstract

Over the last decades, several soft computing techniques have been applied to tourism demand forecasting. Among these techniques, a neuro-fuzzy model of ANFIS (adaptive neuro-fuzzy inference system) 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 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 a 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 become the largest and the fastest growing industry in the world. As tourism became the biggest social and economic phenomena, a huge number of researches were held related to tourism studies. In 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 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 and 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 et al. 2017; Fernandez-Morales et al. 2016).

The tourism sector is one of the most important sources of income, employment, and foreign exchange earnings in

✉ Abdulhamit Subasi
abdulhamit.subasi@utu.fi; absubasi@effatuniversity.edu.sa

Selcuk Cankurt
selcuk.cankurt@sdu.edu.kz

¹ Faculty of Engineering and Natural Sciences, Suleyman Demirel University, Almaty, Kazakhstan

² Institute of Biomedicine, Faculty of Medicine, University of Turku, 20520 Turku, Finland

³ Department of Computer Science, College of Engineering, Effat University, Jeddah 21478, Saudi Arabia

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 in 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 (SAR-IMA), extreme learning machine (ELM), support vector regression (SVR) with particle swarm optimization (PSO), long short-term memory (LSTM) (Wong et al. 2007; Li et al. 2017; Rossello and Sanso 2017; Akin 2015; Zhang et al. 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 and Au 1999), later the artificial neural networks have increasingly been used to forecast demands for tourism (Songa and Li 2008; Law 2000; Pattie and 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 and Li 2008).

In recent years, newly developed soft computing techniques like fuzzy systems have been applied to various fields to implement intelligent information systems, including forecasting (Morabito and Versaci 2003; Anifowose et al. 2013a, b; Anifowose and 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 et al. 1989; Chen and Chen 1993; Kosko 1992; Wang and Mendel 1992). Fuzzy logic and artificial neural networks are complementary technologies and appear to be promising methods 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 the so-called conventional ANFIS. It is a successful hybrid model to implement a universal approximator (Jang and

Sun 1997). However, it cannot achieve 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 a large number of input variables at all. Recently, in a few studies, ANFIS techniques have been applied to tourism forecasting problems (Fernando and Turner 2006; Chen et al. 2010; Hadavandi et al. 2010; Fernando et al. 1999; Karaboga and 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 and Granger 1969); it is a well-established area in the literature of forecasting. A comprehensive review paper of (Clemen 1989) and many other empirical papers, for example, (Ginzburg and Horn 1994; Zhang 2003; Lemke and Gabrys 2010; Lin et al. 2012; Firmino et al. 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 et al. 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 better overall performance than the independent single forecaster. The most important two findings of ensemble methods are to get a more stable and accurate prediction (Polikar 2006). Several studies (Hansen and Salamon 1990; Gheyas and Smith 2011; Qian and Rasheed 2004; Nanni and 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 et al. 2007; Fritz et al. 1984; Chan et al. 2010; Chen 2011a; Andrawis et al. 2011; Shen et al. 2011). However, there has been a few ensemble approaches (Chen and Jie 2011; Cankurt 2016; Zhang et al. 2021) seen in the circumstance of the tourism demand forecasting, which is a recently emerged area in the machine learning.

The purpose of this study is to introduce the conventional ANFIS technique, assess its major strengths and weaknesses, and propose a stacking ensemble model on how the ANFIS model can be applied in multivariate forecasting in the existence of a data set with a 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 since the much number of the relevant input variables will result in a better 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 a 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 Sect. 3. The discussion about the results is given in Sect. 4. And lastly, the discussion and conclusion of the research are given in Sect. 5.

2 Methodology and data set

2.1 Data set

We have employed a subset version of the tourism data used in (Cankurt and Subasi 2016), which covers the monthly time series of Turkey and its top 24 ranked tourism clients. The complete list of the monthly time series is 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, Iran, 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 in 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 of ANFIS is inability 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 a large amount of the input variables. Like many numerical algorithms, ANFIS also suffers from the curse of dimensionality, which denotes such 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 a 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 that should be considered is to keep the number of the inputs as much as smaller. In 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 et al. 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 nonlinear regression (MNLr) models and concluded that the ANFIS model with four input variables outperforms the others. Chen (2011b) proposed a hybrid model of particle swarm optimization (PSO) and ANFIS to predict business failures. He has selected 5 variables out of the 13 input candidates for his proposed model. The remaining eight variables are discarded by conducting principal component analysis (PCA). Boyacioglu (Boyacioglu and Avci 2010) proposed another ANFIS model to predict the stock price index return using nine input variables. Selection is done among 23 financial and macroeconomic variables, based on the review of domain experts and prior researches. Hadavandi et al. (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 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 and 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 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. The original stacking method proposed by Wolpert (1992) is a two-layer model.

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. The number of the newly generated features 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 a large number of inputs (The MathWorks 2013). In the proposed model, the 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 who can deal with a 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

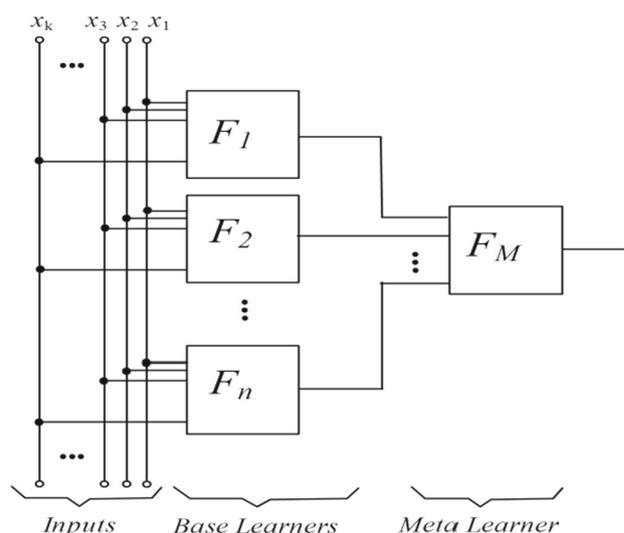


Fig. 1 A schematic view of the proposed ensemble model ($k = 62$, $n = 2$ herein)

section. Therefore, ANFIS becomes possible to work with the data set with a large number of 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 to 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 sets, 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 are 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 modeling; the 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 the meta-learner for our multivariate tourism forecasting task. These models will be used to constitute the 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). The 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:

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

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 Eq. (2) considering the error before calculating for the node (whether hidden or output) (Jones 2008).

$$w_{ij} = w_{ij} + \mu E y_i \quad (2)$$

For the given error (E) and node output (y_i), we multiply by a learning rate (μ) 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 Jang (1993). ANFIS is a hybrid method of neural network and fuzzy system, which is an extension to the 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 optimized. The network applies a combination of the least squares technique and the gradient descent approach of backpropagation for training the parameters of membership functions and fuzzy rules of Sugeno-type fuzzy system (Jang and Sun 1997; Jang 1991, 1993; Sumathi and Paneerselvam 2010).

2.4.3 ANFIS architecture

ANFIS implements a Sugeno-type fuzzy inference system in the form of five-layer neural network architecture. The 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 and Sun 1997; Jang 1991, 1993; Sumathi and Paneerselvam 2010):

2.4.3.1 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) with respect to a node function:

$$O_i^1 = \mu_{A_i}(x) \quad \text{for } i = 1, 2 \tag{3}$$

$$O_i^1 = \mu_{B_{i-2}}(y) \quad \text{for } i = 3, 4 \tag{4}$$

where A_i (or B_{i-2}) 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.

2.4.3.2 Layer 2: product layer In the second layer, every node is fixed node. They are labeled with Π , indicating that output of every node is the product of all the incoming signals:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1, 2 \tag{5}$$

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.

2.4.3.3 Layer 3: normalized layer Every node in this layer is a fixed node. They are labeled with N , indicating normalized firing strengths. The output of the i th node is calculated that is the ratio of the i th rule's firing strength to the sum of all rules' firing strengths:

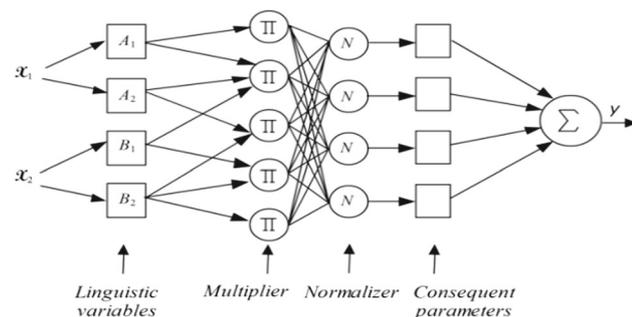


Fig. 2 Neuro-fuzzy network for function approximation

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \tag{6}$$

2.4.3.4 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 \bar{w}_i from layer 3 and a first-order polynomial which is given by:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{7}$$

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

2.4.3.5 Layer 5: total output layer There is only one single node in this layer, which is a fixed node. It is labeled Σ , indicating that the overall output. The overall output as the summation of all incoming signals is given by:

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{8}$$

Nodes of the first and second layers are adaptive, and the others are the type of the fixed nodes (Jang 1991, 1993; Sumathi and 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 , sometimes also denoted 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 and Frank 2005).

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

$$S_A = \frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a})^2, \quad S_P = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})^2, \tag{9}$$

$$S_{PA} = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})(a_i - \bar{a})$$

2.5.1 Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{a_i - p_i}{a_i} \right| \times 100 \tag{10}$$

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

2.5.2 Root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - p_i)^2} \tag{11}$$

Our second accuracy metric will be the 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 and Frank 2005).

2.5.3 Correlation coefficient (R)

$$R = \frac{S_{PA}}{S_P S_A} \tag{12}$$

The final measure which will be employed in our studies is the correlation coefficient (R). It measures the 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 and 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 $[x(t, 1), x(t, 2), \dots, x(t, 62); y(t + 12)]$ and $[x(t, 1), x(t, 2), \dots, x(t, 62); y(t + 6)]$. 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

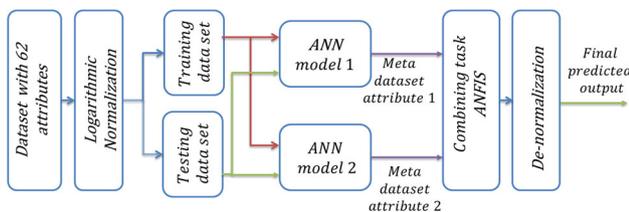


Fig. 3 Experimental setup of the proposed ensemble model

$$y_{t+12} = f_{ANFIS}(f_{ANN1}(\mathbf{X}_t), f_{ANN2}(\mathbf{X}_t)) \text{ and} \tag{13}$$

$$y_{t+6} = f_{ANFIS}(f_{ANN1}(\mathbf{X}_t), f_{ANN2}(\mathbf{X}_t))$$

where d -dimensional input $\mathbf{X} \in R^d$, 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 several feed-forward back-propagation ANN models on the basis of one-layer and three-layer neural networks and the 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 dependent on the diversification of the base learners (Sigletos et al. 2005). To obtain the diversified base learners, we have selected the best ones of the two different categories.

In most empirical studies, the selection of the number of the hidden layer is varied from one to three. One of the used results of Kolmogorov’s theorem for neural networks states that two hidden layers are sufficient for a confident estimate of any complex nonlinear function. Actually, one layer in a network is enough to construct an approximation function (Bishop 1995; Zhang et al. 1998; Aslanargun et al. 2007). To maximize the diversification of the base learners, we have preferred one and three for 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 neurons in the input layer), (2) 75% of input neurons, or (3) 50% of input and output neurons (Lenard et al. 1995; Patuwo et al. 1993; Piramuthu et al. 1994; Efendigil et al. 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 et al. 2003). While we have preferred to use the sigmoid transfer function for the nodes of the hidden layers, the 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 changes in weight and decreases the possibility of any weight oscillation during the training cycle. A learning rate between 0.05 and 0.5 achieves good results in most practical cases (Rumelhart et al. 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 et al. 1986; Palmer et al. 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 and 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 the 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 = 264,956 values, (2) ANN (32, 15, 7) model with $R^2 = 0.957$, MAPE = 8.50% and RMSE = 205,097 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 better forecasting performance, further investigation has been 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-shaped 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. A larger number of the membership functions take longer to train and to generate predictions. Also increasing the number of the membership functions, increases the number of the fitting parameters. That is why, you can use a large number of membership functions, only if you have a large amount of the training data set. But usually, in real-life applications, 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 overtraining (overfitting) 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-month-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).

Table 1 Single Models

Horizon	Type	Model	MAPE (%)	RMSE	R^2
12 months ahead	Single	ANN (32, 15, 7)	8.50	205,097	0.957
		ANN (15)	9.18	264,956	0.928
6 months ahead	Single	ANN (32, 15, 7)	11.24	234,387	0.940
		ANN (15)	14.67	419,409	0.840

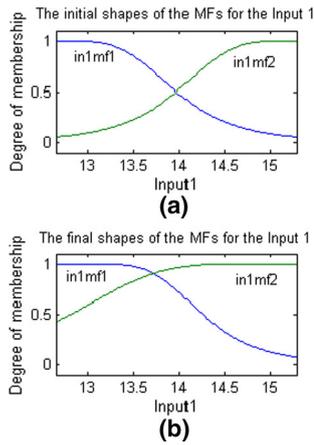


Fig. 4 The membership functions

Figure 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 the checking data set at epoch 115.

Figure 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 several millions.

The performances of the individual models, which are used as the base learner in the proposed ensemble model are reported in Table 1. The results for the proposed ensemble models, which are the combinations of the ANN (32, 15, 7), ANN (15) and ANFIS are reported in Table 2. Results in Tables 1 and 2 show that our proposed ensemble models have demonstrated significantly better forecasting

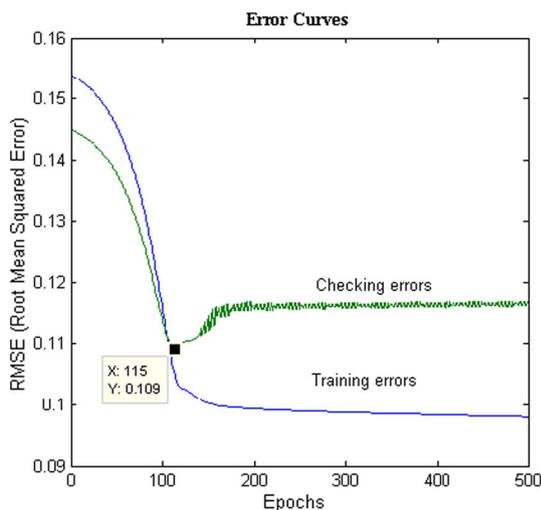


Fig. 5 Training and checking errors

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 their nature that can deal with the imprecise combination rules rather than fixed and exact rules defined by procedures of conventional algorithms.

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 6 and 12 months horizons. The ensemble model employing data set with the 12-month horizon is the best-performing model compared to one employing data set with the six-month 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 for developing the ensemble model is to make applicable our data with all features (62 features for this empirical study) in the case of using conventional ANFIS for the multivariate tourism demand forecasting. While the dimension of our data set is very large, in most

Fig. 6 Comparison of the actual values with the results obtained from ANFIS

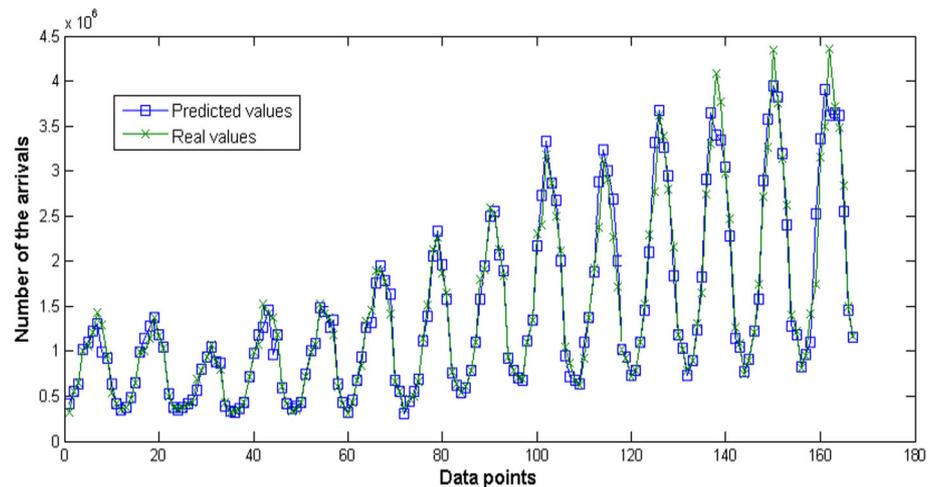


Table 2 The ensemble models

Horizon	Type	Model	MAPE (%)	RMSE	R^2
12 months ahead	Ensemble	$f_{\text{ANFIS}} (f_{\text{ANN}(32,15,7)}, f_{\text{ANN}(15)})$	7.26	192,259	0.969
6 months ahead	Ensemble	$f_{\text{ANFIS}} (f_{\text{ANN}(32,15,7)}, f_{\text{ANN}(15)})$	9.30	218,624	0.952

other practical studies that employ grid partitioned ANFIS, this figure is six or less.

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 through 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 et al. 2007). ANFIS can compensate for 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 the uncertainty, because of the unpredictable factors, such as economic recessions, natural diseases, campaigns 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 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 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 a large amount of the input variables to feed ANFIS, with the 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 the 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 multivariate forecasting using ANFIS.

Funding Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital.

Declarations

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akın M (2015) A novel approach to model selection in tourism demand modeling. *Tour Manage* 48:64–72
- Andrews RR, Atiya AF, El-Shishiny H (2011) Combination of long term and short term forecasts, with application to tourism demand forecasting. *Int J Forecast* 27(3):870–886
- Anifowose F, Abdulraheem A (2011) Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization. *J Nat Gas Sci Eng* 3(3):505–517
- Anifowose FA, Labadin J, Abdulraheem A (2013a) Prediction of petroleum reservoir properties using different versions of adaptive neuro-fuzzy inference system hybrid models. *Int J Comput Inf Syst Ind Manag Appl* 5:413–426
- Anifowose F, Labadin J, Abdulraheem A (2013b) A least-square-driven functional networks type-2 fuzzy logic hybrid model for efficient petroleum reservoir properties prediction. *Neural Comput Appl* 23(1):179–190
- Anifowose FA, Labadin J, Abdulraheem A (2017) Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. *J Pet Sci Eng* 151:480–487
- Archer B (1987) Demand forecasting and estimation. In: Ritchie J, Goeldner C (eds) *Travel, tourism, and hospitality research. A handbook for managers and researchers*. John Wiley & Sons, Inc., New York, 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. *J Stat Comput Simul* 77(1):29–53
- Azar AT (2010) *Adaptive neuro-fuzzy systems*. IntechOpen
- Basak D, Pal S, Patranabis DC (2007) Support vector regression. *Neural Inf Process—Lett Rev* 11(10):203–224
- Bates JM, Granger CW (1969) The combination of forecasts. *Oper Res Q* 20(4):451–468
- Bishop CM (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 Syst Appl* 37(12):7908–7912
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Cankurt S, Subasi A (2016) Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey. *Turk J Electr Eng Comput Sci* 24:3388–3404
- Cankurt S (2016) Tourism demand forecasting using ensembles of regression trees. In: *IEEE 8th international conference on intelligent systems (IS)*, pp 702–708
- Chan CK, Witt SF, Lee Y, Song H (2010) Tourism forecast combination using the CUSUM technique. *Tour Manage* 31(6):891–897
- Chen K-Y (2011a) Combining linear and nonlinear model in forecasting tourism demand. *Expert Syst Appl* 38:10368–10376
- Chen M-Y (2011b) A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Inf Sci* 220:180–195
- Chen Z, Jie Z (2011) Neural network ensemble for chinese inbound tourism demand prediction. *Sci Geograph Sin* 31:1208–1212
- Chen M-S, Ying L-C, Pan M-C (2010) Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system. *Expert Syst Appl* 37:1185–1191
- Chen T, Chen H (1993) Approximation to continuous functionals by neural networks with application to dynamical systems. *IEEE Trans Neural Netw* 4(6):910–918
- Choy KL, Lee WB, Lo V (2003) Design of an intelligent supplier relationship management system: a hybrid case based neural network approach. *Expert Syst Appl* 24:225–237
- Clemen RT (1989) Combining forecasts: a review and annotated bibliography. *Int J Forecast* 5(4):559–583
- Dietterich T (1997) Machine-learning research—four current directions. *AI Magazin* 18:97–136
- Dogru T, Sirakaya-Turk E, Crouch GI (2017) Remodeling international tourism demand: old theory and new evidence. *Tour Manage* 60:47–55

- Efendigil T, Öntüt S, Kahraman C (2009) A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis. *Expert Syst Appl* 36:6697–6707
- Fatai Anifowose JL (2015) Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines. *Appl Soft Comput* 26:483–496
- Fernandez-Morales A, Cisneros-Martínez JD, McCabe S (2016) Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. *Tour Manage* 56:172–190
- Fernando HP, Turner LW (2006) Neuro-fuzzy vs neural network forecasting. In: Whitelaw PA, Barry OG (eds) *CAUTHE 2006: To the City and Beyond*. Footscray, Victoria University. School of Hospitality, Tourism and Marketing, pp 1461–1469
- Fernando H, Reznik L, Turner L (1999) Neuro-fuzzy forecasting of tourism to Japan. In: *Australian tourism and hospitality research conference*, Adelaide, Australia
- Firmino PR, Neto PS, Ferreira TA (2013) Correcting and combining time series forecasters. *Neural Netw* 50:1–11
- Frechling DC (2001) *Forecasting tourism demand: methods and strategies*. Butterworth-Heinemann, Oxford
- Fritz R, Brandon C, Xander J (1984) Combining time series and econometric forecast of tourism activity. *Ann Tour Res* 11:219–229
- Gheyas IA, Smith L (2011) A novel neural network ensemble architecture for time series forecasting. *Neurocomputing* 74:3855–3864
- Ginzburg I, Horn D (1994) Combined neural networks for time series analysis. *Adv Neural Inf Process* 6:224–231
- Hadavandi E, Shavandi H, Ghanbari A (2010) Notice of retraction: hybridization of adaptive neuro-fuzzy inference system and data preprocessing techniques for tourist arrivals forecasting. In: *6th international conference on natural computation*, pp 1692–1695
- Hansen LK, Salamon P (1990) Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 12(10):993–1001
- Haykin S (1999) *Neural networks: a comprehensive foundation*, 2nd edn. Prentice Hall
- Hornik K, Stinchcombe M, White H (1989) Multi-layer feedforward networks are universal approximators. *Neural Netw* 2:359–366
- Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference systems. *IEEE Trans Syst Man Cybern* 23(3):665–685
- Jang J-SR, Sun C-T (1997) *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentice Hall
- Jang J-SR (1991) Fuzzy modeling using generalized neural networks and Kalman Filter algorithm. In: *Proc. of the 9th national conf. on artificial intelligence (AAAI-91)*, pp 762–767
- Jones MT (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 Comput* 24(10):7579–7591
- Kosko B (1992) Fuzzy systems as universal approximators. In: *Proceedings of the IEEE international conference on fuzzy systems*, San Diego, CA, pp 1153–1162
- Law R, Au N (1999) A neural network model to forecast Japanese demand for travel to Hong Kong. *Tour Manag* 20:89–97
- Law R (2000) Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tour Manag* 21:331–340
- Lemke C, Gabrys B (2010) Meta-learning for time series forecasting and forecast combination. *Neurocomputing* 73(10–12): 2006–2016
- Lenard MJ, Alam P, Madey GR (1995) The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decis Sci* 26(2):209–227
- Li X, Pan B, Law R, Huang X (2017) Forecasting tourism demand with composite search index. *Tour Manage* 59:57–66
- Lin C-C, Lin C-L, Shyu JZ, Lin C-T (2012) The ANFIS system for nonlinear combined forecasts in the telecommunications industry. *IJCA J* 37(12):30–35
- Morabito FC, Versaci M (2003) Fuzzy neural identification and forecasting techniques to process experimental urban air pollution data. *Neural Netw* 16:493–506
- Nanni L, Lumini A (2009) An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Syst Appl* 36:3028–3033
- Pai P, Hong W (2005) An improved neural network model in forecasting arrivals. *Ann Tour Res* 32(4):1138–1141
- Palmer A, Montano JJ, Sese A (2006) Designing an artificial neural network for forecasting tourism time series. *Tour Manage* 27:781–790
- Pattie D, Snyder J (1996) Using a neural network to forecast visitor behavior. *Ann Tour Res* 23(1):151–164
- Patuwo E, Hu MY, Hung MS (1993) Classification using neural networks. *Decis Sci* 26(6):749–779
- Piramuthu S, Shaw M, Gentry J (1994) A classification approach using multilayered neural networks. *Decis Support Syst* 11(5):509–525
- Polikar R (2006) Ensemble based system in decisions making. *IEEE Circuits Syst Mag* 6(3):21–45
- 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*. PhD thesis, University of Nottingham
- Rossello J, Sanso A (2017) Yearly, monthly and weekly seasonality of tourism demand: a decomposition analysis. *Tour Manage* 60:379–389
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) *In Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press, Cambridge, pp 318–362
- Schapire RE (1990) The strength of weak learnability. *Mach Learn* 5(2):197–227
- Shen S, Li G, Song H (2011) Combination forecasts of international tourism demand. *Ann Tour Res* 38(1):72–89
- Sigletos G, Paliouras G, Spyropoulos CD (2005) Combining information extraction systems using voting and stacked generalization. *J Mach Learn Res* 6:1751–1782
- Song H, Li G (2008) Tourism demand modelling and forecasting—A review of recent research. *Tour Manag* 29(2):203–220
- Sumathi S, Paneerselvam S (2010) *Computational intelligence paradigms: theory & applications using MATLAB*. CRC Press Taylor and Francis Group, Boca Raton
- Tabari H, Kisi O, Ezani A, Talaee PH (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J Hydrol* 444:78–89
- The MathWorks I (2012) *Fuzzy Logic Toolbox™ User's Guide*. Natick, MA, USA
- The MathWorks I (2013) *Fuzzy logic toolbox™ User's Guide*. Natick, MA, USA
- Wang L-X, Mendel J (1992) Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Trans Neural Netw* 3(5):807–814
- Witten IH, Frank E (2005) *Data mining: practical machine learning tools and techniques*, 2nd edn. Morgan Kaufmann

- Wolpert DH (1992) Stacked generalization. *Neural Netw* 5(2):241–259
- Wong KK, Song H, Witt SF, Wu DC (2007) Tourism forecasting: To combine or not to combine? *Tour Manag* 28:1068–1078
- Zhang GP (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. *Int J Forecast* 14:35–62 (**Prentical Hall**)
- 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. *Int J Tour Res* 23(5):832–845

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.